

ENGINEERING PHYSICS

B.K. Pandey, DPhil

Professor & Head,

Department of Physics and Material Science
M.M.M. University of Technology, Gorakhpur (UP)

S. Chaturvedi

Jahangirabad Educational Trust Group of Institutions, Barabanki (U.P.)



Australia • Brazil • India • Mexico • Singapore • United Kingdom • United States



Engineering Physics

B.K. Pandey

S. Chaturvedi

© 2018 Cengage Learning India Pvt. Ltd.

ALL RIGHTS RESERVED. No part of this work covered by the copyright herein may be reproduced, transmitted, stored, or used in any form or by any means graphic, electronic, or mechanical, including but not limited to photocopying, recording, scanning, digitizing, taping, Web distribution, information networks, or information storage and retrieval systems, without the prior written permission of the publisher.

For permission to use material from this text or product, submit all requests online at
www.cengage.com/permissions

Further permission questions can be emailed to
India.permission@cengage.com

ISBN-13: 978-81-31519-67-7

ISBN-10: 81-31519-67-8

Cengage Learning India Pvt. Ltd.

T-303, 3rd Floor
DAV Complex, Plot 3
Mayur Vihar Phase 1
Delhi-110091

Cengage Learning is a leading provider of customized learning solutions with office locations around the globe, including Australia, Brazil, India, Mexico, Singapore, United Kingdom, and United States. Locate your local office at:

www.cengage.com/global

Cengage Learning products are represented in Canada by Nelson Education, Ltd.

For product information, visit www.cengage.co.in

Printed in India
Eighth Impression 2021

Contents

UNIT I : THEORY OF RELATIVITY

1. Theory of Relativity	3
1.1 Introduction	3
1.2 Concept of Theory of Relativity	3
1.3 Frames of Reference	4
1.3.1 Inertial Frames of Reference	5
1.3.2 Non-inertial Frames of Reference	6
1.4 Galilean Transformation Equations	6
1.4.1 Consequences of Galilean Transformation Equations	8
1.4.2 Newtonian Principle of Relativity	9
1.5 Concept of “Luminiferous Ether” Medium	11
1.6 Need of Special Theory of Relativity	11
1.7 Michelson–Morely Experiment	12
1.7.1 Negative Result of Michelson–Morely Experiment	14
1.7.2 Explanation of Negative Result	15
1.8 Postulates of Special Theory of Relativity	17
1.9 Lorentz Transformation Equations	17
1.10 Inverse Lorentz Transformation	19
1.11 Consequences of Lorentz Transformation Equations	20
1.11.1 Length Contraction	20
1.11.2 Time Dilation	28
1.11.3 Relativity of Simultaneity	33
1.12 Transformation of Velocity Components—Velocity Addition	34
1.12.1 Verification of Constancy of Speed of Light	40
1.13 Variation of Mass with Velocity	40
1.14 Mass–Energy Equivalence	44
1.15 Relativistic Relation between Energy and Momentum	46

1.16 Expression of Relativistic Momentum	46
1.17 Massless Particles and Their Explanation	47
1.18 Formulae and Highlights	61

UNIT II : SOUND WAVE AND ITS APPLICATIONS

2. Oscillation and Waves	73
2.1 Introduction	73
2.2 Simple Harmonic Motion	73
2.2.1 Velocity of Motion	74
2.2.2 Acceleration	75
2.2.3 Time Period	75
2.2.4 Frequency	76
2.2.5 Phase	76
2.3 Damped Harmonic Oscillations	77
2.3.1 Wave Equation of Damped Vibration	77
2.3.2 Analysis of Damped Harmonic Oscillator	78
2.3.3 Important Parameters under Damped Harmonic Oscillations	81
2.3.4 Energy in Damped Harmonic Oscillation	83
2.3.5 Power Dissipation (P)	84
2.3.6 Quality Factor (Q)	84
2.4 Forced Vibration	85
2.4.1 Analytical Treatment of Forced Vibration	86
2.4.2 Equation of Motion of Forced Vibration and Its Solution	86
2.4.3 Amplitude of Forced Vibration	88
2.4.4 Resonance	89
2.4.5 Sharpness of Resonance	90
2.4.6 Average Power in Forced Vibration	91
2.4.7 Power Dissipated	92
2.4.8 Quality Factor in Terms of Band width	93
2.4.9 Quality Factor	94
2.5 Coupled Oscillations	100
2.6 Wave Motion	100
2.6.1 Characteristics of Wave Motion	101

2.7 Wave as Periodic Variation in Space and Time	101
2.8 Types of Waves	102
2.9 Wave Equation	103
2.10 Differential Equation of Wave Motion	104
2.11 Reflection and Transmission of Transverse Waves at a Boundary	105
2.11.1 Special Cases	108
2.12 Reflection and Transmission of Longitudinal Sound Waves at Boundaries	111
2.13 Reflection and Transmission of Sound Intensity	111
2.14 Formulae and Highlights	118
3. Ultrasonic Waves	125
3.1 Introduction	125
3.2 Ultrasonic Waves	125
3.3 Production of Ultrasonic Waves	126
3.3.1 Mechanical Method	126
3.3.2 Piezoelectric Generator	126
3.3.3 Magnetostriction Generator	127
3.4 Detection of Ultrasonic Waves	131
3.5 Measurement of Wavelength of Ultrasonic Waves	132
3.6 Properties of Ultrasonic Waves	133
3.7 Applications of Ultrasonic Waves	133
3.8 Formulae and Highlights	135
4. Non-destructive Testing	141
4.1 Introduction	141
4.2 Non-destructive Testing	141
4.3 Objectives of Non-destructive Testing	142
4.4 Types of Defects	142
4.4.1 Cracking	142
4.4.2 Spalling	143
4.4.3 Staining	143
4.4.4 Construction and Design Defects	143
4.4.5 Honeycombing	144
4.4.6 Dusting	144
4.4.7 Blistering	144

4.4.8 Rain Damage	144
4.5 Methods of Non-destructive Testing	144
4.5.1 Liquid Penetration Method	144
4.5.2 Dye Penetration Method	145
4.5.3 Radiographic Testing	146
4.5.4 Ultrasonic Inspection Method	147
4.5.5 Pulse Echo Method	150
4.5.6 Magnetic Particle Testing	151
4.5.7 Eddy Current Testing	151
4.6 Formulae and Highlights	151
5. Architectural Acoustics	155
5.1 Introduction	155
5.2 Classification of Sound Waves	155
5.2.1 Audible Waves or Sonics	155
5.2.2 Infrasonic Waves or Subsonic	155
5.2.3 Ultrasonic Waves	156
5.3 Noise and Musical Sound	156
5.4 Weber–Fechner Law (Sensation of Loudness)	156
5.5 Loudness Level and Intensity: Bel and Decibel	157
5.6 Basic Requirements for the Acoustically Good Halls	158
5.7 Reverberation	158
5.8 Determination of the Time of Reverberation: Sabine's Law	159
5.9 Derivation of Expression for Reverberation Time (Sabine's Law)	160
5.9.1 Deductions of Sabine	162
5.10 Absorption Coefficient	162
5.10.1 Determination of Absorption Coefficient	162
5.11 Factors Affecting the Architectural Acoustics and Their Remedies	163
5.12 Formulae and Highlights	170

UNIT III : SOLID STATE PHYSICS

6. Elements of Crystallography	177
6.1 Introduction	177
6.2 Crystal Structure	177

6.3	Space Lattice	179
6.4	Basis	179
6.5	Unit Cell	180
6.6	Lattice Parameter	180
6.7	Lattice Vector and Direction	181
6.8	Seven Crystal Systems	181
6.9	Bravais Space Lattice	182
6.10	Symmetry Element of a Crystalline Solid	183
6.10.1	Axis of Symmetry	183
6.10.2	Planes of Symmetry	184
6.10.3	Centre of Symmetry	184
6.11	Coordinates of Lattice Point	184
6.12	Number of Atoms Per Unit Cell	184
6.13	Packing Density or Atomic Packing Factor	186
6.14	Miller Indices	186
6.14.1	Procedure for Finding Miller Indices	187
6.15	Different Crystal Planes with Miller Indices	187
6.16	Important Features of Miller Indices	188
6.17	Direction of a Line	188
6.18	Interplanar Spacing	189
6.19	Atomic Radius in a Cubic System	191
6.19.1	Atomic Radius of Simple Cubic Structure	191
6.19.2	Atomic Radius of Face-Centred Cubic Structure	192
6.19.3	Atomic Radius of Body-Centred Cubic Structure	192
6.20	Angle between Two Planes	193
6.21	Dimensions of Unit Cell	193
6.22	Coordination Number	194
6.23	Interplanar Spacing in Cubic System	194
6.23.1	Simple Cubic Crystal	194
6.23.2	Face-Centred Cubic Crystal	195
6.23.3	Body-Centred Cubic Crystal	196
6.24	Formulae and Highlights	201

7. X-Rays and Compton Effect	207
7.1 Introduction	207
7.2 Production of X-Rays	207
7.3 X-Ray Spectra	208
7.3.1 Continuous X-Ray Spectra	208
7.3.2 Characteristic X-Ray Spectra	210
7.4 Important Properties of X-Rays	211
7.5 Diffraction of X-Rays	211
7.6 Laue's Experimental Demonstration	211
7.7 Bragg's Law	212
7.8 Bragg's Spectrometer	214
7.9 Bragg's Law and Crystal Structure	215
7.10 Compton Effect	216
7.10.1 Expression for Compton Shift	216
7.10.2 Discussion on Compton Effect	219
7.10.3 Importance of Compton Effect	220
7.10.4 Experimental Verification of Compton Effect	220
7.10.5 Direction of Recoiled Compton Electron	222
7.10.6 Kinetic Energy of Recoiled Electron	223
7.11 Crystal Structure of NaCl	225
7.12 Crystal Structure of Diamond	225
7.13 Applications of X-Rays	226
7.14 Formulae and Highlights	241
8. Band Theory of Solids and Semiconducting Materials	251
8.1 Introduction	251
8.2 Basis of the Band Theory of Solids	251
8.3 Electrons in a Periodic Potential of Ion Cores	252
8.3.1 The Bloch Theorem	252
8.3.2 Kronig–Penney Model	253
8.3.3 Allowed Energy Bands	255
8.4 Energy Bands in Solids	256
8.5 Classification of Solids	258
8.5.1 Conductors (Metals)	258

8.5.2	Insulators	259
8.5.3	Semiconductors	259
8.6	Intrinsic Semiconductors	259
8.6.1	Current Conduction in Intrinsic Semiconductors	260
8.7	Extrinsic Semiconductors	261
8.7.1	Donor or N-type Semiconductor	261
8.7.2	Acceptor or P-type Semiconductor	262
8.7.3	Conductivity of Semiconductor Materials	263
8.8	P–N Junction Diode	265
8.8.1	Depletion Layer	266
8.8.2	Behaviour of a P–N Junction under Biasing	267
8.8.3	Voltage Current Characteristics of P–N Junction	268
8.8.4	Applications of P–N Junction Diode	269
8.9	Zener Diode	269
8.9.1	Voltage–Current Characteristics of Zener Diode	270
8.9.2	Applications of Zener Diode	270
8.10	Varactor Diode	271
8.10.1	Applications of Varactor Diode	271
8.11	Light Emitting Diodes (LEDs)	272
8.11.1	Applications of LEDs	272
8.12	Solar Cells	273
8.12.1	Applications of Solar Cells	274
8.13	Photovoltaic Cell	274
8.13.1	Expression for Photovoltaic emf	275
8.14	Photoconductivity in Semiconductors	276
8.14.1	Cut-off Wavelength	276
8.14.2	Effect of Impurity on Photoconductivity	277
8.15	Hall Effect	277
8.15.1	Hall Voltage and Hall Coefficient	277
8.15.2	Applications of Hall Effect	280
8.16	Formulae and Highlights	284
9.	Electrical Conductivity and Electron Theory of Metals	293
9.1	Introduction	293

9.2	Free Electron Concept	293
9.3	Classical Free Electron Theory	294
9.4	Assumptions of Classical Free Electron Theory	295
9.5	Drift Velocity	295
9.6	Expression of Drift Velocity	296
9.7	Relaxation Time	297
9.8	Mean Collision Time	298
9.9	Mean Free Path	299
9.10	Expression for Electrical Conductivity of Metals	300
9.11	Effect of Temperature on Resistivity of Metals	302
9.12	Effect of Impurity on the Resistivity of Metal	302
9.13	Failure of Classical Free Electron Theory	303
9.14	Quantum Free Electron Theory	309
9.15	Assumptions of Quantum Free Electron Theory	309
9.16	Particle in a Three-dimensional Box	309
9.17	Fermi–Dirac Statistics	314
9.18	Filling of Energy Levels	314
9.19	Fermi Energy	317
9.19.1	Expression of Fermi Energy at 0 K (E_{F0})	317
9.20	Density of States [$D(E)$]	318
9.21	Expression of Electrical Conductivity	321
9.22	Temperature Dependence of Resistivity in Metals	323
9.23	Merits of Quantum Free Electron Theory	325
9.24	Formulae and Highlights	329
10.	Dielectrics	341
10.1	Introduction	341
10.2	Dielectric Constant	341
10.3	Dielectric Polarisation	343
10.3.1	Non-polar Dielectric Materials in an Electric Field	343
10.3.2	Polar Dielectric Materials in Electric Field	344
10.4	Electric Polarisation Vector (\vec{P})	345
10.5	Electric Displacement Vector (\vec{D})	346
10.6	Relation between Vectors \vec{E} , \vec{D} , and \vec{P}	347

10.7	Types of Polarisation	348
10.7.1	Electronic Polarisation	348
10.7.2	Ionic Polarisation	349
10.7.3	Orientational Polarisation	350
10.7.4	Space-Charge Polarisation	350
10.7.5	Total Polarisation	351
10.8	Internal Field in Liquids and Solids	360
10.8.1	Lorentz Field	362
10.9	Clausius–Mossotti Equation	362
10.10	Debye Equation	364
10.11	Relation between Dielectric Constant and Refractive Index (Lorentz–Lorentz Formula)	364
10.12	Ferroelectricity	369
10.12.1	Origin of Ferroelectricity	369
10.12.2	Hysteresis in Ferroelectrics	371
10.12.3	Applications of Ferroelectricity	372
10.13	Piezoelectricity	372
10.13.1	Origin of Piezoelectricity	372
10.13.2	Piezoelectricity in Quartz	373
10.13.3	Explanation for the Origin of Piezoelectricity	374
10.13.4	Applications of Piezoelectricity	375
10.14	Frequency Dependence of Dielectric Constant	375
10.15	Dielectric Loss or Loss Tangent	376
10.15.1	Expression for Dielectric Loss	377
10.15.2	Tangent Loss in Terms of Complex Dielectric Constant	379
10.15.3	Dielectric Loss Spectrum	379
10.16	Dielectric Breakdown	380
10.17	Applications of Dielectric Materials	380
10.18	Formulae and Highlights	381
11.	Magnetic Properties of Materials	393
11.1	Introduction	393
11.2	Useful Parameters	393
11.2.1	Magnetic Density (B)	393

11.2.2	Intensity of Magnetisation (I)	393
11.2.3	Magnetic Susceptibility (χ)	393
11.2.4	Relative Permeability (μ_r)	394
11.2.5	Relation between μ_r and χ_m	394
11.3	Magnetic Moment of an Atom: Bohr Magneton	395
11.4	Classification of Magnetic Materials	395
11.4.1	Diamagnetic Substances	396
11.4.2	Paramagnetic Substances	397
11.4.3	Ferromagnetic Substances	398
11.5	Langevin's Theory of Diamagnetism	399
11.5.1	Analysis	399
11.6	Langevin's Theory of Paramagnetism	402
11.6.1	Analysis	402
11.7	Hysteresis (Lagging)	405
11.7.1	Energy Loss Due to Hysteresis	406
11.8	Uses of Hysteresis Curve	407
11.8.1	Construction of Permanent Magnets	407
11.8.2	Electromagnets	408
11.8.3	Construction of Transformer Cores, Telephone Diaphragms, Armature of Dynamos and Motors, and Cores of Chokes	408
11.9	Comparison between Soft Iron and Steel	408
11.10	Formulae and Highlights	409

UNIT IV : OPTICS WITH LASER AND OPTICAL FIBRE

12.	Interference	417
12.1	Introduction (Nature of Light)	417
12.2	Interference of Light	418
12.3	Young's Experiment (Division of Wavefront)	418
12.4	Resultant Intensity due to Superposition of Two Interfering Waves	419
12.4.1	Analysis	419
12.4.2	Conditions for Maxima and Minima	420
12.5	Interference and Conservation of Energy	423
12.6	Determination of Fringe Width in Young's Experiment	423
12.7	Shape of Interference Fringes	425

12.8	Conditions for Interference of Light	427
12.9	Coherent Sources	427
12.9.1	No Interference by Independent Sources	427
12.10	Production of Coherent Sources	427
12.11	Fresnel's Biprism	429
12.11.1	Production of Fringes	429
12.11.2	Experimental Method for Determination of Wavelength	429
12.11.3	Measurements	430
12.11.4	Distance between Virtual Sources	431
12.12	Displacement of Fringes	431
12.13	Phase Change on Reflection (Stokes Treatment)	440
12.14	Interference due to Thin Films (Division of Amplitude)	441
12.14.1	Interference Due to Reflected Rays	442
12.14.2	Interference Due to Transmitted Rays	443
12.15	Interference due to Wedge-Shaped Thin Film: Fringe Width	443
12.15.1	Fringe Width	445
12.16	Formation of Newton's Rings	448
12.16.1	Diameters of Rings	449
12.16.2	Experimental Method for Calculation of Wavelength	451
12.16.3	Determination of Refractive Index of a Liquid by Newton's Ring	452
12.16.4	Newton's Rings with Both Curved Surfaces	453
12.17	Formulae and Highlights	460

13. Diffraction 467

13.1	Introduction	467
13.2	Diffraction of Light	467
13.3	Resultant of n Simple Harmonic Waves	469
13.4	Fraunhofer Diffraction at a Single Slit	471
13.4.1	Directions of Maxima and Minima	472
13.5	Fraunhofer Diffraction due to Double Slit	481
13.6	Plane Diffraction Grating	484
13.6.1	Principal Maxima	485
13.6.2	Minima	485
13.6.3	Secondary Maxima	486

13.7 Missing Orders or Absent Spectra	487
13.8 Determination of Wavelength of Light by Grating	488
13.9 Dispersive Power of a Grating	489
13.10 Difference between Prism and Grating Spectra	489
13.11 Resolving Power of an Optical Instrument: Rayleigh's Criterion of Resolution	501
13.12 Resolving Power of a Grating	502
13.13 Resolving Power of Telescope	503
13.14 Resolving Power of Microscope	504
13.15 Formulae and Highlights	511
14. Polarisation	519
14.1 Introduction	520
14.2 Polarisation of Light	521
14.3 Representation	522
14.4 Plane of Vibration	522
14.5 Plane of Polarisation	522
14.6 Production of Plane Polarised Light	522
14.6.1 Polarisation by Reflection: Brewster's Law	522
14.6.2 Polarisation by Refraction: (Pile of Plates: Malus Law)	523
14.6.3 Polarisation by Double Refraction	525
14.6.4 Polarisation by Scattering	526
14.7 Doubly Refracting Crystals	526
14.8 Calcite Crystal	526
14.8.1 Optic Axis of the Crystal	527
14.8.2 Principal Section of the Crystal	527
14.9 Polarisation by Double Refraction	527
14.10 Nicol Prism	528
14.10.1 Nicol Prism as a Polariser and as an Analyser	528
14.11 Polaroids	529
14.12 Huygens Explanation of Double Refraction	530
14.13 Plane, Circularly, and Elliptically Polarised Light	531
14.14 Superposition of Two Plane Polarised Waves having Perpendicular Vibrations	531

14.15 Retardation Plates	533
14.15.1 Quarter Wave Plate	533
14.15.2 Half Wave Plate	534
14.16 Production and Detection of Plane, Circularly, and Elliptically Polarised Light	535
14.16.1 Plane Polarised Light	535
14.16.2 Circularly Polarised Light	535
14.16.3 Elliptically Polarised Light	535
14.17 Optical Rotation (Rotatory Polarisation)	537
14.18 Fresnel's Theory of Optical Rotation	537
14.19 Specific Rotation	539
14.19.1 Laurent Half-Shade Polarimeter	539
14.19.2 Biquartz Polarimeter	541
14.20 Formulae and Highlights	551
15. Lasers and Holography	559
15.1 Introduction	559
15.2 Atomic Excitation and Energy States	559
15.3 Interaction of External Energy with the Atomic Energy States	560
15.3.1 Absorption	560
15.3.2 Spontaneous Emission	561
15.3.3 Stimulated Emission	562
15.4 Einstein Coefficients	563
15.5 Population Inversion	565
15.6 Pumping Mechanism	566
15.6.1 Optical Pumping	566
15.6.2 Electric Discharge	566
15.6.3 Inelastic Collisions between Atoms	566
15.6.4 Direct Conversion	567
15.6.5 Chemical Reaction	567
15.7 Components of Laser System	567
15.7.1 Active Medium	567
15.7.2 Population Inversion	567
15.7.3 The Optical Resonator	568

15.8	Three-Level Laser	568
15.9	Four-Level Laser	569
15.10	Ruby Laser	569
15.11	Helium–Neon Laser	571
15.12	Carbon Dioxide Laser	572
15.13	Semiconductor Lasers	573
15.14	Applications of Laser	574
15.14.1	Laser Beam Welding	574
15.14.2	Laser Cutting	575
15.14.3	Laser Drilling with Ytterbium Lasers	575
15.15	Holography	576
15.16	Basic Principle of Holography	576
15.17	Construction of the Hologram	576
15.18	Reconstruction of the Image	577
15.19	Important Properties of a Hologram	578
15.20	Types of Holograms	578
15.21	Applications of Holography	579
15.22	Formulae and Highlights	580
16. Optical Fibres		587
16.1	Introduction	587
16.2	The Evolution of Fibre Optic Systems	588
16.3	Advantages of Optical Fibres	588
16.4	Fundamentals of Optical Fibres	589
16.4.1	Core	589
16.4.2	Cladding	590
16.4.3	Jacket	590
16.5	Propagation of Light through Optical Fibre	590
16.6	Fractional Refractive Index Change	592
16.7	Numerical Aperture	592
16.8	Fibre Fabrication	597
16.9	Types of Fibres	597
16.9.1	Step Index Optical Fibre	598
16.9.2	Graded Index Optical Fibre	600

16.10	Numerical Aperture for Graded Index Optical Fibre	603
16.11	V-Number and Cut-off Parameters of Fibres	603
16.12	Comparison of Single-Mode and Multimode Index Fibres	607
16.13	Differences between Step Index and Graded Index Fibres	608
16.14	Propagation Mechanism in Optical Fibres	608
16.15	Communication through Optical Fibres	609
16.15.1	Propagation Mechanism of Meridional Rays	609
16.15.2	Propagation Mechanism of Skew Rays	610
16.16	Optical Fibre Transmission Link	612
16.17	Fibre Attenuation (Losses)	612
16.17.1	Absorption Losses	613
16.17.2	Scattering Losses	614
16.17.3	Bending Losses	616
16.17.4	Dispersion Losses	619
16.18	Attenuation Constant	623
16.19	Fibre Splices and Connectors	623
16.20	Optical Fibre Cables	624
16.21	Optical Cable Design	624
16.22	Applications of Optical Fibres	624
16.23	Formulae and Highlights	625

UNIT V : ELECTROMAGNETISM

17.	Maxwell's Equations and Electromagnetic Waves	635
17.1	Introduction	635
17.1.1	Scalar and Vector Fields	636
17.1.2	Gradient, Divergence, and Curl	636
17.1.3	Gauss Divergence Theorem(Relation between Surface and Volume Integration)	637
17.1.4	Stokes Theorem (Relation between Line Integral and Surface Integration)	637
17.1.5	Poisson's and Laplace's Equations	637
17.2	Fundamental Laws of Electricity and Magnetism	638
17.3	Equation of Continuity	639

17.4	Displacement Current	639
17.4.1	Modified Ampere's Law	640
17.5	Maxwell's Electromagnetic Equations	641
17.5.1	Maxwell's Equations in Differential Form	641
17.5.2	Maxwell's Equations in Integral Form	642
17.5.3	Derivation of Maxwell's Equations	642
17.6	Physical Significance of Maxwell's Equations	647
17.7	Electromagnetic Energy (Poynting Theorem)	648
17.8	Characteristic Impedance or Intrinsic Impedance, or Wave Impedance of the Free Space	650
17.9	Energy Flow in Plane Electromagnetic Wave	652
17.10	Energy Density in Plane Electromagnetic Wave in Free Space	653
17.11	Electromagnetic Wave in Free Space and its Solution	654
17.12	Solution of Plane Electromagnetic Wave: Transverse Nature of Electromagnetic Waves	655
17.13	Depth of Penetration: Skin Depth	657
17.14	Formulae and Highlights	667

UNIT VI : ENGINEERING MATERIALS

18.	Superconductivity	679
18.1	Introduction (A Chronological View)	679
18.2	Superconductivity	680
18.3	Temperature Dependence in Superconducting Materials	680
18.3.1	Temperature Dependence of Resistivity in Superconducting Materials	680
18.3.2	Temperature Dependence of Critical Field (Effect of External Magnetic Field) on Superconductors	681
18.4	Meissner Effect (Effect of Magnetic Field)	687
18.5	Critical Currents	687
18.6	Persistent Current	690
18.7	Types of Superconductors	690
18.7.1	Type I Superconductors or Soft Superconductors	690
18.7.2	Type II Superconductors or Hard Superconductors	691
18.8	Flux Quantisation	692

18.9	London Penetration Depth	692
18.10	BCS Theory (Quantum Mechanical Concept of Superconductivity)	695
18.11	Characteristics of Superconductors	696
18.12	Applications of Superconductors	696
18.13	High-Temperature Superconductors	697
18.14	Josephson Effect	698
18.15	Isotope Effect	699
18.16	Formulae and Highlights	699

19. Nanomaterials and Their Applications **709**

19.1	Introduction	709
19.2	Nanoscience and Nanotechnology	709
19.3	Significance of the Nanoscale	710
19.3.1	Surface Area	710
19.3.2	Quantum Confinement Effect	710
19.4	Fabrication of Nanomaterials	711
19.4.1	Top-Down Process	712
19.4.2	Bottom-Up Process	713
19.5	Quantum Size Effect in Metal or Semiconductor Nanoparticles	715
19.6	Different Types of Nanostructures: (Confinement Dimensions 0-D, 1-D, 2-D, and 3-D)	716
19.6.1	Quantum Wells	718
19.6.2	Quantum Wires	718
19.6.3	Quantum Dots	718
19.7	Unusual Properties of Nanomaterials	718
19.8	Categories of Nanomaterials	719
19.9	Creation and Uses of Buckyballs	719
19.9.1	Fullerene or Buckyball	720
19.9.2	Production of Buckyballs	721
19.9.3	Properties of Buckyballs	721
19.9.4	Uses of Buckyballs	721
19.10	Carbon Nanotubes (CNTs)	722
19.10.1	Types of Carbon Nanotubes	722
19.11	Applications of Nanomaterials	723
19.12	Formulae and Highlights	723

20. New Engineering Materials	729
20.1 Introduction	729
20.2 Metallic Glasses	729
20.3 Types of Metallic Glasses	730
20.3.1 Metal–Metalloid Type Metallic Glasses	730
20.3.2 Metal–Metal Type Metallic Glasses	730
20.4 Preparation of Metallic Glasses	730
20.5 Methods to Synthesise Metallic Glasses	731
20.5.1 Vapour-State Processes	731
20.5.2 Liquid-State Processes	731
20.5.3 Solid-State Processes	731
20.6 Properties of Metallic Glasses	732
20.7 Applications of Metallic Glasses	733
20.8 Bulk Metallic Glasses	733
20.9 Important Applications of Bulk Metallic Glasses	734
20.10 Distinction between Crystals and Glasses	735
20.11 Differences between Amorphous Alloys and Metallic Glasses	735
20.12 Shape Memory Alloys	735
20.13 Phase Change in Shape Memory Alloys	736
20.14 One-Way Memory Effect	738
20.14.1 Macroscopic Scale	738
20.14.2 Microscopic Scale	739
20.15 Two-Way Memory Effect	739
20.16 Pseudoelasticity or the Superelastic Effect	740
20.17 Nickel–Titanium Shape Memory Alloys	741
20.18 Copper-Based Shape Memory Alloys	742
20.19 Biomaterials	743
20.20 Types of Biomaterials	743
20.20.1 Metals	743
20.20.2 Polymers	744
20.20.3 Ceramics	744
20.20.4 Composites	744
20.20.5 Natural Biomaterials	744

20.21 Applications	744
20.22 Biomedical Compatibility of Ti-Al-Nb Alloys	746
20.23 Formulae and Highlights	746

UNIT VII : MODERN PHYSICS AND QUANTUM MECHANICS

21. Elementary Modern Physics	753
21.1 Introduction	753
21.2 Blackbody	753
21.2.1 Blackbody Radiation Spectrum	754
21.3 Wien's Displacement Law	755
21.4 Rayleigh-Jeans Law	755
21.5 Planck's Hypothesis and Radiation Law	755
21.6 Photoelectric Effect	756
21.6.1 Experimental Study	757
21.6.2 Fundamental Laws of Photoelectric Emission	757
21.6.3 Einstein's Photoelectric Equation	758
21.7 Formulae and Highlights	762
22. de Broglie Matter-Waves and Wave-Particle Duality	767
22.1 Introduction	767
22.2 Wave-Particle Duality	767
22.3 de Broglie Hypothesis of Matter-Waves	768
22.4 Expression for Wavelength of Matter-Wave (de Broglie Wavelength)	769
22.5 Properties of Matter-Waves	771
22.6 Detection of Matter-Waves	772
22.7 Nature of Electron	774
22.8 Davisson and Germer Experiment for Matter-Waves	774
22.9 Wave Velocity or Phase Velocity	796
22.10 Need For Wave Packet Representation	797
22.11 Wave Packet	798
22.12 Group Velocity	799
22.13 Expression for Group Velocity	799
22.14 Relation between Group Velocity and Wave Velocity	800
22.15 Relation between Group Velocity and Particle Velocity	801

22.16 Group Velocity of de Broglie Waves	802
22.17 Formulae and Highlights	808
23. Uncertainty Principle and Quantum Mechanics	819
23.1 Introduction	819
23.2 Basis for Uncertainty Principle	819
23.3 Heisenberg's Uncertainty Principle	820
23.3.1 Statement of the Uncertainty Principle	820
23.4 Derivation of Uncertainty Principle	821
23.5 Time-Energy Uncertainty Principle	822
23.6 Experimental Examples of Uncertainty Principle	823
23.6.1 Determination of the Position of a Particle by γ -ray Microscope	823
23.6.2 Diffraction of a Beam of Electrons by a Slit	825
23.6.3 Consequences of Uncertainty Principle	826
23.7 Applications of Uncertainty Principle	826
23.7.1 Non-existence of Electrons in the Nucleus	826
23.7.2 Existence of Protons, Neutrons, and α -particles in the Nucleus	827
23.7.3 Binding Energy of an Electron in an Atom	827
23.7.4 Zero-Point Energy of a Harmonic Oscillator	828
23.7.5 Radiation of Light from an Excited Atom	829
23.8 Wave Function (ψ)	841
23.9 Physical Significance of Wave Function (ψ)	841
23.10 Probability Current Density	842
23.11 Normalization of Wave Function	844
23.12 Schrödinger Wave Equation	845
23.13 Time-Independent Schrödinger Wave Equation	845
23.14 Time-Dependent Schrödinger Wave Equation	847
23.15 Eigenvalues and Eigenfunctions	848
23.16 Applications of Schrödinger Wave Equation	848
23.16.1 Particle in a One-dimensional Box	848
23.16.2 Physical Interpretation of Probability Density Distribution	852
23.17 Energy Eigenvalues of a Particle in a Potential Well of Infinite Depth	854
23.18 Energy Eigenvalues for a Free Particle	856
23.19 Formulae and Highlights	867

24. Thermodynamics	879
24.1 Introduction	879
24.2 Laws of Thermodynamics	879
24.2.1 Zeroth Law of Thermodynamics	879
24.2.2 First Law of Thermodynamics	880
24.2.3 Second Law of Thermodynamics	880
24.2.4 Third Law of Thermodynamics or Nernst Heat Theorem	885
24.3 Thermodynamic Processes	886
24.3.1 Reversible Process	887
24.3.2 Irreversible Process	888
24.4 Thermodynamic Variables	888
24.4.1 Macroscopic Quantities	888
24.4.2 Microscopic Quantities	888
24.5 Extensive and Intensive Variables	889
24.6 Thermodynamic Potentials and Their Interrelations	889
24.6.1 Internal or Intrinsic Energy (U)	889
24.6.2 Helmholtz Function (F)	890
24.6.3 Enthalpy or Heat Content (H)	891
24.6.4 Gibb's Free Energy (G)	892
24.6.5 Gibb's Helmholtz Relation	893
24.6.6 Entropy	894
24.6.7 Entropy with P-V Indicator Diagram	895
24.7 Heat Engine	899
24.8 Efficiency of the Heat Engine	900
24.9 Carnot Cycle	900
24.10 Efficiency of the Carnot Cycle	904
24.11 Carnot's Theorem	905
24.12 Temperature Entropy Diagram (T-S Diagram)	906
24.13 Maxwell's Thermodynamical Relations	910
24.14 Formulae and Highlights	912
25. Semiconductor Physics	927
25.1 Introduction	927
25.2 Energy Bands in Solids	927

25.3	Classification of Solids on the Basis of Energy and Diagram	928
25.3.1	Conductors (Metals)	929
25.3.2	Insulators	929
25.3.3	Semiconductors	930
25.4	Theory of Band Formation in Solids	930
25.5	B and Formation in Pure Semiconductor (Silicon)	931
25.6	Idea of Effective Mass of Electron	933
25.7	B and Structure in Semiconducting Materials with the Concept of Effective Mass	934
25.8	Density of States	935
25.8.1	Density of States for Valence Band and Conduction Band	937
25.9	Fermi Distribution Function	938
25.10	Fermi Level (EF) and its Importance	939
25.11	Carrier Concentration in Intrinsic Semiconductor	940
25.11.1	Concentration of Electrons in the Conduction Band	940
25.11.2	Concentration of Holes in the Valence Band	941
25.11.3	Temperature Dependence of Electron Concentration in Conduction Band	942
25.11.4	Effect of Doping on Band Structure and Carrier Concentration	943
25.12	Semiconductor Statistics (Estimation of Carrier Concentration)	944
25.12.1	Carrier Concentration Statistics for Intrinsic Region	945
25.12.2	Carrier Concentration Statistics for Extrinsic Region	945
25.12.3	Effect of Carrier Concentration on Band Structure	947
25.13	Mobility of Electron and Holes	948
25.14	Extrinsic Semiconductors	949
25.14.1	Donor or N-type Semiconductor	949
25.14.2	Acceptor or P-type Semiconductor	950
25.15	Energy Band Diagram and Position of Fermi Level in an Intrinsic Semiconductor	951
25.16	Energy Band Diagram and Position of Fermi Level in Extrinsic Semiconductor	952
25.17	Mobility	954
25.18	Electrical Conductivity	955

25.19 Expression for Conductivity of Semiconductor Materials	956
25.20 Resistivity	957
25.21 Experimental Determination of Band Gap of Intrinsic Semiconductor	958
25.22 Formulae and Highlights	965
26. Mechanical Properties of Matter	975
26.1 Introduction	975
26.2 Stress	975
26.3 Strain	976
26.3.1 Longitudinal or Tensile Strain	976
26.3.2 Shear or Shearing Strain	976
26.3.3 Volumetric Strain	976
26.4 Behaviour of a Wire under Increasing Load	977
26.5 Hooke's Law	978
26.6 Young's Modulus (Y)	978
26.7 Bulk Modulus (K)	978
26.8 Modulus of Rigidity or Shear Modulus	979
26.9 Poisson's Ratio	980
26.10 Relation between Elastic Constants	980
26.11 Shearing Stress Equivalent to an Equal Linear Tensile stress and Equal Compressive stress at right Angles to Each Other	980
26.12 Relation between Y, K and σ	982
26.13 Relation between Y, h and a	983
26.14 Relation between Y, k and η	984
26.15 Relation between k, η and σ	985
26.16 Limiting Values of Poisson's Ratio (σ)	985
26.17 Determination of Young's Modulus (Y)	985
26.18 Torsion Pendulum	987
26.19 Use of Torsion Pendulum	988
26.19.1 Determination of Moment of Inertia of an Irregular Body	988
26.20 Bending of Beam	989
26.21 Bending Moment	990
26.22 Expression for Bending Moment	990
26.23 Cantilever	991

xxviii ● Contents

26.24 Cantilever Supported at its Ends and Loaded in the Middle	994
26.25 Determination of Young's Modulus by Bending of a Beam	995
26.26 Formula and Highlights	1005

Appendix: Additional Topics

A.1 Gradient, Divergence, and Curl
A.2 Oscillation
A.3 Coupled Oscillations
A.4 Ferromagnetism
A.5 Weiss Law

Bibliography

1051

Index

1053

1

Theory of Relativity

1.1 INTRODUCTION

All the measurements in this universe are relative, for example, the length of a rod depends on the relative difference of its end coordinates. Similarly, the time interval depends on the relative difference between any two instants. In classical mechanics, bodies travel with the velocity much less than the velocity of light. In this domain of physics, the fundamental quantities, viz., space, time, and mass are absolute and invariant. It means that the length of an object is independent of the conditions under which it is measured and the time interval between any two events has the same value for all the observers irrespective of their state of motion. Similarly, the mass of the body does not depend on the velocity of the body.

After the failure of the classical theory for the explanation of many new experimental facts, Albert Einstein established the *special theory of relativity* in the beginning of twentieth century. According to this theory, the fundamental quantities, i.e., space, time, and mass no longer remain constant or invariant but depend on the state of motion. On the basis of the new concept of the theory of relativity, everything in this universe is relative; nothing is absolute. All the states of rest and motion are relative. Position, length, and time are relative. The theory of relativity predicts that the *Newtonian* mechanics is the limiting case of the *special theory of relativity*.

The theory of relativity deals with the way in which observers in relative motion describe the physical phenomenon with respect to one another. The special theory of relativity touches all the branches of modern physics, including high-energy physics, quantum theory, electromagnetic theory, and other branches of science and technology.

In this chapter, the Galilean theory of relativity is discussed briefly along with the *Michelson–Morely experiment*. The special theory of relativity is discussed in detail along with its important *consequences*.

1.2 CONCEPT OF THEORY OF RELATIVITY

The theory of relativity is mainly concerned with the correct measurement of things and *events*. Before realising the data regarding an event, we should be clear about, what is an event? The universe is full of dynamic objects, ranging from giant stars to tiny particles like electrons; nothing is static in this universe. Due to their dynamicity, the objects of the universe may interact in a variety of ways. The interactions of the objects are termed as *events* or *happenings*. Complete information of an event can be obtained by knowing the facts about where the event is occurring (position), what is its time of occurrence (when), and who is recording the data for it. A person who records the data for an event is known as an *observer*.

The data, or the information, for an event depends on the choice of the coordinate system and the observer. For example, an observer in Delhi would locate Taj in Agra with different coordinates as compared to another observer in Mumbai. Note that an event does not *belong* to a particular inertial reference frame. An *event* is just something that happens. Anyone in any reference frame may detect it and assign space-time coordinates to it. In order to understand the above discussion, let us take the following example:

Take a point on the rim of a moving wheel on the ground. For the observer standing on the ground, the path of the point will appear as a cycloid. However, for an observer sitting at the centre of the wheel, the path of the point will be a circle, as it is shown in Figs. 1.1(a) and (b), respectively.

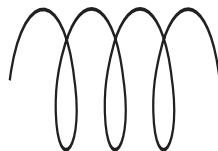


Fig. 1.1(a) Path of point on the rim when observer is on the ground



Fig. 1.1(b) Path of point on the rim when observer is at the centre of the wheel

1.3 FRAMES OF REFERENCE

In order to describe the motion of a body, we should essentially know with respect to what this motion has been measured. If we imagine a coordinate system attached to a rigid body and describe the position of any particle in space relative to it, then such a coordinate system is known as *frame of reference*. The location of a point in space can be described by a position vector drawn from the origin of the coordinate system to the point chosen in space. The position of the observer may or may not coincide with the origin of the coordinate system, but there is no relative motion between the origin of the coordinate system and the observer.

The simplest frame of reference is the Cartesian coordinate system. In this system, the position of a point is described by coordinates (x, y, z) at any instant t . The position vector drawn from the origin to the point is given as

$$\vec{r} = \hat{x}\hat{i} + \hat{y}\hat{j} + \hat{z}\hat{k}$$

For the complete information of any event, we require its position and the time of occurrence in a frame of reference. In this way, we require four coordinates (x, y, z, t) to describe the event. The reference frame employed for this purpose is known as space-time frame of reference. Such a frame of reference is shown in Fig. 1.2.

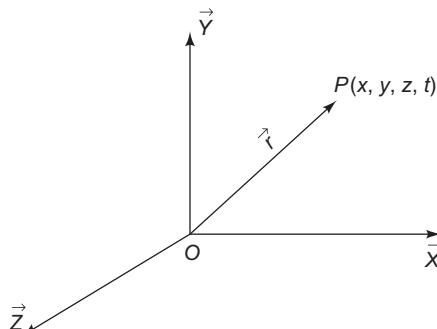


Fig. 1.2 Space-time frame of reference

There are two types of frames of reference:

- (i) Inertial or unaccelerated frames of reference
- (ii) Non-inertial or accelerated frames of reference

1.3.1 Inertial Frames of Reference

The inertial frames of reference are those unaccelerated frames of reference in which Newton's laws hold good. In these frames of reference, a body is not acted upon by an external force. It remains in its state of rest or uniform motion. In an inertial frame of reference, a body does not experience any external force; therefore, its acceleration a is given by

$$\vec{a} = \frac{d^2 \vec{r}}{dt^2} = 0 \quad (\text{Because } \vec{F} = m\vec{a} = 0) \quad (1.1)$$

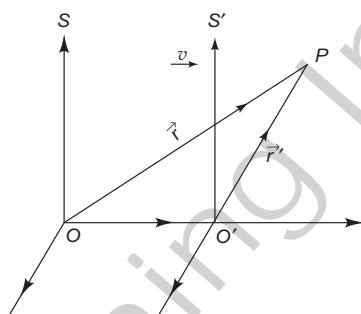


Fig. 1.3 S and S' are inertial frames of reference

For further discussion, let us consider an inertial frame of reference S and another frame of reference S' (Fig. 1.3). Now, S' is moving with constant velocity v with respect to S . Initially, at $t = t' = 0$, S and S' were coinciding with each other. At any instant t , the position vectors for point P observed by the observers O and O' , respectively, can be related by the following expressions:

$$\vec{r} = \overrightarrow{OO'} + \vec{r}'$$

$$= \vec{v}t + \vec{r}'$$

or $\vec{r}' = \vec{r} - \vec{v}t$

Differentiating the above equation with respect to t (twice), we get

$$\frac{d^2 \vec{r}'}{dt^2} = \frac{d^2 \vec{r}}{dt^2} \quad (\text{Because } v \text{ is constant})$$

$$a' = a$$

The above expression implies that the particle at point P experiences the same acceleration in both the frames; it means that if the acceleration of the particle in frame S is zero, then its acceleration in S' will also be zero. On the basis of the above discussion, we may conclude that if S is an inertial frame, then all those frames of reference that are moving with constant velocity with respect to it will be inertial frames of reference.

1.3.2 Non-inertial Frames of Reference

Those accelerated frames of reference, in which Newton's laws do not hold good are called *non-inertial frames of reference*. In non-inertial frames of reference, the body experiences an additional acceleration due to the accelerated motion of the frames of reference.

If the observed acceleration of a body of mass m in an inertial frame is a , then in such a frame, the force experienced by the body is given by

$$F = ma$$

But in the case of accelerated frames of reference, F is not equal to ma ; it means that the acceleration of the body in the non-inertial frames will be other than a .

The earth is an example of a non-inertial frame of reference because it is spinning about its axis, and simultaneously, it is moving in its orbit around the sun. Due to the rotation of the earth, the body which is at rest on the surface of the earth experiences centripetal acceleration, which cannot be neglected, in all the cases.

In order to differentiate the inertial and non-inertial frames of reference, we can take the following example:

Suppose that S is an inertial frame and another frame S' is moving with an acceleration a_0 relative to S . The acceleration of a particle P , on which no external force is acting, will be zero in the frame S , but in frame S' , the observer will find that an acceleration $-a_0$ is acting on it. Thus, in frame S' , the observed force on the particle is $-ma_0$, where m is the mass of the particle. Such a force, which does not really act on the particle but appears due to the acceleration of the frame, is called a *fictitious*, or *pseudo force*. Here, fictitious force on the particle P is $F = -ma_0$.

Hence, the accelerated frame S' is non-inertial with respect to the frame S because in this frame, Newton's law of inertia is not valid.

1.4 GALILEAN TRANSFORMATION EQUATIONS

When a point or an event is observed from two different frames, then it has different coordinates. There are certain equations that can relate these coordinates. These equations are known as *transformation equations*. The equations which relate the coordinates of two inertial frames of reference are known as *Galilean transformation equations*.

