

Second Edition

Engineering Physics



Engineering Physics

Second Edition

Uma Mukherji



**Alpha Science International Ltd.
Oxford, U.K.**

Uma Mukherji
Head, Department of Physics
Smt. Indira Gandhi College of Engineering
Koparkhairane, Navi Mumbai, India

Copyright © 2003, 2007
Second Edition, 2007

Alpha Science International Ltd.
7200 The Quorum, Oxford Business Park North
Garsington Road, Oxford OX4 2JZ, U.K.

www.alphasci.com

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of the publishers.

ISBN-13: 978-1-84265-285-5
ISBN-10: 1-84265-285-0

Printed in India

To my son

Dipu

if you find, it is useful
then my job will be rewarded

MAA

Preface to the Second Edition

This book is intended as a textbook for the students of engineering physics and science. As an experienced teacher of physics/engineering physics, I have tried to select and arrange the material in such a way that the students would not face any difficulty in following the book.

This edition, the second, has been thoroughly revised. Errors and misprints have been corrected and some diagrams have been modified and improved for proper representation of the theory. To fulfil the requirements of students (on the basis of All India Survey), some more topics have been included in the second edition. New topics including conductivity of metal and semiconductor, Hall effect, magnetic materials, polarisation, zone plate, electro-optic effect, photoelasticity, etc. have been added.

I strongly hope that the students from different universities will find this edition of the book much more useful.

I am grateful to all, who suggested newer topics to improve the standard of the book. I am also grateful to Mr N.K. Mehra, M/s Narosa Publishing House who conveyed all those survey reports to me. I am grateful to MOTHER whose silent support gave me strength to complete the revision work. I am also thankful to my husband Prof. Shekhar Mukherji for checking overall presentation and for giving me a new personal computer, so that I could do my work smoothly and nicely.

Suggestions for improvement and information about errors will be gratefully accepted.

Uma Mukherji

Preface to the First Edition

The courses prescribed for engineering physics by various universities are at a very advanced level, and considerable importance is given to every topic from the viewpoint of its application to practical fields. But few books provide all, that is expected of them. The present book attempts to fulfil course objectives. Some of these topics are still at research level, like holography (production of 3-dimensional photograph), fiber-optics communication, superconductivity, and so forth. With reference to such advanced topics, the emphasis in the courses is given to numerical problems. In the present system of examinations, about fifty per cent emphasis is on problem-solving questions. The relevance to such numerical problems and objective-type questions has been taken from daily life experiences.

This book, intended as a textbook for engineering physics and science students, tries to select and arrange the material in such a way that a student would not face any difficulty in understanding the topics.

This book is an outcome of many years of my teaching experience. The text is based on my lectures given to the undergraduate engineering students of Bombay University from 1990 onwards. The subject matter is arranged and discussed so as to make the level of treatment progressively more difficult and the material more logically acceptable. Proper feedback from the students helped me to clarify different topics.

The book is divided into two parts. The first part discusses basic theories like crystallography and crystal-imperfection, thermoelectricity, thermionic-emission, ultrasonic, acoustics and semiconductor.

The second part covers advance topics like thin film interference and diffraction, x-rays, motion of the charged particle in electric and magnetic field, quantum physics and Schrödinger wave equation, laser, holography, fiber-optics, radioactivity and superconductivity.

The purpose of this book is to give the students a clear understanding of topics which are related to their course and day to day life in different ways.

Motivation was there, and the stimulus came from my husband, Prof. Shekhar Mukherji. My students also gave me moral support and proper feedback regarding the standard of this manuscript—as they have used my class notes during last few years.

I am thankful to Mrs Sailaja who did all the typing work, and to Mr Suhas Narkhede for all the cartographic help and to make the matter presentable. I am also thankful to Mr N.K. Mehra, M/s Narosa Publishing House, for bringing out the book in its present form.

Contents

<i>Preface to the Second Edition</i>	vii
<i>Preface to the First Edition</i>	ix

PART 1

1. Crystallography and Crystal Imperfection	3
1.1 Crystal Structure	3
1.2 Definition of Crystal Structure, Lattice Points, Space Lattice, Lattice Plane, Basis, Unit Cell	4
1.3 Lattice Parameter	6
1.4 Lattice Vector and Direction	6
1.5 Seven Crystal System	6
1.6 Bravais Space Lattice	7
1.7 Symmetry Elements of a Crystalline Solid	8
1.8 Coordinates of Lattice Point	9
1.9 Number of Atoms/Molecules per Unit Cell (n)	10
1.10 Coordination Number	11
1.11 Atomic Radius (r)	12
1.12 Packing Density or Atomic Packing Factor	13
1.13 Calculation of Lattice Constant or Crystal Density	14
1.14 Miller Indices	15
1.15 Different Crystal Planes with Miller Indices	16
1.16 Direction of a Line (Lattice Vector)	18
1.17 Interplaner Distance	20
1.18 Interplaner Spacing for Cubic System	21
1.19 Important Features of Miller Indices	24
1.20 Effects on Crystal Due to Presence of Imperfection	25
1.21 Causes of Imperfections in Crystals	25
1.22 Different Types of Crystal Defects	25
1.22.1 Point Defect	25
1.22.2 Dislocation – Line Defect	28
1.22.3 Surface Defects	29
1.22.4 Volume Defects	29
1.23 Diamond Structure	29
1.24 Barium Titanate (BaTiO_3)	31
1.25 Interstitial Voids	32
1.26 Ligancy and Ionic Crystal Structure	34
1.26.1 CsCl-Structure (Ligancy-8) with Cubical configuration	35
1.26.2 NaCl-Structure (Ligancy-6) with Octahedral Configuration	36
1.26.3 ZnS Structure (Ligancy-4) with Tetrahedral Configuration	37

1.26.4 Ligancy-3 with Triangular Configuration	38
<i>Problems</i>	40
<i>Questions</i>	41
2. Thermoelectricity	44
2.1 Seebeck Effect or Thermoelectric Effect	44
2.2 Origin of Seebeck Effect	44
2.3 Thermoelectric Series	45
2.4 Variation of Thermoelectric e.m.f. with Temperature	45
2.5 Thermoelectric Power	47
2.6 Determination of Neutral Temperature (T_n) and Temperature of Inversion (T_i) from Seebeck Coefficient	47
2.7 Seebeck Coefficients	48
2.8 Law of Intermediate Metals	48
2.9 Law of Successive Temperature	49
2.10 Thermoelectric Thermometer	49
2.11 Thermistor	51
<i>Problems</i>	54
<i>Questions</i>	55
3. Thermionic Emission	57
3.1 Introduction	57
3.2 Theoretical Explanation for Thermionic Emission	57
3.3 Richardson Dushman Equation	58
3.4 Thermionic Valve	60
3.5 Diode Characteristics	60
3.5.1 Anode Characteristics or V-I Characteristics of Diode	61
3.5.2 Thermionic Characteristics	63
3.6 Child-Langmuir Law for Space Charge Limited Current	64
3.7 Diode as Rectifier	65
3.7.1 Half wave rectifier	65
3.7.2 Full wave rectifier	66
<i>Problems</i>	67
<i>Questions</i>	68
4. Ultrasonic	69
4.1 Introduction	69
4.2 Production of Ultrasonic Waves	69
4.3 Amplifier	70
4.4 Oscillator	71
4.5 Tuned Circuit	71
4.6 Positive Feedback Amplifier as an Oscillator	72
4.7 Magnetostriction Oscillator	74
4.8 Piezoelectric Effect	76
4.9 Piezoelectric Oscillator	78
4.10 Detection of Ultrasonics	79
4.11 Properties of Ultrasonics	80
4.12 Applications of Ultrasonics	80
4.13 PZT Ceramics	83
4.14 Ultrasonic Transducer	83
<i>Problems</i>	84
<i>Questions</i>	84

5. Acoustics	87
5.1 Basic Requirements for an Acoustically Good Hall	87
5.2 Reverberation and Time of Reverberation	87
5.3 Derivation of Optimum Reverberation Time (R-Time)	88
5.4 Sabine's Formula for Optimum Reverberation Time T (R-Time)	90
5.5 Absorption Coefficient	91
5.6 Measurement of Absorption Coefficient	92
5.7 Factors Affecting Architectural Acoustics and their Remedies	94
<i>Problems</i>	97
<i>Questions</i>	99
6. Semiconductors	100
6.1 Band Theory and Formation of Energy Gap in Different Materials	100
6.2 Metals, Insulators and Semiconductors	102
6.3 Intrinsic Semiconductor	104
6.4 Electrons and Positive Holes in a Semiconductor	104
6.5 Fermi Level and Effect of Temperature on Intrinsic Semiconductor	105
6.6 Fermi Level in an Intrinsic Semiconductor	107
6.7 Extrinsic Semiconductor	109
6.8 Fermi Level in an Extrinsic Semiconductor	112
6.9 Variation of Fermi Level with Temperature in Extrinsic Semiconductor	113
6.10 Electrical Conductivity in Metallic Conductor	113
6.11 Relaxation Time	115
6.12 Conductivity in Intrinsic-Semiconductor	116
6.13 Conductivity in Extrinsic semiconductor	118
6.14 Hall Effect	119
<i>Problems Related Conductivity, Current Density, Hall Effect</i>	121
6.15 $p-n$ Junction without Applied Field	125
6.16 Semiconductor Junction with Applied Voltage	127
6.16.1 Forward Bias	127
6.16.2 Reverse Bias	127
6.17 V-I Characteristics of $p-n$ Junction/Semiconductor-Diode	130
6.18 Semiconductor Diode as Rectifier	132
6.18.1 Half Wave Rectifier	132
6.18.2 Full Wave Rectifier	132
6.19 Flow of Current Across P-N Junction: The Rectifier Equation	134
<i>Problem Related Rectifier Equation</i>	136
6.20 Transistor	136
6.21 Working Principle of Transistor (Under C.B. Configuration)	137
6.22 Transistor Characteristics	141
6.22.1 Common Base (CB) Configuration	142
6.22.2 Common Emitter (CE) Configuration	145
6.23 Current Relation in C.E. Mode	147
6.24 Relation Between α_{bc} and β_{bc}	147
<i>Problems</i>	148
<i>Questions</i>	149
7. Magnetic Materials	151
7.1 Atomic Model and Magnetization	151
7.2 Molecular Field Inside a Magnetic Material	152

7.3	Various Magnetic Materials	153
7.3.1	Diamagnetic Material	153
7.3.2	Paramagnetic Material	154
7.3.3	Ferromagnetic Material	155
7.3.4	Anti-Ferromagnetic Material	156
7.3.5	Ferrimagnetic Material	156
7.4	Ferromagnetic Domain and Magnetization Curve	157
	<i>Questions</i>	158

PART 2

1. Interference, Diffraction and Polarisation

1A. Thin Film Interference of Light 161

1.1	Wave Nature of Light	161
1.2	Interference	162
1.3	Analytical Discussion and Conditions for Constructive and Destructive Interference	162
1.4	Thin Film Interference Due to Reflected Light (Parallel Thin Film)	164
1.5	Thin Film Interference Due to Transmitted Light	166
1.6	Intensity Distribution	167
1.7	Origin of Colours in Thin Film	168
1.8	Necessity of a Broad Source	168
1.9	Fringes Produced by Wedge-shaped Film	169
1.10	Fringe Width	171
1.11	Nature of Interference Pattern	172
1.12	Thickness of the Wedge	172
1.13	Testing the Flatness of Surfaces	173
1.14	Newton's Ring	174
1.15	Determination of the Radius of Curvature of the Plano-Convex Lens, Wave Length of Light Used, and Refractive Index of Some Unknown Liquid	178
1.16	Non-Reflecting Films	179
1.17	Totally Reflecting Film (Sun Control Film)	179
	<i>Problems</i>	180
	<i>Questions</i>	186

1B. Diffraction of Light

187

1.18	Optical Diffraction	187
1.19	Fresnel Half Period Zone	188
1.20	Zone Plate	191
1.21	Comparison Between Zone Plate and Convex Lens	193
	<i>Problems on the Zone Plate</i>	193
1.22	Fresnel and Fraunhofer Diffraction	194
1.23	Fraunhofer Diffraction at Single Slit	194
1.24	Fraunhofer for Diffraction at Double Slit	196
1.25	Fraunhofer Theory of Diffraction Grating	198
1.26	Resultant Intensity	201
1.26.1	For Single Slit	201
1.26.2	For Double Slit	202
1.26.3	For Diffraction Grating	203
1.27	Absent Spectra	205

1.28 Maximum Number of Order of Spectra in a Grating	206
1.29 Dispersive Power of the Grating	206
1.30 Spectrum Formation and Overlapping of Spectrum Lines	207
1.31 Determination of Unknown Wave Length by Diffraction Grating	208
<i>Problems</i>	209
<i>Questions</i>	211
1C. Polarisation of Light	211
1.32 Polarisation of Wave	211
1.33 Polarisation of Light Wave	212
1.34 Production of Polarised Light Wave by Different Polarising Devices	213
1.35 Analyser for Polarised Wave	216
1.36 Superposition of Two Plane Polarised Waves	217
1.37 Quarter Wave Plate and Half Wave Plate	219
1.38 Optical Activity	220
1.39 Electro-optic Effect	221
1.40 Photo-Elasticity	221
<i>Questions</i>	222
2. X-Rays	223
2.1 Discovery of X-Rays	223
2.2 Production of X-Rays	223
2.3 Modern Coolidge Tube	224
2.4 Detection of X-Rays	225
2.5 Origin of X-Rays	225
2.5.1 Line Spectrum or Characteristic Spectrum	225
2.5.2 Continuous Spectrum	227
2.6 Moseley's Law	229
2.7 Comparison Between Bohr's Theory of Atomic Spectra and Moseley's Law	231
2.8 Properties of X-Rays	231
2.9 X-Ray Diffraction and Bragg's Law	231
2.10 Explanation of X-Ray Diffraction	233
2.11 Bragg's Spectrometer and Different Methods for X-Ray Diffraction	234
2.11.1 Laue Method	235
2.11.2 Rotating Crystal Method	236
2.11.3 Powder Method	236
2.12 Uses of X-Ray Diffraction Patterns	237
<i>Problems</i>	237
<i>Questions</i>	240
3. Motion of the Charged Particle in Electric and Magnetic Field	244
3.1 Introduction	244
3.2 Motion of Charged Particle in a Parallel Electric Field	244
3.3 Motion of Electron in Transverse Electric Field	245
3.4 Motion of Electron in Transverse Magnetic Field	247
3.5 Motion of the Charged Particle in Crossed Electric and Magnetic Field and its Application (Cathode Ray Tube)	248
3.6 Experimental Set-up for Thomson Method for Measuring <i>e/m</i> Ratio for Electron	250

3.7	Vertical Deflection due to Electric and Magnetic Field on a Screen and Sensitivity	251
3.7.1	Vertical deflection on the screen at a distance D from the centre of the perpendicular electric field E ,	252
3.7.2	Vertical deflection due to perpendicular magnetic field of magnetic induction B	252
3.8	Electrostatic Focusing	253
3.9	Magnetostatic Focusing and Magnetic Lens	255
3.10	Cathode Ray Oscilloscope (C.R.O.)	257
3.10.1	Construction of C.R.O.	257
3.10.2	Working of C.R.O.	259
3.10.3	Uses of C.R.O.	261
3.10.3.1	Measurement of A.C./D.C. Voltage	261
3.10.3.2	Measurement of Frequency	261
3.10.3.3	Measurement of Phase-difference	262
3.11	Cyclotron	263
3.12	Aston-Mass Spectrograph	266
3.13	Bain Bridge's Mass Spectrograph	269
	<i>Problems</i>	270
	<i>Questions</i>	277
4.	Quantum Physics and Schrödinger Wave Equation	278
4.1	Introduction	278
4.2	Quantum Theory of Radiation	278
4.3	Photo-Electric Effect	279
4.4	Einstein's Photo-Electric Equation	282
4.5	Compton Effect	283
4.6	Wave Nature of Particle and Wave-Particle Duality	285
4.7	de-Broglie's Hypothesis	289
4.8	Davison and Germer Experiment	290
4.9	Properties of Matter Waves and how they Differ from Sound and Light Waves	292
4.10	Heisenberg's Uncertainty Principle	293
4.11	Verification of Heisenberg Principle	294
4.11.1	γ -Ray Microscope Experiment	294
4.11.2	Single Slit Diffraction Method	294
4.12	Electron Microscope	296
	<i>Problems</i>	298
	<i>Questions</i>	302
4.13	The Wave Function	304
4.14	Physical Significance of Wave Function	304
4.15	Development of Wave Equation	306
4.16	Phase Velocity and Group Velocity	306
4.17	Schrödinger Wave Equation (Time Dependent)	308
4.18	Schrödinger (Time Independent) Wave Equation (STIE)	311
4.19	Eigen Value Equation	312
4.20	Stationary State	313
4.21	Applications of Schrödinger Time (STIE) Independent Equation	313
4.21.1	Particle in a One Dimensional Box: One dimensional infinite rectangular potential well	313

4.21.2	Barrier Tunneling: Rectangular finite potential barrier of height V_0	317
4.21.2.1	α -decay	320
4.21.2.2	Tunnel diode	321
<i>Questions</i>		322
5.	Laser, Holography and Fiber Optics	323
5.1	Introduction—Laser	323
5.1.1	Absorption of Radiation	323
5.1.2	Emission of Radiation	323
5.1.2.1	Spontaneous Emission of Radiation	323
5.1.2.2	Stimulated Emission of Radiation and Lasing Action	324
5.2	Population Inversion and Pumping	325
5.3	Operating Principle in Laser	325
5.4	Characteristic Properties of Laser Light	327
5.5	Ruby Laser	327
5.6	Helium-Neon Laser	329
5.7	Molecular Energy Levels	330
5.8	Carbon Dioxide Laser	331
5.9	Dye Laser	332
5.10	Semiconductor Laser or Diode Laser and Light Emitting Diode (LED)	334
5.11	Introduction—Holography	339
5.11.1	Recording of Hologram	340
5.11.2	Reconstruction of the Image	341
5.11.3	Uses of Holography	341
5.12	Introduction—Fiber Optics	342
5.13	Total Internal Reflection	342
5.14	Propagation of Light in Fiber	344
5.15	Different Types of Optical Fiber	346
5.15.1	Cladless Fiber	346
5.15.2	Graded Index Fiber (Grin Fiber)	346
5.15.3	Step Index Fiber	347
5.15.4	Passive and Active Fiber	347
5.16	Types of Rays (Meridional and Skew Rays)	347
5.17	Modes of Propagation in Fiber	348
5.18	Multimode and Monomode Fibers	350
5.19	Dispersion—Inter Modal Dispersion	350
5.20	Losses in Fiber Cable	352
5.21	Photo Detector	352
5.21.1	Mechanism of Photo-Diode	352
5.21.2	Quantum Efficiency and Responsivity	354
5.21.3	Long Wave Length Cut-Off	354
5.22	Components of an Optoelectronic Communication System (Block-diagram)	354
5.23	Uses of Optical Fiber	356
5.24	Advantages of Optical Fiber	357
<i>Problems</i>		357
<i>Questions</i>		358

6. Radioactivity and Nuclear Reactions	361
6.1 Introduction	361
6.2 Electrical Instruments as Detectors/Geiger-Muller Counter	361
6.3 Radioactive Disintegration or Decay	365
6.4 Natural Radio-Isotopes	366
6.5 Law of Radioactive Disintegration	367
6.6 Half Life	367
6.7 Nuclear Force	368
6.8 Packing Fraction	368
6.9 Binding Energy of the Nucleus	369
6.10 Q-Value	371
6.11 Artificial Transmutation or Artificial Disintegration	372
6.12 Rutherford's Experiment	372
6.13 Artificial Radioactivity and Artificial Radio-Isotopes	372
6.14 Radio-Isotopes of Transuranium Elements	374
6.15 Uses of Radio-Isotopes	374
6.15.1 In Biological Field	374
6.15.2 In Medical World	375
6.15.3 In Industrial Field	375
6.15.4 Carbon Dating in Geological Field	375
6.16 Nuclear Reactions	376
6.17 Theory of Nuclear Fission	378
6.18 Nuclear Fission as a Source of Energy	381
6.19 The Controlled Chain Reaction System	382
6.20 Basic Idea for Nuclear Reactor	383
6.21 Critical Size for Nuclear Reactor	384
6.22 Essential Parts of the Nuclear Reactor	384
6.23 Thermal Power Reactor	386
6.24 Breeder Reactor	387
6.25 Thermo-Nuclear Reactions or Nuclear Fusion	388
6.26 Stellar Thermonuclear Reaction	389
6.26.1 Proton-Proton Chain	389
6.26.2 Carbon-Nitrogen Chain	390
6.27 Comparisons Between Fission and Fusion	390
<i>Problems</i>	391
<i>Questions</i>	394
7. Superconductivity	396
7.1 Introduction	396
7.2 Meissner Effect	397
7.3 Theoretical Explanation of Superconductivity: B.C.S. Theory	399
7.4 Pairing Energy Gap in Superconductor	400
7.5 Effect of External Magnetic Field on Superconducting State	402
7.6 London Equation	404
7.7 Properties of Superconductor	407
<i>Questions</i>	407
References	408
Subject Index	409

PART-1

Crystallography and Crystal Imperfection

CRYSTALLOGRAPHY¹

In nature, elements and their chemical compounds are usually found in three states, namely, solid, liquid and gaseous. Different solids have different properties. If the atoms or molecules in a solid are arranged in some regular fashion, then it is known as *crystalline solid*. The majority of solids are crystalline, with the atoms, ions or molecules, of which they are composed of, arranged into regular repeated three-dimensional patterns. *The presence of long-range order is thus, the defining property of a crystal.* Due to lack of this long-range order in the arrangement of their constituent particles, other solids may properly be regarded as *super-cooled liquid* or *amorphous solid*, whose stiffness is due to an exceptionally high viscosity. Glass, pitch, and many plastics are examples of such amorphous (without form) solids. Amorphous solids may exhibit short-range order in their structures. The analogy between an amorphous solid and a liquid is worth pursuing, generally they do not have long range order, as X-ray diffraction indicates that many liquid molecules are continually shifting their positions. Amorphous solids have no sharp melting point whereas for crystalline solid, melting occurs at a precisely defined temperatures.

1.1 Crystal Structure

In a crystalline solid, *each atom or molecule is situated at a definite point in space, at a definite distance from and in a definite angular orientation to other atoms or molecules surrounding it.* Thus, a perfect crystal is considered to be constructed by the infinite regular repetition of identical structural units or building blocks. The actual materials are made up of aggregates of those crystals. Crystals can be *single* or *polycrystalline*. In a single crystal *the orientation of the atoms or molecules are all uniform and continue throughout the entire crystal* as shown in Fig. 1.1. There are no discontinuities, the boundary of that type of crystal is the boundary for the entire crystal, there are no subdivisions, no internal grain boundaries, for example, diamond, ruby, etc. Whereas for polycrystalline material *whole crystal is made up of smaller crystallites* as shown in Fig. 1.2. Each small

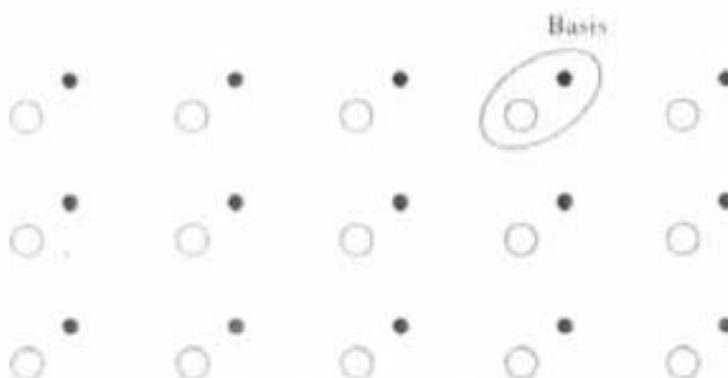


Fig. 1.1 Single crystal structure

crystallite is called a *grain*. These grains are situated side by side to form a whole polycrystalline material (Fig. 1.2), like in quartz. There are grain boundaries between two grains. The orientation of atoms within a particular grain is always uniform, i.e., each grain has the same lattice pattern but grains are generally oriented randomly and thus forming grain boundaries in polycrystalline solid. Single crystals are more preferable than polycrystalline solid, as many properties of solids are best studied with single crystals. Today almost all naturally occurring crystals have been synthesized in the laboratory under controlled conditions. Single crystals may be classified as *elemental crystal* (e.g., Al, Fe, Cu etc.) and *ionic crystal* (e.g., AgCl, CuSO₄ etc.).



Fig. 1.2 Polycrystalline structure with grains and grain boundary

1.2 Definition of Crystal Structure, Lattice Points, Space Lattice, Lattice Plane, Basis, Unit Cell

A crystal is a 3-dimensional body. Regular and periodic arrangement of 3-dimensional pattern of atoms or molecules in-space, is called **crystal structure**, where each and every atom or molecule have the same environment. Within crystal structure, positions of atom or molecule or group of atoms where they are located, represented for simplicity by some points, which are known as '**lattice points**' as shown in Fig. 1.3. A parallel net-like arrangement of lattice points in space (Fig. 1.3), is defined as **space lattice**, which provide the environment around any particular lattice point and is in every way the same as around any other lattice point. Within crystal any plane passing through at least one lattice

point is called **Lattice Plane**. It is essential to distinguish a lattice point, from a crystal structure. A crystal structure like (Fig. 1.1) is formed by associating with every lattice point as shown in (Fig. 1.3), an *unit assembly of atoms or ions* as

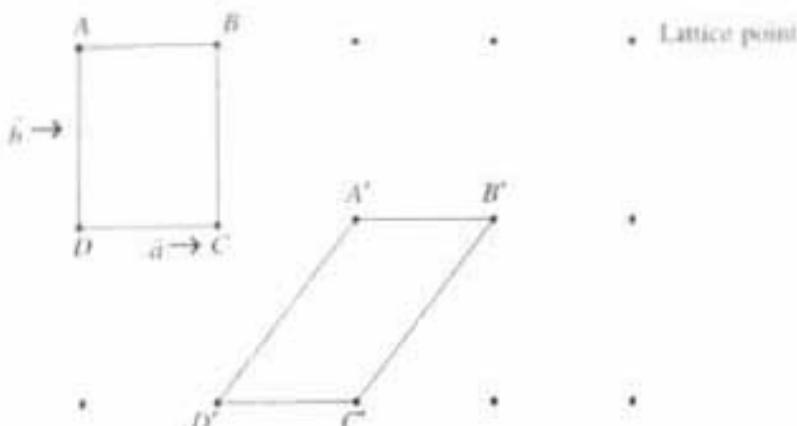


Fig. 1.3 Space lattice and unit cell in two dimension

shown in Fig. 1.4, which is called as **Basis**, they are identical in composition, arrangement and orientation (Fig. 1.4). Hence, *lattice point + basis = crystal structure*.

Unit Cell: The general 2-dimensional lattice (Fig. 1.3) is an infinite array of lattice points, which represent the atomic or molecular positions in the crystal. These points obey the lattice condition that, *every lattice point should have the same environment in the same orientation*.

Let us consider Fig. 1.3, where vectors \vec{a} and \vec{b} are possible choices for the primitive translational vector of the lattice. The general parallelogram ABCD defined by the primitive vectors \vec{a} and \vec{b} , forms a primitive cell which is called the unit cell. Thus, a *unit cell* is defined as the *smallest geometric figure, the repetitions of which gives the actual crystal structure and it should have all the characteristics of the particular crystal*. It can also be defined as the fundamental elementary pattern of a minimum number of atoms or molecules or group of molecules, which represent fully all the characteristics of the crystal.

For the 3-dimensional case, the unit cell is a parallelopiped formed by the basic vectors \vec{a} , \vec{b} , \vec{c} , as concurring edges and including angles α , β , γ as shown in Fig. 1.5. Hence in general, a "unit cell" may be defined as *the smallest volume of a solid from which the entire crystal may be constructed by translational repetitions in 3-dimension and which represent fully all characteristics of the particular crystal*.

The choice of unit cell is not unique. It may be ABCD or A'B'C'D' (Fig. 1.3). It can be constructed in a number of ways, but the unit cell should be chosen in such a way that it conveys all the symmetry of a crystal lattice, by having

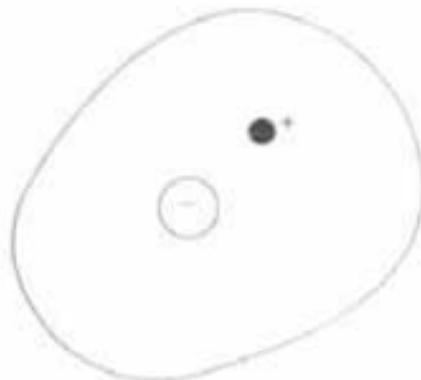


Fig. 1.4 Basis

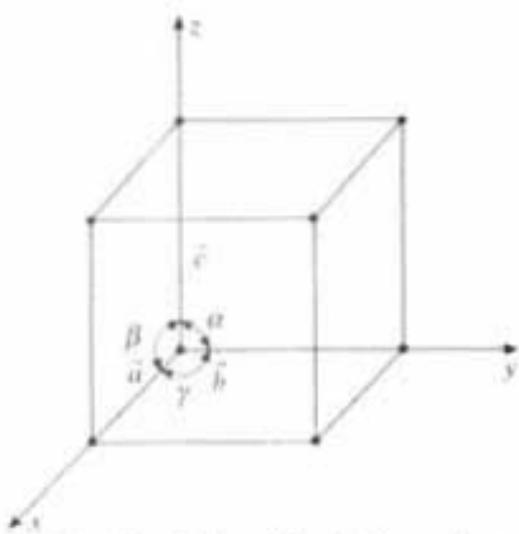


Fig. 1.5 Unit cell in 3-dimension

vectors \vec{a} , \vec{b} , \vec{c} are the basic vectors along x , y , z axes called crystallographic axes. The angles between these axes are called interfacial angles which are α , β , γ between vectors $(\vec{b} - \vec{c})$, $(\vec{c} - \vec{a})$, $(\vec{a} - \vec{b})$. Primitive vectors \vec{a} , \vec{b} , \vec{c} and the interfacial angles α , β , γ are together called 'lattice parameter' of the crystal. To determine the actual size of unit cell, the actual values of vectors \vec{a} , \vec{b} , \vec{c} and the interfacial angles must be known.

1.4 Lattice Vector and Direction

In order to specify a certain point say P (Fig. 1.6), or a certain direction, say OP in a unit cell, we use basic vectors. When vectors \vec{a} , \vec{b} , \vec{c} are the basic vectors of the unit cell, then point P is related to the origin by vector transition.

$$\vec{OP} = \vec{T} = n_1 \vec{a} + n_2 \vec{b} + n_3 \vec{c}$$

where n_1 , n_2 and n_3 are three integers.

shortest possible size, which makes the mathematical calculations easy. For example, ABCD will be the unit cell for Cubical crystal when ($a = b$), not A'B'C'D'. Each atom or molecule in a unit cell is considered as a lattice point. The distance a between two atoms or ions of the same type is the "length of the unit cell", and known as *lattice parameter*.

1.3 Lattice Parameter

In a 3-dimensional case, the unit cell is shown as a parallelopiped. In Fig. 1.5,

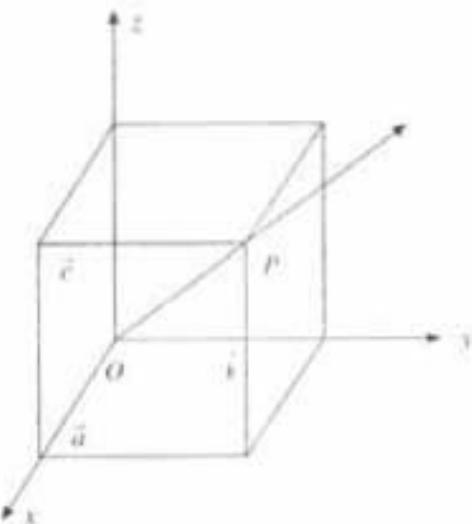


Fig. 1.6 Lattice vector

1.5 Seven Crystal System

According to the different values of the lattice parameter, seven crystal systems can be formed as follows:

S. No.	Crystal System Vectors	Basic	Interfacial Angles
1.	Cubic	$a = b = c$	$\alpha = \beta = \gamma = 90^\circ$
2.	Monoclinic	$a \neq b \neq c$	$\alpha = \beta = 90^\circ \neq \gamma$
3.	Triclinic	$a \neq b \neq c$	$\alpha \neq \beta \neq \gamma \neq 90^\circ$
4.	Tetragonal	$a = b \neq c$	$\alpha = \beta = \gamma = 90^\circ$
5.	Orthorhombic	$a \neq b \neq c$	$\alpha = \beta = \gamma = 90^\circ$
6.	Rhombohedral (trigonal)	$a = b = c$	$\alpha = \beta = \gamma \neq 90^\circ$
7.	Hexagonal	$a = b \neq c$	$\alpha = \beta = 90^\circ, \gamma = 120^\circ$

The shapes of the unit cells are shown in Fig. 1.7. Of these seven crystal systems, the cubic system is the simplest and most symmetrical in which $a = b = c$ and $\alpha = \beta = \gamma = 90^\circ$. It is interesting to note that more than half of the naturally occurring crystals belong to the cubic system.

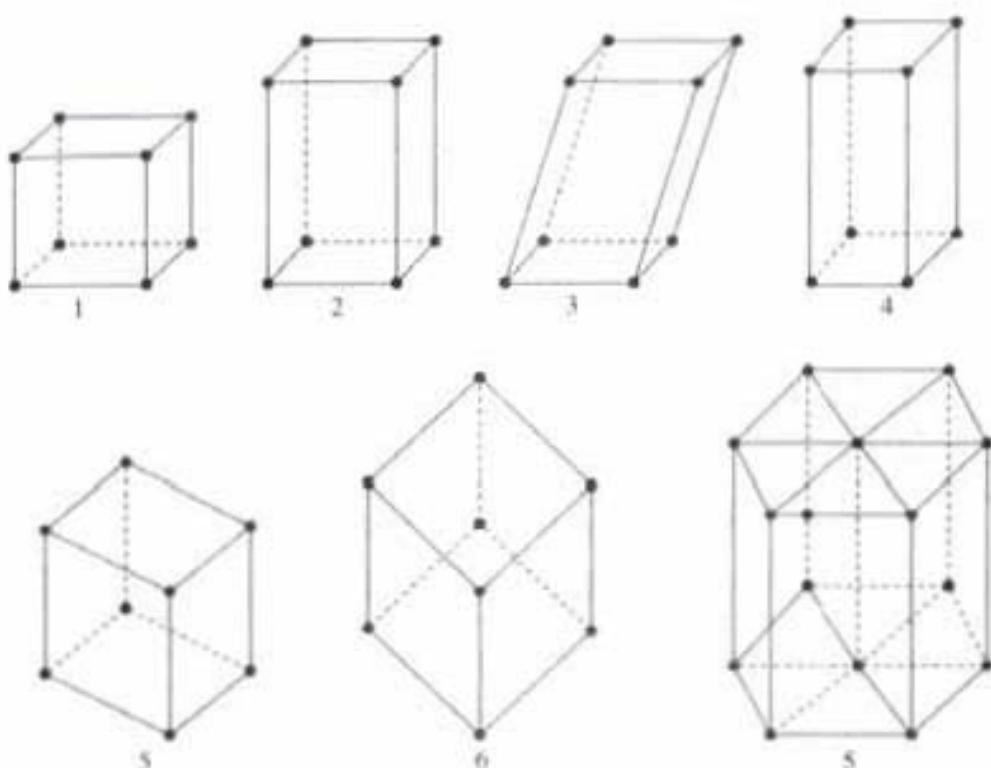


Fig. 1.7 Seven crystal systems

1.6 Bravais Space Lattice¹

In 1945, Bravais showed that there are only 14 ways of arranging points in space so that the environment looks the same from each point. These 14 lattices are called *Bravais space lattices*, each of which has the same collection of symmetry elements at the lattice points.

For cubic system, three types of Bravais lattices are possible, viz.:

- (i) Simple cubic (S.C.) (e.g., palladium)
- (ii) Body-centred cubic (B.C.C.) (e.g., molybdenum)
- (iii) Face-centred cubic (F.C.C) (e.g., Ni, Cu, NaCl)

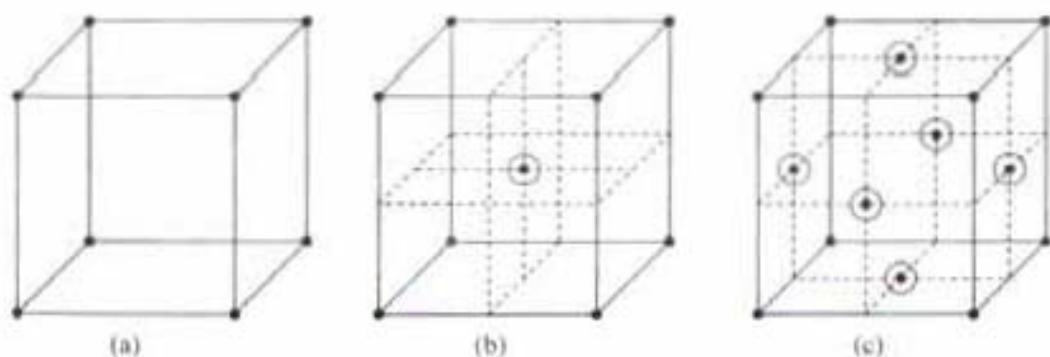


Fig. 1.8 (a) Simple cubic structure (S.C.), (b) Body-centred cubic structure (B.C.C.) and (c) Face-centred cubic structure (F.C.C.)

S. No.	Crystal system	No. of Bravais Lattice
1.	Cubic	3
2.	Tetragonal	2
3.	Orthorhombic	4
4.	Trigonal	1
5.	Monoclinic	2
6.	Hexagonal	1
7.	Triclinic	1
Total		14

1.7 Symmetry Elements of a Crystalline Solid

The main symmetry elements of a crystalline solid are: (1) axes of symmetry, (2) planes of symmetry and (3) centre of symmetry.

1. Axes of symmetry

Consider a crystal is being rotated around an axis. If the crystal occupies two or more identical positions during its rotation through 360° , then that axis is called the *axis of symmetry*.

In a cube, each rotation of 90° brings it into identical or congruent position, i.e., 4 congruent positions in one complete rotation. It is called a *4-fold axis of symmetry*. In general, if a rotation through an angle $(360/n)^\circ$ about an axis brings a figure into congruent position, then that axis is called an *n-fold axis of symmetry*.

There are 13 axes of symmetry for a cube.

- (i) 3 axes are tetrad [4-fold] (Fig. 1.9), which pass through opposite face centres.
- (ii) 4 axes are triad [3-fold] (Fig. 1.10), which pass through diagonally opposite corners.
- (iii) 6 axes are diad (2-fold) (Fig. 1.11), which pass through centres of opposite edges.

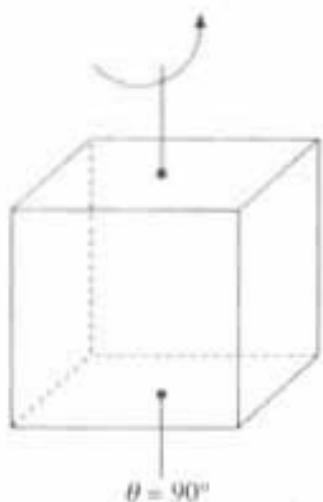


Fig. 1.9 4-fold axes of symmetry

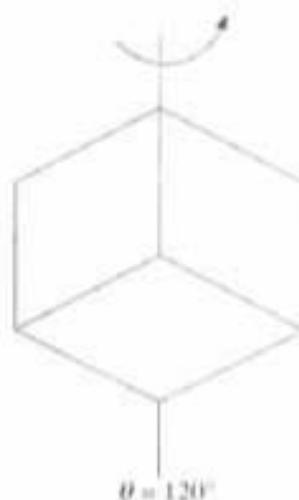


Fig. 1.10 3-fold axes of symmetry

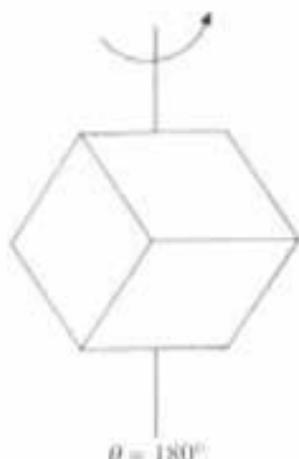


Fig. 1.11 2-fold axes of symmetry

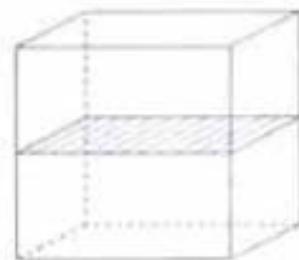


Fig. 1.12 3-parallel plane of symmetry

2. Planes of symmetry

If the crystal is cut into two halves by a plane so that one half becomes the mirror image of the other half, then that plane is called *plane of symmetry*. For cubic system, there are 9 planes of symmetry.

- (i) 3 planes parallel to the faces of the cube (Fig. 1.12);
- (ii) 6 diagonal planes passing through diagonally opposite parallel edges (Fig. 1.13).

3. Centre of symmetry

It is a point in a crystal such that if a line is drawn from any point on the crystal through this point and produces equal distance on the other side of this central point, then it meets an identical point. For cubic system there is one centre of symmetry (Fig. 1.14).

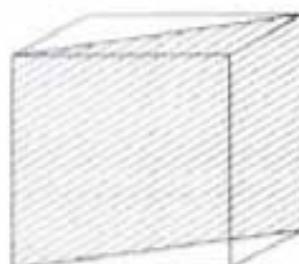


Fig. 1.13 6-diagonal plane of symmetry

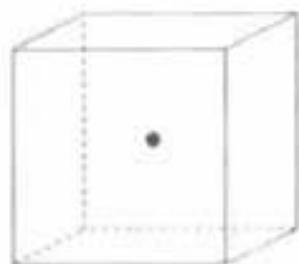


Fig. 1.14 1-centre of symmetry

1.8 Coordinates of Lattice Point

(a) The coordinates of the corner lattice points for S.C., B.C.C. and F.C.C. systems are $(0, 0, 0)$. Any one of these coordinates can be taken as origin. According to that origin, the coordinates of other lattice points are determined (Fig. 1.15).

(b) Body-centred lattice point in B.C.C. system having coordinates $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ (Fig. 1.16).

(c) Face-centred lattice points in F.C.C. system having coordinates $(\frac{1}{2}, \frac{1}{2}, 0)$, $(0, \frac{1}{2}, \frac{1}{2})$, and $(\frac{1}{2}, 0, \frac{1}{2})$ (Fig. 1.17).

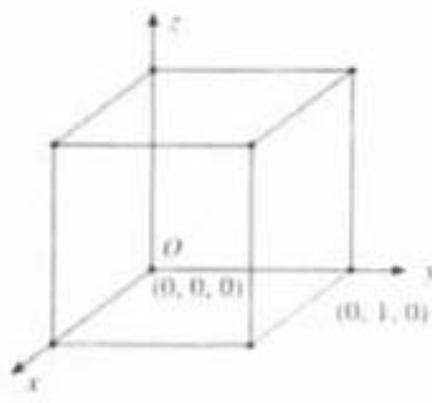


Fig. 1.15 Coordinates of corner lattice points

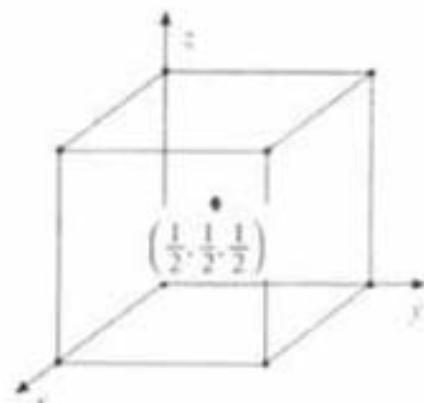


Fig. 1.16 Coordinates of body-centred lattice points

1.9 Number of Atoms/Molecules per Unit Cell (n)

Per unit cell number of atoms in elemental or number of molecules in ionic crystal determines how closely the particular solid is packed up.

1. S.C. Crystal: In S.C. crystal there are 8 atoms, one at each corner. Each corner atom is shared by 8 unit cells. Therefore, the total number of atoms per unit cell $N_C/8 = 8/8 = 1$ (Fig. 1.18).

2. B.C.C. Crystal: In B.C.C. crystal, there is one atom at the centre and 8 atoms at 8 corners. Each corner atom is shared by 8 unit cells. So the total number of atoms per unit cell will be as shown in Fig. 1.19.

$$N_B + N_C/8 = 1 + 1 = 2$$

In ionic crystal, corner and body centred ions should be of same type.

3. F.C.C. Crystal: In F.C.C. crystal, there are 8 atoms at 8 corners. At the centre of 6 faces there are another 6 atoms. Each corner atom is shared by 8 unit cells.

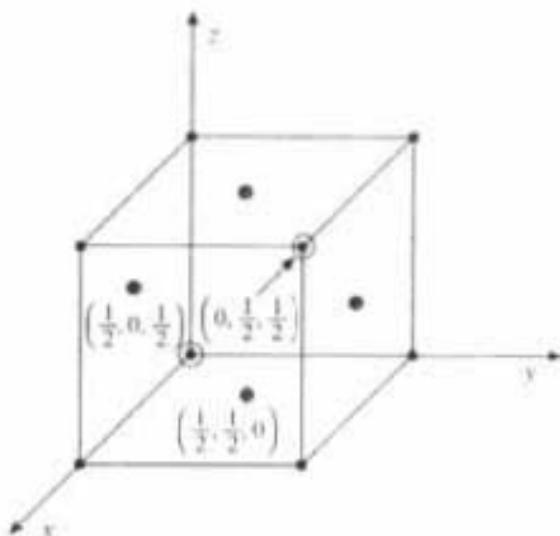


Fig. 1.17 Coordinates of face-centred lattice points

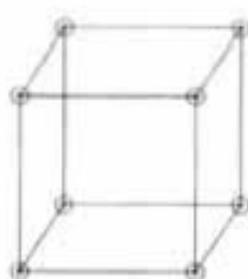


Fig. 1.18 Number of atoms/unit cell in S.C.

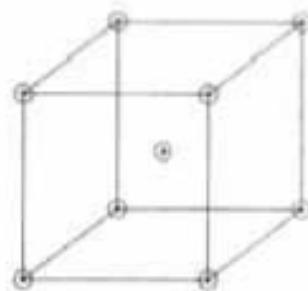


Fig. 1.19 Number of atoms/unit cell in B.C.C.

and each of the 6 face centred atoms is shared by 2 unit cells. Hence, the total number of atoms per unit cell for F.C.C. is equal to (Fig. 1.20).

$$N_C/8 + N_F/2 = 8/8 + 6/2 = 1 + 3 = 4$$

where N_C is the number of corner atoms, N_B the number of body-centred atoms and N_F the number of face-centred atoms.

In ionic crystal, corner and face-centred ions should be of the same type. So in NaCl crystal which is F.C.C. all corners and face centred ions are Na^+ (Fig. 1.20) whereas in between two Na^+ there will be one Cl^- . Each $\text{Cl}^- (N_D)$ which are situated at the middle of 12 edges (i.e. $N_D = 12$) of the unit cell is shared by 4 unit cells as clear from Fig. 1.20, where as one Cl^- ion (i.e. $N_B = 1$) is at the centre of the unit cell. Hence total number of NaCl molecules per unit cell will be

$$(N_C/8 + N_F/2)_{\text{Na}^+} + (N_D/4 + N_B)_{\text{Cl}^-} = (8/8 + 6/2)_{\text{Na}^+} + (12/4 + 1)_{\text{Cl}^-} = (4)_{\text{Na}^+} + (4)_{\text{Cl}^-} = (4) \text{ NaCl molecules}$$

Hence, the number of atoms or molecules per unit cell is 4, which is highest in F.C.C. crystal i.e., F.C.C. crystal is more closely packed than B.C.C. and S.C.

1.10 Coordination Number

In the crystalline solid each atom or ion is surrounded by other atoms or ions. The number of equidistant nearest neighbour of the same type that an atom or ion has in a given structure is called *coordination number*, which will be different for different structures.

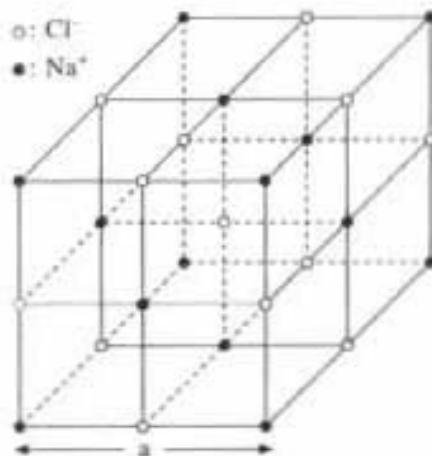


Fig. 1.20 Number of atoms or molecules per unit cell in F.C.C.

1. For S.C. crystal coordination number is 6 (Fig. 1.18). The corner atoms are the nearest neighbours of each other and connected with 8 unit cells. Hence, each corner atom has 4 nearest neighbours in the same plane, one vertically above and one vertically below. *Here nearest neighbour distance is a .*

2. For B.C.C. crystal coordination number is 8 (Fig. 1.19). Here for any corner atom the nearest neighbours are the body centred atoms. Since there are eight surrounding unit cells for any corner atom and each of them have a body centred atom, so the coordination number for any corner atom is 8. Similarly the centre atom is surrounded by 8 corner atoms so coordination number for centre atom is also 8. *Here nearest neighbour distance is $\sqrt{2}a/4$.*

3. For F.C.C. crystal coordination number is 12 (Fig. 1.20). Here the nearest neighbours of any corner atom are face centred atoms of the surrounding unit cells. For any corner atom there are 4 face-centred nearest neighbour in its own plane, 4 in a plane above it, and 4 in a plane below it. *Hence nearest neighbour distance is $a/\sqrt{2}$.*

1.11 Atomic Radius (r)

In calculating atomic radius, it is assumed that the atoms are spheres and are in contact with each other in a crystal. The atomic radius is defined as '*half the distance between the nearest neighbours in a crystal of pure element.*' The atomic radius r is always expressed in terms of cube edge a .

1. S.C. Crystal: The front view of a S.C. crystal is shown in Fig. 1.21. There are 8 atoms situated at 8 corners. In S.C. crystal, corner atoms are the nearest neighbour of each other. So the corner atoms will be in touch with each other. If a is the side of the unit cell and r the radius of the atom, then from Fig. 1.21,

$$2r = a \quad \text{or} \quad r = a/2$$

Thus, nearest neighbour distance in S.C. is $2r = a$.

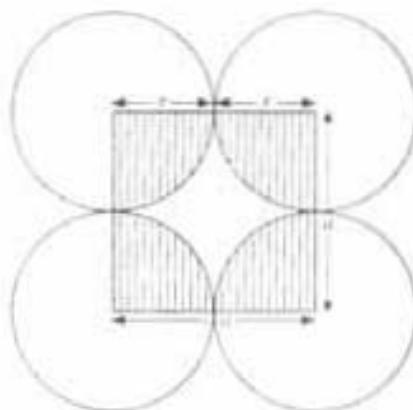


Fig. 1.21 Atomic radius for S.C.

2. B.C.C. Crystal: In B.C.C. crystal corner atom and body centred atom will be in touch, because they are the nearest neighbour. Atomic radius r can be calculated for the B.C.C. crystal in the following way:

Let us consider Fig. 1.22, where

$$(DF)^2 = (DA)^2 + (AF)^2$$

$$(4r)^2 = a^2 + (a^2 + a^2)$$

that is

$$r = \sqrt{3}a/4$$

Hence, nearest neighbour distance in B.C.C. is $2r = \sqrt{3}a/2$.

3. F.C.C Crystal: In F.C.C. crystal corner atom and face-centred atom will be in touch, because they are the nearest neighbour. Atomic radius r can be calculated for an F.C.C. crystal as follows:

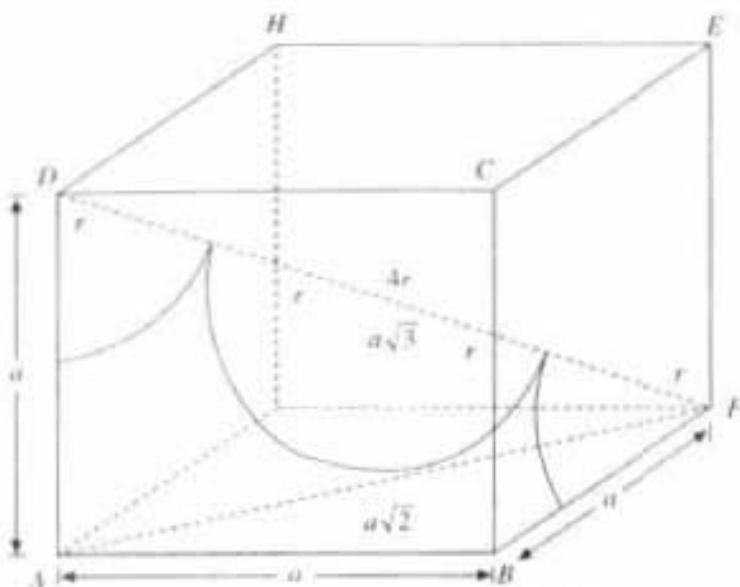


Fig. 1.22 Atomic radius for B.C.C.

Consider Fig. 1.23, where

$$(DB)^2 = (DC)^2 + (BC)^2$$

$$(4r)^2 = a^2 + a^2$$

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

Nearest neighbour distance in F.C.C. is $2r = a/\sqrt{2}$.

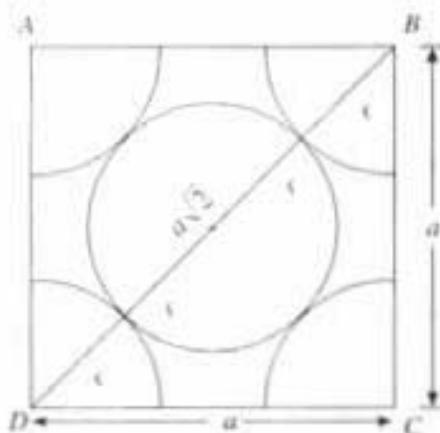


Fig. 1.23 Atomic radius for F.C.C.

1.12 Packing Density or Atomic Packing Factor

In the study of crystal structure, one of the important concepts is packing efficiency of atoms or ions in a crystal. Higher packing efficiency provides efficient packing of atoms in solids. The concept of packing is understood by a parameter known as atomic packing factor of the crystal. *Packing density or atomic packing factor is defined as the ratio of the volume occupied by the atoms per unit cell to the total volume of the unit cell.*

$$\text{A.P.F.} = \frac{\text{Volume occupied by the atoms in the unit cell}}{\text{Volume of the unit cell}} = \frac{n \frac{4}{3}\pi r^3}{a^3} =$$

For ionic crystal, the ionic packing factor is

$$\text{I.P.F.} = \frac{\text{Volume occupied by the ions in the unit cell}}{\text{Volume of the unit cell}}$$

1. S.C. Crystal

No. of atoms/unit cell = 1

Volume of one atom = $4/3 \pi r^3$

Side of the unit cell = $a = 2r$

Volume of the unit cell = a^3

Thus, packing density

$$\text{A.P.F.} = \frac{4/3 \pi r^3}{a^3} = \frac{4/3 \pi r^3}{(2r)^3} = \pi/6$$

2. B.C.C. Crystal

Number of atoms/unit cell = 2

Volume of two atoms = $2 \cdot 4/3 \pi r^3$

Side of unit cell = $a = 4r/\sqrt{3}$

Volume of unit cell = a^3

Hence, packing density

$$\text{A.P.F.} = \frac{2 \cdot 4/3 \pi r^3}{a^3} = \frac{2 \cdot 4/3 \pi r^3}{(4r/\sqrt{3})^3} = \sqrt{3}\pi/8$$

3. F.C.C. crystal

No. of atoms/unit cell = 4

Volume of four atoms = $4 \cdot 4/3 \pi r^3$

Side of unit cell = $a = 4r/\sqrt{2}$

Volume of unit cell = a^3

So, packing density

$$\text{A.P.F.} = \frac{4 \cdot 4/3 \pi r^3}{(4r/\sqrt{2})^3} = \frac{\sqrt{2}\pi}{6}$$

Crystal	Coordination No. [CN]	Atomic Radius (r)	Atoms per Unit Cell (n)	A.P.F.
S.C.	6	$a/2$	1	$\pi/6$
B.C.C.	8	$\sqrt{3}a/4$	2	$\sqrt{3}\pi/8$
F.C.C.	12	$\sqrt{2}a/4$	4	$\sqrt{2}\pi/6$

1.13 Calculation of Lattice Constant or Crystal Density

In the case of cubic crystal, lattice constant is a . If D is the density of the crystal under consideration, then

$$\text{Volume of the unit cell} = a^3$$

$$\text{Mass in each unit cell} = Da^3 \quad (1.1)$$

Further, if there are n number of molecules or atoms (lattice points) per unit cell, M Mol. wt. and N Avogadro number (No. of mols/gm. mole of substance) then

$$\text{Mass of each mol.} = M/N$$

$$\text{Mass in each unit cell} = n \cdot M/N \quad (1.2)$$

From (1.1) and (1.2) we get

$$Da^3 = n \cdot M/N$$

$$a^3 = (nM/ND)$$

$$D = nM/Na^3 \quad (1.3)$$

n = Number of atoms or mol/unit cell.

M = Mol. wt. or Atomic wt.

N = Avogadro number = 6.023×10^{23} No. of mols/gm mols.

= 6.023×10^{26} No. of mols/kg mols,

1.14 Miller Indices

The crystal structure may be regarded as made up of an aggregate of a set of parallel equidistant planes passing through at least one lattice point or a number of lattice points, which are known as "lattice planes". For a given crystal, lattice planes can be chosen in different ways, e.g., (a), (b), (c) shown in Fig. 1.24.

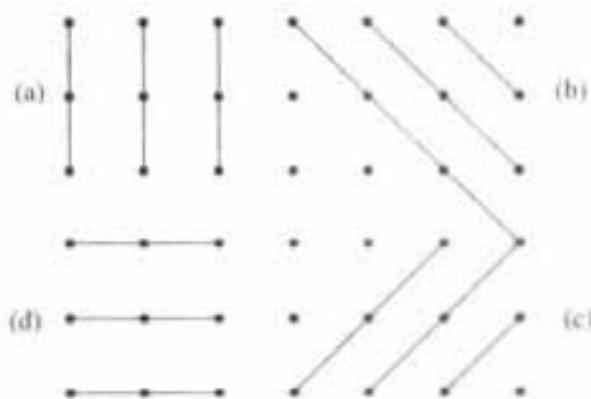


Fig. 1.24 Different lattice planes

In order to designate a lattice plane scientifically, Miller developed a method by three numbers (hkl), which are known as "*Miller indices*". This method is universally employed. Miller Indices can be defined in the following way.

"*Miller Indices*" are the three smallest possible integers h, k, l , which have the same ratios as the reciprocals of intercepts of the concerned plane, on the three axes. The Miller indices for a particular crystal plane are determined in the following way

(1) First, determine the intercepts of the plane on the three coordinate axes say p, q, r .

(2) Second, take the reciprocals of these numbers and by taking L.C.M.

reduce them to the smallest three integers as h, k, l , such that $h : k : l = \frac{1}{p} : \frac{1}{q} : \frac{1}{r}$. The result is then enclosed in parentheses as (hkl) .

(3) Then the concerned plane is called to have the Miller indices (hkl) .

As an example, for a plane whose intercepts are 2, 3, 4, along 3 axes, the reciprocals are $1/2, 1/3, 1/4$. Hence, according to the definition of Miller indices

$$h : k : l = 1/2 : 1/3 : 1/4 = 6 : 4 : 3$$

Thus, Miller indices of the above plane are (643) .

1.15 Different Crystal Planes with Miller Indices

The Miller indices (hkl) denote a single plane or a set of parallel planes. If a plane cuts an axis on the negative side of the origin, the corresponding index will then be negative, as indicated by placing a bar above that index, e.g., $(\bar{h}\bar{k}\bar{l})$. Miller indices for six faces of cubic crystal are shown in (Fig. 1.25(a)):

BCGF (100), CDHG (010), ABCD (001), w.r.t. origin at E.

ADHE ($\bar{1}00$), ABFE ($0\bar{1}0$), EHGF ($00\bar{1}$) w.r.t. origin at C.

These six planes are equivalent by symmetry, which together are denoted by curly brackets around Miller indices as $\{100\}$. Figure 1.25 (b) and (c) shows planes of Miller Indices (101) and (111) , respectively.

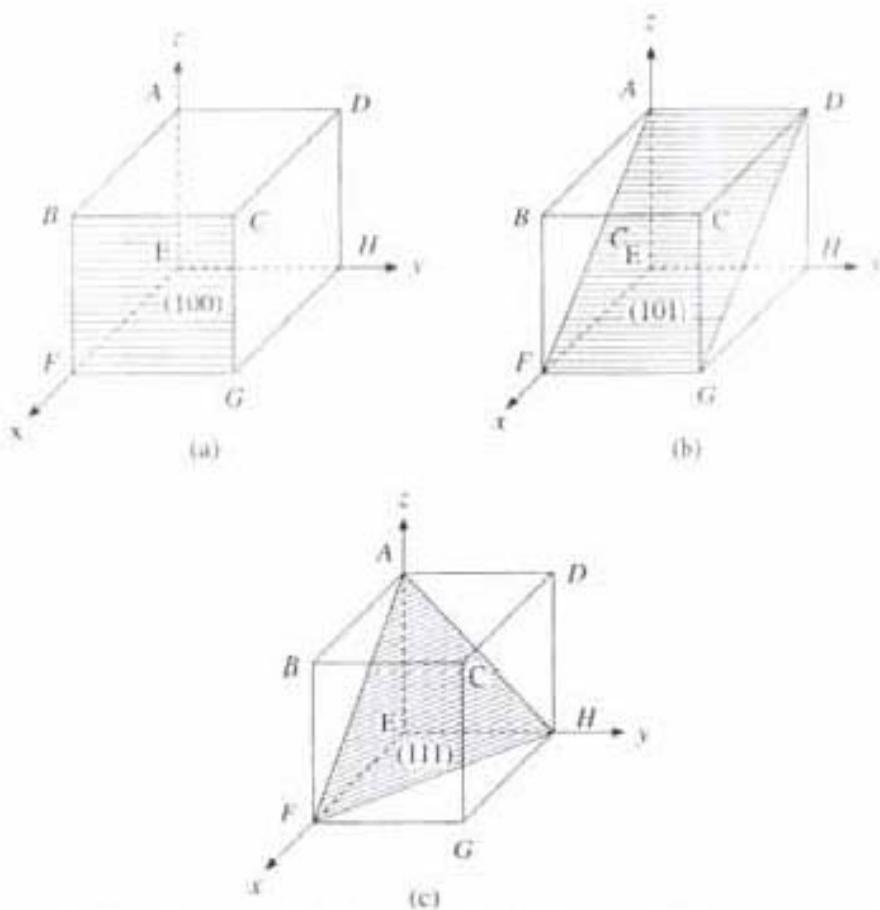


Fig. 1.25 (a) BCGF plane: (100) Miller indices, (b) ADGF plane: (101) Miller indices, (c) AHF plane: (111) Miller indices

It is important to note that parallel planes which are crystallographically equivalent have identical Miller indices, i.e., (010) is equivalent to (020), (111) is equivalent to (222), (110) is equivalent to (220) as shown in Fig. 1.26.

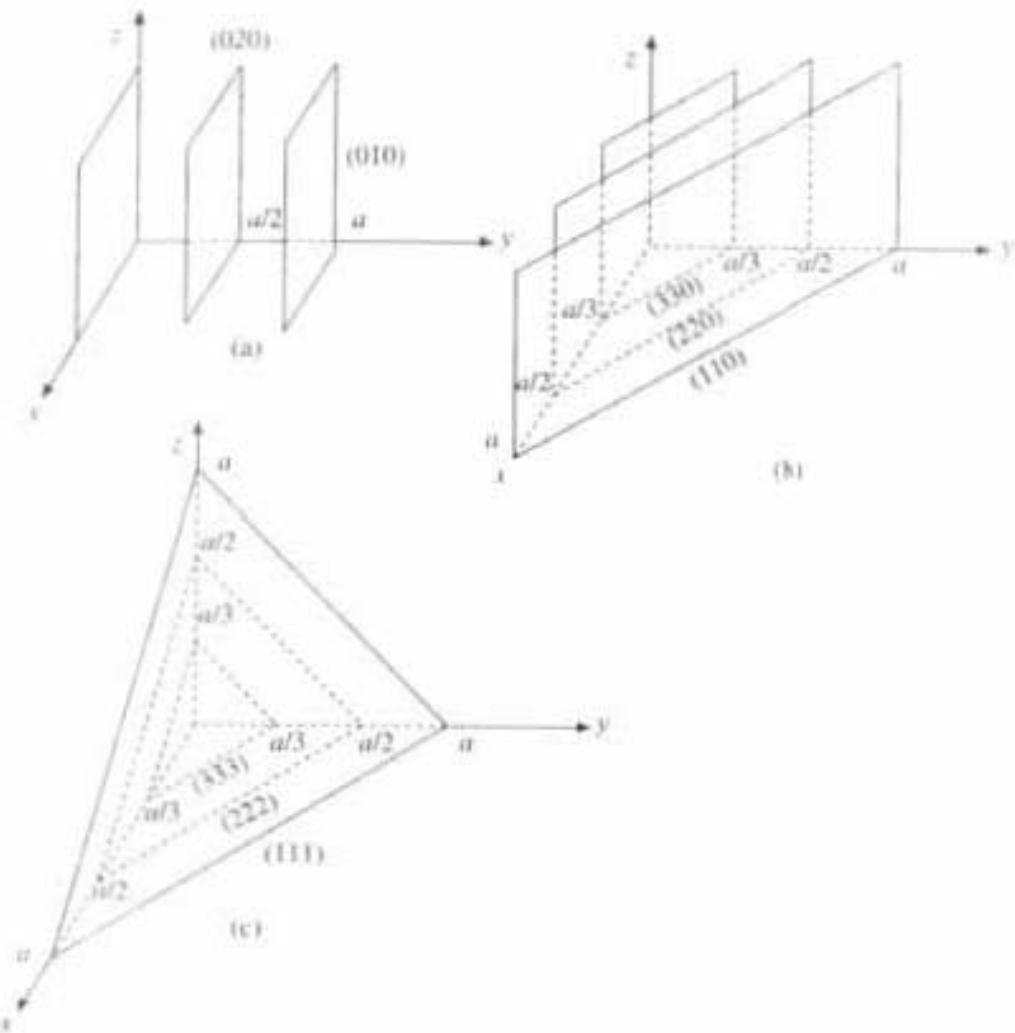


Fig. 1.26 (a) Different (010) Miller planes, (b) Different (110) Miller planes, (c) Different (111) Miller planes

Example 1. Draw the following Miller planes: (a) (121); (b) (12 $\bar{2}$); [c] ($\bar{1}10$);

(1) Miller plane is (121)

Intercepts are 1, 1/2, 1

Hence, the plane ABC is (121) plane (Fig. 1.27(a)).

(2) Miller plane is (12 $\bar{2}$)

Intercepts are 1, 1/2, -1/2

Thus, the plane ABC is (12 $\bar{2}$) plane (Fig. 1.27(b)).

(3) Miller plane is ($\bar{1}10$)

Intercepts are (-1, 1, ∞)

Therefore, the plane ABCD is ($\bar{1}10$) plane (Fig. 1.27(c)).

Example 2. Draw the different (111) planes in the unit cell of simple cubic system.

Figure 1.27 (d) Shows the unit cell of the simple cubic system. The four (111) planes are AFH, CFH, ACH, ACF.

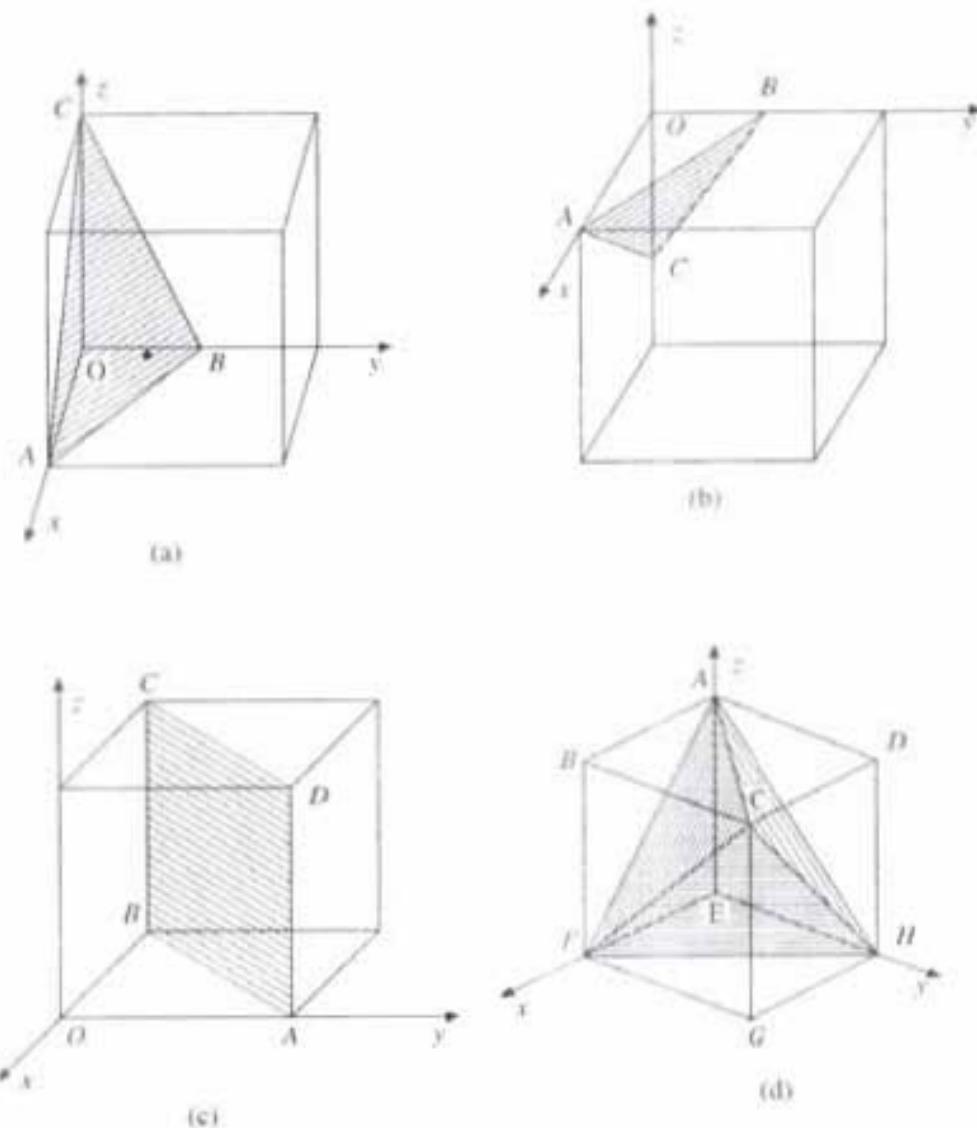


Fig. 1.27 (a) ABC plane: (121) Miller indices, (b) ABC plane: (122) Miller indices, (c) ABCD plane: (110) Miller indices, (d) Four (111) plane in a cube

1.16 Direction of a Line (Lattice Vector)

Direction of a line in space can be specified satisfactorily by giving the coordinates through which each line passes.

The square brackets are usually used to indicate a direction. For cubic system normal to the plane whose Miller indices (hkl) is the direction $[hkl]$. The directions of various lines of Fig. 1.28 are as follows:

$$\begin{array}{lll} OA = [110] & OR = [100] & OP = [\frac{1}{3}, \frac{2}{3}, 1] \text{ or } [123] \\ OB = [010] & OD = [\frac{1}{2}, \frac{1}{2}, 1] \text{ or } [112] & OG = [101] \\ OC = [111] & OE = [1, 0, \frac{1}{2}] \text{ or } [201] & OH = [\frac{1}{2}, 1, \frac{1}{2}] \text{ or } [121] \end{array}$$

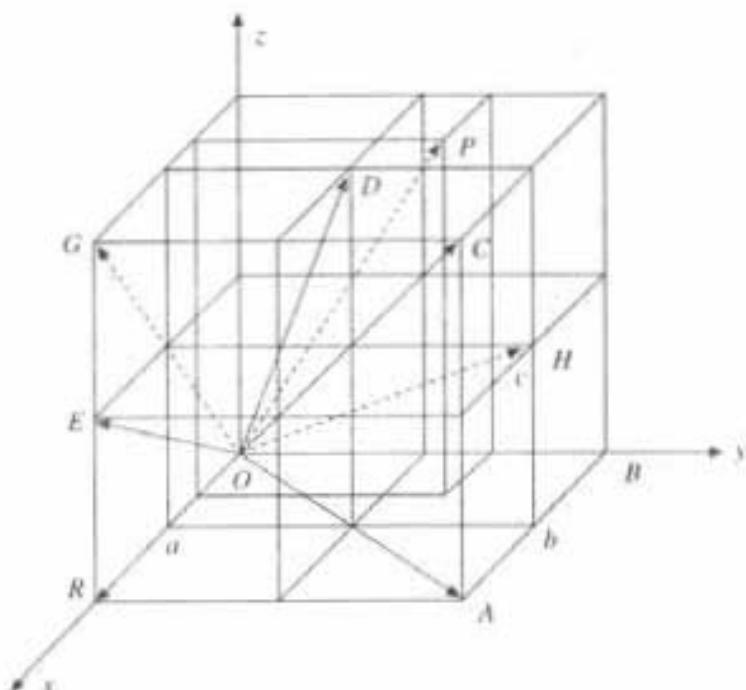


Fig. 1.28 Direction of a line

PROCEDURE TO FIND OUT THE MILLER-DIRECTION [hkl] (SAY FOR EXAMPLE [121]):--

Step 1. Divide all three integers h, k, l by the highest number (say k)

Step 2. So it will be $h/k, k/k, l/k$ i.e. $h/k : k/k : l/k = h : k : l$ (eg. $\frac{1}{2} : \frac{2}{2} : \frac{1}{2} = 1 : 2 : 1$)

Step 3. Draw three planes as shown in the Fig. (1.28) at $h/k, k/k$ & l/k
(eg. at $\frac{1}{2}, \frac{2}{2}, \frac{1}{2}$)

Step 4. Now (H) is the point where those three planes will intersect. Join (H) with origin (O), then (OH) will be the Miller-direction [h k l]. (eg. [121])

[121] direction can be also done by an easy method in the following way:

Step 1. Divide all three integers h, k, l by the highest number (say k)

Step 2. So it will be $h/k, k/k, l/k$, i.e., $h/k : k/k : l/k = h : k : l$

Step 3. From origin (O) go a distance (h/k) (i.e. $\frac{1}{2}$) along x-axis (say O a).
from that point (a) go a distance (k/k) (i.e. 1) along Y-axis (say a b).
from that point (b) go to a distance (l/k) (i.e. $\frac{1}{2}$) along Z-axis (say b c).
then (c) is the point which is same as point (H) as discussed earlier.

Step 4. Join (OH) which is the Miller direction [h k l]. (eg. [121])

Say for another example of the Miller-direction [1 2 3] $\rightarrow 1/3 : 2/3 : 1 = 1 : 2 : 3$, so go $1/3$ along X, $2/3$ along Y and 1 along Z. You will reach to the point (P). Join (P) with origin (O). Then (OP) is the [1 2 3] direction. Same point (P) will come by the 1st method also, as shown in the Fig. 1.28.

1.17 Interplaner Distance

The "Interplaner Spacing or distance" is the perpendicular distance d_{hkl} between the corresponding (hkl) planes, which can be computed by taking the general case as follows. It is also the perpendicular distance from origin to that set of parallel planes (hkl) . Let us consider Fig. 1.29 where the plane ABC of a cubic crystal belongs to a family of planes whose Miller indices are (hkl) . Let the perpendicular ON be drawn from origin to the plane ABC .

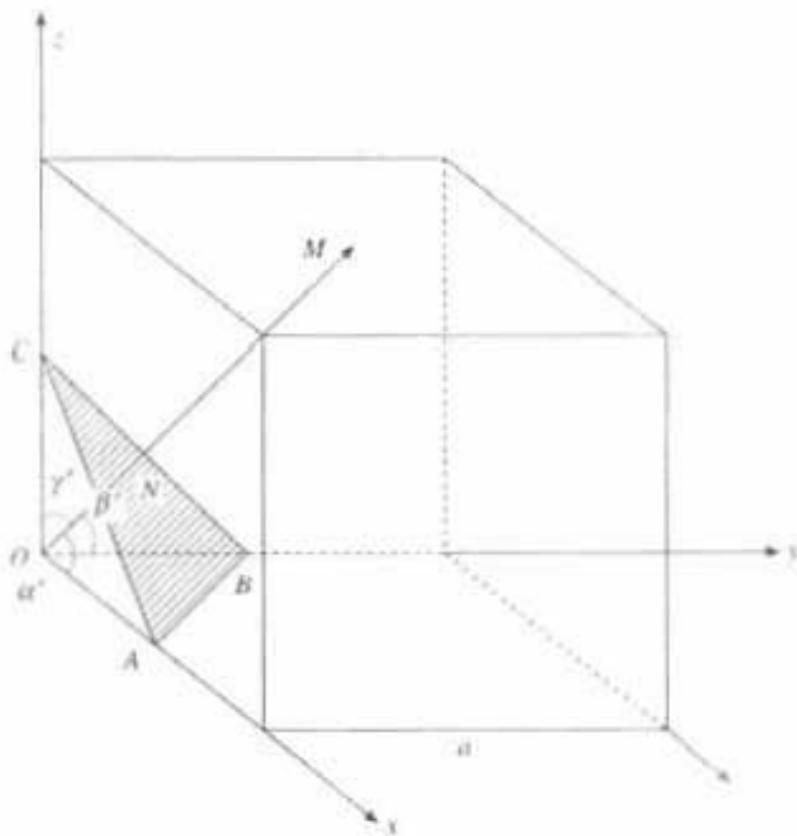


Fig. 1.29 Inter-planer distance

Let us suppose ON makes an angle α' , β' , γ' (Figs. 1.29 and 1.30), with X , Y , Z axes, respectively. The intercepts of plane ABC whose Miller indices are (hkl) on these axes are:

$$OA = a/h, OB = a/k, OC = a/l \quad (1.4)$$

where a = lattice constant.

Now, from Fig. 1.29,

$$\cos \alpha' = ON/OA, \cos \beta' = ON/OB, \cos \gamma' = ON/OC \quad (1.5)$$

The distance ON represents the interplaner spacing d_{hkl} of the family of planes (hkl) . So from (1.4) and (1.5), we get

$$\cos \alpha' = \frac{d}{a/h}, \cos \beta' = \frac{d}{a/k}, \cos \gamma' = \frac{d}{a/l} \quad (1.6)$$

Let us consider Fig. 1.30, where the coordinates of N are (x, y, z) .

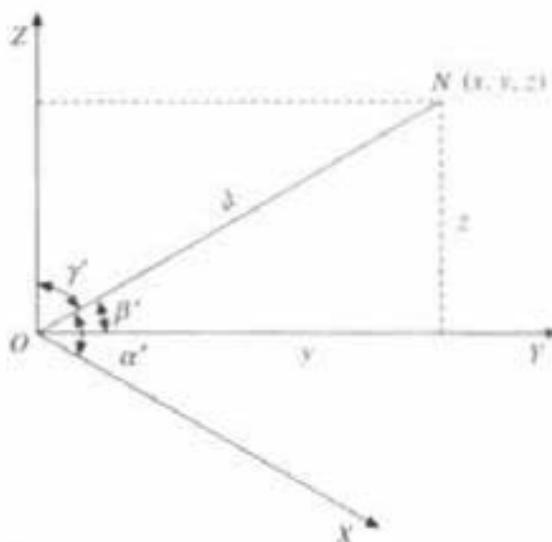


Fig. 1.30 Enlarged version of Fig. 1.29

The line ON makes an angle α' , β' , γ' with X , Y , Z axes, respectively. So

$$(ON)^2 = (d)^2 = (x^2 + y^2 + z^2) \quad (1.7)$$

where $x = d \cos \alpha'$, $y = d \cos \beta'$, $z = d \cos \gamma'$

$$\text{i.e. } d = \sqrt{d^2 \cos^2 \alpha' + d^2 \cos^2 \beta' + d^2 \cos^2 \gamma'} \quad (1.8)$$

$$\text{So, } \cos^2 \alpha' + \cos^2 \beta' + \cos^2 \gamma' = 1 \quad (1.9)$$

Substituting the values of $\cos^2 \alpha'$, $\cos^2 \beta'$ and $\cos^2 \gamma'$ from (1.6), we get

$$(dh/a)^2 + (dk/a)^2 + (dl/a)^2 = 1$$

$$\frac{d^2}{a^2} (h^2 + k^2 + l^2) = 1$$

$$d_{hkl} = al\sqrt{h^2 + k^2 + l^2} \quad (1.10)$$

This is the expression for interplaner spacing in terms of lattice constant a and Miller indices h , k and l .

1.18 Interplaner Spacing for Cubic System

1. S.C. Crystal

$$(i) d_{100} = \frac{a}{\sqrt{h^2 + k^2 + l^2}} = a$$

$$(ii) d_{110} = \frac{a}{\sqrt{h^2 + k^2 + l^2}} = a/\sqrt{2}$$

$$(iii) d_{111} = \frac{a}{\sqrt{h^2 + k^2 + l^2}} = a/\sqrt{3}$$

So,

$$(d_{100} : d_{100} : 111) = 1 : 1/\sqrt{2} : 1/\sqrt{3} = 1 : 0.71 : 0.53$$

$$\left[\frac{1}{d_{100}} : \frac{1}{d_{100}} : \frac{1}{d_{100}} \right]_{SC} = 1 : \sqrt{2} : \sqrt{3}$$

2. F.C.C. Crystal

The (100), (110) and (111) planes for F.C.C. lattice are shown in Figs. 1.31 (a, b, c). Due to the presence of extra lattice point on each face of F.C.C. crystal there will be some additional planes in addition to the original planes in S.C. crystal.

(i) (100) Plane: In F.C.C. crystal (Fig. 1.31a) it is clear that an additional (100) plane arises half-way between two (100) planes. These planes pass through the centres of the front and back faces. Now the interplaner distance of new (100) planes for F.C.C. crystal is,

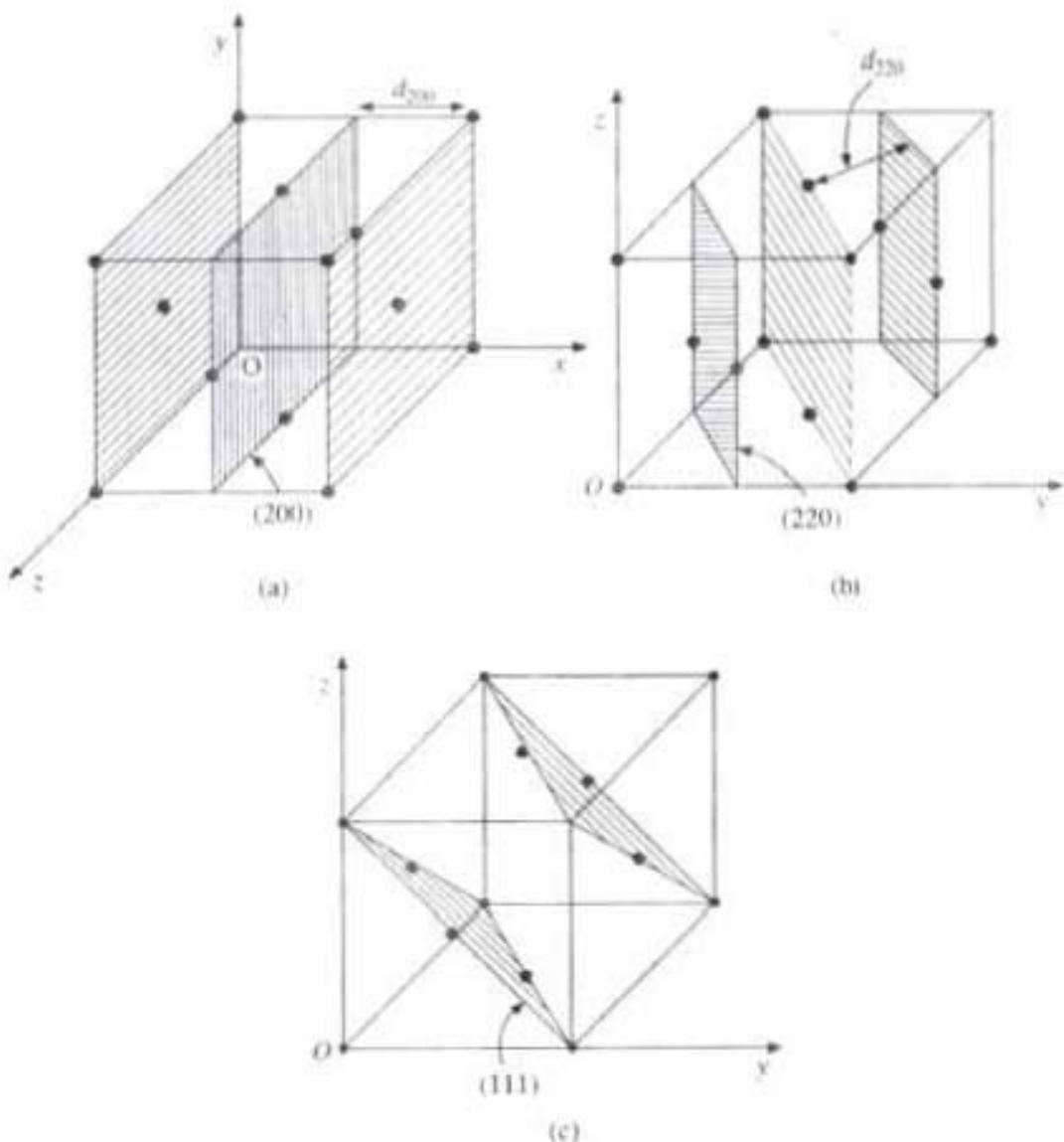


Fig. 1.31 (a) Additional (100) as (200) plane for F.C.C., (b) Additional (110) as (220) plane for F.C.C., (c) Original (111) plane for F.C.C..

$$\text{new } d_{100} = (d_{200}) \text{ F.C.C.} = \frac{1}{2} (d_{100}) \text{ S.C.} = a/2$$

This new (100) plane can be called as (200) plane w.r.t. origin O.

(ii) (110) plane: It is clear from Fig. 1.31(b) that there arises an additional set of (110) planes parallel to the first set. These planes come between two (110) planes. So, the interplaner spacing for new (110) plane for F.C.C. crystal is

$$\text{new } d_{110} = (d_{220}) \text{ F.C.C.} = \frac{1}{2} (d_{110}) \text{ S.C.} = a/2\sqrt{2}$$

This new (110) plane can be called as (220) Plane, w.r.t. Origin O.

(iii) (111) Plane: Figure. 1.31(c) clearly shows that in the case of (111) plane no new plane arises due to face centred point. So,

$$(d_{111}) \text{ F.C.C.} = (d_{111}) \text{ S.C.} = a/\sqrt{3}$$

So,

$$[1/d_{200} : 1/d_{220} : 1/d_{111}]_{\text{F.C.C.}} = 1 : \sqrt{2} : \sqrt{3}/2$$

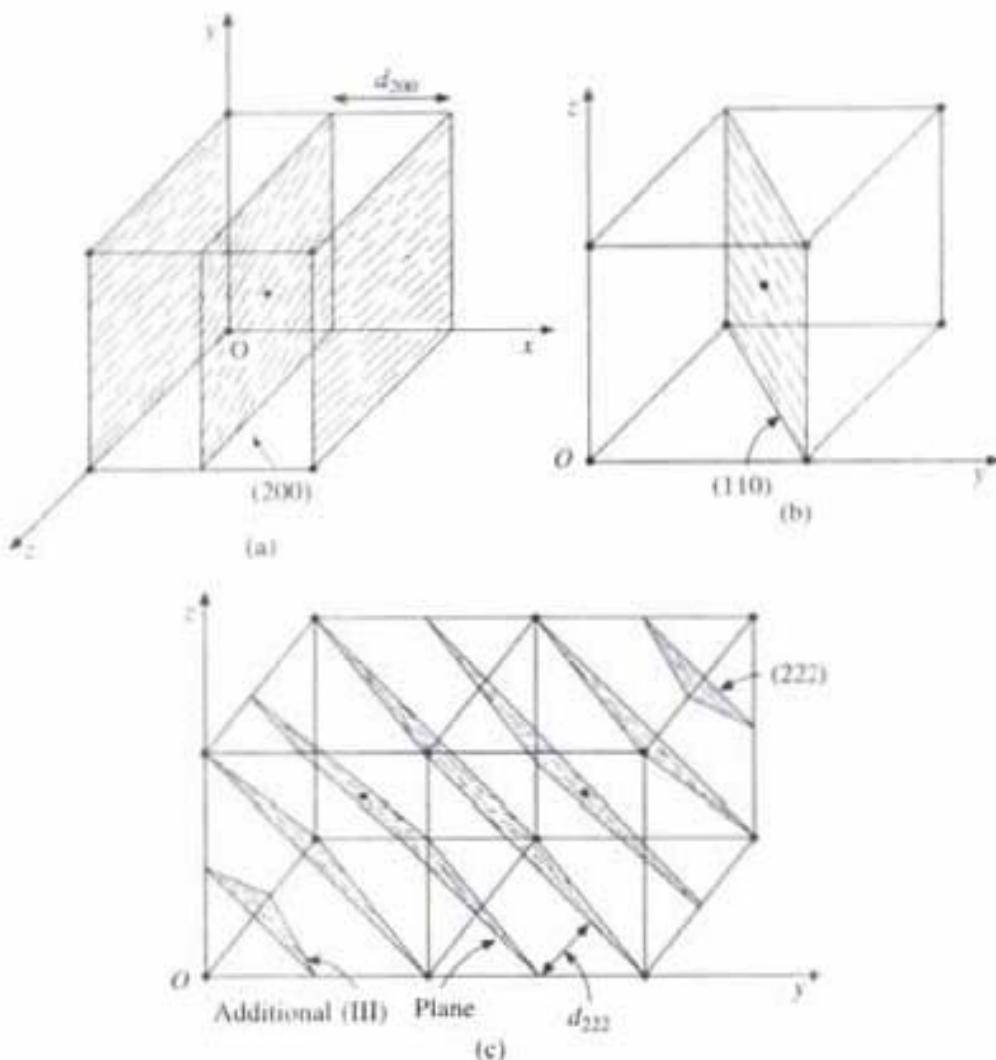


Fig. 1.32 (a) Additional (100) as (200) plane for B.C.C., (b) Original (110) plane for B.C.C., (c) Additional (111) as (222) plane for B.C.C.

3. B.C.C. Crystal

The (100), (110) and (111) planes for B.C.C. are shown in Figs. 1.32 (a), (b), (c). Due to an extra lattice point at the body centre of the B.C.C. crystal, here also some extra additional planes will come.

(i) (100) Plane: From Fig. 1.32(a), it is clear that an additional (100) Plane arises halfway between two (100) planes in B.C.C. So,

$$\text{new } d_{100} = (d_{200}) \text{ B.C.C.} = \frac{1}{2} (d_{100}) \text{ S.C.} = a/2$$

This new (100) plane can be called as (200) plane w.r.t. origin O.

(ii) (d_{100}) Plane: From Fig. 1.32(b), it is clear that no extra (110) plane is possible for B.C.C.. So,

$$(d_{110}) \text{ B.C.C.} = (d_{110}) \text{ S.C.} = a/\sqrt{2}$$

(iii) (111) Plane: From Fig. 1.32(c), it is clear that for B.C.C., additional (111) planes arise midway between two (111) planes. This new (111) plane can be called as (222) plane w.r.t. origin O.

Hence

$$\text{new } d_{111} = (d_{222}) \text{ B.C.C.} = \frac{1}{2} (d_{111}) \text{ S.C.} = a/2\sqrt{3}$$

$$(1/d_{200} : 1/d_{110} : 1/d_{222}) \text{ B.C.C.} = 1 : 1/\sqrt{2} : \sqrt{3}$$

1.19 Important Features of Miller Indices

The important features of Miller indices, particularly for the cubic system are:

- All equally spaced parallel planes have the same index number (hkl).
- A plane parallel to one of the coordinate axes has an intercept of infinity and Miller index will be zero for that axis.
- Miller indices do not define a particular plane but a set of parallel planes.
- It is only the ratio of the indices which is of importance, (422) planes are the same as (211) planes.
- A plane passing through origin is defined in terms of a parallel plane having non-zero intercept.
- For cubic system, normal to the plane with Miller indices (hkl) is the direction [hkl].
- Interplaner distance d between adjacent planes or a set of parallel planes of the Miller indices (hkl) is given by

$$d_{hkl} = a/\sqrt{(h^2 + k^2 + l^2)}$$

CRYSTAL IMPERFECTION

An ideal crystal is one in which each atom has definite equilibrium location in a regular array. Such a crystal is very difficult to obtain in practice. All real

crystals are in some respect imperfect. An imperfection means any deviation from perfect regular lattice or structure.

1.20 Effects on Crystal Due to Presence of Imperfection

Imperfection or defects in the crystal structure means missing atoms, atoms out of place, irregularities in the spacing of rows of atoms, presence of impurities, etc., which have a considerable effect upon the physical properties of the crystal. A number of important properties of solids are controlled far more by the nature of the imperfections than by the nature of the host lattice, which may be only a vehicle for the imperfection. The following are important effects on crystal due to the presence of imperfections:

1. Conductivity of extrinsic semi-conductors may be dictated entirely by the trace amount of chemical impurities, as *n*- or *p*-type semi-conductors.
2. Colour of the crystal arises due to imperfection.
3. Mechanical and plastic properties of solids are usually controlled by dislocation, a special type of imperfection.
4. Diffusion through solids may be accelerated enormously by the presence of impurities.
5. Luminescence of the crystal is nearly always connected with the presence of impurities.
6. Imperfections shows marked influence on electrical, thermal conduction and on hysteresis.
7. Imperfection in some cases produces substitute properties like increase in hardness and strength.

1.21 Causes of Imperfections in Crystals

Imperfections in a crystal arises due to following reasons:

1. Thermal vibrations in a crystal at high temperature.
2. Particle radiation in a nuclear detector.
3. Artificial doping of impurities.
4. Plastic deformation.
5. During growth.
6. By quenching.
7. By severe deformation (hammering, rolling, etc.).

1.22 Different Types of Crystal Defects

1.22.1 Point Defect

The simplest type of imperfection in a solid is the point defect. Point defects are generally of the following types:

- | | |
|-------------------------------|-----------------------------|
| (a) Vacancy | (b) Interstitial |
| (c) Substitutional impurities | (d) Interstitial impurities |
| (e) Schottky defect | (f) Frenkel defect |

The first four types of defects occur in elemental crystals, whereas the last two types occur in ionic crystals.

(a) *Vacancy*: Vacancy is produced due to the removal of an atom from its position in the lattice (Fig. 1.33a). Certain amount of energy is required to break the bond and to free the atom from its position. Usually, there is a reduction in the strength of bonding around the vacancy region. In general, the lattice tends to collapse around the region of vacancy which causes distortion in the lattice. To create vacancy the energy required is of the order of 1-2 eV, which occurs in all elemental crystals as a result of thermal excitation and increases with increase in temperature. The removed atom does not vanish. It travels to the surface of the material.

(b) *Interstitial*: Here an extra atom is fitted into the void between the regularly occupied sites in a crystal (Fig. 1.33b). It is also created by thermal agitation and causes distortion in the lattice and produces local strain. Since in general the size of atom is larger than the void into which it is fitted, so the energy required for interstitial formation is higher than that of vacancy formation.

(c) *Substitutional Impurities*: In this, a foreign atom is occupying a regular site in a crystal lattice (Fig. 1.33c).

(d) *Interstitial Impurities*: Here a foreign atom or ion occupies a place into the void between regularly occupied sites in crystals (Fig. 1.33d). The foreign atom can be fitted well if its size is relatively small compared to the crystal atom.

(e) *Schottky Defect*: In ionic crystals, a specific type of lattice vacancy is often called as Schottky defect. This point imperfection in ionic crystals occurs when a negative ion vacancy is associated with a positive ion vacancy (Fig. 1.33e). It is thus a localised vacancy pair of positive and negative ions. A coupled pair of vacancy of opposite sign is also possible in Schottky defect, which maintains the ionic crystal electrically neutral.

The existence of point defect in a crystal of this type makes possible the diffusion of atoms or ions within it easily. When vacancies are present, diffusion occurs by jumping of an adjacent atom into the vacancy to leave a new vacancy behind, into which another atom or ion may jump. It is believed that the coupled pair of positive and negative ions may diffuse faster than isolated vacancies. In closely packed structure (F.C.C. and H.C.P.) vacancies and Schottky defects occur more. For ionic crystal numbers of ion pair production is

$$n = Ne^{-E_f/2KT} \quad (1.11)$$

where N = no. of lattice sites

K = Boltzman constant

E_f = energy to create a pair of ion vacancy inside crystal lattice

T = temperature in °K

(f) *Frenkel defect*: In ionic crystal, a specific type of interstitial is called Frenkel defect. In interstitial defect, an atom is transferred from a lattice site to an interstitial position. Here for ionic crystal when a negative ion vacancy is associated with an interstitial (-ve) ion or a positive ion vacancy is associated

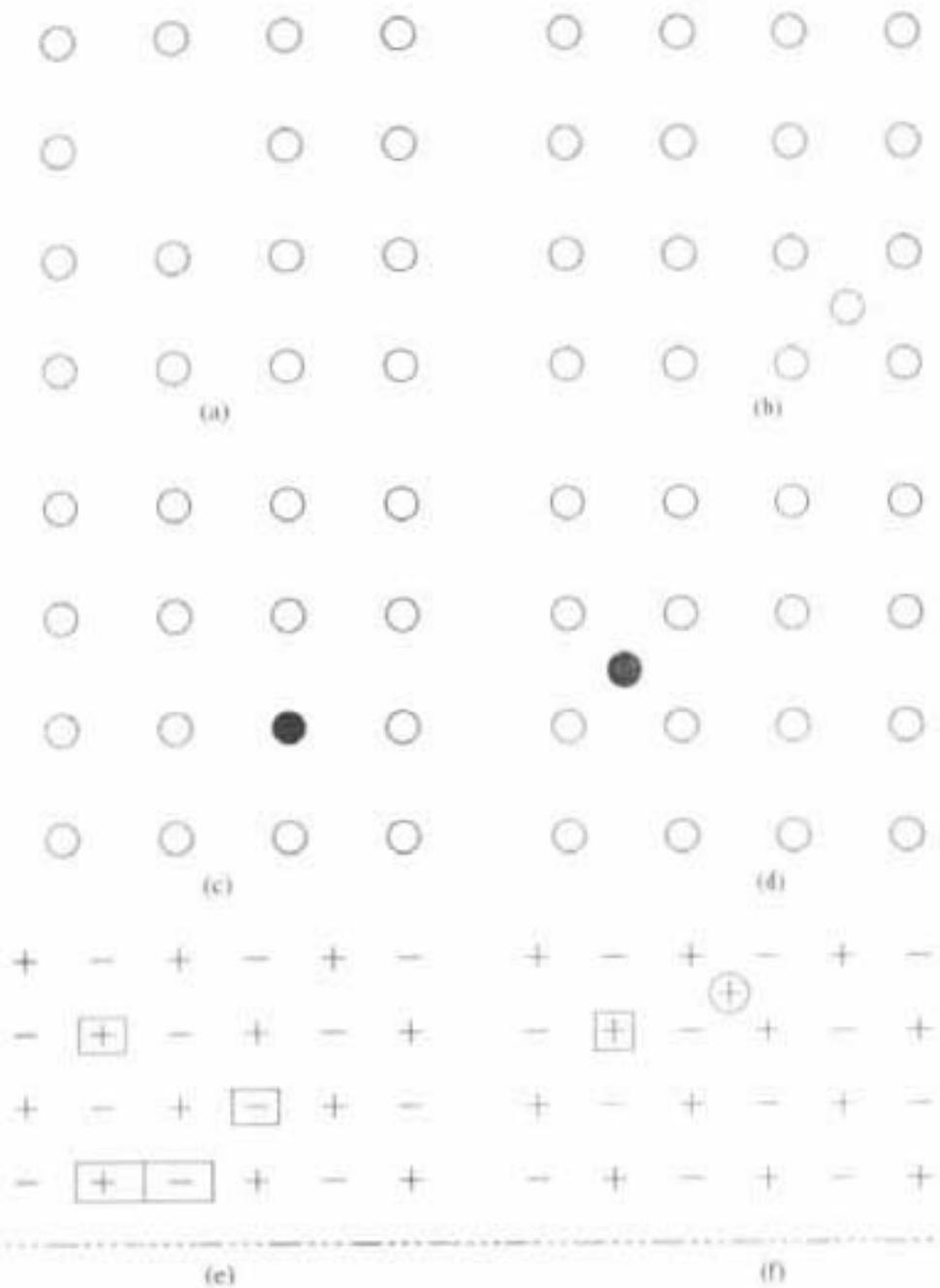


Fig. 1.33 Point defects: (a) vacancy, (b) interstitial, (c) substitutional impurities, (d) interstitial impurities, (e) Schottky defect, (f) Frenkel defect.

with an interstitial (+ve) ion then it is called Frenkel defect (Fig. 1.33f). It maintains the ionic crystal electrically neutral. Such imperfection is created when a positive or negative ion is moved from its normal lattice site to a nearby interstitial positioin. When such type of interstitials are present, diffusion occurs as they migrate through the crystal. As we would expect that diffusion in a solid is strongly temperature dependant, it increases with increase in temperature. In relatively open structure (e.g., B.C.C.) interstitial and Frenkel defects generally occur. Number of Frenkel defects creation is

$$n = (NN')^{\frac{1}{2}} e^{-E/2kT} \quad (1.12)$$

where N = Number of lattice sites,
 N' = Number of interstitial sites

E = energy required to remove an atom from lattice site to an interstitial position

Since the size of (-ve) ion is large compared to (+ve) ion, the (-ve) ion interstitial is very rare.

1.22.2 Dislocation – Line Defect

It is a line defect in which a line of atoms is not in proper position. Dislocations are linear array of atoms where coordination characteristics differ systematically from the ideal one. *These dislocations are perfect within themselves but mismatched with each other.*

1. *Edge Dislocation:* It can be visualised in terms of removal of one layer of atoms and the subsequent accumulation of the array to the defect (Fig. 1.34a). Dislocation is itself indicated by the symbol \perp .

Around the region of the dislocation, the bonding of the atoms are not in normal condition, they are under strain, which is clearly found from Fig. 1.34(a). So under the application of modest pair of forces the crystal can be permanently deformed (Figs. 1.34b and c). Since the dislocation is occurring along the edge,

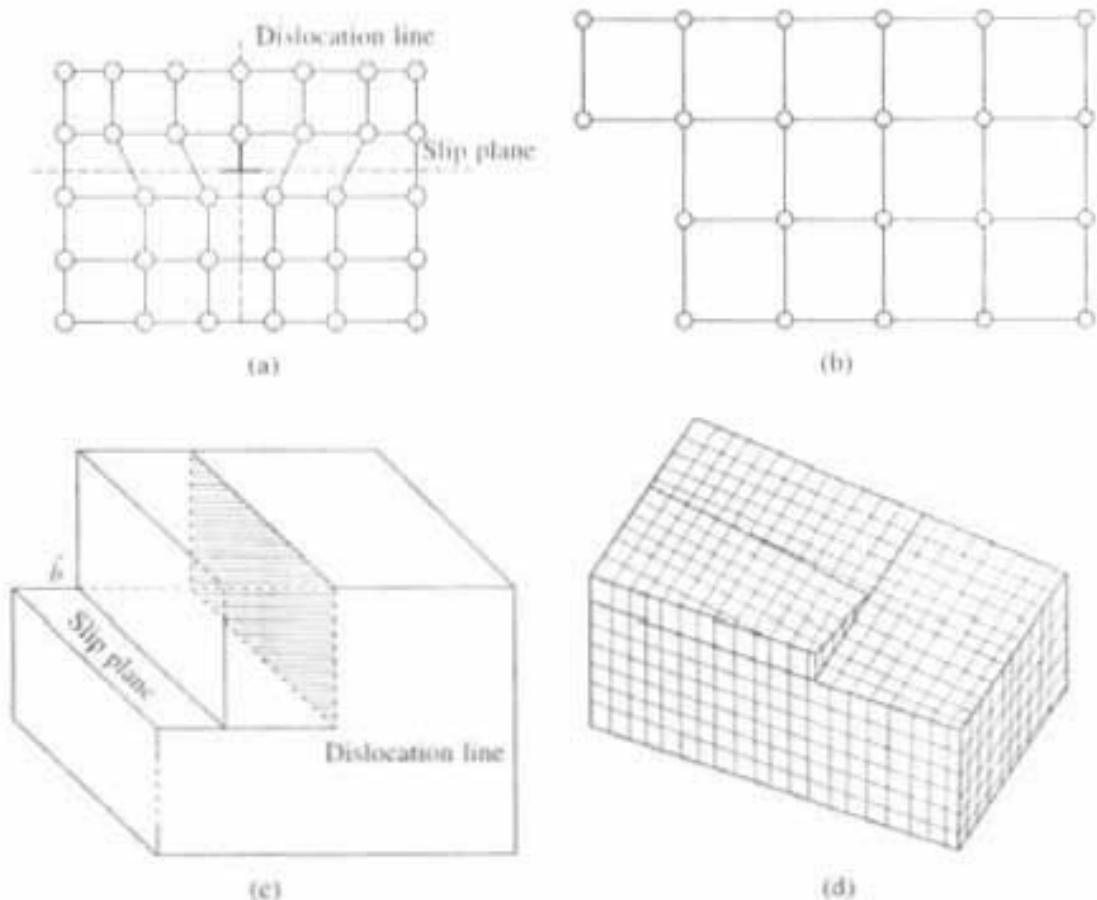


Fig. 1.34 Line defects: (a), (b) and (c) edge dislocation, (d) screw dislocation

therefore, it is called edge dislocation. The entire process is called the slip. The plane along which the slip takes place is called the *slip plane* (Fig. 1.34a, 1.34c). The atoms above the slip plane are in a state of compression and those below the slip are in tension. If the extra half plane is above the slip plane then it is called positive dislocation (+). If the extra plane is below the slip plane then it is called Negative dislocation (-). Figures. 1.34 (a), (b), (c) show positive dislocation. The size of the dislocation is described by the *Burger's vector* \bar{b} (Fig. 1.34c). It measures the relative displacements along slip plane, a part of lattice containing dislocation, over the other parts. In *slip* the atomic bonds holding one layer to the next are broken only one at a time, whereas in a perfect crystal all the bonds between two layers would have to be broken simultaneously for plastic flow to occur, which needs much more force. Thus the presence of dislocation makes it possible to understand why solids are less strong, as they ought to be on the basis of perfect crystal structure.

2. Screw Dislocation: Another kind of dislocation is screw dislocation that can be visualised by imagining that a cut is made, part-way into a perfect crystal and one side of the cut is then displaced relative to the other (Fig. 1.34d) atomic layers spiral around the dislocation which accounts for the name.

3. Mixed Dislocation: In actual crystals dislocations are usually the combination of edge and screw dislocations, which is called mixed dislocation.

1.22.3 Surface Defects

1. These are due to the mismatch of atomic layers for the absence of any layer on the surface. The mismatch may be across the entire crystal or part of it, or between two adjacent crystallites within a poly-crystalline solid.

2. Stacking fault: This results when one atomic plane is piled out of sequence on another, when lattice on either side of the fault is perfect.

3. Grain boundary: Commercial metals are usually poly-crystalline. Each contains large number of crystallites each having the same lattice pattern but differently oriented. The interface between the region of crystalline structure in a poly-crystalline solid is called grain boundary. Along the grain boundary mismatch can occur which gives rise to imperfection (Fig. 1.2).

1.22.4 Volume Defects

These occur due to addition of vacancies in a small region which can either produce voids in a crystal or the presence of foreign atoms of large dimension compared to the actual crystal atoms. Voids produce decreasing density while foreign atoms increase or decrease the density. Both create imperfection in the crystal.

1.23 Diamond Structure

Many elements and compounds can adopt more than one crystal structure because the particular structure determined by the energy band model (Fig. 6.1). So it is possible that the same element or compound behave like a metal when it crystallizes in one form and like a non-metal when it adopts another, for example white tin and grey tin. Similarly one form may be an insulator and the other a semiconductor, like diamond and graphite but fundamentally both are carbon.

The elements carbon, silicon, germanium and tin (gray) all crystallize with diamond structure, which is face-centred cubic structure and the atomic basis is two atoms per one lattice point. Each lattice point corresponds to two identical atoms; one located at $(0, 0, 0)$ and other at $(1/4, 1/4, 1/4)$ (Fig. 1.35(a)), at a point one quarter of the way along the body diagonal. So in a diamond structure eight

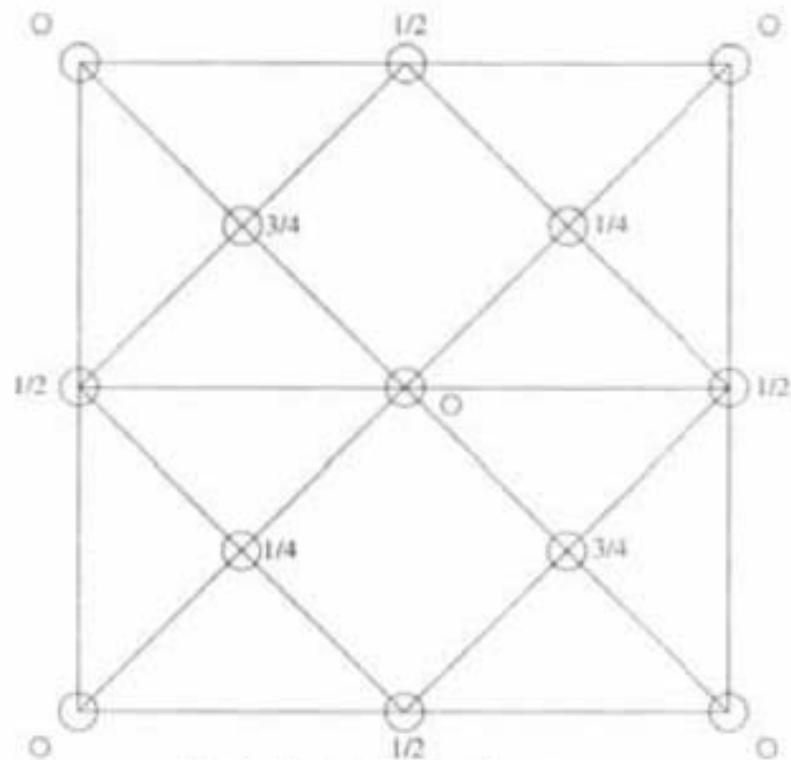


Fig. 1.35 (a) Diamond structure

O → Corner atom

$1/2$ → Face centered atom

$1/4$ → Atom at $(1/4, 1/4, 1/4)$ Position w.r.t. corner atom (O)

$3/4$ → Atom at $(1/4, 1/4, 1/4)$ Position w.r.t. face centered atom ($1/2$)

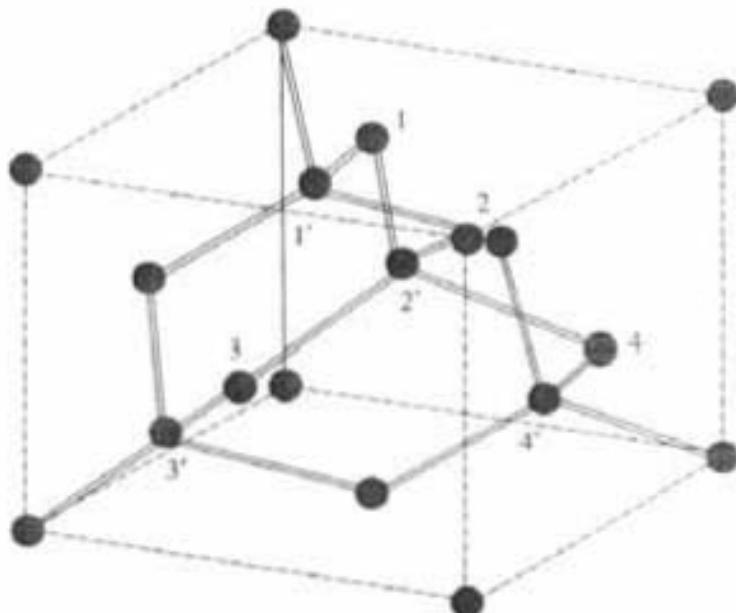


Fig. 1.35 (b) Diamond structure

atoms at eight corners, six face centred atoms at the centre of each of the six faces and four atoms at $(1/4, 1/4, 1/4)$ are positioned on the way along the body diagonal inside the unit cell, in such a way that each of them is linked tetrahedrally with one corner and three face-centred atoms (Fig. 1.35b).

So in diamond structure

(1) Each atom have four nearest neighbours so the coordination number is four which means relatively open structure.

(2) 12 next nearest neighbours.

(3) There are eight atoms per unit cell in diamond crystal structure, 4 from original F.C.C. structure + 4 extra which are entirely inside the cube at $\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$.

(4) Two atoms per basis i.e. per one lattice point, along the body diagonal situated at quarter distance. As they are the nearest neighbour, so they are in touch with each other (Fig. 1.35 a and b). So if d is the atomic diameter and r the atomic radius, then

$$d = 2r = 1/4 \text{ (body diagonal)} = 1/4 (\sqrt{3}a)$$

Hence, the atomic radius $r = \sqrt{3}a/8$ (1.13)

(5) Atomic packing factor for diamond structure is

$$\text{A.P.F.} = \frac{8 \times 4/3\pi(a\sqrt{3}/8)^3}{a^3} = \sqrt{3}\pi/16 = 0.34 \quad (1.14)$$

All the elements having diamond structure have four valence electron in which two electrons will be in s-state while the other two will be in p-state. The bonding between the atoms is strongly co-valent. So we accept that a certain amount of energy is required in order to make the electrons free when they (e.g., C, Si, Ge etc.) take part in conduction. This is in agreement with the fact, that at very low temperature all these elements having diamond structure are insulator. Bonding is strongest in diamond crystal ($Z = 6$). So it is a very hard insulator, having high melting point. Bonding is weakest in gray tin ($Z = 50$). Because, as atomic number increases, the interatomic distances increases, binding force becomes weaker and elements become more metallic where valence electrons become more and more free. Where as for intermediate value of Z , say for Si ($Z = 14$) and Ge ($Z = 32$) elements become semiconductor. As bonding becomes weaker, then with increase of atomic number:

- (a) Forbidden energy gaps will decrease.
- (b) Lattice parameter a will increase.
- (c) Dielectric constant, melting point, and Debye temperature will decrease.

1.24 Barium Titanate (BaTiO_3)

BaTiO_3 is a ferroelectric crystal as shown in Fig. 1.36, which exhibits an electric dipole moment in the absence of an external electric field. In a ferroelectric crystal the centre of positive charge Ti^{4+} of the crystal does not coincide with the centre of negative charge (O^{2-}), due to which a dipole moment will form. This spontaneous dipole moment can be altered by applying sufficiently strong electric field.

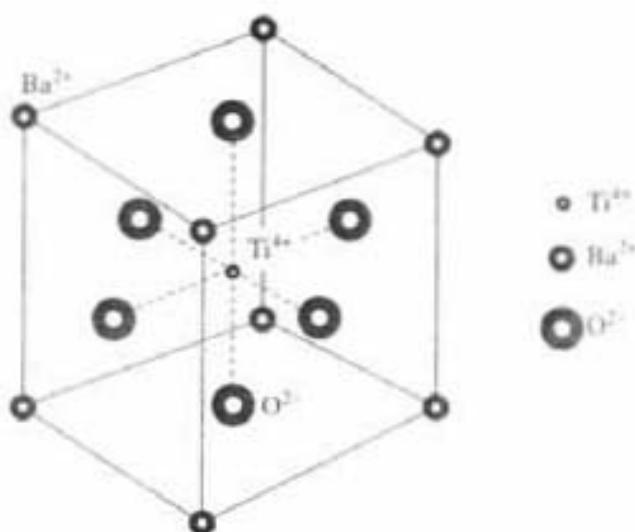
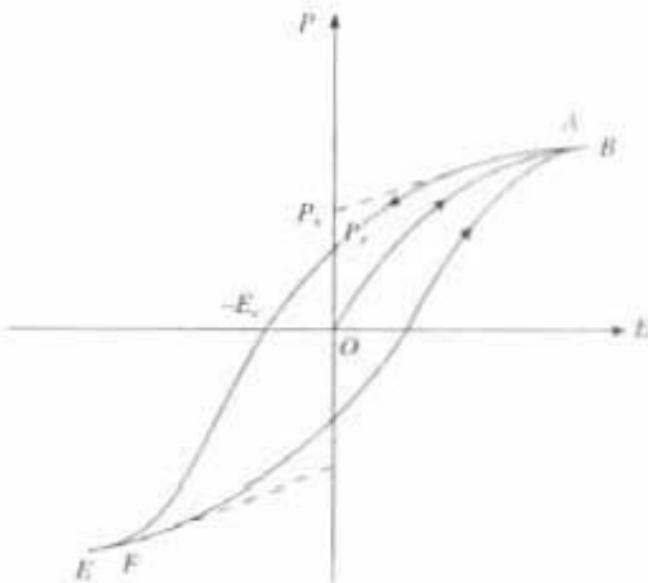


Fig. 1.36 Barium titanate structure

In the non-polarized phase BaTiO_3 has cubic structure, with Ba^{2+} ions at the eight cube corners (Fig. 1.36), O^{2-} ions at the centre of the six faces and Ti^{4+} at the body centre. But below "Curie temperature" ($120^\circ\text{C} = 393^\circ\text{K}$) the structure is not cubic but slightly deformed, when Ba^{2+} and Ti^{4+} ions are displaced relative to O^{2-} ions and thereby developing dipole moment. Crystal structure for BaTiO_3 and variation of direction of polarization with temperature are shown in tabular form in Table 1. Above Curie temperature, BaTiO_3 is non-ferroelectric and has cubic structure.

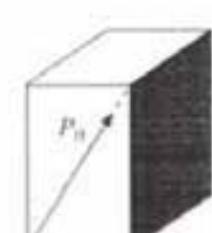
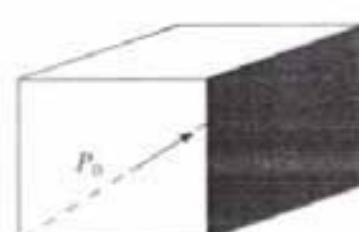
As a ferroelectric crystal, below curie temperature, BaTiO_3 shows the hysteresis loop (Fig. 1.37) which is the sign of a ferroelectric material.

Fig. 1.37 Hysteresis curve for BaTiO_3 below curie temperature

1.25 Interstitial Voids

Any three-dimensional arrangement of spheres of atoms (or ions) give rise to void space between the spheres. These voids are called "interstitial voids". Two main types of interstitial voids in close-packed structures are possible.

Table 1 Change of direction of polarisation with temperature for BaTiO₃

Temperature Region	Direction of Polarisation	Crystal Structure
1. $T > 393^\circ\text{K}$	Non ferroelectric	1. Cubic $a = b = c$ $\alpha = \beta = \gamma = 90^\circ$
2. $T = 393^\circ\text{K}$	 [0 0 1]	2. Tetragonal $a = b \neq c$ $\alpha = \beta = \gamma = 90^\circ$
3. $T = 278-193^\circ\text{K}$	 [1 0 1]	3. Orthorhombic $a \neq b \neq c$ $\alpha = \beta = \gamma = 90^\circ$
4. $T < 193^\circ\text{K}$	 [1 1 1]	4. Rhombohedral $a = b = c$ $\alpha = \beta = \gamma \neq 90^\circ$

1. *Tetrahedral Void:* It is formed between three spheres on a close-packed plane and a fourth sphere on an adjacement plane (either on top or at the bottom), in the cavity space between the three spheres (Fig. 1.38). The name tetrahedral void is derived from the regular tetrahedron with 4 corners and 4 triangular faces of equal size. It is obtained by joining the centres of the four spheres. For every sphere in three-dimensional array, there are two tetrahedral voids. It can be shown from Equ. (1.17) that the largest sphere that can fit in a tetrahedral void $0.225r$, where r is the radius of the spheres of the close-packed array. (Figs. 1.43(a), 1.43(b)).

2. *Octahedral Void:* It is formed with three spheres on a close-packed plane and three more spheres on an adjacement close-packed plane, so that the centres of the three spheres on one plane are directly over the three triangular valleys surrounding the central valley of the first plane and there is no sphere over the central valley (Fig. 1.39). The name octahedral void comes from the regular octahedron (a polyhedron with six corners and eight triangular faces of equal size), obtained by joining the centres of the six spheres. There is one octahedral void per sphere in the three-dimensional array. It can be shown from Equ. (1.16) that the largest sphere that can fit an octahedral void is $0.414r$, where r is the radius of the spheres of the close-packed array (Fig. 1.42a, 1.42b).

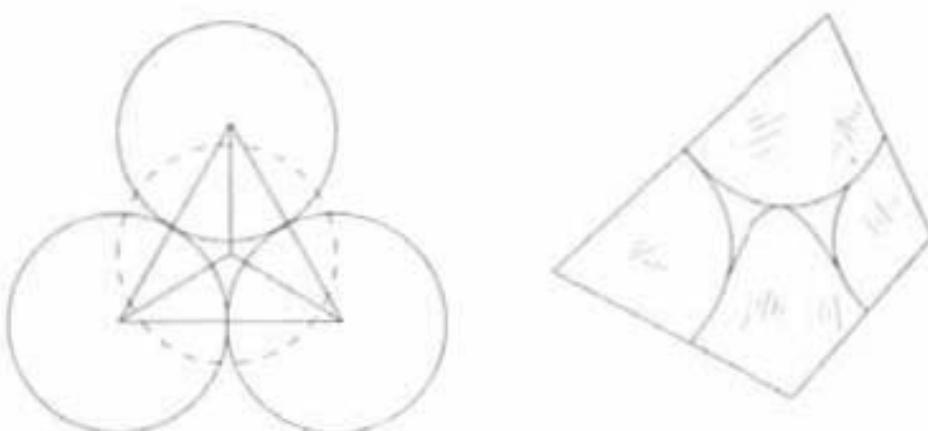


Fig. 1.38 Tetrahedral void

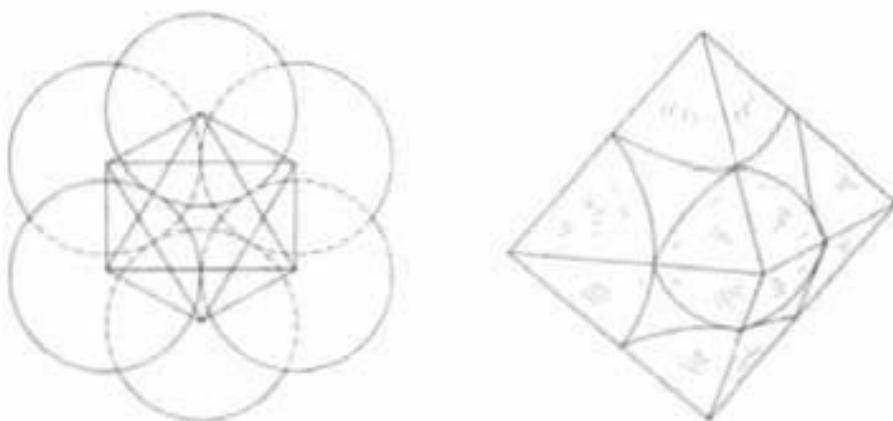


Fig. 1.39 Octahedral void

1.26 Ligancy and Ionic Crystal Structure

In general, the structure of an ionic crystal is determined by the requirement that each ion be in contact with as many ions of the opposite sign as possible, so as to reduce potential energy and to attain maximum ionic bonding energy, which provides the greatest stability. In ionic crystal (+ve) ions are smaller in size than the (-ve) ions because the size of an ion increases on adding extra nuclear electron and shrinks on removing the electron. So unlike the metallic elemental crystal structure, the difference in size of the two types of ions, which form the crystal lattice in ionic crystal, should be taken into account in the geometry of their close-packing.

In a particular ionic crystal structure, the number of (+ve) ions surrounding a central (-ve) ion and vice-versa is called "ligancy". In an ionic crystal ligancy is the number of nearest neighbour of opposite type (i.e. around a Na^+ , how many Cl^- are there). Whereas coordination number is the number of nearest neighbour of the same type (i.e. around a Na^+ , how many Na^+ are there or around Al , how many Al are there). *So for ionic crystal both ligancy and coordination number are possible but for elemental crystal only coordination number is possible.* Thus for stable configuration in ionic crystal, the optimum arrangement in any crystal should be the one, allowing the greatest number of oppositely charged

ions to touch (which give strong bonding), without requiring any squeezing together of ions of same charges (which increase repulsive force).

This can be explained clearly for a binary crystal (AB) of 3-fold ligancy as follows. In Fig. 1.40(a), three B^- ions are in contact with A^+ ion but not with each other. So this is a stable configuration according to the above conditions. In Fig. 1.40(b) size A^+ ion is smaller than in Fig. 1.40(a) and in all B^- ions are just in touch with each other and with A^+ . This configuration is just at the limit of stability or the critical condition for Ligancy-3 and the radius of B^- and A^+ ions in this configuration is called the critical ionic radius (r_c^- and r_c^+). If the size of A^+ ion is still become smaller, Fig. 1.40(c) then none of the B^- ions touch A^+ ion and B^- ions are squeezed together. So the repulsive force is more than attractive force. So this configuration Fig. 1.40(c) becomes unstable and unfavourable for ligancy-3 and more favourable for lower ligancy configuration, say ligancy-2. Thus the relative sizes of the (+ve) and (-ve) ions involved in an ionic crystal, govern the type of the structure that occurs. *So the critical ionic radius ratio for A^+ ion = r_c^+ and B^- ion = r_c^- (Fig. 1.40b), i.e., the ratio r_c^+ / r_c^- , between the radii of the ions in an ionic crystal, can give the answer to why a particular compound crystallizes in a particular form, with particular ligancy number, because ionic radius is an inherent property of the material.*

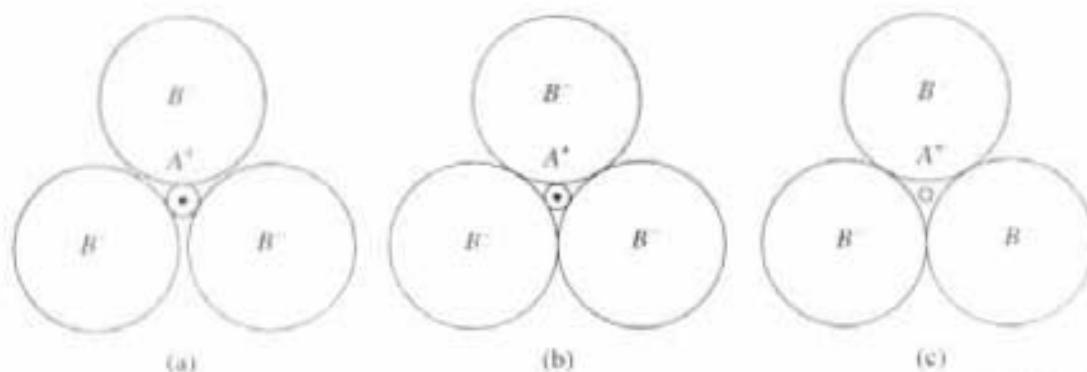


Fig. 1.40 Critical condition for Ligancy-3 and change of ligancy number with ionic radius

1.26.1 CsCl-Structure (Ligancy-8) with Cubical configuration

Let us analyze the situation for CsCl structure of B.C.C. type of ligancy-8 (Fig. 1.41a), with cubic structure configuration, to calculate the critical ionic radius ratio for 8-fold ligancy. Here 8 Cl^- ions of radius r_c^- are placed at 8 corners of a cube around a Cs^+ ion with r_c^+ radius at the centre of the cube, such that (+ve) and (-ve) ions all are in touch in critical condition. So the A^+ to B^- distance (Fig. 1.41b) is $(r_c^+ + r_c^-)$ and the adjacent B^- ions are also just in touch with each other (Fig. 1.41b) at the critical condition. Now B^- to B^- distance is the lattice parameter a , which is given according to geometric construction for B.C.C. structure as from Fig. (1.41b). In B.C.C. structure (-ve) ions at 8 corners will be in touch with (+ve) ion at centre, i.e., all (+ve) and (-ve) ions will be in touch along body diagonal (Fig. 1.41b).

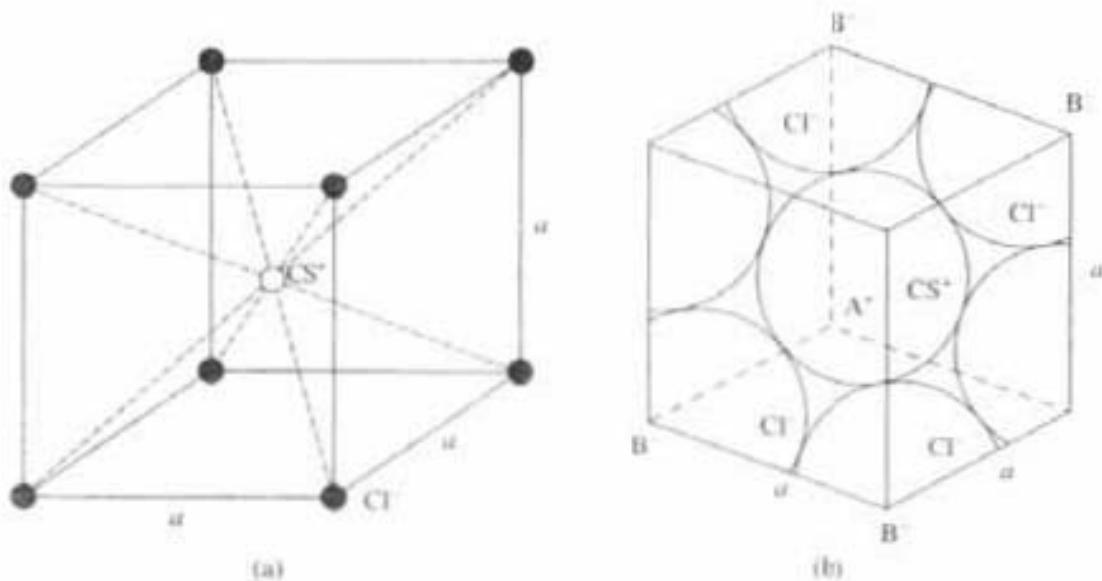


Fig. 1.41 Ligancy-8 configuration

Therefore, the body diagonal (Fig. 1.41b) is

$$\sqrt{3}a = 2(r_c^+ + r_c^-)$$

and lattice parameter is

$$a = r_c^+ + r_c^-$$

i.e.,

$$a = \frac{2}{\sqrt{3}}(r_c^+ + r_c^-) = 2r_c^-$$

which gives the critical ionic radius ratio

$$r_c^+/r_c^- = \sqrt{3} - 1 = 0.729 \quad \text{for Ligancy-8.} \quad (1.15)$$

or

$$r_c^-/r_c^+ = 1.37$$

Thus, in the critical condition, the largest cation of radius $r_c^+ = 0.729 r_c^-$ (where r_c^- is a radius of the anion) can fit exactly into a cubical void in a cubical configuration, where ligancy is 8.

Now if $r_c^+/r_c^- > 0.73$ i.e., r^+ is slightly bigger, as in the case of Fig. 1.40(a) which is also stable. But if the ratio r_c^+/r_c^- is less than 0.73 i.e., r^+ becomes still smaller as in case of (Fig. 1.40c), then that configuration is not stable and unfavourable for ligancy-8 CS.C.I structure, but more favourable for lower ligancy configuration, say ligancy-6.

1.26.2 NaCl-Structure (Ligancy-6) with Octahedral Configuration

NaCl structure having ligancy-6 (Fig. 1.42a), F.C.C.-type structure, with octahedral configurations. Here in critical condition all positive and negative ions are just touching each other but not squeezing, with radius r_c^- for B^- ion and r_c^+ for A^+ ion. 6 Cl^- ions are situated at 6 corner and one Na^+ ion is at the centre of the octahedron Fig. (1.42a). The critical ionic radius ratio will come in the following way from Fig. 1.42(b), which is the square planer cross-section of an octahedron

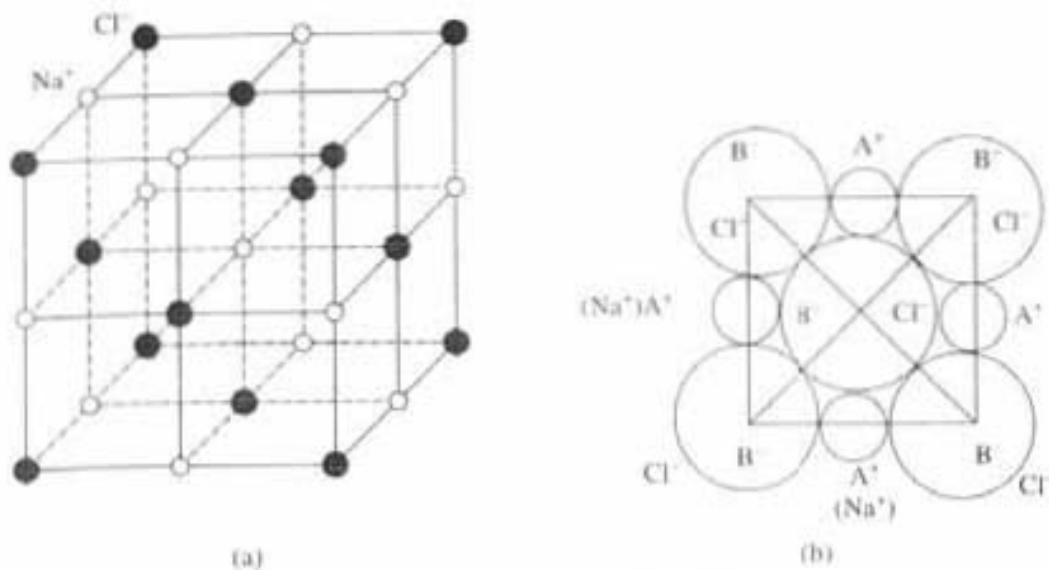


Fig. 1.42 Ligancy-6 configuration

(Fig. 1.39). For F.C.C. structure with octahedral configuration (+ve) and (-ve) ions will be in touch along lattice parameter a and (-ve) ions will be in touch along surface diagonal in critical condition. Hence, from Fig. 1.42(b)

$$\sqrt{2}a = 4r_c^- \text{ and } a = 2(r_c^+ + r_c^-)$$

$$\text{i.e., } 2\sqrt{2}(r_c^+ + r_c^-) = 4r_c^-, 1 + r_c^+/r_c^- = \sqrt{2}$$

$$\text{So we have, } r_c^+/r_c^- = 0.414 \quad \text{for Ligancy-6} \quad (1.16)$$

whereas

$$r_c^-/r_c^+ = 2.44.$$

So the largest cation of radius $r_c^+ = 0.414 r_c^-$, can fit exactly in a octahedral void of octahedral configuration whose ligancy is 6, as discussed in section 1.25.

Thus, the binary ionic compounds (AB) with critical ionic radius ratio r_c^+/r_c^- between 0.73 and 0.414, crystallize into NaCl structure with ligancy of 6. If r_c^+/r_c^- is less than 0.414, then lower ligancy, i.e. 4 is favourable and not 6.

1.26.3 ZnS Structure (Ligancy-4) with Tetrahedral Configuration

If the ratio r_c^+/r_c^- is less than 0.414 for some crystals, then ligancy-6 will become unfavourable and ZnS structure, ligancy-4, the diamond structure with alternate Zinc and Sulphur ion will occur, (Fig. 1.43a). In ZnS, which is having diamond-type structure with regular tetrahedral configuration, the (+ve) Zinc ion will be at the $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$, position which is indicated as (O') in Fig. 1.43(a), at the centre of the tetrahedral void (Fig. 1.38), whereas its four nearest neighbours, four (-ve) sulphur ions indicated as (A⁻, B⁻, C⁻, D⁻) (Fig. 1.43a) will be at each corner of the tetrahedron (Fig. 1.38) and in critical condition all (-ve) sulphur ion will be in touch with each other as well as with the central zinc (+ve) ion. (Fig. 1.43a, 1.43b).

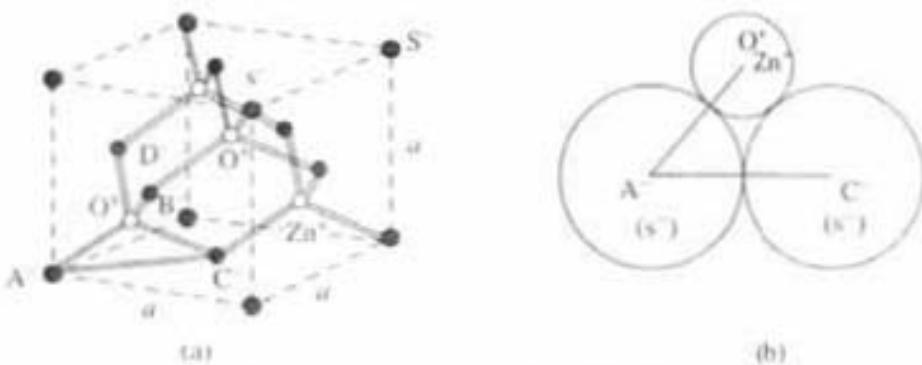


Fig. 1.43 Ligancy-4 configuration

In diamond structure the (-ve) ion A^- i.e. (s^-) at $(0, 0, 0)$ and (+ve) ion O^+ i.e. (zn^+) at $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ position are separated by a distance, one fourth of the body diagonal. So in critical conditions (from Figs. 1.43a and b) we get

$$AO = r_c^+ + r_c^- = \frac{\sqrt{3}a}{4}$$

whereas in critical condition distance between two sulphur (-ve) ions (A^- and C^-) is

$$AC = r_c^- + r_c^- = \frac{\sqrt{2}a}{2} = \frac{a}{\sqrt{2}}$$

or

$$r_c^- = \frac{a}{2\sqrt{2}}$$

where a is the lattice parameter.

Now by taking the ratio of the above two equations we get

$$1 + \frac{r_c^+}{r_c^-} = \frac{\sqrt{3}a}{4} \times \frac{2\sqrt{2}}{a} = \frac{\sqrt{3}}{\sqrt{2}} = 1.225$$

So,

$$\boxed{\frac{r_c^+}{r_c^-} = 0.225} \quad \text{for Ligancy-4} \quad (1.17)$$

Thus for ligancy-4 with tetrahedral configuration, critical ionic radius ratio (r_c^+/r_c^-) = 0.225 or $r_c^-/r_c^+ = 4.44$, may become favourable. So in critical condition the largest cation of radius $r_c^+ = 0.225 r_c^-$ can fit exactly in a tetrahedral void whose ligancy is 4. If A^+ is still smaller as in Fig. 1.40(c), then r_c^+/r_c^- will be less than 0.225, then ligancy-4 will not be possible; lower ligancy-3 will be allowed.

1.26.4 Ligancy-3 with Triangular Configuration

Let us consider Fig. 1.44 which shows critical ionic arrangement for ligancy-3. X, Y, Z are the centres of three (-ve) B ions and O is the centre of (+ve) A ion in critical condition. From ΔXYZ it is clear that

$$XO = r_c^- + r_c^+ \text{ and } XN = r_c^-$$

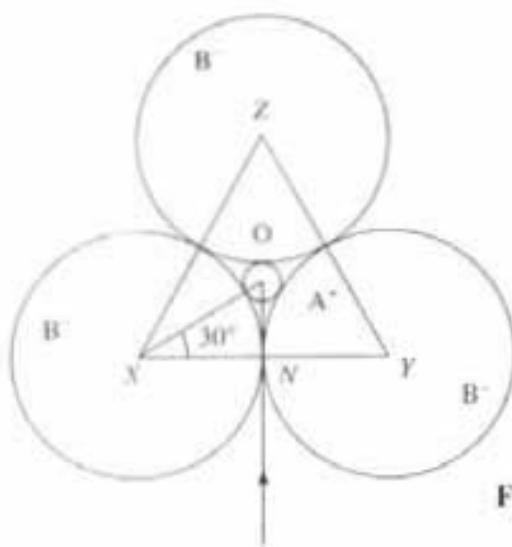


Fig. 1.44 Ligancy-3 configuration

Now $\angle OXY = 30^\circ$ and in ΔOXN , $\cos 30^\circ = \frac{XN}{OX} = \frac{r_c^+}{r_c^+ + r_c^-}$

$$\text{i.e. } 1 + r_c^+ / r_c^- = \frac{1}{\cos 30^\circ} = 1.155$$

$$\text{Or, } r_c^+ / r_c^- = 0.155 \quad \text{or } r_c^- / r_c^+ = 6.45 \text{ for Ligancy-3}$$

Thus, in the critical condition the largest cation of radius $r_c^+ = 0.155 r_c^-$ can fit exactly in a triangular void whose ligancy is 3. If radius of A^+ is still smaller, as in Fig. 1.40(c), then r_c^+ / r_c^- will be less than 0.155, and ligancy-3 will not be possible, lower ligancy-2 will be allowed.

Thus, in this simple approximation of packing consideration (the arrangement of positive and negative ion positions) leads us to accept the various stable structures in terms of critical ionic radius ratio r_c^+ / r_c^- . So we can generalize it as given in Table 2 in the following way.

Table 2

Crystal Structure and Configuration	Ligancy	Critical Radius Ratio
1. FCC or HCP	12	1.0
2. CsCl and CaF ₂ (Cubical structure (B.C.C. type))	8	$1 > r_c^+ / r_c^- > 0.73$
3. NaCl Structure (Octahedral (F.C.C. type))	6	$0.73 > r_c^+ / r_c^- > 0.414$
4. ZnS structure (Tetrahedral (Diamond structure by replacing alternate carbon atoms with zinc and sulphur))	4	$0.414 > r_c^+ / r_c^- > 0.225$
5. Triangular	3	$0.225 > r_c^+ / r_c^- > 0.155$
6. Linear	2	$0.155 > r_c^+ / r_c^- > 0$

Hence, as the critical radius r_c^*/r_i ratio decreases, ligancy decreases. Though there are a number of exceptions but in general, the correlation between the predictions of the above data and observed crystal structures are good.

PROBLEMS

1. Consider a Cu metal having F.C.C. structure. If the inter-planer distance d is 2.08 Å for the set of parallel planes (111), find density of Cu-metal, given that at. wt of Cu = 63.54 and Avogadro's No. = 6.023×10^{23} /g mole

Solution: Density = D

$$D = \frac{nM}{Na^3}$$

where

n = No. of atoms or mol. per unit cell = 4 (for F.C.C.)

M = Molecular wt. = 63.54

N = Av. No. = 6.023×10^{23} mols per gm mol of substance

$$d_{111} = \frac{a}{\sqrt{1^2 + 1^2 + 1^2}} = \frac{a}{\sqrt{3}} = 2.08 \text{ \AA}$$

$$a = \sqrt{3} \times 2.08 \text{ a.u.}$$

$$\text{So } D = \frac{4 \times 63.54}{6.023 \times 10^{23}} \times \frac{1}{(\sqrt{3} \times 2.08)^3 \times (10^{-8})^3} = 9.04 \text{ g/cc}$$

2. Find Miller indices of a set of parallel planes which make the intercepts in the ratio $3a : 4b$, on X and Y axes and are parallel to Z axis. a, b, c , are basic vectors.

Solution: Linear parameter of the intercepts of the plane are in the ratio $3a : 4b : \infty$.

So, the ratio of numerical intercepts are $(3 : 4 : \infty)$

Therefore, the Miller indices are $1/3 : 1/4 : 1/\infty = (430)$

3. In a crystal whose primitives are 1.2 au, 1.8 au and 2 AU, a plane (231) cuts an intercept 1.2 Å along X -axis. Find the length of the intercepts along Y and Z -axes.

Solution: Miller plane (231), where $\bar{a} = 1.2 \text{ \AA}$, $\bar{b} = 1.8 \text{ \AA}$, $c = 2 \text{ \AA}$

So intercepts are $a/2$, $b/3$, $c/1$,

i.e., ABC plane (231) (Fig. 1.45)

But intercepts for ABC plane along X axis is 0.6 Å, not 1.2 Å.

So it is not the required plane.

Intercepts for DEF plane parallel to ABC plane are a , $2b/3$, $2c$

Hence, Miller indices for DEF plane are the reciprocals of the intercepts, i.e. $1 : 3/2 : 1/2 = (231)$

So DEF is also (231) plane (Fig. 1.45).

The intercepts of the DEF plane

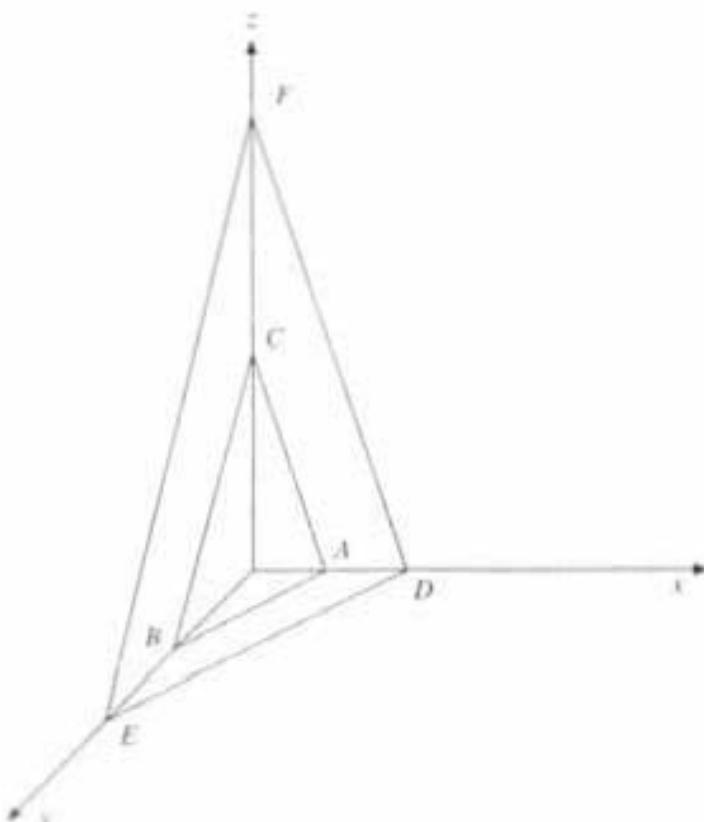


Fig. 1.45

along $X = a = 1.2 \text{ au}$

along $Y = 2b/3 = 2/3 \times 1.8 = 1.2 \text{ au}$

along $Z = 2c = 2 \times 2 = 4 \text{ au}$

Thus, DEF is the required plane, whose intercepts along y and z axes are 1.2 \AA and 4 \AA , respectively.

QUESTIONS

- What is lattice plane? How are the lattice planes described in terms of Miller indices? Show the planes (211) and (111) with a neat diagram in S.C. unit cell.
- Clarify the types of point defects and explain how they are produced?
- Define the terms lattice parameter and unit cell. Deduce the relation for atoms per unit cells and atomic packing factor of cubic system and explain its importance.
- Find atomic radius and atomic packing factor for F.C.C.. Given the lattice constant 3.61 AU.
- What is interplaner spacing? Deduce the relation for interplaner spacing in terms of Lattice constant and Miller indices.
- What is Crystal imperfection? Explain the types of point defects and how are they produced?
- For diatomic crystal Si, calculate the atomic packing factor. Lattice constant of Si is $5.43 \times 10^{-8} \text{ cm}$.

Solution: Si, the diatomic crystal is having diamond structure. So, number of atoms/unit cell = 8

$$\text{Radius of Si atom} = r = \frac{\sqrt{3}a}{8} = \frac{\sqrt{3} \times 5.43 \times 10^{-8}}{8}$$

$$\text{Thus, A.P.F. for Si} = \frac{8 \times \frac{4}{3}\pi r^3}{a^3} = \frac{8 \times \frac{4}{3}\pi \left(\frac{\sqrt{3}}{8}a\right)^3}{a^3} = 0.34$$

8. For S.C., F.C.C. and B.C.C. determine the following;
 - (a) Lattice points per unit cell.
 - (b) nearest neighbour distance.
 - (c) Atomic packing factor
 - (d) Co-ordination Number
9. Copper metal is having F.C.C. structure. If the interplaner distance d is equal to 2.08 Å.u. for the set of parallel lines (LLL) find the density of copper metal given At. wt. = 63.54, Avogadro's number = 6.023×10^{23} per gni-mole.
10. Explain the term interplaner Spacing in a crystal and show that the interplaner spacing of corresponding planes of S.C. of (100), (110), (111) are in the ratio 1 : 0.71 : 0.53.
11. What is crystal imperfection. Explain with diagram. (1) Schottkey defect and (2) Frenkel defect as point defects.
12. With a neat diagram show edge dislocation. Calculate atomic radius of B.C.C., F.C.C., in terms of Lattice constant a .
13. Ni is having F.C.C. structure. Calculate the atomic radius, A.P.F. and mass density. Lattice constant $a = 3.52$ a.u., At. Wt. of Ni = 58.71 and Avo. No. = 6.023×10^{23} .

Solution: Atomic Radius (F.C.C.) = $r = \frac{a\sqrt{2}}{4} = \frac{3.52 \times \sqrt{2}}{4} = 1.244 \text{ \AA}$

$$\text{A.P.F.} = \sqrt{\frac{2}{3}\pi} = 0.74$$

$$\text{Mass density } D = \frac{nM}{N} \cdot \frac{1}{a^3} = \frac{4 \times 58.71}{6.023 \times 10^{23} \times (3.52 \times 10^{-8})^3}$$

$$D = 8939 \text{ kg/m}^3$$

14. Show the planes (101) and $(\bar{1}10)$ in simple cubic crystal with neat diagrams. Calculate interplaner spacing d .
15. In a molybdenum B.C.C. structure, calculate atomic radius r and mass density. Lattice constant $a = 3.15 \text{ \AA}$. At. Wt. of Mo = 95.94 and Avogadro's number $N = 6.023 \times 10^{23}$.
16. What is crystal imperfection? Explain with diagrams various point defects in elemental as well as ionic crystals.
17. Calculate density of GaAs and Si using following data:
At. Wt. of Ga = 69.7, Lattice constant for GaAs = $5.65 \times 10^{-8} \text{ cm}$
At. Wt. of As = 74.9, No. of atoms/unit cell for GaAs = 4
At. Wt. of Si = 28.1, Lattice const. for Si = $5.43 \times 10^{-8} \text{ cm}$, No. of atoms/unit cell for Si = 8
18. What is Ligancy? Show that critical ionic radius ratio r_c^+/r_c^- decreases as Ligancy decreases.
19. Discuss diamond structure and calculate Atomic packing factor for diamond structure.

20. What is the speciality in BaTiO_3 structure? Discuss in detail with diagram. How does its ferroelectric property changes with temperature?
21. Calculate critical radius ratio r_c^*/r_c for Ligancy 6 and 8.
22. Distinguish between amorphous solids, polycrystalline solids and single crystal.
23. With a neat diagram show the following Miller planes and directions.
 Planes : (234) and (111)
 Directions: (011) and (123)
24. Define the terms space lattice, unit cell, lattice axis and atomic basis.
25. A crystal lattice plane (326) makes an intercept of 1.5 Å on X-axis, in a crystal having lattice constants 1.5 Å, 2 Å and 2 Å on X, Y, Z axis, respectively. Find Y and Z intercepts.
26. Show that maximum fraction of the unit cell volume which can be filled by hard sphere in the S.C., B.C.C. and F.C.C. crystal are 0.52, 0.68 and 0.74, respectively.
27. A certain crystal has lattice constant of 4.24 Å, 10 Å and 3.66 Å on X, Y, Z axis respectively. Determine the Miller Indices of two lattice planes of this crystal having inving intercepts of:
 (i) 2.12 Å, 10 Å, 1.83 Å for one plane.
 (ii) 4, 24 Å, ∞, 1.22 Å for the other plane.

Solution: Here lattice constants are $a = 4.24 \text{ \AA}$, $b = 10 \text{ \AA}$, $c = 3.66 \text{ \AA}$

(i) For first plane intercepts are $\frac{a}{2}, b, \frac{c}{2}$

i.e. numerical intercepts are $\frac{1}{2}, 1, \frac{1}{2}$

So Miller indices are (212)

(ii) For 2nd plane intercepts are $a, \infty, c/3$

i.e. numerical intercepts are 1, ∞, 1/3

i.e. so Miller indices are (103)

28. Silver have F.C.C. structure and its atomic radius is 1.441 Å. Find the spacing of the (220) and (111) Planes.

Solution: For F.C.C. atomic radius = $r = \frac{\sqrt{2}a}{4}$

$$\text{so } a = \frac{4r}{\sqrt{2}} = \frac{4 \times 1.441}{\sqrt{2}} = 4.075 \text{ \AA}$$

$$\text{i.e. } d_{220} = \frac{4.075}{\sqrt{2^2 + 2^2 + 0}} = 1.44 \text{ \AA}$$

$$\text{and } d_{111} = \frac{4.075}{\sqrt{3}} = 2.35 \text{ \AA}$$

29. Density of CaF_2 is 3180 kg/m^3 . A unit cell contains four Ca^{++} and eight F^- ions. Atomic wt. of Ca = 40 and F = 19. Calculate the lattice constant of this crystal.
- Solution:* As it was discussed in section 1.9, the number of above mentioned CaF_2 molecules per unit cell is four.

$$\text{so density of the crystal } \text{CaF}_2 = D = \frac{nM}{N_a V}$$

$$\text{i.e. lattice const. } a = \left(\frac{nM}{ND} \right)^{1/3} = \left[\frac{4 \times 59}{6.023 \times 10^{23} \times 3180} \right]^{1/3}$$

$$\text{or } a = 4.976 \times 10^{-10} \text{ m} = 4.976 \text{ \AA}$$

Thermoelectricity

2.1 Seebeck Effect or Thermoelectric Effect

This phenomenon was discovered by Seebeck in 1821 and hence called the Seebeck effect. He found that *if two wires of dissimilar metals are connected, so as to form a closed circuit, then an e.m.f. is developed on it and hence a current flows through it, when two junctions of the two dissimilar metals are maintained at two different temperatures.* The magnitude and direction of this current depends on the nature of the two materials used and upon the difference of temperature between the hot and cold junctions. *This effect is known as thermoelectric effect or Seebeck effect.* A pair of junctions of this kind is called a thermocouple (Fig. 2.1). The e.m.f. generated in a thermocouple is called thermoelectric e.m.f. and current which flows is called thermoelectric current. This thermoelectric e.m.f. will exist and thermoelectric current will flow so-long there is a temperature difference between two junctions. Whenever the temperature of the two junctions becomes equal, e.m.f. will be zero and current stops flowing. With the increase in the temperature difference between the two junctions thermoe.m.f. will increase. Thus, if the temperature of the cold junction is maintained at 0°C , then the e.m.f. developed would be a measure of the temperature of the hot junction.