Case I In order to realise the Galilean transformation equations, let us consider an inertial frame of reference S and another frame of reference S' moving with constant velocity v with respect to S along the x -axis. For our convenience, we assume that initially (at $t = t' = 0$), S and S' are coinciding with each other. After time t , the observer (O) of frame S records the coordinates of point P as (x, y, z, t) , but at the same instant, the observer (O') of frame S' records the coordinates of point P as (x', y', z', t') [Fig. 1.4].

Now, the position vectors of the point P at any instant t can be related by the following equation:

$$\begin{aligned} \vec{r} &= \vec{r}' + \vec{vt} \\ \text{or } \vec{r}' &= \vec{r} - \vec{vt} \\ \text{or } x'\hat{i} + y'\hat{j} + z'\hat{k} &= x\hat{i} + y\hat{j} + z\hat{k} - vt\hat{i} \\ &= (x - vt)\hat{i} + y\hat{j} + z\hat{k} \end{aligned} \tag{1.2}$$

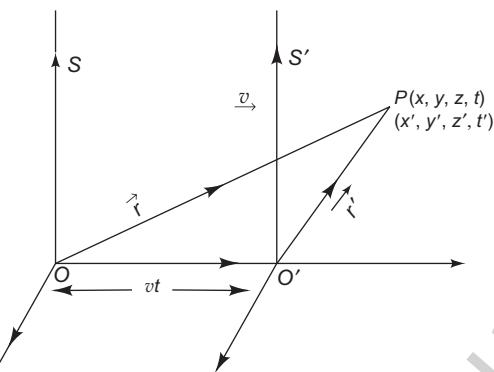


Fig. 1.4 Relative motion of S' with respect to S along x -axis

Expressing in the coordinate form, the above equation can be given as

$$\left. \begin{array}{l} x' = x - vt \\ y' = y \\ z' = z \\ t' = t \end{array} \right\} \quad (1.3)$$

Equation (1.3) gives the transformation equations of coordinates from one inertial frame to another, and hence, they are referred as the *Galilean transformation* equations.

Case II In this case, we will derive the transformation equations between the coordinates of two inertial frames of reference, where the motion of the second frame of reference is along all the directions (x, y, z) simultaneously (Fig. 1.5). Let S and S' be two inertial frames of reference in such a way that S' is moving with constant velocity v with respect to S having its components v_x, v_y , and v_z , along x -, y -, and z -axes, respectively. At any instant t , the observer of frame S records the coordinates of point P as (x, y, z, t) and the observer of frame S' records the coordinates of the same point as (x', y', z', t') .

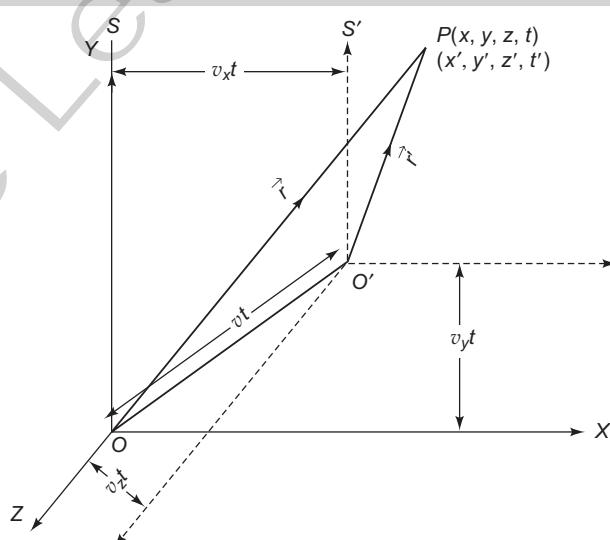


Fig. 1.5 Relative motion of S' with respect to S along all the axes

At the instant t , the radius vectors of point P from S and S' can be related as

$$\begin{aligned}\vec{r} &= \vec{r}' + \overrightarrow{OO'} \\ &= \vec{r}' + \vec{v}t\end{aligned}\quad \text{or} \quad \vec{r}' = \vec{r} - \vec{v}t \quad (1.4)$$

The above equation may be expressed in terms of the following equation:

$$\begin{aligned}x'\hat{i} + y'\hat{j} + z'\hat{k} &= x\hat{i} + y\hat{j} + z\hat{k} - v_x\hat{i}t - v_y\hat{j}t - v_z\hat{k}t \\ \text{or} \quad &= (x - v_x t)\hat{i} + (y - v_y t)\hat{j} + (z - v_z t)\hat{k}\end{aligned}$$

Comparing similar terms of both sides, we get

$$\left. \begin{array}{l} x' = x - v_x t \\ y' = y - v_y t \\ z' = z - v_z t \\ t' = t \end{array} \right\} \quad (1.5)$$

Equation (1.5) gives the Galilean transformation equations for such frames of reference which have their relative motion along all the three axes simultaneously.

1.4.1 Consequences of Galilean Transformation Equations

Important consequences of the Galilean transformation equations are as follows:

- (i) *Transformation of length:* Let S and S' be two inertial frames of reference in which S' is moving with velocity v relative to S along the x -axis. If the coordinates of ends of a rod are (x_1, y_1, z_1) and (x_2, y_2, z_2) in frame S , and (x'_1, y'_1, z'_1) and (x'_2, y'_2, z'_2) in frame S' , respectively, then the length of the rod can be given as

$$L = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

When the rod is observed by an observer of frame S' , then its end coordinates will be (x'_1, y'_1, z'_1) and (x'_2, y'_2, z'_2) , respectively. The length of the rod in S' can be given as

$$\begin{aligned}L' &= \sqrt{(x'_2 - x'_1)^2 + (y'_2 - y'_1)^2 + (z'_2 - z'_1)^2} \\ &= \sqrt{(x_2 - vt - x_1 + vt)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}\end{aligned}$$

Using the Galilean transformation equations in the above expression, we get

$$L' = L$$

It means that the length of the rod remains invariant under the Galilean transformation.

- (ii) *Transformation of velocity (concept of relative velocity):* As discussed in Section 1.4, the position vectors of a point P at any instant, measured from different frames, can be related by the equation

$$\vec{r}' = \vec{r} - \vec{v}t$$

Differentiating the above equation with respect to t , we get

$$\frac{d\vec{r}'}{dt} = \frac{d\vec{r}}{dt} - v \quad (1.6)$$

or $u' = u - v$

where u' and u are the velocities of the body observed in the moving and stationary frames of reference, respectively, and v is the relative velocity between S and S' . Equation (1.6) is also known as the *Galilean law of addition of velocity* because the velocity observed in the stationary frame is equal to the sum of the velocity observed in the moving frame and the velocity by which it is moving.

(iii) *Transformation of acceleration:* From Eq. (1.6), we know that $u' = u - v$.

Differentiating it, we get

$$\frac{du'}{dt} = \frac{du}{dt} - 0 \quad (\because v \text{ is constant})$$

i.e., $a' = a$ (1.7)

In this way, the acceleration is invariant under the Galilean transformation.

1.4.2 Newtonian Principle of Relativity

On the basis of fundamental concepts and experimental observations, Newton pointed out that the basic laws of physics remain identical in all the inertial frames of reference. In other words, we can say that the basic laws of physics remain identical in form in those frames of reference that are connected by the Galilean transformation.

Solved Examples

Example 1.1

Calculate the fictitious force and the total force on a body of mass 6 kg if the frame of reference is moving (i) vertically upwards and (ii) vertically downwards, with an acceleration of 5 m/s^2 .

Solution

$$\begin{aligned}\text{Weight of the body} &= mg \\ &= 6 (-9.8) = -58.8 \text{ N} \\ &= 58.8 \text{ N downwards}\end{aligned}$$

- (i) The fictitious force acting on it, when the frame of reference is moving in upward direction is given as

$$\begin{aligned}-ma_0 \\ = -6 \times 5 = -30 \text{ N} \\ = 30 \text{ N downwards}\end{aligned}$$

Hence, the total force $= 58.8 + 30 = 88.8 \text{ N}$ downwards, i.e., the body appears to be heavier.

- (ii) The fictitious force acting on the body during the downward motion

$$\begin{aligned}= -ma_0 \\ = 6 \times (-(-5)) \\ = 30 \text{ N upwards}\end{aligned}$$

The net force experienced by the body $= 58.8 - 30 = 28.8 \text{ N}$ downwards so that it seems to be lighter.

Example 1.2

The earth, revolving around its own axis, is not an inertial frame of reference. If, however, we take it to be so, what would be the error involved in 1 s in the position of a particle close to its surface? (Given that radius of the earth = 6.4×10^8 cm.)

Solution

The particle which is close to the surface of the earth will experience a centripetal acceleration due to the rotation of the earth about its own axis, whose value can be given as

$$\begin{aligned}\alpha &= \omega^2 R \\ &= \left(\frac{2\pi}{24 \times 60 \times 60} \right)^2 \times 6.4 \times 10^8 \text{ cm/s}^2 \\ &= 3.4 \text{ cm/s}^2 \quad \left(\text{because } \omega = \frac{2\pi}{T} \right)\end{aligned}$$

Thus, the acceleration of the particle will differ by 3.4 cm/s^2 when it will be measured in true inertial frame of reference. Due to this acceleration on the surface of the earth, the particle will be altered by a distance $1/2 at^2$ in time t .

Now, the altered distance S for the particle in 1 s can be given as

$$\begin{aligned}S &= \frac{1}{2} at^2 \\ &= \frac{1}{2} \times 3.4 \times (1)^2 \\ &= 1.7 \text{ cm}\end{aligned}$$

Hence, the error involved in the position of the particle measured in earth's frame will be 1.7 cm in 1 s.

Example 1.3

Using the Galilean transformation, show that the distance between two points (x_1, y_1, z_1) and (x_2, y_2, z_2) remains invariant in two inertial frames.

Solution

Let S' be a frame of reference which is moving with constant velocity v in x -direction relative to a stationary frame S .

Let the coordinates of two points in frame S be (x_1, y_1, z_1) and (x_2, y_2, z_2) , while those in frame S' be (x'_1, y'_1, z'_1) and (x'_2, y'_2, z'_2) .

From the Galilean transformation, we have

$$\begin{array}{l}x'_1 = x_1 - vt, \quad y'_1 = y_1, \quad z'_1 = z_1 \\ \text{and} \quad x'_2 = x_2 - vt, \quad y'_2 = y_2, \quad z'_2 = z_2\end{array}$$

Distance measured in frame S is

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

The distance between these two points in frame S' can be given as

$$d' = \sqrt{(x'_2 - x'_1)^2 + (y'_2 - y'_1)^2 + (z'_2 - z'_1)^2}$$

Substituting the values of $x'_1, y'_1, z'_1, x'_2, y'_2$, and z'_2 in the above equation, we get

$$\begin{aligned} d' &= \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \\ &= d \end{aligned}$$

Hence, the distance between two points remains invariant under the Galilean transformation.

1.5 CONCEPT OF “LUMINIFEROUS ETHER” MEDIUM

The necessity of a medium for the propagation of mechanical waves forced the physicists of the nineteenth century to think about the existence of a medium essential for the propagation of light and other electromagnetic waves in the free space. Therefore, they assumed the presence of ether medium in the entire universe including vacuum. This hypothetical universal ether medium is rigid, invisible, massless, perfectly transparent, perfectly non-resistive, continuous, and stationary solid like steel having a very high elasticity and negligible density. All the objects including the earth, moon, stars, etc., may move freely through this hypothetical ether medium without any disturbance. Thus, it may be concluded that ether provides a universal fixed frame of reference or absolute frame of reference with respect to which all the measurements can be done.

Prior to the twentieth century, the ether hypothesis was an established concept for the successful explanation of the phenomena of interference and diffraction. But this hypothesis was questioned after the failure of the Galilean relativity for the laws of electrodynamics. This interrogative remark on this hypothesis suggests the need of special theory of relativity.

1.6 NEED OF SPECIAL THEORY OF RELATIVITY

The laws of mechanics remain invariant under the Galilean relativity but the laws of electrodynamics do not preserve their form under this theory. The failure of the Galilean theory of relativity in the case of electromagnetism suggests the facts as mentioned below:

- (i) The principle of relativity is meant only for the laws of classical mechanics. It is not applicable for the laws of electromagnetism.
- (ii) The principle of relativity is valid for the laws of classical mechanics as well as for the laws of electrodynamics but the laws of electromagnetism need modification in accordance with the Galilean transformations.
- (iii) The principle of relativity is valid for the laws of classical mechanics as well as for the laws of electromagnetics but Newton's laws of classical mechanics need modification in accordance with the Galilean transformations (Galilean theory of relativity).
- (iv) There is the presence of universal ether medium that provides absolute frame of reference with respect to which all the measurements are performed.

A number of experiments were conducted by different groups of scientists to check the validity of the above options. In this regard, the experiments performed by Fizeau, Michelson and Morely, and Trouton and Noble are of special importance. The results of these experiments are almost identical and discard the presence of ether medium; these results suggest that the speed of light is constant and independent of direction.

1.7 MICHELSON-MORELY EXPERIMENT

According to the electromagnetic theory, the speed of light in vacuum is c (3×10^8 m/s) in all directions. Now, a natural question arises: with respect to which frame of reference is this statement true? If we are considering an inertial frame of reference in which an observer is moving with speed v opposite to or along the direction of the propagation of light, then the speed of light in the frame of observer will be given as

$$c_r = c \pm v$$

It means, there should be an inertial frame in which the speed of light is c in all the directions or there should be an absolute frame of reference. Any frame of reference which has relative motion with respect to the absolute frame is not suitable for this purpose. Therefore, the universal ether medium is proposed for the absolute frame of reference. The Michelson–Morely experiment was performed to determine the speed of earth with respect to this frame so that the effect of observer's motion on the speed of light could be known.

The main objective of the Michelson–Morely experiment was to confirm the presence of a hypothetical ether medium. They were interested to detect the relative motion between the earth and the ether medium. Michelson and Morely repeated their experiment many times in different seasons at different places but they got no relative motion between the earth and the ether. Since it was against the hypothesis of the ether medium, so they declared it as negative result.

A schematic diagram of Michelson–Morely experiment is shown in Fig. 1.6 in which S is the source of monochromatic light, P is the half-silvered glass plate inclined at 45° , M_1 and M_2 are the plane mirrors, which are placed in horizontal and perpendicular directions, respectively, at the same distance from the glass plate P , and T is the telescope used to see the interference pattern due to the superposition of light beam reflecting back from mirrors M_1 and M_2 .

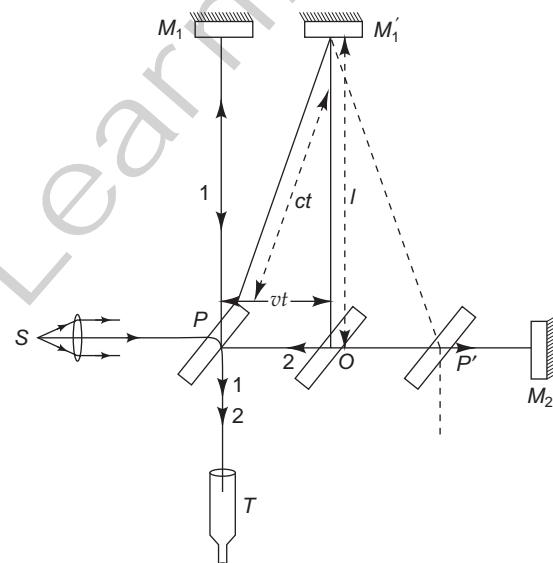


Fig. 1.6 Schematic diagram of Michelson–Morely experiment

A collimated beam of light from the source S is allowed to incident on the half-silvered glass plate P which is inclined at 45° with respect to the incident beam. At the glass plate P , the light beam splits into two beams. One is reflected in perpendicular direction, while the second one is transmitted in horizontal

2

Oscillation and Waves

2.1 INTRODUCTION

To and fro motion of a particle about the equilibrium position is known as *oscillation*. Any motion that repeats itself continuously on a definite path in a definite interval of time is known as *periodic motion*, and this interval of time is called the *time period*. Harmonic oscillations are the periodic motions which are very common in nature. Examples of this motion are motion of the Earth around the Sun, motion of the Moon around the Earth, oscillation of the pendulum of a clock, a mass attached to a spring, and motion of atoms within molecules or in a solid lattice. These oscillations are due to the change in mechanical quantities such as displacement, velocity, acceleration, and density. Thus, these oscillations belong to the mechanical system. The oscillations belonging to electrical system are oscillations of electric and magnetic field vectors in visible light, radio waves, microwaves, etc. A tuned circuit and closed metal cavity oscillate electromagnetically; thus, these are the examples of electrical system. Although these oscillations belong to different systems, these are described by the same fundamental principle and mathematical equations.

Essential condition for a particle to execute a simple harmonic motion is given as $F \propto -y$, where F is the restoring force and y is the displacement. The model of harmonic oscillator is helpful in understanding important phenomenon of physics, e.g., elasticity, acoustics, optics, solid state physics, vibration of crystal lattices, and atomic and molecular physics. Many complex vibrations occurring in physical problems such as vibrations of string or membrane and the vibration of atoms in solids can also be described by simple harmonic vibrations. In fact it is difficult to realise exact simple harmonic motions in the nature. It is because the linear dependence of restoring force and displacement do not hold good in actual systems. Thus, the systems can be treated as simple harmonic oscillators only within certain limits.

In this chapter, we will discuss basic features of different oscillatory systems such as simple harmonic motion, damped harmonic oscillation, forced vibration, and coupled oscillation. We will also focus on qualitative discussion on waves having periodic variation in space and time, wave equation, and reflection and transmission of waves at the boundary of two media.

2.2 SIMPLE HARMONIC MOTION

An oscillatory motion in which the amplitudes are equal on both sides of equilibrium position is known as *simple harmonic motion*. In order to understand it qualitatively, let us consider a particle P moving around a circle with a uniform speed v and N be the foot of perpendicular drawn from its instantaneous

position to the diameter YY' of the circle. A complete oscillation can be described according to the motion of particle on the circumference of the circle along with the movement of N on the diameter YY' , as shown in Fig. 2.1. During anti-clockwise motion of particle P , the foot of perpendicular changes its position at YY' . Let the motion of particle start from X , i.e., initially the particle is at X , and N is at O . When the particle reaches at Y , the foot of perpendicular coincides with its position at Y . As P moves along the circle and reaches from Y to X' , N moves from Y and reaches at O . Further when P travels from X' to Y' , N travels from O to Y' . Finally, when P moves from Y' to X , N reaches at O to complete one revolution. It is clear from the above discussion that during the time in which P completes one revolution, N travels through the diameter YY' twice. This oscillatory motion of N about centre O is called simple harmonic motion, which is defined as the projection of a uniform circular motion on the diameter of a circle.

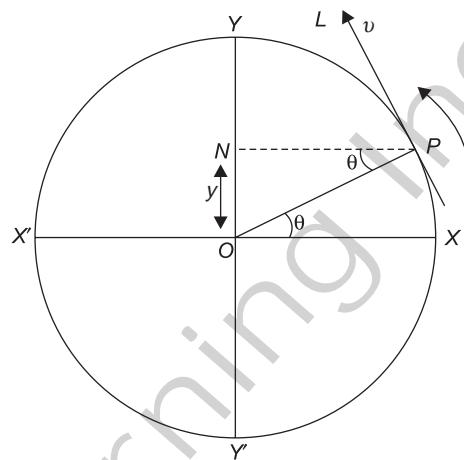


Fig. 2.1 Simple harmonic motion of particle P

If a is the radius of the circle, then θ is the angle made by radius vector by joining instantaneous position of particle and OX . From Fig. 2.1, it is clear that ON is the displacement of N during which particle moves from X to P . This displacement can be given as

$$y = a \sin \theta \quad (2.1)$$

If t is the time during which OP travels through an angle θ , then $\theta = \omega t$, and Eq. (2.1) can be written as

$$y = a \sin \omega t \quad (2.2)$$

Radius a is the maximum displacement on either side of O and is called the amplitude of motion.

2.2.1 Velocity of Motion

The velocity of N can be given by the component of the velocity of P parallel to YY' . From Fig. 2.1, it is clear that the velocity of P is v along PL , i.e., along the tangent of the circle at point P . Its component parallel to YY' can be given as $v \cos \theta$. Now the velocity of N can be given as

$$v_N = v \cos \theta$$

[Because v_N will be perpendicular to PN and is given as $v \sin \angle LPN = v \sin (90 - \theta) = v \cos \theta$]

$$\begin{aligned} &= v \cos \omega t \\ &= v (1 - \sin^2 \omega t)^{1/2} \end{aligned}$$

From Eq. (2.1), $\sin \omega t = \frac{y}{a}$

$$\text{Thus, } v_N = v \left(1 - \frac{y^2}{a^2} \right)^{1/2}$$

$$= \frac{v}{a} (a^2 - y^2)^{1/2}$$

Since $v = a\omega$,

$$v_N = \omega \sqrt{(a^2 - y^2)} \quad (2.3)$$

From Eq. (2.3), it is clear that the velocity of N varies according to the displacement, i.e., the distance of N from O . When the displacement is zero, v_N is maximum, while it is minimum, when y is maximum. This indicates that the velocity of a particle executing simple harmonic motion will be maximum at the mean position and will be zero at the extreme positions.

2.2.2 Acceleration

We know that in circular motion, the centripetal acceleration is v^2/a and is directed towards the centre along PO (see Fig. 2.1). Acceleration of N is given as the component of v^2/a along NO . Now the acceleration of N , i.e., a_N can be given as

$$\begin{aligned} a_N &= -a\omega^2 \sin \theta \\ &= -\omega^2 a \sin \theta \\ \text{or} \quad a_N &= -\omega^2 \cdot y \end{aligned} \quad (2.4)$$

where $y = a \sin \theta$.

Negative sign in Eq. (2.4) shows that the acceleration acts opposite to the direction in which y increases. According to Eq. (2.4), a body is in simple harmonic motion, if its acceleration of motion ($\omega^2 y$) is always directed towards a fixed point at the centre and is directly proportional to its distance from it.

2.2.3 Time Period

The time taken by the moving particle P to complete one oscillation is called *time period* or *periodic time*. If ω is the angular velocity of motion, then by definition

$$\omega = \frac{\text{Angle described in one-revolution}}{\text{Time of one complete revolution}}$$

$$= \frac{2\pi}{T}$$

$$\text{or } T = \frac{2\pi}{\omega} \quad (2.5)$$

From Eq. (2.4),

$$\omega = \sqrt{\frac{a_N}{y}} = \sqrt{\frac{\text{Acceleration}}{\text{Displacement}}}$$

By putting the value of ω in Eq. (2.5), we get

$$T = 2\pi \sqrt{\frac{\text{Displacement}}{\text{Acceleration}}} \quad (2.6)$$

2.2.4 Frequency

Frequency is defined as the number of oscillations made by a moving body in one second. It is expressed as the reciprocal of periodic time, i.e.,

$$n = \frac{1}{T} = \frac{\omega}{2\pi} \quad (2.7)$$

where n is frequency.

2.2.5 Phase

The phase of an oscillating particle determines its position and direction of motion at a particular instant.

If at any instant, two vibrating particles are passing simultaneously through their equilibrium positions in the same direction, then at that instant they are in *same phase*, and if they are passing in opposite directions, then they are in *opposite phase*.

Suppose the time is measured from the instant, when the particle P is at P_0 moving on the reference circle (Fig. 2.2), then $P_0Ox = \phi$, and the displacement y (ON) of N , from the equilibrium position O after a time t , will be given as

$$y = ON = OP \sin(\omega t + \phi) \\ y = a \sin(\omega t + \phi) \quad (2.8)$$

where a is the radius of the circle and ωt is the angle through which the particle P has rotated in moving from P_0 to P in time t . Hence, at the instant t , the phase of N will be measured by the total angle $(\omega t + \phi)$, where ϕ is known as the initial phase or epoch of N . Equation (2.8) is known as the *general equation of a simple harmonic motion*.

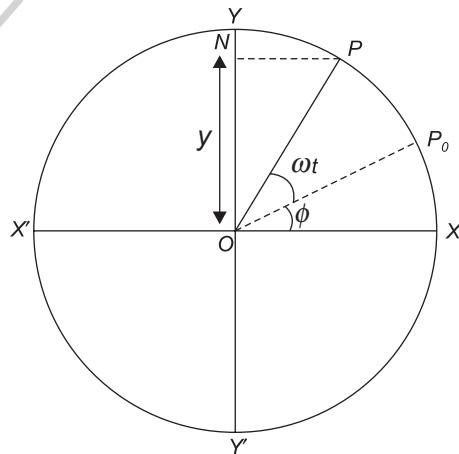


Fig. 2.2 P_0 and P are two different positions of a particle depicted at different instants

2.3 DAMPED HARMONIC OSCILLATIONS

A body executing single harmonic oscillations has constant amplitude only under ideal conditions. It is possible only when there is no dissipative or damping force acting on the body, thrown to execute simple harmonic motion. In all practical cases, it is observed that the vibrating system experiences a resisting force which causes decay of its amplitude.

If the bob of a pendulum is displaced slightly from its mean position, it starts oscillating. The oscillation continues with its natural frequency but the amplitude gradually dies away and finally comes to rest. Decay in the amplitude of oscillation is due to the various resisting forces opposing the motion of the body, which diminishes the amplitude or damps the oscillation. Such vibrations with decreasing amplitude are called damped oscillations. The energy given to the body initially to begin the oscillations is used up in doing work against damping forces. Thus, the energy given to the system decreases with time, and finally the system comes to rest.

2.3.1 Wave Equation of Damped Vibration

If y is the instantaneous displacement from the equilibrium position of a body of mass m , performing damped simple harmonic oscillation, then the forces acting on the body are as follows:

- (i) **Restoring force:** The tendency of this force is that it always tries to bring the body (executing oscillation) back to its mean position and is proportional to the displacement. Thus, it may be written as

$$\begin{aligned} F_R &\propto -y \\ \text{or} \quad F_R &= -ky \end{aligned} \tag{2.9}$$

- (ii) **Damping force:** It is the retarding force proportional to the velocity acting in the opposite direction of motion. Thus, it may be written as

$$\begin{aligned} F_D &\propto -\frac{dy}{dt} \\ \text{or} \quad F_D &= -b \frac{dy}{dt} \end{aligned} \tag{2.10}$$

- (iii) **Force of inertia:** Each body ready to set itself for oscillatory motion has inertia, which can be given as

$$\begin{aligned} F_I &\propto \frac{d^2y}{dt^2} \\ \text{or} \quad F_I &= m \frac{d^2y}{dt^2} \end{aligned} \tag{2.11}$$

At equilibrium, the force of inertia is balanced by restoring force and damping force. Hence, the equation of motion of a body executing damped harmonic vibrations is given as

$$m \frac{d^2y}{dt^2} = -ky - b \frac{dy}{dt}$$

$$\text{or } \frac{d^2y}{dt^2} + \frac{b}{m} \frac{dy}{dt} + \frac{k}{m} y = 0$$

Putting $b/m=2r$ and $k/m = \omega^2$, we obtain

$$\frac{d^2y}{dt^2} + 2r \frac{dy}{dt} + \omega^2 y = 0 \quad (2.12)$$

This is a homogenous linear differential equation of second order for damped harmonic oscillator. The general solution of eq. (2.12) is given as

$$y = a_0 e^{-rt} \sin(\omega t + \phi) \quad (2.13)$$

$$\text{where } \omega = \sqrt{\frac{k}{m} - \left(\frac{b}{2m}\right)^2} \quad (2.14)$$

$$\text{or } \omega = \sqrt{\omega_0^2 - \left(\frac{b}{2m}\right)^2} \quad (2.15)$$

where $\omega_0 = \sqrt{\frac{k}{m}}$ is the natural frequency and ω is the frequency of damped oscillation.

2.3.2 Analysis of Damped Harmonic Oscillator

In order to analyse damped harmonic oscillations

Let us start with equation (2.12) i.e.

$$\frac{d^2y}{dt^2} + 2r \frac{dy}{dt} + \omega_0^2 y = 0$$

where $2r = \frac{b}{m} = \frac{1}{\tau}$ and $\omega_0 = \sqrt{\frac{k}{m}}$ is the natural frequency in absence of damping forces. τ is the relaxation time and r is damping constant.

To solve equation (2.12) let its possible solution be

$$y = A e^{\alpha t} \quad (2.16)$$

using the value of y in equation (2.12) we get

$$A(\alpha^2 + 2r\alpha + \omega_0^2) e^{\alpha t} = 0$$

$$\text{or } \alpha^2 + 2r\alpha + \omega_0^2 = 0$$

Now the roots of α can be given as

$$\alpha = -r \pm \sqrt{r^2 - \omega_0^2}$$

Thus the solution of equation (12.2) are

$$y = A_1 e^{\left(-r+\sqrt{r^2-\omega_0^2}\right)t} \text{ and } y = A_2 e^{\left(-r-\sqrt{r^2-\omega_0^2}\right)t}$$

Hence the general solution of equation (12.2) can be given as

$$\begin{aligned} y &= A_1 e^{\left(-r+\sqrt{r^2-\omega_0^2}\right)t} + A_2 e^{\left(-r-\sqrt{r^2-\omega_0^2}\right)t} \\ \text{or } y &= e^{-rt} \left(A_1 e^{\left(\sqrt{r^2-\omega_0^2}\right)t} + A_2 e^{\left(-\sqrt{r^2-\omega_0^2}\right)t} \right) \end{aligned} \quad (2.17)$$

where A_1 and A_2 are constants, can be determined with initial conditions.

Depending on the values of quantity $\sqrt{r^2 - \omega_0^2}$ we may reach at different conditions of damping oscillations. According to the values of r and ω_0 this quantity may have imaginary, real and zero values. Corresponding to the different values of $\sqrt{r^2 - \omega_0^2}$, the conditions of damped oscillations are discussed below.

1. Under damped case: This condition occurred when $r < \omega_0$

$$\text{it means } \sqrt{r^2 - \omega_0^2} = \sqrt{-1(\omega_0^2 - r^2)} = i\sqrt{\omega_0^2 - r^2} = i\omega \text{ (say)}$$

where $i = \sqrt{-1}$ and $\omega = \sqrt{(\omega_0^2 - r^2)}$ which is a real quantity. Now equation (2.17) can be given as

$$\begin{aligned} y &= e^{-rt} \left[A_1 e^{i\omega t} + A_2 e^{-i\omega t} \right] \\ &= e^{-rt} \left[A_1 (\cos \omega t + i \sin \omega t) + A_2 (\cos \omega t - i \sin \omega t) \right] \\ &= e^{-rt} \left[(A_1 + A_2) \cos \omega t + i(A_1 - A_2) \sin \omega t \right] \end{aligned}$$

since y is a real quantity, $(A_1 + A_2)$ and $i(A_1 - A_2)$ must be real quantities. Clearly, A_1 and A_2 are complex quantities. If $A_1 + A_2 = a_0 \sin \phi$ and $i(A_1 - A_2) = a_0 \cos \phi$ then

$$y = a_0 e^{-rt} \sin(\omega t + \phi) \quad (2.18)$$

This equation is same as equation (2.13) represents a damped harmonic motion, whose periodic time is given by

$$T = \frac{2\pi}{\omega} = \frac{2\pi}{\sqrt{\omega_0^2 - r^2}} \quad (2.19)$$

It is observed that the effect of damping have negligible effect on time period except few extreme cases,

The amplitude of the oscillatory motion equation (2.18) is given by $a = a_0 e^{-rt} = a_0 e^{-t/2\tau}$, where a_0 is the amplitude in the absence of damping. In case of damping the amplitude decreases exponentially with time. Displacement curve for damped harmonic motion is shown in Fig. 2.3.

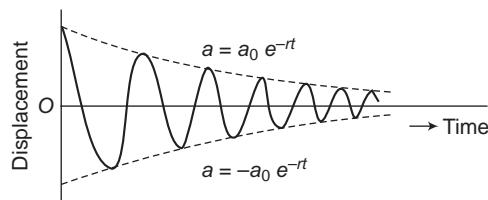


Fig. 2.3 Damped harmonic motion

- Over damped case: If damping is high i.e. $r > \omega_0$ then $\sqrt{r^2 - \omega_0^2} = \beta$ (say), is a real quantity and then from equation (2.17) we have

$$y = (A_1 e^{-(r-\beta)t} + A_2 e^{-(r+\beta)t}) \quad (2.20)$$

As $r > \beta$, both quantities of right hand side decreases exponentially with time and motion is non-oscillatory. Such motion is called dead beats or *aperiodic* and its main application is in dead beat galvanometers. Highly damped oscillation is shown in Fig. 2.4.

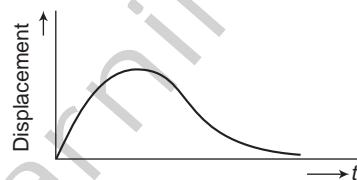


Fig. 2.4 Highly damped oscillation

- Critically damped Case: This case occurs at $r = \omega_0$, at this condition equation (2.17) gives us

$$y = (A_1 + A_2) e^{-rt} = c e^{-rt} \text{ where } c = A_1 + A_2$$

Since there is only one constant, thus it does not provide us the solution of equation (2.12) of second order.

Now let us consider

$$\sqrt{r^2 - \omega_0^2} = h \text{ which is very small}$$

Thus, equation (2.17) gives us

$$\begin{aligned} y &= e^{-rt} [A_1 e^{ht} + A_2 e^{-ht}] \\ &= e^{-rt} [A_1 (1 + ht + \dots) + A_2 (1 - ht + \dots)] \end{aligned}$$

Due to small value of h , higher term will be negligible.

$$\text{Hence } y = e^{-rt} [(A_1 + A_2) + h(A_1 - A_2)t]$$

$$\text{or } y = e^{-rt} (P + Qt) \quad (2.21)$$

where $P = A_1 + A_2$ and $Q = (A_1 - A_2)$

Also from the expression $\frac{dx}{dt}$ we can write

$$\frac{dy}{dt} = e^{-rt} Q + e^{-rt} (-r)(P + Qt)$$

Initially at $t = 0$ the displacement of particle is $y = y_0$ and then velocity is v_0 , using these values in equation (2.21) we get

$$y_0 = P \text{ and } v_0 = Q - rP = Q - ry_0 \text{ or } Q = v_0 + ry_0$$

$$\text{Thus } y = [y_0 + (v_0 + ry_0)t]e^{-rt} \quad (2.22)$$

From equation (2.22) it is clear that initially displacement increases due to the factor $y = e^{-rt} \sin(\omega t + \phi)$. But as time elapses, exponential term becomes dominating and displacement returns to zero. Such motion is known as critically damped or just *aperiodic*. Figure 2.5 shows the critically damped oscillation.

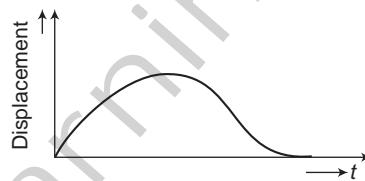


Fig. 2.5 Critically damped oscillation

2.3.3 Important Parameters under Damped Harmonic Oscillations

The secrets of damped harmonic oscillations can be observed in its different parameters as discussed below:

1. **Logarithmic Decrement (λ):** As it is mentioned in equation (2.18) the solution of damped harmonic oscillations can be given as

$$y = e^{-rt} \sin(\omega t + \phi)$$

The maximum value of $\sin(\omega t + \phi)$ is alternately +1 and -1, obviously the time displacement curve of vibrating body lies entirely between the curves $a = a_0 e^{-rt}$ and $a = -a_0 e^{-rt}$ as shown in Fig. 2.3.

If a_n and a_{n+1} are the successive amplitudes then

$$a_n = a_0 e^{-rt} \text{ and } a_{n+1} = a_0 e^{-r(t+T/2)}$$

$$\text{So } \frac{a_n}{a_{n+1}} = \frac{e^{-rt}}{e^{-r(t+T/2)}} = e^{rT/2} = d \text{ (say)} \quad (2.23)$$

where d is a constant for motion, called as decrement, indicating for the reduction in amplitude.

$$\text{Now, } \log_e \left(\frac{a_n}{a_{n+1}} \right) = \log_e d = \frac{rT}{2} = \lambda \quad (2.24)$$

The quantity λ is known as logarithmic decrement and is equal to the ratio of the natural logarithm of the ratio of two successive amplitudes of vibration.

- 2. Damping Coefficient (r):** The value of damping coefficient may be obtained directly from equation (2.24), as $\lambda = \frac{rT}{2} = \frac{r}{2v}$ (where v is the frequency)

$$\text{or } r = 2\lambda v$$

again from equation (2.24) above expression can be given as

$$r = 2.303 \cdot 2v \log_{10} \left(\frac{a_n}{a_{n+1}} \right) \quad (2.25)$$

Thus by knowing v and ratio $\left(\frac{a_n}{a_{n+1}} \right)$, r can be calculated

- 3. Relaxation Time (τ):** Relaxation time is defined as a time in which the amplitude, velocity or total mechanical energy of oscillation falls to $\frac{1}{e}$ (0.368) of its original value. It is denoted by τ .

Sometimes it is also termed as Time constant, Mean life time, Modulus of decay or mean decay time.

According to the definition we may write

$$\text{amplitude after time } (t + \tau) = \frac{1}{e} \text{ amplitude after time } t$$

$$\text{or } \frac{1}{e} = \frac{\text{amplitude after time } (t + \tau)}{\text{amplitude after time } t}$$

$$\text{or } \frac{1}{e} = \frac{a_0 e^{-r(t+\tau)}}{a_0 e^{-rt}}$$

$$\text{or } \frac{1}{e} = e^{-r\tau} = \frac{1}{e^{r\tau}}$$

$$\text{or } r\tau = 1$$

$$\text{or } \tau = \frac{1}{r} = \frac{2m}{b} \quad (2.26)$$

Thus in terms of relaxation time, the amplitude is written as $a_0 e^{-t/\tau}$

3

Ultrasonic Waves

3.1 INTRODUCTION

The sound waves of frequency in the range of 20 Hz to 20 kHz stimulate the human ear and brain to the sensation of hearing. These waves are known as *sonic waves* or *audible waves*. The sound waves having frequency greater than 20 kHz are called *ultrasonic* or *supersonic waves*. Ultrasonic waves have many important applications in the fields of metallurgy, medicines, biology, and other branches of science and technology.

In this chapter, we will discuss the basic aspects of ultrasonic waves, their production, detection, and important applications.

3.2 ULTRASONIC WAVES

We know that sound waves are caused by the vibrations of same or different kinds of particles. Sound waves can be generated in all three forms of matter, viz., solid, liquid, and gas. On the basis of their frequencies, these waves can be broadly divided into the following three classes:

- (i) *Subsonic or infrasonic waves*: The sound waves of frequencies less than the lower limit (i.e., 20 Hz) of audible range are known as subsonic or infrasonic waves. Such waves are also produced during earthquakes.
- (ii) *Sonic or audible waves*: The human ear can perceive the sound waves of frequency ranging from 20 Hz to 20 kHz. The waves of this frequency range are known as *audible waves*. These waves are produced by different music instruments (such as violin and guitar), our vocal cords, vibrating air columns, and from vibrating membranes.
- (iii) *Ultrasonic waves*: The sound waves having frequencies greater than the upper limit of the audible range (i.e., above 20 kHz) are known as *ultrasonic* or *supersonic waves*. Due to their high frequencies, ultrasonic waves cannot stimulate our ear and brain to the sensation of hearing. But some animals such as bats, cats, dogs, and rats can hear and produce ultrasonic waves. Bats emit a continuous series of ultrasonic squeaks during their flying and use it for communication. Ultrasonic waves are also used for communication by dolphins and some other animals.

3.3 PRODUCTION OF ULTRASONIC WAVES

There are several methods to produce the ultrasonic waves. Some common methods are as follows:

- Mechanical method
- Piezoelectric generator
- Magnetostriction generator

3.3.1 Mechanical Method

Mechanically, the ultrasonic waves are produced by Galton whistle, sirens, wedge resonators, vortex generators, hydrodynamic oscillators, vibrating blade transducer, etc. However, these mechanical methods are rarely used to produce the ultrasonic waves due to their limited range of frequency.

3.3.2 Piezoelectric Generator

Use of piezoelectric generator for the production of ultrasonic waves is the most modern method. With the help of these generators, we can produce ultrasonic waves of very high and constant frequencies. Construction and working of piezoelectric generators is described below:

- Principle:* This generator is based on the principle of piezoelectric effect. According to this principle, when certain piezoelectric crystals such as quartz, Rochelle salt, and tourmaline are stretched or compressed along their mechanical axis, the electrical potential is developed along their electrical axis (perpendicular to mechanical axis). Conversely, if the electric charges are placed on flat crystal surface and alternating potential difference is applied along the electric axis, then mechanical stress will be produced along the mechanical axis (perpendicular to the electrical axis). Under the influence of the applied alternating potential, the crystal plate (along its mechanical axis) will expand and contract alternatively. Thus, it starts to vibrate with the frequency of the applied field. When the frequency of the applied field becomes equal to the natural frequency of the crystal (2.5×10^2 kHz– 10^4 kHz), the amplitude of vibrations becomes large due to resonance. Hence, the converse piezoelectric effect is used to produce the ultrasonic waves.
- Construction:* A suitable arrangement of a piezoelectric generator for the production of ultrasonic waves is shown in Fig. 3.1. A slice of a quartz crystal Q is placed between two highly polished metal plates serving as electrodes. The quartz slab is connected to anode A and grid G of a triode valve. Anode A is also connected with an L-C circuit and high-tension battery, shunted through a by-pass capacitor C_1 . This by-pass capacitor is used to stop the high-frequency currents from passing through battery.

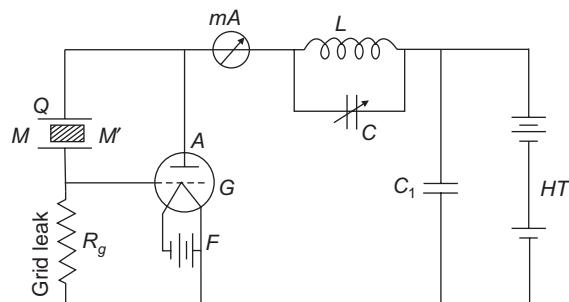


Fig. 3.1 Schematic arrangement of piezoelectric generator

The capacity of variable capacitor C is adjusted in such a way that the frequency of the oscillating circuit becomes equal to the natural frequency of the crystal. The arrangement of the quartz crystal between the plates forms a parallel plate condenser with crystal as the dielectric medium. MM' represents the mechanical axis, while the metal plates represent its electrical axis.

- (iii) *Working:* When the triode valve is operated, a small alternating current flows in its anode circuit. The polarities of anode A and grid G change alternatively. At the same instant, plate A is more negative and G is more positive. Since A and G are connected to the opposite plates of the quartz slab, these plates are oppositely charged. The charges reverse in sign when A becomes more positive and G more negative. Thus, an alternating potential difference is applied on the opposite faces of the crystal and the crystal is set into vibrations along its mechanical axis (MM') producing longitudinal ultrasonic waves in the surrounding medium. At the resonance condition (which is obtained by the varying capacitor), the current in the milliammeter (mA) becomes maximum and the produced vibrations achieve their maximum amplitude. For the production of different range of frequencies of ultrasonic waves, different crystals are used.

The natural frequency of quartz crystal of thickness t for its fundamental mode of vibration is given as

$$v = \frac{1}{2t} \sqrt{\frac{E}{\rho}} \quad (3.1)$$

where E is young's modulus (which is also denoted by Y) and ρ is the density of the material.

For quartz crystal, we have

$$E = 7.72 \times 10^{11} \text{ dyn/cm}^2$$

$$\rho = 2.65 \text{ g/cm}^3$$

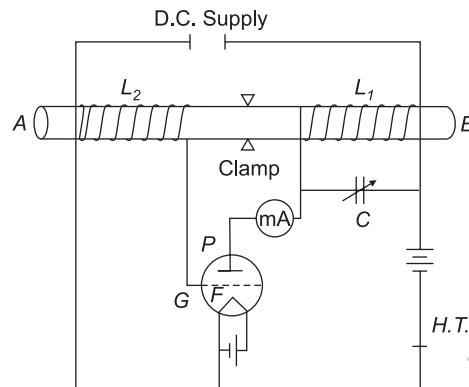
$$t = 0.5 \text{ cm (say)}$$

$$\text{Then, } v = 5.4 \times 10^5 \text{ Hz}$$

3.3.3 Magnetostriction Generator

Magnetostriction generator is another method for producing the ultrasonic waves. It works on the magnetostriction effect discovered by J.P. Joule in 1847. The principle, construction, and working of this generator are described as follows:

- (i) *Principle:* The principle of this generator is based on the manetostriction effect. This effect states that if a rod or tube of ferromagnetic material is placed in a magnetic field parallel to its length, the length changes. This change in length is independent of direction of applied field; it depends only on the magnitude of the field and the nature of the material.
Under the influence of alternating magnetic field, the rod also expands and contracts in length alternately. Hence, the longitudinal compressional waves are produced in the medium around the rod.
- (ii) *Construction:* The experimental arrangement was designed by G.W. Pierce. The schematic diagram of this arrangement is shown in Fig. 3.2. In this arrangement, AB is a rod of ferromagnetic material clamped at the middle. There are two coils L_1 and L_2 on AB . The diameter of the turns of the coils is greater than that of rod so that the rod can move freely lengthwise. Coil L_1 is forming a tuned circuit with the variable condenser C which is also connected to the plate of triode valve through milliammeter. Coil L_2 is connected between the grid and the cathode.

**Fig. 3.2** Schematic arrangement of magnetostriction generator

The frequency of the oscillatory circuit is determined by the values of L_1 and C . If the frequency of the tuned circuit becomes equal to the natural frequency of the rod, then the longitudinal oscillations are set up and ultrasonic waves are produced in the surrounding medium.

- (iii) *Working:* Initially, direct current is passed through the coil to magnetise the rod. Any change in the plate current passing through L_1 causes corresponding change in magnetisation of the rod with the coil L_1 , and therefore the length of the rod changes. These variations in length cause a change in the magnetic flux through the grid coil L_2 . Therefore, by converse magnetostriction effect, an induced emf is set up in L_2 . This induced emf modifies the grid potential of the valve and produces an amplified current change in the plate circuit and in the coil L_1 . In this way, the oscillation of the rod is maintained. In the beginning, the milliammeter indicates alteration when oscillations are set up in the circuit. The variable capacitor C is adjusted in such a way that maximum alterations are produced. In this way, the frequency of oscillating plate current will be equal to the natural frequency of the rod. Hence, the vibrations are most intense and ultrasonic waves are produced. Using this method, we can generate the ultrasonic waves of frequency 250 kHz. At the condition of resonance, the frequency of wave produced by a tank circuit of capacitance C and inductance L is given as

$$v = \frac{1}{2\pi} \sqrt{\frac{1}{LC}} \quad (3.2)$$

Solved Examples

Example 3.1

A quartz crystal of thickness 0.003 m is vibrating at resonance. Calculate the fundamental frequency. (Given that Young's modulus for quartz is 7.9×10^{10} N/m² and its density is 2650 kg/m³.)

Solution

We know that velocity of quartz along the x -direction is given as

$$v = \sqrt{\frac{E}{\rho}}$$

$$\text{Now, } v = \sqrt{\frac{7.9 \times 10^{10}}{2650}} \\ = \sqrt{2.98 \times 10^7} \\ = 5458 \text{ m/s}$$

For fundamental mode of vibration, the thickness should be equal to $\lambda/2$. Hence,

$$\lambda = 2t = 2 \times 0.003 = 0.006$$

$$\text{Now, } u = v \lambda$$

$$\text{or } v = \frac{u}{\lambda} = \frac{5458}{0.006} \\ = 9.096 \times 10^5 \text{ Hz}$$

Note: We can also use the formula $v = \frac{1}{2t} \sqrt{\frac{E}{\rho}}$ for the calculation of frequency.

Example 3.2

A piezoelectric X-cut quartz plate has thickness of 1.6 mm. If the velocity of propagation of sound waves along the x -axis is 5760 m/s, calculate the fundamental frequency of the crystal.

Solution

We know that

$$v = \sqrt{\frac{E}{\rho}}$$

$$\text{and } v = \frac{1}{2t} \sqrt{\frac{E}{\rho}}$$

$$\text{Now } v = \frac{v}{2t}$$

$$\text{Given that } v = 5760 \text{ m/s}$$

$$t = 1.6 \text{ mm} = 0.0016 \text{ m}$$

$$\text{Hence, } v = \frac{5760}{2 \times 0.0016} = 18 \times 10^5 \text{ Hz} \\ = 1.8 \text{ MHz}$$

Example 3.3

Calculate the thickness of a quartz plate needed to produce ultrasonic waves of frequencies (i) 2 MHz and (ii) 30 kHz. (Given that $P = 2650 \text{ kg/m}^3$ and Young's modulus = $8 \times 10^{10} \text{ N/m}^2$.)

Solution

$$\text{We know that } v = \frac{1}{2t} \sqrt{\frac{E}{\rho}}$$

(i) Given that $\rho = 2650 \text{ kg/m}^3$

$$E = 8 \times 10^{10} \text{ N/m}^2$$

and $v = 2 \text{ MHz} = 2 \times 10^6 \text{ Hz}$

$$\begin{aligned}\text{Now } 2t &= \frac{1}{v} \times \sqrt{\frac{E}{\rho}} \\ &= \frac{1}{2 \times 10^6} \times \sqrt{\frac{8 \times 10^{10}}{2650}} \\ &= \frac{1}{2 \times 10^6} \times 5.4944 \times 10^3 \\ \text{or } t &= \frac{5.4944}{4} \times 10^{-3} \\ &= 1.37 \times 10^{-3} \text{ m}\end{aligned}$$

(ii) When $v = 30 \text{ kHz}$

$$2t = \frac{1}{v} \sqrt{\frac{E}{\rho}}$$

$$\begin{aligned}\text{or } t &= \frac{1}{2 \times 30 \times 10^3} \sqrt{\frac{8 \times 10^{10}}{2650}} \\ &= \frac{1}{2 \times 30 \times 10^3} \times 5.4944 \times 10^3 \\ &= \frac{5.4944}{60} \\ &= 9.157 \times 10^{-2} \text{ m}\end{aligned}$$

Example 3.4

Ultrasonic pulse echo method is used to detect possible defects in a steel bar of thickness 40 cm. If the pulse arrival times are 40 and 80 μs , determine the position of defect in the steel bar.

Solution

Let x be the distance of the defect from the end of bar, where the pulse enters. The pulse which arrives first comes from the defect after reflection, and the one which arrives later on, comes from the end after reflection. This means that the pulse takes 40 μs to traverse a distance of $2x$, and 80 μs to cover the distance of 80 cm. Therefore,

$$40 = \frac{2x}{v} \quad (1)$$

$$\text{and } 80 = \frac{80}{v} \quad (2)$$

From Eqs. (1) and (2), we get

$$\frac{80}{40} = \frac{80}{2x}$$

or $4x = 80$
or $x = 20 \text{ cm}$

Example 3.5

A quartz crystal of thickness 0.006 m is vibrating at resonance. Calculate the fundamental frequency. (Given that $E = 7.9 \times 10^{10} \text{ N/m}^2$ and ρ for quartz is 2650 kg/m^3 .)

Solution

$$v = \frac{1}{2t} \sqrt{\frac{E}{\rho}}$$

Given that $t = 0.006 \text{ m}$

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

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

Now,

$$\begin{aligned} v &= \frac{1}{2 \times 0.006} \sqrt{\frac{7.9 \times 10^{10}}{2650}} \\ &= \frac{1}{0.012} \sqrt{29.81 \times 10^6} \\ &= 454.9878 \times 10^3 \\ &= 4.55 \times 10^5 \text{ Hz} \end{aligned}$$

Example 3.6

Calculate capacitance (C) to produce ultrasonic waves of frequency $2 \times 10^6 \text{ Hz}$ with an inductance (L) of 1 H.

Solution

We know that frequency (v) of an LC circuit is given by

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

or $C = \frac{1}{4\pi^2 v^2 L} = \frac{1}{4 \times (3.14)^2 \times (2 \times 10^6)^2 \times 1}$

$$\begin{aligned} &= 6.34 \times 10^{-9} \text{ F} \\ &= 0.00634 \mu\text{F} \end{aligned}$$

3.4 DETECTION OF ULTRASONIC WAVES

Due to high frequency of ultrasonic waves, our ear is not sensitive to it. But some animals such as bats, cats, dogs, and rats can sense these waves. However, we can detect ultrasonic waves by the following methods:

- (i) *Piezoelectric detector*: This detector works on the principle of piezoelectric effect. When a pair of opposite faces of a quartz crystal is subjected to ultrasonic waves, then opposite charges develop on the other pair of opposite faces (perpendicular to mechanical axis). These changes are amplified by a valve amplifier and then detected by suitable means.
- (ii) *Kundt's tube*: Kundt's tube is used to detect the ultrasonic waves of relatively large wavelength. In this method, lycopodium powder is spread in the Kundt's tube and ultrasonic waves are passed in the tube. The powder gets collected at the nodal points of the waves. Now, we can calculate the value of $\lambda/2$ by measuring the distance between two nodes. Thus, by knowing the value of frequency of the wave, velocity of ultrasonic waves can be calculated. For the liquid medium, powdered cake is used in place of lycopodium.
- (iii) *Sensitive flame method*: When a narrow sensitive flame is moved in a medium, where ultrasonic waves are present, it is observed that the flame remains steady at the positions of antinodes and flickers at nodal points, due to change of pressure there. By measuring the distance between the adjacent nodes, we can calculate the value of $\lambda/2$. Thus, by knowing the frequency, we can calculate the velocity of ultrasonic waves.
- (iv) *Thermal detector method*: In this method, platinum wire is moved through the medium of ultrasonic waves. It is observed that the compressions and rarefactions occur at the nodal points; hence, the adiabatic change in temperature takes place. In this way, the resistance of the platinum changes at nodes and remains constant at antinodes. The variations in the resistance of platinum are detected by using sensitive Wheatstone bridge arrangement.

3.5 MEASUREMENT OF WAVELENGTH OF ULTRASONIC WAVES

The wavelength and velocity of ultrasonic waves can be determined with the help of diffraction pattern produced through acoustical grating. When ultrasonic waves are propagated through a liquid, its density varies from layer to layer due to periodic variations in pressure, and the liquid behaves as an acoustical grating. When monochromatic light is passed through this grating, normally, a diffraction pattern is obtained.

For determining velocity and wavelength of ultrasonic waves, an experimental arrangement is set up as shown in Fig. 3.3. In this set-up, a quartz crystal is placed in a glass cell containing the liquid and is connected to an RF oscillator. The oscillator is tuned to resonance frequency. Due to the longitudinal compression and expansion of the crystal, ultrasonic waves are generated, which travel through the liquid and form regions of high and low densities alternatively in the liquid column.

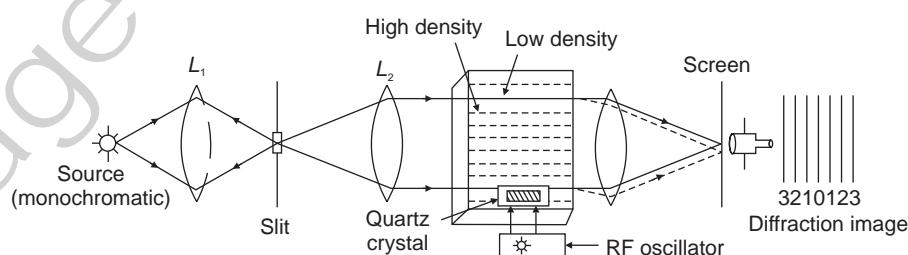


Fig. 3.3 Experimental arrangement for the determination of velocity and wavelength of ultrasonic waves

These regions of high and low densities in the liquid column behave as parallel rulings of a diffraction grating. The region of high density is shown in Fig. 3.3 by broken lines and the region of low density as white region.

When a monochromatic source of light is incident on the liquid column normally, a diffraction pattern is observed at the screen. L_1 is the lens used to confine the light at the slit, lens L_2 is used to ensure that a parallel beam of light passes through the acoustic grating at right angles to the direction of propagation of ultrasonic waves. Now, the diffracted light passing through the grating is received at the telescope in the form of diffraction pattern as shown in Fig. 3.3.

Let θ be the angular separation between the direct image of the slit and the diffracted image of an order n (say). Now, using the theory of diffraction grating, the wavelength and velocity of ultrasonic waves can be calculated. If λ is the wavelength of monochromatic light and $(a + b)$ is the grating element, then for n th order diffraction, we can write

$$(a + b) \sin \theta = n\alpha \quad (3.3)$$

For the acoustic grating, grating element is equal to the wavelength of ultrasonic wave (λ_c). Hence, Eq. (3.3) can be given as

$$\alpha_c \sin \theta = n\alpha \quad (3.4)$$

or
$$\alpha_c = \frac{n\lambda}{\sin \theta}$$

If v_c is the frequency of ultrasonic wave, then the velocity of ultrasonic wave is given as

$$v_c = v_c \cdot \lambda_c = v_c \frac{n\lambda}{\sin \theta} \quad (3.5)$$

3.6 PROPERTIES OF ULTRASONIC WAVES

Some important properties of ultrasonic waves are as follows:

- (i) Ultrasonic waves have high frequencies and therefore they are highly energetic.
- (ii) The speed of ultrasonic waves depends on their frequency; it increases with increase in frequency.
- (iii) Due to their small wavelengths, ultrasonic waves can be transmitted over long distances without any appreciable loss of energy.
- (iv) When ultrasonic waves are passed through a liquid, stationary ultrasonic waves are set up in the liquid due to the reflected and incident waves. This standing wave forms compressions and rarefactions in the liquid medium which act as parallel rulings of a diffraction grating.
- (v) Intense ultrasonic waves produce disruptive effect on liquids by forming bubbles.

3.7 APPLICATIONS OF ULTRASONIC WAVES

Ultrasonic waves have a variety of applications in various branches of science, technology, and industries. Some important applications are mentioned below:

1. Communication:

- (i) Ultrasonic waves are used as SONAR (sound navigation and ranging).
- (ii) They are used for directional signalling.
- (iii) Ultrasonic waves are frequently used for echo depth sounding.

- (iv) They are used for the detection of submarines, icebergs, and other objects in oceans.
 - (v) Using piezoelectric transmitters and receivers, ultrasonic waves are also used for the detection of aircrafts and in burglar's alarm.
2. *Elastic symmetries of the crystal:*
 - (i) Certain crystals produce diffraction images when they are subjected to ultrasonic waves. The elastic symmetries of the crystal can be observed through these diffraction images.
 - (ii) Ultrasonic waves are used for the detection of flows in metals.
 - (iii) These waves are widely used to locate the faults and cracks in metal castings.
 3. *Industrial applications:*
 - (i) Ultrasonic waves are used for cleaning utensils, washings clothes, and removing dust and soot from chimneys.
 - (ii) They are used for drilling and cutting processes in metals. They are also used for soldering of those metals which cannot be soldered by the usual methods.
 - (iii) The welding of metals can be done at room temperature by using ultrasonic waves.
 - (iv) Ultrasonic waves are used in textile industry to improve the dyeing properties of dye suspensions.
 - (v) They are also used in forming of alloys of uniform composition as ultrasonic waves are capable of mixing different densities uniformly.
 4. *Chemical applications:*
 - (i) Ultrasonic waves are used for the preparation of colloidal solutions or emulsions of two immiscible liquids. Nowadays, most of the emulsions such as polishes, paints, food products, and pharmaceutical preparations are done by using ultrasonic mixing.
 - (ii) Ultrasonic waves coagulate dust particles from air or gas. Hence, these waves are used to remove mists, fogs, etc., over the aerodromes.
 - (iii) They are used to break the chain of polymers and molecules.
 - (iv) They are also frequently used in metallurgy.
 5. *Biological and medical applications:*
 - (i) High-intensity ultrasonic waves are frequently used for killing some animals and insects such as rats, frogs, and fishes.
 - (ii) Ultrasonic waves have the capability to kill certain types of bacteria and are, therefore, used to sterilise milk.
 - (iii) These waves are used to destroy diseased tissues. This use of ultrasonic waves is known as *knifeless surgery*.
 - (iv) They are also used to reduce the pain from the neuralgic or rheumatic-affected body parts.
 - (v) They are used for detecting abnormal growth in certain delicate body parts such as brain and heart.
 - (vi) Sonar waves (a type of ultrasonic waves) are used to view soft internal tissues and organs such as the kidney, liver, spleen, pancreas, and urinary bladder which are nearly invisible to X-rays.
 - (vii) These waves are also used for dental cutting and for removing broken teeth without the use of mechanical devices.

4

Non-destructive Testing

4.1 INTRODUCTION

Although the testing is important for all the products before they are being used for specific purposes, it has a great role in civil engineering. It is often necessary to test concrete structure after the concrete has hardened to determine whether the structure is suitable for its designed use. Ideally, such testing should be done without damaging the products (concrete). The tests available for the testing of products range from the completely non-destructive, where there is no damage to the product, through those where the product surface is slightly damaged, to partially destructive tests such as core tests and pullout and pulloff tests, where surface has to be repaired after the test. Non-destructive and partially destructive tests are used to assess the properties such as density, elastic modulus, strength, surface hardness, surface absorption, reinforcement location, overall quality, uniformity, size, and distance from the surface. In some cases it is also possible to check the quality of workmanship and structural integrity by the ability to detect voids, cracking, and delamination.

Non-destructive testing (NDT) is the technique of determining the quality of a product without impairing its properties in any way. This testing provides not only an estimate of the relative strength and overall quality of product or concrete in the structures, but also helps in deciding whether more rigorous tests such as load testing or core drilling at selected locations are required. Non-destructive testing can be applied to both old and new structures. For new structure, the main objective is quality control or the resolution of doubts about the quality of materials or construction. The testing of existing structures is usually related to an assessment of structural integrity or adequacy. With increasing demands of reliability and safety of engineering components, NDT is emerging as an important inter-disciplinary technique and is being used in fields such as nuclear energy, space, and defence.

In this chapter, we will describe NDTs, its objectives, and then focus our attention on different types of defects. Some of the commonly used methods of NDT such as liquid penetration, dye penetration, radiography, ultrasonic inspection, pulse echo system, and visual display units are also described along with their principle of operations.

4.2 NON-DESTRUCTIVE TESTING

Non-destructive testing is the technique which is used to determine the overall quality of product without impairing its properties in any way.

There are various NDT methods which can be broadly classified as those which measure the overall quality of concrete (product), for example, dynamic or vibration methods, such as resonance frequency and ultrasonic pulse velocity tests; and those which involve the measurement of parameters such as surface hardness, rebound, penetration, and pullout strength and are believed to be indirectly related to the compressive strength of concrete. In addition, radiographic, radiometric, nuclear, magnetic, and electrical methods are also available. Since such non-destructive tests are best indirect methods of monitoring the particular characteristic of concrete, and the measurements are influenced by materials, mix, and environmental factors, proper interpretation of the results calls for certain degree of expertise. It is more so when the data on the materials and mix proportions used in the construction are not available as is often the case.

4.3 OBJECTIVES OF NON-DESTRUCTIVE TESTING

The main objective of NDT is to determine the quality of a product without any damage in its structure and properties. In addition to it, some other objectives are mentioned as follows:

- (i) Assessment of the quality of the product or concrete in relation to standard requirements.
- (ii) Assessment of the quality of one element of a structure in relation to other.
- (iii) Quality control of pre-cast units.
- (iv) Monitoring of strength development in relation to formwork removal, cessation of curing, pre-stressing, load application or similar purpose.
- (v) Removing uncertainties about the acceptability of the material supplied, owing to apparent non-compliance with specification.
- (vi) Confirming the doubt concerning the workmanship involved in batching, mixing, placing, compacting, or curing of concrete.
- (vii) Determining the locations of cracks, voids, honeycombing, and similar defects within a concrete structure.
- (viii) Determining the position and condition of reinforcement.
- (ix) Monitoring long-term changes in the properties of concrete and other products.
- (x) Assessing the potential durability of the products.
- (xi) Searching the location of suspected deterioration of concrete or products in a structure due to overloading, fatigue, external or internal chemical attack, fire, explosion, and environmental impact.
- (xii) Assessing any proposed change in a structure for its insurance or for change of ownership.

4.4 TYPES OF DEFECTS

There are many factors due to which a concrete structure is prone to wear and tear and damage. Some common factors are undue loading, earthquakes, and damage due to the environmental effects. In order to repair the damaged structures, it is important to examine the defects and classify them into different categories for their proper treatment. Some important defects are discussed in ensuing sections.

4.4.1 Cracking

It is very common that visually detectable distress appears in any building which needs repair or retrofit. Depending on cause the cracks may be major or minor. Usually, the problem of minor cracks are associated with shrinkage, whereas the major cracks are due to overloading of structure. Cracking by itself does

not warrant repair. Many cracks are inevitable such as those due to shrinkage or around reinforcement in the case of reinforced brickwork. Only those cracks are considered as a sign of distress which are wider than the acceptable limit.

Small cracks can be repaired with the cosmetic treatment such as putty work. For the repair of wide cracks, there is a need of detailed investigation for its real cause. Depending on the crack direction as well as its stability with time on the affected structure, these can be studied under the following points:

(i) *Dormant*: This is the category of cracks which do not increase in size once formed. Usually, these are caused by shrinkage, initial movement of supports, or perious overload. Its repair is not essential.

(ii) *Active*: The size of these cracks changes under load. There is continuous movement in crack size and overloading. It is difficult to repair such cracks. If cause is not treated then new crack next to the repaired one is common.

(iii) *Fine, medium, and wide cracks*: The cracks generally of the size of less than 1 mm are known as fine cracks. The cracks having width between 1 and 2 mm are called medium cracks. If the width of cracks are greater than 2 mm, then it is called wide cracks.

4.4.2 Spalling

De-lamination of the surface of concrete is known as *spalling*. Usually, internal stress or external loads are responsible for spalling. Some important causes for spalling are as follows:

- (i) Concentrated eccentric external load producing highly stressed narrow compression zone.
- (ii) Corrosion of steel embedded in concrete.
- (iii) Freeze-thaw effect of entrapped water.
- (iv) Chemical reactions, efflorescence, and repeated wetting and drying.

Careful examination of the location of spalling can reveal the facts responsible for spalling.

4.4.3 Staining

It is an other important defect occurring in buildings or in structures, which is usually caused by absorption of water, which contains minerals/salts and results in leaching or draining over other components. Staining is mainly caused by efflorescence. In the process of efflorescence, water soluble salts are deposited on the surface of masonry or concrete because the water carrying these salts evaporates. Due to efflorescence, the surface of buildings become ugly and aesthetically unpleasant. It can weaken the constituents of building because of internal crystallisation. The root cause of efflorescence staining and corrosion is the moisture penetration.

4.4.4 Construction and Design Defects

There are many defects which may occur during the construction and designing of a building due to the defective workmanship. Some important deficiencies which are recorded by visual inspection are as follows:

- (i) Choice of wrong wall thickness
- (ii) Out of plumb of walls
- (iii) Defective joint and bonds
- (iv) Lack of movement joints
- (v) Misalignment of joints
- (vi) Failure to connect inserting walls and columns
- (vii) Improper drainage path causing staining.

- (viii) Poor layout that causes excessive torsion.
- (ix) Selection of inadequate diaphragm stiffness for distributing lateral shear.
- (x) Cracking below beams due to inadequate gap for deflection.

4.4.5 Honeycombing

Honeycombing occurs when too much coarse aggregate appears on the surface with some cavities underneath. It occurs as a result of poor compaction, or if a bony mix is used with not enough sand. If it occurs on the surface, it can be re-profiled with a render (thin layer of sand/cement mortar) or a proprietary cement product. If cavities exist below the surface, it is more appropriate to break out to sound and dense concrete and repair it with suitable technique.

4.4.6 Dusting

Dusting is a surface defect which appears as fine powder on the concrete surface and comes off when brushed. Such defect is caused when finishing work is done on the concrete before bleed water has dried out or due to inadequate curing. This defect can be repaired by applying a chemical floor hardener or bounded topping.

4.4.7 Blistering

Blisters occur when the fresh concrete surface is sealed by trowelling trapping air or bleed water under the surface. It can be avoided by delaying trowelling as long as possible and covering to prevent evaporation.

4.4.8 Rain Damage

Heavy rain may cause pitted surface or eroded surface on a concrete structure. Such damage can be avoided by convexing newly placed concrete with plastic sheeting when it rains. Rain damage on the wetted (i.e., newly constructed surface) can be reworked and refinished.

4.5 METHODS OF NON-DESTRUCTIVE TESTING

With increasing demands of reliability and safety of engineering components, NDT is emerging as an important inter-disciplinary technique and is being used in different fields such as research, industry, nuclear energy, space, and defence. There are many techniques used in NDT of materials. Some important techniques are as follows:

- (i) Liquid penetration method
- (ii) Dye penetration method
- (iii) Radiographic testing
- (iv) Ultrasonic inspection method
- (v) Pulse echo method
- (vi) Magnetic particle testing
- (vii) Eddy current testing

4.5.1 Liquid Penetration Method

Liquid penetration method is one of the simplest and low-cost method of testing. This test is used as a quality control test for precast units. In this test, liquid is applied uniformly on the surface to be tested. When liquid comes into contact with dry concrete, it is absorbed by the surface defects such as minute cracks and pinholes, and gets trapped. After sometime, the excess liquid on the surface is wiped off. Now developer is used to record the surface defects.

Developer acts as blotter placed on the surface. In this way, penetrant appears on the surface and the defects are recorded. Liquid penetration method is frequently used in detecting the cracks in turbine blades, smoothness of the surface, minor and middle type surface cracks, etc.

4.5.2 Dye Penetration Method

Dye penetration inspection (DPI) is also known as liquid penetration inspection (LPI). It is the low-cost inspection method used to locate surface breaking defects in all non-porous materials (metals, plastics or ceramics). The penetrant may be applied to all non-ferrous materials; however, for inspection of ferrous components, magnetic particles inspection is preferred. This method is used to detect casting and forging defects, cracks, leaks in new products and fatigue cracks on inservice components.

Principle

This method is based on capillary action, when low surface tension fluid penetrates into clean and dry surface-breaking discontinuities. Penetrant may be applied to the test component by dipping, spraying, or brushing. Penetrant is allowed to be on the surface for sufficient time, the excess penetrant is wiped from the surface which is under test. Now the developer is used to draw penetrant out of the flow, where a visible indication becomes clear (visible) to the inspector. Depending on the nature of dye (fluorescent or nonfluorescent), ultraviolet or white light is used for inspection.

Materials or Penetrants

According to the selection of sensitivity levels, different types of penetrants are used in this method. Visible penetrants are typically red in colour which represent the lowest sensitivity. Fluorescent penetrants contain two or more dyes that fluorescence when excited by ultraviolet radiation. Fluorescent penetrant inspection is performed in a darkened environment, because the excited dyes emit brilliant yellow-green light that contrasts strongly against the dark background. This material is more sensitive to small defects.

Inspection Method (Procedure)

Pre-cleaning is the first step in this method. In this step, the test surface is cleaned to remove dirt, paint, oil, grease, or any loose scale that could either keep penetrant out of a defect, or cause irrelevant or false indications. For the pre-cleaning, some solvent alkaline cleaning steps, vapour degreasing, or media blasting is used. The ultimate goal of this step is to clean the surface where defects are present and open it to the surface, dry and free of contamination.

In the second step, penetrant is applied to the surface under inspection. The penetrant is allowed for sufficient time (nearly 30 min) to soak into any flaws. The dwell time depends on the penetrant and the surface which is under test. It is observed that smaller flaws require a longer penetration time. Due to their incompatible nature, one must be careful not to apply solvent-based penetrant to a surface which is to be inspected with a water-washable penetrant.

In the third step, excess penetrant is removed from the surface under test. There are different removal methods used for different type of penetrants used.

According to the choice of different penetrants, water-washable, solvent removable, lipophilic post-emulsifiable, or hydrophilic post-emulsifiable are most commonly used methods. This process must be performed under controlled conditions so that all penetrants on the surface are removed but penetrants trapped in real defects remain in place.

In the fourth step, developers are applied as the surface under the inspection. After removing the excess penetrant, a white developer is applied to the sample. There are several developers such as non-aqueous

developer, dry developer, water suspendable, and water soluble. The selection of developer is governed by penetrant compatibility and inspection conditions. For the use of non-aqueous wet developer or dry developer, the sample must be dried before its application, while soluble and suspendable developers are applied with the part still wet from the previous step. The developer draws penetrant from defects out the surface to form a visible indication, a process to the action of blotting paper. Any coloured stains indicate the position and types of defects on the surface under inspection.

In this step, suitable radiations are used to examine the defects. For this purpose, visible light with adequate intensity is used for visible dye penetrant. For fluorescent penetrant examinations, ultraviolet radiation of adequate intensity along with low ambient light levels are used. Inspection of the test surface should take place after a 10-min development time. This time delay allows the blotting action to occur.

Limitations

Liquid penetration has the following limitations:

- (i) This test is very sensitive to changes in quality and to correlate with the observed weathering behaviour.
- (ii) This method is not applicable for porous and honeycombed concrete.

4.5.3 Radiographic Testing

Radiographic testing is used to determine the position of cables, voids in grouting, and in situ density of concrete. Depending on the choice of radiations, there are three methods used for testing of concrete, viz., X-ray radiography, γ -ray radiography, and γ -ray radiometry. When the radiation is allowed to incident on the object, it penetrates upto the varying degrees in the object depending on its thickness and composition. The portion which penetrates the object can be recorded on the film. By examining the film in different exposures, the thickness and composition of the member can be found.

Principle

The intensity of a beam of X-rays or gamma rays suffers a loss of intensity while passing through a material. This phenomena is due to the absorption or scattering of the X-rays or gamma rays by the object being exposed. The amount of radiation lost depends on the quality of radiation, the density of the material, and the thickness traversed. The beam of radiation which emerges from the material is usually used to expose a radiation sensitive film so that different intensities of radiation are revealed as different densities on the film.

The relation between the intensity of incident and transmitted radiation is given as

$$I = I_0 e^{-\mu x} \quad (4.1)$$

where I is the intensity of transmitted radiation, I_0 is the intensity of incident radiation, μ is the attenuation coefficient, and x is the thickness of object.

The radiographic inspection of a surface is shown in Fig. 4.1.

Radiations Used

In the radiographic testing, X-rays and γ -rays are used. For steels of 15 mm thickness, 150 kVA X-ray machine is used. For the thickness of the order of 75 mm, 400 kVA X-ray machines are required. Beyond this thickness of steel, γ -rays are used. The commonly used γ -rays sources in industrial applications are CO_{60} , CS_{137} , and Ir_{192} .

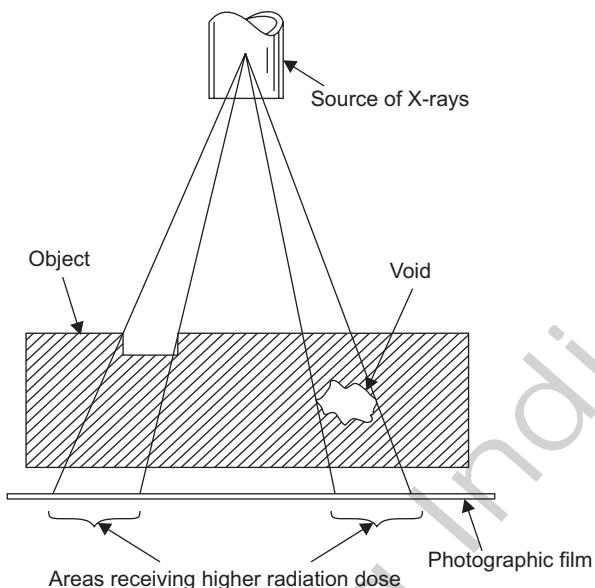


Fig. 4.1 Radiographic method of testing

Inspection Method (Procedure)

There are three steps which are involved in the inspection method of radiographic test. In the first step, suitable radiation (X-ray or gamma ray) is allowed to pass through the structure which is under examination. The choice of radiation depends on the thickness and composition of structure. In the second step, impression or image of voides in grouting, damage in the structure, and in situ density of concrete in structure are obtained on a photographic plate. Thus, the portion of radiation which penetrates through the object is recorded on the film. In the third step, film is examined by different exposures. By the closed observations, the thickness and composition of the member can be found. In reinforced concrete rebars can easily be located. The commonly used γ -ray sources do not use electrical power such as X-ray sources. Due to this reason, it is convenient to use these radiations in fields. Radiography is widely used in testing the quality of welding and casting.

Limitations

There are certain limitations of radiographic testing which are as follows:

- (i) It is expensive.
- (ii) It requires stringent safety precautions.
- (iii) It is also limited by the member thickness (maximum upto 600 mm).

4.5.4 Ultrasonic Inspection Method

An ultrasonic pulse is produced by an electro-acoustical transducer, which is held in contact with one surface of the concrete under test. When the pulse generated is transmitted into the concrete from the transducer using a liquid coupling material such as grease or cellulose paste, it undergoes multiple reflections at the boundaries of the different material phases within the concrete. The first waves to reach the receiving transducer are the longitudinal waves, which are converted into an electrical signal by a

second transducer. Electronic timing circuits enable the transit time T of the pulse to be measured. Hence, the longitudinal pulse velocity is given as

$$v = \frac{L}{T} \quad (4.2)$$

where v is the longitudinal pulse velocity, L is the path length, and T is the time taken by the pulse to traverse that length.

Principle

Basic principle of ultrasonic inspection method is based on the fact that the velocity of ultrasonic waves are different in different medium. The velocity of an ultrasonic pulse through any material depends upon its density, modulus of elasticity, the presence of reinforcing steel, and Poisson's ratio. It is observed that the velocity of ultrasonic wave is comparatively higher when the quality of concrete in terms of density, homogeneity, and uniformity is good. In case of poorer quality, a lower velocity is obtained.

Apparatus Required

The main apparatus required for ultrasonic pulse velocity test are as follows:

- (i) One electrical pulse generator
- (ii) A pair of transducer
- (iii) An amplifier
- (iv) An electronic timing device

Electrical pulse generator is used to produce the pulse of suitable frequency. In the pair of transducer, one is used as transmitter while other is used as receiver. In this method of inspection, transducers with a frequency of 50 to 60 kHz are suitable for most common applications. Suitable amplifier is used to amplify the electrical signal obtained from the conversion of vibration pulses. The electronic timing circuit enables the transit time (T) of the pulse to be measured, from which the ultrasonic pulse velocity (v) can be calculate as

$$v = \frac{L}{T}$$

where L is the path length.

Different Methods

There are three ways of measuring pulse velocity through concrete.

- (i) *Direct method*: The direct method is preferred wherever access to opposite sides of the component is possible [Fig. 4.2(a)]. The receiving transducer detects the arrival of that component of the pulse, which arrives earliest. This is generally the leading edge of longitudinal vibration. Although the direction in which the maximum energy is propagated is at right angles to the face of the transmitting transducer, it is possible to detect pulses which have travelled through the concrete in some other direction.
- (ii) *Semi-direct method*: The semi-direct method is used wherever access to different but not opposite sides of the component is possible [Fig. 4.2(b)]. The semi-direct transmission arrangement has a sensitivity intermediate between direct and indirect methods. Although there may be some reduction in the accuracy of the measurement of the path length, it is generally found to be sufficiently accurate to take this as the distance measured from centre to centre of the transducer faces.

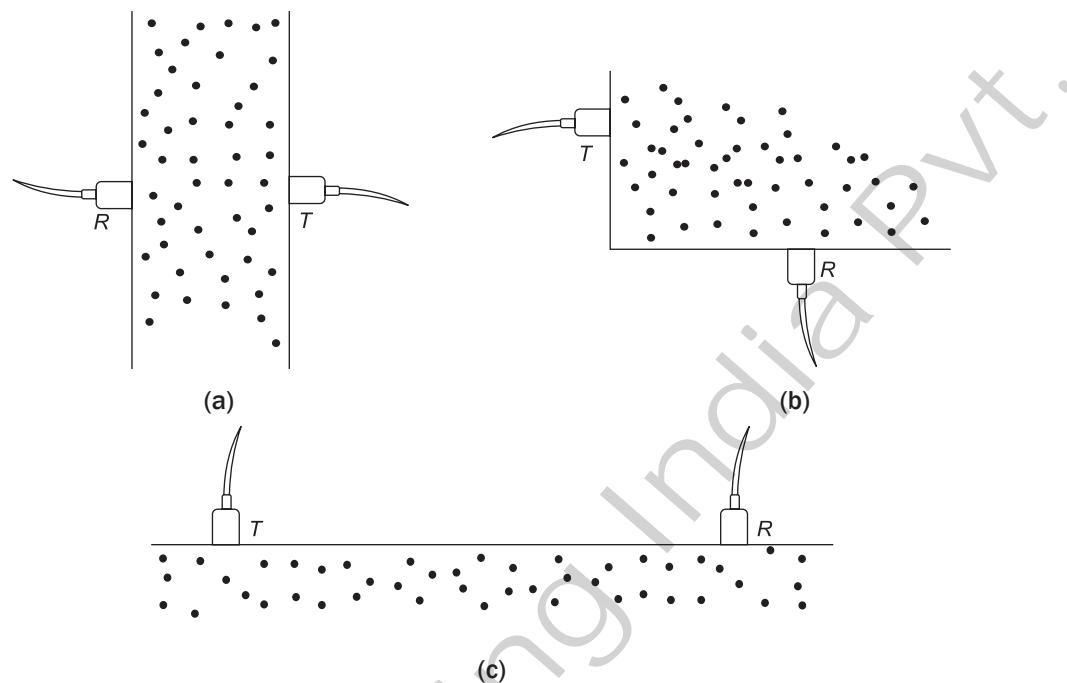


Fig. 4.2 (a) Direct method of ultrasonic pulse velocity test, (b) semi-direct method of ultrasonic pulse velocity test, and (c) indirect method of ultrasonic pulse velocity test

- (iii) *Indirect method:* The surface method or indirect method is least satisfactory and should be used when access to only the surface is possible [Fig. 4.2(c)]. This technique is used when the depth of a surface crack is to be determined or when the quality of the surface concrete relative to the overall quality is of interest.

Inspection Method (Procedure)

In this testing, the ultrasonic waves are made to travel through the test sample. Whenever these encounter a defect in their path, they get reflected. The reflected beam can be detected and from the time delay between the incident and reflected waves, the location of the defect can be ascertained. This technique has been used for most metals and hard non-metals as well as on forgings, weldings, and castings. There have been tremendous advances in recent years in this field. Ultrasonic imaging and ultrasonic holography are helping greatly in the study of defects.

Limitations

There are some limitations of this testing method, which are mentioned as follows:

- (i) Variation of concrete temperature may affect the pulse velocity. It is observed that at temperatures between 30°C–60°C, there can be reduction in the pulse velocity by 5%. Below the freezing temperature, pulse velocity is increased by 7%.
- (ii) The increase in the moisture content in concrete increases the pulse velocity. The influence is pronounced for low-strength concrete than for high-strength concrete.

- (iii) The pulse velocity measured in reinforced concrete in the vicinity of reinforcing bars is usually higher than in plain concrete of the same composition.
- (iv) For highly stressed concrete, pulse velocity may be reduced due to the development of micro-cracks.

4.5.5 Pulse Echo Method

Pulse echo method is used for locating defects such as voids, cracks, fracture, crack growth, metallurgical change, and zone of deterioration. Schematic diagram of this method is shown in Fig. 4.3.

Principle

The principle of this method is based on the interaction of acoustic waves with the internal structure of the object under examination.

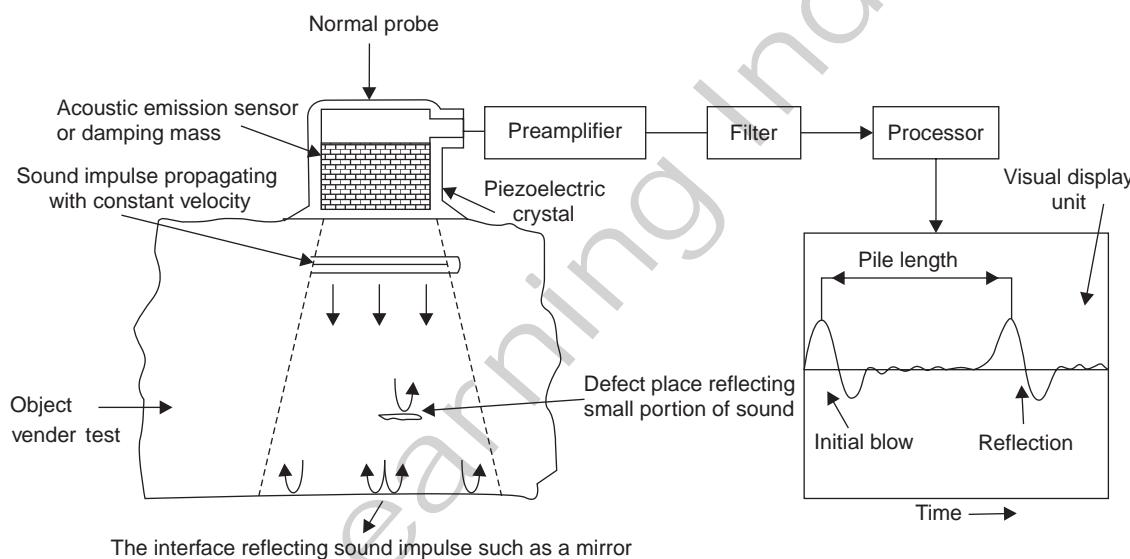


Fig. 4.3 Pulse echo method of testing

Apparatus Required

There are different parts which are used in the set up of pulse echo method; these parts are described as follows:

- (i) *Normal probe*: It is the most important part of this set up. This section serves the purpose of both transmitter and receiver. The piezoelectric crystal used in this part produces the ultrasonic waves in phase with the applied alternating voltage. These pulses are passed through the object under the test. The pulses get their reflection from the defects of object under the examination. Damping mass or acoustic emission sensor receives mechanical oscillations and converts it into the electrical signal.
- (ii) *Preamplifier, filter, and processor*: The signal detected as normal probe needs amplification which is done by preamplifier unit. Sometimes there is need to prevent unwanted signal (if any) which is done by filtering unit. In order to display this signal in terms of desired scale, there is need of its processing for which suitable processor is required.

5

Architectural Acoustics

5.1 INTRODUCTION

The science of architectural acoustics deals with the planning of a building, auditorium, music halls, recording room, etc., with a view to provide the best audible sound to the audience. In other words, we can say that this branch of science deals with the behaviour of sound waves in closed space. In this chapter, we will focus on the important factors which affect the quality of sound waves, reaching to the audience in the closed space.

5.2 CLASSIFICATION OF SOUND WAVES

It is difficult to classify sound waves into different categories, because the reception of sound waves depends on the choice of the receivers of different ability. However, depending on their frequencies, the sound waves can be divided roughly into three classes. These are discussed in the ensuing sections.

5.2.1 Audible Waves or Sonics

Audible waves are that range of sound wave frequency which stimulate the human ear and brain to the sensation of hearing. These waves are confined to the frequency ranges from 20 Hz to 20 kHz. Such waves are also known as *sonics*.

Usually, audible waves or sonics are produced in vibrating bodies such as stretched strings (violin, guitar, etc.), vocal cords, vibrating air column, vibrating plates, and membranes. Sometimes audible waves are also produced by regular periodic or irregular nonperiodic disturbance of the bodies.

5.2.2 Infrasonic Waves or Subsonic

Infrasonic waves are that range of sound wave frequency within which the human ear is not capable of receiving sound energy. Usually, infrasonic waves have their frequency less than the lower limit of audible range (i.e., below 20 Hz).

Infrasonic waves are produced by large sources such as earthquakes or due to any massive disturbance inside the earth and sea. Although human ears are not sensitive to such waves, some animals such as dogs are capable to sense such waves.

5.2.3 Ultrasonic Waves

The sound waves in which the frequencies are above the limits of human audibility, i.e., frequencies greater than 20 kHz are referred to as *ultrasonics* or *supersonics*. Because of high frequencies, ultrasonic waves cannot stimulate our ears. Although some animals such as bats and dogs show their response to it, the wavelengths of ultrasonic waves are very small as compared to the audible sound. Most of the applications of the ultrasonic waves have been possible on account of their small wavelengths. Nowadays, the term *supersonic* is used only to refer to velocities greater than that of sound waves (about 330 m/s). Sound waves of frequency higher than 10^8 Hz are known as *hypersonic waves*.

Ultrasonic waves have many applications in the field of industries, medicines, metallurgy, biology, and other branches of science and technology. Ultrasonic waves are discussed in detail in Chapter 2.

5.3 NOISE AND MUSICAL SOUND

In Section 5.2, we discussed the classification of sound waves according to their frequencies. In another classification, sound can be divided roughly into two broad classes: noise and musical sound. There is no sharp line which can discriminate between noise and musical sound. It is observed that some noises have more or less musical character and some musical sounds are not completely free from noise. Musical sounds are characterised by their regularity and smoothness, while the noise is characterised by its irregularity and roughness. In other words, we can say that the musical sounds are produced by periodic vibrations, whereas noises are characterised by nonperiodic vibrations. There are three musical sounds which are characterised by three features, namely, pitch, intensity, and quality. Brief discussions on these features are given as follows:

- (i) *Pitch*: The pitch of sound wave is a general feature which depends only on the period or frequency of vibration constituting the sound. The pitch of sound wave increases or decreases with its frequency—for the greater frequency, pitch is higher. Pitch is generally specified in two distinct ways. One is scientifically and the other is musically. Scientifically, pitch is specified by the frequency of vibration, whereas musically it is a certain accepted series of sound constituting the musical scale.
- (ii) *Intensity and loudness*: The intensity of sound wave is completely a physical quantity which is proportional to the square of the amplitude of sound wave. The loudness is measured as the degree of sensation, depending on the intensity of sound and the sensitiveness of the organ (ears) receiving it. Although the loudness of a sound depends on the intensity producing it, it is difficult to define a precise relation between these two.
- (iii) *Quality*: It is an important property of sound wave with the help of which we can distinguish a musical sound from others having same pitch and intensity. The different musical instruments can produce musical sounds of different quality. The quality of a musical tone depends upon the number order and relative strength of the harmonics but not upon the difference of phase.

5.4 WEBER—FECHNER LAW (SENSATION OF LOUDNESS)

According to the Weber's law, the increase of stimulus to produce the minimum perceptible increase of sensation is proportional to the pre-existing stimulus. With the help of Weber's law, we can measure the sensation of loudness.

It means that using this law we can measure the change in intensity, capable to produce a noticeable change of sensation, starting from certain initial level of sound intensity.

Using Weber's law, Fechner derived the following relation between the magnitude of sensation and the intensity of stimulus:

$$\delta S = K \frac{\partial E}{E}$$

or $S = K \log E$ (5.1)

where S is the sensation magnitude, E the intensity of stimulus, and K is a constant. Equation (5.1) is known as Weber–Fechner law. Later on, this relation has been verified experimentally by Knudsen.