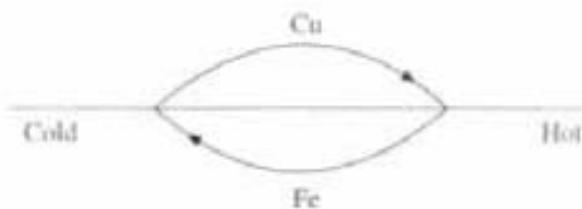


Fig. 2.1 Fe-Cu Thermocouple

2.2 Origin of Seebeck Effect

In atomic terms Seebeck effect can be explained as follows. It is well known that the number of free electrons per unit volume and their average velocity varies from one metal to another, though all are good conductors. Thus at a junction of two dissimilar metals, there may be a tendency of electrons to migrate from one metal to the other across the junction, from higher concentration to lower concentration. But such migration cannot continue for long as it quickly sets up an opposite electric field that prevents any further movement of electrons across

the boundary. As a result a fixed potential arises at the junction between two dissimilar metal, called as *contact potential*. The velocity of an electron in a metal depends to some extent on temperature, as a result the contact potential between any two given metal varies with temperature. In Fig. 2.2 the difference between two contact potentials V and V' when two junctions are kept at two different temperatures, will act to derive current round the circuit. The resultant e.m.f., E will be

$$E = V - V' \quad (2.1)$$

When temperature of two junctions is same, the contact potential is also same across two junctions. As a result there is no resultant e.m.f. in the thermocouple, so no current flows in the circuit. Hence, when one junction of a thermocouple is heated up and other is cooled, then the contact potential at the heated junction becomes more than that of cold junction, and a *potential difference* is developed, which acts as the resultant e.m.f. and is responsible for the flow of current, in the outer circuit.

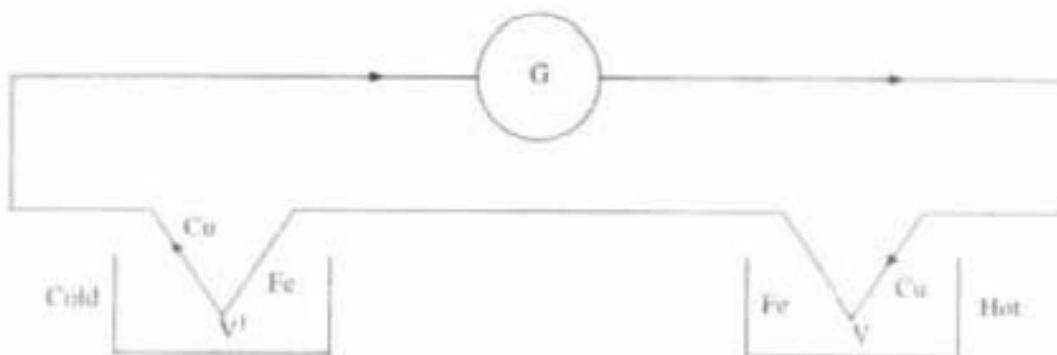


Fig. 2.2 Fe-Cu Thermocouple

2.3 Thermoelectric Series

For a given temperature range, the direction of thermoelectric e.m.f. for different combination of metals may be predicted by arranging the metals in a series. Seebeck arranged the metals in a series in such a way that when any two form a thermocouple, the current will flow at cold junction, from the one occurring earlier to the other occurring later in the series. This is shown in Fig. 2.1. Between the 200° and 100°C thermoelectric series for a selection of metals is antimony, Fe, Cd, Zn, Ag, Au, Sn, Pb, Hg, Mn, Cu, Pt, Co, Ni, bismuth.

The metals to the right side of lead (Pb) are called thermoelectrically positive metals, while those to its left side are called thermoelectrically negative metals.

At the cold junction, the thermoelectric (TE) e.m.f. acts in such a way that it drives current from the metal, earlier in the list towards the metal later in the list. The series, in general, also indicates the magnitude of e.m.f. The further apart two metals are, the greater is the e.m.f. they produce. The largest e.m.f. is obtained with an antimony bismuth thermocouple.

2.4 Variation of Thermoelectric e.m.f. with Temperature

It has been observed that as we increase the temperature difference between two junctions of a thermocouple, thermo e.m.f. will increase (Fig. 2.3) until it reaches

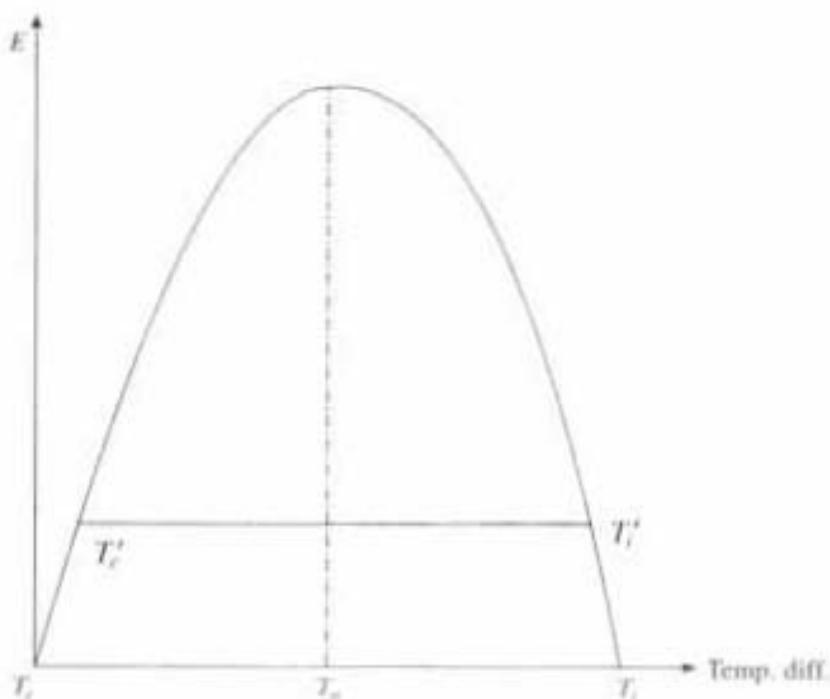


Fig. 2.3 Variation of thermo e.m.f. with temperature

its maximum limit. The temperature of the hot junction T_n , where the thermo e.m.f. becomes maximum is called neutral temperature (T_n) which is constant for a particular pair of metals and different for different pair. After the neutral temperature, the e.m.f. starts decreasing and at a particular temperature of the hot junction, called temperature of inversion (T_i) thermo e.m.f. again becomes zero. After temperature of inversion the direction of thermo e.m.f. will increase in reverse direction, with the increase of temperature further.

Thus it has been seen that when the cold junction is at constant temperature and hot junction temperature is continuously varied then the thermo e.m.f. (E) will vary almost as a parabolic function of the temperature difference between the two junctions, in the following way:

$$E = at + \frac{1}{2}bt^2 \quad (2.2a)$$

This is known as Seebeck equation, when cold junction temperature is $0\text{ }^\circ\text{C}$ and $t\text{ }^\circ\text{C}$ is the temperature of the hot junction. But when cold junction temperature is not zero but at $t_1\text{ }^\circ\text{C}$, then

$$E = a(t_2 - t_1) + \frac{1}{2}b(t_2 - t_1)(t_2 + t_1) \quad (2.2b)$$

where $(t_2 - t_1)$ is the temperature difference between hot junction and cold junction and a, b are known as Seebeck coefficients.

The neutral temperature (T_n) is half-way between the cold junction temperature and the temperature of inversion (T_i). Neutral temperature is constant for a given thermocouple and it is independent of the cold junction temperature (T_c) which is equal to

$$T_n = (T_i + T_c)/2 \quad (2.3)$$

Inversion temperature T_i is a variable, it being as much above the neutral temperature T_n as the cold junction is below it. Thus T_i depends on T_c . As shown in Fig. 2.3 if cold junction temperature changes from T_c to T'_c then inversion temperature T_i will also change from T_i to T'_i . But in that situation also T_n will be same for a particular couple.

2.5 Thermoelectric Power

Total e.m.f. in a thermocouple is a parabolic function of temperature and can be expressed as

$$E = at + \frac{1}{2}bt^2 \quad (2.6)$$

Differentiating equation 2.6, we have thermoelectric power (P) given by

$$P = \frac{dE}{dT} = a + bt \quad (2.7)$$

If T is the temperature of the hot junction and T_0 the temperature of the cold junction which is kept constant, then, $T - T_0 = t$, the temperature difference between two junctions in °C.

From equation 2.7 we conclude that the variation of thermoelectric power with temperature is a straight line (Fig. 2.4) called *thermoelectric diagram*. Such diagrams for different metals shown in Fig. 2.4 clarify that the rate of change of thermo e.m.f. with temperature is positive for thermoelectrically positive metals like Cu and it is negative for thermoelectrically negative metals like Fe.

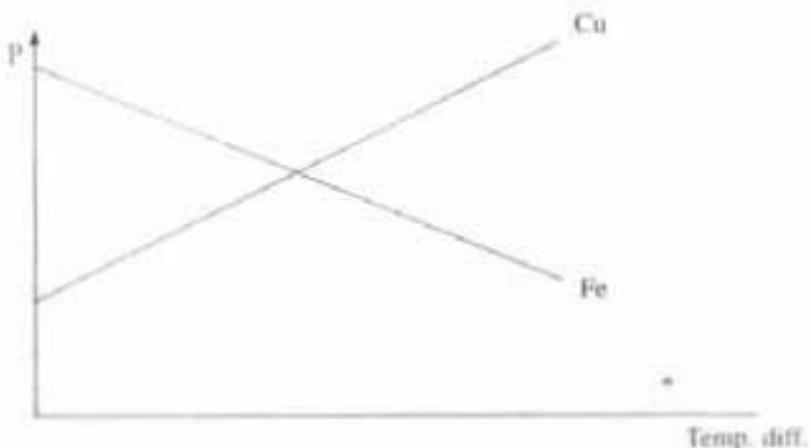


Fig. 2.4 Variation of thermoelectric power with temperature

2.6 Determination of Neutral Temperature (T_n) and Temperature of Inversion (T_i) from Seebeck Coefficient

At neutral temperature (T_n) thermo e.m.f. is maximum, so at $t = T_n$ the rate of change of thermo e.m.f. E with temperature is zero. So,

$$E = at + \frac{1}{2}bt^2$$

at

$$t = T_n$$

$$\frac{dE}{dT} = a + bT_n = 0$$

that is,

$$T_n = -a/b \quad (2.8)$$

At $t = T_i$, thermo e.m.f. E is zero. So,

at

$$t = T_i, E = aT_i + \frac{1}{2}bT_i^2 = 0$$

that is

$$T_i = -2a/b \quad (2.9)$$

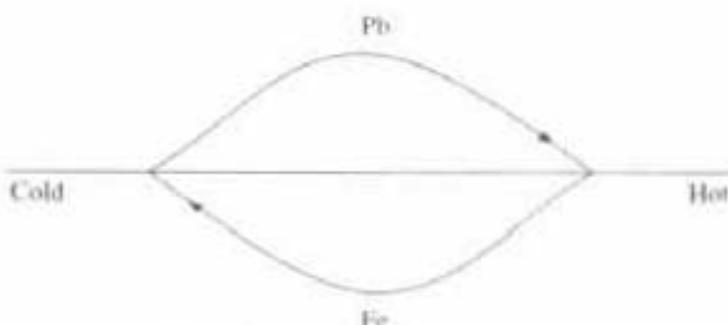


Fig. 2.5 Fe-Pb thermocouple

2.7 Seebeck Coefficients

The values of Seebeck coefficients a and b are generally given with respect to lead (Pb), it can be positive or negative (Fig. 2.5).

For a thermocouple, formed with metals X and Y, Seebeck coefficients can be written as a_{x-y} and b_{x-y} and expressed as

$$a_{x-y} = (a_x - a_y) \quad (2.10)$$

$$a_{(X-Pb)-Y-Pb)} = a_{(X-Pb)} - a_{(Y-Pb)} = (a_x - a_y)$$

Similarly

$$b_{x-y} = (b_x - b_y) \quad (2.11)$$

This Seebeck coefficient's value will be positive when current flows at hot junction from lead to other metals used in thermocouple.

2.8 Law of Intermediate Metals

In practical use the circuit may not contain only the thermocouple metals, it may have additional metals, i.e., Cu or other metals as connecting wires and also in the coil of the galvanometer or voltmeter to complete the circuit. So it is important to know the effect of additional junctions. The effect of additional metals is known from the law of intermediate metals. It states that if *junction of thermocouple of two metals A and B is opened and a third metal C is inserted in between them, the resultant thermo e.m.f. remains the same if only both junctions of metal C are at the same temperature*. This can be explained as follows:

Consider a circuit containing 3 metals A, B, C (Fig 2.6). If in the circuit all junctions are at the same temperature then there will be no possible source of energy that could derive current in the circuit. So algebraic sum of the three

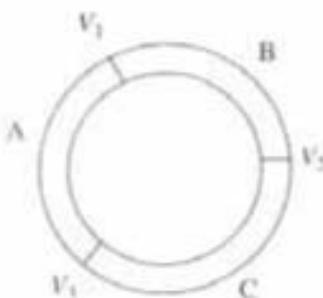


Fig. 2.6 Junction of three metals A, B, C

$$V_1 + V_2 + V_3 = 0 \quad (2.12)$$

Now, if the temperature of junction AB is changed, the contact potential at that point changes to a new value V'_1 but V_2 and V_3 remains same. The thermo e.m.f. E acting in the circuit will then be

$$E = V'_1 + V_2 + V_3 \quad (2.13)$$

But from equation (2.12), $V_2 + V_3 = -V_1$.

$$\text{So, } E = V'_1 - V_1 \quad (2.14)$$

which is the same as we should get, if metal C was eliminated from the circuit by bringing its junctions with A and B into contact. Thus, *breaking the circuit at some point or inserting a third metal, whose temperature is the same as that of the wire at the break, does not alter the total e.m.f. in the circuit.*

2.9 Law of Successive Temperature

For a thermocouple, the e.m.f. generated between temperature difference θ_1 and θ_n is equal to the sum of the e.m.f. for any number of successive steps into which the entire range of temperature has been divided, but that intermediate temperature should be fixed at θ_n during the experiment

$$E_{\theta_1}^{\theta_n} = E_{\theta_1}^{\theta_2} + E_{\theta_2}^{\theta_3} + \dots + E_{\theta_{n-1}}^{\theta_n} \quad (2.15)$$

Between 0° and 100°C , three fixed temperatures can be possible, 0°C , room temperature (say 30°C) and 100°C , which will not change during experiment. So in the laboratory, law of successive temperatures can be easily verified as

$$E_0^{100} = E_0^{30} + E_{30}^{100}$$

2.10 Thermoelectric Thermometer

It is the simplified arrangement to measure the temperature by using thermocouple wires. The complete thermoelectric thermometer outfit consists of:

1. Two elements constituting a thermocouple.
2. Electrical insulation of these wires and protecting tubes.
3. Galvanometer or millivoltmeter or potentiometer for measuring the thermo e.m.f. .
4. Arrangement of controlling cold junction temperature.

Requirements for Different Ranges of Temperature: For low temperatures up to 300°C , the couple of base metals such as Fe-constantan and Cu-constantan are satisfactory, as they develop a large e.m.f./centigrade. For high temperatures, base metals cannot be used, as they get oxidised and melt. Ni-Fe couple may be used upto 600°C , but above that, platinum and alloys of platinum must be used. The useful range of a thermoelectric thermometer is -200 to 1600°C . It may go up to 2100°C . The choice of metals and alloys for constructing the thermocouple depends on the range of temperature to be measured. Within the range, the thermo-e.m.f developed should be large, and to avoid the reversal of e.m.f., the neutral temperature should be far remote from the temperature to be measured.

Construction of Thermoelectric Thermometer: Thermoelectric thermometer construction is shown in Fig. 2.7a. It consists of two elements in the form of a wire, and near the hot junction is insulated with fire-clay. These wires are passed through mica discs to keep them in position. The whole arrangement is enclosed in a porcelain or quartz tube. The free ends of the wires are connected to two terminals C_1 and C_2 . These terminals are connected to the extension leads, leading to the cold junction (Fig. 2.7b). Circuit connection is clearly explained in Fig. 2.7c. The cold junction is thus transferred to a convenient distant place, where a constant low temperature, say 0°C can be maintained because maintaining cold temperature at points C_1 , C_2 (Fig. 2.7a) which is just above the hot junction (say at temperature 2000°C) is not possible. In between these two extension wires, a calibrated galvanometer is connected, from which, for a particular thermo-e.m.f., corresponding temperature of the specimen can be measured.

Working Principle

The body whose temperature is to be measured is brought in contact with the welded hot junction of the thermometer. When the junction attains the temperature of the body, the deflection of the galvanometer is noted. This reading directly gives the temperature of the body, since the galvanometer is calibrated for temperature. By choosing suitable metallic wires, we can measure temperatures between -200°C and 1600°C .

Merits

1. It has a very wide operation range, -200°C to 2100°C .
2. Since the surface area of the hot junction is small and thermal capacity of the junction is small, it can be quickly heated up. Hence, time lag is virtually absent.
3. Rapidly varying temperatures can be measured.
4. It is cheap and easy to construct.
5. It can measure temperature at a point.
6. A very high temperature (say, that of a furnace) can be measured by this method.

Demerits

1. A particular thermometer is not accurate over the wide range.

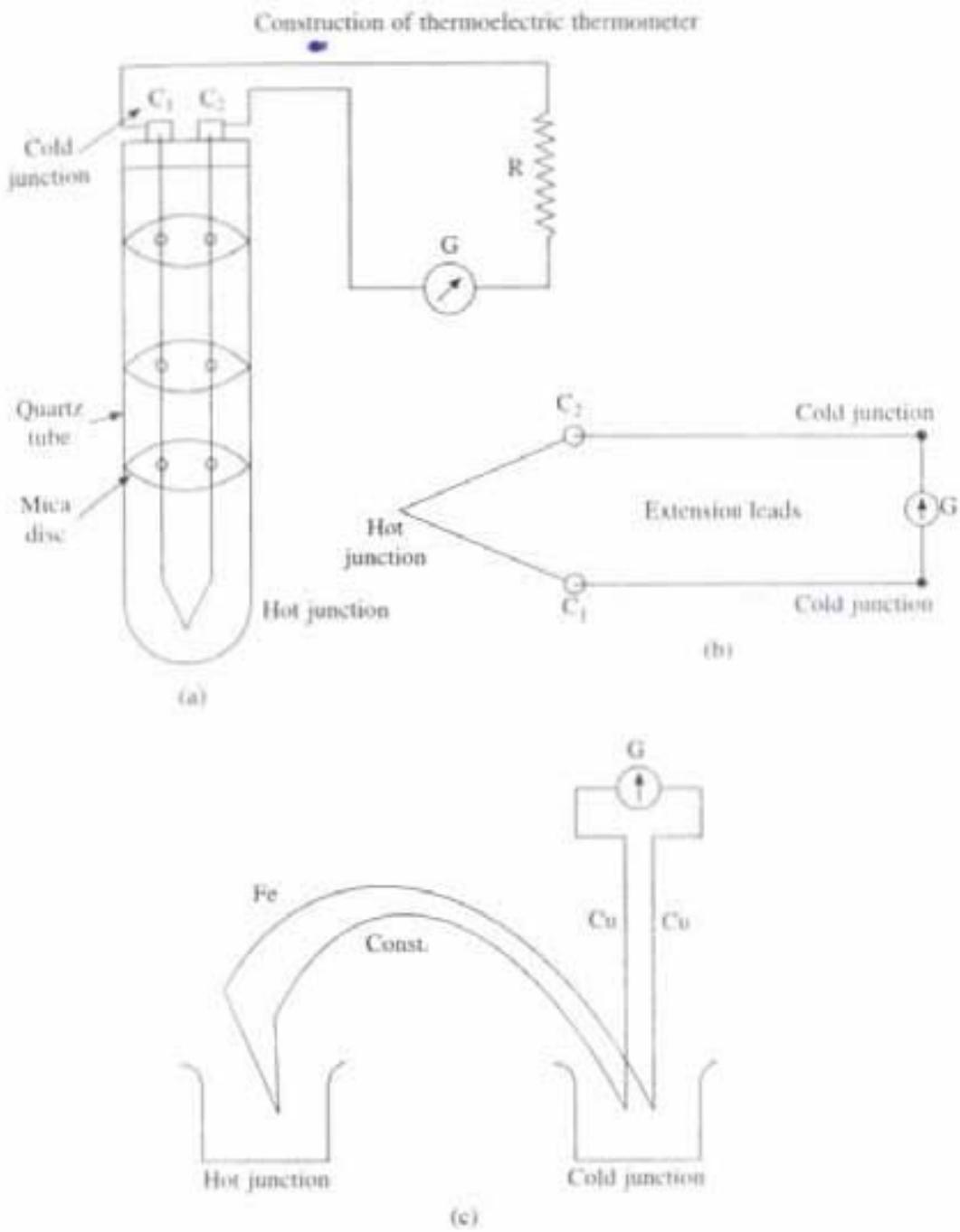


Fig. 2.7 Thermoelectric thermometer

2. There is no theoretical formula that can be extrapolated over a wide range. For different temperature range different formulae are used.
3. Different thermocouples are required for different ranges.

2.11 Thermistor

The electrical resistance of most metals or their alloys generally changes with change in temperature as

$$R_t = R_0 [1 + \alpha (T - T_0)] \quad (2.16)$$

R_t = resistance at temperature T °C

R_0 = resistance at 0°C (Ref. Temp.)

α = temperature coefficient of the resistance per $^\circ\text{C}$

The temperature coefficient of most metals are positive. So with the increase in temperature the resistivity of the metal increases. But in the case of semi-conductor the resistivity generally decreases with increase of temperature. A semi-conductor used in the manner is called thermistor which has extensive application in thermometers, in the measurement of microwave frequency, power, as a thermal delay and in control devices functioned by changes in temperature.

For most metals the resistance increases about $0.4\%/\text{ }^\circ\text{C}$ whereas for a semi-conductor say Si, Ge, resistance decreases 6 to $8\%/\text{ }^\circ\text{C}$. So metals have positive coefficient of resistance in smaller magnitude whereas semi-conductor thermistor has negative coefficient of resistance of higher magnitude.

But the temperature coefficient of thermistor can also be positive. The thermistor with positive temperature coefficient are generally used in temperature measuring instruments. The resistance of thermistor is given by

$$R_t = R_0 \times e^{\beta(1/T - 1/T_0)} \quad (2.17)$$

Temperature coefficient β can be positive or negative depending upon the property of the material (Fig. 2.8). The active material of thermistor is transition metal oxide or mixture of oxides (oxides of manganese, nickel, cobalt, copper, iron, titanium). It is a type of a semi-conductor. Thermistors are manufactured in many shapes and sizes, beads, discs, washers, rods and glass-sealed probes. Miniature beads are well suited for applications where fast response and minimum thermal loading are desired. In the case of glass encapsulation the time of response

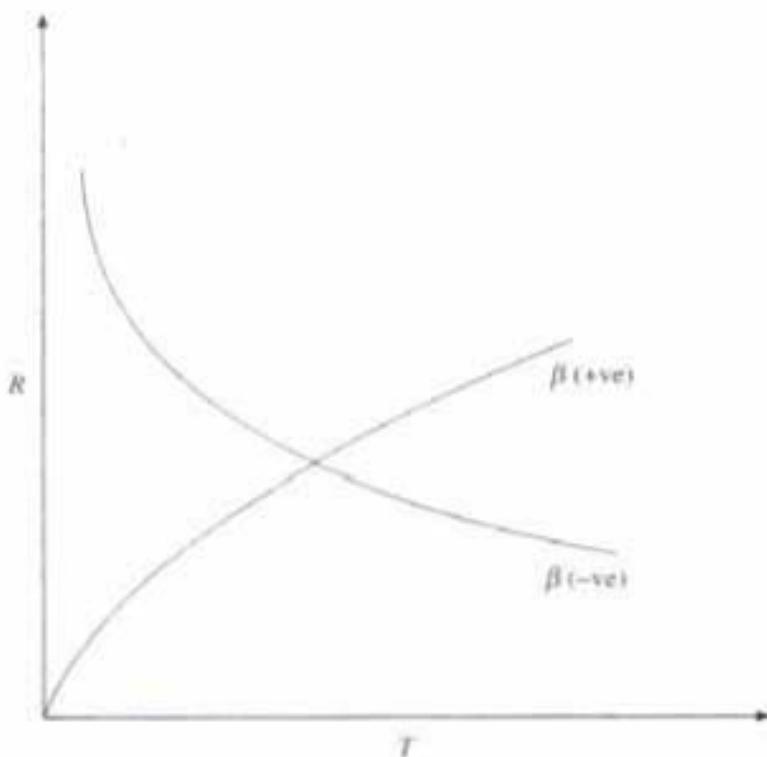


Fig. 2.8 Variation of resistance in a thermistor with temperature

of the thermistor to sudden temperature variation increases significantly, from milliseconds to one second.

Advantages

1. The advantage of using thermistor with high sensitivity is low cost, precision temperature measurement and control.
2. Thermistors are available in a wide range of high resistance values. This makes remote two-wire measurements possible without the need for temperature compensation of the leads, because the resistance of the leads becomes insignificant. Since both the resistivity and temperature coefficient of resistance of copper wire are so much lower than those exhibited by thermistors, there is virtually no limit to the length of the cable that may be used between a sensor and associated instrumentation.
3. Thermistors are available in small sizes and faster response time.
4. They have long-term stability provided they are not exposed to high temperature. Above 300°C, the stability of most thermistor is degraded.

Characteristic curve

The resistance versus temperature for typical thermistor with negative temperature coefficient is shown in Fig. 2.9. The resistance decreases non linearly with temperature coefficient α as

$$\alpha = \frac{1}{x} (dx/d\theta), \quad (2.18)$$

For small range of temperatures linear approximation is valid. The range of temperatures over which thermistors can be used is 0 to 100°C, higher temperatures

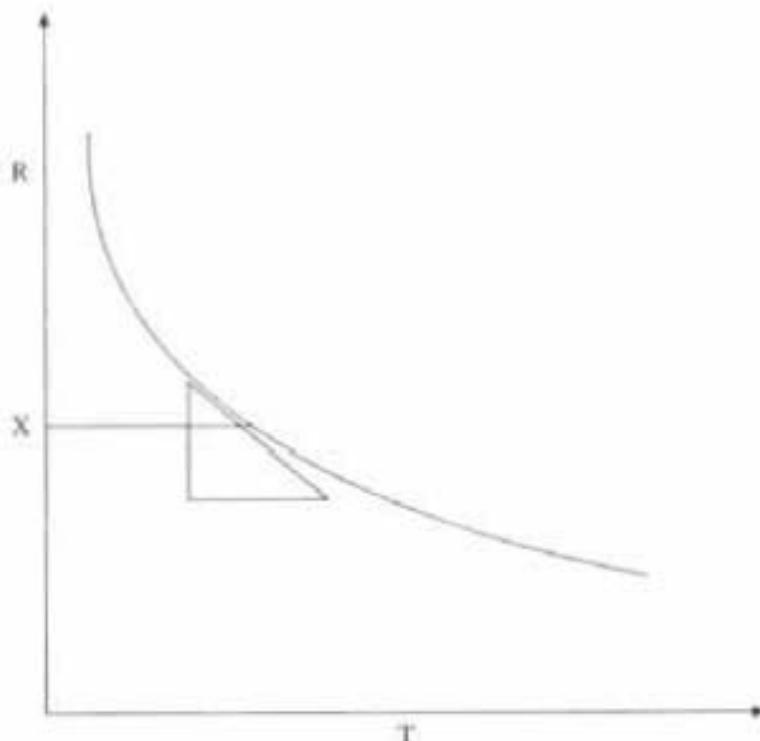


Fig. 2.9 Variation of resistance in a thermistor with temperature

can be measured with same degree of success. The resistance of the thermistor does not jump immediately to a new value when temperature changes but requires a small amount of time to stabilize at the new resistance value. This is expressed in terms of the time constant of the thermistor in a manner similar to the charging of a capacitor in RC-circuit. If the characteristic of the thermistor is known it is possible to relate the current measurement to the actual temperature, and it can be used as temperature measuring device. For measuring very low temperatures near absolute zero a special type of thermistor has been developed.

PROBLEMS

1. For Fe-Cu thermocouple it is observed that thermo e.m.f. is zero when one of the junctions is at 20°C , and the other is at some higher temperature. If the neutral temperature is 285°C , calculate the higher temperature and hence find out temperature of inversion at cold junction temperature of -20°C .

$$\begin{aligned} \text{Solution: } T_n &= 285^{\circ}\text{C} \\ T_c &= 20^{\circ}\text{C} \\ T_n &= (T_i + T_c)/2 \\ T_i &= 550^{\circ}\text{C}. \end{aligned}$$

$$\begin{aligned} \text{Now } T_c &= -20^{\circ}\text{C} \\ T_i &= 2T_n - T_c = 590^{\circ}\text{C} \end{aligned}$$

2. Calculate the neutral temperature of Fe-Ag thermocouple. The values a and b are 16.65 and -0.095 for Fe and 2.86 and 0.017 for Ag, respectively. Calculate also, the thermo e.m.f. of the thermocouple when the junctions are at 0°C and 100°C .

$$\begin{aligned} \text{Solution: } a_{\text{Fe}} &= a_{\text{Fe-Pb}} = 16.65 & a_{\text{Ag}} &= 2.86 \\ b_{\text{Fe}} &= b_{\text{Fe-Pb}} = -0.095 & b_{\text{Ag}} &= 0.017 \\ a_{\text{Fe-Ag}} &= 16.65 - 2.86 = 13.79 \\ b_{\text{Fe-Ag}} &= -0.095 - 0.017 = -0.112 \\ T_n &= -a/b = 13.79/0.112 = 123.125^{\circ}\text{C} \\ E_{\text{Fe-Ag}} &= at + 0.5 bt^2 \\ &= 13.79 \times 100 + 0.5 \times (-0.112) \times (10000) \\ &= 819. \end{aligned}$$

3. Thermoelectric power of Fe is $1734 - 4.87t$ and that of Cu is $136 + 0.95t \mu\text{V}/^{\circ}\text{C}$, where t = temperature in $^{\circ}\text{C}$. Show that e.m.f. of the thermocouple, the junctions of which are at 0°C and 100°C is 0.1307 V.

$$\begin{aligned} \text{Solution: } P &= \left(\frac{dE}{dt} \right)_{\text{Fe}} = a + bt = 1734 - 4.87t \\ P &= \left(\frac{dE}{dt} \right)_{\text{Cu}} = a + bt = 136 + 0.95t \\ a (\text{Fe-Pb}) &= 1734 & a_{\text{Fe-Cu}} &= 1734 - 136 = 1598 \end{aligned}$$

$$a(\text{Cu-Pb}) = 136$$

$$b(\text{Fe-Pb}) = -4.87 \quad b_{\text{Fe-Cu}} = -4.87 - 0.95 = -5.82$$

$$b(\text{Cu-Pb}) = 0.95$$

$$E(\text{Fe-Cu}) = at + 0.5 bt^2$$

$$= 1598 \times 100 + 0.5 \times (-5.82) \times (10000)$$

$$E(\text{Fe-Cu}) = 130700 \mu\text{V}$$

$$= 0.1307 \text{ V}$$

QUESTIONS

- Define the terms neutral temperature, inversion temperature, thermoelectric e.m.f. Explain the importance of neutral temperature and temperature of inversion in the study of thermoelectric e.m.f.
- For Fe-Cu thermocouple, neutral temperature is 285°C, when cold junction temperature is 0°C. Calculate temperature of inversion if cold junction is at 30°C.
- What is thermoelectric effect? State and explain the law of intermediate temperature and law of intermediate metal.
- For Seebeck effect, $E = at + 0.5bt^2$. Show that neutral temperature T_n is $-a/b$ where a and b are Seebeck coefficients.
- For Fe-Cu thermocouple, when junction A is at 273°K, the thermoelectric emf is found to be 0, when the other junction is at 843°K. On further increasing the temperature of junction B, the current is found to be changing its direction of flow. Calculate the temperature at which e.m.f. is maximum and the temperature of inversion of cold junction is 250°K.
- Find inversion temperature for Fe-Cu thermocouple when cold junction is at -25°C and 25°C. Neutral temperature is 175°C, when cold junction is at 25°C.
- Write short notes on: Seebeck coefficient and thermoelectric thermometer.
- Explain thermoelectric power. How does it help to determine a and b ?
- State and explain thermoelectric e.m.f. formulae. How will you find the neutral temperature from it.
- Calculate the thermoelectric e.m.f. in a Sb-Au thermocouple whose junctions are at 25°C and 100°C respectively. The value of a and b are 35.58 and 0.146 for Sb and 2.90 and 0.009 for Au, respectively. All constants are in $\mu\text{V}/^\circ\text{C}$.

Solution: $a_{\text{Sb-Au}} = 35.58 - 2.90 = 32.68$

$$b_{\text{Sb-Au}} = 0.146 - 0.009 = 0.137$$

$$\text{Thermo e.m.f.} = E_{\text{Sb-Au}} = a(t_2 - t_1) + 1/2b(t_2 + t_1)(t_2 - t_1)$$

$$\text{Now } t_1 = 25^\circ\text{C}, t_2 = 100^\circ\text{C}$$

$$\text{So } E_{\text{Sb-Au}} = 32.68 \times 75 + 1/2 \times 0.137 \times 125 \times 75$$

$$E = 2451 + 642.1875 = 3093.1875 \mu\text{V}$$

- Discuss property of thermistor.
- The e.m.f. of Fe-Pb thermocouple, where one junction is at 0°C and other at 100°C is 1185 μV, when the second junction is at 300°C, the e.m.f. is 675 μV.

Similar readings with Ag-Pb couple are 371 μV and 1623 μV , respectively. Calculate the neutral temperature for Fe-Ag thermocouple.

$$\text{Solution: } E_{\text{Fe-Pb}} = 1185 = \alpha t + \frac{1}{2} \beta t^2$$

$$1185 = \alpha(100) + \frac{1}{2} \beta(100)^2$$

$$675 = \alpha(300) + \frac{1}{2} \beta(300)^2$$

$$\text{Therefore } \alpha_{\text{Fe}} = 16.65 \\ \beta_{\text{Fe}} = -0.096$$

$$E_{\text{Ag-Pb}} = 371 = \alpha(100) + \frac{1}{2} \beta(100)^2$$

$$1623 = \alpha(300) + \frac{1}{2} \beta(300)^2$$

$$\alpha_{\text{Ag}} = 2.86$$

$$\beta_{\text{Ag}} = 0.017$$

$$\alpha_{\text{Fe-Ag}} = 16.65 - 2.86 = 13.79$$

$$\beta_{\text{Fe-Ag}} = -0.096 - 0.017 = -0.113$$

$$(T_n)_{\text{Fe-Ag}} = \frac{\alpha}{\beta} = 122.03^\circ\text{C}$$

- (13) Thermoelectric power of Fe is 17.5 $\mu\text{V}/^\circ\text{C}$ at 0°C and zero at 360°C , that of Cu is 5 $\mu\text{V}/^\circ\text{C}$ at 50°C and zero at -50°C . Find the value of e.m.f. for Fe-Cu couple when cold junction is at 0°C and hot junction at neutral temperature.

$$\text{Solution: } P = \alpha + \beta t$$

$$(P)_{\text{Fe}} = 17.5 = \alpha + \beta(0)$$

$$0 = \alpha + \beta(360)$$

$$\text{Therefore } \alpha_{\text{Fe}} = 17.5, \beta_{\text{Fe}} = -0.0486$$

$$(P)_{\text{Cu}} = 5 = \alpha + \beta(50)$$

$$0 = \alpha + \beta(-50)$$

$$\alpha_{\text{Cu}} = 2.5, \beta_{\text{Cu}} = 0.05$$

$$\alpha_{\text{Fe-Cu}} = 17.5 - 2.5 = 15.0$$

$$\beta_{\text{Fe-Cu}} = -0.0486 - 0.05 = -0.0986$$

$$(T_n)_{\text{Fe-Cu}} = \frac{-\alpha}{\beta} = 152.13^\circ\text{C}$$

$$E_{\text{Fe-Cu}} = \alpha(T_n) + \frac{1}{2} \beta(T_n)^2$$

$$E_{\text{Fe-Cu}} = 1140.98 \mu\text{V}$$

Thermionic Emission

3.1 Introduction

The process of emission of electrons from a metallic surface is called electronic emission. If this emission is caused due to heating the metal, it is called as *thermionic emission*.

3.2 Theoretical Explanation for Thermionic Emission

In a metal like copper or tungsten, there are plenty of free electrons, but they cannot leave the metallic surface. At room temperature, ordinary metals do not lose their electrons. This means that a force must exist which prevents the electrons from leaving the metallic surface permanently. For an electron to escape from the metallic surface permanently, it must have sufficient kinetic energy to overcome this force. This force is described as *surface barrier energy* (E_B).

The surface barrier is analogous to the gravitational pull of the earth. Analogous to escape velocity, an electron can come out of the metallic surface, only if its velocity or K.E. is more than a particular value. Modern physics tells us that even at the absolute zero temperature, the velocity or K.E. of all the electron does not reduce to zero. The highest energy that an electron in a metal has at 0°K is called *Fermi level of energy* (E_F). For emission of electron to take place, we have to supply additional energy from outside. The additional energy needed for emission of electron is called *work function* ($E_W = eQ$) of the metal (Fig. 3.1), that is,

$$E_W = E_B - E_F = eQ \quad (3.1)$$

Work function is expressed in terms of energy unit, electron volt. Just as escape velocity, it is different for different metals.

There are many ways to supply additional energy needed for an electronic emission. For example, we can supply additional energy in any of the following ways: heat, electrical field, light, thrust produced by bombarding the surface with electrons.

According to the way we choose the supply of additional energy, the electronic emission is classified as (1) thermionic emission, (2) high field emission, (3) photo-electric emission and (4) secondary emission. Out of these, the first one is most common.

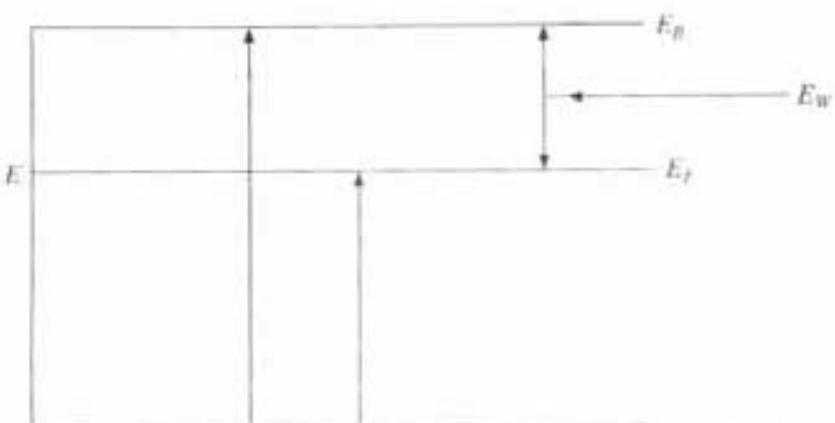


Fig. 3.1 Fermi energy level (E_f) and work function (E_w) of a metal

3.3 Richardson Dushman Equation

Richardson (1901) derived a relation between the number of electrons emitted by a unit area of the metallic surface (J) and the temperature (T) in °K of the emitting materials. Later on, this relation was corrected by Dushman. Richardson in his earlier attempt, derived this equation on a thermodynamic basis. Later on, it was derived on the basis of Fermi-Dirac Statistics.

A. Thermodynamic Derivation

The emission of electrons from the surface of the metal resembles the process of evaporation from the surface of a liquid and the energy, an electron must be given for it to escape, corresponds to the latent heat of evaporation, of the liquid.

On the basis of thermodynamic derivation, Richardson derived the equation for the current density (J) of electrons emitted from a heated filament which depends on the temperature of the filament and work function.

$$J = n_0 e v = A' e T^{1/2} e^{-Q/kT} = AT^{1/2} e^{-Q/kT} \quad (3.2a)$$

where K is the Boltzmann constant, T the temperature in °K, n_0 the number of electrons emitted per unit area per sec from filament with velocity v , A' and A are some constants, where they assumed that work function Q of the metal to be independent of filament temperature T .

When this relation was experimentally verified, it gave the values A and Q , very different from experimental values and thus it was discarded.

Later on Richardson and Dushman showed that Q is a function of filament temperature T , as given by

$$Q = Q_0 + 3/2 KT$$

where Q_0 is the work function at 0°K. With this value of Q , Richardson and Dushman got the accurate relationship as

$$J = AT^2 e^{-Q_0/kT} \quad (3.2b)$$

This thermodynamic derivation has some drawbacks. It is based on the classical theory which is contrary to the modern quantum consideration. Thus thermodynamic derivation is not adopted these days.

B. Quantum Mechanical Derivation

Thermionic emission emission is best explained by Fermi-Dirac Statistics, in which the concept of Fermi energy and Fermi band explain the condition for thermionic emission more clearly. According to this only those free electrons will escape from the metal due to heating which have got energy = surface barrier energy = E_B or greater than E_B . According to quantum consideration the K.E. of the electron is not zero at 0°K .

Highest energy that an electron in metal has at 0°K is E_F , the Fermi level of energy. So for emission to take place electrons must possess the energy (Fig. 3.1)

$$E_W = E_B - E_F = eQ = \text{Work function} \quad (3.1)$$

$E_W = eQ$ = minimum energy necessary to liberate electrons from metals.

So Richardson Dushman equation (3.2b) gives the thermionic current density as

$$J = (4 \times 3.145 \text{ meK}^2)/h^3 \times T^2 e^{-eQ/kT}$$

$$J = AT^2 e^{-eQ/kT} \quad (3.3)$$

where J = emission current density

$A = (4 \times 3.145 \text{ meK}^2)/h^3$ = emission constant

T = Temperature in $^{\circ}\text{K}$

Q = Work function of metal in e.v. (the minimum escape energy of the electron)

$Q = (E_B - E_F)/e$, K = Boltzman constant

E_B = surface barrier energy, E_F = Fermi energy,

e, m = electronic charge and mass

If the area of the emitter surface is S , then the total emission current is

$$I = J \cdot S = AST^2 e^{-b/T} \quad (3.4)$$

where

$$b = eQ/K$$

Equation (3.4) is called Richardson-Dushman equation for thermionic emission. A and b are the constants of Richardson-Dushman equation. Dushman has shown that A is constant for different metals but the value b , i.e., work function varies from metal to metal.

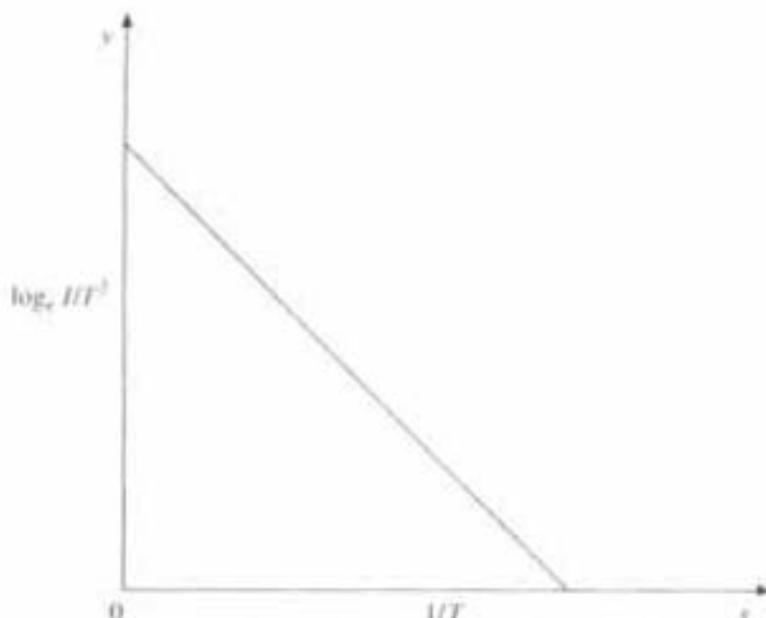
This equation suggests that for easy emission of electrons the metals should have high melting point and low work function.

Determination of the constants

From (3.4) we get

$$\log_e (I/T^2) = \log_e (A \times S) - b/T$$

This is the equation of a straight line, $y = mx + c$. Thus, if we plot the graph between $1/T$ and $\log_e (I/T^2)$ then slope (Fig. 3.2) will give the work function b and intercept along Y-axis will give the value of $\log_e (A \times S)$.

Fig. 3.2 Variation of $\log_e (I/T^2)$ with $(1/T)$

3.4 Thermionic Valve

The Richardson-Dushman equation offers the choice of materials for efficient emission which was put to use for developing electron valve. The first such valve was developed by J.A. Fleming, by using thermionic emission. It has a metal filament as an electron emitter, and a plate as an electron collector, fitted in a glass bulb, which is highly evacuated. If the plate is given a positive potential, the electrons emitted will be attracted to the plate and electric current will flow from the filament to the plate. If negative potential is given to the plate, electrons will not flow, which means it will allow only one-way flow of electrons. That is why it is considered as a valve. Here emitter is called cathode and collector is called anode. Since it has two electrodes, it is called *diode* (Fig. 3.3).

Cathodes or emitters are generally of two types: (1) directly heated or filament types (2) indirectly heated or heater type. The two types of diodes are (1) vacuum diode and (2) gas filled diode.

Vacuum diode consists of directly or indirectly heated cathode. Two possible types of electrode arrangements are: (1) parallel electrodes and (2) cylindrical electrodes.

3.5 Diode Characteristics

In a diode, three different variables are involved:

- Filament current I_F i.e. the current that heats the filament to enable it to emit electrons.
- Anode or plate current I_P , thermionic current that flows to the plate.
- Anode or plate voltage (V_P), rather high and always positive w.r.t. filament.

A graphical relation of any two of these quantities constitutes a characteristic curve.

1. The variation of plate current I_P with filament current I_F for a fixed plate voltage V_P is called *thermionic characteristics*.

2. The variation of plate current I_P with plate voltage V_P for a fixed filament current I_F is called *anode characteristics*.

3.5.1 Anode Characteristics or *V-I* Characteristics of Diode

In a vacuum diode the filament or cathode is connected to L.T. for filament heating. Between cathode and plate a variable D.C. H.T. voltage is connected (Fig. 3.4). When the filament is heated up electrons are emitted from it and if plate voltage is kept at zero (Fig. 3.3), these electrons form an electron cloud around the cathode, known as space charge, and that region is called *space charge region*. Obviously the space charge is negative and as more electrons are emitted from the filament they experience a retarding force. But as the plate voltage is increased, more and more electrons are attracted to the plate. Thus the plate current will be increased gradually. The variation of plate current I_P with plate voltage V_P is shown in Fig. 3.5. This graph is called *V-I* characteristics or anode characteristics of vacuum diode. It can be seen from the graph that at first

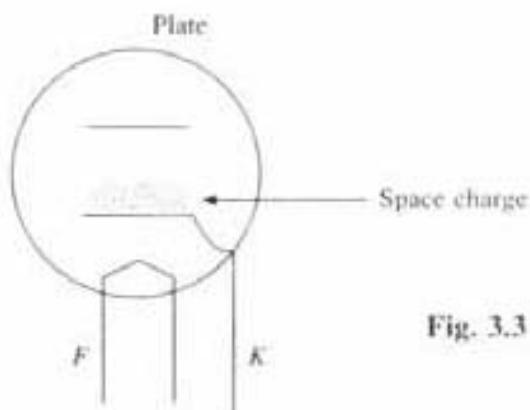


Fig. 3.3 Vacuum diode

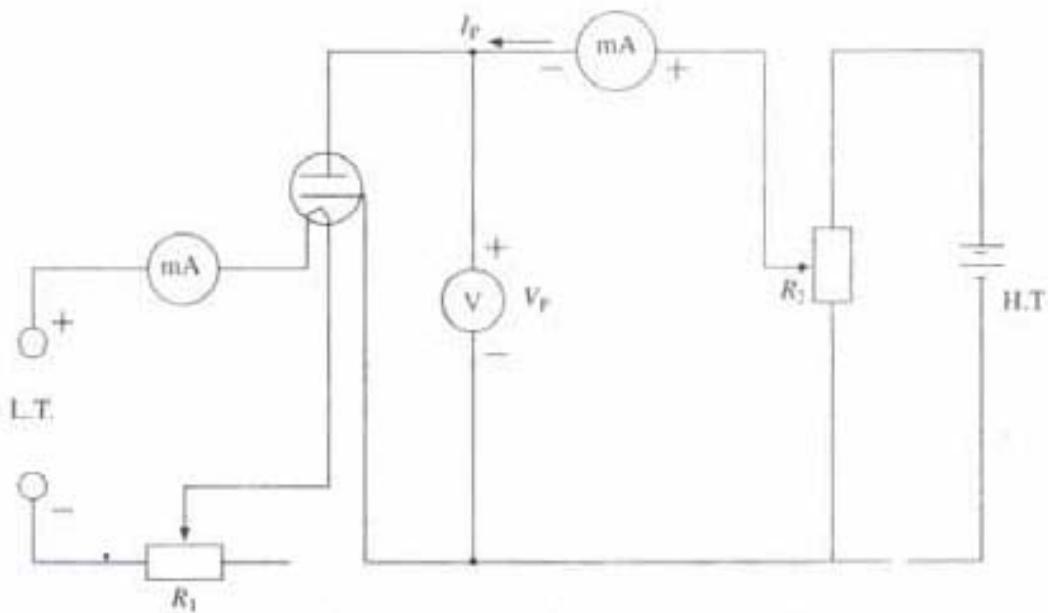


Fig. 3.4 Experimental set up for study the variation of plate current I_P with plate voltage V_P

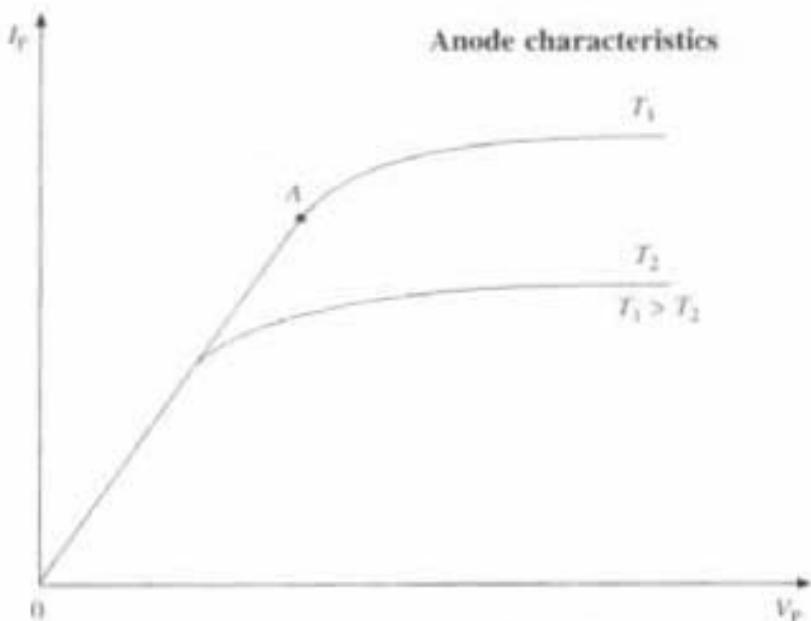


Fig. 3.5 Variation of plate current I_p with plate voltage V_p at particular filament temperature T

the plate current I_p increases rapidly with increase of V_p . A stage is reached when further increase in V_p does not result in much increase in plate current. Soon, the plate current reaches its saturation value. The voltage beyond which current starts saturating is called *saturation voltage*.

Upto saturation point A in Fig. 3.5, space charge exists around the cathode. The current in this region is limited due to presence of space charge. That is why the current in this region is called *space charge limited current*. When the plate voltage is more than the saturation voltage almost all the electrons are collected by the plate. The rate of emission becomes maximum and the collection of electrons by the plate also becomes the same. Therefore, the space charge is exhausted. Plate current becomes almost constant whatever be the value of plate voltage. If we can reduce the temperature of the cathode by reducing filament current, then at reduced temperature, the cathode cannot supply the electrons at as high rate as it was doing at the high temperature. Therefore, the saturation current is now lower. Thus, this saturation portion of the VI curve depends on the cathode temperature. So this saturation current portion is called *temperature limited current*. If the temperature of the cathode is further reduced, the saturation current further reduces. Note that till space charge exists around the cathode, the plate current is decided by this space charge. It does not matter what the temperature of the cathode is. Space charge limited current is given by Child-Langmuir's three half power law as

$$I_p = K_1 V_p^{3/2} \quad (3.5)$$

where K_1 is the constant of proportionality.

Saturation current depends on the temperature of the cathode. Current in the temperature limited region is given by Richardson-Dushman equation

$$I = A S T^2 e^{-eQ/kT} \quad (3.6)$$

where $A = \text{constant}$, $S = \text{area of cathode}$
 $T = \text{temperature in } {}^\circ\text{K}$, $K = \text{Boltzman constant}$
 $Q = \text{work function}$, $e = \text{electronic charge}$

The curve in Fig. 3.5 is called static characteristic curve for a diode since no load was taken in the plate circuit. The static plate resistance r_s and dynamic plate resistance r_d can be obtained from the static characteristic curve as follows:

$$\begin{aligned}\text{Static plate resistance} &= V_p/I_p = r_s \\ \text{Dynamic plate resistance} &= \delta V_p/\delta I_p = r_d\end{aligned}$$

Characteristic curve with a load in the plate circuit is called dynamic characteristic curve of diode.

3.5.2 Thermionic Characteristics

When anode voltage V_p is kept constant and temperature or current in the filament is gradually increased, then the curve between $I_p - I_F$ obtained for different plate voltage V_p is shown in Fig. 3.6.

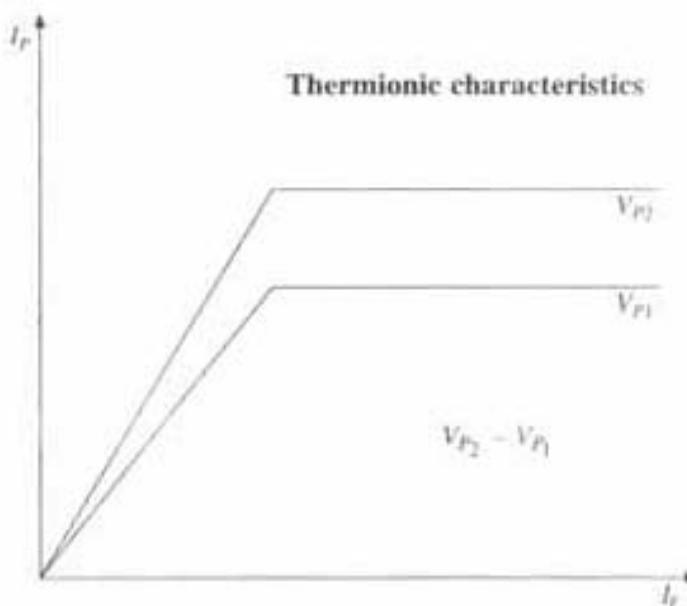


Fig. 3.6 Variation of plate current I_p with filament current I_F with particular plate voltage V_p

These curves again show the saturation in I_p for a particular value of V_p . This is because of the fact that as we increase the emission, the space charge is also increased and initially I_p will increase with increase of I_F but since V_p is constant so power of attraction by the plate is constant. Thus only a limited number of electrons can reach the plate through the space charge. So plate current I_p is saturated after certain value. Thus at the bend of the curve, for a particular plate voltage, the space charge is just sufficient to prevent more electrons from getting through the cloud of electrons already around the filament.

These two characteristics *anode characteristic* and *thermionic characteristic* suggest that it is not possible to obtain unlimited current from the plate, only by increasing the plate potential or only by increasing the filament heating current.

Space charge is always there, playing an important role for controlling the plate current.

3.6 Child-Langmuir Law for Space Charge Limited Current

A mathematical expression for plate current I_p under space charge limited current condition for parallel electrode was first derived by Child (1911). Later, Langmuir (1913) extended this analysis for cylindrical electrodes. For parallel electrode current density is

$$J = \frac{4\epsilon_0}{9} \sqrt{2e/m} \frac{V_p^{3/2}}{d^2}$$

$$J = \frac{K_1}{d^2} V_p^{3/2}$$

where J = current density, under space charge limited condition

d = distance between cathode and plate

e, m = electronic charge and mass

$K_1 = \frac{4\epsilon_0}{9} \sqrt{2e/m}$ = some constant, that depends on the diode

ϵ_0 = permittivity of free space

If S is the area of the cross-section of the electronic beam, then the equation for the plate current under space charge limited region is

$$I_p = J \cdot S = K_1 \cdot S \frac{V_p^{3/2}}{d^2} \quad (3.7)$$

Later on Langmuir modified this expression for cylindrical electrode. The total current passing at right angles through an imaginary cylinder of length l and radius r is

$$I_p = \frac{4\epsilon_0}{9} \sqrt{2e/m} \times \frac{V_p^{3/2}}{r_p^2 \beta^2} \times 2\pi l r_p$$

So total current passes through unit length is

$$I_p/l = (14.66 \times 10^6 \times V_p^{3/2}) / (r_p \beta^2)$$

$$I_p/l = (K_2 \times V_p^{3/2}) / (r_p \beta^2) \quad \text{amps/unit length of axis} \quad (3.8)$$

where $K_2 = 2\pi \times \frac{4\epsilon_0}{9} \sqrt{2e/m} = 2\pi K_1$

β = correction factor

= $\log(r_p/r_k)$, generally taken as 1 because $r_p \gg r_k$

r_p = radius of anode or plate

r_k = radius of cathode.

Equations 3.7 and 3.8 are known as *Child-Langmuir equations* for parallel plate electrodes and cylindrical electrodes, under space charge limited current region. From the equations 3.7 and 3.8, we can say that for a given electrode system

$$I_p \propto V_p^{3/2} \quad \text{i.e.} \quad I_p = K' V_p^{3/2} \quad (3.9)$$

The relation (3.9) is called *Child-Langmuir three half power law* and K' is some other constant.

3.7 Diode as Rectifier

Rectifier is a device which converts AC into DC. As diode allows the current to flow only in one direction so it can be used as a half wave or full wave rectifier.

3.7.1 Half wave rectifier

In Fig. 3.7 a circuit diagram of half wave rectifier is shown. The input AC is connected across the primary of a transformer which has two secondaries S_1 and S_2 . S_2 is of few turns that provides few volts for heating of a filament. The terminal B of S_1 is connected to the plate of the diode and terminal A is connected to the filament through a high resistance R . The output DC voltage is taken across R .

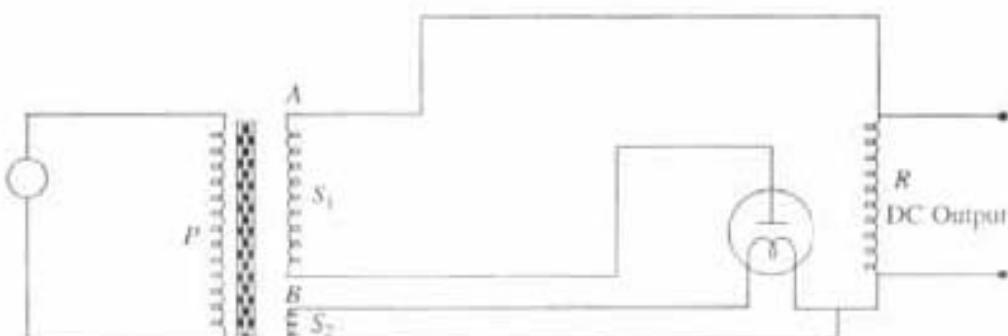


Fig. 3.7 Half wave rectifier circuit connection

As the current passes through the filament it gets heated up and starts emitting electrons. The input alternating voltage is supplied to the plate. For one half cycle the plate becomes positive and for the second half it becomes negative, i.e. when B is positive the plate becomes positive and the electrons emitted by the filament are attracted by the plate and thus a current flows in the plate circuit through R . Now in the second half cycle when B is negative the plate also becomes negative and attraction of the electrons by the plate takes place no more and as a result there is no flow of current in the plate circuit through R . Again in the next cycle the plate becomes positive and a current flows through R . This process is repeated again and again. In this circuit the current flows always in the same direction.

From above it is clear that current flows through R for half the time only when plate is positive and such a circuit is called half wave rectifier. Waveform of the output DC is given in Fig. 3.9

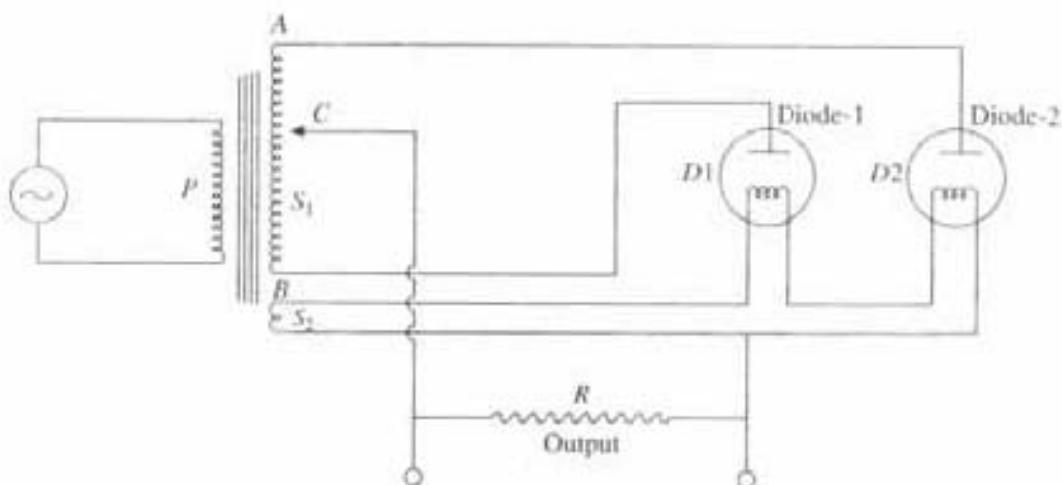


Fig. 3.8 Full wave rectifier circuit connection

3.7.2 Full wave rectifier

The circuit is shown in Fig 3.8. Here we use 2 diodes. AC input is applied across the primary of the transformer which has 2 secondaries S_1 and S_2 . The main secondary S_1 is connected to the plate of the diodes, i.e. the terminal B is connected to the plate of diode 1, and terminal A is connected to the plate of diode 2. Secondary S_2 is connected to the filaments of the diodes. The central point of the main secondary C is connected to the filament through the high resistance R . The output which is direct voltage is obtained across the resistance R .

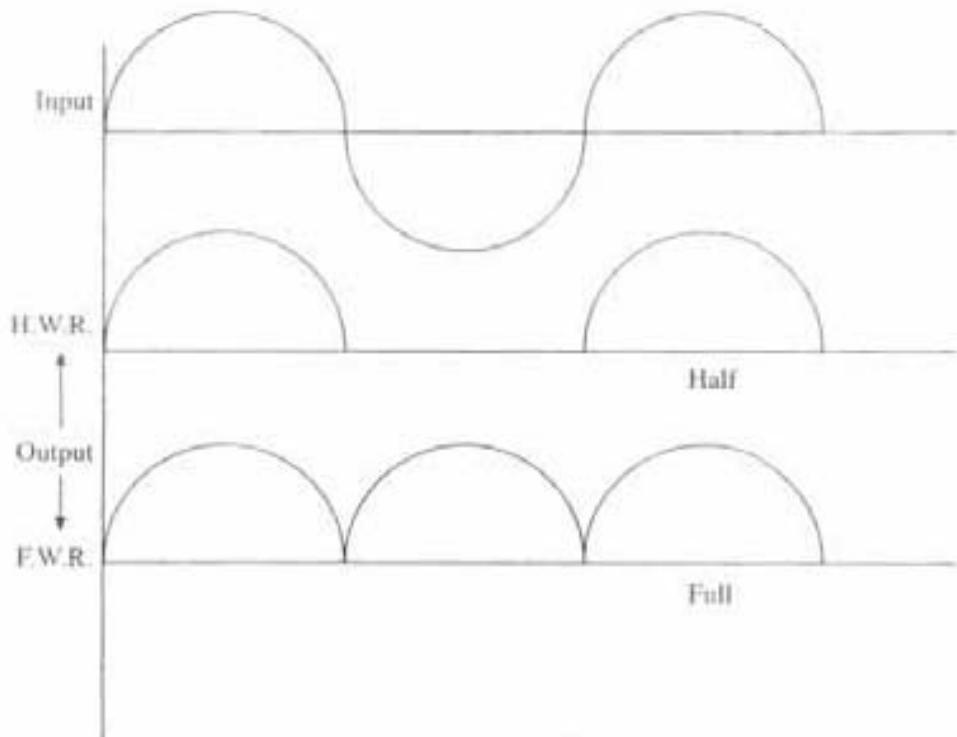


Fig. 3.9 Input and output wave form from half wave and full wave rectifier

Secondary S_2 which has few turns supplies a low voltage to the filament for emission of electrons. As the input voltage is AC due to this, when terminal A is positive then B is negative for first half cycle. Whereas in the next half cycle A

In first-half cycle, when A is positive the plate of second diode is positive, and a current flows through R due to second diode. While in the second half cycle A is negative and B is positive hence the plate of first diode is positive and as a result the plate, a current flows through R due to only first diode and so on.

In this way we have an output which is direct and flows for the whole of the cycle, such a rectifier is called full wave rectifier. The output voltage is shown for half wave and full wave in Fig 3.9. Though the output is DC but still it is fluctuating for the full wave rectifier. So to get smooth DC this output is passed through filter circuit which gives the final, constant direct output.

PROBLEMS

- (i) Tungsten filament of 2 sq mm of surface area is fitted concentrically with a cylindrical anode in vacuum diode. Calculate the maximum obtainable electronic emission current at the filament temperature 2000°K . Given the value of constants A and b in Richardson-Dushman equation for tungsten wire as 60.2×10^4 amp/m 2 K, and 52400 respectively.

Solution: $S = 2 \text{ sq. mm} = 2 \times 10^{-6} \text{ m}^2$, $T = 200^{\circ}\text{K}$

$$A = 60.2 \times 10^4 \text{ amp/m}^2\text{K}, b = Q/K = 52400$$

$$\begin{aligned} I &= AST^2 e^{-b/T} = 60.2 \times 10^4 \times 2 \times 10^{-6} \times (2000)^2 \times e^{-52400/2000} \\ &= 20.145 \times 10^{-6} \text{ amp} = 20.145 \text{ micro amp} \end{aligned}$$

- (ii) Calculate emission current density and number of electrons (n_0) emitted per unit area per second for the above problem.

$$\begin{aligned} \text{Solution: } J &= AT^2 e^{-b/T} = 60.2 \times 10^4 \times (2000)^2 e^{-52400/2000} \\ &= 10.0725 \text{ A/m}^2 \end{aligned}$$

$$n_0 = J/e = \frac{10.0725}{1.6 \times 10^{-19}} = 6.29 \times 10^{19} / \text{m}^2 \text{ sec.}$$

- For certain vacuum diode 20 mA of plate current is drawn at 80 volts of plate voltage, calculate the plate voltage to obtain 30 mA of plate current, assuming that the filament heating current is constant.

Solution: $I_p = KV_p^{3/2}$

$$I_{p1} = 20 = K(80)^{3/2} = KV_{p1}^{3/2} \quad (1)$$

$$I_{p2} = 30 = K(V_{p2})^{3/2} \quad (2)$$

From (1) and (2) we get

$$\frac{I_{p1}}{I_{p2}} = \frac{20}{30} = \frac{K(80)^{3/2}}{K(V_{p2})^{3/2}}$$

$$V_{p2} = \left[\frac{(80)^{3/2} \times 3}{2} \right]^{2/3} = 104.83 \text{ volts}$$

QUESTIONS

1. What is thermionic emission? State the equation of thermion current and explain how it is useful in designing vacuum diode. What is the role of work function in thermoionic emission?
2. Describe the experiment of plate current, plate potential, characteristics of a diode, and hence explain whether it is possible to obtain unlimited plate current by
 - (a) increasing filament current only.
 - (b) increasing plate potential only.
3. Explain the terms 'space charge limited current' in electron diode and hence explain 'three half power law'.
4. (i) What is thermionic emission? State the Richardson-Dushman equation, and hence obtain emission current density and emission current. Given the constants A and b for Richardson-Dushman equation, are $A = 6.0 \times 10^5$ amp/m²K, $b = 5.2 \times 10^4$.
4. (ii) In Richardson-Dushman equation there is a disparity for constant A , between theoretical and experimental value. Give reasons for this.
- 5.(a) Explain the characteristic curve for vacuum diode for variation of plate current with plate voltage at fixed filament current. What part of the current in diode is temperature limited and what part is space charge limited current.
(b) Explain the working of half wave rectifier using diode. Draw neat diagram and show the output wave forms.
- 6.(a) For a vacuum diode the plate current obtained is 20 mA at 100 volts of plate potential. Calculate the plate potential to obtain 25 mA plate current, filament current remains same.
(b) For a vacuum diode show graphically the region of space charge limited current and part of temperature limited current.
7. A tungsten emitter operates at 2500°K. What change in work function expressed in e.v. will cause 20% increase in emission current density?

Ultrasonic

Audio frequency range—20 Hz to 20 kHz

Radio frequency range—550 kHz to 22 MHz

TV frequency range—47 MHz to 230 MHz

Above audio range—Ultrasonic (above 20 kHz)

Below audio range—Infrasonic (below 20 Hz)

4.1 Introduction

Ultrasonic waves refers to sound waves produced by an object vibrating at a frequency higher than the human ear can hear (i.e., above 20 kHz). By using modern techniques it has become possible to produce ultrasonic waves of frequency upto 25 billion Hz, which has a wavelength of 10^{-8} m, comparable with X-ray wavelength. An ultrasonic wave is highly energetic and has extremely short wavelength because of its high frequency and energy. The use of ultrasonics, especially in the field of medicine and in various industries become very important because of its small wavelength and high energy. Due to this it has greater promises in future.

Supersonic means also of very high frequency, but ultrasonic must be differentiated from supersonic, since the former study of affects produced by objects that travel through a medium at a faster speed than the waves they generate, supersonic is essentially confined to aeroplanes, missiles, which fly through air at a speed of the sound in air.

Sound waves having frequency less than the audible range (<20 Hz) are called *Infrasonic*.

4.2 Production of Ultrasonic Waves

Ultrasonic waves cannot be produced by usual methods, like from a diaphragm of a loudspeaker, fed to alternating current. This is due to the fact that at very high frequency the inductive effects of loudspeaker coil is so large that practically no current passes through it. Moreover, the diaphragm of a loudspeaker cannot vibrate at such high frequencies. Hence different methods are specially used for the production of ultrasonic waves. The two important methods, mostly used are *magnetostriiction* and *piezoelectric*. Magnetostriiction method for frequency upto 100 kHz and piezoelectric method for frequencies above 100 kHz.

To understand magnetostriction and piezoelectric oscillators, let us first understand the working principle of amplifier and oscillator and then would discuss the oscillator principle.

4.3 Amplifier

Amplifier is an electronic device, which enables conversion of weak electrical impulses into strong impulses. When triode is used as an amplifier, the grid is normally maintained at a negative potential related to cathode. Because of the negative potential of the grid, less number of electrons are able to reach the plate. As a result the plate current is reduced. Thus the grid potential is able to control the number of electrons reaching the plate, i.e., the plate current. For a given plate voltage, a particular value of grid voltage at which the plate current is cut off is called cut-off bias (Fig. 4.1). The circuit of a vacuum tube amplifier using a triode is shown in Fig. 4.2. In the basic circuit of amplifier, grid becomes negative w.r.t. cathode by means of grid-bias voltage (E_{cc}). The input AC signal e_g to be amplified is connected between the terminal T_1 and T_2 in series with E_{cc} . On the plate side is DC voltage E_{bb} called plate supply voltage, whose positive terminal is connected to the plate through load impedance Z_L . The load impedance may be pure resistance, a complex impedance for a tuned circuit, bias voltage E_{cv} is necessary to maintain the grid always negative w.r.t. cathode (Fig. 4.2).

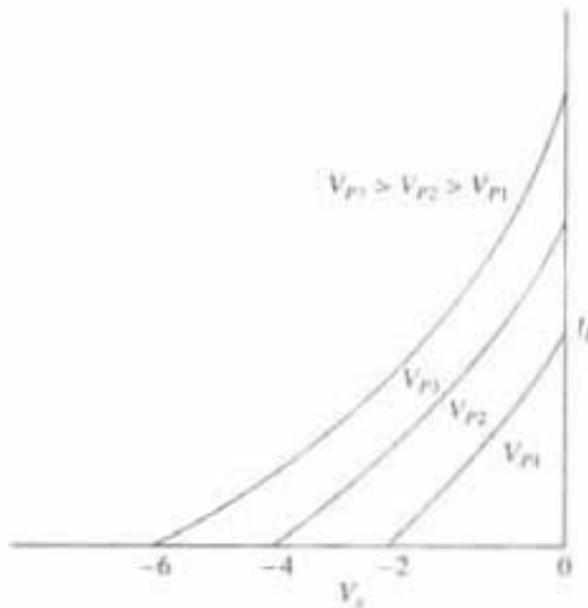


Fig. 4.1 Variation of plate current (I_p) with grid voltage (V_g) and cut-off voltage of the grid for a particular plate voltage (V_p)

In the absence of signal voltage ($e_g = 0$). The grid voltage is constant at E_{cc} , so that constant plate current (I_p) flows in the plate circuit. Due to this plate current a voltage drop ($I_p Z_L$) is produced across the load impedance, i.e., the voltage drop depends upon the plate current, which in turn, is controlled by grid voltage when $e_g \neq 0$. The AC component of plate current flowing through load impedance develops the amplified output voltage.

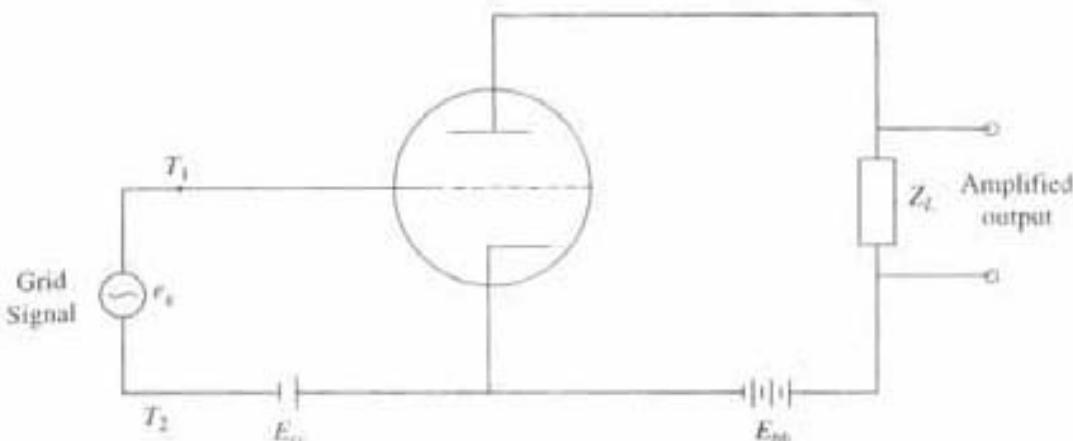


Fig. 4.2 Circuit diagram for amplifier

4.4 Oscillator

Any circuit that generates an alternative voltage is called an oscillator. To generate AC voltage, the circuit is supplied energy from DC source.

Depending upon how oscillations are produced, the oscillators are named as:

- (1) Tuned circuit (L.C.) oscillators
- (2) R.C. oscillators
- (3) Crystal oscillators

Here just to give the idea how oscillators work, the tuned circuit is described.

4.5 Tuned Circuit

An inductor (L) and capacitor (C) are connected in parallel to form a tuned or tank circuit. In Fig 4.3 (a) energy is introduced into this circuit by connecting the capacitor to a DC voltage by throwing S to 1. The capacitor gets charged and there is a voltage across it. When switch S is thrown to 2, current starts flowing in the circuit. The capacitor now starts discharging through the inductor. Since the inductor has the property to oppose the growth of current so the current builds up slowly, ultimately the capacitor is fully discharged. But the magnetic field energy around the inductor coil is maximum and the electron motion being the greatest, i.e., max. current (Fig. 4.3b).

Once the capacitor is fully discharged the magnetic field begins to collapse, and current in the inductor coil is going to decrease. But the self-inductance of inductor coil opposes the decay of current. So a back e.m.f is developed in the inductor, which keeps the current flowing in the same direction. Due to that back e.m.f. capacitor again starts charging, but with opposite polarity (Fig. 4.3c). As charge builds up across the capacitor, the current decreases and magnetic field decreases in the inductor coil. Ultimately once again all energy stored in the capacitor is in the form of potential energy. The capacitor again begins to discharge through inductor (L) (Fig. 4.3d), but this time current flows in the opposite direction. This interchange or oscillation of energy between L and C is repeated again and again. But due to the resistance of the inductor coil and the leakage of the capacitor the amplitude of oscillation decreases continuously (Fig. 4.3e). Frequency of oscillation in L.C. circuit is

$$F = 1/(2\pi\sqrt{LC})$$

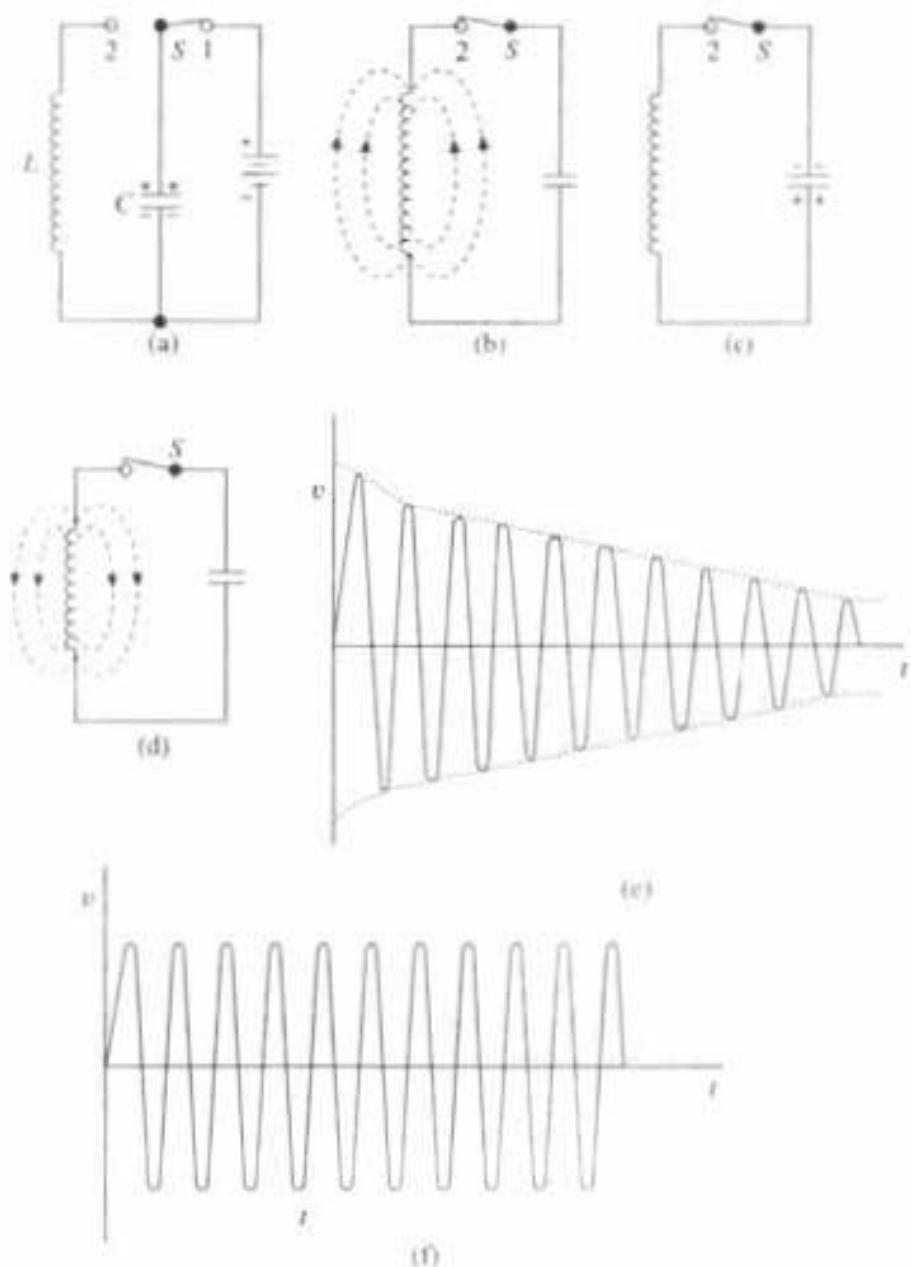


Fig. 4.3 Conversion of voltage from DC to AC in a tuned circuit

In order to get undamped oscillation or sustained oscillation (Fig. 4.3f), the pulse of energy is to be supplied at the right time in each cycle as feedback. Thus the oscillator circuits which give sustained oscillations have the following three features in common:

- (i) There must be an active device (transistor or tube) that works as an amplifier.
- (ii) There must be (+ve) feedback in the amplifier.
- (iii) The amount of feedback must be sufficient to overcome losses.

4.6 Positive Feedback Amplifier as an Oscillator

The main application of positive feedback is in an oscillator. An oscillator generates AC output signal without any input AC signal. A part of output is fed back to input and this feedback signal is the only input. Let us consider Fig. 4.4a, which

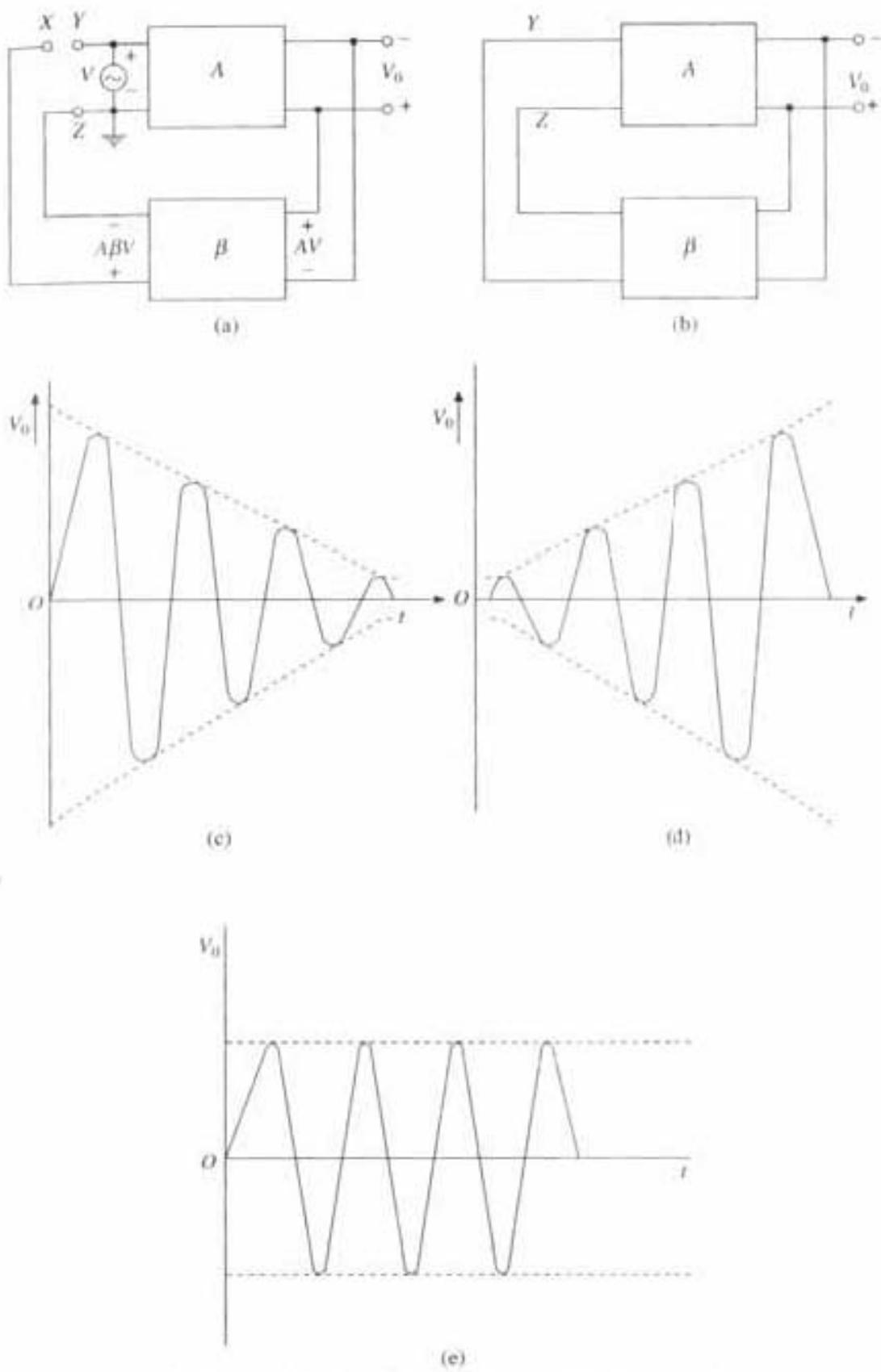


Fig. 4.4 Positive feedback amplifier as an oscillator

shows the block diagram of a feedback circuit where A is the amplification factor of the amplifier and β represents the leakage factor, from the tuned circuit.

The voltage source V derives from the input terminal Y of the internal amplifier

(voltage again A). The amplified signal AV drives the feedback network to produce feedback voltage $A\beta V$ after passing through the tuned circuit with leakage factor β . AV becomes $A\beta V$. This voltage returns to point X. If the phase shifts due to the amplifier and the feedback network is correct, the signal at point X will be exactly in phase with the signal that drives the input terminal Y of the internal amplifier.

Now if we connect points X and Y and remove voltage source V (Fig. 4.4b), the feedback signal now drives directly to the input terminals YZ of the amplifier. Now if A is the amplification factor and β the leakage factor, then, three cases are possible.

- (1) If $A\beta < 1$, $A\beta V < V$ —output signal will die out (Fig. 4.4c)
- (2) If $A\beta > 1$, $A\beta V > V$ —output signal will build up (Fig. 4.4d)
- (3) If $A\beta = 1$, $A\beta V = V$ —output signal amplitude will remain same (Fig. 4.4e)

Hence, to get sustained undamped oscillations, case (3) should be satisfied.

4.7 Magnetostriction Oscillator

To produce ultrasonic waves one of the widely used methods is magnetostriction method. Here magnetostriction effect is utilized. According to this effect, if a rod or tube of ferromagnetic magnetic material is placed into a magnetic field parallel to its length, then length gets changed slightly. The change of length is independent of the sign of the field but depending on the magnitude of the field, nature of the material, its previous treatment, the degree of earlier magnetization and the temperature. It was observed that out of various ferromagnetic materials, nickel and some alloy of nickel shows simple and regular decrease in length with the increase of magnetization and further more the phenomenon is reversible, i.e., compression of the rod along length increases magnetization. Thus if a rod of Ni is placed inside a coil which is carrying an AC current, then it suffers the some change in length for each half cycles of AC current and starts vibrating. An unmagnetized rod vibrates at a frequency twice that of the AC field. On the other hand if the rod is previously magnetized, then change in length follows the same frequency of the AC field.

The natural frequency of the rod is

$$n = \frac{P}{2L} \sqrt{E/D} \quad (4.1)$$

where L is the length of the rod, E is Young's modulus, D the density of the rod materials and P is the harmonic modes 1, 2, 3, etc.

Ordinarily the amplitude of vibrations of the rod is small. But if the frequency of the AC current is same as the natural frequency of the rod, then resonance occurs and the amplitude of vibration is considerably increased. If the applied frequency is of the order of ultrasonic frequency then from the ends of the rod, ultrasonic waves will be emitted.

An experimental arrangement for producing ultrasonic waves is shown in Fig. 4.5. There is a short Ni rod which is clamped at the centre. This rod is permanently magnetized in the beginning by passing a direct current in the coil

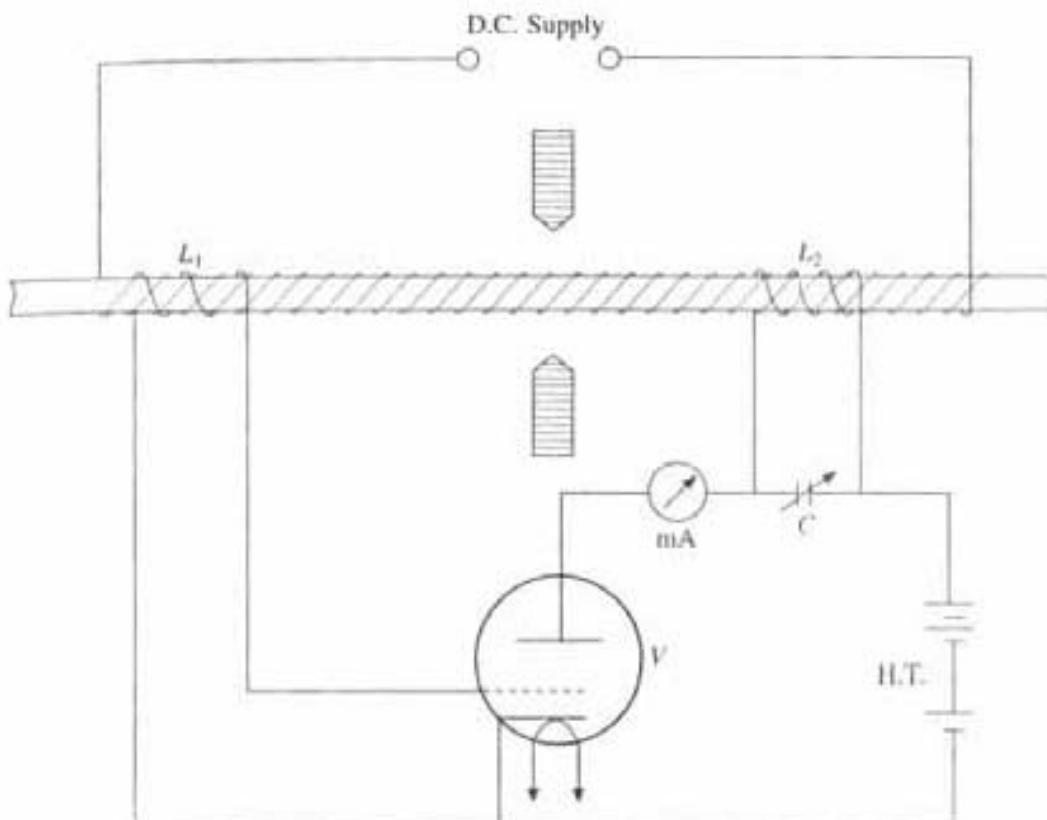


Fig. 4.5 Magnetostriction oscillator

(primary coil) which is wrapped round the rod. There are two other coils L_1 and L_2 (secondary coil) which are wrapped over primary coil as in Fig. 4.5. Coil L_2 is connected in the plate circuit of the valve, while L_1 is connected in the grid circuit and L_1 and L_2 are inductively coupled.

When the plate current passing through a coil L_2 is changed, it causes a corresponding change in the magnetization of the rod (which is already magnetized by steady current), which causes a change in the length of the Ni rod. This change in length gives rise to change of the flux linked with the grid coil L_1 and so a change in the grid voltage. This change of e.m.f. acts on the grid, produces an amplified current change in the plate circuit, i.e., in the coil L_2 , which again causes a change of length of Ni rod. In this way oscillation of the rod is maintained. The oscillation frequency f of the Ni rod is controlled by the variable capacitor C and is given by

$$f = \frac{1}{2\pi\sqrt{LC}} \quad (4.2)$$

If this frequency matches with the natural frequency $f = \frac{P}{2L} \sqrt{E/D}$ of the rod, resonance will occur. By adjusting the length of the rod and the capacity of the condenser, high frequency oscillations of different frequencies are obtained. Frequencies can be extended upto 3×10^5 Hz.

In order to secure maximum transfer of energy the transducer housing is fitted close to the medium, so that the transducer element is in contact with the

medium and transducer is operated at resonance frequency of the driving element. The driving element are Ni tubes.

The *main advantages* are for magnetostriction method:

- (1) The principal advantages of magnetostriction oscillator are simplicity and low cost of construction.
- (2) At low ultrasonic frequency large power output is possible without risk or damage to the oscillator circuit, even under temporary overload.
- (3) The magnetostiction transducers have certain advantages over piezoelectric transducers. They possess large power handling capacity and sharp resonance curve.

The *main drawbacks* are

- (1) Difficulty in imposing permanent polarising magnetisation of the Nickel rod initially.
- (2) Limitation of frequency, it is not possible to use for very higher frequency range.
- (3) Greater dependence of frequency on temperature and breadth of resonance curve, causes changes in elastic constants of ferromagnetic substances with degree of magnetisation.

4.8 Piezoelectric Effect

When mechanical forces like compression or tension are applied to a suitably cut piezoelectric crystal like quartz, rochelle salt, tourmaline in a certain direction then electric charges are developed on the end-faces of the plate perpendicular to these directions. The strains experienced by the plate make the molecules of these materials undergo a change in orientation, which results in the appearance of electric charges. The magnitude of the potential difference developed across the faces of the plate is proportional to the amount of forces applied, i.e., the extent of deformation produced and its polarity will be reverse when the direction of the deformation force is reversed. For example, if the voltage produced by the crystal is positive when compressed then it will be negative when stretched. This phenomenon is known as *direct piezoelectric effect*. This effect is directional and hence the crystal plate should be cut in proper direction relative to its axes of symmetry.

Converse is also true, i.e., if the crystal plate is subjected to electrical stress, mechanical strain in a particular direction will develop, which means that if a direct electric field is applied between two faces of the crystal plate, then it will either contract or extend and if AC field is given, then alternative extension and contraction is produced in the crystal and hence it starts vibrating. This is called *inverse piezoelectric effect*. Frequency of vibration will depend on the applied AC frequency. If the frequency of the AC field is of the order of Ultrasonic frequency and if it matches with the natural frequency of the crystal then resonance will occur and crystal will produce ultrasonic waves.

Quartz crystal belongs to the trigonal system. It has a shape of hexagonal prism with pyramid attached to each end. The axis joining two end points of the

pyramid is called *optic axis (ZZ)*. Figure 4.6(a) shown the ideal quartz crystal in its natural state while Fig. 4.6(b) shows a section of the crystal at right angle to Z axis which is a hexagon. Three lines X_1X_1 , X_2X_2 , X_3X_3 which pass through the opposite corner of the crystal are called *electrical or X-axis* and they are perpendicular to optic axis ZZ. The piezoelectric effect is most marked along electrical axis. Similarly, three lines which are perpendicular to the sides of the

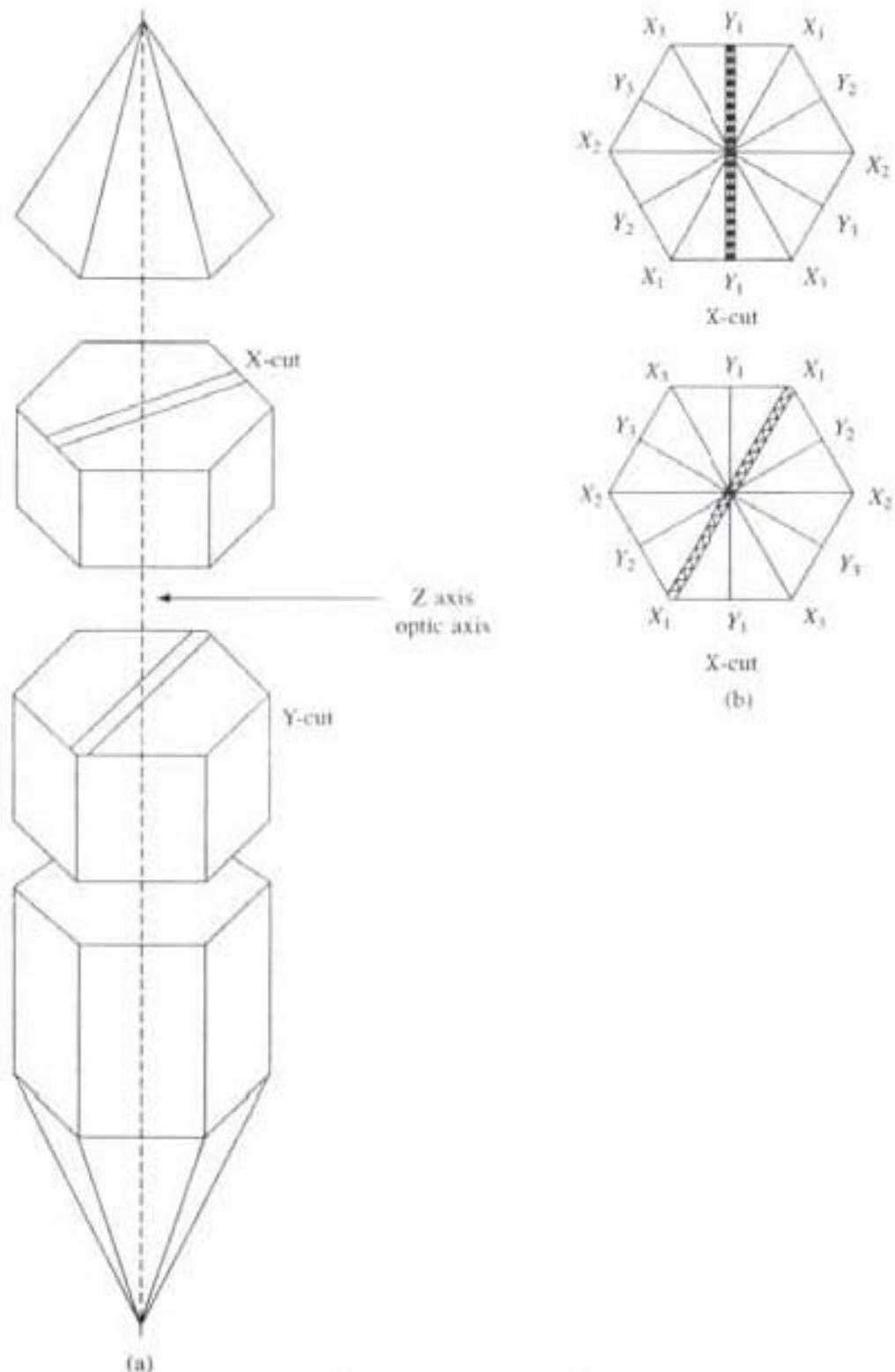


Fig. 4.6 Piezoelectric crystal

hexagon form the *Y-axis or mechanical axis*, they are also perpendicular to the optic axis. The mechanical axis are at right angles to the respective electric axis. Usually the quartz plate required to be cut has its faces parallel to the optic axis and perpendicular to the electric axis or to the mechanical axis. When its largest face is perpendicular to the *X-axis* it is called *X-cut* when its largest face area is perpendicular to the *Y-axis* it is called *Y-cut plate* (Fig. 4.6b).

For a piezoelectric crystal if an electric stress is applied in the direction of *X-axis* or electric axis, mechanical strain will develop in the direction of respective *Y-axis* or mechanical axis.

Suppose there is an *X-cut* crystal plate of thickness *t* and length *L* (along optic axis). When AC voltage is applied across the face of this plate along the electrical axis, alternating stresses and strains are set up both thickness and length-wise. If the frequency of AC voltage happens to be equal to the natural frequency of the vibration of the plate, then resonance will occur.

The frequency of the thickness wise vibration (i.e., along *X-axis*) is

$$n_t = \frac{P}{2t} \sqrt{\frac{E}{D}}$$

Similarly, the frequency of the length-wise vibration is

$$n_l = \frac{P}{2l} \sqrt{\frac{E}{D}}$$

where *P* = 1, 2, 3, etc. for fundamental (1st overtone, 2nd overtone, etc.), *E* the Young's modulus along the appropriate direction and *D* the density of the crystal plate.

4.9 Piezoelectric Oscillator

Piezoelectric effect is utilized for production of ultrasonic waves in the following way. The experimental set up is shown in Fig. 4.7. The high frequency AC voltage which is applied to the crystal is obtained by Hartley oscillatory circuit.

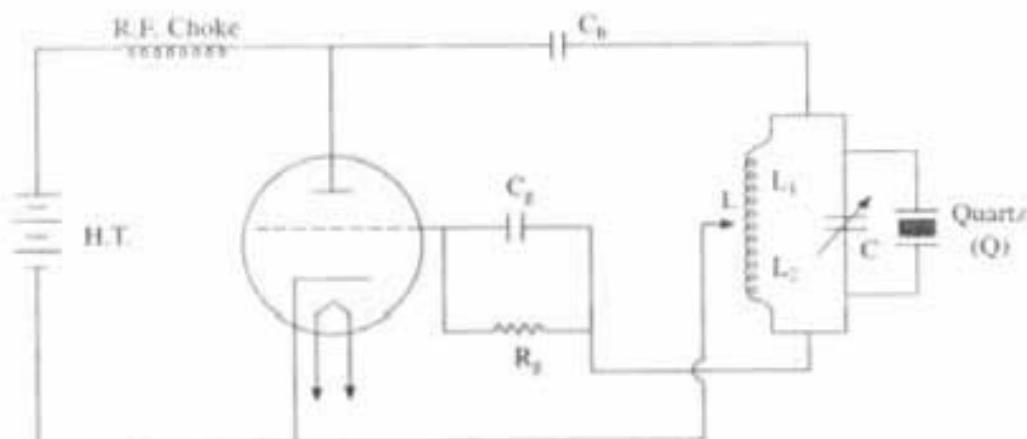


Fig. 4.7 Piezoelectric oscillator

Hartley oscillator consists of a tuned oscillatory circuit (inductance *L* and a variable condenser *C* in parallel). One end of the tuned circuit is connected to the plate while the other end is connected to the grid. Inductor coil *L* is tapped in

such a way that the portion L_1 of the coil is in plate circuit while L_2 is in the grid circuit. The amplified energy in the plate circuit is fed back to the grid circuit by means of inductive coupling. The tapping point is connected to the cathode. The quartz crystal Q is connected parallel to a variable capacitor C .

The Hartley oscillator circuit shown in Fig. 4.7 is parallel feedback circuit, the plate circuit is divided into 2 parallel branches, one to provide a path for direct current and the other for alternative current. RF choke, keeps the alternating current out of DC circuit while the blocking capacitor C_b keeps the direct current out of the AC circuit. The proper grid bias is obtained by means of R_g and C_g . The moment switch is closed the valve starts functioning as an oscillator producing oscillations having frequencies

$$f = \frac{1}{2\pi\sqrt{LC}}$$

The frequency of oscillation can be controlled by the variable capacitor C in such a way that it is tuned with the natural frequency $\left(n_r = \frac{P}{2t} \sqrt{\frac{E}{D}} \right)$ of the crystal. Then resonance will take place and ultrasonic waves will be produced by quartz crystal. Ultrasonic waves upto frequency 50,000 kHz with a moderate size quartz crystal can be produced by this method. However frequency upto 150,000 kHz can be produced by tourmaline crystal.

Advantages of piezoelectric oscillator are as follows:

- (1) For higher frequency range one has to use the piezoelectric oscillator.
- (2) They are small in size and economical.
- (3) These are very sensitive when used in detector.
- (4) The wave form is good.

The disadvantage is its low power handling capacity.

4.10 Detection of Ultrasonics

Ultrasonic waves cannot be detected directly, but some animals like a bat can do so. Ultrasonic waves are detected indirectly by the following ways:

(1) *Piezoelectric detector*: When one pair of the faces of a quartz crystal is subjected to ultrasonics, on the other perpendicular faces, varying electric charges are produced, though they are very small but can be amplified and detected by some suitable means.

(2) *Kundt's tube method*: When ultrasonic waves of relatively large wave length, is almost comparable to audible sound, were passed through Kundt's tube, then lycopodium powder sprinkled in the tube, collects in the form of heaps at nodal points and is blown off at antinodal points from which detection is possible.

(3) *Acoustic diffraction method*: When ultrasonic waves are propagated in a liquid, the density varies from layer to layer due to periodic variation of pressure. If under this condition monochromatic light is passed through the liquid at right angles to the waves, then liquid behaves as diffraction grating. This grating behaves in the same way as a ruled grating. Hence the method can

be used for detecting and for finding the wavelength and velocity of ultrasonic waves in liquid.

4.11 Properties of Ultrasonics

- (1) They are highly energetic.
- (2) Their speed of propagation depends upon their frequency, i.e., it increases with increase in frequency.
- (3) Ultrasonics show very negligible diffraction due to their small wavelength. Hence they can travel over long distances without any loss of energy.
- (4) The liquid through which ultrasonic waves pass, behaves as a diffraction grating under monochromatic light.

4.12 Applications of Ultrasonics

Ultrasonics have found numerous applications in the following fields:

(1) Communication (2) Industry (3) Scientific world (4) Medical world.
They are so useful mainly due to the following reasons:

- (1) At sufficiently high frequency almost parallel beam of plane waves can be propagated.
- (2) As the wavelengths are small, measurements can be made on a small sample without affecting the physical conditions like temperature, density etc.
- (3) Wavelength at very high frequency is of the order of molecular and electronic free path. So it is useful for studying the internal structure of a material.

A. Ultrasonics as Means of Communication

Ultrasonic waves are used for sound signaling. The high frequency sound waves may be readily formed into a beam and set in the desired direction. The beam of sound gets more and more narrow as the radius of the plate of the generator is increased in relation to the wavelength of radiated sound. At 600 Kc/s., a plate of 3.5 cm diameter is enough to confine the sound with an angle of 5 degrees. Such narrow beams can never be attained in the audible frequency range and this low power source can radiate very high intensity of ultrasonic sound waves, which can travel many km in water before being absorbed.

Submarine ultrasonic transmitters have been developed for detecting the presence of icebergs or submarines. A typical submarine ultrasonic transmitter uses piezoelectric resonator which generates powerful ultrasonic waves of frequency about 40 kHz.

These submarine ultrasonic transmitter may be use for the following purposes:

- (1) *For signalling* from ship-to-ship specially in submerged submarines.
- (2) *Determination of the depth of the sea*, position of a ship, position of submerged submarine etc. This uses the so-called echo-principle. The working principle is the same as that of a conventional pulsed radar. Pulses of ultrasonic signals are sent out at short intervals, from a piezo electric transmitter which is typically energised at a frequency of about

described below as *high power application of U-waves*.

and produces noise, but this has been successfully used in industrial devices formed by ultrasonic vibration hamper the propagation of ultrasound waves atmospheres. This high pressure results in emulsification. The cavitation bubbles the pressure around the bubbles becomes very high, about several hundreds of hollow bubbles. But these bubbles soon collapse and at the instant of implosion, ultrasound wave, such that excessive stress breaks a part of the liquid and produces bubbles. Pressure develops at some points in a liquid, when subjected to powerful standing wave formation and these result in the development and implosion of (iv) *Cavitation*: When an ultrasonic transducer is placed in a liquid it produces process of reflection of ultrasonic wave as *low power application*.

(iii) *Thickness of the gauge*: Ultrasonic waves can be used for measuring the thickness of boiler, atomic pile structure where one side is not accessible by the waves.

can convert the ultrasonic waves coming from medium into corresponding light homogeneous in a medium such as metal or plastic. For visual presentation we (ii) *Non-homogeneities*: Ultrasonic waves can be used for the study of non-

of the flaw in the metal as *low power application*.
reflections is recorded by an oscilloscope and is utilised for locating the position the hole and the back surface of the metal. This time interval between two boundary of the solid with air. This results in the reflection of waves both from a homogeneous solid in a straight line, and the waves are totally reflected in the accurately by ultrasonics. The high frequency ultrasonic vibrations propagate in the process of casting. These flaws can be done very conveniently and accurately by ultrasonics. The effects may be as large as cracks or as tiny as cavities produced during

deleets in the materials.
(i) *Flaw detector*: In most engineering applications, the strength of a component plays a significant role. The strength of the material gets reduced if there are deleets in the materials.

B. Industrial Application

distance of a submarine or icebergs is called "Sound Navigation and Ranging" (SONAR).
m 0°C in water (m/sec) ($= 1510 \text{ m/sec}$), S = salinity (g/militer), T = temperature where $V = \text{velocity of sound in } t^{\circ}\text{C}$ in sea water, (m/sec), $V_0 = \text{velocity of sound}$

$$V = V_0 + 1.14S + 4.21T - 0.037T^2 \quad (4.3)$$

The velocity of sound in sea water is measured.

also uses piezo crystal followed by an amplifier and the time interval between the transmitted and received signals is noted by time measuring device. By knowing the velocity of sound in water, the distance can be 40 Kc/s. The returned echo is received in the ultrasonic receiver, which

(1) *Preparation uniform alloy*: This process of cavitation is used for dispersion of metals in molten materials to obtain uniform alloying. By action of *U-waves* it has become possible to disperse different metals such as Na, K, Hg, Pb, Zn, Cu, etc. and their fusible alloys in alcohol, oil, and water.

(2) *Soldering and metal cutting*: The cavitation action of *U-waves* can be used for soldering, metal cutting and drilling processes in metals.

(3) *Ultrasonic cleaning device*: Contaminated articles can be cleaned by irradiating ultrasonically in sequences in suitable solvents. Nowadays, this technique is use for washing textiles. In ultrasonic cleaning the large variation of acoustic pressure actually breaks off the contaminated particles from the surface.

(4) *Preparation of very stable emulsion* of two immiscible liquid such as oil and water.

C. Scientific Applications

(i) *Photographic emulsion*: *U-waves* improve the homogeneity and stability of photographic emulsions. Their colour sensitivity has also been increased by thorough dispersion of the dye in the emulsion under the influence of *U-waves*.

(ii) *Coagulation and crystallisation*: The particles of a suspended liquid can be brought quite close to each other using Ultrasonics so that coagulation may take place which helps in the rate of crystallisation.

(iii) *Ultrasonics in the study of structure of matter*: This is a technique of ultrasonics displaying smaller and smaller inhomogeneities in metal. A limit of the improvement is reached when the wavelength of the *U-wave* becomes comparable to the size of the crystal grain of the test materials for the frequency range of 1–20 Mc/s.. Then multiple reflections and interferences are caused due to the scattering of the *U-waves* at the boundaries of the crystal grain. Then the crystalline material behaves like a 3 dimensional diffraction grating where the diffraction pattern so obtained by *U-waves* reveal the internal symmetry of the crystalline solid, like X-ray diffraction pattern.

D. Medical and Biological Applications by Ultrasonography

(a) *Disease treatment*: The body parts affected due to neuralgic or rheumatic pains, on being exposed to ultrasonics get great relief.

(b) *Detection of abnormal growth*: Cerebral ventricles are explored by *U-waves* for locating abnormal growth in the brain.

(c) *Surgical use*: Surgical uses of *U-waves* includes the selective cutting of the tissues during an operation.

(d) *Dental cutting*: *U-waves* have been found very useful for dental cutting because: (i) they make the cutting almost painless, (ii) they cut the hard material very easily, (iii) they do not require any mechanical device for cutting purpose.

Transducers are actually energy converters, that is, they convert the energy from one transmission system, where it is activated, to second transmission system either in the same form or in another form. An ultrasonic transducer can be used with electronic amplifier for development of intense acoustic fields in solutions and materials at frequency from few thousand Hz to many megahertz. Common crystal, both useful for converting electric energy to mechanical acoustic energy forms of ultrasonic transducers are the magnetostrictive bar and the piezoelectric crystal.

4.14 Ultrasonic Transducer

Ceramic piezoelectric materials can be used for obtaining transducers with curved surfaces by stretching and grinding, in which piezoelectric axis is oriented automatically at every point, normal to the surface during polarization. For quartz crystal in transducer ground with a curve, is not possible. If the transducer ground excessively concave or convex, faces no longer contribute to the radiation of longitudinal waves. So the considerable concentrations of the sound field along the axis of the cylinder can be achieved by using PZT ceramic piezoelectric crystal over natural crystal. For nondestructive testing lead manganobate and lithium sulphate are the best.

PZT ceramic substance is frozen state due to the alignment one axis. PZT ceramic materials close to the curie temperature. PZT ceramic materials thus remain is not reheated randomly earlier, now align along one axis, provided the material PZT ceramic materials become polarized along one axis, provided the material which oriented randomly earlier, now align along one axis and get frozen. So the due to this voltage, the small elementary crystals in the ceramic material, and a direct voltage of few thousand volts per centimeter thickness is applied; off, and a direct voltage of few thousand volts per centimeter thickness is applied; than BaTiO₃, curie temperature 120°C. The material is then permitted to cool PZT ceramic substance is first heated to curie temperature up to 350°C (more than 300°C) and then cooled slowly to room temperature.

Ceramic piezoelectric crystals can be formed in different required shapes. The ground, raw material mixed with binders, is moulded as required by pressing and sintering, above 1000°C and then shaped accurately by grinding as required. PZT ceramics are white to yellowish materials of lower hardness and resistance than quartz, their dull surfaces can be silver plated by baking process.

Ceramic piezoelectric crystals can be formed in different required shapes.

The ground, raw material mixed with binders, is moulded as required by pressing

and sintering, above 1000°C and then shaped accurately by grinding as required.

PZT ceramics are white to yellowish materials of lower hardness and resistance

than quartz, their dull surfaces can be silver plated by baking process.

Ceramic piezoelectric properties can be given by polarizing the

ceramic material mixed with binders, is moulded as required by pressing

and sintering, above 1000°C and then shaped accurately by grinding as required.

PZT ceramics are white to yellowish materials of lower hardness and resistance

than quartz, their dull surfaces can be silver plated by baking process.

Ceramic piezoelectric properties can be given by polarizing the

ceramic material mixed with binders, is moulded as required by pressing

and sintering, above 1000°C and then shaped accurately by grinding as required.

PZT ceramics are white to yellowish materials of lower hardness and resistance

than quartz, their dull surfaces can be silver plated by baking process.

4.13 PZT Ceramics

(c) Biological effect: Small animals like frogs, rats, fishes, etc. can be killed or injured by high intensity ultrasonics. This may be the only one destructive application of U-waves since other applications described above are all nondestructive applications of U-waves.

PROBLEMS

1. Two ships are anchored at some distance away in the deep sea. An ultrasonic signal of 50 kHz is sent from one ship to another by two routes (1) through water with velocity 1372 m/s and (2) through air with velocity 342 m/s. Calculate the distance between the two ships, when time difference to receive two signals is 3 sec.

Solution: $S = Vt$

For air $V_1 = 343$ m/s, $S = V_1 t_1$

For sea water $V_2 = 1372$ m/s, $S = V_2 t_2$

$$t_1 - t_2 = 3 = S/V_1 - S/V_2$$

$$3 = \frac{S(1372 - 343)}{343 \times 1372}, S = 1372 \text{ m.}$$

2. An ultrasonic source of 0.07 MHz sends down a pulse towards the sea bed, which returns after 0.65 sec. The velocity of sound in sea water is 1700 m/sec. Calculate the depth of sea and the wavelength of pulse.

(a) $2d = Vt = 1700 \times 0.65$

$$d = 552.5 \text{ m}$$

(b) $V = n\lambda, \lambda = V/n = 1700/0.07 \times 10^6 = 2.43 \times 10^{-2} \text{ m.}$

$$\lambda = 0.0243 \text{ m.}$$

3. Calculate natural frequency of 40 mm length of pure rod. Given the density of pure iron is $7.25 \times 10^3 \text{ kg/m}^3$ and if the Young's modulus is $115 \times 10^9 \text{ N/m}^2$. Can you use it in magnetostriction oscillator to produce ultrasonic waves?

Solution: $n = \frac{P}{2l} \sqrt{\frac{E}{D}} = \frac{1}{2l} \sqrt{ED} \text{ for } P = 1$

$$l = 40 \text{ mm} = 40 \times 10^{-3} \text{ m}$$

$$E = 115 \times 10^9 \text{ N/m}^2, D = 7.25 \times 10^3 \text{ kg/m}^3$$

$$n = \frac{1}{2 \times 40 \times 10^{-3}} \sqrt{\frac{115 \times 10^9}{7.25 \times 10^3}} = 0.04978 \cdot 10^4 \cdot 10^6 \\ = 49.78 \text{ kHz.}$$

This frequency is more than the audible range. So this rod can be used in magnetostriction oscillator for production of *U*-waves.

QUESTIONS

1. What is direct piezoelectric effect and inverse piezoelectric effect. Explain its important uses (application and importance at least one each).
- 2a. What is ultrasonics? Explain the magnetostriction method of producing ultrasonic energy and hence describe its advantages over the piezoelectric method.
- 2b. Distinguish between low power application and high power application of ultrasonics and hence give one example of each.
- 3a. Discuss the advantages and disadvantages of piezoelectric and magnetostriction

- 4a. Explain the principles of the following:
 (1) magnetostriction effect, (2) piezoelectric effect.
- 4b. Describe in brief the scientific applications of ultrasonic waves.
5. What are the application of ultrasonics. Describe the nondestructive testing application in detail.
- 6a. What is magnetostriction effect? Draw circuit diagram of magnetostriction oscillator. Explain its working.
- 6b. State the application of ultrasonic waves using the principle of (i) Echo sounding and (ii) Cavitation.
- 6c. Calculate the natural frequency of ultrasonic waves using following data:
 Thickness of quartz plate = 5.5×10^{-3} m
 Young's modulus of quartz = 8×10^{10} N/m
 Its density = 2.65×10^3 kg/m³.
7. What are PZT ceramics, and how do they help in the construction of ultrasonic transducers.
- 8.(a) Explain piezoelectric effect. Draw circuit diagram for production of ultrasonic waves using piezoelectric effect. Explain its working
- (b) State the applications of ultrasonic waves using the principle of (i) Cavitation (ii) Nondestructive testing.
- (c) Calculate natural frequency of iron rod of 0.03 m length. The density of iron is 7.23×10^3 kg/m³ and Young's modulus 11.6×10^{10} N/m².
9. An ultrasonic beam of 1 cm wavelength is sent by a ship, returns from the sea bed after 2 sec. If velocity of ultrasonic beam in sea water is 1510 m/sec at 0°C, its salinity at 30°C is 29 g/litre, calculate the depth of the sea bed at 30°C and frequency of ultrasonic beam.

$$\text{Solution: } V = V_0 + 1.14s + 4.21t - 0.037r^2$$

$$V_0 = 1510 \text{ m/sec} = 151000 \text{ cm/sec}$$

$$S = 29 \text{ gm/litre} = \frac{29 \text{ gm}}{1000 \text{ C.C.}} = \frac{29 \text{ gm}}{(10 \text{ cm})^3} = 29 \times 10^{-3} \text{ gm/cc}$$

$$t = 30^\circ\text{C}$$

$$V = 151093 \text{ cm/sec} = 1510.93 \text{ m/sec}$$

$$\text{Depth of the sea} = d = \frac{VT}{2} = 1510.93 \text{ m}$$

$$\text{Frequencies} = v = \frac{V}{\lambda} = 151093 \text{ Hz} = 151.093 \text{ kHz}$$

10. Find echo time of ultrasonic pulse, which travelling with velocity 5.9×10^3 m/sec in mild steel whose correct thickness measured by gauge meter is 18 mm.

$$\text{Solution: } 2d = vt, T = \frac{2d}{v}$$

$$T = \frac{2 \times 18 \times 10^{-3}}{5.9 \times 10^3} = 6.1 \times 10^{-6} \text{ sec}$$

Hence, echo time is 6.1×10^{-6} sec.

11. The ultrasonic flaw detector is used with defective and non-defective steel bar of thickness 40 cms. If pulse arrival times are 30 μs and 80 μs , locate distance at which defect has occurred?

Solution:

$$d_1 = \frac{vT_1}{2}, v = \frac{2d_1}{T_1} = \frac{2 \times 40}{80 \times 10^{-6}} = 10^6 \text{ cm/sec}$$

$$d_2 = \frac{vT_2}{2} = \frac{10^6 \times 30 \times 10^{-6}}{2} = 15 \text{ cm}$$

Hence, defect occurred at a distance of 15 cm.

Acoustics

In the beginning of the 20th century the acoustic properties of halls, theatres, concert halls, etc. received very little or no consideration at all by the architects or building engineers. Very often buildings designed for such public purposes were built in such a way that they were found to be unsatisfactory for the purpose for which they were built. There seemed to be no planned effort to ensure the best reproduction of speech and music. Ultimately, this branch of science also got proper attention for its improvement and for best performance.

5.1 Basic Requirements for an Acoustically Good Hall

The branch of science which deals with the planning of a building or a hall with a view to provide best audible sound to the audience is called the *acoustics of buildings* or *architectural acoustics*. Prof. W.C. Sabine, Harvard University (1911) was the first to scientifically tackle the problems of satisfactory features of good acoustics:

1. The sound heard must be sufficiently loud in every part of the hall and no echoes should be present.
2. The total quality of the speech and music must be unchanged, i.e., the relative intensity of several components of a complex sound must be maintained.
3. For the sake of clarity, the successive syllables spoken must be clear and distinct, i.e., there must be no confusion due to overlapping of syllables.
4. Reverberation should be quite proper.
5. There should be no concentration of sound in any part of the hall.
6. The boundary should be sufficiently sound proof to exclude extraneous noise.
7. There should be no echelon effect.
8. There should be no resonance within the building.

5.2 Reverberation and Time of Reverberation

Reverberation means the prolonged reflection of sound from walls, floors, and ceiling of a room. It also means the gradual dying out of sound. *This prolongation or persistence of sound in a room due to reflection, even when the sound sources have stopped, is called reverberation.* So it is also defined as the *persistence of audible sound after the source has stopped to emit any sound.* Thus time of

reverberation of a room or hall is defined as the time taken by the sound to fall from its average intensity to inaudibility. In 1911, Sabine first defined the "standard reverberation time" by considering all the factors as "the time taken by the sound to fall to one-millionth of its intensity just after the source is cut off." Sabine found that time of reverberation depends upon the size of the hall, loudness of the sound, and upon the kind of music or sound for which the hall is to be used.

In a hall if the reverberation time is too large, then there is overlapping of successive sounds which causes confusion and results in loss of clarity in hearing. On the other hand if the reverberation is very small then the loudness will be inadequate. Thus the time of reverberation for a hall should neither be too large nor too small. It must have a definite value which may be satisfactory for the speakers and the audience. The perfect time for reverberation is called *optimum reverberation time*. The expression of optimum reverberation time was given by Sabine.

5.3 Derivation for Optimum Reverberation Time (R-Time)

Sabine derived the formula for calculating optimum reverberation time for a particular hall. This is based on the assumption that the sound energy in the room is distributed uniformly all around. In derivation of that formula, following steps are to be considered.

1. To calculate, in terms of energy density E , the rate at which the energy is incident upon the walls and other surfaces and hence the rates at which it is being absorbed.
2. To calculate the final steady value of E in terms of the rate of emission of power P of the sound emitting source.
3. To calculate the standard reverberation time T .

Step 1. If ds is a small element of plane wall AB , then the rate of energy received from the source at ds is given by

$$E \cdot ds \cdot c/4$$

where c is the velocity of sound. If a is the coefficient of absorption of the wall AB then

$$\text{Rate of energy absorbed by } ds = Ec \cdot a \cdot ds/4$$

Hence, the total absorption of all the surfaces of the wall where the sound is falling will be

$$E \cdot c \cdot (\sum a \cdot ds)/4 = EcA/4 \quad (5.1)$$

where $\sum ads = A$ = total absorption of all the surfaces on which sound falls.

Step 2. If P is the power output, i.e. the rate of emission of energy from the source and V total volume of the room, then the total energy in the room at the instant when energy is E will be $E \cdot V$.

So,

$$\text{Rate of growth of energy} = \frac{d(EV)}{dt} = V \frac{dE}{dt}$$

But at any instant,

Rate of growth of energy in space = Rate of supply of energy by source – Rate of absorption by all the surfaces

i.e.

$$\frac{VdE}{dt} = P - EcA/4 \quad (5.2)$$

Thus, when steady state is reached, $dE/dt = 0$.

If E_m is the steady energy, then from eqn. (5.2)

$$E_m = 4P/cA \quad (5.3)$$

Now eqn. (5.2) becomes

$$\begin{aligned} \frac{dE}{dt} &= \frac{P}{V} - \frac{cA}{4V} \cdot E \\ \frac{dE}{dt} + \alpha E &= \frac{4P}{cA} \quad \alpha \end{aligned}$$

where

$$\alpha = cA/4V$$

Multiplying both sides by $e^{\alpha t}$ and by integration w.r.t t , we get

$$Ee^{\alpha t} = \frac{4P}{cA} e^{\alpha t} + K \quad (5.4)$$

where K = constant of integration.

Growth of Sound Energy Density in a Hall

The general distribution of sound energy in a hall is

$$Ee^{\alpha t} = \frac{4P}{cA} e^{\alpha t} + K \quad (5.4)$$

where P = power transmitted, A = total absorption by the hall

c = velocity of sound 340 m/sec, $\alpha = cA/4V$, V = volume of hall

t = time, K = constant of integration.

$A = \sum a_s$ = summation of the product of different area(s) and their respective absorption coefficient (a)

In order to understand the growth of sound in a hall, which is controlled by all these factors, we have to know the value of constant K .

At $t = 0$ $E = 0$

i.e. from Eq. (5.4) $K = -4P/cA$

So, $E = 4P/cA (1 - e^{-\alpha t})$

At $t = \infty$, $E = 4P/cA = E_m$ (since $e^{-\infty} = 0$)

So $E = E_m (1 - e^{-\alpha t}) \quad (5.5)$

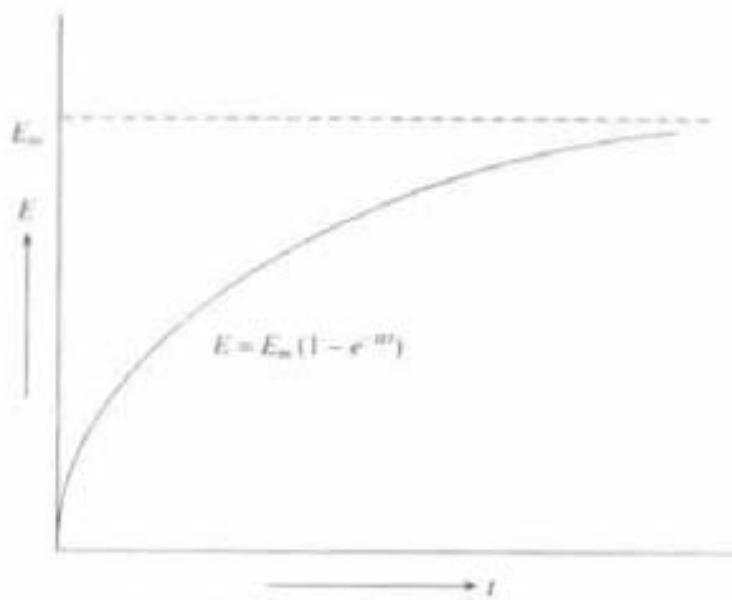


Fig. 5.1 Growth of sound with time

According to the eqn. (5.5) Fig. 5.1 shows the exponential growth of energy with time t , which ultimately reaches the value E_m .

When sound transmitted is from the speaker, it gets reflected, absorbed, and transmitted in the surroundings. If the absorption and transmission is less, then most of the sound energy will be reflected back. Then the total energy in the hall will increase linearly and attain the ultimate energy in a very short time.

But due to the presence of absorbing material in the hall sufficient absorption takes place and ultimately reaches to the maximum value E_m at $t = \infty$.

Decay of Sound Energy

If the source is cut off when $E = E_m$, then $P = 0$ at $t = 0$, and now the decay of sound energy will take place.

Here $P = 0$ at $t = 0$ and $E = E_m$.

So from Eq. (5.4), we get $E_m = K$ i.e.

$$E = E_m e^{-at} \quad (5.6)$$

Equation (5.6) shows that decay of sound energy density with time after the source is cut off. The exponential decay of sound energy is shown in Fig. 5.2. Here since no transmission power (i.e., $P = 0$), so the sound will exponentially decay with time from E_m to 0.

5.4 Sabine's Formula for Optimum Reverberation Time T (R-Time)

Reverberation time T by definition, according to Sabine, is the time taken for the energy density to fall to one millionth of its value after the source is just cut off.

After the sound source is cut off the decay of sound energy density (E) with time (t) is:

$$E = E_m e^{-at}$$

where E_m = steady energy where source is cut off (Fig. 5.2).

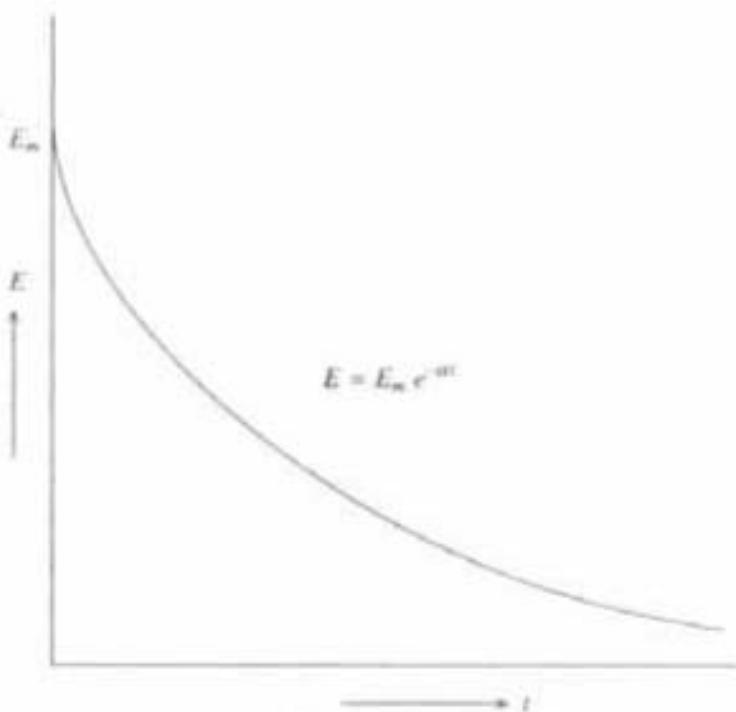


Fig. 5.2 Decay of sound with time

So when $t = T$ = reverberation time, then according to Sabine's definition,

$$\frac{E}{E_m} = e^{-at} = e^{-\alpha T} = \frac{1}{10^6} \rightarrow \alpha T = 6 \log_e 10$$

$$\text{So } T = \frac{6 \times 2.3926}{\alpha} = \frac{4 \times 6 \times 2.3026}{340} \times \frac{V}{A} = \frac{0.161V}{\Sigma as}$$

i.e.,

$$T = 0.161V/\Sigma as$$

This is Sabine's equation for optimum reverberation time. This equation is in general in good agreement with experimental value obtained by Sabine.

Limitations of Sabine's Formula

1. Sabine's formula, as deduced above, is in fact a special case of more general formula put forward by C.F. Eyring. Sabine's formula gives contradictory results in the case of a dead room, when $a = 1$, i.e., in case of complete absorption reverberation time T should be 0, but according to Sabine's formula $T = 0.161 V/s$.

2. This formula does not give correct result for absorption coefficient more than 0.2.

3. It is assumed in the derivation of Sabine's formula that there is uniform energy density and no loss of energy in air, but that is also not practical.

5.5 Absorption Coefficient

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

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

In order to compare the relative efficiency of different absorbing materials, it is essential to first select some standard of absorption, in terms of which all substances can be assessed. Sabine selected a unit area of open window, as standard, because all the sound falling on the open window passes out and none is reflected. Hence, open window is an ideal/perfect absorber of sound. Thus *the absorption coefficient of a material is defined as the reciprocal of its area which absorbs the same sound energy, as absorbed by an unit area of an open window.* Suppose 4 sq m of certain carpet absorbs the same amount of energy, as absorbed by 1 sq m of open window. Then the coefficient of absorption of carpet is $1/4 = 0.25$. The absorption coefficient is measured in open window unit (O.W.U.) or Sabines.

5.6 Measurement of Absorption Coefficient

Following are the two methods of measuring the absorption coefficient.

(1) *Single Source Method:* The absorption coefficient can be measured in terms of reverberation time (R-time). First, the R-time is measured when the absorbing material is not in the room. Let its value be T_1 .

$$\text{So, } \frac{1}{T_1} = A/0.161V = \sum a_s / 0.161V \quad (5.8)$$

where a = average absorption coefficient of R-chamber (reverberation chamber)

$\sum s$ = inside surface area

V = volume of chamber

Now there may be two different ways of measuring the absorption coefficient when an absorber is present inside the room. These are given as follows.

(a) *Absorber Suspended Inside the Room:* Now consider an absorbing material, like stage screen, which is suspended inside the room, and R-time T_2 is measured again. Then

$$\frac{1}{T_2} = \frac{\sum a_s + 2a_2 S_2}{0.161V} \quad (5.9)$$

where a_2 = absorption coefficient of the material of area S_2 , which is suspended inside the room.

Since the absorbing material is suspended inside the room, so absorption will occur from both the sides. So total absorption due to that material is $2a_2 S_2$. Now subtracting coefficient (5.8) from (5.9), we get

$$2a_2 S_2 = 0.161 V \left(\frac{1}{T_2} - \frac{1}{T_1} \right)$$

$$\text{So, } a_2 = \frac{0.161V}{2S_2} \left(\frac{1}{T_2} - \frac{1}{T_1} \right) \quad (5.10)$$

by knowing all other quantities the absorption coefficient a_2 of absorbing material can be measured, which is suspended inside the room.

(b) **Absorber Spread on the Floor or Spread on the Wall:** Now if the absorbing material is spread on the floor of R-chamber like carpet or if it is fixed on the wall like window glass or curtain cloth, then for that type of material, absorption coefficient will be determined in the following way. If R-time = T_3 with that material, whose area is S_3 and absorption coefficient a_3 then

$$\frac{1}{T_3} = \frac{\sum as + a_3 S_3 - a S_3}{0.161V} \quad (5.11)$$

Here, $a S_3$ is subtracted because now surface area S_3 of the floor or wall does not contribute to absorption which is included in $\sum as$, the total absorption by the chamber, and here it is $a_3 S_3$ not twice of that, since absorption takes place from one side only.

So by subtracting Eq. (5.8) from (5.11), we get

$$a_3 = \frac{0.161V}{S_3} \left(\frac{1}{T_3} - \frac{1}{T_1} \right) + a$$

by knowing other quantities absorption coefficient a_3 of carpet or window glass can be found out.

(2) **Double Source Method:** This method requires two sound sources, of powers P and P' . Actual powers need not necessarily to known so long as their ratio is known. The steady sound energy density,

$$E_m = \frac{4P}{cA}$$

So for two sources steady sound energy density is

$$E_m = \frac{4P}{cA} \quad \text{and} \quad E'_m = \frac{4P'}{cA}$$

Let during the time of decay, they reach a value of bare inaudibility E_0 in times T and T' , which is given by

$$E_0 = \frac{4P}{cA} e^{-\alpha T}, \quad E_0 = \frac{4P'}{cA} e^{-\alpha T'}$$

$$\text{i.e., } \frac{P'}{P} = e^{\alpha(T-T')}, \quad \text{where } \alpha = \frac{cA}{4V}$$

$$\text{So } \frac{cA}{4V} \times (T - T') = \log_e \frac{P'}{P}$$

$$A = \frac{4V \log_e (P'/P)}{c(T - T')} = \sum as$$

So average absorption coefficient of the hall is

$$A = \sum a = \frac{4V \log_e (P'/P)}{cS(T - T')} \quad (5.13)$$

where S is the total internal surface area of the hall.

By knowing all the quantities on the R.H.S. of Eq. (5.13) average absorption coefficient A can be find out.

5.7 Factors Affecting Architectural Acoustics and their Remedy

By an acoustically good hall we mean a hall in which every syllable or musical note reaches an audible level of loudness, at every point of the hall, and then quickly dies away to make room for the next syllable or group of notes. Departure from this makes the hall defective acoustically. The following factors affect architectural acoustics.

1. Reverberation: In a hall when reverberation is large, there is overlapping of successive sounds, which results in loss of clarity in hearing. On the other hand if the reverberation is very small, the loudness is inadequate. Thus, the time of reverberation for a hall should neither be too large nor too small. A formula for standard time for reverberation was given by Sabine which is

$$T = 0.161V/A = 0.161V/\sum a s$$

where A is the total absorption of the hall, V the volume, $\sum a$ the average absorption coefficient and s the internal surface area of the hall.

Experimentally it is observed that reverberation time depends upon the size of the hall, loudness of sound, and kind of music for which hall is used. For a frequency of 512 c/s, the best reverberation time is controlled by the following factors:

- (a) Providing windows and ventilators that can be opened or closed to make the value of reverberation time optimum.
- (b) Decorating the walls with pictures and maps.
- (c) Using heavy curtains with folds.
- (d) Walls are lined with absorbent materials such as felt, fiber board, etc.
- (e) Having full capacity of audience.
- (f) Carpeting the floors.
- (g) Providing acoustics tiles.

2. Adequate Loudness: Reverberation time can be made smaller by using absorbing materials, but there is a chance that the intensity of sound is weakened and may go below the level of intelligibility of hearing. Sufficient loudness in every portion of the hall is an important factor for satisfactory hearing. Loudness may be increased by following ways:

(a) To get good loudness, maximum reflection of sound from the stage area is desirable so that there is no loss of sound energy from speakers. This can be done by using large sounding boards behind the speakers facing the audience. Large polished wooden reflecting surfaces immediately above the speakers are also helpful. Use of good quality loudspeakers is also essential.

(b) Low ceilings are also helpful for reflection of sound from stage towards the audience.

3. Effect Due to Shape and Size of the Hall: Reflections from the plane surfaces are quite helpful in increasing the loudness of the sound and for distributing it uniformly, whereas the curved surfaces can be troublesome. Reflections from curved surfaces can be harmful unless designed properly. Curved surfaces on walls, ceilings, or floor produce concentration of sound into particular region, while in some other parts no sound reaches at all (Figs. 5.3a, b)

If the centre of curvature of the ceiling is made twice the height of the room, the bad focusing is removed (Fig. 5.3c). Large size of a hall results in the increase in reverberation time. Un-necessary large size of hall should therefore be avoided.

4. Absence of Echoes: An echo is heard, when direct and reflected sound waves coming from the same source reach to the listener, with a time interval of

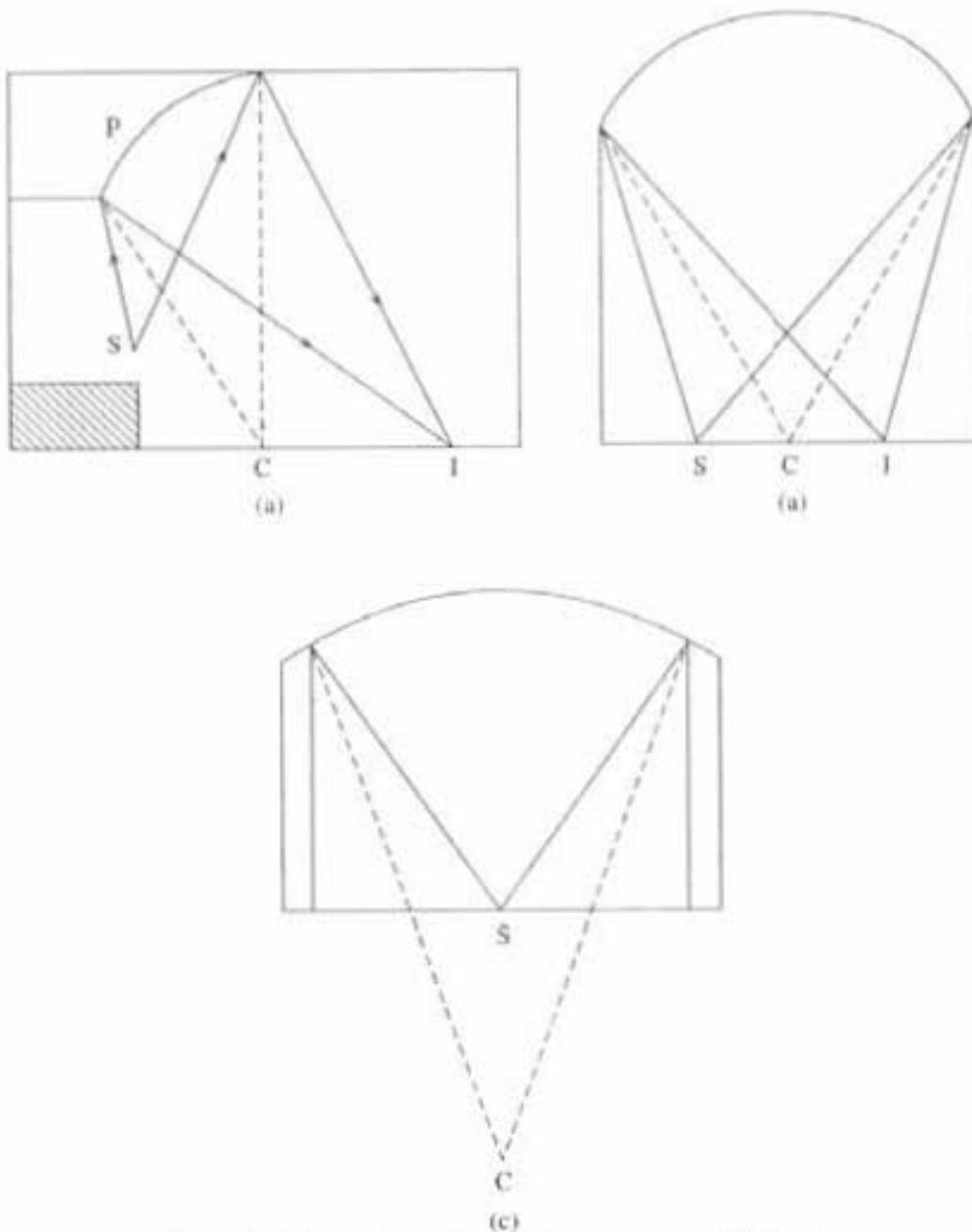


Fig. 5.3 Focussing of sound due to shape of the hall

1/7 sec. If the reflected sound is arriving earlier than that, it helps in arising the loudness, while those arriving later produce echoes, and cause confusion. These should be avoided by suitable arrangement of absorbing material for long distant walls and high ceiling.

5. Echelon Effect: A set of railings or staircase or any regular spacing of reflected surfaces may produce a musical note due to regular succession of echoes of the original sound to listener (Fig. 5.4). This makes original sound to appear confused or unintelligible. So these type of surfaces should be avoided or properly covered with heavy carpets.

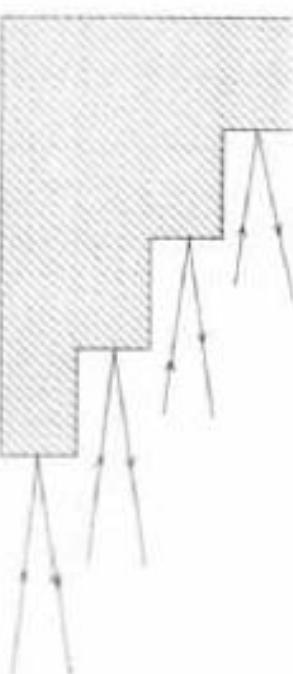


Fig. 5.4 Echelon effect

6. Extraneous Noise: Extra noise from outside or inside, other than music and speech, is called extraneous noise.

(A) *Air-borne noise:* Noises which are coming from outside through open windows, doors, ventilators, are known as air-borne noises and can be controlled by:

- (a) using double or treble windows doors with separate frames and having insulated materials between them.
- (b) using heavy glasses in doors, windows, and ventilators.
- (c) making perfect arrangement of shutting the doors and windows.

(B) *Structure-borne noise:* The noises which are conveyed through the structure of the building are called structure-borne noises. They can be controlled by:

- (a) introducing discontinuity in the path of the sound, like noise from water pipe can be controlled by using rubber at junctions.
- (b) using double walls, with an air space between them.
- (C) *Inside noise:* Inside noise in a big office from typewriters, machinery, etc., causes disturbance that can be controlled by taking proper care.

7. Resonance Within a Building: Structural resonance cause uneven response to frequency and hence can lead to unpleasant effect. Resonant frequency of a big hall is normally well below the audible range because it is proportional to $1/\sqrt{\text{vol.}}$. However, sometimes resonance can be useful. Some halls, which are famous for their acoustic properties have a large area of resonant material in the form of wood panelling in order to improve the loudness of the hall.

PROBLEMS

1. Let the reverberation time be 1.5 sec for an empty hall and 1 sec when a curtain cloth of 20 sq.m is suspended at the centre of the hall. If the dimensions of the hall are $10 \times 8 \times 6$ m, calculate the coefficient of absorption of the curtain cloth.

Solution

For empty hall R-time = $T_1 = 1.5$ sec

With curtain cloth R-time = $T_2 = 1$ sec

Area of absorber = 20 sq.m

Volume of the hall = $10 \times 8 \times 6 = 480$ cu.m

Absorption coefficient of the curtain cloth

$$a = \frac{0.161V}{2s} \left(\frac{1}{T_2} - \frac{1}{T_1} \right) = \frac{0.161 \times 480}{2 \times 20} \left(\frac{1}{1} - \frac{1}{1.5} \right)$$

So $a = 0.63$ Sabines.

2. For an empty assembly hall of size $20 \times 15 \times 10$ cu.m the R-time is 3.5 sec. Calculate the average absorption coefficient of the hall. What area of the wall should be covered by curtains so as to reduce the R-time to 2.5 sec. Given the absorption coefficient of curtain cloth is 0.5.

(a) $T_1 = 0.161V/A$ (1)

$$A = \sum aS = \frac{0.161 \times 3000}{3.5} = 138$$

S = Total inside surface area

$$= 2(20 \times 15 + 15 \times 10 + 20 \times 10) = 1300 \text{ sq.m}$$

$$\Sigma a = 138/1300 = 0.106 \text{ o.w.u}$$

- (b) By using curtain cloth of surface area S_1 whose absorption coefficient is $a_m = 0.5$ on the wall, reverberation time becomes T_2 which is equal to

$$T_2 = \frac{0.161V}{A + a_m S_1 - a S_1} \quad (2)$$

from Eqs. (1) and (2) we get

$$a_m = \frac{0.161V}{S_1} \left(\frac{1}{T_2} - \frac{1}{T_1} \right) + a$$

i.e. $a_m - a = \frac{0.161V}{S_1} \left(\frac{1}{T_2} - \frac{1}{T_1} \right)$

$$S_1 = \frac{0.161V}{(a_m - a)} \left(\frac{1}{T_2} - \frac{1}{T_1} \right)$$

Now $a_m = 0.5, a = \Sigma a = 0.106,$

$T_2 = 2.5, T_1 = 3.5,$

$$V = 3000 \text{ cu.m}$$

$$\text{so } S_1 = 140.1015 \text{ sq.m}$$

where S_1 is the surface area of the curtain cloth spread on the wall due to which R-time is reduced from 3.5 to 2.5 sec. So the area of the wall covered by curtain cloth is 140.1015 sq. m.

3. Calculate the reverberation time for: (i) empty hall. (ii) with 60 persons
 (iii) with full capacity of audience inside the hall from the following data:
 * (Volume of the hall = 1450 cu. m)

Surface	Area (sq m)	Absorption coefficient A
1. Plastered Wall	112	0.03
2. Wooden Floor	130	0.06
3. Plastered Ceiling	170	0.04
4. Wooden Door	20	0.06
5. Seat Cushioned	100 No.s	1.00/seat
6. Persons		4.70/person

Solution

$$\text{Vol. of the wall} = 1450 \text{ cu.m}$$

$$\text{Absorption due to plastered wall} = 112 \times 0.03 = 3.36$$

$$\text{Absorption due to wooden floor} = 130 \times 0.06 = 7.8$$

$$\text{Absorption due to plasterd. ceiling} = 170 \times 0.04 = 6.8$$

$$\text{Absorption due to wooden floor} = 20 \times 0.06 = 1.2$$

$$\text{Absorption due to cushioned chairs} = 100 \times 1.00 = 100$$

$$\text{Total absorption in a empty hall} = \sum_{as} = 119.16$$

Case I: For empty hall

$$\text{Rev. time} = T_1 = \frac{0.161V}{\sum_{as}} = \frac{0.161 \times 1450}{119.16}$$

$$T_1 = 1.959 \text{ sec}$$

Case II: Hall with 60 audience inside

$$\text{Rev. time} = T_2 = \frac{0.161V}{\sum_{as} + 60 \times 4.7} = \frac{0.161 \times 1450}{119.16 + 282}$$

$$T_2 = 0.582 \text{ sec}$$

Case III: Hall with full capacity

$$\text{Rev. time} = T_3 = \frac{0.161V}{\sum_{as} + 100 \times 4.7} = \frac{0.161 \times 1450}{119.16 + 470}$$

$$T_3 = 0.497 \text{ sec}$$

QUESTIONS

1. (a) Enumerate the requirement of reverberation chamber. How will you experimentally determine the coeff. of sound absorption of some absorbing material like window glass. Deduce the formula used?
 (b) Define the term reverberation time. Describe any one of the methods for the determination of the reverberation time?
2. (a) Write down Sabine's formula. Explain the terms involved in it and describe the units of each of them. State the limitations of the formula?
3. (a) Explain how reverberation of hall is affected by (i) its size (ii) nature of its wall surfaces (iii) audience.
 (b) Reverberation time is found to be 3 sec for an empty reverberation chamber and an acoustic sheet of 15 sq.m is suspended at the centre of the reverberation hall. Calculate the coeff. of sound absorption of acoustic sheet if the volume of the chamber 600 cu.m?
4. (a) Define the term coeff. of sound absorption and hence describe any one of the methods in detail to determine the coeff. of sound absorption formula used for it?
 (b) Explain the terms (i) Reflection, (ii) Reverberation, and (iii) Echo of sound energy and then show graphically only the nature of growth and decay of sound energy in a hall due to reverberation.
5. (a) Explain (i) Absorption coeff, (ii) Echelon effect, (iii) Extraneous noises.
 (b) A classroom has dimension $20 \times 15 \times 5$ cu.m. The Rev. time is 3.5 sec. Calculate the total absorption of its surface and the average absorption coeff.?
6. Explain in detail the acoustic demands of a hall?
7. (a) State and explain Sabine's formula for reverberation time?
 (b) State the acoustic requirements of a good auditorium. Explain how can these requirements be achieved?
8. Explain the term Coefficient of sound absorption and reverberation
9. Reverberation time for a cubical chamber 10 wide is 2.68 sec. Calculate its average absorption coeff. If one of the walls is covered with acoustic tiles the reverberation time will be 2.0 sec. Calculate the sound absorption coeff. of acoustic tiles.
10. A cinema hall has a volume of 7500 m^3 . It is required to have reverberation time of 1.5 sec. What would be the total absorption of the hall? What will be the effect if you increase the total absorption of the hall?

Semiconductors

6.1 Band-Theory and Formation of Energy Gap in Different Materials²

X-ray and other studies show, that most metals and semiconductors are crystalline in structure. A crystal consist of space arrays of atoms or molecules or ions. It is built up by regular repetition of some fundamental structural unit in three dimension. Now electronic energy levels which are well known for single free atoms, will no longer applicable in the same way to a crystal. In crystal structure the change in energy levels of the inner electrons will be small, since their closed shell electrons are not affected very much by the presence of neighbouring atoms. But the levels of outer valence electrons are greatly affected since these are the electrons which are participating in chemical bonds. The new energy levels of the outer electrons can be determined by means of quantum mechanics, and it is found that the coupling between the outer shell valence electrons of different atoms in a crystal which are having almost same energy, form a band of closed spaced energy states, instead of the widely separated energy levels of the isolated atoms as shown in Fig. 6.1.

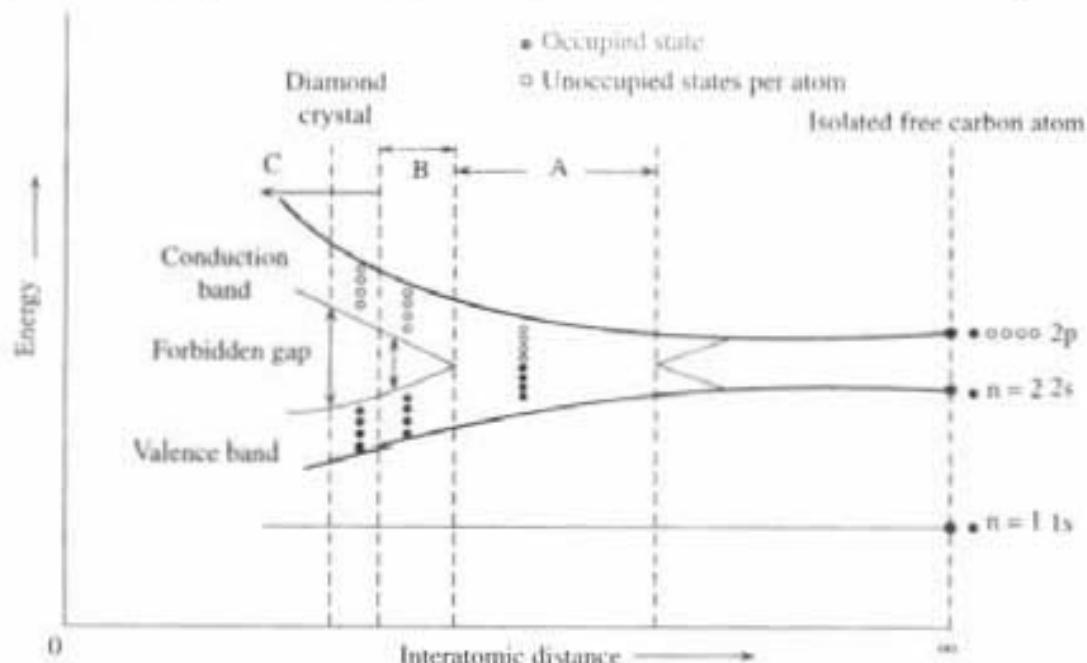


Fig. 6.1 Formation of valence band and conduction band. Region A-Metal, B-Semiconductor, C-Insulator

The width of the energy band depends upon the coupling between the outer electrons of the atoms, i.e. upon the distance between atoms in crystal. If the interatomic distance decrease, the coupling will increase and band width will also increase (Fig. 6.1). It is a theoretical curve, because it is not possible to vary continuously the interatomic distance in a crystal, as it is shown in Fig. 6.1.

Formation of the bands in different materials can be explained more clearly in the following way. In free carbon atom ($Z = 6 \rightarrow 1s^2, 2s^2, 2p^2$) innermost shell $1s$ is completely filled up. In $n = 2$ shell, out of 8 electronic states 4 are filled up and 4 are vacant. As inter atomic distance decreases these two outer discrete sub shells ($2s, 2p$) of $n = 2$, get broadened out and ultimately overlap with each other (region A, Fig. 6.1). In case of silicon (Si) ($Z = 14 \rightarrow 1s^2, 2s^2, 2p^6, 3s^2, 3p^2$) as inter atomic distance decreases $n = 1$ and $n = 2$ shell will be unaffected as they completely filled up but overlapping takes place in outermost $n = 3$ shell between $3s$ and $3p$. In $n = 3$ shell, out of 8 electronic states 4 filled up and 4 vacant.

Similarly for germanium (Ge) ($Z = 32 \rightarrow 1s^2, 2s^2, 2p^6, 3s^2, 3p^6, 3d^{10}, 4s^2, 4p^2$) $n = 1, n = 2, n = 3$ shells will be unaffected but the overlapping will occur in $n = 4$ shell between $4s$ and $4p$. Now in $n = 4$ shell, out of 8 electronic states 4 filled up and 4 vacant. So always overlapping takes place in the outermost sub-shells where as inner shells are unaffected in crystalline materials when interatomic distance is very small.

Now if the inter atomic distance is further decreased, a split occurs between the upper four vacant states and lower four filled up states in outermost shell (for $n = 2$ in carbon, $n = 3$ for Si, $n = 4$ level for Ge crystal and so on). *The upper part with four vacant states formed Conduction Band (C.B.),* this C.B. is generally empty for Insulator and partially filled up for Conductor. *The lower part with four filled up states formed Valence Band (V.B.).* This V.B. is band is either completely filled up or partially filled up. *The gap arises in between C.B. and V.B. where no electronic energy level are permitted is called Forbidden energy gap (E_g) (Fig. 6.1 and 6.2).*

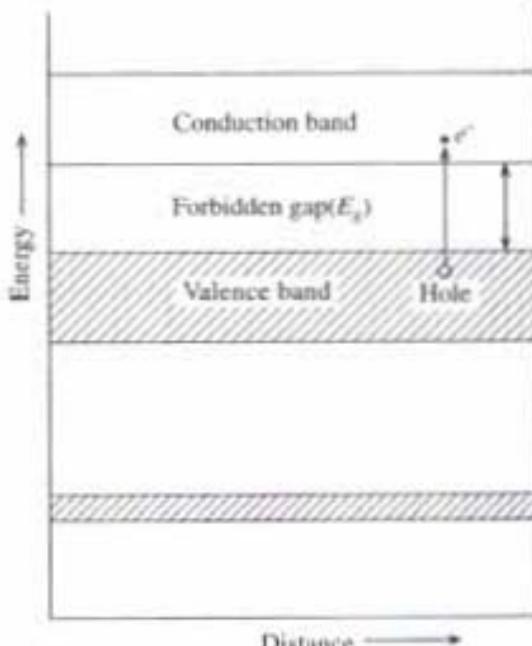


Fig. 6.2 Valence band (V.B.) and conduction band (C.B.) with forbidden gap E_g for a semiconductor

Now as the inter atomic distance decreases further, the forbidden energy gap (E_g) will gradually increase. When the forbidden energy gap ($E_g = 0$), which shows the overlapping region A in Fig. (6.1), and represents metal (eg. Cu, Al, Sb etc.). The region B in Fig. (6.1), where the forbidden energy gap is small (E_g around 1 eV) represent Semiconductor (eg. Si, Ge, GaAs etc.). The region C in Fig. (6.1), where the forbidden energy gap is very big (E_g around 6 ev) represent the Insulator (eg. diamond).

Four electrons from 2s, 2p or 3s, 3p or 4s, 4p in the valence band determine the valency of carbon, Si or Ge respectively and they are participating in covalent bonds. In diamond crystal the four valence electrons being shared with neighbouring atoms, are closely bound to the nuclei and so their energy level in V.B. is lower (more negative) than the energy level of the corresponding states of free carbon atoms (Fig. 6.1), similarly for Si and Ge.

Now if it is possible to remove one electron from the covalent bond in valence band of diamond crystal or in Si or Ge, then it may acquire sufficient energy to move into an energy state of the conduction band. Figure 6.2 shows the passage of an electron from the valence band to conduction band. The minimum energy that must be imparted to the electron to remove it from the covalent bond, is the energy difference between top of the valence band to bottom of the conduction band. This is the width of the forbidden gap 6 eV for diamond, which is quite large and bonding is also very strong. So it is not possible to remove an electron from V.B to C.B for diamond, due to which it acts as an *Insulator*. But if the electron can reach to the conduction band, as it happens for Si and Ge, it is free and can move within the material from atom to atom, because the conduction band is empty. The electrical and thermal properties of a material are due to movement of the electron in the conduction band, so in case of materials like Si ($E_g = 1.1$ eV), Ge ($E_g = 0.76$ eV), by breaking the covalent bond in V.B at room temperature some electrons can reach to the C.B and give rise to conductivity. Due to that they are called as *Semiconductor*.

6.2 Metals, Insulators and Semiconductors

A very excellent conductor of electricity is called metal. Poor conductor is an insulator and a substance whose conductivity lies in between insulators and metals is called semiconductor. A material may be placed in any one of these three classes, depending upon its energy-band structure.

Metals: All metals are generally very good conductor of electricity because band structure of a metal crystal contains no forbidden gap ($E_g = 0$), conduction and valence bands overlap with each other (Fig. 6.3(a), Region A in Fig. 6.1).

So under the influence of an external electric field, electrons are easily moved from valence band to conduction band and they are free to move about in the crystal from atom to atom and this constitutes the conduction current (J_c). With increase of temperature, conductivity of metals will decrease due to increase of resistance.

Insulators: The requirements for a crystalline substance to be an insulator at room temperature are, the valence band must be full and the conduction band must be normally empty and conduction band should be separated from valence band by a large forbidden energy gap ($E_g = 6$ eV) at room temperature (Fig. 6.3(b), Region C in Fig. 6.1). In diamond, the covalent bonding is so strong, that the energy which can be supplied to an electron from an external applied field becomes very small than that required to free the electrons from valence band. As there are no free electrons available in C.B., so conduction cannot be possible. But as the temperature increases some electrons may get enough energy to cross the large forbidden gap and take part in conduction. So conductivity of good insulator like diamond may increase with temperature, but the chances are very rare.

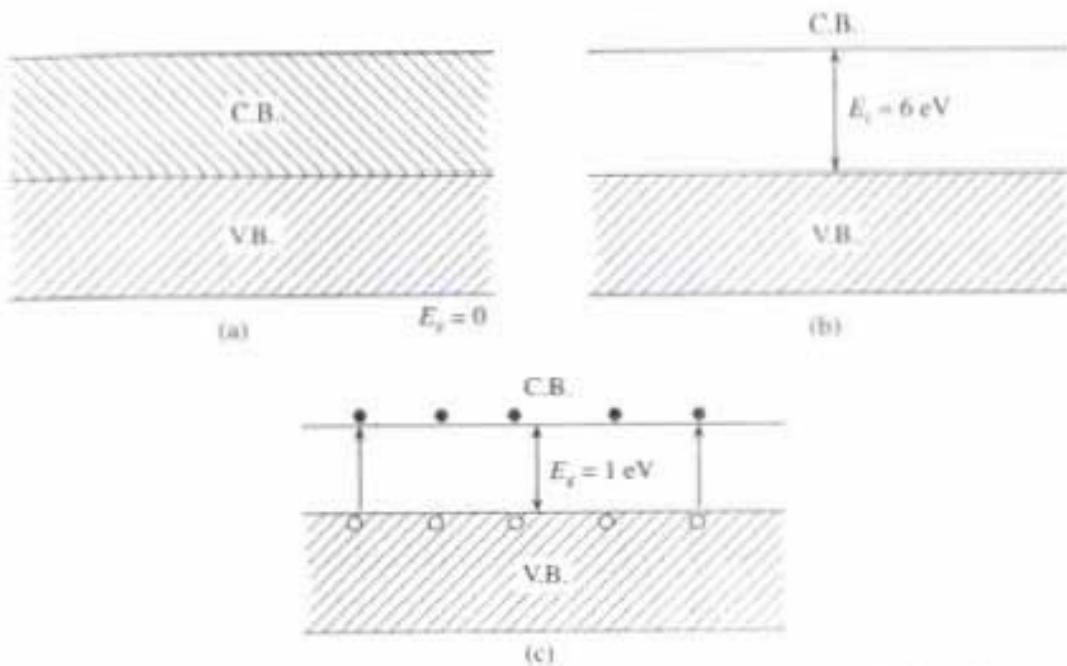


Fig. 6.3 Energy level diagram for, (a) metal; (b) insulator; (c) semiconductor

Semiconductor: Semiconductors have the same type of energy band structure as an insulator. But an insulator has a forbidden gap which is so wide (Region C, Fig. 6.1), that very few electrons can cross it at room temperature, whereas a semiconductor possesses a narrow energy gap (Fig. 6.3c, Region B in Fig. 6.1), which allows a considerable amount of conduction at room temperature. Graphite, a crystalline form of carbon but having a crystalline symmetry which is different from diamond, has such a small value of E_g , that it is almost a semiconductor. Germanium and silicon (Fig. 6.3c) are the two most widely used semiconducting elements having the forbidden gap as 0.76 and 1.1 eV respectively at room temperature 27°C (= 300°K). The conductive property of semiconductors lie between a conductor (like copper) and an insulator (like diamond). *The conductivity of semiconductors increases with increase of temperature but at 0°K semiconductors act as an insulator.* Because covalent bonds are very strong and the valence electrons of semiconductors (Fig. 6.4a) cannot break away the covalent bond without receiving the external energy, and since no free electrons are available

in C.B. at 0°K , so it behaves as an insulator. But as temperature increases more and more bonds will be broken and conductivity will increase with temperature for semiconductor.

6.3 Intrinsic Semiconductor

When the conductivity in crystalline semiconductors is only due to breaking of covalent bonds of the electrons in the valence band, then that substance is said to be an *intrinsic semiconductor* (Figs. 6.3c, 6.4).

6.4 Electrons and Positive Holes in a Semiconductor

Whenever a covalent bond of Si or Ge crystal is broken, then a free electron (e) moves from the valence band to the conduction band of a semiconductor, and it leaves behind an unfilled electronic state in the valence band (Figs. 6.2, 6.3c and 6.4b). *The absence of an electron in valence band is called a positive hole (p).* The word 'holes' in a semiconductor, therefore refers to the empty energy levels

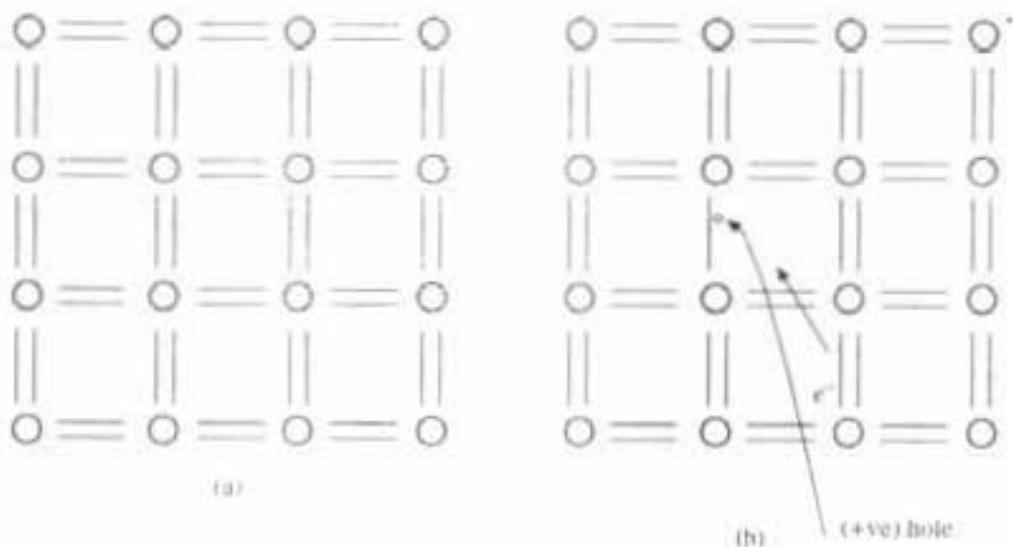


Fig. 6.4 Formation of holes in V.B. and movement of holes in V.B., which give rise to hole-current

in an otherwise filled valence band. So whenever there is an extra electron (n) in the conduction band there is a simultaneous production of a hole (p) in the valence band (Fig. 6.2, 6.3c). The number of electrons in C.B. and simultaneously, number of holes in V.B. will increase with increase of temperature. *But for intrinsic semiconductor number of electrons in C.B. (n) and no. of holes in V.B. (p) are always same.* So mathematically, we can say

$$n = p = n_i \quad (6.1)$$

where n_i is the number of charge carriers in intrinsic semiconductor. The positive hole, regarded as the active particle in the valence band, in the same way that the electron is considered to be active particle in conduction band. This leads to the concept that conductivity in semiconductors is caused not only by the motion of electrons in the conduction band which gives rise to the conduction current in C.B. but also by the motion of the positive hole which gives rise to hole current in the V.B. The movement of the hole in the valence band is understood from Figs. 6.4b and 6.4c. Hole positions can be changed from one place to other by interchanging their position with electrons (Figs. 6.4b and 6.4c). Thus positive holes are also physical entities whose movement constitute a flow of current, which is called "Hole Current". Since movement of hole is opposite to that of electron, so the direction of hole current is same as direction of current flow. So in semiconductors, there are two charge carriers, electrons and holes. But in metals electrons are the only charge carrier because there is no question of breaking the bonds, and thereby forming the holes.

6.5 Fermi Level and Effect of Temperature on Intrinsic Semiconductor

As we have seen earlier that at absolute zero all the electronic states of the valence band are full and those of the conduction band are empty, so a semiconductor becomes an insulator at 0°K. But as the temperature increases some electrons from the valence band get sufficient energy and become free. They move to the conduction band and take part in conduction and give rise to conductivity in the semiconductor (Fig. 6.3c).

Now we want to discuss the phenomenon with a quantum mechanical point of view. Classically all electrons have zero energy at 0°K, but quantum mechanically all electrons are not having zero energy at 0°K. The maximum energy that electrons may possess at 0°K is the Fermi energy E_F . So quantum mechanically the electrons actually have energies extending from 0 to E_F at absolute zero temperature (Fig. 6.5).

Now in order to know how many of the electronic energy states in the valence band and conduction band will be occupied at different temperatures, we introduce a *Fermi-factor*, $F(E)$, which is the number that expresses the 'Probability that a state of a given energy (E) is occupied by an electron under condition of thermal equilibrium. This number has a value between zero and unity and is a function of energy and temperature (T) as'.

$$F(E) = \frac{1}{1 + \exp [(E - E_F)/kT]} \quad (6.2)$$

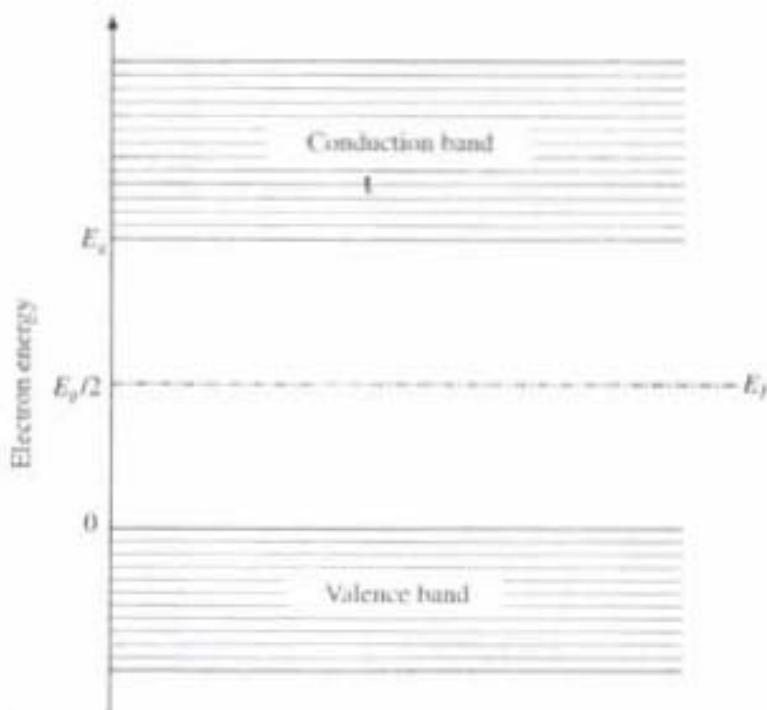


Fig. 6.5 Intrinsic semiconductor with fermi level $E_F = E_g/2$

where K = Boltzman constant,

T = temperature in $^{\circ}\text{K}$ and

E_F = fermi level of energy in eV, maximum energy that electron may possess at 0°K .

The Fermi factor is independent of the energy density of states, it is the probability that the states occupied at that level, irrespective of the number of states actually present. That is, it is the fractional occupancy of possible states.

The variation of Fermi factor $F(E)$ with energy at different temperature for intrinsic semiconductor is as shown in Fig. 6.6, where full line curve shows $F(E)$ at 0°K and dotted line curve at $T^{\circ}\text{K}$.

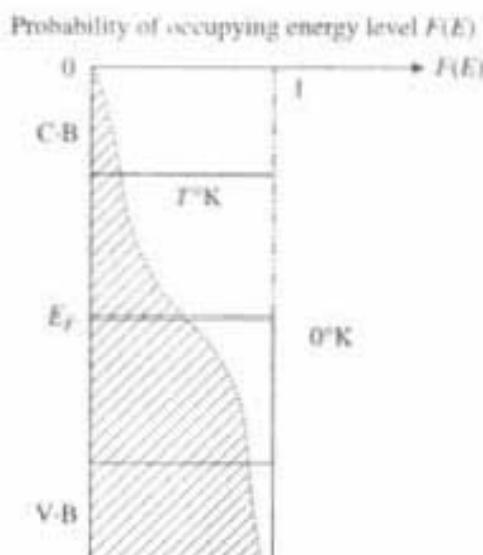


Fig. 6.6 Variation of fermi factor $F(E)$ with temperature

When $T = 0^\circ\text{K}$ $F(E)$ can take two values

$$(1) \quad E > E_F$$

$$F(E) = \frac{1}{1 + \exp(\infty)} = 0, \text{ (because } e^{\infty} = \infty)$$

$$(2) \quad E < E_F$$

$$F(E) = \frac{1}{1 + \exp(-\infty)} = 1 \text{ (because } \exp(-\infty) = 0)$$

Thus at $T = 0^\circ\text{K}$, $F(E) = 1$, when $E < E_F$, which means all the levels below E_F i.e. V.B. Fig. (6.5, 6.6) are filled up by electron and $F(E) = 0$, when $E > E_F$ i.e. all level above E_F i.e. C.B. Fig. (6.5, 6.6) are empty. Because at 0°K no heat energy is present, so no covalent bonds are being broken and the semiconductor behaves as an insulator.

But when $t = T^\circ\text{K}$ then at $E = E_F$, $F(E)$ becomes

$$F(E) = \frac{1}{1 + \exp(0)} = \frac{1}{2} \quad (\because e^0 = 1)$$

which means that when the temperature is not 0°K but some higher value say $T = 1000^\circ\text{K}$, then some covalent bonds will be broken and some electrons will be available in C.B. whereas some electron vacancy, i.e. hole, will be available at V.B. This is shown by the dotted line in Fig. 6.6.

6.6 Fermi Level in an Intrinsic Semiconductor

The position of Fermi level in an intrinsic semiconductor can be calculated in the following way: It should be noted that the Fermi level will be somewhere in the forbidden gap, since nearly all the valance band states are occupied, and very few electrons will be found in the conduction band, at room temperature.

Now the total number of occupied states [$N(E)$] by electrons, with energy between E and $(E + dE)$ is equal to the product of the available number of states [$S(E)$] and the probability of their occupancy [$F(E)$], that is,

$$N(E)dE = S(E)F(E)dE \quad (6.3)$$

where $N(E)$ = total number of occupied states in C.B. by electrons, that is the density of electron in C.B. $S(E)$ = Number of available states and $F(E)$ = Probability of occupancy of those states.

Similarly the number of holes in the V.B. within the energy E and $(E + dE)$ is

$$P(E)dE = S(E)[1 - F(E)]dE \quad (6.4)$$

Consider energy level E_1 in C.B. and E_2 in V.B. which are symmetrically placed about the centre of the energy gap (E_g) (Fig. 6.7). For an energy range dE , the number of electrons in the C.B. is

$$N(E_1)dE = S(E_1)F(E_1)dE$$

Similarly, the number of holes in V.B. is

$$P(E_2)dE = S(E_2)[1 - F(E_2)]dE$$

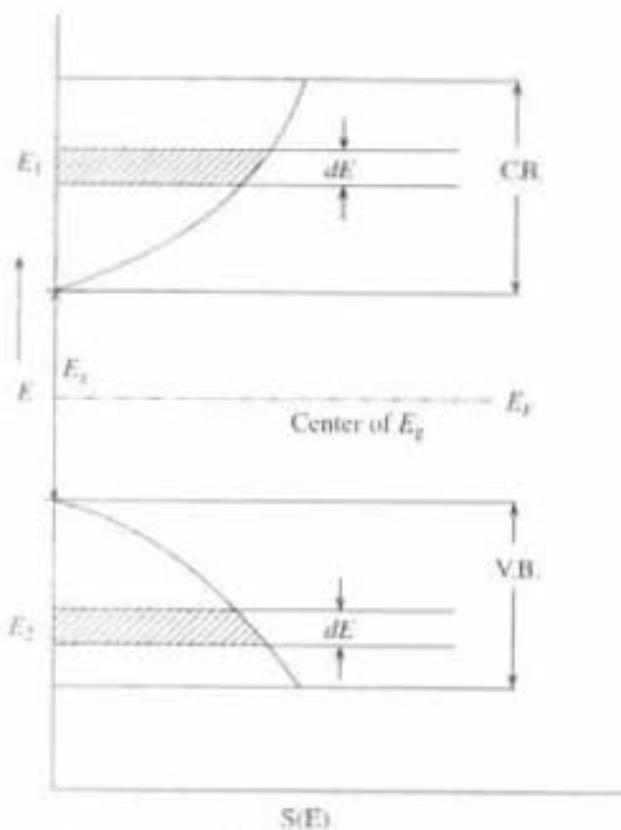


Fig. 6.7 Variation of available number of states $[S(E)]$ with energy (E) in valence and conduction band.

Assuming for simplicity $S(E_1) = S(E_2)$, the number of available states in C.B. and V.B. are almost same. Hence,

$$\frac{N(E_1)}{P(E_2)} = \frac{F(E_1)}{[1 - F(E_2)]} \quad (6.5)$$

Under certain condition, the following approximation to the Fermi factor can be made. Assuming that, at room temperature energy level E_1 in C.B. is far above the Fermi level (E_F), say at 300°K , then the factor

$$e^{\frac{E_1 - E_F}{KT}} \gg 1$$

So $F(E_1) = \frac{1}{1 + e^{(E_1 - E_F)/KT}} = e^{-(E_1 - E_F)/KT}$

or $F(E_1) = e^{(E_F - E_1)/KT} \quad (6.6)$

Similarly assuming E_2 level in V.B. which is far below from Fermi level, then

$$1 - F(E_2) = 1 - \frac{1}{1 + e^{(E_2 - E_F)/KT}}$$

where $E_F - E_2 \gg KT$ at room temp. so the above expression becomes

$$[1 - F(E_2)] \approx 1 - \left[1 - \exp \frac{(E_2 - E_F)}{KT} \right]$$

$$[1 - F(E_2)] = \exp \frac{(E_2 - E_F)}{KT} \quad (6.7)$$

Thus, the above expression (6.5) becomes

$$\frac{N(E_1)}{P(E_2)} = \frac{\exp [(E_F - E_1)/KT]}{\exp [(E_2 - E_F)/KT]} \quad (6.8)$$

by assuming $E_1 - E_F \gg KT$ and $E_F - E_2 \gg KT$ at room temperature.

For intrinsic semiconductor the number of electrons is equal to number of holes (Eqn. 6.1) i.e.,

$$n_i = p_i$$

So,

$$\frac{N(E_1)}{P(E_2)} = \frac{n_i}{p_i} = 1$$

$$\text{Hence from Eqn. (6.8)} \quad \frac{E_F - E_1}{KT} = \frac{E_2 - E_F}{KT}$$

or

$$E_F = \frac{E_1 + E_2}{2} \quad (6.9)$$

where E_1 and E_2 were taken symmetrically about the centre of the forbidden gap (E_g). Above equation shows that the Fermi level lies at the centre of the forbidden gap for intrinsic semiconductor (Figs. 6.5 and 6.7) and it is independent of the temperature.

6.7 Extrinsic Semiconductor

The electrical conductivity of intrinsic semiconductor is very small. To increase the conductivity of intrinsic semiconductor a small percentage of trivalent or pentavalent atoms are added to the pure semiconductor in the process of crystallization which is called doping and results the impure semiconductor, being called as extrinsic semiconductor. The conductivity of extrinsic semiconductor is much higher, say for example 12 times, than intrinsic semiconductor when an impurity is added 1 part in 10^6 . The impurity atom has a size which is almost of the same order of the host lattice. Since percentage of impurity atoms is very small so every impurity atom is surrounded by a normal lattice site. So basic structure of the crystal will not get altered after doping. There are two types of extrinsic semiconductors: (1) *n*-type semiconductor and (2) *p*-type semiconductor.

(1) *n*-type semiconductor

If a pentavalent atom is introduced in the intrinsic semiconductor say germanium, the situation will be as shown in Fig. 6.8(a). Then impurity pentavalent atoms, say antimony ($Z = 51$) or phosphorous ($Z = 15$) or arsenic ($Z = 33$), will displace some of the germanium atoms in the original crystal lattice and four of the five valence electron of the impurity atom will occupy four covalent bonds and the fifth one will be almost free, as it does not find a place in the covalent bond.

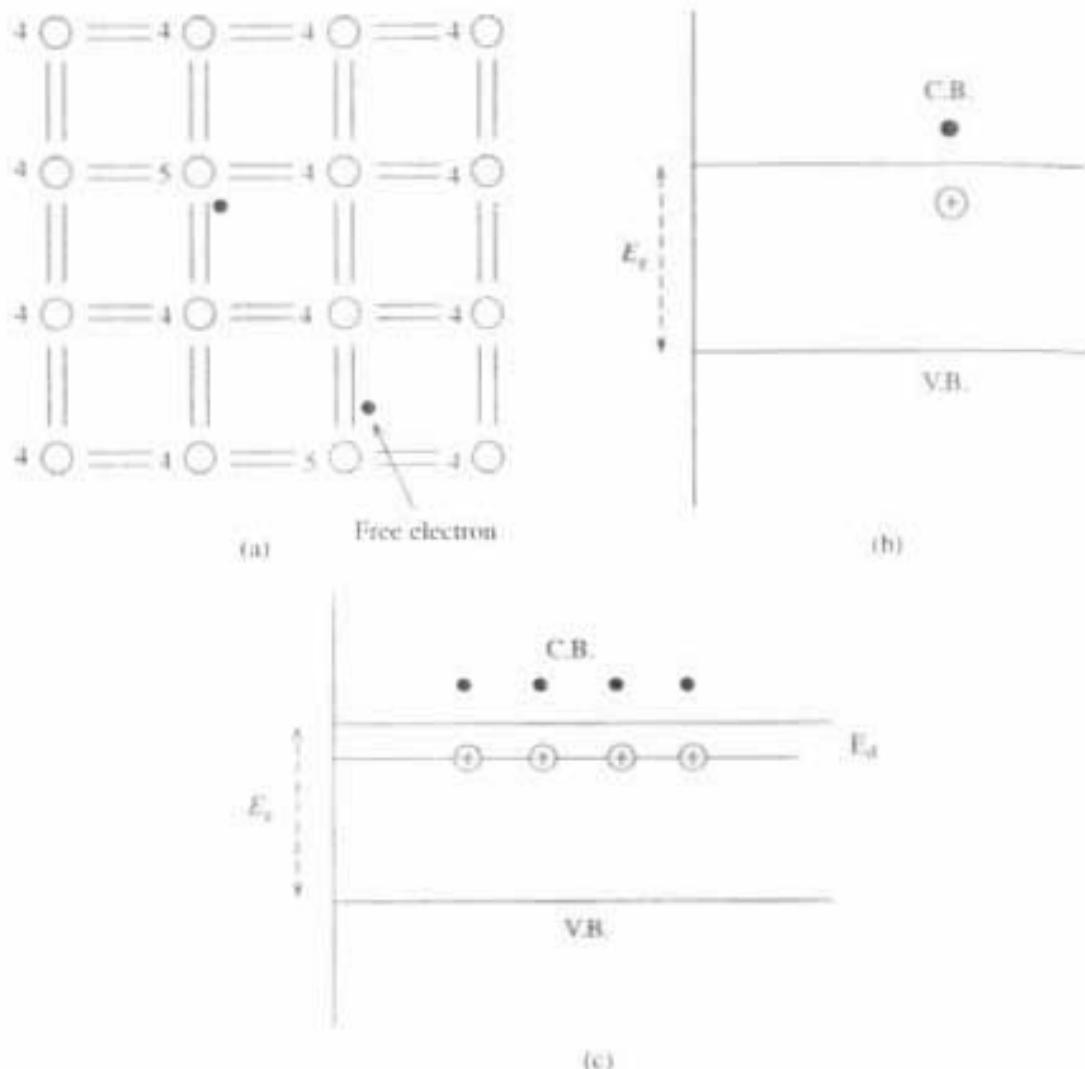


Fig. 6.8 *n*-type semiconductor with donor impurity and donor level (E_d)

Every impurity atom thus contributes one almost free electron without creating a hole, and those free electrons act as carrier of current. A very small amount of energy is required to free the fifth electron from the atom, of the order of 0.01 eV for Ge and 0.05 eV for Si. Since those pentavalent impurities donate excess electron carrier, so they are called donor or *n*-type impurity, and the crystal doped with donor impurity is called *n*-type semiconductor. Each ionized donor atom has a net charge of (+e) and is bound or immobile (Fig. 6.8b).

When the donor impurities are doped to an intrinsic semiconductor Ge or Si, then additional discrete levels are introduced just below the conduction band in the forbidden gap (Fig. 6.8c). These new additional levels will be discrete and known as donor level E_d (Fig. 6.8c) because the additional impurity atoms are situated far apart in the crystal structure and hence their interaction is small. Since very small amount of energy (0.01 eV for Ge and 0.05 eV for Si) is required to free the electron from donor level to conduction band, therefore almost all the fifth electron of the donor material are raised to the conduction band at room temperature, leaving positively ionised donor atoms on the donor level.

(2) *p*-type semiconductor

When a trivalent impurity (boron $Z = 65$, gallium $Z = 31$, or indium $Z = 49$) is added to the intrinsic semiconductor, only three of the covalent bonds can be filled up and the vacancy that exists in the fourth bond constitutes a hole (Fig. 6.9a) without making an electron free. These holes can accept electrons from V.B. and thus make available positive carriers hole in the valence band. So these trivalent impurities are called acceptor because every trivalent atom accept electron, intrinsic semiconductor with acceptor impurity is called *p*-type semiconductor. Each acceptor atom has a net charge ($-e$) and is bound or immobile (Fig. 6.9b). When acceptor impurity is doped to an intrinsic semiconductor Ge or Si, then the additional discrete energy levels are introduced just above the valence band Fig. (6.9c) in the forbidden gap, which is called acceptor level (E_a). Since a very small amount of energy is necessary for an electron to leave the valence band and occupy the acceptor energy level, so almost all the vacancies of the acceptor material are filled up by electrons from V.B. and formed negatively ionised acceptor atom. It thus creates a larger number of holes in the valence band which constitute the largest number of hole carrier in the valence band in the extrinsic semiconductor.

Thus by doping the impurity atoms in the intrinsic semiconductor not only conductivity is increased but it also serves to produce a semiconductor in which

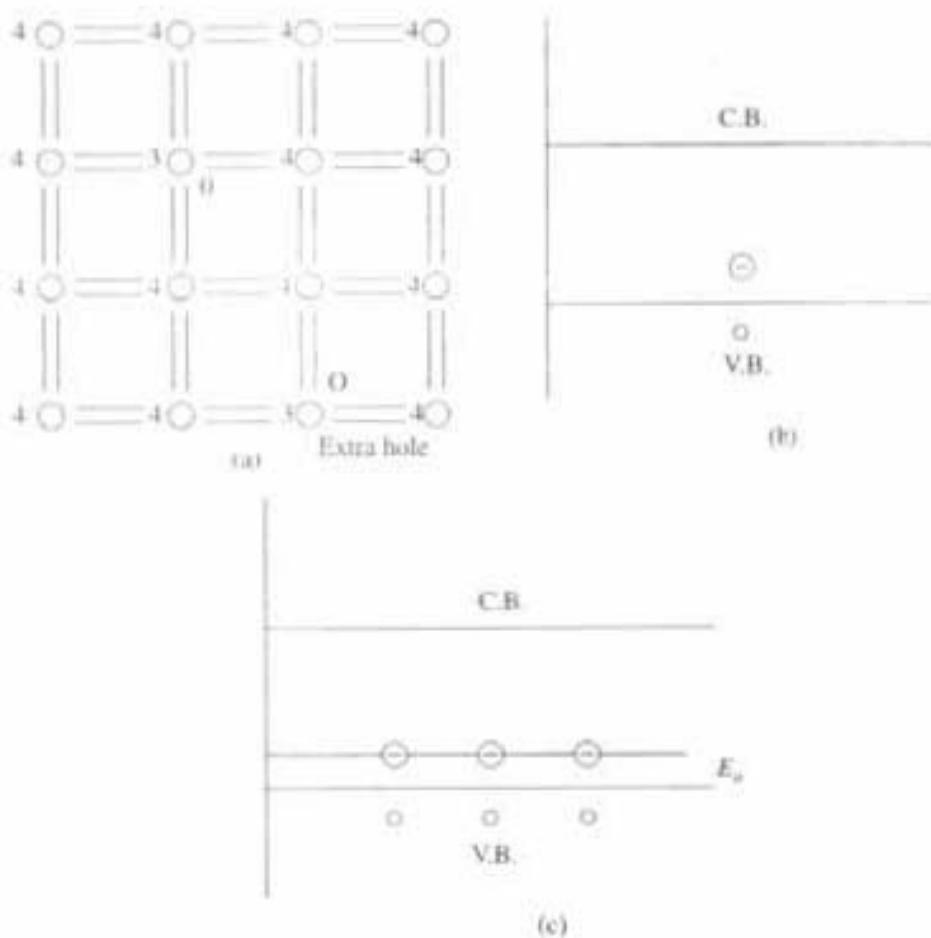


Fig. 6.9 *p*-type semiconductor with acceptor impurity and acceptor level (E_a)

the charge carriers, are mainly electrons or holes. Electrons are majority carriers and holes are minority carriers in *n*-type semiconductor whereas in *p*-type semiconductor the majority carrier are holes and minority carriers are electrons.

6.8 Fermi Level in an Extrinsic Semiconductor

In an intrinsic semiconductor the number of electrons is equal to number of holes ($n_i = p_i$). But in *n*-type extrinsic semiconductor number of electrons are increased due to the doping of pentavalent atom ($n_e > n_i$) and number of holes

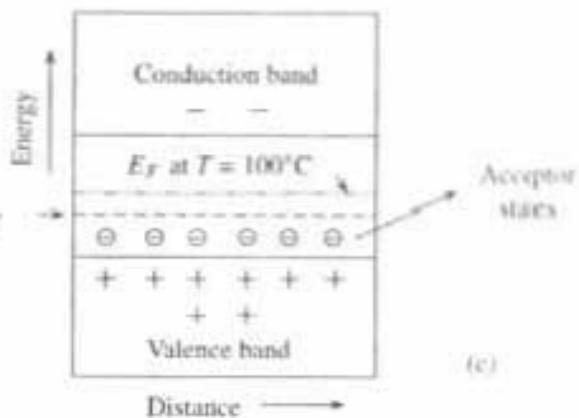
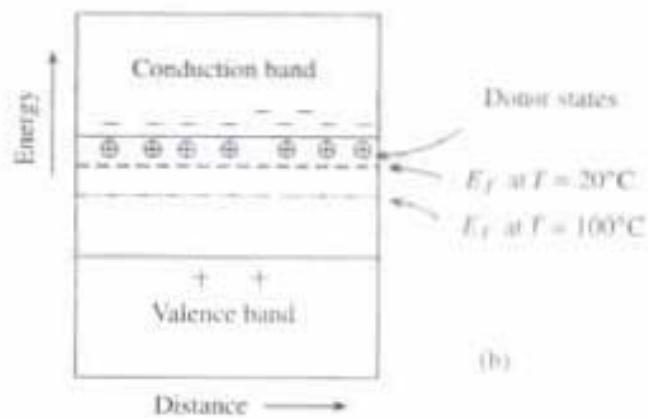
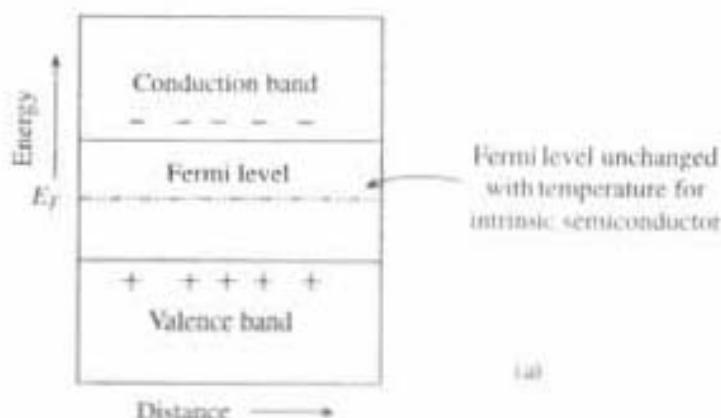


Fig. 6.10 Position of fermi level (E_F) for (a) intrinsic; (b) *n*-type and (c) *p*-type semiconductor at different temperatures

are decreased ($p_e < p_i$) than which would be available for intrinsic semiconductor. The number of holes are decreased because of the large number of electrons present which increase the rate of recombination of electrons with holes. With similar argument it can be shown that in *p*-type extrinsic semiconductor the number of holes are increased due to the doping of trivalent impurity atom ($p_e > p_i$) and number of electrons decreased ($n_e < n_i$) than which would be available for intrinsic semiconductor. Now since the Fermi level (E_F) is a measure of the probability of occupancy of the allowed energy states by the electron, so for intrinsic semiconductor when $n_i = p_i$, Fermi level (E_F) is at the center of the forbidden gap (Fig. 6.7). But for *n*-type semiconductor, since $n_e > p_i$, it is clear that E_F must move upward closer to conduction band (Fig. 6.10b), to indicate that many of the energy states in that band are filled by the donor electrons, and a few holes exist in the valence band. Similarly, for *p*-type semiconductor since, $p_e > n_i$, so E_F must move downward from the center of the forbidden gap closer to the valence band (Fig. 6.10c).

6.9 Variation of Fermi Level with Temperature in Extrinsic Semiconductor

For an intrinsic semiconductor $n_i = p_i$ and as temperature increases both n_i and p_i will increase. Thus, the fermi level (E_F) will remain approximately at the center of the forbidden gap. Thus for intrinsic semiconductor Fermi level is independent of temperature (Fig. 6.10a).

But in an extrinsic semiconductor it is different. Say for *n*-type material the electrons are coming from two different sources. Some electrons come from donor state which are easily separated from their parent atom and they do not vary much as the temperature is increased. The other electrons in the conduction band are present because of the breaking of a covalent bond. Such intrinsic effect by breaking the bonds will increase in number as temperature is raised. So with increase of temperature proportion of the total number of conduction electron from valence band by breaking the bonds will increase. Thus *as the temperature rises the material becomes more and more intrinsic and Fermi level moves down closer to the intrinsic position, i.e., at the centre of the forbidden gap* (Fig. 6.10b).

Similarly *for p-type semiconductor as temperature increases the material also becomes increasingly intrinsic and the Fermi level rises up until it approaches the center of the gap as in the intrinsic semiconductor* (Fig. 6.10c). Thus both *n*-type and *p*-type material become more and more like intrinsic material at high temperature. This places a limit on the operating temperature of an extrinsic semiconductor device.

6.10 Electrical Conductivity in Metallic Conductor

All conductors containing a periodic three dimensional array of heavy tightly bound (+ve) ions surrounded by electron clouds, which can move about quite freely due to thermal agitation. But the average velocity of the electron clouds due to thermal agitation will be zero in the absence of the external E-field.

Now according to Coulomb's law in the presence of an external E-field electrons will experience an uniform force $\vec{F} = e\vec{E}$ Newton, in the direction

opposite to the direction of the E -field which accelerates the electrons. But at the same time due to the collision of the electrons with (+ve) ions, they loose their energy and ultimately electrons will acquire a steady state of average drift velocity $\langle \bar{v}_d \rangle$. Average drift velocity is directly related to the E -field as

$$\langle \bar{v}_d \rangle = -\mu_e \bar{E} \text{ m/sec} \quad (6.10)$$

where μ_e is called the mobility of the electron in a given material whose unit is $\text{m}^2/\text{volts.sec}$. Average drift velocity $\langle \bar{v}_d \rangle$ of electron is in a direction opposite to the direction of E -field (\bar{E}), so (-ve) sign has come in Eqn. (6.10). $(\mu_e)_{Al} = 0.0012$, $(\mu_e)_{Cu} = 0.0032 \text{ m}^2/\text{volt.sec}$. The drift of the electrons in a conductor due to external E -field cause an electron flow which is called as Conduction current. Conduction current density (\bar{J}_c) or drift current density (\bar{J}_{dr}) of negatively charged electrons can be written as

$$\bar{J}_{dr} = \bar{J}_c = -ne \langle \bar{v}_d \rangle = -\rho_e \langle \bar{v}_d \rangle \text{ Amp.} \quad (6.11)$$

where ρ_e ($= ne$, n is the electron density) is a free-electron charge density of negative value. Now by substituting the value of $\langle \bar{v}_d \rangle$ from above eqn. (6.10), conduction current density becomes

$$\bar{J}_{dr} = \bar{J}_c = +\rho_e \mu_e \bar{E} = \sigma \bar{E} \text{ Amp.} \quad (6.12)$$

where $\sigma = \rho_e \mu_e$ is called the conductivity of the material whose unit is mho/m which is 3.82×10^7 for aluminum and 5.80×10^7 for copper. Above equation (6.12) is called the point form of Ohm's Law, which states that current density (J_c) will be constant so long E -field (\bar{E}) is constant.

The conductivity is a function of temperature. Resistivity which is inverse of conductivity increases linearly with temperature and decreases also linearly with temperature. But for some material resistivity drops suddenly to zero at a particular temperature of few Kelvin, which are called as *super conductor*. Aluminum become superconductor below $1.14^\circ \text{ Kelvin}$.

Higher temperature cause crystalline lattice vibration higher, which cause the higher collision rate of the electrons with ions (lattice), due to which drift velocity will decrease. So for a given E -field strength-lower drift velocity $\langle \bar{v}_d \rangle$, lower mobility (μ_e), results the conductivity (σ) lower at higher temperature.

From the point form of Ohm's Law eqn. (6.12), the well known Ohm's law in circuit theory can be derived in the following way. Lets consider a conductor as in Fig. (6.11a), with area A and length L . The applied voltage between two ends of the conductor is $V = EL$

i.e.

$$J_c = \sigma E = \sigma V/L \quad (6.13)$$

Now the conduction current through this conductor is

$$I_c = J_c A = \sigma V A / L$$

$$\text{i.e., } V = I_c L / \sigma A \quad (6.14)$$

where as well known Ohm's Law states that



Fig. 6.11a Current flow in a conductor

$$V = I_c R \quad (6.15)$$

By comparing above two equations (6.14) and (6.15) we get

$$\text{Resistance } (R) = L/\sigma A \text{ ohms} \quad (6.16)$$

6.11 Relaxation Time

Steady State

In the presence of external D.C. source electrons in a conductor will experience an acceleration, directed opposite to the applied E -field, which is expressed as $a = -eE/m$. So the rate of increase of velocity with time i.e. acceleration due to applied E -field (\bar{E}) can be written as

$$\left[\frac{d}{dt} v \right]_{\text{field}} = -eE/m \quad (6.17)$$

where v is the velocity of the electron under E -field (\bar{E}). Due to this acceleration conduction current density ($J_c = -nev$) will also increase in a particular value of E -field (\bar{E}), which is contradicts the steady state condition or the point form of the Ohm's Law Eqn (6.12) ($J_c = \sigma \bar{E}$), which states that current density (J_c) remains constant as long as E -field (\bar{E}) remains constant.

The above contradiction will go in the following way. The electron acceleration in the presence of external E -field (\bar{E}) will counter balance by the collision of the electron with the lattice (+ve) ion core and ultimately electrons will acquire an average drift velocity $\langle v_d \rangle$ as we have discussed earlier. The average distance traveled by the electron between two collision with lattice core is called the *Mean free path* (λ), and the time taken between two collision is called *Relaxation Time* (τ).

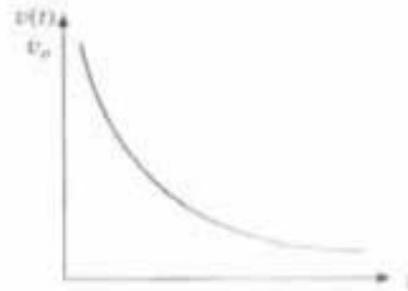


Fig. (6.11b) Change of vel. with time

Now in the absence of the external E -field (\bar{E}) the velocity of the electron at ($t = 0$) say (v_0). As the time passes due to the collision with (+ve) ion core velocity of the electron will decrease with time. So at any instant of time velocity $v(t)$, of the electron can be written as

$$v(t) = v_0 e^{-vt} \quad (6.18)$$

which is shown graphically in Fig. (6.11b).

The rate of decrease of velocity with time due to collision is

$$\left[\frac{d}{dt} v \right]_{\text{collision}} = -\frac{1}{\tau} v \quad (6.19)$$

So from above two Eqns. (6.17 and 6.19) in steady state of current flow with average drift velocity $\langle \bar{v}_d \rangle$ in a conductor we get

$$\left[\frac{d}{dt} v \right]_{\text{field}} + \left[\frac{d}{dt} v \right]_{\text{collision}} = \frac{d}{dt} \langle \bar{v}_d \rangle = 0 \quad (6.20)$$

$$\text{or } eE/m = -\frac{1}{\tau} \langle \bar{v}_d \rangle = 0 \text{ i.e. } \langle \bar{v}_d \rangle = -\frac{(e\tau)}{m} E = -\mu_e E \quad (6.21)$$

where $\left(\mu_e = \frac{e\tau}{m} \right)$ is the mobility of the electron, as it was given in Eqn. (6.21) whose unit is $m^2/\text{volt.sec}$. Now by substituting the value of $\langle \bar{v}_d \rangle$ in the above Eqn. (6.11), we get the expression of conduction current density (J_c) in metal as follows

$$J_c = -ne \langle \bar{v}_d \rangle = \frac{ne^2 \tau}{m} E = ne \mu_e E = \sigma E \text{ Amp/m}^2$$

$$\text{i.e. } \text{Conductivity of metal } (\sigma = 1/\rho) = \frac{ne^2 \tau}{m} = ne \mu_e \text{ mho/m} \quad (6.22)$$

6.12 Conductivity in Intrinsic-Semiconductor

In semiconductor since there are two charge carriers electron and hole, so in semiconductor conductivity and mobility depends upon temperature, purity of the material and on the fact whether electron or hole motion takes place. For semiconductor (μ_n) is the mobility of electron in conduction band, where as (μ_p) is the mobility of hole in valance band. For semiconductor mobility of electron (μ_n) is greater than mobility of hole (μ_p) because electrons in the C.B. is more free than the holes in the V.B.

Lets consider the semiconductor having an electron density in C.B. is (n) per unit volume and the charges which are available for conduction is (ne), and these charges will move with average drift velocity $\langle v_d \rangle$ under the E -field (E_s). So under this situation conduction current density will be $J_c = -ne \langle v_d \rangle$

Now if the conductivity of the material due to electron in C.B. is (σ_n), then from Ohm's law conduction current density will be

$$J_c = \sigma_n E_s = -ne \langle v_d \rangle \quad (6.23)$$

So from above eqn. (6.23) and by using eqn. (6.21) we get

$$\sigma_n = J_c/E_s = -ne \langle v_d \rangle / E_s = ne \mu_n = ne^2 \tau_n / m_n \quad (6.24a)$$

where m_n is the mass associated with electron in C.B. and τ_n is the relaxation time for electron in C.B. Similarly conductivity of hole in V.B. will be

$$\sigma_p = J_c/E_s = pe \langle v_d \rangle / E_s = pe \mu_p = pe^2 \tau_p / m_p \quad (6.24b)$$

where m_p is the mass associated with hole in V.B. and τ_p is the relaxation time for hole in V.B. Since in semiconductor both electron and hole are taking part in conduction in C.B. and V.B. respectively, so the total conductivity will be

$$\sigma = \sigma_n + \sigma_p = e(n\mu_n + p\mu_p) = e^2 \left(\frac{n\tau_n}{m_n} + \frac{p\tau_p}{m_p} \right)$$

For intrinsic semiconductor number of electron in C.B. is equal to the number of hole in V.B. i.e. $n = p = n_i$

So conductivity for *intrinsic semiconductor* (σ_i) will be

$$\sigma_i = n_i e(\mu_n + \mu_p) = n_i e^2 \left(\frac{\tau_n}{m_n} + \frac{\tau_p}{m_p} \right) \quad (6.25)$$

Effect of Temperature on Conductivity for Intrinsic semiconductor and calculation of Energy Band Gap (E_g):

For intrinsic semiconductor number of electron and number of holes are very sensitive to the temperature. Because as temperature increases more numbers of covalent bonds will be broken and production of free electron and hole pairs will also increase, which will increase the conductivity of the intrinsic semiconductor. It can be shown that for any intrinsic semiconductor the intrinsic carrier concentration (n_i) and hence conductivity (σ) increases with temperature (T) as follows:

$$n_i = AT^{3/2} e^{-E_g/2KT}, \quad (6.26)$$

where $A = \left[\frac{2\pi m}{h^2} \right]^{3/2} \left[\frac{m_e m_p}{m^2} \right]^{3/4}$, m = mass of electron in free space, K = Boltzman constant, T = Temperature in °K and E_g = Forbidden energy gap in intrinsic semiconductor

The conductivity of intrinsic semiconductor Ge, Si is found to be increases approximately 6 to 8% per degree increase of temperature in the following way

$$\sigma_T = \sigma_0 e^{-E_g/2KT} \quad (6.27a)$$

whereas resistivity will decrease with the temperature as

$$\rho_T = \rho_0 e^{E_g/2KT} \quad (6.27b)$$

This property is used in *Thermistor*, as it was already discussed in section 2.11 in Part 1 which has extensive application in thermometers, in the measurement of micro-wave frequency power as a thermal relay, and in control devices actuated by changes in temperature.

Resistance of the semiconductor (thermistor) at T °K from above Eqn. (6.27b) and (2.17) from Part 1 can be written as

$$R_T = R_{T_0} e^{E_g/2K} \left(\frac{1}{T} - \frac{1}{T_0} \right) \quad (6.28a)$$

where R_{T_0} = Resistance at T_0 °K and σ_0, τ_0 are some constant.

$$\text{Now } \ln R_T = \ln R_{T_0} + E_g/2K \left[\frac{1}{T} - \frac{1}{T_0} \right] \quad (6.28b)$$

which is a equation of a straight line (Fig. 6.12) having slope $E_g/2K$. So one can calculate Energy Band Gap (E_g) for an intrinsic semiconductor like Si or Ge from the slope of the straight line as drawn between $\ln R_T$ and $1/T$, in Fig. 6.12.

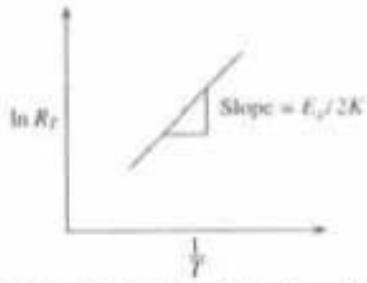


Fig. 6.12 Variation of $\ln R_T$ with $\frac{1}{T}$

6.13 Conductivity in Extrinsic Semiconductor

For extrinsic semiconductor according to the doping one carrier predominates over other at room temperature and that will become the majority charge carrier.

Consider a semiconductor which is doped by trivalent as well as pentavalent atom. The density of trivalent atom is (N_a) and that of pentavalent atom is (N_d). Where as the density of electron in C.B. is (n) and density of hole in V.B. is (p). Now after donate extra electron to the C.B. each pentavalent atom become (+ve) donor ion where as by accepting electron from V.B. each trivalent atom become (-ve) acceptor ion. But as a whole crystal must be electrically neutral. So from the charge neutrality condition it can be written as

$$e(n + N_a) = e(p + N_d) \quad \text{i.e.} \quad n + N_a = p + N_d \quad (6.29)$$

Since for intrinsic semiconductor $n = p = n_i$, or $np = n_i^2$

$$\text{i.e.} \quad n = n_i^2/p \quad \text{and} \quad p = n_i^2/n$$

So for p type semiconductor above eqn. (6.29) becomes $\frac{n_i^2}{p} + N_a = p + N_d$

which is a quadratic equation

$$\text{which gives} \quad p = -\frac{(N_d - N_a) + \sqrt{(N_d - N_a)^2 + 4n_i^2}}{2}$$

$$\text{similarly for } n\text{-type semiconductor} \quad n = \frac{(N_d - N_a) + \sqrt{(N_d - N_a)^2 + 4n_i^2}}{2}$$

In the above quadratic equation (-ve) sign hasn't taken because n and p should be (+ve).

But when % of doping of donor atom is more i.e. $N_d \gg N_a$ and $(N_d - N_a) \gg 2n_i$ then

$$n = N_d - N_a \quad \text{and} \quad p = n_i^2/(N_d - N_a) \quad (6.30)$$

similarly when % of doping of acceptor atom is more i.e. $N_a \gg N_d$ and $(N_a - N_d) \gg 2n_i$ then

$$p = N_a - N_d \quad \text{and} \quad n = n_i^2/(N_a - N_d) \quad (6.31)$$

in such materials *majority concentration* approximately equal to $1/N_d - N_a$, which is *temperature independent*, whereas *minority concentration varies with (ii)* and *changes rapidly with temperature*. When $N_a = N_d$, then the material will become again intrinsic.

For n -type semiconductor $N_a = 0$ and $N_d \gg 2n_i$ then $n = N_d$ and $p = n_i^2/N_d$
 For p -type semiconductor $N_d = 0$ and $N_a \gg 2n_i$ then $p = N_a$ and $n = n_i^2/N_a$
 So the conductivity for n -type extrinsic semiconductor will be

$$\sigma_n = ne \mu_n + pe \mu_p = N_d e \mu_n = \frac{ne^2 \tau_n}{m_n} \quad (6.32)$$

Similarly the conductivity for *p*-type extrinsic semiconductor will be

$$\sigma_p = ne \mu_n + pe \mu_p = N_a e \mu_p = \frac{pe^2 \tau_p}{m_p} \quad (6.33)$$

6.14 Hall Effect

In semiconductor physics it is often necessary to know whether a semiconductor material is *p*-type or *n*-type? Measurement of conductivity cannot give the answer, because it can not possible to distinguish between (+ve) hole and (-ve) electron conduction. *But Hall effect help us to determine whether a specimen is p-type or n-type. It can distinguish between two types of carrier as well as it helps to determine the density of two different charge carriers.*

In 1897 Hall discovered that when a current (*I*) was passed through a slab of a semiconducting material in the presence of transverse magnetic field, a small potential difference was established in a direction perpendicular to the direction of current flow as well as to the magnetic field.

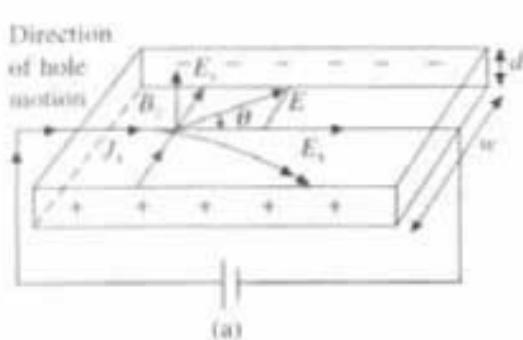


Fig. 6.13(a) Hall effect in *P*-type semiconductor

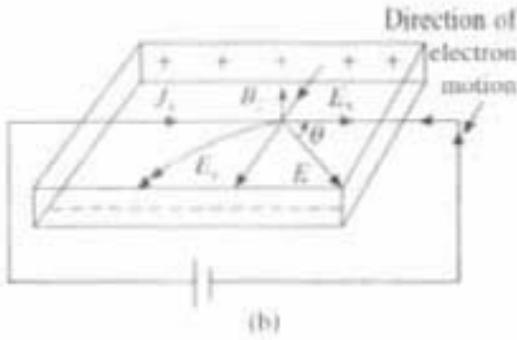


Fig. 6.13(b) Hall effect in *N*-type semiconductor

Let us consider a *P*-type semiconductor as shown in the Fig. 6.13(a). Since in the *p*-type semiconductor charge carrier is (+ve) hole, so the direction of hole flow is same as direction of current density (\vec{J}_x), which is produced in the semiconductor bar due to the Electric field (\vec{E}_x). Now the Transverse Magnetic field of flux density (B_z) applied along *Z*-axis (perpendicular to the plane of the paper and in a direction out of the paper). Since holes are flowing under the *E*-field (\vec{E}_x), so they will move with a drift velocity (v_{dx}) and under the influence of transverse *M*-field (\vec{B}_z) they will experience a Lorentz force as

$$\vec{F} = ev_{dx} \vec{B}_z \quad (6.34)$$

Due to this Lorentz force holes will be deflected downward. This Lorentz force tend to drive the holes towards the front surface of the block as shown in the Fig. 6.13(a). So there is an excess of hole in the front surface, so it shows (+ve) polarity, where as there will be a deficiency of hole automatically in the back surface, due to which that shows the (-ve) polarity. So an *E*-field (\vec{E}_y) will be formed along (+ve) direction of *Y*-axis, the corresponding voltage is called Hall voltage (V_{Hp}). This *E*-field (\vec{E}_y) produces a force on the electron in the upward direction i.e. along (+ve) *Y*-axis as (eE_y), which will be counter balance by the downward Lorentz force. So in equilibrium

$$eE_y = F = ev_{dx} B_z \text{ i.e. } E_y = B_z v_{dx} \quad (6.35)$$

The current density due to hole in *p*-type semiconductor from Eqn. (6.11) will be

$$J_s = pe v_{dx} \text{ i.e. } v_{dx} = J_s / pe \text{ i.e. } E_y = \frac{B_z J_s}{pe} \quad (6.36)$$

According to the Fig. 6.13(a), the *Hall voltage* V_{HP} developed across the front and back surfaces of the semiconductor slab of width (*w*) is

$$V_{HP} = E_y w = \frac{B_z J_s w}{pe} \quad (6.37)$$

Above Eqn. (6.37) can also be written as

$$R_{HP} = 1/pe = \frac{E_y}{B_z J_s} \quad (6.38)$$

Where the factor (R_{HP}) is called the **Hall co-efficient** for *P*-type semiconductor. Hall co-efficient can also be written in another form by considering the total current flow (*I*) through a semiconductor bar Fig. 6.13(a) of thickness (*d*) and width (*w*) i.e. of area ($A = wd$).

So total current *I* = Current density \times Area

$$\text{i.e. } I = J_s \cdot wd \text{ or } J_s = I/wd \quad (6.39)$$

Now by substituting this value of J_s in the Eqn. (6.38) for R_{HP} , we will get for *p*-type semiconductor

$$\text{Hall co-efficient } R_{HP} = 1/pe = \frac{E_y}{B_z J_s} = \frac{V_{HP}/w}{B_z I/wd} = \frac{V_{HP}d}{IB_z} \quad (6.40)$$

$$\text{and Hall Voltage } V_{HP} = E_y w = R_{HP} I B_z / d \quad (6.41)$$

Similarly for *n*-type semiconductor as shown in the Fig. 6.13(b), Hall voltage will be along (-ve) *Y*-axis, so for *n*-type semiconductor

$$\text{Hall Voltage } V_{Hn} = -E_y w = -R_{Hn} I B_z / d \quad (6.42)$$

Hall co-efficient

$$R_{Hn} = -1/ne = -\frac{E_y}{B_z J_s} = -\frac{V_{Hn}/w}{B_z I/wd} = -\frac{V_{Hn}d}{IB_z} \quad (6.43)$$

Now from the Eqns. (6.40) and (6.43) the measurement of current (*I*), the magnitude and sign of Hall voltage (V_H), for any M-flux density (B_z) gives the sign of the charge carrier i.e. whether it is *p*-type or *n*-type as well as the hole density (*p*) or electron (*n*) can also be calculated.

In case of semiconductor there will be some modification in the drift velocity (v_{dx}) of the charge carrier under *E*-field, since they are not free charge carrier as it was assumed. So the modified expression for Hall co-efficient will be

$$R_{HP} = 3\pi/8 (1/pe) \text{ and } R_{Hn} = -3\pi/8 (1/ne) \quad (6.44)$$

Hall angle: From Fig. (6.13a) it is clear that the net *E*-field in the semiconductor is the vector sum of \vec{E}_x and \vec{E}_y . The angle (Θ) between the resultant *E*-field (\vec{E}) with (+ve) *X*-axis is known as Hall angle. So from Fig. (6.13a) we get

$$\tan \Theta = E_y / E_x = \frac{B_z J_x}{peE_x} \quad (6.45)$$

Now current density from Eqn. (6.12) for *p*-type material $J_x = \sigma_p E_x = pe\mu_p E_x$. Now by combining the above equation for J_x with eqn. (6.45), we get

$$\tan \Theta = \frac{B_z \sigma_p}{pe} = \frac{B_z}{pe} \sigma_p = \mu_p B_z \quad (6.46)$$

$$\text{i.e. } 1/pe = R_{HP} = \mu_p / \sigma_p \text{ or } \mu_{HP} = \mu_p = \sigma_p R_{HP} \quad (6.47)$$

This μ_{HP} is now called as Hall mobility for *p*-type semiconductor.

Now when the semiconductor doped with trivalent as well as pentavalent impurity atoms of different percentage then both the type *p* as well as *n*-type will be present in the semiconductor material and it will be nearly same as intrinsic material. In that case Hall co-efficient will be

$$R_H = \frac{E_y}{J_x B_z} = \frac{3\pi}{8} \frac{1}{e} \frac{p\mu_p^2 - n\mu_n^2}{(p\mu_p + n\mu_n)^2} \quad (6.48)$$

Comparison of the above equation (6.48) with Eqns. (6.40, 6.41 & 6.42, 6.43) shows that Hall co-efficient (R_H) and Hall voltage (V_H) are small for near intrinsic material as above in comparison to the material which is highly doped as extrinsic material.

Problems Related Conductivity, Current Density, Hall Effect, For Metal and Semiconductor:—

1. An electron in a cathode ray tube falls through a potential difference of 1000 volts. If the electron initially started from rest, what is its final velocity?

$$\text{Solution} \quad eV = \frac{1}{2} mv^2 \quad \text{or} \quad v = \sqrt{\frac{2eV}{m}}$$

$$\text{i.e. } v = \sqrt{\frac{2 \times 1.6 \times 10^{-19} \times 1000}{9.1 \times 10^{-31}}} = 1.88 \times 10^7 \text{ m/sec}$$

2. Find the *E*-field required to establish a current density 1A/m^2 of Copper plate whose conductivity $\sigma = 5.8 \times 10^5 \text{ mho/cm}$

$$\text{Solution From Eqn. (6.12) } J_c = \sigma E \quad \text{or} \quad E = J_c / \sigma$$

$$\text{i.e. } E = \frac{1}{5.8 \times 10^5} = 0.1724 \times 10^{-7} \text{ V/m}$$

3. Find the magnitude of the *E*-field intensity in a sample of Silver having ($\sigma = 6.17 \times 10^7 \text{ mho/m}$) and ($\mu_e = 0.0056 \text{ m}^2/\text{V.s}$), if (a) drift velocity is 1 mm/sec. (b) the current density is 10^7 A/m^2 (c) the sample is a cube 3 mm on a side, carrying a total current of 80 amp. (d) sample is cube 3 mm on a side, having a p.d. of 0.5 mV between opposite faces.

Solution

- (a) From Eqns. (6.12, 6.11) we get

$$J_c = \sigma E = -\rho_e \mu_e E = -\rho_e v_d$$

where $v_d = 1 \text{ mm/sec.} = 1 \times 10^{-3} \text{ m/sec.}$

So from above equation we get $\sigma E = -\mu_e \rho_e E$

i.e. $\rho_e = -\sigma/\mu_e = -(6.17 \times 10^7/0.0056)$

So $J_c = -\rho_e V_d = -(6.17 \times 10^7/0.0056) \times 10^{-3} = -1.10 \times 10^7 \text{ A/m}^2$

i.e. $E\text{-field (E)} = J_c/\sigma = -0.1786 \text{ V/m}$

(b) $J_c = 10^7 = 6.17 \times 10^7 \times E$

i.e. $E\text{-field (E)} = 0.1621 \text{ V/m}$

(c) Current (I) = 80 Amp and $J_c = I/A = 80/(9 \times 10^{-6}) = 8.88 \times 10^6$

i.e. $E\text{-field (E)} = J_c/\sigma = 0.144 \text{ V/m}$

(d) $E\text{-field (E)} = V/d = 0.5 \times 10^{-3}/3 \times 10^{-3} = 0.166 \text{ V/m}$

4. An aluminium conductor with conductivity $3.82 \times 10^7 \text{ mho/m}$, is 1000 ft long has circular cross section with diameter 0.8 inch. If there is a D.C. voltage of 1.2 V applied between the ends, the calculate (a) the current density (b) total current and (c) power dissipation

Solution

$$\text{Length (L)} = 1000 \text{ ft} = 1000 \times 12 \times 2.54 \times 10^{-2} = 304.80 \text{ m}$$

$$\text{Radius (r)} = 0.4 \text{ inch} = 0.4 \times 2.54 \times 10^{-2} = 1.016 \times 10^{-2} \text{ m}$$

(a) Current density (J_c) = $\sigma E = \sigma V/L = \frac{3.82 \times 10^7 \times 1.2}{304.80} = 1.504 \times 10^5 \text{ A/m}^2$

(b) Total current (I_c) = $J_c \times A = 1.504 \times 10^5 \times 3.14 \times (1.016 \times 10^{-2})^2 = 48.8 \text{ Amp}$

(c) Power dissipation (P) = $I \times V = 48.8 \times 1.2 = 58.56 \text{ watt}$

5. A rod of p -type Ge 6 mm long, 1 mm wide, and 0.5mm thick has an electrical resistance of 120 ohm. Now (a) what is p -type impurity concentration, by neglecting minority carrier concentration? (b) What portion of the conductivity is due to holes and electrons? Given that

$$n_i = 2.5 \times 10^{19}/\text{m}^3, \mu_n = 0.39 \text{ m}^2/\text{V.s}, \mu_p = 0.19 \text{ m}^2/\text{V.s}$$

Solution (a) Hole concentration = $p = N_a$ and $\sigma = \sigma_p = N_a e \mu_p$

Now $\sigma_p = \frac{1}{\rho} = \frac{L}{RA} = \frac{6 \times 10^{-3}}{120 \times 0.5 \times 10^{-6}} = 100 \text{ mho/m}$

where ρ is the resistivity ohm/m.

i.e. for Eqn. (6.25) $p = N_a = \sigma_p/e\mu_p = \frac{100}{1.6 \times 10^{-19} \times 0.19}$

Majority concentration i.e. Hole concentration = $p = 328 \times 10^{19}/\text{m}^3$

(b) From Eqn. (6.29) Minority concentration = $n = n_i^2/N_a = \frac{(2.5 \times 10^{19})^2}{328 \times 10^{19}}$

i.e. $n = 0.019 \times 10^{19}$

Portion of minority conductivity $\sigma_n = ne\mu_n = 0.019 \times 10^{19} \times 1.6 \times 10^{-19} \times 0.39 = 0.0118 \text{ mho/m}$

so the proportion of conductivity hole : electron = $\frac{N_a e \mu_p}{ne \mu_n} = \frac{\sigma_p}{\sigma_n}$

i.e. $\frac{\sigma_p}{\sigma_n} = 100/0.0118 = 8434.5 : 1$

6. Calculate the current produced in a small Ge plate of area 1 cm^2 and of thickness 0.3 mm, when a p.d. of 2V is applied across the faces. Given the concentration of free electron is $2 \times 10^{19}/\text{m}^3$ ($= n_i$) and mobility of electron and hole are $\mu_n = 0.36$ and $\mu_p = 0.17 \text{ m}^2/\text{V.S.}$

Solution

$$\text{Current } I = J \times A = \sigma E \times A$$

where

$$E = V/d \text{ and } \sigma = n_i e(\mu_p + \mu_n)$$

So

$$I = \frac{n_i e(\mu_p + \mu_n) \times V \times A}{d}$$

$$= \frac{2 \times 10^{19} \times 1.6 \times 10^{-19} (0.17 + 0.36) \times 2 \times 10^{-4}}{0.3 \times 10^{-3}}$$

$$I = 1.1306 \text{ Amp.}$$

7. The resistivity of Si is $6.3 \times 10^4 \text{ ohm-m}$ at 300°K. Calculate the intrinsic carrier density. Given that $\mu_n = 0.14$ and $\mu_p = 0.05 \text{ m}^2/\text{V.S.}$

Solution

$$\sigma = n_i e (\mu_n + \mu_p)$$

$$\sigma = \frac{1}{\rho} = n_i e (\mu_n + \mu_p) \text{ (where } \rho \text{ is the resistivity)}$$

So

$$n_i = \frac{1}{\rho e (\mu_n + \mu_p)} = \frac{1}{6.3 \times 10^4 \times 1.6 \times 10^{-19} \times 0.19}$$

$$n_i = 1.915 \times 10^{15}/\text{m}^3$$

8. A resistance between opposite face of a 1 mm cube of a semiconductor is 1.5 ohm. A Hall field of 0.6 V/m developed when a current 120 mA is carried between opposite faces and 0.05 wb/m² M-field induction is applied perpendicularly to it. Calculate Hall voltage, Hall co-efficient Hall mobility, Hall angle, and density of charge carrier by assuming single carrier concentration.

Solution $\sigma_{\text{sp}} = \frac{I}{RA} = \frac{10^{-3}}{1.5 \times 10^{-6}} = 0.66 \times 10^3 \text{ /}\Omega \cdot \text{m}$ (a) From Eq. (6.37) Hall voltage = $V_{H_P} = E_1 d = 0.6 \times 10^{-3} \text{ volt}$ (b) From Eq. (6.40) Hall co-eff = $R_{H_P} = \frac{V_{H_P} d}{IB} = \frac{0.6 \times 10^{-3} \times 10^{-3}}{120 \times 10^{-3} \times 0.05} = 10^{-4} \text{ m}^2/\text{c}$ (c) From Eq. (6.47) Hall mobility = $\mu_{H_P} = \sigma_{\text{sp}} R_{H_P} = 0.066 \text{ m}^2/\text{V.S.}$ (d) From Eq. (6.46) Hall angle = $\theta = \tan^{-1}(\mu_{H_P} B_z) = 0.19^\circ$ (e) From Eq. (6.38) Density of charge carrier = $p = \frac{1}{R_{H_P} e} = 625 \times 10^{20}/\text{m}^3$

9. What is concentration of holes in Si which is having a donor concentration $1.4 \times 10^{24}/\text{m}^3$ and intrinsic concentration $1.4 \times 10^{19}/\text{m}^3$. Consider concentration of electron = donor concentration. Find also the ratio of electron concentration to the hole concentration.

Solution

$$n = N_d = 1.4 \times 10^{24}$$

$$n_i = 1.4 \times 10^{19}$$

$$\text{Now } np = n_i^2 \text{ so } p = \frac{n_i^2}{n} = \frac{n_i^2}{N_d} = 1.4 \times 10^{12}/\text{m}^3$$

$$\text{So concentration of hole} = p = 1.4 \times 10^{12}/\text{m}^3$$

Now electron-hole concentration ratio = n/p

$$\text{i.e. } n/p = 1.4 \times 10^{24}/1.4 \times 10^{12} = 10^{12}$$

10. A specimen of semiconductor has a Hall co-eff. of $3.66 \times 10^{-4} \text{ m}^2/\text{C}$ and resistivity of $8.93 \times 10^{-3} \text{ ohm-m}$. In a Hall effect expt. a M-flux density of 0.5 wb/m^2 is used. Find the Hall angle, mobility and density of charge carrier by assuming single charge carrier concentration.

Solution $R_{H\rho} = 3.66 \times 10^{-4} \text{ m}^2/\text{C}$ and $\rho = 8.93 \times 10^{-3} \Omega\text{m}$

$$(a) \text{ Density or charge carrier } = p = \frac{1}{R_{H\rho} e} = 171 \times 10^{20}/\text{m}^3$$

$$(b) \text{ Hall mobility } = \mu_{H\rho} = \sigma R_{H\rho} = \frac{R_{H\rho}}{\rho} = 0.041 \text{ m}^2/\text{V.S.}$$

$$(c) \text{ Hall Angle } = \theta = \tan^{-1}(\mu_{H\rho} B_z) = 1.17^\circ$$

11. Ge has a donor type impurity added to the extent of one atom per 10^8 Ge atom. What effect does this have on the conductivity of the material at 300° K , when $\mu_n = 3900 \text{ cm}^2/\text{v.s.}$, $\mu_p = 1900 \text{ cm}^2/\text{v.s.}$, No. of atom of Ge/cc = 4.42×10^{22} and $n_i^2 = 6.25 \times 10^{36}/(\text{cc})^2$.

Solution

$$\text{Intrinsic conductivity} = \sigma_m = n_i e(\mu_n + \mu_p)$$

$$\text{i.e. } \sigma_m = 2.5 \times 10^{13} \times 1.6 \times 10^{-19} (5800) = 0.0232/\Omega \cdot \text{cm}$$

Now 1 donor atom per 10^8 Ge atom doped

$$\text{So donor concentration} = N_d = \frac{4.42 \times 10^{22}}{10^8} = 4.42 \times 10^{14}/\text{cc}$$

$$\text{So extrinsic conductivity} = \sigma_{ex} = N_d e \mu_n$$

$$\text{i.e. } \sigma_{ex} = 4.42 \times 10^{14} \times 1.6 \times 10^{-19} \times 3900 = 0.269/\Omega \cdot \text{cm}$$

$$\text{So } \sigma_{ex} = 10 \times \sigma_m$$

i.e. Due to doping conductivity increased by 10 times.

12. What is the probability of an electron being thermally promoted to the C.B. in diamond at room temperature 27°C . The band gap energy is 5.6 eV in diamond and $K = 1.38 \times 10^{-23} \text{ J/K}$.

Solution

From Eqn. (6.2), probability of electron to be in C.B. = Fermifactor $F(E_c)$, where

$$F(E_c) = \frac{1}{1 + e^{(E_c - E_F)/kT}}$$

$$\text{Now } E_c = 5.6 \text{ eV i.e. } E_F = \frac{E_c}{2} = 2.8 \text{ eV}$$

$$\text{So } E_c - E_F = 2.8 \text{ eV and } KT = 1.38 \times 10^{-23}(27 + 273) \text{ J}$$

$$\text{i.e. } KT = 4.14 \times 10^{-21} \text{ J} = 2.6 \times 10^{-2} \text{ eV}$$

$$\text{i.e. } e^{(E_c - E_F)/kT} \gg 1$$

$$\text{So } F(E_c) = e^{(E_F - E_C)/KT} = e^{-2.8/2.6 \times 10^{-2}} = e^{-107.7} = 1.68 \times 10^{-47}$$

i.e. $F(E_c) = 1.68 \times 10^{-47} \rightarrow$ It's very negligible for an insulator like diamond, so diamond can not take part in conduction.

13. A bar of *n*-type Ge of size $10 \text{ mm} \times 1 \text{ mm} \times 1 \text{ mm}$ is mounted in a *M*-field 0.2 T. The electron density in the bar is $7 \times 10^{21}/\text{m}^3$. If 1 mV is applied across the long end of the bar, then determine the current through the bar, the Hall co-eff. and Hall voltage between Hall electrodes placed across the short end of the bar, when $\mu_e = 0.39 \text{ m}^2/\text{v.s.}$

Solution

(a) By using equation (6.11 and 6.12) Current = $I = J \times A = \sigma E \times A = ne\mu_e E \times A$

where

$$E = V/L$$

$$\text{So } I = \frac{7 \times 10^{21} \times 1.6 \times 10^{-19} \times 0.39 \times 10^{-3} \times 10^{-6}}{10 \times 10^{-3}}$$

$$I = 4.37 \times 10^{-5} \text{ Amp.}$$

$$(b) \text{Hall co-eff} = R_{H_e} = -\frac{1}{ne} = -8.9 \times 10^{-4} \text{ m}^3/\text{c}$$

$$(c) \text{Hall voltage} = V_{H_e} = -R_{H_e} IB_z/d$$

$$\text{i.e. } V_{H_e} = -8.9 \times 10^{-4} \times 4.37 \times 10^{-5} \times 0.2/10^{-3}$$

$$V_{H_e} = 7.75 \times 10^{-6} \text{ volt}$$

6.15 *p-n* Junction without Applied Field

Semiconductor junctions are commonly formed between a *p*-and *n*-type material, as shown in Fig. 6.14(a). But the junctions cannot be made by placing two pieces of semiconducting materials side by side, because surface film and other irregularities would produce major discontinuities in the crystal structure. The *p-n* junction can be made by diffusing the donor impurities into one side and acceptor impurities into the other side of a single crystal of a semiconductor, either by alloying process or growing a crystal with an impurity density which is a function of distance. Such a system of *p-n* junction is shown in Fig. 6.14(b).

Just after *p-n* junction is formed electrons on *n*-side of the junction would have an average energy higher than those on *p*-side as shown in Fig. 6.14(a), since Fermi level E_{F_n} for *n*-side is higher than Fermi level E_{F_p} for *p*-side. So there would be transfer of energy and electrons will flow from *n*-side to *p*-side, until the Fermi level in the two sides does line up as shown in Fig. 6.14(c). Electrons diffuse from *n* to *p* side which leave (+ve)ly ionised atom on *n*-side and produce (-ve)ly ionised atom on *p*-side near *p-n* junction due to diffusion as shown in Fig. 6.14(b). Similarly when holes diffuse from *p* to *n* side then they leave (-ve)ly ionised atom on *p* side and produce (+ve)ly ionised immobile atom on *n* side near *p-n* junction. Those ionised atoms are immobile, since they are bound in the crystal lattice close to the *p-n* junction, therefore there are very few mobile charge carriers available near *p-n* junction. So this region is referred to as the **charge depletion region** or **space charge region**. Due to this space charge region in the immediate neighbourhood of the junction, a restoring force i.e. a contact potential difference is set up, which is called **barrier potential (V_B)**. The physical distance from one side of the barrier to the other side, i.e., the width of the depletion region is referred to as the **width of the barrier**. The difference of

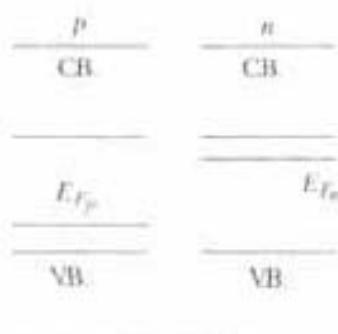


Fig. 6.14(a)

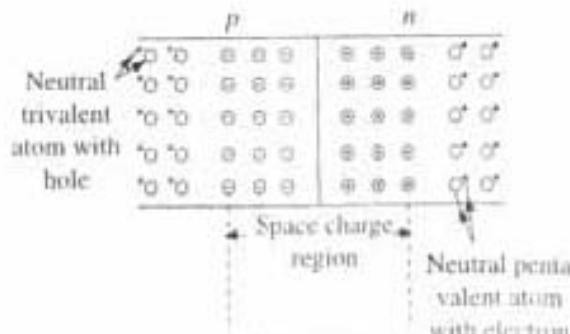


Fig. 6.14(b)

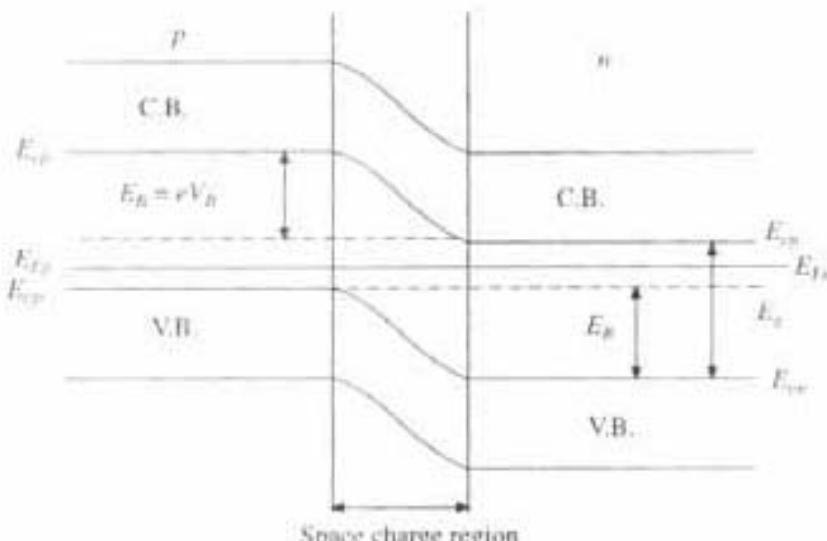


Fig. 6.14(c)

Fig. 6.14 (a) Energy levels for separate *p* and *n* sides; (b) Formation of space-charge or depletion region in *p-n* junction; (c) Energy level diagram of *p-n* junction under open circuit condition with barrier potential energy $E_B = eV_B$

contact potential between two sides of the barrier is referred to as *Height of the Potential Barrier* ($E_B = E_{cp} - E_{cn} = eV_B$) as shown in Fig. 6.14(c). Due to this potential barrier, no further electron diffusion is possible from *n* to *p*-side, as well as no hole diffusion is possible from *p* to *n*-side. Formation of potential barrier can be explained more clearly with energy level diagram as follows.

In Fig. 6.14(c) the energy level diagram for open CKT *p-n* junction has been shown. As soon as *p-n* junction is formed the Fermi level becomes constant throughout the specimen at equilibrium. After that no further electron and hole diffusion is possible. But in *n*-type region Fermi level (E_{F_n}) will be close to conduction band edge E_{cn} , whereas in *p*-type (E_{F_p}) will be close to valence band edge E_{vp} . So C.B. and V.B. of *p*-type region in *p-n* junction cannot be at the same levels as those of *n*-type region as it was before in Fig. 6.14(a). Hence the energy level diagram of *p-n* junction will be changed as shown in Fig. 6.14(c), where a shift in energy level E_B will appear as barrier potential energy, which is equal to

$$E_B = E_{cp} - E_{cn} = E_{vp} - E_{vn} = eV_B$$

This energy gap $E_B = eV_B$ (≈ 0.5 eV at room temperature) represents the potential energy of the electron at the p - n junction. Because of this contact potential difference V_B across a junction, an electron at the bottom of C.B. in the p -type region is having higher potential energy $E_B = eV_B$ than the electron at the bottom of C.B. of n -type region. ***This potential barrier resists the flow of electrons from n to p side.*** So in an open circuited p - n junction, i.e., without external voltage, net flow of current is zero.

6.16 Semiconductor Junction with Applied Voltage

When an external voltage is applied to a p - n junction it is found that the voltage current characteristic is non-linear (Fig. 6.15c). It permits the easy flow of current in one direction whereas the flow of current in opposite direction is restricted. This property is used to consider the p - n junction as *semiconductor diode and as rectifier*.

6.16.1 Forward Bias: The important effects due to forward bias situation are as follows:

(1) If p -type material is made +ve and n -type -ve by an external voltage supply junction, then it is called forward biased (Fig. 6.15a). In this situation the resistance of the device is very low and current flows easily (Fig. 6.15c).

(2) In forward bias conduction the width of the space charge region will decrease Fig. (6.15a) due to the repulsion of holes and electrons from +ve and -ve polarity of the battery towards the p - n junction.

(3) In the presence of forward bias voltage V , electrons from n to p side are moving in the direction of the applied external field, so barrier potential V_B will be reduced to $V_f = (V_B - V)$ and the C.B. and V.B. of the p -side have been also lowered by an amount V electron volts (Fig. 6.16a). The Fermi level of the p -type material (E_{Fp}) which coincides with Fermi level for n -type material (E_{Fn}) as in Fig. 6.14(c) is now also lowered by V electron volt w.r.t. its counter-part in the n -type material (E_{Fn}). This favours the condition of easy flow of electron from n -type to p -type side (Fig. 6.15c).

(4) In the forward biased conduction, majority charge carriers, i.e., large number of electrons will diffuse from n - to p -side as well as large number of holes will diffuse from p to n side. The deficiency of electrons in n -side will be compensated by the electron supply from (-ve) terminal of the external source. Similarly in the p -type material a new hole is created and compensates for the loss of holes, when an electron is coming out by breaking the covalent bond in a crystal and enters the (+ve) terminal of the external source. Thus in the forward bias condition there is a continuous current flow in the circuit. If forward bias voltage (V) increases, more diffusion occurs so more current (I) will pass through the junction (Fig. 6.15c).

(5) The flow of minority carriers remains uninfluenced due to forward bias, which gives rise to a very small current in the direction opposite to that of the current due to forward bias.

6.16.2 Reverse Bias: The following are the effects of reverse bias situation:

1. If *n*-type material is made positive w.r.t. *p*-type material, the junction is said to be reverse biased (Fig. 6.15b) and its resistance is very high which allows a very feeble current, almost zero, through the junction (Fig. 6.15c).

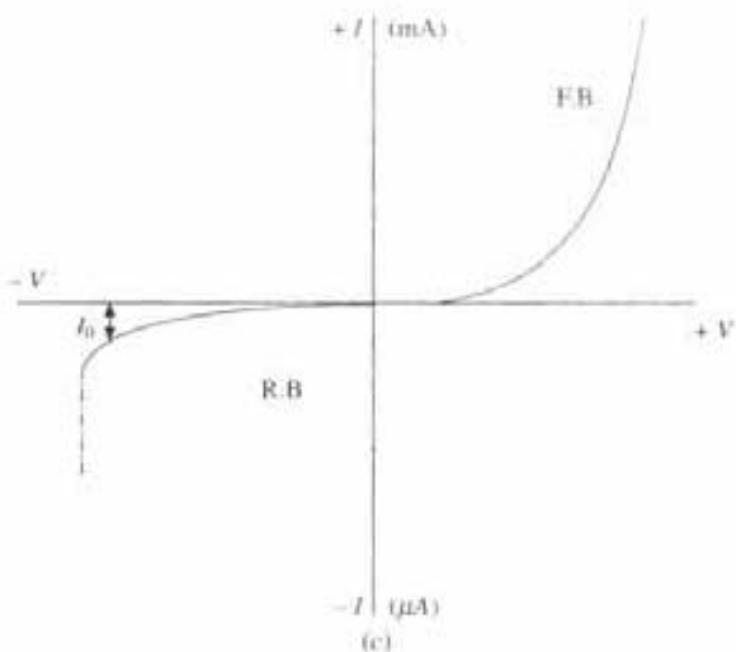
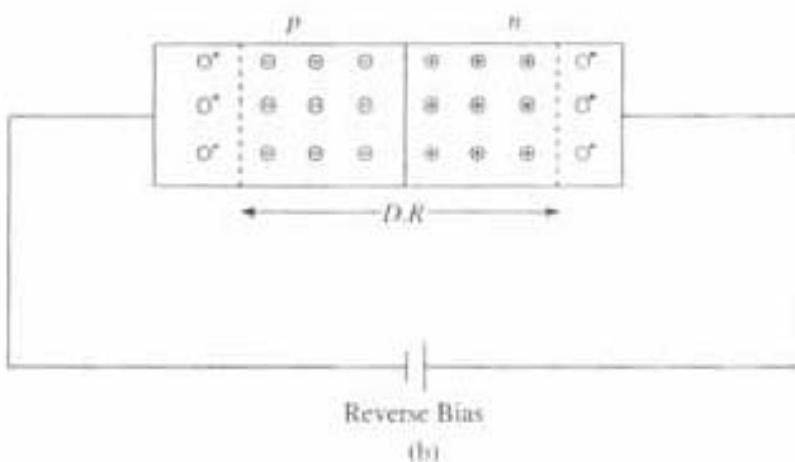
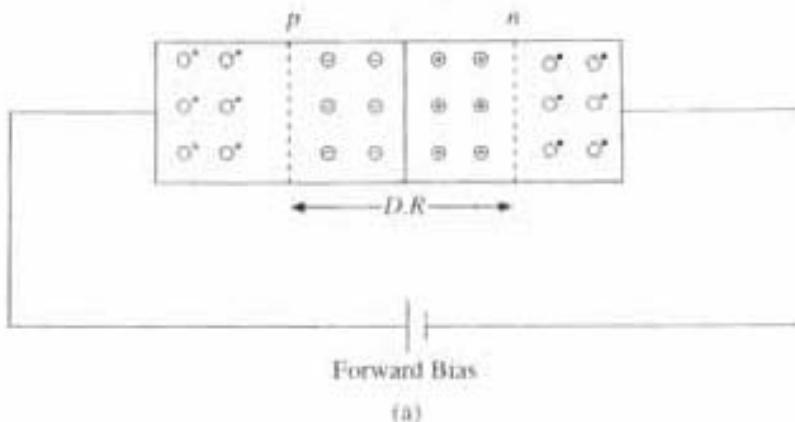


Fig. 6.15 (a) *p-n* junction under forward bias; (b) *p-n* junction under reverse bias; (c) Current voltage (*I-V*) relation in *p-n* junction under forward and reverse bias

2. When reverse bias is applied, the free electrons in the *n*-region and holes in the *p*-region are attracted away from the junction under action of the reverse voltage *V* due to which the width of the depletion region increases (Fig. 6.15b).

3. Due to the presence of reverse voltage *V* the potential barrier E_B will increase to $E_B = V_B + V$ (Fig. 6.16b). Because in the reverse biased condition

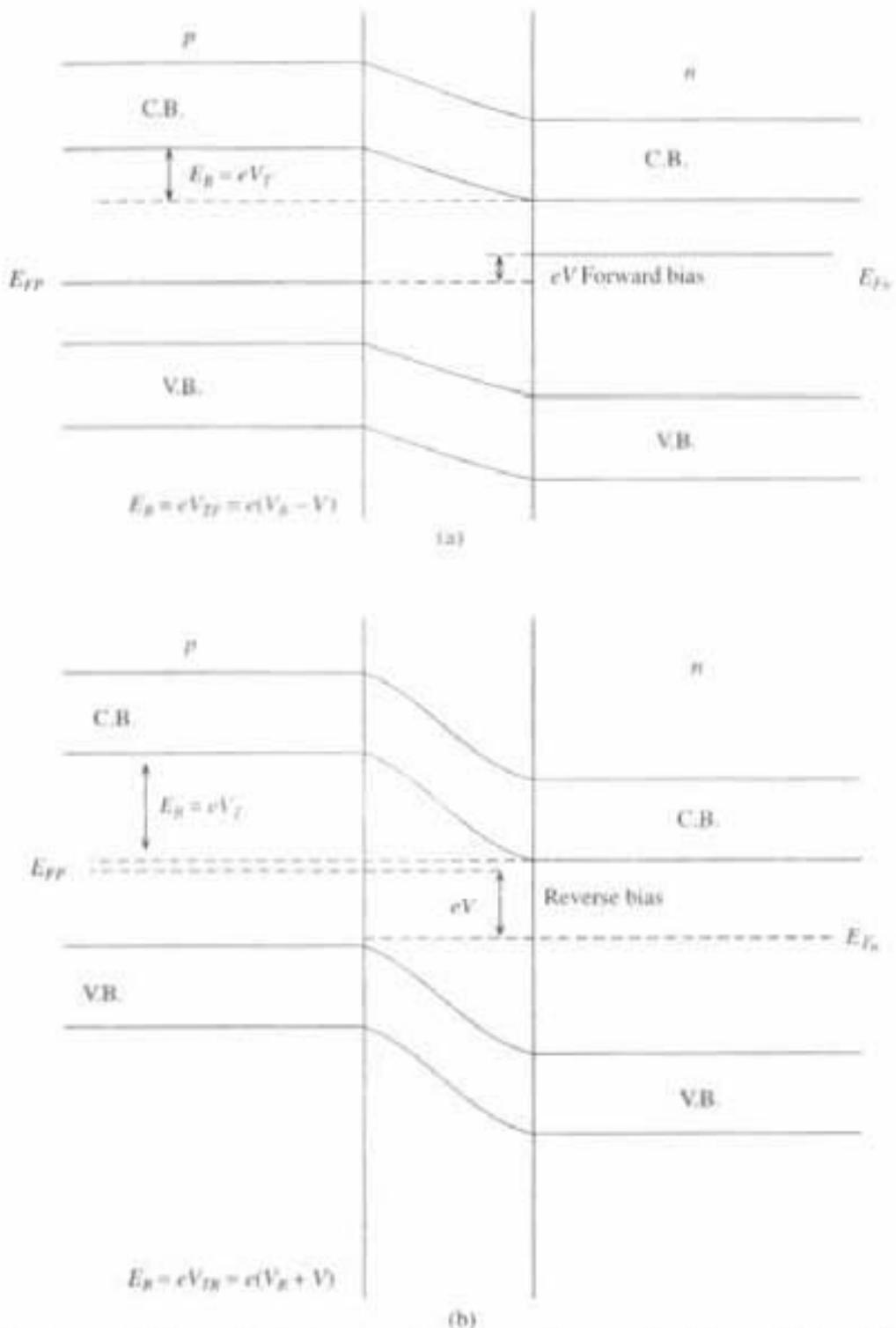


Fig. 6.16 (a) Energy level diagram for *p-n* junction under forward bias voltage (*V*), with decreased barrier potential energy (E_B). [$E_B = eV_F = e(V_B - V)$]; (b) Energy level diagram for *p-n* junction under reverse bias voltage (*V*), with increased barrier potential energy (E_B). [$E_B = eV_F = e(V_B + V)$]

when electrons are moving from *n* to *p* region they are moving in a direction opposite to the direction of electric field. So barrier potential will increase by an amount V volt (Fig. 6.16b). Increased barrier potential resists the flow of electrons and holes across the junction which establishes a high resistance.

4. The increased barrier potential makes it difficult for the majority carriers to diffuse across the junction. But this barrier potential for reverse bias condition helps the minority carriers in crossing the junction and gives rise to a small current. This current is very small in comparison to forward bias current because the number of minority carriers is very small. Since the generation of minority carriers depends on temperature only, so the current flow due to minority carriers remains same, it does not depend on reverse bias voltage so this current is called reverse saturation current (I_0) (Fig. 6.15c). It is very small, of the order of nano amperes for silicon and of the order of micro amperes for germanium diode.

5. The fermi level for *p*-type material (E_{F_p}) is also raised w.r.t its counterpart E_{F_n} in *n*-type material by an amount V -electron volts, which will resists the condition of flow of electron from *n* to *p* side (Fig. 6.16b).

From the above discussion it is clear that under the forward bias situation *p-n* junction has a low resistance and allows current to flow across the junction but in reverse biased situation it has high junction resistance which will not allow the current to flow across the junction, this behaviour is similar to vacuum diode.

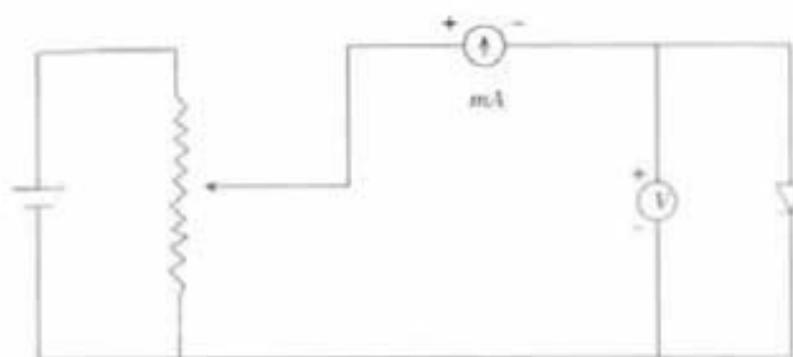
So *p-n* junction is often called as semiconductor diode whose symbol is 

6.17 V-I Characteristics of *p-n* Junction/Semiconductor-Diode

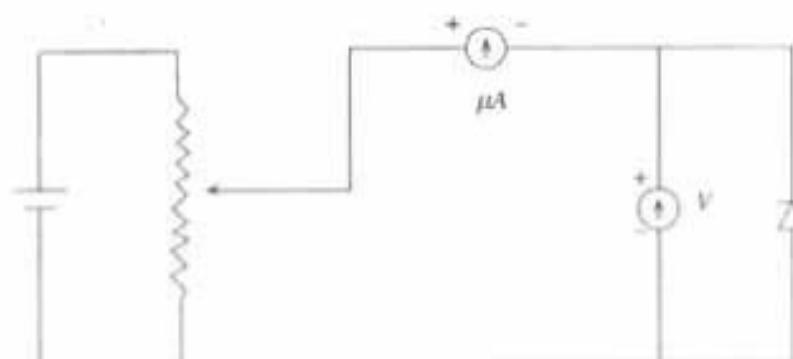
The variation of current against voltage across *p-n* junction is called the semiconductor diode characteristic. The *V-I* characteristics across *p-n* junction for forward bias and reverse bias are as shown, in the circuit connection shown in Figs. 6.17(a) and (b), respectively.

At zero applied voltage the potential barrier will not allow any current to flow across the junction. As forward bias V is applied across the *p-n* junction, current increases exponentially with voltage except for a small region in the neighbourhood of the origin called "Cutin or Threshold Voltage" (V_c) (0.2 for Ge, and 0.6 V for Si) below which the current is very small, almost zero, because the applied voltage is less than the barrier potential at the junction. When the applied voltage exceeds the barrier potential, i.e., beyond the cutin voltage, current rises very rapidly with applied voltage. The point from where the current increases rapidly is called *knee voltage*. After knee voltage, current increases almost linearly with voltage, having a small constant resistance R_f (Fig. 6.17c).

The range of forward current over which a semiconductor diode is operated is many orders of magnitude larger than the reverse saturation current. So in order to display the forward and reverse characteristics together we need a different current scale. As shown in Fig. 6.17(c) the current scale for forward bias is millamps whereas for reverse bias it is microamps. At a particular reverse voltage called *breakdown voltage* a large reverse current flows, this region of the reverse characteristics is called breakdown region.



(a)



(b)

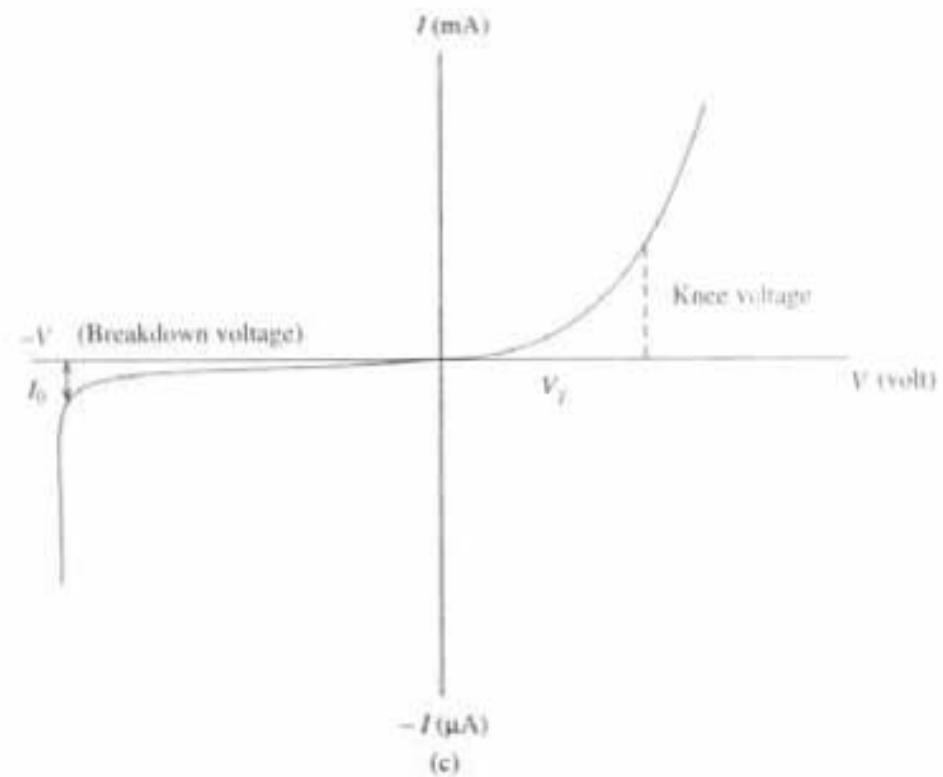


Fig. 6.17 Experimental set up for $p-n$ junction under forward bias, reverse bias, and $I-V$ characteristics for both

The static and dynamic resistances of the semiconductor diode are:

(a) *Static resistance*

$$r_s = V/I \text{ ohm}$$

In the forward bias condition the ratio of voltage to current is called static resistance (r_s). It varies widely as operation point on the characteristics.

(b) Dynamic resistance

$$r_d = \delta V / \delta I \text{ ohm}$$

In forward bias condition the reciprocal of the slope of the junction characteristics is called dynamic resistance (r_d). It is not constant, rather it depends on the operation point and decreases as forward bias voltage increases.

For forward bias condition, r_s and r_d comes out to be small, of the order of few ohms. But in reverse bias condition static resistance is very high, of the order of mega ohms. Dynamic resistance is much higher.

6.18 Semiconductor Diode as Rectifier

Almost all electronic components need D.C. power supply. Rectifier is a device with converts A.C. to D.C. and allows the current to flow only in one direction. Thus a device which offers a low resistance to the current in one direction but high resistance in the opposite direction is called a rectifier. The rectifying device is usually a semiconductor diode or vacuum diode. Since semiconductor diode or $p-n$ junction has infinite resistance in reverse bias condition but a small and constant resistance R_f in the forward bias condition, so it can be used as a rectifier. It can be used as half wave or full wave rectifier as follows.

6.18.1 Half Wave Rectifier

The connection for half wave rectifier is shown in Fig. 6.18(a) and input, output wave forms are shown in Fig. 6.18(c).

The input A.C. is connected to the primary, one end of secondary S is connected to p side of $p-n$ junction other end is connected to n -side of $p-n$ junction through a load resistance R_L . For positive half cycle of the A.C. input p is positive w.r.t. n , then $p-n$ junction is in the forward biased condition which allows the current to flow through the junction. But for negative half cycle of the A.C. input p is negative w.r.t. n , then the junction will be in the reverse biased condition which will not allow the current to flow through the junction as well as through R_L . Hence diode does not conduct and there is no current in the circuit. Thus there is a current I through R_L for the positive half cycle of input A.C. voltage V , which gives output voltage $V_o = I R_L$. But for negative half cycle of V , there is no current through R_L so output voltage across R_L is zero. Wave form for input voltage (V_i) and output voltage (V_o) are given in Fig. 6.18(c). As the circuit gives unidirectional current only for positive half cycle so it is called half wave rectifier. Output can be smoothed with the help of filter circuit.

6.18.2 Full Wave Rectifier

Full wave rectifier consists of two diodes. The circuit diagram is shown in Fig. 6.18(b), and input and output wave forms are shown in Fig. 6.18(c).

Full wave circuit comprises two half wave circuits. They are connected as shown in Fig. 6.18(b). The two ends of the secondary S_1 and S_2 of the transformer

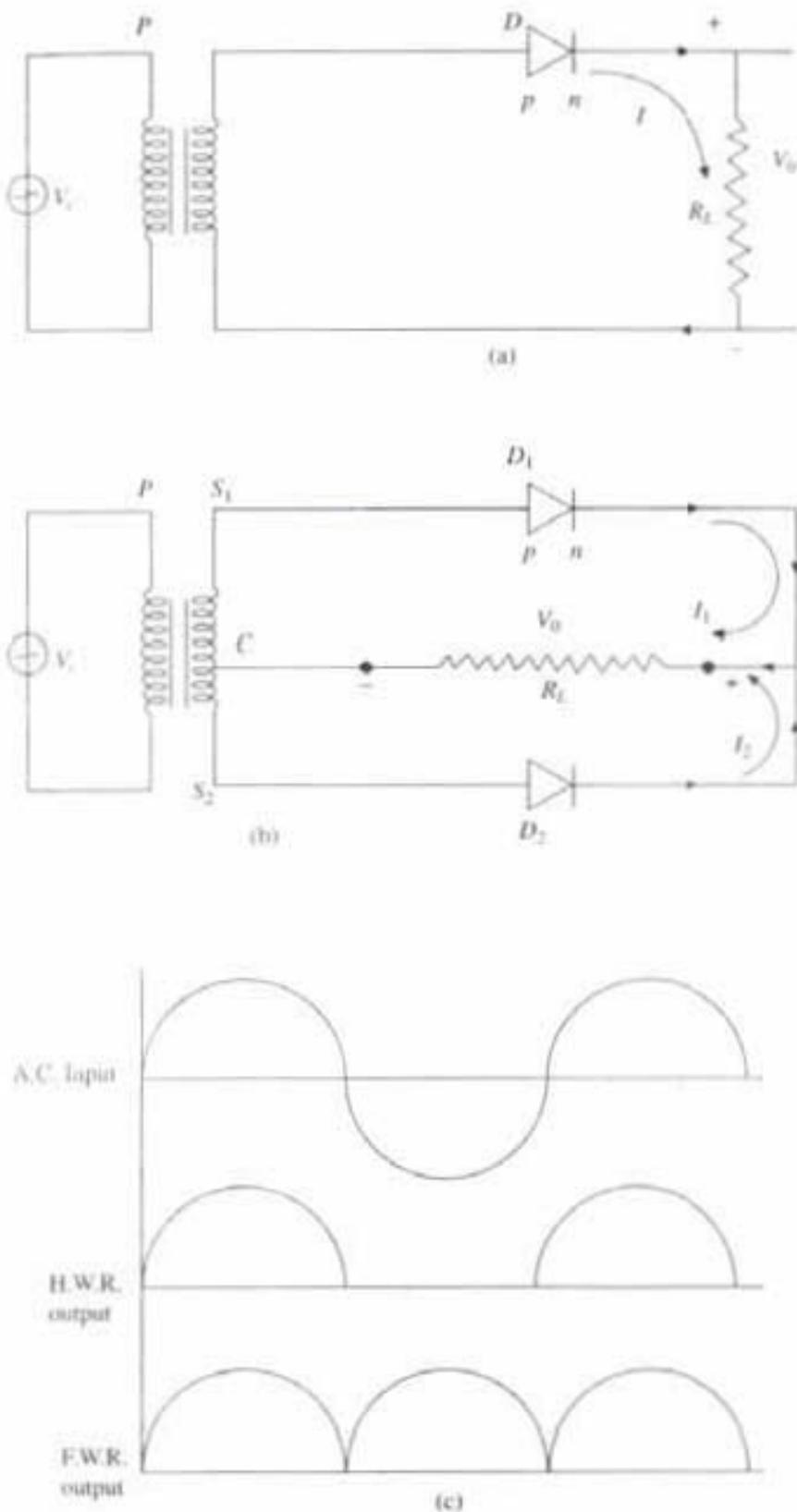


Fig. 6.18 Semiconductor diode as half wave rectifier, full wave rectifier and output wave form for both

are connected to the p -sides of two semiconductor diode whereas centre point C of the secondary is connected to the n -sides of two diode through load R_L in such a way that conduction takes place through one diode D_1 during one half cycle of

the A.C. power cycle when S_1 is positive and through the other diode D_2 during the second half cycle of a.c power supply, when S_2 is positive. The direction of current flow in two loop is as shown in Fig. 6.18(b), from which it is clear that the direction of current flow through load R_L is same for both the loops. Thus, current to the load is the sum of these two currents

$$I = I_1 + I_2$$

output voltage across load R_L is $V_0 = I \times R_L$.

As the circuit gives unidirectional current for both the cycles so it called full wave rectifier. Output from full wave rectifier is as shown in Fig. 6.18c, which can be smoothed with the help of filter circuit.

6.19 Flow of Current Across P-N Junction: The Rectifier Equation

The current flow across $p-n$ junction quantitatively can be discussed in the following way. Fig. 6.19(a) shows the energy bands of a $p-n$ junction under forward bias ($+V$) condition, when the effective barrier potential height is ($V_T = V_R - V$). It also shows the current flow by the majority and minority charge carrier across the junction in C.B. and V.B. There are the following four possibilities of current flow across the junction and arrows give the direction of the particular current flow.

1. J_1^- = current density due to flow of majority charge carrier (n) from n to p side in C.B.
2. J_2^- = current density due to flow of minority charge carrier (n) from p to n side in C.B.
3. J_1^+ = current density due to flow of majority charge carrier (p) from p to n side in V.B.
4. J_2^+ = Current density due to flow of minority charge carrier (p) from n to p side in V.B.



Fig. 6.19(a) Different current flows across forward bias P-N junction

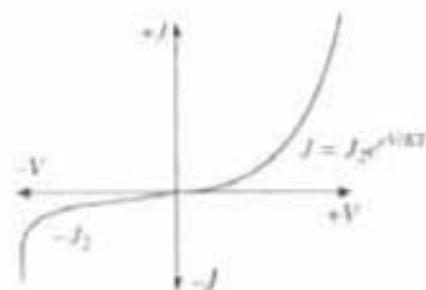


Fig. 6.19(b) I-V characteristics forward and reversed

So under forward bias ($+V$) net current flow through $p-n$ junction towards right in C.B. as well as in V.B. according to the Fig. 6.19(a) will be

$$J = (J_1^- - J_2^-) + (J_1^+ - J_2^+) = (J_1^- + J_1^+) - (J_2^+ + J_2^-) \quad (6.49)$$

where $(J_1^- + J_1^+)$ is the sum of the majority current densities for electron and hole. $(J_2^+ + J_2^-)$ is the sum of the minority current densities for electron and hole $= J_2$.

Now as the barrier potential increases current flow through *p-n* junction decreases. So the majority current density in C.B. for the flow of electron across the junction towards right can be written mathematically as

$$J_1^- = C_1 e^{-eV_T/KT} \quad (6.50)$$

Similarly majority current density in V.B. for the flow of hole across the junction

$$J_1^+ = C_2 e^{-eV_B/KT} \quad (6.51)$$

Now in the open circuit, i.e., when no external bias voltage is given, then ($V = 0$) and ($V_T = V_B$) and there will not be any current flow across the *p-n* junction towards right either in C.B. or in V.B. So from the above eqn. (6.49) we get

$$J = (J_1^- - J_2^-) + (J_1^+ - J_2^+) = 0$$

$$\text{i.e. } C_1 e^{-eV_B/KT} - J_2^- = 0 \quad \text{i.e. } C_1 = J_2^- e^{eV_B/KT} \quad (6.52)$$

$$\text{similarly } C_2 e^{-eV_B/KT} - J_2^+ = 0 \quad \text{i.e. } C_2 = J_2^+ e^{eV_B/KT} \quad (6.53)$$

So in general when forward bias voltage (V) is given then by substituting the value of the constants C_1 and C_2 in the eqns. (6.50) and (6.51), we get

$$J_1^- = C_1 e^{-eV_T/KT} = J_2^- e^{e(V_B - V_T)/KT} \quad (6.54)$$

Similarly

$$J_1^+ = C_2 e^{-eV_T/KT} = J_2^+ e^{e(V_B - V_T)/KT} \quad (6.55)$$

Thus the forward bias current from Eqn. (6.49) across the *p-n* junction towards right will be

$$J = (J_1^- + J_1^+) - (J_2^+ + J_2^-)$$

$$J = [J_2^- e^{e(V_B - V_T)/KT} + J_2^+ e^{e(V_B - V_T)/KT}] - [J_2^+ + J_2^-]$$

Now V_B = Barrier potential, and $V_T = (V_B - V)$ = Reduced barrier potential under Forward bias (V), which gives ($V_B - V_T = V$). So the above equation now becomes

$$J = (J_2^+ + J_2^-) [e^{e(V_B - V_T)/KT} - 1] = J_2 [e^{eV/KT} - 1] \quad (6.56)$$

where $J_2 = (J_2^+ + J_2^-)$ = Reverse saturation current. (6.57)

Case-1 Forward-Bias

Now in the *forward bias*, applied voltage (V) increases (+ve)ly, so $e^{eV/KT} \gg 1$, so we can neglect the last term in the eqn. (6.56). So the eqn. (6.56) becomes under forward bias condition as follows:

$$J = J_2 e^{eV/KT} \quad (6.58)$$

So as the forward bias voltage (V) increases, current density (J) towards right across the $p-n$ junction will increase, which is the rectifier equation (6.58) for forward bias and $I-V$ characteristic for forward bias is as shown in the Fig. 6.19(b).

Case-1 Reverse-Bias

For the *reverse bias*, applied voltage (V) increases (-ve)ly, so $e^{-eV/KT} \ll 1$. So Eqn. (6.56) becomes

$$J = -J_2 \quad (6.59)$$

which is the reverse saturation current due to minority carrier only flowing towards left across $p-n$ junction, it is very negligible as shown in the $I-V$ characteristic Fig. 6.19(b) for reverse bias.

PROBLEM RELATED RECTIFIER EQUATION

1. Current flow in $p-n$ junction at 300°K is 0.2 micro amp when a large reverse bias is applied. Find current flow when forward bias of 0.1 volt is applied. Given that $K = 1.38 \times 10^{-23} \text{ J/K}$

Solution $J_2 = \text{Reverse saturation current} = 0.2 \times 10^{-6} \text{ amp}$.

Forward bias current according to the Eqn. (6.56)

$$J = J_2(e^{eV/KT} - 1) = 0.2 \times 10^{-6} e^{eV/KT} \text{ (as } e^{eV/KT} \gg 1\text{)}$$

$$J = 0.2 \times 10^{-6} e^{1.38 \times 10^{-23} \times 0.1 / 1.38 \times 10^{-23}} = 9.49 \times 10^{-6} \text{ amp} = 9.49 \text{ micro-amp.}$$

6.20 Transistor

A junction transistor is constructed by placing two $p-n$ junction of silicon or germanium crystal back to back. That is, either a n -type silicon is sandwiched in between two layers of p -type silicon, i.e., called $p-n-p$ transistor or a p -type silicon is placed in between two n -type materials which is called $n-p-n$ transistor. The invention of the transistor, completely changed the industrial electronics field because it is very handy, cheap, can be operated at very low voltage, does not require any heater or filament, has a long life, shock proof etc. In 1947, the transistor was invented in Bell laboratory by John Bardeer, Walter Brattain and William Shockley.

Figures 6.20 (a) and (b) represent $n-p-n$ and $p-n-p$ transistor, three parts of a transistor called emitter, base and collector.

At the time of manufacturing a transistor the base region is made lightly doped whereas emitter is heavily doped and the collector is doped in between the emitter and base. In order to dissipate more heat, collector region is made larger than the emitter region.

There are two $p-n$ junctions in the transistor. One is between emitter and base, other is between collector and base. Because of this transistor is called "bi-polar junction transistor" or BJT. The arrow on the emitter lead shows the direction of conventional emitter current flow (I_E). Figures 6.21(a) and 6.21(b) show the symbols of both $n-p-n$ and $p-n-p$ transistors. In $n-p-n$ transistor the arrow is

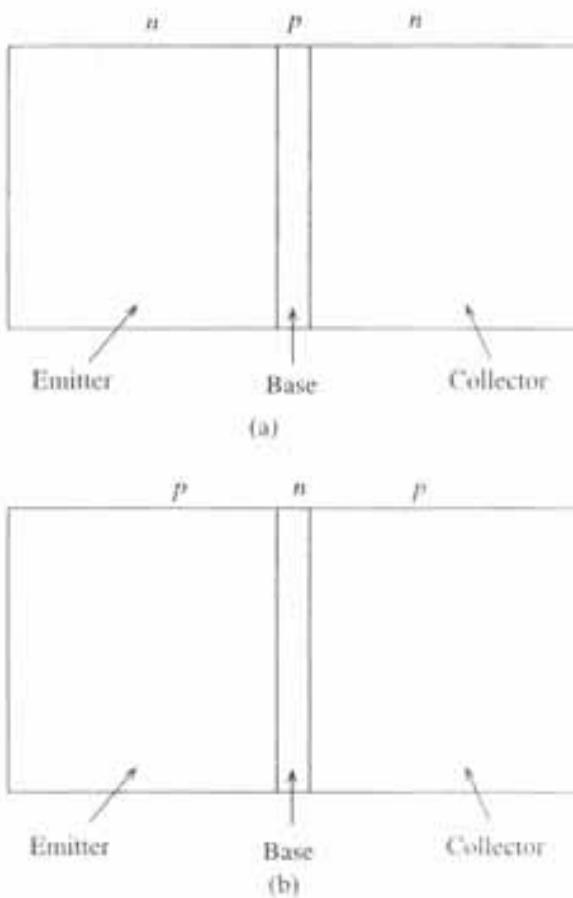


Fig. 6.20 (a) *npn* and *pnp* transistor; (b) *n-p-n* and *p-n-p* transistor

outward, because here electrons are the majority charge carrier but for *p-n-p* transistor the holes are majority charge carrier, therefore the arrow has inward direction. It is convenient to assume in both types, the emitter current (I_E), base current (I_B), and collector current (I_C) will be positive when the current flows into the transistor and negative when the current is coming out of the transistor. The symbols V_{EB} , V_{CB} and V_{CE} denote emitter-base, collector-base and collector-emitter voltage, respectively.

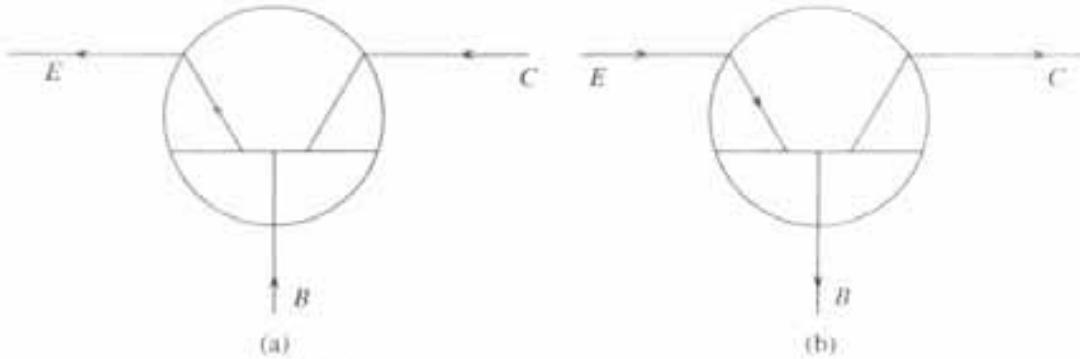


Fig. 6.21 Usual notation for *npn* and *pnp* transistor.

6.21 Working Principle of Transistor (Under C.B. Configuration)

For active operation of transistor the emitter-base voltage V_{EB} is made forward

bias whereas collector-base voltage V_{CB} is made reverse bias. The magnitude of forward bias V_{EB} will be made small in comparison to the magnitude of reverse bias voltage V_{CB} .

Since $p-n-p$ transistors are used more than $n-p-n$ transistors because they are easier to manufacture, the working principle of $p-n-p$ transistor is explained with energy level in Figs. 6.22(a) and 6.22 (b).

Similarly, the $n-p-n$ transistor can also be explained with the help of Figs. 6.23(a) and 6.23 (b) where all current and voltage polarities are the reverse of those as in the case of a $p-n-p$ transistor.

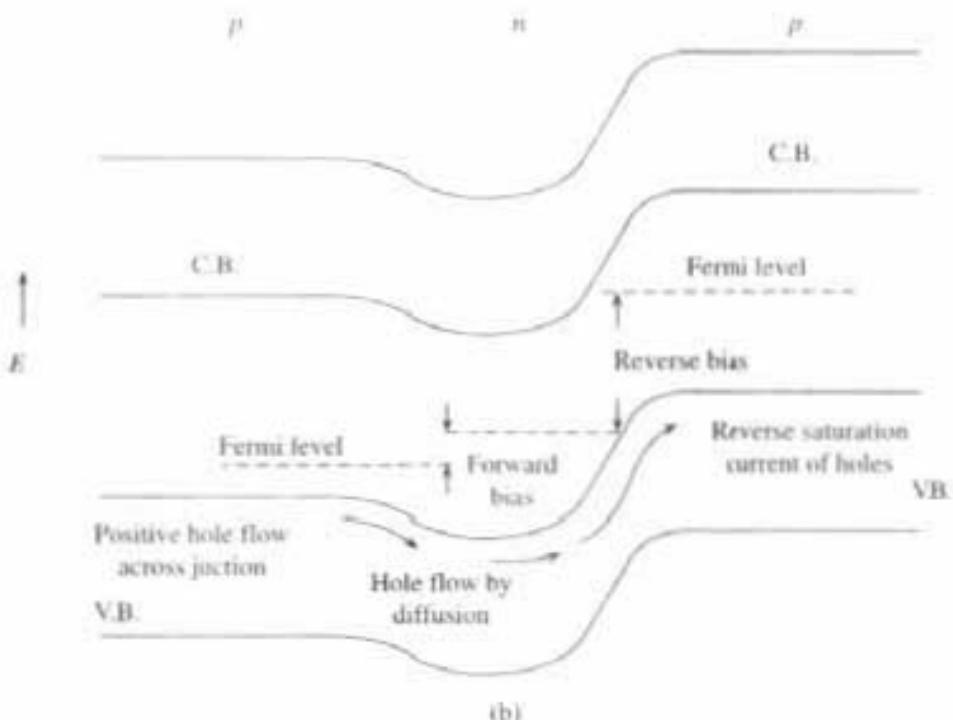
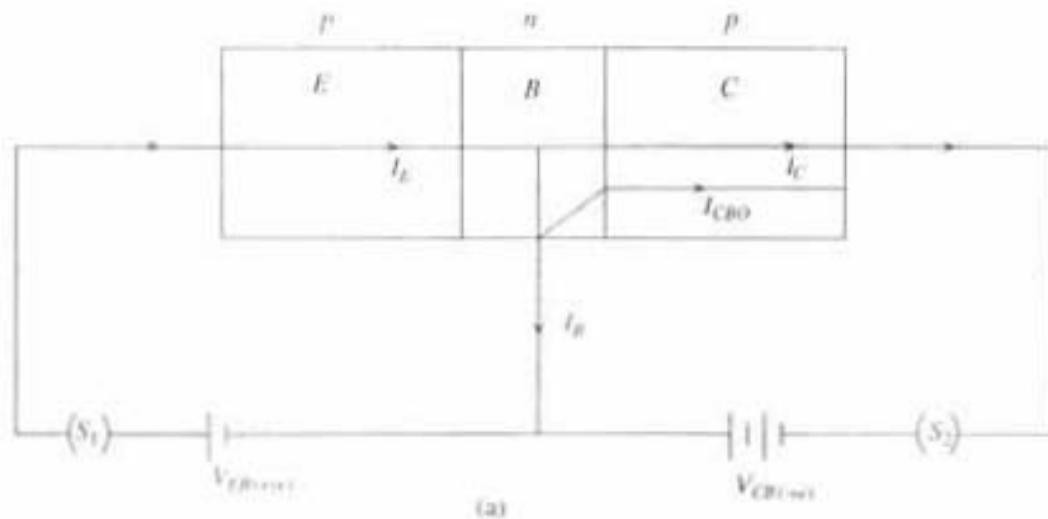


Fig. 6.22 Working procedure for hole current flow in V.B. for pnp transistor and energy level wise its explanation

Since the transistor has two junctions of emitter and collector with base, so there are four possible ways of biasing of those two junctions.

Condition	Emitter Junction	Collector Junction	Region of Operation
1. F.R.	Forward biased	Reverse biased	Active
2. F.F.	Forward ..	Forward ..	Saturation
3. R.R.	Reverse ..	Reverse ..	Cut off
4. R.F.	Reverse ..	Forward ..	Inverted

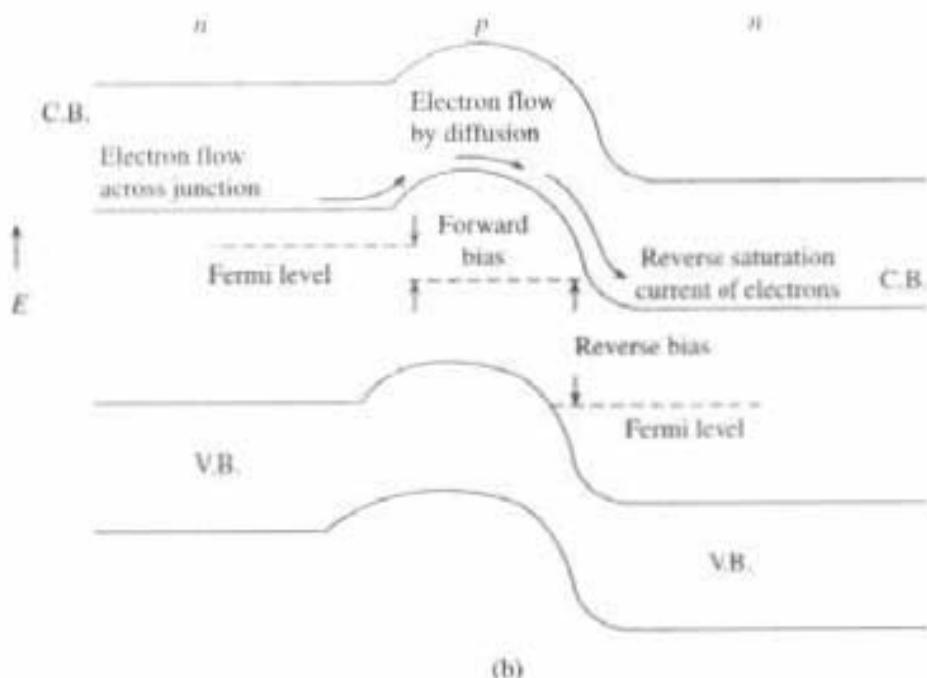
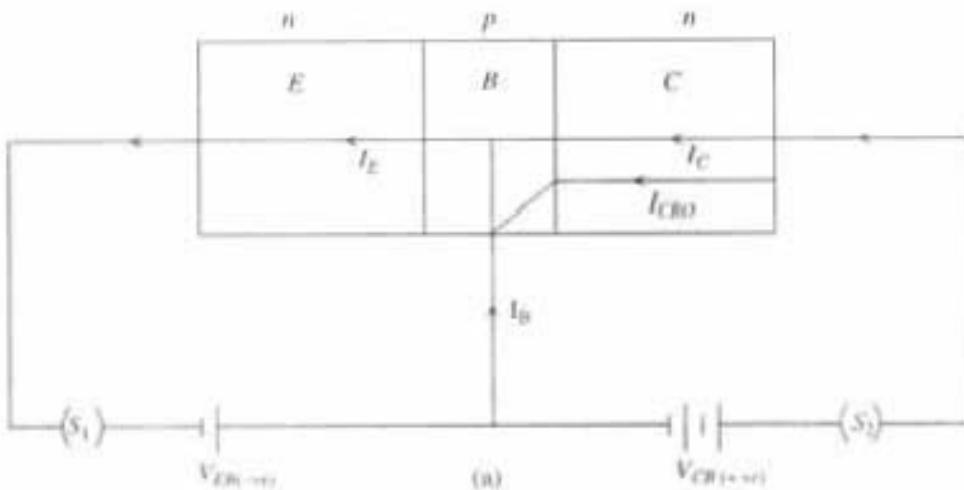


Fig. 6.23 (a), (b) Working procedure for electron current flow in C.B. for *n-p-n* transistor and energy level wise its explanation

Out of those four biasing operations, *active operation of the transistor is the most important, where emitter-base junction is always forward biased and collector-base junction is always reverse biased.*

Figure 6.22(a) shows *p-n-p* transistor under active operation. When the switches S_1 and S_2 are open both *p-n* junctions are unbiased. So no current will flow. Now as soon as S_1 and S_2 are on, emitter base will be forward bias and collector base will be reverse bias. Since emitter base is forward bias so in *p-n-p* transistor holes, the majority carrier, will diffuse from *p* to *n*, i.e. from emitter to base constitute the emitter current I_E .

Collector-base junction is made reverse biased which offers very high resistance for the majority carriers but it is helpful for minority carriers. In *p-n-p* transistor once holes cross the emitter junction thus become minority carriers in the base region. Thus, due to the diffusion of holes under forward bias condition from emitter to base, the minority carrier density has been increased in the base region. Since the base region is very lightly doped, so a very small percentage of holes are recombined with electrons, and remaining holes are now attracted by the reverse bias (V_{CB}) voltage as a minority carrier, towards collector region and constitutes collector current I_C (Fig. 6.22a).

In the base region for *p-n-p* transistor a small percentage of diffused holes from emitter to base recombine with electrons. So there is a shortage of electrons in the base region of *p-n-p* transistor. This shortage of electrons will be filled-up by the supply of electrons from (-ve) polarity of emitter-base (V_{EB}) voltage. This contributes the base current I_B .

Thus, we can write

$$I_E = I_C + I_B \quad (6.60)$$

The collector current I_C is always less than emitter current I_E by an amount base current I_B .

The reverse bias across the collector junction thus attracts the minorities carrier hole from base region in *p-n-p* transistor. As in Fig. 6.22(a) minority carrier hole will diffuse from base to collector constitute collector current (I_C). Now when emitter is open circuit, then $I_E = 0$, in this situation base and collector will be act as a reverse bias diode, then the collector current I_c will be equal to leakage current or reverse saturation current I_{CBO} , due to flow of minority carrier from collector to base. I_{CBO} means the current between the collector and base when emitter is open.

The current gain α_{dc} in C.B. configuration can be written as

$$\alpha_{dc} = I_C/I_E \quad (6.61)$$

$$I_C = \alpha_{dc} I_E \quad (6.62)$$

where

$$I_E = 0, I_C = I_{CBO}$$

Therefore, in general, equation (6.62) can be written as

$$I_C = \alpha_{dc} I_E + I_{CBO} \quad (6.63)$$

Under emitter-base forward bias condition in *p-n-p* transistor, holes will diffuse

from emitter to base region. So there is a shortage of holes in the emitter region. New holes can be created and compensate for the loss of holes in emitter region when electrons are coming out by breaking the covalent bond in the crystal and enter the +ve terminal of emitter base voltage source. Thus in the active region when V_{EB} is forward bias and V_{CB} is reversed bias, there is a continuous current flow by holes in $p-n-p$ transistor. The direction of hole flow and its energy level wise explanation is given in Fig. 6.22(b).

Thus from the direction of conventional current flow as shown in Fig. 6.22(a) for $p-n-p$ and Fig. 6.23(a) for $n-p-n$ transistor and from sign convention of transistor it is clear that,

For $pnp \cdot I_E = (+ve)$

$I_C, I_B = (-ve)$

$I_{CBO} = (-ve)$

$V_{EB} = (+ve)$

$V_{CB} = (-ve)$

For $npn \cdot I_E = (-ve)$

$I_C, I_B = (-ve)$

$I_{CBO} = (-ve)$

$V_{EB} = (-ve)$

$V_{CB} = (+ve)$

6.22 Transistor Characteristics

In Figs. 6.22(a) and 6.23(a) pnp and npm transistors, respectively, are connected in *common base* (CB) configuration, where *base* is the common for input and output. In the same way by making *emitter* or *collector* common, we can get *common emitter* (CE) or *common collector* (CC) configuration, respectively, as shown in Fig. 6.24 for pnp transistor. But in all configurations the emitter junction is always forward biased and collector junction is always reverse biased in *Active Region* of the transistor. Here we will discuss in details only CB and CE configurations, which are mostly used for practical purposes.

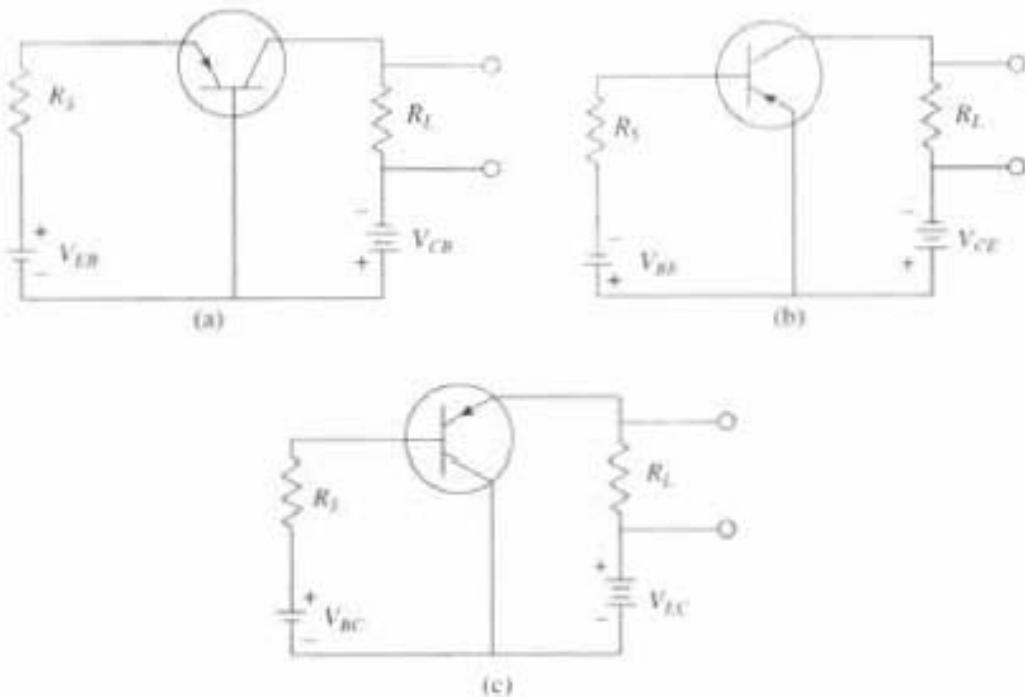


Fig. 6.24 Three PnP transistor connections: (a) CB mode (b) CE mode (c) CC mode

The relation between various currents and voltages in a transistor can be represented by curves which are known as static characteristics curves. The two most important characteristic curves for any configuration are the input and output characteristics curves. *Input characteristics curves relate the input current with input voltage for a constant output voltage. The output characteristic curves relate the output current with output voltage for constant input current.*

6.22.1 Common Base (CB) Configuration

The circuit connection for CB mode of a pnp transistor are as shown in Fig. (6.25a) and for npn transistor shown in Fig. (6.25b) Emitter base voltage (V_{EB}) and the output voltage (V_{CE}) are controlled by variable resistor. According to the polarity convention that all current that flows into the transistor is taken to be positive and that which comes out of the transistor is taken as negative. So for pnp transistor (Fig. 6.25a, 6.26a, 6.27a), I_E is positive, I_C and I_B are negative. Forward bias V_{EB} is positive and reverse bias V_{CB} is negative. For npn transistor (Fig. 6.25b, 6.26b, 6.27b) all current and voltage polarities are just reverse of those as given for pnp transistor, i.e., I_E is negative, I_C , I_B are positive, V_{EB} is negative and V_{CB} is positive.

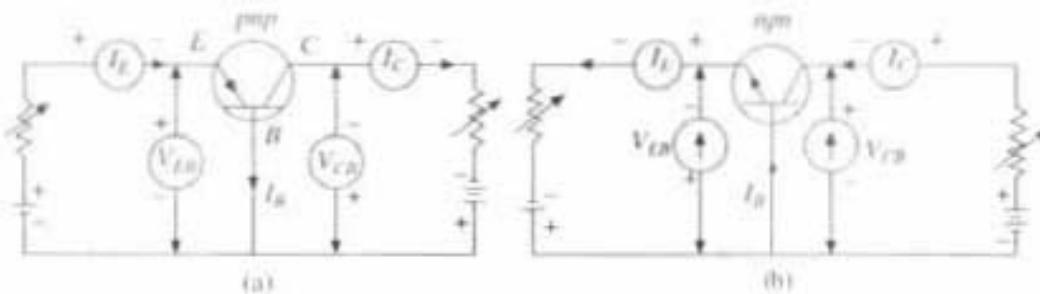


Fig. 6.25 Circuit connection for pnp and npn under C.B. Mode

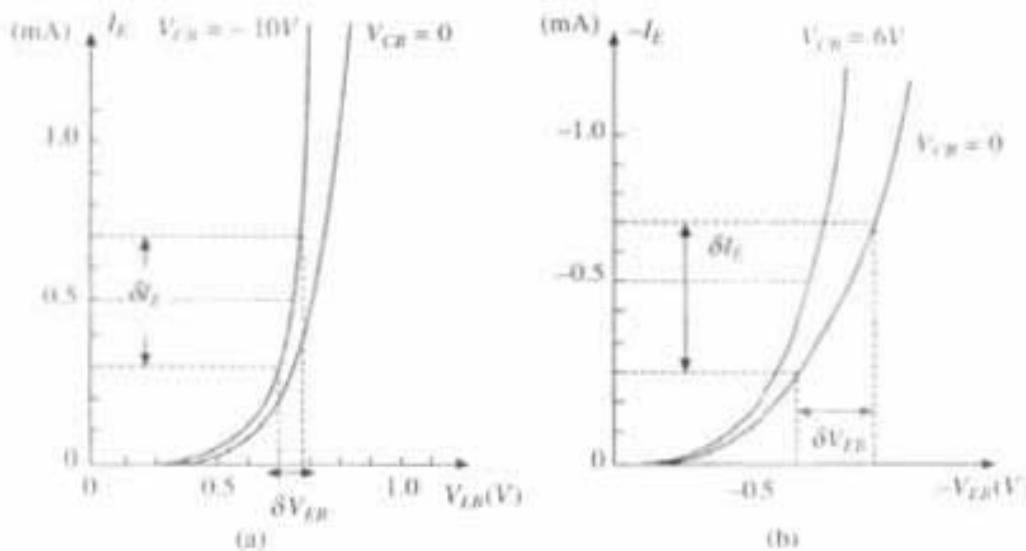


Fig. 6.26 Input characteristics for pnp and npn transistor under C.B. mode

Input Characteristics

The input characteristics curves for pnp transistor are plotted between I_E and V_{EB} for different values of V_{CB} (Fig. 6.26a). For $V_{CB} = 0$, it is just like junction diode characteristics in the forward bias direction. There is also cutin or threshold voltage V_t below which the emitter current is very small. When V_{CB} is not equal to zero, then the emitter current is higher for a given value of V_{CB} than the case when $V_{CB} = 0$, because now collector attracts the minority carrier more from base and hence base can attract more holes from the emitter. From the slope of input characteristics we can calculate dynamic input resistance (R_I) of the transistor as

$$R_I = \left. \frac{\delta V_{EB}}{\delta I_E} \right|_{V_{CB} = \text{Constant}} \quad (6.64)$$

This dynamic input resistance R_I is very small, between 20 and 100 Ω because the curve is not linear. As V_{EB} increases value of R_I decreases with the operating point. *Input characteristic of npn transistor will be similar to Fig. 6.26(a) as shown in Fig. 6.26(b) only both I_E and V_{EB} will be negative and V_{CB} will be positive.*

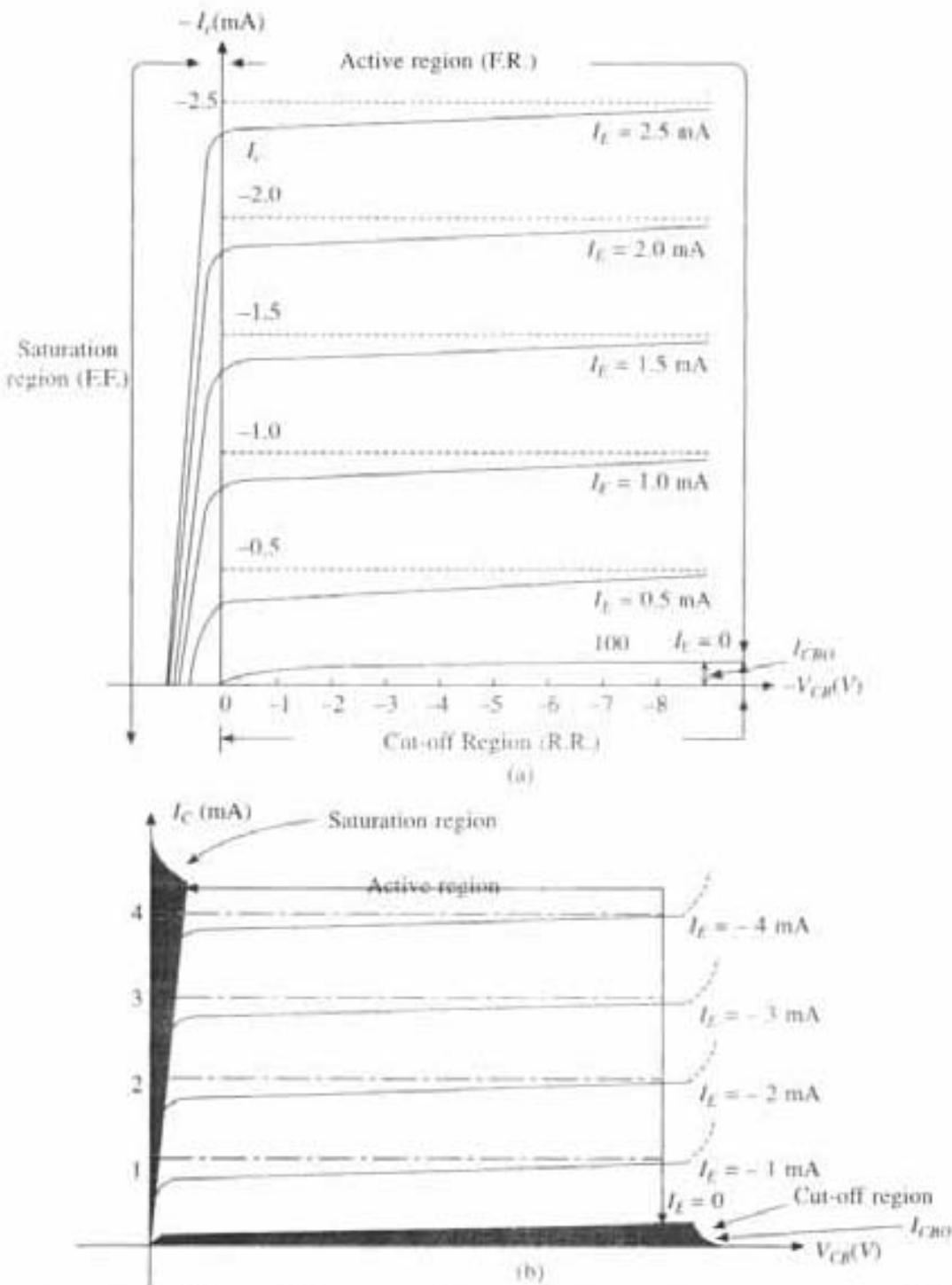
Output Characteristics

The output characteristics for same pnp transistor is shown in Fig. 6.27(a) and for npn Fig. 6.27(b). Here I_C is plotted against V_{CB} for constant values of I_E . The different important points of output characteristics are:

1. From the description of the working principle of the transistor it is known that I_C is slightly less than I_E . This is true in *active region* when emitter base is forward bias and collector base reverse bias. This is also seen from the output characteristics. The difference between I_E and I_C at any point is the base current I_B .
2. The output curves are nearly horizontal lines provided V_{CB} is reverse bias. So it behaves as a constant current source which means the transistor have very high output resistance (R_O), whereas input dynamic resistance (R_I) is very small.
3. *Saturation Region* If V_{CB} becomes forward biased, instead of reverse biased, collector current I_C sharply decreases to zero for all values of I_E . This region when emitter and collector (F.F.) both forward biased is called saturation region here I_C does not depend on I_E .
4. When $I_E = 0$, then I_C is not zero, and very small reverse saturation current I_{CBO} exists (Fig. 6.27), due to the minority carrier flow.
5. *Cut Off Region* The region below and to the right of $I_E = 0$ characteristics, for which emitter and collector junctions (R.R.) are both reverse biased is called cut off region.

Dynamic output resistance of the transistor is obtained from the output characteristics (Fig. 6.27) as

$$r_0 = \left. \frac{\delta V_{CB}}{\delta I_C} \right|_{I_E = \text{Constant}} \quad (6.65)$$

Fig. 6.27 Output characteristics for *pnp* and *npn* transistor under CB mode

Since $\delta I_C \ll \delta V_{CB}$ so the value of output resistance is very high.

In *CB* mode the current gain of the transistor is defined as α_{dc} , where

$$\alpha_{dc} = I_C/I_E = 0.98 \dots \quad (6.66)$$

From the output characteristics we can see at any point $I_E > I_C$ but the difference between I_E and I_C is very small so α_{dc} is less than but very close to unity (0.98). *Output characteristic of n-p-n transistor is shown in Fig. 6.27(b), same as pnp transistor only I_C and V_{CB} will be positive and I_E will be negative.*

6.22.2 Common Emitter (CE) Configuration

In *CE* configuration the emitter is the common for input and output. The circuit connection for *pnp* transistor is as shown in Fig. 6.28.

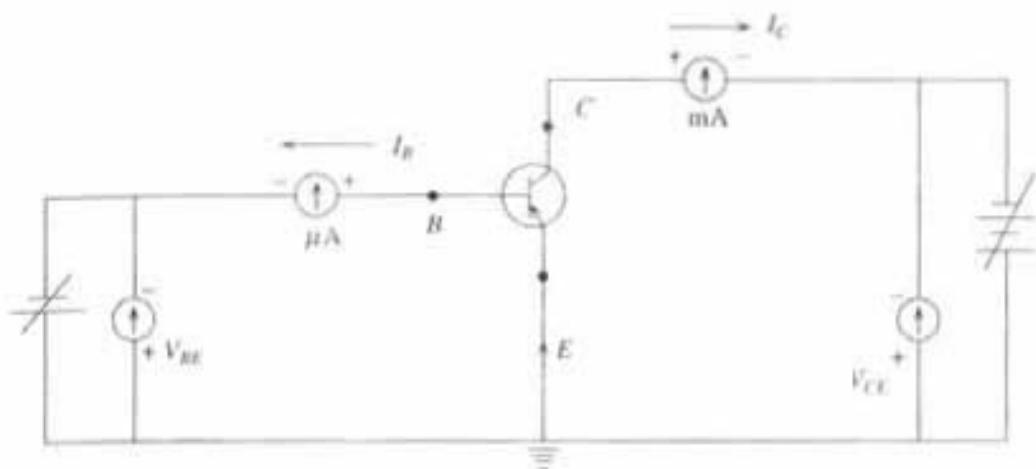


Fig. 6.28 Circuit connection for *pnp* under *CE* mode

The base emitter voltage \$V_{BE}\$ and collector emitter voltage (\$V_{CE}\$) is regulated by resistors. According to polarity convention in *CE* mode for *pnp* transistor \$I_B\$ and \$I_C\$ is (-ve) and \$I_E\$ is (+ve), \$V_{BE}\$ is (-ve) and \$V_{CE}\$ is also (-ve). These are just opposite for *npn* transistor.

Input Characteristics

Input characteristic \$I_B\$ is plotted against \$V_{BE}\$ for given value of \$V_{CE}\$ as shown in Fig. (6.29a). These curves are almost similar to those in *CB* configuration. The change in output voltage \$V_{CE}\$ does not have much effect on current \$I_B\$. The dynamic input resistance can be obtained from the inverse of the slope of the input characteristics curves as

$$r_i = \left. \frac{\delta V_{BE}}{\delta I_B} \right|_{V_{CE} \text{ const.}} \quad (6.67)$$

Since \$\delta V_{BE} \ll \delta I_B\$ so the input dynamic resistance value is very small. Input characteristics of *n-p-n* transistor will be same as Fig. 6.29(a), only \$I_B\$ and \$V_{BE}\$ will be positive and \$V_{CE}\$ will also be positive.

Output Characteristics

In output characteristics of *CE* mode for *pnp* transistor, variation of \$I_C\$ with \$V_{CE}\$ is studied for given value of \$I_B\$ (Fig. 6.29b). The different important points of the output characteristics are:

1. For *CE* mode \$I_C\$ increases slowly as \$V_{CE}\$ increases. The slope of this curve i.e. the output resistance of *CE* mode circuit is much lower than *CB* mode circuit. In *CE* mode current gain is defined as

$$\beta_{ac} = I_C / I_B \quad (6.68)$$

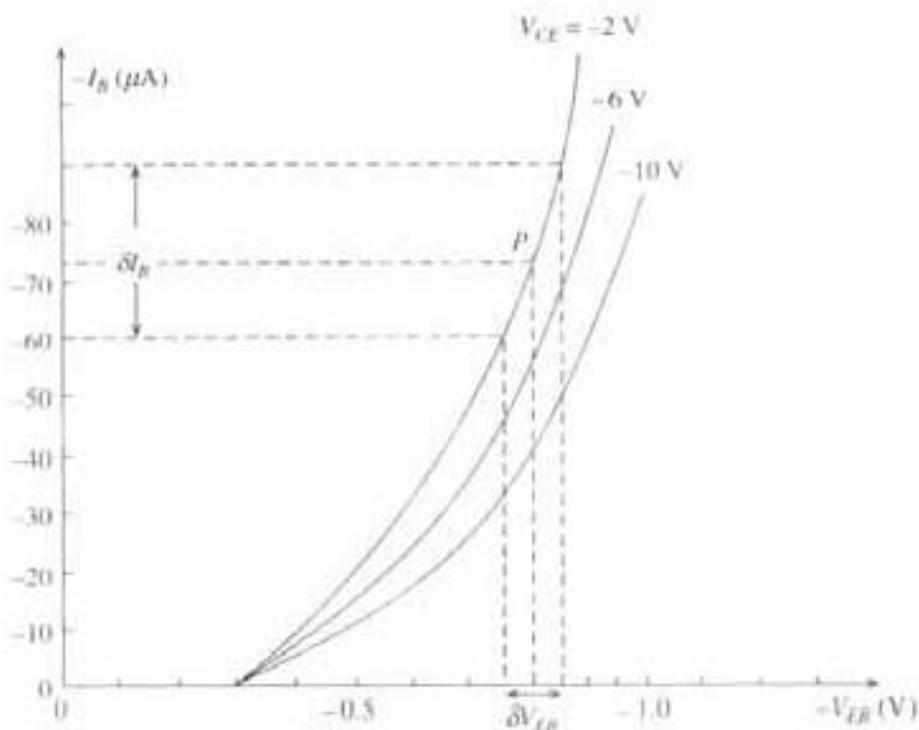


Fig. 6.29a Input characteristic for pnp transistor under CE mode

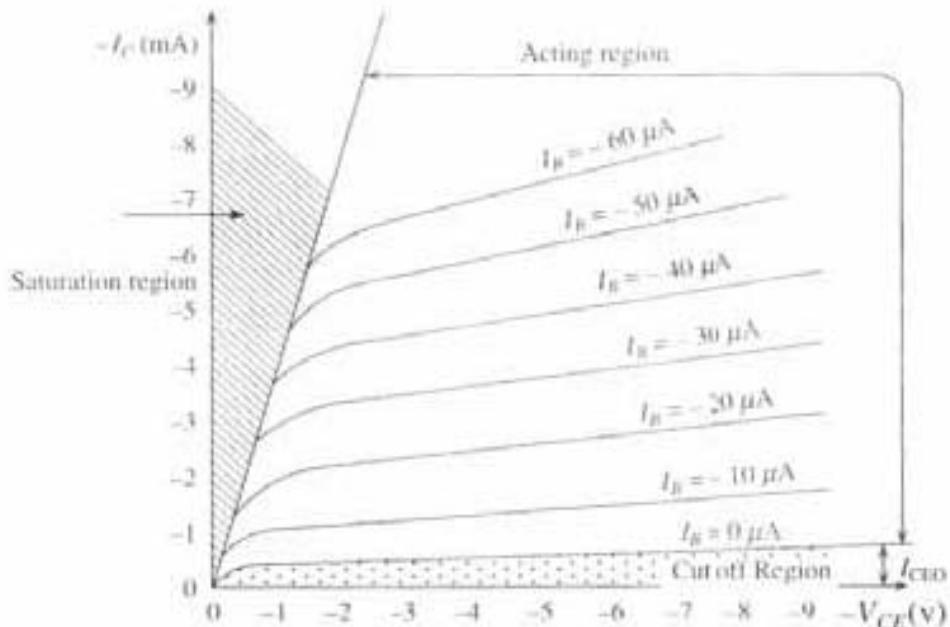


Fig. 6.29b Output characteristics for pnp transistor under CE mode

Now for each output characteristic curve I_B constant but I_C increases with V_{CE} , i.e. β_{dc} increases with V_{CE} .

2. When V_{CE} is very close to zero, say below few tenths of volts, I_C decreases rapidly as V_{CE} decreases. This occurs because V_{CE} drops below the value of V_{BE} which means that both the junctions of the transistor are forward biased. This region is called saturated region. Here I_C is not dependent on I_B .

3. The collector current I_C is not zero when $I_B = 0$. It has a value I_{CEO} , the reverse leakage current.

4. From output characteristics we can determine the following terms:

$$(1) \text{ output dynamic resistance } = r_o = \left. \frac{\Delta V_{CE}}{\Delta I_C} \right|_{I_B = \text{const}}$$

$$(2) \quad \beta_{dc} = \left. \frac{I_C}{I_B} \right|_{V_{CE} = \text{const}}, \quad \beta_{ac} = \left. \frac{\Delta I_C}{\Delta I_B} \right|_{V_{CE} = \text{const}}$$

6.23 Current Relation in C.E. Mode

For C.B. configuration I_E is the input current and I_C the output current. The relation between input and output current is as earlier

$$I_E = I_C + I_B \quad (6.60)$$

and

$$I_C = \alpha_{dc} I_E + I_{CBO} \quad (6.63)$$

Since $I_{CBO} \ll I_C$ so d.c. current gain in CB mode is

$$\alpha_{dc} = I_C/I_E \quad (6.61)$$

Now in C.E. configuration I_B is the input current and I_C the output current. Now in order to find out the relation between I_B and I_C we can proceed as follows:

In C.E. configuration, current gain is

$$\beta_{dc} = I_C/I_B \quad (6.68)$$

$$I_C = \beta_{dc} I_B$$

Here also reverse saturation current I_{CEO} is due to minority carrier with flow from collector to emitter when base is open i.e. $I_B = 0$.