5.5 LOUDNESS LEVEL AND INTENSITY: BEL AND DECIBEL

The audible frequencies within which the ear can respond is very large, having range from 20 Hz to 20 kHz. In this range of frequency, our ears can distinguish a large gradation of intensity usually from 1 to 10^{12} . Since this range of sound is very wide so instead of using simple calculation to deal it, logarithmic scale is preferred. The usual unit of sound intensity in logarithmic scale is *bel*.

In order to define bel or decibel, let us consider P_1 and P_2 as the output powers of two sources of sound. Now the gain regarding to these two sources is defined as

$$\text{gain} = \log_{10} \left(\frac{P_2}{P_1} \right) \text{bel} \quad (5.2)$$

Bel is the large unit of sound intensity. It can be expressed in terms of smaller unit decibel (abbreviated as dB). As from its name, it is clear that decibel is 1/10th of the bel. In terms of decibel, Eq. (5.2) can be written as

$$\text{gain} = 10 \log_{10} \left(\frac{P_2}{P_1} \right) \text{decibel} \quad (5.3)$$

As the output power is proportional to the square of amplitude, if A_1 and A_2 represent the corresponding amplitudes, then

$$\text{gain} = 10 \log_{10} \left(\frac{A_2}{A_1} \right)^2 \text{decibel}$$

or $\text{gain} = 20 \log_{10} \left(\frac{A_2}{A_1} \right) \text{decibel}$ (5.4)

If we are selecting a system which has its gain as 1 dB, then Eq. (5.3) can be given as

$$10 \log_{10} \frac{P_2}{P_1} = 1$$

or $\log_{10} \frac{P_2}{P_1} = 0.1$

or $\frac{P_2}{P_1} = 1.26$ (5.5)

According to Eq. (5.4),

$$20 \log_{10} \left(\frac{A_2}{A_1} \right) \text{dB} = 1$$

or

$$\log_{10} \left(\frac{A_2}{A_1} \right) = 0.05$$

or

$$\frac{A_2}{A_1} = 1.12 \quad (5.6)$$

Thus the gain of 1 dB is equivalent to the power amplification of 1.26 or amplitude amplification of 1.12.

5.6 BASIC REQUIREMENTS FOR THE ACOUSTICALLY GOOD HALLS

Prior to 1900, the architects and building engineers had no consideration about the acoustical problems. Sometimes it was observed that building was found to be unsatisfactory for the purpose of which it was built. The Fogg Art Museum Hall of Harvard University was so defective acoustically that the speaker could hardly make his words intelligible to his audience. In order to remove the acoustical problems of this hall, Prof. Wallace C. Sabine conducted a series of experiments in 1911. On the basis of his experimental findings, scientifically, he talked about the problem of satisfactory speech and music in a hall and laid down the following basic requirements for the acoustically good halls.

- (i) The sound should be sufficiently loud and intelligible at every point in the hall.
- (ii) There should be no echoes.
- (iii) Relative intensities of several components of a complex sound must be maintained.
- (iv) For the sake of clarity the successive syllable spoken must be clear and distinct.
- (v) There should be neither concentration of sound nor zone of silence in any part of the hall.
- (vi) Undue noise and effect of resonance should be avoided.
- (vii) The reverberation should neither be too large nor too small.
- (viii) There should be resonance within the building.

5.7 REVERBERATION

It is observed that when sound waves are produced in a big hall, they spread out, scatter, and strike the surfaces of the various objects in the hall such as ceiling, walls, and floor. Some of the sound energy is reflected and some absorbed and thus the distraction of sound energy in the hall continues to be affected. The reflected portions travel back into the hall and reunite to form objectionable echoes and affects good hearing. Thus, an audience in a hall receives the sound in two ways—directly from the source and after suffering reflections. The quality of the sound received by the audience is a combined effect of these two sets of waves. There is a time gap between the sound waves directly received from the source and the waves received by successive reflections. Due to this reason, sound received by audience does not cease immediately but it persists for some time even after the source is stopped. This persistence of audible sound after the source has ceased to emit sound is called *reverberation*.

The time during which sound persists in the hall is called reverberation time of sound. It is measured from the time that the original sound is produced, or in the case of a continuous note from the time, that the source stopped sounding, to the time at which the sound becomes inaudible.

The reverberation upto a certain extent is, of course, desirable as it adds a pleasing characteristic to the acoustical qualities of a room, especially for giving richness to music, but too much reverberation is most undesirable. In this way we can conclude that obtaining the right value of reverberation is the secret of good acoustics.

The variation of sound density in a hall with time has been shown in Fig. 5.1. From the figure it is clear that the average growth and fall of sound density for a spoken syllable within an auditorium is exactly similar to the growth and decay curve of the electric current in an inductive circuit. Initially, when a source emits sound continuously, the listener receives an increasing sound intensity. After a few time, the energy emitted per second from the source becomes equal to the energy absorbed or dissipated by the room. At this stage, the resultant energy does not increase further with the time, i.e., it attains average steady value. Now if the original sound ceases, the sound heard does not stop instantaneously, because the listener continues to receive the successive reflections till the minimum audibility level.

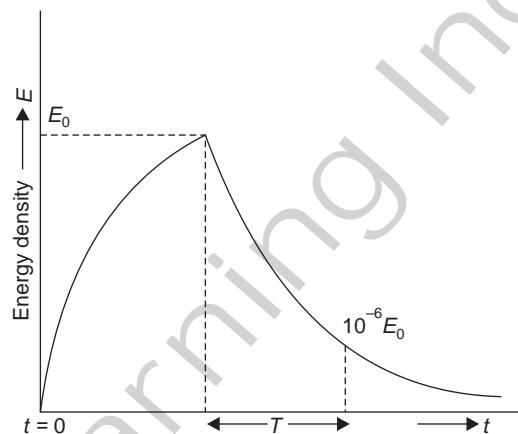


Fig. 5.1 Growth and decay of sound density

The time in which the energy density of sound falls to 10^{-6} times of its steady value, after the source is cut off, is known as the *standard time of reverberation*.

5.8 DETERMINATION OF THE TIME OF REVERBERATION: SABINE'S LAW

W.C. Sabine (1868–1919) focussed his study on the subject of architectural acoustics scientifically, particularly with reference to reverberation time, which is one of the most important factor affecting speech and music in a room. Initially, Sabine used the following accessories in his experimental setup to measure the time of reverberation.

- Standard source of sound waves having initial intensity of 10^6 times that of the human audible intensity in a small hall. An organ pipe of frequency 512 Hz is used for this purpose.
- A stop watch.

During his experiment, a stop watch was started at the instant when the source of sound ceased, and it was stopped at the instant when the sound ceased to be audible. Later on more sophisticated method was used by Sabine for more accurate measurement.

In his experimental findings, Sabine concluded the following points:

- The reverberation time (T) of room varies inversely as A , where A is the effective absorbing surface area.
Thus,

$$T \propto \frac{1}{A} \quad (5.7)$$

- The reverberation time (T) of room is directly proportional to the volume of the room.

$$T \propto V \quad (5.8)$$

From Eqs. (5.7) and (5.8), we get

$$T \propto \frac{V}{A}$$

or $T = K \frac{V}{A}$

where K is a constant.

If a is the absorption coefficient per unit area and s denotes a small area on either of the walls of the room or that of an absorbing material, then reverberation time (T) can be given as

$$T = K \frac{V}{\Sigma as} \quad (5.9)$$

Sabine determined the value of K experimentally, which is approximately equal to 0.05, and hence we get

$$T = \frac{0.05V}{\Sigma as} \quad (5.10)$$

5.9 DERIVATION OF EXPRESSION FOR REVERBERATION TIME (SABINE'S LAW)

In order to derive the Sabine's expression for reverberation time, let us consider a_1, a_2, a_3, \dots , etc., be the absorption coefficients at each reflection of the surfaces s_1, s_2, s_3, \dots , etc., in a room. The average value of the absorption coefficient \bar{a} is given as

$$\begin{aligned} \bar{a} &= \frac{a_1 s_1 + a_2 s_2 + a_3 s_3 + \dots}{s_1 + s_2 + s_3 + \dots} \\ &= \frac{\sum a_i s_i}{S} \end{aligned}$$

or $\sum a_i s_i = s \bar{a} \quad (5.11)$

where S is the total area of the surface. On the basis of statistical data, Jaeger suggested that the sound travels an average distance ($4V/S$) between two successive reflections, where V is the volume of the room and S is the total surface area. If v is the velocity of sound, then the time between any two successive reflections will be ($4V/Sv$). Now the average number of reflections in time t is given as ($Svt/4V$). In a single reflection, if \bar{a} is the absorption coefficient, then the reflected fraction will be given as $(1 - \bar{a})$. After two reflections, the fraction reflected is $(1 - \bar{a})(1 - \bar{a})$, i.e., $(1 - \bar{a})^2$. Similarly, after $(Svt/4V)$ reflections, the fraction of reflected sound will be given as $(1 - \bar{a})^{Svt/4V}$.

If I_0 is the initial intensity of sound I_t after time t [i.e., after $(Svt/4V)$ reflections], then we can write

$$I_t = I_0 (1 - \bar{a})^{Svt/4V} \quad (5.12)$$

If $t = T$ (reverberation period), then from the definition of reverberation time, we know that

$$\frac{I_t}{I_0} = 10^{-6} \quad (5.13)$$

At $t = T$, using Eqs. (5.12) and (5.13), we get

$$10^{-6} = I_0 (1 - \bar{a})^{SvT/4V} \quad (5.14)$$

Taking log on both sides of Eq. (5.14), we get

$$\begin{aligned} \log_e 10^{-6} &= (SvT/4V) \log_e (1 - \bar{a}) \\ \text{or } T &= \log_e 10^{-6} \times \frac{4V}{Sv \log_e (1 - \bar{a})} \\ &= 2.3026 \log_{10}(10^{-6}) \left[\frac{4V}{Sv \log_e (1 - \bar{a})} \right] \end{aligned}$$

Using $v = 1120$ ft/s (sound velocity), we get

$$\begin{aligned} T &= \frac{2.3026 \times (-6) \times 4V}{S \times 1120 \log_e (1 - \bar{a})} \\ &= -\frac{55.26V}{1120 \times S \times \log_e (1 - \bar{a})} \\ \text{or } T &= -0.05 \left(\frac{55.26V}{S \times \log_e (1 - \bar{a})} \right) \end{aligned} \quad (5.15)$$

Equation (5.15) is also known as *Eyring's formula*.

In Eq. (5.15), \bar{a} is less than 1. Now expanding $\log_e (1 - \bar{a})$ as series, we get

$$T = \frac{0.05V}{-S \left[-\bar{a} - \frac{\bar{a}^2}{2} - \frac{\bar{a}^3}{3} \dots \right]} \quad \left\{ \text{Because } \log_e (1 - \bar{a}) = -\bar{a} - \frac{\bar{a}^2}{2} - \frac{\bar{a}^3}{3} \dots \right\}$$

Since a is much less than 1, we can neglect the higher-power terms, and hence we get

$$\begin{aligned} T &= \frac{0.05V}{-S(-\bar{a})} \\ &= \frac{0.05V}{S\bar{a}} \\ \text{or } T &= \frac{0.05V}{\sum a_i s_i} \end{aligned} \quad (5.16)$$

Equation (5.16) is also known as *Franklin's formula*, where $\sum a_1 s_1 = S\bar{a}$ is the effective absorbing surface, S is the total area in square feet, and V is the volume of room in cubic feet. If we compare Eqs. (5.9) and (5.16), we can write

$$\frac{KV}{\sum a_1 s_1} = \frac{0.05V}{\sum a_1 s_1}$$

Thus, $K = 0.05$ when the dimensions are taken in FPS system. The value of K will be equal to 0.165 if the units will be taken in MKS system. Thus, Eq. (5.16) in MKS system can be given as

$$T = \frac{0.165V}{\sum a_1 s_1} \quad (5.17)$$

5.9.1 Deductions of Sabine

Two important deductions were made by Sabine from a series of measurements:

- (i) The duration of reverberation is almost independent of the position of the source and of the absorber in the room.
 - (ii) The effect of any given amount of absorbent is also independent of its position in the room.
- Suitable times of reverberation for different events in a hall are given in Table 5.1.

Table 5.1 Time of reverberation in seconds at 500 Hz

Event	10^4 cu. ft.	$1.5 \times 10^4 \text{ cu. ft.}$	$2 \times 10^4 \text{ cu. ft.}$	$3 \times 10^4 \text{ cu. ft.}$
Choir	1.0	1.0	1.2	1.25
Orchestra	0.8	0.85	0.9	1.0
Speech or music	0.6	0.65	0.7	0.75

5.10 ABSORPTION COEFFICIENT

The sound absorption coefficient of a surface is defined as the ratio of absorption of sound energy by the surface to that of an equal area of a perfect absorber such as an open window. Thus, the absorption coefficient of any surface is given as

$$\alpha = \frac{\text{Sound energy absorbed by the surface}}{\text{Sound energy absorbed by an open window of same area}}$$

5.10.1 Determination of Absorption Coefficient

If T_1 is the reverberation time for an audio frequency source of sound in an empty room and Σas is the absorption due to the walls, ceiling, and floor of the room, then

$$T_1 = \frac{0.05V}{\Sigma as}$$

When a certain amount of absorbing material of area S and absorption coefficient α is introduced into the room, then the time of reverberation T_2 (say) is determined as

$$T_2 = \frac{0.05V}{\Sigma as + \alpha S}$$

Now from the above equations, we can write

$$\frac{1}{T_2} - \frac{1}{T_1} = \frac{\alpha S}{0.05V}$$

or $\alpha = \frac{0.05V}{S} \left[\frac{1}{T_2} - \frac{1}{T_1} \right]$ (5.18)

Sabine used an open window as his standard unit of absorption expressing other results in terms of open window units. The absorption coefficients for some materials are given in Table 5.2.

Table 5.2 Absorption coefficients of various materials

Materials	<i>a</i> (absorption coefficient)
Treetax	0.49
Lime plaster	0.02–0.04
Hard plaster	0.01–0.03
Red cloth	0.17
Felt (1/2" thick)	0.27
Asbestos	0.146
Fibre boards 1/2" thick	0.30–0.35
Acoustic plasters	0.27
Wood panelling	0.01–0.02
Porous rubber sheet	0.05–0.1
Audience per person	4.7
Chairs	0.17–0.21
Cushions $\left(2\frac{3}{4} \text{ sq. ft.} \right)$	1.5–1.8

5.11 FACTORS AFFECTING THE ARCHITECTURAL ACOUSTICS AND THEIR REMEDIES

Acoustically, good hall means a room (or close area) in which every syllable or musical note reaches an audible level of loudness at every point of hall and then quickly dies away to make room for the next syllable or group of notes. If in a hall, these conditions are not ensured, then hall will be acoustically defective. The factors affecting the architectural acoustics and their remedies are mentioned below:

1. *Control of reverberation:* Reverberation in a hall is caused by repeated reflections. However, it can be reduced by increasing the ways of absorption of sound in the room. We can achieve it by taking the following steps:
 - (i) By opening the window and ventilators. Because the windows and ventilators are perfect absorber, these reflect no sound wave.

- (ii) By covering the walls of hall by good absorbent such as draperies, acoustic plaster, and Celotex asbestos.
 - (iii) By using the chairs with cushions in the hall. Because these cushions provide high degree of absorption.
 - (iv) By ensuring sufficient numbers of audience in the hall. Because the cloths of audience are good absorber of sound wave. One person is equivalent to about 0.5 m^2 area of an open window. Due to their loose clothing ladies are better absorbers.
 - (v) By using the acoustic tiles in the hall.
2. *Adequate loudness*: In an acoustically good hall, each separate syllable should be heard sufficiently loud at every point in the hall. Because sufficient loudness in every portion of the hall is an important factor for satisfactory hearing, the sufficient loudness in a hall can be ensured by
- (i) using large sounding boards behind the speaker and facing the audience.
 - (ii) low ceilings are also helpful in reflecting the sound energy towards the audience.
 - (iii) using suitable loudspeakers for providing additional sound energy.
3. *Uniform distribution of sound*: The sound energy (waves) produced by source should be distributed uniformly in the hall. There should be neither silence region (poor audibility) nor focussing of sound energy in the hall. For uniform distribution of sound energy in the hall, following points are important:
- (i) There should be no curved surface (viz., concave, spherical, cylindrical, or parabolic) on the walls, ceiling, and at the floor of the hall. If such surfaces are present, they should be covered with absorbent materials.
 - (ii) The shape of hall should be parabolic and the place of speaker should be at its focus.
 - (iii) Low ceiling is also helpful in uniform distribution of sound.
 - (iv) *Freedom from resonance*: Sometimes, the window-pane sections of wooden portions and the wall lacking in rigidity are shown in resonant vibrations, with the sound produced in the room. It is also observed that the enclosed air in the hall also causes resonance. Such resonant vibrations should be suitably damped.
 - (v) *Absence of echoes*: When the time interval between the sound wave received directly and that after reflection is about $1/7$ second, then echoes are produced. Such disturbing echoes should be avoided or weakened as much as possible by absorption. The problem of echoes can be avoided by covering distant walls and high ceiling with absorbent materials.
 - (vi) *Absence of echelon effect*: Any regular spacing between reflecting surfaces may produce a musical note due to the regular succession of echoes of the original sound to the listener. Such effect is known as echelon effect in the sound waves. Sometimes this effect makes the original sound unintelligible. Such problems can be removed using a stair carpet.
 - (vii) *Extraneous noise and sound insulation*: Acoustically, good halls should be free from noise. Noise is the unwanted sound which may be produced due to high frequency of sound or intensity of sound or due to both. In general there are three types of noise which create much trouble to the audience. These are as follows:
 - (a) Air-borne noise
 - (b) Structure-borne noise
 - (c) Inside noise

In general, air-borne noise reaches the hall from the outside, through the open windows, doors, and ventilators. Air-borne noise can be minimised by taking the following steps:

6

Elements of Crystallography

6.1 INTRODUCTION

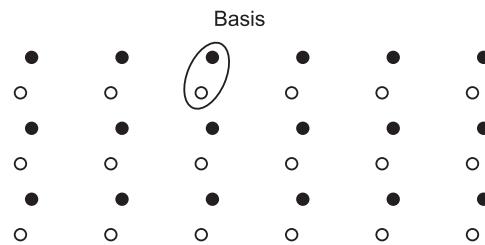
Usually all the elements and their chemical derivatives are found in three states viz, solid, liquid, and gaseous. At some specific conditions, sometimes they may lie in the fourth state of matter known as *plasma*, but it is rare. It is observed that nature loves symmetry due to which, starting from tiny particles like atoms and molecules to the giant particles like planet and stars, all are arranged in a regular fashion and obey their universal laws. The word *crystallography* is the combination of two words: *crystal* and *graphein*, where crystal has its usual meaning while the Greek word *graphein* means *to write*. Hence, crystallography means writing about the complete crystal. If the atoms and molecules in a solid are arranged in some regular fashion, then it is known as *crystalline solid*. The basic building blocks of a crystalline solid are stacked in a regular manner, just like the soldiers do on the parade ground, and thus form a three-dimensional pattern. Thus, the presence of long-range order and periodicity are the important properties of a crystal. In some crystals, there is a lack of long-range order in the arrangement of their constituent particles like atoms and molecules, and they are known as supercooled liquids or *amorphous solids*. The stiffness of such solids is due to their exceptionally high viscosity. The common examples of amorphous solids are glass, pitch, and plastics. The amorphous solids have no sharp melting point, whereas for crystalline solid, melting occurs at a definite temperature.

During his study on the properties of cathode rays in 1885, Professor Wilhelm Conrad Roentgen found that some highly penetrating radiations were coming from the discharge tube. Since the properties of these radiations were unknown at that time, he named these rays as X-rays. Later, X-rays were thoroughly investigated and it was established that they are electromagnetic waves. They have short wavelength, lying in the range of 0.01 Å to 10 Å. These X-rays are most commonly used to investigate crystal structure.

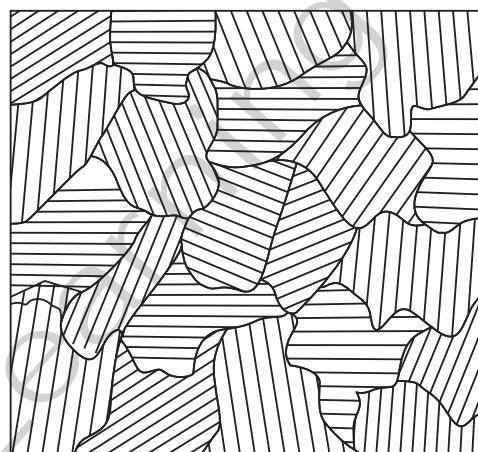
In this chapter, we will study about the basics of different crystal lattices, plane and directions of cubic system, and related parameters of crystal structures.

6.2 CRYSTAL STRUCTURE

In a crystalline solid, each basic building block (atoms or molecules) has a fixed position in space. It is at a definite distance and a definite angular orientation to the other atoms or molecules lying in its surroundings. In a perfect crystal, there is an infinite regular repetition of such identical structural units. In a single crystal, the position and orientation of atoms or molecules are uniform and continuous throughout the entire crystal (Fig. 6.1). The examples of single crystal are diamond, ruby, etc.

**Fig. 6.1** Structure of single crystal

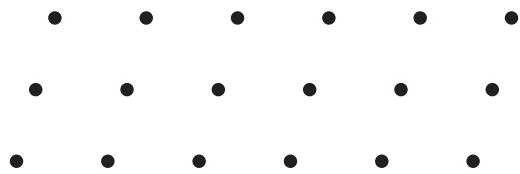
There is another class of material known as *polycrystalline* in which the whole crystal is made up of smaller crystallites, as shown in Fig. 6.2. Each constituent of the small crystallite is called *grain*. These grains are situated side by side to form the whole crystal, like in quartz. Within the boundary of a grain, there is same lattice pattern but grains are generally oriented randomly, and thus, form the *grain boundaries* in the polycrystalline solid. Since many properties of the solids are best studied with single crystals, so these are preferable than the polycrystalline solids. There are two types of single crystals known as *elemental* and *ionic*. The examples of elemental single crystals are Al, Fe, Cu, etc., while the examples of ionic single crystals are AgCl, CuSO₄, etc.

**Fig. 6.2** Grain boundaries of different grains in a polycrystalline solid

In order to describe a crystal structure, it is must that the learner should know what are lattice, lattice translation vector, and basis. Without the knowledge of these important facts, it is very difficult to describe the crystal structure. For the better understanding of these fundamentals of the crystal structure, we have discussed them in the following sections.

6.3 SPACE LATTICE

In order to describe the arrangement of atoms in a crystal, it is always convenient to describe the arrangement of imaginary points in space which have a definite relationship with the atoms of the crystals. Actually, these set of points provide a base for the crystal structure. Such three-dimensional arrangement of imaginary points in space that have identical points in its surroundings is known as *space lattice*. A two-dimensional space lattice is shown in Fig. 6.3.

**Fig. 6.3** Space lattice

In the above discussion, it is mentioned that a point is surrounded by identical points which means that the lattice has same appearance when viewed from a point r' in the lattice as when viewed from any other point r' with respect to some arbitrary origin. It is possible on the cost of repetition of pattern unit in all directions by means of translation operation of vector T . Pattern unit may be considered as basic building block of lattice, containing a small group of points. Translation operation vector T is given as:

$$\vec{T} = n_1 \vec{a} + n_2 \vec{b} + n_3 \vec{c} \quad (6.1)$$

where n_1 , n_2 , and n_3 are arbitrary integers and vectors \vec{a} , \vec{b} , \vec{c} are known as *fundamental translation vectors*. With respect to the other point \vec{r}' , Eq. (6.1) can be given as

$$\vec{r}' = \vec{r} + \vec{T} = \vec{r} + n_1 \vec{a} + n_2 \vec{b} + n_3 \vec{c} \quad (6.2)$$

The translation vectors are also known as *crystal axis*, or *basis vectors*.

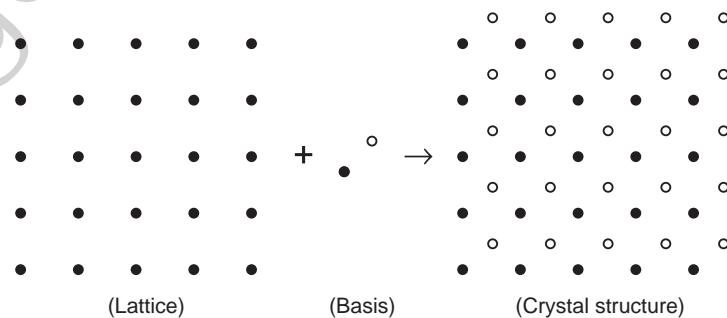
6.4 BASIS

A crystal structure is always described in terms of the atoms arranged in the lattice. A *space lattice* is defined as an infinite three-dimensional array of imaginary points in space which have identical surroundings. Thus, in order to obtain a crystal structure, an atom or a group of atoms must be placed on each lattice point in a regular fashion. Such an atom or a group of atoms placed on each lattice point is called the *basis*, and it acts as a building unit, or a structural unit, for the complete crystal structure. Thus, the crystal structure is generated by the combination of lattice with a basis. Thus, we can express a crystal structure as

$$\text{Space lattice} + \text{Basis} = \text{Crystal structure}$$

From the above discussion, it is clear that a lattice is a mathematical concept, whereas the crystal structure is a physical concept.

The generation of a crystal structure from a two-dimensional lattice and a basis is shown in Fig. 6.4. The basis consists of two atoms, represented by \circ and \bullet . The crystal structure is obtained by placing the basis on each lattice point such that the centre of the basis coincides with the lattice point.

**Fig. 6.4** Crystal structure generated from lattice and basis

6.5 UNIT CELL

A *unit cell* may be defined as the smallest unit of the lattice which, on continuous repetition, generates the complete lattice. Both primitive and non-primitive translation vectors may be used to construct a unit cell. Accordingly a unit cell is named as a primitive unit cell or a non-primitive unit cell. Some unit cells are shown in Fig. 6.5, out of which the parallelogram $ABCD$ represents a two-dimensional primitive cell, whereas the parallelograms $EFGH$ and $KLMN$ represent non-primitive cells. The smallest volume is known as *primitive unit cell*.

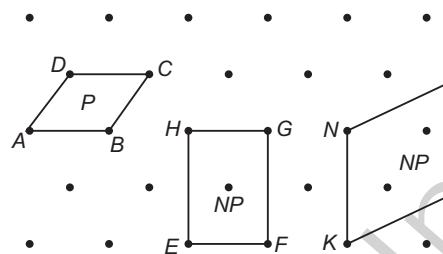


Fig. 6.5 Primitive (P) and nonprimitive (NP) unit cells of 2-D lattice

In a primitive cell, all the lattice points belonging to it lie at its corners. Thus, the effective number of lattice points in a primitive cell is one. In a non-primitive cell, lattice points lie at the corner as well as at other locations, both inside and on the surface of the cell. Thus, the effective number of lattice points in a non-primitive cell is greater than one. The distance between two atoms or ions of same type is known as the length of the unit cell.

6.6 LATTICE PARAMETER

In a three-dimensional structure, a unit cell is usually represented by a parallelepiped as shown in Fig. 6.6. As mentioned in Fig. 6.6, vectors \vec{a} , \vec{b} , and \vec{c} along x -, y -, and z -axes, respectively, are called *crystallographic axes*. The angles between these axes are called *interfacial angles*, which are measured as α , β , and γ . Primitive vectors \vec{a} , \vec{b} , and \vec{c} and the interfacial angles α , β , and γ together are known as *lattice parameter* of a crystal. The actual size of the unit cell can be determined with the help of vectors \vec{a} , \vec{b} , and \vec{c} and interfacial angles α , β , and γ (i.e., lattice parameter).

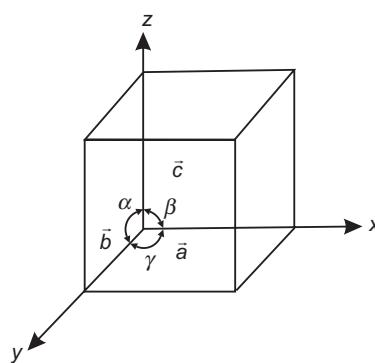


Fig. 6.6 Three-dimensional structure of unit cell

6.7 LATTICE VECTOR AND DIRECTION

Any particular point or specific direction in a unit cell can be determined with the help of basic vectors \vec{a} , \vec{b} , and \vec{c} . For example, if anyone wants to determine the position and the direction of point P in the unit cell shown in Fig. 6.7, then it can be expressed in terms of the basic vectors \vec{a} , \vec{b} , and \vec{c} as

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

where n_1 , n_2 , and n_3 are three integers.

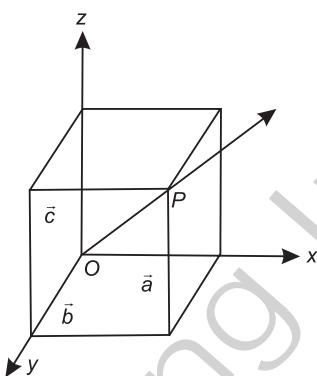


Fig. 6.7 Position and direction of a point with the help of vectors \vec{a} , \vec{b} , \vec{c}

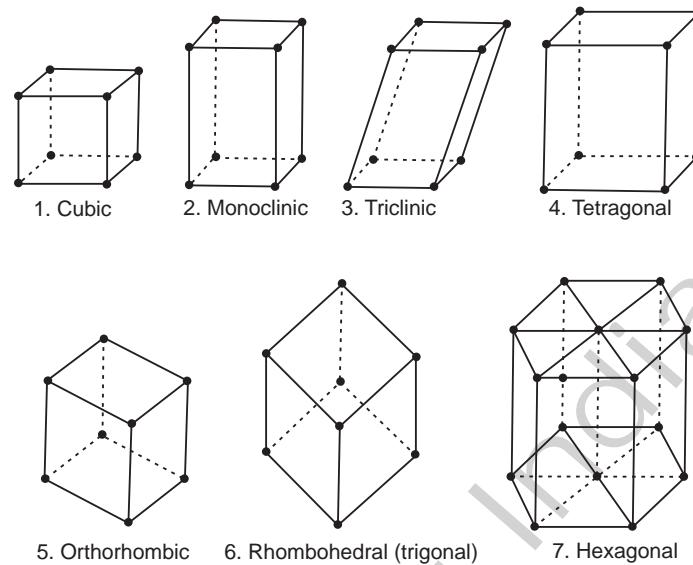
6.8 SEVEN CRYSTAL SYSTEMS

Corresponding to the different values of the lattice parameters, the crystal systems can be classified into seven types. Important information corresponding to these crystal systems is given in Table 6.1.

Table 6.1 Seven crystal systems and their lattice parameters

S. No.	Crystal System	Lattice parameters	Examples
1.	Cubic	$a = b = c$ $\alpha = \beta = \gamma = 90^\circ$	Cu, Ag, Fe, Na, NaCl
2.	Tetragonal	$a = b \neq c$ $\alpha = \beta = \gamma = 90^\circ$	β -Sn, TiO ₂
3.	Orthorhombic	$a \neq b \neq c$ $\alpha = \beta = \gamma = 90^\circ$	Ga, Fe ₃ C (cementite)
4.	Rhombohedral or Trigonal	$a = b = c$ $\alpha = \beta = \gamma \neq 90^\circ$	As, Sb, Bi
5.	Hexagonal	$a = b \neq c$ $\alpha = \beta = 90^\circ, \gamma = 120^\circ$	Mg, Zn, Cd, NiAs
6.	Monoclinic	$a \neq b \neq c$ $\alpha = \gamma = 90^\circ \neq \beta$	CaSO ₄ · 2H ₂ O (gypsum)
7.	Triclinic	$a \neq b \neq c$ $\alpha \neq \beta \neq \gamma \neq 90^\circ$	K ₂ Cr ₂ O ₇

Simple sketch of all the seven crystal systems are given in Fig. 6.8.

**Fig. 6.8** Seven crystal systems

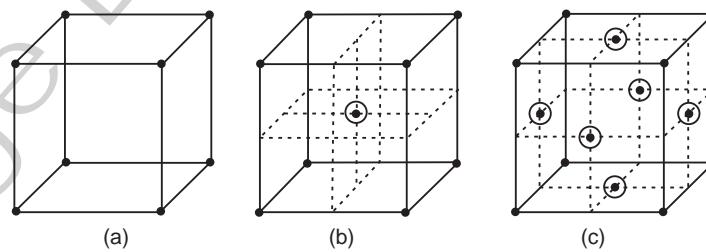
6.9 BRAVAIS SPACE LATTICE

During 1945, Bravais suggested that there are only 14 ways of arranging points in space. In such a situation, the system appears the same from each point. These arrangements give 14 lattices, which are called *Bravais space lattices*, each of which has the same collection of symmetry elements at the lattice points.

In the case of cubic system, there are three types of possible Bravais lattices which are as follows:

- (i) Simple cubic (SC)
- (ii) Body-centred cubic (BCC)
- (iii) Face-centred cubic (FCC)

The above mentioned Bravais lattices for cubic system are shown in Fig. 6.9.

**Fig. 6.9** (a) Simple cubic structure (SC), (b) body-centred cubic structure (BCC), and (c) face-centred cubic structure (FCC)

In the case of tetragonal system, there are following two types of Bravais lattices:

- (i) Simple tetragonal (ST)
- (ii) Body-centred tetragonal (BCT)

In the case of orthorhombic system, there are following four types of possible Bravais lattices:

- (i) Simple orthorhombic (SO)
- (ii) Body-centred orthorhombic (BCO)
- (iii) End-centred orthorhombic (ECO)
- (iv) Face-centred orthorhombic (FCO)

In the case of rhombohedral or trigonal system, there is only one possible Bravais lattice known as simple rhombohedral (SO).

In the case of hexagonal system, there is only one possible Bravais lattice known as simple hexagonal (SH).

Monoclinic system has two possible bravais lattices known as simple monoclinic (SM) and end-centred monoclinic (ECM). Triclinic crystal system has only one possible lattice known as simple triclinic (ST). Thus, the total number of possible Bravais lattices corresponding to all the systems are 14.

6.10 SYMMETRY ELEMENT OF A CRYSTALLINE SOLID

There are three main symmetry elements used to describe a crystalline solid.

- (i) Axis of symmetry
- (ii) Plane of symmetry
- (iii) Centre of symmetry

6.10.1 Axis of Symmetry

During the rotation of a crystal around an axis, if it occupies two or more identical positions in a complete rotation (i.e., 360°), then that axis is called the *axis of symmetry*.

When a cube is rotated around the axis passing through the opposite faces, then each rotation of 90° repeats the identical or congruent position, i.e., four congruent positions in one complete rotation. It is known as *four-fold axis of symmetry*. In general, if in a crystal n , the identical positions are repeated around an axis in one complete rotation, then that axis is called an n -fold axis of symmetry.

There are 13 axes of symmetry for a cube, which are as follows:

- (i) 3 axes have four-fold symmetry known as *tetrad*. These axes pass through opposite face centres as shown in Fig. 6.10 (a).
- (ii) 4 axes have three-fold symmetry known as *triad*. These axes pass through diagonally opposite corners as shown in Fig. 6.10 (b).
- (iii) 6 axes have two-fold symmetry known as *diad*. These axes pass through the centres of opposite edges as shown in Fig. 6.10 (c).

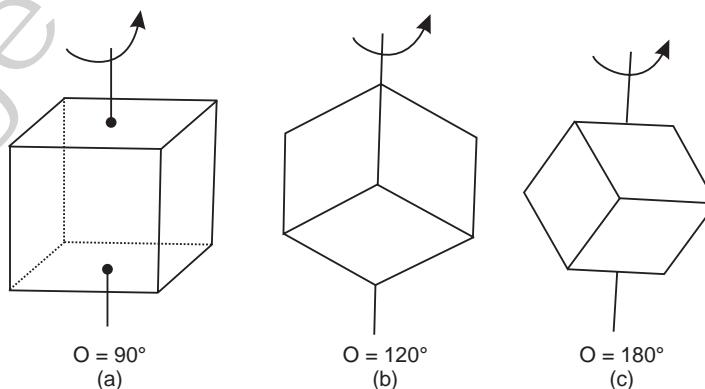


Fig. 6.10 (a) 4-fold axes of symmetry, (b) 3-fold axes of symmetry, and (c) 2-fold axes of symmetry

6.10.2 Planes of Symmetry

If a plane is able to cut a crystal into two halves in such a way that one half becomes the mirror image of the other half, then that plane is known as the *plane of symmetry*. There are 9 planes—3 are parallel to the face of the cube [Fig. 6.11(a)] and 6 are diagonal planes passing through diagonally opposite parallel edges [Fig. 6.11(b)].

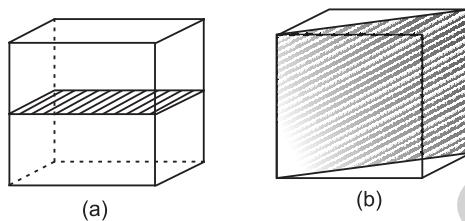


Fig. 6.11 (a) Parallel plane of symmetry and **(b)** diagonal plane of symmetry

6.10.3 Centre of Symmetry

It is defined as a point in a crystal such that if a line is drawn from any point on the crystal through this point and an equal distance is produced on the other side of this central point, then it meets an identical point. There is only one centre of symmetry for cubic system (Fig. 6.12).

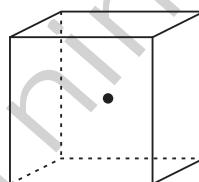


Fig. 6.12 Centre of symmetry in cubic system

6.11 COORDINATES OF LATTICE POINT

The coordinates of the corner lattice points for simple cube, body-centred cube, and face-centred cube systems may be considered as (0,0,0). The coordinates of any one of such points can be considered as the origin. Considering the origin, the coordinates of the other lattice points can be determined [Fig. 6.13(a)]. The coordinates of the body-centred lattice point in BCC system are (1/2, 1/2, 1/2) [Fig. 6.13(b)]. Similarly, the coordinates of the face-centred lattice points in FCC system are (1/2, 1/2, 0), (0, 1/2, 1/2), and (1/2, 0, 1/2) [Fig. 6.13 (c)].

6.12 NUMBER OF ATOMS PER UNIT CELL

The number of atoms per unit cell gives the idea of how the atoms are packed in the solid.

In the simple cubic crystal, there are 8 atoms—one at each corner, and each corner atom is shared by 8 unit cells. Thus, the total number of atoms per unit cell is given as $N_c/8 = 8/8 = 1$ as shown in Fig. 6.14 (a), where N_c is the number of corner atoms.

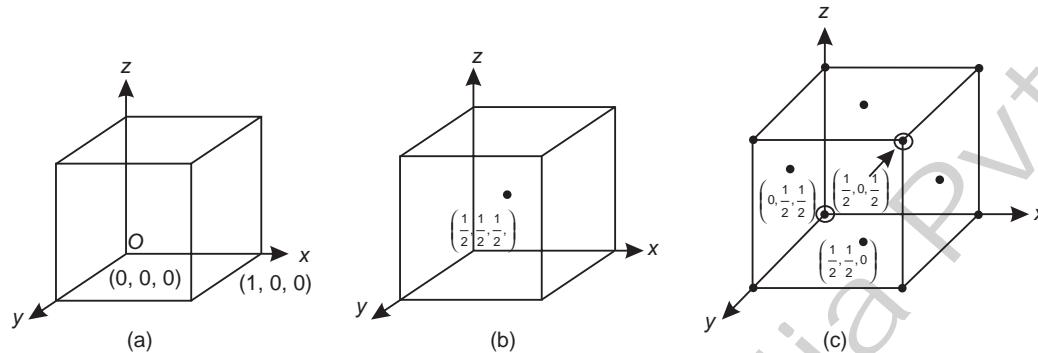


Fig. 6.13 (a) Coordinates of corner lattice points, (b) coordinates of body-centred lattice points, and (c) coordinates of face-centred lattice points

In the body-centred cubic crystal, there is one atom at the centre of the cell and 8 atoms are at the corner. Similar to SC crystals, each corner atom is shared by 8 unit cells. Therefore, the total number of atoms per unit cell is given as

$$N_B + \frac{N_C}{8} = 1 + 1 = 2$$

where N_B is the number of body-centred atoms. The arrangement of the atoms per unit cell in a body-centred crystal is shown in Fig. 6.14 (b).

In the face-centred cubic crystal, there are 6 atoms at the centre of faces (one face has one atom) and 8 atoms at the 8 corners. Each corner atom is shared by 8 unit cells and each of the 6 face-centred atoms are shared by 2 unit cells. Thus, the total number of atoms can be given as

$$\frac{N_C}{8} + \frac{N_F}{2} = \frac{8}{8} + \frac{6}{2} = 1 + 3 = 4$$

where N_F is the number of face-centred atoms. The arrangement of the atoms in the face-centred cubic crystal is shown in Fig. 6.14 (c).

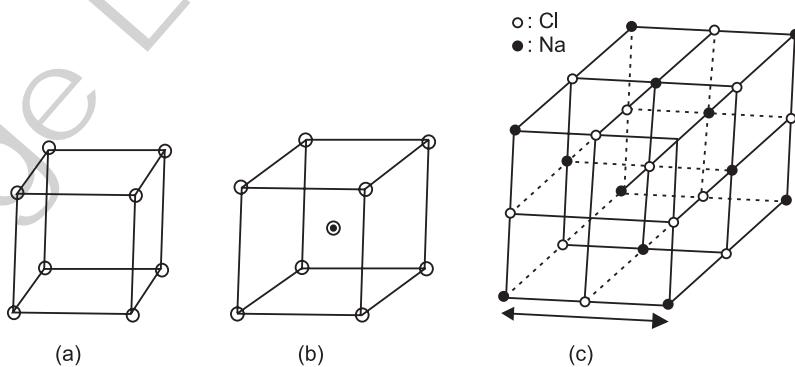


Fig. 6.14 (a) Number of atoms per unit cell in simple cubic system, (b) number of atoms per unit cell in body-centred cubic system, and (c) number of atoms per unit cell in face-centred cubic system

6.13 PACKING DENSITY OR ATOMIC PACKING FACTOR

The packing efficiency of atoms or ions in a crystal gives an important concept in the study of crystal structure. Higher packing density provides efficient packing of atoms in solids. How the atoms or ions are packed in a crystal can be defined 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 in the unit cell to the total volume of the unit cell.

$$\text{Atomic packing factor} = \frac{\text{Volume occupied by the atoms in the unit cell}}{\text{Volume of the unit cell}}$$

In the case of ionic crystal, the ionic packing factor is expressed as

$$\text{Ionic packing factor} = \frac{\text{Volume occupied by the ions in the unit cell}}{\text{Volume of the unit cell}}$$

Atomic packing factors for SC crystal, BCC crystal, and FCC crystal are given as $\pi/6$, $\sqrt{3}\pi/8$, and $\sqrt{2}\pi/6$, respectively.

6.14 MILLER INDICES

A crystal lattice consists of a large number of parallel equidistant planes passing through the lattice points. These planes are known as *lattice planes*. For a given crystal, the lattice planes can be chosen in different ways as shown in Figs. 6.15 (a), (b), (c), (d), and (e).

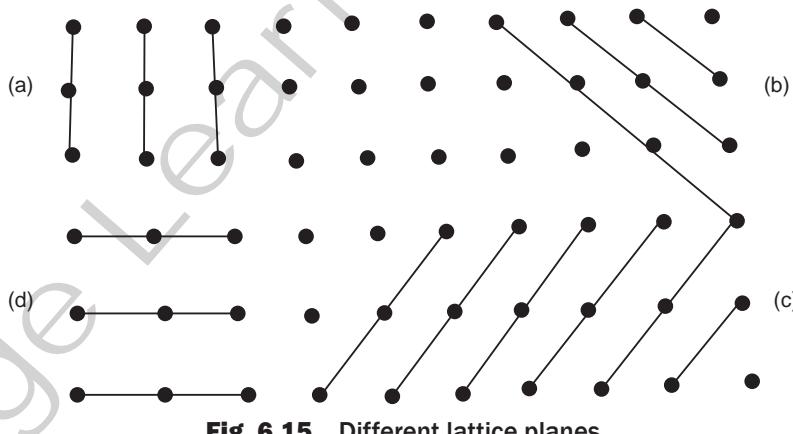


Fig. 6.15 Different lattice planes

Miller introduced a system to designate a plane in a crystal. He suggested that it is more useful to describe the orientation of a plane by the reciprocal of its numerical parameters rather than by its linear parameters. The reciprocals when appropriately converted to whole numbers are called *Miller indices*. In other words, *Miller indices* are defined as three smallest possible integers which have the same ratios as the reciprocal of intercepts of the plane concerned on the three axes.

7

X-Rays and Compton Effect

7.1 INTRODUCTION

During his study on the properties of cathode rays in 1885, Professor Wilhelm Conrad Röntgen found that some highly penetrating radiations were coming from discharge tube. Since the properties of these radiations were unknown at that time, he named these rays as X-rays. Later, X-rays were thoroughly investigated, and it was established that they are electromagnetic waves. They have short wavelength, lying in the range from 0.01 \AA to 10 \AA .

In this chapter, we shall discuss X-rays and their diffraction, Bragg's law and its applications, and Compton effect.

7.2 PRODUCTION OF X-RAYS

When electrons with high value of kinetic energy collide with the target metal of high melting point and large atomic weight, X-rays are produced. A gas-filled tube, Coolidge tube, and betatrons are generally used for the production of X-rays.

In 1913, Dr Coolidge introduced a new type of tube which is now most commonly used to produce X-rays (Fig. 7.1).