Hence

$$I_C = \beta_{dc} I_B + I_{CEO} \quad (6.69)$$

6.24 Relation Between α_{dc} and β_{dc}

The D.C. Current gain of transistor in CE configuration is β_{dc} , D.C. current gain of transistor in CB configuration is α_{dc} . The relation between them is as follows:

From equations (6.63) and (6.60) we get

$$\begin{aligned} I_C &= \alpha_{dc} (I_C + I_B) + I_{CBO} \\ I_C (1 - \alpha_{dc}) &= \alpha_{dc} I_B + I_{CBO} \\ I_C &= \frac{\alpha_{dc}}{1 - \alpha_{dc}} I_B + \frac{I_{CBO}}{1 - \alpha_{dc}} \end{aligned} \quad (6.70)$$

Comparing (6.69) and (6.70), we get

$$\beta_{dc} = \frac{\alpha_{dc}}{1 - \alpha_{dc}}, \quad \alpha_{dc} = \frac{\beta_{dc}}{1 + \beta_{dc}}, \quad I_{CEO} = \frac{I_{CBO}}{1 - \alpha_{dc}} \quad (6.71)$$

When transistor is used as an amplifier then input will be A.C. signal instead of d.c. Then

$$\alpha_{AC} = \frac{\delta I_C}{\delta I_E}, \quad \beta_{dc} = \frac{\delta I_C}{\delta I_B}$$

Thus

$$\alpha_{AC} = \frac{\beta_{AC}}{1 + \beta_{AC}}, \quad \beta_{AC} = \frac{\alpha_{AC}}{1 - \alpha_{AC}}$$

or in general

$$\boxed{\alpha = \frac{\beta}{1 + \beta}, \quad \beta = \frac{\alpha}{1 - \alpha}} \quad (6.72)$$

PROBLEMS

1. If the forward current in the case of a junction diode is 60 mA when the voltage across it is 1.4 V and 85 mA then the voltage is 1.5 V. Find static and dynamic resistance.

Solution Static Resistance, $R_s = V/I = 1.4/(60 \times 10^{-3}) = 23.3 \Omega$
 $= 1.5/(85 \times 10^{-3}) = 17.6 \Omega$

Dynamic Resistance $R_d = dV/dI = (1.5 - 1.4)/(85 - 60) \times 10^{-3} = 6.6 \Omega$

2. For a transistor used in CB mode the emitter current is 1 mA and the base current is 0.02 mA. Find α and β .

Solution $I_E = I_C + I_B$
 $I_C = I_E - I_B = 1 - 0.02 = 0.98 \text{ mA}$
 $\beta = I_C/I_B = 0.98/0.02 = 49$
 $\alpha = \beta/(1 + \beta) = 49/50 = 0.98$
 $\alpha = I_C/I_E = 0.98/1 = 0.98$

3. (i) Show that $I_{CEO} = 1/(1 - \alpha) I_{CBO}$.
(ii) The α factor of a transistor is 0.99. Find β . If $I_{CBO} = 0.5 \mu\text{A}$, find I_{CEO} .

Solution $\beta = \alpha/(1 - \alpha) = 0.99/(1 - 0.99) = 99$

$$I_{CEO} = 1/(1 - \alpha) I_{CBO} = [1/(1 - 0.99)] \times 0.5 = 50 \mu\text{A}$$

4. For a transmitter in C.E. mode, collector current changes from 2 to 4.5 mA when base current is changed from 40 to 80 μA , keeping collector-emitter voltage constant. Calculate α and β .

$$\Delta I_C = 2.5 \text{ mA}$$

$$\Delta I_B = 40 \mu\text{A}$$

$$\beta_{AC} = \Delta I_C / \Delta I_B = 62.5$$

$$\alpha_{AC} = \beta_{AC} / (1 + \beta_{AC}) = 0.9843$$

5. In the C.B. mode the emitter current is 1 mA and the output current is 0.04 mA less. Find current gain.

Solution $I_E = 1 \text{ mA}$.

$$I_C = 1 - 0.04 = 0.96 \text{ mA}$$

Current gain $\alpha = I_C/I_E = 0.96$

6. In C.B. mode, current gain is 0.96. The voltage drop across 1 K resistance connected in the collector circuit is 1.5 V. Find the base current.

Solution

$$V = I_C \cdot R$$

$$I_C = V/R = 1.5/1 \times 10^3 = 1.5 \text{ mA.}$$

Current gain = $I_C/I_E = \alpha$

$$I_E = I_C/\alpha = 1.5/0.96 = 1.5625$$

Now

$$I_E = I_C + I_B$$

$$I_B = I_E - I_C = 1.5625 - 1.5$$

$$I_B = 0.0625 \text{ mA}$$

QUESTIONS

- What is potential barrier? How is it formed in *p-n* junction?
- Discuss forward bias and reverse bias characteristics of semiconductor diode.
- What is the difference between a metal, insulator and semiconductor from the energy level point of view?
- Discuss *n-p-n* and *p-n-p* transistor working principle in terms of carrier movements across junctions. Why it is called BJT?
- Discuss C.B. and C.E. mode configuration for *p-n-p* transistor, with circuit diagram.
- Show that current relation in C.B. mode is

$$I_C = \alpha_a I_E + I_{CEO}$$

and current relation in C.E. mode is

$$I_C = \beta_a I_E + I_{CBO}$$

7. Prove that

$$I_{CEO} = \frac{I_{CBO}}{1 - \alpha_a}$$

and

$$\alpha = \frac{\beta}{1 + \beta}$$

- Describe the formation of energy bands in crystalline solid. Define valence band, conduction band, and forbidden gap in energy band structure. Explain hole as positive charge carrier.
- Explain how depletion layer is formed in *p-n* junction and explain different types of current flowing across *p-n* junction.
- Describe four biasing conditions of a transistor.
- Explain how energy bands are formed in a solid, as isolated atoms are brought together to form the solid. Classify solids into conductors, insulators, and semiconductors in terms of their energy band and hence explain how their conductivity depends on temperature.

12. Derive the equation for conductivity for metal as well as for semiconductor.
13. Explain how Hall effect shows whether holes or electrons predominate in a semiconductor and derive the expression for Hall co-efficient, Hall angle and Hall voltage for P-type as well as for N-type semiconductor.
14. Derive Rectifier equation for P-N junction with proper diagram.
15. Calculate the density of donor atom which has to be added to intrinsic Ge to produce n-type material of conductivity 0.19 ohm cm . It is given that the mobility of electrons in n-type semiconductor is $3250 \text{ cm}^2/\text{volt. sec.}$.
16. An n-type Ge has a donor density of 10^{17} per c.c. . It is arranged in a Hall effect experiment where $B = 0.5 \text{ wb/m}^2$ and current density $J = 500 \text{ amp./m}^2$. What is the Hall voltage if the specimen is 3 mm thick?
17. The resistivity of intrinsic Ge at 300° K is 0.43 ohm cm . Calculate the electron and hole concentration when $\mu_e = 0.38$ and $\mu_h = 0.19 \text{ m}^2/\text{volt. sec.}$

Magnetic Materials

7.1 Atomic Model and Magnetization

Difference in behaviour of various types of material in a magnetic field can be accurately analyzed with the help of quantum mechanics, which is out of the scope of this text. But simple atomic model which assumes that there is a central positive nucleus surrounded by electrons in various orbits gives reasonable qualitative results.

An electron rotating in a orbit around nucleus Fig. (7.1), is analogous to a small current loop which will experience a torque in an external M-field induction (\bar{B}_o). That torque tries to align the magnetic dipole moment (\bar{m}) produced by the orbiting electron, along the external M-field induction (\bar{B}_o). *Thus the resultant M-field induction (\bar{B}), at any point in the material would be greater than it would be at that point, if the material were not present.*

An orbiting electron around the nucleus as shown in the Fig. (7.1), behaves as a magnetic dipole having a magnetic moment (\bar{m}), directed into the plane of the paper and perpendicular to the plane of the paper. Electrons are not only rotating around the nucleus, they are also spinning about themselves. In addition to the electrons movement, nucleus itself also has a spin motion associated with it. Although this nuclear spin has a very negligible effect on the overall magnetic property of the material. Thus each atom can be characterized by the superposition of the magnetic dipole moments corresponding to the electrons orbital motion, electron spin motion and nuclear spin motion. So the effective magnetic dipole moment can be written as

$$\bar{m} = I A \hat{n} \text{ amp. m}^2 \quad (7.1)$$

Where I = current in the loop, A = area of the electron orbital loop, \hat{n} = unit normal vector, normal to the plane of the loop.

On macroscopic scale the vector \bar{M} is defined as the magnetization vector, = magnetic dipole moments per unit volume

i.e.

$$\bar{M} = N \bar{m} \text{ amp./m} \quad (7.2)$$

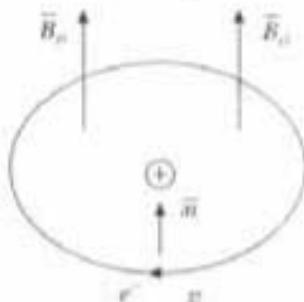


Fig. 7.1 Orbital motion of electron

where N = Number of molecules per unit volume

and $\bar{m} = IA\hat{n}$ = Average dipole moment per molecule amp. m²

Thus magnetization vector (\bar{M}) have the same unit amp/m as M-field intensity (\bar{H}). So resultant M-field induction (\bar{B}) in the presence of a magnetic material will be

$$\bar{B} = \mu_0(\bar{H} + \bar{M}) \quad (7.3)$$

In most of the non-ferro magnetic material, magnetization vector (\bar{M}) is proportional to the M-field intensity (\bar{H}) and in the same direction of (\bar{H}). Thus we can write

$$\bar{M} = \chi_m \bar{H} \quad (7.4)$$

where (χ_m) is the constant of proportionality called as *magnetic susceptibility* of the material. This relation applied for large class of paramagnetic and diamagnetic materials. So the above eqn. (7.3) now becomes

$$\bar{H} = \frac{\bar{B}}{\mu_0} - \chi_m \bar{H}$$

or $\bar{B} = \mu_0(1 + \chi_m)\bar{H} = \mu_0(\bar{H} + \bar{M}) = \mu H \quad (7.5)$

where the coefficient (μ) is called the *magnetic permeability*, which is written as

$$\mu = \mu_0(1 + \chi_m)$$

or $\frac{\mu}{\mu_0} = \mu_r = (1 + \chi_m)$ or $\chi_m = (\mu_r - 1) \quad (7.6)$

where dimension less quantity (μ_r) is called as *relative permeability*.

7.2 Molecular Field Inside a Magnetic Material

Molecular field inside a magnetic material may be calculated in the following way. Lets consider a magnetic material of arbitrary shape, which is uniformly magnetized with magnetization vector (\bar{M}). Now lets consider an imaginary spherical cavity in the magnetic material, at the centre of which the molecular field is to be calculated. The material outside that cavity is treated as continuum from the macroscopic point of view, whereas inside the cavity the dipoles are treated as individual dipoles or dipole groups not as continuum.

Total internal molecular field at the centre of the cavity can be written as

$$\bar{H}_m = \bar{H} + \bar{H}_1 + \bar{H}_2 \quad (7.7)$$

where \bar{H} = External macroscopic M-field intensity applied to the specimen.

\bar{H}_1 = Contribution to the M-field intensity at the centre of the cavity due to surface pole density ($\sigma_m = M_n$) on the cavity surface, where M_n is the normal component of magnetization vector \bar{M} .

\bar{H}_2 = Contribution to the M-field intensity due to various dipoles inside the cavity.

Now if there are large number of dipoles inside the cavity, which are oriented parallelly but randomly distributed in position, then from cubical symmetry consideration $\bar{H}_2 = 0$, in general.

The surface contribution will come out as

$$\bar{H}_1 = \bar{M}/3 \quad (7.8)$$

So from eqn (7.7) the internal molecular field inside a magnetic material can be written as

$$\bar{H}_m = \bar{H} + \bar{M}/3 \quad (7.9)$$

and internal molecular flux density

$$\bar{B}_m = \mu_0 \bar{H}_m \quad (7.10)$$

Equations (7.9 and 7.10) give the internal molecular field inside a magnetic material in terms of applied macroscopic M -field intensity (\bar{H}) and the magnetization vector (\bar{M}). For most of the diamagnetic and paramagnetic materials the second term (i.e., $M/3 = \chi_m H/3$) is negligibly small (i.e., $\bar{H}_m = \bar{H}$), but for ferromagnetic materials that correction is very important, as it will show spontaneous magnetization even in the absence of the external applied M -field, (i.e., $\bar{H}_m = \bar{M}/3$, even though $\bar{H} = 0$).

7.3 Various Magnetic Materials

Each atom contains different components of magnetic dipole moments due to electron orbital motion, electron spin motion and nuclear spin motion as discussed above in section (7.1), and their total effect determines the overall magnetic characteristics of the material and provide its general magnetic classification. Following are the different types of magnetic materials.

7.3.1 Diamagnetic Material

In diamagnetic materials the orbital motion of the electron and spin motion of the electron produce the M -field opposite to each other, so net magnetic dipole moment of each atom (\bar{m}) is zero and net M -field induction (\bar{B}) inside diamagnetic materials is also zero, in the absence of the external M -field. Thus an external M -field cannot produce any torque on the atom, and no realignment takes place.

But in the presence of external M -field induction (\bar{B}_o) as shown in Fig. (7.1) the orbiting electron either speeds up or slows down in its orbit depending upon whether the orbital magnetic moment is anti parallel or parallel to the applied M -field induction (\bar{B}_o), but in either case the change in orbital magnetic moment is in a direction opposite to the applied M -field induction (\bar{B}_o). So the overall magnetic induction (\bar{B}) is weakened by the presence of the diamagnetic material.

For a diamagnetic substance containing N dipoles per unit volume, the magnetization vector can be written as

$$\bar{M} = N\bar{m}_{ind} = -\frac{Ne^2}{4m} R^2 \bar{B}_m = -\frac{Ne^2}{4m} R^2 \mu_0 H_m$$

where

$\bar{B}_m = \mu_0 \bar{H}_m$, N = no. of dipoles per unit vol., R = radius of the electron orbit loop. For diamagnetic material $\bar{H}_m = \bar{H}$, so Diamagnetic Susceptibility (χ_{dm}) will be

$$\chi_{dm} = M/H = -\frac{Ne^2}{4m} R^2 \mu_o \quad (7.11)$$

i.e. $\chi_{dm} = M/H = -\frac{Ne^2}{4m} R^2 \mu_o < 1 \text{ (-ve)} \quad (7.12)$

Thus diamagnetic substance having ($\chi_{dm} < 1$), (-ve), and temperature independent. So the magnetic vector (\bar{M}) is also (-ve) in Eqn. (7.5), which proves that *magnetic induction (\bar{B}) is weakened by the presence of diamagnetic material*. Diamagnetism is the result of Lenz's law operating on an atomic scale. Due to the application of a M -field, the electronic current in each atom are modified in such a way that they tend to weaken the effect of this applied field. Since it is a consequence of magnetic moment induced in the atom by the external M -field, so in this sense *all material are diamagnetic*. It is particularly present in materials that consists entirely of atoms or ions with closed electronic shells. Due to that paramagnetic contribution gets cancelled out. In other materials diamagnetism is frequently masked by stronger paramagnetic or ferromagnetic behaviour. When an external M -field is applied to the diamagnetic substances it expels the magnetic lines of forces due to Lenz's Law, as shown in the Fig. (7.2). $\bar{B} = 0$, for diamagnetic substances. So $\bar{B} = \mu_o$ ($\bar{H} + \bar{M}$) = 0 i.e. $\chi_{dm} = M/H = (-ve)$

Examples of diamagnetic substances are Cu, Ag, Zn, Bi, Au, C, Cd, Hg, Pb, Si, etc.

7.3.2 Paramagnetic Material

In paramagnetic materials the orbital magnetic moment and spin magnetic moment of the electron are not cancelling each other, then atom as a whole has a magnetic moment (\bar{m}_p), but random orientation of the atoms in a material makes the overall average magnetic moment zero. Due to that in the absence of the external M -field those materials will not show any magnetic property. But in the presence of external M -field a torque will apply to each atomic magnetic moment and try to align them along with applied M -field. *So the magnetic induction (\bar{B}) is strengthened by the presence of these type of materials*, which are called as paramagnetic materials. Since the diamagnetic effect is still here which decreases M -induction (\bar{B}), so the net increase in (\bar{B}) is small. So in this case paramagnetic susceptibility (χ_{pm}) is small but positive. With increase of the thermal energy random orientation of the atomic dipoles will increase. *So with the increase of temperature overall M -induction (\bar{B}) will decrease, i.e., (χ_{pm}) is inversely proportional to absolute temperature (T).* Generally when the free atoms or ions have partly filled inner shells or odd number of electrons, then they possess paramagnetic behaviour.

For all practical purposes except for temperature near absolute zero, magnetization vector (\bar{M}) for paramagnet material will become

$$\bar{M} = \frac{Nm_p^2 \mu_o \bar{H}_m}{3KT} \quad (7.13)$$



Fig. 7.2 Expulsion of M -lines

For Paramagnetic materials $\bar{H}_m = \bar{H}$ which gives the paramagnetic susceptibility

$$\chi_{pm} = M/H = \frac{Nm_p^2\mu_0}{3KT} \text{ small (+ve)} \quad (7.14)$$

where ($\bar{H}_m = \bar{H}$) is the M -field intensity inside the paramagnetic material or internal molecular field, (N) is the number of atoms or molecules per unit volume, (m_p) is the atomic or molecular dipole moment, which is in the range of few Bohr magneton (1 Bohr magneton = $eh/4\pi m$, h = Planck's constant), (K) is the Boltzmann's constant, (T) is the absolute temperature in degree Kelvin.

Above eqn (7.14) can also written as

$$\chi_{pm} = C/T \quad (7.15)$$

where the constant $C = \frac{Nm_p^2\mu_0}{3K}$ is known as *Curie constant*, and the eqn. (7.15) is known as *Curie's Law*, which shows that the paramagnetic susceptibility (χ_{pm}) is inversely proportional to the absolute temperature (T).

Examples of paramagnetic materials are Pt, Mg, Al, Cr, Mn, Sn, V, W, etc.

7.3.3 Ferromagnetic Material

In ferromagnetic material each atom or molecule has a strong dipole moment (m_p) due to uncompensated electron spin moment, which are very nearly aligned even in the absence of an external M -field, it is also having partly filled inner electronic shell. The cause of this alignment is the internal molecular field (\bar{H}_m) which, according to eqn. (7.9), does not vanish when ($\bar{H} = 0$), unless \bar{M} vanishes simultaneously.

The generalized version of internal molecular field (\bar{H}_m), eqn. (7.9) can be written as

$$\bar{H}_m = \bar{H} + \gamma \bar{M} \quad (7.16)$$

where $\gamma = 1/3$, according to the simple theory. But for ferromagnetic materials (γ) is having a large value of the order of ($\approx 10^3$).

So for ferromagnetic materials when ($\bar{H} = 0$), i.e., in the absence of the external M -field, the internal molecular field will still be there which is called as *spontaneous magnetization*, in that case above eqn. (7.16) modified as

$$\bar{H}_m = \gamma \bar{M} \quad (7.17)$$

The molecular field \bar{H}_m as given in eqn. (7.17) is known as Weiss Molecular Field.

Weiss-Heisenberg theory predicts how magnetization changes in ferromagnetic material with temperature. The theory showed that ferromagnetism is the limiting case of paramagnetism in an extremely large M -field. A temperature which is called as *Curie temperature (T_c)*, above which spontaneous magnetization vanishes and ferromagnetic material becomes ordinary paramagnetic material.

Now when $T \gg T_c$, magnetisation vector (\bar{M}) for ferromagnetic material can be written as

$$\bar{M} = \frac{N\mu_0 m_p^2 \bar{H}_m}{3KT} = \frac{N\mu_0 m_p^2}{3KT} (\bar{H} + \gamma \bar{M}) \quad (7.18)$$

$$\text{i.e. } \bar{M} \left(1 - \frac{\gamma N \mu_o m_p^2}{3KT} \right) = \left(\frac{N \mu_o m_p^2}{3KT} \right) \bar{H}$$

So Ferro magnetic Susceptibility (χ_{fm}) can be written as

$$\chi_{fm} = M/H = \frac{N \mu_o m_p^2}{3KT} / \left[1 - \left(\frac{\gamma N \mu_o m_p^2}{3KT} \right) \right]$$

$$\text{i.e. } \chi_{fm} = \frac{(N \mu_o m_p^2 / 3K)}{T - \gamma (N \mu_o m_p^2 / 3K)} = \frac{C}{T - \Theta} > 1 \quad (7.19)$$

Where $C = N \mu_o m_p^2 / 3K$ = Curie Constant, and $\Theta = \gamma N \mu_o m_p^2 / 3K$ is the Paramagnetic Curie temperature, which is slightly higher than ($\Theta > T_c$) Curie temperature (T_c) as shown in the Fig. (7.3) though theoretically they are same.

The above eqn. (7.19) is called as *Curie Weiss Law*, which is valid above Curie temperature (T_c), when ferromagnetic material behave as paramagnetic material.

Examples of ferromagnetic materials are Fe, Ni, Co etc.

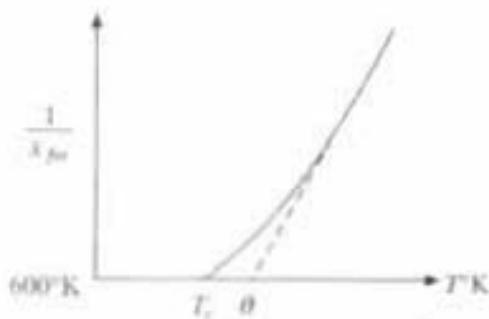


Fig. 7.3 Paramagnetic Curie temperature

7.3.4 Anti-Ferromagnetic Material

In anti-ferromagnetic material, forces between adjacent atoms cause the atomic dipole moments to line up in an anti-parallel way as shown in the Fig. (7.4b), which cause the net magnetic moment zero and they are affected slightly by the presence of external H -field. Mn, Cr exhibit antiferromagnetism at room temperature. MnS, Cr_2O_3 , MnO are example of antiferromagnetic ionic compound. Anti-ferromagnetic materials don't have much technical importance.



Fig. 7.4 (a) Ferromagnet; (b) Anti ferromagnet and (c) Ferrimagnet

7.3.5 Ferrimagnetic Material

Ferrimagnetic substances have most general ordered spin structure, contains both spin up and spin down dipole moment components, but of unequal magnitude as shown in the Fig. (7.4c), due to which there is a net non-zero magnetic moment in one of these directions. They are called ferrimagnet or ferrites. Ferrimagnetic materials are ceramic materials, are good electrical insulators. Ferrites mainly consists of Fe_2O_3 combined with one or more oxides of divalent metals, generally represented as $M\text{Fe}_2\text{O}_4$ or $M\text{O}\cdot\text{Fe}_2\text{O}_4$ e.g. $\text{FeO}\cdot\text{Fe}_2\text{O}_3$.

$Zn_0.5Fe_2O_4$ etc. Composite ferrites which are solid solution of one ferrite in another, like Ni-Zn shows best magnetic properties. Garnets like $Y_3Fe_5O_{12}$ Yttrium iron garnet (YIG) etc. constitute another group of ferrimagnetic materials. They have considerable technical importance because in addition to the relatively large magnetization, they are very poor conductor of electricity, several orders of magnitude less than semiconductors, have very high resistance. Due to that they can be used for high frequency applications, where the reduced currents (eddy currents) lead to the lower ohmic losses in the transformer.

7.4 Ferromagnetic Domain and Magnetization Curve

Now according to the above section (7.3.3) ferromagnetic materials should be magnetized nearly saturation at temperatures below Curie temperature (T_C). But say for example (Fe), which is a ferromagnetic material exists both in magnetized or unmagnetized condition. The reason for this is that a ferromagnetic material breaks up into number of domains. Domains may have different sizes and shapes, ranging from one micrometer to several centimeter. Each domain contains permanent dipoles and inter atomic forces cause these moments to line up in parallel over a region which contains a large number of atoms. These regions are called as *domain*, as shown in the Fig. (7.5a). The presence of domains was first postulated by Weiss in 1907. Although each domain is magnetized to saturation according to the above section (7.3.3), but the direction of the magnetization in various domains are randomly oriented as shown in the Fig. (7.5a). The region between two domains is called the domain wall. So due to overall cancellation net magnetization is zero, in the absence of any external M -field.

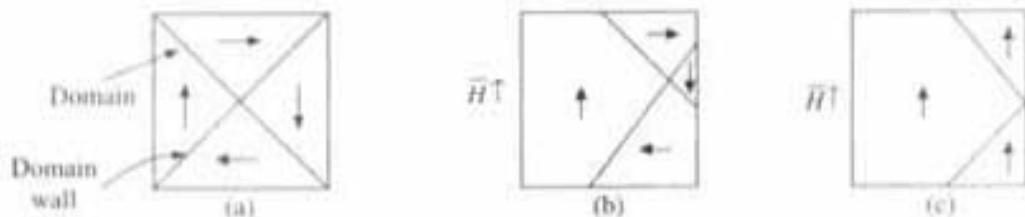


Fig. 7.5 With increase of applied M -field (\bar{H}), change of directions of moments in different domains.

Now due to the presence of the external M -field, there is an increase in the volume of the domains whose moments are favourably oriented relative to the external M -field intensity (\bar{H}), at the expense of neighbouring domains whose moments are unfavourably oriented as shown in the Fig. (7.5b), by the process of domain wall motion. Now if the external M -field is small enough then this process of domain wall motion is reversible with the removal of the external field. But as the M -field intensity (\bar{H}) becomes stronger domain wall motion continues but in an irreversible manner. During this stage movements often takes place in large jump, which is known as *Barkhausen jump*, due to which some irregularity may take place in the magnetization curve. With further increase of external field, domain wall motion is accompanied by domain rotation, i.e., the alignment of magnetization in the individual domain, along the direction of the external field (\bar{H}) takes place as shown in Fig. (7.5c) and thereby magnetize the

whole material to a saturation value (\bar{H}_m), along the direction of the external M -field intensity (\bar{H}).

Different specific portions of the whole process are shown in the Fig. (7.6a) along the curve (OA), which is called as *magnetization curve*, where M -field induction (\bar{B}) versus M -field intensity (\bar{H}) is drawn and it is called as (BH) curve. Now after reaching the saturation value (\bar{H}_m) at (A), when the external field (H) is removed gradually (i.e., $\bar{H} \rightarrow 0$), a complete random domain alignment is not usually attained, a residual or remanent dipole field still remains in the material. That is to say that the *magnetic state of the material is a function of its magnetic history* which is called as *hysteresis*.

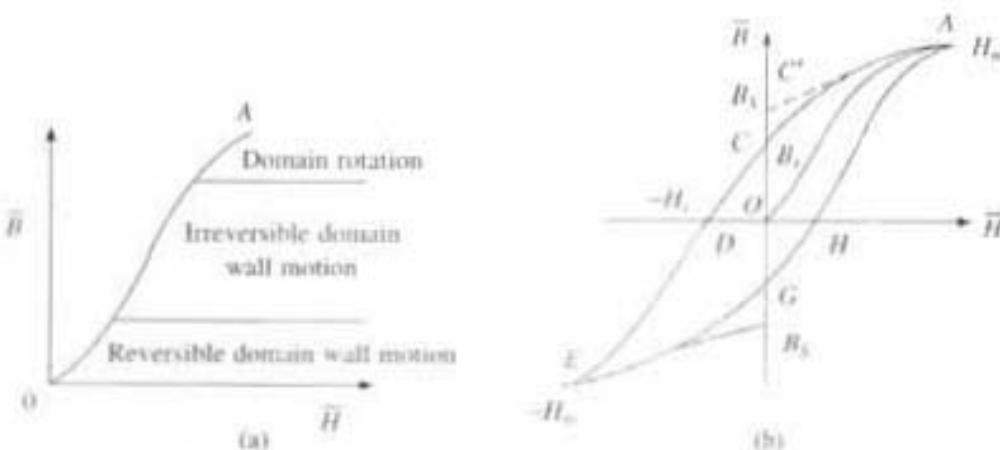


Fig. 7.6 (a) Magnetization curve (b) Hysteresis loop

Now consider the Fig. (7.6b) after reaching the saturation point (A) when ($\bar{H} = \bar{H}_m$), if M -field intensity (\bar{H}) starts reducing then effects of hysteresis starts, as it is shown in the Fig. (7.6b). Then it will not retrace its original curve (i.e., AO) but decreases along a path (AC). At the point (C) when ($\bar{H} = 0$), then ($\bar{B} = \bar{B}_r$) not zero, which is called as *remanence, retentivity or residual flux density*. The point C' (\bar{B}_s) shows the spontaneous magnetization when ($\bar{H} = 0$). To reduce the magnetic induction ($\bar{B} = 0$) at the point (D), it is required that a reversed or (-ve) value (\bar{H}_c) should be applied, which is known as *coercive force*. By changing (H) to ($-\bar{H}_m$) along (DE) the value of ($-\bar{B}_m$) results at the point (E). By returning to ($\bar{H} = 0$) at the point (G) when ($\bar{B} = -\bar{B}_r$). Finally by the process of completing the cycle from ($\bar{H} = 0$) to ($\bar{H} = \bar{H}_m$) along the curve (GHA), form a closed loop as shown in the Fig. (7.6b), which is known as *hysteresis loop*.

QUESTIONS

- Derive the relation between \bar{B} , \bar{H} , and M .
- Derive the expression of molecular field inside a magnetic material.
- Differentiate between different magnetic materials like Diamagnetic, Paramagnetic and Ferro magnetic material.
- Explain spontaneous magnetization, ferromagnetic domain and magnetization curve for ferromagnetic material.

PART-2

Interference, Diffraction and Polarisation

1A Thin Film Interference of Light

1.1 Wave Nature of Light

The phenomena of interference, diffraction, polarisation can be explained by the wave motion of light. This motion is periodic in nature. At the end of definite interval of time, the particles find themselves in the same position and having same velocity. These movements are handed on from one particle to the next in a definite phase relationship. This type of periodic movement of particles is called simple harmonic motion (SHM).

At any instant of time t the displacement of the particle in it is given by

$$y = a \sin \omega t = a \sin \frac{2\pi t}{T} \quad (1.1)$$

where ω is the angular frequency ($= 2\pi/T$) and T the time period.

Obviously the displacement varies with time. Consider a particle P (Fig. 1.1) at a distance x from starting point 0. If v be the velocity of the particle, then the particle P starts vibrating some time later than 0, given by $t_1 = x/v$ and its displacement at any instant of time t would be the same as that of 0, t_1 sec earlier, and it is represented by

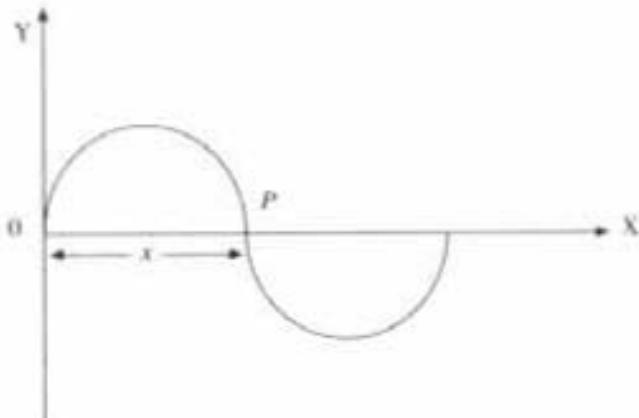


Fig. 1.1 Wave propagation of light

$$y = a \sin \frac{2\pi}{T} (t - t_1) = a \sin 2\pi \left(\frac{t}{T} - \frac{x}{vT} \right)$$

Now $vT = \lambda$ = wavelength.

Thus

$$y = a \sin 2\pi(t/T - x/\lambda) \quad (1.2)$$

Thus displacement of particle P at any instant of time t is

$$y = a \sin (\omega t - \delta) \quad (1.3)$$

$\delta = 2\pi x/\lambda$ = phase difference between the particles at 0 and P ; P lags behind 0 in phase δ .

1.2 Interference

When light waves from two coherent sources mix up and cross each other's path then a modification of intensity of light will occur in the region of crossing. This modification in the distribution of light energy obtained by the superposition of two or more waves is called *interference*.

1.3 Analytical Discussion and Conditions for Constructive and Destructive Interference

Let S_1 and S_2 be two point sources of light separated by a distance d , the monochromatic waves of wavelength λ are sent by them (Fig. 1.2).

Let us assume that waves are starting from S_1 and S_2 , coherent in nature, i.e., they are, in same phase, same amplitude, same state of polarisation and travelling nearly along the same direction. So they can produce an interference pattern.

Let the instantaneous displacements y_1 and y_2 at P which is at a distance x_1 from S_1 and x_2 from S_2 , due to two waves, be given by

$$y_1 = a \sin 2\pi(t/T - x_1/\lambda)$$

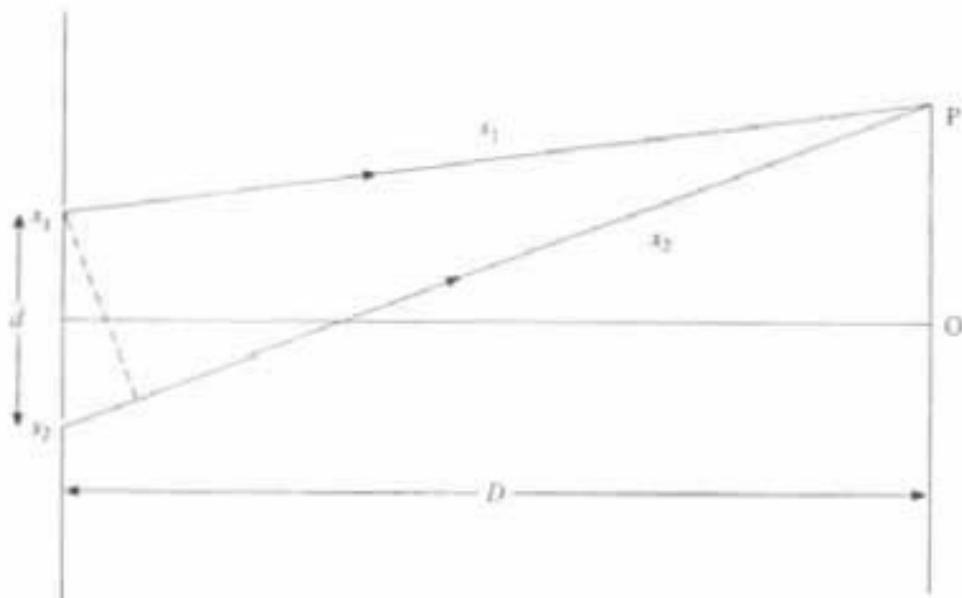


Fig. 1.2 Superposition of two light rays

$$y_2 = a \sin 2\pi(t/T - x_2/\lambda)$$

The "phase difference" between these waves when they arrive at P is

$$\delta = \frac{2\pi}{\lambda} (x_2 - x_1) = \frac{2\pi}{\lambda} \Delta \quad (1.5)$$

where $\Delta = (x_2 - x_1)$ is called "path difference" between those two waves.

From the principle of superposition of waves

$$\begin{aligned} y &= y_1 + y_2 \\ &= a [\sin 2\pi(t/T - x_1/\lambda) + \sin 2\pi(t/T - x_2/\lambda)] \\ &= 2a \sin 2\pi \left[t/T - \frac{(x_1 + x_2)}{2\lambda} \right] \cos 2\pi \left[\frac{x_2 - x_1}{2\lambda} \right] \\ y &= 2a \cos \delta/2 \sin [2\pi(t/T - x_1/\lambda) - \delta/2] \end{aligned} \quad (1.6)$$

Hence displacement at P is also S.H.M. of the same angular frequency but of amplitude $2a \cos \delta/2$ and phase lagging $\delta/2$.

Intensity at a point is proportional to the square of the amplitude. So,

$$\begin{aligned} \text{Intensity} &= I \propto (2a \cos \delta/2)^2 \\ &= I \propto 4a^2 \cos^2 \delta/2 \\ &= I \propto 2a^2 (1 + \cos \delta) \end{aligned} \quad (1.7)$$

Case I: Bright Fringes

When $\cos \delta = 1, I = I_{\max}$

i.e., the condition for constructive interference or bright fringes in terms of phase difference is

$$\delta = 0, 2\pi, \dots, 2n\pi$$

where $n = 0, 1, 2, \dots$

In terms of path difference the condition for bright fringes is

$$\frac{2\pi}{\lambda} (x_2 - x_1) = 2n\pi$$

$\text{Path difference} = \Delta = (x_2 - x_1) = n\lambda$

(1.8)

when $n = 0, 1, 2, \dots$

Case II: Dark Fringes

When $\cos \delta = -1, I = I_{\min} = 0$

So the condition for dark fringes in terms of phase difference is

$$\delta = \pi, 3\pi, \dots, (2n \pm 1)\pi$$

where $n = 0, 1, 2, \dots$

So in terms of path difference condition for destructive interference or dark fringes is

$$2\pi/\lambda(x_2 - x_1) = (2n \pm 1)\pi$$

path difference = $\Delta = (x_2 - x_1) = (2n \pm 1)\lambda/2$	(1.9)
--	-------

when $n = 0, 1, 2, \dots$

INTERFERENCE BY THIN FILM

Newton and Hooke observed and developed an interference phenomenon due to multiple reflections from the surface of thin transparent materials. In daily life, we are familiar with the colours produced by the thin film of oil on the surface of water, by the thin film of a soap bubble and also in thin films of mica. Young was able to explain this phenomenon on the basis of interference between light reflected from the top and the bottom surface of a thin film. It has been observed that the interference in the case of thin film takes place due to: (1) reflected lights and (2) transmitted light from the top and bottom surface of the thin film.

I.4 Thin Film Interference Due to Reflected Light (Parallel Thin Film)

Consider a transparent parallel film (Fig. I.3) of thickness t and refractive index μ . A ray SA incident on the upper surface of the film is partly reflected along AR_1 and partly refracted along AB . At B , part of it is internally reflected along BC and finally emerges out along CR_2 parallel to AR_1 . This will continue further in the same way.

Thus for a single incident ray a number of parallel reflected rays like AR_1 , CR_2 from the front surface and back surface of the thin film can be obtained. Similarly a number of parallel rays can be obtained from the transmitted portion of the incident rays like BT_1 , DT_2 . These parallel reflected rays having effective path difference in between themselves which cause interference, because they are all coherent.

Determination of Optical Path Difference

Draw CN normal to AR_1 and AM normal to BC . The angle of incidence is i and angle of refraction is r . Produce CB backward to meet AE at P . Here $\angle APC = r$ (Fig. I.3). So from $\triangle ABP$, $AB = BP$ and $AE = EP = t$.

The optical path difference between AR_1 and CR_2 is

$$\Delta = \mu[(AB + BC)] - AN$$

Here $\mu = \sin i / \sin r = \frac{AN/AC}{CM/AC}$ (from $\triangle ACN$ and $\triangle CAM$)

$$\mu = AN/CM \quad (\therefore AN = \mu CM)$$

$$\begin{aligned} \Delta &= \mu(AB + BC) - \mu CM = \mu(AB + BC - CM) \\ &= \mu(PB + BC - CM) = \mu(PC - CM) \quad (\because AB = BP) \end{aligned}$$

$\Delta = \mu PM$

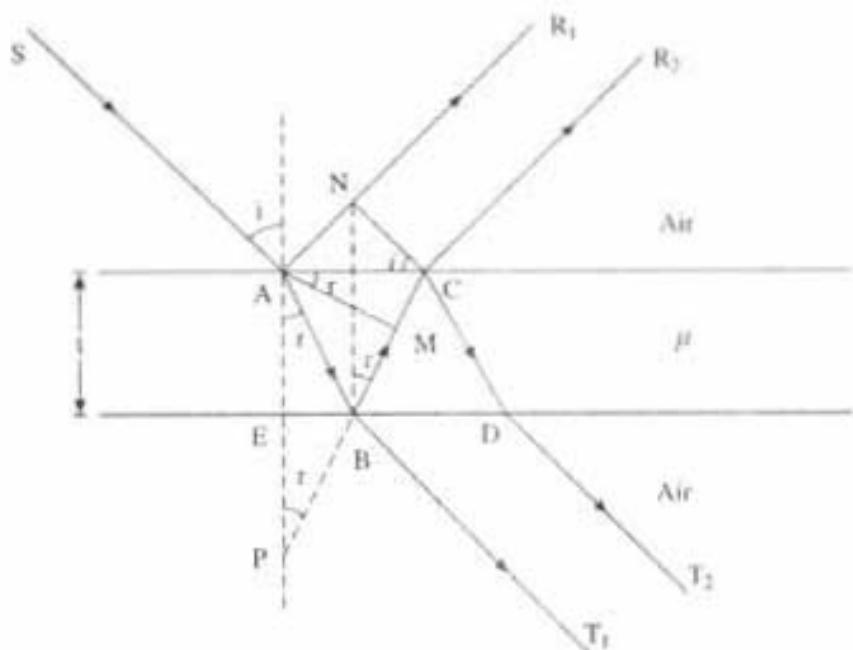


Fig. 1.3 Optical path difference in parallel thin films for reflected beams

In ΔAPM , $\cos r = PM/AP$, $PM = AP \cos r = (AE + EP) \cos r$

$$PM = 2t \cos r \quad (\because AE = EP = t)$$

So

$$\boxed{\Delta = \mu PM = 2\mu t \cos r} \quad (1.10)$$

Above equation in the case of reflected light does not represent the correct path difference between two reflected rays AR_1 and CR_2 but only the apparent. It has been established *on the basis of electromagnetic theory that, when light is reflected at point A, from the surface of an optically denser medium (Air medium interface) a phase change π , equivalent to path difference $\lambda/2$, occurs.*

Therefore, the correct path difference in this case will be

$$\Delta = 2\mu t \cos r \pm \lambda/2 \quad (1.11)$$

Condition for bright band in thin film for reflected light:

If the path difference, from eqn. (1.8) be $\Delta = n\lambda$, where $n = 0, 1, 2, 3, 4, \dots$ etc. Then constructive interference takes place in case of thin film and film appears bright when

$$2\mu t \cos r \pm \lambda/2 = n\lambda$$

Therefore

$$\boxed{2\mu t \cos r = (2n \pm 1)\lambda/2} \quad (1.12)$$

Condition for dark band in thin film for reflected light:

If the path difference, from eqn. (1.9) be $\Delta = (2n \pm 1)\lambda/2$.

where $n = 0, 1, 2, \dots$ etc. Then destructive interference takes place and film appears dark, when

$$2\mu r \cos r \pm \lambda/2 = (2n \pm 1)\lambda/2$$

$$2\mu r \cos r = n\lambda \text{ or } (n+1)\lambda$$

Here n is an integer only. So $(n+1)$ can also be taken as n

$$2\mu r \cos r = n\lambda \quad (1.13)$$

The condition for bright and dark band in thin film interference (eqns. 1.12 and 1.13) are just opposite to the general conditions (eqns. 1.8 and 1.9) for interference. That is, due to the extra phase difference π occurs at reflection at A , from the surface of denser medium. It should be remembered that the interference pattern will not be perfect because intensity of the rays AR_1 and CR_2 will not be the same and their amplitudes are different. The amplitudes will depend on the amount of light reflected and transmitted through the films. There is a small difference in the amplitudes of the rays AR_1 and CR_2 . Therefore the intensity never vanishes completely and perfect dark fringes will not be observed for these rays alone. But in the case of multiple reflection, the intensity of the minimum will be zero.

1.5 Thin Film Interference Due to Transmitted Light

Figure 1.4 shows the geometry of the transmitted light. Due to simultaneous reflection and refraction we obtain two transmitted rays BR and DQ . These rays have originated from same point source S hence they have a constant phase difference and are in a position to produce sustained interference when combined. The effective path difference between the rays BR and DQ is the path difference

$$\Delta = \mu(BC + CD) - BN$$

But

$$\mu = \sin i / \sin r$$

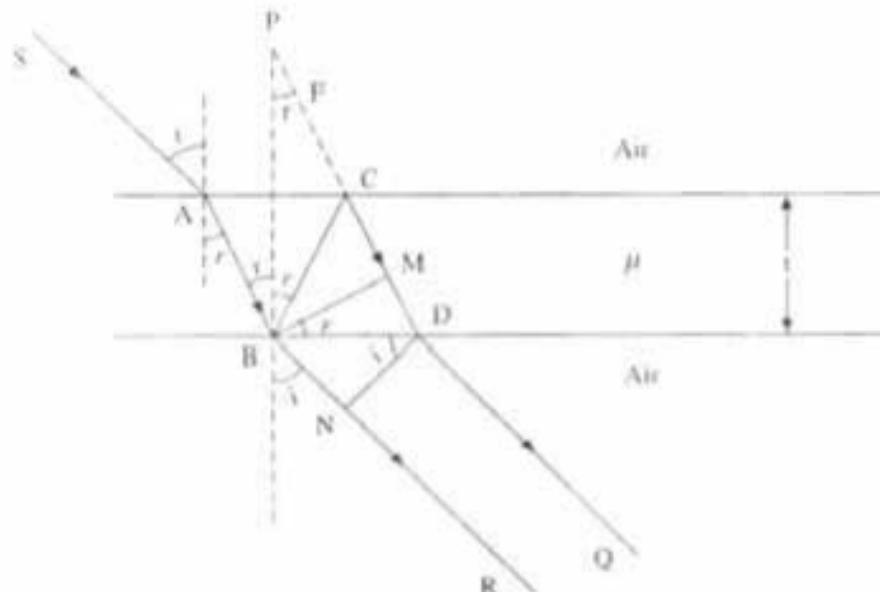


Fig. 1.4 Optical path difference in parallel thin films for transmitted beams

From ΔBMD and ΔBDN (where BM and DN are perpendicular on CD and BR , respectively, we get)

$$\mu = \frac{(BN/BD)}{(MD/BD)} = BN/MD, BN = \mu MD$$

$$\text{Thus } \Delta = \mu(BC + CD - MD) = \mu(CP + CD - MD) \quad (\because BC = CP) \\ \Delta = \mu(PM)$$

But from the ΔBPM

$$\cos r = PM/PB, PM = PB \cos r$$

$$\text{i.e. } PM = 2t \cos r \quad (\because PB = 2t)$$

So, path difference between BR and CQ is

$$\boxed{\Delta = \mu PM = 2\mu t \cos r}$$

Here it should be remembered that *inside the film, reflection at different points (say at B and C), takes from the surface of the rarer medium (air). Thus no extra phase change π takes place in this case.*

So condition for bright band in transmitted light is

$$\boxed{2\mu t \cos r = n\lambda} \quad (1.14)$$

Condition for dark intransmitted light is

$$\boxed{2\mu t \cos r = (2n \pm 1)\lambda/2} \quad (1.15)$$

From Eqns. (1.12, 1.13) and (1.14, 1.15) it is clearly found that conditions of bright and dark for reflected side and transmitted side are just complementary to each other.

1.6 Intensity Distribution

It has been found that for normal incidence, about 4% of the incident light is reflected and 96% is transmitted. The intensity of the transmitted (I_t) beam is given by

$$I_t = \frac{I_0}{1 + \frac{4r^2}{(1-r^2)^2} \sin^2 \delta/2} \quad (1.16)$$

and

$$I_0 = I_r + I_t$$

Here δ is the phase difference, r^2 the reflection coefficient, I_r the intensity of reflected beam and I_0 the incident intensity. When I_t is minimum $\sin^2 \delta/2$ can get maximum value 1.

For reflectance 4% gives, $r^2 = 0.04$.

Putting this value of r^2 in eqn. (1.16) we get

$$I_{t \min} = 0.8521 I_0 = 85.21\%, I_{t \max} = 100\% \rightarrow \text{Transmitted side}$$

$$I_{r \max} = I_0 - I_{t \min} = 14.79\%, I_{r \min} = 0\% \rightarrow \text{Reflected side}$$

1. In the reflected side, the intensity of the interference maxima will be 14.79% of the incident intensity and the intensity of the minima will be zero.
2. In the transmitted side, the intensity of the maxima will be 100% and the intensity of the minima will be 85.21%.

It means the visibility of the fringes is much higher in the reflected system than in the transmitted system because of dark fringe intensity being absolutely zero. Thus, the fringes are more sharp in reflected light, because contrast between dark and bright fringes are much better in reflected light than transmitted light. So, for thin film experiment in the laboratory, reflected side will be better than transmitted side.

1.7 Origin of Colours in Thin Film

From our daily experience we know that when oil is spread over water, then (1) different colours will be seen, i.e., the thin film will appear as coloured and (2) those colours will also change according to the angle of vision. Now the question is why?

(1) When sunlight which consists of seven colours (seven λ 's) is incident on the thin film, then some of them will satisfy constructive condition (eqn. 1.12) and some of them will satisfy destructive condition (eqn. 1.13) in interference after repeated reflection from the thin film. Since all seven colours (seven λ 's) cannot satisfy the constructive condition at a time for a particular thin film of particular thickness t and refractive index (R.I.) μ , so as a result the thin film will appear as coloured, due to missing of some wave lengths, out of seven.

(2) Now at a particular point of the thin film and for a particular position of the eye, the angle of refraction r will (Eqn. 1.12) be also fixed, so for that particular r (the angle of vision), thickness t and R.I. μ , a particular wave length (λ) will satisfy constructive condition (Eqn. 1.12) and only that colour will be seen in that angle of vision. If angle of vision, i.e., r is changed, then some other (λ) will be seen. Hence, colour will change with angle of vision, and due to the different position of eye one can see different colours.

(3) The condition for minima and maxima is reverse in the case of transmitted light. So those colours which are absent or suppressed in reflected light appear as intense colours in transmitted light, so they are complementary to each other.

1.8 Necessity of a Broad Source

Interference fringe obtained in the case of Fresnel's Lloyd's mirror are produced by two coherent sources. The source should be narrow there. But in the case of interference in thin films, the narrow source limits the visibility of the film.

Consider first a thin film and a narrow source S Fig. 1.5(a). Ray 1 produces interference fringes because 3 and 4 reach the eye, whereas ray 2 meets the surface at a different angle and gets reflected along 5 and 6 which are not reaching the eye. Therefore only portion A of the film is visible and not the rest.

But in the case of extended source of light (Fig. 1.5b), the incident rays 1 and 2 meet the surface at A and B, and are reflected along 3, 4 and 5, 6. All of them reach the eye. So in the case of broad source of light, the rays incident at different angles on the film are accommodated by the eye, and the field of view is large.

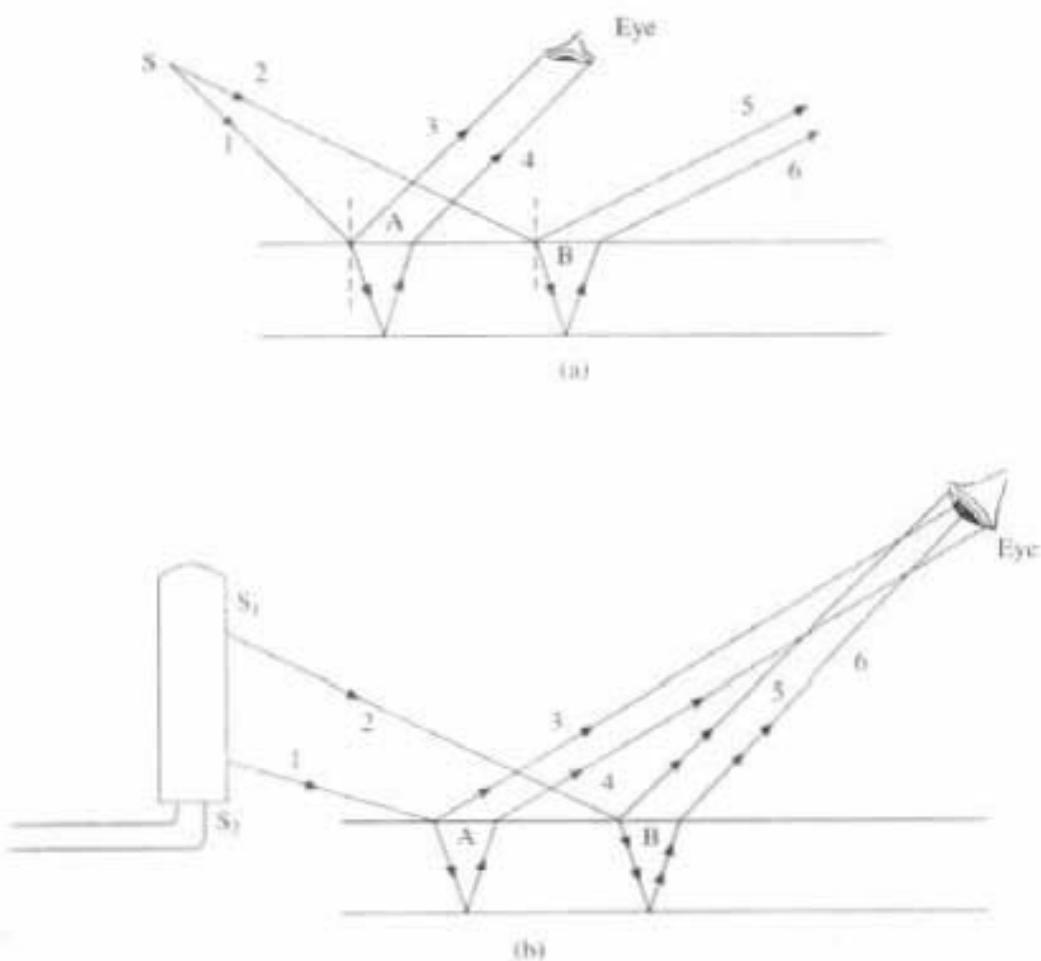


Fig. 1.5 (a) Effect of pointed source of light on thin film; (b) Effect of broad source of light on thin film

1.9 Fringes Produced by Wedge-shaped Film

Let us consider two plane surfaces OH and OH_1 , inclined at an angle α enclosing a wedge-shaped air film. The thickness of the air film t is increased gradually from O to H (Fig. 1.6). Ray AB incident on the wedge film with angle of incidence i . Part of which is reflected along BR and part will refract along BC with angle of refraction r . At the point C it is again reflected with angle of reflection $(r + \alpha)$ along CD and finally emerges as DR_1 . DF and DE are the perpendiculars drawn from D to BR and BC . Angle between two perpendiculars BQ and CQ is α . BC produces backward meet DP at P , so $\angle CPD = \angle CDP = (r + \alpha)$ and from $\triangle CPD$, $CD = CP$ and $DM = MP = t$.

When the air film is viewed with reflected monochromatic light, a system of equidistant interference fringes are observed which are parallel to the edge of the wedge. The interfering rays BR and DR_1 do not enter the eye parallel to each other but they appear to diverge from a point S near the film. Thus the interference that takes place at S is virtual. The effect of interference will be best when the angle of incidence is small.

In order to consider the interference between the waves BR and DR_1 , we first calculate the path difference between them.

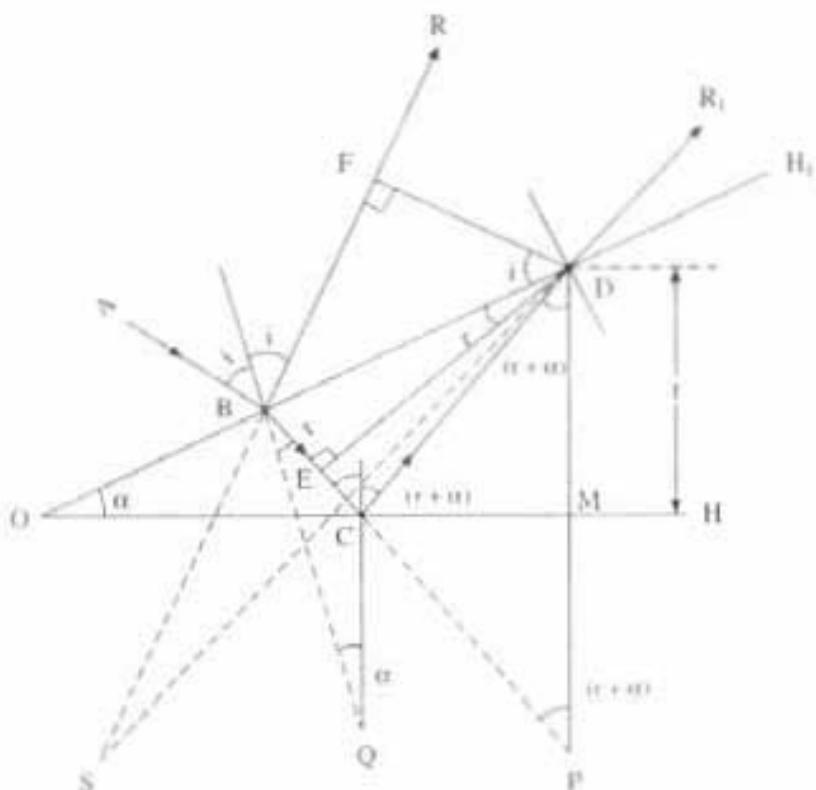


Fig. 1.6 Optical path difference in wedge films for reflected beams

The optical path difference between BR and DR_1 is

$$\Delta = \mu(BC + CD) - BF \quad (1.17)$$

where $\mu = \sin i / \sin r = BF/BE$ i.e. $BF = \mu BE$

since from $\triangle BFD$, $\sin i = BF/BD$ and from $\triangle DEB$, $\sin r = BE/BD$.

$$\text{So } \Delta = \mu(BE + EC + CD - BE) = \mu(EC + CP) = \mu EP \quad (1.19)$$

From $\triangle EPD$ $\cos(r + \alpha) = EP/PD$

$$EP = PD \cos(r + \alpha) = 2t \cos(r + \alpha) \quad (1.20)$$

So optical path difference $\Delta = 2\mu t \cos(r + \alpha)$ (1.21)

From electromagnetic theory of light it is well known that due to reflection (at the point B) from the surface of optically denser medium an extra phase difference π or a path difference $\lambda/2$ occurs.

So for **constructive interference**, effective optical path difference is

$$2\mu t \cos(r + \alpha) \pm \lambda/2 = n\lambda$$

$2t \cos(r + \alpha) = (2n \pm 1)\lambda/2$ (1.22)

where $n = 0, 1, 2, 3, \dots$

For **destructive interference**, effective optical path difference is

$$2\mu t \cos(r + \alpha) \pm \lambda/2 = (2n \pm 1)\lambda/2$$

$$2\mu t \cos(r + \alpha) = n\lambda \quad (1.23)$$

where $n = 0, 1, 2$.

At $r = 0$, $n = 0$ i.e. the zero order fringe will be dark, it will come at the junction of the two plates. When the angle α will be zero, then wedge film becomes parallel film and Eqns (1.22 & 1.23) will become Eqns (1.12 & 1.13)

1.10 Fringe Width

Fringe width of the fringes occurring due to thin film interference can be calculated as follows:

Suppose that n th bright fringe occurs at P_n a distance X_n from 0. Thickness of the air film at P_n is t_n . Applying the condition for bright fringe (Fig. 1.7), we get from equation (1.22)

$$2\mu t \cos(r + \alpha) = (2n + 1)\lambda/2$$

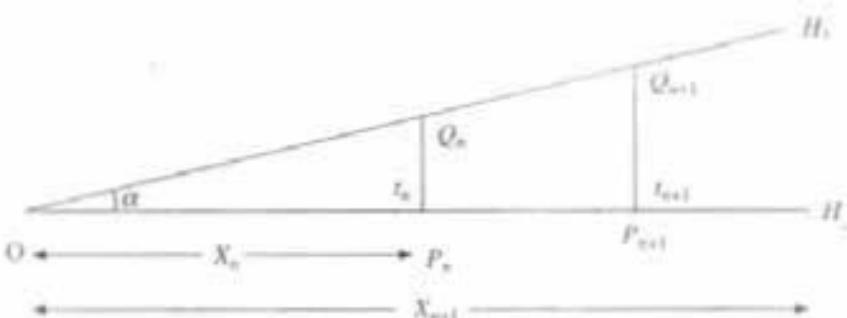


Fig. 1.7 Fringe width in wedge film

Here for normal incidence

$$r = 0$$

$$2t_n \cos(\alpha) \mu = (2n + 1)\lambda/2$$

Now from Fig. 1.7

$$t_n = X_n \tan \alpha \quad (1.24)$$

So above equation becomes

$$2X_n \mu \tan \alpha \cos \alpha = (2n + 1)\lambda/2$$

$$2X_n \mu \sin \alpha = (2n + 1)\lambda/2 \quad (1.25)$$

if $(n + 1)$ bright fringe occurs at P_{n+1} at distance X_{n+1} from 0 then

$$\begin{aligned} 2(X_{n+1}) \mu \sin \alpha &= [2(n + 1) + 1]\lambda/2 \\ &= (2n + 3)\lambda/2 \end{aligned} \quad (1.26)$$

Subtracting these two equations, we get

$$2(X_{n+1} - X_n) \mu \sin \alpha = \lambda$$

So fringe width

$$\beta = (X_{n+1} - X_n) = \frac{\lambda}{2 \sin \alpha \mu} \quad (1.27a)$$

When α will be taken in degree. But if α is very small, then $\sin \alpha = \alpha$.

Then

$$\beta = \lambda / 2\alpha\mu \quad (1.27b)$$

In that case α should be taken in radians and not in degrees.

1.11 Nature of Interference Pattern

From the expression of fringe width we have seen that

$$\beta = \frac{\lambda}{2\alpha\mu}$$

Thus, fringe width β depends only on the wavelength λ , R.I. μ and angle of the wedge α . So when monochromatic light is used, and for air $\mu = 1$, then for a particular wedge angle α all the fringes will come of equal thickness.

For normal incidence $r = 0$ and α is very small, so $\cos(r + \alpha) = 1$. So the condition for bright and dark is

and

$$2\mu t = (2n \pm 1)\lambda/2 \quad (1.28)$$

$$2\mu t = n\lambda$$

Which are dependent only on t for monochromatic rays and for particular value of μ . Now since in the case of wedge-shaped film t remains constant in the direction parallel to the edge of the wedge and increases linearly along the length of the wedge, hence straight parallel fringes of equal width, parallel to the edge of the wedge are obtained. Order (n) of the fringes will increase with increase of thickness (t) as in equation (1.28). Such fringes are called fringes of equal thickness and eqn. (1.27) is called the condition for fringes of equal width.

1.12 Thickness of the Wedge

It is possible to measure the diameter of a thin wire or thickness of very thin foil by the wedged-shaped film.

The wedge is formed by placing the specimen whose thickness or diameter is to be measured, in between two optically flat glass plates, at one end, while the other end should be joined (Fig. 1.8).

According to the thin film interference theory for dark band the condition is

$$2\mu t \cos(r + \alpha) = n\lambda$$



Fig. 1.8 Thickness of the wedge

Since angle of incidence i is very small, so angle of refraction r is also very small and angle of the wedge α is again small. Therefore

$$\cos(r + \alpha) = 1, \mu = 1 \text{ for air}$$

Hence

$$2t = n\lambda$$

Now if l is the length of the wedge, n the number of dark fringes occurred in the length l and β the fringe width, then

$$n = l/\beta$$

Hence

$$2t = l\lambda/\beta$$

$$\text{Thickness of wedge} = \text{Thickness of the specimen} = \boxed{l = \frac{l\lambda}{2\beta}} \quad (1.29)$$

So thickness of any very thin specimen can be determined by using the interference technique in wedge-shaped film.

1.13 Testing the Flatness of Surfaces

If a wedge is formed between two surfaces OA and OB which are perfectly plane, then the air film gradually increased in thickness from O to A . Due to this interference fringes will occur, which are of equal thickness and parallel to the edge of the wedge. Since each fringe is the locus of the points at which the thickness (eqn. 1.28) of the film has a constant value (Fig. 1.9), so this phenomena can be used as an important application of interference. If these fringes are not of equal thickness it means that the surfaces OA and OB are not perfectly plane.

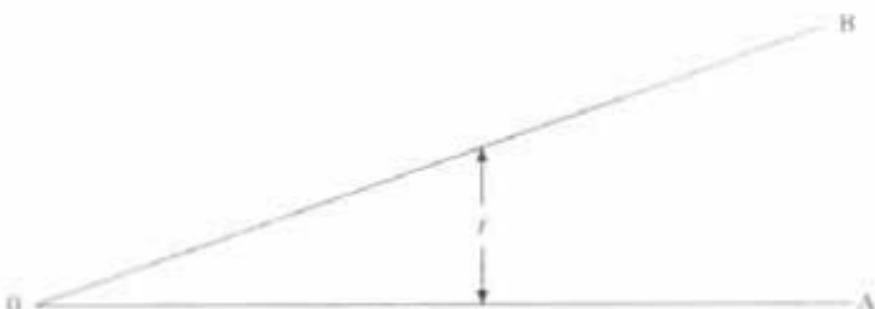


Fig. 1.9 Wedge film between two non-optically flat glass plates

In order to check the optical flatness of any surface say OB , the standard method is to take an optically plane surface OA and the surface OB . The fringes are observed in the field of view.

If they are not of equal thickness (Fig. 1.10), then the surface OB is not optically flat. If they are of equal thickness, then the surface OB is optically flat. If OB is not optically flat then the surface OB should be polished and process should be repeated. When the fringes of equal width are observed that means the surface OB has now become optically flat (Fig. 1.10).

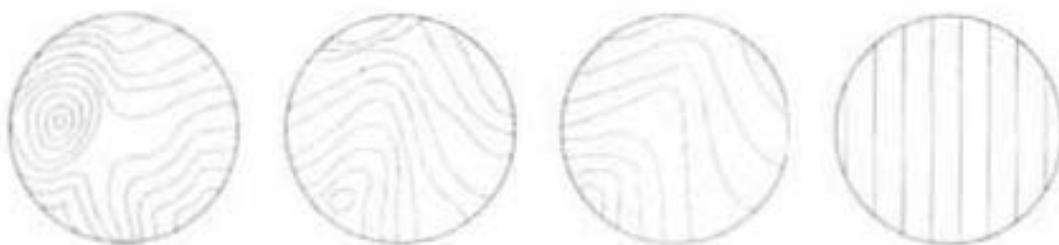


Fig. 1.10 Types of formation of fringes between two non-optically flat glass plates

1.14 Newton's Ring

When a plano-convex lens of large focal length is placed on a plane glass plate, a thin film of air is enclosed between the lower surface of the lens and the upper surface of the plate. Thickness of the air film is very small at the point of contact and gradually increases from the centre to outwards. In this type of thin film interference fringes produced with monochromatic light are circular. Since the convex side of the plano-convex lens has a spherical surface, so the thickness of the air film will be constant over a circle of particular radius, with the point of contact O as centre. Thus concentric dark and bright rings will be produced with the point of contact O as centre. As thickness of the air film (t) increases, orders (n) of the circular rings will also increase, according to equation (1.28). When $t = 0$, n will be zero (eqn. 1.28), so central spot will be dark. These rings are called Newton's rings (Figs. 1.14 and 1.15).

When viewed with white light, concentric circles with different colours with dark as centre are found. When monochromatic light is used, bright and dark concentric rings are produced in the air film, with dark as centre.

S is a source of monochromatic light at the focus of the lens (Fig. 1.11). A parallel beam of light falls on the glass plate G at 45° . The glass plate G reflects a part of the incident light towards the air film enclosed by the lens L and the glass plate P . The reflected beam from the air film is viewed with a microscope (Fig. 1.11).

Interference takes place and alternate dark-bright circular rings are produced. This is due to interference between the light reflected from the lower surface of the lens and the upper surface of the glass plate P (Fig. 1.12).

Theory: Newton's Rings by Reflected Light

Let us suppose the radius of curvature of the lens is R and the air film of thickness t_n is at a distance $OQ = r_n$, from the point of contact O (Fig. 1.13) where n th fringe will be formed. In the case of reflected light the condition for bright fringe is

$$2\mu t_n \cos r = (2n - 1)\lambda/2$$

where $n = 1, 2, 3, \dots$ [Here $n \neq 0$, since the central band ($n = 0$) should be dark]. For normal incidence $r = 0 \cos r = 1$

So

$$2\mu t_n = (2n - 1)\lambda/2 \quad (1.30)$$

where $n = 1, 2, 3, \dots$, except zero.

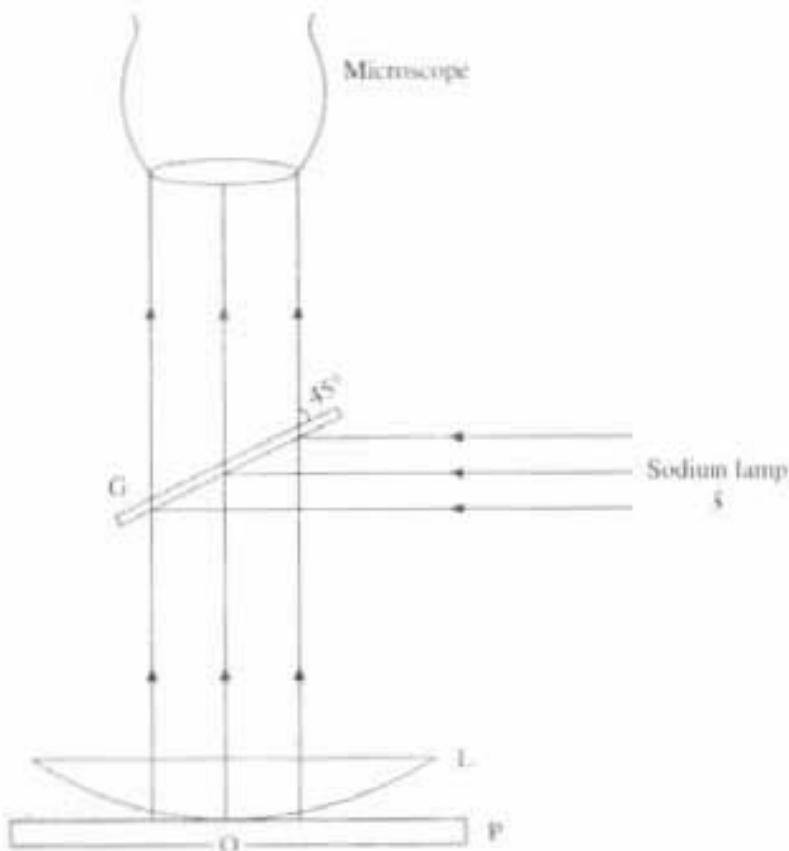


Fig. 1.11 Newton's rings experimental set up

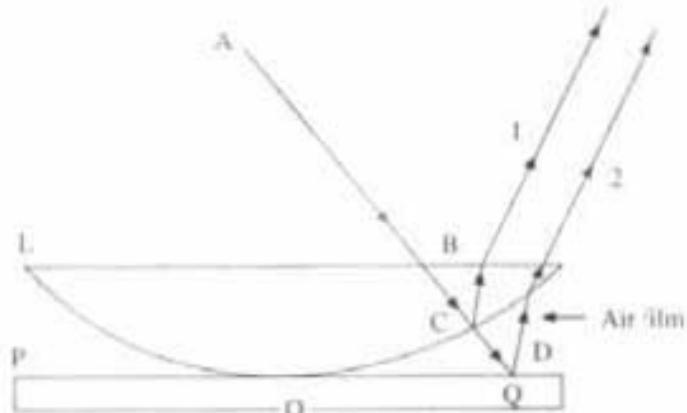


Fig. 1.12 Formation of Newton's ring

Condition for dark fringe for normal incidence is

$$2\mu t_n = n\lambda \quad (1.31)$$

where $n = 0, 1, 2, \dots$

$n = 0$ when $t = 0$ at point O.

At the point of contact O, $n = 0$, i.e., the central band of Newton's ring for reflected light is dark (Fig. 1.14). Since the condition for dark and bright is reverse in the case of transmitted light. So for transmitted light the central band will be white (Fig. 1.15).

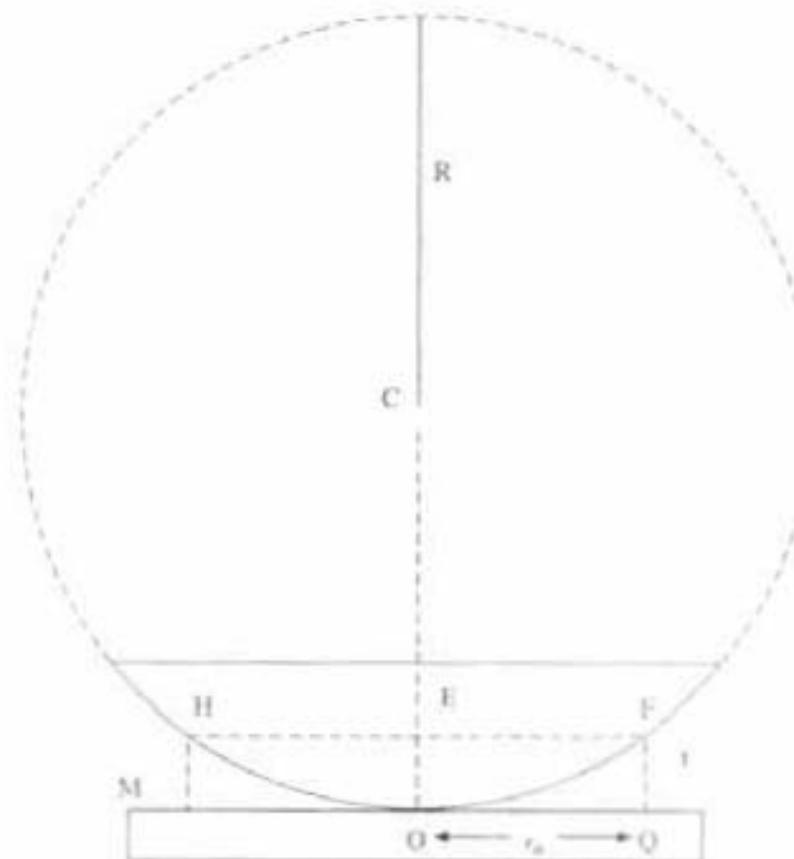


Fig. 1.13 Formation of Newton's ring

Newton's ring in white light

When white light is used instead of monochromatic sodium light then equation (1.30) will be satisfied at a different value of t (thickness of air film), i.e., for different values of the order (n) with different values of λ . So the rings will appear coloured. But at $t = 0$ all $\lambda = 0$, thus the central band ($n = 0$) will be dark, for reflected light, and as thickness (t) of the air film increases, different order (n) of the rings will appear with different colour (λ). At the same time overlapping of different orders (n) with different λ also will take place. So colour of a particular ring cannot be distinguished very accurately.

Now from the theory of circle (Fig. 1.13)

$$EF \times HE = OE \times (2R - OE)$$



Fig. 1.14 Newton's ring observed in reflected light



Fig. 1.15 Newton's ring observed in transmitted light

But $EF = HE = r_n$ = radius of n th ring and

$$OE = FQ = t_n$$

$$\text{So } 2R - OE = 2R - t_n = 2R \quad (\text{since } t_n \ll R)$$

$$\text{So } r_n^2 = 2Rt_n$$

$$t_n = \frac{r_n^2}{2R}$$

Substituting this value of t_n in equation (1.30), we get, for the n th bright ring

$$\begin{aligned} \frac{2\mu r_n^2}{2R} &= (2n-1)\lambda/2 \\ r_n^2 &= \frac{(2n-1)R\lambda}{2\mu} \end{aligned} \quad (1.32)$$

Replacing $r_n = D_n/2$ we get, diameter of the n th bright ring (D_n)

$$D_n^2 = \frac{4(2n-1)\lambda R}{2\mu}$$

i.e.

$$D_n \propto \sqrt{2n-1} \quad (1.33)$$

Thus the diameter of the n th bright rings are proportional to the square root of odd natural number, i.e. $(2n-1)$.

Similarly for the n th dark ring from equation (1.31)

$$r_n^2 = 2Rt_n = n\lambda R/\mu$$

$$D_n^2 = 4n\lambda R/\mu$$

i.e.,

$$D_n \propto \sqrt{n} \quad (1.34)$$

Thus the diameter of n th dark rings are proportional to the square root of natural number n .

When $\mu = 1$, for the n th dark ring

$$r_n^2 = n\lambda R \quad (\text{where } n = 0, 1, 2, \dots) \quad (1.35)$$

which implies that the radii of the rings varies as square root of natural numbers. Thus, the rings will become closer to each other as radius of the rings increases, i.e., fringe width goes on decreasing progressively with ring number, so the rings become thinner and thinner as ring numbers will increase, which can be proved in the following way:

$$\left. \begin{array}{l} r_1 \propto \sqrt{1} = 1 \\ r_2 \propto \sqrt{2} = 1.414 \end{array} \right\} \begin{array}{l} \text{Fringe width i.e., } (r_2 - r_1) \text{ is proportional to} \\ 0.414 \text{ between 1st and 2nd dark ring} \end{array}$$

$$\left. \begin{array}{l} r_{19} \approx \sqrt{19} = 4.359 \\ r_{20} \approx \sqrt{20} = 4.472 \end{array} \right\} \dots, \text{whereas fringe width i.e., } (r_{20} - r_{19}) \text{ between 19th and 20th dark ring is proportional to } 0.113$$

Thus the fringe width (β) of the Newton ring is inversely proportional to the square root of the natural n , i.e., $\beta \propto 1/\sqrt{n}$, as ring number increases thickness of the ring decreases, i.e., fringe width decreases.

So

$$\beta \propto \frac{1}{\sqrt{n}} \quad (1.35a)$$

1.15 Determination of the Radius of Curvature of the Plano-Convex Lens, Wave Length of Light Used, and Refractive Index of Some Unknown Liquid

Diameter of n th and $(n+m)$ th dark band are

$$D_n^2 = 4n\lambda R/\mu$$

$$D_{m+n}^2 = 4(n+m)\lambda R/\mu$$

Thus the radius of curvature of plano-convex lens is

$$R = \mu(D_{m+n}^2 - D_n^2)/4m\lambda \quad (1.36)$$

By using the formula (1.36) the radius of curvature of plano-convex lens or wave length of monochromatic light or the refractive index of some liquid can be found out when other factors are given.

Generally the refractive index of some unknown liquid can also be found in the following way:

From equation (1.34) we have seen that the diameter of the n th dark ring is inversely proportional to refractive index μ .

Hence, in order to find the refractive index of some liquid first measure the diameter of n th dark ring with air, say D_{1n} . Then

$$D_{1n}^2 = 4n\lambda R$$

With the same setting, now put the liquid in between glass plate and plano-convex lens, then diameter of n th dark ring will be reduced to D_{2n} such that

$$D_{2n}^2 = 4n\lambda R/\mu$$

Taking the ratio of 1st and 2nd of diameter of the n th dark ring with air and liquid, one gets

$$D_{1n}^2 / D_{2n}^2 = \mu \quad (1.37)$$

the refractive index of the unknown liquid.

1.16 Non-Reflecting Films

One of the practical applications of the thin film interference phenomenon is the making of non-reflecting films, which are used in cameras, binoculars and telescopes. The reflection from the camera lens or a prism can be decreased to a minimum by coating a thin film of proper thickness. So that images formed by the lens or prism will be bright and comparatively less exposure time would be needed to photograph the object.

In order to make the non-reflecting film, we choose a coating material whose refractive index will be in between that of air and glass. Here we need total transmission of light from object through the camera lens to the photographic plate. Hence, the rays should satisfy the destructive condition for reflected side. Here (Fig. 1.16), from the points A and B, in both cases, reflection is occurring from the surface of denser medium. So, extra path difference $\lambda/2 + \lambda/2$ or extra phase difference $\pi + \pi = 2\pi = 0$ will occur as per electromagnetic theory of light. Thus, in this case the condition of destructive interference is

$$2\mu t = (2n + 1)\lambda/2 \quad (1.38)$$

i.e., optical thickness $\mu t = (2n + 1)\lambda/4$, where n is the natural number 0, 1, 2, 3,

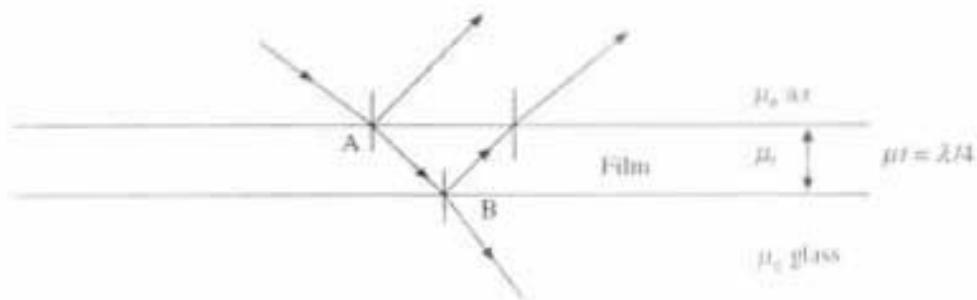


Fig. 1.16 Non-reflecting film ($\mu_a < \mu_f < \mu_g$)

Now, in order to make non-reflecting film over glass, the minimum optical thickness of the film (μt) should be $\lambda/4$. In this case, the thickness $\lambda/4$ of the film will be true for one particular wave length. For this purpose, the wave length is selected from yellow-green portion of the spectrum where the eye is most sensitive. Therefore, if the incident light is white, some light is reflected at shorter and longer wave lengths and the reflected light has a purple hue. This is a very useful method to decrease the reflection to a minimum value. Such thin coating of a liquid of thickness are known as non-reflecting films. For practical purpose, durable layers are made by evaporating calcium or magnesium fluoride on the surface in vacuum, or by chemical treatment with acids which leave a thin layer of silicon on the surface of the lens or prism.

1.17 Totally Reflecting Film (Sun Control Film)

We can use the same principle for sun control in the window glass of cars, buses, or in sun glasses. Here we need total reflection of the sun rays from the glass, i.e., the majority of the rays will satisfy the condition for brightness on the reflected side.

Here, we choose a film whose R.I. is in between air and glass, so that reflection will occur at point A and B (Fig. 1.17) from the surface of denser media. The total phase change will be $(\pi + \pi) = 2\pi = 0$. Thus, by the condition of complete constructive interference the path difference will be

$$2\mu t = n\lambda \quad (1.39)$$

Therefore, if the minimum optical thickness of the film be $\mu t = \lambda/2$, then it will satisfy the complete constructive interference condition and majority of the sun light will be reflected back from the surface of the glass. Here also we have to consider the λ in yellow-green range where human eye is most sensitive.

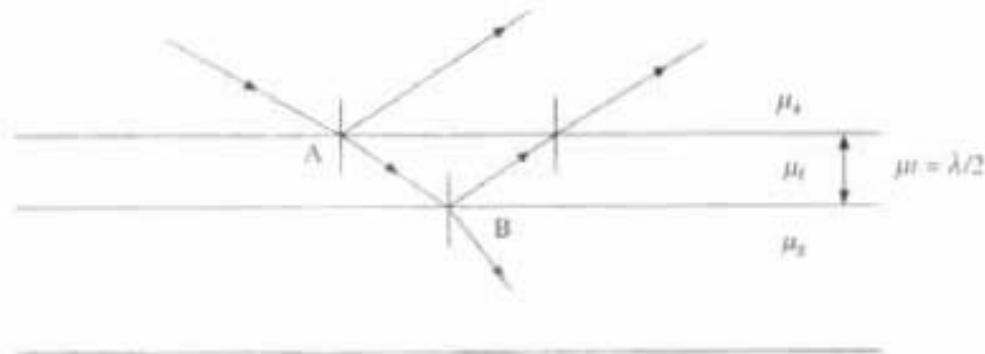


Fig. 1.17 Totally reflecting film ($\mu_1 < \mu_t < \mu_2$)

PROBLEMS

Note: For bright condition

Use the equations

- (1) $2\mu t \cos(r + \alpha) = (2n + 1)\lambda/2$ when $n = 0, 1, 2, \dots$ for minimum thickness $n = 0$
- (2) $2\mu t \cos(r + \alpha) = (2n - 1)\lambda/2$ when $n = 1, 2, 3, \dots$ except zero for minimum thickness $n = 1$
- (3) For minimum thickness n will be one

For dark condition when the equation is $2\mu t \cos(r + \alpha) = n\lambda$

1. A glass wedge ($\mu = 1.5$) of angle 0.01 radian is illuminated by monochromatic light of 6000 \AA falling normally on it. At what distance from the edge of the wedge, will the 10th fringe be observed by reflected light?

Solution: $\alpha = 0.01$ radian, $n = 10$, $\lambda = 6000 \times 10^{-8} \text{ cm}$

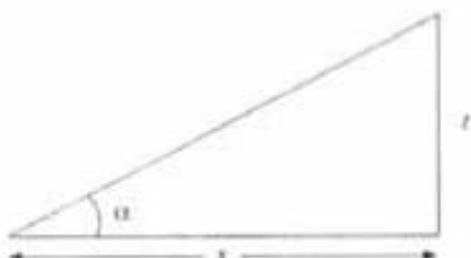
For dark fringe

$$2\mu t \cos \alpha = n\lambda$$

But

$$\tan \alpha = \alpha = t/x$$

$$t = x \tan \alpha$$



(a)

$$2\mu x \sin \alpha = 2\mu x \alpha = n\lambda \quad (\because \alpha \text{ is small, so } \sin \alpha = \alpha)$$

$$\text{i.e., } x = \frac{(10 \times 6000 \times 10^{-8})}{2\mu\alpha} = (10 \times 6000 \times 10^{-8}) / (2 \times 0.01 \times 1.5)$$

or

$$x = 0.02 \text{ cm}$$

2. White light falls normally on a soap film of thickness 5×10^{-5} cm. What wavelength within the visible spectrum will be strongly reflected, given that μ for soap film = 1.33, and range of visible spectrum $\lambda = 4000$ to 7000 \AA ?

Solution: For constructive interference of the reflected light, the condition is:

$$2\mu t \cos r = (2n + 1)\lambda/2, \text{ where } n = 0, 1, 2, \dots$$

For normal incidence, $r = 0$, $\cos r = 1$.

$$\text{Thus, } 2\mu t = (2n + 1)\lambda/2$$

Now, $t = 5 \times 10^{-5}$ cm, and $\mu = 1.33$.

(a) For zeroth order $n = 0$,

$$\lambda = \lambda_1 = 4\mu t = 4 \times 1.33 \times 5 \times 10^{-5} = 26600 \times 10^{-8} = 26600 \text{ \AA}$$

(b) For first order $n = 1$,

$$\lambda = \lambda_2 = 1/3 (4\mu t) = 1/3 \times 1.33 \times 4 \times 5 \times 10^{-5} = 8866 \times 10^{-8} \text{ cm} = 8866 \text{ \AA}$$

(c) For second order $n = 2$,

$$\lambda = \lambda_3 = 1/5 (4\mu t) = 5320 \times 10^{-8} \text{ cm} = 5320 \text{ \AA}$$

(d) For third order $n = 3$,

$$\lambda = \lambda_4 = 1/7 (4\mu t) = 3800 \times 10^{-8} \text{ cm} = 3800 \text{ \AA}$$

Out of these wave lengths only 5320 \AA lies within the visible range, which is strongly reflected.

3. Newton's rings are obtained with reflected light of wave length 5000 \AA , if the diameter of 10th dark ring is 0.5 cm, calculate the radius of curvatures of the lens and hence find out the radius of 50th dark ring.

Solution: Diameter of n th dark ring is

$$D^2 = 4nR\lambda$$

$$D_{10}^2 = 4 \times 10 \times R \times 5000 \times 10^{-8} = (0.5)^2$$

$$R = \frac{(0.5)^2}{4 \times 10 \times 5000} \times 10^8 = 1.25 \times 10^6 \times 10^8$$

$$R = 125 \text{ cm}$$

$$D_{50}^2 = 4 \times 50 \times 125 \times 5000 \times 10^{-8} = 1.26 \times 10^9 \times 10^{-8}$$

$$D_{50}^2 = 1.26$$

$$D_{50} = 1.12249$$

$$r_{50} = 0.56 \text{ cm}$$

4. White light falls at an angle of 45° on a parallel soap film of refractive index 1.33. At what minimum thickness of the film will it appear bright yellow ($\lambda = 5000 \text{ \AA}$) in the reflected light?

Solution: Angle of incidence = $i = 45^\circ$

$$\mu = \frac{\sin i}{\sin r}$$

$$\sin r = \sin i / \mu = 0.7071 / 1.33$$

$$\sin r = 0.53$$

$$r = 32.11^\circ$$

For bright fringe

$$2\mu t \cos r = (2n + 1)\lambda/2$$

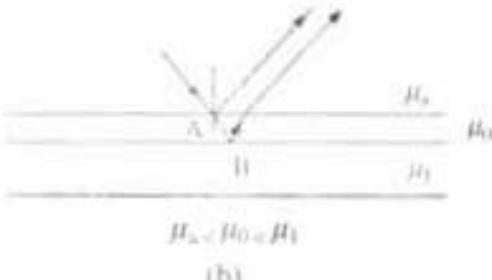
For minimum thickness, $n = 0$

$$2\mu t \cos r = \lambda/2$$

$$t = \frac{5000}{2} \times 10^{-8} \times \frac{1}{2 \times 1.33 \times 0.847} = 1109 \times 10^{-8}$$

$$t = 1109 \text{ \AA}$$

5. An oil drop of vol. 0.2 cc is dropped on the surface of a liquid in a tank of surface area of 1 sq.m. The oil spreads uniformly over the surface of the liquid in the tank and forms a thin film. White light falls normally on the film. It is found that 5500 \AA is the most intensified wave length in the reflected light. What is the reasonable value of the refractive index of oil which is less than that of liquid?



Solution: Since R.I. of oil $\mu_0 < \mu_1$ (R.I. of liquid) and R.I. of air $\mu_a < \mu_0$, so both the reflection that occur at A and B are from the surface of denser medium. So using condition for the brightness for minimum thickness (i.e., $n = 1$) is

$$2\mu t \cos r = \lambda$$

For normal incidence $r = 0$

$$2\mu t = \lambda$$

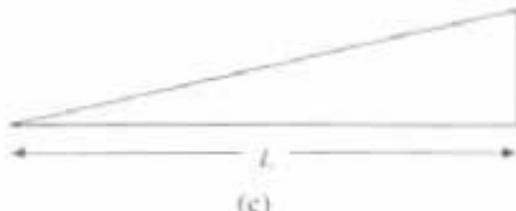
$$t = \text{Thickness of film} = 0.2 / 100 \times 100 \text{ cm} = \text{Volume/Area}$$

$$\text{Hence, R.I. of oil} \quad \mu_0 = \lambda / 2t$$

$$= \frac{5500 \times 10^{-8} \times 10^4}{2 \times 0.2}$$

$$\mu_0 = 13750 \times 10^{-4} = 1.3750$$

6. Light of wave length 6300 \AA is incident normally on a thin wedge-shaped film of index of refraction 1.5. There are ten bright and nine dark fringes over length of the film. How much does the film thickness change over this length?



(c)

Solution: Within the length or the wedge (L) there are 10 bright and 9 dark fringes.

$$\text{Condition for dark is } 2\mu t = n\lambda$$

$$\text{Condition for bright is } 2\mu t = (2n \pm 1)\lambda/2$$

Thus, when $t = 0$ zeroth order dark band will come and at the edge of the wedge 10th bright band will come.

Thus, thickness of the air film is

$$2\mu t = (2n - 1)\lambda/2$$

where n is the number of bright fringe = 10

$$2\mu t = 19\lambda/2$$

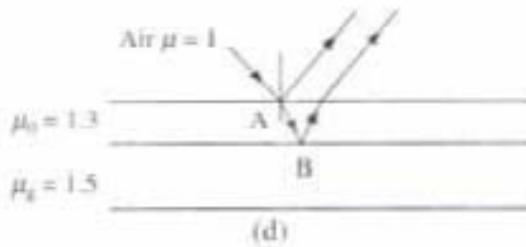
$$t = \frac{19 \times 6300 \times 10^{-8}}{2 \times 1.5 \times 2} = 19950 \times 10^{-8} \text{ cm} = 19950 \text{ \AA}$$

$$t = 0.00001995 \text{ cm} = 19950 \text{ \AA}$$

7. A plane wave of monochromatic light falls normally on a uniformly thin film of oil which covers a glass plate. The wavelength of the source can be varied continuously. Complete destructive interference is obtained for wave lengths 5000 and 7000 \AA and for no other wave lengths in between. Find the thickness of the layer.

R.I. of glass = 1.5

R.I. of oil = 1.3



(d)

Solution: Here reflections occur at A and B, in both cases from the surface of the denser medium.

Thus the condition for destructive interference in reflected side is

$$2\mu r \cos r = (2n \pm 1)\lambda/2 \text{ for } n\text{th minimum}$$

$$r = 0, \text{ hence } 2\mu t = (2n + 1)\lambda/2 \text{ where } n = 0, 1, 2, \dots$$

and $2\mu = [2(n+1)+1]\lambda_2/2$ for $(n+1)$ th minimum

$$\text{Therefore } \lambda_1 = \frac{2 \times 2\mu}{(2n+1)} = 7000 \text{ Å}, 1/\lambda_1 = \frac{(2n+1)}{4\mu} \quad (1)$$

$$\lambda_2 = \frac{2 \times 2\mu}{[2(n+1)+1]} = 5000 \text{ Å}, \frac{1}{\lambda_2} = \frac{(2n+3)}{4\mu} \quad (2)$$

Subtracting (1) from (2), we get

$$\begin{aligned}\frac{1}{4\mu} &= (1/\lambda_2 - 1/\lambda_1) = \frac{\lambda_1 - \lambda_2}{\lambda_1 \lambda_2} \\ r &= \frac{2\lambda_1 \lambda_2}{4\mu(\lambda_1 - \lambda_2)} = \frac{5000 \times 7000}{2 \times 1.3 \times 2000} \\ r &= 6.730 \times 10^3 \text{ Å}\end{aligned}$$

$$r = 6730 \text{ Å}$$

8. In Newton's ring's experiment, the diameter of the 16th ring changes from 1.40 to 1.27 cm, when a liquid is introduced between the lens and the glass plate. Calculate the R.I. of the liquid.

Solution: Diameter or n th dark ring is

$$D_n^2 = \frac{4n \times R \times \lambda}{\mu}$$

when $\mu = 1$

$$D_n^2 = 4n \times \lambda R = (1.40)^2 \quad (1)$$

when $\mu = \mu_l$

$$D^2 = \frac{4n \lambda R}{\mu_l} = (1.27)^2 \quad (2)$$

Dividing (1) and (2), we get

$$\frac{(1.40)^2}{(1.27)^2} = \mu_l$$

$$\mu_l = 1.2152$$

9. Interference fringes are produced with monochromatic light falling normally on a wedge-shaped film of cellophane whose R.I. is 1.40. The angle of wedge is 10 sec of an arc and the distance between the successive fringes is 0.5 cm. Calculate wave length of the light used.

Solution: $\alpha = 10$ sec of an arc

Total angle of an arc $= 180^\circ = \pi$ radian

$$\alpha = 10 \text{ sec of an arc} = \frac{\pi \times 10}{60 \times 60 \times 180} = \frac{3.14 \times 10}{3600 \times 180} \text{ radian}$$

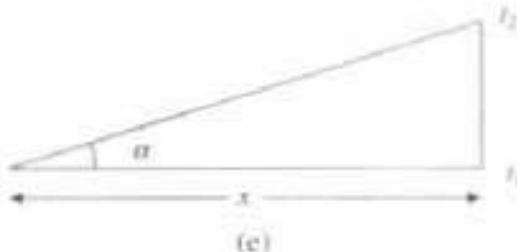
Fringe width $\beta = 0.5 \text{ cm}$, $\mu = 1.40$

$$\text{Hence } \beta = \frac{\lambda}{2\alpha\mu}$$

$$\lambda = 2\beta\alpha\mu = \frac{2 \times 0.5 \times 3.14 \times 10 \times 1.4}{60 \times 60 \times 180}$$

$$\lambda = 6790 \text{ Å}$$

10. A square piece of cellophane film with R.I. 1.5 has a wedge-shaped section such that its thickness at two opposite sides are t_1 and t_2 . If with a light of $\lambda = 6000 \text{ \AA}$, the number of fringes appearing in the film is 10, calculate the difference $t_2 - t_1$.



Solution: Condition for dark fringe is $2t = n\lambda$.

The number of fringes is 10 within the distance x when thickness is changed from t_1 to t_2 .

$$\text{Fringe width} \quad \beta = \frac{\lambda}{2\alpha\mu}$$

$$\beta = x/10 \text{ and } \alpha = \frac{t_2 - t_1}{x}$$

$$\text{Now} \quad \alpha = \frac{\lambda}{2\beta\mu}$$

$$t_2 - t_1 = \frac{\lambda x}{2\beta\mu} = \frac{\lambda \cdot x \cdot 10}{2 \cdot x \cdot \mu} = \frac{10\lambda}{2\mu}$$

$$t_2 - t_1 = \frac{10 \times 6000 \times 10^{-8}}{2 \times 1.5} = 2 \times 10^{-3} \text{ cm}$$

11. A parallel beam of light ($\lambda = 5890 \times 10^{-8} \text{ cm}$) is incident on a thin glass plate ($\mu = 1.5$), such that the angle of refraction into the plate is 60° . Calculate the smallest thickness of the glass plate which will appear dark by reflection.

Solution: Condition for dark $\rightarrow 2\mu t \cos r = n\lambda$

$$t = \frac{n\lambda}{2\mu \cos r}$$

For $n = 1$:

$$t = \frac{5890 \times 10^{-8}}{2 \times 1.5 \times \cos 60^\circ}$$

$$t = 3926.6 \times 10^{-8} = 3926 \text{ \AA}$$

12. Light containing two wavelengths λ_1 and λ_2 falls normally on a convex lens of radius of curvature R , resting on a glass plate. If the n th dark ring due to λ_1 coincides with $(n+1)$ th dark ring due to λ_2 , prove that the radius of the n th dark ring due to λ_1 is $\sqrt{\lambda_1 \lambda_2 R / (\lambda_1 - \lambda_2)}$

Solution: For dark ring $D_n^2 = 4n\lambda R/\mu$, for air $\mu = 1$

For wave length λ_1 , $D_n^2 = 4n\lambda_1 R$

For wave length λ_2 , $D_{n+1}^2 = 4(n+1)\lambda_2 R$

$$\text{Hence } 4n\lambda_1 R = 4(n+1)\lambda_2 R \quad \text{or} \quad n = \lambda_2(\lambda_1 - \lambda_2)$$

Radius of the n th dark ring due to λ_1 is $r_n = \sqrt{n\lambda_1 R} = \sqrt{\frac{\lambda_1 \lambda_2 R}{\lambda_1 - \lambda_2}}$

$$r_n = \sqrt{n\lambda_1 R} = \sqrt{\frac{\lambda_1 \lambda_2 R}{\lambda_1 - \lambda_2}}$$

QUESTIONS

- Calculate the path difference for reflected rays in thin film and hence obtain the condition for constructive interference and destructive interference with monochromatic light.
- Explain how the fringes of constant thickness are formed and hence describe the method to use them for testing the flatness of a glass plate.
- Explain how one can minimize the reflection of light incident on the surface of a glass plate by using interference technique. Derive the necessary formulae.
- Show that the fringe width in case of Newton's rings is inversely proportional to the (i) radius of the dark ring and (ii) square root of natural number n .
- Explain the theory of wedge-shaped films. Describe a method for determination of diameter of a thin wire using wedge-shaped film.
- A soap film of refractive index $4/3$ and of thickness 1.5×10^{-3} cm is illuminated by white light incident at an angle of 45° . The light reflected by it is examined by a spectroscope which finds a dark band corresponding to a wave length of 5×10^{-5} cm. Calculate the order of interference band.
- A wedge-shaped air film is formed between two glass plates by placing a paper at one of the sides. When illuminating this film with light of wave length 6000 \AA , 10 fringes are seen in a centimeter. If light is incident normally, find the angle of the wedge.
- (i) Describe experimental set up for Newton's ring. How is it used to determine radius of curvature of convex lens? Derive necessary formula.
(ii) In a Newton's ring experiment, the diameter of 5th ring was 0.336 cm and the diameter of 15th ring was 0.590 cm. Find the radius of curvature of the plane convex lens, if the wave length of light used is 5880 \AA .
- Newton's ring arrangement is used with a source emitting two wave lengths $\lambda_1 = 6.0 \times 10^{-5}$ cm and $\lambda_2 = 4.5 \times 10^{-5}$ cm and it is found that the n th dark ring due to λ_1 coincides with $(n+1)$ th dark ring due to λ_2 . If the radius of curvature of the curved surface of the lens is 90 cm, find the diameter of the n th dark ring for λ_1 .
- Two glass plates enclose a wedge-shaped air film, touching at one edge and are separated by a wire of 0.05 mm diameter and distance of 15 cm from the edge. Calculate the fringe width. Monochromatic light of $\lambda = 6000 \text{ \AA}$ from a broad source falls normally on the film.
- Explain the formation of colour on thin film in sunlight with proper mathematical expression, also explain properly why do colours change with angle of vision and why colours which are present in reflected light are absent in transmitted light?
(Hints: Explain from equations (1.12) and (1.13) for seven different λ and different value of n ; see section 1.7.).

12. A soap film ($\mu = 1.33$) on a wire loop held in air appears black at its thinnest portion when viewed by reflected light. On the other hand a thin oil film ($\mu = 1.20$) floating on water ($\mu = 1.33$) appears bright in its thinnest portion, when similarly viewed from the air above. Explain these phenomena.

[*Hints:* For soap film on a wire loop the condition of dark is $2\mu t = \lambda$ (equation 1.13) for normal incidence so when t is very small it will appear dark. But for oil ($\mu = 1.20$) on water ($\mu = 1.33$) the condition for bright is $2\mu t = \lambda$ (equation 1.39) so for thinnest portion it will appear bright. Because in first case medium are rarer-denser-rarer, whereas in second case mediums are rarer-denser-more denser, so conditions are opposite.]

13. Wedge film produces fringes of equal thickness whereas in Newton's ring, the thickness of the rings gradually decreases with the increasing ring number. Explain with proper expression.

[*Hints:* Explain from equations (1.35a)]

14. Explain why the system of Newton's rings observed by transmitted light is complimentary to that observed by reflected light. What will be the nature of the Newton's ring, if the experiment is performed in the sunlight.

[*Hints:* In sunlight central spot of Newton's ring will appear dark ($t = 0, n = 0$) but other rings will be coloured where different orders will also overlap. (For different t , different λ , consider equations 1.30, 1.31 and Section 1.14.)

1B. DIFFRACTION OF LIGHT

1.18 Optical Diffraction

It is known from common experience that if an obstacle is placed in the path of light, a sharp shadow is produced, which proves the rectilinear propagation of light on the basis of Newton's corpuscular theory. But it has also been observed that when a beam of light passes through a very small opening Fig. 1.18(a), it spreads over to some extent into the region of the geometrical shadow area. This can be explained by considering the wave nature of light like sound wave, when one would expect the bending of the light beam around the edge of an opaque obstacle and the illumination of the geometrical shadow area to some extent.

According to the Huygen's wave theory, each progressive wave produces secondary wavelets and these secondary wavelets say M_1N_1 originating from various points of the original spherical wave front MN (Fig. 1.18), mutually

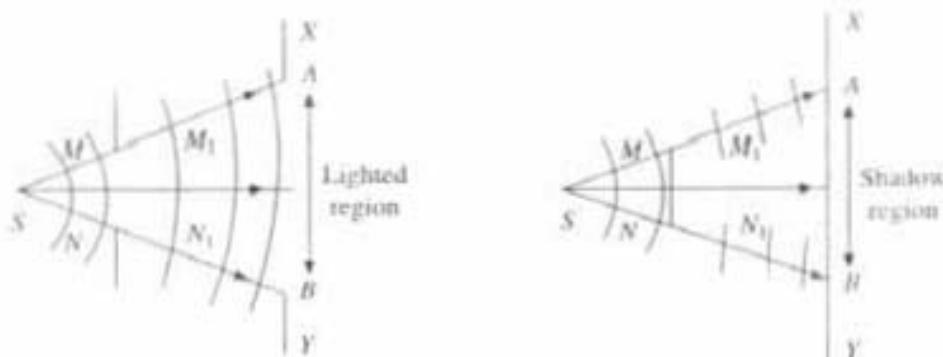


Fig. 1.18 (a) Light from a point source (s) passing through an aperture MN
 (b) Light from a point source (s) obstructed by an obstacle MN

interfere with each other, which are not blocked by the small obstacle. Due to which when a small obstacle Fig. 1.18(b), whose size is comparable to the wave length of the light used is placed in the path of the light, then light rays bend round the edge of that obstacle and the shadow formed by the small obstacle are not very sharp. This phenomena of bending of light round the edge of an obstacle or the encroachment of light within geometrical shadow area is called diffraction.

But in general the illumination in the geometrical shadow area due to point source not commonly observed, because firstly the light sources are not point sources and for an extended source of light, every point of the extended surface forms its own diffraction pattern and after that they will overlap with each other. So no single diffraction pattern can be identified. Secondly the obstacles which are generally used, are of very large size compared to the wave length of light used, so practically no diffraction occurs.

Augustin Jean Fresnel in 1815 satisfactorily explained the diffraction phenomena, i.e., the bending of light round the edge of a small obstacle and also the rectilinear propagation of light by combining the Huygens theory of secondary wavelets with the principle of interference.

1.19 Fresnel Half-period Zone^{3,4}

According to Fresnel, the resultant effect at an external point due to a wave front will depend on the following factors:

- (1) A wave front can be divided into a large number of strips or zones which are called as Fresnel's zones of small area and the resultant effect at any point will be the combined effect from all secondary waves emanating from the various zones.
- (2) The effect at a point due to any particular zone will depend on the distance of the point from that zone.
- (3) The effect at a point will also depend on the obliquity of the point with reference to the zone under consideration, with the increase of the obliquity effect will decrease.

Let us consider a plane wave front W_1W_2 perpendicular to the plane of the paper propagating along Z-direction (Fig. 1.19) and P is an external point at a perpendicular distance b from W_1W_2 . In order to determine the resultant effect at the point P due to disturbances reaching from the different points of the wave front, Fresnel's method, consists in dividing the whole wave front in a number of half-period elements or zones called as Fresnel's half-period zones, and then find the combined effect from all zones at P .

With P as centre draw spheres of radii $b + \lambda/2, b + 2\lambda/2, b + 3\lambda/2$ etc. These sphere will intersect the wave front W_1W_2 in concentric circles as shown in the (Fig. 1.19a, b). The radii of the n th circle from the above Fig. 1.19(a) will be

$$\begin{aligned} R_n &= [(b + n\lambda/2)^2 - b^2]^{1/2} = \left(b^2 + 2bn\lambda/2 + \frac{n^2\lambda^2}{4} - b^2 \right)^{1/2} \\ &= \sqrt{nb\lambda} [1 + n\lambda/4b]^{1/2} \end{aligned}$$

Or

$$R_n = \sqrt{nb\lambda} \quad (1.40)$$

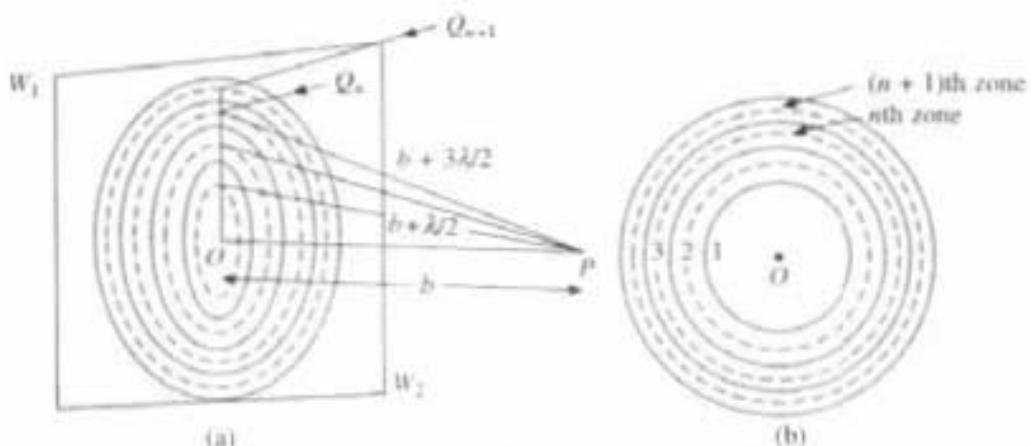


Fig. 1.19 Construction of Fresnel Half period zone

where the term $n^2\lambda^2/4$ is neglected in comparison to $nb\lambda$, because $b > \lambda$. The area of the annular region between two consecutive concentric circles will be

$$\pi R_{n+1}^2 - \pi R_n^2 = \pi[(n+1)b\lambda - nb\lambda] = \pi b\lambda \quad (1.41)$$

Thus the areas of all the annular regions are having the value ($\pi b\lambda$), which is not depends on n , as n increases thickness of zone decreases. *The annular region between the (n - 1)th and the nth circles is known as nth half-period zone.*

Now according to the Fig. 1.19(a), the path difference between any two consecutive zones say $Q_{n+1}P - Q_nP = \lambda/2$, which corresponds to phase difference $\left(\delta = \frac{2\pi}{\lambda} \cdot \frac{\lambda}{2} = \pi\right)$ of (π) i.e. they are out of phase. Thus the secondary waves reaching to the point P from different zones are continuously out of phase and in phase with reference to the central or the first half-period zone.

Let M_1, M_2, M_3, M_4 , etc. represent the amplitudes of vibration of the ether particles at P due to secondary waves from the 1st, 2nd, 3rd, etc. half-period zones. Now firstly as we consider earlier that as the zones going out ward from O, the obliquity increases and hence amplitudes from different zones will be continuously in decreasing order (Fig. 1.20). Secondly due to the phase difference (π) between any two consecutive zones if the displacement of the ether particle due to odd number zones is in positive direction, then due to even number of zones the displacement will be in negative direction at the same instant as shown in the (Fig. 1.20). Since the amplitude is gradually decreasing as zone numbers are increasing, so amplitude of any zone can be approximately taken as mean of the amplitudes due to the zones preceding and succeeding it, as for example

$$M_2 = \frac{M_1 + M_3}{2} \quad (1.42)$$

So the resultant amplitude at the point P will be

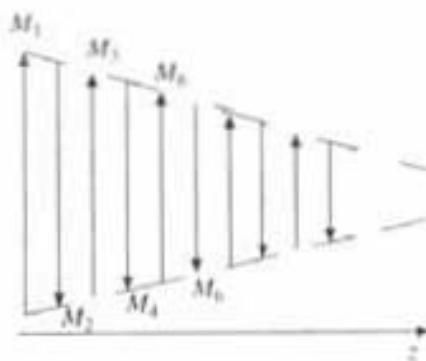


Fig. 1.20 Gradual decrease of amplitude

$$A = M_1 - M_2 + M_3 - M_4 + \dots + (-1)^{n+1} \quad (1.43)$$

Now the above equation can also be written as

$$A = \frac{M_1}{2} + \left[\frac{M_1}{2} - M_2 + \frac{M_3}{2} \right] + \left[\frac{M_3}{2} - M_4 + \frac{M_5}{2} \right] + \text{etc}$$

which can be transformed in the following way by using the Eqn. (1.42)

$$A = \frac{M_1}{2} + \frac{M_n}{2} \text{ when } n \text{ is odd}$$

$$\text{and } A = \frac{M_1}{2} + \frac{M_{n-1}}{2} - M_n \text{ when } n \text{ is even.}$$

Now when $n \rightarrow \infty$, then M_n and $M_{n-1} \rightarrow 0$, then the resultant amplitude due to the whole wave front $W_1 W_2$ will be

$$A = \frac{M_1}{2} \quad (1.44)$$

and the intensity at the point P due to whole unobstructed wave front will be

$$I \propto A^2 \propto \left[\frac{M_1}{2} \right]^2 \propto M_1^2 / 4 \quad (1.45)$$

Thus the resultant intensity at P due to unobstructed whole wave front is only one-fourth of that due to 1st half-period zone alone. Here only half the area of the 1st half-period zone is effective in producing the illumination at the point P . So when an obstacle of the size of half the area of the 1st half-period zone is placed at O , then that will screen off the effect of whole wave front and the intensity at P due to rest of the wave front will be zero.

Now while considering the rectilinear propagation of light, generally the size of the obstacle used is far greater than the area of the 1st half-period zone and hence bending of light rays around the edge of the obstacle, i.e., the diffraction effect cannot be noticed. But if the size of the obstacle less than the area of the 1st half-period zone ($= \pi b \lambda$) i.e., when the size of the obstacle is comparable to the wave length (λ) of the light used then it is possible to see illumination in the geometrical shadow region, which proves the diffraction phenomena.

The above analysis can be used to study the diffraction of a plane wave by a circular aperture. Let the point P at a distance b of a circular aperture, now as the radius (r) of the circular aperture increases from zero to the radius of the 1st half-period zone ($R_1 = \sqrt{b\lambda}$) the intensity at point P will be maximum, as the resultant amplitude is (M_1). But as the radius of the aperture increases more than R_1 , intensity at P starts decreasing and when the radius of the aperture is ($R_2 = \sqrt{2b\lambda}$) the intensity at P is minimum as the amplitude ($M_1 - M_2$) $\rightarrow 0$. So for an circular aperture whenever the distance

$$OP = b = R_1^2 / \text{odd multiple of wave length } (\lambda), \text{ point } P \text{ will be brightest} \quad (1.46)$$

$$\text{and } OP = b = R_1^2 / \text{even multiple of wave length } (\lambda), P \text{ will be darkest} \quad (1.47)$$

1.20 Zone Plate

Zone plate is a specially constructed screen such that light obstructed from every alternate zone as shown in the (Fig. 1.21a). Radii of the zones are proportional to the square roots of the natural numbers according to the eqn. (1.40). The correctness of Fresnel's method in dividing the whole wave front W_1W_2 (Fig. 1.19, 1.21) into half-period zones can be verified with its help. Here even number of zones are blackened so that they will obstruct the light wave.

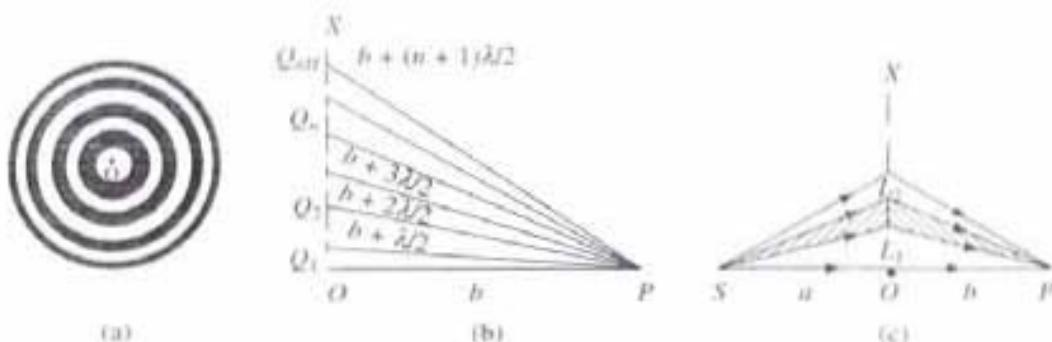


Fig. 1.21 (a) Zone plate (b) Zone plate with broad source (c) Zone plate with point source

Let us consider a point P at a perpendicular distance b from the zone plate as shown in the (Fig. 1.19, 1.21b). If the source is at a large distance from the zone plate, a bright spot will be observed at P . As the distance of the source from zone plate is large, the incident wave front can be taken as a plane wave front w.r.t. the small area of the zone plate. According to the (Fig. 1.21), even number of zones cut off the light and hence the resultant amplitude at the point P due to the partially obstructed wave front from the zone plate will be

$$A = M_1 + M_3 + M_5 + \dots + M_n \quad (n \text{ is odd}) \quad (1.48)$$

Since the amplitudes from the even number of the zone plates are obstructed by the zone plate (i.e., M_2, M_4, M_6 etc. = 0). So it is interesting to note that as the even-numbered zones are blackened, the intensity at P according to eqn. (1.48) is much greater than that when the whole wave-front is exposed to the point P , since as we have discussed in earlier section (Fig. 1.19 and eqn. 1.44), the resultant amplitude at P due to an unobstructed wave front W_1W_2 is

$$A = M_1 - M_2 + M_3 - M_4 + \dots + (-1)^{n+1} = \frac{M_1}{2} \quad (\text{when } n \rightarrow \infty) \quad (1.44)$$

Now when a parallel beam of white light is incident on the zone plate, different colours come to focus at different points along the line (OP) (Fig. 1.21). Thus, the function of a zone plate is similar to that of convex lens. From the (Fig. 1.21b) it's clear that the zone plate is giving the convergence effect of light beams for the odd numbered zones at the point P like convex lens. The focal length of the zone plate according to the (Fig. 1.21) will be from eqn. (1.40)

$$F_n = b = Rn^2/\lambda \quad (1.49)$$

Thus a zone plate has different foci for different wave-lengths, which shows that the zone plate suffer from considerable chromatic aberrations.

The lens formula like convex lens, i.e., the relation between object distance, image distance and the focal length can also be obtained for a zone plate in the following way. Let us consider a point source at S . XO is the section of the zone plate perpendicular to the plane of paper and P is position of bright image on the screen as shown in the (Fig. 1.21c). The distance from the source S to the zone plate is a and the distance of the point P from the zone plate is b . $OL_1, OL_2(R_1, R_2 \text{ etc.})$ are the radii of 1st, 2nd, etc. half-period zones. The position of the screen is such, that from one zone to the next there is an increasing path difference of $\lambda/2$.

Thus

$$SO + OP = SP = a + b$$

$$SL_1 + L_1 P = a + b + \lambda/2$$

$$SL_2 + L_2 P = a + b + 2\lambda/2$$

i.e.

$$SL_1 + L_1 P - SP = \lambda/2$$

where L_1 is on the periphery of the 1st circle of the zone plate and the radius of the 1st circle is R_1 , so from the (Fig. 1.21c)

$$SL_1 + L_1 P - SP = \sqrt{a^2 + R_1^2} + \sqrt{b^2 + R_1^2} - (a + b) = \lambda/2$$

i.e. $\lambda/2 = a \left[1 + \frac{R_1^2}{2a} \right] + b \left[1 + \frac{R_1^2}{2b} \right] - (a + b)$ (, $a \gg R_1$ and $b \gg R_1$)

$$\lambda/2 = \frac{R_1^2}{2} \left[\frac{1}{a} + \frac{1}{b} \right]$$

i.e.

$$\left[\frac{1}{a} + \frac{1}{b} \right] = \lambda/R_1^2 = \frac{1}{F}$$

Now by considering sign convention ($a = -ve$ but $b = +ve$)

i.e.

$$\left[\frac{1}{b} - \frac{1}{a} \right] = \lambda/R_1^2 = \frac{1}{F} \quad (1.50)$$

or in general

$$\left[\frac{1}{b} - \frac{1}{a} \right] = n\lambda/R_n^2 = \frac{1}{F_n} \quad (1.51)$$

which is similar to the well-known lens equation

$$\left[\frac{1}{v} - \frac{1}{u} \right] = \frac{1}{F} \quad (1.52)$$

where a and b are the object and image distance and $F_n = R_n^2/n\lambda$ are the focal lengths for different wave lengths. Thus zone plate act as a converging lens with number of foci which depends on the number of the zones (n) and different wave lengths (λ).

1.21 Comparison Between Zone Plate and Convex Lens

It is well known that for a particular wave length of light convex lens have one focal length, because

$$\frac{1}{F} = (\mu - 1) \left[\frac{1}{R_1} - \frac{1}{R_2} \right] \quad (1.53)$$

where F is the focal length, μ is the R.I. of the material of the lens and R_1, R_2 are the radii of curvature. Since the R.I. of violet rays is more than R.I. of red light ($\mu_v > \mu_r$), so violet light will focus nearer to the lens than red light, where as from Eqn. (1.51) it is clear in zone plate focal length of violet will be more than focal length of red as ($\lambda_v \ll \lambda_r$).

While constructing the zone plate, it is constructed according to the (Fig. 1.21) for a particular value of b . But during the experiment with zone plate the distance b , i.e. the image distance (Fig. 1.21) i.e., the position of P can be changed. Now as the screen is moved nearer to the zone plate b will reduce, the area of the half-period elements according to eqn. (1.41) ($\pi b \lambda$) will also reduce, so more number of half-period elements can be present on each zone which is already constructed on the zone plate. Now if, P_m is the position of the image when $(2m - 1)$ half-period elements can be present on each constructed zone on the zone plate, then focal length of the zone plate will be

$$F_m = \frac{R_1^2}{(2m - 1)n\lambda} \quad (1.54)$$

Say for a particular position of the screen each zone contain three half period elements (i.e. $m = 3$), then the focal length for the 1st half-period zone (i.e. $n = 1$) will be

$$F_1 = \frac{R_1^2}{(2m - 1) \cdot 1 \cdot \lambda} \Rightarrow F_1 = R_1^2 / \lambda, F_2 = R_1^2 / 3\lambda, F_3 = R_1^2 / 5\lambda$$

which will be on the line (OP) (Fig. 1.21c) as P_1, P_2, P_3 and out of all these three P_1 will be most intense, since at this point resultant amplitude is (M_1), where as at P_2 resultant amplitude is ($M_1 - M_2$) and at the point P_3 resultant amplitude is ($M_1 - M_2 + M_3$).

Problems on the Zone Plate

- Assume a plane wave of wave length 5000 Å is incident on a circular aperture of radius 0.5 mm. Calculate the position of the brightest and darkest spot on the axis.

Solution:

From Eqn (1.46 and 1.47) for brightest spot $R_1^2 = b\lambda$

So distance on the axis is $b = OP = R_1^2 / \lambda = 0.0025/5 \times 10^{-5} = 50$ cm

For darkest spot $b = OP = R_1^2 / 2b\lambda = R_1^2 / 2 \times 5 \times 10^{-5} = 25$ cm

- With a zone plate for point source of light ($\lambda = 6000$ Å), the strongest and the next strongest images are formed on the axis at 30 cm and 6 cm, respectively, from the zone plate. Both the images are on the same side and on the other side

of the source. Now calculate (a) the distance of the source from the zone plate, (b) the radius of the 1st zone, (c) principal focal length.

Solution

$$\text{From Eqn. (1.54) for a zone plate } F_w = \frac{R_n^2}{(2m-1)n\lambda}$$

For 1st zone $n = 1$, and for strongest image $m = 1$ for next strongest $m = 2$, so $F_1 = R_1^2/\lambda$ and $F_2 = R_1^2/3\lambda$

$$\text{Now from Eqn. (1.51)} \left[\frac{1}{b} - \frac{1}{a} \right] = n\lambda/R_1^2 = \frac{1}{F_r}$$

$$\frac{1}{30} - \frac{1}{a} = \lambda/R_1^2 \text{ and } \frac{1}{6} - \frac{1}{a} = 3\lambda/R_1^2$$

i.e., distance of the source from the zone plate = $a = -30 \text{ cm}$

By substituting the value of a in the above equation

$$R_1^2 = \frac{6 \times 10^{-5}}{\left[\frac{1}{30} + \frac{1}{30} \right]} = 90 \times 10^{-5} \quad \text{i.e., } R_1 = 0.003 \text{ cm}$$

By substituting the value of R_1 and λ , the principal focal length (F_1) will be $F_1 = R_1^2/\lambda = 90/6 = 15 \text{ cm}$.

1.22 Fresnel and Fraunhofer Diffraction

Diffraction phenomenon was correctly explained by Fresnel as it is due to the mutual interference of secondary wavelets originating from various points of the wave front, which are not blocked off by the obstacle. Here he applied the Huygen's principle of secondary wavelets with the principle of interference. Diffraction phenomena can be divided into two different classes.

(1) *Fresnel's diffraction* is where the source and screen both are placed at finite distances from the obstacle having sharp edge. No lenses or mirrors are used to make the rays parallel or convergent. The incident wave front to the obstacle is either spherical or cylindrical. But when the source is placed at a large distance then the spherical wave front will become as an plane wave front, which we have already discussed in sections 1.19 to 1.21.

(2) *Fraunhofer diffraction* is where the source and the screen or telescope, through which diffraction pattern is observed are placed at infinity or effectively at infinity, i.e., incident rays become parallel with a lens and the diffracted beam is focused on the screen with another lens. Now onwards we will discuss Fraunhofer Diffraction.

1.23 Fraunhofer Diffraction at Single Slit

Let a monochromatic parallel beam of light be incident on the slit AB of width a (Fig. 1.22). The secondary waves travelling along same direction come to focus at M_0 . If the width of the slit a is comparable with the wave length of the monochromatic light used, then secondary waves will also diffract at an angle θ and come to focus at P_1 . Now consider two diffracted beams AA_1, BB_1 from the

two extreme points of the single slit (Fig. 1.22). If the path difference AY between two diffracted rays AA_1 and BB_1 is multiple of λ then P_1 will be dark. It is so, because when $AY = \lambda$ then $OD = \lambda/2$, where O is the middle point of the slit. Thus, the path difference between BB_1 and OD is $\lambda/2$, similarly path difference between OD and AA_1 is again $\lambda/2$, and when path difference is of the order of $\lambda/2$, then these two rays will give destructive interference at P_1 . Thus, for every point in the upper half OA , there is a corresponding point in the lower half OB and they destroy each other by interference in between themselves. Hence the condition for *diffraction minima* is

$$\Delta = \text{Path difference} = AY = a \sin \theta = p\lambda \quad (1.50)$$

where $p = 1, 2, 3, \dots$

$$\text{and } \delta = \text{phase difference} = \frac{2\pi}{\lambda} (a \sin \theta) \quad (1.50a)$$

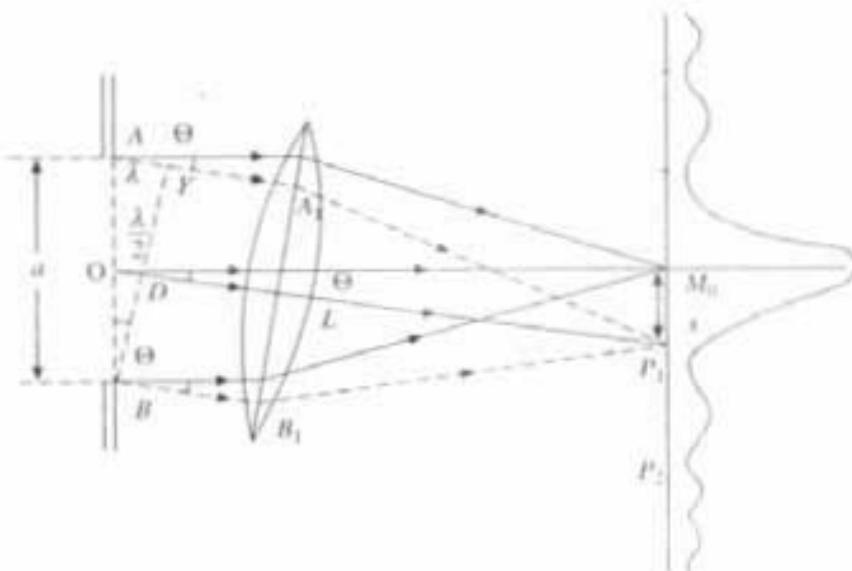


Fig. 1.22 Diffraction and diffraction pattern from a single slit

The condition for *diffraction maxima* is

$$\text{Path difference } AY = a \sin \theta = (2p \pm 1)\lambda/2 \quad (1.50b)$$

The diffraction pattern due to single slit contains a central bright maximum at M_0 (Fig. 1.22), followed by secondary maxima and minima on both the sides of M_0 . Intensity distribution on the screen is as shown in Fig. (1.22). The secondary maxima are of much less intensity and that intensity falls off rapidly from the point M_0 outwards. Secondary minima will come when the path difference between two diffracted beams AA_1 and BB_1 is $\lambda, 2\lambda, 3\lambda$, etc.

If the lens L in Fig. (1.22) is very near to the slit or the screen is far away from the lens L then

$$\sin \theta = x/f$$

where f is the focal length of the lens L and $x = M_0P_1$ (Fig. 1.22). But from Fig. (1.22) it is clear from the triangle ABY that

$$\sin \theta = AY/a = \lambda/a$$

So,

$$x/f = \lambda/a$$

$$x = \frac{\lambda f}{a}$$

where x is the distance of the first secondary minima P_1 from the central maxima M_0 . So the width of the central maximum is $2x$ where

$$2x = \frac{2f\lambda}{a} \quad (1.50c)$$

From the equation (1.50c), we can conclude

(1) Width of the central maximum is proportional to λ . When λ is more for red then width of the central maximum will be more. (2) Width of central maximum is inversely proportional to a , width of the slit. So if the slit is narrow, width of the central maximum will be more. (3) Diffraction pattern consists of alternate dark and bright bands with monochromatic light. With white light central band will be white because all waves will superimpose at $\theta = 0$, at central maximum. Rest of the diffraction band will be coloured because θ depends on λ . (4) When width of the slit is large, θ will be small, so width of the central maximum will be small and secondary minima and maxima will be very close to central maximum. But if a is small then they are all well separated.

1.24 Fraunhofer for Diffraction at Double Slit

Let monochromate light fall on a double slit Fig. (1.23) AB and CD , parallel to each other and perpendicular to the plane of the paper. The width of each slit is a and the opaque portion is b . Let the wave front be incident on the surface HK and all the secondary waves travelling in a direction parallel to OM_0 , come to focus at M_0 . So M_0 is the position of central maximum.

Here, the diffraction pattern has to be considered in two parts. Since the slits are very narrow, so each slit will give rise to diffracted rays. These diffracted rays then interfere with one another and give final pattern. *Thus the final pattern is a superposition of interference pattern on diffraction pattern (Fig. 1.24).* If for

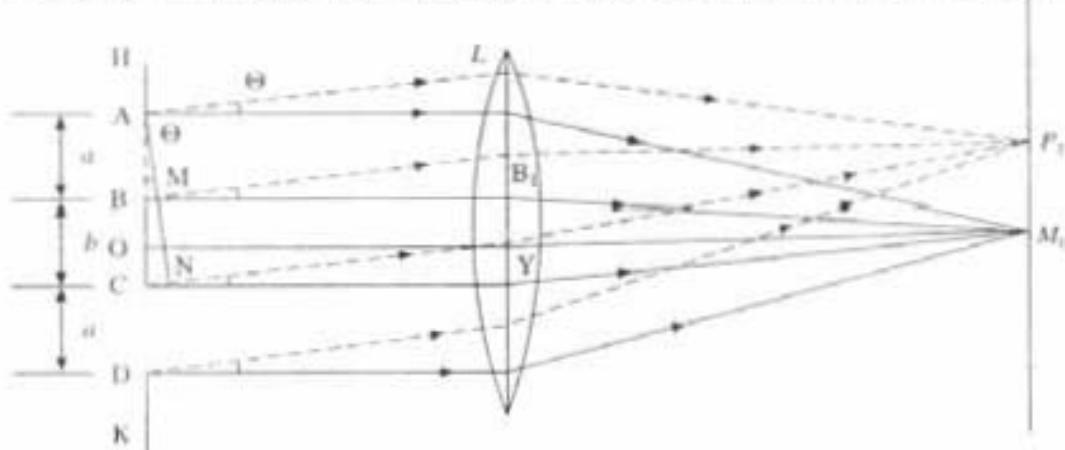


Fig. 1.23 Diffraction from double slit

particular direction θ , interference maxima and diffraction minima condition is satisfied simultaneously, then that spectrum of particular order (say M_1 when $2a = b$, in the Fig. 1.24) will be absent in that direction (Fig. 1.24).

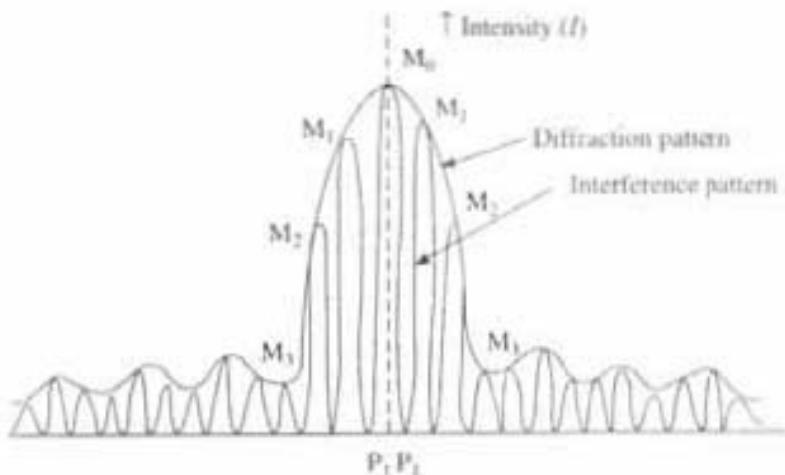


Fig. 1.24 Diffraction pattern due to double slit for $2a = b$

Interference Maxima and Minima

Let us consider the secondary waves diffracted in a direction inclined at an angle θ , with initial direction. So the path difference between the consecutive interfering rays AL and CY (Fig. 1.23), from two consecutive slits is

$$CN = (a + b) \sin \theta$$

Now if this path difference is equal to odd multiple of $\lambda/2$, θ will give direction of interference minima. So the condition of *interference minima* for double slit is

$$(a + b) \sin \theta = (2n \pm 1)\lambda/2 \quad (1.51a)$$

where $n = 1, 2, 3, \dots$ gives the 1st, 2nd, 3rd order interference minima.

On the other hand if it is even multiple of $\lambda/2$, it will give interference maxima. So the condition for *interference maxima* is

$$(a + b) \sin \theta = 2n\lambda/2 = n\lambda \quad (1.51b)$$

where $n = 0, 1, 2, \dots$ gives central, 1st, 2nd, ... order interference maxima (Fig. 1.24). So point P_i in Fig. 1.23 may give the interference maxima or minima position, depending upon the path difference.

Diffraction Maxima and Minima

Let us consider the secondary waves diffracted in a direction inclined at an angle θ , with the initial direction of incident light. The path difference between two diffracted beam AL and BB_1 emanating from the extremities of the first slit is (Fig. 1.23)

$$BM = a \sin \theta$$

Now if this path difference is even multiple of λ , then θ will give direction of diffraction minimum, which can be explained in the same way, as it was done for single slit. Consider the wave front AB to be made up of two halves, if the path difference between the corresponding points of the upper and the lower half is equal to $\lambda/2$, then all the secondary waves will interfere destructively. Thus, the direction θ , may give the direction of diffraction minima. Then point P_1 (Fig. 1.23) may give diffraction minima position. Hence, in general, the condition for *diffraction minima* is,

$$BM = a \sin \theta = p\lambda \quad (1.52a)$$

where $p = 1, 2, \dots$, except zero. On the other hand,

$$\text{when } BM = a \sin \theta = (2p \pm 1)\lambda/2 \quad (1.52b)$$

then the direction θ will give the direction of *diffraction maxima*.

Figure (1.24) shows the boundary curve, which represents the intensity distribution of diffraction pattern due to double slit, and the inner lines represent the intensity distribution due to the interference between the diffracted lights from both the slits. *The equally spaced interference pattern is superimposed on the diffraction pattern of the double slit.* The region originally occupied by the central maximum of the single slit diffraction pattern is now superimposed by the equally spaced interference maxima and minima. The intensity of the central maximum for double slit is four times than the intensity of the central maximum for the single slit, as we will see in the following discussions. The intensity of the other secondary maxima are gradually decreasing on the two sides of the central maximum. The intensity distribution due to diffraction by double slit with $2a = b$ is as shown in Fig. 1.24, where the region of the central and secondary maxima diffraction pattern due to double slit is superimposed by the equally spaced interference maxima of low intensity. For the case $2a = b$ there will be five interference maxima in the central diffraction maximum. Third interference maxima (M_3) will be in a position of first diffraction minima, so it will be absent [see for details, section 1.27, case (3)].

1.25 Fraunhofer Theory of Diffraction Grating

One of the most important and interesting applications of diffraction is the diffraction grating. It consists of a large number of extremely narrow parallel slits separated by equal opaque space. There are two types of grating: transmission grating and reflecting grating.

Transmission gratings are prepared by ruling fine lines extremely close together on the surface of a glass plate by using a fine sharp diamond point. In that, diamond scratches act as opaque space and the transparent space in between two lines act as slit.

Figure (1.25) represents the section of a grating whose slits are perpendicular to the plane of the paper. Only four slits have been shown in Fig. (1.25) but an actual grating contains thousands of them per centimeter. The width of each slit is a and distance between two slits is b . The distance $(a + b)$ is known as *grating element*. If number of slits per centimeter of the grating is N then $(a + b) = 1/N$ cm. But if $N = 15,000$ line, per inch, then

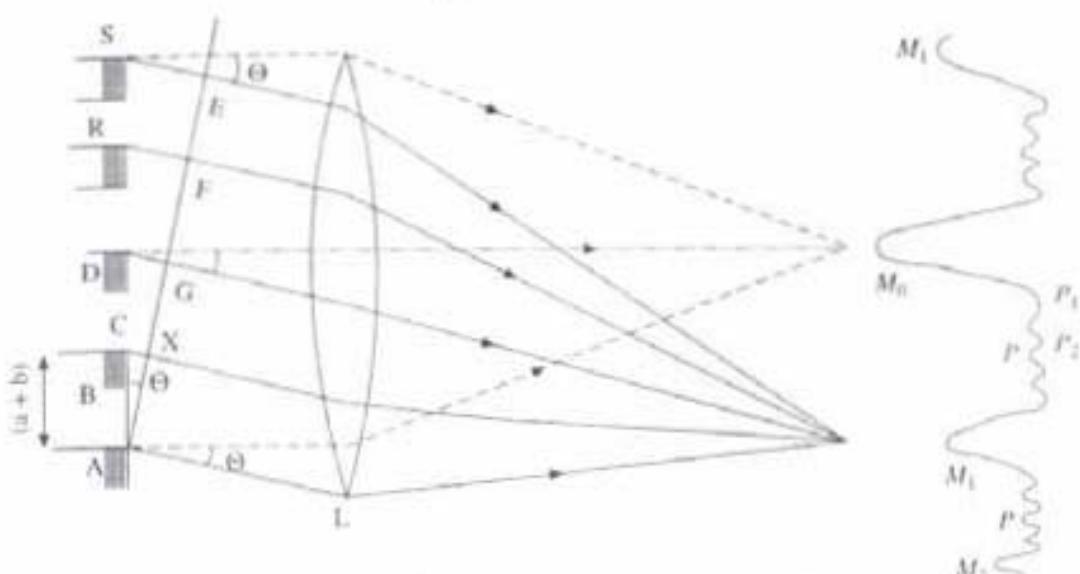


Fig. 1.25 Diffraction and diffraction pattern from diffraction grating

$$(a + b) = \frac{2.54}{N} = \frac{2.54}{15,000} = 1.6 \times 10^{-4} \text{ cm} \quad (\because 1 \text{ inch} = 2.54 \text{ cm})$$

$$\text{i.e.} \quad a = b = 0.8 \times 10^{-4} \text{ cm} = 8000 \text{ \AA}$$

So the wave length of sodium light ($\approx 6000 \text{ \AA}$) is comparable to the obstacle size ($= b$), which is essential for diffraction.

To get the Fraunhofer diffraction pattern in a transmission grating, the incident wave front must be plane and the diffracted light is collected on the screen with the help of the lens L . Let a parallel beam of monochromatic light of wave length λ be incident on the transmission grating. The problem of finding the intensity of light transmitted by grating involves the principles of interference and diffraction. Since the slits are very narrow, comparative to the wave length of light, so each gives rise to diffracted rays. These diffracted rays then interfere with one another to produce the final pattern. Thus, the final pattern is a superposition of interference pattern on diffraction pattern (Figs. 1.25 and 1.26) consisting of:

1. Widely-spaced diffraction maxima (known as primary maxima) such as M_1, M_2, \dots etc. (Fig. 1.25) on both the sides of central maximum M_0 .
2. A large number of faint subsidiary maxima and minima in between two primary maxima on both the sides of M_0 (Fig. 1.25).

Condition for Primary Maxima

Let a plane wave front be incident on the grating surface normally. By Huygen's principle, each of these points in the slits sends secondary wavelets in all directions. Most of the secondary waves proceed straight and when collected by the lens L produce a strong bright point at M_0 . It is known as central maximum. Next, consider the light diffracted in the direction making an angle θ with that of incident beam. The collecting lens L also is properly rotated such that those diffracted rays interfere with each other and get focused at M_1 . The intensity at M_1 will depend upon the path difference between secondary waves originating

from the corresponding points A and C of two neighbouring slits. In Figure 1.25 $AB = a$ and $BC = b$. So the path difference between two interfering secondary waves from A and C from two consecutive slits is

$$\text{Path difference} = \Delta = CX = AC \sin \theta = (a + b) \sin \theta$$

The point M_1 will be of maximum intensity, if this path difference is equal to $n\lambda$. Thus for a particular value of diffracted angle $\theta = \theta_1$ when $CX = \lambda$, we will get 1st order primary maxima at M_1 . If the diffracted angle θ is increased to θ_2 such that path difference CX between two consecutive interfering rays from A and C is 2λ then that will give 2nd order primary maxima at M_2 and so on. Hence the general condition of *interference primary maxima* is

$$\text{Path difference} = \Delta = CX = (a + b) \sin \theta_n = n\lambda \quad (1.53a)$$

where $n = 0, 1, 2, 3, \dots$ and θ_n = angle of diffraction for n th. Primary maxima. For $\theta = 0$, ($n = 0$) give the direction of central maximum M_0 .

Condition of secondary minima

When path difference $CX = \lambda$ and angle of diffraction θ_1 , 1st order primary at M_1 will come (Fig. 1.25). Now if the angle of diffraction θ_1 be increased by $d\theta_1$, such that the path difference CX between two consecutive interfering secondary rays for A and C increases by λ/N , then the path difference between the secondary waves for two extreme points A and S of the grating surface which contain N number of slits will be, $\lambda/N \times N = \lambda$. Now assume that the whole wave front is to be divided into two halves, then the path difference between the corresponding points of the two halves will be $\lambda/2$, and all the secondary waves will interfere destructively. Thus $(\theta_n + d\theta_1)$ will give the direction of the first secondary minima after n th primary maximum. Similarly if for higher value of $d\theta = d\theta_2$, CX is increased by $2\lambda/N$, then the path difference between two extreme rays of the grating surface from A and S will be $2\lambda/N \times N = 2\lambda$. Then whole wave front will be divided into four parts, and 2nd secondary minima will be produced and so on. Thus, in general for gradually increasing value of $d\theta$ from θ_n the path difference between two consecutive rays from A and C goes on increasing as

$\lambda/N, 2\lambda/N, \dots, \frac{(N-1)\lambda}{N}$ and produces at these angles the 1st, 2nd, 3rd ...

$(N-1)$ th order secondary minima after n th primary maximum. So the number of secondary minima in between two primary maxima is $(N-1)$ and number of secondary maxima is automatically $(N-2)$.

Thus, the condition for 1st secondary minima after n th primary maxima is

$$(a + b) \sin (\theta_n + d\theta_1) = n\lambda + \lambda/N \quad (1.53b)$$

whereas the condition for 1st secondary minima after central maximum is

$$(a + b) \sin (0 + d\theta_1) = 0 + \lambda/N = \lambda/N$$

Similarly, the condition for 2nd secondary minima after 1st primary maxima is

$$(a + b) \sin (\theta_1 + d\theta_2) = \lambda + 2\lambda/N$$

1.26 Resultant Intensity

Resultant intensity distribution for single slit, double slit and diffraction grating, will be discussed one by one; because all are related to each other.

1.26.1 For Single Slit

The final intensity at any point may be considered by applying the theory of Fraunhofer diffraction at a single slit (Fig. 1.22). The wavelets proceeding from all points in a slit along the direction θ , are equivalent to a single wave starting from the middle point of the slit. If the width of the single slit is divided into n equal parts and the amplitude of the wave from each part is a' , then the phase difference between any two consecutive waves from these n -parts would be

$$1/n \text{ (total phase)} = 1/n \left(\frac{2\pi a \sin \theta}{\lambda} \right) = x \text{ (say)}$$

Now by using the method of vector addition the resultant amplitude A of the actual wave from a single slit is

$$A = a' \frac{\sin nx/2}{\sin x/2} = a' \frac{\sin \left(\frac{\pi a}{\lambda} \sin \theta \right)}{\sin \left(\frac{\pi a}{n\lambda} \sin \theta \right)}$$

Hence, $A = a' \frac{\sin \alpha}{\sin \alpha/n} = a' \frac{\sin \alpha}{\alpha/n}$ ($\because \frac{\alpha}{n}$ is very small)

$$A = na' \sin \alpha/\alpha$$

where

$$\alpha = \frac{\pi a \sin \theta}{\lambda}$$

Thus, the intensity I at any point for single slit will be

$$I = A^2 = (na')^2 \left(\frac{\sin^2 \alpha}{\alpha^2} \right)$$

or

$$I = I_0 \sin^2 \alpha/\alpha^2 \quad (1.54)$$

when $I_0 = (na')^2$.

Now for central maximum M_0 , ($\theta = 0$, $\alpha = 0$)

$$\lim_{\alpha \rightarrow 0} \frac{\sin \alpha}{\alpha} = 1$$

Hence the intensity of central maximum at M_0 due to single slit (Fig. 1.22), is

$$I = \lim_{\alpha \rightarrow 0} I_0 \left(\frac{\sin^2 \alpha}{\alpha^2} \right) = I_0 \text{ which is maximum} \quad (1.54a)$$

1.26.2 For Double Slit

In double slits, there we have two central diffracted waves, each one from the middle point of the two slits. So the path difference between two consecutive rays from two consecutive slits is $(a + b) \sin \theta$ (Fig. 1.23)

Therefore

$$CN = (a + b) \sin \theta$$

So the corresponding phase difference is

$$\frac{2\pi}{\lambda} (a + b) \sin \theta = 2\beta$$

Now the final intensity in a direction θ , of 2-diffracted rays each of amplitude $A = na' (\sin \alpha / \alpha)$ and having common phase diffraction $2\beta = \frac{2\pi}{\lambda} (a + b) \sin \theta$, can be calculated in the following way:

Again by using the method of vector addition of amplitude, the resultant amplitude from 2-slits in a direction θ will be,

$$A' = A \frac{\sin 2\beta/2}{\sin \beta/2} = A \frac{\sin 2\beta}{\sin \beta}$$

$$A' = na' \frac{\sin \alpha}{\alpha} \frac{\sin 2\beta}{\sin \beta}$$

Thus the intensity I at any point (Fig. 1.19) for double slit is

$$I = (A')^2 = (na')^2 \frac{\sin^2 \alpha}{\alpha^2} \frac{\sin^2 2\beta}{\sin^2 \beta}$$

$$I = I_0 \frac{\sin^2 \alpha}{\alpha^2} \frac{\sin^2 2\beta}{\sin^2 \beta} \quad (1.55)$$

For central maximum in double slit $\theta = 0$.

Therefore, $\alpha = 0$, $\beta = 0$ which gives

$$\lim_{\alpha \rightarrow 0} \frac{\sin^2 \alpha}{\alpha^2} = 1 \text{ and } \lim_{\beta \rightarrow 0} \frac{\sin^2 2\beta}{\sin^2 \beta} = (2)^2 = 4 \text{ (by using L'Hospital's rule)}$$

Hence the intensity of central maximum at M_0 due to double slit is

$$I = 4I_0 \quad (1.55a)$$

when $\alpha = 0$ and $\beta = 0$.

Thus, the intensity of the central maxima in case of double slit is four times than that of single slit.

The intensity distribution shown in Fig. (1.24) corresponds to $2a = b$ where a is the width of the slit and b the opaque spacing between two slits. Thus the pattern due to diffraction at a double slit, consists of a diffraction pattern due to the individual slits of width a each, which is super-imposed by the interference pattern maximum and minimum of equal spacing due to the interference of those

two diffracted beams. The spacing of the interference maximum and minimum is dependent on the values of a and b .

1.26.3 For Diffraction Grating

Now in diffraction grating, there are N number of slits, so we have N -central diffracted wave, one each from the middle points of the slits. So the path difference (CX) between two consecutive rays from two consecutive slits is $(a + b) \sin \theta$ (Fig. 1.25), therefore the corresponding phase difference is,

$$\frac{2\pi}{\lambda} (a + b) \sin \theta = 2\beta$$

$$\boxed{\beta = \frac{\pi}{\lambda} (a + b) \sin \theta}$$

Again by using the method of vector addition of amplitude, the resultant amplitude from N -slits in a direction θ will be

$$A^* = na' \frac{\sin \alpha}{\alpha} \frac{\sin N(2\beta/2)}{\sin (2\beta/2)}$$

i.e.,

$$A^* = na' \frac{\sin \alpha}{\alpha} \frac{\sin N\beta}{\sin \beta}$$

and $\boxed{\text{Intensity} = I = A^{*2} = I_0 \frac{(\sin \alpha)^2}{\alpha^2} \frac{(\sin N\beta)^2}{(\sin \beta)^2}} \quad (1.56)$

The maximum intensity will come for $\alpha = 0$ and $\beta = 0$, i.e. from equations (1.54 and 1.55) when $\theta = 0$, which gives the position of central maximum M_0 , when $\alpha \rightarrow 0, \frac{\sin \alpha}{\alpha} = 1$, and when $\beta \rightarrow 0, \sin N\beta / \sin \beta = N$.

So the *maximum intensity at central maximum M_0* in case of diffraction grating from equation (1.56) will be

$$\boxed{I = N^2 I_0} \quad (1.56a)$$

Thus, general intensity distribution for transmission grating can be written as

$$\boxed{I = I_0 (\sin^2 \alpha / \alpha^2) (\sin^2 N\beta / \sin^2 \beta)} \quad (1.57)$$

Here, in equation (1.57) the factor $\sin^2 \alpha / \alpha^2$ is responsible for the diffraction effect due to single slit and the additional factor $\sin^2 N\beta / \sin^2 \beta$ is responsible for the interference effects of the diffracted rays from N -slits.

Primary Maxima or Principal Maxima

The resultant intensity I will be maximum when denominator of equation (1.57) is zero, i.e., when $\sin \beta = 0$

$$\beta = \pm n\pi \text{ where } n = 0, 1, 2, 3 \dots$$

Now

$$\beta = \pm n\pi$$

Hence, for n th order primary maxima condition is $\beta = \pm n\pi$, i.e.

$$\frac{\pi(a+b)\sin\theta}{\lambda} = \pm n\pi$$

$$(a+b)\sin\theta_n = \pm n\lambda \quad (1.58)$$

Where $n = 0$ corresponds to central maximum M_0 and $n = 1, 2, 3, \dots$ corresponds to 1st order, 2nd order, primary maxima M_1, M_2 . The \pm sign in eqn. (1.58) shows that there are two principle maxima of the same order lying on the both sides of the central maximum. The intensity goes on decreasing as the order of maxima increases. Resultant intensity distribution is shown in Fig. (1.26).

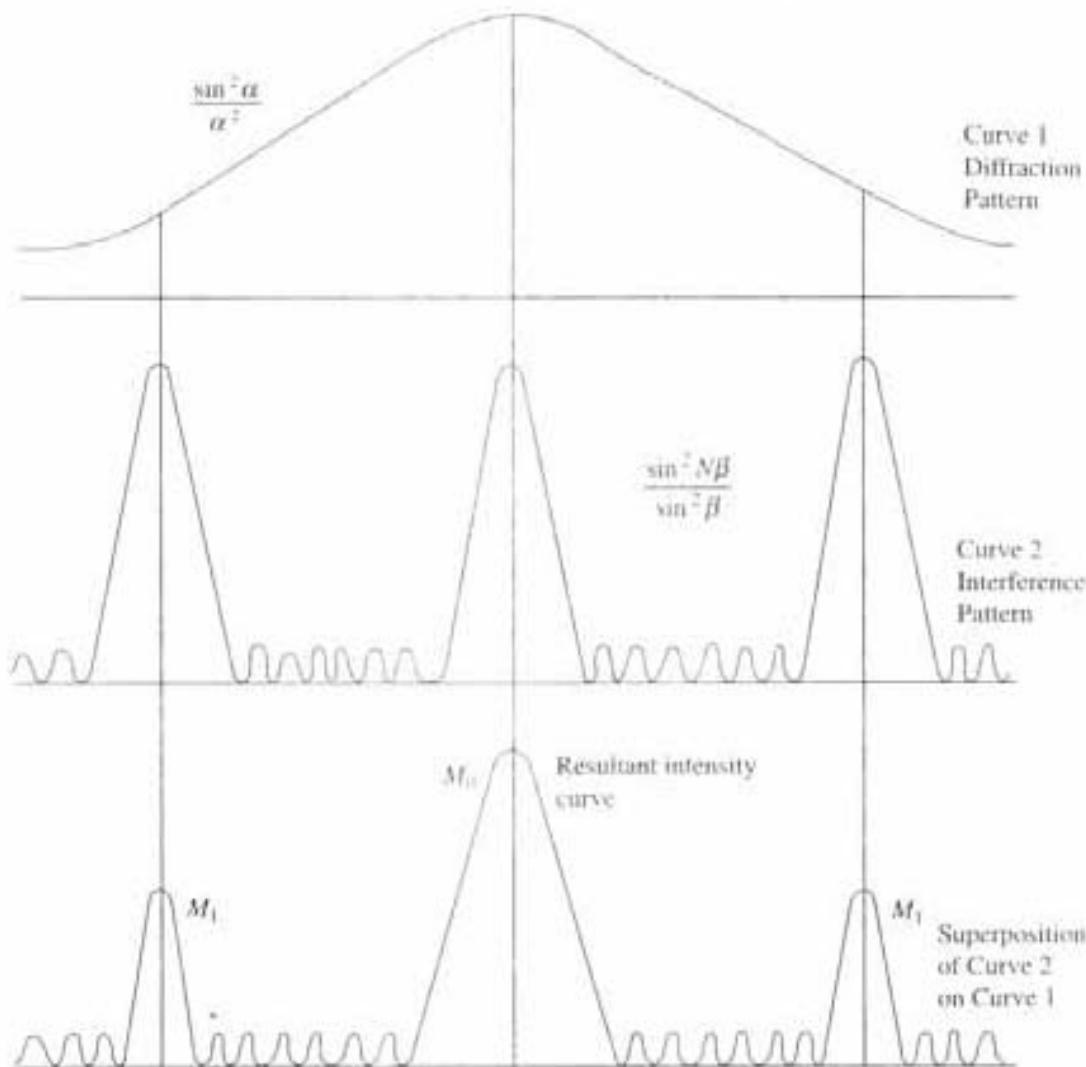


Fig. 1.26 Resultant intensity distribution of grating spectra

Subsidiary or Secondary Minima and Maxima

Condition of primary maxima from eqn. (1.57) is $\beta = \pm n\pi$. But when $\beta = \pm n\pi$, then numerator of Eqn. (1.57) is also zero, because $\sin N\beta = 0$. Then the resultant intensity I will be minimum. So the condition for minima is

$$N\beta = 0, \pi, 2\pi, \dots, \pm K\pi$$

or

$$N\beta = \pm K\pi$$

But then the quotient (equation 1.57) becomes 0/0 which is indeterminate. Therefore $N\beta = \pm K\pi$ gives the condition for minimum intensity for all values of K , except 0, N , $2N, \dots, nN$. Because for those values of K , $\sin \beta$ will be zero, which gives the primary maxima. Thus K can get the value 1, 2, ..., $(N - 1)$ and there are $(N - 1)$ subsidiary minima and $(N - 2)$ subsidiary maxima in between central (M_0) and 1st order (M_1) primary maxima, and in between any two primary maxima.

As N increases, the intensity of subsidiary maxima decreases relative to primary maxima and becomes negligible when N becomes very large.

Intensity of the Primary Maxima

For $\beta = 0, \pi, 2\pi, \dots, n\pi$ etc., we get from equation (1.57)

$$\frac{\sin N\beta}{\sin \beta} = \frac{0}{0}$$

which is indeterminate.

Now by applying La-Hospital's rule, the limiting value of this can be calculated as follows:

$$\lim_{\beta \rightarrow \pm n\pi} \frac{\sin N\beta}{\sin \beta} = \lim_{\beta \rightarrow \pm n\pi} \frac{d/d\beta(\sin N\beta)}{d/d\beta(\sin \beta)} = \lim_{\beta \rightarrow \pm n\pi} \frac{N \cos N\beta}{\cos \beta} = \pm N$$

Hence

$$\lim_{\beta \rightarrow \pm n\pi} \left\{ \frac{\sin^2 N\beta}{\sin^2 \beta} \right\} = N^2$$

Thus, the resultant intensity of the primary maxima from equation (1.57) is

$$I = I_0 N^2 (\sin^2 \alpha / \alpha^2) \quad (1.59)$$

These maxima (M_1, M_2, \dots) on both sides of the central maximum (M_0) are quite intense. But their intensity goes on decreasing as the order number increases, due to the superposition of interference pattern, shown clearly in Fig. (1.26).

1.27 Absent Spectra

The condition for interference maxima from equation (1.58) is

$$(a + b) \sin \theta = n\lambda$$

Similarly, condition for diffraction minima from equation (1.50) is

$$a \sin \theta = p\lambda$$

where n and p are integers, $n = 0, 1, 2, \dots$ and $p = 1, 2, 3, \dots$. If the value of a and b are such that both the equations are satisfied simultaneously for some particular direction θ , then although the diffracted rays from all the slits of the grating reinforce each other, but for each slit, rays from the upper and lower halves destroy each other and hence the resultant intensity is zero, which means that the spectrum of that particular order will be absent in that direction.

By taking the ratio of above two equations we get

$$\frac{(a+b) \sin \theta}{a \sin \theta} = \frac{n}{p}$$

i.e.,

$$\frac{a+b}{a} = \left(\frac{n}{p} \right) \quad (1.60)$$

Case 1. If $a = b$, i.e., when the width of the slit is equal to opacity, then $n = 2p$ ($p = 1, 2, 3, \dots \neq 0$) so that the 2nd, 4th, 6th etc. orders of the primary maxima will be missing.

Case 2. If $b = 0$, i.e. $n = p$ then except central maximum ($n = 0$) all other orders of primary maxima (M_1, M_2, \dots) will be zero. The resultant diffraction pattern will be similar to that due to single slit.

Case 3. When $b = 2a$, i.e. $n = 3p$ so that the 3rd, 6th 9th orders of primary maximum will be missing. This is shown in Fig. (1.24).

1.28 Maximum Number of Order of Spectra in a Grating

The condition for n th order principal maxima for grating is

$$(a+b) \sin \theta_n = n\lambda$$

θ_n can get maximum value 90° , so the maximum possible order is

$$n_{\max} = \frac{(a+b) \sin 90^\circ}{\lambda} = \frac{(a+b)}{\lambda} \quad (1.61)$$

If grating element $(a+b) < 2\lambda$

$$n_{\max} < 2\lambda/\lambda < 2$$

i.e., only 1st order is possible to see.

For a grating with 15,000 lines/inch

$$(a+b) = 2.54/15,000 \text{ cm} \quad (\because 1 \text{ inch} = 2.54 \text{ cm})$$

$$(a+b) = 1.6 \times 10^{-4} \text{ cm}$$

For sodium light, $\lambda = 5893 \text{ \AA} = 0.5893 \times 10^{-4} \text{ cm}$.

$$\text{So } n_{\max} = \frac{1.6 \times 10^{-4}}{0.5893 \times 10^{-4}} = 2.7$$

Thus, 2nd order is the highest order possible to see.

1.29 Dispersive Power of the Grating

The change in the angle of diffraction per unit of wave length of light used is defined as the *dispersive power* of the grating. The condition for n th order principal maxima is

$$(a+b) \sin \theta = n\lambda$$

From this equation it is clear that for a particular value of grating element $(a+b)$, λ is directly proportional to θ . Now differentiating above condition w.r.t. λ we get

$$(a + b) \cos \theta d\theta = n d\lambda$$

or

$$\text{Dispersive power} = \frac{d\theta}{d\lambda} = \frac{n}{(a + b) \cos \theta} = \frac{nN}{\cos \theta}$$

(1.62)

where

$$N = \frac{1}{(a + b)} = \frac{\text{Number of lines}}{\text{Unit length of grating}}$$

For small values of θ , $\cos \theta = 1$. So the angular dispersion $d\theta$ is directly proportional to $d\lambda$. Such a spectrum is called *normal spectrum*. Thus, for eqn. (1.62), it is clear that the dispersive power $d\theta/d\lambda$ depends directly on: (i) n -the order of the spectrum, and (ii) N -the number of lines/unit length of the grating.

Here dispersive power $d\theta/d\lambda$ gives the angular dispersion per unit change of wave length. But when the diffracted beam is recorded on photographic plate, then we need linear dispersion, not angular. If f is the focal length of the lens used to record diffraction pattern on the photographic plate, and dl is the separation of two lines having wave length λ and $\lambda + d\lambda$ then $d\theta = dl/f$.

Now, the linear dispersion per unit wave length from equation (1.62) is

$$\frac{dl}{d\lambda} = \frac{fd\theta}{d\lambda} = \frac{nf}{(a + b) \cos \theta} = \frac{nfN}{\cos \theta}$$
(1.63)

Hence the linear separation dl between wave length λ and $\lambda + d\lambda$ in the n th order is

$$dl = \frac{nf}{(a + b) \cos \theta} d\lambda$$

(1.64)

1.30 Spectrum Formation and Overlapping of Spectrum Lines

From equation of n th order primary maxima

$$(a + b) \sin \theta_n = n\lambda$$

It is clear that for a particular order n , the light of different wave lengths will be diffracted in different directions. For light of longer wave length (red) will be diffracted more than shorter wave length (violet). Thus, if we use white light instead of monochromatic light (say Na), then each wave length will give its own principal maximum in each order. So we get spectrum for each order primary maxima, red being the outermost and violet being the innermost position. But, for central maximum for which $n = 0$, $\theta = 0$ for all wavelengths which gives white central maximum M_0 .

Now if the angle of diffraction θ in equation

$$(a + b) \sin \theta = n\lambda$$

be the same for wavelength λ_1 in first order ($n = 1$), wavelength λ_2 in 2nd order ($n = 2$), and wavelength λ_3 in 3rd order ($n = 3$). Then

$$(a + b) \sin \theta = 1 \cdot \lambda_1 = 2\lambda_2 = 3\lambda_3 = \dots$$

Hence, the first order spectrum of wave length λ_1 will be overlapping with 2nd order spectrum of wave length λ_2 and 3rd order spectrum of wave length λ_3 in the same direction θ , i.e., at the same position of the screen.

1.31 Determination of Unknown Wave Length by Diffraction Grating

Diffraction grating is very much used in the laboratory for the accurate measurement of wave length of light given by a source by using spectrometer. Experimental arrangement is as shown in Fig. (1.27).

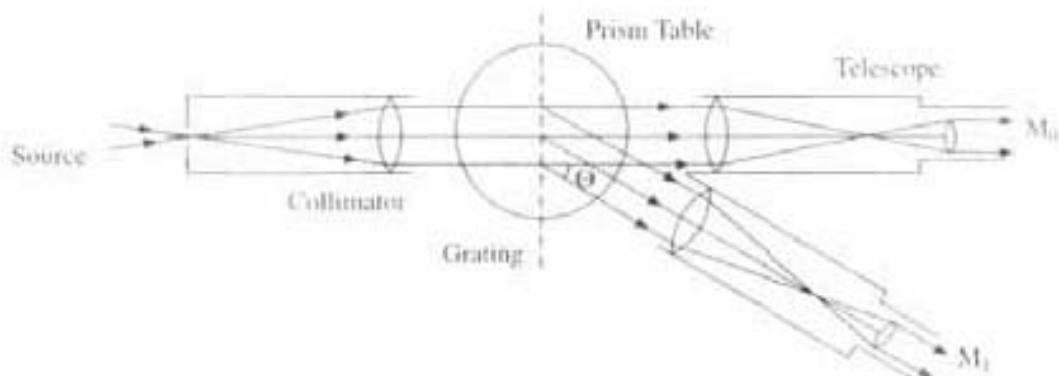


Fig. 1.27 Experimental set up for determination of unknown wave diffraction length by grating

Initial adjustments of the spectrometer are made and it is adjusted for parallel rays by Schuster's method. The slit of the collimator is illuminated by monochromatic light, generally Na-light and telescope is adjusted in such a way that the image of the slit is obtained at the position of the vertical cross-wire in the field of view of the telescope. Now the axes of collimator and telescope are in the same line. The telescope is rotated by 90° and clamped so that collimator and the telescope are now at perpendicular position. Now place the transmission grating at the centre prism table such that grating surface is perpendicular to prism table. Then the Prism table is rotated in such a way that the image of the collimator slit is reflected from the grating surface and obtained at the centre of the field of view of the telescope. This means that the parallel light from the collimator falls at an angle 45° on the grating surface and reflected rays go to telescope because collimator and telescope are initially kept at 90° apart. Now the position of the prism table is noted and adding 45° to this reading, the prism table is again rotated through 45° to the new position so that at that particular position grating surface is normal to the incident beam.

The wave length of the Na-light can be determined by measuring the angle of diffraction of the 1st order (M_1) i.e., θ_1 and 2nd order maxima (M_2) i.e., θ_2 from the central maximum (M_0) on both sides. Then

$$(a + b) \sin \theta_n = n\lambda$$

For 1st order $\theta_n = \theta_1$, $n = 1$ and for 2nd order $\theta_n = \theta_2$, $n = 2$

$$(a + b) = \text{Grating element} = 1/N$$

N = Number of lines/Unit length of the grating.

Now we can calculate λ , the wave length of Na-light or any monochromatic light.

If the light is not monochromatic, say white, then for each other primary maxima spectrum will come, except for central. In that case in order to find out wave length of any particular colour, the diffracting angles are noted for the 1st and 2nd order and using the equations given above, the wave length of that particular colour can be calculated. Overlapping spectral orders can be avoided by using suitable colour filters.

As the diffracting angles are large for usual grating having 15,000 lines/inch, so they can be measured accurately. The number of lines/inch (or cm) is given by the manufacturers, so $(a + b)$ can be calculated. As the method does not involve any small measurements so an accurate value of wave length λ can be obtained for a source of light whose wave length is not known.

PROBLEMS

- A parallel beam of monochromatic light is allowed to be incident normally on a plane transmission grating having 1250 lines/cm and a 2nd order spectral line is observed to be deviated through 30° . Calculate the wave length of the spectral line.

Solution

$$(a + b) \sin \theta_2 = n\lambda$$

$$(a + b) = 1/1250, n = 2, \theta_2 = 30^\circ$$

$$\text{Therefore, } \lambda = \sin 30^\circ / 1250 \times 2 = 2 \times 10^{-4} \text{ cm.}$$

- How many orders will be visible if the wave length of the incident radiation is 5893 \AA and the number of lines on the grating is 2540/inch?

Solution

$$(a + b) \sin \theta_n = n\lambda$$

$$n_{\max} = (a + b)/\lambda \quad (\because \sin \theta_n = 1)$$

$$(a + b) = \frac{2.54}{2540} = 10^{-4} \text{ cm}, \lambda = 5893 \times 10^{-8}$$

$$n_{\max} = 10^{-3}/5893 \times 10^{-8} = 16.96$$

Hence maximum order will be 16.

- The separation of the two Na-lines (mean $\lambda = 5893 \text{ \AA}$) in the 2nd order spectrum of a transmission grating containing 5000 lines/cm is 2.5 minutes for normal incidence.
(i) What is the difference in wave length of the two yellow lines? (ii) What will be the linear separation if they are recorded by a camera of focal length 30 cm?

Solution

- (i) Dispersive power = $d\theta/d\lambda = n/(a + b) \cos \theta$

$$d\lambda = \frac{(a + b) \cos \theta}{n} d\theta$$

$$\cos \theta = \sqrt{1 - \sin^2 \theta} = \sqrt{1 - [n\lambda/(a + b)]^2} \quad \left[\because \sin \theta = \frac{n\lambda}{(a + b)} \right]$$

Now, $(a + b) = 1/5000$, $n = 2$, $\lambda = 5893 \times 10^{-8}$ cm.
 $d\theta = 2.5 \text{ min} = (3.14 \times 2.5)/(180 \times 60) = 7.27 \times 10^{-4}$ radian.

$$d\lambda = \frac{\sqrt{1 - (2 \times 5893 \times 10^{-8} \times 5000)^2}}{5000 \times 2} \times 7.27 \times 10^{-4}$$

$$d\lambda = \frac{\sqrt{1 - 0.3473}}{10,000} \times 7.27 \times 10^{-4} = 5.87 \times 10^{-8} \text{ cm} = 6 \text{ \AA}$$

(i) Hence difference of wave length between two yellow lines of sodium is 6 \AA
(ii) Linear dispersion = $dl/d\lambda = f d\theta/d\lambda$

$$\text{Linear separation} = dl = f d\theta = 30 \times 7.27 \times 10^{-4} = 2.18 \times 10^{-2} \text{ cm}$$

4. What particular spectra would be absent if the width of the opacities be double that of the transparencies in a grating?

Solution

$(a + b) \sin \theta = n\lambda$. . . Interference maxima

$a \sin \theta = p\lambda$. . . Diffraction minima

When both conditions are satisfied for a particular value of θ for particular value of a and b such that $b = 2a$, then some spectra will be absent

$$\text{Now, } (a + b)/a = n/p$$

$$\text{So, } 3a/a = n/p \text{ (as } b = 2a)$$

$$\text{i.e., } n = 3p \text{ (where } p = 1, 2, \dots)$$

So, 3rd, 6th, 9th etc. spectra will be absent.

5. A simple grating spectrometer, the grating of which has 6000 lines/cm, has its telescope replaced by a camera whose lens is of 30 cm focal length. Find the linear separation for spectra lines of wave lengths 5770 \AA and 5460 \AA in 2nd order. (*Hint*: Use equation (1.54))
6. A diffraction grating used at normal incidence gives a line (5400 \AA) in a certain order superimposed on the violet line (4050 \AA) of the next higher order. If the angle of the diffraction is 30° , how many lines/cm are there in the grating?

Solution

$$(a + b) \sin \theta_n = n\lambda$$

Now, n th order of λ_1 is superimposed on $(n + 1)$ th order of λ_2 for a particular value of $\theta = 30^\circ$. So,

$$(a + b) \sin 30^\circ = n \times 5400 \times 10^{-8} = (n + 1) \times 4050 \times 10^{-8}$$

$$n\lambda_1 = (n + 1)\lambda_2$$

$$n = \lambda_2/(\lambda_1 - \lambda_2) = 4050/(5400 - 4050) = 3$$

$$1/(a + b) = N = \text{number of lines/cm}$$

$$N = 1/(a + b) = \sin \theta/n\lambda = \sin 30^\circ/(3 \times 5400 \times 10^{-8})$$

$$N = 3086 \text{ lines/cm.}$$

QUESTIONS

1. If the light of wave length λ is incident on a grating of spacing d at an angle ψ and if θ is the angle of diffraction, then show that the condition of diffraction maxima is

$$d(\sin \psi + \sin \theta) = n\lambda$$

2. Discuss the phenomenon of Fraunhofer diffraction at a single slit and obtain the condition for first minimum in the diffraction pattern. Also obtain the linear distance of first minima from central maximum.
3. Give construction of a plane diffraction grating for N number of slits. Obtain the condition for first secondary minima after the central maximum. Also write the general number of secondary minima and number of secondary maxima in between two primary maxima.
4. What is the fundamental criteria between wave length of light used and obstacle size in order to get diffraction?
5. Instead of Na-light if white light is used for diffraction experiment, what will be the major change in central maximum and primary maxima?
(Hints: In sunlight central maxima will be white because $(a + b) \sin \theta_n = n\lambda$. $\theta = 0$, $n = 0$, all λ 's will meet at central maximum, but θ_n will be different for different λ , so all primary maxima will be a colour band.)
6. Explain why (i) increasing the number of slits ' N ' in a diffraction grating, (ii) decreasing wave length, (iii) increasing grating spacing ($a + b$); increases the intensity of the maxima of the grating spectra?
(Hints: Explain from equation 1.57).
7. What is zone plate and how is it made? Explain also how a zone plate acts like a convergent lens having multiple foci. Derive an expression for its focal length.
8. Explain the half-period zone in relation to a plane wave front. Show that the amplitude due to a complete wave front at a point is half of what would be caused by the 1st zone.
9. What is the radius of the 1st zone in a zone plate of focal length 20 cm for light of wave length 5000 Å?

1C. POLARISATION OF LIGHT

1.32 Polarisation of Wave

When a periodic disturbance travels through a material medium, like in water or in a stretched string, particles in its path are set into vibration leading to a wave propagation through that medium. There are two types of vibration are possible which are called modes of vibration.

First is the *Longitudinal mode of vibration*, in which particles of the medium vibrating along the path of the wave propagation. Second is the *Transverse mode of vibration*, in which particles of the medium vibrating in any path which is perpendicular to the path of the wave propagation.

In longitudinal mode of vibration there is only one possible direction of vibration of the particles is possible that is along the direction of wave propagation.

But in case of transverse mode of vibration, there are infinite directions of vibrations are possible for the particles which are perpendicular to the direction of wave propagation.

Now if a restriction is imposed on the different possible directions of vibration of the particles perpendicular to the path of a wave propagation, so that their vibrations are confined only to a single plane, i.e., only along one direction which is perpendicular to the direction of wave propagation, then that outgoing wave is called *polarised wave*.

For a longitudinal wave there is only one possible direction of vibration of the particle of the medium is possible that is along the direction of wave propagation, i.e., the wave is already polarised along the direction of wave propagation, so no other polarisation is possible. But for transverse wave there are infinite possible directions, all are perpendicular to the direction of wave propagation, so transverse wave can be polarised in an infinite way.

Now when the vibration of the particles of the medium perpendicular to the path of the transverse wave propagation occurs in all possible directions with equal favour, then the outgoing wave is called *unpolarised*. If the vibration of the particles occur in only one plane or direction which is perpendicular to the wave propagation then the outgoing wave is called *plane polarised wave or linearly polarised*. When the vibration of the particles occur in more than one perpendicular plane and if it will occur in any perpendicular plane or direction with more favour than another, then that outgoing wave is called *partially polarised*.

1.33 Polarisation of Light Wave

Interference and diffraction phenomena proved the wave nature of light, and after the well-known Maxwell's electromagnetic theory of light, light wave is well known as transverse electromagnetic wave.

For a plane electromagnetic wave there is an *E*-field vector (\vec{E}) and *M*-field vector (\vec{B}) which are perpendicular to each other, as well as both of them perpendicular to the direction of propagation (β = propagation vector) of the light wave. That is why the light wave called as Transverse electromagnetic wave. For a linearly polarised light wave propagating along *Z*-axis, the *E*-field and *M*-field vectors are generally written as

$$\vec{E}_x(z, t) = E_0 \cos(\omega t - \beta z), \quad \vec{E}_y = \vec{E}_z = 0 \quad (1.65)$$

and

$$\vec{B}_y(z, t) = B_0 \cos(\omega t - \beta z), \quad \vec{B}_x = \vec{B}_z = 0$$

where

$$\beta = \text{propagation vector} = \omega/v = \omega/\sqrt{\mu_0 \epsilon_0}$$

$$\omega = \text{Angular frequency} = 2\pi f$$

$v = c = 1/\sqrt{\mu_0 \epsilon_0}$ = velocity of light wave in free space (air) = 3×10^8 m/sec,
 $\mu_0, \epsilon_0 \Rightarrow$ permeability and permitivity of free space (air).

In general, an ordinary light beams coming from a lamp or from sunlight are random in phase and unpolarised, i.e., electric vector (\vec{E}) and magnetic vector (\vec{B}) though they are perpendicular to each other (in a plane, transverse to the direction of propagation), but keeps on changing their directions in a random

manner as shown in (Fig. 1.28a). When such a beam incident on a polarising device (polariser), the emergent beam of light will become linearly polarised with its electric vector (\vec{E}) and magnetic vector (\vec{B}) oscillating in particular direction as shown in the (Fig. 1.28b). Different types of polarising devices are possible which will impose the restriction, so that it will allow to pass only one component of the electric vector and magnetic vector. The direction of electric vector and magnetic vector of the emergent light beam depends on the orientation of the polariser.



Fig. 1.28 (a) Unpolarised beam moving along Z-axis, when E -vector or M -vector continuously changing its direction in xy plane. (b) Linearly polarised light moving along z -axis, when E - or M -vector which are orthogonal to each other, oscillates only along a particular direction

1.34 Production of Polarised Light Wave by Different Polarising Devices

(A) Polarisation by Double Refraction or Birefringence^{3, 4, 5}

Double Refraction or Birefringence

When a beam of monochromatic light is incident on a plane transparent medium like glass, it gives rise to a single refracted ray, as the refractive index of glass is same everywhere within the glass material. That type of materials are called singly refracting material or the *isotropic material*.

But in many crystalline optical material such as calcite (CaCO_3), quartz (SiO_2), KDP (Potassium di-hydrogen phosphate- KH_2PO_4) when light incident on them, they give rise to two refracted rays. One is called ordinary ray (O -ray), which follows the Snell's law of refraction and other is called extraordinary ray (E -ray). The velocity of (O -ray) is same in all directions, so the refractive index. But the velocity of (E -ray) is different in a different direction. So the refractive index also will vary with direction for (E -ray). Those crystalline materials are called *an-isotropic material*. This an-isotropy arises due to different types of arrangement of atoms in different direction within the material. So the speed of light as well as refractive index in such materials depend on the direction. *This phenomena is called double refraction or Birefringence and those type of crystals are called doubly refracting or Birefringent crystals.*

Those an-isotropic crystals crystallize in different forms, but generally they can be reduced into a rhombohedron as shown in the (Fig. 1.29a). At two opposite corners A & H of the rhombohedron all the angles are obtuse. Now if a line drawn through A & H , for a crystal whose three edges are equal, then that line will make

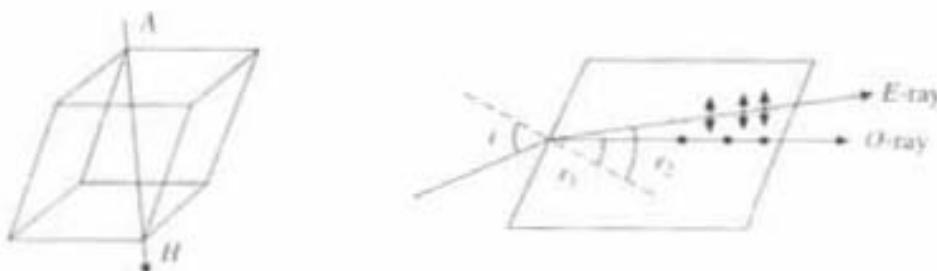


Fig. 1.29 (a) Doubly refracting crystal Calcite; (b) Split up of O -ray & E -ray in Calcite

equal angle with each of the three edges. That line or that direction is called *optic axis*. Any line parallel to AH is also optic axis, i.e., optic axis is only a fixed direction inside the doubly refracting crystal not a fixed straight line, like AH.

Uni-axial and Bi-axial crystal

Crystals which are having one direction of optic axis, they are known as uni-axial crystal (e.g., calcite, quartz are uni-axial).

Crystals which are having two directions of optic axis they are known as bi-axial crystals (e.g., borax, mica are bi-axial).

Following are the different important features regarding double refraction or Birefringence:

1. If the ray of light incident on the doubly refracting crystal along optic axis or parallel to the optic axis then it will not break into two rays, i.e., double refraction is absent along optic axis.
2. For the other directions as shown in the (Fig. 1.29b) incident light will break into (E -ray and O -ray), which will travel with different speeds, both are plane polarised and have mutually orthogonal planes of polarisation. The dots (perpendicular component) represent the O -ray and line (parallel component) represent the E -ray in (Fig. 1.29b). So when an unpolarised light incident normally on such doubly refracting crystal it resolved into two independent waves (E -ray and O -ray) which travel with different velocities.
3. Both (E -ray and O -ray) are linearly or plane polarised. The vibration corresponding to the O -ray is always perpendicular to the principal section of the crystal, whereas that for E -ray is parallel to the principal section of the crystal. *Principal section is the plane in a crystal containing direction of propagation and the optic axis.*
4. If the velocity of (E -ray) is everywhere more than (O -ray), i.e., $V_e > V_o$, then that type of crystal is called as *Negative crystal* (e.g., calcite, whose R.I. for O -ray is more than R.I. for E -ray i.e., $n_o > n_e$). Whereas if the velocity of (O -ray) is everywhere more than (E -ray) i.e., $V_o > V_e$, then that type of crystal is called as *Positive crystal* (e.g., quartz, whose R.I. for E -ray is more than R.I. for O -ray i.e. $n_e > n_o$).
5. When the light beam incident perpendicular to the optic axis of a doubly refracting crystal, then also E -ray and O -ray will not split up, they will move in the same direction but with different speeds.
6. For negative crystal like calcite, velocity of E -ray is maximum at right

angle to the optic axis. Whereas for positive crystal like Quartz velocity of *E*-ray is minimum right angle to the optic axis.

This property of double refraction or birefringence can be used to get polarised light. As we have seen that when light incident on an anisotropic crystal then incident light will split up into *E*-ray and *O*-ray, when both are plane polarised and their plane of vibrations are orthogonal to each other. Now a selective absorption of one is a simple method for eliminating one of them, which is known as dichroism. Crystal like tourmaline has different co-efficient of absorption for the two linearly polarised light, into which incident beam split up. Out of two one of the beam gets absorbed quickly and the other component passes through without much attenuation. Thus if unpolarised beam incident on tourmaline crystal, the emergent beam will be linearly polarised.

B. Polarisation by Reflection

Brewster performed a series of experiments to study the polarisation by reflection at surface of different medium. He found that an *ordinary light* become completely polarised when it gets reflected from a transparent medium, at a particular incident angle *i*, which is known as angle of polarisation or Brewster's angle.

Brewster's law states that when a ray of light reflected from a surface is polarised, then the tangent of the polarising angle is equal to the refractive index of the material of medium.

So according to the (Fig. 1.30), Brewster's law states that

$$\tan i = \mu \quad (1.66)$$

Where *i* = polarising angle of incidence and μ = refractive index of the medium. Now from the Fig. (1.30)

$$\tan i = \mu = \sin i / \sin r_1 = \sin i / \cos i$$

$$\text{i.e.} \quad \sin r_1 = \cos i = \sin (90 - i)$$

$$\text{i.e.} \quad r_1 = 90 - i, \text{ but by law of reflection } i = r. \text{ So } r + r_1 = 90^\circ$$

Hence the angle between reflected and refracted beam is 90° , as is shown in Fig. 1.30.

So the consequence of the Brewster's law is that when the light is incident on the surface of the crystal at a polarising angle, the reflected and refracted rays will be at right angle to each other.

Thus if an unpolarised beam incident on the crystal at the polarising angle, then reflected beam will be plane polarised whose plane of vibration is perpendicular to the plane of incidence as shown in (Fig. 1.30). The transmitted beam will be partially polarised as it is having the component richer in vibration in the plane of incidence.

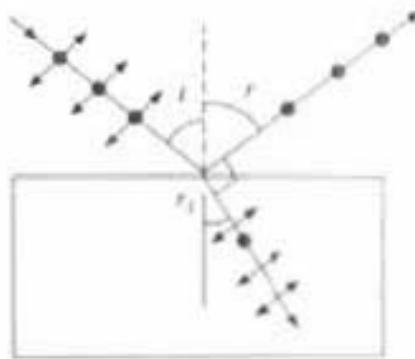


Fig. 1.30 Brewsters Law

C. Polarisation by Total Internal Reflection

A Nicol prism is a device which acts as a polariser, made from double refracting

calcite crystal to produce plane polarised light by total internal reflection from a layer of Canada Balsam. As the *E*-ray and *O*-ray split up in the double refracting crystal like calcite, which are moving with different velocity due to their different refractive index. Now if a layer of a material is sandwiched in between two halves of the properly cut calcite crystal, whose R.I. (1.55) lies in between the R.I. of *E*-ray (1.486) and R.I. of *O*-ray (1.658), then for the component (*O*-ray) the incidence on that layer will be at rarer medium and for the other component (*E*-ray) it will be at denser medium. This principle is used in Nicol prism, which consists of calcite crystal cut in such a way that for a particular ray, sandwich material will be a rarer medium and the angle of incidence for that ray will be greater than the critical angle as shown in the (Fig. 1.31a). Then this particular component of ray will be eliminated by total internal reflection (the phenomena of total internal reflection explained clearly in Fiber Optic Section, Chapter 5), whereas the other component of ray will be passes away. Fig. (1.31a) shows the properly cut calcite crystal in which a layer of Canada Balsam is introduced in such a way, that *O*-ray undergoes total internal reflection and *E*-ray passes away. Thus outgoing ray become plane polarised.

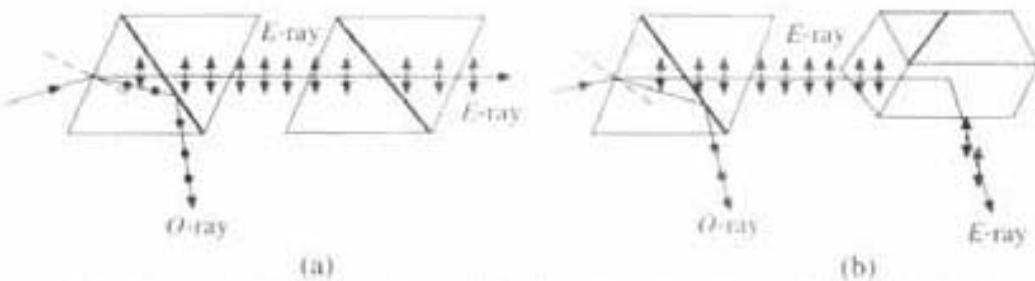


Fig. 1.31(a) Two Nicols are parallel to each other; (b) Crossed Nicols

1.35 Analyser for Polarised Wave

A. Crossed Nicol Prism

As shown in the (Fig. 1.31a, 1.31b), when two Nicols placed adjacent to each other, then one of them acts as a polarizer and other one act as an analiser. When the position of two Nicols are parallel to each other only *E*-ray passes through the both the Nicols (Fig. 1.31a). But when the 2nd prism is gradually rotated the intensity of the out going *E*-ray will gradually decrease and when the two Nicols are crossed (Fig. 1.31b), then no light will come out from the 2nd prism. Because when the *E*-ray enters the 2nd prism at this crossed position, *E*-ray will act as an *O*-ray and is totally internally reflected by the Canada Balsam, so no light comes out of 2nd prism. Therefore the 1st prism produces plane polarised light (act as an-Polariser) and the 2nd prism detects its plane of polarisation (act as an-Analyser). The combination of 1st and 2nd prism is called as *polariscope*.

B. Malus Law

Let us consider a polariser which has pass axis parallel to *Y*-axis (pass axis means, it will only allow to pass the plane polarised beam which are parallel to *Y*-axis). When an unpolarised beam incident on the polariser then the transmitted wave will be plane polarised and plane of vibration of that plane polarised

transmitted beam will be parallel to Y-axis. Now if the polariser say it can be a Nicol prism or a polaroid rotated about Z-axis, then for unpolarised incident beam there will not be any change of intensity of the emergent beam. But if a partially polarised beam incident on this polariser, then as the polariser rotated about Z-axis then the intensity of emergent beam will change but it will never be zero.

Now let us consider a plane polarised beam with electric vector of amplitude (E_0) incident on the polariser whose plane of vibration (say AB) makes an angle Θ with the Y-axis, as shown in the (Fig. 1.32). Then the amplitude of the emergent beam from the Polaroid will be $E_0 \cos \Theta$ and the intensity of the emerging beam will vary according to the following equation

$$I = I_0 \cos^2 \Theta \quad (1.67)$$

where I_0 is the intensity of the incident beam. Above equation (1.67) represent the Malus law. Thus if a plane polarised beam incident on a polaroid, and if the polaroid rotates about Z-axis, then the intensity of the emergent beam will vary according to the Malus law. Say the polaroid is rotated in clockwise direction. Then the intensity of the emergent beam will increase, and it will be maximum when the pass axis is parallel to AB. Further rotation will decrease the intensity and it will be zero when the pass axis is perpendicular to AB. If it will further rotate then again it will be maximum at the position AC and again it will be zero at a perpendicular position before it reaches to its original position.

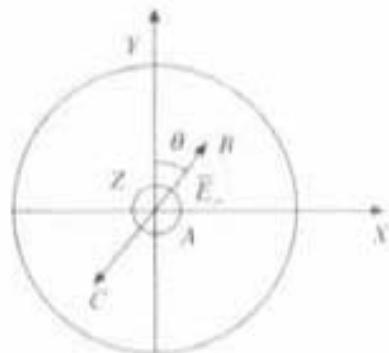


Fig. 1.32 Incidence of plane polarised light on a analyser of pass axis along Y

1.36 Superposition of Two Plane Polarised Waves and Production of Circularly and Elliptically Polarised Light

Let us consider a monochromatic beam incident on a Nicol prism after passing through the Nicol it become plane polarised. Then allow that plane polarised beam to incident normally on a uniaxial doubly refracting crystal (calcite or quartz) whose incident face has been cut, parallel to its optic axis, i.e., the plane polarised light incident normally to the optic axis of the double refracting crystal. The plane of vibration of the plane polarised beam incident on the doubly refracting crystal are as shown in the (Fig. 1.33). In that situation as we have discussed in the section (1.34A), the incident beam will be split

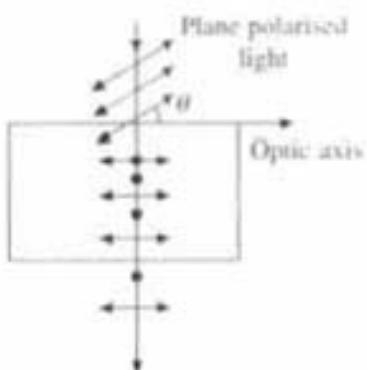


Fig. 1.33 Incidence of plane polarised beam on Nicol prism

up into E-ray and O-ray, when both are plane polarised and their plane of vibrations are orthogonal to each other as well as perpendicular to the optic axis. Now according to the feature number-Five in section 1.34A E-ray and O-ray will move in the same direction but with different speeds. So as they proceed through the doubly refracting crystal the path difference (Δ) as well as phase difference $(\delta = \frac{2\pi}{\lambda} \Delta)$ will gradually increase between them.

Now in order to find out the state of polarisation of the resultant field, we have to consider the superposition of those two waves. Let us consider the time variation of the electric-field vector at any arbitrary plane perpendicular to the direction of propagation (say Z-axis) according to the eqn. (1.65) with ($z = 0$). This can be written as say for O-ray

$$E = X = a \cos (\omega t)$$

and for E-ray, $E' = Y = b \cos (\omega t - \delta) = b(\cos \omega t \cos \delta + \sin \omega t \sin \delta)$

i.e.
$$Y = b[\cos \omega t \cos \delta + \sqrt{(1 - \cos^2 \omega t)} \sin \delta]$$

$$Y = (b/a) X \cos \delta + b[\sqrt{1 - (X/a)^2} \sin \delta]$$

or
$$[Y - (b/a) X \cos \delta]^2 = b^2 \sin^2 \delta [1 - (X/a)^2]$$

By rearranging the equation will come as follows

$$\frac{X^2}{a^2} + \frac{Y^2}{b^2} - \frac{2XY}{ab} \cos \delta = \sin^2 \delta \quad (1.68)$$

which is a equation of an asymmetrical ellipse. So for any arbitrary value of phase difference (δ), the emergent light from the doubly refracting crystal will be in general *Elliptically polarised*, i.e., the tip of the resultant electric field vector (E) will move in an elliptical path. In the same way as if we rotated the end of a stretched string on the circumference of a circle (or an ellipse), then each point of the string will move in a circular (or an elliptical) path.

Special Cases

(1) when $\delta = n\pi$ (where $n = 0, 1, 2, 3$ etc.) $\rightarrow \sin \delta = 0, \cos \delta = \pm 1$, then the eqn. (1.68) becomes

$$\left[\frac{X}{a} - (-1)^n \frac{Y}{b} \right]^2 = 0$$

i.e.
$$Y/X = (-1)^n b/a \quad (1.69)$$

which is an equation of a straight line. So when ($\delta = n\pi$), the resultant wave will be plane polarised.

(2) When $\delta = (n + 1/2)\pi$ (for $n = 0, 1, 2$ etc.) $\rightarrow \cos \delta = 0, \sin \delta = 1$. Then the eqn. (1.68) becomes

$$\frac{X^2}{a^2} + \frac{Y^2}{b^2} = 1 \quad (1.70)$$

which represents an equation of a symmetrical ellipse when ($a \neq b$). Then also the resultant wave will be elliptically polarised in a symmetrically way.

But when $\delta = (n + 1/2)\pi$ and $a = b$, then the eqn. (1.68) becomes

$$X^2 + Y^2 = a^2 \quad (1.71)$$

which is an equation of a circle. So the resultant wave will be circularly polarised.

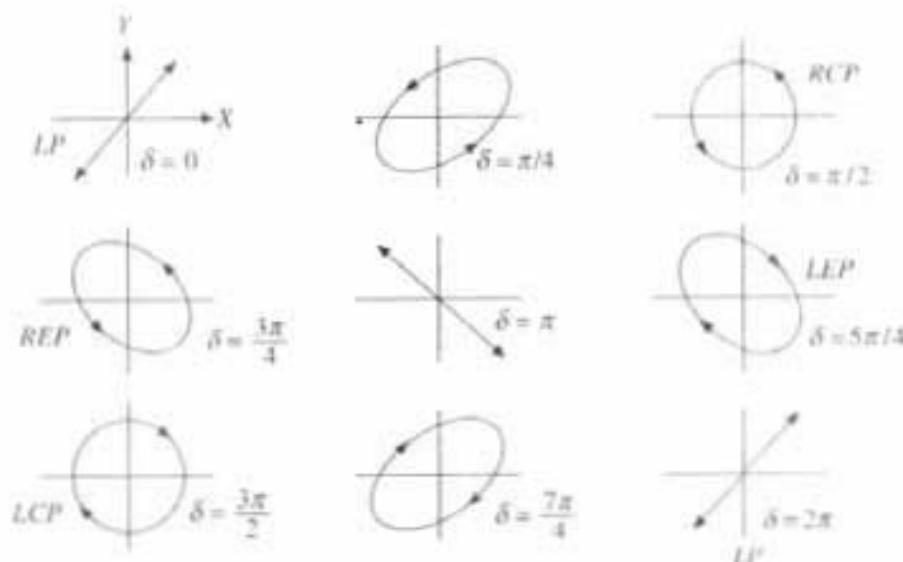


Fig. 1.34 Resultant states of polarizations and directions of polarization for different value of (δ) corresponding to ($a = b$) due to superposition of two plane polarized light

States of polarisation and direction of polarisation for different value of phase difference (δ) corresponding to ($a = b$ in eqns 1.69, 1.70, 1.71) are shown in the (Fig. 1.34). Now according to the convention, when the tip of the electric vector rotates on a circumference of a circle in anticlockwise direction and the propagation is in + Z-direction, then the wave is called right circularly polarised (RCP) light, whereas when it will rotates in a clock wise direction, then it is called as left circularly polarised light (LCP), which are shown in the (Fig. 1.34).

1.37 Quarter Wave Plate and Half Wave Plate

(A) Quarter-Wave Plate

As we have discussed in above section (1.36) that when a plane polarised light with plane of vibration are as shown in the (Fig. 1.33), incident normally on a uniaxial doubly refracting crystal (calcite or quartz), whose incident face has been cut parallel to its optic axis, it splits up into E and O-ray. They will move along same diection but with different speeds. So as the thickness of the doubly refracting crystal increases the path difference as well as phase difference will increase between E-ray and O-ray.

Now for a particular thickness (t) of the doubly refracting crystal when the path difference between E-ray and O-ray becomes $\lambda/4$, then that plate is called as Quarter wave plate.

For Positive crystal Quartz $\rightarrow (n_e - n_o) t = \lambda/4$ or $t = \lambda/4(n_e - n_o)$

For Negative crystal Calcite $\rightarrow (n_o - n_e) t = \lambda/4$ or $t = \lambda/4(n_o - n_e)$

For a positive crystal like Quartz with Sodium light ($\lambda = 5893 \text{ \AA}$), $t \rightarrow 0.000857 \text{ mm}$
And for a positive crystal like Calcite with Sodium light ($\lambda = 5893 \text{ \AA}$), $t \rightarrow 0.0164 \text{ mm}$

(1.72)

So when a plane polarised light with its plane of vibration inclined at any arbitrary angle (Θ) to the optic axis as shown in the Fig. (1.33) is incident on a quarter wave plate, it produce path difference ($\lambda/4$) i.e. phase difference ($\delta = \pi/2$) between E -ray and O -ray. Due to the super position of them (according to Eqn. 1.68) in general emergent light will be Elliptically Polarised. But for ($\Theta = 45^\circ$) (i.e. $a = b$, in the Eqn. 1.70) emergent light will be Circularly Polarised as shown in Fig. (1.34).

(B) Half-Wave Plate

Now for a particular thickness (t) of the doubly refracting crystal when the path difference between E -ray and O -ray becomes $\lambda/2$, then that plate is called as Half wave plate.

For Positive crystal Quartz $\rightarrow (n_e - n_o) t = \lambda/2$ or $t = \lambda/2(n_e - n_o)$ (1.73)

For Negative crystal Calcite $\rightarrow (n_o - n_e) t = \lambda/2$ or $t = \lambda/2(n_o - n_e)$

So when a plane polarised light with its plane of vibration inclined at an angle 45° to the optic axis as shown in the Fig. (1.33) is incident on a half-wave plate, it produce path difference ($\lambda/2$) i.e. phase difference ($\delta = \pi$) between E -ray and O -ray. Due to the super position of them (according to Eqn. 1.69) emergent light will be again Plane Polarised, but the direction of polarisation of the Linear Emergent light will be rotated through 90° , in comparison to the direction of vibration of incident Linear light as shown in the Fig. (1.34).

1.38 Optical Activity⁵

Certain crystals and liquids have the ability to rotate the plane of polarisation of light as it passes through them. That property is called Optical Activity and those mediums are called Optically Active medium. For example when a linearly polarised light beam propagates through an Optically active medium like sugar solution then plane of polarisation rotates. The angle of rotation directly depends on the distance traversed through the solution as well as on the concentration of sugar solution. When linearly polarised light propagating along the optic axis of a quartz crystal which is also optically active, the plane of polarisation gets rotated.

Optical Activity can be explained in the following way. In Optically active medium the velocity of right-circularly polarised light (RCP) is different from the velocity of left circularly polarised light (LCP), because of their different R.I. along right handed direction (n_r) and along left handed direction (n_l). When the particular medium rotates the plane of polarisation in Clock wise direction, then that medium is called Right Handed or Dextro-rotatory (where $n_r > n_l$), whereas when it rotates in anti-clockwise then that is called Left handed or Laevo-rotatory (where $n_r < n_l$).

Now a Plane polarised light which is incident on the optically active medium

can be resolved into two circularly polarised light (RCP and LCP). When these two oppositely directed circularly polarised light propagate through an optically active medium with different speeds, a path difference as well as a phase difference will come between them, and that difference will go on increasing directly with the distance traversed through the medium as well as with the concentration of the solution. So ultimately for a particular thickness of the medium as well as for the particular concentration of the solution, particular value of phase difference will come. This is responsible for a rotation of the plane of incident plane polarised beam, which results from the recombination of (RCP) and (LCP) light when they will come out from the optically active medium as an emergent beam.

1.39 Electro-optic Effect⁵

When an E-field is applied across an optical medium, the distribution of the electrons within the medium will get disturbed. So due to this applied E-field an Isotropic medium may become an an-isotropic. So the polarizability and hence RI of the medium changes an-isotropically due to the applied E-field and it may become an optically active medium, which is known as Electro-Optic Effect.

The result of this electro-optic effect may introduce new optic axes into naturally doubly refracting crystal like (KDP), or to make a naturally isotropic crystal like (GaAS) into doubly refracting crystal by introducing a new optic axis.

This property of electro-optic effect is widely used in optical modulator for fiber optics communication, like POCKEL Modulator and KERR modulator. In the same way *Magneto-Optic effect* results the optical activity in some medium due to the application of M-field. Magneto-optic effect is used in Faraday's modulator for fiber optics communication.

1.40 Photo-elasticity

The stress analysis in girder, beams etc. can be studied by the Optical technique, which is known as *Photo-elasticity*. When a doubly refracting crystal is placed in between the two crossed nicols as discussed section (1.35), the field of view is covered by an interference pattern because of the phase difference between emergent beams. Now some materials like celluloid, glass, bakelite etc. when they are free from strain, do not exhibit the property of double refraction. But those materials exhibit the property of double refraction when they are subjected to some internal stresses.

So when such strained material is placed between two crossed nicols an interference pattern is observed. For stresses at different points, photographs of the interference pattern can be taken. The stress is maximum where the interference pattern will be close. So the stress produced in a beam or in a girder when loaded, can be estimated by this method. Generally a model of the girder made of a transparent plastic material is placed in between the crossed nicols and then loaded. The stresses produced at different points can be analyzed and estimated. Since the stress and strain relation can be estimated by this optical technique, so this is called as *Photo-elasticity*.

QUESTIONS

1. Explain what do you mean by polarisation of light? Distinguish between polarised light and unpolarised light.
2. Describe the construction of Nicol prism and show how it acts as a polariser and as an analyzer?
3. Explain the phenomenon of double refraction in calcite crystal.
4. Give the construction theory of quarter wave plate and half-wave plate.
5. Give a different method of producing plane polarised light.
6. What do you mean by plane polarised light, elliptically and circularly polarised light? Describe how you will produce circularly polarised light and distinguished it from unpolarised light?
7. What will be the state of polarisation of the emergent light when
 - (a) A beam of circularly polarised light is passed through a quarter wave plate
(Ans. Plane polarised)
 - (b) A beam of plane polarised light is passed through a quarter wave plate
(Ans. Elliptically polarised)
 - (c) A beam of elliptically polarised light is passed through a quarter wave plate
(Ans. Plane polarised)
 - (d) A beam of plane polarised light is passed through a quarter wave plate such that the vibrations of the plane polarised light falling on the quarter wave plate make an angle of 45 degree with the optic axis.
(Ans. Circularly polarised)
 - (e) A right-circularly polarised beam incident on a calcite half-wave plate.
(Ans. Left circularly polarised)
8. Explain the principle of rotation of plane of polarisation by an optically active medium.
9. Plane polarised light passes through a quartz plate with optic axis parallel to the faces. Calculate the least thickness for the plate for which emergent beam will be (a) plane polarised, (b) circularly polarised. Given that ($n_p = 1.553$, $n_o = 1.544$, $\lambda = 5500 \text{ \AA}$). [Hints.—(a) Use eqn. (1.73), (b) Use eqn. (1.72).]
10. Explain photo-elasticity and its use.

X-Rays

2.1 Discovery of X-Rays

Rontgen in 1895 discovered X-rays, when he was studying the phenomenon of discharge through rarefied gases. He found that when the pressure in the discharge tube is reduced to 0.001 mm of mercury and an electric discharge is passed between cathode and anode, the glass wall behind the cathode begins to glow with a greenish-yellow colour. During this experiment he also observed that a fluorescent screen placed close to the discharge tube continued to remain fluorescent even if the discharge tube was completely covered with a black thick paper. After performing a series of experiments, Rontgen concluded that when a beam of fast moving electrons strikes a solid target, an invisible, highly penetrating radiation is produced. Because of their unknown nature Rontgen called these radiation as X-rays. Early investigations of the general properties of these invisible rays indicated that they might be of the same nature as visible and ultraviolet light, but of very short wave length (10^{-5} – 10^{-9} cm). The first attempts to reflect, refract and diffract them were, however, unsuccessful.

In 1912, Prof. Laue got an ingenious idea which solved the problem to a very great extent. He suggested that the failure in reflecting, refracting, diffraction, etc., of X-rays was merely due to the requirement of a proper medium. What was good for ordinary and even ultraviolet light was not suitable for X-rays on account of their short wave length. But *internal structure of crystals might form sufficiently refined medium for X-rays, since ordinary X-ray has wave lengths between 10^{-8} and 10^{-9} cm while the average distance between atoms in a solid is between 10^{-8} and 10^{-9} cm. So a crystal can act as a closely packed three dimensional grating, where atoms/molecules will act as opaque spaces and interatomic or intermolecular distance will act as open slits, in three-dimensional space.* Experiments carried out on this idea by Laue himself and others, were crowned with remarkable success. These were followed by other research which proved that X-rays could be effectively reflected, refracted, diffracted, polarised and scattered. Some of these phenomenon can be explained by the classical theory of wave nature of X-rays but some can be explained by the quantum nature of X-rays.

2.2 Production of X-Rays

X-rays are produced when fast moving electrons strike a target material of high

atomic weight like tungsten kept in vacuum. The basic requirements for X-rays production are (1) a source of electrons, (2) effective means for accelerating the electrons and (3) a target of suitable materials of high atomic weight.

There are different types of devices for the production of X-rays, namely (1) gas tubes, (2) Coolidge tube, (3) betatron. These are based on the different important stages of perfection of technique in the production of X-rays.

Here we discuss the Coolidge tube in detail.

2.3 Modern Coolidge Tube

The modern Coolidge tube (Fig. 2.1) designed by Coolidge, is widely used for commercial and medical purposes. It consists of a highly evacuated hard glass bulb containing a cathode and anticathode or anode. Electrons are produced thermionically from a tungsten filamentary Cathode *F* which is heated to incandescence either by a storage battery or by a low voltage A.C. current from step-down transformer *T*₂ (Fig. 2.1). These electrons are focused on the target *T* with the help of a molybdenum cylindrical shield *S* which surrounds *F* kept at a negative potential to the filament. The electrons are accelerated to very high speed (upto 10% of velocity of light) by a D.C. potential difference (of about 50 to 100 kV) applied between *F* and anode (*A*). This high D.C. potential is obtained from a step-up transformer *T*₁ whose output is converted into direct current by full wave rectifier and a suitable filter. The target *T* consists of a copper block in which a piece of tungsten or molybdenum is fitted. The requirements of anode are:

1. High atomic weight—to reduce hard X-rays.
2. High melting point—so that it is not melted due to the bombardment of fast moving electrons which cause lot of heat.
3. High thermal conductivity—to carry away the generated heat.

The target *T* is placed at an angle 45° to the electrons beam. Being a very good conductor of heat, copper helps to conduct heat efficiently to the external cooling fins or water-cooling system. Under the bombardment of the target by so many electrons, most metals will melt. That is why metals like tungsten, platinum and

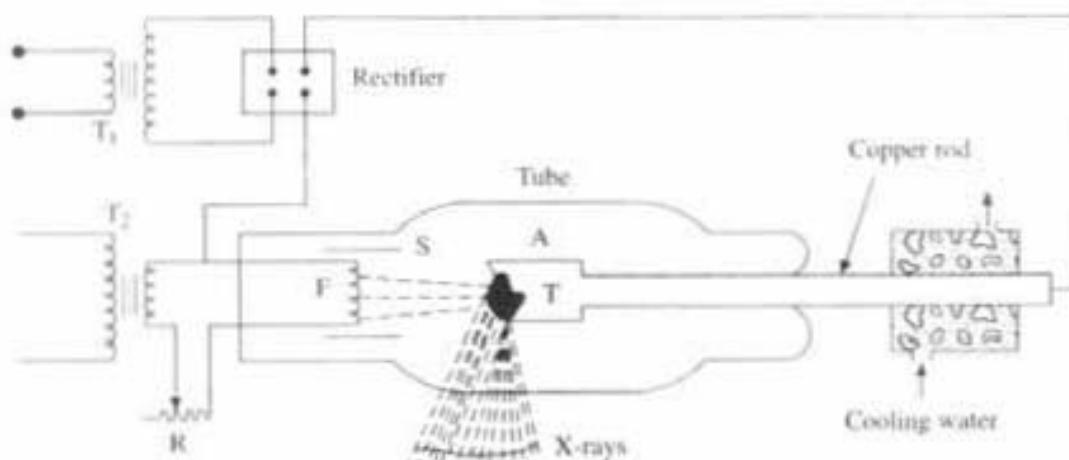


Fig. 2.1 Coolidge tube for production of X-ray

molybdenum, etc. are used as target material, which have high-melting points and also have a high atomic weight. When fast moving electrons from cathode are suddenly stopped by the tungsten target situated in their path, they experience a considerable negative acceleration. Such quickly decelerated charges radiate energy in the form of electromagnetic pulses which constitutes X-rays. Here, one thing should be remembered that only a small percentage of the electron energy at the target is converted into X-rays. While the rest is dissipated as heat. Due to this heating the target is heated up. In order to save the target it is constantly cooled by cooling arrangement.

Control of Intensity and Quality

The intensity of X-rays depends on the number of electrons striking the target. This number is determined by the temperature of the electron emitting filament, which is itself proportional to the heater current. Hence by controlling the filament current with the help of the rheostat R , thermionic emission and hence intensity of X-rays can be controlled. The quality of X-rays is measured in terms of their penetrating power which is dependent on the potential difference between filamentary cathode and the anode. Greater the accelerating voltage, higher the speed of the striking electrons and consequently, more penetrating X-rays produced. It is customary to refer to highly penetrating X-rays (i.e., those possessing high frequency), as hard X-rays and to those less penetrating (i.e., of low frequency) as soft X-rays, obviously the quality of penetrating power of X-rays can be controlled by varying the potential difference between cathode and anode.

It will be noticed from above explanation that in Coolidge X-ray tube, it is possible to achieve separate control of the intensity and quality of X-rays independent of each other.

2.4 Detection of X-Rays

X-rays are detected by some of their properties such as (a) action on photographic plates, (b) production of fluorescence in certain substance, (c) ionisation of gases and (d) penetration through matter.

2.5 Origin of X-Rays

X-rays are produced when high speed electrons strike some material object of high atomic number. It is found that nearly 99.8% of the energy of the electrons goes into the heating of the target material, only a few number of fast moving electrons produce X-rays by using their kinetic energy in the following two ways:

2.5.1 Line Spectrum or Characteristic Spectrum

Some of the high velocity electrons while penetrating the interior of the atoms of the target material, knock out the tightly bound electrons in the innermost shell (like K , L -shell etc.) of the atoms. Now the vacancies so created may be filled up by the electrons from higher shell, i.e. higher level electron jump to fill up the created vacancies. Due to this electronic transitions the energy difference is radiated in the form of X-ray of very small but definite wave lengths (and high frequency). These wave lengths constitute the line spectrum, which depends on

the characteristic of the material of the target. So these line spectra are also called *characteristic spectra*. Now since they are coming at a particular line of the spectra, hence the name. When the jump takes place from valence electron to some level which are not innermost then that electronic transition produces less energetic visible spectra.

Figure 2.2(a) shows the case when the high velocity incident electron knocks off one electron from K-shell. As shown in Fig. 2.2(b) and (c) this vacancy in K-shell is filled up by nearby electron in the L-shell. During this jump an X-ray radiation is emitted whose frequency is given by

$$\Delta E_1 = |E_K - E_L| = hf_1 \quad (2.1)$$

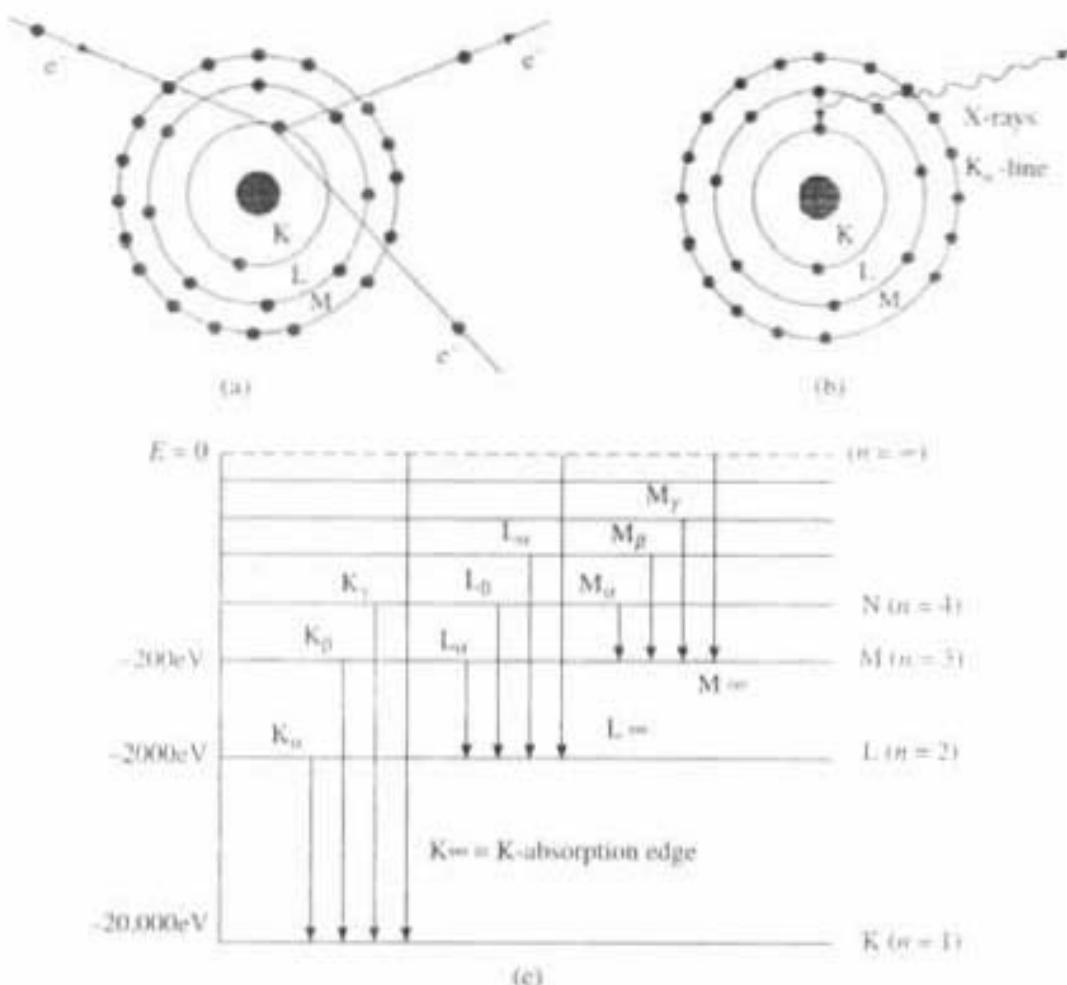


Fig. 2.2 Origin of characteristic spectrum K_α, K_β etc. due to electron jump from higher level

where E_K is the energy required to remove an electron from the K-shell and E_L is that required for L-shell. Since this energy difference is comparatively very large, the X-rays emitted have very large energy content and hence are highly penetrating. This X-ray radiation corresponds to K_α line of K-series. If, however, this vacancy in K-shell is filled up by an electron from M-shell, the X-rays emitted would be still more energetic and would consequently possess still higher frequency because,

$$\Delta E_2 = |E_K - E_M| = hf_2 \quad (2.2)$$

is more than $\Delta E_1 = |E_K - E_L| = hf_1$. This radiation with frequency f_2 corresponds to K_β line of K-series (Figs. 2.2 (b) and (c)). These K_α , K_β lines are the most intense lines in the X-rays spectrum. Similarly, when the incident electron carries somewhat lesser amount of energy, then it removes the electron from L-shell and consequent transition from other outer orbit say M , N will produce X-rays of frequency lower than that of K-series. These transitions give rise to the L-series of X-rays spectrum, L_α , L_β , L_γ (Figs. 2.2(b), (c)).

Spectral lines of M series are produced in a similar way as shown in the energy-level diagram (Fig. 2.2c). So the *characteristics of the line spectra or characteristic spectra are*

1. It has discrete spectral lines of K-series, L-series, etc.
2. *The energy difference $\Delta E_1 = E_K - E_L$ or $\Delta E_2 = E_K - E_M$ are the characteristic of the particular material used as a target in the X-ray tube, which are responsible for the frequency of K_α , K_β line. So these line spectra of X-rays are called characteristic spectra.* These frequencies will be different for different element.
3. X-ray spectra of different elements are found similar to each other. All are having K , L , M series but it was noticed by Moseley that the *frequencies of lines for every series from the element of higher atomic number is greater than that produced by an element of lower atomic number.*

He found the relationship between frequency and atomic number as

$$\sqrt{f} \propto (z - b)$$

where z is the atomic number and b the nuclear screening constant.

2.5.2 Continuous Spectrum

The continuous spectrum is emitted, when a few fast moving electrons penetrate deep into the interior of the atom and suffer repeated losses of energy due to the diversion of their path from original path, in the attractive Coulomb force field of nuclei of target atoms. This process of losing energy by radiation is called *Bremsstrahlung* (braking-or slowing-down radiation). This energy appears as a continuous X-rays spectrum. The continuous distribution shows that all values of energy losses are probable, which constitutes the *continuous spectrum*.

If the striking electron has its velocity reduced from v to v' (Fig. 2.3) then its loss of energy is $(\frac{1}{2}mv^2 - \frac{1}{2}mv'^2)$. This must be equal to energy of the X-ray photons emitted.

$$\frac{1}{2}m(v^2 - v'^2) = hf$$

The highest frequency of the emitted X-rays corresponds to this case, when the electron is completely stopped, i.e., $v' = 0$ then

$$\frac{1}{2}mv^2 = hf_{\max} = ch/\lambda_{\min} \quad (\because c = f\lambda) \quad (2.3)$$

When an electron is accelerated by a potential V , then kinetic energy is

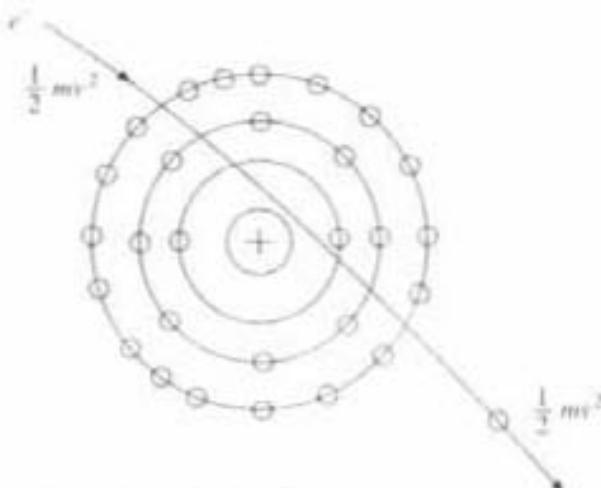


Fig. 2.3 Origin of continuous spectrum

$$\frac{1}{2}mv^2 = eV \quad (2.4)$$

by equating (2.3) and (2.4), we get

$$\frac{ch}{\lambda_{\min}} = eV, \lambda_{\min} = \frac{ch}{eV} \quad (2.5)$$

Now $e = 1.602 \times 10^{-19}$ C, $c = 3 \times 10^8$ m/sec velocity of lights and $h = 6.62 \times 10^{-34}$ joule sec.

Substituting the value of e , h and c , λ_{\min} becomes

$$\lambda_{\min} = \frac{12,400 \times 10^{-10}}{V} \text{ m} = \frac{12,400}{V} \text{ Å}$$

$$\text{i.e. } 1 \text{ Å} = 10^{-10} \text{ m} \quad (2.6)$$

Equation (2.6) shows that λ_{\min} for continuous spectra does not depend upon the target material but it is inversely proportional to the applied voltage V , between cathode and anode. Thus the *characteristic of continuous X-ray spectrum are:*

1. It consists of continuous range of frequencies upto maximum frequency f_{\max} which corresponds to λ_{\min} (Fig. 2.4b).
2. It is observed from the curves (Fig. 2.4b) that there is a shift of the maximum intensity position towards the shorter wave length side as voltage V increases.
3. The intensity of continuous spectrum is directly proportional to the atomic number of the material at a constant applied voltage.

Figure 2.4(a) shows the intensity distribution of X-rays from 30 kV X-ray tube with molybdenum (Mo) and tungsten (W) target. The curve for tungsten (W) shows the continuous spectrum of X-ray, whereas curves for molybdenum (Mo) target show two sharp lines K_{α} and K_{β} at specific wave lengths, superimposed on the continuous spectrum.

K_{α} , K_{β} lines for tungsten are obtained at 75 kV operating voltage. Thus the

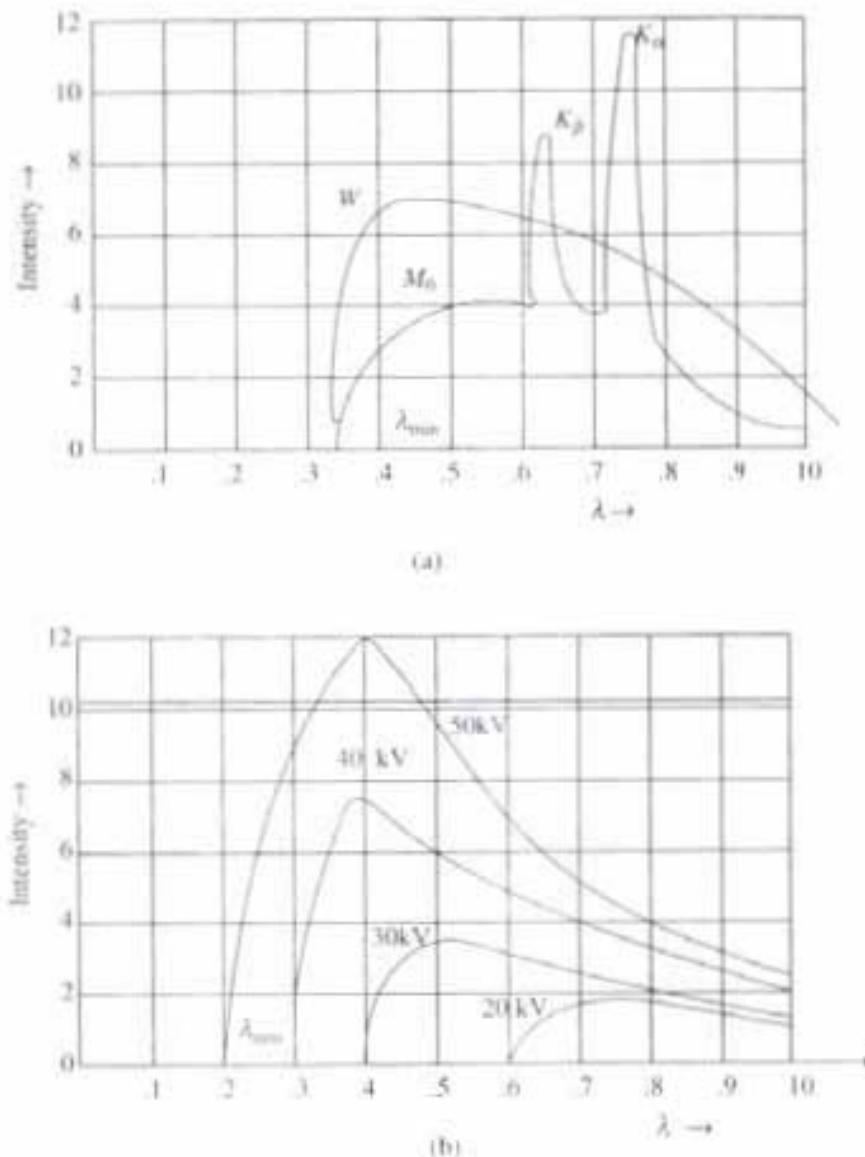


Fig. 2.4 (a) Characteristic spectrum; (b) Continuous spectrum with different excitation voltage

number of characteristic lines present in the X-ray spectrum depends both on the nature of the target material and the excitation voltage. Figure (2.4b) shows that for different acceleration voltage potential X-rays produced are of continuous wave lengths and the λ_{\min} is inversely proportional to applied voltage. It has a value $12400/V$.

2.6 Moseley's Law

In 1913–14 Moseley found out that the characteristic X-ray spectrum of different elements are remarkably similar to each other, all are having K -, L -, M -series. But there is a difference in frequency for each series. The frequency of lines in every series produced from an element of higher atomic number, is greater than that produced by an element of lower atomic number. It is due to the fact that binding energy of electrons goes on increasing as we increase the atomic number.

so larger amount of energy is required to liberate an electron from the K , L or M shell of that element whose atomic number is larger. Frequency for K line for copper ($Z = 29$) is 1.94×10^{18} Hz, whereas that for molybdenum ($Z = 42$) is 4.16×10^{18} Hz. The relationship between the frequency and atomic number is given by Moseley as

$$f \propto (Z - b)^2 \text{ or } \sqrt{f} = a(Z - b) \quad (2.7)$$

where Z is the atomic number and a and b are constants for a particular series but vary from one series to another. The constant b is called *nuclear screening factor*. For K-series $b = 1$ and it is having higher values for L , M series. So the Moseley's law states that frequency of a spectral line in the characteristic X-ray spectrum varies directly as the square of the atomic number of the target element (Fig. 2.5).

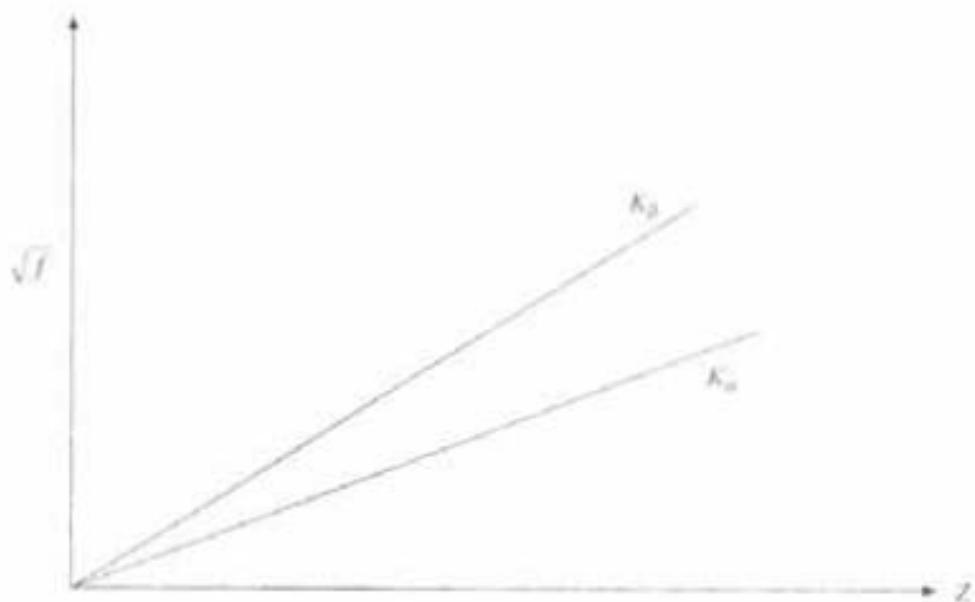


Fig. 2.5 Variation of square root of frequencies (square root of f) with atomic number (Z).

An exact form of Moseley's Law is

$$f = \frac{c}{\lambda} = R c (Z - b)^2 \left[\frac{1}{n_1^2} - \frac{1}{n_2^2} \right] \quad (2.8)$$

where R is the Rydberg's constant n_1 , n_2 are the energy levels between which the transition occurs, and c the velocity of light.

The shortest wave length and highest energy associate with a shell is called the *absorption edge wave length of that shell*. The corresponding energy will be liberated when an electron jumps from infinite ($n = \infty$) to that particular shell. So frequency of K-absorption edge is

$$f_{K\infty} = R c (Z - b)^2 \left[\frac{1}{1^2} - \frac{1}{\infty^2} \right]$$

2.7 Comparison Between Bohr's Theory of Atomic Spectra and Moseley's Law

It can be shown that Moseley's law is in good agreement with Bohr's theory for hydrogen atom. Moseley generalised Bohr's theory for all elements. One can compare the work of Bohr and Moseley in a quantitative way, so that it strengthens both their views.

According to Bohr's theory the frequency of the spectral lines of hydrogen is

$$f = R_H Z^2 \left[\frac{1}{n_1^2} - \frac{1}{n_2^2} \right] \quad (2.9)$$

This bears some resemblance to Moseley's law

$$f = R c (Z - b)^2 \left[\frac{1}{n_1^2} - \frac{1}{n_2^2} \right]$$

where Rydberg's constant $R_H = R_c$ and Z in Bohr's theory is replaced by $(Z - b)$ in Moseley's law.

Now if we remove one of the two K-shell electrons of an element to infinity, the remaining outer electrons will be attracted to the nucleus not by the nuclear charge (Ze) as Bohr assumed but by a charge $(Ze - e)$. One may think of remaining K-electron as 'screening' the outer electrons from the full nuclear attraction. Thus 'b' term for Moseley's law comes out to be 1 for K-series which is called *nuclear screening factor*. So Moseley's Law is in good agreement with Bohr's theory of hydrogen atom. From this law Moseley first proved that it is the atomic number and not the atomic weight of an element which determines its characteristics and position in the periodic table.

2.8 Properties of X-Rays

- (1) X-rays are electromagnetic waves of very small wave length of the order of 1 Å, 0.1 Å. They travel with velocity of light and are invisible to eyes.
- (2) They are very energetic and highly penetrating, due to which they can pass through wood, flesh, etc. which are opaque to ordinary light.
- (3) Under suitable condition X-rays can reflect, refract, interfere, diffract and get polarised like ordinary light.
- (4) X-rays can ionise the gas through which it passes.
- (5) They cause fluorescence in many substances.
- (6) X-rays have destructive effect on living tissue.
- (7) Photo-electric effect is also observed with X-rays.
- (8) X-rays are not deflected by electric or magnetic fields.

2.9 X-Ray Diffraction and Bragg's Law

The crystal structure can be studied by X-ray diffraction method. Diffraction would depend upon the crystal structure and on the wave length of light. An optical wave length such as 5000 Å, when superimposed on a crystal, will scatter elastically by the individual atoms of the crystal, result ordinary optical reflection.

But when X-ray radiation of wave length 10^{-8} to 10^{-9} cm is incident on the crystal whose lattice constant (i.e., the distance between two atoms) is of the order of 10^{-7} to 10^{-8} cm, then wave length of X-ray and lattice constant become comparable to each other, which gives rise to *X-ray diffraction*.

X-ray diffraction is same as ordinary diffraction which means—if the obstacle size is comparable with the wave length of radiation, then they divert from the original path and diffraction takes place. Since the wave length of X-rays is comparable with the interatomic distance or interplaner distance of the crystal, so when X-rays are incident on the crystal surface then the systematic arrangement of the atoms in the crystalline solid in three dimensions will act as a three dimensional grating, where atoms act as an obstacle and interatomic or interplaner distance act as an open slit, just like ordinary grating. Due to which diffraction can take place.

W.L. Bragg presented a simple explanation of the X-ray diffraction from a crystal. Let us suppose X-ray wave are incident on the crystal at an angle θ , where θ is measured from the crystal plane and known as 'Glancing angle' (Fig. 2.6). Incident X-rays are specularly reflected (i.e., angle of incident = angle of reflection) or scattered in all directions from the parallel planes of atoms in the crystal. Which is actually treated as X-ray diffraction. The major difference between reflection or ordinary scattering with diffraction is that in diffraction, interplaner distance or interatomic distance is responsible mainly, according to the eqn. (2.10), which is not the case for ordinary reflection or scattering. Since the refractive index of the crystal is very nearly equal to unity, so there is practically no bending due to refraction when X-rays are entering or leaving the crystal.

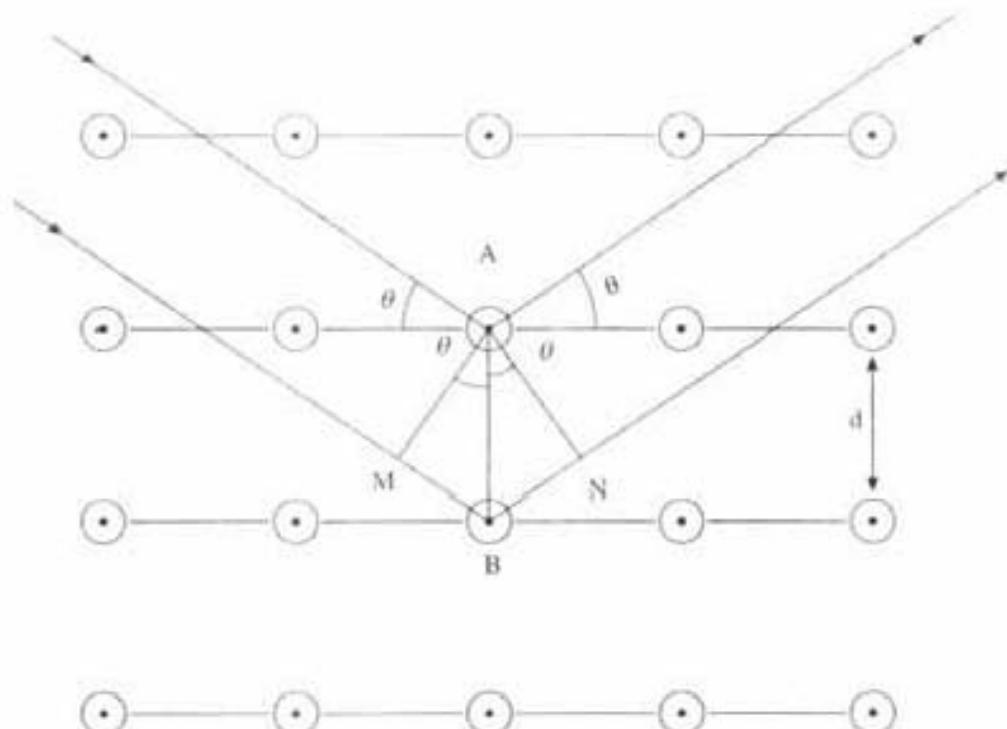


Fig. 2.6 X-ray diffraction which obey Bragg's law

In crystals there are certain planes which are particularly rich in atoms in three dimensional space, and the combined diffraction/scattering of X-rays from these planes can be treated as reflection from these planes. Each such plane reflects only a very small fraction of the incident radiation and only when the reflected X-rays from parallel planes of atoms interfere constructively, then only effective diffracted beams are found at an angle θ (Fig. 2.6). Here, we treat elastic scattering in which the energy of X-ray is not changed on reflection. Thus, the crystal behaves as a three-dimensional reflection grating with X-rays, of wave length (10^{-8} to 10^{-9} cm), which is comparable with lattice constant (10^{-7} to 10^{-8} cm).

Consider parallel lattice planes spaced d apart (Fig. 2.6), the X-ray radiation is incident in the plane of the paper. The path difference for rays reflected from adjacent planes is, (from Fig. 2.6)

$$MB + BN = 2d \sin \theta$$

where θ is the glancing angle. Constructive interference of the radiation from successive planes occurs when the path difference is an integral multiple n of wave length λ , so that

$$2d \sin \theta = n\lambda \quad (2.10)$$

where $n = 1, 2, 3, \dots$ etc. for first order, second order maxima. This relation is known as *Bragg's Law*. Although the reflection from each plane is specular, but only for certain values of θ the reflections from all parallel planes add up in phase to give a strong reflected/diffracted beam. For other angles, the reflection from different planes are out of phase and they will not reinforce. Bragg's law is a consequence of the periodicity of the lattice. The law does not refer to the arrangement of atoms on the 'basis' associated with each lattice point. The composition of the basis determines the relative intensity of the various order n of diffraction from a given set of parallel planes. Bragg's reflection can occur only for wave length $\lambda \leq 2d$. So we cannot use visible light, since interplaner distance $d = 10^{-7}$ cm.

2.10 Explanation of X-Ray Diffraction

Bragg's law, $2d \sin \theta = n\lambda$, will be valid for the reflection of some of the incident X-rays from one of the various sets of parallel crystal planes which are particularly rich in atoms. Two such planes are shown in Fig. (2.7) by double lines. Now the interplaner distance d is different for different sets of plane oriented differently θ to the incident X-ray beam. So d and θ (Fig. 2.7) should be matched with λ , in order to satisfy Bragg's law. *Monochromatic X-rays of wave length λ striking a three-dimensional crystal at an arbitrary angle of incident θ , will not in general be diffracted. To satisfy the Bragg's law requires an accident, and to create the accident i.e. to match λ with d and θ , it is necessary to scan-in either wave length or angle. The standard methods for X-ray diffraction used in crystal structure analysis by Bragg's spectrometer is based on this requirement.*

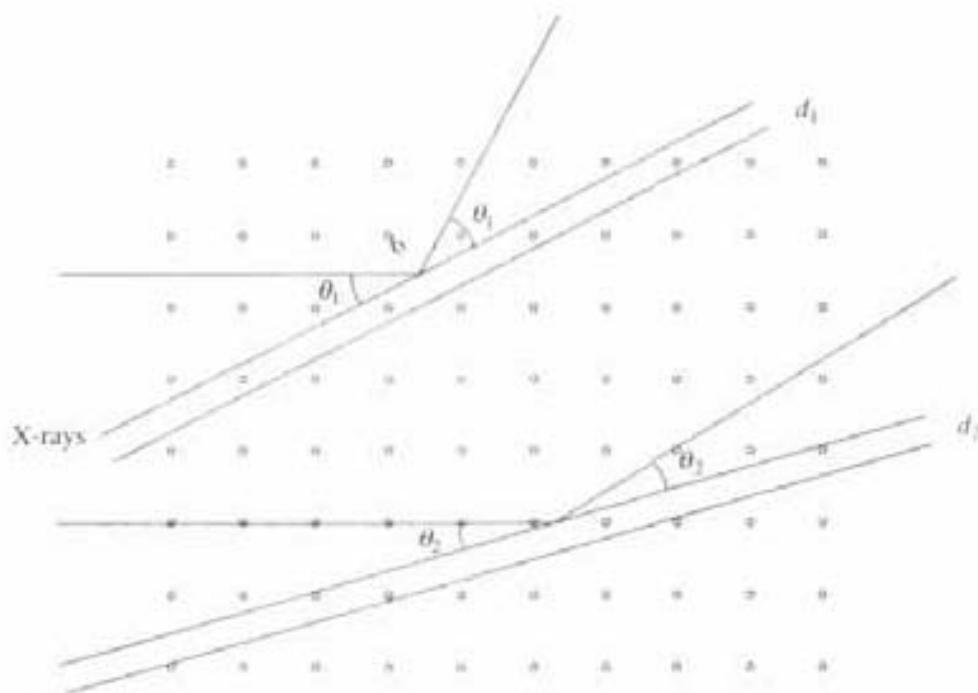


Fig. 2.7 X-ray diffraction from different set of lattice planes having different atomic concentration

2.11 Bragg's Spectrometer and Different Methods for X-Ray Diffraction

A schematic diagram of Bragg's X-ray spectrometer is shown in Fig. 2.8. The X-rays from a tube collimated into a narrow beam by slit S_1 impinge on the crystal C on the turn table, which acts as a three dimensional diffraction grating. The angular position of the crystal can be changed by the rotation of the turn table about a vertical axis passing through its centre and its rotation can be read from the circular scale. After diffraction, the X-rays enter an ionization chamber D , which absorbs X-rays strongly. The electrometer E records the intensity of ionisation in the chamber or it can be recorded by a photographic film.

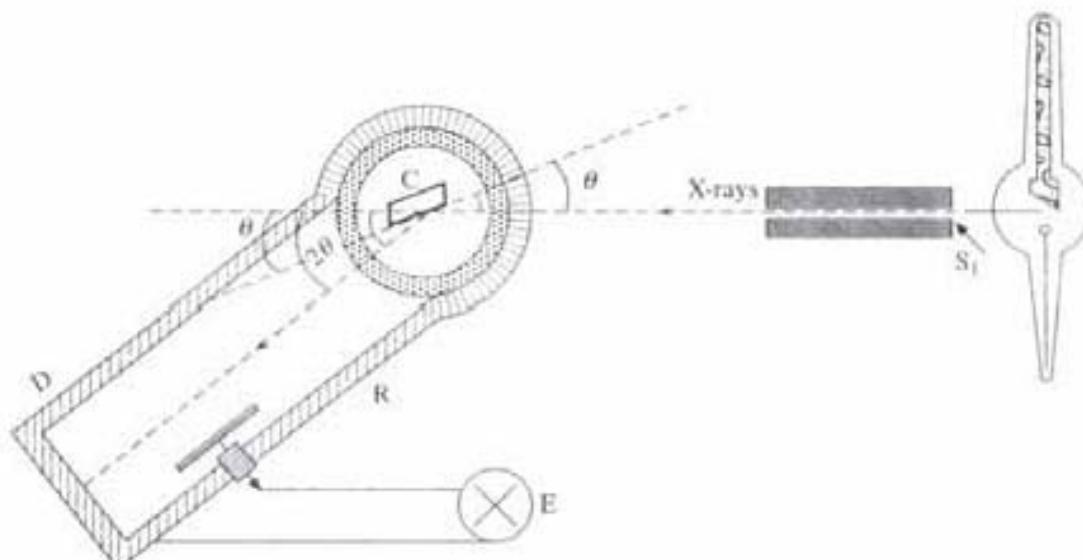


Fig. 2.8 Bragg's spectrometer

The turn table and the arm '*R*' of the ionisation chamber are so linked together that when the turn table, i.e., the crystal rotates through an angle θ , the arm *R*, i.e., ionisation chamber rotates through 2θ . So during the scanning of glancing angle θ at different values, beam is always diffracted into the ionization chamber. Ionisation passes through beams at certain sharply defined angles (Fig. 2.9). A_1, B_1, C_1 represent first order X-ray spectrum line, A_2, B_2, C_2 , for 2nd order. Angles and intensities ratio of these lines are same in first and 2nd order. As we increase the order, intensity will decrease. There are several methods for X-ray diffraction study. Here the standard three methods are discussed.

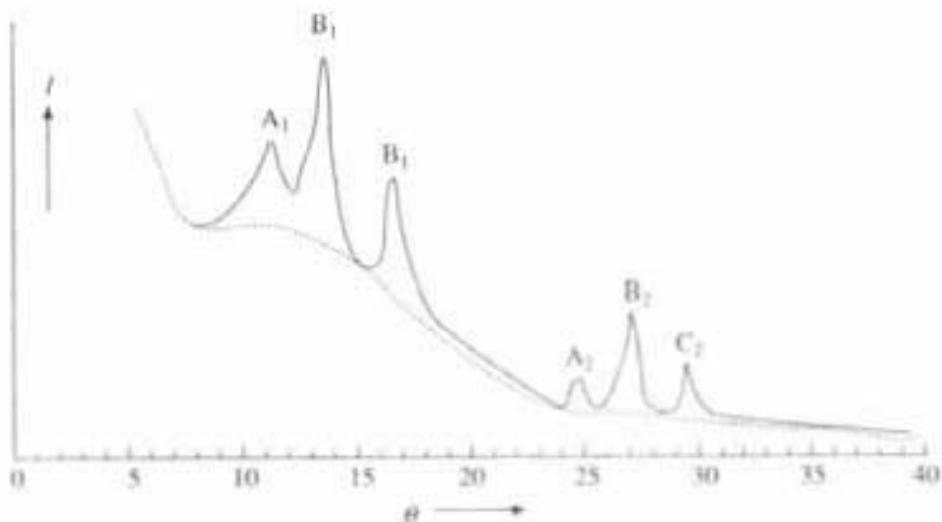


Fig. 2.9 Change of ionisation current (I) with glancing angle θ . Peaks showed that only for particular values of θ , X-ray diffraction is occurring

2.11.1 Laue Method

In Laue method (Fig. 2.10) a single crystal is stationary in a beam of X-rays of continuous wave length. The crystal selects and diffracts the discrete values of λ for which interplanar spacing d and glancing angle θ satisfy Bragg's law. A source is used that gives a beam of X-rays over a wide range of wave lengths. A pinhole arrangement produces a well collimated beam. Flat film receives the diffracted beams. The diffraction pattern consists of series of spots. That pattern will show the symmetry of the crystal. If a crystal has four-fold axis of symmetry parallel to the beam, then Laue pattern will show four-fold symmetry (Fig. 2.10).

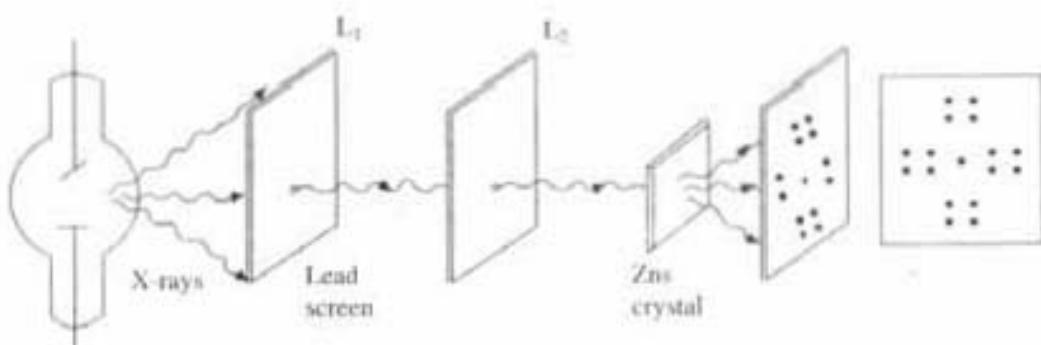


Fig. 2.10 Laue method for X-ray diffraction

2.11.2 Rotating Crystal Method

In rotating crystal method a single crystal is rotated about a fixed axis in a beam of monochromatic X-rays. The variation in angle θ brings different atomic planes into position for reflection. The beam is diffracted from a given crystal plane when in the course of rotation θ satisfies the Bragg's law. Diffracted rays are recorded on a photographic film mounted in a cylindrical holder concentric with a rotating spindle crystal mount.

2.11.3 Powder Method

In the Powder method (Fig. 2.11) the incident monochromatic radiation strikes a finely powdered specimen or a finely grained polycrystalline specimen contained in a thin-walled capillary tube. The distribution of crystallite orientations will be random, i.e., nearly continuous. The powder method is convenient precisely because a single crystal is not required which is difficult to get and it is costly. Diffracted rays go out from individual crystallites, that happen to be oriented with planes making an incident angle θ with the beam which satisfies the Bragg's equation. Moreover each set will give out not only first order but higher orders also. Since all orientations are equally likely, the reflected rays will form a cone where axis lines up along the direction of the incident beam and whose semivertical angle is 2θ for that particular set of planes. The cones intercept the film in a series of concentric rings (Fig. 2.11), when photographic film mounted in a cylindrical camera which is concentric with powder crystal specimen.

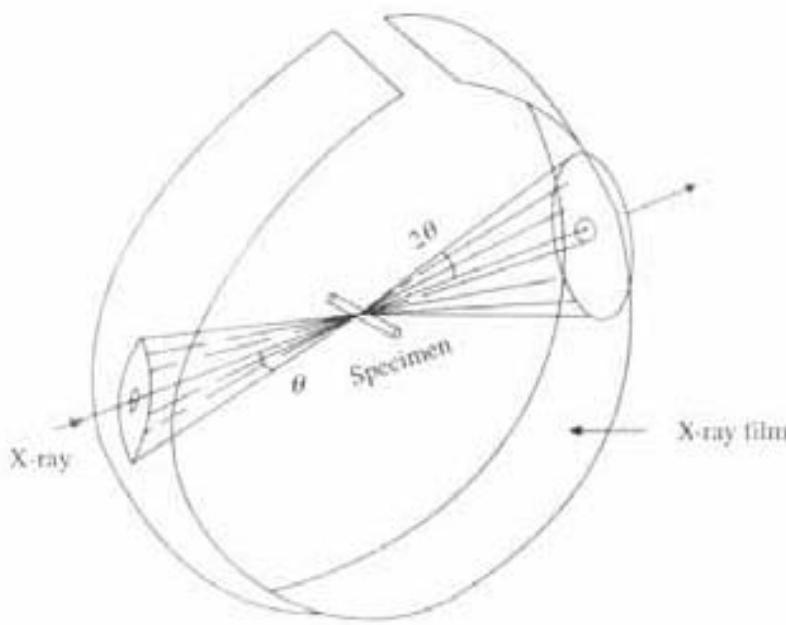


Fig. 2.11 X-ray diffraction by powder method

If the distances between symmetrical lines on the photograph are L_1, L_2, L_3 etc. (Fig. 2.12) and diameter of cylindrical film is D then

$$\frac{L_1}{\pi D} = \frac{4\theta_1}{360^\circ}, \quad \theta_1 = \frac{90^\circ}{\pi D} L_1$$

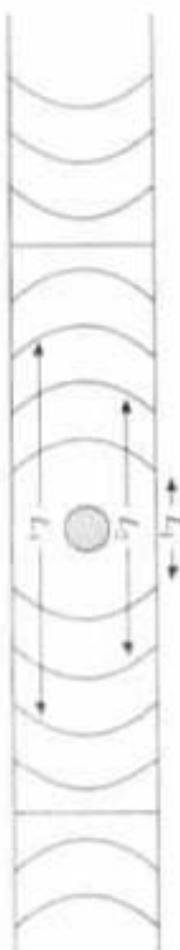


Photo Film

Fig. 2.12 Calculation of interplaner distance (d) from the X-ray diffraction photograph of Powder method

Hence

$$\theta_1 = KL_1$$

where $K = 90^\circ/\pi D$.

Similarly, $\theta_2 = KL_2$, $\theta_3 = KL_3$ etc. Using these values of θ in Bragg's law equation (2.10), interplaner spacing d can be calculated, as d_1 , d_2 , d_3 for different set of lattice planes in a particular crystal.

2.12 Uses of X-Ray Diffraction Patterns

X-ray diffraction patterns are widely used for the following purposes:

1. In determining the microscopic structure of matter in solids, for example, in metallurgy for the study of mechanical properties.
2. In locating as well as determining relationship between different planes in a crystal lattice which are rich in atoms.

PROBLEMS

1. Calculate the smallest glancing angle at which K-copper line of 1.549 \AA will be reflected from crystal having atomic spacing of 4.255 \AA . What is the highest order of reflection that can be observed with this radiation?

(a) From Bragg's law

$$2d \sin \theta = n\lambda$$

where $n = 1, 2, \dots$, etc.

When θ is the smallest then $n = 1$

$$2d \sin \theta = \lambda$$

$$2 \times 4.255 \times \sin \theta = 1.549, \sin \theta = 0.1802$$

$\theta \approx 10.487^\circ$ the smallest glancing angle.

(b) Maximum value of $\sin \theta = 1$

$$\text{So for highest order } 2d = n\lambda, n = 2d/\lambda = \frac{2 \times 4.255}{1.549}, n_{\text{max}} = 5.494$$

Since fraction is meaningless so highest order of Bragg's reflection is five.

2. An X-ray tube operates on 60 kV potential difference. The minimum wave length

of the continuous X-ray is found to be 0.194 Å. Assuming all the electric energy is converted into X-rays, calculate the value of the Planck's constant. Given that

$$\text{Charge of electron} = 1.6 \times 10^{-19} \text{ C}$$

$$\lambda_{\min} = \frac{ch}{eV}$$

$$h = \text{Planck's constant} = \frac{\lambda_{\min} \times eV}{c}$$

$$h = \frac{0.194 \times 1.6 \times 10^{-19} \times 10^3 \times 60 \times 10^{-10}}{3 \times 10^8}$$

$$= 6.208 \times 10^{-34} \text{ J.S.}$$

3. In a crystal with simple cubic-structure the lattice-parameter is 3 Å with a certain beam of monochromatic X-rays. The glancing angles for the (110) plane in first order Bragg's reflection is 12.5°. Find the wave length of the incident X-ray beam. Find the maximum orders possible from these plane.

$$(a) \quad 2d \sin \theta = n\lambda$$

$$2 \times \frac{3}{\sqrt{2}} \times \sin(12.5) = 1 \times \lambda, \lambda = \frac{1.298}{\sqrt{2}} = 0.9178 \text{ Å}$$

$$\left(\text{Since } d_{110} = \frac{a}{\sqrt{1^2 + 1^2 + 0}} = \frac{3}{\sqrt{2}} \right)$$

(b) Maximum value of $\sin \theta = 1$, then $n = n_{\max}$

$$2d = n_{\max} \lambda, n_{\max} = 2d/\lambda = \frac{2 \times 3 \times \sqrt{2}}{1.298 \times \sqrt{2}} = 4.62$$

Therefore

$$n_{\max} = 4.62$$

Since fraction is meaningless so highest order for Bragg's reflection is Four

4. A X-ray tube operates on 9100 volts. The shortest wave length of this X-ray beam falls on a rock-salt crystal in a Bragg's spectrometer. It is found that for this wave length the first order Bragg's reflection occurs when the glancing angle is 14°. From the first principle, calculate the Planck's constant from the above data. Given the interplaner spacing $d = 2.81 \text{ Å}$, charge of electron $= 1.6 \times 10^{-19} \text{ C}$, velocity of light $= 3 \times 10^8 \text{ m/sec}$

$$2d \sin \theta = n\lambda$$

$$2 \times 2.81 \times \sin(14) = 1 \times \lambda_{\min}, \lambda_{\min} = 1.3596 \text{ Å}$$

Now

$$\lambda_{\min} = \frac{ch}{eV}, h = \frac{\lambda_{\min} \times eV}{c}$$

$$h = \frac{1.3596 \times 1.6 \times 10^{-19} \times 9100 \times 10^{-10}}{3 \times 10^8}$$

$$\text{Planck's Const.} = h = 6.598 \times 10^{-34} \text{ J.S.}$$

5. If comparing the wave lengths of two monochromatic X-rays lines, it is found that line A gives a first order Bragg's reflection maximum at a glancing angle at 30° to the smooth face of a crystal. Line B of known wave length of 0.97 Å gives

a third order reflection maximum at a glancing angle of 60° with the same face of the same crystal. Find the wave length of the line A.

$$\text{For line } A \quad 2d \sin \theta = \lambda A \quad (\because n = 1), \quad 2d \sin 30^\circ = \lambda A \quad (1)$$

$$\text{For line } B \quad 2d \sin \theta = 3\lambda B, \quad 2d \sin 60^\circ = 3 \times 0.97 \quad (2)$$

From Eqn. (2) we get $d = 1.68 \text{ \AA}$

$$\text{So from Eqn. (1)} \quad \lambda A = 1.68 \times 2 \times \sin 30^\circ = 1.68 \text{ \AA}$$

6. X-rays of unknown wave length give first order Bragg's reflection at glancing angle of 21.7° with (111) planes of Cu having F.C.C. structure. Find the wave length of these X-rays if the lattice constant for Cu is 3.615 \AA .

$$2d \sin \theta = n\lambda, \quad 2 \times \frac{3.615}{\sqrt{3}} \times \sin (21.7) = \lambda \quad (\because n = 1)$$

$$\lambda = \frac{2.673 \text{ \AA}}{\sqrt{3}} = 1.5433 \text{ \AA}$$

$$\left[\text{Since } d_{111} = \frac{a}{\sqrt{h^2 + k^2 + l^2}}, \quad d_{111} = \frac{a}{\sqrt{3}} \right]$$

7. The radiation from an X-ray tube operated at 50 kV are diffracted by a cubic *kcl* crystal of molecular mass 74.6 and density $1.99 \times 10^3 \text{ kg/m}^3$. Calculate (i) the short wave length of the spectrum from the tube and (ii) glancing angle for first order reflection from the principal planes of the crystal for that wavelength.

$$(i) \quad \lambda_{\min} = \frac{12,400}{V} \text{ \AA} = \frac{12,400}{50,000} = 0.24 \text{ \AA}$$

$$(ii) \quad \text{Now} \quad a^3 \rho = \frac{nm}{N}$$

where a is the lattice constant n the number of atoms per unit cell, m the mol. wt., N the Avogadro's number, ρ the density of the crystal.

$$m = 74.6, \quad N = 6.02 \times 10^{23}/\text{kg. mol.}$$

kcl is a F.C.C. crystal so $n = 4$

$$a^3 \times 1.99 \times 10^3 = \frac{4 \times 74.6}{6.02 \times 10^{23}}$$

$$a = 6.28 \times 10^{-10} \text{ m}$$

For *kcl* ionic crystal, $d = a/2$

$$d = 3.14 \times 10^{-10} \text{ m}$$

For Bragg's law $2d \sin \theta = n\lambda = \lambda_{\min}$ $(\because n = 1)$

$$2 \times 3.14 \times 10^{-10} \sin \theta = 0.248 \times 10^{-10}$$

$$\sin \theta = 0.395, \theta = 2.258^\circ$$

8. The K_α X-ray from palladium are reflected in the first order from a certain crystal at the following glancing angles: 5.4° from the (100) planes; 7.6° from (110) planes and 9.4° from (111) planes. Show how this data could be used to determine the type of cubic lattice structure possessed by the crystal.

From Bragg's law, we have

$$2d \sin \theta = n\lambda = \lambda \text{ for } n = 1$$

$$d_{100} = \frac{\lambda}{2 \sin(5.4^\circ)}, d_{110} = \frac{\lambda}{2 \sin(7.6^\circ)}, d_{111} = \frac{\lambda}{2 \sin(9.4^\circ)}$$

$$d_{100} = \frac{\lambda}{2 \times 0.0941}, d_{110} = \frac{\lambda}{2 \times 0.1323}, d_{111} = \frac{\lambda}{2 \times 0.1633}$$

$$1/d_{100} : 1/d_{110} : 1/d_{111} = 0.0941 : 0.1323 : 0.1623$$

$$= 1 : \sqrt{2} : \sqrt{3}$$

This relation is valid for simple cubic systems so the above crystal is simple cubic.

9. An X-ray analysis of a crystal is made with monochromatic X-rays of wave length 0.58 Å. Bragg's reflections are obtained at angles of (a) 6.50° (b) 9.15° and (c) 13°. Calculate the interplaner spacing of the crystal.

QUESTIONS

- Explain X-ray diffraction. Give a short account of Bragg's spectrometer. Deduce the formulae used.
- What is continuous and characteristics X-rays spectrum? Explain their origin.
- (i) State and explain Moseley's law. Explain its significance.
(ii) Discuss how Moseley explained the characteristic X-ray spectrum on the lines similar to line spectra of Hydrogen-like atom?
- The smallest glancing angle of a beam of X-rays incident on a crystal for which reflection is observed with a Bragg's Spectrometer is 9.30°. The crystal is then rotated by a certain angle and the detector is rotated by a corresponding angle to observe the 2nd order reflected beam. Calculate the angle through which the detector must be rotated in order to detect the 2nd order reflected beam.

Solution

$$2d \sin \theta_1 = \lambda \quad (1)$$

$$2d \sin \theta_2 = 2\lambda \quad (2)$$

Dividing (2) by (1) we get

$$\frac{\sin \theta_2}{\sin \theta_1} = 2, \sin \theta_2 = 2 \sin \theta_1 = 2 \sin 9.30^\circ$$

$$\text{Thus } \sin \theta_2 = 0.3232, \theta_2 = 18.85^\circ$$

When crystal rotates through an angle $\theta = \theta_2 - \theta_1 = 9.55^\circ$, then detector rotates through an angle $2\theta = 19.10^\circ$ to detect 2nd order.

5. When an electron is transferred in an atom from the L-shell, to K-Shell the X-rays emitted have a wave length of 0.788 Å. What is the atomic number of this atom?

Screening constant $b = 1$

Rydberg's constant $R = 1.097 \times 10^7 \text{ m}$

Solution: Frequency $f = \frac{c}{\lambda} = R c (z - b)^2 \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right)$

Here $n_1 = 1$ K-shell, $n_2 = 2$ L-shell

$$(z - b)^2 = \frac{1}{\lambda R} \left(\frac{1}{1 - 1/4} \right) = \frac{1}{\lambda R} \cdot \frac{4}{3}$$

$$(z - 1)^2 = \frac{4}{3} \times \frac{1}{1.097 \times 10^{-7} \times 0.788 \times 10^{-10}} = 15.42 \times 10^2$$

$$(z - 1) = 39.27, z = 40.27 \approx 40$$

6. The spacing between the principal planes in a crystal of NaCl is 2.820 Å. It is found that a first order Bragg's reflection of monochromatic beam of X-rays occurs at an angle of 10° . (a) what is the wave length of the X-rays in Å. (b) at what angle would a 2nd order reflection occur?
7. What is the shortest wave length that can be emitted by the sudden stopping of an electron when it strikes a screen of a T.V. tube operating at 10,000 V?
8. (a) At what p.d. must an X-ray tube operate to produce X-rays with minimum wave length of 1 Å? (b) What is the maximum freq. of X-rays produced in a tube operating at 20 kV?

(Hint.

$$\hbar v_{\max} = \frac{ch}{\lambda_{\min}} = eV, v_{\max} = \frac{eV}{h}$$

$$V = \frac{ch}{e\lambda_{\min}} = \frac{12,400}{\lambda_{\min}}$$

9. X-rays from a certain cobalt target tube are composed of strong K-series of cobalt and weak K-lines due to impurities. The wave length of the K_{α} lines are 1.785 Å for Cobalt and 2.285 Å and 1.537 Å for the impurities. (a) Using Moseley's law, calculate the atomic number of each of the two impurities. (b) What element are they? (for K-series $b = 1$ in Moseley's law, At. No. of cobalt = 27)

Solution

$$\lambda_{K_{\alpha}} = \frac{1}{R} \frac{1}{(Z - 1)^2} \left[\frac{1}{\left(\frac{1}{1^2} - \frac{1}{2^2} \right)} \right]$$

Z = At. No. of cobalt

Z_1 = At. No. of impurity 1

Z_2 = At. No. of impurity 2

Now

$$(Z - 1)^2 = \frac{1}{R \lambda_{K_{\alpha}}} \left(\frac{4}{3} \right) = \frac{4}{3R} \frac{1}{1.785} \quad (1)$$

$$(Z_1 - b)^2 = \frac{1}{R \lambda_{K_{\alpha_1}}} \left(\frac{4}{3} \right) = \frac{4}{3R} \frac{1}{2.285} \quad (2)$$

$$(Z_2 - b)^2 = \frac{1}{R \lambda_{K_{\alpha_2}}} \left(\frac{4}{3} \right) = \frac{4}{3R} \frac{1}{1.537} \quad (3)$$

Dividing (2) by (1) and (3) by (1), we get

$$\frac{(Z_1 - 1)^2}{(26)^2} = \frac{1.785}{2.285}, \quad (Z - 1)^2 = 528.07877, \quad Z_1 = 23.9$$

$Z_1 = 24 \rightarrow$ chromium

$$\frac{(Z_2 - 1)^2}{(26)^2} = \frac{1.785}{1.537}, \quad (Z_2 - 1)^2 = 785.0748, \quad Z_2 = 28.019$$

$Z_2 = 28 \rightarrow \text{Copper}$

10. The K-absorption edge of tungsten is 0.178 \AA and the average wave lengths of the K-series lines are $K_{\alpha} = 0.210 \text{ \AA}$, $K_{\beta} = 0.184 \text{ \AA}$ and $K_{\gamma} = 0.179 \text{ \AA}$
 (a) Construct the X-ray energy level diagram of tungsten. (b) What is the least energy required to excite the L-series? (c) What is the wave length of the $L\alpha$ line? (d) If a 100 KeV electron struck the tungsten target in a tube, what is the shortest X-ray wave length it could produce?

Solution:

$$\lambda_{\min} = \frac{ch}{eV} = \frac{12,400}{V} \text{ \AA}$$

where

V = Operating voltage

Electrons strike the tungsten target with an energy eV electron volts.

$$\text{So } \lambda_{\min} = \frac{12,400}{\text{energy in eV of electron}} \text{ \AA}$$

K-absorption edges = $\lambda_{K\alpha} = 0.178 \text{ \AA}$

Now minimum energy needed to remove K-shell electron is

$$\frac{12,400}{0.178} = 69662 \text{ eV}$$

Thus, K-shell Binding energy is

$$E_K = -69662 \text{ eV}$$

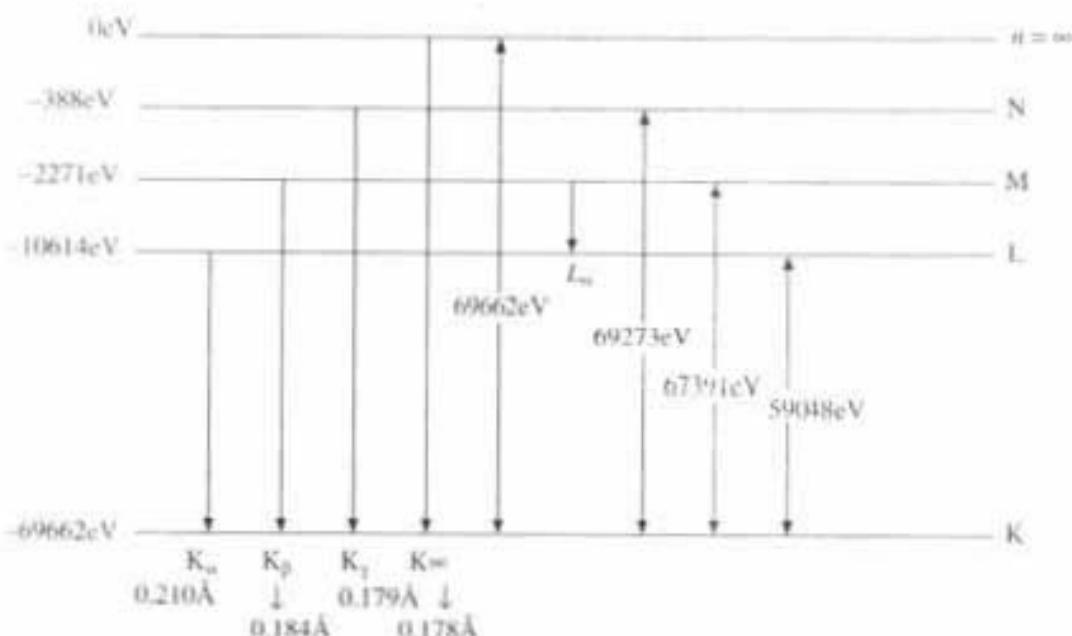


Fig. 2.13

$$\lambda_{K\alpha} = 0.210 \text{ \AA}, \text{ Energy of } K_\alpha \text{ radiation} = \frac{12,400}{0.210} = 59048 \text{ eV}$$

So L-shell is at $E_L = -69662 + 59048 = -10614 \text{ eV}$

$$\lambda_{K\beta} = 0.184 \text{ \AA}. \text{ So energy of } K_\beta \text{ radiation} = 12400/0.184 \text{ \AA} = 67391 \text{ eV}$$

Now M-Shell is at $E_M = -69662 + 67391 = -2271 \text{ eV}$

$$\text{Thus, Energy of } L_\alpha \text{ radiation} = 10614 - 2271 = 8343 \text{ eV}$$

Therefore

$$\lambda_{L\alpha} = 12,400/8343 = 1.487 \text{ \AA}$$

$$\lambda_{K\gamma} = 0.179 \text{ \AA} \text{ So energy of } K_\gamma \text{ radiation} = 12,400/0.179$$

$$= 69273.743 \text{ eV}$$

So N-Shell is at $E_N = -69662 + 69273 = -388 \text{ eV}$

- (a) Energy level diagram of tungsten is shown in Fig. (2.13).
- (b) Binding energy of L-series is $E_L = -10614 \text{ eV}$. So least energy required to excite the L-series is 10614 eV.
- (c) $\lambda_{L\alpha} = 1.487 \text{ \AA}$
- (d) $\lambda_{\min} = 12,400/100 \times 10^3 = 0.124 \text{ \AA}$

So shortest wave length emitted is 0.124 \AA

11. The radiation from a X-ray tube operated at 80 kV is incident on a sheet of tungsten. Using the energy levels of tungsten of problem (10), find the maximum kinetic energies of the electrons ejected from K, L, M shell in eV.

Solution

$$E_K = -69662 \text{ eV}$$

$$E_L = -10614 \text{ eV}$$

$$E_M = -2271 \text{ eV}$$

Incoming electron are having energy 80 keV because of operating voltage. So the K.E. of the electron ejected from K-Shell

$$= 80,000 - 69662$$

$$= 10,338 \text{ eV} = 10.4 \text{ keV}.$$

K.E. of electron ejected from L-Shell

$$= 80,000 - 10614$$

$$= 69386 \text{ eV} = 69.386 \text{ keV}$$

K.E. of electron ejected from M-Shell = $80,000 - 2271 = 77729 \text{ eV} = 77.729 \text{ keV}$

12. Explain the statement "crystal act as three dimensional reflection grating with X-ray". Why is X-ray diffraction said to occur as an accident?
13. What is the origin of the cut-off wavelength λ_{\min} ? Why is it an important clue for photo nature of X-rays?
14. Why does the characteristic X-ray spectrum vary in a systematic way from element to element but the spectrum in the visible range does not?
15. What are characteristic X-rays of an element? How can they be used to determine the atomic number of an element?

Motion of the Charged Particle in Electric and Magnetic Field

3.1 Introduction

The motion of a charged particle is affected by electric, magnetic or gravitational fields. When this particle is subjected to electric or magnetic fields, it experiences a force, and hence is accelerated. The trajectory can be determined precisely by Newton's law, provided that the forces acting on the particle are known. This property makes it possible to make radiation detecting instruments and particle accelerators. Charge of an electron e has been found by numerous experiments. The direct measurement of the mass m of an electron cannot be made but by using the effect on negatively charged electron due to electric and magnetic fields, one can measure e/m ratio for an electron, from which mass of an electron can be calculated. This chapter discusses most of the cases, by considering charged particle as electron but it will be equally valid for other charged particles like proton, α -particle, deuteron, etc. and also for ionised atoms.

3.2 Motion of Charged Particle in a Parallel Electric Field

Consider an electron is situated between two parallel plates of large surface area contained in an evacuated envelope (Fig. 3.1). A difference of potential V is applied between the two plates, the direction of the electric field between the plates is as shown in Fig. 3.1. The distance between the two plates is d . If this distance d is small compared to the dimensions of the plate then the uniform electric field between the plates is

$$E = V/d \quad (3.1)$$

and lines of force are pointing along negative X-axis.

Electrons starting from the -ve plate are attracted by the +ve plate and since there is no force along the y and z direction so the electrons emitted from -ve plate are accelerated along x -direction towards the +ve plate. The electric force acting on the electron of charge e and mass m is

$$eE = ma_x, a_x = eE/m \quad (3.2)$$

Hence,

$$a_x = e/m \cdot V/d \quad (3.3)$$

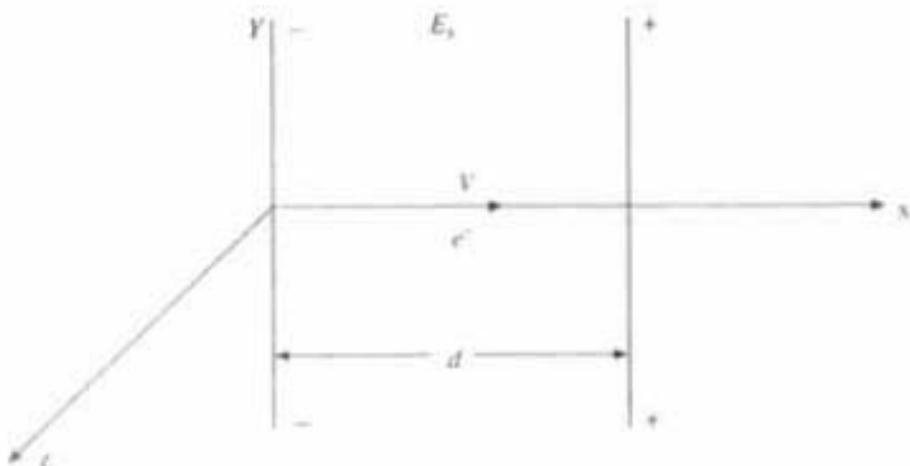


Fig. 3.1 Charged particle in parallel electric field

Thus the acceleration of the electrons along x -axis is constant, for a particular value of V and d .

If the electrons strike the +ve plate with an velocity v_t , which is called as *terminal velocity*, then the kinetic energy T of the electron will be

$$T = 1/2 mv_t^2 \quad (3.4)$$

which is equal to work done W , where

$$\begin{aligned} W &= \text{Force} \times \text{Displacement} \\ &= eE \times d = eV/d \times d = eV \end{aligned} \quad (3.5)$$

i.e.

$$1/2 mv_t^2 = eV$$

So the terminal velocity

$$v_t = \sqrt{2eV/m} \quad (3.6)$$

Equation (3.5) represents a new unit of energy or work called the electron volt eV. One electron volt is defined as the kinetic energy of a particle carrying a charge e coulomb and accelerated through a potential difference of 1 volt.

$$1 \text{ eV} = 1.602 \times 10^{-19} \text{ Joule.}$$

3.3 Motion of Electron in Transverse Electric Field

Let us suppose that an electron enters in the region between two parallel plates separated by a distance d Fig. (3.2), with an initial velocity (v_x) in +ve x -direction. An uniform electric field E_y is given between these two plates, the length of the field (E_y) is L . The direction of the electric field along -ve y -axis, no other fields existing in this region.

The motion of the charged particle under this transverse field E_y can be studied as follows:

Initial velocities along x , y , z directions are

$$\begin{aligned} u_x &= v_x \\ u_y &= 0 \quad \text{when } t = 0 \\ u_z &= 0 \end{aligned} \quad (3.7)$$

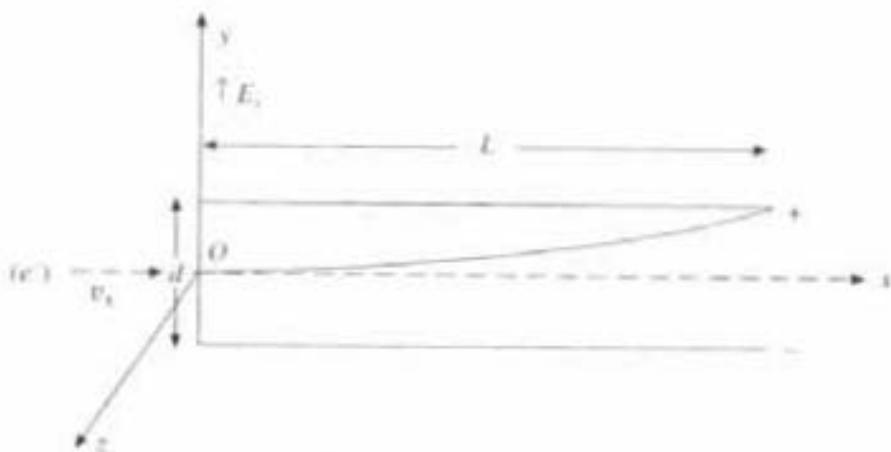


Fig. 3.2 Charged particle in transverse electric field

Since there is no force along x and z , so velocity along z remains zero and velocity along x -axis remains v_i and the distance travelled along x by any time t is

$$x = v_i t, \quad t = x/v_i \quad (3.8)$$

But there is constant electric field along y -axis so there is a constant acceleration along y -axis and the velocity at any instant of time t is

$$\begin{aligned} v_y &= u_y + a_y t \\ u_y &= a_y t \quad (\because u_y = 0) \end{aligned}$$

$$v_y = \frac{eE_y}{m} \frac{x}{v_i} = \frac{e}{m} \frac{V}{d} \frac{x}{v_i} \quad (3.9)$$

The distance travelled along y -axis by the time t within the applied electric field (E_y) is

$$y = u_y t + \frac{1}{2} a_y t^2, \quad y = \frac{1}{2} a_y t^2 \quad (\because u_y = 0) \quad (3.10)$$

When force acting on the electron of charge e along $+y$ -axis is

$$ma_y = eE_y$$

So,

$$a_y = eE_y/m = eV/dm \quad (3.11)$$

These equations indicate that in the region between the plates the electron is accelerated upwards with the velocity component v_y varying from point to point, whereas velocity component v_x remains unchanged in the passage of the electron between the plates. The resultant velocity will be

$$v = \sqrt{v_x^2 + v_y^2} \quad (3.12)$$

The path of the particle in the electric field with respect to the point O is

determined by combining equations (3.8) and (3.10), the variable t being eliminated. This gives

$$y = \frac{1}{2} a_y t^2 \quad (3.13)$$

which shows that the particle moves in a parabolic path in the region between the charged plates.

The angular deflection of the electron from original path produced by the transverse electric field at any instant of time t is given by

$$\tan \theta = \frac{v_y}{v_x} = \frac{eE_y}{m} \frac{x}{v_i^2} = \frac{eV}{dm} \frac{x}{v_i^2} \quad (3.14)$$

When L is the length of the plates where the electric field is given, then the displacement of the electron due to transverse electric field, after travelling length $x = L$, i.e. at the end point of the applied electric field area is obtained by using equations (3.11) and (3.13) as

$$y = \frac{1}{2} a_y t^2$$

$$y = \frac{1}{2} \left(\frac{eE_y}{m} \right) \left(\frac{L}{v_i} \right)^2 \quad (3.15)$$

It's an equation of a parabola, so it will move in a parabolic path. After the end of the applied electric field area, electron emerges out tangentially to the parabolic path at the point of emergence (see Section 3.7).

3.4 Motion of Electron in Transverse Magnetic Field

Let us consider an electron with initial speed v entering a uniform transverse magnetic field acting downward along z axis, i.e. perpendicular to plane of the paper (Fig. 3.3). Since magnetic field is perpendicular to initial velocity v and so to its motion at every instant, so no work is done on the electron. That means its kinetic energy will not increase and so its speed remains unchanged within the magnetic field. The force acting on the electron due to transverse magnetic field is

$$f_m = Bev$$

Since B , e , v are constants in magnitude so f_m is constant in magnitude and perpendicular to the direction of motion. Due to this type of force, electron will be continuously deflected given by Flemings left hand rule and finally results in motion in a circular path with constant speed within the transverse magnetic field. For the downward magnetic field, movement of the charged particle will be clockwise, whereas it will be anticlockwise for upward magnetic field. It is analogous to the problem of a mass tied to a rope and twirled around with constant speed. This force which is actually the tension in the rope, remains constant in magnitude and is always directed towards the centre of circle, and so is normal to the motion, which is known as centripetal force.

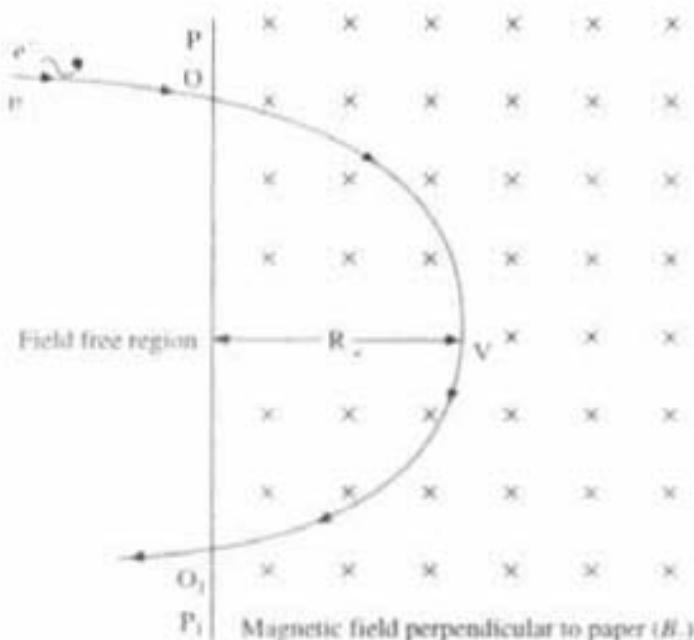


Fig. 3.3 Charged particle in transverse magnetic field

So in the transverse magnetic field when the electron of mass m is moving in a circular path of radius R with constant speed v has a radial acceleration towards the centre, then the magnetic force acting on it is

$$f_m = mv^2/R \quad (3.16)$$

i.e.

$$f_m = Bev = mv^2/R, \quad R = mv/eB \quad (3.17)$$

Equation (3.17) shows that in the transverse magnetic field, an electron will move in a circular path of radius R . After the magnetic field, the electron emerges out tangentially to the arc at the point of emergence. From eqn. (3.17) we can get

$$e/m = v/RB$$

which can be used for determining the e/m ratio for electron.

If T is the time period to complete one revolution then

$$T = 2\pi R/v = 2\pi Rm/BeR = (2\pi/B)(m/e)$$

So T is independent of v and R . The resonance frequency

$$n = 1/T = Be/2\pi m \quad (3.18)$$

3.5 Motion of the Charged Particle in Crossed Electric and Magnetic Field and its Application (Cathode Ray Tube)

Consider an electron moving with an initial velocity v entering in a crossed electric and magnetic field. The initial direction of velocity of the electron is along X-axis, electric field E applied along y . The direction of electric lines of force is along the (-Y-axis). Magnetic field of flux density B is applied along Z , the direction of magnetic lines of force is along (+Z-axis), the circle represents

the area of uniform magnetic field (Fig. 3.4). A screen is placed at a distance D ($= OS$) from the centre of electric and magnetic field.

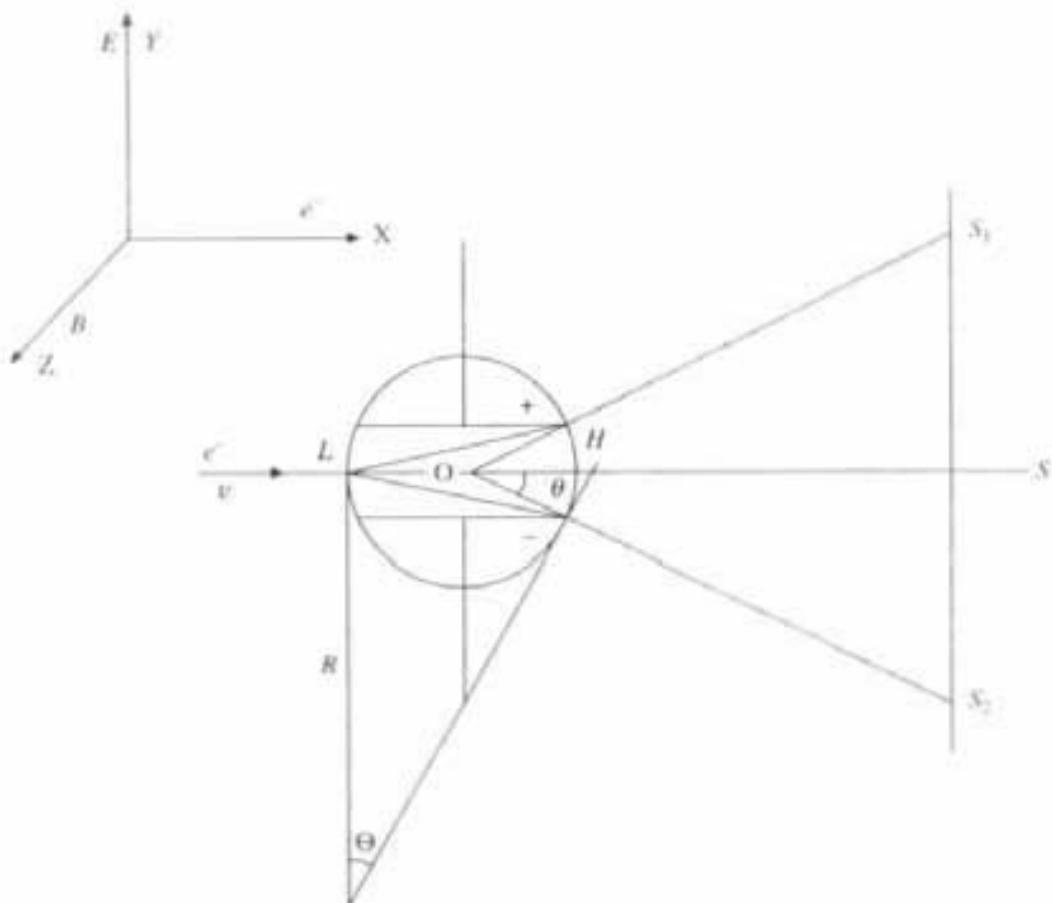


Fig. 3.4 Charged particle in crossed electric and magnetic field

Within these crossed electric and magnetic fields if the electron is made to travel along X direction then due to electric field alone it will deflect upwards along Y -axis (Fig. 3.4). Whereas due to magnetic field alone it will deflect downward along Y -axis. After crossing the magnetic field or electric field electron will move tangentially to the arc at point of emergence. So if S is the striking position of electron on the screen when there is no field, then due to electric field alone it will deflect by an amount SS_1 upwards (Fig. 3.4). Whereas due to magnetic field alone it will deflect by an amount SS_2 downward, just in the opposite direction. Now if force acting on the electron by electric and magnetic field simultaneously are same, then there will not be any deflection of the electron beam. Force acting on the electron due to electric field is eE . Whereas force acting due to magnetic field is Bev . If these two forces are equal then $SS_1 = SS_2$. So there will not be any deviation of the electron from its original path. In that condition

$$eE = Bev \quad (3.19)$$

i.e.

$$v = E/B$$

But

$$E = V/d$$

where d is the distance between two plates, and V the voltage applied between these two plates.

So

$$v = V/dB$$

When the electron is moving in a transverse magnetic field then it will travel a circular path of radius R . Then

$$mv^2/R = Bet$$

i.e.

$$v = BRe/m$$

(3.20)

From (3.19) and (3.20) we can get

$$v = V/dB = BRe/m$$

$$e/m = \frac{V}{dB^2 R}$$

(3.21)

where V and d are known. Now by recording the deflection φ of the magnetometer needle, due to the applied magnetic field, B can be calculated from the equation $B = H \tan \varphi$, where H is the horizontal component of earth's magnetic field. R can be calculated from Fig. 3.4 in the following way

$$\tan \theta = SS_2/OS = LH/R, \quad R = \frac{LH \times OS}{SS_2} \quad (3.22)$$

where OS is the distance between the centre of the deflecting plates and the screen. R is the radius of curvature of the curved path due to magnetic field. LH is the distance between two magnets used to apply magnetic field. So, everything is known in eqn. (3.21), from which e/m can be determined. This principle is used in Thomson method for measuring e/m ratio for electron.

3.6 Experimental Set-up for Thomson Method for Measuring e/m Ratio for Electron

The apparatus consists of a discharge tube (Fig. 3.5) called cathode ray tube (C.R.T.). C is cathode and A is anode. When a high $p.d$ is applied between cathode and anode, cathode rays are produced. These rays pass along X-axis through metal diaphragms maintained at anode potential. So, a fine pointed electronic beam is obtained. This beam strikes the C.R.T. screen normally and produces fluorescence at S .

P and Q are two parallel plates of length L , between which electric field can be applied, separated by a distance d . Initially when there is no electric and magnetic field then the fluorescent spot will be at S . Now by applying electric field along Y-axis, perpendicular to the motion of the electron (along X-axis), its deflection of the spot occurs, which is deflected upward at S_1 . Next by placing the C.R.T. between two pole pieces of a magnet, a downward magnetic field can be applied simultaneously along Z-axis, perpendicular to the motion of the electron (X-axis), as well as perpendicular to the direction of electric field

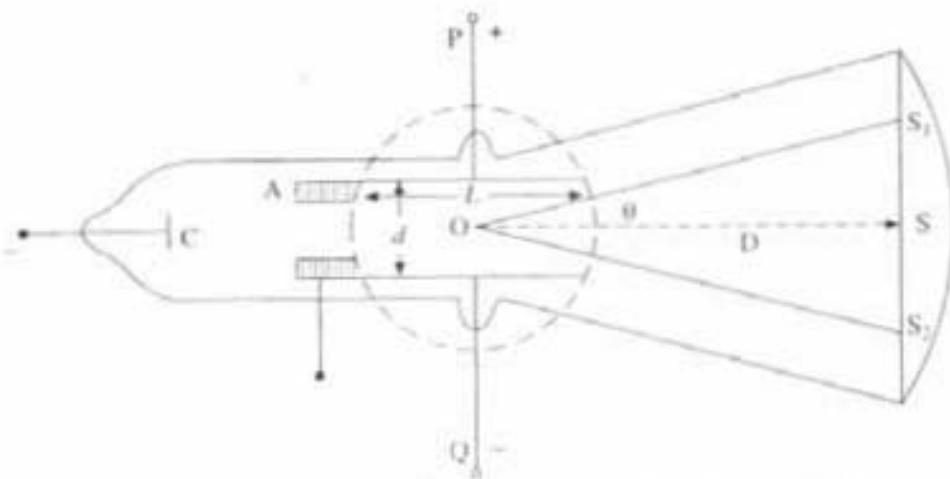


Fig. 3.5 Thomson method for measuring e/m ratio for electron

(Y-axis). The dotted circular area represents the region of uniform magnetic field. The direction of deflection will be downward. Since crossed electric and magnetic fields are acting on the electron beam simultaneously, then their effect will cancel each other and the spot will come back to its original position S . In these conditions one can calculate e/m ratio of electron by using eqns. (3.19) and (3.21), with the help of eqn. (3.22).

3.7 Vertical Deflection Due to Electric and Magnetic Field on a Screen and Sensitivity

When an electron is moving with initial velocity v_i perpendicular to the electric field ($E_i = V/d$), then the vertical deflection on the screen, which is at a distance D from the centre (O) of the applied electric field, can be calculated in the following way:

If accelerating anode voltage is V_A , then the energy of the electron moving with initial velocity v_i from the accelerating anode is

$$\frac{1}{2}mv_i^2 = eV_A$$

i.e.,
$$v_i = \sqrt{\frac{2eV_A}{m}}$$

From Fig. 3.5 we get vertical deflection.

$$y = SS_1 = OS \tan \theta = D \tan \theta$$

Now

$$\tan \theta = \frac{v_y}{v_x} = \frac{\frac{e}{m} \frac{V}{d} \frac{L}{v_i}}{v_i} = \frac{e}{m} \frac{V}{d} \frac{L}{v_i^2} \quad (3.23)$$

Because

$$v_y = \frac{e}{m} \frac{V}{d} \frac{L}{v_i} \text{ (from Eqn. (3.9))}$$

where (from Fig. 3.5) L = length of the plate P, Q in C.R.T. where voltage V is applied.

d = distance between two plates in C.R.T. where vertical voltage V is applied.

$D = OS$ = distance from O to the screen.

3.7.1 Vertical deflection on the screen at a distance D from the centre of the perpendicular electric field E_y

$$Y = D \tan \theta = D \frac{e}{m} \frac{V}{d} \frac{L}{v_i^2} = D \frac{e}{m} \frac{V}{d} \frac{Lm}{2eV_A}$$

$$Y = \frac{1}{2} D \frac{L}{d} \frac{V}{V_A} \quad (3.24)$$

Electrostatic deflection sensitivity: It is the deflection per unit applied voltage, i.e.

$$\frac{Y}{V} = \frac{1}{2} \frac{DL}{dV_A}$$

3.7.2 Vertical deflection due to perpendicular magnetic field of magnetic induction B

From eqns. (3.22) and (3.20), we get

$$Y = SS_2 = \frac{OS \cdot LH}{R} = (OS \cdot LH \cdot B \cdot e)/mv \quad \text{or} \quad Y = D \frac{BeL}{mv}$$

$$\text{Now } v = v_i = \sqrt{\frac{2eV_A}{m}} \quad \text{So} \quad Y = DBL \sqrt{e/2mV_A} \quad (3.25)$$

where $D = OS$ = distance from centre O of the magnetic field to screen.

$L = LH$ = length of the applied magnetic field.

Magnetostatic deflection sensitivity: It is the deflection per unit magnetic induction

$$\frac{Y}{B} = DL \sqrt{e/2mV_A} = \frac{DL}{v} \frac{e}{m}$$

Magnetic and Electrostatic Focusing

We have already seen that the motion of the charge particles can be controlled by proper electric and magnetic field. So it may also possible to focus the diverging beam of charge particles originating from the sources. These are essential features for some very important scientific equipments, such as Cathode ray oscilloscope, Cyclotron, Mass spectrometer, Electron microscope, etc. Now we will discuss how electrostatic and magnetostatic focussing is possible for charge particles.

3.8 Electrostatic Focusing

In electrostatic focusing a series of cylindrical anodes A, B at increasing potentials V_1, V_2 are issued (where $V_2 > V_1$).

Both the anodes are kept highly positive with respect to grid G and cathode, and separated by a gap. A high-equipotential ring R is placed in the gap between two anodes A and B to get the lens action. The shape of the equipotential lines on the cylindrical surface of anode A and B will be curved as shown by full lines in Fig. 3.6(a). The electric lines of force are always acting at right angles (dotted curve Fig. 3.6a) to the equipotential lines. The resultant electric field will be as indicated by the dotted lines (Fig. 3.6a). The curved equipotential planes form an electrostatic lens. Due to the effect of which the diverging beam of electrons from the source, say O (Fig. 3.6a) gets diverted from its original path and focussed at point I . The electron source is O and the electrons travel from a region A of low potential V_1 at velocity v_1 to a region B of high potential V_2 and velocity v_2 . Energy of the particle in region A is

$$\frac{1}{2}mv_1^2 = eV_1 \quad \text{or} \quad v_1 = \sqrt{(2eV_1/m)}$$

and that is in the region B is

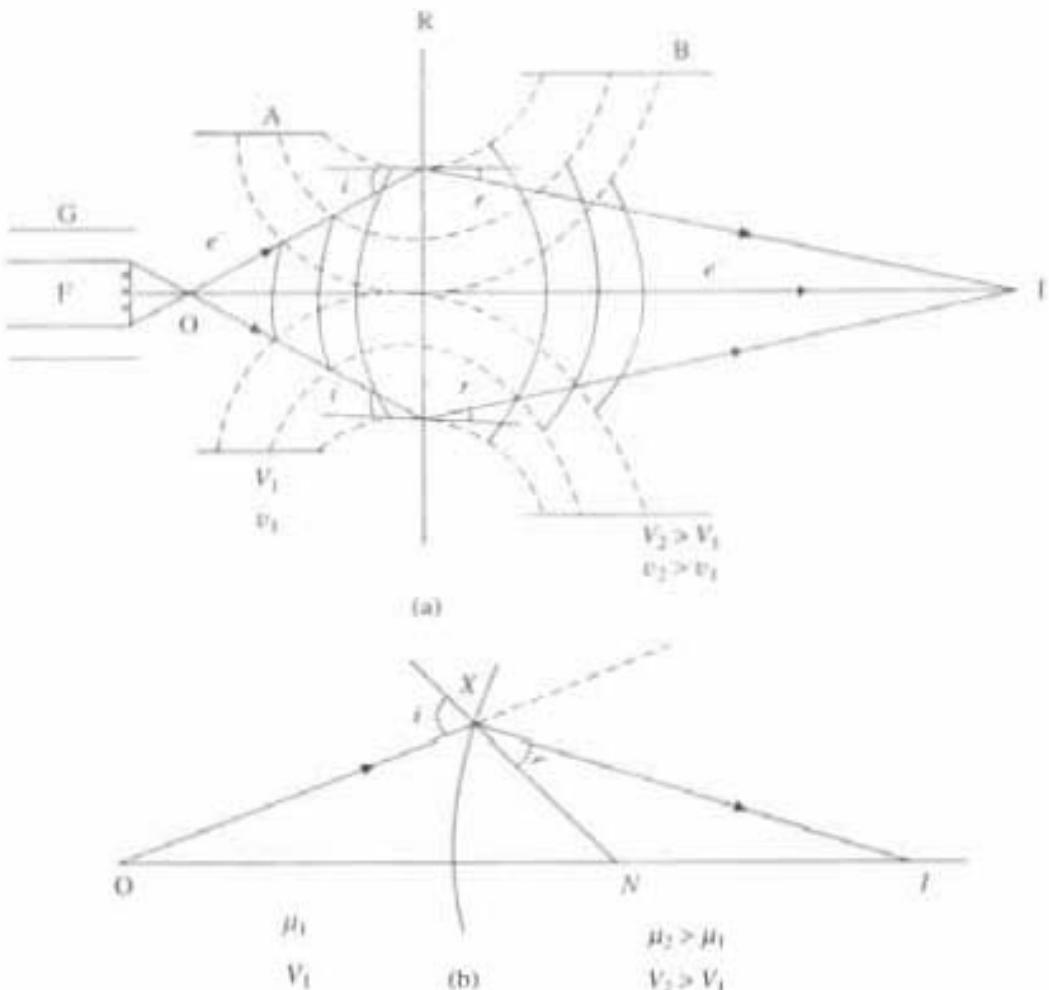


Fig. 3.6 (a) Electrostatic focusing; (b) Optical focusing

$$\frac{1}{2}mv_2^2 = eV_2 \quad \text{or} \quad v_2 = \sqrt{(2eV_2/m)}$$

i.e.

$$v_1/v_2 = \sqrt{(V_1/V_2)} \quad (3.26)$$

when $V_2 > V_1$, then $v_2 > v_1$.

This is similar to the refraction of a beam of light at a convex spherical surface, when the beam is travelling from a rarer to a denser medium (Fig. 3.6b). Thus, when diverging electron beam moves from lower +ve potential area to higher +ve potential area, it is diverted towards the central position, i.e. going to be focused. This can be mathematically shown as follows.

Explanation for Electrostatic Focusing

Let us consider Fig. 3.7, where XY be the equipotential interface like R in Fig. 3.6(a) between region A and B , which are of the uniform electrostatic +ve potential V_1 and V_2 where $V_2 > V_1$.

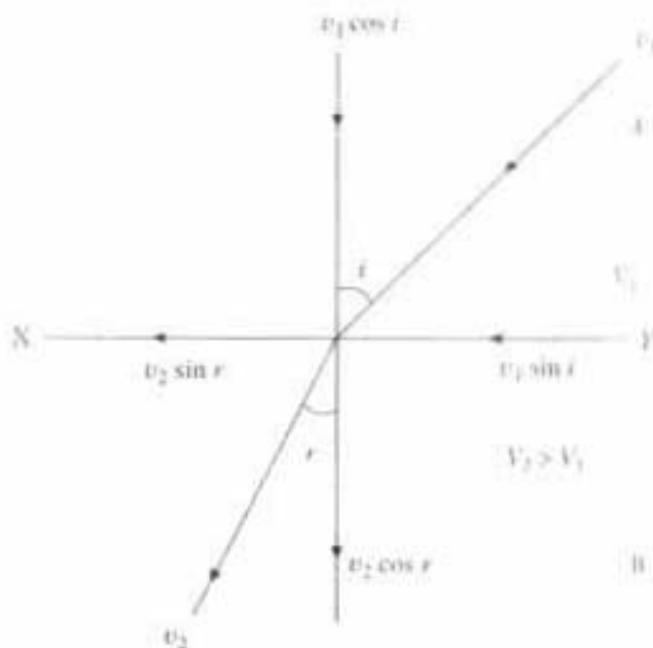


Fig. 3.7 Explanation of electrostatic focusing

Electrons moves with a velocity v_1 in region A and with v_2 in region B . If i and r are the angle of incidence and angle of refraction of the electron beam on the interface XY , then we can resolve the velocity v_1 and v_2 into parallel and perpendicular components to the surface (Fig. 3.7). Now XY is an equipotential surface so the parallel velocity components $v_1 \sin i$ and $v_2 \sin r$ will be equal; there should not be any change. So,

$$v_1 \sin i = v_2 \sin r$$

$$v_2/v_1 = \sin i/\sin r = \sqrt{V_2/V_1} \quad (3.27)$$

i.e., from equation (3.27) it is clear that when $V_2 > V_1$ then $i > r$, the electron beam

will divert towards the normal, towards the central position. At the same time since $i > r$ and $v_2 > v_1$, so the perpendicular component $v_2 \cos i > v_1 \cos i$, which means electron beam will be accelerated as well as diverted towards the central position, i.e. going to be focussed when moving from lower potential area to higher potential area. This is similar to refraction of light (Fig. 3.6b), which is the main feature of electrostatic focusing.

The focal length of the optical lens is fixed by the curvature of the lens but the focal length of the electrostatic lens can be changed by altering the relative size and potentials of the cylindrical surfaces A and B. This kind of electrostatic focusing is widely used in cathode ray oscilloscope.

3.9 Magnetostatic Focusing and Magnetic Lens

Let O is the position of the electron inside the uniform magnetic field of intensity \bar{B} (Fig. 3.8a) moving with a velocity v along OP making an angle θ with \bar{B} . Velocity v can be resolved into two components $v \cos \theta$ along \bar{B} and $v \sin \theta$ perpendicular to \bar{B} . If there is only one component $v \sin \theta$ then electron will be moving in a circular orbit of radius r such that

$$mv^2 \sin^2 \theta / r = Bev \sin \theta$$

i.e.,

$$v \sin \theta = Ber/m$$

Now time taken by the electron to move in a complete circle of radius r Fig. 3.8(a), with velocity $v \sin \theta$ is

$$t = 2\pi r/v \sin \theta = 2\pi rm/Ber$$

$$t = \frac{2\pi m}{Be} \quad (3.28)$$

But when the electron possesses both the velocity component $v \sin \theta$ as well as $v \cos \theta$, then it will not move in a circular path, but it will move in a helical path as shown in Fig. 3.8a) along OX and pitch of the helix ($l = OR$) will be the distance travelled by the electron parallel to the magnetic field by the time t . So

$$\begin{aligned} l &= v \cos \theta \times t \\ &= v \cos \theta 2\pi/B(e/m) \\ l &= 2\pi mv \cos \theta / Be. \end{aligned} \quad (3.29)$$

Explanation for Magnetostatic Focusing

From equation (3.28) it is clear that t is independent of θ , v and r , so the electrons starting from point O can diverge in any direction, with different velocities but will be forced to focus by the uniform field at the point R ($OR = l$) after particular time interval t . OP gives the direction of motion in the absence of the magnetic field. Thus the magnetic field just acts as a lens to bring the divergent electron beam into a focus. From equation (3.29) it is found that focal length l ($= OR$) depends on v and $\cos \theta$. But if velocity v is large, divergence θ also will be large, but $\cos \theta$ will be small. So, again focal length l will be constant for a

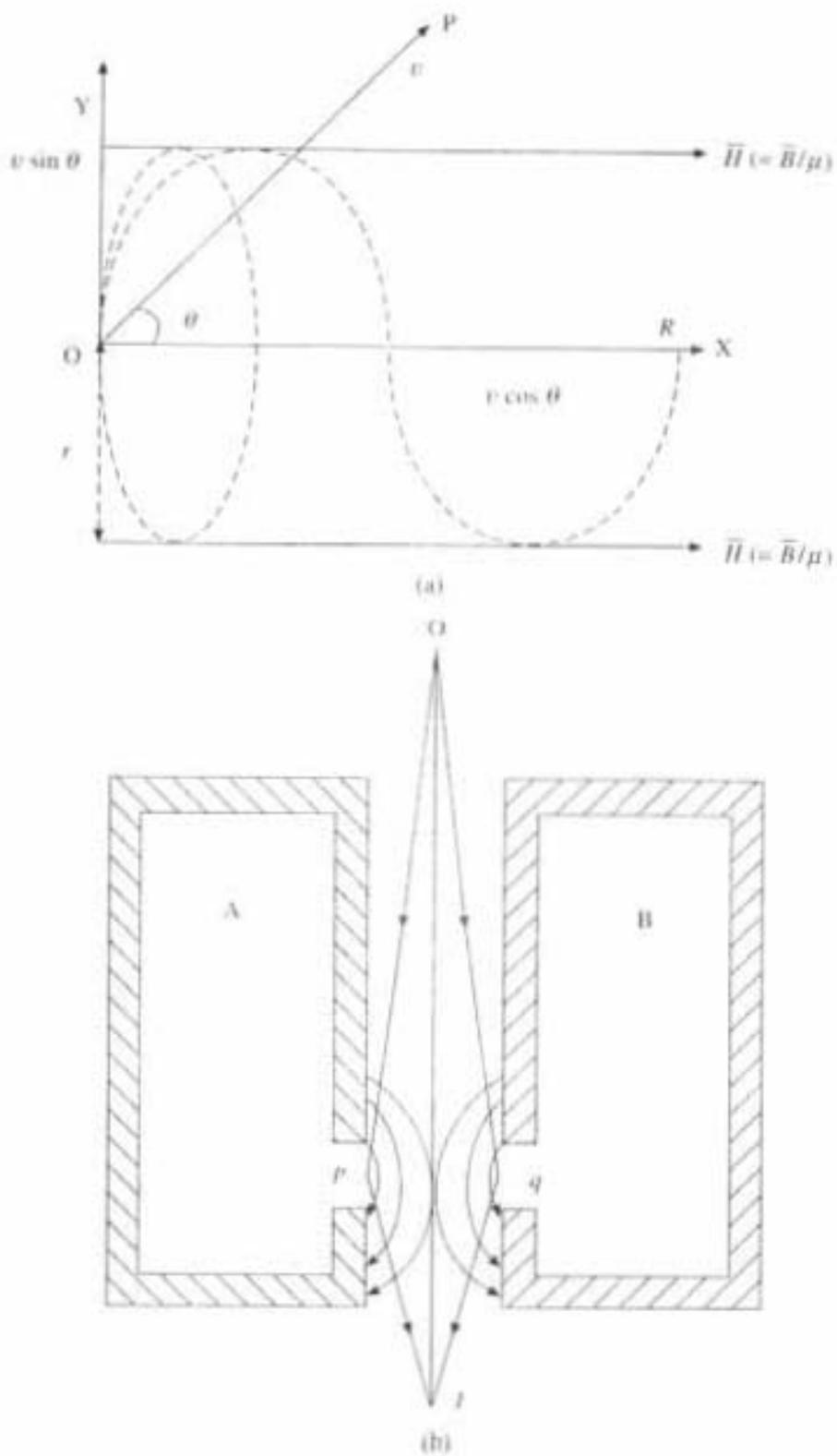


Fig. 3.8 Magnetostatic focusing

particular value of \bar{B} . Focal length $OR = l$ can be altered only by changing the magnetic induction B . If the electrons make two or three rotation before reaching R then more than one focal point will occur at distances $l, 2l, 3l$, etc. Thus the focal length of magnetic lens can be varied by varying the intensity of the magnetic field (\bar{B}), which is not possible for optical lens.

Different types of magnetic lens are possible. "Busch" first proposed that a non-uniform magnetic field along the axis of a short solenoid also has the properties of a magnetic lens. In Fig. (3.8b) AB is the section having soft iron shielded coil of electromagnet with a narrow gap in the form of a ring. p and q represent the gaps at diametrically opposite positions, and there is a strong magnetic field across the gap which is symmetrical about the axis OI of the electromagnetic coil. Arrangement is made in such a way that O is source of electron which are focused to the point I due to magnetic lens (Fig. 3.8b). The focal length of magnetic lens is of very small order of few cm. This is the required criteria for getting large magnification in *Electron Microscope*, where an intense and very fast moving beam of electrons are focused and mainly magnetic lens focusing system is used. Details of Electron Microscope and magnetic lenses are discussed in section 4.12.

3.10 Cathode Ray Oscilloscope (C.R.O.)

One of the important application of the electrostatic focusing and motion of the charged particles in electric and/or magnetic field is cathode ray oscilloscope. This oscilloscope is one of the most important tools for the development of electronic circuits and for modern electronics. C.R.O. is a device that allows us to display the amplitude of electrical signals. Whether they be voltage or current, they can be displayed primarily as a function of time on C.R.O. screen. It depends on the movement of an electron beam, which will be visible by allowing the beam to impinge on a phosphor surface. If electron beam is deflected in either of two orthogonal axes X and Y, the luminous spot can be used to create two-dimensional displays. Along X-axis, spot is deflected at a constant rate relating to time, whereas a long Y-axis spot is deflected in response to an input stimulus such as voltage. This produces time-dependent variation of the input voltage, which is very important for the design and development of electronic circuit.

3.10.1 Construction of C.R.O.

Construction and working of C.R.O. are shown in Fig. 3.9(a) and (b). The electron beam is generated from a conventional heated cathode surrounded by the control grid. Grid is maintained at negative potential w.r.t. cathode to give focusing effect. After grid, the first accelerating anode and focus electrode follow which provide electrostatic focusing, as well as fast accelerating voltage. Beyond the focusing electrode, the electron beam passes through two vertical deflection plates (Y-plates). The beam at this point is not fully focused. The beam will be further focused after deflection to provide a fine spot. After vertical deflection the beam passes through a scan expansion lens (focus element) that increases the amount of beam bending in the vertical plane. This beam is then deflected by two horizontal plates (X-plates) and passes through another electron lens which provides additional focusing. The beam is accelerated to the final velocity by a quadrupole scan expansion lens, which provides not only an increase in electron velocity, but adds to the scan angle without distorting or defocusing the electron beam. Now the electron beam impinges on the C.R.O. screen which is coated with a layer of fluorescent material, phosphor.

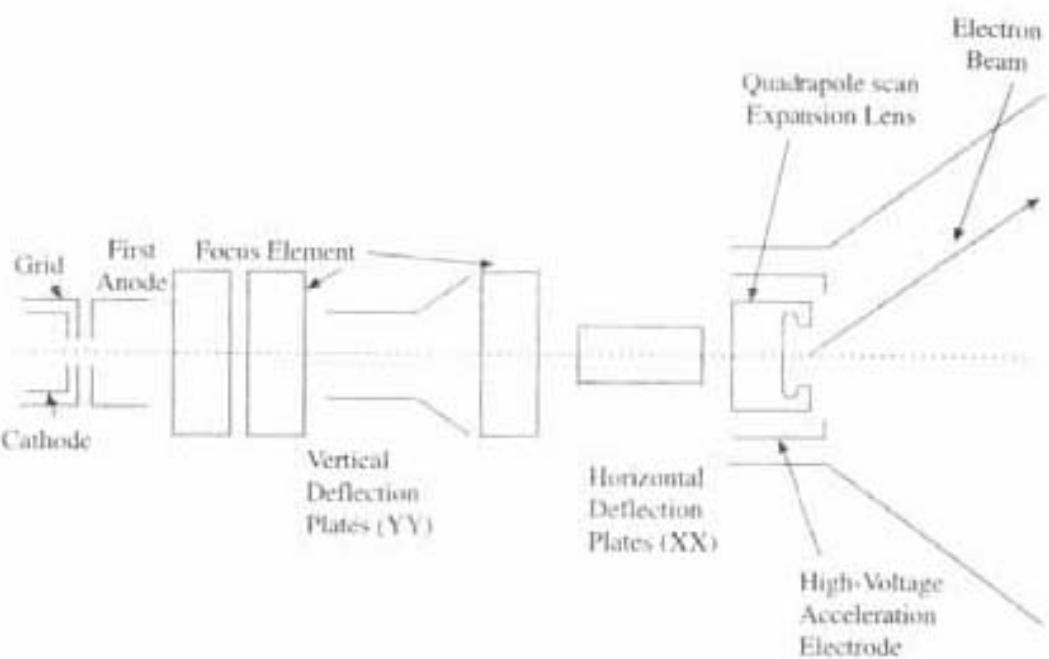


Fig. 3.9(a) Internal components of C.R.O.