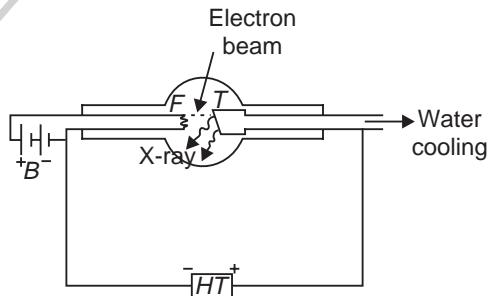


Fig. 7.1 Coolidge tube producing X-rays

In this tube, filament F is heated to produce electrons by thermionic emission. These electrons are accelerated by means of applied potential difference through HT, and then incident on target T. These

fast moving electrons, after striking the target, produce X-rays. Since most of the kinetic energy of electrons is absorbed by the target metal in the form of heat, a cooling system is required to maintain the temperature. The intensity of X-rays (the number of electrons emitted per second) is controlled by the filament battery, and the quality (penetrating power) of X-rays is controlled by the high-tension battery.

7.3 X-RAY SPECTRA

When the intensities of different wavelengths of X-rays coming out from any source are plotted with wavelength, in general, two types of spectra, continuous and characteristic spectra, are obtained.

7.3.1 Continuous X-Ray Spectra

Continuous X-ray spectra consist of radiations of all possible wavelengths within a certain range. The intensity versus wavelength curve for different value of accelerating voltage is found as shown in Fig. 7.2.

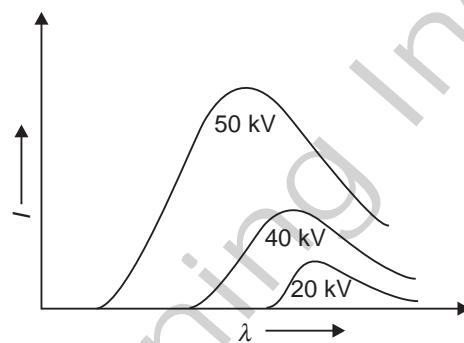


Fig. 7.2 Continuous X-ray spectra

The total intensity of continuous X-ray radiations for a given target is proportional to V^2 . The minimum wavelength λ_{\min} is independent of the nature of the target and is directly related to accelerating potential (V) as

$$Ve = hV_{\max} = \frac{hc}{\lambda_{\min}} \quad (7.1)$$

Origin of continuous X-ray spectra: As it is discussed in preceding sections that when high-velocity electrons hit a suitable target, X-rays are produced. As the target atom is of large atomic weight, the positive charge of the nucleus is completely shielded by the surrounding electrons (Fig. 7.3).

When incoming electron reaches near to the target atom, it experiences a strong repulsive coulomb force and is suddenly slowed down and also suffer deflection in path. During retardation the electron emits a pulse of electromagnetic radiation of energy $h\nu$. Due to different losses in velocity of the incident electrons, radiations of all possible wavelengths are emitted which give a continuous spectrum of X-rays. The continuous radiation is also termed as *bremsstrahlung radiations* or *braking radiations*.

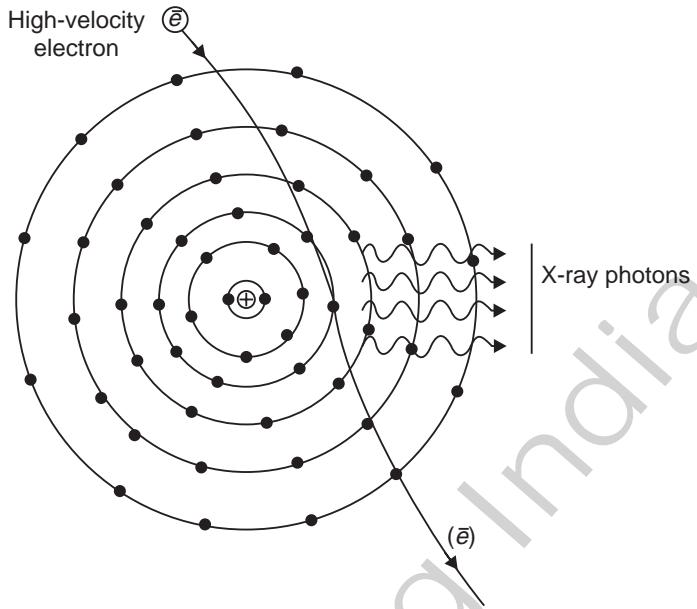


Fig. 7.3 Interaction of high-velocity electron with target atom to produce continuous X-ray spectra

In order to calculate the wavelength of emitted X-ray photon, let us consider an electron whose charge is e , accelerated by a potential V . This electron will gain an energy eV , which must be equal to its kinetic energy. Thus,

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

or $v = \sqrt{\left(\frac{2eV}{m}\right)}$ (7.2)

where v is the velocity of striking electron of mass m .

If the entire energy of electron is converted into the energy of X-rays, then

$$\frac{1}{2}mv^2 = eV = h\nu$$

or $\nu = \frac{eV}{h}$ (7.3)

ν is the maximum value of frequency of X-ray photon. We know that

$$c = \nu\lambda$$

Hence, $\lambda_{\min} = \frac{c}{\nu_{\max}} = \frac{hc}{eV}$ (7.4)

7.3.2 Characteristic X-Ray Spectra

Characteristic X-ray spectra consist of a few peaks superimposed on the continuous spectrum. The number of peaks and their corresponding wavelengths are the characteristic of material used as target. These peaks generally occur in groups and are named as K peaks, L peaks, and so on from the shorter wavelength side to the longer wavelength side. Characteristic X-ray spectra is shown in Fig. 7.4.

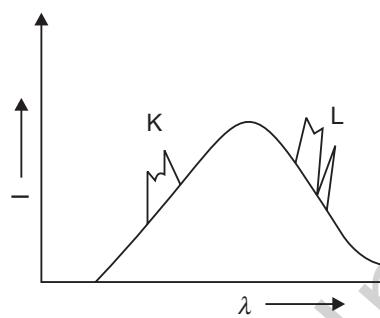


Fig. 7.4 Characteristic X-ray spectra

Origin of characteristic X-ray spectra: Sometimes it happens that the incident high-velocity electron penetrates through the outer shells of the target atom and strikes with an electron in the inner shell (Fig. 7.5). The collision is so strong that an electron from the inner shell is knocked out. Definitely a vacancy is created and to fill it, an electron from higher energy state jumps into this vacancy. The difference in energy is emitted as X-rays photon of energy $\Delta E = h\nu$. This process further creates a vacancy in higher state which is again filled up by an other electron jumping from the next higher state.

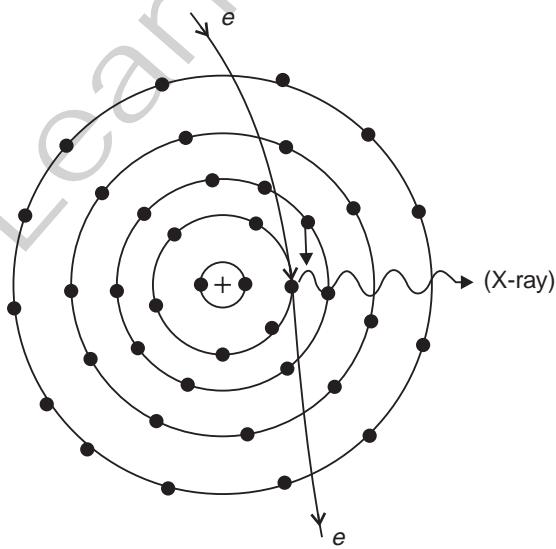


Fig. 7.5 Origin of characteristic X-ray spectra

In this way, another photon of X-ray is emitted out with slightly less energy. The process goes on till the highest shell is reached. Thus, we get a series of lines giving characteristic spectra electrons which after striking the target produce X-rays. Since most of the kinetic energy of electrons is absorbed by the target metal in the form of heat, a cooling system is required to maintain the temperature. The intensity of X-rays (the number of electrons emitted per second) is controlled by filament battery, and the quality (penetrating power) of X-rays is controlled by the high-tension battery.

7.4 IMPORTANT PROPERTIES OF X-RAYS

Some important properties of X-rays are as follows:

- (i) X-rays are electromagnetic radiations like light. Thus, they possess all the properties of electromagnetic radiations and exhibit phenomena such as reflection, refraction, diffraction, and interference.
- (ii) X-rays have very short range of wavelength, lying between 0.01 \AA and 10 \AA .
- (iii) X-rays are absorbed by the material.
- (iv) X-rays are not deviated by either electric or magnetic field.
- (v) X-rays ionise the gas through which they pass.
- (vi) X-rays can penetrate the solid materials.
- (vii) X-rays affect photographic plate.
- (viii) Secondary radiations arise when X-rays strike a metal.
- (ix) X-rays have harmful effects on human tissues. Therefore, a continuous exposure to these rays must be avoided.
- (x) X-rays cause fluorescence.

7.5 DIFFRACTION OF X-RAYS

When X-rays were discovered, their nature was not clear to the scientists. Some scientists considered them as high-speed particles while others considered them as electromagnetic waves of short wavelength. For testing the wave nature of X-rays, scientists tried to obtain the diffraction pattern of X-rays. But due to short wavelength of X-rays, man-made transmission gratings (ruled with 6000 lines per inch) could not produce any appreciable amount of diffraction with X-rays. Thus, it was concluded that the diffraction of X-rays with man-made transmission gratings is practically impossible. As we know, the diffraction pattern of a wave is obtained only if the aperture is of the order of the wavelength of the diffracting wave. Therefore, transmission gratings having 40 million lines per centimetre are required for diffracting X-rays.

In 1913, German physicist Max Von Laue suggested that atoms in a natural crystal are regularly arranged at equal distances. The separation between successive layers is of the order of the wavelength of X-rays. Thus, natural crystals can be used as grating for diffraction of X-rays. However, crystals act as space gratings instead of plane gratings because the diffraction centres in the crystals are not in one plane.

7.6 LAUE'S EXPERIMENTAL DEMONSTRATION

In 1913, Laue, in collaboration with W Friedrich and P Knipping, became successful in obtaining diffraction pattern of X-rays passing through a three-dimensional crystal grating. The experimental arrangement of Laue's demonstration is shown in Fig. 7.6. X-rays are produced from the Coolidge tube and are

collimated into fine pencil beam by passing through pin holes, S_1 and S_2 . The beam is allowed to pass through a properly oriented crystal of zinc sulphide or sodium chloride, which is acting as a grating. Diffraction pattern is obtained at photographic plate. This pattern consists of a central spot, surrounded by a series of fainter spots in a definite pattern. The symmetrical pattern of spots is called *Laue's pattern* or *Laue's spots*.

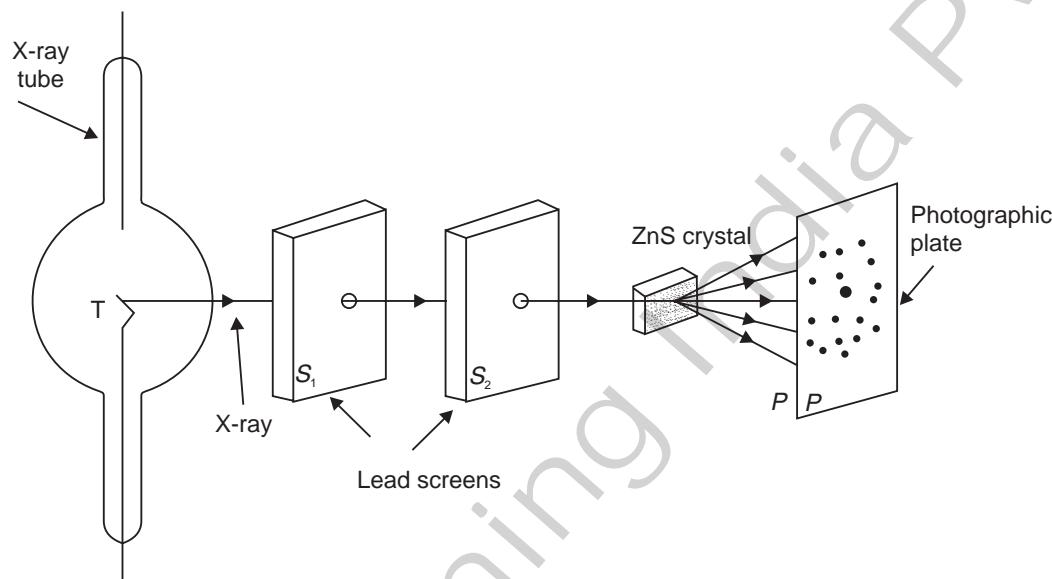


Fig. 7.6 Laue's experiment

Laue's spots prove that X-rays are electromagnetic waves. The proper explanation of Laue's spots was given by W.L. Bragg. He said that these spots correspond to the constructive interference between the rays reflected from the various sets of parallel crystal planes, thereby satisfying the equation, $2d \sin \theta = n \lambda$, where θ is the glancing angle, d is the inter-atomic space, and λ is the wavelength of X-rays. Laue's experiment established the following two important facts:

- (i) X-rays are electromagnetic waves of short wavelength.
- (ii) In crystals, atoms are arranged in a three-dimensional lattice.

7.7 BRAGG'S LAW

Bragg considered that a crystal is made up of a number of parallel planes on which atoms are arranged in a regular fashion. When X-rays are allowed to fall on the crystal, reflections from different planes take place. These reflected rays interfere and produce diffraction pattern. The diffraction of X-rays through a crystal is shown in Fig. 7.7.

Let us consider a narrow beam of X-rays of wavelength λ . It is allowed to incident on a crystal of parallel lattice planes having inter-atomic (lattice planes) separation of d . Let us assume the glancing angle to be θ . We know that incident rays are reflected from various parallel planes of atoms in the crystal. In the experiment, only a small fraction of radiation is reflected. The diffraction pattern of reflected radiation is observed only when reflections from various planes of atoms interfere constructively.

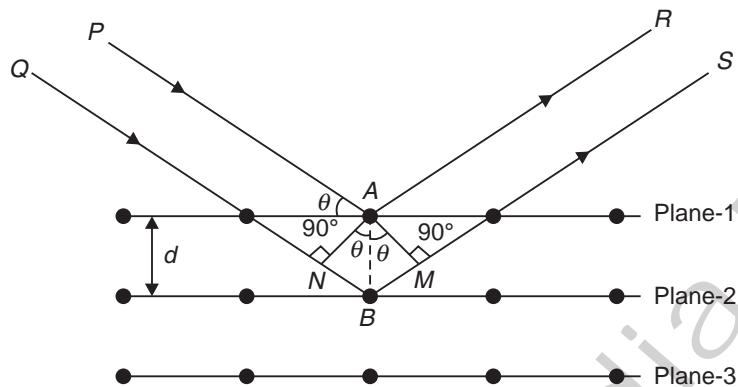


Fig. 7.7 Diffraction of X-rays through a crystal lattice

Let us suppose that ray PA is reflected from atom A in the direction AR and ray QB is reflected from atom B in the direction BS (Fig. 7.7). Now, if the path difference between AR and BS satisfies the condition of constructive interference, then a diffraction pattern will be observed. In order to realise the condition of constructive interference, let us first calculate the path difference between the rays reflected along AR and BS . Let us draw perpendiculars AM on BS and AN on QB . Now, the path difference between the reflected rays AR and BS can be given as $(NB + BM)$. From Fig. 7.7, it is clear that

$$NB = BM = d \sin \theta$$

Hence, the path difference can be given as

$$(NB + BM) = 2d \sin \theta$$

If this path difference is an even multiple of $\lambda/2$ (i.e., $n \lambda$), then the two rays will reinforce each other and produce an intense spot. Thus, the condition of reinforcement can be given as

$$2d \sin \theta = n \lambda \quad (7.5)$$

where $n = 1, 2, 3$, etc., for the maxima of first order, second order, third order, and so on. Equation (7.5) is known as *Bragg's law*.

For the same crystal and the same incident X-rays, different maxima can be obtained corresponding to different directions (angles). The conditions for these maxima are mentioned below:

$$\text{First maxima: } \sin \theta_1 = \frac{\lambda}{2d}$$

$$\text{Second maxima: } \sin \theta_2 = \frac{2\lambda}{2d}$$

$$\text{Third maxima: } \sin \theta_3 = \frac{3\lambda}{2d}$$

and so on.

It is worth mentioning here that the intensities go on decreasing as the order of the spectrum increases.

7.8 BRAGG'S SPECTROMETER

For the structural studies of crystals, W.H. Bragg and his son W.L. Bragg devised a spectrometer in which a crystal is used as a reflection grating instead of a transmission grating. With the help of their spectrometer, interatomic separation in the crystal can be calculated.

Experimental arrangement: Experimental arrangement of Bragg's spectrometer is shown in Fig. 7.8. It consists of a source of X-rays S , slits S_1 and S_2 , and crystal C mounted on a prism table. A round scale, consisting of two vernier scales V_1 and V_2 to note down the angle, is attached with the prism table. An ionisation chamber is attached with the prism table, along with a galvanometer.

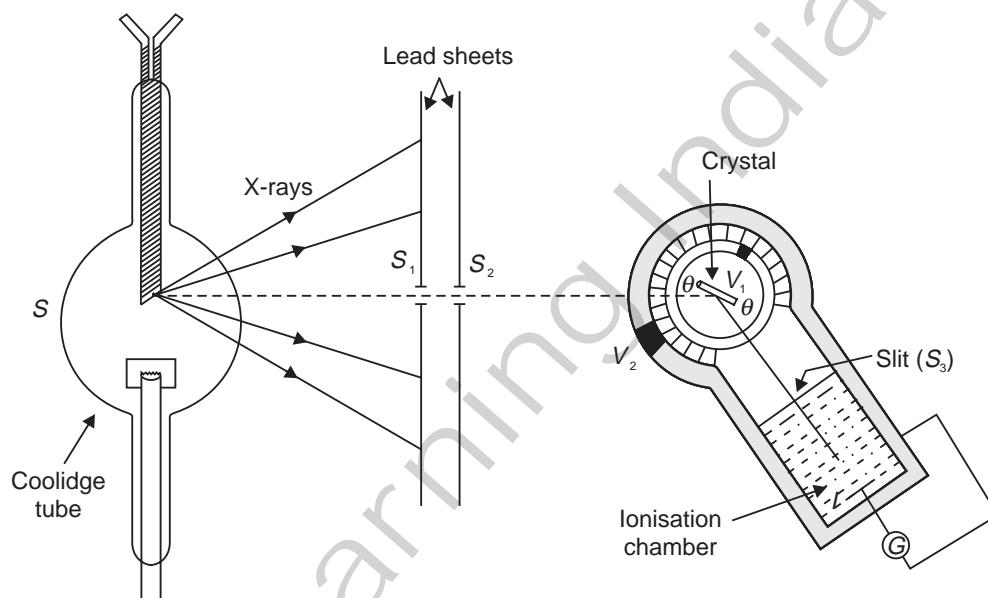


Fig. 7.8 Bragg's spectrometer

Working: X-rays from the Coolidge tube are passed through slits S_1 and S_2 to obtain a fine pencil of beam. Now, this narrow beam is allowed to fall on the crystal C mounted on the prism table, capable of rotating about a vertical axis passing through the centre of the prism table. The X-ray beam after reflection from the crystal face enters the ionisation chamber I through slit S_3 . The ionisation chamber I is filled with ethyl bromide and is capable of rotating with the prism table or crystal. The reading of second vernier scale gives the position of the ionisation chamber. The position of crystal and ionisation chamber is arranged in such a way that a rotation by angle θ in the position of crystal produces the rotation of 2θ in the position of ionisation chamber. Due to this arrangement, the X-ray beam reflected from the surface of the crystal is always received in the ionisation chamber. The intensity of X-rays in terms of ionisation current is observed for different values of glancing angle. The resulting ionisation current is observed through a galvanometer.

The graph plotted between the ionisation current I and glancing angle θ for the sodium chloride crystal is shown in Fig. 7.9.

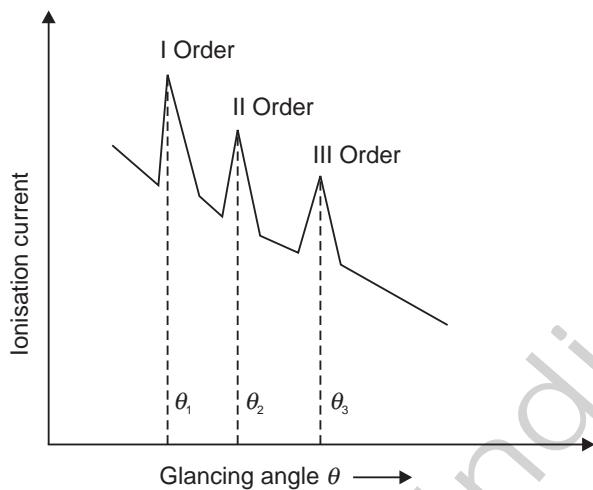


Fig. 7.9 Plot between ionisation current (I) and glancing angle (θ) for NaCl

Analysis: It is clear from the graph (Fig. 7.9) that for certain values of glancing angle θ , the ionisation current I increases abruptly. The number of peaks in the curve corresponds to those glancing angles which satisfy Bragg's equation, i.e.,

$$2d \sin \theta = n \lambda$$

$$\text{For } n = 1, \lambda = 2d \sin \theta_1$$

$$n = 2, 2\lambda = 2d \sin \theta_2$$

$$n = 3, 3\lambda = 2d \sin \theta_3$$

Hence, $\sin \theta_1 : \sin \theta_2 : \sin \theta_3 = 1 : 2 : 3$. Thus, out of θ , λ , and d , if two are known, the value of the third one can be calculated. From the experimental observations, the following facts have been established:

- (i) As the order of spectrum increases, the intensity of the reflected X-rays decreases.
- (ii) The ionisation current does not fall to zero for any value of glancing angle θ , but it does attain a maximum value for certain glancing angles. It indicates that there is continuous spectrum over which the characteristic line spectrum is superimposed.

7.9 BRAGG'S LAW AND CRYSTAL STRUCTURE

The crystal structure of potassium chloride (KCl) can be analysed with the help of Bragg's law. It is observed that in this crystal, maxima of reflected X-rays occur at glancing angles 5.22° , 7.30° , and 9.05° , respectively, using three different reflecting planes.

From Bragg's law, we have

$$2d \sin \theta = n \lambda$$

If $n = 1$, i.e., for the first order, the above equation becomes

$$2d \sin \theta = \lambda$$

or $\frac{1}{d} = \frac{2\sin\theta}{\lambda}$

or $\frac{1}{d_1} : \frac{1}{d_2} : \frac{1}{d_3} :: \sin 5.22^\circ : \sin 7.30^\circ : \sin 9.05^\circ$

or $\frac{1}{d_1} : \frac{1}{d_2} : \frac{1}{d_3} :: 0.091 : 0.127 : 0.157$

or $\frac{1}{d_1} : \frac{1}{d_2} : \frac{1}{d_3} :: 1 : 1.4 : 1.73$

On the basis of the above result, we can conclude that KCl is a simple cubic crystal. Hence, we can analyse the different types of crystal structure using Bragg's law.

7.10 COMPTON EFFECT

Compton effect gives direct and conclusive evidence in support of the particle nature of electromagnetic radiations. Compton explained this effect by treating X-rays as a stream of photons (or particles which scattered after the collision of electrons).

Statement: When a monochromatic beam of X-rays (or the electromagnetic radiations of short wavelength) of wavelength λ is allowed to incident on scattering material, the scattered beam contains radiations of longer wavelength λ' in addition to the radiations of incident wavelength λ . The difference between λ' and λ , i.e., $\lambda' - \lambda$, is known as *Compton shift*, and this effect is called *Compton effect*.

The Compton shift does not depend on the wavelength of incident radiation and nature of scattering material. It depends on the scattering angle only.

7.10.1 Expression for Compton Shift

In order to derive an expression for change in wavelength of the scattered X-rays, Compton made some basic assumptions as given below:

- Incident X-rays or electromagnetic radiations consist of photons of same energy, i.e.,

$$E = h\nu$$

Momentum (P) of electron can be calculated from the relativistic energy expression,

$$E^2 = P^2c^2 + m_0^2 c^4$$

where m_0 is the rest mass of the particle. Since we are taking photons into consideration, the rest mass of photons will be considered in the above expression. But we know that the rest mass of photon is zero, hence the above expression becomes

$$E^2 = P^2c^2 + 0$$

or $E = Pc$

or $P = \frac{h\nu}{c}$ (7.6)

- Scattered electrons are treated as free particles because the energy of X-ray photons is very large (~ 10 keV) as compared to the binding energy of the electrons in the atom.

8

Band Theory of Solids and Semiconducting Materials

8.1 INTRODUCTION

The free electron theory was successful in explaining various electronic and thermal properties of metals such as heat capacity, thermal conductivity, electrical conductivity, magnetic susceptibility, etc. But there are several other properties for which this model becomes helpless. For example, this theory fails to give the answer of *this* question that “why some materials are conductors while some are insulators.” It also cannot account for the variation of resistivity with temperature for the latter type of materials. Furthermore, this theory is unable to explain the properties of semiconductors and many other transport properties, specially magnetotransport.

The main reason for the failure of the free electron theory was the oversimplified assumption according to which the conduction electron in a metal experiences a constant or a zero potential due to the ion cores and hence, it is free to move throughout the crystal. After the failure of this theory, more attention was given to the motion of an electron in a crystal lattice and its response to an applied electric field. A focussed study on the motion of an electron in the crystal lattice concludes that the potential due to ion cores is not constant and may change with the position of the electron in the crystal at the same time. Some potential is also contributed by the other electrons present in the crystal. Thus, the actual nature of the potential under which an electron moves in the crystal is very complicated.

During its motion in the crystal, an electron experiences a periodic potential with period equal to the lattice constant. This periodic potential experienced by the electron forms the basis of the band theory of solids. According to the band theory of solids, the electrons in crystals are arranged in energy bands which are separated by energy gap regions where no wavelike electron orbitals exist. Such forbidden regions are called *energy band gaps*. The formation of these bands is due to the interaction of the conduction electron waves with the ion cores of the crystal.

In this chapter, we will discuss the classification of solids on the basis of the band theory of solids. Proper attention will be given to the energy band structure of conductors, semiconductors, and insulators. We will also describe the types of diodes, LED, solar cell, photovoltaic cell, photo conductivity, and Hall's effect.

8.2 BASIS OF THE BAND THEORY OF SOLIDS

A focussed study on the thermal, electrical, and magnetic properties of materials suggests that the band theory of solids is due to the following reasons:

- (i) Some solids are good conductors of electricity, some are semiconductors, and others are insulators.
- (ii) There is variation in resistivity with temperature, especially, in case of semiconductors.
- (iii) The potential experienced by an electron during its motion through the crystal is periodic, instead of being constant or zero.

In view of the above facts, it is clear that the motion of an electron in periodic potential needs more emphasis to establish the band theory of solids. The simplest quantum mechanical view of the potential experienced by an electron in passing through the crystal is perfectly periodic as shown in Fig. 8.1. The distribution of potential between the nuclei is a series of humps and increases at the boundary of the solids. The periodicity of the potential is due to the coulomb interaction between the moving electrons and the periodic charge distribution arising from the positive ions situated on the lattice sites. Hence, the periodicity of the potential is equal to the periodicity of the lattice.

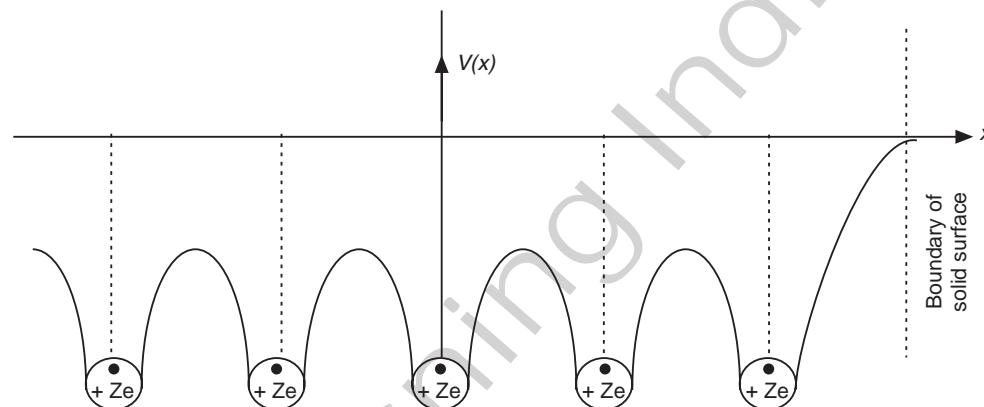


Fig. 8.1 One-dimensional representation of potential experienced by an electron in a perfectly periodic crystal lattice

8.3 ELECTRONS IN A PERIODIC POTENTIAL OF ION CORES

The behaviour of an electron in periodic potential is described by constructing the electron wave functions using one-electron approximation. In this approximation, the total wave function for the system is obtained from the combination of wave functions, each one of which involves the coordinates of one electron. Thus, the net field experienced by an electron is assumed to be the resultant of the field due to the ion cores and the average field due to the charge distribution of all other electrons. The motion of an electron in the ion core can be described by the Kronig–Penney model and the wave function associated with the electron is given by Bloch functions. The analysis of the behaviour of an electron in the periodic lattice in the view of Kronig–Penney model and Bloch theorem yields the following results:

- (i) There are allowed energy bands separated by forbidden regions or band gaps.
- (ii) The electronic energy functions $E(k)$ are periodic in the wave vector k .

8.3.1 Bloch Theorem

According to the free electron theory, an electron experiences constant potential (V_0), throughout the crystal lattice. One-dimensional Schrödinger wave equation for such electron is given as

$$\frac{\partial^2 \Psi}{\partial x^2} + \frac{2m}{\hbar^2} (E - V_0) \Psi = 0 \quad (8.1)$$

where ψ is the wave function associated with the electron of mass M and total energy E . The solution of this equation can be given as

$$\psi(x) = e^{\pm ikx} \quad (8.2)$$

and $(E - V_0) = \frac{\hbar^2 k^2}{2m} = \frac{p^2}{2m} = E_k$ (8.3)

Now, if we consider that the electron is moving through one-dimensional periodic potential $V(x)$, then the Schrödinger wave equation can be given as

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{2m}{\hbar^2} [E - V(x)] \psi = 0 \quad (8.4)$$

Since the potential is periodic with period equal to the lattice constant a , so $V(x)$ can be given as

$$V(x) = V(x + a)$$

The solution of Eq. (8.4) can be given by the Bloch theorem or Floquet's theorem, according to which, "the solution of Eq. (8.4) are the plane waves similar to Eq. (8.2), which are modulated by the function $u_k(x)$ having the same periodicity as that of the lattice." Hence, the solution of Eq. (8.4) will be of the form given as

$$\psi(x) = e^{\pm ikx} u_k(x) \quad (8.5)$$

with $u_k(x) = u_k(x + a)$ (8.6)

The wave functions of the type given in Eq. (8.6) are called the Bloch functions.

8.3.2 Kronig–Penney Model

Kronig–Penney model illustrates the behaviour of electrons in a periodic potential, considering simply one-dimensional model of the periodic potential as shown in Fig. 8.2. In this consideration, the problem is simplified by replacing the periodic potential with a square well of period $(a + b)$. At the bottom of the well, i.e., for $0 < x < a$, the electron is assumed to be very close to the nucleus, where the potential energy is zero; but outside the well, i.e., for $-b < x < 0$, the potential energy is assumed to be V_o . Although, this approximation does not match exactly with the potential offered by the ion cores in the lattice, yet it is able to describe the important features of an electron during its motion in the periodic lattice.

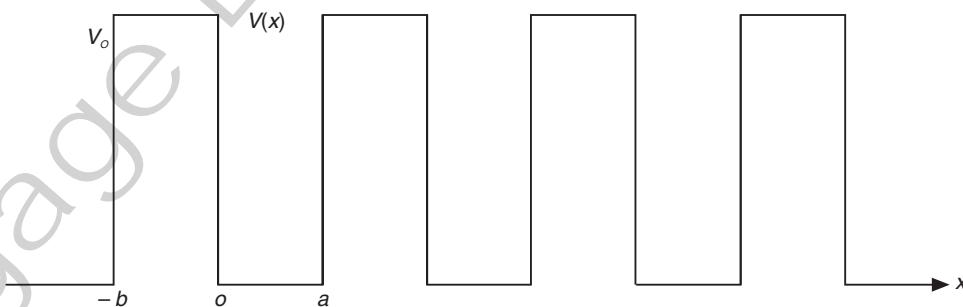


Fig. 8.2 One-dimensional periodic potential due to ion cores

The wave functions for the electron can be obtained by solving the Schrödinger wave equation for the regions $0 < x < a$ and $-b < x < 0$.

Schrödinger wave equations for these regions are given as

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{2m}{\hbar^2} E \psi = 0 \quad \text{for } 0 < x < a \quad (8.7)$$

and $\frac{\partial^2 \psi}{\partial x^2} + \frac{2m}{\hbar^2} (E - V_0) \psi = 0 \quad \text{for } -b < x < 0 \quad (8.8)$

The solutions of these equations have the form of a plane wave, modulated by the periodicity of the lattice. The actual Bloch solutions are given as

$$\psi_k(x) = u_k(x) \exp(\pm ikx) \quad (8.9)$$

where $u_k(x)$ is a periodic function with period $(a + b)$.

In order to derive the relation for the allowed values of electron energies during the motion of the electron in the lattice, Kronig–Penney assumed that

- (i) the energy of the electron (i.e., E) is less than V_o ,
- (ii) the solutions of Schrödinger wave equations are the Bloch functions, and
- (iii) the wave functions and their first derivatives are continuous throughout the crystal lattice.

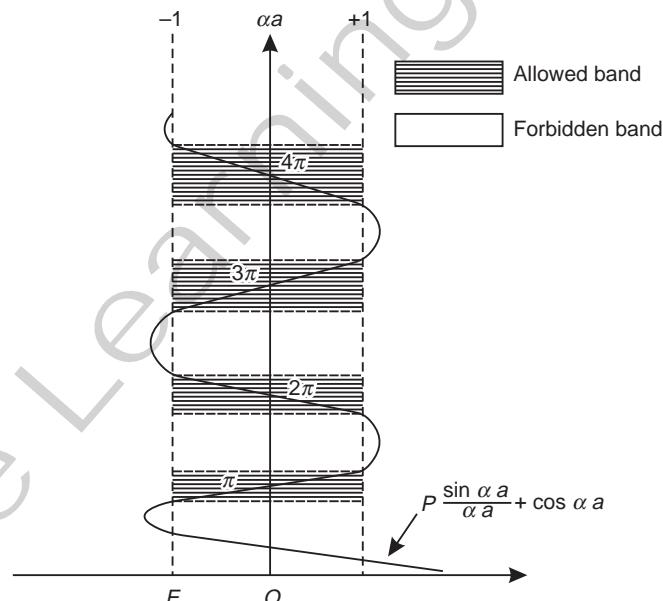


Fig. 8.3 Plot of $P(\sin \alpha a / \alpha a) + \cos \alpha a$ versus αa , showing the allowed (shaded) and forbidden energy regions

They further assumed that the product of the height and the width of the potential barrier V_{ob} , remains finite; when V_o tends to ∞ , b tends to zero. The permissible values of electron energies are governed by the relation

$$\text{where } P \left(\frac{\sin \alpha a}{\alpha a} \right) + \cos \alpha a = \cos K a \quad (8.10)$$

$$\text{and } P = \frac{ma V_o b}{\hbar^2} \quad (8.10 \text{ a})$$

$$\alpha^2 = \frac{2mE}{\hbar^2}$$

Here, m is the mass of the electron.

8.3.3 ALLOWED ENERGY BANDS

The allowed energy levels (bands) of an electron in a periodic potential of ion cores can be obtained by the proper interpretation of Eq. (8.10).

On the left-hand side of this equation, there is $\cos \alpha a$, whose values lies between +1 and -1. It means that this equation is satisfied only for those values of αa for which its right-hand side lies between +1 and -1. Such values of αa , therefore, represent the allowed solutions, while the other values of αa are not allowed. Figure 8.3 shows the oscillatory curve which is the plot of the left-hand side of Eq.(8.10) (i.e., $P(\sin \alpha a/\alpha a) + \cos \alpha a$) versus αa . The limits of +1 and -1 imposed by $\cos \alpha a$ are represented by the vertical dashed lines. The allowed and the forbidden bands are also represented between these two limits, by shaded and unshaded regions, respectively.

The following conclusions can be drawn from the plot shown in Fig. 8.3.

- In the energy spectrum of an electron, there are alternate regions of allowed and forbidden energy bands.

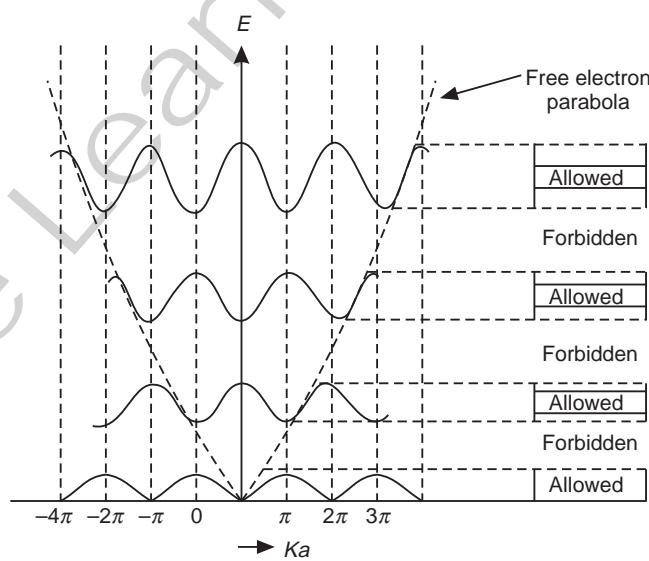


Fig. 8.4 Plot between energy and Ka showing the free electron parabola with allowed and forbidden regions

- (ii) The width of the energy bands increases with the increase in energy.
- (iii) The width of the allowed energy bands decreases with the increase in the binding energy of the electrons (i.e., value of P). When $P \rightarrow \infty$, then the bands are compressed into discrete energy levels. If we plot a curve between energy (i.e., $E = K^2 \hbar^2 / 2m$) and Ka , then a dashed line parabola is obtained as shown in Fig. 8.4.

Discontinuities in the plot (Fig. 8.4) occur for $\cos Ka = \pm 1$, i.e., when $Ka = \pm n\pi$, where $n = 0, 1, 2, \dots$. The allowed energy bands are shown at the right side of Fig. 8.4, which are separated by the forbidden energy bands.

8.4 ENERGY BANDS IN SOLIDS

In an isolated atom, there are discrete energy levels as $1s, 2s, 3s, \dots$ which are occupied by the electrons of the atom, as shown in Fig. 8.5(a). If all the atoms of a solid are assumed to be isolated from one another, then they have completely identical electronic schemes of their energy levels. When these isolated atoms are brought together to form a solid, various interactions occur between the neighbouring atoms. During this interaction, the higher energy levels are affected considerably, and consequently, the energy levels of the outer shells are slightly altered without violating Pauli's exclusion principle. Hence, the single energy levels of atoms split into large number of energy levels as shown in Fig. 8.5(b). Since, there are many atoms in a solid, so the separation between the split energy levels is very small. Hence, the split energy levels are almost continuous and are said to form an energy band.

There are energy bands in a solid corresponding to the energy levels in an atom. An electron in a solid can have only those discrete energies that lie within these energy bands. These energy bands are known as the allowed energy bands, which are generally separated by some energy gap known as the forbidden energy bands or not allowed energy bands. Energy band occupied by the valence electrons is known as the valence band, and the energy band which is empty or occupied by the conduction electrons is known as the conduction band.

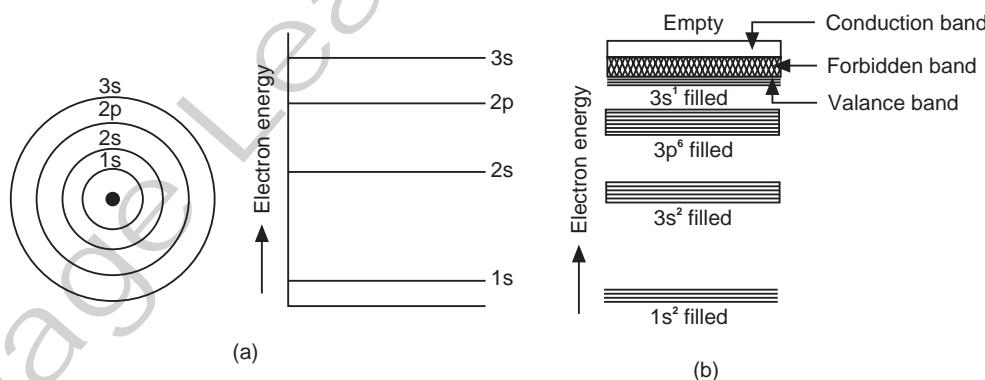


Fig. 8.5 (a) Electrons of an isolated atom in discrete energy levels and **(b)** energy bands formed due to the interaction between different sodium atoms in a solid

In order to clarify the concept of valence band and conduction band, let us take an example of sodium crystal. This crystal has many closely packed sodium atoms, which is usually referred to as crystal lattice. Each sodium atom has eleven electrons with electronic configuration $1s^2, 2s^2, 2p^6$, and $3s^1$. In

in this electronic configuration, all the subshells are filled except $3s^1$. The outermost electron $3s^1$ of this atom is called valence electron. When more number of atoms are brought together to form a solid, then the outer energy levels are split to form an energy band (i.e., large number of split energy levels closely spaced). In this band, there are two regions: one region is occupied by the valence electrons known as, valence band, whereas the other region is either empty or occupied by the conduction electrons (called conduction band) as shown in Fig. 8.5(b). Valence band is also known as the highest occupied band, whereas the conduction band is the lowest unoccupied band. The forbidden energy gap between the conduction band and the valence band is known as the energy band gap.

Solved Examples

Example 8.1

Using the Kronig–Penney model, show that for $P \ll 1$, the energy of the lowest energy band is

$$E = \frac{\hbar^2 P}{ma^2}$$

Solution

From Eq. (8.10), the energy of the lowest band corresponds to $K = \pm \pi/a$, i.e., when

$$P\left(\frac{\sin \alpha a}{\alpha a}\right) + \cos \alpha a = \pm 1$$

If we take only magnitude of the right side, then we get

$$\frac{P}{\alpha a} (\sin \alpha a) = 1 - \cos \alpha a$$

$$P\left(\frac{\sin \alpha a}{\alpha a}\right) + \cos \alpha a = \pm 1$$

or $\frac{2P}{\alpha a} \sin\left(\frac{\alpha a}{2}\right) \cos\left(\frac{\alpha a}{2}\right) = 2 \sin^2\left(\frac{\alpha a}{2}\right)$

For $P \ll 1$, we can write

$$\tan\left(\frac{\alpha a}{2}\right) = \frac{P}{\alpha a} = \tan\left(\frac{P}{\alpha a}\right)$$

or $\frac{\alpha a}{2} = \frac{P}{\alpha a}$

or $\alpha^2 = \frac{2P}{a^2}$

From Eq. (8.10 a), we know that

$$a^2 = \frac{2mE}{\hbar^2}$$

$$\text{or } \frac{2P}{a^2} = \frac{2mE}{\hbar^2}$$

$$\text{or } E = \frac{\hbar^2 P}{2ma^2}$$

8.5 CLASSIFICATION OF SOLIDS

On the basis of the forbidden energy gap, solids are broadly classified into following three categories:

- (i) conductors (metals), (ii) insulators, and (iii) semiconductors.

The electrical properties of the materials can be explained by the concept of energy bands. Depending on the nature of band occupation by electrons and the width of forbidden bands, some important characteristics of above said class of solids are discussed below.

8.5.1 Conductors (Metals)

In case of conductors, there is no forbidden gap between the valence band and the conduction band. It is observed that the valence band overlaps with the conduction band in metals as shown in Fig. 8.6. There are sufficient number of free electrons, available for electrical conduction. Since, conduction band and valence band overlap in case of metals, so the electrons from valence band freely enter the conduction band. Most important point in conductors is that due to the absence of forbidden gap between the conduction band and the valence band, there is no chance for the presence of holes. Hence, the total current in conductors is simply due to the flow of electrons only.

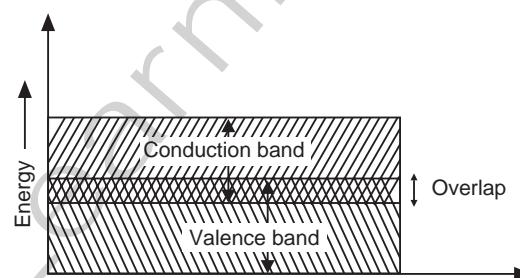


Fig. 8.6 Conduction band and valence band in conductors

For further discussion, let us consider the example of sodium, which has electronic configuration as $1s^2, 2s^2, 2p^6, 3s^1$. From this configuration, it is clear that $1s, 2s, 2p$ bands in sodium are completely filled, while the outermost ($3s$) band has only one electron and can accommodate one more electron. Thus, the outer valence band is half-filled. Under the influence of external electric field, electrons of valence band acquire sufficient kinetic energy to move towards the conduction band, where they can move freely to participate in the process of conduction.

Thus, we can conclude that in conductors, there is overlapping between valence and conduction bands. Hence, valence band itself behaves as the conduction band. Due to this reason, metals in general have high electrical conductivity.