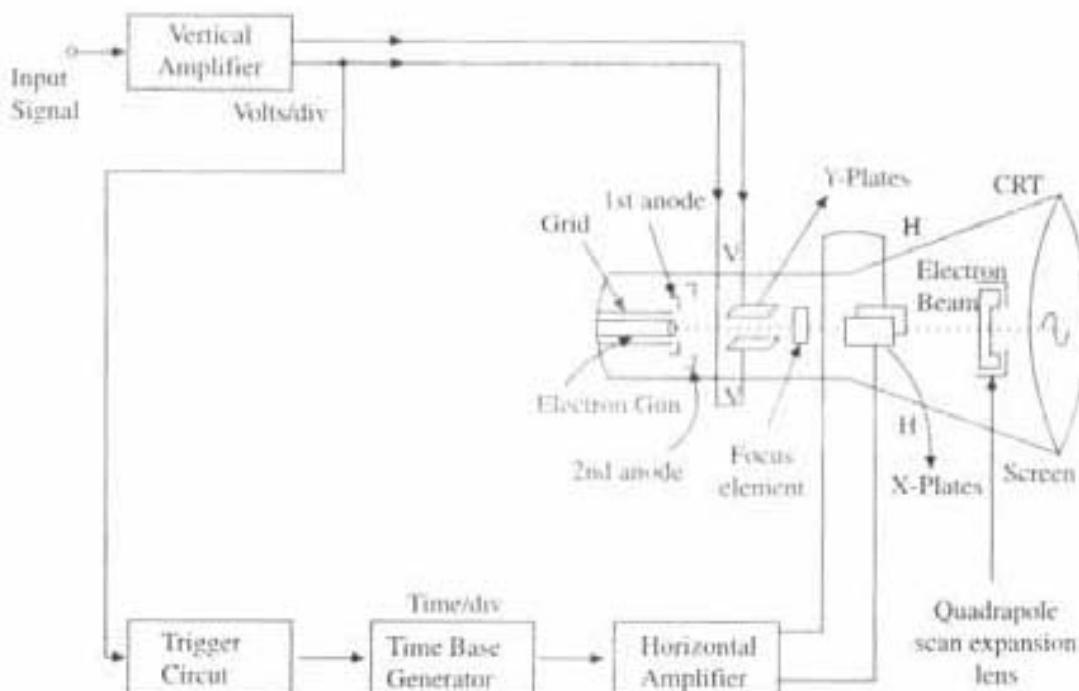


Fig. 3.9 (b) Working procedure of C.R.O.

The amount of light given off by the phosphor screen depends on the amount of energy that is transferred to phosphor by the electron beam. For a fast oscilloscope it is desirable to accelerate the electron beam to the greatest amount possible, while on the other hand a greater electron beam velocity will make it more difficult to focus the beam. So main purpose in the construction of good C.R.O. is the matching between these two. First the electron beam is accelerated, then focused and then deflected and, after deflection is further accelerated to the desired final velocity.

3.10.2 Working of C.R.O.

The heart of the oscilloscope is the cathode ray tube (C.R.T.). If a steady p.d., i.e., D.C. voltage is applied to X-plates Fig. 3.9(b), the beam is deflected along X-axis. If the d.c. voltage is applied to Y-plates the beam will deflect along Y-axis. The distance between the centre and the new position of the spot is proportional to the D.C. voltage applied.

But instead of D.C. voltage, if an a.c. voltage is applied on the vertical deflecting plates, then the spot of light is drawn out as a vertical line and the length of the line is proportional to the maximum value of A.C. voltage.

For displaying exact wave form of the A.C. signal on the C.R.O. screen, horizontal deflection should be synchronized with the vertical input, such that horizontal deflection starts at the same point of the input vertical signal Fig. 3.10(b) where the time period T of the sweep voltage is equal to or is some multiple of the time period of the A.C. signal on the vertical plates. For this, a synchronizing or triggering circuit is used as shown in Fig. 3.9(b). This circuit is the link between the vertical input and the horizontal time base generator.

Vertical deflection system

The function of the vertical deflection is rather straight-forward; it must provide an amplified signal to a proper level that will provide a usual deflection of the electron beam, without introducing any appreciable distortion into the system.

The order of amplification in a vertical side adjusted by attenuator, Volts/div. knob is in the common 1-2-5 sequence. Say for example 1, 2, 5 volt, 10, 20, 50 mV etc. per centimeter.

Horizontal deflection system

It is the general purpose of most C.R.O. to deflect the spot horizontally at a constant rate relative to time, which is often referred to as linear sweep. Since the voltage used across horizontal plates is used to sweep the electron beam across the screen, it is called as sweep voltage. The horizontal deflection system consists of time base generator, a trigger circuit, and a horizontal amplifier (Fig. 3.9(b)). The time base generator or sweep generator controls the rate at which the beam is scanned across the screen of the C.R.O. and is adjusted from the front panel by adjusting time/div. knob. The trigger circuit, ensures that the horizontal sweep starts at the same point of the vertical input signal as in Fig. 3.10(b). The horizontal amplifier is similar to vertical amplifier.

Time base generator

The sweep generator uses the charging characteristic of a capacitor C to generate linear rise time voltages to feed the horizontal amplifier. Figure 3.10(a) shows a capacitor of capacitance C being charged from a constant current source. The rate of voltage rise is

$$\frac{\text{Change of voltage}}{\text{Time}} = \frac{I}{C}$$

Thus, there is a gradual increase of voltage across C , when the critical voltage is reached, the capacitor returned to zero voltage quickly by discharging through

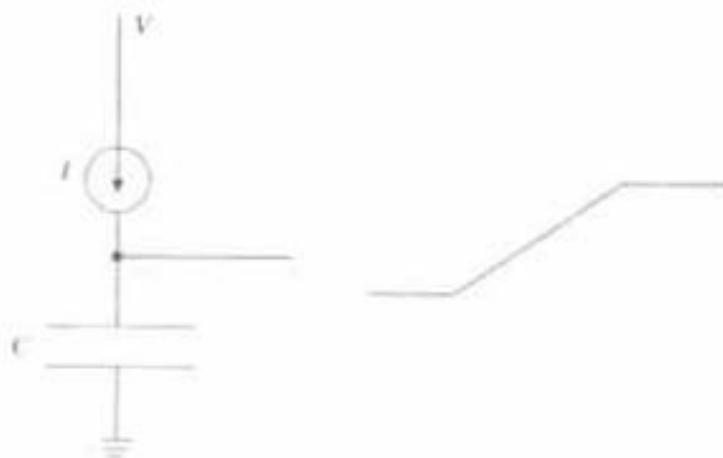


Fig. 3.10 (a) Working procedure of time base generator

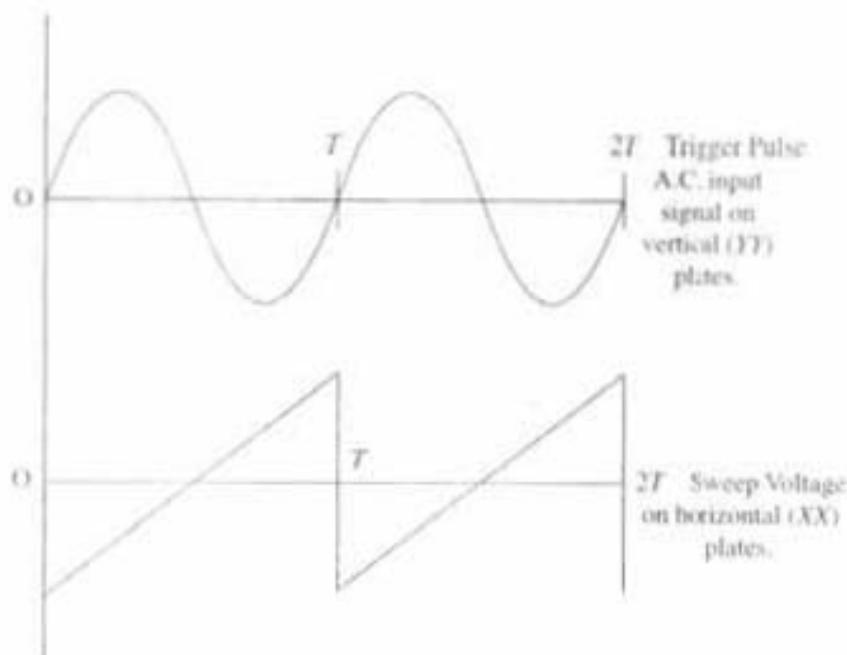


Fig. 3.10 (b) Synchronisation of trigger pulse and sweep voltage

the transistor (Fig. 3.10b). So the electrons are deflected linearly with time in the horizontal direction for a time T . Then the beam returns to its starting point (Fig. 3.10b) on the screen very quickly, as the sawtooth voltage rapidly falls to its initial value at the end of each period.

Because the rate of increase of voltage in capacitor can be varied by adjusting either current I or capacitance C , so the sweep rate (time period T) can be controlled. The sweep generator or time/base knob follows the same 1-2-5 sequence as in the vertical system, as 0.1, 0.2, 0.5 ms, 10, 20, 50 microsec etc. per cm.

The relationship between the sweep generator and the trigger pulse, which represent the same point of the input wave form is shown in Fig. 3.10(b). If a sinusoidal voltage is impressed across the vertical deflection plates (YY) then, simultaneously, the sweep voltage is impressed across the horizontal deflection plates (XX) through time base generator. The sinusoidal voltage, which itself would give rise to a vertical line, will now be spread out and will appear as a sinusoidal

trace on C.R.O. screen. The pattern will appear stationary only if the time T is equal to or is of some multiple of, the time for one cycle of the wave on the vertical plates. It is then necessary that frequency of sweep circuit be adjusted to synchronise with the frequency of applied input signal.

3.10.3 Uses of C.R.O.

C.R.O is used as a tool for display of the waveforms as well as for making many kinds of measurement quickly and almost accurately. It is not particularly a very accurate measuring instrument but is best used where only an approximate and quick measurement is required. Its great value lies in its versatility and applicability to measure a wide range of A.C. and D.C. voltage. As a test instrument, it is often required to measure voltage whose frequency and magnitude are totally unknown.

In the C.R.O's most common mode of usage of measuring time varying signals, the unknown signal is applied to the vertical (YY) plates and a time base to the horizontal (XX) plates. In this mode of operation, the display on the C.R.O. screen is in the form of a graph with amplitude of the unknown signal on the vertical axis and time on the horizontal axis.

The two basic measurements one can make with C.R.O. are amplitude and time. All most all other measurements are based on these two measurements. So the following parameters can be measured by C.R.O. are as follows.

3.10.3.1 Measurement of A.C./D.C. Voltage

C.R.O. is a voltage-measuring device. Since voltage is shown as an amplitude on C.R.O. screen for A.C. signal along vertical axis (Fig. 3.11), which can be controlled by volts/div. knob. So amplitude of unknown A.C. voltage is

$$\text{Amplitude (V)} = \text{Amp.} \times \text{Volts/div. Knob position} = \text{cm.V/cm} = \text{Volts} \quad (3.30a)$$

In case of D.C. voltage, the distance between the spot before and after the D.C. voltage applied is proportional to the applied voltage, which is also controlled by volts/div knob.

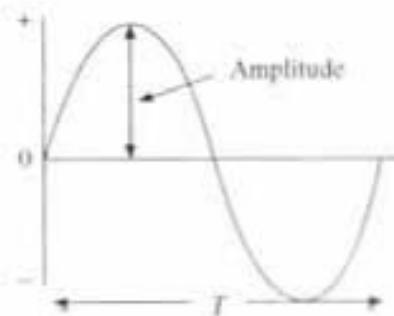


Fig. 3.11 Measurement of Amplitude

3.10.3.2 Measurement of Frequency

(a) C.R.O. is also commonly used for obtaining an approximate measurement of frequency especially in circuit test and fault diagnosis applications. Time period for a particular signal can be measured on C.R.O. screen along horizontal axis (Fig. 3.11), which can be controlled by time/div. knob, i.e., by time base generator.

Hence time period of unknown a.c. signal is

Time period (T) = Distance between two successive cycles \times Time base knob

Position = cm. sec/cm = sec.

$$\text{Frequency} \quad (f) = 1/T = \text{sec.} = \text{Hz} \quad (3.30b)$$

The alternative way of using C.R.O. to measure frequency is to generate Lissajous patterns. These are produced by using C.R.O. in $X-Y$ mode, where one can bypass the time-base, and apply a known reference frequency sinusoidal signal directly to vertical (Y) plates and an unknown frequency a.c. sinusoidal signal to the horizontal (X) plates. The resulting Lissajous pattern Fig. 3.12a, on the screen depends on the frequency ratio of two signals. Pattern will be stable only when the frequency ratio of two signals is a whole number. So steady pattern can be obtained by adjusting the reference frequency and then unknown frequency can be calculated according to the frequency ratio for the particular obtained pattern, according to Fig. 3.12a.

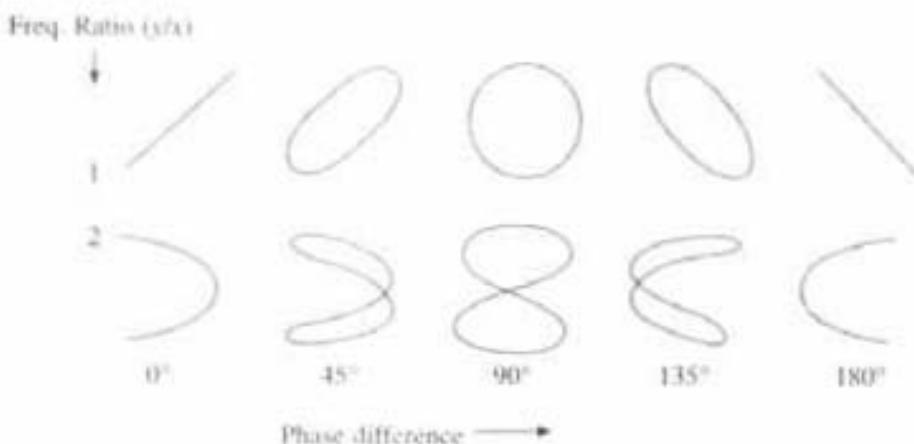


Fig. 3.12(a) Lissajous Pattern

3.10.3.3 Measurement of Phase-difference

(a) When two sinusoidal waveform started at different time, then a phase difference or phase shift will come between them (Fig. 3.12b), which is expressed in terms of degree. Phase shift between two a.c. signals of same amplitude, same frequency

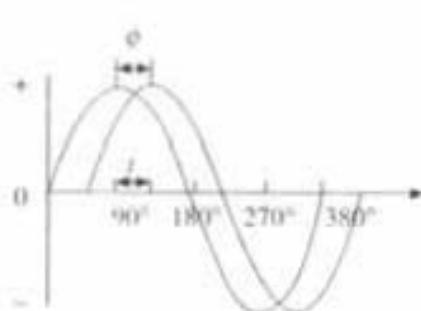


Fig. 3.12 (b)

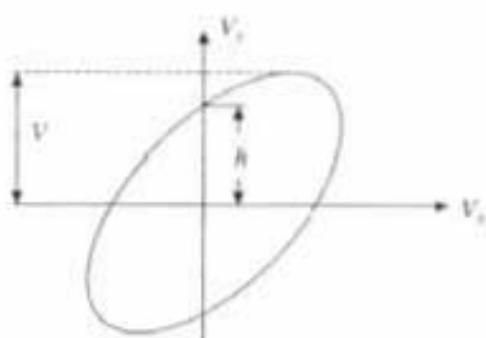


Fig. 3.12 (c)

Measurement of Phase-difference

can be measured, with dual trace C.R.O. when two a.c. signal with a phase difference (ϕ) are applied in two different channel.

If the horizontal distance between the same points of the two waveforms from the Fig. 3.12b is (in time) 't', then phase difference between them will be

$$\text{Phase shift} = \phi = (t/T) 360 \text{ degree} \quad (3.30c)$$

When T = Time period of the wave.

(b) Phase shift between two-a.c. signals can be calculated also by Lissajous pattern (Fig. 3.12a), by using C.R.O. in $X-Y$ mode. When two sinusoidal signal with same amplitude and same frequency but with a phase difference (ϕ) are applied to the X and Y plates of $X-Y$ plotter, the plot obtained is in general ellipse as shown in Fig. 3.12a, 3.12c. If X and Y a.c. input signals are given as

$$V_x = V \sin \omega t$$

$$V_y = V \sin (\omega t + \phi)$$

Then at $t = 0$, $V_x = 0$, $V_y = V \sin \phi$

Now from the Fig. 3.12c, when $V_x = 0$, $V_y = h$

Hence by equating above two equations for V_y , we get

$$V \sin \phi = h$$

or

$$\sin \phi = h/V$$

$$\text{Thus} \quad \text{Phase difference} = \phi = \sin^{-1} (h/V) \quad (3.30d)$$

As a special case, when $h = V$, then phase shift $\phi = 90^\circ$ (3rc Lissajous pattern, in Fig. 3.12a, which is a circle i.e. $h = V$).

Now one can calculate phase difference between any two a.c. sinusoidal signals by observing Lissajous pattern, when C.R.O. is on $X-Y$ mode.

3.11 Cyclotron

The magnetic resonance accelerator or cyclotron is the most famous of all accelerators which accelerates charged particles to as high a value as 40 MeV, which are homogeneous in energy and emerge as a parallel beam. These highly energetic particles can be directly used as projectile in a nuclear reaction.

The cyclotron consists of two flat semicircular metal boxes, called 'dees' or D 's because of their shape. These hollow chambers have their diametric edges parallel and slightly separated from each other (Fig. 3.13). A source of ions is located near the mid-point of the gap between the dees. The dees are connected to the terminals of a radio frequency oscillator. So that a high frequency alternating potential of several million cycles per second is applied between the dees which act as electrodes. In this way, the potential between the dees is made to alter rapidly and the electric field in the gap is directed first towards one dee and then towards the other. The dees are enclosed within but insulated from a larger metal box containing gas at low pressure and the whole apparatus is placed between the poles of a strong electromagnet which provide a magnetic field perpendicular to the plane of the dees. Electric field inside dees is zero because of electrical shielding. When an ion with a positive charge q is emitted from the source, it is accelerated by the electric field between the dees towards the dee which is

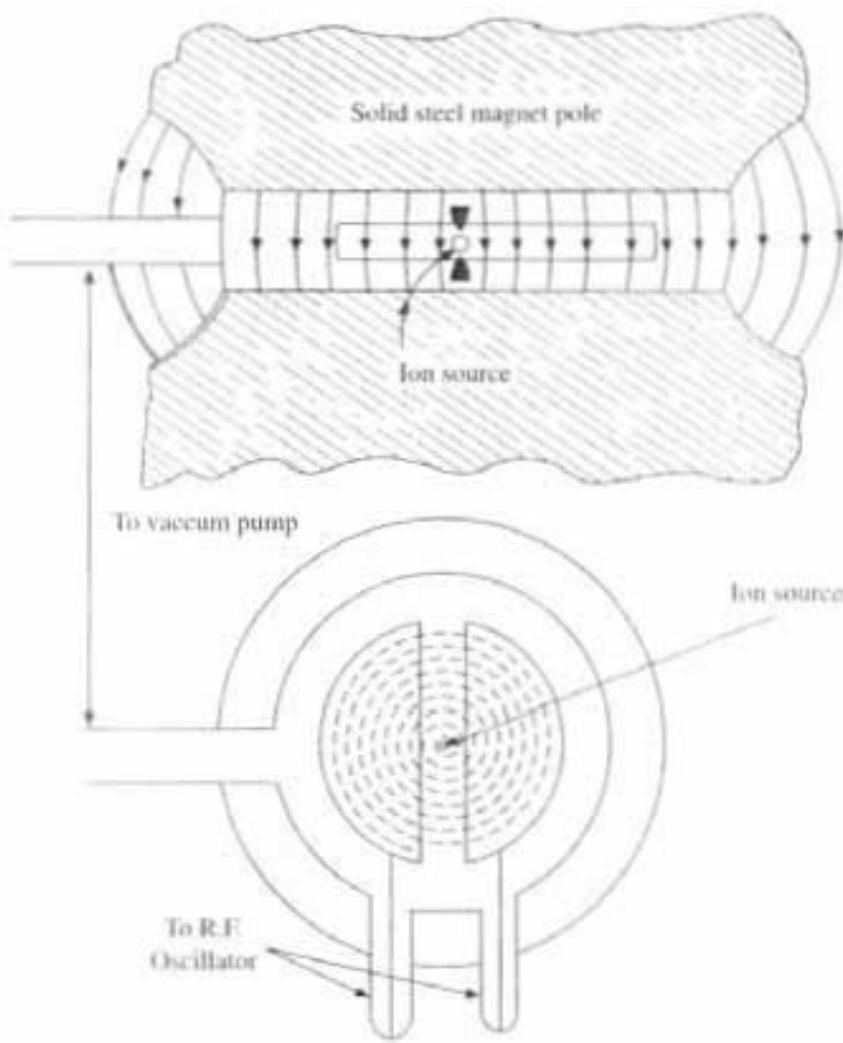


Fig. 3.13 Cyclotron working procedure

negative at that instant. Since there is a uniform magnetic field B acting at right angles to the plane of the dees, the ion travels in a circular path within the dee, of radius r_1 , given by from eqn. (3.17) as

$$r_1 = mv_1/Bq$$

where m is the mass of the ion, B the strength of the magnetic field, and v_1 speed of the ion. While the ion is inside the dee, its speed stays constant, but after describing a semicircle through the dee, the ion reaches the gap, where it is again subjected to the applied potential difference. Now if the time taken by the ion to complete the semicircle is equal to the time taken by the oscillator to change its polarity, then the ion is again accelerated towards the other dee with an increased speed v_2 . It then moves in a semicircle with larger radius r_2 . Thus by adjusting the oscillation frequency and strength of the magnetic field, the charged particles always keep in phase with the changes of electric field polarities and then the speed of the ion goes on increasing as it is moving from one dee to the other in a semicircle of larger and larger radius. Each time the ions cross the gap, they receive an additional kick, so the energy will steadily increase until it will reach the periphery of the dee. Now they can be brought out of the chamber and can be

used directly as highly energetic particles in the study of nuclear reactions and in the production of radio nuclides.

During this rotation angular velocity of the ion will remain same

$$\omega = v/r = Bq/m$$

The maximum energy of the ion produced by a cyclotron can be calculated in the following way. Due to circular motion of the ion from eqn. (3.17) we get

$$Bqv = mv^2/R$$

or

$$v = \frac{BqR}{m}$$

where R is the radius at which the particles leave the machine. When an ion has been accelerated by p.d. V , then energy is

$$qV = 1/2 mv^2$$

So the energy of the outgoing particle from cyclotron is

$$E = 1/2 mv^2 = \frac{B^2 q^2 R^2}{2m} \quad (3.31a)$$

The value of the magnetic field strength cannot be chosen arbitrarily but depends on the frequency of the applied alternating voltage to the dee, and must be chosen so as to give resonance. *For resonance, the time taken by an ion, with an angular velocity ω , to describe a semicircle within the dee must be equal to half the time period T of oscillation of the applied A.C. field, that is*

$$\pi r = vT/2$$

or

$$\pi/v = T/2 \quad (\because \omega = v/r)$$

and the Resonance Frequency

$$n = \frac{1}{T} = \frac{Bq}{2\pi m} \quad (3.31b)$$

Due to this resonance, cyclotron is also called as Magnetic Resonance Accelerator. *Unless resonance will occur, the acceleration process in cyclotron will stop. So this resonance frequency (n) is the most important parameter in a cyclotron.*

From Eqn. (3.31a) we see following things:

(1) The energy acquired by the particle is independent of the voltage applied. When the voltage is small, it makes large number of turns to reach the periphery. It is the property of the cyclotron to accelerate charged particles to relatively high energy by means of small applied voltage.

(2) It is also observed that for same value of magnetic field (B), if the radius of the dees is doubled, energy of the ion will be increased by four times.

(3) The energy acquired by the particle is inversely proportional to its mass. So for massive particles like α -particle, higher magnetic field strength is necessary.

So for accelerating different particle, different type arrangement is necessary.

Mass-Spectroscope

Uses of Mass spectrometers fall into two main divisions. For accurate determination of atomic masses there is the high precision *Mass-Spectrograph* and for the accurate comparison of relative isotopic abundances of stable elements there is *Mass Spectrometer*.

The +ve ions of charge (q) emerging from the exit slit of the ion sources may have different energy depending upon the type of the source and a range of masses of isotopes (M) present. The mass spectrograph must be able to focus ions of the same mass at the same place on the detector i.e. either photographic plate or electrical detector. It is vital that the instrument is capable of focusing ions of different energies (velocity focusing) and different direction of motion (space focusing) at the same point or line on the detector. This combination of *double focusing* is a characteristic of all the best instruments but it is not possible to achieve good focusing for a wide range of ionic energies and directions. Here we will discuss (1) Aston and (2) Bainbridge mass spectrograph.

3.12 Aston-Mass Spectrograph

For detailed analysis of the isotopic masses Aston designed a mass spectrograph in 1919. The schematic diagram for the Aston mass-spectrograph is as shown in Fig. 3.14. Positive rays from the source are made to pass through two very thin parallel slits S_1, S_2 . This fine collimated beam containing particles of wide range of velocities is subjected to perpendicular electric field E_x , maintained between the plates $P(+)$ and $P(-)$. Due to perpendicular electric field positive ions not only deflect by an angle θ towards negative plate but are also dispersed, by an angle $d\theta$. (Fig. 3.14), because the amount of deflection will be greater for the positive ions which is having smaller velocity, than the particles having higher velocity. Since smaller velocity particles are staying longer time in the electric field. The divergent beam is now limited by passing through slit D and enters into a perpendicular magnetic field B_z indicated by circle perpendicular to the plane of the paper. This magnetic field B_z is of such sign that it bends the beam in a direction opposite to that caused by electric field E_x . It will again not only deflect the beam in opposite direction by an angle ϕ but also make it to converge.

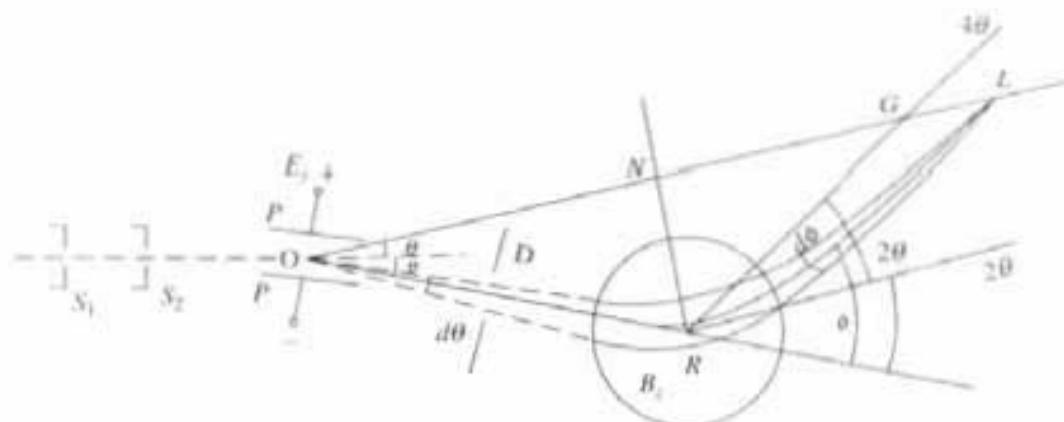


Fig. 3.14 Aston Mass Spectrograph

The slowly moving ions being deflected more than the faster ones, by an angle $d\phi$. The path of the slow-moving ions, therefore, intersect those of the faster moving ions at point L . Thus if the instrument is properly designed, ions having same value of q/M but slightly different energies can be brought to a single focus at L on a photographic plate, other ions with different value of q/M are brought to a focus at different points on the photographic plate. So the isotopes of the same element which are having same charge q but different atomic masses M will focus at different points on the photographic plate. Since the ions having same value of q/M but of slightly different velocity are brought to focus at a point, hence this method is called velocity focusing method. Actually for a particular values of q/M the focus is a line not a point. Thus, for different value of q/M , photographic recording will be series of lines similar to optical line spectrum. Due to this similarity the series of lines for different value of q/M , it is called Mass-spectrum and the apparatus as an Mass-spectrograph. The mass-spectrum of germanium is shown in Fig. 3.15. Ge 72, Ge 73 etc. The relative abundance of the isotopes is measured from the densities of the photographic images they produced.

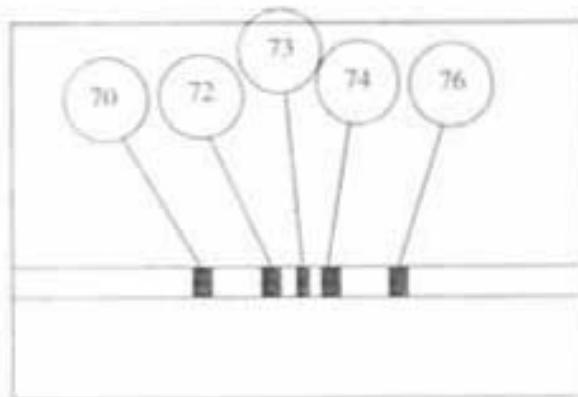


Fig. 3.15 Relative abundance of different isotopes of Ge

Condition for focusing

The condition for such velocity focusing can be derived as follows. Let us consider a beam of group of ions having same value of q/M but moving with different velocities. Let θ and Φ be the mean angle of deviation by the electric and magnetic fields and $d\theta$ be the dispersion angle due to electric field and $d\Phi$ be the convergence angle due to magnetic field.

If S is the linear displacement of the ions and L is the length of their path in the electric field (Fig. 3.14) then,

$$\theta = S/L \quad (3.32)$$

But

$$S = \frac{E_y \times qL^2}{2Mv^2}$$

$$\left[\because S = \frac{1}{2} a_y t^2 \text{ when } a_y = \frac{qE_y}{M}, t = \frac{L}{v}, \text{ So, } S = \frac{1}{2} \frac{qE_y L^2}{Mv^2} \right] \quad (3.33)$$

$$\text{So } \theta = \frac{S}{L} = \frac{E_i q L}{2 M v^2}$$

$$\text{or } \boxed{\theta = C \times q / M v^2} \quad \text{where } \boxed{C = (1/2) \times E_i \times L} \quad (3.34)$$

Similarly if S' is the linear displacement and L' is the length in the magnetic field then,

$$\phi = S' / L' \quad (3.35)$$

$$\text{But } S' = \frac{B_e q L'^2}{2 M v}$$

i.e., Electric field effect is cancelled by magnetic field effect, so

$$q E_i = B_e q v$$

$$\text{i.e., } E_i = B_e v]$$

Now by analogy with Eqn. (3.33)

$$S' = \frac{1}{2} \left[\frac{B_e q v L'^2}{M v^2} = \frac{B_e q L'^2}{M v} \right]$$

$$\text{So, } \phi = \frac{S'}{L'} = \frac{B_e q L'}{2 M v}$$

$$\text{or } \boxed{\phi = \frac{D q}{M v}} \quad \text{where } \boxed{D = \frac{1}{2} B_e L'} \quad (3.36)$$

C and D are constants depending on the strength and distribution of the electric and magnetic fields.

Now the rate of change of angular dispersion w.r.t. velocity in electric fields is from Eq. (3.34)

$$\frac{d\theta}{dv} = - \left(\frac{2 C q}{M v^3} \right) = -2 \frac{\theta}{v}$$

and that in magnetic field is from Eqn. (3.36)

$$\frac{d\Phi}{dv} = - \left(\frac{D q}{M v^3} \right) = - \frac{\Phi}{v}$$

$$\text{So } \boxed{\frac{d\theta}{\theta} = 2 \frac{d\Phi}{\Phi}} \quad (3.37)$$

Let us suppose the electric and magnetic fields acts as if they concentrate at O and R respectively. From the point O the beam begins to diverge and from the point R divergent beams begin to converge, and let $OR = a$ and $RL = b$, where L is the focus.

Now if magnetic field is not present, then the dispersions produced by electric field at a distance $(a + b)$ is given by $(a + b) d\theta$. As this divergence is canceled by magnetic field acting at R , so that the group of beams are brought to a focus at L at a distance b from R . So the condition of focusing is given by

$$(a + b) d\theta = bd\phi$$

$$(a + b)/b = d\phi/d\theta = \phi/2\theta = 1 + a/b$$

So

$$b/a = 2\theta/(\phi - 2\theta) \quad (3.38)$$

This is the equation of a straight line OL drawn from O making an angle 2θ with the direction of the beam deviated by the electric field (Fig. 3.14). From Eqn (3.38) one can find the most convenient position of the photographic plate to record the focus of the convergent beams. When $\phi = 4\theta$, $b = a$ i.e. $OR = RL$ from Eqn. (3.38), at that position, say at point G where photographic plate can be placed, on the line OL , such that $GN = NO$. N is the foot of the perpendicular drawn from R on the line OL . At G , ϕ is being equal to 4θ as shown in Fig. 3.14, which can be taken as a point of reference. The photographic plate containing this point of reference G will not only receive the trace of the selected group of ions of the same value of q/M , but also be in fair focus for ions of different values of q/M over a range (Fig. 3.15).

3.13 Bain Bridge's Mass Spectrograph

The schematic diagram of Bain bridge's mass-spectrograph is as shown in Fig. 3.16, has a source of positive ions of charge q , situated above the slit S_1 . The

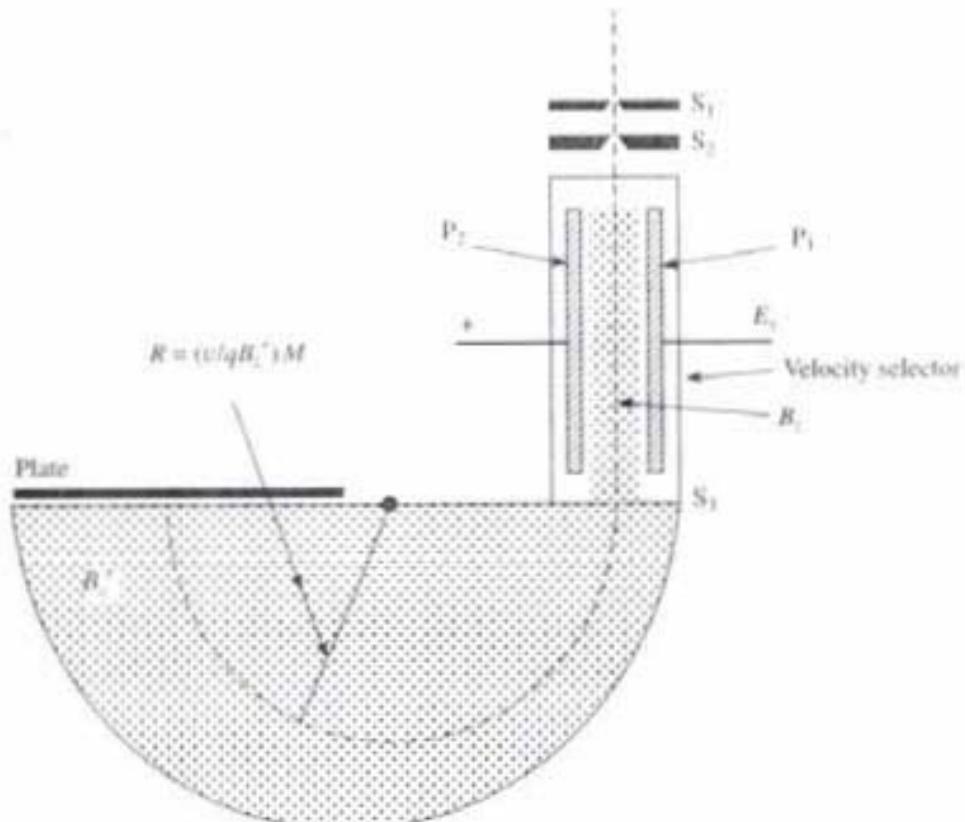


Fig. 3.16 Bain-Bridge's mass spectrograph

ions under study pass through the slits S_1 and S_2 and move down into the electric field (E_z) between two plates P_1 and P_2 . In that region there is also magnetic field of induction B_z , dotted area, perpendicular to the paper. Thus the ions enter in a region of crossed electric and magnetic fields, and if force acting by both fields on the ions are equal, then

$$E_z q = B_z \cdot qv$$

$$v = E_z / B_z$$

(3.39)

i.e. those ions whose velocity, $v = E_z / B_z$, only they pass undeviatedly through this region, but ions with other velocities are stopped by the slit S_3 . Thus, all the ions which emerge from S_3 have the same velocity v . This region of crossed field is called a *velocity selector*. Below S_3 the ions enter a region of another magnetic field B_z' , perpendicular to the paper but with no electric field. Here ions move in a circular path of radius R such that

$$B_z' qv = Mv^2/R$$

$$M = \frac{B_z' qR}{v} = \frac{B_z B_z' qR}{E_z}$$

(3.40)

i.e.

$$R \propto M$$

Assuming equal charges on each ion, since B_z' and v are same for all ions so the masses of the ions are proportional to the radii of their paths. So the ions of different isotopes are converged at different points on the photographic plate. The selective abundance of the isotopes is measured from the densities of the photographic images they produced. Figure 3.15 shows the mass spectrum of germanium, Ge 70, Ge 72, Ge 73, Ge 74, Ge 76.

PROBLEMS

1. An α -particle enters in a homogeneous magnetic field perpendicular to its velocity. The angular momentum of α -particle is now 1.33×10^{-22} kg. m²/sec. The induction of magnetic field is 0.025 Wb/m². Find K.E. of the particle. Given that the mass of α -particle = 6.68×10^{-27} kg and charge of α -particle = 3.2×10^{-19} Coulomb.

$$\text{Angular momentum} = L = mr^2w = 1.33 \times 10^{-22} \text{ kg. m}^2/\text{sec} = Iw$$

where w = angular velocity = $v/r = B \cdot q/m$

B = magnetic induction = 0.025 Wb/m²

q = charge of α -particle = 3.2×10^{-19} C

m = mass of α -particle = 6.68×10^{-27} kg

1eV = 1.6×10^{-19} J

I = Moment of inertia

$$E = \frac{1}{2} Iw^2 = \text{K.E.} = \frac{1}{2} (Iw) w$$

Angular momentum = $L = I \cdot w = 1.33 \times 10^{-22} \text{ kg.m}^2/\text{sec}$

$$E = \frac{1}{2} \times 1.33 \times 10^{-22} \times \frac{2.5 \times 10^{-7} \times 3.2 \times 10^{-19}}{6.68 \times 10^{-27}}$$

$$E = 0.796 \times 10^{-16} \text{ J} = 497.7 \text{ eV}$$

2. For a cyclotron of 0.35 m radius and freq. of potential applied is $1.38 \times 10^7 \text{ Hz}$, obtain the value of induction of magnetic field for synchronous acceleration of protons, and hence calculate the max. energy of the accelerated protons. Mass of protons = $1.67 \times 10^{-27} \text{ kg}$, charge of protons = $1.6 \times 10^{-19} \text{ Coulomb}$.

$$\text{Frequency of oscillation} = n = \frac{Bq}{2\pi m}$$

$$(i) \text{ So } B = 2\pi n \cdot m/q$$

$$B = \frac{2 \times 3.14 \times 1.38 \times 10^7 \times 1.67 \times 10^{-27}}{1.6 \times 10^{-19}} \\ = 0.045 \times 10^{-1}$$

$$B = 0.045 \text{ wb/m}^2$$

$$(ii) \text{ Maximum energy of proton} = \frac{1}{2} mv^2 = \frac{B^2 q^2 R^2}{2m} = E$$

$$\text{So } E = \frac{(0.045)^2 (1.6 \times 10^{-19})^2 (0.35)^2}{2 \times 1.67 \times 10^{-27}}$$

$$1 \text{ eV} = 1.6 \times 10^{-19} \text{ J}$$

$$E = 0.0768 \times 10^{-11} \text{ Joule} = 0.048 \times 10^6 \text{ eV}$$

$$E = 4.8 \text{ MeV}$$

3. (a) An electron accelerated by a potential diff. of 1000 volts enters at right angles into a uniform magnetic field of induction $1.19 \times 10^{-3} \text{ Wb/m}^2$. Find (i) the radius of the electron trajectory in the magnetic field. (ii) The angular momentum of the electron.

Mass of electron = $9.1 \times 10^{-31} \text{ kg}$

Charge of electron = $1.6 \times 10^{-19} \text{ Coulomb}$

Due to the p.d. V electron is accelerated.

$$\text{So } eV = 1/2 mv^2 \quad (1)$$

Due to transverse magnetic field B electron moves in a circular path of radius R .

So

$$mv^2/R = BeV \quad (2)$$

where B = magnetic field = $1.19 \times 10^{-3} \text{ Wb/m}^2$

m = mass of electron = $9.1 \times 10^{-31} \text{ kg}$

V = p.d. = 1000 volt

e = charge of electron = $1.6 \times 10^{-19} \text{ C}$

From (1) $v = \sqrt{2eV/m} = 18.752 \times 10^6 \text{ m/sec}$

From (2) $R = mv/Be$

$$\text{So } R = \frac{9.1 \times 10^{-31} \times 18.752 \times 10^6}{1.19 \times 10^{-19} \times 1.6 \times 10^{-19}} = 89.62 \times 10^{-3} \text{ m} = 8.962 \text{ cm}$$

Angular momentum = $L = mR^2\omega = mR^2v/R$

$$L = mvR = 15293.043 \times 10^{-28} \text{ kg m}^2/\text{sec.}$$

4. An electron moving horizontally with a velocity of $1.7 \times 10^7 \text{ m/sec}$ enters a vertical uniform electric field of $3.4 \times 10^4 \text{ volts/meter}$ acting downwards. Find the vertical displacement of the electron in the electric field at a time when its horizontal displacement is 3 cm. Find also the magnitude and direction of the magnetic field required to neutralize this vertical displacement.

Charge of electron = $1.6 \times 10^{-19} \text{ C}$

Mass of electron = $9.1 \times 10^{-31} \text{ kg}$

$v_x = 1.7 \times 10^7 \text{ m/sec}$

$E_y = 3.4 \times 10^4 \text{ v/m}$

Distance travelled along X-axis by this time

$$t = v_x / x, t = x/v_x = \frac{3 \times 10^{-2}}{1.7 \times 10^7} = 0.176 \times 10^{-8} \text{ sec}$$

Distance travelled along Y-axis by the time $0.176 \times 10^{-8} \text{ sec}$ is

$$y = 1/2 a_y t^2$$

$$a_y = eE_y/m = \frac{1.6 \times 10^{-19} \times 3.4 \times 10^4}{9.1 \times 10^{-31}} = 0.5978 \times 10^{16} \text{ m/sec}^2$$

$$y = 1/2 \times 0.5978 \times 10^{16} \times (0.176 \times 10^{-8})^2 = 9.26 \times 10^{-3} \text{ m}$$

$$y = 0.926 \text{ cm}$$

The deflection of the spot due to electric field is $9.26 \times 10^{-3} \text{ m}$. This can be neutralized by magnetic field when

$$ev_x = B_z ev_y, B_z = E_y/v_x$$

$$B_z = \frac{3.4 \times 10^4}{1.7 \times 10^7} = 2 \times 10^{-3} = 0.002 \text{ Wb/m}^2$$

So magnitude of applied magnetic field is 0.002 Wb/m^2 and its direction will be perpendicular to electric field as well as perpendicular to the motion of the electron. Since vertical electric field E_y is acting downwards, so B_z will be upward.

5. The radius of a cyclotron dee is 1 m and the applied magnetic field is 0.5 Wb/m^2 . It is used to accelerate protons. Calculate the freq. of oscillating voltage needed to maintain resonance acceleration. What is the maximum energy acquired by the proton in eV?

Mass of proton = $1.67 \times 10^{-27} \text{ kg}$

Charge of proton = $1.6 \times 10^{-19} \text{ C}$

Resonance frequency = $n = Bq/2\pi m$

$$n = \frac{0.5 \times 1.6 \times 10^{-19}}{2 \times 3.14 \times 1.67 \times 10^{-27}}$$

$$n = 0.0763 \times 10^8$$

$$n = 7.63 \times 10^6 \text{ Hz} = 7.63 \text{ MHz}$$

Maximum energy of proton = $E = B^2 q^2 R^2 / 2m$

$$E = \frac{(0.5)^2 (1.6 \times 10^{-19})^2 (1)^2}{2 \times 1.67 \times 10^{-27}} = 0.1916 \times 10^{-11} \text{ J}$$

$$E = 0.1197 \times 10^8 \text{ eV} = 11.97 \text{ MeV}$$

6. An α -particle with charge 3.2×10^{-19} coulomb and mass 6.68×10^{-27} kg is injected into a transverse magnetic field of 1.5 Wb/m^2 with velocity $7.263 \times 10^6 \text{ m/s}$. (a) Calculate the force on the particle. (b) Calculate the periodic time and hence resonance frequency.

Force acting on the particle is

$$F = Bqv = 1.5 \times 3.2 \times 10^{-19} \times 7.263 \times 10^6$$

$$F = 34.86 \times 10^{-13} \text{ Newton}$$

$$\text{Periodic time } T = 2\pi/B \cdot m/q = \frac{2 \times 3.14}{1.5} \times \frac{6.68 \times 10^{-27}}{3.2 \times 10^{-19}}$$

$$T = 8.7396 \times 10^{-8} \text{ sec}$$

$$\text{Frequency } n = 1/T = 1.44 \times 10^7 \text{ Hz} = 11.44 \text{ MHz}$$

7. In a cyclotron the frequency of the oscillator is $1.2 \times 10^7 \text{ Hz}$. (a) calculate the flux density B at which cyclotron will attain the resonance for proton. (b) calculate the energy of the proton beam if the radius of the cyclotron is 0.5 m. (c) calculate the energy for the α -particle. Given the charge of α -particle = $2 \times 1.6 \times 10^{-19} \text{ C}$, mass of α -particle = $6.68 \times 10^{-27} \text{ kg}$.

Charge of proton = $1.6 \times 10^{-19} \text{ C}$, mass of proton = $1.67 \times 10^{-27} \text{ kg}$

(a) Resonance frequency $n = Bq/2\pi m$

$$\text{So } 2\pi nm/q = \frac{2 \times 3.14 \times 1.2 \times 10^7 \times 1.67 \times 10^{-27}}{1.6 \times 10^{-19}}$$

$$B_p = 0.7869 \text{ Wb/m}^2$$

$$(b) \quad \text{Energy of proton } E_p = \frac{B_p^2 q_p^2 R^2}{2m_p}$$

$$E_p = \frac{(0.787)^2 (1.6 \times 10^{-19})^2 (0.5)^2}{2 \times 1.67 \times 10^{-27}}$$

$$\approx 1.1875 \times 10^{-12} \text{ J}$$

$$E_p = 7.42 \text{ MeV}$$

$$(c) \quad \text{Energy of } \alpha\text{-particle is } E_\alpha = \frac{B_\alpha^2 q_\alpha^2 R^2}{2 \times m_\alpha}$$

$$\beta_\alpha = \frac{2\pi nm_\alpha}{q_\alpha}$$

$$= 1.573 \text{ Wb/m}^2$$

q_α = charge of α -particle = $2 \times 1.6 \times 10^{-19}$ C

m_α = mass of α -particle = 6.644×10^{-27} kg

$$E_\alpha = \frac{(1.573)^2 (1.6 \times 2 \times 10^{-19}) (0.5)^2}{2 \times 6.68 \times 10^{-27}} = 0.474 \times 10^{-11} \text{ J}$$

$$E_\alpha = 29.6 \text{ MeV.}$$

8. A mixed beam of electrons and α -particles accelerated by p.d. of 20 kV, enters in a uniform magnetic field of 0.05 Wb/m^2 in a direction at right angle to the field. Calculate the linear separation of the electron beam, from α -particle beam, on a common boundary wall.

Given that, charge of electron = 1.6×10^{-19} C

Mass of electron = 9.1×10^{-31} kg

Charge of α -particle = $2 \times 1.6 \times 10^{-19}$ C

Mass of α -particle = 6.68×10^{-27} kg.

$$\text{K.E. of the particle is } \frac{1}{2} mv^2 = qV, v = \sqrt{\frac{2qV}{m}}$$

Radius of circular path traversing in uniform magnetic field is

$$R = \frac{mv}{Bq} = \frac{mv}{Bq} \sqrt{\frac{2qV}{m}}$$

$$R = \frac{1}{B} \sqrt{\frac{2mV}{q}}$$

$$R_1 = R_{\text{electron}} = \frac{1}{0.05} \sqrt{\frac{2 \times 9.1 \times 10^{-31} \times 20000}{1.6 \times 10^{-19}}} = 95.4 \times 10^{-4} \text{ m}$$

$$R_2 = R_{\alpha\text{-particle}} = \frac{1}{0.05} \sqrt{\frac{2 \times 6.68 \times 10^{-27} \times 20000}{2 \times 1.6 \times 10^{-19}}} = 57.8 \times 10^{-2} \text{ m}$$

So the linear separation between them on common boundary wall after traversing semicircular path is

$$S = 2R_2 - 2R_1$$

$$S = 113.7 \text{ cm}$$

9. Electrons accelerated by a p.d. of 200 V enter in an electric field at an angle 60° with the normal to the interface of higher electric field region, where due to the

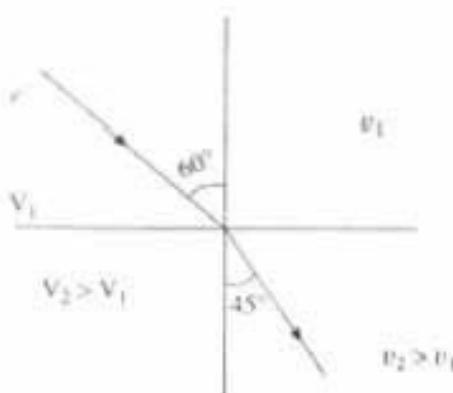


Fig. 3.17

electrostatic focusing effect, it makes an angle 45° with the normal to the interface of these two regions. Find the p.d. between these two regions.

$$V_1 = 200 \text{ V}$$

$$i = 60^\circ$$

$$r = 45^\circ$$

From electrostatic focussing condition we get

$$\frac{\sin i}{\sin r} = \frac{v_2}{v_1} = \sqrt{\frac{V_2}{V_1}}$$

Since the energy of electron is $1/2 mv^2 = eV$

$$\text{So } \sqrt{V_2} = \sqrt{200} \frac{\sin 60^\circ}{\sin 45^\circ}$$

$$\sqrt{V_2} = 17.32 \text{ volts } V_2 = 299.978 \text{ volts}$$

So the p.d. between these two regions is

$$V_2 - V_1 = 99.98 = 100 \text{ volts}$$

10. A charged water drop is just floating in a uniform electric field of 250 V/m. Find the charge on it. Given that

$$\text{Radius of the drop} = 10^{-8} \text{ m}$$

$$\text{Density of water} = 10^3 \text{ kg/m}^3$$

When the charged water drop is just floating in the electric field then the weight of the water drop is counter balanced by the force acting on it due to electric field, so

$$F = mg = qE$$

Now mass of the water drop is $m = \text{Vol.} \times \text{Density}$

$$m = \frac{4}{3}\pi(10^{-8})^3 \times 10^3$$

Thus weight of the water drop

$$mg = \frac{4}{3}\pi(10^{-8})^3 \times 10^3 \times 9.8$$

So in balance condition,

$$mg = \frac{4}{3}\pi(10^{-8})^3 \times 10^3 \times 9.8 = q \times 250$$

$$q = 0.164 \times 10^{-21} \text{ C}$$

11. A single ionised Ge atom enters a Bainbridge mass spectrograph with a velocity $5 \times 10^5 \text{ m/s}$. Calculate the radii of the paths followed by two most abundant isotopes of masses 72 and 74 when the magnetic flux density is 0.3 Wb/m^2 . Calculate also the linear separation of the lines on photographic plate for those two isotopes.

$$v = 5 \times 10^5 \text{ m/s}$$

$$B = 0.3 \text{ Wb/m}^2$$

$$M_{72} = \text{Mass of single ionized Ge}^{72} \text{ atom} = \frac{72}{6.025 \times 10^{26}} \text{ kg}$$

$$M_{72} = 72 \times 1 \text{ a.m.u.}$$

$$1 \text{ a.m.u.} = 1/6.025 \times 10^{26} = 1.66 \times 10^{-27} \text{ kg} = 1.66 \times 10^{-24} \text{ g}$$

Thus

$$\frac{M_{72}v^2}{R_{72}} = Bev$$

Now

$$R_{72} = \frac{M_{72}v}{Be}$$

$$R_{72} = 124.48 \times 10^{-3} \text{ m}$$

$$R_{72} = 1.245 \text{ m}$$

Now R is proportional to M .

Hence

$$\frac{R_{72}}{R_{74}} = \frac{M_{72}}{M_{74}}$$

$$R_{74} = \frac{R_{72}}{M_{74}}, M_{74} = \frac{R_{72}}{72}, R_{74} = 1.278 \text{ m}$$

Now, the linear separation of two lines on photographic plate is $2(R_{74} - R_{72}) = 0.033 \text{ m}$.

12. In a C.R.T. length of the applied magnetic field is 5 cm along the axis, the distance of the screen from the end of the magnetic field is 0.3 m. Calculate the flux density B required to produce a deflection of 1 cm on the screen if anode voltage is 1000 volts.

$$Solution \quad v = Dl \frac{e}{m} \frac{B}{v} = DlB \sqrt{\frac{e}{2mv_A}}$$

$$l = 5 \text{ cm}, D = 0.3 + l/2 = 0.3 + 2.5 \times 10^{-2} \text{ m}$$

$$B = \frac{v}{Dl} \sqrt{\frac{2mV_A}{e}}$$

$$= \frac{0.01}{0.325 \times 0.05} \sqrt{\frac{2 \times 9.1 \times 10^{-31} \times 1000}{1.6 \times 10^{-10}}}$$

$$B = 65.6 \times 10^{-6} \text{ wb/m}^2$$

13. An electron is accelerated through a p.d. of 150 volts. This electron is injected into a transverse electric field produced by the application of 20 volt to a pair of parallel plates of length 10 cm and 1 cm apart. A screen is placed at 50 cm apart from the centre of the applied electric field. Now calculate (a) velocity of the electron reaching the field, (b) acceleration due to deflecting field, (c) Final velocity attained by deflecting field (d) angle of deflection, (e) deflection on the screen and (f) deflection sensitivity.

$$Solution \quad (a) v_i = \sqrt{\frac{2eV_A}{m}} = 7.26 \times 10^6 \text{ m/sec}$$

$$(b) av = \frac{e}{m} \frac{V}{d} = 3.516 \times 10^{14} \text{ m/sec}^2$$

$$(c) v_x = a_i t = id \frac{l}{V_s} = 4.8 \times 10^6 \text{ m/sec}$$

$$(d) \tan \theta = \frac{v_x}{v_i} \Rightarrow \theta = \tan^{-1} \frac{v_x}{v_i} = 33.47^\circ$$

$$(e) Y = D \tan \theta = 0.33 \text{ m}$$

$$(f) S = \frac{Y}{V} = 0.0165 \text{ m/volt.}$$

QUESTIONS

1. Describe the construction, working and use of C.R.O.
2. Give the construction, working and theory of a cyclotron. Obtain an expression for the energy of the particles emerging from a cyclotron.
3. Describe the effect of magnetic field on a moving charge particle. Give any one application of this effect.
4. Describe the effect of a perpendicular electric field on the motion of charged particles. Derive the appropriate formula for linear deflection.
5. Show that the kinetic energy of the particle in a cyclotron is independent of the voltage applied.
6. An electron beam passes through a magnetic field of 2×10^{-3} Wb/m² and an electric field 3.4×10^4 V/m, both acting simultaneously at the same point. If the path of the electron remains undeviated calculate the speed of the electrons. If the electric field is removed, what will be radius of the electron's path?
 $m = 9.0 \times 10^{-31}$ kg, $e = 1.6 \times 10^{-19}$ C.
7. An Electron is accelerated through a potential difference 150 volts. This electron is injected into a uniform transverse electric field, produced by the application of 20 volt, to a pair of parallel plates of length 10 cm. Calculate the deflection produced in the electron path due to transverse electric field, on a screen placed at 50 cm apart from the centre of applied electric field.
8. Describe the principle of electrostatic and magnetostatic focusing. Describe their application in some instrument.
9. Discuss Aston and Bainbridge mass spectrograph. What is the use of velocity selector in mass spectrograph.
10. Double ionised Ge Atom enters in a Bainbridge mass spectrograph with velocity 6.3×10^5 m/sec. Calculate the radius of the paths of Ge⁷⁰ and Ge⁷⁶ when the magnetic induction is 0.02 Wb/m².
11. An electron with horizontal velocity v_1 is injected normal to an electric field in the Y-direction set up between two parallel conductor plates. Calculate vertical acceleration a_y , the transit time t , the vertical velocity v_y , the deflection θ produced in electron path and its vertical displacement S .
12. If energy of accelerated proton in cyclotron is 1.0 MeV and voltage applied between the dees is 1000 volts, then how many rotations a proton will make to achieve this energy? Assume that protons are just introduced in two dees. Given mass of proton = 1.67×10^{-27} kg and charge 1.6×10^{-19} C.

Solution:

$$\text{Initial velocity} = v_1$$

$$\text{Final velocity} = v_{\max}$$

$$\text{Initial energy } \frac{1}{2} mv_1^2 = qV = \frac{1.6 \times 10^{-19} \times 1000}{1.6 \times 10^{-19}} = 1000 \text{ eV}$$

$$\text{Final energy } \frac{1}{2} mv_{\max}^2 = 1.0 \text{ MeV}$$

$$\frac{v_{\max}^2}{v_1^2} = \frac{10^6}{10^3} = 1000$$

$$v_{\max}^2 = 1000 v_1^2$$

i.e.

$$v_{\max} = 31.62 v_1 = 32 v_1$$

Thus, total number of rotation done by Proton to achieve velocity v_{\max} from v_1 is 32 times from one dee to other.

So number of rotation/dee to achieve 1 MeV energy is $= \frac{32}{2} = 16$.

Quantum Physics and Schrödinger Wave Equation

QUANTUM PHYSICS

4.1 Introduction

The motion of a particle which can be observed directly or through microscope can be explained by classical mechanics but in the case of atomic particles like electron, proton classical mechanics failed to do so.

According to classical mechanics, the radiation of all wave lengths is continuous from hydrogen atom, whereas experimentally it is observed that the spectrum of hydrogen atom consists of discrete (quantized) set of lines and there are different series like Lyman series, Balmer series, etc. Thus classical theory failed to explain the discrete spectrum of hydrogen atom, which can be explained by quantum theory.

4.2 Quantum Theory of Radiation

Quantum theory of radiation was first proposed by Max Planck in 1901 in connection with thermal radiations emitted from a black body. Einstein extended this in 1905 to cover all radiations including light rays, γ -rays, etc. The study of the nature of radiation was originally made in connection with the nature of light and how it travelled from one point to another. In this connection, Newton proposed Corpuscular theory towards the end of 17th century, according to which light consists of minute fast moving elastic particles of small mass called as corpuscles. But the phenomena of interference, diffraction, polarisation etc. could not be explained on the basis of corpuscular theory. So it was superseded by Huygen's wave theory of light, where it is considered that light travels in the form of waves in a hypothetical medium called *ether*. Huygen's wave theory of light successfully explained the phenomena of interference, diffraction, polarisation, etc. This theory was followed in 1864 by Maxwell's electromagnetic theory. But still certain other experimental facts, e.g. photo-electric effect, Compton effect, emission and absorption of light etc. which could not be explained on the basis of above mentioned theories, that gave birth to the quantum theory of light.

The Quantization Energy

In 1901 Max Planck while studying black body radiations, had come to the

conclusion that the absorption or emission of thermal energy is not a continuous process, as like classical theory but takes place in discrete amounts. He assumed that matter is composed of a large number of oscillating particles, each can vibrate with a characteristic frequency. While according to the classical theory the particles can have any value of frequency. Now these atomic oscillators may not emit or absorb any amount of energy E , but only certain amount of energies chosen from a discrete set, defined by

$$E = nh\nu, \quad n = 1, 2, 3, \dots \quad (4.1)$$

where n is the quantum number h the Planck's constant, first time introduced in physics, and ν the frequency of oscillator.

Thus we can say the energy of an atomic oscillator is quantized and that the integer n is called the *quantum number*.

In case of large-scale oscillator, such as a swinging pendulum, our daily experience suggests that a pendulum can oscillate with any reasonable total energy and not only with some discrete energies. This is because friction causes the pendulum amplitude to decay, so the energy dissipated in a perfectly continuous way and not in jumps. Actually jumps are there, but they are just too small for us to detect. *Quantum theory must agree with classical theory in the limit of large quantum number, when $n \rightarrow \infty$*

Following Planck's idea in 1905, Einstein also quantized the radiant energy of light by postulating that:

- (i) Light energy is not emitted continuously but intermittently, by indivisible amount of radiant energy which is called as '*quanta*'.
- (ii) These quanta travel in space in the form of bundles or packets called *photons* which carry a definite amount of energy. The energy carried by each photon is given by

$$E = h\nu$$

where ν is the frequency of radiation.

Obviously the photons or quanta of high frequency γ -ray, X-ray, have large amount of energy whereas low frequency photon or radio waves have comparatively less amount of energy.

4.3 Photo-Electric Effect

Photo-electric effect is an example of a radiation matter interaction. The emission of electrons from a metal surface, when electromagnetic radiation is incident upon it, is called "*photo-electric emission*" and the effect is called "*photo-electric effect*". The emitted electrons are called "*photo-electrons*". *The basic theory of photo-electric effect is the quantum theory, where quantized amount of energy ($h\nu$) is absorbed by the metal surface from incident photon, and gives rise to the emission of the Photo-electrons.*

This phenomenon was first observed by Hertz, when ultraviolet light falls on zinc plate. Afterwards R.A. Milliken investigated (1868-1953) the effect with a number of alkali metals over a wide range of light frequencies and was given the Nobel Prize in 1923.

Experimental Verification of Photo-electric Effect

Figure 4.1 shows the typical apparatus used to study photo-electric effect. Light of frequency ν falls on a metal surface E and if the frequency is high enough, the light will eject electrons from the surface. If we set up a suitable potential difference V between metal surface E and the collector C , we can collect these photo-electrons as we call them, and measure them as a photo-electric current i_p .

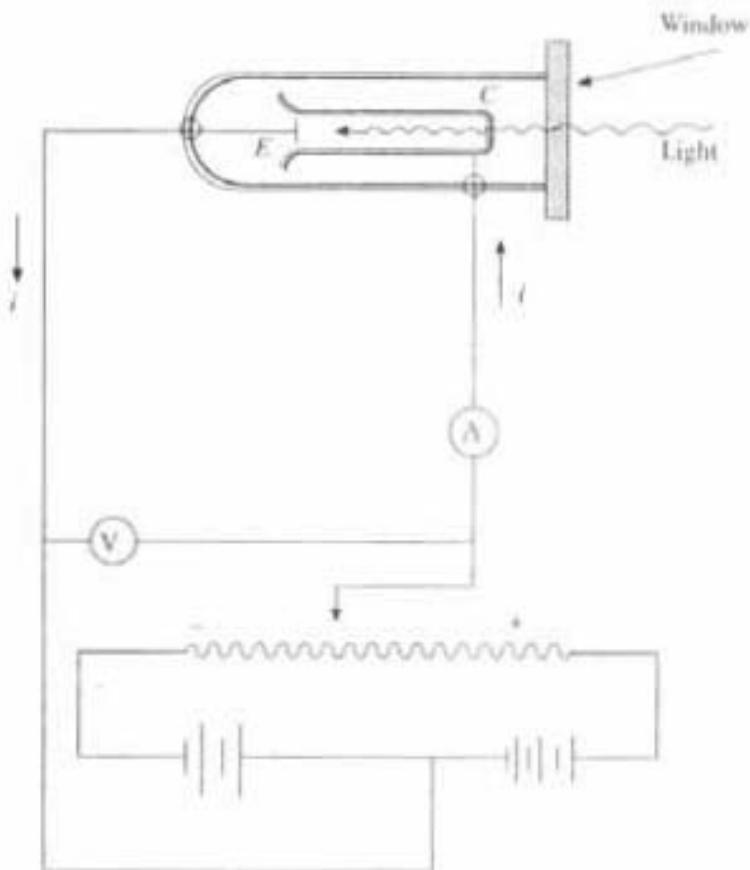


Fig. 4.1 Experimental set-up for photo-electric effect

Figure 4.2(a) shows the photo-current as a function of potential difference V . We see that if V is +ve and large enough, the photo-electric current reaches a constant saturation value, at which all photo-electrons ejected from E and get collected by C .

If we reduce V to zero and then reverse it, the photo-electric current does not immediately drop to zero, because the electrons emerge from emitter E with non-zero velocities, some of them will reach the collector C even though the potential difference opposes their motion. However if we make the reverse potential difference large enough, then there will be a value called V_0 , the *stopping potential*, at which Photo-electric current i_p drops to zero (Fig. 4.2a). This potential difference multiplied by the electronic charge e , gives the maximum kinetic energy K_{\max} of the most energetic emitted photoelectron.

$$K_{\max} = eV_0 \quad (4.2)$$

The following characteristics of photo-electric emission is observed from Figs. (4.2a, b, c).

- It shows that the stopping potential V_0 (Fig. 4.2a) and K_{\max} is independent of the intensity of the incident light I .
- Stopping potential is different for incident light of different frequencies. It increases as frequency of incident light (ν) increases (Fig. 4.2b). Hence maximum energy K_{\max} of photoelectrons also increases with frequency of incident radiation, but it is independent of intensity of radiation I .
- Photo-electric current increases as intensity of light I increases (Fig. 4.2a) for a given frequency, but always there is a saturation. That saturation value is higher, if intensity is higher (Fig. 4.2a).
- Photo-electric emission is instantaneous. The timelag between turning on the source of radiation and emission of photoelectrons is very small, less than 10^{-9} sec.

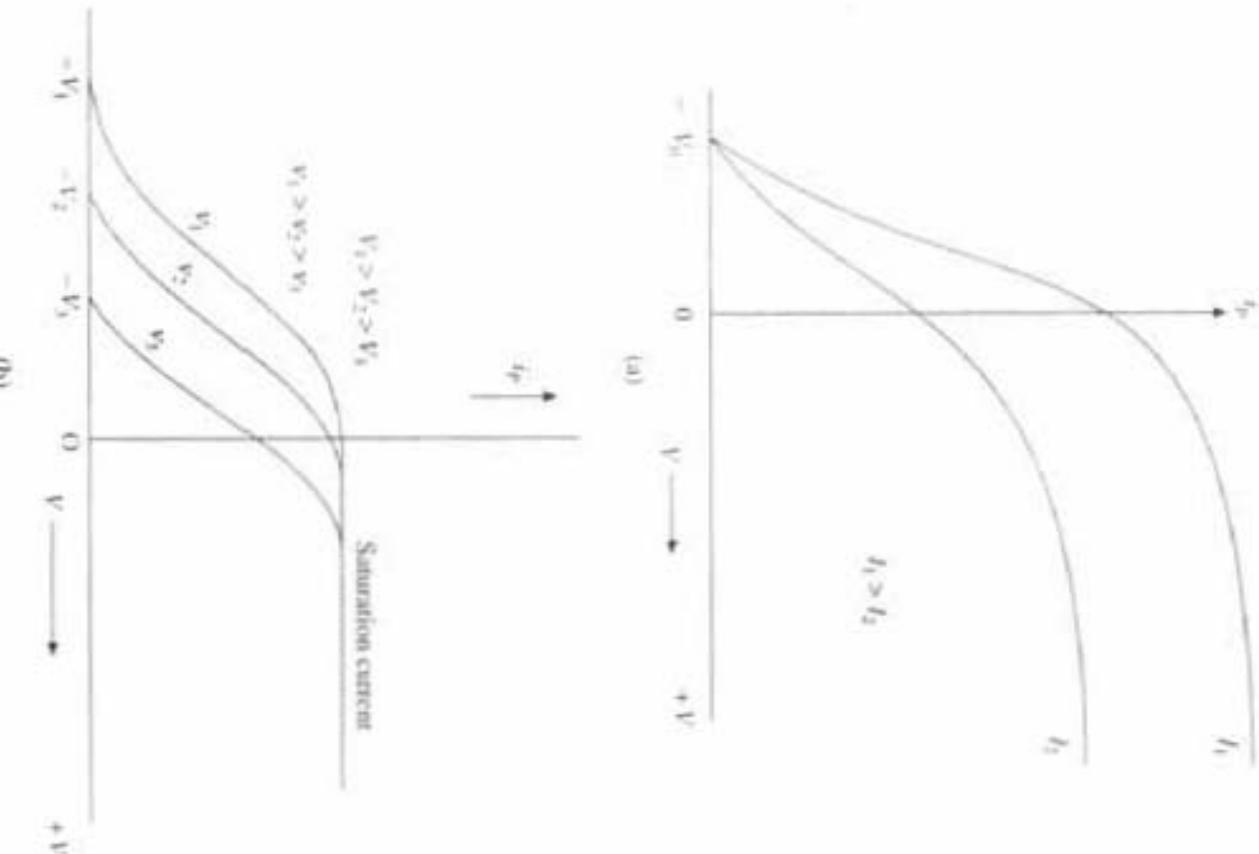


Fig. 4.2 Different characteristics of photo-electric emission

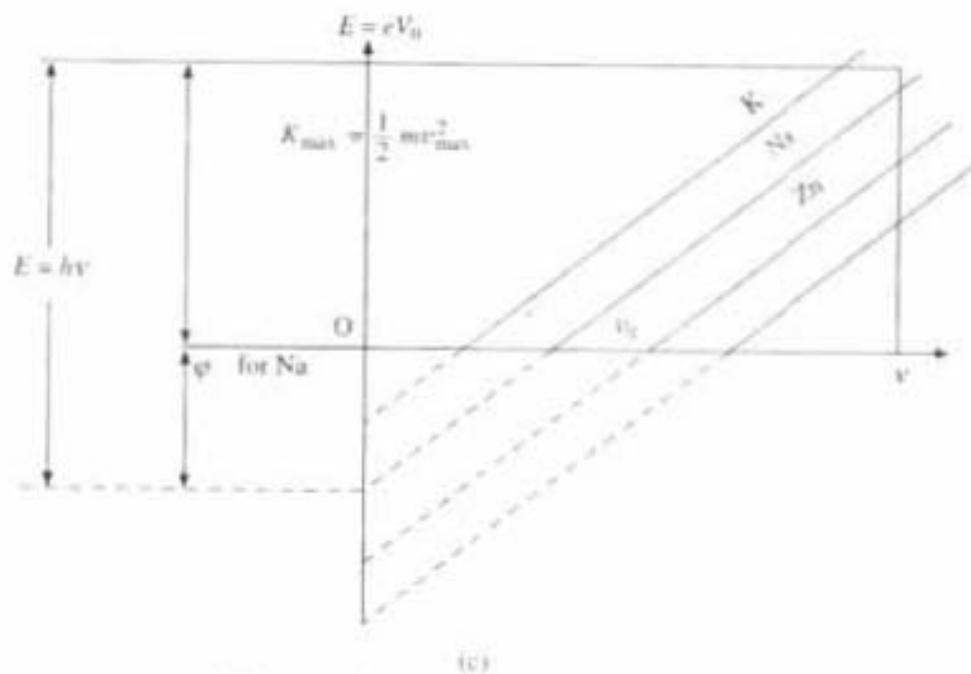


Fig. 4.2 Different characteristics of photo-electric emission

5. For a given surface of emitter E , there exists a certain minimum frequency v_0 of radiation, called threshold frequency, which is different for different metals (Fig. 4.2c). When the incident frequency of light is not equal to or smaller than that of threshold frequency v_0 , then there will not be any photo emission, it does not matter what is the intensity of light.
6. The rate at which photoelectrons are emitted from the surface is independent of its temperature, so the photoelectrons effect is entirely different from the thermionic emission.

4.4 Einstein's Photo-Electric Equation

According to the quantum theory of radiation, a beam of radiation of frequency v consists of photon each of energy $E = hv$. According to Einstein who gave an explanation of photo-electric effect in 1905 by using Planck's quantum theory, that when radiation falls on a metal surface, the photon in the beam is completely absorbed by an electron in the emitter E . That photon in the incident radiation not only give a part of his energy to the electron, but part of the energy of the absorbed photon is used by the electron to become free itself from the metal surface, the balance energy will appear as kinetic energy of the outgoing photo-electron. That K.E. of the photoelectron will be maximum for the least tightly bound electron. The energy required to make free the least tightly bound electron from the metal surface is called 'work function' of the metal (φ) (Fig. 4.2c), it is different for different metal.

Hence if a photon of energy hv emits electrons from a metal surface with maximum K.E., $K_{\max} = 1/2 mv_{\max}^2$, then we can write;

$$hv = 1/2mv_{\max}^2 + \varphi \quad (4.3)$$

This is known as 'Einstein's equation' which explain photo-electric effect. Here

ϕ is the work function of the metal. Since ϕ is a constant, for the metal, so there is a minimum frequency called (Fig. 4.2c) *threshold frequency* $v_0(\phi = hv_0)$ of radiation, below which no photo emission from the metal surface is possible. So equation (4.3) can be written as

$$hv = 1/2 mv_{\max}^2 + hv_0$$

i.e. $K_{\max} = eV_0 = (1/2) mv_{\max}^2 = hv - hv_0 = h(v - v_0)$ (4.4)

Einstein's equation explains all the characteristics features of the photo-electric effects as follows:

1. The above equation shows that the maximum K.E. of the electron emitted from the surface and show the stopping potential V_0 , ($K_{\max} = eV_0$), depends upon the frequency of the radiation, not on the intensity of radiation (Figs. 4.2a, b).

2. When the intensity of the incident radiation is increased, the number of photons reaching the surface also increases. Hence the photo-electric current increases with the intensity of radiation (Fig. 4.2a).

3. The release of a photo-electron from a surface depends upon the chance that a photon from the incident beam hitting the electron and supply the necessary energy to escape from the metal surface. So the emission is instantaneous.

4. It is seen from equation (4.4) that if $v < v_0$ (threshold frequency), no photo-electron can be free, so no photo emission from the surface is possible whatever be the intensity of the radiation (Fig. 4.2c).

5. Since in different materials the electrons are bound to the surface with different energy, so the energy required to free the electrons will be different for different metals. That is why the threshold frequency (v_0) is different for different metals (Fig. 4.2c).

6. We have $(1/2) mv_{\max}^2 = K_{\max} = eV_0$, where V_0 is the stopping potential.

Thus,

$$eV_0 = hv - hv_0 \quad (4.4)$$

hence the plot of eV_0 versus v is a straight line (Fig. 4.2c), which is different for different materials.

4.5 Compton Effect

Compton effect is another example of the interaction of radiation with matter. But here the radiation should be of very high frequency or low wave length region, say X-ray or γ -ray, otherwise it will not be prominent. In 1923 Compton observed that when a monoenergetic beam of high frequency radiation say X-ray or γ -ray is scattered by a substance, then the scattered radiations contain radiation of two components—one having same frequency, same wave length (λ) and other having lower frequency and higher wave length λ' . The change of wave length of the scattered radiation is called *Compton shift* and the effect is called *Compton effect*. It is Purely a quantum effect. Compton shift only depends on the angle θ at which the scattered beams are observed, not on the wave length of the incident radiation nor on the scattering materials.

According to Compton, the phenomenon of scattering is due to the elastic collision between two particles, the photon of incident radiation and the electron

of the scatterer. When the incident photon of energy $E = h\nu$ collides with the electron of the scatterer, which is assumed to at be rest, it gives some of its energy to that electron, i.e. the incident photon losses some energy and hence emerges with lower energy $E' = h\nu'$ since $\nu' < \nu$ so $\lambda' > \lambda$ since $\nu = \frac{c}{\lambda}$.

Expression for Compton Shift

When a particle move with a velocity v which is comparable to velocity of light c , then relativistic mass correction required according to Einstein as

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}}$$

where m_0 = rest mass of the particle.

Now energy from Einstein's relation is

$$E = h\nu = mc^2 = \sqrt{p^2c^2 + m_0^2c^4}$$

where

$$p = mv = \text{momentum} = \frac{E}{c^2}v$$

Now, for photon rest mass $m_0 = 0$.

Hence

$$E = pc = h\nu \quad (4.5)$$

and momentum

$$p = \frac{h\nu}{c} = \frac{h}{\lambda} \quad (\because c = \nu\lambda)$$

Figure 4.3 shows that a photon of energy $h\nu$ and momentum $h\nu/c$ with wave length λ collides with an electron which is initially at rest. Now if m_0 is the rest mass of the electron, then rest energy is m_0c^2 . Since the electron is initially at rest so its momentum will be zero.

Now, after collision photon is scattered at an angle θ with increased wavelength λ' (i.e. with less energy, less frequency ν') and the electron moves off with velocity v at an angle ϕ , as shown in Fig. 4.3. Thus the momentum of the electron after collision is

$$p_e = mv = \frac{m_0v}{\sqrt{1 - v^2/c^2}} \quad (4.6)$$

The energy of the recoil electron is $E = \sqrt{p_e^2c^2 + m_0^2c^4}$.

Now resolving the linear momentum of the photon and electron along and at right angle to the incident direction OX and applying the principle of conservation of linear momentum, we get

$$\frac{h\nu}{c} = \frac{h\nu'}{c} \cos \theta + p_e \cos \phi$$

$$0 = \frac{h\nu'}{c} \sin \theta - p_e \sin \phi$$

which we modified as

$$cp_e \cos \phi = h\nu - h\nu' \cos \theta$$

$$cp_e \sin \phi = h\nu' \sin \theta$$

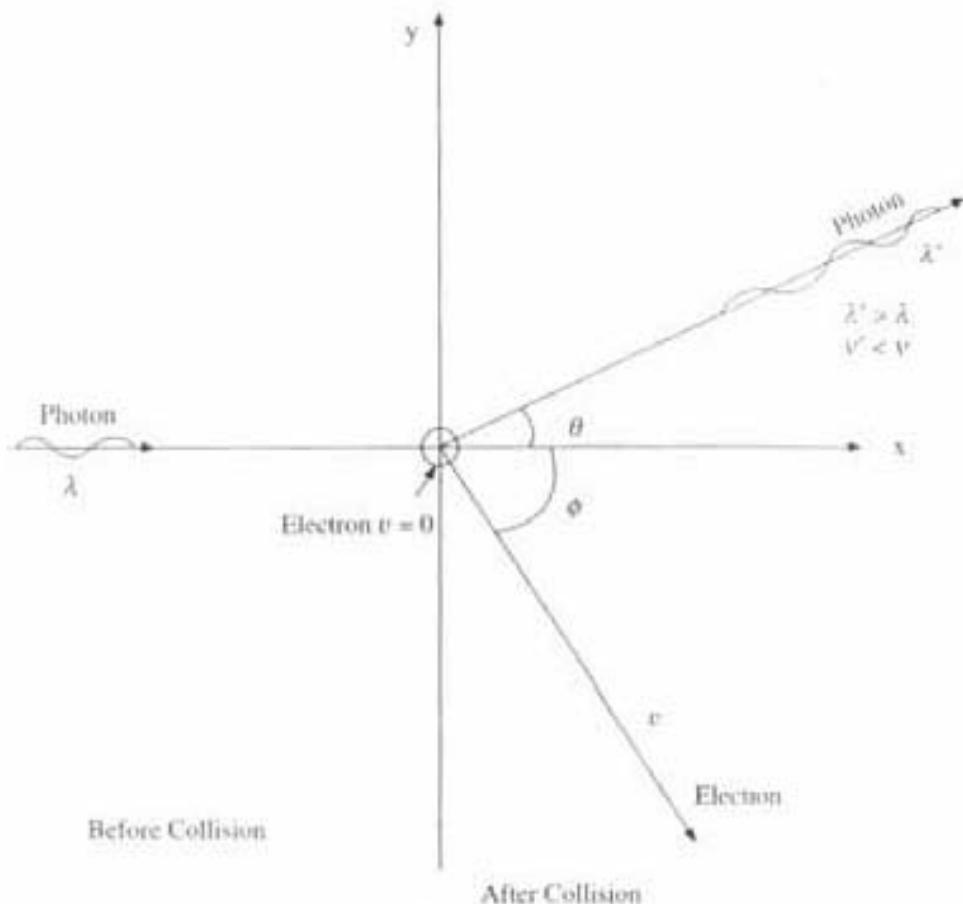


Fig. 4.3 Compton shift

Now by squaring and adding we get

$$c^2 p_e^2 = (hv)^2 + (hv')^2 - 2(hv)(hv') \cos \theta \quad (4.7)$$

The energy ($hv - hv'$) lost by photon in collision with electron is gained by electron as K.E. Thus

$$\text{K.E. of the electron} = hv - hv'$$

Hence total energy of recoil electron

$$E = \text{K.E.} + \text{Rest mass energy} = (hv - hv') + m_0 c^2$$

and also total energy of the electron from Eqn. (4.6) is

$$E = \sqrt{p_e^2 c^2 + m_0^2 c^4}$$

Hence, by equating above two equations, we get

$$\sqrt{p_e^2 c^2 + m_0^2 c^4} = (hv - hv') + m_0 c^2$$

on squaring we get

$$\begin{aligned} p_e^2 c^2 + m_0^2 c^4 &= (hv - hv')^2 + 2(hv - hv') m_0 c^2 + m_0^2 c^4 \\ p_e^2 c^2 &= h^2 v^2 + h^2 v'^2 - 2hv hv' + 2(hv - hv') m_0 c^2 \end{aligned} \quad (4.8)$$

Equating equations (4.7) and (4.8), we get

$$\begin{aligned} 2(hv - hv')m_0c^2 &= 2h^2vv'(1 - \cos \theta) \\ \left(\frac{1}{v'} - \frac{1}{v}\right) &= \frac{h}{m_0c^2}(1 - \cos \theta) \\ \frac{\lambda'}{c} - \frac{\lambda}{c} &= \frac{h}{m_0c^2}(1 - \cos \theta) \\ \Delta\lambda = (\lambda' - \lambda) &= \frac{h}{m_0c}(1 - \cos \theta) \end{aligned} \quad (4.9)$$

which is the expression for *Compton shift*. From the expression (4.9) it is clearly found that Compton shift depends only on angle of scattering θ ; not on the incident wave length nor on scattering material. From the above expression we have seen that for $\theta = 90^\circ$, Compton Shift $\Delta\lambda = h/m_0c$, which is called *Compton wave length*, which for an electron of rest mass m_0 is equal to 0.024 Å. For scattering angle $\theta = 180^\circ$, Compton shift $\Delta\lambda$ will be greatest, and is given by 0.048 Å, which is called *head on collision*, the incident photon being reversed in direction.

Some Important Points Regarding Compton Shift

1. The expression of Compton Shift for higher frequency radiation as we have seen is $\Delta\lambda = \frac{h}{m_0c}(1 - \cos \theta)$.

Now if the incident radiation is of lower frequency region say, in visible or microwave, or radio frequency range, where λ is very large compare to $\Delta\lambda$, in that case $\lambda' = \lambda + \Delta\lambda = \lambda$. So Compton effect is in-significant in such cases. In X-ray or γ -ray $\Delta\lambda$ is comparable to λ , so Compton shift is measurable and very significant mainly for higher frequency radiation.

2. Secondly, the *Compton effect is purely quantum effect*, as $h \rightarrow 0$, $\Delta\lambda \rightarrow 0$, so it cannot be explained on the basis of classical physics.

Experimental Verification of Compton Effect

In Compton experiment, a beam of X-rays with sharply defined wave length λ falls on the target T as shown in Fig. 4.4(a). For various angle of scattering θ , the intensity of the scattered X-rays is measured as a function of wavelength. Figure 4.4(b) shows the Compton's experimental result. For molybdenum K_{α} -line (0.708 Å), the wave length of the modified line observed at 90° was 0.730 Å. This gives $\Delta\lambda = \lambda' - \lambda = 0.022$ Å, which is in agreement with the theoretical value 0.024 Å. Figure 4.4(b) shows the wave length spectra of photons scattered at various angle from a graphite scatterer.

The scattering of photon and ejection of electron are simultaneous process. This is verified by Bothe and Geiger, who set up two ionisation counters very close together. One counter was sensitive to electrons while the other one was sensitive to X-rays. The large number of simultaneous responses in the two counter showed that the process of scattering of photon and recoil of electron are simultaneous.

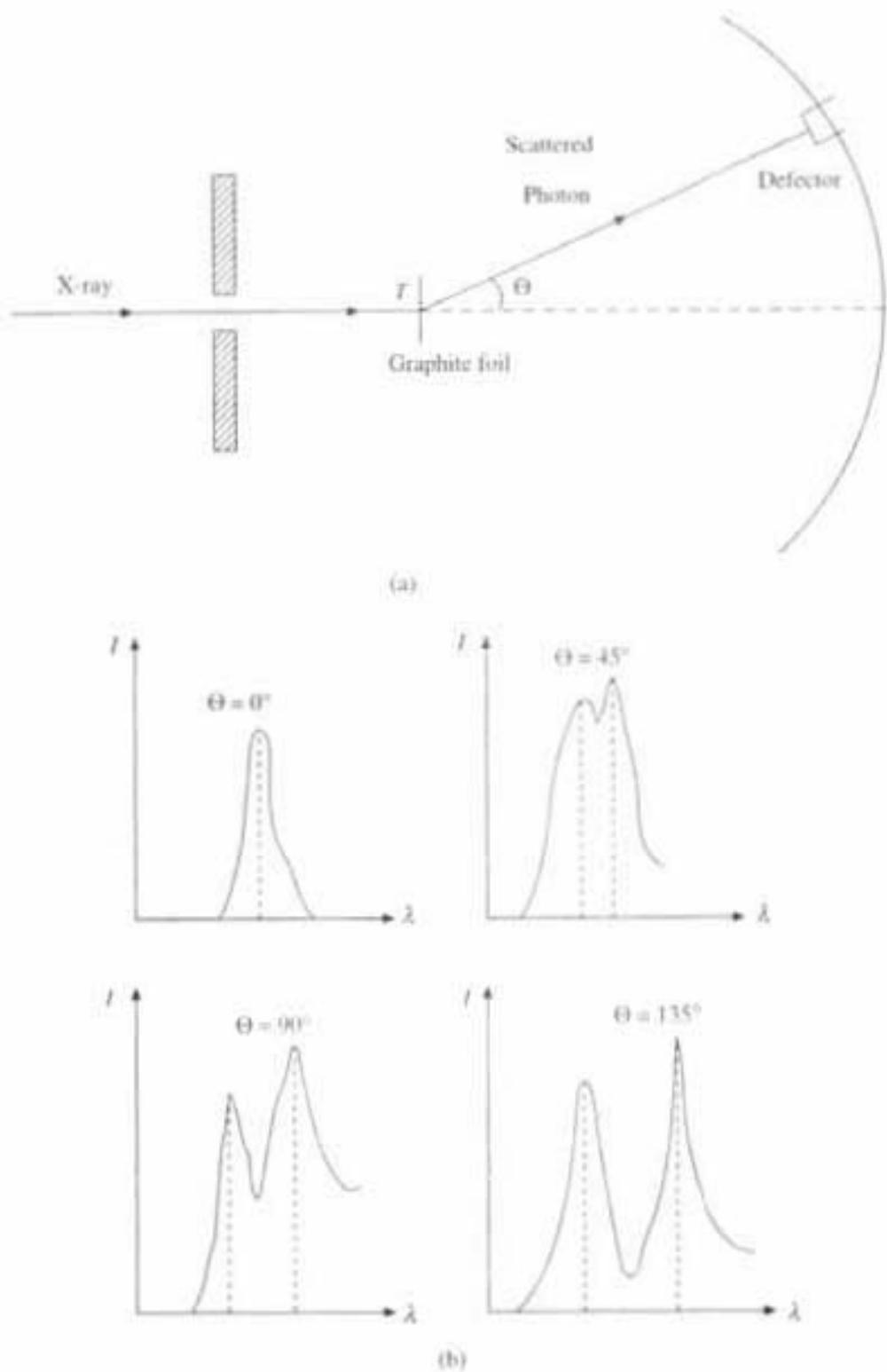


Fig. 4.4 Experimental set up and results for Compton shift for different scattering angle θ .

Explanation of Unmodified Wavelength in Compton Effect

The presence of the peak in Fig. 4.4(b) for which the wave length does not change on scattering, can be explained as follows: This peak results from the collision between photons and the electron that, instead of being nearly free, are tightly bound in an ionic core in the scattering target. During photon collision

that bound electron behaves like a very heavy free electron. This is because the collision between them must be considered as taking place between the incident photon and the entire atom of heavy mass say $M_0 \gg m$ (mass of nearly free electron). This would give $\Delta\lambda = \frac{h}{M_0 c}(1 - \cos \theta)$. So the Compton shift for tightly bound electron is immeasurably small. So during collision some photons are scattered by the electrons which are rendered free, these photons are scattered with modified wavelength λ' , whereas some photons get scattered from electrons which are tightly bound to the nucleus, these photons remain unchanged (λ) in the wave length, which gave rise to the unmodified peak in Fig. (4.4b).

4.6 Wave Nature of Particle and Wave-Particle Duality

The idea of wave nature of particle came first from the observation of dual nature of electromagnetic radiation. In interference, diffraction experiment, radiation behaves like wave, whereas in Photo-electric effect, Compton effect, black-body radiation, it behaves as particles. Of course, radiation cannot exhibit its particle and wave properties simultaneously. But due to this dual property it seems that the wave theory and particle theory are complementary to each other. Say for example, experimentally it is very difficult to observe interference effect with high frequency γ -rays, i.e., wave description of γ -rays becomes less probable at these frequencies. Photon or particle description would be more natural for high frequency photon. Whereas if the frequency is less the wave nature is more probable. Still the dual nature of radiation was not easily accepted, because of the apparently contradictory aspects of the two natures:

- A wave is specified by its: (1) frequency, (2) wave length (3) phase or wave velocity, (4) amplitude and (5) intensity. Moreover a wave spreads out and occupies a relatively large region of space.
- A particle is specified by: (1) mass (2) velocity (3) momentum and (4) energy. Moreover, a particle occupies a definite position in space and hence it is very small.

In view of the above, it is rather difficult to accept the conflicting ideas that radiation is a 'wave' which is spread out over space and also a 'particle' which is localized at a point in space. But this acceptance is essential, if one satisfactorily explains the result of experiments, which can be performed with radiation of different range of frequencies, from very high to low values.

We have already seen in section 4.5 that the wave and particle nature of photons are related by the equation (4.5)

$$\text{i.e.} \quad E = h\nu = pc \quad (4.5.1)$$

$$\text{or} \quad p = \frac{h\nu}{c} = \frac{h}{\lambda} \quad (\because c = \nu\lambda) \quad (4.5.2)$$

where h is the Planck's constant.

In these equations (4.5.1) and (4.5.2), two particle like quantities, energy (E) and momentum (p) of photon appear on left. Whereas two wave like quantities,

frequency (v) and wave length (λ) appear on right. These two equations express the phenomena of wave particle duality of radiation quanta i.e. photon. Sometimes photon behaves as particle with energy E and momentum p and sometimes as wave with frequency v and wave length λ .

4.7 de-Broglie's Hypothesis

Waves and particles are the only two modes through which energy can propagate in nature. Since nature is symmetric, so matter and waves must be symmetric. If radiation displays wave like, as well as particle like properties, matter can also display particle like, as well as wave like properties. On the basis of this argument, de-Broglie put forward his hypothesis as: *particle whose velocity is comparable to c, the velocity of light can exhibit wave like properties and the wavelength of the wave associated with particle of mass 'm' moving with velocity v (momentum $p = mv$) can be expressed as*

$$\lambda = h/p = h/mv \quad (4.10)$$

These waves associated with particles are called as *Matter Waves* and the wave length as *de-Broglie's wave length*.

de-Broglie's Wavelength of an Electron

When an electron is accelerated by a p.d. V then the kinetic energy of the electron is

$$eV = 1/2 mv^2, mv = \sqrt{2meV}$$

1. From de-Broglie's theory, wavelength associated with electron is

$$\lambda = h/mv = h/\sqrt{2meV} = 12.25/\sqrt{V} \text{ Å} \quad (4.11)$$

by substituting the value of h , e and m .

2. Wave number: $n = 1/\lambda = \sqrt{\frac{2meV}{h}} \text{ m}^{-1}$

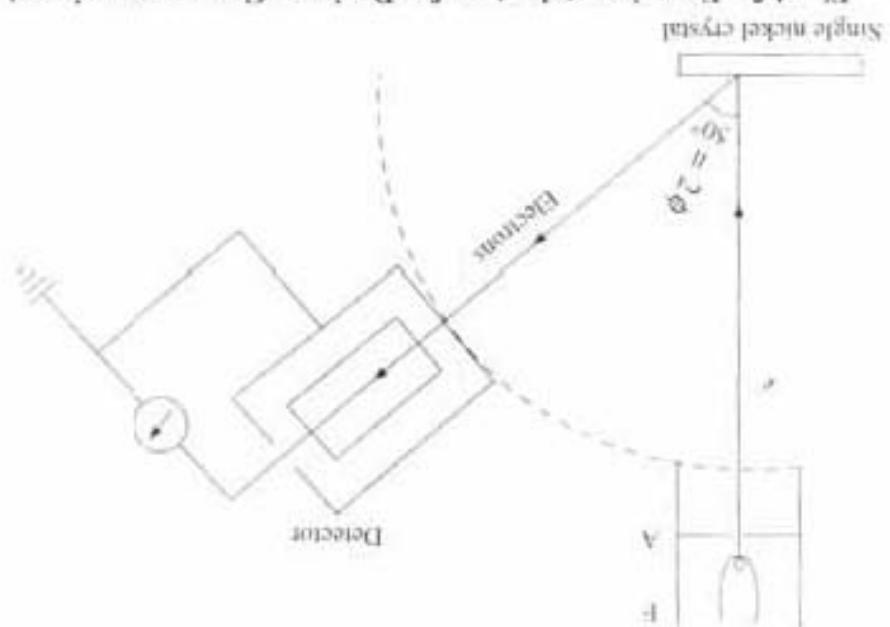
3. Wave Velocity: The wave velocity U of the electron wave can be found by the photon analogy

$$E = hv = eV$$

$$\begin{aligned} v &= \frac{eV}{h} = \frac{2meV}{h^2} \cdot \frac{h}{2m} \\ &= \frac{h}{2m} \cdot \frac{1}{\lambda^2} \quad \left[\because \text{from equation 4.11 } \frac{1}{\lambda} = \sqrt{\frac{2meV}{h}} \right] \end{aligned}$$

$$\text{So Wave velocity of electron} = U = v\lambda = \frac{h}{2m} \cdot \frac{1}{\lambda^2} \cdot \lambda = \frac{h}{2m\lambda} \quad (4.12)$$

Fig. 4.5 Experimental set up for Davisson-Germer experiment



The scattered electron beam can be detected by the electron detector which could be rotated around the nickel crystal through a large angle. From equations (4.11) and (4.13) it is clear that for a given accelerating voltage V , the maximum intensity of the scattered beam will show in a particular direction. Thus, the electron detector D then can be set at any desired angle ϕ that allows the intensity of the scattered beam to be measured. Figure 4.6(a) shows the famous result of Davisson-Germer experiment.

$$2d \sin \theta = n\lambda \quad (4.13)$$

Electrons emitted by a heated filament and accelerated through a.p.d. V volts, emerge from the electron gun with energy eV and are allowed to strike a single crystal of nickel (Fig. 4.5). These electrons, acting like a wave, were diffracted off the crystal and detected by a screen. The diffraction pattern was observed on the screen. This pattern was a series of alternating bright and dark bands. The spacing between the bright bands was measured to be 0.1225 nm . This value of d is of the order of the interplanar distance in the crystal. This fact suggests that it is possible to show the existence of electron wave with the help of electron diffraction, by using crystal as diffraction grating as it was used for X-rays. Experiment based on this idea was performed by Davisson and Germer in 1927.

$$\text{Now if } V = 100 \text{ volts, then } \lambda_{elec} = \frac{\sqrt{100}}{12.25} = 1.225 \text{ Å}$$

$$\lambda_{elec} = \frac{\sqrt{V}}{12.25} \text{ Å}$$

Electron wave length from the de-Broglie's hypothesis is

4.8 Davisson and Germer Experiment

Hence, the wave velocity of electron is not constant, as $\lambda \propto V^{-1/2}$ (accelerating voltage) changes, wave velocity will also change.

Qualitative Explanation

The existence of the peak in the intensity of the scattered electron beam shows qualitatively the validity of de-Broglie's hypothesis. Because it is analogous to the X-ray diffraction by a crystal due to constructive interference of waves diffracted by periodic arrangement of atoms in the planes of the crystal. *It cannot be explained by treating electron as particles. Particles will be reflected by the atomic planes but they cannot exhibit interference, only waves can interfere.*

Quantitative Explanation

Thus the strong beam of electrons (at $2\phi = 50^\circ$ and $V = 54$ volts) is scattered due to its wave like property from the atomic planes in the Nickel crystal (Fig. 4.6b). Here from Fig. 4.6(b) it is clear that the scattering planes in the crystal are included at angle ($\phi = 25^\circ$) to the top surface of the crystal. Constructive interference will occur when it will satisfy Bragg's law equation (4.13), i.e.,

$$2d \sin \theta = n\lambda$$

Now glancing angle $\theta = 90^\circ - 25^\circ = 65^\circ$ from Fig. 4.6(b), and $d = D \sin \phi$, from Fig. 4.6(b), where D = interatomic distance for Ni crystal = 2.15×10^{-10} m and $\phi = 25^\circ$ from Fig. 4.6(b).

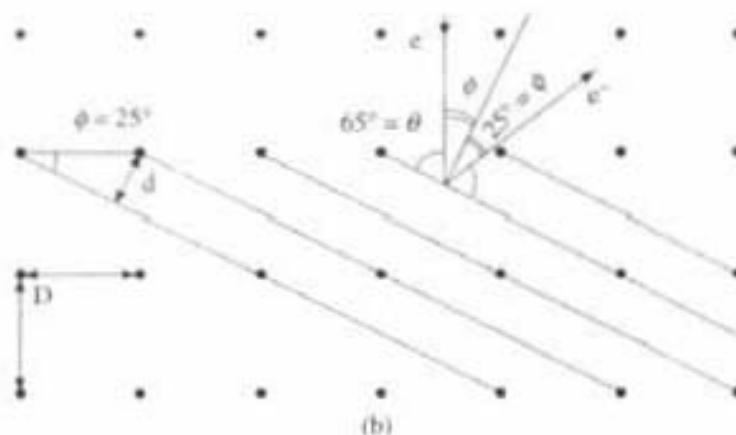
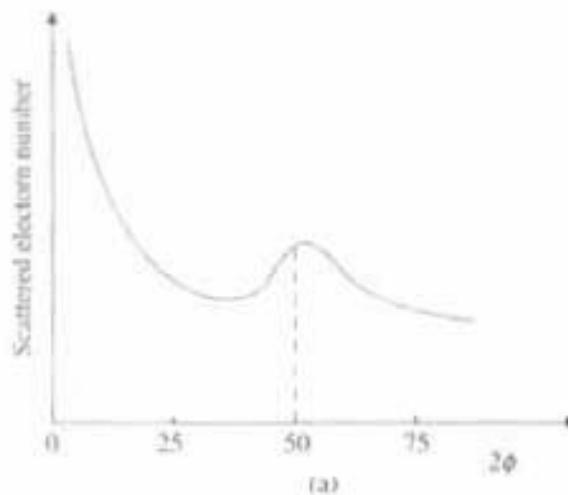


Fig. 4.6 (a) Variation of scattered electron number with scanning angle; (b) Electron diffraction occurring from lattice plane of Nickel crystal

Hence, interplaner distance d of the scattering planes will be

$$d = D \sin \phi = 2.15 \times 10^{-10} \sin 25^\circ = 0.909 \times 10^{-10} \text{ m}$$

The glancing angle θ is 65° as shown in Fig. 4.6(b). Thus, by using Bragg's formula, we have for first order diffraction

$$2d \sin \theta = 1 \times \lambda$$

Therefore $\lambda_{\text{rec}} = 2 \times 0.909 \times 10^{-10} \times \sin 65^\circ = 1.65 \text{ \AA}$ (4.14)

The theoretical value of de-Broglie's wave for electron at operating voltage 54 volts from equation (4.11) is

$$\lambda_e = \frac{12.25}{\sqrt{54}} = 1.67 \text{ \AA} \quad (4.15)$$

The close agreement between the experimental and theoretical values shows that electrons have wave like nature and it confirms the de-Broglie's hypothesis of matter waves.

4.9 Properties of Matter Waves and how they Differ from Sound and Light Waves

Matter waves are neither electromagnetic waves like light waves nor like sound waves which propagate through the media due to the variation of pressure. It is a new type of wave called pilot waves, in the sense that they pilot or guide the matter particle. Properties of matter waves and how they differ from the usual waves, we will discuss now.

1. For de-Broglie's wave length $\lambda = h/\sqrt{2meV}$. Hence, larger the mass of the matter particle or larger accelerating voltage V , shorter the wave length associated with it, not constant.

2. The phase or wave velocity of matter wave is $U = v\lambda$. Now

$$E = hv \quad \text{from photon analogy}$$

Also $E = mc^2 = hv \quad \text{i.e.} \quad v = mc^2/h$

de-Broglie's wavelength $\lambda = h/mv$

\therefore wave velocity $U = v\lambda = mc^2/h \cdot h/mv = c^2/v$

Since the speed v of the particle cannot exceed the velocity of light c , so it is obvious that the wave velocity of the matter wave (U) is more than the velocity of light, which is contradictory with the postulates of special theory of relativity.

3. The velocity of light is constant, whereas that of matter wave (U) is not constant since it depends on velocity (v) of material particle.

4. Light viewed as corpuscle or particle has an energy and associated mass. It can be located precisely at any point. Only difference between the light viewed as particle and other particle of matter is that, the rest mass of light photon $m_0 = 0$, since it is moving with a velocity of light c , which is not so for other matter particles.

5. If the wave length of the matter wave is large, i.e. frequency is small, then

its wave description is more probable, it appears like a wave. But if the wavelength is small, frequency is large then it behaves like a particle and so particle nature will be more probable, say for example, γ -ray and γ -particle.

6. Quantum mechanically the variable quantity which characterises the matter wave, is called wave function (ψ). It can be represented at distance x at a time t as

$$\psi(x, t) = A \sin 2\pi(t/T - x/\lambda)$$

The probability of finding the particle at point (x) at time t is proportional to $|\psi|^2$.

This wave function of matter wave will vary in space and time, like the variation of electromagnetic field for light wave propagation and variation of pressure in the medium for sound wave propagation.

7. The wave velocity of matter wave is also given by $U = h/2m\lambda$ (Eqn. 4.12). It shows that wave velocity depends inversely on wave length λ , even if the particle is moving in vacuum. This is quite different behaviour from that of light, which moves with same velocity c regardless of wave length.

8. Matter wave is not a physical phenomenon. It is rather a symbolic representation of what we know about the particle. So it is called a wave of probability.

4.10 Heisenberg's Uncertainty Principle

The concept of the dual particle and wave nature of matter leads to an important result. According to classical mechanics, a particle has a position and momentum which can, in principle, be determined with any desired accuracy. But a wave is extended throughout a region of space, and the location of an electron, considered as a wave, seems to present some difficulty. The question arises; if an electron is described sometimes as wave, sometimes as particle, is it possible to know exactly where an electron is in space at some given instant? The answer to this question is given by Heisenberg's uncertainty principle, which states that '*The position and momentum of a particle cannot be determined simultaneously with highest accuracy*'.

In a mathematical form it states that

$$\Delta x \cdot \Delta p \geq \hbar$$

$$\Delta t \cdot \Delta E \geq \hbar$$

i.e. the product of uncertainties in measurement of two variables, in position (Δx) and momentum (Δp) or in the time (Δt) and energy (ΔE) must be at least of the order of \hbar ($= h/2\pi$) where h is the Planck's constant. Here Δx , Δp are the errors in the determination of position and momentum, ΔE and Δt are the errors in determination of energy and time. It means that if position is measured accurately (i.e. Δx is very very small) then measurement of momentum becomes correspondingly inaccurate by at least the amount $\Delta p = \hbar/\Delta x$ and vice versa.

Similarly if a system remains in a particular state of motion not longer than a time Δt , then energy of the system in that state is uncertain by at least the amount $\Delta E = \hbar/\Delta t$ and vice versa.

4.11 Verification of Heisenberg Principle

4.11.1 γ -Ray Microscope Experiment

The meaning of the uncertainty principle can be illustrated by an idealized experiment. Consider the following attempt to measure simultaneously the position and momentum of an electron with an imaginary microscope of very high resolving power. It has been shown in wave optics that the *limit of resolution* of a microscope is given approximately by

$$\Delta x = \lambda/2 \sin A \quad (4.16a)$$

Now *resolving power* (R.P.) is proportional to $\frac{1}{\Delta x}$.

$$\text{Thus,} \quad \text{R.P. of Microscope} \propto \frac{1}{\Delta x} \propto 1/\lambda \quad (4.16b)$$

where Δx is the distance between two points which can just be resolved by the microscope, λ is the wave length of the light used, and A is the half angle of the lens used (Fig. 4.7a). Then Δx is the uncertainty in the determination of the position of the electron: to make Δx as small as possible, the wavelength (λ) used to illuminate the electron must be very small, i.e., either X-ray or γ -ray must be used. Now the position of an electron can be measured with a microscope, only if at least one photon is deflected from its initial direction into the microscope. As a result of this displacement, the contribution to the X-component of the momentum will be uncertain. Since the scattered photon can enter the microscope anywhere within the angle A . Thus, the uncertainty in the X-component of momentum of electron is

$$\Delta p_x = p \sin A - (-p \sin A) = 2p \sin A = \frac{2h}{\lambda} \sin A$$

Therefore, the product of the uncertainty in the simultaneous measurements of position and momentum of an electron is

$$\Delta x \cdot \Delta p_x = \frac{\lambda}{2 \sin A} \cdot 2h/\lambda \sin A = h$$

$$\text{Hence} \quad \Delta x \cdot \Delta p_x = h \geq \hbar$$

which is the uncertainty principle.

4.11.2 Single Slit Diffraction Method

Let us consider the electron beam with momentum p incident on the slit of width Δy . If Δy is comparable to the wave length of the electron beam, then the electrons will diffract according to single slit diffraction pattern and form a central maximum M_0 and two secondary minima at P_1 and P_2 (Fig. 4.7b). According to the diffraction theory the first order diffraction minima condition from equation (1.50) is

$$\Delta y \sin \theta = \lambda \quad \text{i.e., } \Delta y = \lambda / \sin \theta$$

where Δy = uncertainty in position of the electron before being diffracted, which are having momentum p only along X-axis.

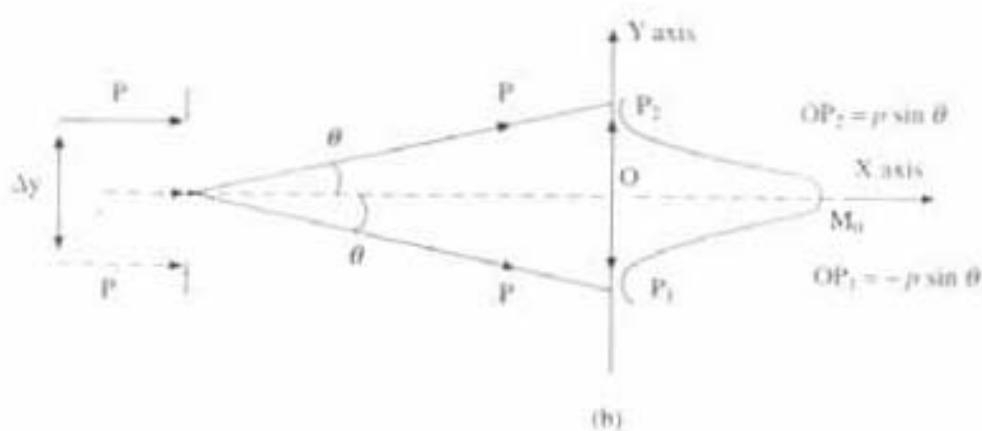
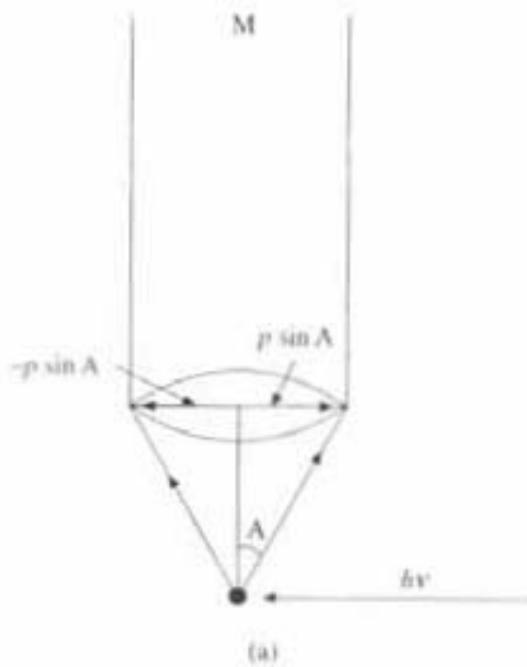


Fig. 4.7 Verification of Heisenberg uncertainty principle by (a) γ -ray experiment; (b) single slit diffraction method

Now before diffraction, electrons have momentum p only along X-axis but after diffraction they are having momentum along Y-axis also. Now due to diffraction when they are forming first order secondary minima at P_1 and P_2 then the component of momentum along Y-axis is $p \sin \theta$ and $-p \sin \theta$. So the uncertainty in momentum is

$$\Delta p_y = p \sin \theta - (-p \sin \theta) = 2p \sin \theta = \frac{2h}{\lambda} \sin \theta$$

Thus,

$$\Delta y \Delta p_y = \frac{\lambda}{\sin \theta} \cdot \frac{2h}{\lambda} \sin \theta = 2h$$

$$\therefore \Delta y \Delta p_y \geq h \quad \left(\text{where } h = \frac{h}{2\pi} \right)$$

which proves uncertainty principle.

4.12 Electron Microscope

An electron microscope in principle is similar to an optical microscope. In optical microscope visible light is used and focussing is done by suitable optical lenses. Whereas in an electron microscope a beam of electrons is used in place of ordinary light and focussing is done by electrostatic and electro-magnetic fields. Like an optical microscope, its purpose is also to magnify extremely minute objects. But magnifying power is more than 200 times greater than that of optical microscope, some times it is 10^5 times greater or more.

The *limit of resolution* for optical microscope i.e. the linear separation between two closest objects which can be resolved by Δx , according to equation (4.16a) and (Fig. 4.7a), is

$$\Delta x = \lambda/2 \sin A = \lambda_0/2\mu \sin A = \lambda_0/2 \text{ N.A}$$

where $\mu \sin A$ is the numerical aperture (N.A.) and $\lambda = \lambda_0/\mu$, when μ is the refractive index of the medium, λ_0 the wave length in vacuum, λ is the wave length in a medium of R.I. μ .

The *resolving power* of the microscope is reciprocal of Δx , i.e. the resolving power of an instrument is inversely proportional to the wave length of radiation (λ) used for illuminating the object under study. Thus *higher magnification and resolving power can be obtained by utilising waves of shorter wave length*.

According to de-Broglie's hypothesis, fast moving electrons which behave like a wave with extremely short wave length ($\leq 1 \text{ \AA}$), can be used instead of ordinary light in order to get higher magnification. The development of electron microscope was based on the fact that an electron accelerated through potential difference V volts is associated with wave lengths, $\lambda = (12.25/\sqrt{V}) \text{ \AA}$.

When $V = 60,000 \text{ V}$, $\lambda = 5 \times 10^{-10} \text{ cm} = 0.05 \text{ \AA}$, which is 10^5 times smaller in comparison to optical light. Obviously, therefore it will give much greater resolving power and magnification of the order of 100,000 times greater than that of optical microscope. But N.A. in the electron microscope is smaller than that of optical microscope. By using magnetic lenses electron beam can be focused, just as a lens focuses light, as we have discussed in sections 3.8 and 3.9, as electrostatic and magnetic focusing. In an optical microscope, objects having diameter less than 2000 \AA cannot be resolved, whereas in an electron microscope resolving power is 10 \AA , which is 200 times more than that of an optical microscope. Resolution of this order permits observations in the range of molecular dimensions.

Construction Principle of Electron Microscope

Magnetic electron microscope of transmission type which is used widely was first built by Ruska in 1934.

After that several modifications were done and very high degree of perfection has been reached in modern design. Figure 4.8(a) and (b) give the schematic diagrams of electron microscope and optical projection microscope, respectively. Magnetic lenses are used in the electron microscope.

An electron microscope is essentially a very large modified cathode ray tube, where magnetic lenses are used. The functions of the various parts in an electron microscope is similar to those of an optical microscope. An electron gun emits electrons which are accelerated by potential of about 60 kV. The microscope

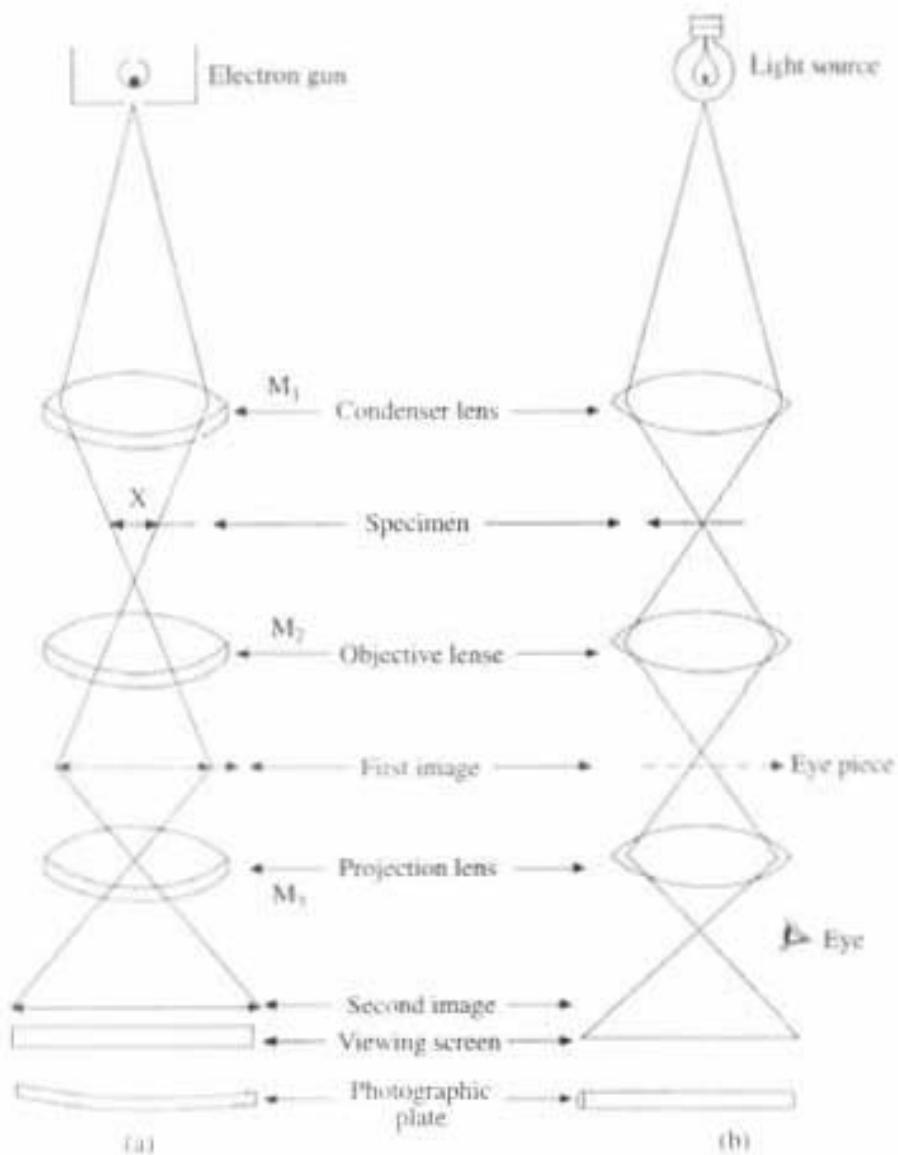


Fig. 4.8 (a) Electron microscope; (b) Optical microscope

enclosed in a rigid metal frame and high degree of vacuum is maintained with a high speed diffusion pump. Magnetic condenser lens M_1 concentrates the beam of electrons on the object specimen X which is to be magnified. Some of the electrons in the beam, which strike the object are stopped. Others scattered by the object traverse to the magnetic objective lens M_2 , which makes the electrons to diverge so that the first magnified image is formed. After that third magnetic projector lens M_3 focuses a small section of the magnified beam either in a viewing screen, usually made of a fluorescent material or on a suitable photographic plate for a permanent record. It is again possible to enlarge the photographs 4 to 5 times without loss of details, which gives pictures of about 500,000 or more times larger than the specimen. The method of focusing of electron beam consists of altering the intensity of the magnetic field in the coil of electromagnetic lens until a sharp image is seen on the screen. *In an electron microscope observer looking at the screen, sees not the magnified object but the magnified shadow of the object.* In addition to the higher degree of magnification power, there is another advantage, due to the great depth of focus of electrons, one gets magnification in three dimensions.

Uses of Electron Microscope

An electron microscope has a variety of uses in medicine, industry, crystal structure, analysis of functions of micro-organisms, etc. due to its high order of magnification, great resolving power and increased depth of focus.

In an electron microscope, the object is destroyed by electron bombardment. One of the method of overcoming this difficulty is to soak the specimen in a solution of osmium salts, so that even if the object burns due to electron bombardment, a mineral skeleton is still left for observation.

Study of textile fibers, purification of lubricating oils, composition of paper and paints, testing the surface of synthetic chemical plastics etc. are done using an electron microscope.

In the field of medicine and pathology, use of an electron microscope is significant. Study of different types of bacteria and viruses can be done in much greater detail with the help of electron microscope.

In the study of crystal structure analysis, electron microscope is used, with the help of electron diffraction unit in addition to the main instrument. This permits one to get a pattern of the atom as a set of concentric circles, from such pattern the constituent atoms in the material can be determined so that crystal structure can be identified and analysed.

PROBLEMS

1. An electron is confined to a spherical box of diameter 10^{-8} m. Calculate the minimum uncertainty in its velocity.

$$\text{Mass of electron} = 9.1 \times 10^{-31} \text{ kg}$$

$$\text{Planck's constant} = 6.625 \times 10^{-34} \text{ J.sec}$$

From uncertainty principle

$$\Delta x \cdot \Delta p \geq \hbar$$

Here $\Delta x = 10^{-8}$ m

$$\text{Thus } \Delta p \geq \hbar/\Delta x = \frac{6.625 \times 10^{-34}}{2 \times 3.14 \times 10^{-8}} = 1.055 \times 10^{-26} \text{ kgm/sec}$$

$$p = mu$$

Therefore, minimum uncertainty in velocity is

$$\Delta v = \Delta p/m = \frac{1.055 \times 10^{-26}}{9.1 \times 10^{-31}} = 0.116 \times 10^5 \text{ m/sec}$$

2. Find the K.E. and velocity of proton associated with de-Broglie's wavelength of 0.2865 \AA . Given that

$$\text{Mass of proton} = 1.67 \times 10^{-27} \text{ kg}$$

$$\text{Charge of proton} = 1.6 \times 10^{-19} \text{ C}$$

$$\text{Planck's constant} = 6.625 \times 10^{-34} \text{ J.sec}$$

de-Broglie's wavelength of proton is

$$\lambda = h/p = h/mv$$

$$\lambda = 0.2865 \text{ Å} = 0.2865 \times 10^{-10} \text{ m}$$

$$m = 1.607 \times 10^{-27} \text{ kg}$$

$$h = 6.625 \times 10^{-34} \text{ J.sec}$$

$$v = h/m\lambda = \frac{6.625 \times 10^{-34}}{1.67 \times 10^{-27} \times 0.2865 \times 10^{-10}}$$

$$v = 13.847 \times 10^5 \text{ m/sec} = 13.847 \text{ km/sec.}$$

$$\text{K.E. of proton} = 1/2 mv^2$$

$$= 1/2 \times 1.67 \times 10^{-27} \times (13.847 \times 10^5)^2$$

$$= 160.09 \times 10^{-21} \text{ J}$$

$$= 100 \times 10^{-2} \text{ eV} = 1 \text{ eV}$$

3. What is the wavelength of a beam of neutrons having: (1) an energy 0.025 eV? (2) an electron and photon each have a wavelength of 2 Å, what are their momentum and energy?

$$1. \quad \text{K.E. of neutron} = 0.025 \times 1.6 \times 10^{-19} \text{ J} = 0.04 \times 10^{-19} \text{ J}$$

$$\therefore \frac{1}{2} mv^2 = 0.04 \times 10^{-19}, v^2 = \frac{0.08 \times 10^{-19}}{1.676 \times 10^{-27}}$$

$$v = 2.185 \times 10^5 \text{ m/s}$$

Now, de-Broglie's wavelength, or neutron is

$$\lambda_n = \frac{h}{mv} = \frac{6.625 \times 10^{-34}}{1.676 \times 10^{-27} \times 2.185 \times 10^5}$$

$$\lambda_n = 1.8085 \times 10^{-10} \text{ m} = 1.808 \text{ Å}$$

2. Momentum of electron and Photon is $p = h/\lambda$

$$\therefore p = \frac{6.625 \times 10^{-34}}{2 \times 10^{-10}} = 3.311 \times 10^{-24} \text{ kgm/s}$$

$$(1) \text{ Velocity of electron} = \frac{p}{m} = \frac{3.311 \times 10^{-24}}{9.1 \times 10^{-31}} = 3.638 \times 10^6 \text{ m/s}$$

$$\begin{aligned} \text{Energy of electron} &= \frac{1}{2} mv^2 = \frac{1}{2} p v \\ &= \frac{3.311 \times 10^{-24} \times 3.638 \times 10^6}{2 \times 1.6 \times 10^{-19}} \\ &= 37.646 \text{ eV} \end{aligned}$$

(2) Energy of Photon is

$$E = h\nu = \frac{hc}{\lambda} = \frac{6.625 \times 10^{-34} \times 3 \times 10^8}{2 \times 10^{-10}}$$

$$= 9.9375 \times 10^{-16} \text{ J}$$

$$E = 6.2109 \times 10^3 \text{ eV.}$$

4. A charged particle accelerated by a p.d. of 200 V has a de-Broglie wave length equal to 0.0202 \AA . Find the mass of this particle and say which particle is it?

Charge of the particle = $1.6 \times 10^{-19} \text{ C}$

Planck's constant = $6.625 \times 10^{-34} \text{ J.sec}$

When the particle is accelerated by a p.d. V then

$$eV = \frac{1}{2} mv^2, mv = \sqrt{2meV}$$

Hence, de-Broglie's wavelength $\lambda = h/p$.

$$\lambda = h/mv = h/\sqrt{2meV}$$

$$m = \frac{h^2}{2\lambda^2 eV} = \frac{(6.625 \times 10^{-34})^2}{2 \times (0.0202 \times 10^{-10})^2 \times 1.6 \times 10^{-19} \times 200}$$

Thus, mass of the particle = $m = 1.68 \times 10^{-29} = 1.68 \times 10^{-27} \text{ kg}$. Hence, it is a proton.

5. Calculate the de-Broglie wavelength of neutron whose energy is 1 eV.

Mass of neutron = $1.676 \times 10^{-27} \text{ kg}$

Charge of electron = $1.6 \times 10^{-19} \text{ C}$

Planck's constant = $6.622 \times 10^{-34} \text{ J.sec}$

Energy of neutron = 1 eV = $1.6 \times 10^{-19} \text{ J}$

$$\frac{1}{2} mv^2 = 1.6 \times 10^{-19} \text{ Joule}$$

$$v^2 = \frac{2 \times 1.6 \times 10^{-19}}{1.676 \times 10^{-27}}$$

$$v^2 = 1.909 \times 10^8, v = 1.38 \times 10^4 \text{ m/sec.}$$

Hence, de-Broglie's wavelength is

$$\lambda = h/mv = \frac{6.625 \times 10^{-34}}{1.676 \times 10^{-27} \times 1.38 \times 10^4}$$

$$\lambda = 2.86 \times 10^{-11} \text{ m} = 0.286 \times 10^{-10} \text{ m}$$

$$\lambda = 0.286 \text{ \AA.}$$

6. Explain why electron cannot exist in the nucleus of radius 10^{-14} m .

According to the uncertainty principle

$$\Delta x \cdot \Delta p \geq \hbar$$

where Δx = uncertainty in the position of a particle.

Δp = uncertainty in its momentum.

Suppose that the electron can stay in the nucleus. The uncertainty Δx in the position of an electron is roughly the same as the diameter of the nucleus which is of the order of $2 \times 10^{-14} \text{ m}$. Then

$$\Delta p = \hbar / \Delta x = \frac{1.0545 \times 10^{-34}}{2 \times 10^{-14}} = 0.5 \times 10^{-20}$$

From the uncertainty in the momentum, it is possible to get a rough estimate of energy of an electron in the nucleus. Relativistic formula must be used because the electron moves very rapidly in the nucleus, as will be seen.

Now from the well known Einstein's relativistic relation we know

$$E = mc^2 = \text{K.E.} + \text{Rest Mass Energy} = \frac{1}{2}mv^2 + m_0c^2$$

when the velocity of the particle is very high then the mass of the particle will change with velocity as

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}}$$

where m = mass of particle with velocity v

m_0 = rest mass of the particle

c = velocity of light.

Therefore,

$$E = mc^2$$

$$E = \frac{m_0c^2}{\sqrt{1 - v^2/c^2}}$$

$$E^2 = p^2c^2 + m_0^2c^4$$

Now if it is assumed that $\Delta p = p$, then

$$p = 0.5 \times 10^{-20}$$

and

$$\begin{aligned} p^2c^2 &= (0.5 \times 10^{-20})^2 (3 \times 10^8)^2 \\ &= 2.25 \times 10^{-24} \text{ J} \end{aligned}$$

This value is greater than the term $m_0^2c^4 = (9 \times 10^{-31})^2 (3 \times 10^8)^4$

i.e.

$$m_0^2c^4 = 6.5 \times 10^{-27} \text{ J}$$

So we can neglect the 2nd term, i.e. $= 10^{-27}$. Then

$$E^2 = 2.25 \times 10^{-24}$$

$$E = 1.5 \times 10^{-12} \text{ Joule}$$

$$E = \frac{1.5 \times 10^{-12}}{1.6 \times 10^{-19}}$$

$$E = 1 \times 10^7 \text{ eV}$$

$$E = 10 \text{ MeV}$$

Thus according to this result, a free electron confined within a space as small as the nucleus would have to have kinetic energy of the order of 10 MeV and a velocity greater than 0.999 c . But experimentally the electron emitted by radioactive nuclei have never been found to have kinetic energies greater than about 4 MeV or of that order, which is smaller than the calculated value from the uncertainty principle based on Proton-electron hypothesis. Although this calculation is a rough one, but similar results are obtained from more rigorous calculations and in view of large discrepancy it seems improbable that nuclei can contain free

electrons. If we consider the uncertainty relation $\Delta x \cdot \Delta p \geq h$ instead of \hbar , then velocity of electron will come 60 MeV, which is very much bigger than experimental value, which proves electron cannot stay in nucleus.

7. Calculate De-broglie's wavelength associated with following particles and comments on the result.

- Mass of 60 micrograms travelling with velocity 80 m/s
- Mass of 8×10^{-27} kg travelling with velocity 13×10^4 km/s
- $m = 60 \text{ microgram} = 60 \times 10^{-6} \times 10^{-3} = 60 \times 10^{-9} \text{ kg}$
and $v = 80 \text{ m/s}$. So momentum $p = mv = 4.8 \times 10^{-6} \text{ kg. m/s}$

$$\text{So De-broglie's wave length } \lambda = \frac{h}{p} = \frac{6.625 \times 10^{-34}}{4.8 \times 10^{-6}} = 1.38 \times 10^{-28} \text{ m}$$

$$(ii) m = 8 \times 10^{-27} \text{ kg}, v = 13 \times 10^4 \text{ km/s} = 1.3 \text{ m/s}$$

$$\text{So De-broglie's wave length } \lambda = \frac{h}{P} = \frac{6.625 \times 10^{-34}}{10.4 \times 10^{-27}} = 6.37 \times 10^{-8} \text{ m} = 6.37 \text{ Å}$$

Comments: In first case for a normal particle of 60 microgram mass, De-broglie's wave is so small which is beyond our perception. Due to that it is not apparent for daily life.

8. Calculate kinetic energy of an electron whose De-broglie's wavelength is same as that of 100 kev X-ray photon. Given that $\hbar = 6.63 \times 10^{-34} \text{ J-S}$, $c = 3 \times 10^8 \text{ m/s}$, $e = 1.6 \times 10^{-19} \text{ C}$, $m = 9.1 \times 10^{-31} \text{ kg}$. Energy of photon = $E = 100 \text{ kev} = 100 \times 10^3 \times 1.6 \times 10^{-19} \text{ J} = h\nu = \frac{hc}{\lambda}$

$$\text{i.e., wave length of photon} = \lambda = \frac{6.63 \times 10^{-34} \times 3 \times 10^8}{1.6 \times 10^{-19}} = 12.43 \times 10^{-12} \text{ m}$$

$$\text{So De-broglie's wave length of electron} = \lambda = \frac{h}{mv} = 12.43 \times 10^{-12} \text{ m}$$

$$\text{i.e., velocity of electron} = v = \frac{h}{m\lambda} = 0.0586 \times 10^9 \text{ m/s}$$

$$\text{So kinetic energy of electron} = \frac{1}{2} mv^2 = 1.56 \times 10^{-15} \text{ J} = 9.76 \text{ KeV}$$

QUESTIONS

- State and explain Heisenberg's uncertainty principle. Derive the expression for de-Broglie's wave length of matter wave.
- Describe the Davisson-Germer experiment for the confirmation of wave nature of electron.
- How matter waves differ from light and sound waves? Give difference only.
- Prove that electron cannot exist in nucleus of an atom. Given radius of atom is 10^{-14} m .
- Calculate the de-Broglie wavelength of a proton whose kinetic energy is equal to the rest energy of an electron. Mass of proton is 1836 times greater than that of an electron. Given that

Velocity of light = $3 \times 10^8 \text{ m/s}$, Mass of electron = $9.1 \times 10^{-31} \text{ kg}$
Planck's constant = $6.62 \times 10^{-34} \text{ J. sec}$

Solution: Rest energy of electron = $m_0 c^2$

$$\text{K.E. of proton} = \frac{1}{2} m v_p^2$$

$$9.1 \times 10^{-31} \times (3 \times 10^8)^2 = 1/2 \times 9.1 \times 10^{-31} \times 1836 \times v_p^2$$

$$v_p^2 = \frac{2 \times (3 \times 10^8)^2}{1836}, \quad v_p = 9.9 \times 10^6 \text{ m/s}$$

de-Broglie's wavelength of proton

$$\lambda_p = \frac{h}{mv_p} = \frac{6.62 \times 10^{-34}}{9.1 \times 10^{-31} \times 1836 \times 9.9 \times 10^6}$$

$$\lambda_p = 4.00 \times 10^{-14} = 0.0004 \times 10^{-10} \text{ m}$$

$$\lambda_p = 0.0004 \text{ Å}$$

6. Write down Einstein's equation for photo-electric effect. What is threshold frequency?
7. State Heisenberg's uncertainty principle and illustrate it by an experiment on diffraction at a single slit.
8. Why is an electron microscope so powerful than ordinary microscope? Compare working principle between electron and optical microscopes with diagram.
9. Compute the minimum uncertainty in the location of a mass of 2 gm, moving with a speed of 1.5 m/s and minimum uncertainty in the location of an electron moving with a speed of 0.5×10^8 m/s, given that the uncertainty in the momentum p for both is

$$\Delta p = 10^{-3} p, \text{ Mass of electron} = 9.1 \times 10^{-31} \text{ kg}$$

$$h = 6.625 \times 10^{-34} \text{ mks-unit (J, sec)}$$

10. Is an electron a particle? Is it a wave? Explain your answer by giving relevant experimental evidence.
11. An electron microscope was made by using short wave length of electrons to provide high resolution. Then is it possible to make a proton microscope and a neutron microscope? Explain your answer in each case.
12. "If an electron is localised in space, its momentum becomes uncertain. If it is localised in time, its energy becomes uncertain". Explain this statement.
13. Why is resolving power of an electron microscope 10^6 times higher than that of an optical microscope? Explain with proper mathematical expression.
14. If the particles listed below have same energy, which has the shortest wave length: electron, α -particle, neutron or proton?
15. A proton and electron have the same speed, which has the longer wave length?
16. Why is the wave nature of matter not more apparent in our daily observation?
(*Hints:* Since in our daily observation mass of the particle is very large in comparison to electron or proton mass, so de-Broglie's wave length will come very small, less than atomic size, which is beyond our perception so it has not become apparent in our daily life.)
17. Find the uncertainty in the location of a particle, in terms of the de-Broglie's wave length λ , so that the uncertainty in its velocity is equal to its velocity itself.

18. Explain Compton effect with a neat diagram.
19. During photo-electron collision, Compton shift is found to be 0.012 \AA . If wave length of the incident photon is 0.05 \AA , find out the angle between the direction of motion of the scattered photon and of the recoil electron.

SCHRÖDINGER WAVE EQUATION⁶

4.13 The Wave Function

The de-Broglie's relation between momentum and wave length, which is known experimentally to be valid for both photons and other high velocity particles, suggests that it might be possible to use concentrated bunches of waves to describe localized particles of matter and quanta of radiation, or moving particles can be viewed as a wave and one can measure its wave length. Now the question is: what is the quantity whose variation in time and space makes up this wave?

To fix our ideas we shall consider a wave function $\psi(x, t)$, that depends on space co-ordinates x, y, z , and time t . In the case of waves on a stretched string, the wave function $\psi(x, t)$ represents the transverse displacement y of the string. For sound wave it represents the differential pressure Δp and for the electromagnetic wave it represents the electric field vector E and magnetic field vector H . So for matter waves representing the particles we can also introduce in the same way the wave function $\psi(x, t)$. Say a conduction electron is moving through copper wire or an electron moving about the nucleus of a hydrogen atom, whatever be the case, if we know the wave function $\psi(x, y, z, t)$ for every point in space and for every instant of time, we will come to know all about the behaviour of the particle.

A typical form for a concentrated bunch of waves which we shall call a 'wave packet' is as shown in Fig. 4.9, where $\psi(x, t)$ is plotted against x for a particular time t . The average wave length λ and approximate extension of the packet is as shown in Fig. 4.9. This quantity $\psi(x, t)$, is assumed to have three basic properties. First, it can interfere with itself, so that it can account for the diffraction experiment. Second, it is large in magnitude where the particle or photon is likely to be, and small elsewhere. And third $\psi(x, t)$ describes the behaviour of a single particle or photon, not the statistical distribution of a number of such quanta.

4.14 Physical Significance of Wave Function

When electromagnetic wave travels through free space then according to Maxwell's equation the arrangement of electric and magnetic fields that varies in space and time and the rate per unit area of which, energy is transported by the wave, is proportional to E^2 , where E is the amplitude of electric field vector. In the same way when a photon is moving with the speed of light then, according to Einstein, the rate of energy transferred is proportional to the average number of photon per unit volume of the beam, where each photon is having an energy $h\nu$. Now from the connection between wave and particle picture of radiation, we can say that the square of the electric field intensity is a direct measure of the average density of photon.

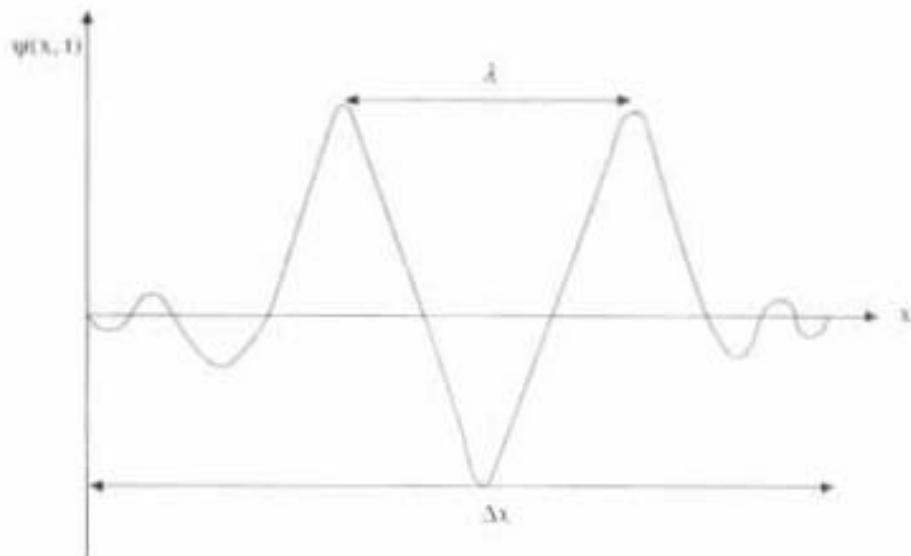


Fig. 4.9 Wave packet

Normalisation condition:

Now if $\psi(x, t)$ is the amplitude of the matter wave at any point in space then the particle density of that point may be taken as proportioned to $\psi^2(x, t)$. This $\psi^2(x, t)$ is the measure of the particle density. Max Born suggested a new idea about the physical significance of $\psi(x, t)$ which is generally accepted now-a-days. According to Max Born, if at instant t , a measurement is made, to locate the particle associated with the wave function $\psi(x, t)$ then the probability $P(x, t) dx$, that the particle will be found at a co-ordinate x and $x + dx$ is

$$\begin{aligned} P(x, t) dx &= \psi^*(x, t)\psi(x, t) dx \\ P(x, t) dx &= |\psi(x, t)|^2 dx \end{aligned} \quad (4.19)$$

where $\psi^*(x, t)$ is the complex conjugate of $\psi(x, t)$, because $\psi(x, t)$ is complex in general but probability must be real and non-negative.

Hence, $P(x, t) = |\psi(x, t)|^2 = \psi^*(x, t)\psi(x, t)$

is called the '*probability density*', which is the probability of locating the particle per unit interval, around x on the x -axis.

In three dimensional case probability density is written as

$$P(r, t) = \psi^*(r, t)\psi(r, t) = |\psi(r, t)|^2$$

This means that $P(r, t)dx dy dz = P(r, t)dv$, is to be the probability of finding a particle in its volume element $dx dy dz = dv$ about its point r at time t .

Probability of finding the particle somewhere in the region must be unity, which gives the condition

$$\int_{-\infty}^{+\infty} |\psi(r, t)|^2 dv = 1 \quad (4.20)$$

which $\psi(r, t)$ must satisfy. This condition is called *normalisation condition* and a wave of $\psi(r, t)$ which satisfies it, is called *normalised wave function*.

4.15 Development of Wave Equation

According to de-Broglie's theory, a particle of mass m is always associated with a wave whose wave length is given by $\lambda = \frac{h}{mv}$. If the particle has wave properties, it is expected that there should be some sort of wave equation, which describes the behaviour of the particle. Earlier we have discussed that $\psi(x, y, z, t)$, is the wave function which represents the wave displacement of the particle whose co-ordinates are (x, y, z) at time t .

Now we re-write the two equations $E = hv$ and $p = h/\lambda$ in terms of universal constant $\hbar = h/2\pi$, as follows:

$$p = \frac{h}{\lambda} = \frac{h}{2\pi} \frac{2\pi}{\lambda} = \hbar k \quad \text{where } k = \frac{2\pi}{\lambda} \quad (4.21)$$

$$E = hv = \frac{h}{2\pi} (2\pi v) = \hbar w, \quad \text{where } w = 2\pi v \quad (4.22)$$

where k is called *propagation vector* and w is called *angular frequency* of the wave.

A one dimensional wave function $\psi(x, t)$, that represents a particle of completely undetermined position travelling in the positive x -direction, with precisely known momentum p and K.E. E , would then be expected to have one of the following forms or some linear combination of them, as it is given classically for the solution of Maxwell's, equation for \vec{E} and \vec{H}

$$\boxed{\begin{aligned} \psi(x, t) &= A \cos(kx - wt) = A \cos 2\pi \left(\frac{x}{\lambda} - vt \right) \\ \psi(x, t) &= A \sin(kx - wt) = A \sin 2\pi \left(\frac{x}{\lambda} - vt \right) \\ \psi(x, t) &= Ae^{\pm i(kx - wt)} = Ae^{\pm i2\pi \left(\frac{x}{\lambda} - vt \right)} \end{aligned}} \quad (4.23)$$

4.16 Phase Velocity and Group Velocity

The velocity of propagation of the wave is called its phase velocity or wave velocity and is shown by

$$\boxed{v_p = \frac{w}{k}} \quad (4.24)$$

It is the velocity with which the phase of the wave or say a crest or a trough travels (Fig. 4.10).

The de-Broglie's matter waves can be described as a concentrated bunch of waves, as wave packets or wave group, to describe the localised particles of matter (Fig. 4.10) and quanta of radiation, as follows from diffraction experiments like those of—Davisson and Germer and of Thomson. Now the question is whether the wave group or wave packet travels with the individual wave velocity (v_p) or not?

Alternate constructive and destructive interference of the two individual waves $\psi_1(x, t)$ and $\psi_2(x, t)$ produces a slowly moving envelope or group or packet of waves as shown in Fig. 4.10. The velocity of the wave packet is called the group velocity (v_g), which is given as:

$$v_g = \frac{dw}{dk} \quad (4.25)$$

where dw is the difference of angular frequency between waves ψ_1 and ψ_2 and dk the difference of propagation vector between waves ψ_1 and ψ_2 .

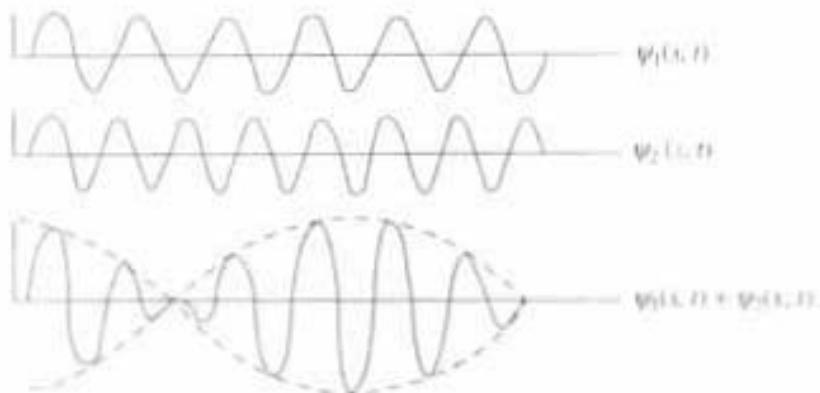


Fig. 4.10 Formation of wave packet from number of individual waves

Now $E = \hbar w$ and $p = \hbar k$, from which we can calculate the value of v_g , the group velocity.

Energy and momentum are also related by

$$E = \frac{p^2}{2m} \quad \text{and} \quad p = mv$$

$$\hbar w = \frac{1}{2m} \hbar^2 k^2 = \frac{\hbar^2 k^2}{2m}$$

Therefore

$$w = \frac{\hbar k^2}{2m}$$

$$dw = \frac{2\hbar k}{2m} dk$$

Now

$$v_g = \frac{dw}{dk} = \frac{\hbar k}{m} = \frac{p}{m} = \frac{mv}{m} = v$$

i.e.

$$v_g = v \quad (4.26)$$

Hence, Group velocity = Particle velocity, which concludes that *de-Broglie's wave group or wave packet has the same velocity as the particle*.

From the relativistic equation also we can come to the same conclusion. According to theory of relativity;

$$E^2 = p^2 c^2 + m_0^2 c^4$$

where m_0 = rest mass of the particle.

Therefore

$$\hbar^2 w^2 = \hbar^2 k^2 c^2 + m_0^2 c^4$$

Now group velocity

$$v_g = \frac{dw}{dk}$$

Differentiating above equation we get

$$\hbar^2 2w dw = \hbar^2 2k dk c^2 \quad \text{or} \quad \frac{dw}{dk} = \frac{k}{w} c^2 = \frac{\hbar k}{\hbar w} c^2$$

But

$$E = \hbar w = mc^2$$

Thus

$$\frac{dw}{dk} = \frac{\hbar k}{mc^2} c^2 = \frac{\hbar k}{m} = \frac{p}{m} = v$$

$$v_g = dw/dk = v$$

i.e. group velocity (v_g) is the same as particle velocity (v).

Whereas phase velocity (v_p) or wave velocity is

$$v_p = \frac{w}{k} = \frac{mc^2}{\hbar} = \frac{\hbar}{mv}$$

$$\left(\because E = \hbar w = mc^2, w = \frac{mc^2}{\hbar}, p = \hbar k, k = \frac{p}{\hbar} = \frac{mv}{\hbar} \right)$$

i.e.

$$v_p = \frac{c^2}{v} > c$$

which shows that the phase velocity of the particular matter wave (v_p) is more than the velocity of light. But according to the theory of relativity, no particle can travel at a speed more than c , the speed of light. Hence phase velocity or wave velocity (v_p) has no relation to the speed of the particle. Thus V_p has no physical significance.

From above information we have seen that the wave velocity or phase velocity (v_p) is greater than particle velocity v as well as more than c , the velocity of light, it does not mean that de-Broglie's wave get away from the particle. In fact the entire wave packet (i.e. the particle now) travels with group velocity (v_g), which is same as particle velocity (v), whereas, individual waves emposing the wave packet, travel with the wave velocity (v_p), which is different from group velocity v_g . Group velocity (v_g) is the same as particle velocity (v).

4.17 Schrödinger Wave Equation (Time Dependent)

In 1926 Erwin Schrödinger inspired by de-Broglie's concept, derived a remarkably successfully theory on wave mechanics. Its central feature is a differential equation now known as *Schrödinger equation*, which governs the variation of wave function ψ in space and time, which describes the behaviour of particle for a wide range of problems in the Atomic field.

Drawing analogy with classical wave equation (4.23), we assume the wave associated with a free particle can be represented by

$$\begin{aligned}\psi(x, t) &= Ae^{\pm 2\pi} \left(\frac{x}{\lambda} - vt \right) \\ &= Ae^{\pm i(kx - \omega t)}\end{aligned}\quad (4.27)$$

where $\psi(x, t)$ = displacement at time t at a distance x from origin

A = amplitude or maximum displacement from the mean position

λ = wave length

v = frequency of vibration

k = propagation vector $= \frac{2\pi}{\lambda} = \frac{p}{\hbar}$

w = angular frequency of the wave $= 2\pi v$

By differentiating the above equation twice we get

$$\frac{d^2\psi(x, t)}{dx^2} + \frac{4\pi^2}{\lambda^2} \psi(x, t) = 0 \quad (4.28)$$

Now if we consider the motion of electron, then the above equation can be applied in the following way. The total energy of the electron is

$$E = \text{K.E.} + \text{P.E.} = 1/2 mv^2 + V$$

or

$$mv^2 = 2m(E - V)$$

Now de-Broglie's wave length is

$$\lambda = \frac{\hbar}{mv} = \frac{\hbar}{\sqrt{2m(E - V)}}$$

Hence, equation (4.28) becomes

$$\frac{d^2\psi(x, t)}{dx^2} + \frac{8\pi^2 m}{\hbar^2} (E - V) \psi(x, t) = 0$$

Now by substituting $\hbar = \frac{h}{2\pi}$, above equation becomes,

$$\frac{d^2\psi(x, t)}{dx^2} + \frac{2m}{h^2} (E - V) \psi(x, t) = 0$$

(4.29)

Now from equation (4.27) we get

$$\frac{\partial\psi}{\partial t} = -2\pi iv\psi = -iw\psi = -\frac{iE}{\hbar}\psi \quad \text{and} \quad \frac{\partial\psi}{\partial x} = i\frac{2\pi}{\lambda}\psi = ik\psi = \frac{ip}{\hbar}\psi$$

so $\frac{\partial\psi}{\partial t} = -i\frac{E}{\hbar}\psi$ and $\frac{\partial\psi}{\partial x} = i\frac{p}{\hbar}\psi$

i.e.

$$E = i\hbar \frac{\partial}{\partial t}$$

and

$$\bar{p} = -i\hbar \frac{\partial}{\partial x}$$

Thus, in one dimensional case

$$E = i\hbar \frac{\partial}{\partial t} \text{ is a differential operator for energy } E \text{ and is a scalar quantity.} \quad (4.30)$$

$$\vec{p} = -i\hbar \frac{\partial}{\partial x} \text{ is a differential operator for momentum } \vec{p}, \text{ and is a vector quantity.} \quad (4.31)$$

By using these two differential operator, equation (4.29) becomes

$$\frac{-\hbar^2}{2m} \frac{\partial^2 \psi(x, t)}{\partial x^2} + V\psi(x, t) = i\hbar \frac{\partial \psi(x, t)}{\partial t} \quad (4.32)$$

We now assume that this equation holds good even when potential energy term V is a function of x and t . Thus we have

$$\boxed{\frac{-\hbar^2}{2m} \frac{\partial^2 \psi(x, t)}{\partial x^2} + V(x, t) \psi(x, t) = i\hbar \frac{\partial \psi(x, t)}{\partial t}} \quad (4.33)$$

This is the *Schrödinger time dependent* wave equation, for the one dimensional case. In three dimensional case *Schrödinger time dependent* wave equation can be modified as

$$\boxed{\frac{-\hbar^2}{2m} \nabla^2 \psi(r, t) + V(r, t) \psi(r, t) = i\hbar \frac{\partial \psi(r, t)}{\partial t}} \quad (4.34)$$

where $\nabla^2 = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right)$ is called the Laplacian operator.

$$\psi(r, t) = \psi(x, y, z, t)$$

$$V(r, t) = V(x, y, z, t)$$

From the normalisation condition (4.20) we know that

$$\int_{-\infty}^{+\infty} P(r, t) dv = \int_{-\infty}^{+\infty} |\psi(r, t)|^2 dv = 1 \quad (4.20)$$

Now for all physically acceptable problems the wave function $\psi(r, t)$ should be normalised and well behaved. It should follow the following conditions:

1. The wave function $\psi(r, t)$ must be finite, single valued and continuous. In particular, $\psi(r, t) \rightarrow 0$ as $r \rightarrow \pm\infty$, otherwise the integral $\int_{-\infty}^{+\infty} \psi^*(r, t) \psi(r, t) dv$ will be infinite and normalisation will not be possible.
2. $\frac{\partial \psi(r, t)}{\partial r}$ must be finite, single valued and continuous.

If the above conditions are not satisfied, the expectation value of the position vector of a particle $\langle r \rangle = \int \psi^*(r, t) r \psi(r, t) dV$ will not be finite and single valued, they may be absurd. So the wave function $\psi(r, t)$ must be well behaved for all physical problems.

4.18 Schrödinger (Time Independent) Wave Equation (STIE)

Schrödinger time dependent wave equation (4.32) is a fundamental equation of wave mechanics, but from the practical point of view sometimes we have to be concerned with a simpler equation, which is called *Schrödinger time independent wave equation*.

In case when the P.E. $V(r)$ is not dependent on time, then we need *Schrödinger time independent wave equation* (STIE), which can be written as

$$\frac{-\hbar^2}{2m} \nabla^2 \psi(r, t) + V(r) \psi(r, t) = i\hbar \frac{\partial \psi(r, t)}{\partial t} \quad (4.35)$$

The solution of the above equation can be written in a separable form as

$$\psi(r, t) = \psi(r) \varphi(t) \quad (4.36)$$

where $\psi(r)$ is a function of space variable (r) only and $\varphi(t)$ is a function of time variable (t) only.

Hence,

$$\nabla^2 \psi(r, t) = \nabla^2 \psi(r) \varphi(t)$$

and

$$\frac{\partial \psi}{\partial t}(r, t) = \psi(r) \frac{\partial \varphi(t)}{\partial t}$$

by substituting these in equation (4.35) we get

$$\frac{-\hbar^2}{2m} \varphi(t) \nabla^2 \psi(r) + V(r) \psi(r) \varphi(t) = i\hbar \psi(r) \frac{\partial \varphi(t)}{\partial t}$$

Dividing throughout by $\psi(r) \varphi(t)$ we get

$$\frac{1}{\psi(r)} \left[\frac{-\hbar^2}{2m} \nabla^2 \psi(r) + V(r) \psi(r) \right] = \frac{i\hbar}{\varphi(t)} \frac{\partial \varphi(t)}{\partial t} \quad (4.37)$$

The left hand side of equation (4.37) is only dependent of r , the space variable, whereas the right hand side of equation (4.37) is only dependent on time variable t , so by the method of separation of variable, both sides will be equal to the same separation constant, which we can call as E . Then the equation for $\varphi(t)$ is readily integrated to give

$$\varphi(t) = e^{-iEt/\hbar} \quad (4.38)^*$$

by substituting equation (4.38) on equation (4.37) the equation of $\psi(r)$ becomes

$$\left[\frac{-\hbar^2}{2m} \nabla^2 + V(r) \right] \psi(r) = E \psi(r)$$

(4.39)

This is called *Schrödinger Time Independent Wave Equation (STIE)*. Its solution $\psi(r)$ determined the space dependance of $\psi(r, t)$ through the Schrödinger time dependent (4.34) wave equation as

$$\psi(r, t) = \psi(r)e^{-iEt/\hbar} \quad (4.40)$$

4.19 Eigen Value Equation

The Schrödinger time independent equation (STIE) (4.39) can be written as

$$H\psi(r) = E\psi(r) \quad (4.41)$$

where H is the *Hamiltonian operator* $= \left[-\frac{\hbar^2}{2m} \nabla^2 + V(r) \right]$, operates on the function $\psi(r)$. The effect of the Hamiltonian operator (H) on wave function is to reproduce the same function $\psi(r)$ multiplied by a constant E . Such an equation is called an *Eigen Value Equation*. The function $\psi(r)$ on which Hamiltonian Operator (H) acts is called an *Eigen function* of the operator, and the constant E with which the Eigen function gets multiplied is called an "*Eigen Value*".

Physical Concept of Eigen Value Equation

The Schrödinger time independent wave equation (4.39) and its solution $\psi(r)$ are essentially similar to the time independent wave equation of classical wave motion and its solutions. In classical wave, certain frequencies arise naturally in the course of their solutions. Now since $E = hv$, in wave mechanics certain particular energies will arise naturally. This is called "quantisation of energy" and these particular allowed values of energies are to be said as characteristics or proper or "eigen values (E)". Corresponding to each allowed eigen value E , there is only one function $\psi(r)$ obtained as solution of STIE equation (4.39). This function $\psi(r)$ is called as characteristics function or "eigen function".

One should clearly remember the difference between the eigen function $\psi(r)$ and the wave function $\psi(r, t)$. The wave function is

$$\psi(r, t) = \psi(r)e^{-iEt/\hbar}$$

whose space dependance part is determined by the eigen function $\psi(r)$.

Just as the wave function $\psi(r, t)$ must be well behaved, Eigen function $\psi(r)$ must also be well behaved. Since $\psi(r, t) = \psi(r)e^{-iEt/\hbar}$, then function $\psi(r, t)$ will misbehave if $\psi(r)$ misbehaves. Thus,

- (1) eigen function $\psi(r)$ must be finite, single valued and continuous.
- (2) $\frac{\partial \psi(r)}{\partial r}$ must be finite, single valued and continuous.

Above two conditions should be fulfilled for all physical problems so that expectation values of the position vector $\langle r \rangle$ will be finite, single valued, and consistant, they should not be absurd.

4.20 Stationary State

In a particular state if the probability distribution function

$$P(r, t) = \psi^*(r, t) \psi(r, t)$$

is independent of time, then the state of the system is called *stationary state*. Now if the particle may be described by just one of the wave function $\psi_n(r, t)$ with energy eigen value E_n , then

$$\psi(r, t) = \psi_n(r, t) = \psi_n(r) e^{-iE_n t/\hbar}$$

Then the probability distribution for the particle is

$$\begin{aligned} P(r, t) &= \psi^*(r, t) \psi(r, t) \\ &= \psi_n^*(r) e^{iE_n t/\hbar} \psi_n(r) e^{-iE_n t/\hbar} \end{aligned}$$

Hence

$$P(r, t) = \psi_n^*(r) \psi_n(r) \quad (4.42)$$

which is independent of time or called *stationary*. Therefore, the relative probability of finding the particle, does not depend on time. In this case, the particle is said to be in a *stationary state* or *eigen state* of the potential $V(r)$.

4.21 Applications of Schrödinger Time (STIE) Independent Equation

4.21.1 Particle in a One Dimensional Box: One dimensional infinite rectangular potential well

Consider a particle say electron, of mass m which can move freely in a one dimensional box of length L . The potential energy (V) at the bottom of the box is zero but as the electron cannot escape from the box, so the impenetrable wall of the box is represented by a potential $V = \infty$. Now the particle is trapped in the region $0 < x < L$ of the X -axis. This region is called potential well, as shown in Fig. 4.11(a).

Thus,

$$V(x) = 0 \text{ for } 0 \leq x \leq L$$

$$= \infty \text{ for } x < 0 \text{ and } x > L$$

In the present case we have to solve STIE for the region $0 < x < L$, where $V = 0$. Hence, we get from equation (4.39)

$$\frac{-\hbar^2}{2m} \frac{\partial^2 \psi(x)}{\partial x^2} - E\psi(x) = 0$$

or
$$\frac{\partial^2 \psi(x)}{\partial x^2} + k^2 \psi(x) = 0 \quad \text{where } k = \sqrt{\frac{2mE}{\hbar^2}} \quad (4.43)$$

A suitable solution of above Schrödinger equation is

$$\psi(x) = A \sin kx + B \cos kx \quad (4.44)$$

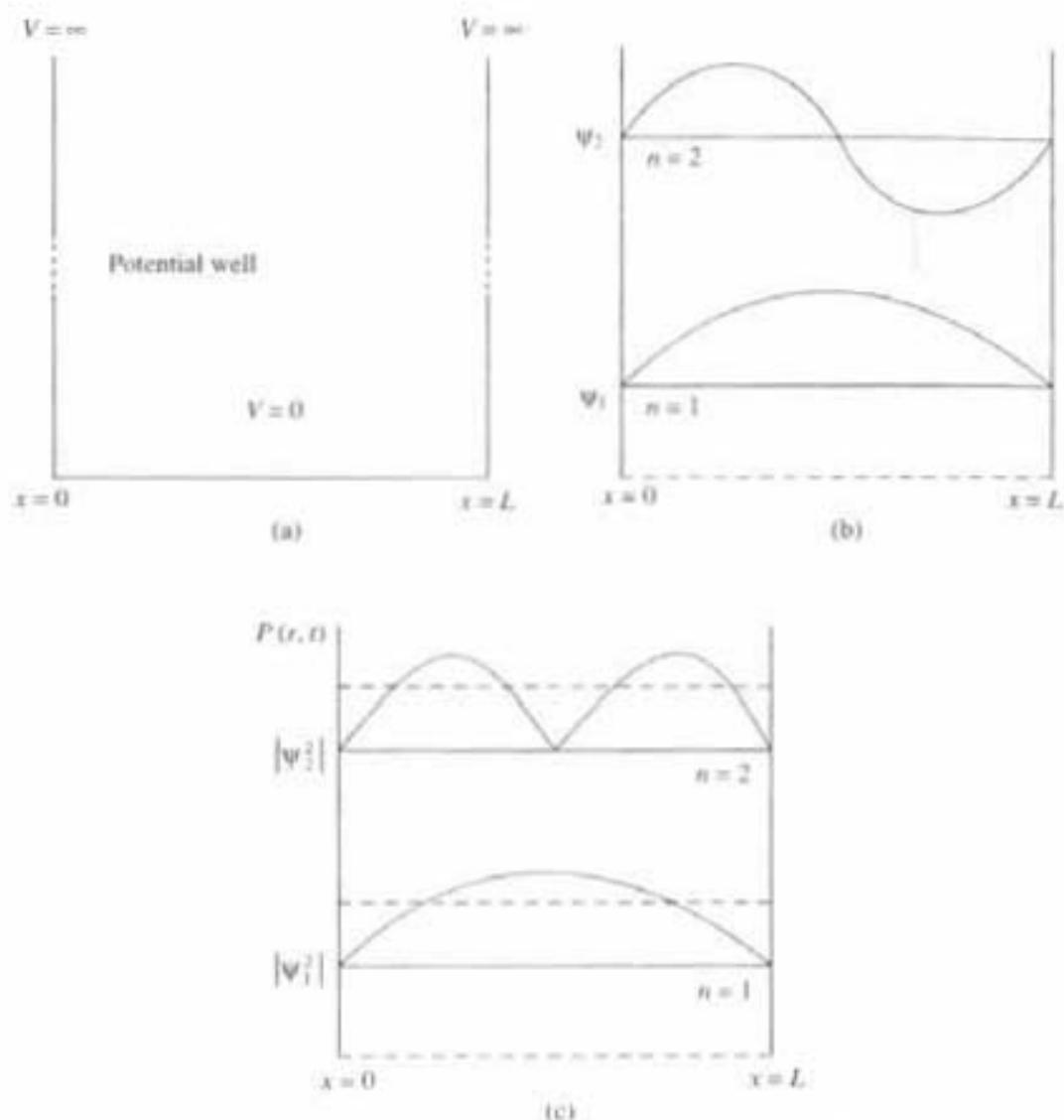


Fig. 4.11 (a) Infinite potential well; (b) Wave function $\psi(x)$ for first two eigen functions in infinite potential well; (c) Probability density $|\psi(x)|^2$ in infinite potential well

The boundary conditions are

1. $\psi(x) = 0$ at $x = 0$
2. $\psi(x) = 0$ at $x = L$

By applying these two boundary conditions in equation (4.44) we get

$$0 = A \sin 0 + B \cos 0, \text{ i.e. } B = 0$$

Next $0 = A \sin kL$ which gives $kL = n\pi$.

or

$$k = \frac{n\pi}{L} \quad (4.45)$$

Equating equations (4.43) and (4.45), we get

$$\sqrt{\frac{2mE}{\hbar^2}} = \frac{n\pi}{L}$$

or

$$E = E_n = \left(\frac{\pi^2 \hbar^2}{2mL^2} \right) n^2 \quad (4.46)$$

where $n = 1, 2, 3 \dots$, the principal quantum numbers.

This equation shows that the particle in the one dimensional box, or in a one dimensional infinite potential well, cannot have any arbitrary energy. That trapped particle can only have discrete energies given by

$$\begin{array}{ll}
 25E_1 & E_5 \quad n = 5, E = E_5 = \left(\frac{\pi^2 \hbar^2}{2mL^2} \right) 5^2 = 25E_1 \\
 \\
 16E_1 & E_4 \quad n = 4, E = E_4 = \left(\frac{\pi^2 \hbar^2}{2mL^2} \right) 4^2 = 16E_1 \\
 \\
 9E_1 & E_3 \quad n = 3, E = E_3 = \left(\frac{\pi^2 \hbar^2}{2mL^2} \right) 3^2 = 9E_1 \\
 \\
 4E_1 & E_2 \quad n = 2, E = E_2 = \left(\frac{\pi^2 \hbar^2}{2mL^2} \right) 2^2 = 4E_1 \\
 \\
 E_1 & E_1 \quad n = 1, E = E_1 = \left(\frac{\pi^2 \hbar^2}{2mL^2} \right) 1^2 \\
 \\
 & E = 0
 \end{array}$$

Fig. 4.12 Discrete energy levels for a particle bound in a one-dimensional infinite potential well

and so on, as shown in Fig. 4.12. Thus the energy of the particle in infinite well is *quantised*.

Since the particle can exist only in discrete levels, the energy that is emitted or absorbed by the particle can also have some certain discrete values, corresponding to the difference in energy levels. This speciality of transition energy is one of the most striking features of atomic physics. $E_1, E_2 \dots$ etc. are called the

energy eigen values. The lowest possible energy is $E_1 = \left(\frac{\pi^2 \hbar^2}{2mL^2} \right)$, not zero,

according to wave mechanics or quantum mechanics. This is called '*zero point energy*', it is the lowest energy that the particle have if it is bound in a one dimensional box of width L . The particle cannot have zero kinetic energy, because from Heisenberg uncertainty principle, the uncertainty in position of the trapped particle in a box is $\Delta x = L$, hence Δp as well as velocity of the particle and their kinetic energy cannot be zero.

Figure 4.11(b) shows the plot of $\psi(x)$, for first two eigen functions of the particle in the infinite well and Fig. 4.11(c) shows the corresponding probability density $[P(r, t)] = |\psi(x)|^2$, i.e. the probability of finding the particle in a box.

In the mode of oscillation with $n = 1$, the probability density is highest at $x = L/2$ and the smallest next to the walls. In $n = 2$ mode, probability density is minimum at $x = L/2$ and maximum at $L/4$ and $3L/4$ as shown in Fig. 4.11(c). The horizontal straight line shows the classical expectation value for the probability density.

In case of completely free particle say, electron in a metal, it possesses the kinetic energy $E = 1/2 mv^2$. The de-Broglie's wave length associated with the particle is $\lambda = \frac{h}{mv}$, the wave number is given by $k = \frac{2\pi}{\lambda}$. Now, combining $E = \frac{mv^2}{2}$ and $\lambda = \frac{h}{mv}$ with wave number k , we get

$$v = \frac{h}{m\lambda} \quad \text{and} \quad \lambda = \frac{2\pi}{k}$$

Hence $E = \frac{1}{2} mv^2 = \frac{mh^2}{2m^2\lambda^2} = \frac{\hbar^2 k^2}{2m4\pi^2}$

or

$$E = \frac{\hbar^2 k^2}{2m} \quad (4.47)$$

The above equation shows a continuous parabolic relationship between E and k , for a free particle, say for an electron in a metal, as shown in Fig. 4.13. This means, *for free particle the energy distribution is continuous, whereas for bound particle the energy is quantised* as we have seen earlier in Fig. 4.12. This can be explained in the following way.

For bound particle in a one dimensional box of width L , where L has a definite but small value, the energy levels from Eqn. (4.46) as shown in Fig. 4.12 are widely spaced. But for free particle $L = \infty$, so from equation (4.46) the spacing of energy levels becomes infinitesimally small. So they appear almost like a continuous spectrum with energy $\frac{\hbar^2 k^2}{2m}$, as shown in Fig. 4.13.

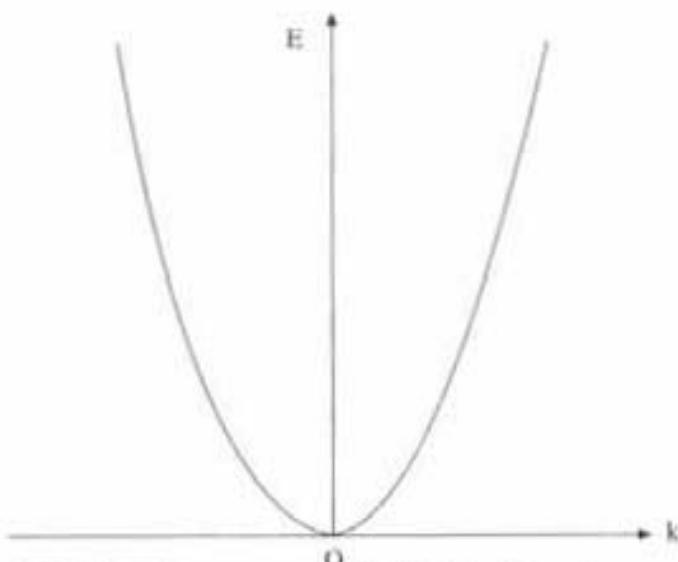


Fig. 4.13 Continuous energy distribution for a free particle

4.21.2 Barrier Tunneling: Rectangular finite potential barrier of height V_0

Let us consider a particle (from a beam of mono-energetic particles), incident from the left on the rectangular finite potential (Fig. 4.14a). The finite potential is described by

$$\begin{aligned} V(x) &= 0 && \text{For } x < 0 && \text{Region-I} \\ V(x) &= V_0 && \text{For } 0 \leq x \leq L && \text{Region-II} \\ V(x) &= 0 && \text{For } x > L && \text{Region-III} \end{aligned}$$

where E is the energy of the incident particle from left on the rectangular potential barrier.

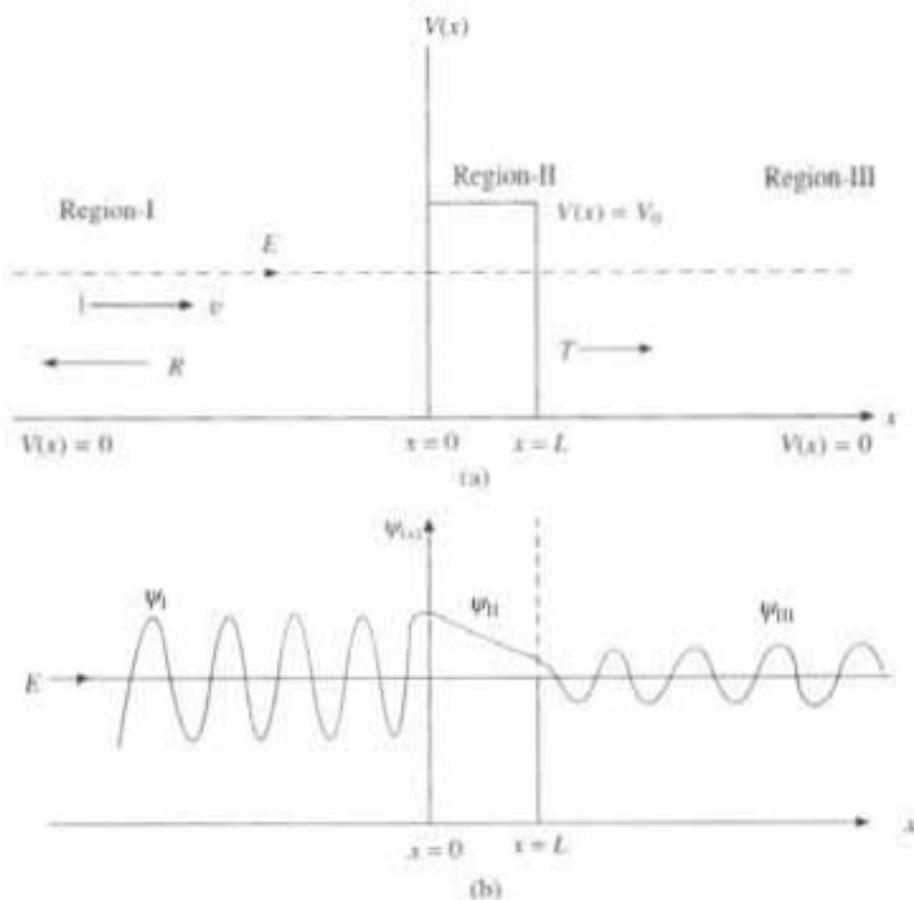


Fig. 4.14 Barrier tunnelling

Classically if $E > V_0$, the particle can enter region-II from region-I, but its energy will be slowed down to a value $(E - V_0)$ and it will travel with velocity

$$v = \sqrt{\frac{2(E - V_0)}{m}}. \quad \text{The particle is thus transmitted into region-II.}$$

But if $E < V_0$, the particle cannot transmit to the region-II, due to potential $V_0 > E$ it will be then reflected back to region-I again without loss of speed.

Quantum mechanically/wave mechanically we can treat this problem nicely and a major quantum surprise awaits for us, if we consider a more realistic case

of a well of finite depth, instead of infinite depth as it was discussed in section 4.21.1. Quantum mechanical result is obtained by solving STIE for the particle in region I and II, as shown in Fig. 4.14.

In region I, $V(x) = 0$. Hence, STIE becomes as Eqn. (4.43)

$$\frac{\partial^2 \psi(x)}{\partial x^2} + \frac{2mE}{\hbar^2} \psi(x) = 0$$

or $\frac{\partial^2 \psi(x)}{\partial x^2} + k_1^2 \psi(x) = 0$ where $k_1 = \sqrt{\frac{2mE}{\hbar^2}}$ (4.48)

The solution of equation (4.48) in region I is

$$\psi_I(x) = Ae^{ik_1 x} + Be^{-ik_1 x} \quad (4.49)$$

where the first term is the incoming wave moving along (+ve) x direction whereas second term is the reflected wave moving along (-ve) x direction. Equation (4.49) is an oscillating Sine wave as shown in region-I in Fig. (4.14b).

In region II, $V(x) = V_0$. Hence, STIE Eqn. (4.39) becomes

$$\frac{\partial^2 \psi(x)}{\partial x^2} + \frac{2m}{\hbar^2} (E - V_0) \psi(x) = 0$$

$$\frac{\partial^2 \psi(x)}{\partial x^2} + k_2^2 \psi(x) = 0$$

where

$$k_2 = \sqrt{\frac{2m(E - V_0)}{\hbar^2}} \quad (4.50)$$

Solution of equation (4.50) is

$$\psi_{II}(x) = Ce^{ik_2 x} \quad (4.51)$$

no wave is reflected back from region II, so there is no need to take $e^{-ik_2 x}$ term.

Case I: When $E > V_0$, $k_2 = \frac{2m(E - V_0)}{\hbar^2}$ is real, so the eigen function $\psi_I(x)$ and $\psi_{II}(x)$ and their derivatives must be continuous at the boundary $x = 0$, which is common for both the regions.

Hence $\psi_I(x)|_{x=0} = \psi_{II}(x)|_{x=0}$ and $\left. \frac{\partial \psi_I(x)}{\partial x} \right|_{x=0} = \left. \frac{\partial \psi_{II}(x)}{\partial x} \right|_{x=0}$

Therefore $[Ae^{ik_1 x} + Be^{-ik_1 x}]_{x=0} = [Ce^{ik_2 x}]_{x=0}$

which gives

$$A + B = C$$

and

$$ik_1 [Ae^{ik_1 x} - Be^{-ik_1 x}]_{x=0} = [ik_2 Ce^{ik_2 x}]_{x=0}$$

or

$$k_1(A - B) = k_2 C \quad \text{or} \quad A - B = \frac{k_2}{k_1} C$$

by combining above two equations, we get

$$A = C \left[\frac{k_1 + k_2}{2k_1} \right], B = C \left[\frac{k_1 - k_2}{2k_1} \right] \quad (4.52)$$

Thus, $\frac{B}{A} = \frac{(k_1 - k_2)}{(k_1 + k_2)}$

Now the probability of reflection or reflection coefficient is

$$R = \frac{\text{Reflected probability flux} = B^* B v}{\text{Incident probability flux} = A^* A v}$$

where v is the velocity of incoming particle in region I.

Now, $R = \frac{B^* B}{A^* A} = \frac{B^2}{A^2} = \frac{(k_1 - k_2)^2}{(k_1 + k_2)^2}$

or $R = \left[\frac{1 - \sqrt{\frac{E - V_0}{E}}}{1 + \sqrt{\frac{E - V_0}{E}}} \right]^2 \quad (4.53)$

Since the incident particle gets either reflected or transmitted, so the probability of reflection plus probability of transmission is equal to unity.

Therefore

$$R + T = 1$$

$$T = 1 - R = 1 - \left[\frac{(k_1 - k_2)}{(k_1 + k_2)} \right]^2 = \frac{4 \sqrt{\frac{E - V_0}{E}}}{\left[1 + \sqrt{\frac{E - V_0}{E}} \right]^2} \quad (4.54)$$

Case II: $E < V_0$

Classically we have already seen that if $E < V_0$, then the particle cannot transmitted to the region II, it will be reflected back again in region I. No transmission is possible.

But now quantum mechanical treatment of that problem will give different result, which is described as follows. This is the beauty of the quantum mechanical treatment.

When $E < V_0$, then $k_2 = \sqrt{\frac{2m(E - V_0)}{\hbar^2}}$

is an imaginary quantity. Say $k_2 = ir$, where r is real. Therefore solution in region II will be

$$\psi_{II}(x) = Ce^{ik_2 x} = Ce^{i(r)x} = Ce^{-rx} \quad (4.55)$$

This is an exponentially damped wave as shown in Fig. (4.14b). Although damped but the eigen function $\psi_{II}(x)$ has penetrated into the classically forbidden region II. The probability density in region-II is given by

$$p_{II} = \psi_{II}^*(x)\psi_{II}(x) = C^* Ce^{-2rx} \quad (4.56)$$

This decreases rapidly with increasing x as shown in Fig. (4.14b) in region-II, but there is a finite probability of finding the particle in region II. This is in contrast to the classical result, where it is completely impossible to find the particle in region II, when $E < V_0$. Penetration to the classically forbidden region is one of the most striking feature of wave mechanics or quantum mechanics.

The value of the transmission co-efficient (T) is very sensitive to the thickness and height of the barrier and to the factor k_2 , the wave number.

Now for this case when $k_2 = ir$,

$$\frac{B}{A} = \frac{k_1 - ir}{k_1 + ir} \quad \text{and reflection Co-efficient} \quad R = \frac{BB^*}{AA^*} = 1$$

Thus the incident particle is always reflected, in agreement with classical result. But we see again from equation (4.52),

$$C = \frac{2k_1 A}{k_1 + k_2} = \frac{2k_1 A}{k_1 + ir} \quad \text{i.e.} \quad C^* C = \frac{4k_1^2}{k_1^2 + r^2} |A|^2$$

This means C -the amplitude of the transmitted wave in region-II is not zero unless r and hence V_0 is infinite from equation (4.50), so there is a finite probability of finding the particle in the classically forbidden region $x > 0$ for finite potential barrier. This is called *barrier tunnelling* phenomena. Thus in wave mechanics or quantum mechanics there is a finite chance that the electron can penetrate the barrier and continue its motion to the right as shown in Fig. 4.14b.

In region-III $V(x) = 0$, same as in region-I. So the solution of STIE in region-III will be of same type as in region-I like Eqn. (4.49)

$$\text{i.e.} \quad \psi_{\text{III}}(x) = D e^{ik_1 x} \quad (4.57)$$

$\psi_{\text{III}}(x)$ is also oscillatory but with smaller amplitude as shown in Fig. (4.14b). It is also so drawn that at each boundary ($x = 0$ and $x = L$), $\psi(x)$ and $d\psi/dx$ become continuous. Probability density in region-III will be

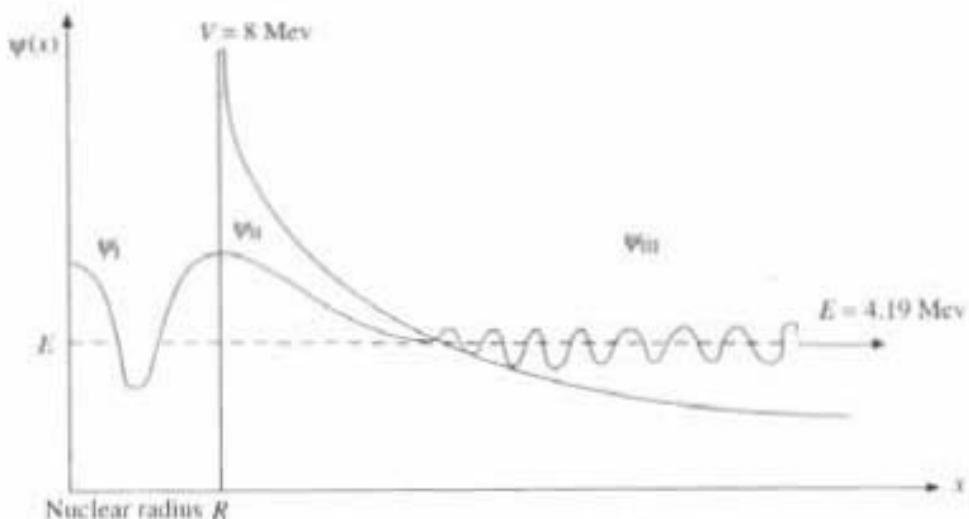
$$P_{\text{III}} = D^* D \neq 0$$

This means that the incident particle with $E < V_0$ has finite probability to tunnelled through the rectangular barrier potential and arrived in region III as a free particle with smaller amplitude.

Thus we have seen that if the rectangular barrier potential had a finite height, $\psi(x)$ is not zero on the other side of the barrier. Which means the particle could tunnel through the classically excluded region and come out on the other side, which is called the "*barrier penetration*" and practically this can clearly explain the working condition of tunnel diode and α -decay from U^{238} .

4.21.2.1 α -decay

The practical example of this tunnel effect is observed in the α -decay of U^{238} nucleus, which then gets converted into thorium²³⁴ nucleus. The K.E. of the outgoing α -particle is 4.19 MeV, whereas it overcomes the electrostatic barrier potential height of 8 MeV, due to the attractive force of the nucleus of radius R (Fig. 4.15), which is not possible to explain by classical mechanics. The very

Fig. 4.15 α -decay from U-238 nucleus

well known problem of α -decay can be explained only by wave mechanical *barrier tunnelling phenomena*. According to wave mechanics eigen function $\psi_I(x)$ is oscillatory within attractive potential, within nucleus ($0 < x < R$) as shown in Fig. (4.15), then exponentially decaying in the barrier region $\psi_{II}(x)$ and again oscillatory $\psi_{III}(x)$ outside the nucleus after penetration, with smaller amplitude.

4.21.2.2 Tunnel diode

Current flow under small forward bias condition in *tunnel diode* is also possible, which can also be explained by this *barrier penetration phenomenon*. In tunnel diode, heavily doped *n*-type semiconductor is separated by heavily doped *p*-type semiconductor by a very thin depletion layer as shown in Fig. 4.16. Fermi level E_{Fn} is located inside C.B. for *n*-type material and E_{FP} inside V.B. for *p*-type material. Whatever be the temperature, the Fermi levels (E_{FP} and E_{Fn}) will line up in the absence of external voltage. Therefore tunneling can take place where the two bands overlap in their energy ranges. When a small forward bias is applied, the electron, can tunnel through the barrier potential, from the C.B. of *n*-side to V.B. of *p*-side. In the barrier depletion region, electron energy (E) is less than the barrier potential height (V_0), of finite value, as shown in Fig. (4.16).

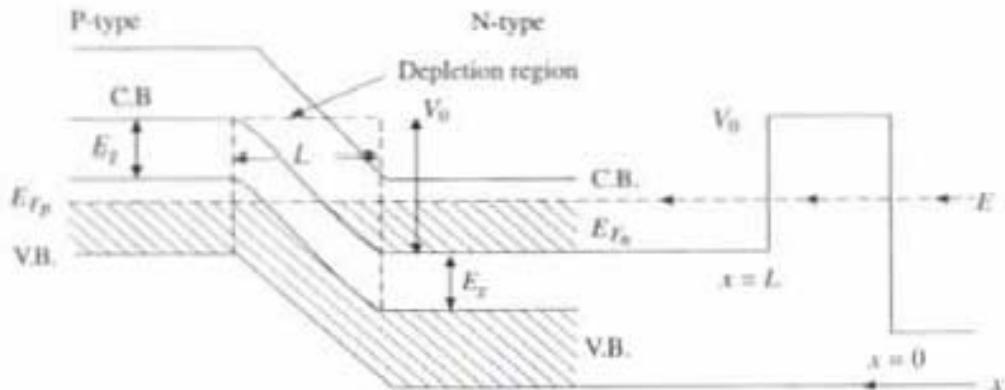


Fig. 4.16 Tunnel diode.

So classically the electrons do not have enough energy to overcome the potential barrier, so they remain trapped in the *n*-side, under very small forward bias condition. But quantum mechanically electrons can be tunneled through the bottom of the C.B. on the *n*-side region into the similar energy state at the top of the V.B. of *p*-side region. In a similar way holes can be tunneled in the opposite direction. Thus under small forward bias condition current flow in the tunnel diode can be explained quantum mechanically, which is not possible classically. The electron eigen function is oscillatory in the *n*-region, it is exponentially decaying in the barrier region and again becomes oscillatory after penetration with smaller amplitude, in the other side of the barrier, i.e. in *p*-region, just as in Fig. 4.15. Tunneling current is zero when there is no applied voltage, since E_{Fn} and E_{Fp} would linedup. With increase of forward bias, tunneling current will increase but upto certain value, *after that as forward voltage is increased tunneling ceases, since there is no available states on the opposite side i.e. on p side, to receive them*. This property of tunnel diode makes it useful as an amplifier, an oscillator or as a switch. Thus the current flow through the tunnel diode can be rapidly turned on or off by controlling the height of the barrier potential, i.e. by varying the external bias voltage. This can be done with a very short response time of the order of 10^{-11} second, so that the device is suitable for applications where speed of response is critical. This action may be called as *quantum switching* which is based on tunnelling.

QUESTIONS

1. Give Max-Born's interpretation of wave function.
2. Explain the physical interpretation of wave function. Explain what is meant by normalisation of wave function?
3. Explain eigen functions and eigen values.
4. State the requirements of Schrödinger wave function and derive time independent schrödinger equation for a particle moving in a finite potential V .
5. What is called barrier tunnelling? Explain tunnel diode operation with neat diagram.

Laser, Holography and Fiber Optics

LASER

5.1 Introduction—Laser

Principle of *laser* (light amplification by stimulated emission of radiation) is similar to *Maser* (Microwave Amplification by Stimulated Emission).

The light emitted from a conventional light source is said to be incoherent because the radiation emitted from different atoms do not in general, bear any definite phase relationship with each other. On the other hand, the light emitted from a laser has a very high degree of coherency and is almost perfectly parallel to each other. The theory of laser was proposed in 1958 by Townes and Schawlow in U.S.A. The first laser was fabricated by Maiman in 1960.

In laser the fundamental processes of interaction are as follows:

5.1.1 Absorption of Radiation

We know that light is absorbed or emitted by particles during their transition from one energy level to another. In general, at normal condition all the atoms are situated in the ground state (E_1) (Fig. 5.1a). Thus, when the energy $h\nu = (E_2 - E_1)$ is supplied from outside in the favourable condition, the atom will absorb that energy ($E_2 - E_1$) and go to the upper energy state E_2 . This process is called *absorption of radiation* (Fig. 5.1a). If more energy is supplied from outside, atom can absorb more energy and go to upper energy state E_3 , E_4 , etc.

5.1.2 Emission of Radiation

Atoms which are in the excited state always try to come down to the ground state by emitting the extra energy, in form of emission of radiation. Atoms can emit radiation in two ways:

5.1.2.1 Spontaneous Emission of Radiation

Consider Fig. 5.1(b) where the atom initially is in the excited state E_2 . It may spontaneously emit energy $h\nu_1 = E_2 - E_1$ and come down to state E_1 . Now the photons having energy $h\nu_1 = E_2 - E_1$ are emitted in all possible directions and in random phases, i.e., the emitted light is highly incoherent in the case of spontaneous emission. If the atom initially is in excited state E_3 , then it will emit energy $h\nu_2 =$

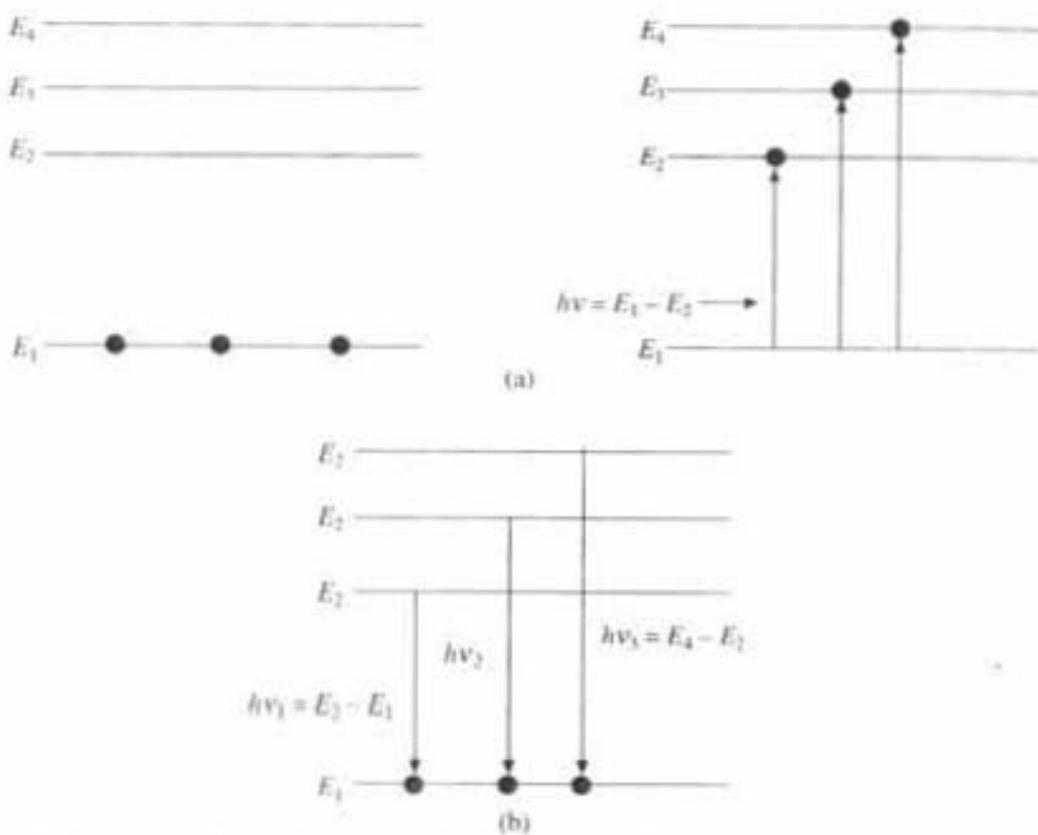


Fig. 5.1 (a) Absorption of radiation; (b) Spontaneous emission of radiation

$E_3 - E_1$ by spontaneous emission or it may emit $h\nu_3 = E_4 - E_1$, when atoms will come down from excited state E_4 to E_1 . Thus these spontaneously emitted photons are:

- (1) not monochromatic, because they are having different energy $h\nu_1$, $h\nu_2$, etc.
- (2) highly incoherent, because they are emitted at random phase
- (3) not unidirectional, because they are emitted in all possible directions.
- (4) so emitted energy/unit area is very less. Example of spontaneous emission are sunlight, ordinary light from incandescent lamp, tube, etc.

5.1.2.2 Stimulated Emission of Radiation and Lasing Action

In stimulated emission, transition of atom from higher excited state to lower state is strictly restricted between two energy levels. Out of that the higher energy level is generally "metastable", which is having longer life time ($= 3 \times 10^{-3}$ sec), in comparison to ordinary state and transition from *metastable state* to lower state is not automatically possible, according to quantum mechanical selection rule. So spontaneous emission from metastable state is not possible. Here atoms are initially at the excited metastable state E_2 (Fig. 5.2). Now if a photon having a specific energy $h\nu = E_2 - E_1$ collides with the excited atom in metastable state E_2 , then this collision can force or stimulate the atom to drop into the lower level E_1 . Due to this transition a new photon of energy $h\nu = E_2 - E_1$ will be emitted, in addition to the incident photon having the same energy. *This type of emission is called stimulated emission because it is stimulated by the incident photon.*

The most important point about stimulated emission is that the emitted photon is exactly identical to the incident photon as regards energy, direction of travel, phase and state of polarisation, etc. Now we have two identical photons starting

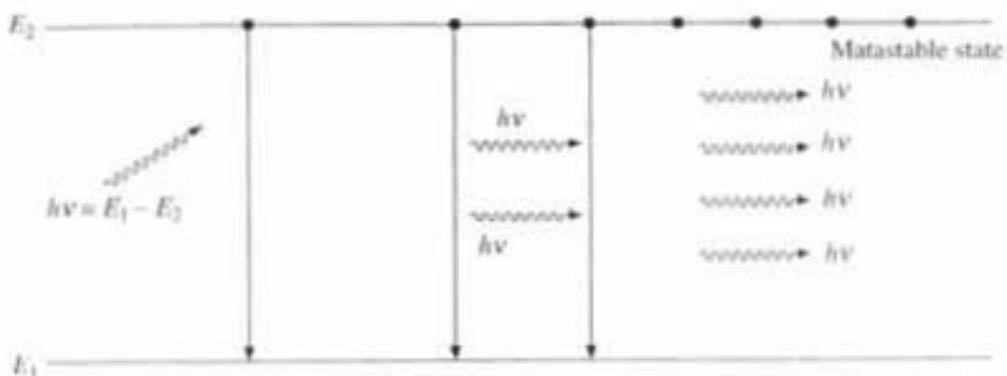


Fig. 5.2 Stimulated emission of radiation

from one photon of selected frequency ν . If these two photons could be made to (Fig. 5.2) collide with more excited atoms in level 2, then we would get multiple of identical photons as 4, 8, 16, and so on. Such process of increasing the number of photons is indeed possible in laser, which is called *lasing action*. Thus photons emitted by stimulated emission in laser are:

- (1) monochromatic because all are having same energy $h\nu$.
- (2) highly coherent because there is a particular time difference between two collision.
- (3) highly unidirectional otherwise they will be stopped by laser tube (Fig. 5.3).
- (4) the emitted energy/unit area is very high

Now in order to get lasing action, the following are essential:

1. A large number of excited atoms in level 2.
2. Photons should not be lost to an appreciable extent by absorption in atoms in the lower energy level E_1 or in any other way.

5.2 Population Inversion and Pumping

In general the number of atoms in the ground state is greater than the number of atoms in the excited states. But in order to get lasing action, it is required that we must have large number of atoms in the excited state than in the ground or lower state. Thus for stimulated emission to take place, we must artificially increase the population of the atoms in higher energy state. This is called as *population inversion*.

Population inversion is achieved by supplying energy to the system, so that atoms will absorb a requisite amount of energy and go from the lower level to higher level. This supply of energy to the medium is called *pumping*. A system in which population inversion is achieved is called an *active system*.

5.3 Operating Principle in Laser

Once population inversion is achieved, spontaneously emitted photons can initiate the process of amplification because a large number of active centres or atoms in the required higher energy state are available. In order to make large number of collisions between the photons present and the active centres, active medium is to be placed between two highly reflecting mirrors M_1 and M_2 (Fig. 5.3).

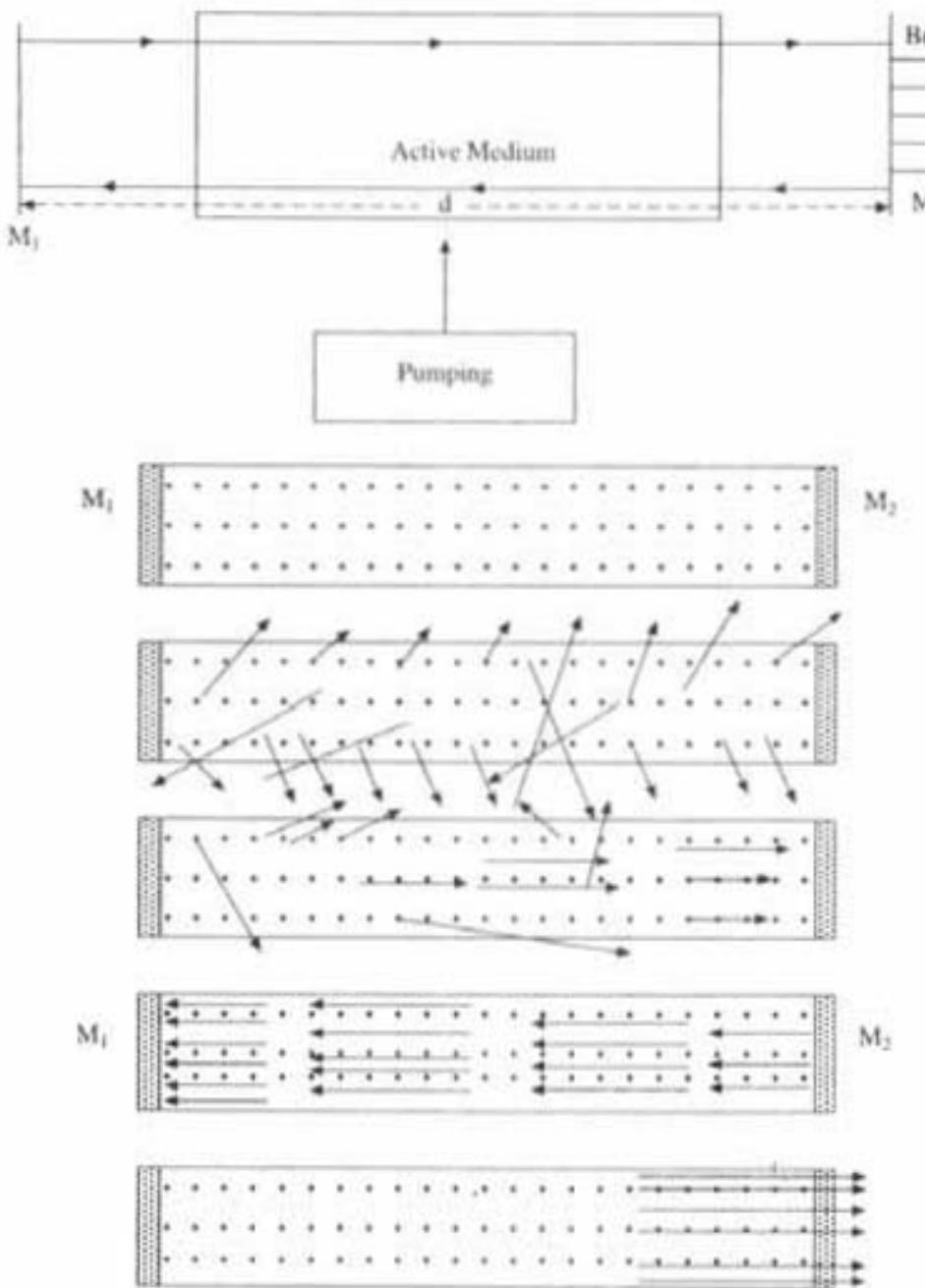


Fig. 5.3 Experimental set up for starting lasing action

which is known as resonator. Thus in between the two mirrors, waves propagate along both directions and interfere with each other. Due to that stable standing wave pattern will be formed, condition for that being the cavity length ($= d$) must be integral multiple of $\lambda/2$. Thus $d = n\lambda/2$, i.e.

i.e.

$$\lambda = 2d/n, \text{ where } n = 1, 2, 3 \dots$$

Therefore, resonant frequency $= f = c/\lambda = n(c/2d)$.

Those photons that travel parallel to the axis of the arrangement will travel to and fro between the two mirrors on account of their related reflections. Every time they pass through the active medium, some of them collide with excited atoms and give rise to stimulated emission of photons of same frequency, that way the number of photons goes on increasing enormously and lasing action

starts. If one of the mirrors is made partially transparent then a very powerful parallel beam of photons will come out of it. Photons which are not parallel to the central axis will be lost due to interaction with the wall surface.

Thus the following basic requirements, which must be satisfied for laser operation are:

- (1) There must be an active medium which emits radiation in the required region of the electromagnetic spectrum
- (2) Population inversion must be created within the medium; this, in turn requires the existence of suitable energy levels (i.e. E_1 , E_2) associated with lasing transition for pumping.
- (3) For true laser oscillation there must be optical feedback at the ends of the medium to form resonant cavity in order to get highly powerful and collimated, monochromatic, coherent laser beam.

5.4 Characteristic Properties of Laser Light

The most important of the properties of laser beam are its high degree of:

(1) *Directionality*: Photons which are moving parallel to the central axis can only be sustained. Others will be lost. Thus the beam emerges along the direction of the axis. They are almost perfectly parallel to each other.

(2) *Monochromaticity*: This property is due to two reasons: (a) Only an electromagnetic wave of fixed frequency (determined by the energy difference $E_2 - E_1$) can be amplified. (b) two mirror arrangement forms a resonant cavity, only its resonant frequencies which depend upon the distance between two mirror, can be sustained.

(3) *Coherency*: It is highly coherent since the amplification is due to stimulated emissions. Because of the high degree of coherency associated with laser beam, it is used in many diverse areas like holography, data processing, communications through fiber optic cable, etc.

(4) *Brightness*: Due to the directionality property, brightness, that is the power emitted per unit surface area per unit solid angle from laser source is very large. If laser beam is focused by a convex lens at some region then the intensity will be tremendous at that place. This high intensity can be used effectively to 'drill' through a metallic target. Now in medical world for surgery, it is also used extensively.

5.5 Ruby Laser

Ruby laser is made from a single cylindrical crystal of ruby, whose ends are flat and one of the ends is completely silvered and the other end is partially silvered.

In the original set up of Maiman, a flash lamp filled with xenon gas was connected to a capacitor (Fig. 5.4a), which was charged to a few kilovolts. The energy stored in the capacitor was discharged through the Xenon lamp in a few millisecond. This results in a power, which is a few megawatts and coded as *Optical pumping* since it is given by light energy. Some of this energy is absorbed by the active centres of the ruby rod resulting in their excitation and subsequent lasing action. Ruby consists of aluminium oxide (Al_2O_3) with some of the aluminium ions (Al^{+3}) replaced by chromium ions (Cr^{+3}). These chromium ions are the active centres in the ruby laser. For a good lasing action, the ruby crystal consists of about 0.05% of chromium.

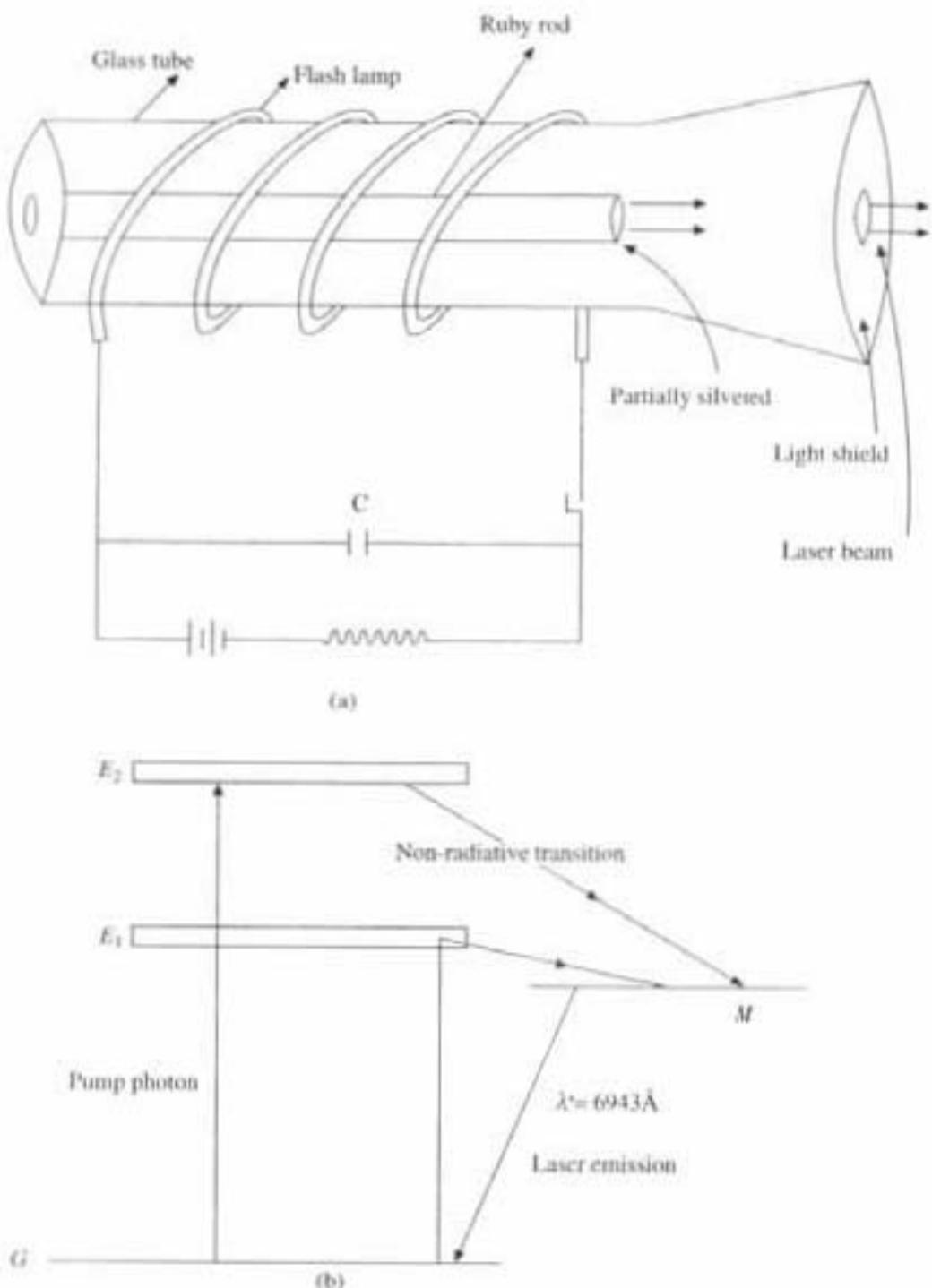


Fig. 5.4 (a) Experimental set up for ruby laser; (b) Energy level diagram shows different intermediate steps for laser action in ruby laser

The energy states of the chromium ions are shown in Fig. 5.4(b). The chief characteristic of the energy levels of a chromium ion is the fact that the bands levelled as E_1 and E_2 are the ordinary energy bands having a lifetime $\leq 10^{-8}$ sec, whereas the state marked as M has a long lifetime which is $= 3 \times 10^{-3}$ sec. A state which is characterised by such a long life is termed as a *metastable state*. From metastable state atoms cannot automatically loose energy by emission of radiation. Transition from metastable state to ground state is forbidden by selection rules of quantum mechanics, which govern such transitions. Thus, spontaneous emission

from metastable state (M) is not possible. Chromium ion in its ground state can absorb photon, from the xenon lamp, of energy ($E_1 - G$) and make a transition to the state E_1 , it could also absorb a photon of energy ($E_2 - G$) and make a transition to E_2 state. In either case it immediately makes a non-radiative transition (giving energy to the crystal) to the metastable state M and since the state M has a very long lifetime, the number of atoms in this state keep on increasing and one may achieve a population inversion between the state M and G . Once population inversion is achieved, light amplification, that is laser action, can take place by any spontaneously emitted photon.

Chromium ions decay from metastable state (M) to ground state (G) emitting laser light of wave length 6943 Å (red). This light can be made, to pass again and again, through a ruby rod by using mirrors at its ends and gets amplified. Ruby laser is also called *solid state laser*, due to solid ruby rod.

In ruby laser optical pumping is not continuous. It is in the form of pulses, because the Xenon lamp does not operate continuously. Correspondingly the laser output is also in the form of strong pulses of very short duration. Due to that it is also called as pulsed laser. Normally pulse duration is $\approx 10^{-6}$ S and the energy per pulse is $\approx 10^{-2}$ J, the average power in this solid state laser in each pulse is ≈ 10 KW. Ruby laser is a three level laser.

5.6 Helium-Neon Laser

The He-Ne laser consists of a mixture of He and Ne of about 10:1, placed inside a long narrow discharge tube (Fig. 5.5a). The pressure inside the tube is about 1 mm of mercury. The gas system is enclosed between a set plane mirror or a set of concave mirrors, so that a resonator is formed. One of the mirrors is of very high reflectivity while the other is partially transparent, so that energy may be coming out of the system, here Ne atoms are the active centres.

When the discharge passes through the gas mixture, then accelerated electrons collide with the helium atoms and excite them to different higher energy state (Fig. 5.5b) which is known as electrical pumping. Some of the helium atoms are excited to the levels F_2 and F_3 , which happen to be *metastable state*, so the helium atoms excited to those levels spend a long time before getting de-excited. Thus gradually a large number of helium atoms will accumulate in the states F_2 and F_3 . Now some of the excited states of neon E_6 and E_4 (Fig. 5.5b), correspond approximately to the same energy as the excited energy level, F_2 and F_3 of helium. So when excited helium atoms from metastable states F_2 , F_3 collide with unexcited neon atoms, an energy exchange takes place. The neon atoms are thus raised directly to the levels E_4 and E_6 and helium atom will be de-excited to the ground state. Because of the metastable nature of F_2 and F_3 level of helium atom, discharge through the gas mixture continuously populates the neon atom, excited to the energy levels E_4 and E_6 directly. This helps to create population inversion between levels E_6 , E_4 and the lower lying levels E_5 and E_3 in Ne. Hence once population inversion has been achieved, any spontaneously emitted photon can trigger laser action. The various transitions (Fig. 5.5b) lead to emission at wave lengths 3.39 μm , 1.15 μm and 6328 Å, the first two are in the infrared region, whereas 6328 Å is the well known red light from He-Ne laser. This

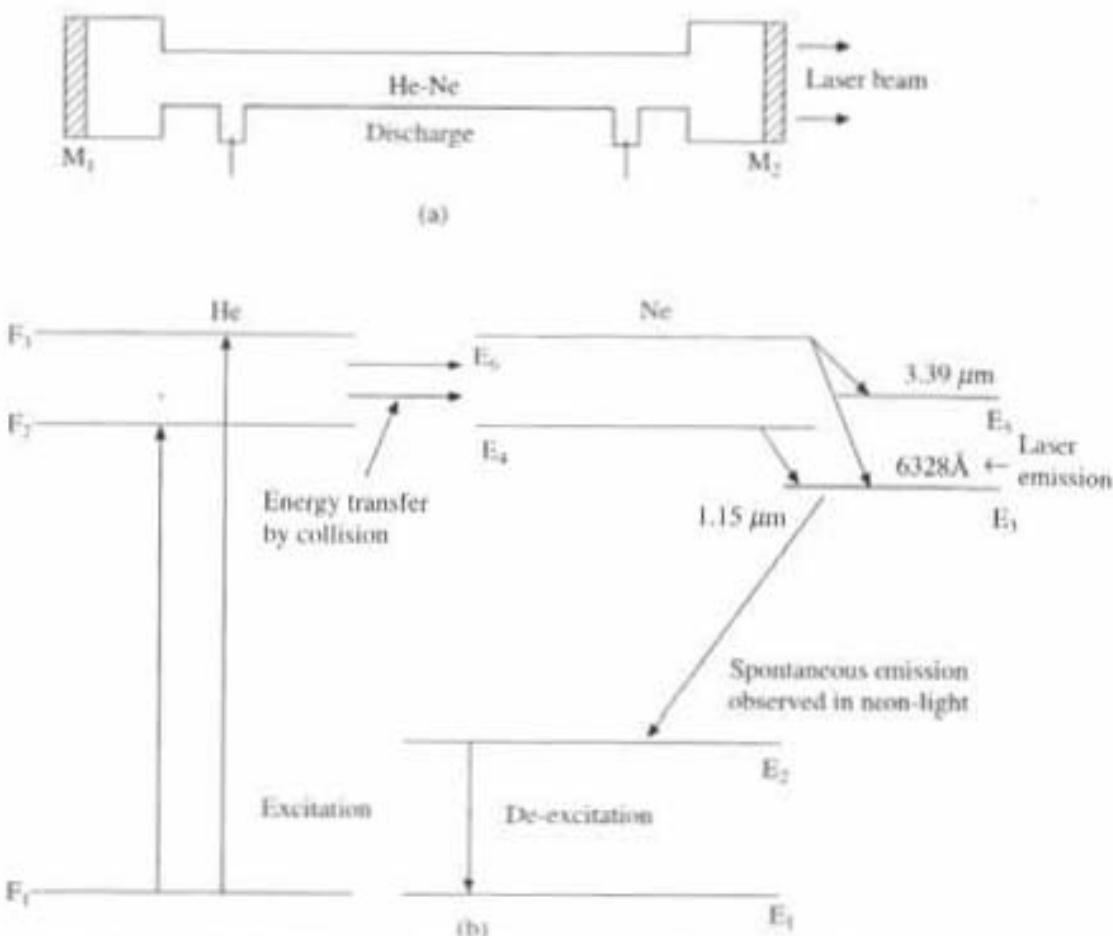


Fig. 5.5 (a) Experimental set up for He-Ne laser; (b) Energy level diagram shows different steps for Laser action in He-Ne laser

radiation is amplified by the process of stimulated emission, making use of the reflecting mirrors (Fig. 5.5a). Here pumping is called *Electrical pumping* since electrical energy is used for pumping.

He-Ne laser normally requires 5 to 10 W of excitation power and produces 0.5 to 50 mW of laser output. It is also called gas-laser, since it is originating from a He-Ne gas mixture.

Gas-lasers are, in general, found to emit light which is more directional and more monochromatic, this is because of the absence of the effects such as crystalline imperfection, thermal distortion and scattering which are present in solid state lasers. Gas-lasers are capable of operating continuously without the need for cooling. So this is also called as continuous output laser or CW laser. He-Ne laser is a four level laser.

5.7 Molecular Energy Levels

Energy level diagram as we have discussed for Ruby or He-Ne laser are due to *electronic motions*. But when a molecule say CO_2 , is formed, then in addition to the *electronic motions*, atoms in the molecules may vibrate in *different modes of vibration or vibrate about various axes*. Thus, the molecular energy levels are thus classified not only by electronic levels but also by vibrational and rotational levels. Each *electronic level* is split up into various *vibrational sub levels* due to the one-dimensional motion of individual atoms about a center of mass, in

molecules in the crystal, as shown in Fig. 5.6. Each *vibrational level* is again further subdivided into *rotational sub levels* estimated from the moment of inertia of the molecule about various axes. Suppose the molecule has linear dimensions of order a . Then the energy E_e associated with the electronic motion of a valence electron is of the order of

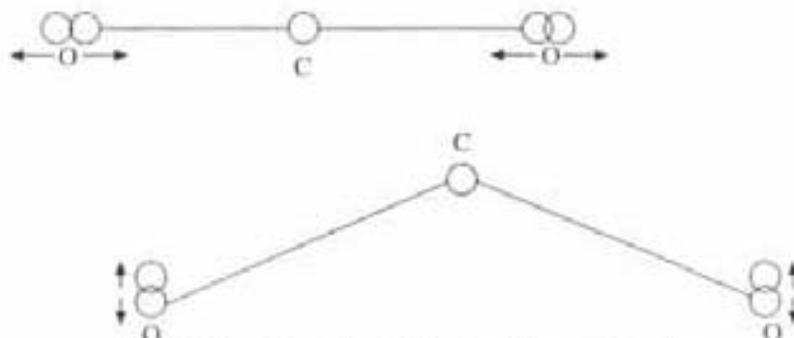


Fig. 5.6 Vibration of atoms in a molecule

$$E_e = \hbar^2/ma^2$$

where m is the electric mass. When a is of order of few Å unit, then this corresponds to transition frequencies (between various electronic levels) in the visible and ultraviolet region. Energy E_v associated with vibrational motion in a molecule is of the order of

$$E_v = (m/M)^{1/2} E_e$$

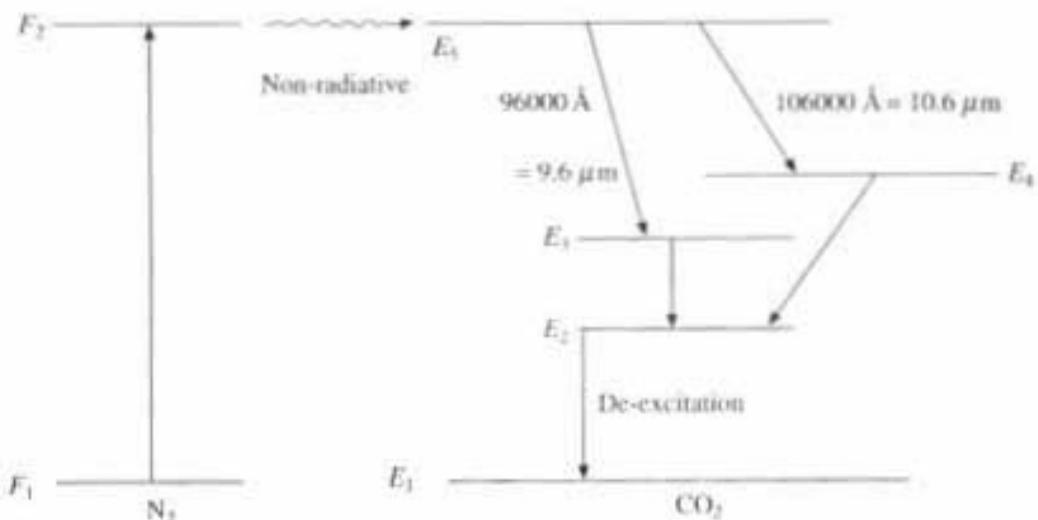
where M is the molecular mass, since molecular mass M is 100 times bigger than electron's mass m , so E_v is hundred times smaller than E_e and corresponds to transitions (between various vibrational levels) in the near infrared region. Rotational energy E_r associated with rotational motion is of the order of Ma^2 so that in a molecule

$$E_r = (m/M) E_e$$

which is about 10^4 times smaller than E_e and 100 times smaller than E_v , which corresponds to transition (between various rotational levels) in the far infrared region. In CO_2 laser and Dye Laser, laser action takes place between various vibrational and rotational levels which we will discuss now in detail.

5.8 Carbon Dioxide Laser

In carbon dioxide laser, laser action takes place between rotational sub level of a particular vibrational level to the rotational sub level of lower vibrational level. Then population inversion is achieved through a gas discharge. The efficiency of CO_2 laser is generally increased due to the addition of N_2 gas to CO_2 gas, when discharge is passed through the gas mixture, nitrogen molecules are excited to the upper vibrational level F_2 . Figure 5.7 shows some energy levels of N_2 and CO_2 gas. The upper vibrational level E_5 of CO_2 gas is on the same level of F_2 of N_2 gas, thus due to the collision of unexcited CO_2 molecule, with excited N_2 molecule, which are at the metastable F_2 level, energy transfer takes place and CO_2 molecules, raised directly to the upper vibrational level E_5 . This creates

Fig. 5.7 CO_2 laser

population inversion between the level of E_5 and lower energy levels E_4 and E_3 in CO_2 molecule. Now laser action takes place between rotational sub level E_5 of upper vibrational level of CO_2 molecule to rotational sub level E_4 and E_3 of lower vibration level and leads to emission at wave lengths 10.6 and 9.6 μm . which are in the infrared region.

The CO_2 gas-laser is much more efficient than other gas-lasers because the levels taking part in the laser transitions are the vibrational rotational levels of the lowest electronic level. These levels are very close to the ground level so de-excitation, from the lowest laser level to the ground level, involves very small amount of energy loss which is not contributing to the output laser beam. Hence large portion of the input energy is converted into output laser, resulting in a very high efficiency.

From several watts to several hundred watts power output is possible from CO_2 laser. High power CO_2 laser can be used in industry for drilling, welding, cutting, etc. CO_2 laser can also be useful in open air communication systems because the wave length from CO_2 laser fills in a band where atmospheric attenuation is small, it can also be used in an optical radar system.

5.9 Dye Laser⁴

Dyes used in the laser are organic substances which absorb in the near ultraviolet, visible or near infrared regions of the electromagnetic spectrum. These substances are most commonly solid which are dissolved in various organic or inorganic solvents like water, ethyl alcohol, etc. These dye lasers are now of great interest because of their broad absorption and fluorescent spectrum and possibilities of getting a cheap and simple tunable laser. One can also choose any region of the visible spectrum from the large range of dyes.

Various vibrational sub levels of different electronic states of the dye molecule are responsible for the laser action in dye laser. Fig. 5.8 shows the different electronic singlet levels S_0 , S_1 , S_2 , each of which are subdivided into vibrational levels and each vibrational level is again subdivided into various rotational level. This leads to each electronic level containing a large number of vibrational, rotational sub levels, and this dense collection of sub levels, leads to a near

continuum of states. This dense collection of sub levels is responsible for broad absorption and emission spectrum of dye molecule

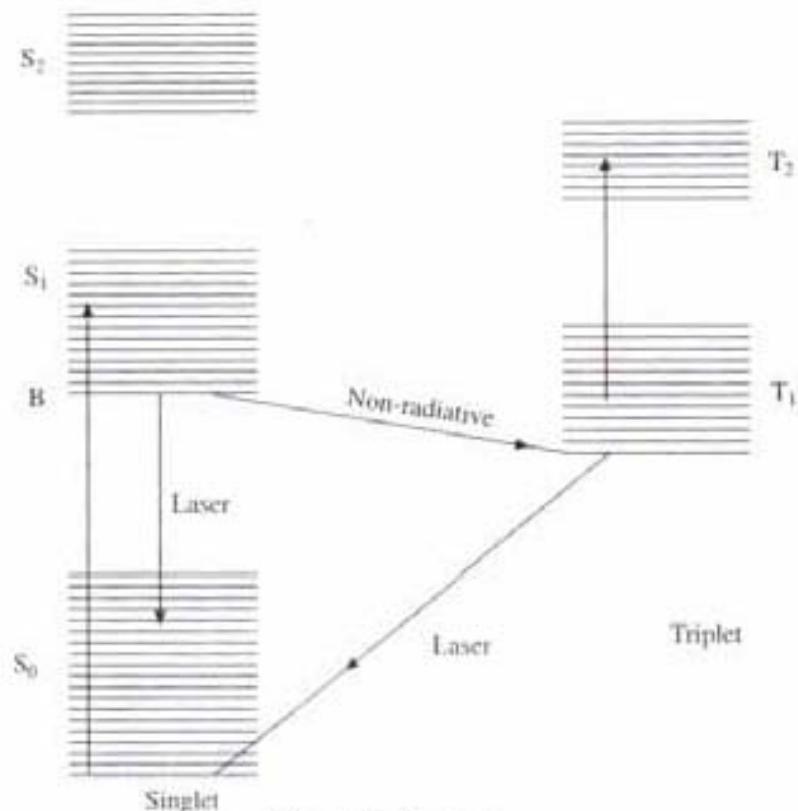


Fig. 5.8 Dye laser

Due to the absorption of light, dye molecules get excited from ground state S_0 , to various vibrational rotational level of the higher electronic level S_1 . Now, by the thermal redistribution in level S_1 , most of the dye molecules drop to the lowest vibrational level B of S_1 . Fluorescence comes as Laser radiation when the dye molecules make a transition from level B to any higher vibrational sublevel of S_0 . The lifetime of upper laser level B is of the order of 10^{-9} sec, which is much smaller than 10^{-3} sec, the lifetime of upper metastable laser level for solid state laser. Vibrational rotational sub levels for different electronic singlet levels as S_0 , S_1 , and S_2 and triplet levels as T_1 , T_2 etc. for dye molecule are shown in Fig. 5.8.

Dye molecules also make non radiative transition from B to triplet state T_1 . This transition is very important to maintain laser action in dye laser. Because the absorption spectrum of $T_1 \rightarrow T_2$ usually overlaps the emission spectrum of $S_1 \rightarrow S_0$, thus leading to a loss at the wave length corresponding to the laser emission, due to transition from level B to any level of S_0 . These losses can even prevent the larger oscillation. For good laser action, one must reach the threshold level before a significant number of molecules have dropped to level T_1 from S_1 . T_1 acts as an upper laser level and the laser transition takes place from T_1 to lower lying sub level of S_0 . Due to this, an exciting source with very fast rise time, either flash lamps or pulsed laser, is necessary.

Dye laser tuning over 500 Å is possible by replacing one of the resonance cavity with a diffraction grating (of reflection type), which is adjusted so that the

first order reflected beam for the desired wave length will be along the axis of the resonance cavity. By rotating the grating, wave length can be altered, as by using Bragg's law. Emission bands of dye laser are usually consisting of very large frequency range ($= 10^{-13}, 10^{-14}$ Hz or $100 \text{ \AA}, 1000 \text{ \AA}$). By using mode locking technique one can obtain mode locked, extremely short (i.e. 10^{-12} sec) optical pulse with the additional freedom of wave length tuning.

5.10 Semiconductor Laser or Diode laser and Light Emitting Diode (LED)

Across the junction of the semiconductor diode electrons and holes are injected under forward bias conditions as minority carrier (Fig. 5.9, 5.10). The excess minority carrier diffuse and recombine with majority carrier. In the absence of an external applied voltage no current flows, as potential barrier ($E_B = eV_B$), prevents the net flow of carrier and Fermi level for p and n side will coincide (Fig. 5.11a).

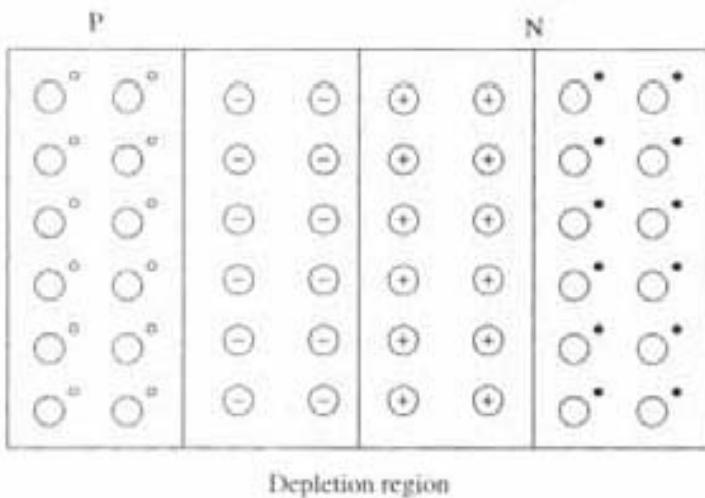


Fig. 5.9 Formation of depletion region in P-N junction

Spontaneous Emission

When external voltage is applied as forward bias (Fig. 5.10), electrons start flowing from n to p and holes from p to n . These minority carriers are effectively injected across the junction by the application of the external voltage and current flow through the device as they continuously diffuse away from the interface.

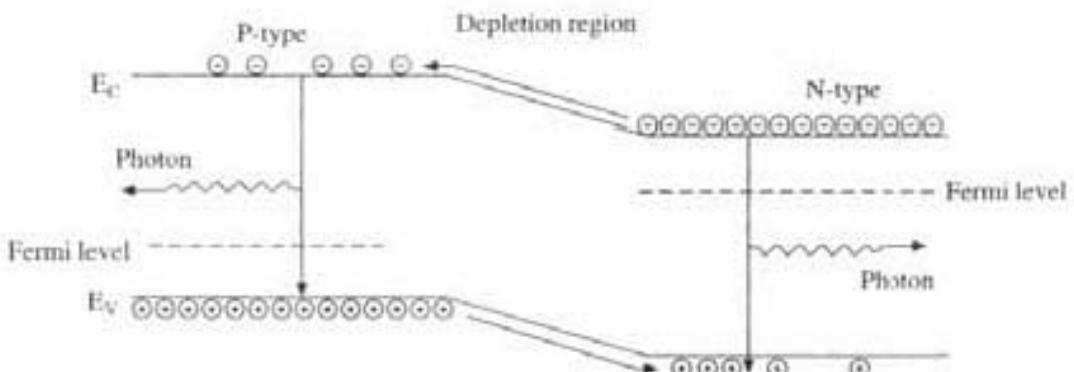


Fig. 5.10 Recombination of carriers under forward bias in P-N junction

This situation, in suitable semiconductor material, allows the *carrier recombination with the emission of light, in the depletion region*, which leads to LED (Light Emitted Diode) and semi-conductor laser. Since here forward biased P-N junction i.e., semi conductor diode gives rise the Laser Action, so it is called as semiconductor laser or Diode Laser.

The increased concentration of minority carriers in the opposite type region, under forward biased *P-N* junction leads to the recombination of carriers with the emission of photons, across the band gap in the depletion region is as shown in Fig. 5.10. Where measure of recombination can be defined as

$$\eta_i = \text{Internal quantum efficiency} = \frac{\text{No. of photons emitted}}{\text{No. of electrons injected}} \leq 1 \quad (5.1a)$$

which is unit less.

Energy released by this electron-hole recombination in deplection region is approximately equal to the band gap energy E_g . Excess carrier population is, therefore, decreased by recombination, which may be radiative or non-radiative. In non-radiative recombination, energy is released as lattice vibration and thus heat. Whereas in band to band radiative recombination the energy is released with the creation of photon which is giving *spontaneous emission* of photons. This is the mechanism by which light is emitted from an LED. So due to spontaneous emission of photon under forward bias condition *P-N* junction act as LED. But for the injection laser or semiconductor laser, 'Stimulated Emission' is required in addition to an optical cavity and mirror effects, to provide feedback of photons.

In LED, generally the energy released is equal to the band gap energy E_g . So the frequency emitted is

$$h\nu = E_g = \frac{hc}{\lambda} \quad (5.1b)$$

For GaAs, $E_g = 1.43$ eV. Hence, $\lambda = 0.86 \mu\text{m} = 8600 \text{\AA}$.

This 8600\AA is in infrared region, so that LED is called IRED. This spontaneous emission of light from within the diode structure is known as electroluminescence (i.e., optical emission due to electrical field). This amount of radiative recombination and emission mainly stay within the structure, depending upon the semi-conductor material.

Stimulated Emission and Lasing in Semiconductor Laser

In semiconductor laser both sides of *P-N* junction are heavily doped as p^+ and n^+ . Due to that, population inversion is already achieved. Heavy p doping causes the Fermi level to enter into the V.B. in *P*-side, whereas heavy n doping causes the Fermi level to enter into the C.B. in *N*-side. These are called *quasi-Fermi-level*. Figure 5.11(a) shows the open circuit condition, when Fermi level for p^+ and n^+ side will be on the same line. Thus a P^+-N^+ junction provides the active medium. To get laser action, we need population inversion and optical feed back, where population inversion has already been achieved due to heavy doping. To obtain stimulated emission, there must be a region of the device, where there are many excited electrons and holes present together. That is possible only in the *depletion region*.

Mechanism of Stimulated Emission in Diode-Laser/Semiconductor Laser

Figure 5.11(a) shows the P^+-N^+ junction energy levels under open circuit. When a forward bias nearly equal to band gap voltage is applied (Fig. 5.11b), then there is direct conduction possible. Now when photons with energy E_g which is less than the separation energy of the quasi-Fermi level E_q ($E_g < E_q$) (where $E_q = E_{F_{n^+}} - E_{F_{p^+}}$) is incident on the P^+-N^+ junction, then those photons cannot be absorbed, because the necessary C.B. is already occupied, so those photons can induce a downward transition of an electron from C.B. to V.B. and thus stimulating the emission of another photon and this can be continued. Thus the basic condition for *stimulating emission* in semiconductor laser is dependent of the quasi-Fermi level separation energy (E_q) as well as band gap (E_g) energy, and it is expressed as

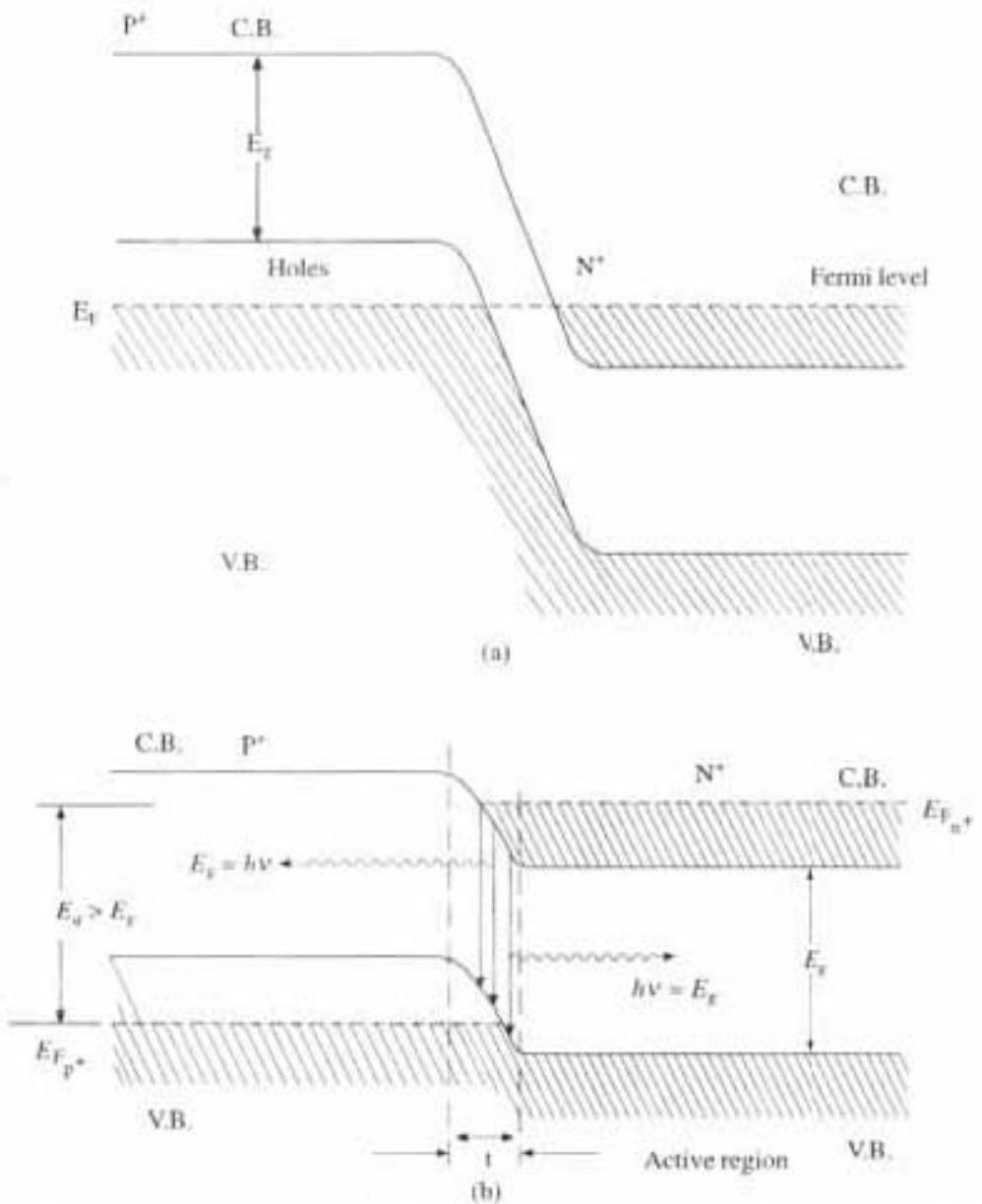


Fig. 5.11 Heavily doped P^+-N^+ junction under open circuit and forward bias condition, where carrier recombination under forward bias gives rise to stimulated radiation transition and acts as diode-laser

$$E_g = E_{F_{n^+}} - E_{F_{p^+}} > h\nu > E_g$$

The region near the depletion layer, that contains large number of electrons and holes simultaneously is called "active region" (Figs. 5.14, 5.15b), of thickness t . Thus any radiation of energy (frequency ν) which is more than E_g (i.e. $E_g = h\nu_g$) confined to the active region will be amplified. Thus heavy doping (or called degenerative doping) of P^+-N^+ junction provide stimulated emission and act as "Semiconductor laser". Therefore when extremely high current (threshold current J_{th}) is passed through the P^+-N^+ junction under high-forward bias voltage, spontaneous emission becomes coherent stimulated emission. Thus above threshold current, P^+-N^+ junction starts *laser action* and becomes *semiconductor or diode laser* (Figs. 5.12 and 5.13c). Whereas ordinary doped $P-N$ junction gives only

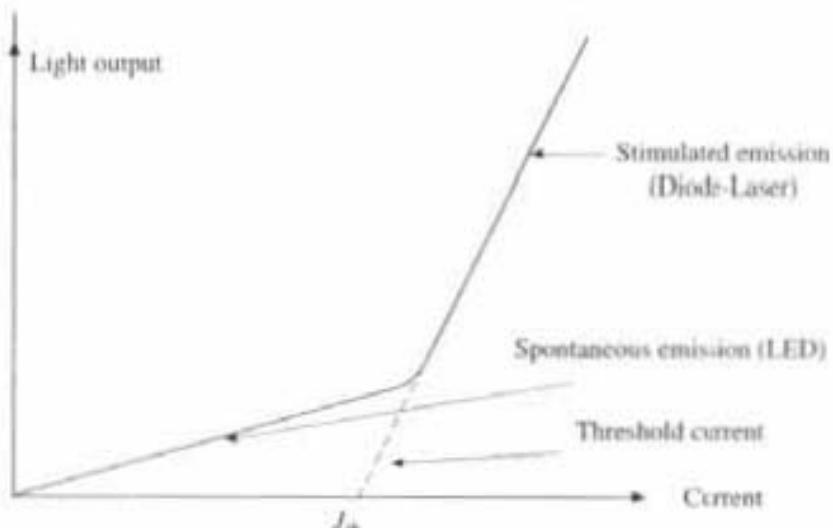


Fig. 5.12

spontaneous emission and acts as LED (Figs. 5.13a). The radiation generated within the active region of thickness t spreads out into the surrounding lossy semiconductor (say GaAs). There is infact some confinement of the radiation within a region of thickness d (Fig. 5.14), called mode volume or recombination region. Since the additional carriers are present due to the diffusion of electrons and holes as minority carrier in the active region, i.e., depletion region, so the

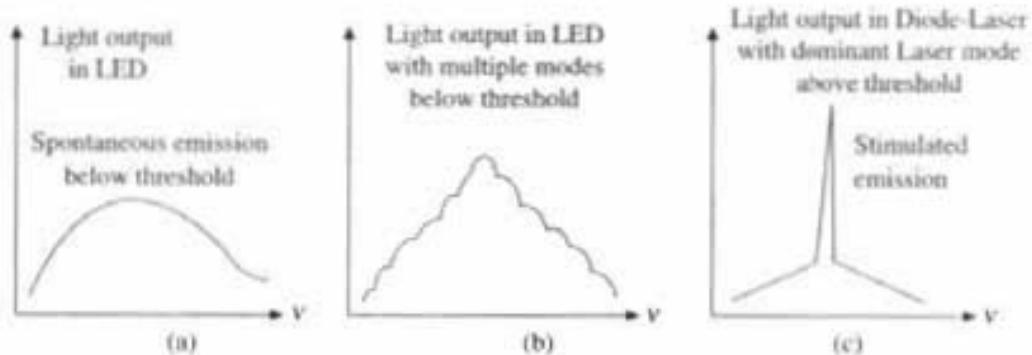


Fig. 5.13 Variation of light output with frequency for P^+-N^+ junction show sudden increase in light output after threshold current, when it acts as diode-laser, and before threshold current it acts as LED

refractive index of the active region will be more than the surrounding, thereby forming a *dielectric wave guide*, i.e. confinement of radiation within active region. In general the difference in R.I. between centre and neighbouring region is very small (0.1 to 1%) (Fig. 5.14), generally 0.02%, so the wave guiding is very inefficient and the radiation extends some way beyond the active region, thereby forming the mode volume (Fig. 5.14), and that way somehow they are almost lost. Due to this small difference in R.I. between center and neighbouring region 'homojunction laser' diode only works just well enough, to allow laser action as a result of vigorous pumping. Because of this reason Homo-junction Lasers can only be operated in the "pulsed mode", at room temperature, because for very high threshold pumping, current density required is of the order of 400 A/mm^2 . This difficulty can be overcome in "Hetero-junction" or "double hetero-junction" diode laser. Where GaAs is sandwiched between two layers of GaAlAs, which has wider energy gap and lower R.I. than GaAs. Threshold pumping current density in Hetero-junction level is considerably less, of the order of 10 A/mm^2 .

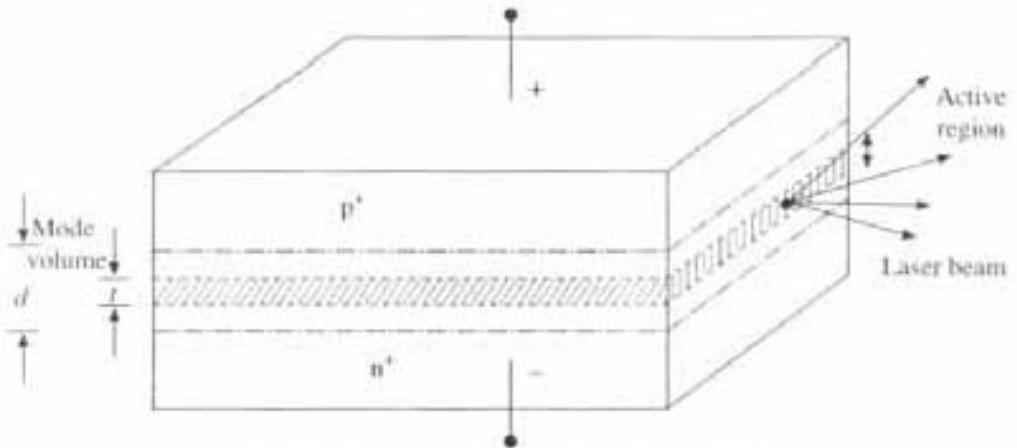


Fig. 5.14 Semiconductor laser construction

Above threshold current a small increase in current will give very high light (Fig. 5.12) output power, i.e. very high internal quantum efficiency (η_i), which is also defined as

$$\eta_i = \frac{\text{Rate of emission of photon}}{\text{Rate of supply of electron}} = \frac{R_p}{R_e} = \frac{R_p e}{I_p} \quad (\because I_p = eR_e) \quad (5.2a)$$

i.e.

$$R_p = \eta_i I_p / e$$

where I_p is the forward current.

Now power radiated by LED or diode laser can be written as

$$P_o = R_p h\nu = \frac{\eta_i I_p h\nu}{e} = \frac{\eta_i I_p hc}{\lambda e} \quad (5.2b)$$

Comparison Between Semiconductor Laser and LED

1. LED can operate at lower current densities, below threshold current, but it is essential for a semiconductor laser, it starts lasing action above threshold current only (Fig. 5.12).

2. From the LED, emitted photon have random phases because it is caused by

Spontaneous Emission and due to that the device is an incoherent optical source. But semiconductor laser is a coherent source of light, since it is due to stimulated emission of light.

3. The energy of the emitted photon is roughly equal to E_g of the semiconductor material, which gives wider spectral line width in LED (Fig. 5.13a). But in Diode-laser the line width is very narrow (Fig. 5.13c).

4. LED supports many optical modes within its structure and is therefore used as multimode source Fig. (5.13b). Semiconductor laser is used as Monomode source (Figs. 5.13c).

5. LED is having low value of internal quantum efficiency. The internal quantum efficiency for semiconductor laser is very high.

6. In semiconductor laser, power output increases somewhat linearly with current density (Fig. 5.12), similar to the spontaneous emission of LED, until a threshold is reached. After threshold, brightness increases suddenly to a very high order of magnitude, which is accompanied by a drastic narrowing of spectral width (Fig. 5.13c).

7. Light emission from LED due to spontaneous emission is very weak, it is confined only within the structure, whereas light output from semiconductor laser is very intense.

In semiconductor diode-laser since the laser emission is restricted to a very thin region ($\approx 1 \mu\text{m}$) around the junction (Fig. 5.14), in active region, so the laser beam usually emerges over a wide angle. Here in semiconductor laser, laser emission occurs between two bands of energies and not between two well defined energy levels. So the emission is not as monochromatic as in other laser. Dimension of semi-conductor laser is small and it is very efficient because there is a direct conversion of electrical current to light energy. Due to that there is a direct modulation scheme. By modulating diode current, output beam can also be modulated.

Due to these special properties of semiconductor laser, they have wide range of applications. They are having a great promise as source for light in optical fiber communication systems. The emission wave length of some semiconductor lasers like GaAs or GaInAsP, match with wave length region where optical fibers show very little attenuation. They are expected to find applications in optical radar equipment, space communications etc.

Above mentioned semiconductor laser operated continuously at very low temperature ($\approx 77^\circ\text{K}$). At room temperature there operation may produce laser in pulsed form. This problem was recently overcome by using hetero-junction lasers. Heat produced during laser action is also dissipated away efficiently by cementing the diode to a tin plate which acts as a heat sink.

HOLOGRAPHY

5.11 Introduction—Holography

An ordinary photograph represents a two-dimensional recording of a three-dimensional scene. It records only intensity distribution and hence the phase distribution which prevailed at the plane of the photograph is lost. Thus one

cannot change the perspective of the image in the photograph by viewing it from different angle. *Holography* is the method developed by Gabor in 1947, in which one not only records the amplitude but also the phase of the light waves, due to which holography has three-dimensional effect, and well known as *three dimensional photography*.

The basic technique of holography is as follows. When an observer is looking at some object before him, an object wave W_1 (scattered or reflected from object) reaches the observer's eye, and he sees the object. If the observer tilts his head, another object wave W_2 can reach his eye, so he now notices some other object, behind the first object. In this way the observer obtains a proper perspective of the scene around him. Now suppose the object mentioned above is removed from the scene, but somehow we manage to reconstruct the object wave W_1 , W_2 etc., so that these waves reach the observer's eye as before in the appropriate positions of the eye. Then the observer will now see quite realistic images of objects, as if they were there before him. Such a reconstruction of object waves to give three dimensional pictures or images of objects is called *holography*. Therefore, *holography is the method developed, by Gabor which gives three-dimensional picture of an object. Reproduction of three dimensional images is done in two stages. The first stage is to 'freeze' (or record) the different object waves by using the phenomenon of interference of light waves. The second stage is to 'unfreeze' (or reconstruct) the object waves by using diffraction of waves.*

5.11.1 Recording of Hologram

A laser beam is divided by a beam splitter into two beams A and B (Fig. 5.15). The transmitted beam B illuminates the object, whose *hologram* is to be recorded. The light scattered or reflected by the different parts of the object with different phases forms the *object wave*. This object wave falls on the photographic plate as shown. The reflected beam A which is called *reference wave* also falls on the

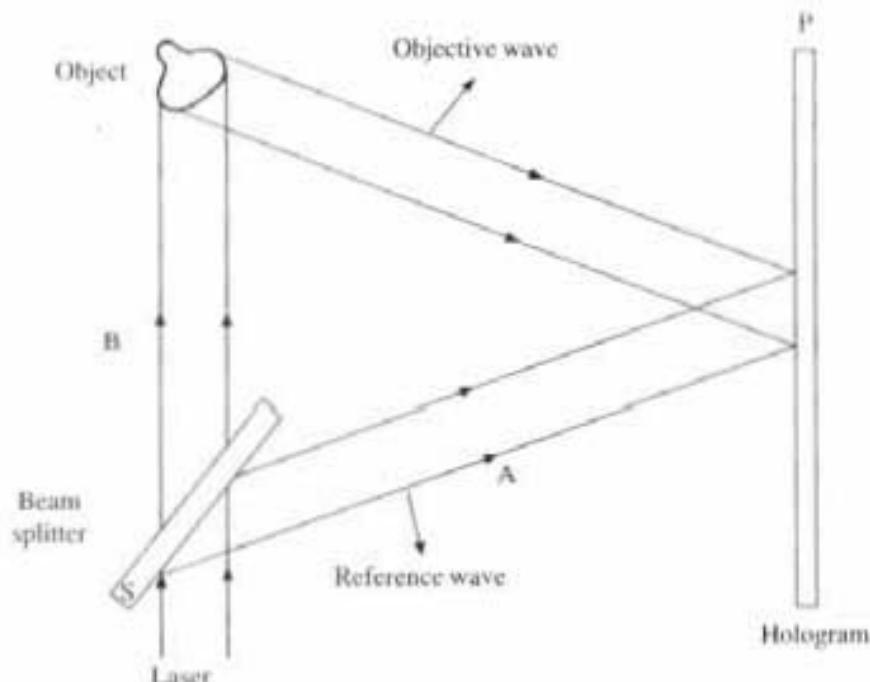


Fig. 5.15 Construction of a hologram

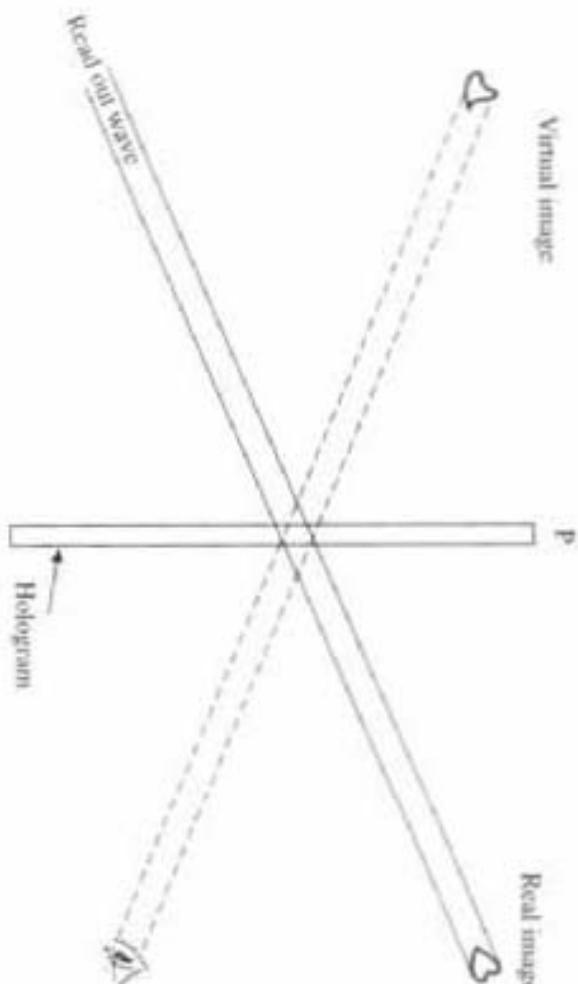


Fig. 5.16 Reconstruction of the image

photographic plate (P). The superposition of object wave and the reference wave produces an interference pattern, which is recorded on the photographic plate. The plate is developed and fixed. It is then called a *hologram*. It is just an interference pattern and bears no resemblance to the actual object. But since it is an interference pattern, so it stores all the information about the amplitudes and phases of the object waves from the object. The hologram thus records complete information about the surface of the object. In a hologram, interference pattern is recorded photographically and made into transparency and fixed like a zone plate.

5.11.2 Reconstruction of the Image

The hologram is now illuminated by a laser beam exactly the same way as the reference wave (A) used in its construction (compare Figs. 5.15 and 5.16.) This beam is now called as *read out wave*. The read out wave interacts with the hologram (i.e. interference pattern) and due to diffraction two images, one real, one virtual are produced on two sides of the hologram. Here the hologram acts just as a kind of grating. The virtual image appears at the original position occupied by the object in the construction of the hologram. The other image (real) is symmetrically situated but on other side of the hologram, it can be photographed by placing light sensitive medium at the position where the real image is formed.

The virtual image is seen by looking through the hologram as if it were a window. The image will appear in complete three dimensional form. It is possible to see the other side of the object as if the object is really before the observer. But since it is a virtual image, nobody can touch it and it is not possible to photograph it. The holograph is therefore said to contain, in the form of its interference pattern, all information about the geometrical characteristics of the object.

5.11.3 Uses of Holography

1. Holography is used in the microscopic study of some transient events. Instead of doing prolonged examination of some small specimen suspended in the medium, if one can take a short exposure hologram of the specimen, then the reconstructed holographic image can be studied for the same purpose at any time.

2. Holography is also useful to provide a high capacity system for image storage and re-examination.

FIBER OPTICS

5.12 Introduction—Fiber Optics

The term fiber optics was first introduced by N.S. Kapans as the "*art of the active and passive guidance of light in the ultraviolet, visible and infrared regions of the spectrum, along transparent fibers through pre-determined paths*". In most of the application of fiber optics or optoelectronics, the optical beams are confined laterally to a finite region in space. Special optical elements are used to confine and allow propagation of such optical modes. An important structure used in optical system is the layered structure or the *wave guide structure*. As the name implies, *these structures are used to confine the optical waves in a well defined region and guide their propagation*. The layered structures can be made from non-crystalline materials or from crystalline materials. For example glass is utilized to produce optical fibers used in optical communication while semiconductor wave guides are used in semiconductor laser.

The confinement of the optical wave in space is done by a variation in the dielectric constant of the guide in space. The variation could be in single dimension to produce planer wave guide (Fig. 5.17a). Where the dielectric constant varies in two dimension, then that produces (Fig. 5.17b) linear or rectangular wave guide. Wave guide produced from dielectrics and semiconductor grown epitaxially, are usually planer or rectangular wave guide form. When the dielectric constant varies cylindrically (Fig. 5.17c), as in optical fiber, then cylindrical wave guide is formed. Optical fibers came in three main categories depending upon the ingredients used such as of silica, glass and plastic.

Optical fiber is a cylindrical wave guide system through which the optical wave can propagate. The principle by which light wave travel through the fiber is "*Total Internal reflection*", without loss of incident intensity. The basic structure consists of a core at the centre and a cladding layer on the outside as shown in Fig. (5.18). The core is in a cylinder of transparent dielectric material with R.I. n_1 and the cladding is another dielectric material with lower R.I. n_2 . The fibers have several classifications based on the index profile and the core size as shown in Figs. (5.18a, b, c). The core size determines how many modes of the optical wave can propagate through the fiber. An optical cable consists of several of such optical fibers.

In some optical fibers, over and above cladding there is another dielectric material which is called Buffer, whose refractive index n_3 is lower than n_2 (i.e. $n_1 > n_2 > n_3$).

5.13 Total Internal Reflection

If a ray of light is incident at an angle i in denser medium (R.I. n_1), the refracted rays will bend away from the normal in rarer medium (R.I. n_2) and Snell's law is written as

$$\frac{\sin i}{\sin r} = \frac{n_2}{n_1} \quad (5.3)$$

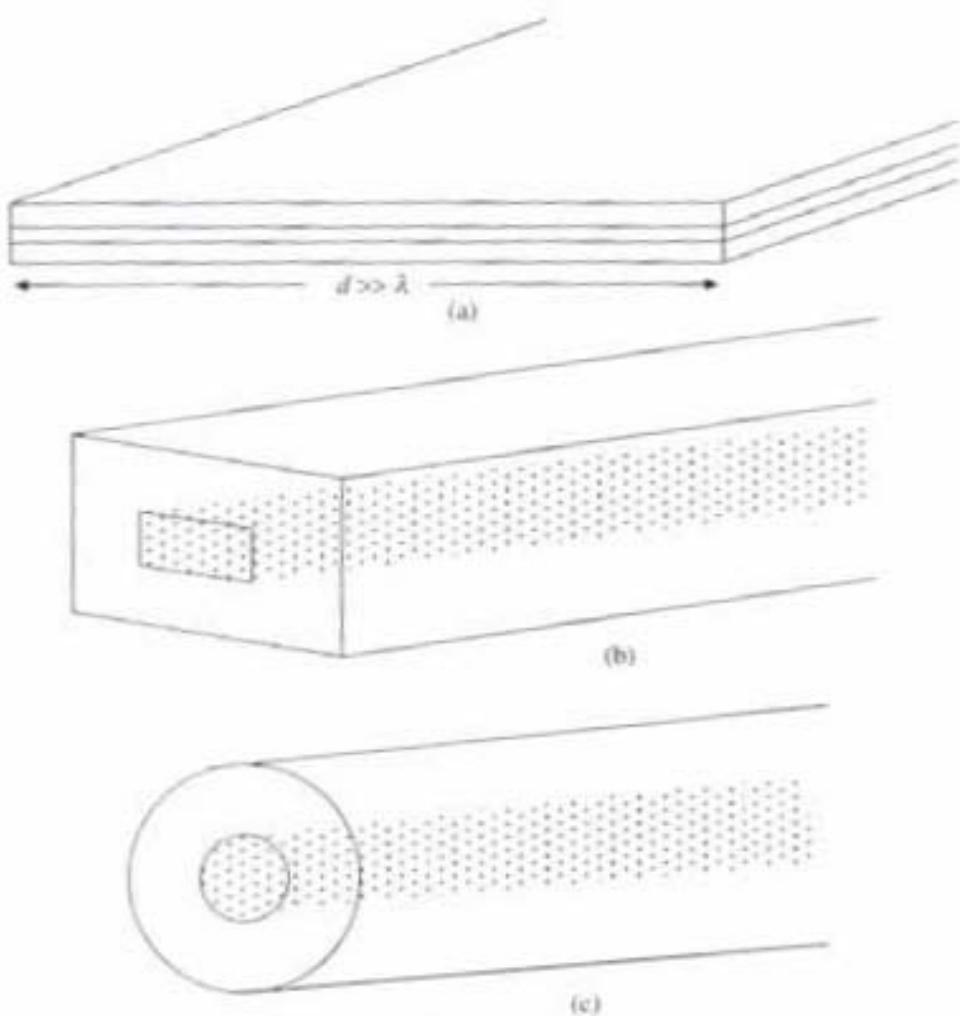


Fig. 5.17 Planer, rectangular and cylindrical wave guide

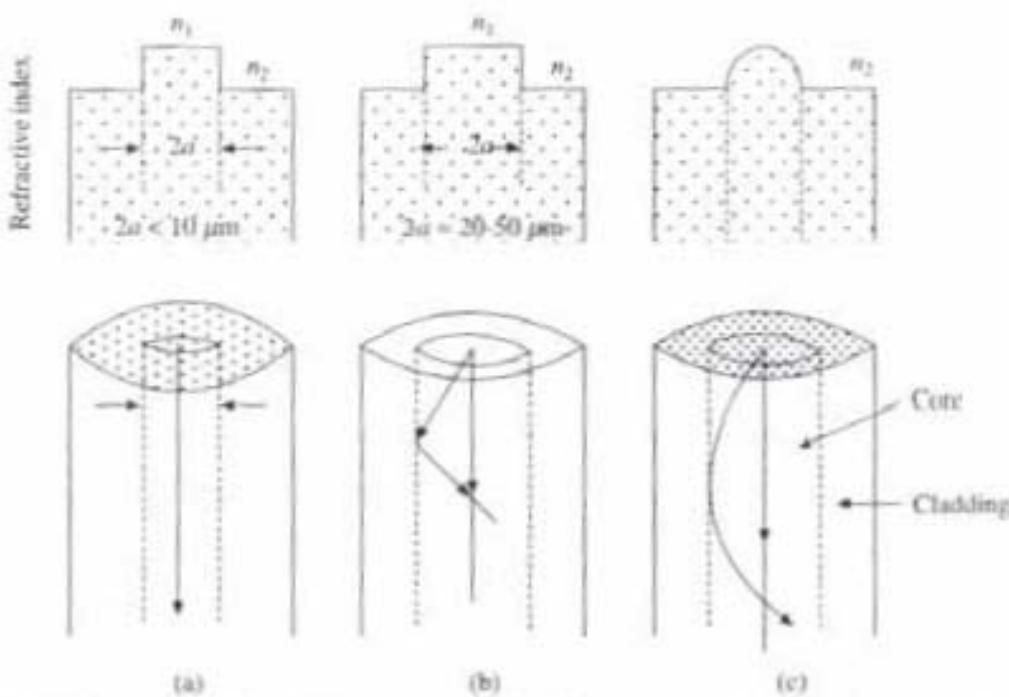


Fig. 5.18 Step index: (a) single mode, (b) multimode and (c) graded index multimode

If a ray of light is incident at an angle say i , when the angle of refraction $r = 90^\circ$ then

$$\sin i_c = n_2/n_1 \quad (\because \sin r = \sin 90^\circ = 1) \quad (5.4)$$

and angle i_c is known as *critical angle*. The value of the critical angle depends upon the refractive index of the denser medium. Now if the light in denser medium is incident at an angle $i > i_c$ then light will be reflected back into the denser medium. This phenomenon is called *total internal reflection*, which is used in optical fiber communication.

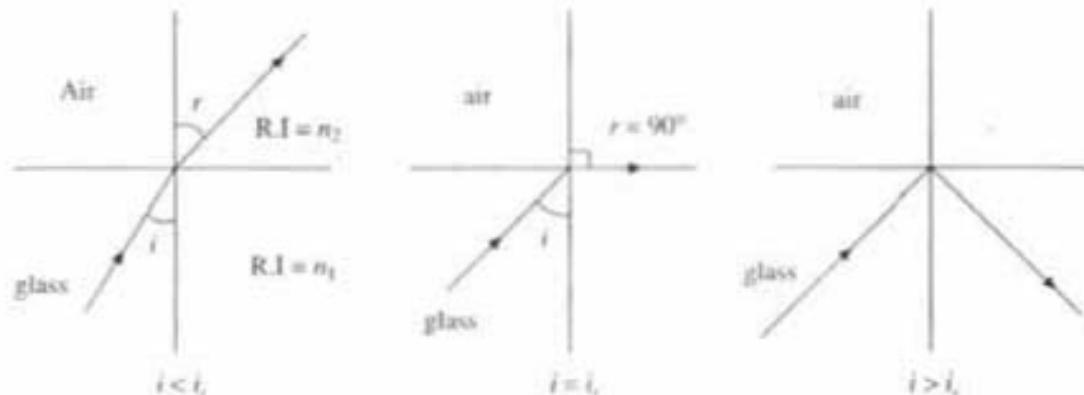
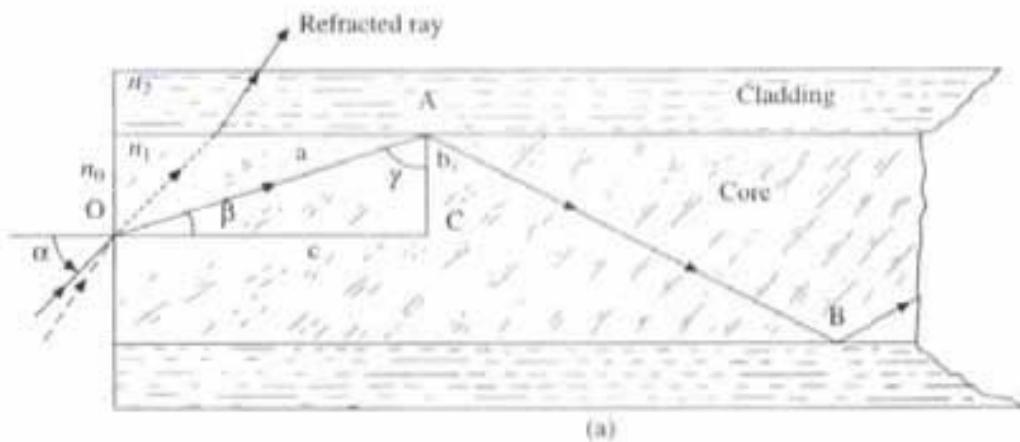


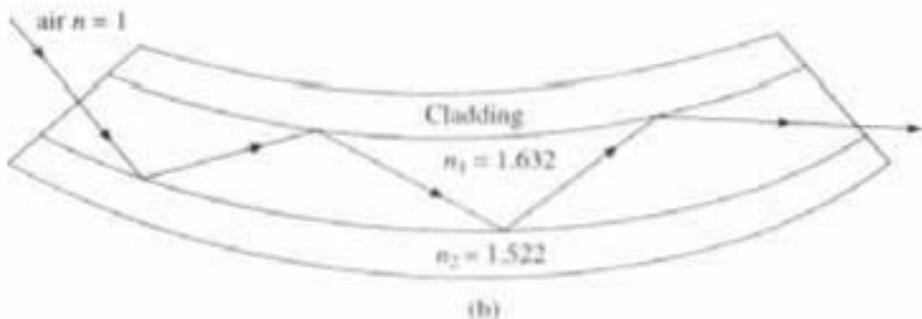
Fig. 5.19 Total internal reflection

5.14 Propogation of Light in Fiber

Consider an optical fiber with cladding (Fig. 5.20a), where n_0 is the refractive index of air, n_1 the index of core and n_2 the index of cladding ($n_1 > n_2 > n_0$). Suppose a beam of light is incident into the optical fiber at an angle α with



(a)



(b)

Fig. 5.20 Light propagation through straight and curve step index optical fiber with cladding

normal at O , and after refraction it will travel the path along OA , through core region by making an angle β with normal at O , at air-core boundary and incident on core-cladding boundary at A , at an angle γ with normal to core-cladding interface.

Now if this ray OA makes an angle $\gamma \geq i_c$ with normal at A , the glass-cladding interface then it will suffer total internal reflection and will travel along AB . In a similar way the beam will continuously suffer total internal reflection until it emerges out of the glass core at the other end. Even if the fiber curves gradually, light will travel its entire length (Fig. 5.20b) and arrive at the other end. Thus a fiber can transmit light energy quite efficiently. There may be about 1,000 reflecting points like A, B per meter. The loss per reflecting point is as low as 0.04% due to absorption etc.

For total internal reflection to takes place

$$\gamma \geq i_c, \sin \gamma = n_2/n_1 \quad (\text{from equation 5.4})$$

Hence $\sin i_c = \sin \gamma = n_2/n_1 = c/a$ (from Fig. 5.20a) (5.5)

i.e. critical angle

$$i_c = \gamma = \sin^{-1} \frac{n_2}{n_1} \quad (5.6)$$

From the $\triangle OAC$ in (Fig. 5.20a) $OA = a, OC = c, AC = b$

Therefore $a^2 = b^2 + c^2, 1 = b^2/a^2 + c^2/a^2$

$$b^2/a^2 = 1 - c^2/a^2 = 1 - n_2^2/n_1^2$$

Thus

$$b/a = \sqrt{1 - n_2^2/n_1^2} = \sin \beta \quad (5.7)$$

Now by applying Snell's law at point O , we get $n_0 \sin \alpha = n_1 \sin \beta$ (5.8)

i.e. $n_0 \sin \alpha = n_1 \sqrt{1 - n_2^2/n_1^2} = \sqrt{n_1^2 - n_2^2}$

Which is called *numerical aperture (N.A.) of the fiber*.

Hence

$$\text{N.A.} = n_0 \sin \alpha = \sqrt{n_1^2 - n_2^2} \quad (5.9)$$

For efficient transfer of light, there should be no refraction at A, B , etc. the glass-cladding interface, otherwise light energy will be lost through walls. From Fig. 5.20(a), it is clear that the angle of incidence γ at the glass-cladding interface will depend on α and β , and if $\gamma \geq i_c$, then only total internal reflection is possible, which helps the propagation of light through optical fiber. So condition for total internal reflection requires that, the angle of incidence for all the rays at the entrance should be less than the *acceptance angle* of the fiber. *Acceptance angle is the maximum angle of incidence (α) for which total internal reflection is possible and fiber can conduct light to the other end*. If the angle of incidence is more than α then that ray will go to cladding after refraction from the core, (Fig. 5.20). Which means, that information will be lost. The acceptance angle α is given by the N.A. expression (eqn. 5.9) as

$$\alpha = \sin^{-1} \left(\frac{\sqrt{n_1^2 - n_2^2}}{n_0} \right) \quad (5.10)$$

Thus, the quantity ' $n_0 \sin \alpha$ ' called the 'numerical aperture' (NA) of the fiber, is a measure of the ability of the fiber to accept light for transmission. Usually $n_1 = n_2$ and for air $n_0 = 1$.

$$\text{Thus, } N.A. = n_0 \sin \alpha = \sin \alpha = \sqrt{n_1^2 - n_2^2} = n_1 \sqrt{1 - n_2^2/n_1^2} = n_1 \sqrt{2\Delta} \quad (5.11)$$

where

$$\Delta = \frac{n_1^2 - n_2^2}{2n_1^2} = \frac{n_1 - n_2}{n_1}$$

is called relative refractive index difference when $n_1 = n_2$.

The dielectric cladding on glass core reduces scattering loss, protects core from absorbing external optical disturbances and provides mechanical strength to main core glass fiber. Sometimes there is buffer coating over cladding which adds further strength to the main fiber and protects fiber from mechanical vibrations and impact.

5.15 Different Types of Optical Fiber

5.15.1 Cladless Fiber

Some optical fibers used in practice which do not have cladding. It shows a ray of light travels through modern type of cladless fiber.

5.15.2 Graded Index Fiber (Grin Fiber)

Grin index fiber is a special technique in which the core of the fiber is made with varying refractive index which is a function of core radius. The refractive index is greatest at the centre of the core and decreases continuously towards periphery, but core R.I. remains always greater than cladding R.I. The rays travel along a sinusoidal path (Figs. 5.18c and 5.21a) through this type of graded index fiber.

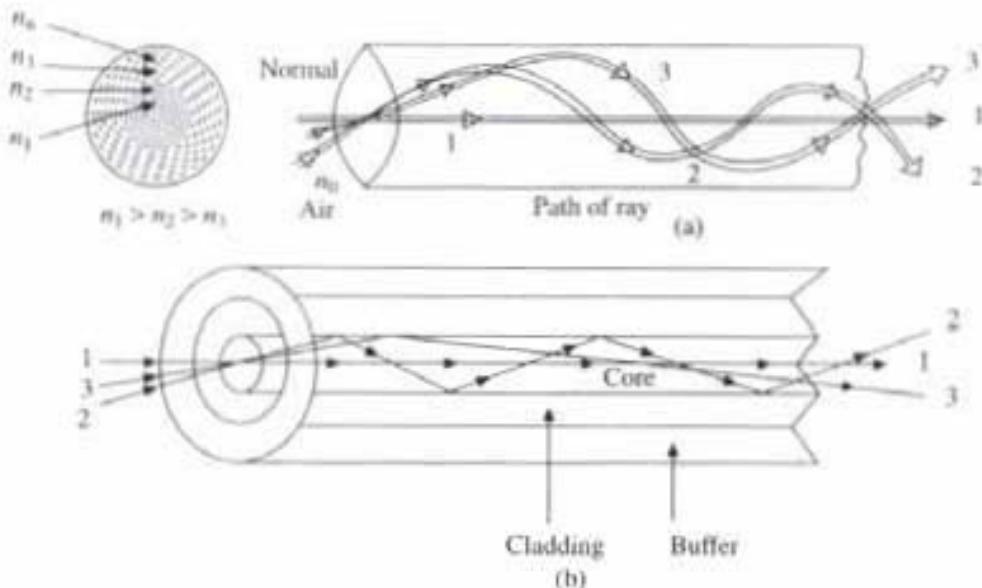


Fig. 5.21 Light propagation in multi-mode graded index and step index fiber

These fibers called as wave guides are very much used for telecommunications and other applications employing lasers.

5.15.3 Step Index Fiber

In this type of fiber (Figs. 5.21b and 5.18a, b) central core is made up of glass refractive index (n_1) between 1.5 and 1.6. This core is surrounded by cladding material like fused quartz R.I. ($n_2 = 1.425$) or plastic ($n_2 = 1.25$) or glass R.I. $n_2 < n_1$. Around the cladding there is a buffer coating.

5.15.4 Passive and Active Fiber

The fibers described above are called passive fibers. A *passive fiber* merely guides light incident on it from an external source. An *active fiber* emits light as well as guides part of it along its length. Scintillating fibers and lasing fibers are examples of active fiber.

5.16 Types of Rays (Meridional and Skew Rays)

1. *Meridional Ray*: The plane containing the central axis of the fiber is called meridian plane. Meridional rays remain in the meridian plane which can be divided as follows: (a) bound rays that propagate along axis (Figs. 5.22a, b) and (b) unbound rays that are not bound along axis and are refracted out of the core and absorbed in cladding or buffer.

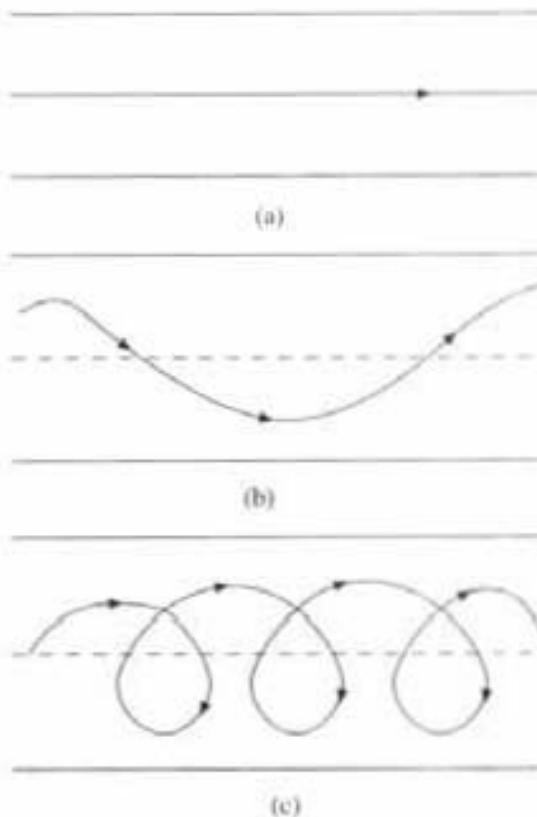


Fig. 5.22 (a), (b) Meridional and (c) skewed rays propagation through fibre optics cable light

2. Skew Rays: The rays which enter the fiber obliquely, which are not confined to a plane and follow a helical path around the axis are known as skew rays. They are soon scattered out of the core at bends and irregularities and thus do not contribute significantly in optical communication (Fig. 5.22c).

5.17 Modes of Propagation in Fiber

In Fig. (5.20) we have drawn only one ray, that is called a ray diagram, but in reality an infinite number of such rays (Figs. 5.21a, b) all slightly displaced from each other, also can be propagated at a time which are called *different modes of propagation*. They will follow different paths in the fiber at the time of propagation as shown in Fig. (5.23).

Now when light is described as an electromagnetic wave, it consists of a periodically varying electric field \vec{E} and magnetic field \vec{H} which are perpendicular to each other. The modes which are shown in Fig. (5.23) are called *transverse modes* which illustrate the case when the electric field (\vec{E}) is perpendicular to

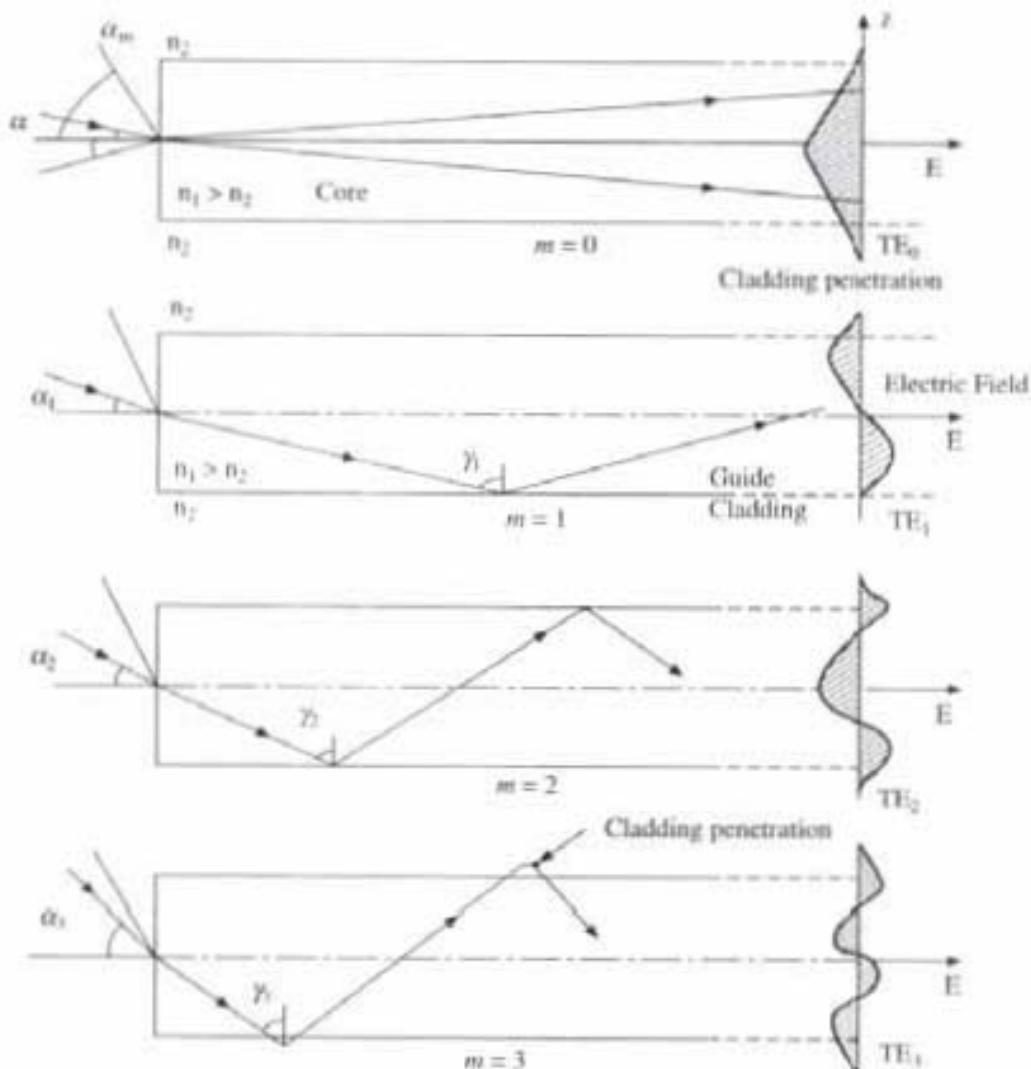


Fig. 5.23 Different modes of light propagation through step index fibre, where α_m = acceptance angle, γ_c = critical angle and $\pi/2 > \gamma_1 > \gamma_2 > \gamma_3 \dots > \gamma_c$ whereas $\alpha_0 < \alpha_1 < \alpha_2 < \alpha_3 \dots < \alpha_m$

the direction of propagation (Z) and hence $\bar{E}_Z = 0$, but a corresponding component of \bar{H} is in the direction of propagation. So these modes are called transverse electric (TE_m) modes. Similarly, when a component of electric field \bar{E} is in the direction of propagation but $\bar{H}_Z = 0$, the modes formed are called transverse magnetic (TM_m) modes. The mode number m are incorporated into this nomenclature by referring to TE_m or TM_m modes, etc. Each value of m is associated with a distinct wave pattern or mode as shown in Fig. (5.23), within the wave guide they follow different paths. For each value of m there is a value of $\gamma(\gamma_m)$. γ_m can take values in the range γ_c to $\pi/2$, which implies that the smaller the value of γ_m , i.e. γ_m closer to γ_c , higher the mode number and vice-versa. From Fig. 5.23 it is clear that the higher order modes, carry along with them a higher portion of mode energy within the cladding. This part of the wave which is carried in the cladding is referred as the "evanescent wave".

(1) For free space communication, the total electromagnetic field lies in the transverse plane, where both $\bar{E}_Z = 0$ and $\bar{H}_Z = 0$, which is called transverse electromagnetic wave (TEM). TEM wave occur in metallic conductor like coaxial cable but not in optical wave guide.

(2) For planar guide, generally $TE(E_Z = 0)$ and $TM(H_Z = 0)$ modes are obtained.

(3) The cylindrical wave guide as in optical fiber, is bounded in two dimensions, rather than one as in plane wave guide. Thus, two integers l and m are necessary in order to specify the modes, in contrast to the single integer (m) required for planar guide. So, for cylindrical wave guide we refer TE_{lm} and TM_{lm} modes. The TE_{lm} and TM_{lm} modes corresponds to meridional rays (Figs. 5.22a, b) which are travelling within the fiber. However, hybrid modes where \bar{E}_Z and \bar{H}_Z are non zero also occur within the cylindrical wave guide. These modes which results from skew rays propagation (Fig. 5.22c), within the fiber are designated as HE_{lm} and EH_{lm} depending upon whether the components of \bar{H} or \bar{E} make larger contribution to the transverse field. These skew rays are called leaky rays and corresponding modes are called leaky modes, not bound modes because they can propagate only appreciable distance before being lost from the fiber.

In most practical wave guides, the refractive indices of core and cladding differ from each other by only a few percent and in that case the full set of modes (i.e., EH , HE and TE) can be approximated by a single set called *linearly polarised (LP_{lm}) modes*. An LP_{lm} mode in general has m field maxima along a radius vector and $2l$ field maxima round a circumference as shown in Fig. 5.24.

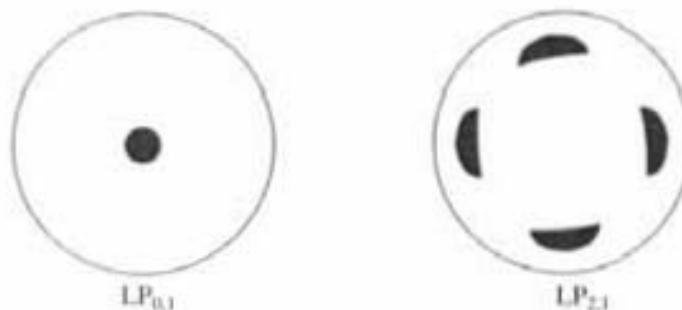


Fig. 5.24 Linearly polarised modes

5.18 Multimode and Monomode Fibers

Multimode step index fibers allow the propagation of a finite number of guided modes along the channel. The number of guided modes is dependent upon the physical parameters, i.e., the refractive index difference between core and cladding (i.e., $N.A. = \sqrt{n_1^2 - n_2^2}$), core radius of the fiber a and the wave length (λ) of the transmitted light, are all included in the 'normalized frequency V ' for the fiber where $V = \frac{2\pi}{\lambda} \cdot a \cdot (N.A.)$ = Normalized frequency. The total number of guided modes for step index fiber is related to V by the approximate expression

$$M_s = V^2/2 \quad (5.12)$$

where as for graded index fiber $M_s = V^2/4$

Monomode or Single Mode Fiber

It can be shown that when $V < 2.405$, only the LP_{01} mode can propagate, i.e. fiber becomes single mode fiber. Theoretically, LP_{01} mode will propagate no matter how small the V value, i.e., no matter how small the core radius a .

Since

$$V = \frac{2\pi}{\lambda} \cdot a \cdot (N.A.) \quad (5.13)$$

Hence

$$a = \frac{V\lambda}{2\pi(N.A.)}$$

When $V < 2.405$ then

$$a < \frac{2.405\lambda}{2\pi(N.A.)}, \quad \text{or} \quad a < \frac{2.405\lambda}{2\pi\sqrt{n_1^2 - n_2^2}} \quad (5.14)$$

for single mode fiber.

In general in single mode fibers cores tend to be only a few microns in diameter, but it has an advantage from several point of views if core can be made as large as possible. From the above expression, we see that this may be done by reducing the N.A. value, i.e., by making core and cladding R.I. very close to each other. One of the major advantages of using single mode fiber is that inter model dispersion no longer exists, and so they are capable of carrying very large bandwidth signal for long-distance communication.

5.19 Dispersion-Inter Modal Dispersion

(1) For multimode step index fiber, when the ray of different modes follow different paths, like some rays (AA' , Fig. 5.25a) which propagate along the fiber axis and some of which propagate along the trajectories of the oblique rays (say CC' or BB') as shown in Fig. 5.25(a). Then there is a difference in time taken for different modes to travel a distance l . Now the R.I. of a medium is simply a measure of velocity of propagation v of the light in the medium.

$$\text{Now } n = \text{R.I.} = \frac{v_a}{v_m} \quad \text{or} \quad v_m = \frac{v_a}{n} = \frac{c}{n}$$

where $v_a (= c)$ is the velocity of light in air and v_m the velocity of light in medium.

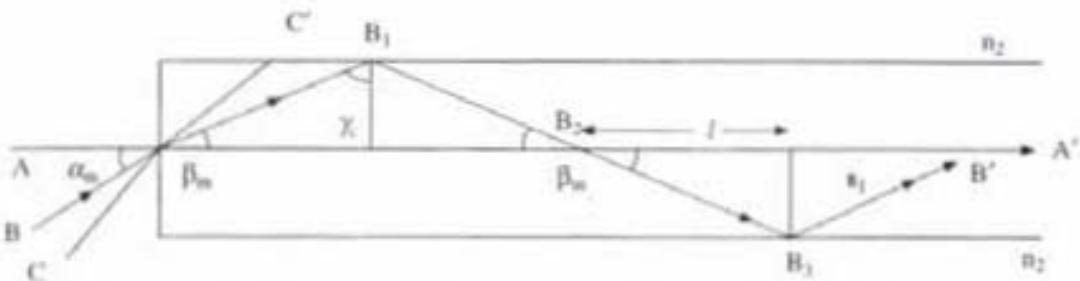


Fig. 5.25(a) Multipath time dispersion for multimode step index optical fiber

Now the axial ray AA' travel an axial distance l along the fiber in time

$$t_1 = n_1 l/c \quad (5.15)$$

whereas the oblique ray BB' travels the same axial distance l along the path B_2B_3 in the time

$$t_2 = \frac{n_1}{c} \frac{l}{\cos \beta_m} = \frac{n_1}{c} \frac{l}{\cos (90 - \gamma_c)} = \frac{n_1}{c} \frac{l}{\sin \gamma_c} = \frac{n_1^2 l}{n_2 c} \left(\because \sin \gamma_c = \frac{n_2}{n_1} \right) \quad (5.16)$$

Thus if two rays (AA' and BB') are launched together they will be separated on arrival of the other end by *dispersion time* ΔT , where

$$\Delta T = t_2 - t_1 = \frac{n_1 l}{c} \left(\frac{n_1}{n_2} - 1 \right)$$

or

$$\Delta T = \frac{n_1}{n_2} \frac{l}{c} \Delta n \quad (\text{where } \Delta n = n_1 - n_2) \quad (5.17)$$

Thus a pulse containing rays at all possible angles spreads out during propagation by an amount given by

$$\frac{\Delta T}{l} = \frac{n_1}{n_2} \frac{\Delta n}{c} \quad (5.18)$$

This is known as *multipath time dispersion* of the step index fiber, which causes pulse broadening.

(2) For graded index multimode fiber, different modes will follow different paths, sinusoidally. The rays making larger angles (Fig. 5.25b) traverse longer path but they do so in lower R.I. area and hence they travel with higher speed of propagation. Since R.I. changes gradually in graded index fiber, so there is a self focusing mechanism that leads to a smaller value of pulse dispersion than that of

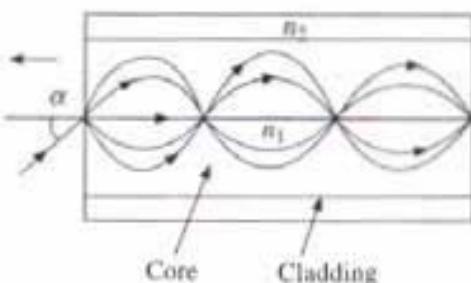


Fig. 5.25(b) Less multipath time dispersion for graded index fibre

step index fiber. Thus, for long distance communication graded index fiber will be more useful than step index fiber, which causes less time dispersion.

5.20 Losses in Fiber Cable

The losses in fiber, i.e., attenuation, are mainly caused by absorption and scattering.

(1) Absorption is basically a material property, at all wave lengths it occurs, when electronic transitions take place within the material and are followed by non-radiative relaxation processes. As a result there is an increase of thermal energy in the material.

(2) Scattering is also partly a material property but is also caused by the imperfection in the fiber geometry. It occurs when the mode of propagation of light is changed, such that some of the optical energy leaves the fiber.

(3) Losses due to bending: Tight bends of the fiber similarly cause some of the light not to be internally reflected but to propagate into the cladding and be lost.

4. Loss due to imperfection: The presence of imperfection in fiber material also cause power losses from the fiber optic cables.

5. Connectors and splices which required to connect two fibers temporarily or permanently also cause power loss for long distance communication cables.

5.21 Photo Detector

After information carried by fiber optics cable, it should be detected by photo-detector. The most common semiconductor photo detector is the photo diode or photo-transistor. The photo detector detects the optical power falling upon it and converts the variation of this optical power into a correspondingly varying electric current. In optical fiber communication most commonly used system is IM/DD, where Intensity of the optical carrier, Modulated by biasing current in LED/Diode Laser (IM) and that variation of optical power Directly Detected in photo-detector (DD).

5.21.1 Mechanism of Photo-Diode

Figure 5.26(a) shows the basic structure and operation of a extrinsic *P-N* junction silicon photo diode. In normal operation a sufficiently large reverse bias voltage is applied across the device, so that depletion region as well as barrier potential will be large. Now when the light incident on the reverse biased *P-N* junction, then if the energy of the light quanta ($h\nu$) which is called as photon, is more than the band gap energy (E_g , i.e. $h\nu > E_g$), then the absorbed photon excite an

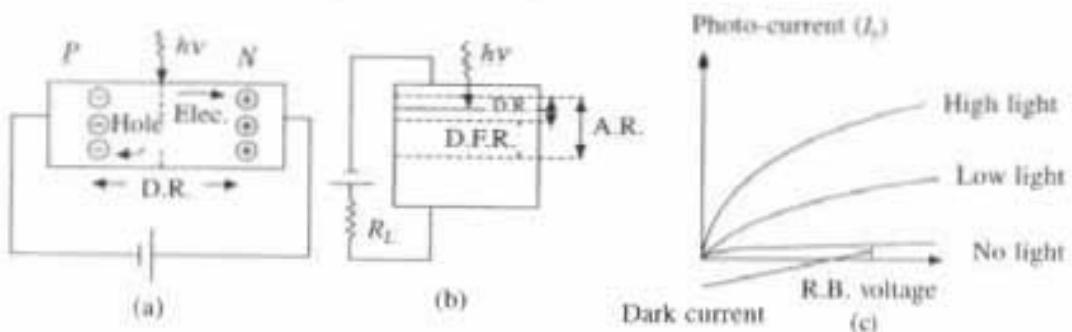


Fig. 5.26 Working principle of Photo-Diode

electron from V.B. to C.B. and leaving a hole in V.B. The electron-hole pairs created due to incident light, which are absorbed in the depletion region called as *photo carriers*.

As electron and hole flow in the depletion region (D.R.) in opposite direction under reverse bias voltage, some of the electron-hole pairs will recombine and hence lost. The distance they will traverse before they recombine is called *diffusion length* (L_n and L_p , for electron and hole respectively) and the time it takes for an electron and hole to recombine is known as carrier lifetime (T_n and T_p). The carrier lifetime and diffusion length can be related as

$$L_n = (D_n T_n)^{1/2}, L_p = (D_p T_p)^{1/2}$$

where D_n and D_p are the elec. and hole diffusion coefficient, respectively, unit of which is cm^2/sec .

Photons can be absorbed in depletion region (D.R.) as well as in diffusion region (D.F.R.) also, which is in general indicated as *absorption region* (A.R.) in Fig. 5.26b. So this region becomes a void of free photo carriers. Now within the depletion region these electron-hole pair cannot recombine, because of the main electrostatic field exists due to the depletion region voltage. Any photon induced electron-hole pair therein will be separated and drawn out (in opposite directions) by the combined effect of equivalent depletion region voltage as well as reverse bias voltage (notice that these two voltage polarities are in the same direction). This effect leading to a photo-current flow in the external circuit (Fig. 5.26a, 5.26c), as photo carriers drift away across the depletion region. Typical out put characteristics for R.B. P-N photo diode are shown in Fig. 5.26(c). Different operating conditions are giving from no light to high light level. When there is no incident light, then also a small reverse current will be there, which is known as dark current here.

If the electron-hole pair are produced outside the absorption region (A.R.) then they have higher probability to recombine, in which case no photo current will be available. It is, therefore, necessary to make the top layer as thin as possible and effective depletion region as wide as possible in order to maximize the efficiency of the photo-diode as photo detector, which is referred as quantum efficiency. Quantum efficiency of the photo-diode can be increased if one goes from ordinary photo-diode to PIN photo diode or avalanche photo-diode (APD) or reach through avalanche photo-diode (RAPD). Idea behind PIN, APD or RAPD is to increase effective depletion region which will increase quantum efficiency. Details of those are not discussing here.

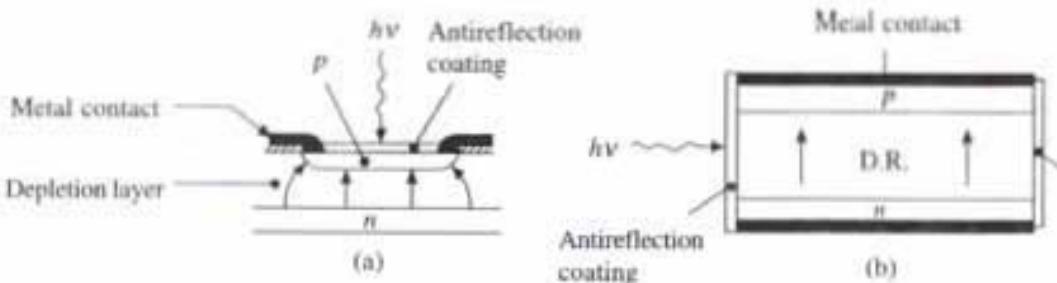


Fig. 5.27 Front illuminated and side illuminated Photo-Diode

There are generally two different structures possible for photo-diode, one is front illuminated (Fig. 5.27a) operating at 0.8 to 0.9 μm range, other is side illuminated Fig. (5.27b). Operating at higher wavelength 1.09 μm range.

5.21.2 Quantum Efficiency and Responsivity

For a photo detector quantum efficiency (η) and response speed or responsivity (R), are two important characteristics. These parameters depend on the material band gap, operating wave length, % of doping and thickness of P and N region. To achieve high quantum efficiency, the depletion layer must be thick enough to permit a large fraction of incident light to be absorbed. Due to that PIN or APD photo-diode is better than ordinary photo-diode. Quantum efficiency and responsivity can be defined as follows:

$$\text{Quantum efficiency } \eta = \frac{R_e}{R_p} = \frac{\text{Rate of electron-hole pair generated (elec./sec)}}{\text{Rate of incident photon (photon/sec)}}$$

Thus Internal quantum efficiency as given in eqn. (5.2a) $\eta_i = \frac{1}{\eta}$

$$\text{so } \eta = \frac{I_p/c}{P_o/hv} = 1/\eta_i \quad (5.19)$$

Responsivity $= R = I_p/P_o = \text{Outgoing photo current}/\text{Incident optical power}$

$$R = I_p/P_o = \eta e/hv = \eta e\lambda/hc \text{ A/watt} \quad (5.20)$$

$$\text{and incident optical power } P_o = \frac{I_p h c}{\eta e \lambda} = \frac{\eta_i I_p h c}{e \lambda} \quad (5.21)$$

where I_p = photo current, e = charge of electron, P_o = incident optical power, $h\nu$ = energy of photon and c = velocity of light.

5.21.3 Long Wave Length Cut-Off

It is essential when considering intrinsic absorption process, that the energy of the incident photon ($h\nu$) should be greater than the band gap energy (E_g) of the material used as photodiode, otherwise electron-hole pair will not generate, i.e.

$$h\nu \geq E_g \text{ or } hc/\lambda \geq E_g \text{ or } \lambda \leq \lambda_c \leq hc/E_g \quad (5.22)$$

Thus there will be threshold wave length for detection, which is commonly known as long wave length cut-off (λ_c). If the incident wave length of photon (λ) is more than λ_c , then photo detector will not work.

5.22 Components of an Optoelectronic Communication System (Block-Diagram)

The *block diagram* for fiber optics communication is shown in Fig. (5.28). The information to be transmitted (data, voice, music etc.) is first coded by a coder into a carrier wave of the optical signal. This requires a proper modulator and an optical source. The most *prominent source* to be used in optical communication

is *LED or semiconductor laser*, while the *detector* is *PIN or avalanche photo-diode or photo-transistor*. The optical beam that carries the information is characterised by its amplitude, frequency, intensity. All these parameters can be modulated to code the information properly. The signal is next coupled to an optical fiber. As the signal passes through the fiber at some stage it may need to be switched to another channels. This requires proper switching elements. Once data reaches the desired point, it is detected by the optical detector. The general optical signal is amplified, decoded by decoder (demodulator) as electrical signal and received by the receiver. Repeater unit may be necessary over relatively long-distance communication to counter the effects of fiber transmission losses and dispersion. In a repeater the signal is detected, amplified and then reemitted. A separate power supply line must be provided for all these units (Fig. 5.28).

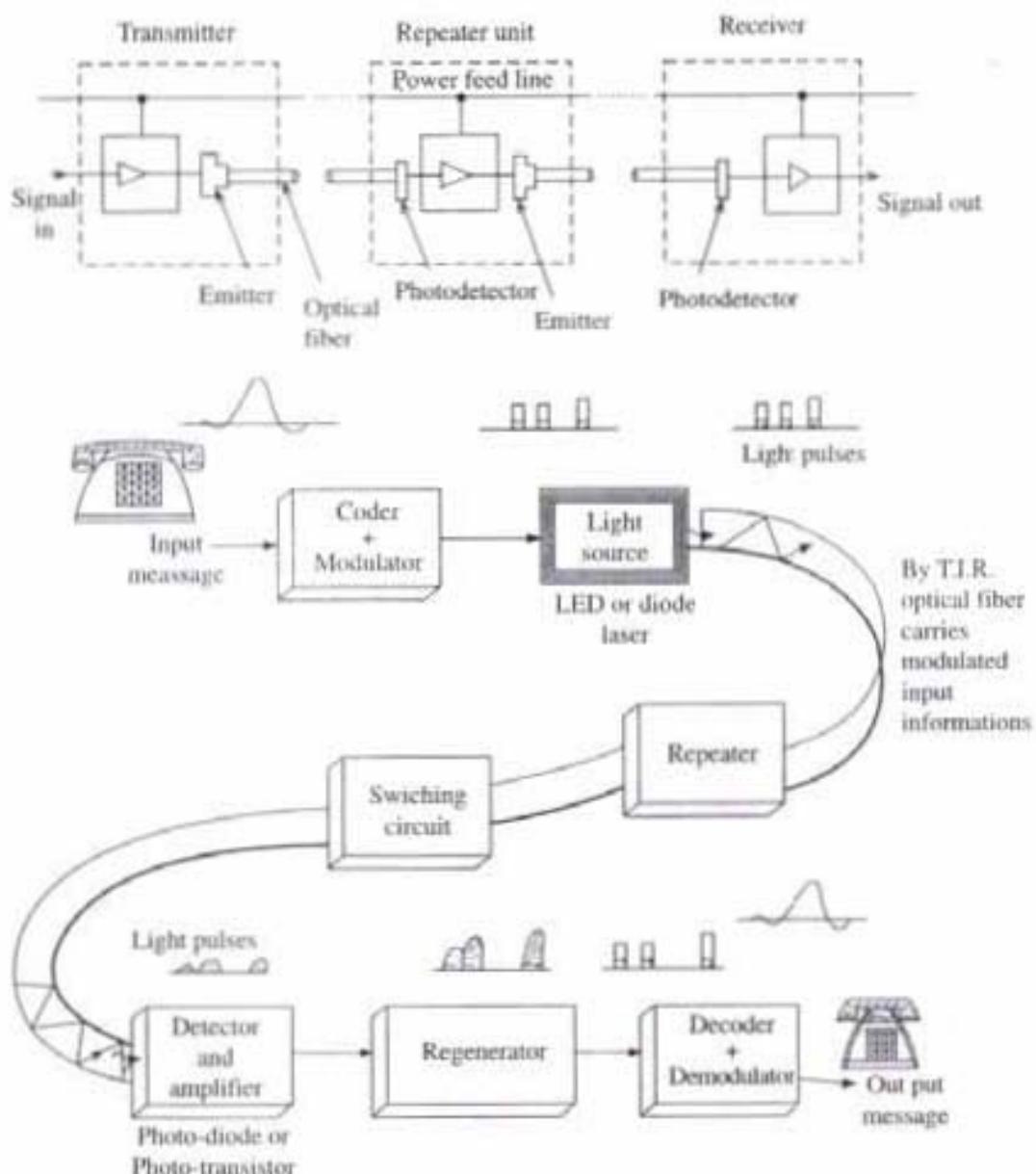


Fig. 5.28 Block diagram for optical communication system through optical fiber cable

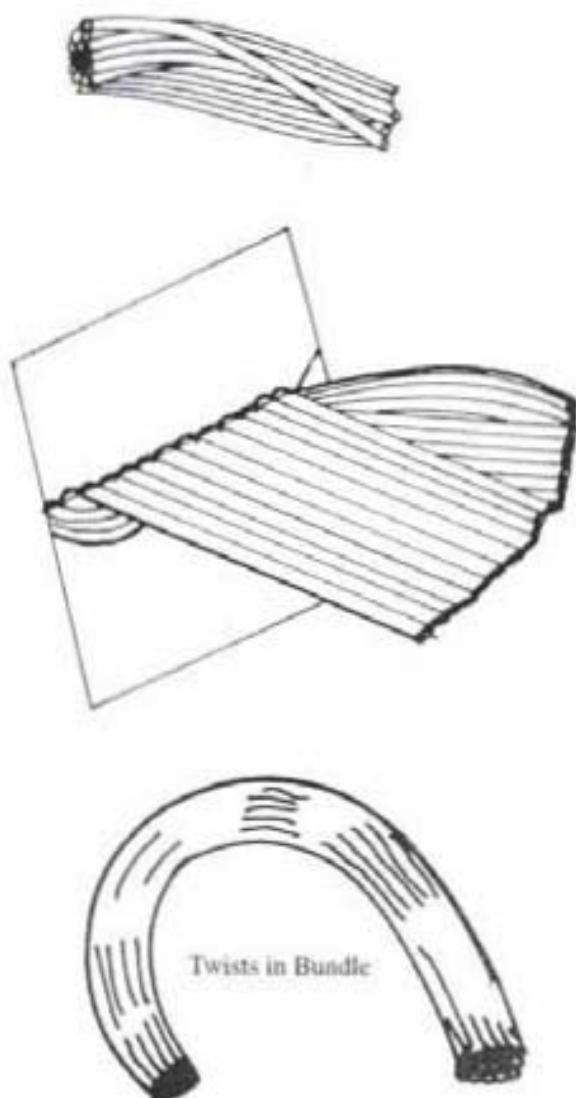


Fig. 5.29 Bundle of optical fiber

An optical communication system is quite similar in concept to electronic (microwave) communication system (10^9 Hz) only the frequency range is different (10^{15} Hz) and optical fiber instead of metallic fiber.

5.23 Uses of Optical Fiber

A single optical fiber can transmit light energy very efficiently (Figs. 5.20 and 5.21). But in order to transmit an image of an object we need assembly or bundles of fine glass fibers or plastic fibers because each fiber transmits rays from a very small region of the object (Fig. 5.29). The quality of the image depends upon the diameter of the fiber. It can be as small as 10^{-6} m. A bundle may consist of thousands of individual fibers.

Application of fiber optics in medicine are wide ranging. With the help of fiber optics, it is possible to study the interior of parts of the human body which cannot be viewed directly. It is also useful in the study of tissues and blood vessels far below the skin. For this purpose, a bundle of fibers is enclosed in a hypodermic needle. These flexible fiber bundles used in medicine are called "Endoscopes". Endoscopes are employed to make observations inside the human body, at a place which can be approached only along a curved path.

5.24 Advantages of Optical Fiber

(A) 1. Ability to transport light along flexible paths. This enables viewing of areas which could not be observed otherwise.

2. Ability to accept and transport light without loss of intensity.
3. Ability to emit light as well as guide it along a path.
4. Ability to propagate and couple different discrete waveguide modes.

(B) 1. Fiberoptic communication system have a large band width 16 Hz. The band width is the maximum rate at which information can be transmitted.

2. For under water communication it is very useful.
3. High security against tapping.
4. They are less expensive than copper wire
5. Glass fibers have no fire risk and there are no short circuit problems.
6. High capacity—one laser beam can carry upto 1,00,000 telephone conversations or 100 different color T.V. programmes.

This extraordinary capacity to handle communication traffic is mainly due to the fact that laser light can be switched on and off much faster than radio waves or microwaves used in communication.

PROBLEMS

1. Consider a multimode step index fiber with $n_1 = 1.53$ and $n_2 = 1.50$, and $\lambda = 1 \mu\text{m}$. If the core radius is $50 \mu\text{m}$ then calculate the normalised frequency of the fiber (V) and the number of guided mode.

$$\text{Solution} \quad V = \frac{2\pi a (\text{N.A.})}{\lambda} = \frac{2 \times 3.14 \times 50 \times 10^{-6}}{1 \times 10^{-6}} \times \sqrt{(1.53)^2 - (1.50)^2}$$

$V = 94.72$ = Normalised frequency

$$\text{Total number of guided mode} = M_s = \frac{V^2}{2} = 4486.$$

2. In order to make the above fiber single mode what will be the core radius a ?

Solution For single mode fiber

$$a \leq \frac{2.405 \lambda}{2\pi(\text{N.A.})} = \frac{2.405 \times 1 \times 10^{-6}}{2 \times 3.14 \times \sqrt{(1.53)^2 - (1.50)^2}}$$

$$a < 1.27 \times 10^{-6} \text{ m}$$

$$a < 1.27 \mu\text{m}.$$

3. The numerical aperture of an optical fiber is 0.5 and core refractive index 1.54.
(1) Find R.I. of cladding. (2) Calculate the change in core cladding refractive index per unit R.I. of the core.

Numerical aperture of optical fiber is

$$(A) \quad \text{N.A.} = \sin \alpha = \sqrt{n_1^2 - n_2^2}$$

n_1 = R.I. of core

n_2 = R.I. of cladding

N.A. = 0.5, $n_1 = 1.54$

Therefore $(0.5)^2 = (1.54)^2 - n_2^2$, $n_2 = 1.456$

Hence, R.I. of cladding = 1.456

(B) Change of core cladding R.I. per unit R.I. of the core.

$$\text{R.I. of the core} = \frac{n_1 - n_2}{n_1} = \frac{1.54 - 1.456}{1.54} = 0.0542$$

4. Numerical aperture of a fiber is 0.5 and core refractive index is 1.48. Find cladding refractive index and acceptance angle

$$\text{Numerical aperture} \quad \sin \alpha = \sqrt{n_1^2 - n_2^2} = 0.5$$

$$\text{R.I. of core} \quad = n_1 = 1.48 \quad \text{So } n_2 = 1.393$$

$$\text{Hence, R.I. of cladding} \quad n_2 = 1.393.$$

$$\text{Acceptance angle} = \alpha = \sin^{-1} (\text{N.A.}) = 30^\circ.$$

QUESTIONS

- Explain the operation of He-Ne laser with diagram showing essential components.
- Explain the construction of optical fiber and their use.
- Describe the process of obtaining three dimensional image by the holography method.
- What is N.A. of an optical fiber? What is its significance?
- Describe the expression for N.A. in terms of R.I. of the core and the cladding of an optical fiber.
- Light travelling in air strikes a glass plate at a glancing angle 33° . While striking the glass plate, part of the beam is reflected and part is refracted. If the refracted and reflected beams make an angle 90° , with each other, then: (1) what is the refractive index of the glass and (2) what is the critical angle for that glass?

Solution From Fig. 5.29(a)

$$i = 57^\circ$$

$$r = 33^\circ$$

(1) Thus, the refractive index of glass

$$\mu_g = \frac{\sin i}{\sin r} = 1.5398$$

(2) For $i = i_c$

$$r = 90^\circ$$

$$\text{Hence} \quad \frac{\sin i_c}{\sin r} = \frac{\mu_a}{\mu_g} \quad (\text{from Fig. 5.29b})$$

$$\text{i.e.} \quad \sin i_c = \frac{1}{\mu_g} = \frac{1}{1.5398}$$

$$i_c = 40.5^\circ$$

- Discuss CO₂ laser and give some of its practical applications.
- Give block diagram for optical communication through fiber optic cable. What are the advantages of optical communication?
- What is monomode and multimode optical fiber? How is N.A. involved with them? For long distance communication which is preferable?
- What is holography? How does it differ from ordinary photographic technique?

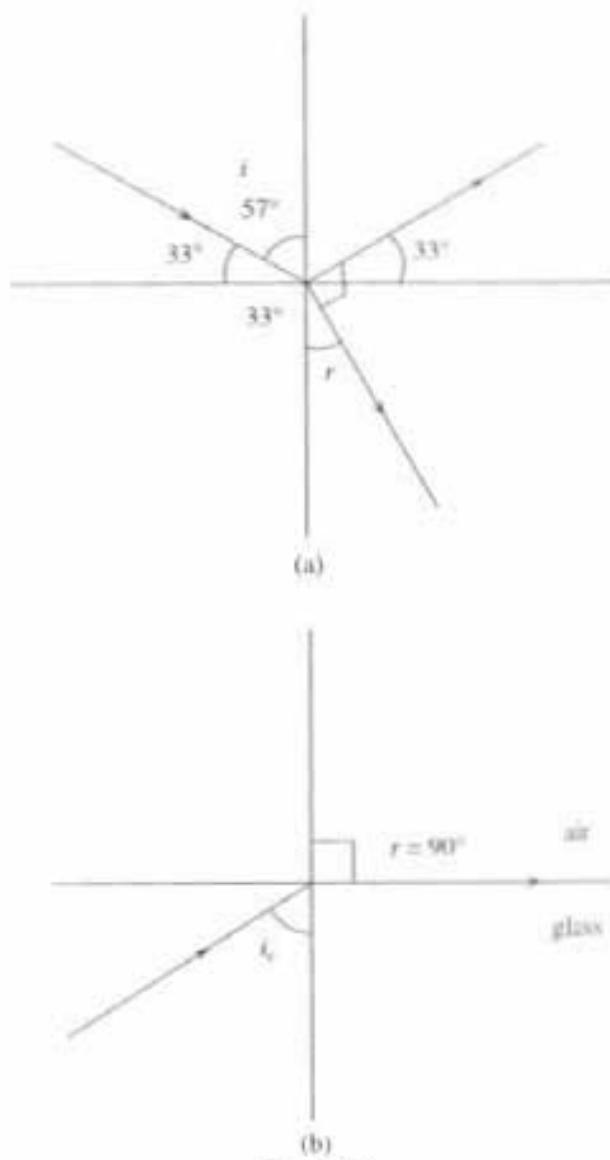


Fig. 5.29

How is interference and diffraction related with construction and reconstruction of the hologram. Explain in detail with suitable diagram

11. In what way laser light and starlight are similar? In what way do they differ?
12. What are approximate wave length range for X-ray, laser and ordinary light (visible)? Which way are they similar and which way do they differ?
13. Why are X-ray and laser so powerful than ordinary visible light?
14. With suitable diagram show light propagation in step index and graded index fiber. Explain how the graded index fiber produces less dispersion than step index fiber? For long-distance communication which will be preferable?
15. How is interference connected in holography "to freeze" the object and diffraction is connected "to unfreeze" the object?
16. What are the similarities and dissimilarities between X-ray, laser and ordinary light?
17. When 3×10^{11} photons each with wave length of $0.85 \mu\text{m}$ are incident on a photo diode, on average 1.2×10^{11} electrons are generated. Determine quantum efficiency and responsivity.

$$\text{Solution} \quad \text{quantum Efficiency} = \frac{\text{Re}}{\text{Rp}} = 1.2 \times \frac{10^{11}}{3} \times 10^{11} = 0.4$$

$$\text{Responsivity} = \eta e\lambda/hc = \frac{0.4 \times 1.602 \times 10^{-19} \times 0.85 \times 10^{-6}}{6.626 \times 10^{-34} \times 3 \times 10^8}$$

Hence

$$R = 0.274 \text{ A/W}$$

18. A silicon photo diode has quantum efficiency of 65% with photon energy $1.5 \times 10^{-19} \text{ J}$. Its band gap energy is 0.67 ev. Calculate

- (a) Long wave length cut off (λ_c)
- (b) Responsivity (R)
- (c) Incident power required to obtain a photo current $2.5 \mu\text{A}$ (P_o)?