8.5.2 Insulators

A class of solids behaves as insulator if it satisfies the following two conditions:

- it has even number of valence electrons per atom and
- the valence band and the conduction band are separated by a large energy gap compared with kT .

The band diagram of insulators is given in Fig. 8.7. Actually, due to the large band gap, electrons cannot jump from the valence band to the conduction band. In insulators, the valence electrons are bound tightly to their parent atoms. For example, in case of glass, the valence band is completely filled at 0 K and energy band gap between the valence band and the conduction band is of the order of 10 eV. An external electric field cannot provide this amount of energy to an electron in solid. Therefore, such solids behave as insulators because no free electrons are available to conduct electricity. Hence, we can conclude that the insulators are those solids which have full valence band and an empty conduction band with a large energy band gap between the valence band and the conduction band.

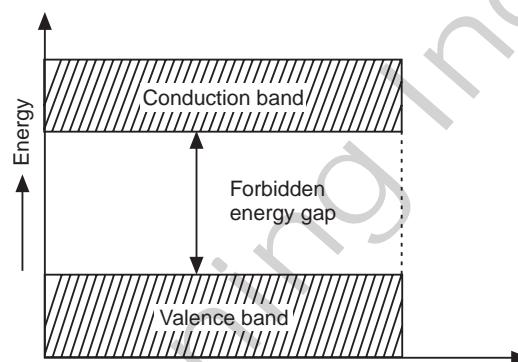


Fig. 8.7 Conduction band and valence band in insulators

8.5.3 Semiconductors

In case of semiconductors, the energy band gap (forbidden gap) between the filled valence band and the empty conduction band is small as compared to the insulators and more as compared to the conductors. Usually, this energy band gap is of the order of 1 eV. Energy band structure of semiconductors is shown in Fig. 8.8. Germanium and silicon are the common examples of semiconductors. In germanium, the forbidden energy gap is 0.7 eV, whereas in case of silicon, it is 1.1 eV. Due to the small energy band gap between the valence band and the conduction band, the electrons of valence band can be thermally excited to the conduction band, where they are able to conduct small current on application of electric field to the crystal. At 0 K temperature, a semiconductor behaves as an insulator, but at room temperature, a semiconductor has the following peculiar properties:

- A partially filled conduction band
- A partially filled valence band
- A very narrow energy band gap (≈ 1 eV)

8.6 INTRINSIC SEMICONDUCTORS

Natural pure form of a semiconductor is known as *intrinsic semiconductor*. In intrinsic semiconductors, conductivity is mainly due to the thermally generated charge carriers, because in such materials, the

electron-hole pairs are generated due to the thermal excitation. The electrons reaching the conduction band, due to thermal excitation leave equal number of vacancies or holes in the valence band (in intrinsic semiconductor). The number of free electrons is always equal to the number of holes.

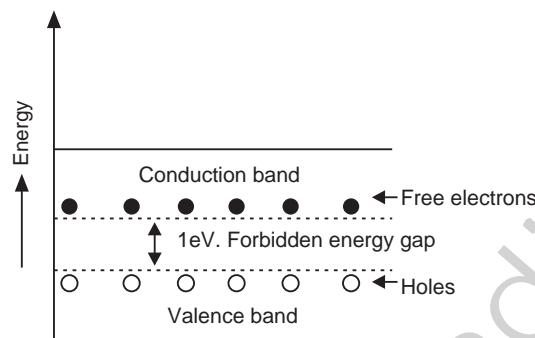


Fig. 8.8 Conduction band and valence band in semiconductors

8.6.1 Current Conduction in Intrinsic Semiconductors

In order to describe the current conduction mechanism in intrinsic semiconductors, let us take the example of germanium. The electronic configuration of germanium is $1s^2, 2s^2, 2p^6, 3s^2, 3p^6, 3d^{10}, 4s^2, 4p^2$, which shows that the outermost orbit contains four electrons. Thus, a germanium atom has four valence electrons, i.e., it is a tetravalent element. Each of the four valence electrons in a germanium atom is shared with the valence electrons of four adjacent germanium atoms making four electron pairs as shown in Fig. 8.9(a). Hence, all valence electrons of each germanium atom are used to form covalent bond with their neighbouring atoms and no electron is free to participate in the process of conduction, particularly at 0 K. In this way, pure germanium behaves as an insulator at 0 K. As the temperature increases, valence electrons acquire sufficient energy to break their covalent bond and become free to participate in the conduction process. Such electrons are known as *free electrons* or *conduction electrons*.

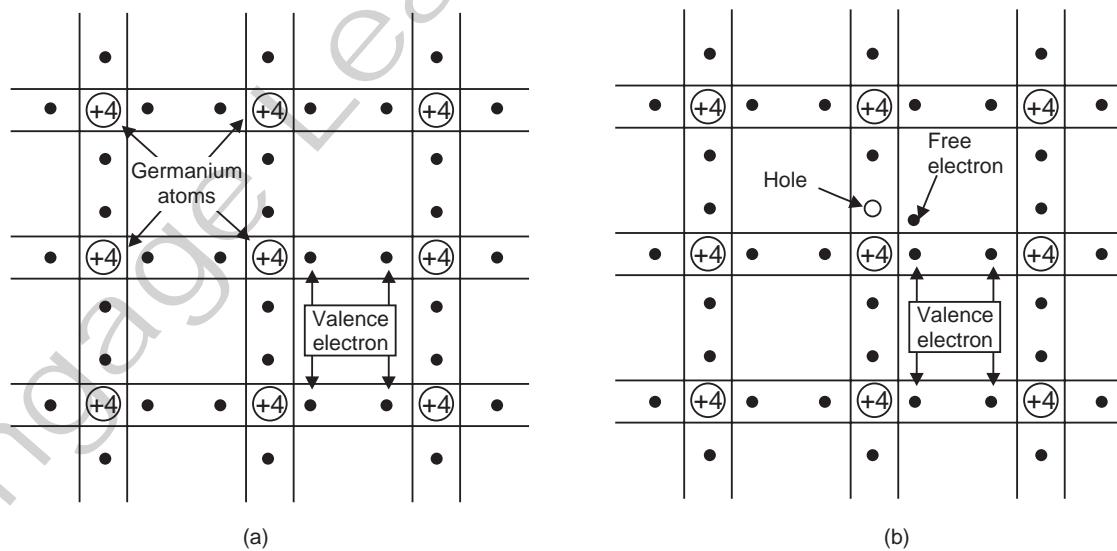


Fig. 8.9 (a) Two-dimensional representation of germanium crystal and (b) two-dimensional representation of germanium with an electron-hole pair

9

Electrical Conductivity and Electron Theory of Metals

9.1 INTRODUCTION

The physical properties of a metal are governed by its elementary charged particle known as *electron*. The mobility of electrons in the metals decides the electrical, thermal, and other properties of metals. In order to explain most of the properties of metals, the theory of free electron model was proposed by Drude, in 1900, soon after the discovery of electron. In this theory, the valence electrons of metal are supposed to be completely free from the atoms. The basis for this theory is the picture of metals as a lattice (regular arrangement which remains fixed) of positive ions through which the electron gas can move. The number of electrons in such a gas is equal to the number of valence electrons. If the metal is not subjected to the external field, the valence electrons oscillate equally in all directions among the atoms just like the molecules in a gas. The statistics of this random electron velocity will be analogous to the statistics of the molecular motion in a gas.

Drude theory of free electron model was initially used to explain the electrical and the thermal conductivities of metals. Later on, Lorentz investigated the implications of this hypothesis exhaustively and he assumed free electrons as a perfect gas (different from ideal gas molecules because they are charged particles), which in equilibrium obey the Maxwell–Boltzmann statistics. This theory is known as *Drude–Lorentz theory*. There are several impressive successes and several remarkable failures in the credit of this theory. In order to meetout the failures of Drude–Lorentz theory, the quantum theory of free electrons was introduced.

In this chapter, we will focus our discussion on the electrical conductivity in metals using free electron theory along with its failure and successes. In the same chapter, we will also introduce the quantum theory of free electrons and its applications to discuss the density of states, electrical conductivity of metals, and its temperature dependence.

9.2 FREE ELECTRON CONCEPT

The motion of electrons in a solid is responsible for its physical properties. Due to this reason, the clear concept regarding the state of electrons in a solid is very important. In order to explain the properties of solid in the context of electrons, the free electron model was suggested by Drude.

The concept of free electron can be explained by taking the example of sodium crystal, in which each atom contains 11 electrons. According to the electronic configuration of the sodium atom, 10 electrons

are tightly bound to the nucleus to form an ion of net positive charge $+ |e|$ and the eleventh electron (also known as the valence electron) moves in the orbit around the ion. When large number of such atoms are brought together to form a solid, the outer orbits of different atoms overlap and interact with each other. Due to the extensive overlapping of outer orbits, the valence electron of a particular atom is no more associated with its parent atom. Under these circumstances, the electrons of outermost orbit of each atom (which are constituting solid) are free to move in between the ion cores of the crystal. These free electrons may behave as the molecules of gas up to some extent (not exactly) and hence, these are known as free electron gas. Thus, the metal crystal can be pictured as composed of positive metal ions with free electron gas of valence electrons of the atoms. In this consideration, a natural question arises that how the crystal holds these free electrons? The answer of this questions can be given in terms of electrostatic attraction between positively charged ions and negatively charged free electrons. Actually, this electrostatic force holds the free electron within the crystal lattice. Since the negative charge of free electrons is equal to the positive charge of ions in the crystal, so the crystal remains electrically neutral.

9.3 CLASSICAL FREE ELECTRON THEORY

During 1900, Drude proposed free electron theory for the metals using some basic assumptions. According to this theory, the valence electrons of atoms, constituting a crystal are free to move inside the metal just like the molecules or atoms in a perfect gas. Thus, free electrons are referred to as free electronic gas and the theory is accordingly named as free electron gas model. Later in 1909, H. A. Lorentz suggested that electrons constituting the electron gas obey Maxwell–Boltzmann statistics under the equilibrium condition. Combining the idea of Lorentz with the assumption of Drude, the constituted theory is known as Drude–Lorentz theory. Since the theory is based on the classical ideas of Maxwell–Boltzmann statistics, so it is called the classical free electron theory.

According to this theory, valence electrons (conduction electrons) in a metal find themselves in the field of all ions and other valence electrons in the crystal. The mutual repulsion between the electrons is ignored and it is assumed that a uniform electrostatic attraction exists between electrons and ion core throughout inside the crystal. Although conduction electrons are free to move inside the crystal, these are not free to move out of the metal at room temperature; rather they are confined to the volume of the metal. Thus, we can conclude that the potential energy of a stationary electron inside a metal is less than the potential energy of an identical electron just outside it. This energy difference V_0 acts as a potential barrier to stop the valence electrons from leaving the surface of the metal. The mobility of free electrons is equivalent to the motion of a particle in a potential box of potential height V_0 as shown in Fig. 9.1.

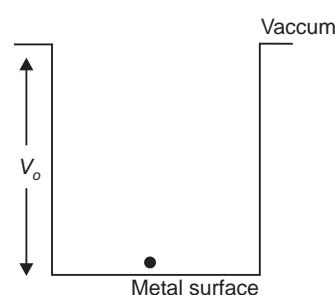


Fig. 9.1 Potential well showing the difference in potential energy of an electron at the metal surface and just outside it

This theory is able to prove the validity of Ohm's law, which relates the electric current with the electric field. It is also successful to derive the relation between electrical and thermal conductivities (the Wiedemann–Franz law). But this theory completely failed to explain the heat capacity and the paramagnetic susceptibility of conduction electrons. On the basis of this theory, we cannot predict the temperature dependence of resistivity and occurrence of long mean free paths at low temperature. In order to resolve the failure of classical free electron theory, Sommerfeld's quantum theory is introduced. In this theory, Fermi–Dirac statistics is used for the electron gas instead of classical Maxwell–Boltzmann statistics.

9.4 ASSUMPTIONS OF CLASSICAL FREE ELECTRON THEORY

Drude and Lorentz proposed the classical free electron theory for metals using some important assumptions, which are as follows:

- (i) Metal structure consists of positive ion cores, with the valence electrons moving freely among these ion cores.
- (ii) The behaviour of free electrons (valence/conduction electrons) moving inside a metal is similar (not exactly same) to that of atoms or molecules in a perfect gas.
- (iii) Mutual repulsion among the free electrons is ignored.
- (iv) Free electrons are negatively charged particles while the atoms or molecules of a perfect gas are mostly neutral.
- (v) Free electrons obey the Pauli's exclusion principle.
- (vi) The free electrons experience constant and uniform electrostatic field due to the ions cores.
- (vii) Total energy of free electrons inside a metal is almost equal to their kinetic energy.
- (viii) The total negative charge of electrons is equal to the positive charge of ion cores and hence, the metal remains electrically neutral.
- (ix) The electrostatic attraction between free electrons and ion cores is responsible for holding the electrons within the surface of metals.
- (x) The valence electrons constituting the electron gas obey Maxwell–Boltzmann statistics under the equilibrium condition.
- (xi) The movement of free electrons inside a metal is equivalent to the motion of free electrons in a potential well.

9.5 DRIFT VELOCITY

In classical free electron theory of metals, free electron gas model is used to explain the electrical conductivity of metals. At the condition of thermal equilibrium, electrons in a solid are in the state of random motion. The kinetic energy of an electron during its random motion is given as

$$\frac{1}{2}mc\bar{c}^2 = \frac{3}{2}k_B T \quad (9.1)$$

$$\text{or } \bar{c} = \sqrt{\frac{3k_B T}{m}} \quad (9.2)$$

where k_B is Boltzmann's constant and \bar{c} is the root mean square velocity at temperature T .

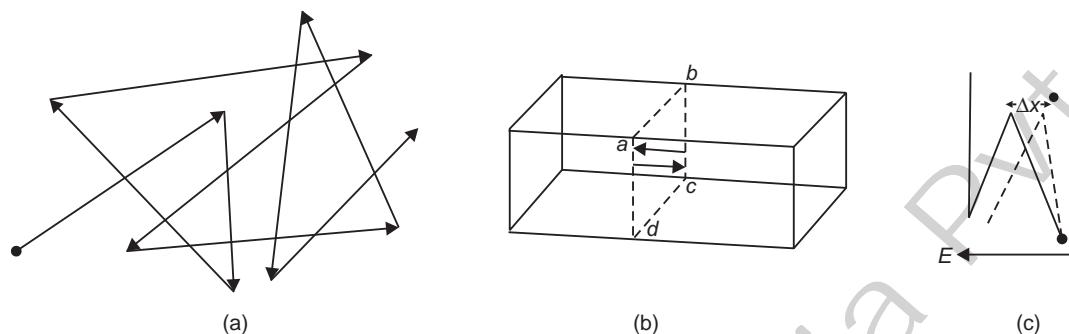


Fig. 9.2 (a) Motion of electron in absence of external field, (b) view of plane through which same number of electrons are passing left to right and vice versa, and (c) motion of an electron under the influence of external fields

The behaviour of an electron's motion with and without external fields is shown in Figs. 9.2 (c) and (a), respectively. Actually, in the absence of external field, electrons undergo random motion and they constitute no net current. In order to illustrate it further, let us consider a plane $abcd$ as shown in Fig. 9.2 (b). It is observed that many electrons are passing through this plane from left to right and almost same number of electrons are passing through this plane from right to left. Thus, the net current constituted by the electrons in the solid is zero.

Under the influence of external potential difference applied across the solid, the equilibrium condition is disturbed. Due to the applied electric field, electrons are accelerated, but very soon their motion is hindered due to their interaction with the atoms of the crystal lattice. Thus, the magnitude and the direction of the velocity of electrons are changed suddenly. At the steady state condition, the electrons have net movement in the direction opposite to the applied electric field [Fig. 9.2 (c)]. This net movement of electrons is known as *drift*, and the corresponding mean velocity is called *drift velocity* (v_d). Electrons drift in a particular direction and constitute a current known as *drift current*, which is more often called *conduction current*.

9.6 EXPRESSION OF DRIFT VELOCITY

The strength of electric current in a conductor is measured in terms of the magnitude of electric charge flowing per second through a cross section of the conductor. Let us consider an electric field which is applied at the ends of a metal under the influence of which, free electrons start moving with the drift velocity v_d in the direction opposite to the applied field as shown in Fig. 9.3.

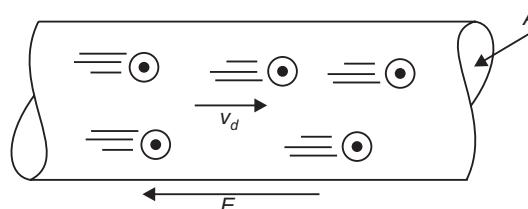


Fig. 9.3 Drift velocity of electrons under the influence of external field

If q charge is passing through this metal in time t , then the constituted current i can be given as

$$i = \frac{q}{t} \quad (9.3)$$

Let the cross sectional area of metal be A and n be the free electrons per unit volume, then the number of electron passing per second through the cross section of metal is $nA v_d$. Now, the charge of electrons passing through the metal in time t second is given as

$$q = (nA v_d t) \times e$$

By putting the value of q in Eq. (9.3), we get

$$i = neAv_d$$

We know that the current density is given as

$$j = \frac{i}{A} = nev_d$$

$$\text{or } v_d = \frac{i}{ne} = \frac{i}{neA} \quad (9.4)$$

Equation (9.4) gives the relation between the current and the drift velocity.

9.7 RELAXATION TIME

Relaxation time is a measure of time required to relax an electron when the externally applied field is removed.

If an electric field E_x is applied to a specimen, then all the electrons are accelerated in the x -direction with the acceleration a_x which is given as

$$a_x = -\left(\frac{eE_x}{m}\right)$$

$$\text{or } \frac{dv_x}{dt} = -\left(\frac{eE_x}{m}\right)$$

If $\langle v_x \rangle$ is the average velocity of all n electrons, then

$$\frac{d}{dt} \langle v_x \rangle = -\left(\frac{eE_x}{m}\right) \quad (9.5)$$

Due to the electron lattice and other constraints of crystal lattice scattering, it is very difficult to solve the above equation to obtain the value of v_x . In order to solve this equation, let us consider that the average velocity of the electron at an instant $t = 0$ is $\langle v_x \rangle_0$ (tends to zero). In other words, we can say that the velocity of electrons becomes random in the crystal.

Let us suppose that the process of randomisation follow the simplest law of decay, according to which

$$\langle v_x \rangle = \langle v_x \rangle_0 e^{-t/\tau} \quad (9.6)$$

τ is known as relaxation time. In order to define the relaxation time, let us put $t = \tau$. Now, Eq. (9.6) becomes

$$\begin{aligned}\langle v_x \rangle &= \langle v_x \rangle_0 e^{-1} \\ &= \frac{\langle v_x \rangle_0}{e}\end{aligned}\quad (9.7)$$

Hence, the relaxation time is defined as the time taken by the drift velocity to decay to $1/e$ of its initial value.

Differentiating Eq. (9.6), we get

$$\begin{aligned}\frac{d}{dt} \langle v_x \rangle &= -\frac{\langle v_x \rangle_0 e^{-t/\tau}}{\tau} \\ \text{or } \frac{d \langle v_x \rangle}{dt} &= -\frac{\langle v_x \rangle}{\tau} \quad (\text{because } \langle v_x \rangle = \langle v_x \rangle_0 e^{-t/\tau})\end{aligned}\quad (9.8)$$

Equation (9.8) is used to obtain the electron lattice term without knowing the details of the collision process. Relaxation of electrons to the equilibrium is shown in Fig. 9.4.

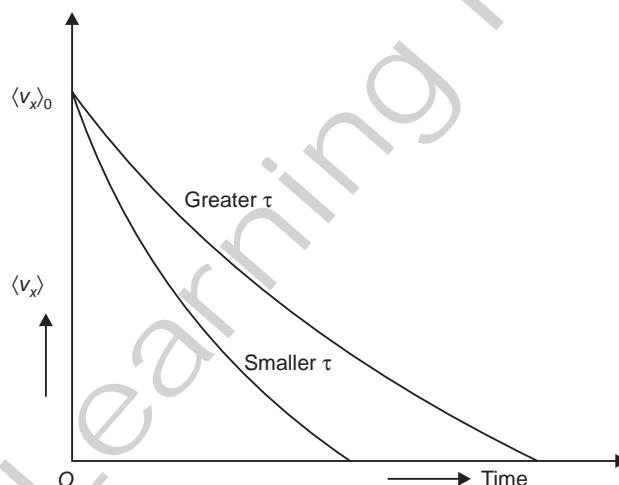


Fig. 9.4 Relaxation of electrons to equilibrium after the removal of electric field

9.8 MEAN COLLISION TIME

The average time taken by an electron between any two consecutive collisions during its motion in the crystal lattice is known as the *mean collision time*.

During the motion of electrons in the crystal lattice, it is assumed that the electron collisions are elastic. It means that their energy is conserved before and after the collision. But it is observed that the conductor warms up when current passes through it. This implies that a certain part of energy is transferred to the lattice during electron-lattice collision. Since the fraction of energy transferred to the crystal is very small, so it can be neglected and collision may be considered the elastic one.

Let us consider $\langle v_x \rangle$ as the average velocity of electron along the X-axis before the collision which become completely random after the collision, i.e.,

$$\langle v_x \rangle_{\text{after}} = 0 \text{ and } \langle v_x \rangle_{\text{before}} = \langle v_x \rangle$$

Now, the net change in average velocity before and after the collision is given as

$$\langle v_x \rangle_{\text{after}} - \langle v_x \rangle_{\text{before}}$$

or $0 - \langle v_x \rangle = -\langle v_x \rangle$

Suppose that the probability of an electron making a collision in time dt is dt/τ_c , i.e., on an average, there are $1/\tau_c$ collisions per second. Now, the rate of change of average velocity $\langle v_x \rangle$ may be given as

$$\frac{d}{dt} \langle v_x \rangle = -\frac{\langle v_x \rangle}{\tau_c} \quad (9.9)$$

Comparing Eqs. (9.8) and (9.9), we get $\tau = \tau_c$, i.e., relaxation time is equal to the mean time between collisions.

In the above derivation, it is assumed that the average scattered angle is 90° (the average of angles 0° to -180°) and after each collision, electrons have no memory of what went before. If on an average, velocity direction changes by θ on collision (Fig. 9.5), then the change of $\langle v_x \rangle$ is on an average, $-\langle v_x \rangle (1 - \cos \theta)$.

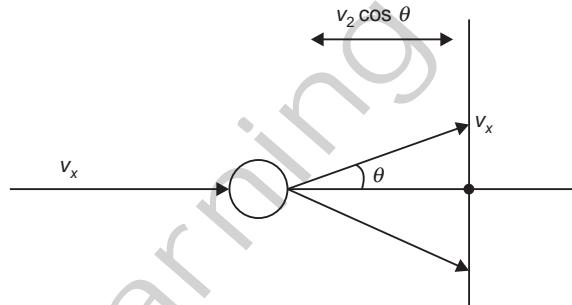


Fig. 9.5 Scattering of electron at an average angle θ . The change in any v_x is therefore on an average, $v_x (1 - \cos \theta)$

Hence,

$$\tau = \frac{\tau_c}{[1 - \langle \cos \theta \rangle]} \quad (9.10)$$

where τ is the relaxation time and τ_c is the mean time between collisions.

9.9 MEAN FREE PATH

The mean free path λ of an electron is an average distance which the electron travels without suffering any collision during its motion in the crystal lattice. The mean free path λ of an electron is given as

$$\lambda = \bar{c} \tau_c \quad (9.11)$$

where \bar{c} is the root mean square velocity of the electron and τ_c is the collision time.

Mean free path depends on the obstacles present in its path.

For a given concentration of obstacles, the collision time (τ_c) is inversely proportional to the velocity \bar{c} of the electrons. From Eq. (9.10), it is clear that the relaxation time (τ) is proportional to the collision

time (τ_c) and hence, inversely proportional to the velocity \bar{v} . Since the relaxation time (τ) determines the conductivity of metals, so it should have a single value decided by fermi energy E_F . The proof of this statement can be given on the basis of the quantum theory of free electrons.

9.10 EXPRESSION FOR ELECTRICAL CONDUCTIVITY OF METALS

Let us consider a metal specimen of cross sectional area A and length l . If the potential difference V is applied across the solid as shown in Fig. 9.6, then it establishes an electric field E , which is given as

$$E = \frac{V}{l} \quad (9.12)$$

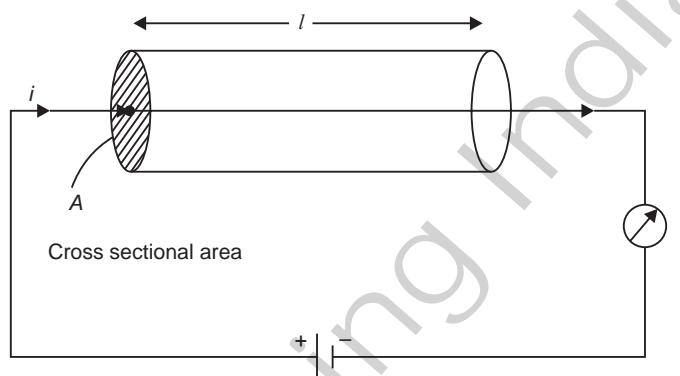


Fig. 9.6 Schematic circuit diagram for determining the conductivity of a solid

This electric field accelerates the mobile charges and causes a flow of electric current through the specimen. If Q charges are passing through the cross sectional area A in time t , then current i can be given as

$$i = \frac{\theta}{t} \quad (9.13)$$

The ratio of the intensity of the electric field E at any point within the conductor and the current density j at that point is called resistivity or specific resistance of the conductor and is represented by ρ .

$$\text{Thus, } \rho = \frac{E}{j} = \frac{EA}{i} \quad \left(\text{because } j = \frac{i}{A} \right) \quad (9.14)$$

Putting the value of E in Eq. (9.14), we get

$$\begin{aligned} \rho &= \frac{V}{i} \cdot \frac{A}{l} \\ &= R \left(\frac{A}{l} \right) \end{aligned} \quad (9.15)$$

where $R = V/i$ is known as the resistance of specimen.

The reciprocal of resistance is known as conductance (g) and the reciprocal of resistivity is known as conductivity (σ).

These are expressed as

$$g = \frac{1}{R} = \frac{A}{\rho l} \quad (9.16)$$

and $\sigma = \frac{1}{\rho} = \frac{l}{RA}$ (9.17)

The unit of conductance is mho, also called siemens (S). The unit of conductivity is $(\Omega \cdot \text{m})^{-1}$ or mho/m or S/m.

By putting the value of ρ in Eq. (9.17) from Eq. (9.16), we get

$$\begin{aligned}\sigma &= \frac{1}{E/j} \\ &= \frac{j}{E}\end{aligned}$$

or $j = \sigma E$ (9.18)

i.e., current density = conductivity \times electric field.

If n is the concentration of free electrons available in the unit volume of the specimen, then total number of electrons in the solid is given as

$$N = nAl$$

Now, the total charge present in the solid is expressed as

$$Q = Ne = neAl$$

Now, the current through the specimen is given as

$$\begin{aligned}i &= \frac{Q}{t} = \frac{neAl}{t} \\ &= neAv_d\end{aligned}$$

Now, the current density (j) given as

$$j = \frac{i}{A} = nev_d \quad (9.19)$$

From Eq. (9.18),

$$\sigma = \frac{j}{E} = \frac{nev_d}{E} \quad (9.20)$$

$$\sigma = ne\mu \quad (9.20)$$

$$\mu = \frac{v_d}{E} \quad (9.21)$$

μ is known as the mobility of electrons and it is defined as the drift velocity per unit electric field. The unit of mobility is $\text{m}^2/\text{V}\cdot\text{s}$.

9.11 EFFECT OF TEMPERATURE ON RESISTIVITY OF METALS

When the temperature of a metallic wire is raised, its electric resistance increases. This can be explained as below.

The resistance R of a wire of length l , cross sectional area A , and the relaxation time (τ) of free electrons within the wire is given as

$$R = \frac{m}{ne^2\tau} \frac{l}{A} \quad (9.22)$$

where n is the number of free electrons per unit volume of wire, m is the mass, and e is the charge of an electron for the given specimen. From Eq. (9.22),

$$R \propto \frac{l}{\tau}$$

The relaxation time (τ) is the average time between two successive collisions of an electron with the positive ions. The relaxation time (τ) is expressed in terms of λ and \bar{c} (root mean square velocity) as

$$\tau = \frac{\lambda}{\bar{c}} = \frac{\lambda}{v_{rms}} \quad (\text{Because } \bar{c} = v_{rms})$$

$$\therefore R \propto \frac{l}{\tau} \propto \frac{\bar{c}}{\lambda} \quad (9.23)$$

As the temperature of the specimen rises, the root mean square velocity (\bar{c}) of its free electrons increases ($\bar{c} \propto \sqrt{T}$) and the mean free path (λ) decreases. Hence, the resistance R of the wire increases, i.e., the resistivity of the material of the specimen $\rho = R \frac{A}{l}$ increases. In this way, we conclude that with the increase in the temperature, the electrical conductivity of metallic specimen decreases.

If the resistance of the wire at 0°C is R_o and at $t^\circ\text{C}$ it is R_t , then the value of R_t can be obtained by the following formula:

$$R_t = R_o (1 + \alpha t) \quad (9.24)$$

where α is a constant which is called the “temperature coefficient of resistance” of the material of the given specimen. From Eq. (9.24), we can write

$$\alpha = \frac{R_t - R_o}{R_o \times t} \text{ per } {}^\circ\text{C} \quad (9.25)$$

If $R_o = 1 \Omega$ and $t = 1^\circ\text{C}$, then $\alpha = (R_t - R_o) / (R_o \times t)$. It implies that if the resistance of the wire at 0°C is 1Ω , then on raising its temperature to 1°C , the increase in its resistance will be equal to the “temperature coefficient of resistance” of the material of the given specimen.

9.12 EFFECT OF IMPURITY ON THE RESISTIVITY OF METAL

The addition of small amounts of impurities leads to a considerable increase in the resistivity. For example, brass (63% Cu and 40% Zn), which is widely used as a contact material in power plugs and sockets,

10

Dielectrics

10.1 INTRODUCTION

Dielectrics are basically non-conductors of electricity. These materials usually do not contain any free charge carriers for conduction. They have positive and negative charges tightly bound together. Under the influence of external electric field, the charges of dielectrics are affected (i.e., positive and negative charges are displaced from their equilibrium) and they form electric dipoles, which are responsible for the dielectric behaviour of these materials. Dielectrics may be solids, liquids, or gases. In addition to these, sometimes high vacuum can also be used as loss-less dielectric which has its relative dielectric constant unity.

Although the dielectrics are very much similar to the insulators, there is a slight difference between dielectrics and insulators. The main function of dielectric materials is to store electric energy, while the function of the insulators is to obstruct the flow of current.

When dielectric materials are placed in a capacitor, their capacitance increases. This is due to the polarisation of dielectric materials, which causes electric dipoles. The polarisation charges induced on the two faces of dielectric substrate produce their own electric field, which opposes the externally applied electric field; hence, the resultant electric field within the dielectrics is reduced. Reduction in the electric field between the plates (within dielectrics) increases the capacitance of the capacitor. Every dielectric material has a characteristic *dielectric strength*, which is defined as the maximum value of electric field that it can tolerate without its dielectric breaking down.

In this chapter, we shall discuss electric and dielectric properties of non-metals, various sources of molecular polarisation, frequency dependence of dielectric constant, dielectric loss, and its important applications. We shall also consider two important properties of certain insulators: piezoelectricity and ferroelectricity.

10.2 DIELECTRIC CONSTANT

Michael Faraday discovered that the capacitance of a capacitor increases if the space between the conductor's plates is filled with a dielectric material. If C_0 is the capacitance of condenser with air filled between its plates and C is the capacitance when the space between the two plates is completely filled by the dielectric material, then the dielectric constant of filled dielectric material is defined as

$$K = \epsilon_r = \frac{C}{C_0} \quad (10.1)$$

where ϵ_r is the known as the *relative permittivity* or *dielectric constant* of the medium. Sometimes, it is also denoted by K . It has been found that dielectric constant ϵ_r is independent of the shape and size of the capacitor; it is solely a characteristic of the particular dielectric medium used.

Let us consider a parallel plate capacitor with metal plates of area A and separation d . If there is air medium between the plates of the capacitor, then its capacity C_0 can be expressed as

$$C_0 = \frac{\epsilon_0 A}{d} \quad (10.2)$$

where ϵ_0 is the permittivity of free space between the plates of capacitor.

If the space between the plates is filled with an insulating substance (dielectric material) of permittivity ϵ , then the capacitance of same capacitor will be given as

$$C = \frac{\epsilon A}{d} \quad (10.3)$$

Here, $\epsilon = K \epsilon_0$, where K is a constant having its value more than unity and is also known as relative permittivity (ϵ_r).

Now, Eq. (10.3) can be written as

$$\begin{aligned} C &= \frac{K \epsilon_0 A}{d} \\ \text{or } C &= \frac{\epsilon_r \epsilon_0 A}{d} \end{aligned} \quad (10.4)$$

From Eqs. (10.2) and (10.4), we can write

$$K = \epsilon_r = \frac{C}{C_0} = \frac{\epsilon}{\epsilon_0} \quad (10.5)$$

Using the following relations in Eq. (10.5),

$$\begin{aligned} q &= CV \\ \text{or } C &= \frac{q}{V} = \frac{q}{Ed} \end{aligned}$$

we get

$$K = \epsilon_r = \frac{C}{C_0} = \frac{E_0}{E} = \frac{V_0}{V} \quad (10.6)$$

where E_0 and V_0 are the electric field and potential difference, respectively, between the plates of the capacitor without any dielectric medium, while E and V are the electric field and potential difference, respectively, with dielectric medium. Thus, the ratio of the capacitance of a capacitor with dielectric to the capacitance of the same capacitor without dielectric is defined as *dielectric constant*.

10.3 DIELECTRIC POLARISATION

Dielectric polarisation of non-conductors of electricity can be discussed in the context of polar and non-polar molecules. Based on the presence of these molecules in the dielectric materials, they are classified into two categories discussed in the following paragraphs.

10.3.1 Non-polar Dielectric Materials in an Electric Field

When a dielectric material is placed in an electric field, its electric properties are modified. The positive and negative charges of the molecules experience electrostatic forces in opposite directions, i.e., the negative charges are attracted by positively charged plate and vice versa. Therefore, the centres of gravity of the two charges are separated from each other. The molecules thus acquire an induced electric dipole moment in the direction of the field. This process is known as the *polarisation of dielectrics*. The dielectric materials that are polarised only when they are placed in an electric field are known as *non-polar dielectrics*. The process of polarisation in non-polar dielectric materials is shown in Fig. 10.1.

It is clear from Fig. 10.1(b) that induced charges appear in such a way that the electric field set up by them (E') opposes the external electric field E_0 . The resultant field E in the dielectric is the vector sum of E' and E_0 , i.e.,

$$E = E_0 - E' \quad (10.7)$$

The weakening of electric field within the dielectric is illustrated in Fig. 10.2(a) and (b). Fig. 10.2(a) shows the original field. In Fig. 10.2(b), not all the lines of force leaving the positive plate of the capacitor penetrate the dielectric; some of them terminate on the charges induced on the dielectric.

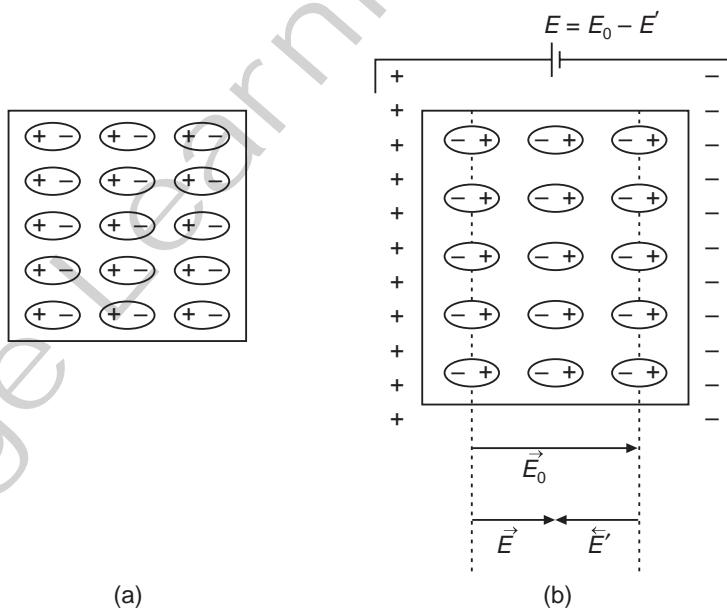


Fig. 10.1 (a) Random distribution of plus and minus charges in non-polar dielectric and **(b)** dielectric placed in electric field and induced charges

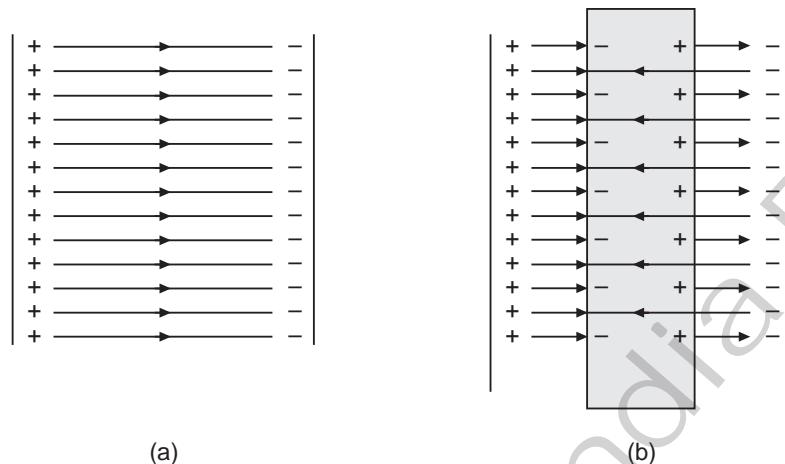


Fig. 10.2 (a) Original field and (b) modified field due to dielectric materials

The reduction in the magnitude of the electric field from E_0 to E causes a reduction in the potential difference between the plates of the capacitor.

10.3.2 Polar Dielectric Materials in Electric Field

The dielectric materials which have permanent dipole moments with their random rotation in the absence of external electric field are known as *polar dielectrics*. Dipoles of polar dielectrics are shown in Fig. 10.3. When external electric field is applied to the polar dielectrics, the partial alignment of dipoles takes place as shown in Fig. 10.3(b). In the absence of external electric field, the dipoles have their random orientation but due to thermal agitation, etc., some dipole moments may be contributed by the existing dipoles. The dipole moment of polar dielectric in an electric field is given as

$$p = p_p - p_i \quad (10.8)$$

where p_p is the permanent dipole moment and p_i is the induced dipole moment (due to orientation of dipoles).

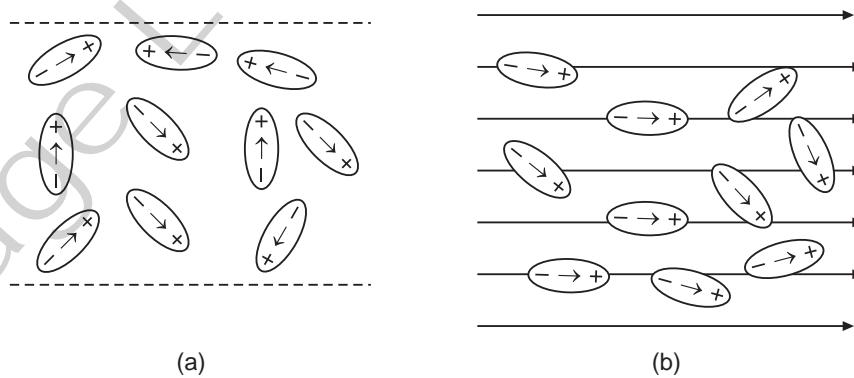


Fig. 10.3 (a) Random orientation of dipoles and (b) alignment of dipoles in presence of external field

On the basis of the above discussion for polar and non-polar dielectrics, we can conclude that the dipoles are induced in non-polar dielectrics in the presence of electric field, while the dipole moment increases in polar dielectrics due to re-orientation of existing dipoles.

The non-polar molecules have symmetrical structure and zero dipole moment. Some examples of non-polar molecules are H_2 , N_2 , O_2 , CO_2 , and benzene. In the case of polar molecules, the centres of gravity of positive and negative charges do not coincide with each other; instead they are slightly displaced. Such molecules have unsymmetrical structures and have a permanent dipole moment. Some examples of polar molecules are H_2O , HCl , CO , N_2O , and NH_3 .

10.4 ELECTRIC POLARISATION VECTOR (\vec{P})

The polarisation vector \vec{P} is defined as the induced surface charge per unit area, i.e., surface density of bound charges in dielectric medium.

$$\vec{P} = \frac{Q'}{A} \quad (10.9)$$

The induced charge Q' appears when the dielectric is polarised and this is the reason we call Q'/A as electric polarisation \vec{P} .

If we multiply the numerator and denominator of Eq. (10.9) by d , then we get

$$\vec{P} = \frac{Q'd}{Ad}$$

The product $Q'd$ is the induced electric dipole moment due to the polarisation of dielectric slab. Q' charge is induced on one surface and $-Q'$ on the opposite surface, separated by the thickness of the slab. Since Ad is the volume of slab, so the above equation can be given as

$$\vec{P} = \frac{\text{Induced dipole moment}}{\text{Volume}} = \frac{Q'd}{V} \quad (10.10)$$

The polarisation vector (\vec{P}) is defined as the induced dipole moment per unit volume. Since electric dipole is a vector, polarisation (\vec{P}) is also a vector.

It has been experimentally observed that the electric dipole moment (\vec{p}_m) induced in each molecule is proportional to the induced electric field \vec{E}_m acting on the molecule (which is produced due to the displacement of charges). Now, we can express it mathematically as

$$\vec{p}_m \propto E_m$$

or $\vec{p}_m = \alpha E_m$

where α , a constant, is called *molecular polarisability*. If N is the number of molecules per unit volume of dielectric material, then the total induced dipole moment per unit volume or the electric polarisation (\vec{P}) is given as

$$\vec{P} = N \vec{p}_m = N\alpha \vec{E}_m \quad (10.11)$$

10.5 ELECTRIC DISPLACEMENT VECTOR (\vec{D})

If a charge q_1 is sending q_1 lines of forces received by an area $4\pi r^2$, then the number of lines forces received by the unit area is called *flux density* or *electric displacement vector* (\vec{D}). The lines of forces from charge q_1 are shown in Fig. 10.4.

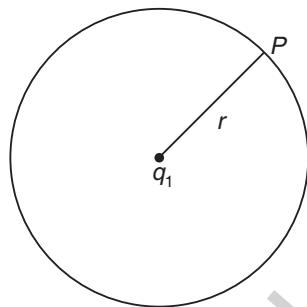


Fig. 10.4 Lines of forces from charge q_1

From the above definition, the displacement vector can be expressed as

$$\vec{D} = \frac{q_1}{4\pi r^2} = \frac{q_1}{A} \quad (10.12)$$

In order to express the electric field produced by charge q_1 , let us consider another charge q_2 at a distance r from q_1 . If the force experienced by charge q_2 is \vec{F} , then the electric field is expressed as

$$\vec{E} = \frac{\vec{F}}{q_2}$$

On putting the value of \vec{F} , we get

$$\begin{aligned}\vec{E} &= \frac{\vec{F}}{q_2} = \frac{q_1 q_2}{4\pi \epsilon r^2 q_2} \\ &= \frac{q_1}{4\pi \epsilon r^2} \\ &= \frac{1}{\epsilon} \cdot \frac{q_1}{4\pi r^2}\end{aligned}$$

Using Eq. (10.12) in the above equation, we get

$$\begin{aligned}\vec{E} &= \frac{\vec{D}}{\epsilon} \\ \text{or} \quad \vec{D} &= \epsilon \vec{E} \\ &= \epsilon_0 \cdot \epsilon_r \vec{E}\end{aligned} \quad (10.13)$$

where ϵ_r is relative permittivity for free space, $\epsilon_r = 1$ and, hence, the above expression can be given as

$$\vec{D} = \epsilon_0 \vec{E} \quad (10.14)$$

The unit of displacement vector or electric flux density is coulomb/metre².

10.6 RELATION BETWEEN VECTORS \vec{E} , \vec{D} , AND \vec{P}

If a dielectric slab is placed between the plates of parallel plate capacitor, then the resultant electric field within the dielectric material can be given as

$$\vec{E} = \vec{E}_0 - \vec{E}'$$

where \vec{E}_0 is the applied field and E' is the induced field.