Solution (a) Band gap energy $= E_g = 0.67 \text{ ev} = 0.67 \times 1.6 \times 10^{-19} \text{ J} = hc/\lambda_c$

$$\text{Hence Cut off wave length } \lambda_c = \frac{6.626 \times 10^{-34} \times 3 \times 10^8}{0.67 \times 1.6 \times 10^{-19}}$$

$$\text{i.e. } \lambda = 1.854 \times 10^{-6} \text{ m}$$

$$(b) R = \frac{I_p}{P_o} = \eta e/hv = \frac{0.65 \times 1.602 \times 10^{-19}}{1.5 \times 10^{-19}}$$

$$R = 0.694 \text{ A/W}$$

$$(c) \text{ Now } R = \frac{I_p}{P_o} = 0.694 = \frac{2.5 \times 10^{-6}}{P_o}$$

$$\text{i.e. } P_o = 3.60 \times 10^{-6} \text{ W}$$

19. Find how many Photons are required per second to Produce laser beam of 3 mw power, when wave length of laser light is 6943 \AA .

Solution:—

$$\text{Power } P = 3 \text{ mw} = 3 \times 10^{-3} \text{ W or J/sec}$$

$$P = \frac{n h v}{\text{time}} = \left(\frac{n h c}{\lambda} \right) / \text{time J/sec}$$

where n = number of Photon

$$h v = \frac{hc}{\lambda} = \text{energy of one photon}$$

$$\text{So } n = \frac{P \lambda}{hc} = \frac{3 \times 10^{-3} \times 6943 \times 10^{-10}}{6.625 \times 10^{-34} \times 3 \times 10^8} \text{ (since time is one second)}$$

$$n = 1.048 \times 10^{16}$$

20. Internal critical angle at core-cladding interface of a step index fiber is 80.6° . calculate maximum acceptance angle when cladding R.I. is $n_2 = 1.48$.

Solution:— Internal critical angle $i_c = \sin^{-1}(n_2/n_1) = 80.6^\circ$

$$\text{So maximum acceptance angle } \alpha_m = \sin^{-1} \sqrt{n_1^2 - n_2^2} = 14.13^\circ$$

Radioactivity and Nuclear Reactions

6.1 Introduction

Radioactivity was originally discovered by Becquerel in 1896. This phenomenon is confined almost entirely to the heaviest element. All the naturally occurring radioactive elements lie in the range of atomic number $Z = 81$ to 92 . Becquerel discovered that uranium gave out some type of radiations which were highly penetrating, could affect photographic plate wrapped in thick black paper, ionize gases and cause scintillation on fluorescent screen. After two years, a French couple Pierre Curie and Marie Curie found that a substance polonium from uranium ore, also emitted radioactive radiations which was many millions of times more active in emission than uranium. They named this element as 'radium'. The substances which emit these types of radiations are known as 'radioactive substances' and the phenomenon of spontaneous emission of radiation from radioactive substances is called 'radioactivity'. Uranium, polonium, radon, ionium, thorium, actinium etc. are radioactive substances.

It was also found that the radioactive behaviour was unaffected by physical and chemical changes. Radioactivity is essentially a nuclear phenomenon. It is a drastic process because the element changes in kind due to this emission of radiation. It is a spontaneous and an irreversible self-distracting activity.

Radioactive substances are usually found to emit either α -rays or β -rays, sometimes emission of γ -rays accompany these α and β decay.

1. α -ray is doubly-ionised positively charged helium atoms (${}_{2}He^4$), with both of their electrons removed.

2. β^- ray consists of electrons (${}_{-1}e^0$) with (-ve) charge and almost zero mass.

β^+ -rays consists of positron (${}_{+1}e^0$) with (+ve) charge, same in magnitude as an electron but with opposite sign and almost zero mass same as electron.

3. γ -ray or photons is an electromagnetic wave having frequency higher than that of X-rays. They have no charge.

6.2 Electrical Instruments as Detectors/Geiger-Muller Counter

Ionization chamber, proportional counter, and Geiger-Mueller counter, each of these three detectors, used to detect radioactive radiation, is based on the production of ionization in a gas and the separation and collection of the ions by means of

an electrostatic field. The differences in the three systems can be explained with the aid of Fig. 6.1 which shows a cylindrical conducting chamber containing a central conducting electrode located on the axis of the chamber and insulated from it. The chamber is filled with a gas at a pressure of one atmosphere or less. A voltage V is applied between the wall and the central electrode through resistance R shunted by the capacitor C ; the central electrode is at a positive potential relative to that of the chamber wall.

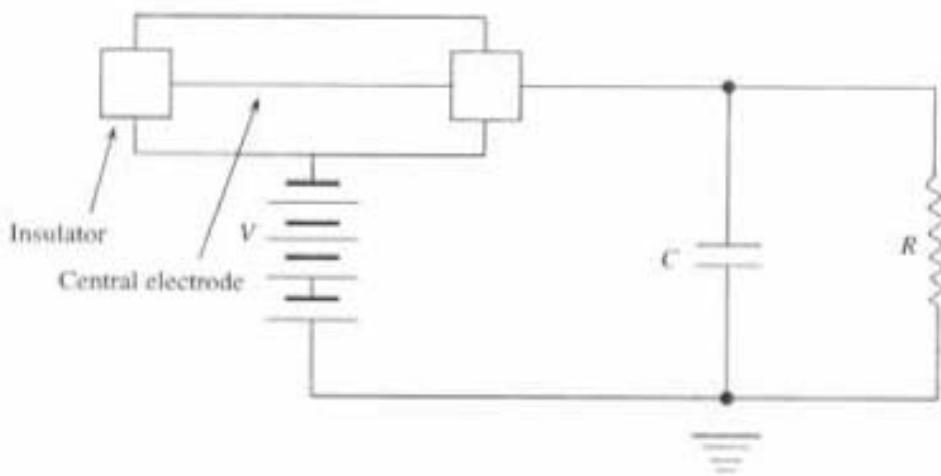


Fig. 6.1 Nuclear detector for detecting α and β rays.

Let us suppose that some ionization occurs in the gas because of the passage of a charged particle. Each ion pair consists of a positive ion and an electron. Now for a given initial ionization, how many ion pairs are collected or how many electrons reach the central electrode as the applied voltage is varied, is shown in Fig. (6.2), where curves of total ion collection are plotted as functions of applied voltage. For convenience, the logarithm to the base 10 of the number ($\log n$) of ion pairs has been used as the ordinate. If there is no voltage across the electrodes, the ions will recombine and no charge will appear on the capacitor. As the voltage is increased, say to a few volts, then there is competition between the loss of ion pairs by recombination and the removal of ions by collection on the electrodes. Some electrons will reach the central electrode. At a voltage V_1 of the order of some 10 volts, the loss of ions by recombination becomes negligible and all electrons reach the central electrode; $\log n$ reaches the value unity. As voltage is increased from V_1 , n stays constant until a voltage V_2 is reached which may be some tens or hundreds of volts depending on the conditions of the experiment. The region between V_1 and V_2 , in which the number of ion pairs collected is independent of the applied voltage and the curve is horizontal, is called the *ionization chamber region*.

When voltage is increased above V_2 , n increases above 10 because of a phenomenon called gas multiplication or gas amplification. The electrons released in the primary ionization acquire enough energy to produce additional ionization, when they collide with gas molecules, and n increase roughly exponentially with V . Each initial electron produces a small 'avalanche' of electrons with most of these secondary electrons liberated close to the central electrode. If number of

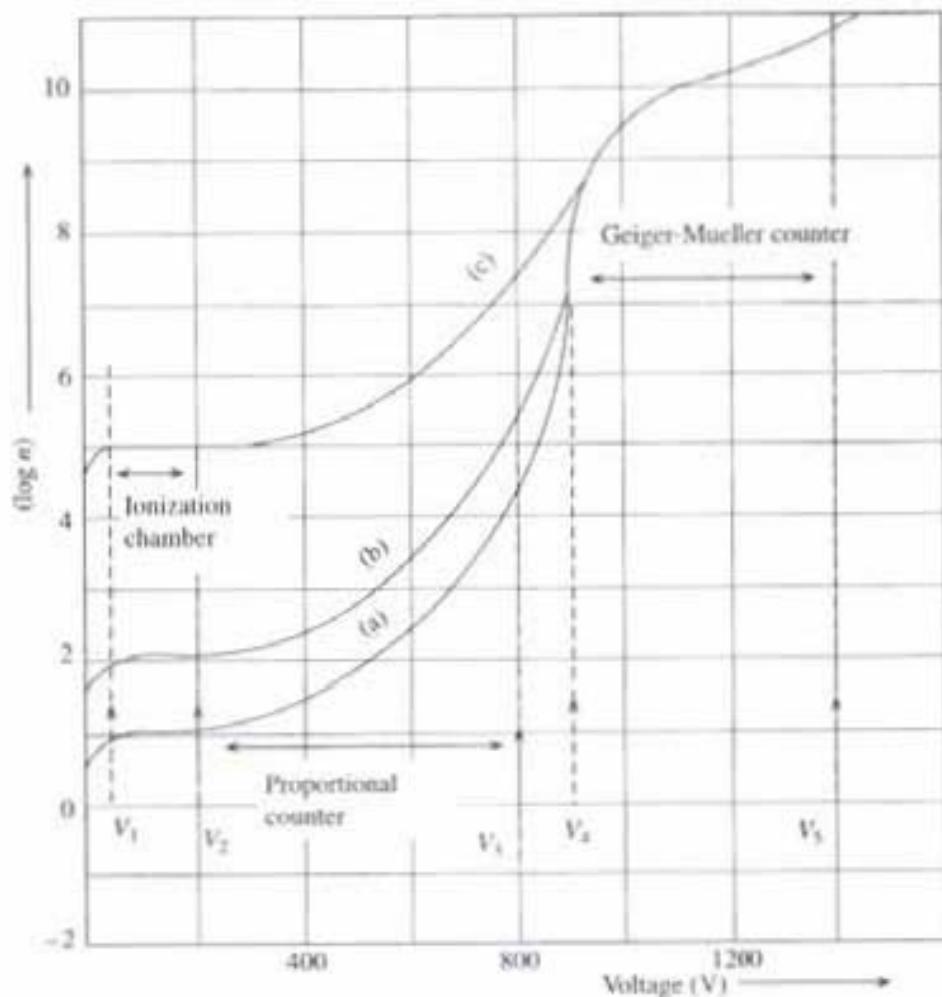


Fig. 6.2 Variation of total ion pair produced ($\log n$) with applied voltage (V)

ion pairs are formed initially is more, curve b results; it is parallel to curve a in the ionization chamber region and is separated from it by one unit on the $\log n$ scale. The behaviour of the two curves above V_2 is interesting. For some range of voltages, upto V_3 , each electron acts independently and gives its own avalanche, not being affected by the presence of the other electrons. Curves a and b continue to be parallel with a difference between them of one unit on the $\log n$ scale. Between V_2 and V_3 , the number of ion pairs collected is then proportional to the initial ionization. This is the region of *proportional counter operation*.

Above V_3 , the gas multiplication effect continues to increase very rapidly, and as more electrons produce avalanches, the latter begin to interact with one another; the positive ion space charge of one avalanche inhibits the development of the next avalanche. The discharge of initial electrons in curve b is affected with a less number of initial electrons which increases less rapidly than curve a ; curves a and b approach each other and eventually meet at V_4 . The region between V_3 and V_4 is the *region of limited proportionality*. Above V_4 the charge collected becomes independent of the ionization initiating it, and curves a and b becomes identical. Gas multiplication increases the total number of ions to a value that is limited by the characteristics of the chamber and the external circuit. The region above V_4 is the *region of Geiger-Mueller counter operation*. It ends at a voltage

V_5 where the discharge tends to propagate itself indefinitely; V_5 marks the end of the useful voltage scale, the region above being that of *continuous discharge*.

If the initial ionization is very large, curve *c* is obtained. It is similar to curves *a* and *b*, and parallel to them between V_2 and V_3 , the proportional counter region ends sooner, instead of at V_3 . Thus, the extent of that region depends on the initial ionization.

As a result of the behaviour of the ions of the gas in the electrostatic field, three detection instruments have been developed.

A. The ionization chamber which operates at voltages in the range V_1 to V_2 , is characterized by complete collection, without gas amplification, of all the electrons initially liberated by the passage of the particle. Subject to certain conditions, it will give a pulse proportional to the number of these electrons. Fig. 6.1 schematically shows a chamber which could be used as an ionization chamber between 10 and 200 volts, approximately. The numbers in Fig. 6.1 would be appropriate for a counter with an outer cylinder 1 cm in radius, a wire central electrode 0.01 cm in radius, and filled with gas to a pressure of a few centimeters of Hg (e.g. argon at about 6 cm of Hg plus a little alcohol). The counter length would be about 10 cm. The electrodes of an ionization chamber may also be parallel plates.

B. The Proportional counter which operates in the voltage region V_2 to V_3 is characterized by a gas multiplication independent of the number of initial electrons. Hence, although gas multiplication is utilized, the pulse is always proportional to the initial ionization. The use of this counter permits both the counting and energy determination of particles which do not produce enough ions to yield a detectable pulse in the region V_1 to V_2 . The proportional counter, therefore, offers advantages for pulse-type measurements of beta radiation, an application for which ionization chambers are not sensitive enough.

C. Geiger-Mueller Counter (G-M Counter or G-M Tube). The Geiger-Muller (G-M) counter, named after its inventors is a very efficient and accurate device for counting individual *a*, *b* or *g* particles emitted from radioactive substances, it operates in the voltage region V_4 to V_5 and is characterized by the spread of the discharge throughout the entire length of the counter, resulting in a pulse independent of the initial ionization. It is specially useful for counting lightly ionizing particles such as *b* particles or *g*-rays. The G-M counter usually consists of a fine wire (e.g. tungsten) mounted along the axis of the tube which contains a gas at a pressure of about 2 to 10 cm of Hg. The tube may be a metal such as copper, or a metal cylinder may be supported inside a glass tube (Fig. 6.3a), a mixture of 90% Argon and 10% ethyl alcohol is suitable. A p.d. which may be between 800 to 2000 volts, is applied between cylinder and wire which acts as anode. The value of this voltage is adjusted to be somewhat below the breakdown potential of the gaseous mixture. When the ionising particle passes through the tube, it ionizes the gaseous mixture, thereby releasing a number of ion-pairs. The (+ve) and (-ve) ions so produced are drawn towards the appropriate electrodes and the resulting pulse of a current through the tube produces a voltage pulse (10 mV) across *R*. An electronic pulse amplifier

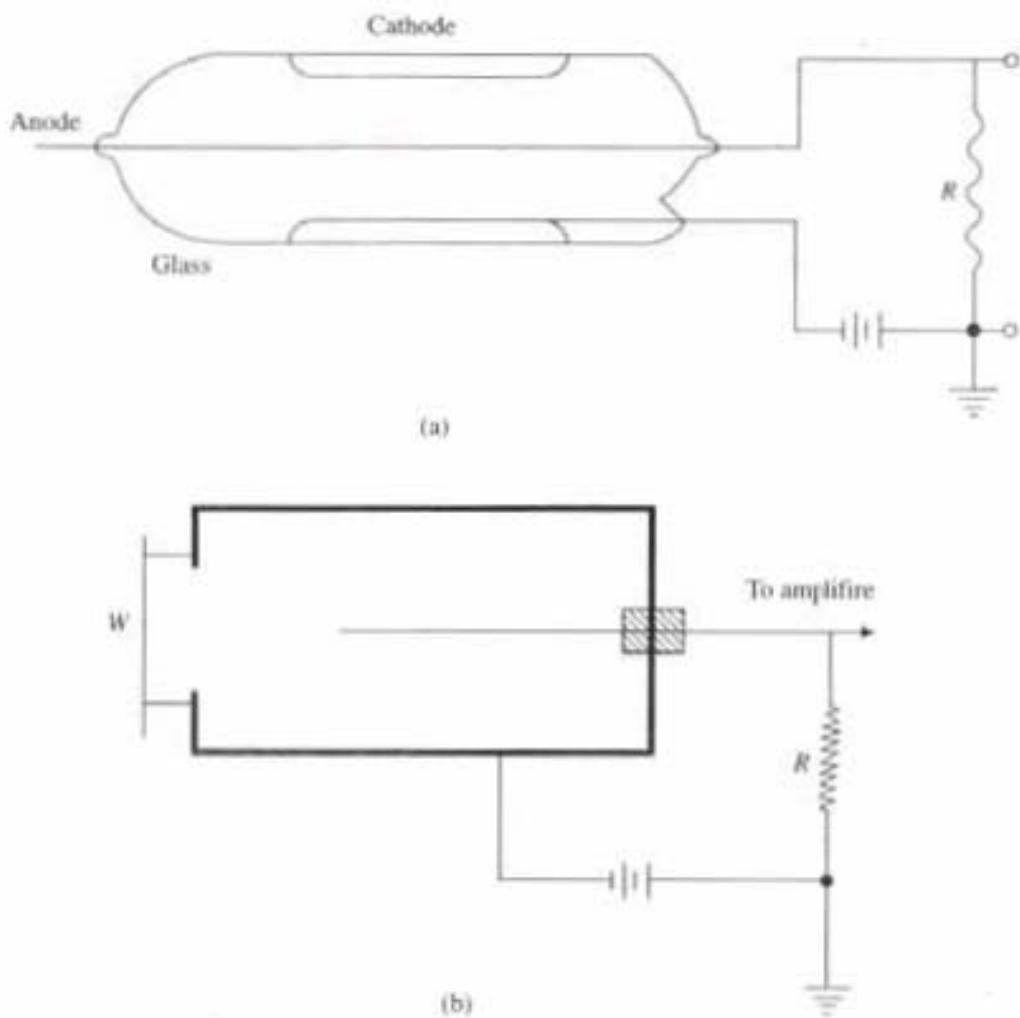


Fig. 6.3 Geiger-Mueller counter and Geiger-Mueller tube

accepts these pulse voltages and amplifies them to about 5 to 10 V. This amplified output is then applied to a relay type recorder or to a counting-rate meter, so the rate of coming of β - or γ -rays can be measured. For very high pulse rates, a suitable scaler is used in conjunction with the G-M tube.

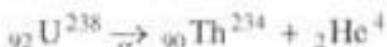
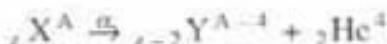
G.M. counter tube discussed above is suitable for β - and γ -rays measurement, it is not suitable for α -particles, because the walls of the glass tube are too thick for α -particles to penetrate. For detection of α -particles, the Geiger Point Counter shown in Fig. 6.3(b) is used. It consists of an axially mounted fine wire inside a tube which is maintained at a (+ve) potential w.r.t. the tube. The tube contains dry air at atmospheric pressure and has an extremely thin foil window at one end. The α -particles to be detected enter the tube through the window (W).

6.3 Radioactive Disintegration or Decay

Atoms of heavy elements like uranium, thorium, polonium and radium etc. are constantly breaking up into fresh, radioactive atoms with the emission of α -, β - or γ -rays, from their nuclei. In the process, the original (parent) atom disappears and gives rise to new (daughter) atom. These new atoms are also, in general, radioactive and hence spontaneously break up, in their turn, thereby leading to a long chain of different radioactive elements in the form of a radioactive series

until an inactive element (usually lead) is reached. For naturally occurring radioactive elements the three well-known series are *uranium-radium series*, *Thorium series* and *actinium series*. This spontaneous breaking up of the nucleus is known as *radioactive disintegration*.

For example an element X of mass number A and atomic number Z while emitting an α -particle ($_2\text{He}^4$) is converted into another element Y having mass number less than 4 and atomic number less than 2.



Similarly when an element disintegrating by the emission of β^- -rays, it is turned into another new element whose atomic number (number of proton) will increase by 1 whereas mass will be the same.



6.4 Natural Radio-Isotopes

It was found generally that the product of radioactive decay is itself radioactive, and that each of the decay products behave chemically in a different manner from its immediate parent and its daughter product.

Isotopes: Let us consider the following radioactive disintegrations:



where ${}_{88}\text{Ra}^{226}$ is the product of disintegration of Ionium. The chemical properties of ${}_{90}\text{Ionium}^{230}$ turned out to be similar to those of naturally occurring radioactive element Thorium (${}_{90}\text{Th}^{232}$), in fact the two elements when mixed could not be separated chemically even with extremely sensitive methods. Moreover, ionium and thorium were found to be spectroscopically identical. Thus the pair of elements ${}_{90}\text{Ionium}^{230}$ and ${}_{90}\text{Th}^{232}$ which were chemically identical but differed in atomic weight are called *isotopes*. Isotopes means the same position in periodic table. *Thus if two atomic species ${}_z\text{X}^A$ and ${}_z\text{Y}^{A'}$ which have different atomic masses (A, A') inspite of being different element and have same atomic number (Z) and identical chemical properties, then those different species are called isotopes.* ${}_{92}\text{U}^{238}$ and ${}_{92}\text{U}^{234}$ are natural Isotopes, similarly ${}_{90}\text{Ionium}^{230}$, ${}_{90}\text{Th}^{232}$, ${}_{90}\text{UX}_1^{234}$ are all natural isotopes.

Isobars: The following elements ${}_{90}\text{UX}_1^{234}$, ${}_{91}\text{UX}_2^{234}$, ${}_{92}\text{U}^{234}$ all have same mass number (A) but atomic number (Z) is different so they are chemically different element. *These elements which are having same mass number (A) but different atomic number (Z) are called isobars.*

6.5 Law of Radioactive Disintegration

From the experimental observation Rutherford and Soddy formulated the theory of radioactive change. The rate at which a particular radioactive material disintegrates or decays is a constant, that is almost completely independent of all physical and chemical conditions. The intensity of the radioactivity which is called *Activity*, is proportional to the no. of atoms which disintegrates per unit time. If dN is the average no., that will decay in a small time interval dt , it is found to be proportional to the no. of atoms N present at time t . That is the rate of disintegration.

$$-dN/dt \propto N$$

$$dN/dt = -\lambda N = A/C \quad (6.1)$$

where λ is the const. of proportionality, known as *disintegration const.* or *decay const.* It is the characteristics of a particular radioactive isotope and independent of all external conditions like temperature, pressure (-ve) sign indicates that there is a decrease in the original no. of atoms in present element. In Eqn. (6.1) A is the activity and C is the detection coefficient. N and A are always interchangeable by the relation $A/C = -\lambda N$.

Now from Eqn. (6.1) we get $dN/N = -\lambda dt$.

If at $t = 0$, $N = N_0$ and at $t = t$, $N = N$ then by integration within the limit 0 to t , we get

$$\int_{N_0}^N \frac{dN}{N} = -\lambda \int_0^t dt$$

$$[\log N]_{N_0}^N = -[\lambda t]_0^t, \frac{\log N}{\log N_0} = -\lambda t$$

Hence

$$\frac{N}{N_0} = e^{-\lambda t}$$

$$N = N_0 e^{-\lambda t} \quad (6.2)$$

Equation (6.2) shows that the no. of atoms of a given radioactive substance decreases exponentially with time (Fig. 6.4). This graph shows that N is theoretically never zero but $N \rightarrow 0$ as $t \rightarrow \infty$. This theoretical decay equation is experimentally proved for the decay of many radioactive elements.

6.6 Half Life

A radioactive nuclide may be characterized by the rate at which it disintegrates. From Eqn. (6.2) and Fig. 6.4 it is clear that infinite time will be taken by a radioactive substance to disappear completely. So the quantity which is generally used to characterize radio nuclides is the *half life* ($T_{1/2}$), the time in which the number of radioactive atoms are reduced to one half of its original value i.e.

$$N/N_0 = 1/2$$

In general

$$N/N_0 = e^{-\lambda t}$$

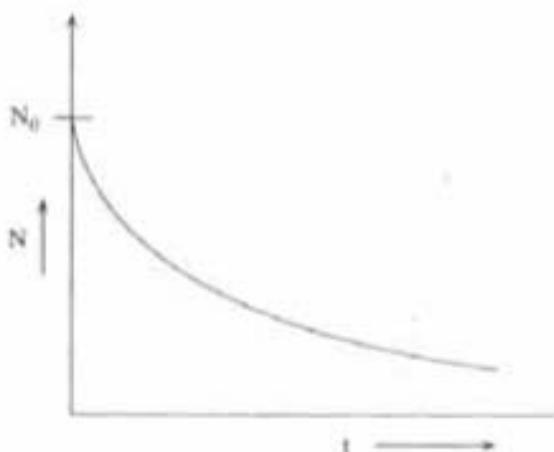


Fig. 6.4 Radioactive decay

when $t = T_{1/2}$ then $N/N_0 = e^{-\lambda T_{1/2}} = 1/2$

$$\begin{aligned}\lambda T_{1/2} &= \log_2 = \log 2 \times 2.3026 \\ &= 0.3010 \times 2.3026 = 0.693\end{aligned}$$

$$T_{1/2} = \frac{0.693}{\lambda} \quad (6.3)$$

6.7 Nuclear Force

The nucleus constitutes of positively charged protons and neutral neutrons. Because of the positively charged protons, there must be repulsive electrostatic forces between the protons, tending to push the nucleons apart. But it is apparent from the small size and large density of the nucleus that, nuclear forces must be very strong in comparison with the forces between the nucleus and the extra nuclear electrons. Hence if stable complex nuclei are to exist, there must be attractive forces, specifically nuclear forces between a proton and a neutron, between two neutrons or between two protons. They seem to be more complex than the gravitational or electromagnetic forces of classical physics. The nuclear attractive forces must be very strong at distances of the order of nuclear radius i.e., *nuclear forces are short range forces*. Outside the nucleus they decrease very rapidly. The magnitude of the nuclear forces is one million times greater than the force needed to separate an extra nuclear electron from the atom.

In 1935 Yukawa predicted the existence of a particle, now called a "meson", having a rest mass between the electron and nucleon masses. He proposed a theory to classify the nuclear forces with the help of the meson and is referred to as "meson theory of nuclear forces".

NUCLEAR BINDING ENERGY

6.8 Packing Fraction

The existence of a slight but systematic departure of the atomic masses from the whole number in every case was first studied by Aston. He gave the following formula for checking the stability of the atom.

Packing fraction = $f = \frac{\text{Actual atomic mass} - A}{A}$

$$\boxed{\text{Packing fraction} = f = \frac{M_{z,A} - A}{A}} \quad (6.4)$$

The packing fraction is the mean gain or loss per particle in atomic mass, in order to pack the elementary particle to form a stable nucleus. The variation of f with mass no. shown in Fig. 6.5. From the graph it is clear that except for smaller and higher values of mass no., the graph is having almost flat minimum with (-ve) values of f . When f is (-ve), that means mass no. A is larger than actual atomic mass, which implies the stability of the nucleus. When f is (+ve) that means the nucleus is unstable.

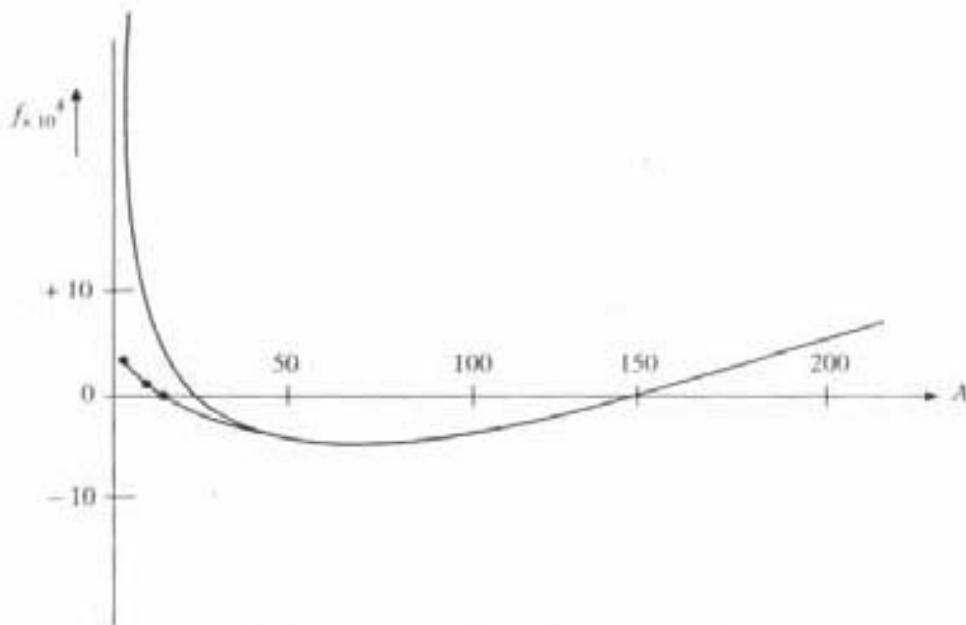


Fig. 6.5 Variation of atomic packing fraction (f) with mass number (A)

6.9 Binding Energy of the Nucleus

It might at first be supposed that the mass of an atom should be the sum of the masses of its constituent particles. But a survey of the atomic masses shows that the atomic mass is less than the sum of its constituent particles in the free state. To account for this difference in mass, Einstein's mass energy relation is used. The grouping together of the elementary particles (protons and neutrons) into a stable nucleus, involves a certain interchange of energy. This 'energy' is the "Binding energy" (B.E.) of the nucleus i.e. the energy which is either liberated or absorbed when the constituent particles are bound together into a nucleus.

If ΔM is the decrease in mass when a number of protons, neutrons and electrons combine to form an atom, then the above principle states that an amount of energy equal to

$$\boxed{\Delta E = \Delta M c^2} \quad (6.5)$$

is released in the process. The same amount of energy would be needed to break the atom into its constituent particles. Conversely if ΔM is the increase in mass

due to the combination of protons, neutrons and electrons to form the atom then the amount of energy $E = \Delta Mc^2$ will be absorbed in this process and when the atom will break, this much amount of energy will be released. The difference in mass, ΔM is called the *mass defect*, it is that amount of mass which would be converted into energy. Thus the energy equivalence of the mass defect is therefore a measure of the binding energy of the nucleus.

Let the constituents of the atom be Z -protons, Z -electrons and $(A - Z)$ neutrons. Masses of Z -protons and Z -electrons are equivalent to mass of Z -hydrogen atoms. So the mass defect can be written as

$$\Delta M = ZM_H + (A - Z)M_n - M_{Z,A}$$

where M_H = mass of hydrogen atom in a.m.u.

M_n = mass of neutron in a.m.u.

$M_{Z,A}$ = atomic mass, the actual mass of the atom in a.m.u.

Since the mass of an electron is very negligible so the main contribution to the mass of the atom is due to the nucleus, i.e., due to the mass of protons and mass of neutrons. So we can write the mass defect as

$$\Delta M = ZM_p + (A - Z)M_n - M_{Z,A}$$

where M_p is the mass of proton in a.m.u.

Since 1 a.m.u. is equivalent to 931.48 MeV, so B.E. of the nucleus is

$$\text{B.E.} = \Delta Mc^2 = 931.48 [ZM_p + (A - Z)M_n - M_{Z,A}] \quad (6.6)$$

Binding energy per nucleon is

$$\text{B.E./A} = 931.48 \left[\frac{Z}{A} M_p - \left(1 - \frac{Z}{A}\right) M_n - \frac{M_{Z,A}}{A} \right]$$

A graph of the B.E. per nucleon as a function of the mass no. is shown in Fig. (6.6). With the exception of He^4 , C^{12} , O^{16} the values of the B.E./nucleon lie on or close to a single curve. The graph gives us the following results.

1. B.E./nucleon for some very light nuclides are very small, say for H, He etc.
2. B.E./nucleon rises sharply and reaches a maximum value of 8.8 MeV in the neighbourhood of $A = 50$.
3. The maximum is quite flat, and the B.E./nucleon is still 8.4 MeV at $A = 140$.
4. For higher mass no., the value B.E./A decreases to about 7.6 MeV at Uranium at $A = 238$. Two ends of the packing fraction curve shows that packing (Fig. 6.5) fraction f is (+ve) for light and heavy nuclides due to which B.E./A is also less for light and heavy nuclides. It is this decrease in B.E./A which is the fundamental cause of release of energy in the fusion of light elements and in splitting (fission) of a heavy element.
5. Over a considerable range of mass numbers [where packing fraction f is (-ve)] the B.E./A has an average value of about 8 MeV.

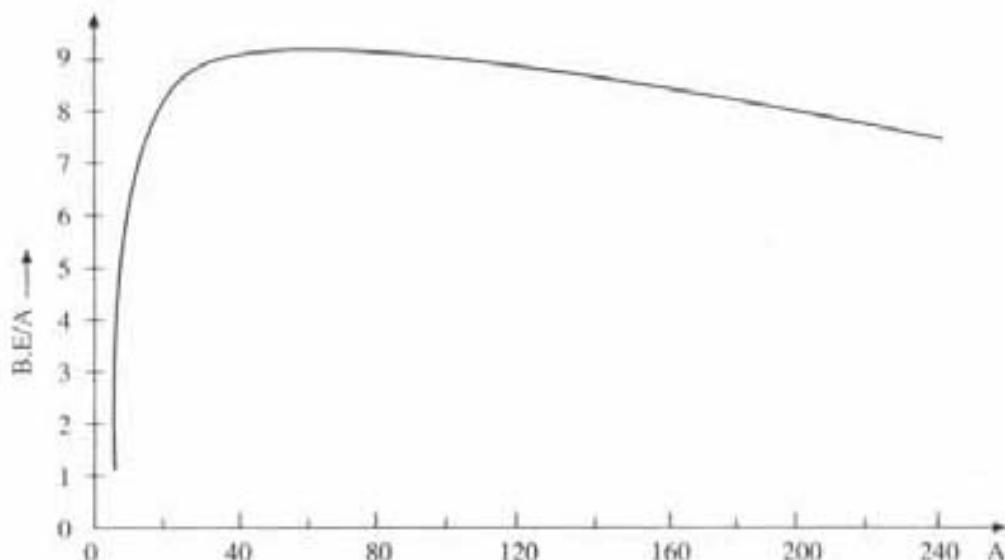


Fig. 6.6 Variation of binding energy per nucleon (B.E./A) with mass number A

6. The (+ve) value of the B.E./A for all values of A shows the stability of the nuclei. This shows that there will be attractive nuclear forces between the nucleons. And since the nuclei does not collapse, these forces must become repulsive for very close distance between nucleons.
7. The value of B.E./A vary in a very peculiar manner from nucleus to nucleus for very light nuclides. There is a rapid increase in B.E./A for light nuclei with a notable peak at $A = 4n$ (say for carbon-12, oxygen-16). This reflects the peculiar stability of the α -particle structure.
8. Above $A > 20$ the B.E./A is remarkably constant (upto 10%). This reflects an important property of the nuclear forces, called Saturation Characteristics. This indicate that B.E. is nearly proportional to A i.e. the nuclear constituents interact with only a limited number of their neighbour or the nuclear forces are saturated.
9. Discontinuities in the B.E. curve were observed by several workers when Z or N equal to the magic number 2, 8, 20, 50, 82, 126 where B.E. is very high.

6.10 Q-Value

A nuclear reaction can be analysed quantitatively in terms of the masses and energies of the nuclei and the particles involved. The analysis of nuclear reactions is one of the main sources of information about nuclear properties. The analysis is same as chemical reaction except that the relativistic relation between mass and energy must be taken into account. Consider a nuclear reaction represented by the equation



where X is the target nucleus, x the bombarding particle, Y the product nucleus and y is the product particle. It will be assumed that the target nucleus X is initially at rest so that it has no kinetic energy. Since the total energy of a particle

of atom is the Sum of the rest energy and the Kinetic energy, the statement that the total energy is conserved in the nuclear reaction means that

$$(E_x + m_x c^2) + M_X c^2 = (E_Y + M_Y c^2) + (E_z + m_z c^2) \quad (6.9)$$

here m_x, M_x, m_y, M_y represent the masses of the incident particle, target nucleus, product particle and product nucleus, respectively. The E 's represent the kinetic energies. *Here excited state of product nucleus, and the emission of more than one particle is not taken into account.*

Now Q represents the difference between the Kinetic energy of the products of the reaction and that of the incident particle

$$Q = E_Y + E_z - E_x = (M_X + m_x - M_Y - m_y)c^2 \quad (6.10)$$

The quantity Q is called the "energy balance" of the reaction or, more commonly, the *Q-value*. If the value of Q is positive, kinetic energy of the products is greater than that of the reactants; the reaction is then said to be '*exothermic*'. The total mass of the reactants is greater than that of the products in this case. If the value of Q is negative, the reaction is called '*Endothermic*'. It is apparent from Eqn. (6.10) that the analysis of nuclear reactions can, therefore be used to obtain information about the masses of nuclei, about particle energies, or about Q -values, depending on what information is available, and which quantities can be measured.

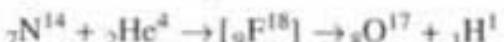
6.11 Artificial Transmutation or Artificial Disintegration⁷

The fact that radioactive atoms undergo spontaneous disintegration led to speculation about the possibility of causing the disintegration of the ordinary inactive nuclides. When by some artificial means it is possible to convert one type of atom into another type then that process is known as artificial transmutation or artificial disintegration. This process is entirely different from spontaneous disintegration.

6.12 Rutherford's Experiment

In 1919 earliest experiments on artificial nuclear disintegration were performed by Rutherford. The disintegration was caused by α -particles from the "RaC" and resulted in the emission of a highly energetic proton from ordinary nitrogen nucleus.

The disintegration of the N atom due to the bombardment of α -particle may be represented by an equation analogous to those used for chemical reactions.



The α -particle (${}_2\text{He}^4$) is captured by Nitrogen atom as it hits, and a new unstable compound nucleus $[{}_{9}\text{F}^{18}]$ is formed which emits proton. This transmutation can be represented by the abbreviated notation $\text{N}^{14} (\alpha, p)\text{O}^{17}$.

6.13 Artificial Radioactivity and Artificial Radio-Isotopes

In 1934 Curie and Joliot while studying the effect of a particle bombardment on

the light elements like Al, B, Mg observed that the product of the induced nuclear transmutation are also radioactive. The elements bombarded by α -particles continued to emit radiation (Proton, Neutron), as expected from the known (α, p), (α, n) reactions even when the source of α -particles had been removed. In addition to these particles, positive electrons i.e. positrons were also observed. (+ve) electron or positron is the fundamental particle whose rest mass is the same as that of an electron and whose charge has the same magnitude but opposite sign from that of the electron. It was also observed that the rate of emission of positrons gradually decreased with time, as in the case of natural radioactive elements ($N = N_0 e^{-\lambda t}$). They therefore concluded that the product nucleus formed in the (α, n) reaction in each case was an unstable nuclide, i.e. the product of nuclear transmutation is radioactive which then disintegrates with the emission of a positron. This process in which a stable nucleus changed into "radioactive isotope" by some artificial means is called "*artificial or induced radioactivity*". This process is entirely different from the spontaneous or natural radioactivity. The whole process for Al can be explained as follows:



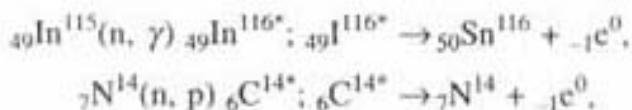
The Al nucleus after capturing the α -particle is converted into an isotope of Phosphorous (${}_{15}^{\text{P}} \text{P}^{30*}$) with the ejection of a neutron. The Phosphorous (${}_{15}^{\text{P}} \text{P}^{30*}$) so produced is not stable and is a radio-isotope which disintegrates spontaneously and forms a stable isotope of silicon (${}_{14}^{\text{Si}} \text{Si}^{30}$) with the emission of a positron (${}_{+1}^{\text{e}} \text{e}^0$).

Some other reactions for artificial radion activities are:

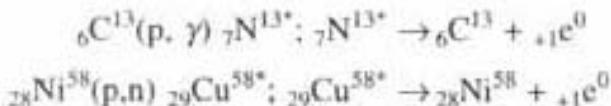
1. ${}_{-5}^{\text{B}} \text{B}^{10} + {}_2^{\text{He}} \text{He}^4 \rightarrow {}_7^{\text{N}} \text{N}^{14} \rightarrow {}_7^{\text{N}} \text{N}^{15*} + {}_0^1 \text{n}^1$
 ${}_7^{\text{N}} \text{N}^{15*} \rightarrow {}_6^{\text{C}} \text{C}^{15} + {}_{+1}^{\text{e}} \text{e}^0$
2. ${}_{11}^{\text{Na}} \text{Na}^{23} + {}_2^{\text{He}} \text{He}^4 \rightarrow {}_{13}^{\text{Al}} \text{Al}^{26*} + {}_0^1 \text{n}^1$
3. ${}_{-7}^{\text{N}} \text{N}^{14} + {}_2^{\text{He}} \text{He}^4 \rightarrow {}_9^{\text{F}} \text{F}^{17*} + {}_0^1 \text{n}^1$
 ${}_9^{\text{F}} \text{F}^{17*} \rightarrow {}_8^{\text{O}} \text{O}^{17} + {}_{+1}^{\text{e}} \text{e}^0$

All the elements with * mark are the artificial radio-isotopes of that respective element, ${}_{+1}^{\text{e}} \text{e}^0$ is the positron whereas ${}_{-1}^{\text{e}} \text{e}^0$ is referred to as electron.

The study of the radioactivity of the artificially made atomic nuclei, used to clarify the concepts of constitution and stability of nuclei. The artificial radio nuclides will decay either by electron emission, or positron emission or orbital electron capture depends on the energy available for the disintegration. The nuclear reactions induced by protons, neutrons, neutrons and photons can also result in radioactive products. There is some correlation between the radioactivity of an artificial nuclide and the transformation by means of which it is produced. Electron emission is common for activities produced by (n, γ), (n, p), (n, α) and (d, p) reactions. Since these reactions decrease the +ve charge to mass ratio of the nucleus.



Positron emission is common for activity produced by (p, γ), (p, n), (α, n), (d, n) and (γ, n) reactions, which increase the (+ve) charge to mass ratio of the nucleus.

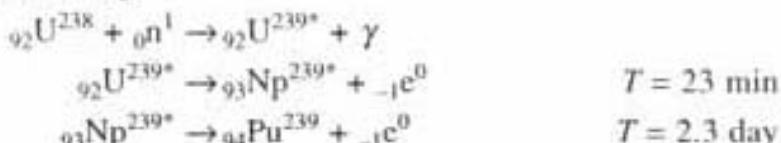


The (α, p), (p, α) and (d, α) reactions usually lead to stable products.

This discovery of artificial or induced radioactivity by Joliot and Curie, in 1934 started a new line of research and hundreds of radioactive nuclides have now been made by various methods. Say for example Iodine has only one stable species ${}_{53}^{127}\text{I}$ containing 53 protons and 74 neutrons. There are however 17 artificially radioactive isotopes possible, containing from 68 to 86 neutrons. The investigation of the radiations from the natural and artificial radio nuclides has shown that, the nucleus has energy levels analogous to the atomic energy levels. Nuclear spectroscopy, which deals with the identification and classification of these levels, is an important source of information about the structure of the nucleus. At present nearly 1200 radio nuclides have been identified.

6.14 Radio-Isotopes of Transuranium Elements

All elements which have values of Z greater than 92 are called transuranium elements because they lie beyond ${}_{92}^{238}\text{U}$ in the periodic table. The first of these, Neptunium was detected when a beam of slow neutron is incident on ${}_{92}^{238}\text{U}$ then the common radio-isotope ${}_{92}^{239}\text{U}$ is formed which gives β^- emission to produce neptunium (${}_{93}^{239}\text{Np}$) ($Z = 93$). Neptunium emits β^- particles spontaneously and changes into Plutonium (${}_{94}^{239}\text{Pu}$), the 94th element. In this way new members of the periodic table are increasing.



6.15 Uses of Radio-Isotopes

Radio-isotopes are used in different fields. The important applications of radio-isotopes depends on the fact that the chemical properties of the radio-isotope of a given element are essentially identical. But a radio-isotope can be detected easily by its radioactivity and stable isotope by a mass spectrometer. Radio-isotopes are used in different fields, as follows:

6.15.1 In Biological Field

- When a plant grows, it absorbs Phosphorus both from the soil and from fertilizer. In order to know, what is the effect of fertilizer on the plant, it is important to know what proportion of Phosphorus comes from each source. If radio-isotope is used in fertilizer the exact proportion will come out easily.

2. Similarly, for blood transfusion it is helpful to detect whether the given blood is suitable for the patient or not.

6.15.2 In Medical World

1. Radio isotopes have been used to locate and detect the presence of tumours, particularly brain tumours which are difficult to detect.

2. Radio sodium has been used to study cases of restricted circulation of blood.

3. Pumping action of the heart has been studied by using radio-sodium or radio-iodine.

6.15.3 In Industrial Field

Radio isotopes are used in industrial field.

1. Alloys are frequently subjected to different treatments such as annealing, quenching etc. Radio-isotopes have been used to find the effect of such treatment.

2. The self diffusion in metal, i.e. movement of atoms in metal within the crystal lattice has been studied by radio-isotopes.

3. Uniformity of mixing during blending of different materials, has been achieved by labelling one of the constituents with radioactive tracer.

6.15.4 Carbon Dating in Geological Field

Carbon has three important isotopes 12, 13, 14 having relative abundances of 98.89%, 1.11% and zero, respectively. Of these C¹⁴ is unstable, it disintegrates by β^- emission as



The basis assumption of the carbon dating method is that the relative abundance of carbon isotopes has remained unchanged for last few thousand years. When a living organism dies, its intake of carbon ceases and the amount of radioactive C¹⁴ in it decreases continuously. A determination of its present C¹⁴ content per gram of substance, together with a knowledge of its half life, yields the age of the sample. The concentration of C¹⁴ in all living plants is the same due to the fact that the CO₂ taken from the atmosphere contains a constant quantity of C¹⁴. But when a living plant dies it no longer takes in CO₂ and starts to decay by β^- emission with half life ($T_{1/2}$) of the value 5730 ± 30 yrs. By that way one can calculate the time of its death, as follows

We know

$$N = N_0 e^{-\lambda t}$$

$$\lambda = \text{decay const.} = \frac{0.693}{T_{1/2}}, \quad T_{1/2} = 5730 \pm 30 \text{ yrs.}$$

where N = present activity counts of C¹⁴ of the old specimen

N_0 = activity at the time of death

t = time passed after death

N_0 , the original count number at the time of death, is the same as present count rate of the β^- emission from any living matter. Hence, t can be determined.

Whenever other reliable methods for determining the age of such samples were available, the two methods give consistent result. Ages of such samples extend upto 5000 years.

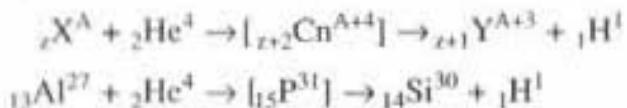
6.16 Nuclear Reactions

Depending upon the nature of projectile particle and outgoing particle, the various nuclear reactions are classified as follows:

(a) Transmutation by α -Particle

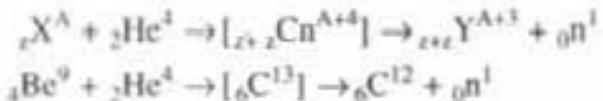
1. ($\alpha - p$) reaction

The general ($\alpha - p$) reaction is



2. ($\alpha - n$) reaction

The general ($\alpha - n$) reaction is

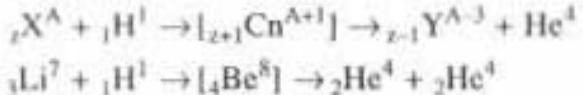


This ($\alpha - n$) reaction is well known for the detection of a new particle neutron (${}_0 n^1$) with large mass and no charge, done by Chadwick in 1932.

(b) Transmutation by Proton

1. ($p - \alpha$) reaction

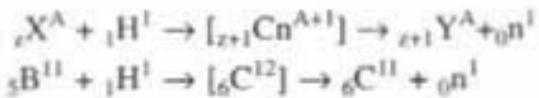
The general (p, α) reaction is



This ($p - \alpha$) reaction provided one of the earliest quantitative proof's of the validity of Einstein's mass energy relationship.

2. ($p - n$) reaction

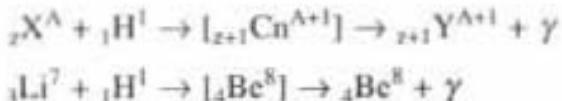
The general type is



In ($p - n$) reactions, the mass change is usually (-ve), so the reactions are endothermic.

3. ($p - \gamma$) reaction

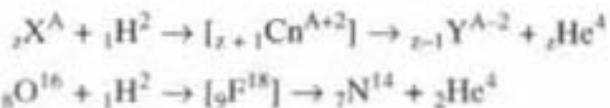
The general type is



(c) Transmutation by Deutrons

1. ($d - \alpha$) reaction

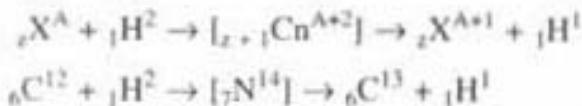
The general type is



The mass change is usually in ($d - \alpha$) reaction (+ve) so the reactions are in general exothermic.

2. ($d - p$) reaction

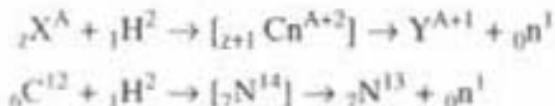
The general type is



Here in ($d - p$) reaction isotopes will form. The Q value for these reactions is usually (+ve) so the reactions are in general exothermic.

3. ($d - n$) reaction

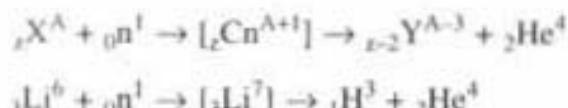
The general type is



(d) Transmutation by Neutron

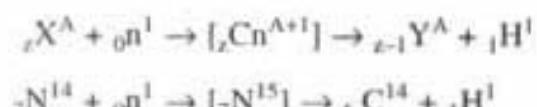
1. ($n - \alpha$) reaction

The general type is



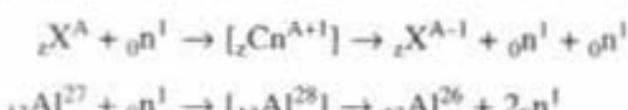
2. ($n - p$) reaction

The general type is



3. ($n - 2n$) reaction

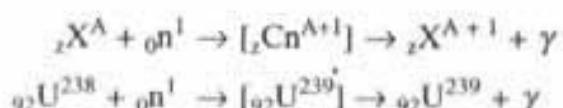
The general type is



The mass change in ($n - 2n$) reaction is always (-ve) so the reaction is in general endothermic.

4. ($n - \gamma$) reaction

The general type is



It is the commonest process which results from neutron captured in radioactive capture. Compound nucleus thus formed emits one or more γ -rays and final nucleus is an isotope of the target nucleus. The Q value is always (+ve). Thus the $(n - \gamma)$ reaction is exothermic with excess energy being carried away by γ -rays.

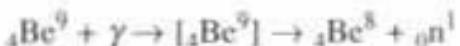
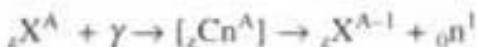


The radioactive capture of slow neutron often results in product nuclei, which are also radioactive, so this reaction is one of the most important source of artificial radioactive nuclides.

(e) Transmutation by Photon

1. $(\gamma - n)$ reaction

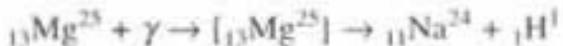
The general type is



Q value is (-ve) so the reaction is endothermic.

2. $(\gamma - p)$ reaction

The general type is



It has been seen that same nuclides can be transmuted in a number of different ways, depending upon the particle used to produce transmutation and upon the particle given off during the reaction.

6.17 Theory of Nuclear Fission⁷

The integration process in which heavy nucleus after capturing a neutron, splits up into two lighter nuclei of nearly equal masses is called fission of the nucleus or *nuclear fission* for example



When this fission occurs, a large amount of energy is released with the emission of neutrons, which produce further fission. This energy is known as nuclear energy or fission energy.

Some of the properties of nuclear forces (saturation, short range) are analogous to the properties of the forces which hold a liquid drop together. Hence a nucleus may be considered to be analogous to a drop of incompressible fluid of very high density ($\approx 10^{14}$ g/cc). This idea has been used in 1939 by Bohr and Wheeler, together with other classical idea such as electrostatic repulsion and surface tension, to explain *fission process with the help of liquid drop model*.

In the spherical liquid drop nucleus, the shape of the drop depends on a

balance involving the surface tension forces and the coulomb repulsive forces. If energy is added to the drop, as in the form of the excitation energy resulting from the capture of a slow neutron, oscillations are set up within the drop. These tend to distort the spherical shape, so that the drop may become ellipsoidal in shape. The surface tension forces tend to make the drop return to its original shape. While the excitation energy tends to distort the shape still further. If the excitation energy is sufficiently large, the drop may attain the shape of a dumb bell. The coulomb repulsive forces may then push the two "bells" apart until they split into two small drops. Heavier nuclei (large A) have larger volume and they are less stable, slightest provocation from the outside (Fig. 6.7) cause ultimate disruption of the nucleus into two or more fragments which fly apart at high speeds. This is known as *fission* and a *large amount of energy is released in fission process*. The product nuclides must be in highly excited states at the instant of fission. Some of this energy is released almost immediately and carried away by the prompt neutrons. The remaining excess energy will be released by neutrons, or α -particles or β -decays or γ -ray emission.

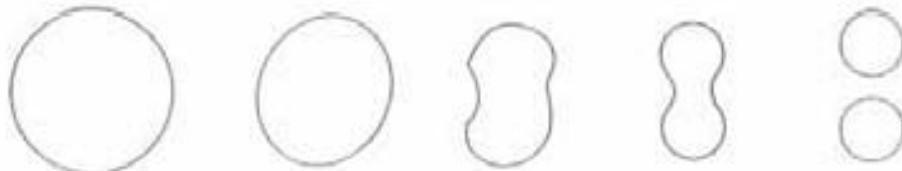


Fig. 6.7 Liquid drop model for nuclear fission

However if the excitation energy is not large enough the ellipsoid will return to the spherical shape. In that case excitation energy is given out in the form of γ -rays and the process becomes one of the radioactive capture process rather than fission process.

The value of critical deformation energy E_{critical} was first calculated by Bohr and Wheeler on the liquid drop model and from that they derived the condition for spontaneous process as $Z^2/A \geq 45$. This shows U^{238} requires more excitation energy than U^{235} to initiate fission.

It is now known that fission can be produced for various heavy nuclei under different conditions. Uranium nuclei, when bombarded with neutrons undergoes fission and it splits into two nuclei of intermediate atomic weight. It was found that slow neutron cause fission of U^{235} but not of U^{238} , fast neutrons with energies greater than one MeV, cause fission of both U^{235} and U^{238} . Thorium and Pa^{231} undergo fission but only with fast neutrons. Some heavy nuclei have been found to undergo "Spontaneous fission". In this process, the nucleus divides automatically in the ground state without bombardment by particles from outside.

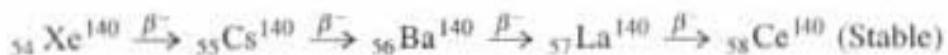
Fission Products

Due to fission the heavy nuclei can be split up in many different ways. The Uranium nucleus is capable of splitting in over 40 different ways. About 97% of the U^{235} nuclei undergoing fission yield products which falls into two groups, a "light" group ($A = 85$ to 104) and a 'heavy' group ($A = 130$ to 149), which is clear from following reactions.



After fission, the fission products are in a highly excited state and have too many neutrons for stability. Such an unstable nucleus might be expected to approach stability by: either (1) electron emission or (2) if the excitation energy is high enough, by ejection of one or more neutron; it has been found experimentally that both of these processes occur.

1. Fission fragments which decay by electron emission. They start a short radioactive series involving the successive emission of electrons until they reach to a static stable product. These series are called fission decay series. In above equation Mo^{95} and La^{139} are the stable products. An example of fission series of one fission product is as follows



2. Neutron emission in fission: It is thought that in fission, compound nucleus (U^{236}), first split into two fragments, each of which has too many neutrons for stability, and also excess energy (6 MeV or more) needed to expel a neutron. The excited, unstable nucleus consequently ejects one or more neutrons with γ -rays within a very short time after its formation. This idea is consistent with experimental result.

Energy Release in Fission

One of the most striking properties of the fission process is the magnitude of the energy released per fission, which is about 200 MeV as compared with several MeV for other nuclear reaction. An estimate of the amount of energy released per fission can be made by considering the binding energy curve (Fig. 6.6). The value of the average binding energy per particle has a broad maximum of about 8.4 MeV in the range of mass numbers from 80 to 150 and it has been shown that nearly all of the fission products has mass numbers in this range. The average B.E. per particle is about 7.5 MeV in the neighbourhood of uranium. Hence average B.E. per particle is about 0.9 MeV greater in the fission products, than in the compound nucleus (U^{236}) and the excess, 0.9 MeV, is liberated in the fission process. The total amount of energy released per fission should be roughly equal to the product of the number of particles (236) multiplied by the excess B.E./particle (0.9 MeV) i.e.

$$236 \times 0.9 = 200 \text{ MeV.}$$

The total energy released per fission can also be calculated from the nuclear masses of (U^{236}) and a typical pair of fission products in the following way. Considering the fission equation



$$\text{Atomic mass of } \text{U}^{235} = 235.124 \text{ a.m.u.}$$

$$\text{Atomic mass of } {}_0\text{n}^1 = 1.009 \text{ a.m.u.}$$

$$\text{Total mass of reactants} = 236.133 \text{ a.m.u.}$$

Atomic mass of La ¹³⁹	= 138.955 a.m.u.
Atomic mass of Mo ⁹⁵	= 94.946 a.m.u.
Atomic mass of 2 ₀ n ¹	= 2.018 a.m.u.

Total mass of products = 235.919 a.m.u.

From above two calculations we get "Q-value" and amount of energy released per fission, can be calculated from equation (6.10) as

$$\begin{aligned} Q &= (\text{Mass of the reactants} - \text{Mass of the products}) \times c^2 \\ &= (236.133 - 235.919) \times 931 \text{ MeV} \end{aligned}$$

Since Q is positive so the reaction is exothermic and the amount of energy released per fission is

$$E = 0.214 \times 931 = 198 \text{ MeV} \approx 200 \text{ MeV}$$

where

$$1 \text{ a.m.u.} = 931 \text{ MeV}$$

Although there are at least 30 different ways in which the compound nucleus (U^{236}) can divide, but the mass excess is approximately the same for all these processes and 200 MeV is the average amount of energy released per fission, as calculated in this way which is in good agreement with experimental value.

6.18 Nuclear Fission as a Source of Energy

The large amount of energy released in fission, together with the emission of more than one neutron, has made it possible to use the fission process as a source of energy. The emission, on the average, of 2.5 neutrons in the fission of U^{235} nucleus permits a chain reaction in which these neutrons produce more fissions and more neutrons and so on. Under some conditions, the numbers of fission and neutrons increases exponentially with time because each fission produces more neutrons than the one absorbed, and the amount of energy released can become enormous. The time interval between successive generation of fissions can be very small, fraction of a second and the energy released in the chain reaction can take the form of an explosive; the result is an "*atom bomb*". This is called *uncontrolled fission*. Under other conditions the chain reaction can be controlled, and a steady state can be attained in which just as many neutrons are produced per unit time as are used up. This is called "*controlled fission*". The rate at which fission occur and energy is released is kept constant, and the result is a chain reaction pile, or *Nuclear Reactor*, which can be used as a source of neutrons or of nuclear power.

Nuclear energy: The amount of energy released by the complete fission of 1 kg of Uranium can be calculated in the following way:

$$\text{Energy released/Fission} = 200 \text{ Mev}$$

So total energy released per fission = $E = 200 \text{ Mev}$

$$= 200 \times 1.6 \times 10^{-19} \times 10^6 = 3.2 \times 10^{-11} \text{ Joules or Watt sec.}$$

Now Avogard's Number = $N = 6.02 \times 10^{26}$ per kg Atom

Thus energy released by the fission of 1 kg Atom of U^{235} or 235 kg of U^{235} is

$$N \times E = 6.02 \times 10^{26} \times 3.2 \times 10^{-11} = 1.93 \times 10^{16} \text{ Joules}$$

Hence amount of energy released from 1 kg ($= m$) of U^{235} is

$$m N E/A = 1 \times 1.93 \times 10^{16}/235 = 8.2 \times 10^{13} \text{ Joules or watt sec}$$

$$\text{So Total Power output per day} = P = \frac{m N}{A} \frac{E}{T}$$

$$= \frac{1 \times 6.02 \times 10^{26} \times 3.2 \times 10^{-11}}{235 \times 24 \times 60 \times 60} = 9.48 \times 10^8$$

where A = Atomic mass of $\text{U}^{235} = 235$

i.e. $P = \frac{m N}{A} \frac{E}{T} = 9.48 \times 10^8 \text{ Watt} = 1000 \text{ M watt (approx)}$ (6.11)

Hence 1 kg U^{235} by complete fission can supply approx. 1000 M watt of power/day.

Electricity Equivalence: If the power output by fission is changed to electricity, at a 30% conversion efficiency ($= \eta$).

$$\text{Then electric power generated from 1 kg of } \text{U}^{235} = P = \eta \frac{m N}{A} \frac{E}{T} \quad (6.12)$$

i.e. Electric power generated from 1 kg of U^{235} with 30% efficiency $= P = 1000 \times \eta = 1000 \times 0.3 = 300 \text{ M watt/Day}$

This is equivalent to the thermal power plant which consumes about 300 tonnes of coal per day, that is the advantage of nuclear power over thermal power.

Nuclear Power

As discussed above, the fission of one U^{235} nucleus releases energy $3.2 \times 10^{-11} \text{ J}$ or Watt.sec. Hence, fission rate of U^{235} for producing one watt of electric power $= 1/3.2 \times 10^{-11} = 3.1 \times 10^{10}$ fission/second.

Hence time required for one fission $= 1/3.1 \times 10^{10} = 3.2 \times 10^{-9} \text{ sec.}$

Since this fission rate is very large and large amount of fission energy is released in an extremely short interval of time ($= 10^{-9} \text{ sec}$), so its production leads to a very powerful explosion unless its rate of production is controlled.

6.19 The Controlled Chain Reaction System

Controlled fission chain reaction are essential for nuclear reactor. The achievement of a chain reaction with uranium depends on a favourable balance among four competing processes.

1. Fission of uranium nuclei, with the emission of more neutrons than are captured.

2. Non-fission capture of neutrons by uranium.

3. Non-fission capture of neutrons by other materials.

4. Escape of neutrons without being captured.

If the loss of neutrons by the last three processes is less than or equal to the

surplus neutron produced by the first, a chain reaction occurs; otherwise it does not.

One of the most important factor in making favourable condition for neutron balance is the size of the sample. If the size of the sample is less, then neutron will go out of the sample without producing fission, on the other hand if size is too big then production rate or neutron will be more than escape, so explosion takes place. Thus, there is a certain size called the "critical size" of the sample uranium, which is necessary for the production of neutron in controlled chain reaction. It should be just equal to their loss by non-fission capture and escape and that helps to maintain a chain reaction. The calculation for the critical size of the sample is one of the main problems in reactor design theory.

6.20 Basic Idea for Nuclear Reactor⁷

A nuclear reactor is a device in which nuclear fission can take place at a controlled rate.

A nuclear reactor contains an amount of fissionable natural uranium called fuel, large enough, so that the loss of neutron by leakage may be neglected. Natural uranium contains 98.28% abundance of U^{238} and 0.72% abundance of U^{235} . If some spontaneous fission occurs, it will release about 2.5 fast neutrons with average energy 2 Mev. The *fission cross-section* (probability) of two isotopes of uranium U^{238} and U^{235} for these fast neutrons are almost same. But since abundance of U^{238} is very large, so the fission neutrons cause further fissions which will mainly occur in U^{238} . These neutrons can suffer (i) fission, (ii) radiative capture and (iii) elastic or inelastic scattering. Since the last two processes are more probable so chain fission reaction is not possible in natural Uranium. $_{94}Pu^{239}$ will be produced in a non-fission capture of neutrons by U^{238} as shown in Fig. 6.8.

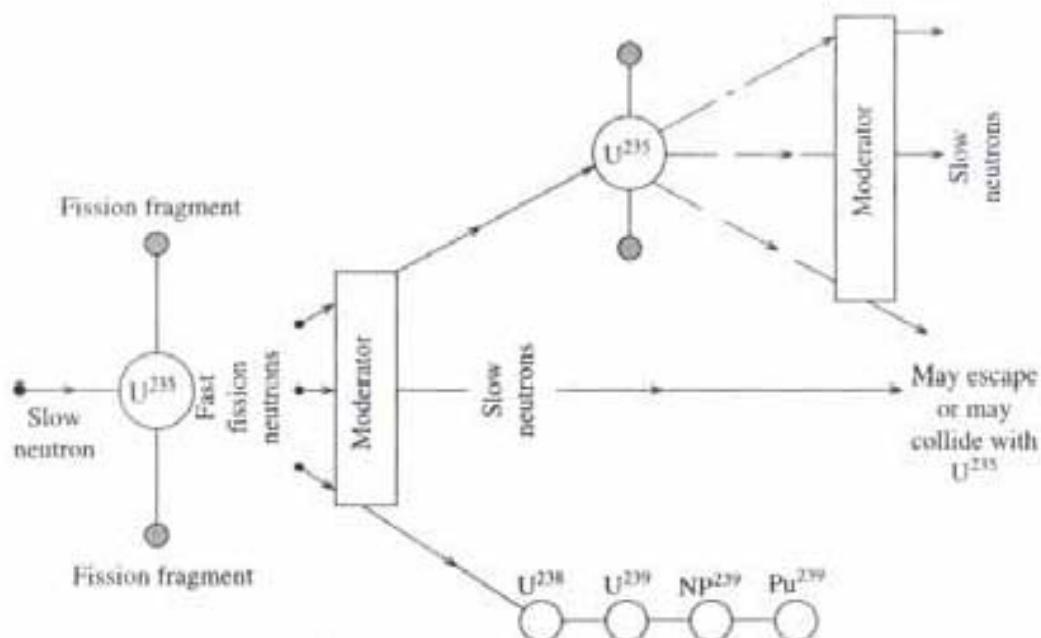


Fig. 6.8 Different steps occurring in fission process

A chain reaction in natural uranium can be made possible by using moderator (i.e. neutron slowing material) along with natural uranium. The function of the moderator is to slow down the fast fission neutrons by collisions (i.e. to thermalise them).

The fission cross-section of U^{238} for slow or thermal neutrons is practically zero whereas that of U^{235} is very large. Hence a self propagating chain reaction becomes possible.

Heavy water (D_2O) and graphite have been used successfully as moderators with natural uranium because they are light, have low atomic number, they considerably reduce the speed of neutrons at each collision and they have low cross-section for neutron absorption. When D_2O is the moderator, uranium can be dissolved in it. This type of reactor is known as "*homogeneous reactor*". When graphite is used as a moderator, uranium is in the form of rods which are distributed in a regular way throughout the graphite thereby forming a lattice. This kind of reactor in which fuel is uranium and moderator graphite are separated is called "*heterogeneous reactor*".

6.21 Critical Size for Nuclear Reactor

The need for a favourable neutron balance sets certain conditions on the system where chain reaction is occurring. The most important factor is the size. If Uranium is distributed in a regular way throughout the assembly, the neutron production depends on the volume of the system (i.e. proportional to r^3), whereas the probability of escape of neutron, depends on surface area (i.e. proportional

to r^2). Now if the system is very small, then the surface to volume ratio $\left(= \frac{1}{r} \right)$ is large, that means the rate of non-fission capture and escape of neutron is more than production rate of neutron, in that case chain reaction cannot sustain. Hence it depends on the size of the system. For larger size, greater will be its radius, so then escape rate will be smaller as compared to production rate, and chain reaction will sustain. But if the size is very big then production rate of neutron will be much bigger than escape or absorption rate, which will result in an explosion. Thus for effective nuclear reactor we need a critical size of the system for which production of neutron by fission will be just equal to their loss by non-fission capture and escape.

6.22 Essential Parts of the Nuclear Reactor

Nuclear reactor is a device where the nuclear fission is occurring under a self sustaining controlled chain reaction and is used as a high energy power supply. The essential parts of the nuclear reactor are:

(1) *Reactor Core*: It is the main part of the reactor where fissionable material, called reactor fuel like U^{235} , U^{238} , Pu^{239} are used. Due to nuclear reaction, huge amount of heat is generated. Reactor core generally has a shape of right circular cylinder with a diameter of few meters. Generally fuel elements are made up of rods of uranium. In general, reactor core has fuel elements, moderator, control rods and cooling material, housed in a pressurised vessel (Fig. 6.9).

(2) *Reactor Moderator*: The function of moderator is to slow down the fast

neutron. It should have high boiling point, large scattering cross-section, small absorption cross-section and low atomic number. Commonly used moderators are heavy water (D_2O), graphite etc.

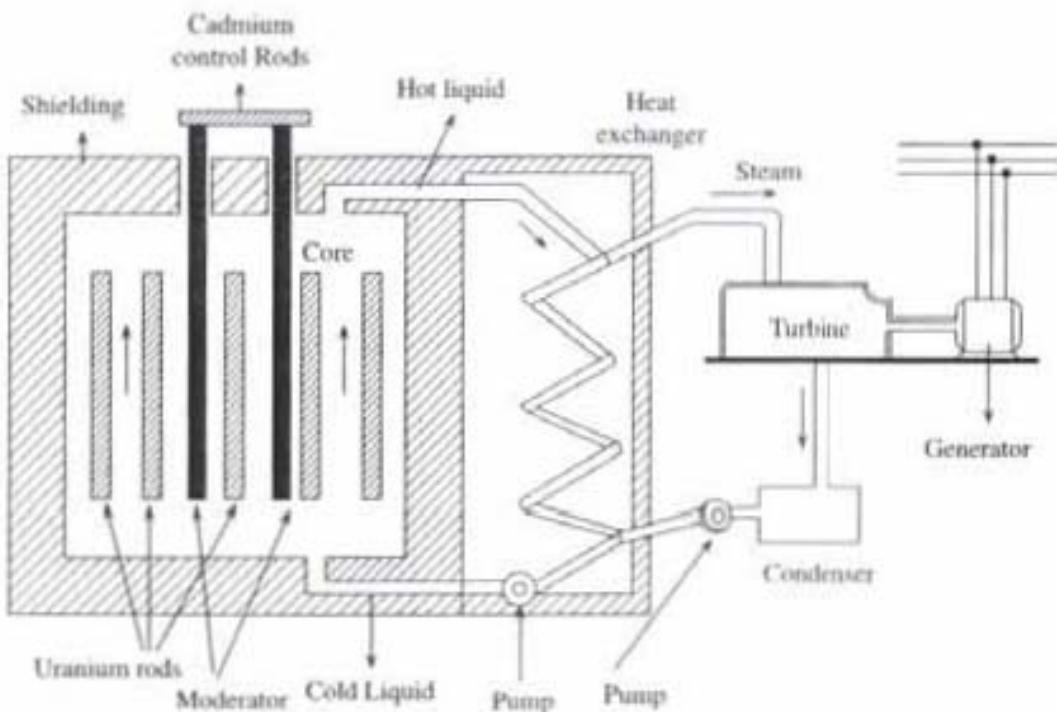


Fig. 6.9 Nuclear reactor

(3) *Reactor Control*: The control of the thermal reactor can be explained by a quantity K_e , known as effective multiplication factor, which controls the condition of reaction at steady state. The multiplicative factor K_e tells how the neutron concentration would increase from generation to generation, if no neutron leaked out from the system. In order to get a controlled chain reaction in a reactor, multiplication factor K_e should be combined in some way with the quantity which tells something about neutron leakage, and that is called Effective Multiplication Factor (K_e), which is defined as

$$K_e = \frac{P}{A + L}$$

where P = production rate of neutrons

A = absorption rate of neutrons

L = leakage rate of neutrons

$K_e = 1$ steady state, critical condition

$K_e > 1$ super critical

$K_e < 1$ sub critical

Now, as discussed earlier in Sec. 6.21, production rate is proportional to (r^3) the volume of the reacting system, whereas the leakage or escape rate is proportional to r^2 , the surface area of the reacting system. Therefore

$$K_e \propto \frac{r^3}{r^2} \propto r$$

where r is radius of the reacting system. Thus effective neutron multiplication factor (K_e) depends on r , i.e. the size of the nuclear reacting system. Now to make effective multiplication factor $K_e = 1$, for maintaining controlled chain reaction in a nuclear reactor, we need a critical size of the system for which production of neutron by fission will be just balanced by the non-fission capture and escape.

Now

Production rate of neutrons P = Rate of fission process $F \times$ Average number of neutron emission per fission N

i.e.

$$P = NF$$

Hence,

$$K_e = \frac{NF}{A + L} = \frac{N(F/A)}{1 + (L/A)}$$

The ratio F/A depends upon the amount of fissionable and non-fissionable material and on their cross-sections of fission and neutron capture.

In a nuclear reactor, fission rate can be controlled by control rods, generally made up of Cadmium or Boron steel. By adjusting the position of the control rods in the core (Fig. 6.9), rate of absorption of excess neutrons can be adjusted i.e. F/A can be controlled and thus we can maintain the value of $K_e = 1$, to get controlled fission. If $K_e < 1$, fresh fissionable material should be added.

(4) *Reactor Reflector:* The leakage of neutrons, from the reactor can be reduced by surrounding the reactor with a reflector, made of weak neutron absorber; graphite is often used for this purpose. The reflector knocks the neutrons which leave the reactor back to it, and the decreased leakage, reduces the critical size of the reactor, with a resulting saving of fuel. The reactor system including reflector, must be enclosed in a shield, usually of concrete, to reduce the intensity of the reduction of neutrons and γ -rays leaving the reactor.

(5) *Reactor Coolant:* The material used to remove heat produced by fission as fast as it is liberated, is known as *reactor coolant*. The coolant is generally pumped through the reactor in the form of liquid or gas. It is circulated throughout the reactor so as to maintain a uniform temperature. The materials proposed for coolant can be (1) ordinary or heavy water, (2) liquid metals, (3) organic liquids (hydro-carbons), (4) gases. Both ordinary water and heavy water are good coolants, as they can serve both as moderator and coolant, but due to their low boiling point, they should be pressurised. The coolant should have the following properties:

1. High B.P., (2) High specific heat, (3) Easy to pump, (4) Cheap, (5) Chemically stable.

When the energy generated in a reactor is to be converted into electric power, it is transferred from the coolant to a working liquid, to produce steam or a hot gas. The regulating vapour or the gas can be used to generate power by means of turbines.

6.23 Thermal Power Reactor

Power reactors are those which are primarily intended for the production of usable power. They transform the heat which is produced from the energy released

in fission into electricity. The heat is transferred to a fluid passing through the core of the reactor. The hot fluid is then passed through a heat exchanger where energy is transferred and utilized by more or less conventional means.

The basic principle of one type of power reactor is shown in Fig. 6.9. A quantity of enriched uranium in the form of a pure metal, forms the centre of the heat energy source. The fission rate and hence the rate of production of heat energy is controlled by keeping the value of $K_e = 1$, which is done by the movement of cadmium rods. For decreasing the temperature, the cadmium rods are pushed little further to absorb more neutrons thereby reducing the fission rate (F/A). For raising the temperature, they are pulled out control rods a little further. When the reactor becomes subcritical fresh fissile material is added.

Liquid sodium is used as a primary coolant because not only is it a dense material with good thermal conductivity but it also has very low absorption cross-section for thermal neutrons. Sodium is usually alloyed with potassium to lower the melting temperature. The energy absorbed by this primary coolant is transferred to water in a heat exchanger. The steam that is generated drives a turbine which is coupled to an electric generator. The electric power so produced can be used for lighting cities or running factories or for driving ships and submarines or large planes.

6.24 Breeder Reactor⁷

The possibility of the development of a power industry, based on nuclear fission, raises the question of the availability and cost of fissile material. Generally, the fissile materials considered have been U^{235} , Pu^{239} , and U^{233} , none of which are available naturally in large amounts. It would be inefficient to get power on a large scale from the fission of U^{235} because only 0.7% or less of the available uranium could be used directly. Plutonium (Pu^{239}) has been made by the fission of U^{235} in a reactor containing U^{238} , but this process is expensive. Similarly, U^{233} , which does not occur in nature, can be made in a reactor containing Th^{232} . Although the problem of availability and cost of nuclear fuels is a serious one, a unique and remarkable solution is possible because of one of the nuclear properties of the fission process. This property, is the emission on an average, of more than two neutrons per fission.

Suppose, for the sake of example, that 3 neutrons are emitted per fission. One of these is needed to keep the chain reaction going on to induce fission in another fuel atom, say an atom of U^{235} . Of the two neutrons that are left, one may be used to convert an atom of U^{238} to one of plutonium, leaving one last neutron. If this last neutron could be used to produce another atom of plutonium (Pu^{239}) from another atom of U^{238} , then one atom of U^{235} would have been used and two atoms of Pu^{239} would be made, leaving a profit of one atom of fissile material. In other words, *more fissile material would be produced than consumed*. It is inevitable that some neutrons should be lost by leakage or by absorption, but so long as more than two neutrons are produced in fission there is a possibility that more nuclear fuel may be produced than is consumed. This process is called "*Breeding*", if it could be made to work, a reactor could make new fuel for itself and, in addition, a stockpile of fissile material could be built up for use in new reactors.

In the process of breeding, fissionable material depends on two materials, one of which is "fissionable" and the other which is 'fertile'. Uranium-238 and Pu²³⁹ form such a pair, and Th²³² and U²³³ form another pair. Consider the first pair and suppose that enough Pu²³⁹ is available to achieve a chain reaction in a system which can be used to produce power. The thermal fission of Pu²³⁹ yields 3 neutrons, so that if some of these neutrons could be absorbed in U²³⁸ to form more Pu²³⁹, a breeding cycle might be achieved. In this case, *U²³⁸ is the fertile material because although it is not itself usefully fissionable, but it can be converted into a good fissionable material*, and it is conceivable that a reactor system could be built in which more plutonium is produced than is consumed. The nuclear property that determines whether the possibility of breeding exists, is not the value of the number of neutrons emitted per fission (v) but the number of neutron produced per fission absorbed in the fuel (η), because the non-fission capture of neutrons by the fuel can be large enough to interfere seriously with breeding. The value of η for plutonium and thermal neutrons is 2.08 ± 0.02 and turns out to be only slightly greater than two. This result seems to make impractical the breeding of Pu²³⁹ in a thermal reactor. It is expected, on theoretical grounds, that the ratio of the fission cross-section to the total absorption cross-section should be closer to unity at high neutron energies than for slow neutrons, and that η should be close to v in a fast reactor. Consequently, breeding of Pu²³⁹ should be possible in a fast neutron, Pu²³⁹-U²³⁸ system. A reactor, the Experimental Breeder Reactor, or EBR, has been built to study the feasibility of power production and breeding in such a system.

The value of η for U²³³ and thermal neutrons is 2.28 ± 0.02 , which makes the breeding of U²³³ from Th²³² in thermal reactor system feasible. Research and development work are in progress on such systems. One design involves the use of a solution of U²³³ O₂SO₄ in D₂O; another involves the use of a solution of U²³³ Bi in liquid bismuth, with graphite as moderator. The U²³³-Th²³² system could also be used in a fast reactor, but a greater value of v for Pu²³⁹ favours the use of the latter material in fast breeder reactors.

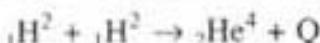
If breeding could be achieved, U²³⁸ and Th²³² could be converted into fissionable materials. The resulting increase in available nuclear fuels would be a step in the direction of economic nuclear power. Much information is needed concerning the cross sections for reactions of fast and intermediate energy neutrons with fissionable and fertile materials, and much work is needed to be done in this field.

6.25 Thermo-Nuclear Reactions or Nuclear Fusion

Nuclear fusion is the process of combining or fusing two lighter nuclei into a stable and heavier nuclide. When this process occurs among the light elements, energy is usually released because the mass of the product nucleus is less than the sum of the masses of the nuclei which are fused. This property of the light elements is shown in the binding-energy curve (Fig. 6.6) in which B.E./A rises rapidly with mass number in the region of small mass numbers. Under appropriate conditions, fusion reactions can liberate vast amount of energy, amounts much greater than that released in an *atom bomb explosion due to fission process*. Fusion reaction takes place at very high temperature (10^9°C), so it is also called

thermonuclear reaction. This *thermonuclear reaction forms the basis of the so called thermonuclear explosion or hydrogen bomb explosion due to fusion process.*

For example when two deuterons are brought together to form an α -particle the following reaction takes place with the liberation of 24 MeV energy.



For fusion to take place, the component nuclei must be brought to within 10^{-12} cm. In order to approach so closely, they should be imparted high energies so that two positively charged deuterons may overcome the repulsive force and fused together. The only practical way of fusing the nuclei together is to raise their temperature to a excessively high value of the order of 10^{9°C . *Fusion reaction can take place at stellar temperature i.e. on sun, not on earth.*

6.26 Stellar Thermonuclear Reaction

The amount of energy radiated by sun is $= 10^{26}$ J/sec. Its age is $= 5 \times 10^9$ yrs. Hydrogen and helium together form 90% by weight of sun's matter and the temperature at the centre of the sun is $3 \times 10^7^\circ\text{C}$. Fusion reaction is therefore continuously taking place in sun which is the source of this stellar energy.

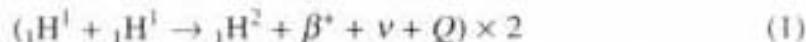
Two sets of thermonuclear reactions have been proposed as sources of energy in the sun.

- (1) Proton-proton chain
- (2) Carbon-nitrogen chain

Here we discuss proton-proton chain in detail.

6.26.1 Proton-Proton Chain

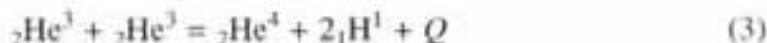
Bethe in 1939 suggested that in proton-proton chain, proton is continuously transformed into helium. In the first step two protons fuse together to form a deuteron, releasing a positron and energy.



The deuteron then combines with another proton and forms helium (He^3) isotopes.

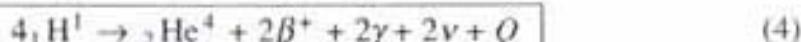


Two ${}_2\text{He}^3$ then fuse together to produce ${}_2\text{He}^4$ nuclei



In Eqn. (1) ν is the "neutrino" produced in β -decay, an electrically neutral particle having a very small mass, and compared to electron, the neutrino mass may even be zero. For the third reaction to occur, each of the first two reactions must occur twice.

The overall effect of the above steps is



The above reaction (4) is exothermic and amount of energy released by this process can be calculated as

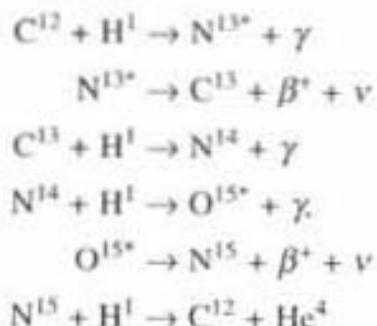
Atomic mass of 4 proton or 4 hydrogen atoms	= 4.03258 a.m.u.
Atomic mass of 1 helium atom	= 4.00387 a.m.u.
Difference in mass between reactants and products	= 0.02871 a.m.u.
Q is positive, so energy released/fusion = 26.7 MeV = 42.7×10^{-13} J	

Since one gram of the sun's matter contains about 2×10^{23} protons. Thus the available energy supply would be 21.31×10^{10} J/g. Such a vast amount of energy will be released from only one gram of sun's matter. The above reactions of the proton-proton chain occur extremely slowly which is fortunate because the sun can then have a reasonably long life.

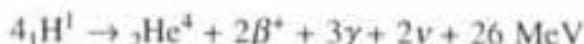
6.26.2 Carbon-Nitrogen Chain

In carbon and nitrogen chain, we found them to have the remarkable property that they could be formed into a cycle in which carbon and nitrogen nuclei are not used up, but are regenerated. These nuclei act as a catalyst in a series of reactions in which 4 protons are converted into a helium nucleus and about 26 MeV of energy are liberated per reaction.

The sequence of reaction is



The overall process will give



6.27 Comparisons Between Fission and Fusion

Fission	Fusion
1. Huge amount of energy released.	1. Huge amount of energy released.
2. A heavy nucleus U^{235} is split up into two nuclei (e.g. Mo^{95} , La^{139})	2. Two lighter nuclei are fused together (e.g. ${}_1H^2$, ${}_1H^3$).
3. The process is possible at room temperature.	3. This process is only possible at very high temp. ~ millions °K.
4. The link of this process is neutron.	4. The link in this process is proton.
5. The energy/nucleon is ~ 0.85 MeV.	5. The energy/nucleon ~ 6.75 MeV.
6. Availability of atoms used (U^{235} , U^{238}) is limited and costly.	6. Availability of atoms used (${}_1H^2$, ${}_1H^3$) is unlimited and cheap.
7. Many isotopes are formed as product particles.	7. In general isotopes are not formed.

PROBLEMS

1. The bombardment of lithium isotope ${}_3\text{Li}^7$ at rest, by protons produces two α -particles per reaction, no other particles are formed. If the energy of each α -particle produced is 9.15 MeV, calculate the energy of the bombarding proton. Given that

Mass of ${}_3\text{Li}^7$ = 7.018232 a.m.u.

Mass of α -particle = 4.003874 a.m.u.

Mass of proton = 1.008145 a.m.u.

1 a.m.u. = 931 MeV

Solution

The reaction is ${}_3\text{Li}^7 + {}_1\text{H}^1 \rightarrow {}_2\text{He}^4 + {}_2\text{He}^4$ (1)

Atomic mass of ${}_3\text{Li}^7$ = 7.018232 a.m.u.

Atomic mass of α -particle = 4.003874 a.m.u.

Atomic mass of proton = ${}_1\text{H}^1$ = 1.008145 a.m.u.

So the mass defect is

$$\begin{aligned}\Delta M &= 7.018232 + 1.008145 - (4.003874) \times 2 \\ &= 8.026377 - 8.007748 \\ \Delta M &= 0.018629 \text{ a.m.u.}\end{aligned}$$

and Q value = $0.018629 \times 931 = 17.3436 \text{ MeV}$ (2)

Now $Q = E_r + E_p - E_i$, (3)

where E_r = Recoil energy of the product nucleus.

E_r = K.E. of the product particle.

E_i = K.E. of the incident particle.

K.E. of the target nucleus is considered as zero, and in equation (1) the product nucleus and product particle is same i.e. two α -particles. So K.E. of incident proton from Eqns. (2) and (3) is

$$E_i = 2E_r - Q$$

Energy of each α -particle = 9.15 MeV = E_r

So $2E_r$ = K.E. of two α -particles = 18.30 MeV

So E_i = K.E. of incident proton = $18.30 - 17.3436$

K.E. of incident proton = 0.9564 MeV.

2. In an uncontrolled fission reaction 1 milligram of U^{235} fissions out in one microsecond. Calculate the power of explosion. Given that the energy produced per fission is 200 Mev.

Solution : According to the Eqn. 6.11, power output

$$P = \frac{m N}{A} \frac{E}{T}$$

where m = Mass of U^{235} used = 10^{-3} gm, N = Avogard's Number = 6.02×10^{23} per gram Atom,

E = Energy released per fission in Joules = $200 \times 10^6 \times 1.6 \times 10^{-19} = 3.2 \times 10^{-11}$ Joules

A = Atomic mass of U^{235} = 235, T = Time in sec. = 10^{-6} sec.

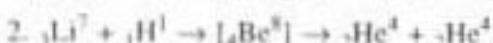
$$\text{So power of explosion} = P = \frac{10^{-3} \times 6.02 \times 10^{23} \times 3.2 \times 10^{-11}}{235 \times 10^{-6}}$$

$$= 8.197 \times 10^{13} \text{ watt}$$

3. Complete the following nuclear reaction.



Solution



4. A nuclear reactor produces 100 Mw of useful electric power. The thermal efficiency of the reactor is 40%. Assuming energy released per fission to be 200 Mev, find the mass of U^{235} in kg.consumed in one day.

Avogard's No. = 6.03×10^{26} per kg. atom.

Solution : According to the Eqn. 6.12, power output with 40% efficiency

$$P = \eta \frac{mN}{A} \frac{E}{T}$$

where η = efficiency = 0.4, m = Mass of U^{235} used = ??, N = Avogard's Number = 6.03×10^{26} per kg Atom,

E = Energy released per fission in Joules = $200 \times 10^6 \times 1.6 \times 10^{-19} = 3.2 \times 10^{-11}$ Joules, $P = 100$ M watt

A = Atomic mass of U^{235} = 235, T = Time in sec. = $24 \times 60 \times 60$ sec.

$$\text{So mass of } \text{U}^{235} \text{ used} = m = \frac{PAT}{\eta NE} = \frac{(100 \times 10^6 \times 235 \times 24 \times 60 \times 60)}{(0.4 \times 6.03 \times 10^{26} \times 3.2 \times 10^{-11})}$$

i.e. $m = 0.263 \text{ kg} = 263 \text{ gm}$

5. When a hydrogen bomb explodes, a thermonuclear reaction for the formation of ${}_2\text{He}^4$ occurs from deuterium (${}_1\text{H}^2$) and tritium (${}_1\text{H}^3$).

1. Write the equation of the nuclear reaction.
2. Find the energy per reaction liberated.

The masses are

$${}_1\text{H}^2 = 2.01474 \text{ a.m.u.}$$

$${}_1\text{H}^3 = 3.01700 \text{ a.m.u.}$$

$${}_0\text{n}^1 = 1.008986 \text{ a.m.u.}$$

$${}_2\text{He}^4 = 4.003880 \text{ a.m.u.}$$

$$1 \text{ a.m.u.} = 931 \text{ MeV}$$

Solution:

(1) The thermonuclear reaction is hydrogen bomb explosion is



$$\begin{array}{ll} \text{(2) Atomic mass of } {}_1\text{H}^2 & = 2.01474 \text{ a.m.u.} \\ \text{Atomic mass of } {}_1\text{H}^3 & = 3.01700 \text{ a.m.u.} \end{array}$$

$$\text{Total mass of reactant} = 5.03174 \text{ a.m.u.}$$

$$\begin{array}{ll} \text{Atomic mass of } {}_2\text{He}^4 & = 4.003880 \text{ a.m.u.} \\ \text{Atomic mass of } {}_0\text{n}^1 & = 1.008986 \text{ a.m.u.} \end{array}$$

$$\text{Total mass of product} = 5.012866 \text{ a.m.u.}$$

$$\begin{aligned} \text{Therefore} \quad Q \text{ value} &= 5.03174 - 5.012860 \\ &= 0.018874 = 17.572 \text{ MeV.} \end{aligned}$$

Thus, amount of energy released per reaction is 17.572 MeV.

6. Calculate Binding energy per nucleon for ${}_3\text{Li}^7$ atom.

$$\text{Mass of Li atom} = 7.01818 \text{ a.m.u.}$$

$$\text{Mass of Hydrogen } ({}_1\text{H}^1) \text{ atom} = 1.0081 \text{ a.m.u.}$$

$$\text{Mass of Neutron } ({}_0\text{n}^1) = 1.009 \text{ a.m.u.}$$

$$1 \text{ a.m.u.} = 931 \text{ MeV}$$

Solution

B.E./nucleon of ${}_3\text{Li}^7$ is

$$\begin{aligned} &= 1/7 (3 M_{\text{H}} + 4 M_n - \text{Mass of } {}_3\text{Li}^7) \\ &= 1/7 (3 \times 1.0081 + 4 \times 1.009 - 7.01818) \\ &= 1/7 \times 0.4212 = 6.0171 \times 10^{-3} \text{ a.m.u.} = 6.0171 \times 931 \times 10^{-3} \text{ MeV} \\ &\text{B.E./nucleon of } {}_3\text{Li}^7 = 5.60196 \text{ MeV.} \end{aligned}$$

7. Calculate power output of a nuclear reactor which consumes 10 kg of U^{235} per day. Energy released per fission is 200 Mev.

Solution : According to the Eqn. 6.11, power output

$$P = \frac{mN}{A} \frac{E}{T}$$

where m = Mass of U^{235} used = 10 kg, N = Avogard's Number

$$= 6.02 \times 10^{26} \text{ per kg. Atom.}$$

$$E = \text{Energy released per fission in Joules} = 200 \times 10^6 \times 1.6 \times 10^{-19}$$

$$= 3.2 \times 10^{-11} \text{ Joules}$$

$$A = \text{Atomic mass of } \text{U}^{235} = 235, T = \text{Time in sec.} = 24 \times 60 \times 60 \text{ sec.}$$

$$\text{So power output} = P = \frac{10 \times 6.02 \times 10^{26} \times 3.2 \times 10^{-11}}{235 \times 24 \times 60 \times 60}$$

$$= 9.48 \times 10^9 \text{ watt}$$

QUESTIONS

- What are radio-isotopes? How are they produced? Give a short account of industrial applications of radio-isotopes.
- Distinguish between the fusion and fission reactions. Which one of them is more economical? Why?
- A nuclear reactor produces 100 MW of electric power. The thermal efficiency is 40%. Assuming 200 MeV of energy per U²³⁵ fission, calculate the amount of natural uranium fuel (in kg) consumed per day. Natural Uranium contains 0.7% of U²³⁵.

Solution: Follow the same steps as done in problem 4. The amount of U²³⁵ required for the production of 100 MW power is 0.261 kg. Now Natural Uranium contains 0.7% of U²³⁵. Hence, the amount of natural uranium consumed/day

$$= \frac{100 \times 0.261}{0.7} = 37.286 \text{ kg.}$$

- A nuclear reactor containing U²³⁵ is operating at power level of 2 MW. Calculate the mass of U²³⁵ consumed/day by this reactor. Energy/Fission = 200 Mev. Avogard's number = $6.03 \times 10^{26}/\text{kg atom}$.

Solution: According to the Eqn. 6.11, power output

$$P = \frac{mN}{A} \frac{E}{T}$$

where m = Mass of U²³⁵ used = ??, N = Avogard's Number = 6.02×10^{26} per kg. Atom,

E = Energy released per fission in Joules = $200 \times 10^6 \times 1.6 \times 10^{-19} = 3.2 \times 10^{-11}$ Joules, P = 2 M watt, T = Time in sec = $24 \times 60 \times 60$, Atomic mass of U²³⁵ = 235

$$\text{So } \text{mass of U}^{235} \text{ used} = m = \frac{PAT}{NE}$$

$$= (2 \times 10^6 \times 235 \times 24 \times 60 \times 60) / (6.03 \times 10^{26} \times 3.2 \times 10^{-11})$$

$$\text{i.e. } m = 2.1 \times 10^{-3} \text{ kg} = 2.1 \text{ gm}$$

- Find the binding energy per nucleon of:

(i) ${}_6\text{C}^{12}$ and (ii) ${}_{92}\text{U}^{238}$

Mass of ${}_6\text{C}^{12}$ = 12.00380 a.m.u., Mass of ${}_1\text{H}^1$ = 1.00898 a.m.u.

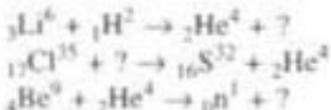
${}_{92}\text{U}^{238}$ = 238.12522 a.m.u. ${}_{1}\text{H}^1$ = 1.007599 a.m.u.

1 a.m.u. = 931 MeV.

- What is the Q value of a nuclear reaction? Calculate it for the reaction taking place in Rutherford's experiment on artificial disintegration of nitrogen by bombardment with α -particle. Relevant masses in a.m.u. are:

$$\begin{array}{ll} {}_7\text{N}^{14} = 14.007515 & {}_2\text{He}^4 = 4.003837 \\ {}_8\text{O}^{17} = 17.004533 & {}_1\text{H}^1 = 1.008142 \end{array}$$

7. Complete the following nuclear reactions:



8. The atomic mass of ${}_2\text{He}^4$ is 4.003837. (1) What is its binding energy? (2) What is its binding energy per nucleon?
9. The Binding energy of ${}_{10}\text{Ne}^{20}$ is 160.6 MeV. Find its atomic mass.
10. When ${}_{92}\text{U}^{235}$ undergoes fission, about 0.1% of the original mass is released as energy. How much energy is released when 1 kg of ${}_{92}\text{U}^{235}$ undergoes fission?
11. How much ${}_{92}\text{U}^{235}$ must undergo fission per day in a nuclear reactor that provides energy to 100 megawatt (10^8 W) electric power plant? Assume perfect efficiency.
12. When coal is burned, about 3.26×10^7 J/kg of heat is liberated. How many kg of coal would be consumed per day by a conventional coal field which produces 100 MW electric power plant? Compare the result with nuclear reactor.

Solution: out put power 100 Mw = 100×10^6 w

$$\text{Power} = \frac{\text{Energy}}{\text{Time}}$$

So total energy liberated = power \times time

$$= 10^8 \times 24 \times 60 \times 60 = 864 \times 10^{10} \text{ J}$$

$$\text{i.e. amount of coal used Per day} = \frac{864 \times 10^{10}}{3.26 \times 10^7} = 265 \times 10^3 \text{ kg}$$

i.e. 265×10^6 gm coal will produce 100 Mw power/day.

Comments: As we have seen in problem 4, that in a nuclear reactor with 100% efficiency, 105 gm of U^{235} will produce 100 Mw power/day. So in thermal power plant 2×10^6 times more coal will be required to produce same power output/day. Thus nuclear reactor is more useful for higher rate of power supply.

13. Describe the construction and working of G.M. tube.
 14. What is Q-value of a Nuclear Reaction?

Superconductivity

7.1 Introduction

As temperature decreases conductivity increases, but it was found for some materials that suddenly at a particular temperature (T_c °K) called *transition temperature*, conductivity becomes infinite, i.e., resistance suddenly falls to zero (Fig. 7.1). Such materials are called *superconductors*. But a superconductor is not merely a good conductor getting better, the mechanism that causes superconductivity is different from the mechanism that causes conductivity. Many good conductors like Ag, Au, Cu are not superconductors. There is an *interesting apparent paradox, the higher the resistivity at room temperature the more likely it is, that metal will be a superconductor when cooled.* The example of some superconductive elements are given below, it can be a metal or alloy or semiconductor. Superconductivity is also observed in some organic substances.

Table I

Metal	T_c	Alloy	T_c	Semiconductor
Tin (Sn)	3.75°K	Nb ₃ Sn	18.1°K	Si, Ge (Thin film or
Tantalum (Ta)	—	YBa ₂ Cu ₃ O ₇	90°K	under high pressure)
Niobium (Nb)	9.46°K	Ti Ba ₂ Ca ₂ Cu ₃ O ₁₀	125°K	
Molybdenum (Mo)	—			
Cd (Cadmium)	0.56°K			
Al (Aluminium)	1.19°K			
Hg (Mercury)	4.16°K			
Pb (Lead)	7.22°K			

Superconductivity results from strong coupling between conducting electrons and the lattice. For normal conduction there is a weak coupling between valence electrons and the lattice. Because resistivity arises due to electron-phonon (lattice) interaction. So a good conductor means that the electron-phonon interaction is less, i.e., coupling is weak. But on the contrary for superconductors electron-phonon interaction is high, coupling is very strong. This superconducting effect was first observed by Kamer-Lighonen in 1911 at Leiden. As the list given

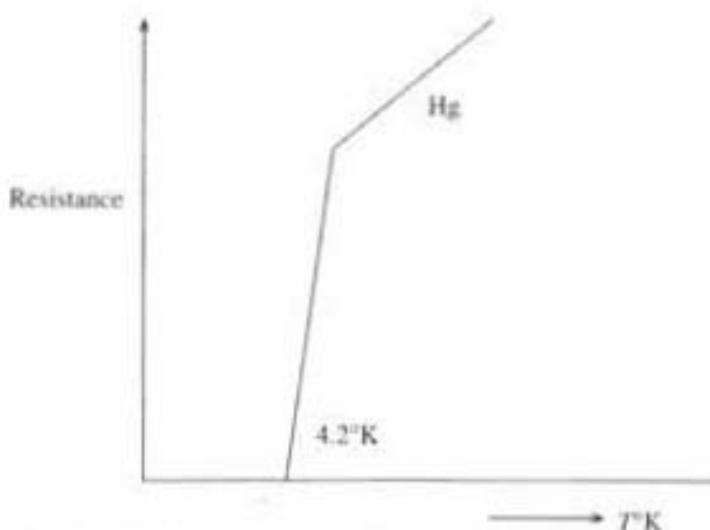


Fig. 7.1 Variation of resistance with temperature

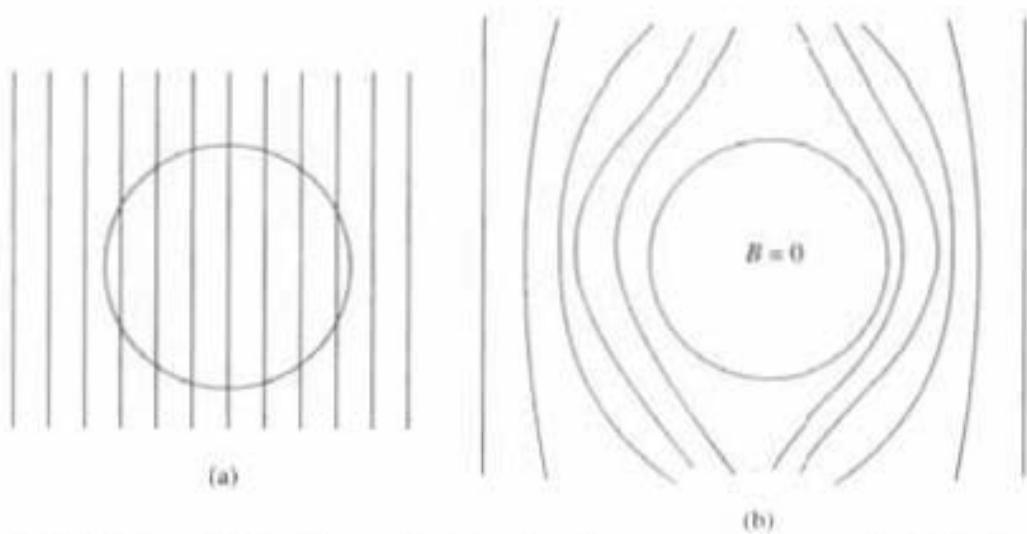


Fig. 7.2 (a) Penetration of magnetic flux through normal conductor; (b) Exclusion of magnetic flux through superconductor

above in Table 1, the temperature (T_c °K) which is called *transition temperature* is different for different metals or alloys. Many of them become superconductors under special conditions like high pressure, thin film, amorphous state etc. as mentioned in the above table. The electrical resistance in superconducting state becomes zero, strong evidence for that is found at the Massachusetts Institute of Technology (MIT), where an induced current of several hundred amperes in a superconducting lead ring showed no change in the magnitude of the current, for a period of more than one year, which is only because of its zero resistance. But in some metals superconductivity has not been observed even below 1°K say for Li, Na, K, etc. Below 1°K say at 0.08–0.09°K they are found to be still normal conductors.

7.2 Meissner Effect

It has been observed that superconductivity not only consists of zero electrical resistance, the magnetic property of the superconductors are also dramatically

change at critical temperature. It was first experimentally observed by W. Meissner at R. Ochsenfeld that when a bulk sample is cooled in a magnetic field through the transition temperature for superconductivity, the magnetic flux originally present in the specimen is pushed out of the specimen. It is also observed that it does not matter how the magnetic field is established—by an external magnet or by a current flowing in the superconductor itself. This effect is called “flux exclusion” or commonly the *Meissner effect*. Thus inside the bulk superconducting specimen the magnetic induction (\bar{B}) is zero (Fig. 7.2b). As it was discussed in Part-I Chapter-7 Eqn. (7.3) that

$$\bar{B} = \mu_0 (\bar{H} + \bar{M})$$

Thus when

$$\bar{B} = 0$$

Then

$$\mu(\bar{H} + \bar{M}) = 0$$

Or

$$\chi_m = M/H = -1$$

where \bar{H} is the applied magnetic field and \bar{M} the magnetization.

Thus, when

$$\bar{B} = 0$$

Then magnetic susceptibility

$$\boxed{\chi_m = M/H = -1} \quad (7.1)$$

Thus *magnetic susceptibility in a superconductor is found to be negative*, which is referred to as *perfect diamagnetism*, which means that *a bulk superconductor in a weak magnetic field acts as a perfect diamagnet*. This effect is known as “Meissner effect”, which is a unique magnetic property of a superconductor and of central importance for the characterization of the superconducting state in contrast to a perfect conductor, only with zero resistivity.

The four well-known Maxwell's equations under static condition, which are the basic laws of electricity and magnetism given as;

$$\vec{\nabla} \cdot \vec{E} = Q_v / \epsilon_0 \quad \text{or} \quad \vec{\nabla} \cdot \vec{D} = Q_v \quad (1)$$

($\because \epsilon_0 \vec{E} = \vec{D}$ = Electric induction)

$$\vec{\nabla} \cdot \vec{B} = 0 \quad \text{or} \quad \vec{\nabla} \cdot \vec{H} = 0 \quad (2)$$

$$\vec{\nabla} \times \vec{B} = \mu_0 \vec{J} \quad \text{or} \quad \vec{\nabla} \times \vec{H} = \vec{J} \quad (3)$$

($\because \mu_0 \vec{H} = \vec{B}$ = Magnetic induction)

$$\vec{\nabla} \times \vec{E} = \partial \vec{B} / \partial t \quad (4)$$

where \vec{E} and \vec{H} are electric field and magnetic field vectors, Q_v is the volume charge density, \vec{J} the current density, ϵ_0 the permittivity and μ_0 is the permeability of free space.

Now from Maxwell's 4th equation, we get

$$\vec{\nabla} \times \vec{E} = \text{Curl } \vec{E} = - \partial \vec{B} / \partial t$$

and from Ohm's law, we get $V = IR$ or $|\vec{E}| = \frac{V}{d} = \frac{\vec{J} \times A \times R}{d}$

$$\text{or } \vec{E} = \rho \vec{J} \left(\because \rho = \frac{R \times A}{d} = \text{Resistivity} \right) \quad (7.2)$$

From equation (7.2), we see that if resistivity ρ goes to zero for superconducting state while current density \vec{J} is finite, then electric field \vec{E} must be zero. Now from Maxwell's equation (4) we get when $\rho = 0$, \vec{E} becomes zero which gives

$$\partial \vec{B} / \partial t = 0 \quad \text{i.e.} \quad \vec{B} = \text{constant} \quad (7.3)$$

which is the case of a perfect conductor (Fig. 7.2a).

Thus the above result predicts from zero resistivity ($\rho = 0$) for perfect conductor, that the magnetic flux \vec{B} is constant not zero, it cannot change on cooling through the transition temperature.

The Meissner effect contradicts this result and suggests that *perfect diamagnetism ($\vec{B} = 0$) and zero resistivity ($\rho = 0$) are two independent essential properties of the superconducting state which is different from a hypothetical perfect conductor (where $\rho = 0$)*. Meissner effect in superconducting state can be verified by London equation, discussed latter in this chapter, in section 7.6.

7.3 Theoretical Explanation of Superconductivity: B.C.S. Theory⁸

The superconducting state is known to be an ordered state of the conduction electrons of the metal. The order is in the formation of loosely associated pair of electrons. At temperature below the transition temperature the electrons are arranged in an order, but they become disordered above the transition temperature. The nature and origin of the ordering was explained successfully by quantum mechanical treatment of superconductivity. This was first given by Barden Cooper, and Schrieffer, who were awarded the Nobel prize in physics in 1972, according to their name it is called as B.C.S. theory.

When an electron is moving through a lattice it attracts the (+ve) ion core towards it, and changes the charge density in its vicinity, i.e., in that position, the crystal has more (+ve) charge density than usual and hence this region attracts more other electrons. In this way, indirect interaction proceeds between electrons, when one electron strongly interacts with the lattice and deforms it. Thus there is an attractive force between two electrons over repulsive force through (+ve) ion core. This forms *a bound system through strong coupling between two oppositely spinned electrons* and no transfer of energy takes place from this bound system to lattice ion. Now when an external electrical field is applied to the substance in superconducting state the pairs of electrons gain additional K.E., acquire some increase in momentum and give rise to a current which continues without changing its value as there is no transfer of energy from the electron pair bound system to the lattice. *This can explain the paradoxical fact, that in a good conductor the electron-phonon (+ve ion lattice core) coupling is poor, which often makes them poor superconductors.* Whereas those metals which have higher resistivity, i.e., the more interaction between electron and phonon at room temperature, they will be more likely to be a good superconductor when cooled.

These electron pairs are called *Cooper pair* and the above theory is called

B.C.S. theory. Electrons in the Cooper pair have opposite spin, so total spin of the Cooper pair is zero and it also has many attributes of *Bosons* (spin zero particle). Single electrons are not bosons they have $\pm 1/2$ spin, they are called *Fermions*. For a superconducting element when temperature increases from very low value, then at transition temperature (T_c) the Cooper pair (Boson) break up into two electrons (Fermions) and it behaves like normal material.

7.4 Pairing Energy Gap in Superconductor

The Cooper pairs have the binding energy Δ called the pairing energy which is typically in the range of 10^{-4} to 10^{-3} eV. The critical temperature T_c for most of the superconductors is between $1 - 10^0$ K as given in Table 1 corresponding to KT_c in the same range of 10^{-4} to 10^{-3} eV (where K is the Boltzmann's constant). Thus the critical temperature is directly related to the pairing energy for a superconductor. It has value $E_g = 4KT_c$. This energy gap separates superconducting electrons, which are available below the gap E_g , from the normal electrons, of which very few are available above the gap (Fig. 7.5b). Gap energy is detected in experiments on heat capacity, infrared absorption and tunnelling. Above T_c the Cooper pairs are broken and material becomes normal conductors.

The attractive interaction between electrons is responsible for the formation of electron pair. The binding energy of the Cooper pair introduces a pairing energy gap 2Δ in conduction band between bound state and excited state. Figs. (7.3a, b) show conduction band for normal and superconductor. The energy gap in superconductor is of an entirely different nature than the energy gap in insulator. In insulator the gap is tied to the lattice, in superconductor the gap is tied to the Fermi gas.

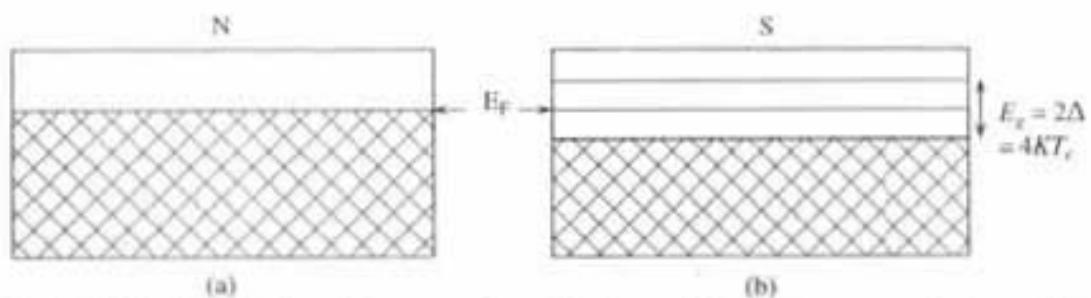


Fig. 7.3 Conduction band for normal conductor and that for superconductor with a pairing energy gap 2Δ , around $E = E_F$

It is energetically favourable for electrons near the Fermi-energy level (E_F) in a superconductor to bind together in Cooper pairs. As a result the density of states $n(E)$ decreases to zero within the interval $\pm \Delta$ of E_F , with corresponding increase in $n(E)$ just above and below (E_F) (Fig. 7.4b). Figures 7.4(a, b) show the resulting density of the states $n(E)$ for normal and superconductors with the pairing energy gap 2Δ for superconductor. Above T_c the density of states of the superconductor (Fig. 7.4b), will be like normal conductor (Fig. 7.4a). The pairing energy gap begins to open as the superconductor is cooled below T_c , the gap energy increases as the temperature decreases, reaches its maximum as T approaches

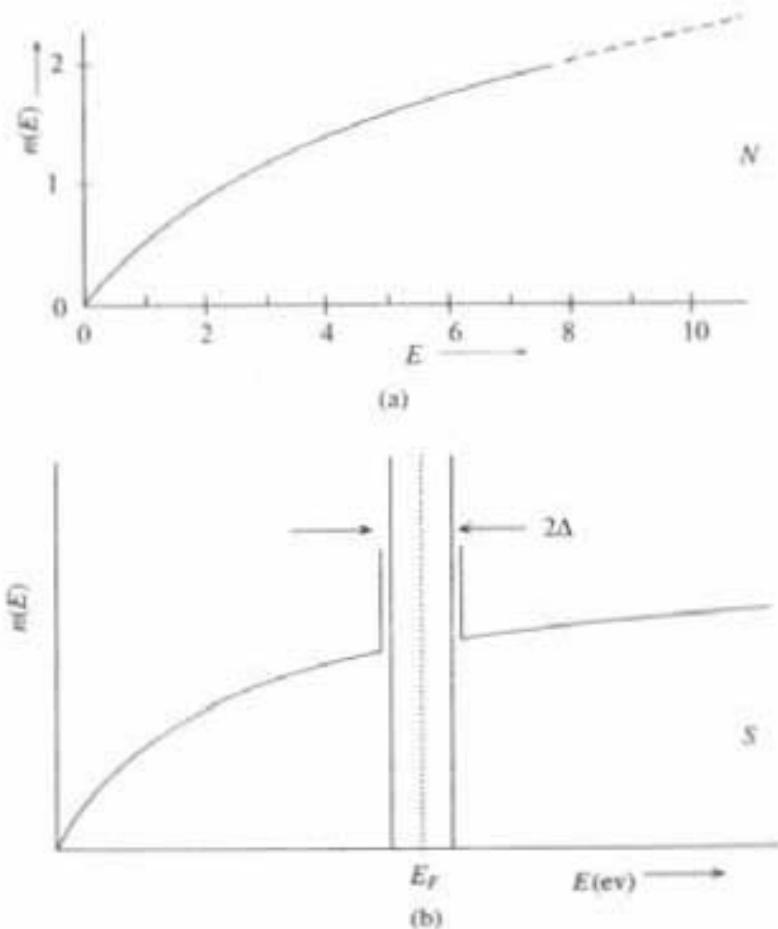


Fig. 7.4 Variation of density of states $n(E)$ with energy (E) for normal conductor, and that for superconductor with a pairing energy gap 2Δ , around $E = E_F$

to 0°K . Whereas, as temperature increases energy gap decreases, and at T_c it becomes zero.

The occupation probability of electron states in superconductor can be found out in the following way,

$$\text{Occupation probability} = n(E) \cdot F(E) = N_0(E).$$

where $F(E) = \frac{1}{1 + e^{(E-E_F)/kT}}$ is the Fermi factor, which is the probability that a state of energy E is filled up by electron at $T^{\circ}\text{k}$ as we have discussed earlier, in part-1 chapter-6, section 6.5.

The Fermi Dirac distribution [$F(E)$] function with energy E in Fig. 7.5(a) shows, that all the levels below E_F are filled up by electrons at $T = 0^{\circ}\text{K}$. Figure 7.5(b) shows a high occupation probability [$N_0(E) = n(E) \cdot F(E)$] of the superconducting states just below the energy gap. Above the gap a small density of normal (unpaired states) occurs. Figure 7.5(b) is obtained by superimposing Fig. 7.4(b) over Fig. 7.5(a).

The energy gap for superconductor can be obtained by electrons tunnelling method. The transition in zero magnetic field from the superconducting state to normal conducting state is reversible and it is a 2nd order phase transition. The transition between vapour to liquid is also reversible but it is 1st order phase

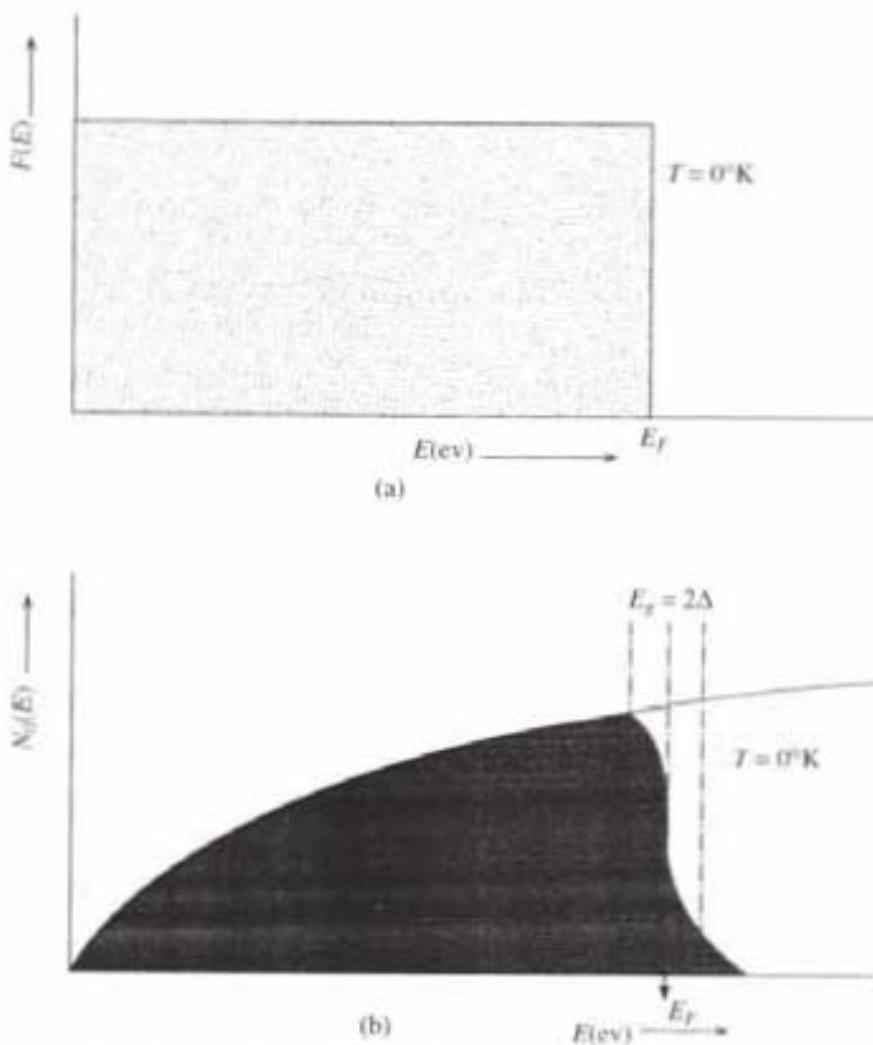


Fig. 7.5 (a) Variation of fermi factor $F(E)$ with energy (E) at 0°K ; (b) Variation of occupational probability $N_0(E)$ with energy E for superconductor

transition. 1st order phase transition is characterised by latent heat and discontinuity of the energy gap. Whereas at 2nd order phase transition there is no latent heat but there is discontinuity of heat capacity and the energy gap decreases to zero at temperature increase to T_c .

The penetration depth, coherence length, emerges as natural consequences of B.C.S. theory. London equation is obtained for magnetic field that vary slowly in space. The Meissner effect is also obtained in a natural way.

7.5 Effect of External Magnetic Field on Superconducting State

It is possible to destroy superconductivity by the application of sufficiently strong magnetic field. The critical or threshold value of the applied field for the destruction of superconductivity H_c is a function of temperature (T). The critical magnetic field H_c will be zero at $T = T_c$. The critical field varies with temperature according to the parabolic law as

$$H_c = H_0 [1 - (T/T_c)^2]$$

where H_0 is the critical field at absolute zero and T_c is the transition temperature.

The variation of H_c for different elements are shown in Fig. (7.6). A specimen is superconducting below the curve and normal above the curve.

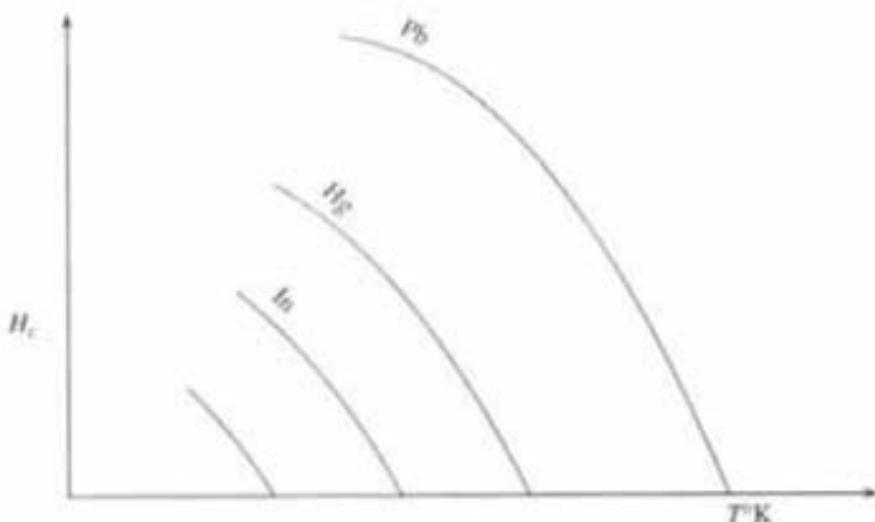


Fig. 7.6 Variation of (H_c) critical magnetic field, which destroys superconductivity for different elements

The critical magnetic field which causes a superconductor to become normal, is not necessary to be external applied field, it may arise as a result of electric current flowing in the conductor. In a long superconducting wire of radius R , superconductivity may be destroyed when current I exceeds the value I_c , which at the surface of the wire will produce critical field H_c , given by $I_c = 2\pi R H_c$; which is known as *Silsbee's rule*.

Type I and Type II Superconductors

There are two types of superconductors, I and II. In a bulk specimen of type I superconductor, the superconductivity can be completely destroyed by the application of external magnetic field H . Type I superconductors are called ideal superconductors.

In Fig. 7.7 magnetisation versus applied magnetic field for a bulk specimen exhibits a complete Meissner effect (perfect diamagnetism). Normal state will be restored when applied magnetic field $H > H_c$; this is called type I superconductor or ideal superconductor or soft superconductor (Fig. 7.7). The value of H_c are always too low for type I superconductors, to have any technical application in coils for superconducting magnets. In Fig. 7.7, $-\mu_0 M$ is plotted on the vertical scale, (-ve) value of M corresponds to diamagnetism ($\because M = \chi H$, χ the magnetic susceptibility is -ve for diamagnetism). Above H_c the magnetisation is too small to be seen on this scale, and specimen becomes normal conductor.

A type II superconductor tends to be alloys or transition metals with very high values of electrical resistivity in the normal state. A type II superconductor has two critical fields $H_{c1} < H_c < H_{c2}$. Type II superconductivity specimens have superconducting electrical properties upto field H_{c2} (Fig. 7.8), but between H_{c1} and H_{c2} the magnetic flux density $B \neq 0$, the flux starts to penetrate the specimen at H_{c1} lower than H_c and the Meissner effect is said to be incomplete. In the region between H_{c1} and H_{c2} , the superconductor is threaded by flux

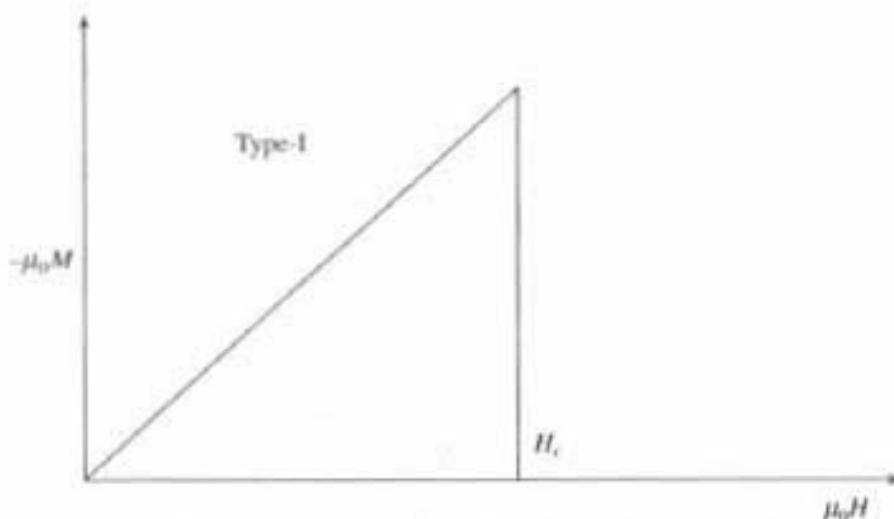


Fig. 7.7 Variation of magnetisation ($-\mu_0 M$) with applied magnetic field ($\mu_0 H$), for type-I superconductor. Negative value of M corresponds to diamagnetism, when $H > H_c$ specimen become normal conductor

lines and is said to be in the *vortex state*. H_{c2} is 100 times or more than H_{c1} . above H_{c2} the specimen is a normal conductor. Type II superconductor is called as *non-ideal or hard superconductor*. Between H_{c1} and H_{c2} , type II material does not exhibit a perfect Meissner effect, nor does it obey the London equations quantitatively. For technical use when one needs a material which remains superconducting in large magnetic fields one must use type II superconductor, Nb_3Sn is an example of type II superconductor.

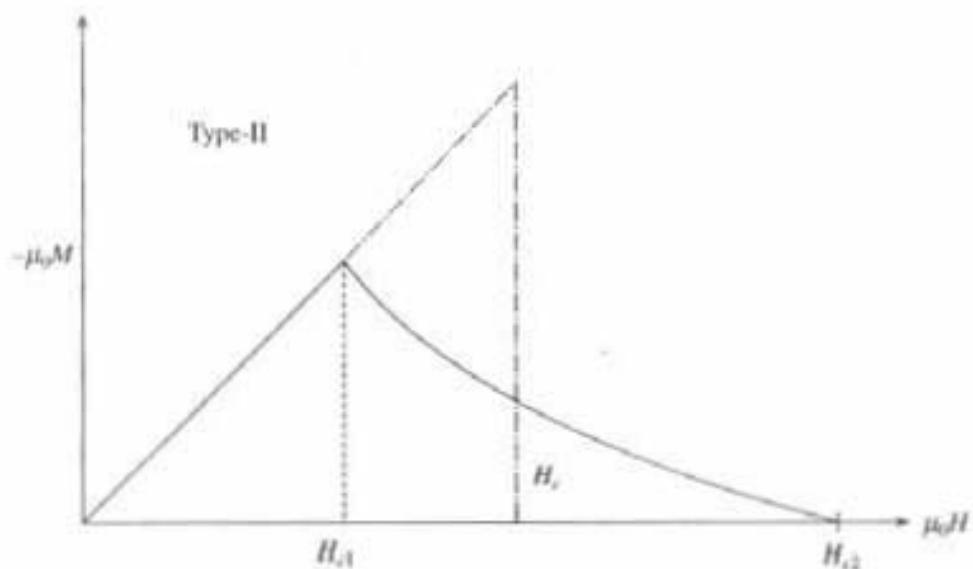


Fig. 7.8 Variation of ($-\mu_0 M$) with ($\mu_0 H$) for type-II superconductor, where between H_{c1} and H_{c2} superconducting element is said to be in a vortex state. Above H_{c2} specimen become normal conductor

7.6 London Equation

A theoretical understanding of the phenomenon associated with superconductivity can be understood in several ways. Certain results follow directly from thermodynamics. Many important results can be described by phenomenological

conditions, London equation is one of the most important phenomenological equation which explains the Meissner effect for the superconducting state.

As studied earlier in section 7.2, we cannot described the unique characterization of the superconducting state, i.e., $\vec{B} = 0$, in the applied magnetic field only by considering the superconductor as a medium of zero resistivity. Hence it is necessary to modify the Ohm's law drastically, in order to describe the zero resistivity and Meissner effect simultaneously in the superconducting state, without modifying the Maxwell's equations.

Ohm's law Eq. (7.2) can also be written as

$$\vec{J} = \sigma \vec{E} \quad (\because \sigma = \frac{1}{\rho} = \text{Conductivity}, \vec{J} = \text{Current density}) \quad (7.4)$$

Since in a superconducting state electrons behave differently, so it was assumed that in superconducting state, the current density \vec{J} is directly proportional to the vector potential \vec{A} of local magnetic field, where the magnetic induction $\vec{B} = \vec{\nabla} \times \vec{A}$ i.e. $\vec{B} = \text{Curl } \vec{A}$, and electric field intensity $\vec{E} = -\partial \vec{A} / \partial t$.

So

$$\vec{J} \propto \vec{A} \quad (7.5)$$

where

$$\vec{B} = \text{Curl } \vec{A} = \vec{\nabla} \times \vec{A} \quad (7.6)$$

The constant of proportionality in eqn. (7.5), has taken in S.I unit as $-1/\mu_0 \lambda_L^2$, where λ_L is a constant with dimensions of length, and of value $\lambda_L^2 = m/\mu_0 ne^2$. Hence Eqn. (7.5) becomes

$$\boxed{\vec{J} = -\frac{\vec{A}}{\mu_0 \lambda_L^2} = -\frac{ne^2}{m} \vec{A}} \quad (7.7)$$

instead of $\vec{J} = \sigma \vec{E}$ as in Ohm's law (eqn. 7.4), where n is the free electron density, e the electronic charge, m the electronic mass and \vec{A} is the magnetic vector potential. Equation (7.7) is called the *London Equation* from the name H. London who proposed this.

London equation can be derived in a superconducting state in the following way also : In a superconducting state resistivity $\rho = 0$ so the electrons can be freely accelerated, the acceleration is given by

$$\frac{d\vec{v}}{dt} = \frac{e\vec{E}}{m} \quad (7.8)$$

where \vec{v} is the velocity of the free electrons. Now by multiplying both sides by ne , we get

$$\frac{d}{dt} (ne\vec{v}) = \frac{ne^2}{m} \vec{E} = \frac{ne^2}{m} \frac{(-\partial \vec{A})}{\partial t}$$

Now $ne\vec{v} = \vec{J}$ = current density. Hence by integrating the above equation w.r.t. t we get

$$\boxed{\vec{J} = -\frac{ne^2}{m} \vec{A} = -\frac{1}{\mu_0 \lambda_L^2} \vec{A}}$$

which is same as eqn. (7.7), the London equation, where constant of integration is taken to be zero.

Meissner effect in superconducting state can be explained by London equation in the following way:

By taking curl on both sides of Eq. (7.7), it becomes

$$\vec{\nabla} \times \vec{J} = -\frac{ne^2}{m} \vec{\nabla} \times \vec{A} = -\frac{\vec{\nabla} \times \vec{A}}{\mu_0 \lambda_L^2} = -\frac{\vec{B}}{\mu_0 \lambda_L^2} \quad (7.9)$$

Now by considering Curl of the Maxwell's equation (3) under static condition as

$$\vec{\nabla} \times (\vec{\nabla} \times \vec{B}) = \mu_0 (\vec{\nabla} \times \vec{J})$$

or $\vec{\nabla} \times (\vec{\nabla} \times \vec{B}) = \vec{\nabla} (\vec{\nabla} \cdot \vec{B}) - \nabla^2 \vec{B} = \mu_0 (\vec{\nabla} \times \vec{J})$

by using vector identity. But $\vec{\nabla} \cdot \vec{B} = 0$ from Maxwell's equation (2).

Thus, $\vec{\nabla} \times (\vec{\nabla} \times \vec{B}) = -\nabla^2 \vec{B} = \mu_0 (\vec{\nabla} \times \vec{J}) \quad (7.10)$

By substituting equation (7.9) in (7.10), we get

$$\nabla^2 \vec{B} = \frac{\vec{B}}{\lambda_L^2} = \text{Constant} \quad (7.11)$$

This equation accounts for the *Meissner effect*, as it does not allow a solution uniform in space, which means that a uniform magnetic field cannot exist in a superconductor. Though uniform magentic field \vec{B} exists outside the superconductor, it gradually decreases exponentially to zero according to the solution of the equation (7.11) as

$$\vec{B}(x) = \vec{B}(0) e^{-x/\lambda_L}$$

where λ_L is the depth of penetration of the magenetic field known as "*London penetration depth*" having dimension of length (Fig. 7.9) and of magnitude as

$$\lambda_L^2 = \frac{m}{\mu_0 n e^2}$$

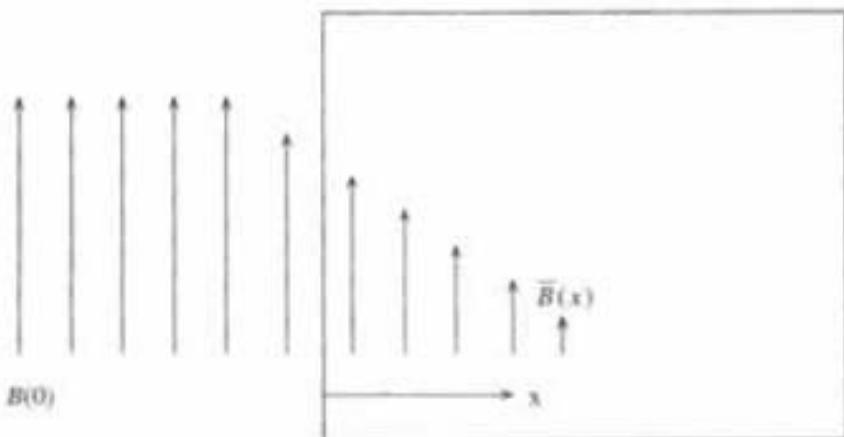


Fig. 7.9 Penetration depth

An applied external magnetic field \vec{H} will penetrate a thin film, if the thickness is much less than λ_L , i.e., in a thin film Meissner effect is incomplete. But for bulk specimen of superconductor, Meissner effect is complete. Thus, the shape

of the specimen has very important effect, which are simple only when the specimen has the shape of a long cylinder whose axis is parallel to the applied magnetic field.

7.7 Properties of Superconductor

1. At room temperature, superconducting material have greater resistivity than other.
2. The presence of trace amount of pure magnetic substance in the superconductor severely lowers the transition temperature.
3. By increasing the pressure the transition temperature can be increased for ultra low transition temperature specimen.
4. In superconducting state there is no change in the crystal structure, as found by X-ray diffraction method.
5. Entropy increases on going from superconducting state to normal state because entropy is a measure of the disorder and superconducting state is more ordered than the normal state.
6. For an ideal or type-I superconductor, there is a marked drop in thermal conductivity when the superconductivity sets in. For non-ideal or type II superconductor, in some cases there is an increase of thermal conductivity in superconducting state.
7. It has been observed that critical temperatures of superconductor varies with isotopic mass. The relation is

$$M^{1/2} T_c = \text{Constant}$$

or

$$T_c \propto 1/\sqrt{M}$$

where M is isotopic mass. The dependence of T_c on the isotopic mass verifies that lattice vibration and hence electron-lattice interactions are deeply involved in superconductivity. There is no other reason for the superconducting transition temperature to depend on the number of neutrons in the nucleus.

QUESTIONS

1. What is a superconductor? Explain critical temperature T_c and critical magnetic field H_c in a superconductor.
2. What is Meissner effect? Explain with diagram.
3. Explain the statement "Superconductor in a weak magnetic field acts as a perfect diamagnet".
4. What is London equation and penetration depth?
5. Give some examples of the elements and alloys which can be superconductor at a particular temperature.
6. State whether superconductors are in general good or bad conductors at normal situation? At what condition they become superconductor?
7. Explain B.C.S. theory for superconductivity?
8. Explain the effect of energy gap in the formation of superconducting state.

References

1. *Introduction to Solid State Physics* – C. Kittel-5th Edition-Wiley Eastern Ltd., New Delhi.
2. *Transistors*–Dennis Le Croissette – Prentice-Hall of India, Private Ltd., New Delhi, 1975.
3. *A Text Book of Optics*–Subrahmanyam Brij Lal-S. Chand & Company Ltd., New Delhi, 1993.
4. *Optics*–Ajay Ghatak – Tata-McGraw Hill Publishing Company Ltd., 1992.
5. *Optoelectronics* – J. Wilson – J.F.B. Hawkes – Prentice-Hall of India Private Ltd, New Delhi, 1996.
6. *Quantum Mechanics*-Leonard I. Schiff-Mc-Graw Hill-International Student Edition.
7. *Nuclear Physics*–Irving Kaplan–Addison-Wesley Publishing Company Inc. Indian Student Edition, Narosa Publishing House.
8. *Physics*, 4th Edition, Vol.-2, Extended-Halliday/Resnick/Krane-John Wiley and Sons, New York.

Subject Index

- Absent spectra, 205
Absorption coeff., 88, 91–93
Absorber, 92
Absorption of radiation, 323
Acceptance angle, 345
Acoustically good hall, 87
Acoustics, 87–96
Advantage of optical fiber, 357
Alpha (α) decay, 320
Amplifier, 70
Analyzer, 216, 217
Application of Ultrasonic, 80–82
Architectural acoustics, 87
Artificial radioactivity and isotopes, 372
Aston mass spectrograph, 266
Atom bomb, 381
Atomic model and magnetization, 151
Atomic packing factor, 13
Atomic radius, 12

B.C.C. Crystal, 7–14, 23, 24, 35
B.C.S. theory for superconductor, 399
Bain Bridge mass spectrograph, 269
Band, band gap, 100–103
Barium titanate, 31–33
Barkhausen jumps, 157
Barrier Penetration, 320–322
Barrier potential, 125–126
Barrier tunnelling, 317–322
Basis, 04, 05
Biaxial crystal, 214
Binding energy of nucleus, 369–371
Birefringence, 213
Bohr's theory, 231
Bragg's law, 231–233
Bragg's spectrometer, 234
Braking, 227
Bravis space lattice, 07
Breeder reactor, 387
Bremsstrahlung, 227
Brewster's angle, 215
Brewster's law, 215

C.B. mode, 141–144
C.C. mode, 141
C.E. mode, 141, 145–147
C.R.O. 257–263
Carbon dating, 375
Carbondioxide laser, 331
Causes of imperfection, 25
Chain reaction, 382–384
Characteristic X-ray spectrum, 225–227,
 229
Characteristics of Laser, 327
Charged particle, 244–270
 in crossed electric & magnetic field,
 248–251
 in parallel electric field, 244
 in transverse electric field, 245
 in transverse magnetic field, 247
Child-Langmuir law, 64
Circularly Polarized light, 217–219
Cladding, 344
Coordinates of lattice point, 09
Coercive force, 158
Colours in thin film, 168
Comparison between fission & fusion, 390
Comparison between zone plate and
 convex lens, 193
Components of optical communication, 354
Compton effect, 283–288
Conductivity, 113–118
 metal, 113
 extrinsic semiconductor, 118
 intrinsic semiconductor, 116
Constructive interference, 162–167, 170,
 171
Continuous X-ray spectrum, 227–229
Controlled chain reaction, 382–388
Coolidge's tube, 224
Co-ordination number, 11, 34
Core, 344
Critical angle, 344
Critical magnetic field on
 Superconductor, 402–404

- Critical size for nuclear reactor, 383-386
 Cross-section fission, 383
 Cross-section neutron absorption, 384
 Crystal density, 14, 15
 Crystal elemental, 04
 Crystal imperfection, 24-29
 Crystal ionic, 04
 Crystal plane, 16
 Crystal structure, 3-8
 body centred, 07
 face centred, 07
 simple cubic, 07
 Cubical Configuration, 35
 Curie temperature, 32, 156
 Curie's law, 155
 Curie-Wiess law, 156
 Cutin voltage, 130
 Cyclotron, 263
- Davison-Germer experiment, 290
 De-Broglie's hypothesis, 289-293
 Decay- radioactive, 365
 Deflection sensitivity, 252
 Density of a crystal, 14, 15
 Depletion region, 125
 Destructive interference, 162-167, 170, 171
 Detection of ultrasonic, 79
 Detection of X-ray, 225
 Determination of wave length by grating, 208
 Dextrorotatory crystal, 220
 Diamagnetic materials, 153
 Diamond structure, 29-31
 Dichroism, 215
 Diffraction, 187-209
 central maximum, 199
 double slit, 196-198
 grating, 198-200
 primary maxima, 199
 resultant intensity, 201-205
 single slit, 194-196
 X-rays, 231-237
 Diode as rectifier, 65, 132
 Diode characteristics, 60-63, 130
 Direction of a line, 18, 19
 Discovery of X-ray, 223
 Dislocation, 28
 edge, 28
 screw, 29
 Dispersion time, 351
 intermodal, 350
- Dispersive power of grating, 206
 Domain - ferromagnetic, 157
 Domain wall, 157
 Double Refraction, 213
 Dye laser, 332-334
 e/m measurement by Thomson method, 250
 Effective neutron multiplication factor (K_e), 385, 386
 Eigen function, 312
 Eigen value, 312
 Einstein's photo-electric equation, 282
 Electro- Optic effect, 221
 Electron microscope, 296
 Electrostatic focusing, 253-255
 Elemental crystal, 04
 Elliptically polarized light, 217-219
 Emission of radiation, 323
 Energy band 100-103
 Energy band gap, 100-103, 117
 Etchlon effect, 96
 External magnetic field on superconductor, 402-404
 Extra ordinary ray, 213, 214
 Extrinsic semiconductor, 109-113
- F.C.C. Crystal 7-14, 22, 36
 Fermi factor, 105
 Fermi level, 105
 extrinsic semiconductor, 112
 intrinsic semiconductor, 107
 variation with temperature, 113
 Ferroelectric crystal, 32
 Ferromagnetic domain, 157
 Ferromagnetic materials, 155
 anti ferro magnetic materials, 156
 ferri magnetic materials, 156
 Fiber graded index, 346
 Fiber optics communication, 354-356
 Fiber step index, 344, 346, 347
 Fission, 378-388
 Fission cross-section, 383
 Focusing, 252-257
 electrostatic, 253-257
 magnetostatics, 255-257
 Fraunhofer diffraction, 194
 Frenkel defect, 26, 27
 Fresnel's diffraction, 188, 194
 Fresnel's Half period Zone, 188
 Fringe width, 171
 Fringes in wedge film, 169-172
 Fusion, 388-390

- Gamma ray experiment, 294
 Gauge meter, 85
 Geiger-Muller counter, 361–365
 Graded index fiber, 346
 Grating, 198–200
 absent spectra, 205
 dispersive power, 206
 fraunhofer theory, 198
 maximum number of order possible, 206
 measurement of unknown wave length, 208
 overlapping of spectrum, 207
 resultant intensity, 201–205
 spectrum formation, 207
 Group velocity, 306–308

 Half life, 367
 Half period zone, 188
 Half wave plate, 220
 Hall effect, 119–121
 Hall angle, 120
 Hall coefficient, 120
 Hall voltage, 120
 Height of the potential barrier, 125
 Heisenberg Uncertainty principle, 293
 verification, 294–295
 Helium-neon laser, 329
 Hole, 104
 Hole current, 105
 Holography, 339–341
 Huygen's Wave Theory, 187
 Hydrogen bomb, 389
 Hysteresis loop, 158

 Important features of Miller indices, 24
 Insulator, 103
 Intensity distribution, 167, 204
 Interference-thin film, 164–180
 constructive, 164–167, 170, 171
 destructive, 164–167, 170, 171
 due to reflected light, 164, 169, 170, 171
 due to transmitted light, 166
 Intermediate metal, 48
 Internal quantum efficiency (η_i), 338
 Interplaner distance, 20
 spacing, 20
 Interstitial defect, 26
 Interstitial void, 32–34
 Intrinsic carrier concentration, 105
 Intrinsic semiconductor, 104–109

 Introduction to laser, 323
 fiber-optics, 342
 holography, 339
 superconductivity, 396
 Inversion temperature, 46–48
 Ionic crystal, 04

 Laser, 323–339
 carbon dioxide, 331
 characteristics, 327
 dye, 332
 He-Ne, 329
 operating principle, 325
 population inversion, 325
 ruby, 327
 semiconductor or diode Laser, 334–339
 Lasing action, 324, 325
 Lattice parameter, 06
 plane, 04, 5, 15
 point, 04
 space lattice, 04
 vector and direction, 18, 19
 Law of intermediate metal, 48
 successive temperature, 49
 Laue method, 235
 LED, 334, 335, 338
 Left Circularly Polarized light (LCP), 219, 220
 Laevorotatory Crystal, 220
 Ligancy (8, 6, 4, 3), 34–40
 Light propagation in fiber, 344
 Line spectrum of X-ray, 225–227
 Linearly Polarised light, 212–218
 Lissajous Pattern, 262
 London equation, 404–407
 Long wavelength Cut-off, 354
 Losses in fiber, 352

 Magnetostatic focusing, 255–257
 Magnetic materials, 151–158
 Magnetization curve, 158
 Magneto-striction effect, 74
 oscillator, 74
 Malus Law, 216, 217
 Mass spectrograph, 266–270
 Aston, 266
 Bain bridge, 269
 Mean free path (λ), 115
 Measurement of absorption co-eff, 92, 93
 Meissner effect, 397–399
 Meridional ray, 347

- Metal, 102
 Metastable state, 324
 Miller indices, 15–24
 different lattice plane, 16
 direction, 18, 19
 Mobility, 114, 116
 Moderator, 384, 385
 Modes of propagation in fiber, 348
 Molecular field inside a Magnetic material, 152
 Moseley's law, 229–231
 Multimode, monomode, 343, 350

 Natural radio isotopes, 366
 Nature of interference pattern, 172
 Necessity of broad source, 168
 Negative Crystal, 214
 Neutral temperature, 46–48
 Newton's ring, 174–178
 diameter of bright ring, 177
 diameter of dark ring, 177
 Nicol prism, 216
 as analyser, 216
 as polariser, 216
 Nonreflecting film, 179
 Normalisation condition, 305, 310
 n-p-n transistor, 137–147
 CB, CE, Characteristics, 141–147
 n-type semiconductor, 109, 112, 113
 Nuclear force, 368
 fission, 378–388
 fusion, 388–390
 reaction, 376–378
 reactor, 383–388
 screening factor, 230
 Number of Atoms/Molecules per unit cell, 10
 Numerical aperture, 345, 346

 Octahedral configuration, 36
 Open window unit, 92
 Optic axis, 213, 214
 Optical Activity, 220
 Optical diffraction, 187–209
 Optical fiber communication, 354, 355
 Optimum reverberation time, 87–91
 Ordinary ray, 213, 214
 Origin of color in thin film, 168
 Origin or seebeck effect, 44
 Origin of X-ray, 225–229

 Oscillator, 71, 74, 78
 Overlapping of spectrum, 207

 Packing fraction, 368, 369
 Pairing energy gap in superconductor, 400–402
 Parallel film, 164–167
 Paramagnetic Materials, 154
 Phase difference, 263
 Phase velocity, 306–308
 Photo electric effect, 279–283
 Photo elasticity, 221
 Photo Detector, 352, 353
 Photon, 279, 282, 283–286, 323–326
 Physical significance of wave function, 304
 Piezo electric crystal, 76
 effect, 76
 oscillator, 78
 p-n junction, 125–136
 p-n junction Forward bias, 127, 131, 134–136
 p-n junction Reverse bias, 127, 131, 134–136
 as LED, Diode laser, 334–339
 as rectifier, 132
 reverse bias, 127, 131, 134–136
 V-I characteristics, 130–131
 with applied voltage, 127
 without applied voltage, 125
 p-n-p transistor, 137–147
 characteristics-CB, CE, 141–147
 Polarisation, 211–221
 circularly polarised light, 217, 218
 elliptically polarised light, 217, 218
 linearly polarised light, 212, 215, 216, 218
 Polariser, 213–216
 Polarising Angle, 215
 Polariscope, 216
 Polaroid, 217
 Population inversion, 325
 Positive crystal, 214
 Powder method, 236
 Power reactor, 386–387
 Primary maxima, 199
 Principal plane, 214
 Production of X-ray, 223
 Propagation of light in fiber, 344
 Properties of matter wave, 292
 Properties of superconductor, 407

- Properties of X-ray, 231
 p-type semiconductor, 111–113
 Pumping electrical, 330
 optical, 327
 PZT ceramics, 83
- Q-value, 371, 372
 Quantum efficiency(η), 354
 Quantum physics, 278–322
 Quantum theory of radiation, 278
 Quarter wave plate, 219
 Quasi Fermi level, 336, 336
 Quenching, 25
- Radio activity, 361–374
 artificial, 372
 Radio isotopes artificial, 372
 Radio isotopes natural, 366
 Radio isotopes uses, 374
 Reactor Nuclear, 383–388
 breeder, 387
 heterogeneous, 384
 homogeneous, 384
 thermal power 386
 Rectifier, 65, 132–136
 Rectifier equation, 134–136
 Refractive index, 342
 Relaxation time (τ), 115
 Residual flux density (B_r), 158
 Resolving power of electron microscope, 296
 Resonance within a Building, 96
 Responsivity, 354
 Resultant intensity, 201–205
 double slit, 202
 grating, 203
 single slit, 201
 Reverberation, 87–91
 Reverberation time, 87–91
 Sabine's formula, 90
 Richardson-Dushman equation, 58
 Right Circularly Polarized light (RCP), 219, 220
 Rotating crystal method, 236
 Ruby laser, 327
 Rutherford's experiment, 372
- Sabine's formula, 90
 Schottky defect, 26
 Schrödinger wave equation, 304–322
 applications of STIE, 313–322
- time dependent, 308
 time independent (STIE), 311
 S.C. Crystal, 7–14 21
 Secondary Minima and Maxima, 204
 Seebeck effect, 44, 45
 co-efficient, 46–48
 origin, 44
 Semiconductor, 100–150
 conductivity, 116–119
 electron, hole, 104
 extrinsic, 109–113
 Fermi level, 105–109, 112–113
 Fermi factor, 105
 formation of energy band, 100
 intrinsic, 104–109
 I-V characteristics, 61, 128, 131
 n-type, 109–113
 n-p-n transistor, 137–147
 p-type, 111–113
 p-n junction, 125–134
 p-n-p transistor, 137–147
 rectifier, 132–136
 transistor characteristics, 141–147
 Semiconductor laser, 334–339
 Sensitivity, 252
 Seven crystal system, 06
 Silsbee's rule, 403
 Skew rays, 347
 Sonar, 81
 Space charge, 61
 Space charge limited current, 62
 Space charge region, 61
 Space lattice, 04
 Spectrograph, 266, 269
 Spectrometer, 208, 234
 Spectrum formation in grating, 207
 overlapping of spectrum, 207
 Spontaneous emission, 323
 Spontaneous magnetization, 155
 Stationary states, 313
 Stellar thermonuclear reaction, 389, 390
 Stimulated emission, 324
 Sun control film, 179
 Superconductivity, 396–407
 B.C.S theory, 399
 cooper pair, 399
 critical magnetic field, 402–404
 London equation, 404–407
 Meissner effect, 397–399
 pairing energy gap, 400
 penetration depth, 406

- properties, 407
- transition temperature, 396
- Superposition of two plane polarised wave, 217
- Temperature limited current, 62
- Temperature neutral, 46–48
 - inversion, 46–48
- Testing the flatness, 173
- Tetrahedral configuration, 37
- Thermal power reactor, 386
- Thermister, 51, 117
- Thermo nuclear reaction, 388–390
- Thermoelectric effect, 44–56
 - power, 47
 - series, 45
 - thermometer, 49
- Thermoelectricity, 44–56
- Thermion emission, 57–67
 - characteristics, 60–63
 - rectifier, 65
 - valve, 60
- Thickness of the wedge, 172
- Thin film interference, 164–186
 - constructive, 164–167
 - destructive, 164–167
 - intensity, 167
 - nature of interference pattern, 172
 - origin of color in thin film, 168
 - reflected light, 164
 - transmitted light, 166
- Three dimensional photograph, 340
- Threshold voltage, 130
- Time dispersion, 351
- Timebase generator, 258–260
- Total internal reflection, 216, 342–344
- Totally reflecting film, 179
- Transistor, 136–147
 - C.B.C.E, C.C mode, 141–147
 - characteristics, 141–147
 - p-n-p, n-p-n, 137–147
- Transition temperature, 396
- Transuraneum elements, 374
- Triangular configuration, 38
- Trigger circuit, 258–260
- Trigger pulse, 260
- Tuned circuit, 71
- Tunnel diode, 321
- Ultrasonic, 69–83
 - application, 80–82
 - detection, 79
 - oscillator, 74, 78
 - production, 69, 74, 78
 - properties, 80
- Ultrasonic transducer, 83
- Uniaxial crystal, 214
- Unpolarized light, 213
- Uses of optical fiber, 356
- Uses of radio isotopes, 374, 375
- Uses of x-ray diffraction pattern, 237
- Verification of Heisenberg principle, 294–295
- Vertical deflection, 251, 252
 - electric field, 251, 252
 - magnetic field, 252
 - sensitivity, 252
- V-I characteristics, 61, 128, 130
- Vortex state, 404
- Wave function, 304
 - development of wave equation, 306
 - physical significance, 304
 - Schrödinger wave equation, 304–322
- Wave nature of light, 161
- Wave nature of particle, 288
 - de Broglie's hypothesis, 289
 - Devisor Germer experiment, 290
- Wave packet, 307
- Wave Particle duality, 288
- Wedge film, 169–173
- X-ray, 223–237
 - characteristic spectra, 225–227, 229
 - continuous spectra, 227–229
 - detection, 225
 - discovery, 223
 - origin, 225–229
 - production, 223
- X ray diffraction, 231–237
 - Bragg's law, 231–233
 - Bragg's spectrometer, 234
 - different methods for x-ray diffraction, 234–237
 - explanation for x-ray diffraction, 233
 - uses of X-ray diffraction, 237
- Zone plate, 191–194
- Zone plate comparison with convex lens, 193
- Zone half-period, 188