From the Gauss law of electrostatics, we have

$$\epsilon_0 \oint \vec{E}_0 \cdot d\vec{s} = Q$$

$$\text{or } \vec{E}_0 = \frac{Q}{\epsilon_0 A}$$

Similarly,

$$\vec{E}' = \frac{Q'}{\epsilon_0 A} \text{ and } \vec{E} = \frac{Q}{K\epsilon_0 A}$$

By putting the values of \vec{E} , \vec{E}_0 , and \vec{E}' in the above equation, we get

$$\begin{aligned} \frac{Q}{K\epsilon_0 A} &= \frac{Q}{\epsilon_0 A} - \frac{Q'}{\epsilon_0 A} \\ \text{or } \frac{Q}{A} &= \epsilon_0 \left(\frac{Q}{K\epsilon_0 A} \right) + \frac{Q'}{A} \end{aligned} \quad (10.15)$$

From Eq. (10.12), we have

$$\vec{D} = \frac{Q}{A}$$

and from Eq. (10.9), we have

$$\vec{P} = \frac{Q'}{A}$$

Using these values in Eq. (10.15), we get

$$\vec{D} = \epsilon_0 \left(\frac{Q}{K\epsilon_0 A} \right) + \vec{P}$$

or $\vec{D} = \epsilon_0 \vec{E} + \vec{P}$, where $\vec{E} = \frac{Q}{K \epsilon_0 A}$ (10.16)

or $\vec{D} = \left(\epsilon_0 + \frac{\vec{P}}{\vec{E}} \right) \vec{E}$ (10.17)

On the basis of the above discussion, we can conclude the following:

- \vec{D} is associated with free charge only. The lines of \vec{D} begin and end on free charges. \vec{D} is constant throughout the capacitor.
- \vec{P} is associated with polarisation (induced) charge only. The lines of \vec{P} begin and end on polarisation (induced) charges. \vec{P} is zero except inside the dielectrics.
- \vec{E} is associated with all charges.
- In isotropic media, \vec{P} and \vec{E} are in the same direction. However, for non-isotropic media, \vec{P} and \vec{E} , in general, are not in the same direction.
- \vec{E} is reduced inside the dielectric where there are fewer lines.
- The displacement vector \vec{D} has no clear physical meaning. It is introduced only because it enables one to calculate fields in the presence of dielectrics without knowing the details of prior distribution of polarisation (induced) charges.

10.7 TYPES OF POLARISATION

Dielectric polarisation is nothing but the displacement of charged particles under the action of the electric field they are subjected to. Some microscopic mechanisms occur in the presence of external electric field in the dielectric material leading to its polarisation. These mechanisms can be grouped into four sections as given below:

- Electronic polarisation
- Ionic polarisation
- Orientational polarisation
- Space-charge polarisation

10.7.1 Electronic Polarisation

Electronic polarisation is the result of displacement of positive and negative charges of an atom in the presence of external electric field. Usually, in an atom, protons and neutrons are concentrated at its centre to form the nucleus, and the electrons revolve around the nucleus. Hence, the positive and negative charges of the atom are balanced by each other as shown in Fig. 10.5(a). When the same atom is subjected to an external electric field, the nucleus and the electron experience Lorentz forces in opposite directions. Therefore, nucleus and electron clouds are pulled apart as shown in Fig. 10.5(b) and a coulomb force develops between them, which tends to counter at the displacement. Experimentally it is observed that the order of the magnitude of displacement between nucleus and electron cloud is 10^{-17} m corresponding to the electric field of 30 kV/m.

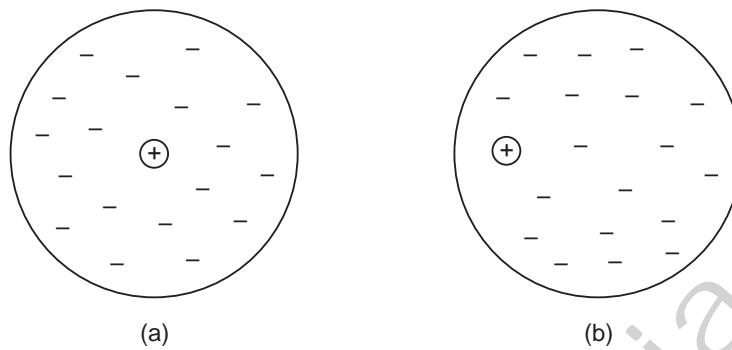


Fig. 10.5 (a) Charge distribution in an atom in the absence of field and (b) charge distribution in an atom in the presence of field

As the nucleus and the centre of electron cloud are separated by a certain distance, dipole moment is created in each atom. The induced dipole moment (\vec{p}) is proportional to externally applied electric field (\vec{E}), i.e.,

$$\vec{p}_e \propto \vec{E}$$

If there are N atoms in the dielectric, then

$$\vec{P}_e \propto N \vec{E}$$

or
$$\vec{P}_e = \alpha_e N \vec{E} \quad (10.18)$$

where α_e is the electronic polarisability. Electronic polarisability is independent of temperature.

10.7.2 Ionic Polarisation

The dielectric materials having ionic bonds, such as NaCl, show ionic polarisation. Polarisation in such ionic crystals arises on account of the ions displaced from their equilibrium positions by the force of the applied electric field. Let us consider the case of sodium chloride in which the ions (i.e., Na^+ and Cl^-) are at a distance of r_0 in the absence of external electric field [Fig. 10.6(a)]. When this sample of sodium chloride is subjected to an external electric field, then each Na^+ ion is displaced in one direction and each Cl^- ion in the opposite direction.

The displacement between ions causes an increase or decrease in distance between the atoms, depending on the location of the ion pair. This process leads to a net dipole moment. Similar to electronic polarisation, this polarisation is also time independent.

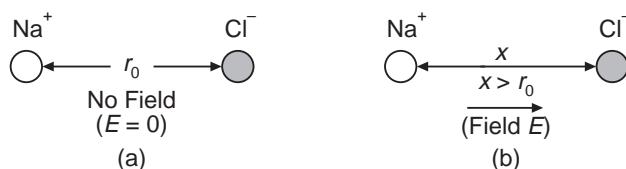


Fig. 10.6 (a) Ionic bond without external electric field and (b) ionic bonds with external electric field

It has been observed that the induced dipole moment due to ionic polarisation is proportional to the applied electric field, i.e.,

$$p_i \propto E$$

or $\bar{p}_i = N\alpha_i E$

where α_i is the ionic polarisability. If there are N molecules, then the ionic polarisation can be given as

$$p_i = N\alpha_i E \quad (10.19)$$

For most of the materials, the ionic polarisability is very less than the electronic polarisability. It is generally observed that

$$\alpha_i = 0.1 \alpha_e \quad (10.20)$$

10.7.3 Orientational Polarisation

Orientational polarisation occurs in polar substances. These substances exhibit dipole moment even in the absence of external electric field. Due to the random orientations of dipoles or molecules, the net dipole moment is zero. The random orientation of dipoles is shown in Fig. 10.7 (a). When such materials are subjected to an external electric field, the permanent molecular dipoles rotate about their axis of symmetry to align with the applied field, which exerts a torque on them. This type of polarisation is known as *orientational polarisation*. The aligned dipoles along the direction of field are shown in Fig. 10.7(b).

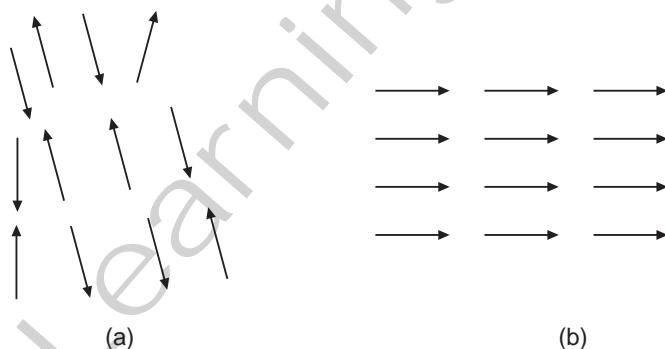


Fig. 10.7 (a) Orientation of dipoles in the absence of electric field and (b) orientation of dipoles in the direction of applied electric field

In electronic and ionic polarisation processes, the force due to the externally applied field is balanced by elastic binding forces, but no such forces exist for orientational polarisation. At thermal equilibrium with no external electric field, the permanent dipoles contribute no net polarisation since they are randomly oriented. With the application of external electric field, dipole alignment is largely offset by thermal agitation. The orientational polarisation is strongly temperature dependent; it decreases with increase in the temperature. Orientational polarisability is denoted by α_o .

10.7.4 Space-Charge Polarisation

Sometimes, due to the application of electric field to the dielectric material, charges accumulate at the electrodes or at the interphases in the multiphase dielectric due to sudden change in conductivity [Fig.

11

Magnetic Properties of Materials

11.1 INTRODUCTION

It is well known that when a potential difference is applied across the ends of a conducting wire, electric current flows due to the motion of electrons. Biot–Savart law and other related experiments showed the presence of magnetic field in the vicinity of a current-carrying wire.

A microscopic view of an atom gives the basis for the origin of the field of magnetism. An *atom* consists of a positively charged nucleus, surrounded by a number of electrons. These electrons have two types of motion, orbital motion about the nucleus and spin motion about their own axes. Accordingly, these motions of electrons produce their own magnetic fields. Thus, electrons moving around the nucleus possess a dipole magnetic moment due to their orbital as well as spin motions. In normal state, these magnetic moments are randomly oriented, and hence, the resultant magnetic moment of an atom is zero. When we apply an external magnetic field, the magnetic moments tend to align themselves in the direction of external magnetic field and the material is then said to be *magnetised*.

11.2 USEFUL PARAMETERS

11.2.1 Magnetic Density (*B*)

When a magnetic material is placed in an external magnetising field (*H*), the number of lines of force inside the material, crossing per unit area of the material, is called *magnetic flux density* (*B*) and the phenomenon is called *magnetic induction*. Its unit is weber/m² or tesla.

11.2.2 Intensity of Magnetisation (*I*)

When a magnetic material is placed in an external magnetic field, it acquires some magnetic moment (*M*). The acquired magnetic moment per unit volume is known as the *intensity of magnetisation* (*I*) and is given as

$$I = \frac{M}{V} \quad (11.1)$$

11.2.3 Magnetic Susceptibility (χ)

It has been observed that the intensity of magnetisation (*I*) is directly proportional to the intensity of the magnetising field (*H*), i.e.,

$$\begin{aligned} I &\propto H \\ \text{or} \quad I &= \chi_m H \\ \text{or} \quad \chi_m &= \frac{I}{H} \end{aligned} \tag{11.2}$$

where χ_m is the *magnetic susceptibility* of the material. It is a dimensionless quantity and is a characteristic of the material.

11.2.4 Relative Permeability (μ_r)

The capability of the magnetic lines of force to penetrate the material or a medium is called *permeability* (μ) of that material or medium.

The magnetic flux density (B) is related with magnetising field as

$$\begin{aligned} B &\propto H \\ \text{or} \quad B &= \mu H \end{aligned} \tag{11.3}$$

where μ is a constant of proportionality called *permeability*. If the medium is vacuum or air, the above expression takes the form

$$B_0 = \mu_0 H \tag{11.4}$$

where μ_0 is called *absolute permeability*.

Dividing Eq. (11.3) by Eq. (11.4), we get

$$\frac{B}{B_0} = \frac{\mu}{\mu_0} = \mu_r \tag{11.5}$$

where μ_r is the *relative permeability* of the medium or material.

11.2.5 Relation between μ_r and χ_m

The magnetic flux density (B), magnetising field (H), and the intensity of magnetisation (I) are related as follows:

$$B = \mu_0 (H + I) \tag{11.6}$$

Using Eq. (11.3), we get

$$\begin{aligned} \mu H &= \mu_0 (H + I) \\ &= \mu_0 H \left(1 + \frac{I}{H}\right) \end{aligned}$$

Using Eq. (11.2), we get

$$\mu = \mu_0 (1 + \chi_m)$$

$$\frac{\mu}{\mu_0} = 1 + \chi_m$$

Now, using Eq. (11.5), we get

$$\mu_r = 1 + \chi_m \tag{11.7}$$

11.3 MAGNETIC MOMENT OF AN ATOM: BOHR MAGNETON

We know that in an atom, electrons revolve around the nucleus. Due to their motion, they produce electric current, and hence, a revolving electron also produces a magnetic field. This makes an atom equivalent to a magnetic dipole whose magnetic moment can be calculated as follows:

From the basic definition of *current*, i.e., the rate of flow of charge, we can write, for an electron having charge e and time period T to complete its one revolution, the equivalent current would be

$$i = \frac{e}{T}$$

If r is the radius of the orbit and v is the speed of the electron, we can write

$$T = \frac{2\pi r}{v} \quad \left(\text{Time} = \frac{\text{distance}}{\text{speed}} \right)$$

$$\therefore i = \frac{ev}{2\pi r}$$

Now, magnetic moment is given by

$$M = iA = \frac{ev}{2\pi r} \cdot \pi r^2 = \frac{evr}{2}$$

From Bohr's third postulate, the angular momentum of an electron is given by

$$mvnr = \frac{nh}{2\pi} \quad (\text{where, } n = 1, 2, 3, \dots)$$

$$\text{or} \quad vr = \frac{nh}{2\pi m}$$

$$\begin{aligned} \text{Thus,} \quad M &= \frac{e}{2} \cdot \frac{nh}{2\pi m} \\ &= n \left(\frac{eh}{4\pi m} \right) \end{aligned}$$

This shows that magnetic moment of an atom is the integral multiple of $eh/4\pi m$. It is also clear that the minimum value of magnetic moment is obtained for $n = 1$ and is equal to $eh/4\pi m$, called *Bohr magneton*. Its value is $9.28 \times 10^{-24} \text{ A-m}^2$.

11.4 CLASSIFICATION OF MAGNETIC MATERIALS

On the basis of their magnetic properties, different materials are classified as follows:

- (i) Diamagnetic substances
- (ii) Paramagnetic substances
- (iii) Ferromagnetic substances

11.4.1 Diamagnetic Substances

Diamagnetism occurs in those materials in which individual atoms do not possess any net magnetic moment as their orbital and spin magnetic moments add vectorially to become zero. The atoms of such materials, however, acquire an induced dipole moment when they are placed in an external magnetic field. The diamagnetic materials are *Type I superconductors* as they exhibit perfect conductivity and perfect diamagnetisation when cooled to very low temperatures. The superconductor repels a magnet and, in turn, is repelled. Such perfect diamagnetism in superconductors is called *Meissner effect*.

Some important properties of diamagnetic substances are as follows:

- When suspended in an uniform magnetic field, diamagnetic substances align their longest axis at right angles to the field as shown in Fig. 11.1.

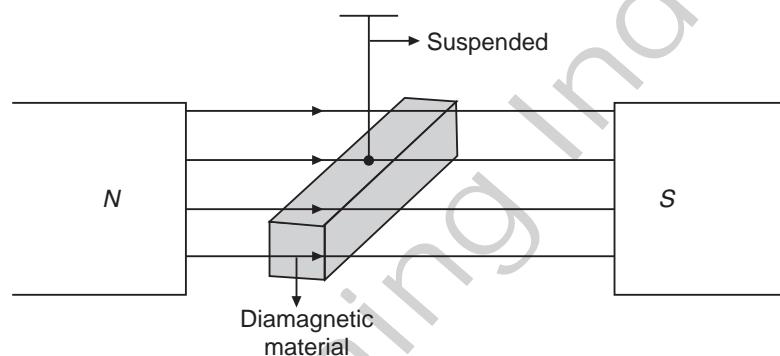


Fig. 11.1 Freely suspended diamagnetic bar

- In a non-uniform magnetic field, these substances move from stronger parts of the field to the weaker parts. For example, when a diamagnetic liquid is put in a watch glass placed on the two pole pieces of an electromagnet and current is switched on, the liquid accumulates on the sides. It has been observed that on increasing the distance between the poles, this effect is reversed.
- A diamagnetic liquid in a U-shaped tube is depressed when subjected to a magnetic field (Fig. 11.2).

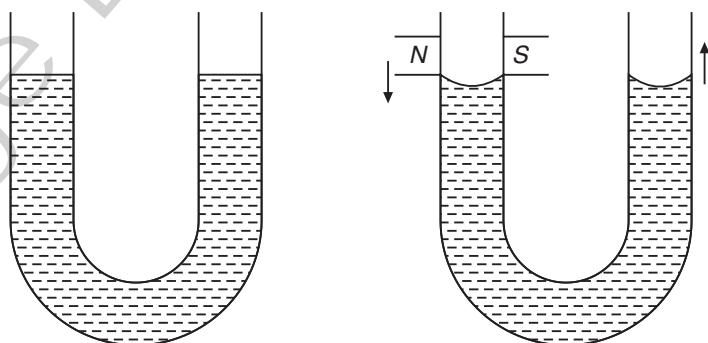


Fig. 11.2 Diamagnetic liquid

- (iv) The lines of force do not prefer to pass through a diamagnetic material, since its ability to permit the passage of magnetic lines of force through it is less (Fig. 11.3).

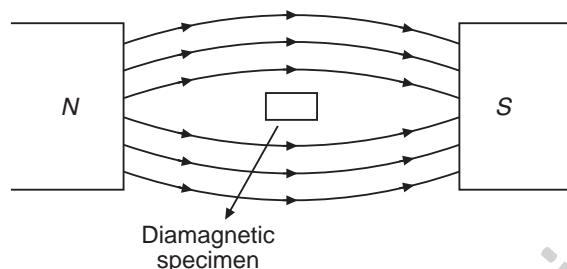


Fig. 11.3 Lines of force for diamagnetic material

- (v) The permeability of the diamagnetic substances is less than one, i.e., $m_r < 1$.
 (vi) Diamagnetic substances lose their magnetisation as soon as the magnetising field is removed.
 (vii) These substances cannot be easily magnetised and so their susceptibility is negative.
 Some examples of diamagnetic substances are bismuth, antimony, copper, gold, quartz, mercury, water, alcohol, air, hydrogen, etc.

11.4.2 Paramagnetic Substances

Paramagnetic substances are very feebly attracted by magnets. In a paramagnetic material, the atomic dipole moments are initially, randomly oriented in space. When an external field is applied, the dipoles align with the field as shown in Fig. 11.4. The vector sum of the individual dipole moments is no longer zero.

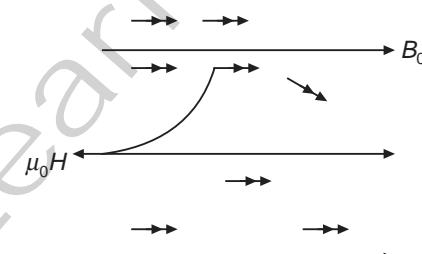


Fig. 11.4 Alignment of dipoles

Some important properties of paramagnetic substances are as follows:

- (i) The paramagnetic substances develop a weak magnetisation in the direction of the field.
 (ii) When a paramagnetic rod is suspended freely in a uniform magnetic field, it aligns itself in the direction of magnetic field (Fig. 11.5).

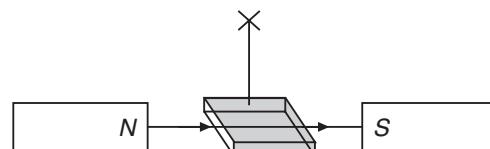


Fig. 11.5 Freely suspended paramagnetic rod

- (iii) The lines of force prefer to pass through the material rather than through air, as their permeability is greater than one, i.e., $\mu_r > 1$ (Fig. 11.6).

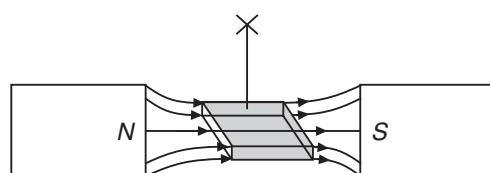


Fig. 11.6 Lines of force for paramagnetic rod

- (iv) As soon as the magnetic field is removed, the paramagnetic substances lose their magnetisation.
 (v) In a non-uniform magnetic field, the paramagnetic substances move from weaker parts of the field to the stronger parts (i.e., they accumulate in the middle of the magnetic field).
 (vi) A paramagnetic liquid in a U-shaped tube placed between two poles of a magnet gets elevated (Fig. 11.7).

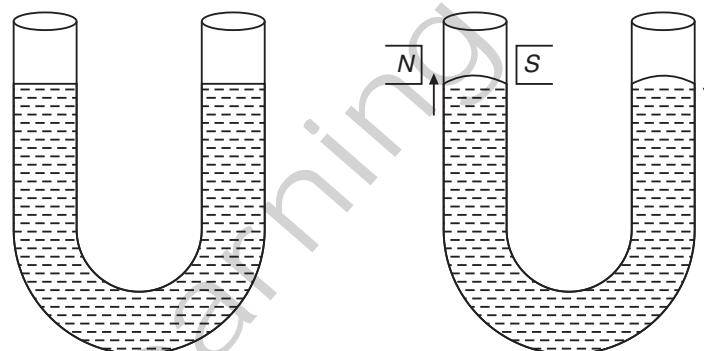


Fig. 11.7 Paramagnetic liquid

- (vii) The magnetisation of paramagnetism decreases with the increase in temperature. This is because the thermal motions of the atoms tend to disturb the alignment of the dipoles.

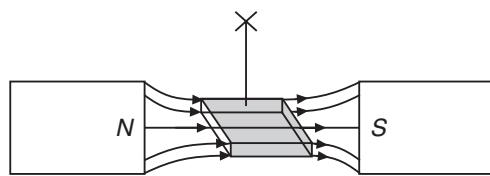
Some examples of paramagnetic substances are aluminium, platinum, chromium, manganese, copper sulphate, oxygen, etc.

11.4.3 Ferromagnetic Substances

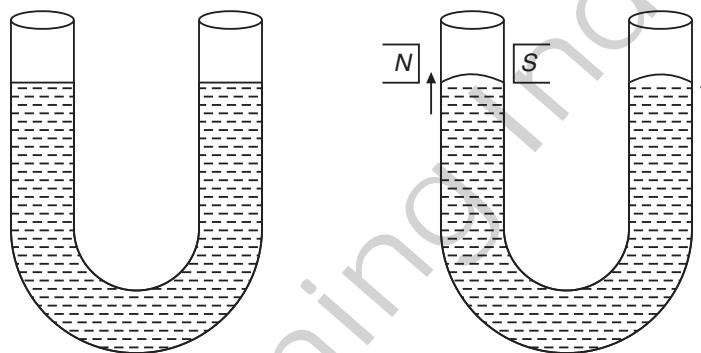
Ferromagnetism occurs in materials in which atoms have permanent magnetic dipole moments. The strong interaction between neighbouring atomic dipole moments keeps them aligned even when the external magnetic field is removed.

Some important properties of ferromagnetic materials are as follows:

- These substances get strongly magnetised in the direction of the field.
- The lines of force prefer to pass through the material rather than through air, as their permeability is greater than one, i.e., $\mu_r > 1$ (Fig. 11.8).

**Fig. 11.8** Lines of force for ferromagnetic rod

- (iii) In a non-uniform magnetic field, ferromagnetic substances move from weaker parts of the field to the stronger parts (i.e., they accumulate in the middle of the magnetic field).
- (iv) A ferromagnetic liquid in a U-shaped tube placed between the two poles of a magnet is elevated (Fig. 11.9).

**Fig. 11.9** Ferromagnetic liquid

- (v) For ferromagnetic materials, m_r is very large, and therefore, their susceptibility, i.e., χ_m is positive.

11.5 LANGEVIN'S THEORY OF DIAMAGNETISM

According to Langevin (1905), in diamagnetic substances, the electrons revolving in different orbits are paired ($+1/2$ and $-1/2$ spin). Now, it is known that a revolving electron must have a dipole moment. Hence, in case of paired electrons, the vector sum of the total magnetic moment will be zero. When an external magnetic field is applied, the electronic orbits of the atom start precessional motion about the direction of the applied field. This precession is known as *Larmor precession*. Other than the precessional motion, in the presence of Lorentz force, the angular velocity of one of the electrons increases while others' decrease.

11.5.1 Analysis

Let us assume an electron revolving in a circular orbit of radius r with an angular frequency ω_0 . The centripetal force acting on the electron (f_e) is

$$f_e = m\omega_0^2 r = \frac{mv_0^2}{r} \quad (11.8)$$

where $v_0 (= r\omega_0)$ is the speed of the electron in the orbit.

In the presence of an external magnetic field of flux density B applied perpendicular to the plane of the orbit, the *Lorentz force* (F_m) on the electron is given by

$$F_m = e(\vec{v} \times \vec{B}) \quad (11.9)$$

This force is at right angle to the motion of the electron, and the direction is determined by Fleming's left-hand rule as shown in Fig. 11.10.

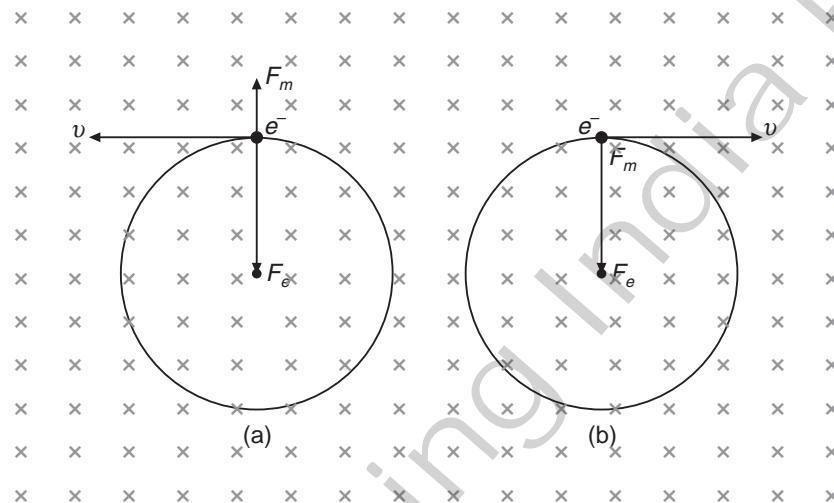


Fig. 11.10 Motion of electron in the magnetic field

Now, using Faraday's law of electromagnetic induction, the emf induced due to the application of external magnetic field is

$$\epsilon = -\frac{d\phi}{dt} \quad (11.10)$$

This emf accelerates the electron, i.e.,

$$m \frac{dv}{dt} = eE \quad (11.11)$$

According to definition, the induced emf is related to the corresponding field as

$$\epsilon = \oint E \cdot dl \text{ or } \epsilon = E \cdot 2\pi r \text{ or } E = \frac{\epsilon}{2\pi r} \quad (11.12)$$

Hence,

$$m \frac{dv}{dt} = \frac{e}{2\pi r} \cdot \left(-\frac{d\phi}{dt} \right)$$

$$m \frac{dv}{dt} = -\frac{e}{2\pi r} \cdot \frac{d\phi}{dt} \quad (11.13)$$

When the magnetic field is applied, its value varies from zero to the value B . Hence, the corresponding change in flux is

$$d\phi = \pi r^2 B - 0 = \pi r^2 B \quad (\phi = BA)$$

Hence, from Eq. (11.13), we get

$$mdv = -\frac{e}{2\pi r} \cdot d\phi$$

$$\begin{aligned} dv &= -\frac{e}{2\pi rm} \pi r^2 B \\ &= -\frac{eBr}{2m} \end{aligned}$$

Therefore, change in angular velocity, or angular frequency, is

$$d\omega = \frac{dv}{r} = -\frac{eB}{2m} \quad (11.14)$$

This change in frequency is called *Larmor frequency*, and it produces a change in the magnetic moment.

However, the magnetic moment is given by

$$\begin{aligned} M &= iA = \text{frequency} \times \text{charge} \times \text{area} \\ &= \frac{\omega}{2\pi} \times e \times \pi r^2 \\ &= \frac{er^2\omega}{2} \end{aligned} \quad (11.15)$$

Thus, change in magnetic moment is

$$\begin{aligned} \Delta M &= \frac{er^2}{2} \Delta\omega = \frac{er^2}{2} \left(-\frac{eB}{2m} \right) \\ \Delta M &= \frac{e^2 r^2 B}{4m} \end{aligned} \quad (11.16)$$

Equation (11.16) gives the expression for change in magnetic moment for only one electron. To consider an atom, the projection of radius (r) must be taken into account such that the plane of the orbit is perpendicular to the applied field or vice versa. Let the projection of r be r_1 . Hence, the total change in magnetic moment due to randomly oriented orbits is

$$M = -\frac{e^2 B}{4m} \sum r_1^2 \quad (11.17)$$

If the field is along the z -axis (perpendicular to the $x-y$ plane), then we have

$$r_1^2 = x^2 + y^2$$

For spherically symmetric atom, we have

$$\bar{x}^2 = \bar{y}^2 = \frac{\bar{r}^2}{3} \text{ (say)}$$

where bars on x and y denote mean values.

Hence,

$$r_1^2 = \bar{x}^2 + \bar{y}^2 = \frac{2\bar{r}^2}{3}$$

Therefore, Eq. (11.17) gives

$$\begin{aligned} M &= -\frac{e^2 B}{4m} \sum \frac{2}{3} \bar{r}^2 \\ M &= -\frac{e^2 B}{6m} \sum \bar{r}^2 \end{aligned} \quad (11.18)$$

Since, $I = M/V$ and $B = \mu H$, the intensity of magnetisation for n number of molecules per unit volume is

$$I = -\frac{e^2 n B}{6m} \sum \bar{r}^2$$

Hence, the susceptibility is given by

$$\begin{aligned} \chi_m &= \frac{I}{H} = \frac{-e^2 n \mu H}{6m H} \sum \bar{r}^2 \\ \text{or } \chi_m &= -\frac{n \mu e^2}{6m} \sum \bar{r}^2 \end{aligned} \quad (11.19)$$

Equation (11.19) shows that the susceptibility of diamagnetic materials is negative and is independent of temperature and field strength.

11.6 LANGEVIN'S THEORY OF PARAMAGNETISM

Langevin derived expression for susceptibility of paramagnetic gas on the basis of kinetic theory of gases. When subjected to an external magnetic field, the magnetic dipoles tend to align themselves along the direction of the field, but due to thermal agitation, they remain at an angle with the field.

11.6.1 Analysis

Let μ_m , the magnetic moment of a molecule, make an angle θ with the direction of external magnetic field B . The potential energy of the dipole along the direction of field is

$$U = -\mu_m B \cos \theta \quad (11.20)$$

Now, using the theory of statistical mechanics, the number of molecules (dn) having energy U between angles θ and $\theta + d\theta$ at absolute temperature T is given by (Fig. 11.11)

12

Interference

12.1 INTRODUCTION (NATURE OF LIGHT)

Upto the middle of the seventeenth century it was generally believed that light consisted of a stream of corpuscles. These corpuscles could penetrate transparent materials and were reflected from the surfaces of opaque materials. When they entered the eye, they caused the sensation of light.

Huygen in 1670 showed that the laws of reflection and refraction could be explained on the basis of wave theory. The wave theory could however not be immediately accepted. It was objected that if light were a wave motion, one should be able to see around corners, since wave can bend around obstacles in their path. We know now that the wavelengths of light waves are so small that the bending, which actually does take place, is so small that it is not ordinarily observed. However, the bending of light waves around the edges of an object (diffraction) was noted by Grimaldi and interpreted by Hooke in terms of a wave picture, but its significance was not recognised at that time.

In about 1827, Young's experiments enabled him to measure the wavelength of the waves, and Fresnel showed that the rectilinear propagation of light along with the diffraction effects observed by Grimaldi, could be explained if light was considered as waves of short wavelengths.

In 1873, Maxwell suggested that light consisted of electromagnetic wave of extremely short wavelengths. After some years, Hertz succeeded in producing short wavelength waves of electromagnetic origin and showed that they possessed all the characteristics of light wave, such as refraction, reflection, polarisation, etc. This Maxwell's electromagnetic theory of light was experimentally justified. The electromagnetic theory, however, failed to account for the phenomenon of photoelectric emission.

Einstein (1905) postulated that energy in a light beam was concentrated in small packet or photons. The photoelectric effect thus consisted in the transfer of energy from a photon to an electron. Compton, in 1921, determined the motion of a photon and a single electron, both before and after a *collision* between them, and found that they behaved like material bodies having kinetic energy and momentum, both of which were conserved during collision. The photoelectric effect and Compton effect thus demanded a return to a corpuscular theory of light.

Thus, the present scenario is to accept the fact that light is dualistic in nature. The phenomenon of light propagation may best be explained by the electromagnetic wave theory, while the interaction of light with matter in the processes of emission and absorption is a corpuscular phenomenon.

12.2 INTERFERENCE OF LIGHT

When two waves of the same frequency having a constant initial phase difference traverse simultaneously in a medium superpose each other, the resultant intensity of light is not distributed uniformly in space. This modification in intensity is called *interference*. At some points of the medium the superposition takes place in such a way that the resultant intensity is greater than the sum of intensities of individual waves. This type of interference is called *constructive interference*, while at some points of the medium the resultant intensity is found to be less than the sum of the intensities of individual waves. This is classified as *destructive interference*.

12.3 YOUNG'S EXPERIMENT (DIVISION OF WAVEFRONT)

Thomas Young, in 1801, demonstrated the phenomenon of interference of light as shown in Fig. 12.1. The sunlight was first allowed to pass through a pinhole S , and then through two pinholes S_1 and S_2 placed at a sufficient distance away from S and the light was received on a screen. The two sets of spherical waves emerging from S_1 and S_2 interfered with each other and a few coloured fringes of varying intensity were observed on the screen.

As an improvement of the original arrangement, the pinholes S_1 and S_2 are replaced by narrow slits and sunlight by a monochromatic light. The interfering waves are then cylindrical and a number of alternate dark and bright fringes running parallel to the length of the slits are observed on the screen.

Explanation

On the basis of wave theory of light, the formation of dark and bright fringes can be explained. The cylindrical wavefronts starting from S falls on S_1 and S_2 . According to Huygens principle, S_1 and S_2 become centres of secondary wavelets, i.e., two cylindrical wavefronts emerge out from S_1 and S_2 . Their radii increase as they move away from S_1 and S_2 , so that they superimpose more and more on each other. At points where crest (or trough) due to one falls on a crest (or trough) due to the other, the resultant amplitude is the sum of the amplitudes due to each wave separately. The intensity, which is proportional to the square of the amplitude, at these points is therefore a maximum. This is the case of constructive interference. At points where a crest due to one falls on a trough due to the other, the resultant amplitude is the difference of the amplitudes due to separate waves and the resultant intensity is a minimum. This is the case of destructive interference.

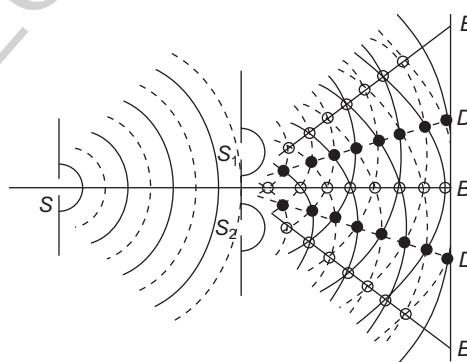


Fig. 12.1 Young's experiment showing the phenomenon of interference

In Fig. 12.1, the solid arcs indicate the crests while the dotted arcs indicate the troughs. The solid lines are the loci of the points of maximum intensity and are called *anti-nodal lines*. The broken lines are the loci of the points of minimum intensity and are called *nodal lines*. However, actually these lines are hyperbola. The intersections of these lines on the screen give the positions of bright and dark fringes which occur alternately at equal distances.

Young's experiment demonstrates both the diffraction of light waves at the slits and the interference between the light waves emerging from the slits.

12.4 RESULTANT INTENSITY DUE TO SUPERPOSITION OF TWO INTERFERING WAVES

Let S be a narrow slit illuminated by a monochromatic light source, and S_1 and S_2 are two similar parallel slits very close together and equidistant from S (Fig. 12.2).

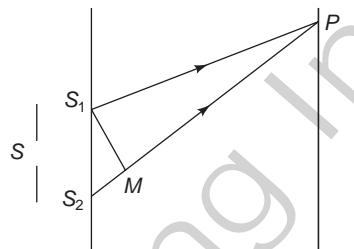


Fig. 12.2 Superposition of two waves

Let the waves from S reach at S_1 and S_2 in the same phase such that the waves proceed as if they started from S_1 and S_2 . We have to find out the resultant intensity at P on a screen placed parallel to S_1 and S_2 .

12.4.1 Analysis

Let a_1 and a_2 be the amplitudes at P due to the waves from S_1 and S_2 , respectively. The waves reaching at P will have different paths S_1P and S_2P . Hence, they will superimpose with a phase difference δ given by

$$\begin{aligned}\delta &= \frac{2\pi}{\lambda} \times \text{path difference} \\ &= \frac{2\pi}{\lambda} \times (S_2P - S_1P) = \frac{2\pi}{\lambda} \times S_2M\end{aligned}\quad (12.1)$$

where λ is the wavelength of light used.

The individual displacements at P will then be represented by

$$y_1 = a_1 \sin \omega t$$

$$y_2 = a_2 \sin (\omega t + \delta)$$

where $\omega/2\pi$ is the common frequency of the two waves.

Due to the principle of superposition, the resultant displacement

$$\begin{aligned}y &= y_1 + y_2 \\ &= a_1 \sin \omega t + a_2 \sin (\omega t + \delta) \\ &= a_1 \sin \omega t + a_2 (\sin \omega t \cos \delta + \cos \omega t \sin \delta)\end{aligned}$$

$$\begin{aligned}
 &= a_1 \sin \omega t + a_2 \cos \delta \sin \omega t + a_2 \sin \delta \cos \omega t \\
 &= (a_1 + a_2 \cos \delta) \sin \omega t + (a_2 \sin \delta) \cos \omega t.
 \end{aligned}$$

Let us make a change in constants as

$$a_1 + a_2 \cos \delta = R \cos \theta \quad (12.2)$$

$$\text{and} \quad a_2 \sin \delta = R \sin \theta \quad (12.3)$$

where R and θ are new constants. This will give

$$\begin{aligned}
 y &= R \cos \theta \sin \omega t + R \sin \theta \cos \omega t \\
 y &= R \sin(\omega t + \theta)
 \end{aligned} \quad (12.4)$$

Hence superposition of two simple harmonic waves will not change the frequency and nature of resultant wave.

Squaring and adding Eqs. (12.2) and (12.3) will give the resultant amplitude R as

$$\begin{aligned}
 R^2 \cos^2 \theta + R^2 \sin^2 \theta &= (a_1 + a_2 \cos \delta)^2 + (a_2 \sin \delta)^2 \\
 \Rightarrow R^2 &= a_1^2 + a_2^2 + 2a_1 a_2 \cos \delta
 \end{aligned} \quad (12.5)$$

Therefore, resultant intensity at P is

$$\begin{aligned}
 I &\propto R^2 \\
 \Rightarrow I &= kR^2
 \end{aligned}$$

where k is the constant of proportionality whose value is found to be unity if the medium of travelling of the waves is same. That is,

$$\begin{aligned}
 I &= R^2 \\
 \text{or} \quad I &= a_1^2 + a_2^2 + 2a_1 a_2 \cos \delta^*
 \end{aligned} \quad (12.6)$$

12.4.2 Conditions for Maxima and Minima

It is clear from Eq. (12.6) that at point P , the intensity will be a maximum whenever

$$\cos \delta = +1$$

$$\text{or} \quad \delta = 2n\pi, \quad n = 0, 1, 2, \dots$$

So, path difference $(S_2 P - S_1 P) = n\lambda$ [from Eq. (12.1)]

Thus from Eq. (12.6), we have

$$\begin{aligned}
 I_{\max} &= a_1^2 + a_2^2 + 2a_1 a_2 \\
 &= (a_1 + a_2)^2 > a_1^2 + a_2^2, \text{ i.e., } I_{\max} > I_1 + I_2
 \end{aligned}$$

Similarly, the resultant intensity will be a minimum whenever

$$\cos \delta = -1$$

$$\text{or} \quad \delta = (2n + 1)\pi, \quad n = 0, 1, 2, \dots$$

So, path difference $(S_2 P - S_1 P) = (2n + 1)\lambda/2$.

* This expression gives the resultant intensity at P due to two waves of amplitudes a_1 and a_2 and having a phase difference δ between them.

Then from Eq. (12.6), we get

$$\begin{aligned}I_{\min} &= a_1^2 + a_2^2 - 2a_1 a_2 \\&= (a_1 - a_2)^2 < a_1^2 - a_2^2, \text{ i.e., } I_{\min} < I_1 - I_2\end{aligned}$$

Thus, as we move on the screen, the path difference between the two waves gradually changes and there is a variation in the intensity of light being alternately maximum and minimum. This is called *interference pattern*.

Solved Examples

Example 12.1

The path difference between the two interfering rays at a point on the screen is $1/8$ th of the wavelength. Find the ratio of the intensity at this point to that at the centre of a bright fringe.

Solution

The intensity at any point is given by

$$I = a_1^2 + a_2^2 + 2a_1 a_2 \cos \delta$$

Here $a_1 = a_2 = a$.

$$\begin{aligned}\text{Therefore, } I &= a^2 + a^2 + 2a^2 \cos \delta \\&= 2a^2(1 + \cos \delta)\end{aligned}$$

At the centre, phase difference $\delta = 0$

$$\therefore I_0 = 2a^2(1 + \cos \theta) = 2a^2(1 + 1) = 4a^2$$

At the point, where the path difference is $\lambda/8$,

$$\begin{aligned}\text{Phase difference} &= \frac{2\pi}{\lambda} \times \text{path difference} \\&= \frac{2\pi}{\lambda} \times \frac{\lambda}{8} = \frac{\pi}{4} \\ \therefore I_1 &= 2a^2 \left(1 + \cos \frac{\pi}{4}\right) = 2a^2 (1 + 0.707) \\&= a^2 \times 3.414\end{aligned}$$

Therefore,

$$\frac{I_1}{I_0} = \frac{3.414a^2}{4a^2} = 0.853$$

Example 12.2

Two coherent sources whose intensity ratio is 100:1 produce interference fringes. Find the ratio of maximum intensity to minimum intensity in the interference pattern.

Solution

The ratio of maximum intensity to the minimum intensity is given by

$$\frac{I_{\max}}{I_{\min}} = \frac{(a_1 + a_2)^2}{(a_1 - a_2)^2} \quad (1)$$

$$\text{Here, } \frac{I_1}{I_2} = \frac{a_1^2}{a_2^2} = \frac{100}{1}$$

$$\therefore \frac{a_1}{a_2} = \frac{10}{1}$$

$$\Rightarrow a_1 = 10a_2 \quad (2)$$

Substituting the value of a_1 from Eq. (2) in Eq. (1), we get

$$\frac{I_{\max}}{I_{\min}} = \frac{(10a_2 + a_2)^2}{(10a_2 - a_2)^2} = \frac{(11)^2}{(9)^2} = \frac{121}{81}$$

Example 12.3

In an interference pattern, the amplitude of intensity variation is found to be 5% of the average intensity. Calculate the relative intensities of the interfering sources.

Solution

It is given that the intensity variation is 5% of the average intensity. So, if the average intensity is 100, then the maximum intensity is 105 and the minimum intensity is 95.

$$\therefore \frac{I_{\max}}{I_{\min}} = \frac{(a_1 + a_2)^2}{(a_1 - a_2)^2} = \frac{105}{95}$$

$$\text{or } \frac{(a_1 + a_2)}{(a_1 - a_2)} = 1.051$$

$$\text{or } (a_1 + a_2) = 1.051(a_1 - a_2) = 1.051a_1 - 1.051a_2$$

$$\text{or } 0.051a_1 = 2.051a_2$$

$$\text{or } \frac{a_1}{a_2} = \frac{2.051}{0.051} = 40$$

$$\therefore a_1 = 40a_2$$

Therefore, the ratio of the intensities of interfering sources is

$$\frac{I_1}{I_2} = \frac{a_1^2}{a_2^2} = \frac{(40a_2)^2}{(a_2)^2} = \frac{1600}{1}$$

$$\text{or } I_1 : I_2 = 1600 : 1$$

Example 12.4

Two coherent sources of intensity ratio α interfere. Prove that in the interference pattern,

$$\frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} = \frac{2\sqrt{\alpha}}{1+\alpha}$$

Solution

Let I_1 and I_2 be the intensities and a_1 and a_2 be the corresponding amplitudes of the two coherent sources.

$$\begin{aligned} \therefore \quad \frac{I_1}{I_2} &= \frac{a_1^2}{a_2^2} = \alpha \\ \text{or} \quad \frac{a_1}{a_2} &= \sqrt{\alpha} \\ \therefore \quad a_1 &= \sqrt{\alpha} a_2 \end{aligned} \tag{1}$$

Now,

$$\begin{aligned} \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} &= \frac{(a_1 + a_2)^2 - (a_1 - a_2)^2}{(a_1 + a_2)^2 + (a_1 - a_2)^2} = \frac{2a_1 a_2}{a_1^2 + a_2^2} \\ &= \frac{2(a_1/a_2)}{\left(\frac{a_1}{a_2}\right)^2 + 1} \\ &= \frac{2\sqrt{\alpha}}{1+\alpha} \end{aligned}$$

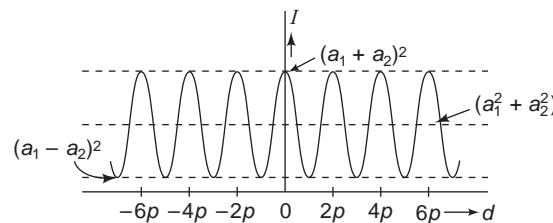
12.5 INTERFERENCE AND CONSERVATION OF ENERGY

When we draw a curve between resultant intensity (I) and the phase difference δ , it is a periodic curve with its all highest and lowest points at $(a_1 + a_2)^2$ and $(a_1 - a_2)^2$, respectively, which are the maximum and minimum intensities.

The average intensity between $\delta = 0$ and $\delta = 2\pi$ is given by

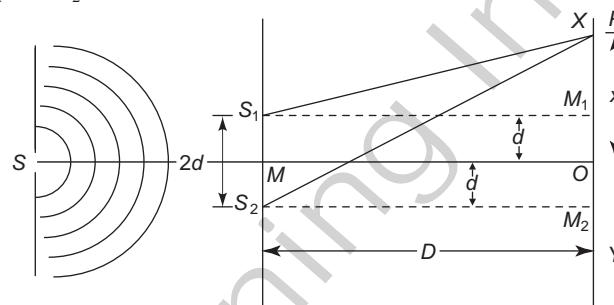
$$\begin{aligned} I_{\text{average}} &= \frac{\int_0^{2\pi} Id\delta}{\int_0^{2\pi} d\delta} = \frac{\int_0^{2\pi} (a_1^2 + a_2^2 + 2a_1 a_2 \cos \delta) d\delta}{\int_0^{2\pi} d\delta} \\ &= \frac{2\pi(a_1^2 + a_2^2)}{2\pi} = a_1^2 + a_2^2 = I_1 + I_2 \end{aligned}$$

Thus, the average intensity is equal to the sum of the individual intensities, i.e., whatever energy apparently disappears at minima is actually present at the maxima. Thus, the phenomenon of interference is consistent with the law of conservation of energy (Fig. 12.3).

**Fig. 12.3** Resultant intensity and phase difference graph

12.6 DETERMINATION OF FRINGE WIDTH IN YOUNG'S EXPERIMENT

Let S be a narrow slit illuminated by monochromatic light and S_1 and S_2 be two parallel slits very close together and equidistant from S . The light waves from S_1 and S_2 produce an interference pattern on a screen XY placed parallel to S_1 and S_2 as shown in Fig. 12.4.

**Fig. 12.4** Analytical treatment of Young's experiment

From Fig. 12.4,

$$\begin{aligned}(S_2 P)^2 &= (S_2 M_2)^2 + (P M_2)^2 \\&= D^2 + (x + d)^2 \\&= D^2 \left[1 + \frac{(x + d)^2}{D^2} \right]\end{aligned}$$

$$\therefore S_2 P = D \left[1 + \frac{(x + d)^2}{D^2} \right]^{1/2}$$

Since $D \gg (x + d)$, the binomial expansion up to two terms will give

$$\begin{aligned}S_2 P &= D \left[1 + \frac{(x + d)^2}{D^2} \right]^{1/2} \\&= D + \frac{(x + d)^2}{2D}\end{aligned}$$

Similarly, $S_1P = D + \frac{(x-d)^2}{2D}$

$\therefore S_2P - S_1P = \frac{2xd}{D}$

Now for maxima or bright fringes, the path difference is given by

$$S_2P - S_1P = n\lambda, \text{ where } n = 0, 1, 2, \dots$$

or $\frac{2xd}{D} = n\lambda$

or $x = n \frac{D\lambda}{2d}$

and for dark fringes, i.e., for minima the path difference is given by

$$S_2P - S_1P = (2n+1) \frac{\lambda}{2}, \text{ where } n = 0, 1, 2, \dots$$

$$\frac{2xd}{D} = (2n+1) \frac{\lambda}{2}$$

or $x = \frac{D}{2d} (2n+1) \frac{\lambda}{2}$

Now let x_n and x_{n+1} denote the distances of n th and $(n+1)$ th bright fringes, then

$$x_n = n \frac{D\lambda}{2d}$$

and $x_{n+1} = (n+1) \frac{D\lambda}{2d}$

\therefore Spacing between n th and $(n+1)$ th bright fringe or successive bright fringes is

$$x_{n+1} - x_n = (n+1) \frac{D\lambda}{2d} - n \frac{D\lambda}{2d} = \frac{D\lambda}{2d}$$

It is independent of n . Hence, spacing between any two consecutive bright fringes is same. Similarly, it can be shown that spacing between two consecutive dark fringes will also be $D\lambda/2d$.

The spacing between any two consecutive bright and dark fringes is called the fringe width (β). Thus,

$$\beta = \frac{D\lambda}{2d} \quad (12.7)$$

12.7 SHAPE OF INTERFERENCE FRINGES

At point P on the screen (Fig. 12.5), there will be a maximum or minimum, respectively, for intensity according as

$$S_2P - S_1P = n\lambda$$

or $S_2P - S_1P = (2n+1) \frac{\lambda}{2}$

Thus, for a given value of n , one can write

$$S_2 P - S_1 P = \text{Constant} = \Delta \text{ (say)}$$

Let the coordinates of S_1 , S_2 , and P be $(d, 0)$, $(-d, 0)$, and (x, y) , respectively. Then,

$$\begin{aligned} S_2 P &= [(x+d)^2 + y^2]^{1/2} \text{ and } S_1 P = [(x-d)^2 + y^2]^{1/2} \\ \therefore &= [(x+d)^2 + y^2]^{1/2} - [(x-d)^2 + y^2]^{1/2} = \Delta \\ (x+d)^2 + y^2 &= \left\{ [(x-d)^2 + y^2]^{1/2} + \Delta \right\}^2 \end{aligned}$$

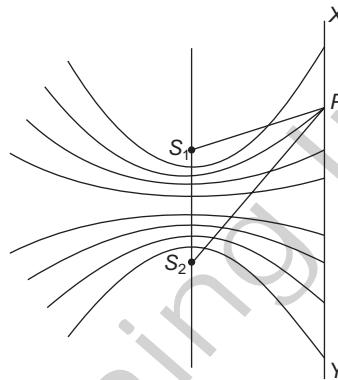


Fig. 12.5 Shape of interference fringes

$$x^2 + d^2 + 2xd + y^2 = x^2 + d^2 - 2xd + y^2 + \Delta^2 + 2[(x-d)^2 + y^2]^{1/2}\Delta$$

$$4xd = \Delta^2 + 2\Delta[(x-d)^2 + y^2]^{1/2}$$

$$2\Delta[(x-d)^2 + y^2]^{1/2} = 4xd - \Delta^2$$

Squaring on both sides will give

$$\begin{aligned} 4\Delta^2[(x-d)^2 + y^2] &= \Delta^4 + 16x^2d^2 - 8\Delta^2xd \\ \Rightarrow 4\Delta^2x^2 + 4\Delta^2d^2 - 8\Delta^2xd + 4\Delta^2y^2 &= \Delta^4 + 16x^2d^2 - 8\Delta^2xd \\ \Rightarrow 4x^2(4d^2 - \Delta^2) - 4\Delta^2y^2 &= 4\Delta^2d^2 \\ \Rightarrow x^2(4d^2 - \Delta^2) - \Delta^2y^2 &= \Delta^2d^2 \\ \Rightarrow x^2\left(\frac{4}{\Delta^2} - \frac{1}{d^2}\right) - \frac{y^2}{d^2} &= 1 \\ \Rightarrow \frac{x^2}{1/(4/\Delta^2 - 1/d^2)} - \frac{y^2}{d^2} &= 1 \end{aligned} \tag{12.8}$$

which is the equation of a hyperbola with S_1 and S_2 as foci. Thus, the shape of the fringes in Young's experiment is a hyperbola.

13

Diffraction

13.1 INTRODUCTION

According to geometrical optics when a plane wave is incident on a long narrow slit (Fig. 13.1), the region PQ of the screen XY is illuminated and the remaining portion (known as the geometrical shadow) is found absolutely dark. However, if the width of the slit is not very large compared to the wavelength, then the light intensity in the region PQ is not uniform and there is also some intensity inside the geometrical shadow.

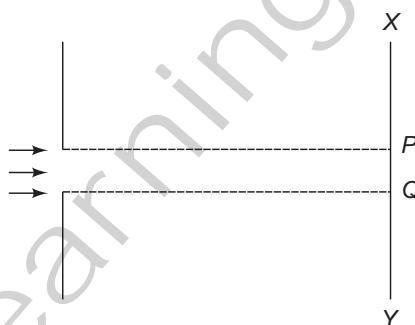


Fig. 13.1 A plane wave incident on the slit will form a sharp shadow in region PQ

Again, if the width of the slit is made smaller, a large amount of energy reaches the geometrical shadow. This spreading out of a wave on passing through a narrow opening is referred to as *diffraction* and the intensity distribution obtained on the screen is known as the *diffraction pattern*. We will show that the spreading out pattern decreases with decrease in wavelength. In fact if the orders of light wavelengths are very small, the effects due to diffraction are not readily observed.

We should also mention that there is not much of a difference between the phenomenon of interference and diffraction. In fact, interference corresponds to the situation when superposition of waves occurs due to the waves coming out from a number of point sources and diffraction corresponds to the situation when the waves from a source—circular or rectangular aperture—superimpose amongst themselves.

13.2 DIFFRACTION OF LIGHT

If an opaque obstacle or aperture is placed between a light source and a screen, a sufficiently dark shadow is obtained on the screen. This shows the rectilinear propagation of light. However, if the size of the

aperture is made comparable to the wavelength of light, the light bends round the corners of the aperture and enters the geometrical shadow. This bending of light is called *diffraction of light*. As a result of diffraction, the edges of the shadow are not sharp, but the intensity is distributed in a manner depending upon the nature of aperture.

The diffraction phenomena are usually divided into two categories: (i) Fresnel diffraction and (ii) Fraunhofer diffraction.

In the Fresnel class of diffraction, the source of light or screen on which diffraction pattern is observed, or usually both, is at finite distance from the diffraction obstacle or aperture. In this case, the incident wavefront is either spherical or cylindrical [Fig. 13.2(a)].

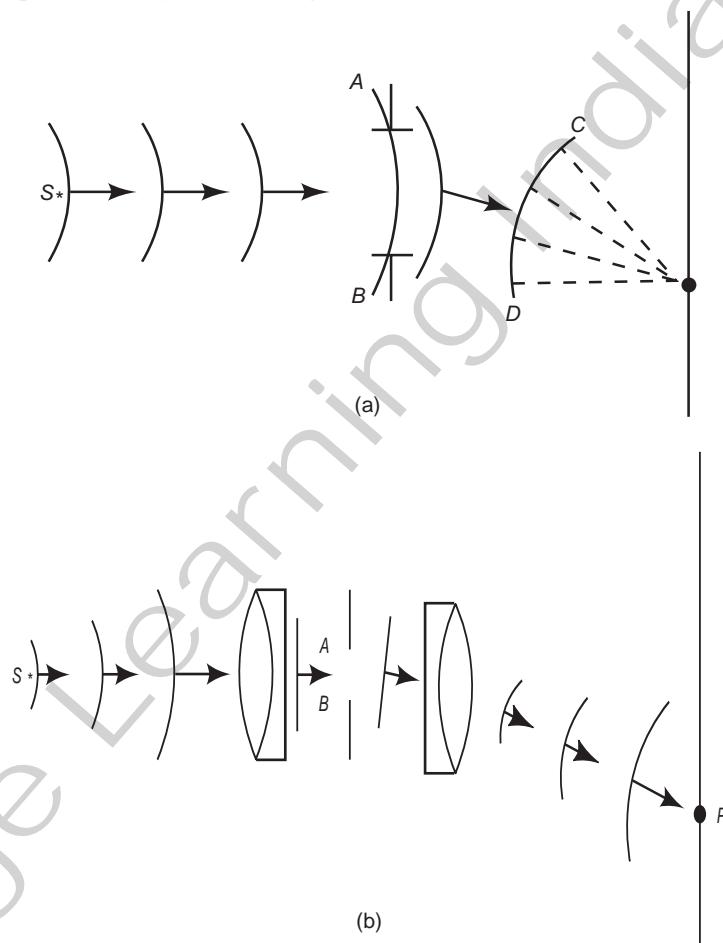


Fig. 13.2 (a) Fresnel diffraction and (b) Fraunhofer diffraction

In the Fraunhofer class of diffraction, the source of light and the screen are effectively at infinite distances from the diffracting obstacle. This is achieved by placing the source and the screen in the focal planes of the two lenses. In this case, the incident wavefront is the plane [Fig. 13.2(b)].

The above distinction creates an impression that a plane wavefront is essential for Fraunhofer diffraction. This is however not necessary in case of Fresnel's diffraction. The essential requirement for the Fraunhofer diffraction is that the pattern must be observed in the plane which is conjugate to the plane in which the source of light lies.

13.3 RESULTANT OF n SIMPLE HARMONIC WAVES

Let us consider n simple harmonic vibrations having equal amplitude a and common phase difference d between successive vibrations. To find the initial phase and amplitude of the resultant vibration, a polygon is constructed as shown in Fig. 13.3. The side OQ of the polygon represents the resultant amplitude R , and the angles δ which this side makes with the first vibration represents the phase of the resultant vibration.

Resolving the amplitude in two perpendicular directions, we get

$$R \cos \delta = a[1 + \cos d + \cos 2d + \dots + \cos (n-1)d] \quad (13.1)$$

$$R \sin \delta = a[0 + \sin d + \sin 2d + \dots + \sin (n-1)d] \quad (13.2)$$

Multiplying Eq. (13.1) by $2 \sin(d/2)$, we get

$$2R \cos \delta \sin \frac{d}{2} = a \left[2 \sin \frac{d}{2} + 2 \cos d \sin \frac{d}{2} + 2 \cos 2d \sin \frac{d}{2} + \dots + 2 \cos (n-1)d \sin \frac{d}{2} \right]$$

Using formula $2 \cos A \sin B = \sin(A+B) - \sin(A-B)$, we have

$$\begin{aligned} 2R \cos \delta \sin \frac{d}{2} &= a \left[2 \sin \frac{d}{2} + \left(\sin \frac{3d}{2} - \sin \frac{d}{2} \right) + \left(\sin \frac{5d}{2} - \sin \frac{3d}{2} \right) \right. \\ &\quad \left. + \dots + \left(\sin \left\{ n - \frac{1}{2} \right\} d - \sin \left\{ n - \frac{3}{2} \right\} d \right) \right] \\ &= a \left[\sin \frac{d}{2} + \sin \left(n - \frac{1}{2} \right) d \right] \\ &= 2a \sin \frac{nd}{2} \cos \frac{(1-n)d}{2} \end{aligned}$$

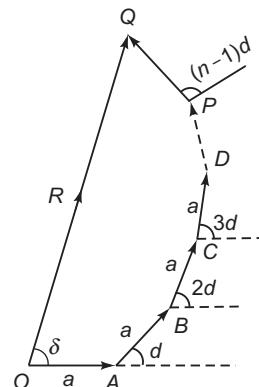


Fig. 13.3 Resultant of harmonic waves

$$\begin{aligned}
 &= 2a \sin \frac{nd}{2} \cos \frac{(n-1)d}{2} \quad \left\{ \because \cos(-\theta) = \cos \theta \right\} \\
 &\quad \left(\text{using formula } \sin C + \sin D = 2 \sin \frac{C+D}{2} \cos \frac{C-D}{2} \right) \\
 \text{or} \quad R \cos \delta &= \frac{a \sin(nd/2)}{\sin(d/2)} \cos \frac{(n-1)d}{2} \tag{13.3}
 \end{aligned}$$

Similarly, multiplying Eq. (13.2) by $2 \sin(d/2)$ and simplifying, we get

$$R \sin \delta = \frac{a \sin(nd/2)}{\sin(d/2)} \sin \frac{(n-1)d}{2} \tag{13.4}$$

Squaring and adding Eqs. (13.3) and (13.4), we get

$$\begin{aligned}
 R^2 &= \frac{a^2 \sin^2(nd/2)}{\sin^2(d/2)} \\
 \text{or} \quad R &= \frac{a \sin(nd/2)}{\sin(d/2)} \tag{13.5}
 \end{aligned}$$

Dividing Eq. (13.4) by Eq. (13.3), we get

$$\begin{aligned}
 \tan \delta &= \tan \frac{(n-1)d}{2} \\
 \text{or} \quad \delta &= \frac{(n-1)d}{2} \tag{13.6}
 \end{aligned}$$

n is the number of rays and hence infinitely large, while a and d are infinitesimally small.

$$\text{Let } \frac{nd}{2} = \alpha \text{ (finite quantity)}$$

Then from Eq. (13.5),

$$\begin{aligned}
 R &= \frac{a \sin \alpha}{\sin(\alpha/n)} \\
 &= \frac{a \sin \alpha}{\alpha/n} \quad \left(\because \frac{\alpha}{n} \text{ is very small, } \therefore \sin\left(\frac{\alpha}{n}\right) = \frac{\alpha}{n} \right) \\
 &= \frac{n a \sin \alpha}{\alpha}
 \end{aligned}$$

na is still finite. Let $na = A$, which gives

$$R = \frac{A \sin \alpha}{\alpha} \tag{13.7}$$

Also, Eq. (13.6) will give

$$\delta = \frac{(n-1)d}{2} = \frac{nd}{2} \quad (\text{as } n \text{ is very large})$$

$$\delta = \alpha \quad (13.8)$$

Equations (13.7) and (13.8) give the resultant amplitude and phase, respectively.

13.4 FRAUNHOFER DIFFRACTION AT A SINGLE SLIT

Let a parallel beam of monochromatic light of wavelength λ be incident normally upon a slit having width e (Fig. 13.4). The diffraction pattern obtained on the screen consists of a central bright band, having alternate dark and bright bands of decreasing intensity on both sides.

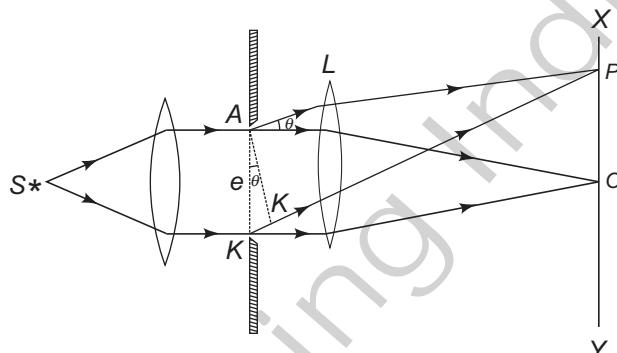


Fig. 13.4 Fraunhofer diffraction at a single slit

Explanation

According to Huygens wave theory, each point in AB sends out secondary wavelets in all directions. The rays which do not diffract are focussed at O while those diffracted through an angle θ are focussed at P . We have to find out the resultant intensity at P .

Analysis

The path difference between the wavelets from A and B in the direction θ is

$$BK = AB \sin \theta = e \sin \theta$$

The corresponding phase difference = $\frac{2\pi}{\lambda} e \sin \theta$.

Let the width of the slit AB be divided into n equal parts, and the amplitude of each wave coming out from these n parts be a . Then the phase difference between the waves from any two consecutive parts is

$$d = \frac{1}{n} \left[\frac{2\pi}{\lambda} e \sin \theta \right]$$

Hence, the resultant amplitude at P will be

$$R = \frac{a \sin(nd/2)}{\sin(d/2)} = \frac{a \sin(\pi e \sin \theta / \lambda)}{\sin(\pi e \sin \theta / n \lambda)}$$

Let $\pi e \sin \theta / \lambda = \alpha$, then

$$\begin{aligned} R &= \frac{a \sin \alpha}{\sin(\alpha/n)} = \frac{a \sin \alpha}{\alpha/n} \quad [\because \alpha/n \text{ is very small}] \\ &= \frac{n a \sin \alpha}{\alpha} \end{aligned}$$

As $n \rightarrow \infty$, $a \rightarrow 0$, but the product na remains finite.

Let $na = A$, then

$$R = \frac{A \sin \alpha}{\alpha}$$

Therefore, resultant intensity at P is

$$I = R^2 = \frac{A^2 \sin^2 \alpha}{\alpha^2} \quad (13.9)$$

13.4.1 Directions of Maxima and Minima

It is clear from Eq. (13.9) that for intensity to be maximum, the value of $(\sin \alpha / \alpha)$ should also be maximum.

Here, $\lim_{\alpha \rightarrow 0} \frac{\sin \alpha}{\alpha} = 1$ (using L' Hospital's rule)

$\therefore \alpha = 0$ corresponds to maxima

or $\frac{\pi e \sin \theta}{\lambda} = 0$

or $\sin \theta = 0$

or $\theta = 0$

i.e., in the same direction as that of the incident light. It is also clear from Eq. (13.9) that intensity is minimum when

$$\begin{aligned} &\sin \alpha = 0 \quad (\text{but } \alpha \neq 0) \\ \text{or } &\alpha = \pm m\pi \quad \left(\begin{array}{l} m \neq 0 \\ m = 1, 2, 3, \dots \end{array} \right) \\ \text{or } &\frac{\pi e \sin \theta}{\lambda} = \pm m\pi \\ &e \sin \theta = \pm m\lambda \end{aligned} \quad (13.10)$$

This equation gives the directions of the first, second, and third minima by putting $m = 1, 2, 3, \dots$

Now it is obvious that due to these minima some other maxima called secondary or subsidiary maxima will also be present in the pattern, whose intensity distribution can be analysed by putting

$$\begin{aligned} \frac{dI}{d\alpha} &= 0 \\ \text{or } &\frac{d}{d\alpha} \left[A^2 \frac{\sin^2 \alpha}{\alpha^2} \right] = 0 \end{aligned}$$

$$\begin{aligned} \text{or } & A^2 \left[\frac{\alpha^2 2 \sin \alpha \cos \alpha - 2 \alpha \sin^2 \alpha}{\alpha^4} \right] = 0 \\ \text{or } & \alpha \cos \alpha - \sin \alpha = 0 \\ \text{or } & \alpha - \tan \alpha = 0 \end{aligned} \quad (13.11)$$

This equation is solved graphically by plotting the curves $y = \alpha$ and $y = \tan \alpha$.

The first of these curves is a straight line through the origin making an angle 45° while the second is a discontinuous curve having a number of branches (Fig. 13.5). The points of intersection of the two gives the values of α satisfying Eq. (13.11).

These values are approximately given out as $\alpha = \frac{3\pi}{2}, \frac{5\pi}{2}, \frac{7\pi}{2}, \dots$

Substituting the approximate values of α in Eq. (13.11), we get the intensities of various maxima. Thus, the intensity of the central or principal maximum is

$$I_0 = A^2 \left(\frac{\sin 0}{0} \right)^2 \approx A^2$$

Similarly, the intensity of first subsidiary maximum is

$$I_1 \approx A^2 \left\{ \frac{\sin (3\pi/2)}{3\pi/2} \right\} = \frac{4A^2}{9\pi^2} \approx \frac{A^2}{22}$$

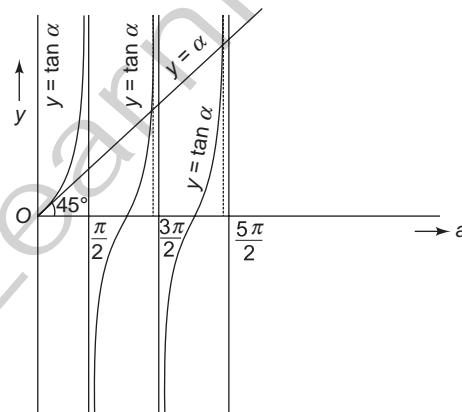


Fig. 13.5 The plot of $y = \alpha$ and $y = \tan \alpha$

and that of the second subsidiary maximum is

$$I_2 \approx A^2 \left\{ \frac{\sin (5\pi/2)}{5\pi/2} \right\} = \frac{4A^2}{25\pi^2} \approx \frac{A^2}{61}$$

and so on. Thus, the intensities of the successive maxima are in the ratio

$$1 : \frac{4}{9\pi^2} : \frac{4}{25\pi^2} : \frac{4}{49\pi^2} : \dots$$

or $1 : \frac{1}{22} : \frac{1}{61} : \frac{1}{121} \dots$

Thus, the diffraction pattern consists of a bright principal maximum in the direction of incident light having alternate minima and weak subsidiary maxima of rapidly decreasing intensity on either sides of it. The intensity pattern is shown in Fig. 13.6.

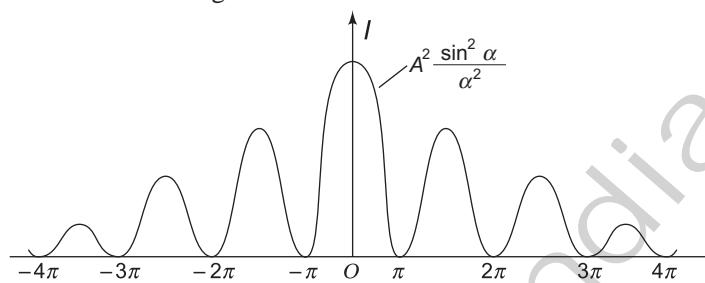


Fig. 13.6 The intensity of diffraction pattern

Solved Examples

Example 13.1

A light of wavelength 6000 \AA falls normally on a straight slit of width 0.10 mm . Calculate the total angular width of central maximum and also the linear width as observed on a screen placed 1 m away.

Solution

The condition for central maximum is

$$e \sin \theta = \lambda$$

where θ is the angular half width.

Given $\lambda = 6000 \text{ \AA} = 6 \times 10^{-5} \text{ cm}$, and $e = 0.01 \text{ cm}$.

$$\therefore \sin \theta = \frac{6 \times 10^{-5}}{0.01} \\ = 6 \times 10^{-3}$$

Since θ is very small, $\sin \theta \sim \theta$. Therefore,

$$\theta = 6 \times 10^{-3} \text{ rad}$$

The angular width of the central maximum is

$$\begin{aligned} &= 2\theta \\ &= 2 \times 6 \times 10^{-3} \\ &= 12 \times 10^{-3} \\ &= 1.2 \times 10^{-2} \text{ rad} \end{aligned}$$

Let Y be the linear half width of central maximum on the screen placed at a distance $d = 1.0 \text{ m} = 100 \text{ cm}$.

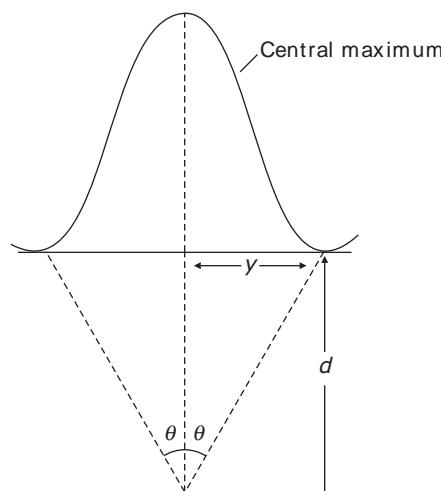


Fig. 1 The intensity

$$Y = \theta d^*$$

$$= 6 \times 10^{-3} \times 100 = 0.6 \text{ cm}$$

Therefore, the linear width of the central maximum on the screen

$$= 2Y$$

$$= 2 \times 0.6$$

$$= 1.2 \text{ cm}$$

Example 13.2

A single slit of width 0.14 mm is illuminated normally by monochromatic light and diffraction bands are observed on a screen 2 m away. If the centre of the second dark band is 1.6 cm from the middle of the central bright band, deduce the wavelength of light.

Solution

Given $e = 0.14 \text{ mm} = 0.014 \text{ cm}$, $d = 2 \text{ m} = 200 \text{ cm}$, $m = 2$, $Y = 1.6 \text{ cm}$, and $\lambda = ?$

In a diffraction pattern due to a single slit, minimum is given by

$$e \sin \theta = m\lambda$$

$$\sin \theta = \frac{m\lambda}{e} \quad \text{or} \quad \sin \theta = \frac{2 \times \lambda}{0.014}$$

Since $\sin \theta$ is very small, so $\sin \theta \approx \theta$.

$$\theta = \frac{2 \times \lambda}{0.014} \quad (1)$$

$$\begin{aligned} \text{But} \quad \theta &= \frac{Y}{d} \\ \theta &= \frac{1.6}{200} \end{aligned} \quad (2)$$

From Eqs. (1) and (2),

$$\begin{aligned} \frac{2\lambda}{0.014} &= \frac{1.6}{200} \\ \lambda &= \frac{1.6 \times 0.014}{2 \times 200} \\ &= 5600 \times 10^{-8} \text{ cm} \\ &= 5600 \text{ Å} \end{aligned}$$

Example 13.3

Light of wavelength 5000 Å is incident normally on a single slit. The central maximum spreads out at 30° on both sides of the direction of incident light. Calculate the width of the slit.

* $\tan \theta = \theta = \frac{Y}{d}$ (θ is very small)

Solution

The direction of minima in Fraunhofer diffraction due to single slit is given by

$$\alpha = \pm m\pi$$

or $\frac{\pi}{\lambda} e \sin \theta = \pm m\pi \quad \text{where } m = 1, 2, 3, \dots \quad (1)$

Therefore, the angular spread of the central maximum on either side of the incident light is

$$\sin \theta = \frac{\lambda}{e} \quad (\text{for } m = 1, \text{ the position of first minimum}) \quad (2)$$

Given $\theta = 30^\circ$ and $\lambda = 5000 \text{ \AA} = 5 \times 10^{-5} \text{ cm}$.

$$\sin \theta = \frac{\lambda}{e}$$

On putting the given values, we have

$$\begin{aligned} \sin 30^\circ &= \frac{5 \times 10^{-5}}{e} \\ \Rightarrow e &= \frac{5 \times 10^{-5}}{\sin 30^\circ} = \frac{5 \times 10^{-5}}{(1/2)} \\ &= 2 \times 5 \times 10^{-5} \\ &= 10 \times 10^{-5} \\ &= 10^{-4} \text{ cm} \end{aligned}$$

Therefore, the width of the slit is 10^{-4} cm .

Example 13.4

In Fraunhofer diffraction due to a single slit, the screen is placed 2 m away from the slit. If the slit width is 0.2 mm and the first minimum lies 5 mm on either side of the central maximum, find the wavelength of incident light.

Solution

Given $e = 0.2 \text{ mm} = 0.02 \text{ cm}$, $Y = 5 \text{ mm} = 0.5 \text{ cm}$, $d = 2 \text{ m} = 200 \text{ cm}$, and $\lambda = ?$

The direction of minima is given by

$$e \sin \theta = \pm m\lambda \quad (\text{where } m = 1, 2, 3, 4)$$

For first minimum, $m = 1$, therefore

$$e \sin \theta = \lambda$$

If θ be small then $\sin \theta \approx \theta$. Hence,

$$\begin{aligned} e \theta &= \lambda \\ \Rightarrow \theta &= \frac{\lambda}{e} \text{ radian} \\ &= \frac{\lambda}{0.02} \text{ radian} \quad (1) \end{aligned}$$

14

Polarisation

14.1 INTRODUCTION

Optical phenomena such as interference and diffraction could be explained on the basis of Huygens' wave theory. However, these phenomena could not explain the nature of light wave: longitudinal or transverse.

If we move one end of a string up and down, then a transverse wave is generated [Fig. 14.1(a)]. Each point of the string executes a sinusoidal oscillation in a straight line and the wave is, therefore, known as a *linearly polarised wave*. It is also known as a plane polarised wave because the string is always confined to the $x-z$ plane.

In general, the string can be made to vibrate in any plane containing the z -axis. If one rotates the end of the string on the circumference of a circle, then each point of the string will move in a circular path [Fig. 14.1(b)]; such a wave is known as a *circularly polarised wave*.

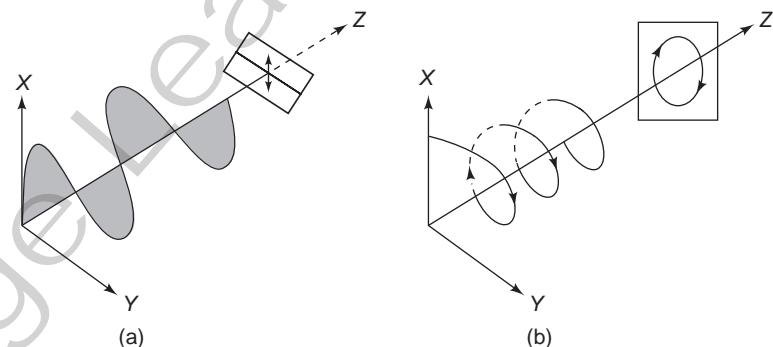


Fig. 14.1 (a) A linearly polarised wave on a string with the displacement confined in $x-z$ plane and **(b)** the displacement corresponding to a circularly polarised wave

Now consider a long narrow slit placed in the path of the string as shown in Fig. 14.2(a). If the length of the slit is along the direction of the displacement, then the entire amplitude will be transmitted as shown in Fig. 14.2(a). On the other hand, if the slit is at right angles to the direction of the displacement, then almost nothing will be transmitted to the other side of the slit [Fig. 14.2 (b)]. This is because of the

fact that the slit allows only the component of the displacement which is along the length of the slit to pass through. We should mention that if a longitudinal wave was propagating through the string then the amplitude of the transmitted wave would have been the same for all orientations of the slit. Thus, the change in the amplitude of the transmitted wave with the orientation of the slit is due to the *transverse character* of the wave.

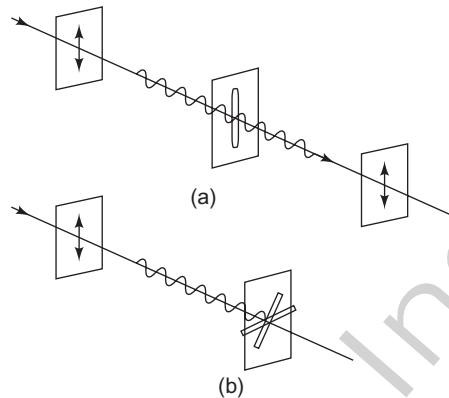


Fig. 14.2 Slit allows only that component of displacement to pass through which is along the length of the slit

In fact, an experiment very similar to the experiment discussed above proves the transverse character of light waves. For this we must define an unpolarised wave. If the plane of vibration of a transverse wave is changed in a random manner in very short intervals of time, then such a wave is known as *unpolarised wave*.

If an unpolarised wave falls on slit S_1 (Fig. 14.3), then the displacement associated with transmitted wave will be along the length of the slit and a rotation of the slit will not affect the amplitude of the transmitted wave although the plane of polarisation of the transmitted wave depends on the orientation of the slit. Thus, the transmitted wave will be linearly polarised and the slit S_1 is said to act as a polariser. If the polarised beam falls on another slit S_2 (Fig. 14.3), then by rotating the slit S_2 we obtain a variation of the transmitted amplitude, the second slit is said to act as an analyser.

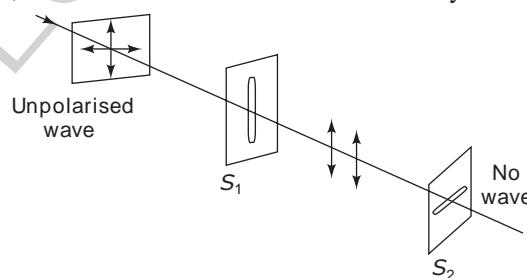


Fig. 14.3 Effect on the intensity of light due to rotation of slit S_2

14.2 POLARISATION OF LIGHT

When ordinary (unpolarised) light is incident normally on a pair of parallel tourmaline crystal plates P_1 and P_2 (Fig. 14.4) cut parallel to their crystallographic axes, the emergent light shows a variation in the

intensity as P_2 is rotated. The intensity is found to be maximum when the axis of P_2 is parallel to that of P_1 [Fig. 14.4(a)] and minimum when at right angles [Fig. 14.4(b)]. This shows that the nature of light emerging from P_1 is not symmetrical about the direction of propagation of light, *but its vibrations are confined only to a single line in a plane perpendicular to the direction of propagation*. Such light is called *plane polarised* or *linearly polarised light*.

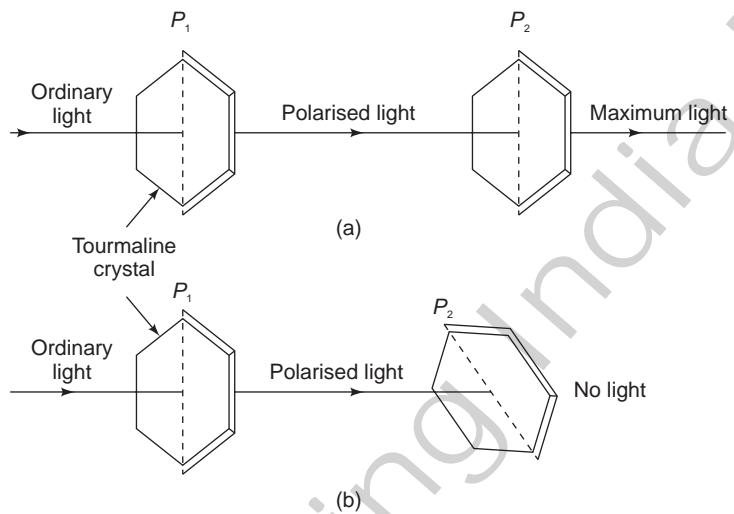


Fig. 14.4 Polarisation of light

According to the electromagnetic theory of light, a light wave consists of electric and magnetic vector vibrating in mutually perpendicular planes, both being perpendicular to the direction of propagation of light. The electric field vector acts as the light vector. Hence, the plane polarised light is the light in which the light vector vibrates along a fixed straight line in a plane perpendicular to the direction of propagation.

14.3 REPRESENTATION

In an unpolarised beam of light all directions of vibrations at right angles to that of the propagation of light are possible; hence, it is represented by a star [Fig. 14.5(a)].

In a plane polarised beam of light, the vibrations are along a straight line. If the vibrations are parallel to the plane of the paper, they are represented by arrows [Fig. 14.5(b)]. If they are along a straight line perpendicular to the plane of the paper, they are represented by dots [Fig. 14.5(c)].

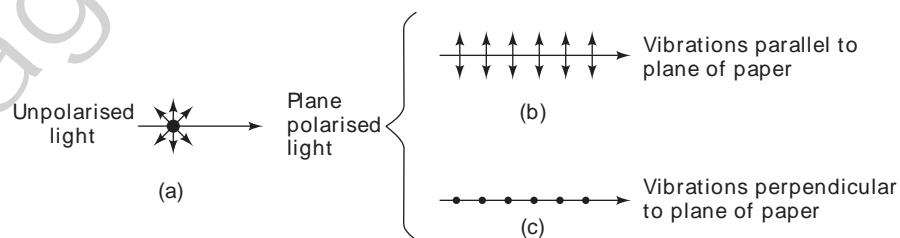


Fig. 14.5 (a) Unpolarised light and (b, c) linearly polarised light

14.4 PLANE OF VIBRATION

The plane containing the direction of vibration and the direction of propagation of light is called the plane of vibration.

14.5 PLANE OF POLARISATION

The plane passing through the direction of propagation and containing no vibration is called plane of polarisation.

14.6 PRODUCTION OF PLANE POLARISED LIGHT

The plane polarised light can be produced by the following methods:

- Polarisation by reflection
- Polarisation by refraction
- Polarisation by double refraction
- Polarisation by scattering

14.6.1 Polarisation by Reflection: Brewster's Law

Malus in 1808 discovered that when unpolarised light is incident on the surface of any transparent material, the reflected and refracted beams are partially plane polarised. The degree of polarisation depends on the angle of incidence, and at a particular angle called the *polarising angle*, it is found to be maximum. This also depends on the nature of the reflecting surface and the wavelength of light.

Brewster discovered that there is a simple relation between the polarising angle (p) and refractive index (μ) of the material relative to the medium. This is called Brewster's law and is given by

$$\mu = \tan p$$

A direct deduction that can be made with the help of this law is that when light is incident at the polarising angle, the reflected beam is at right angles to the refracted beam.

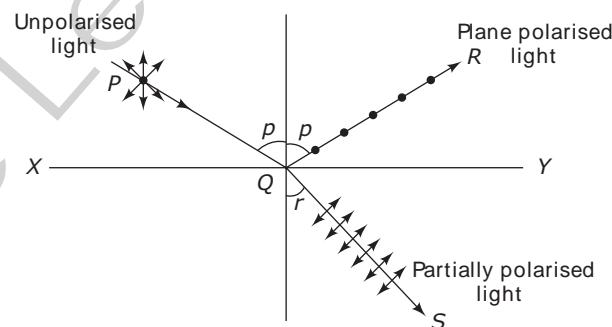


Fig. 14.6 Brewster's law

Let a beam PQ (Fig. 14.6) of unpolarised light be incident on the surface XY of a transparent material at the polarising angle p . Then, from Brewster's law,

$$\mu = \tan p = \frac{\sin p}{\cos p} \quad (14.1)$$

By Snell's law,

$$\mu = \frac{\sin p}{\sin r} \quad (14.2)$$

Comparing Eqs. (14.1) and (14.2), we get

$$\begin{aligned}\frac{\sin p}{\cos p} &= \frac{\sin p}{\sin r} \\ \cos p &= \sin r \\ \sin(90 - p) &= \sin r \\ 90^\circ - p &= r \\ r + p &= 90^\circ\end{aligned}\quad (14.3)$$

This is the case when reflected and refracted beams are at right angles to each other.

14.6.2 Polarisation by Refraction: (Pile of Plates: Malus Law)

When an unpolarised light is incident on the upper surface of a glass slab at the polarising angle p (Fig. 14.7) a small fraction is reflected, which is completely plane polarised with vibrations perpendicular to the plane of incidence, while the rest is refracted and is only partially polarised having vibrations both in the plane of incidence and perpendicular to the plane of incidence. The refracted light is incident at the lower face at an angle r , where r is the angle of refraction at the upper face.

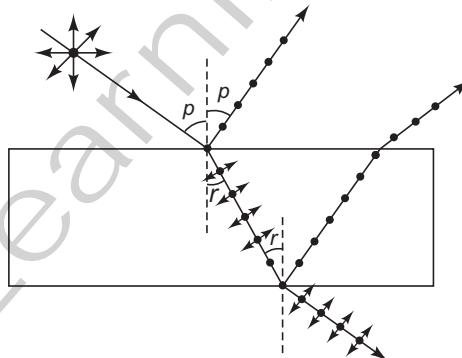


Fig. 14.7 Partial and plane polarised light

Now,

$$\begin{aligned}\tan r &= \frac{\sin r}{\cos r} = \frac{\sin r}{\sin(90^\circ - r)} \\ &= \frac{\sin r}{\sin p} = \mu_a \quad (\because p + r = 90^\circ)\end{aligned}$$

Thus, by Brewster's law, r is the polarising angle for the reflection at the lower surface of the plate. Hence, the light reflected at the lower face is completely plane polarised, while that refracted into the air is partially plane polarised.

Hence, if a beam of unpolarised light is incident at the polarising angle on a pile of plates as shown in Fig. 14.8, the sequential refracted beams become poorer and poorer in the perpendicular component. If a good number of plates is used, the refracted beam will contain mostly the parallel component, i.e., it will almost be plane polarised. The greater the number of the plates, the more pure is the refracted plane polarised light.

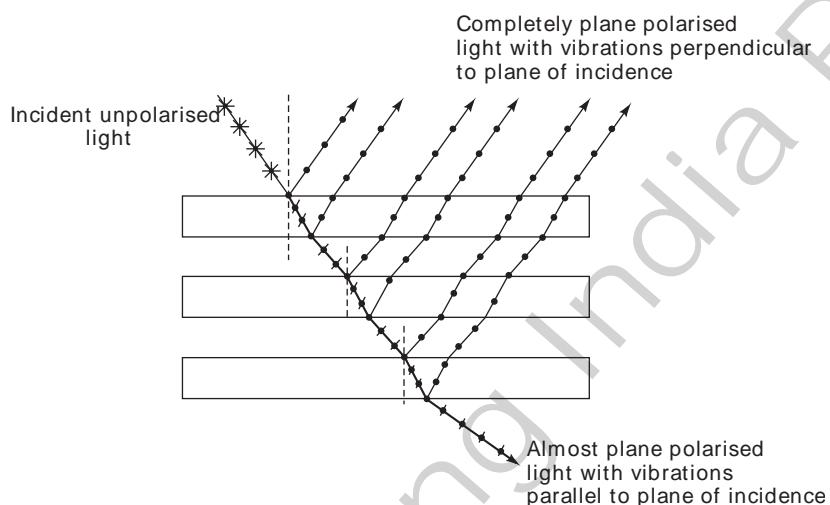


Fig. 14.8 Sequential refraction

Malus Law

When a completely plane polarised light beam is incident on an analyser, the intensity of the emergent light varies as the square of the cosine of the angle between the planes of transmission of the analyser and the polariser. This is known as *Malus law* and holds for a combination of reflecting planes, Nicol prisms, and polaroids, but not for the pile of plates, because in this case, the light emerging from the polariser is not completely plane polarised.

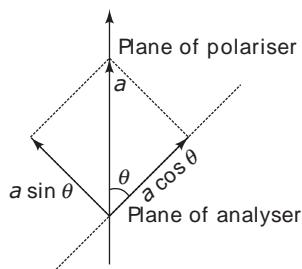


Fig. 14.9 Variation of intensity according to θ between analyser and polariser

Proof

Let a be the amplitude of the plane polarised light transmitted by the polariser and θ be the angle between the plane of transmission of the polariser and that of the analyser. As shown in Fig. 14.9, only the component $a \cos \theta$ is transmitted by the analyser. Therefore, the intensity of light emerging from the analyser is given by

$$I_\theta = a^2 \cos^2 \theta = I \cos^2 \theta \quad (14.4)$$

where $I = a^2$, the intensity of the incident polarised light. I_θ is obviously one-half the intensity of the unpolarised light striking the polariser, provided no losses due to absorption are taken into account and $\theta = 45^\circ$.

If the plane of polariser and analyser are parallel, i.e., $\theta = 0^\circ$ or 180° , then

$$I_\theta = I$$

and if they are perpendicular, i.e., $\theta = 90^\circ$, then

$$I_\theta = 0$$

14.6.3 Polarisation by Double Refraction

It is found that both the ordinary and extraordinary rays are plane polarised having vibrations perpendicular to each other. The polarisation of light by double refraction in calcite can be observed as follows:

- (i) When the principal sections of the two crystals are parallel, two images O_1 and E_1 are seen [Fig. 14.10(a)]. The O -ray from the first crystal passes undeviated through the second crystal and emerges as O_1 -ray. The E -ray from the first crystal passes through the second crystal and emerges as E_1 -ray. Hence, the images O_1 and E_1 are seen separated by a distance equal to the sum of the two displacements found in each crystal, if used separately.

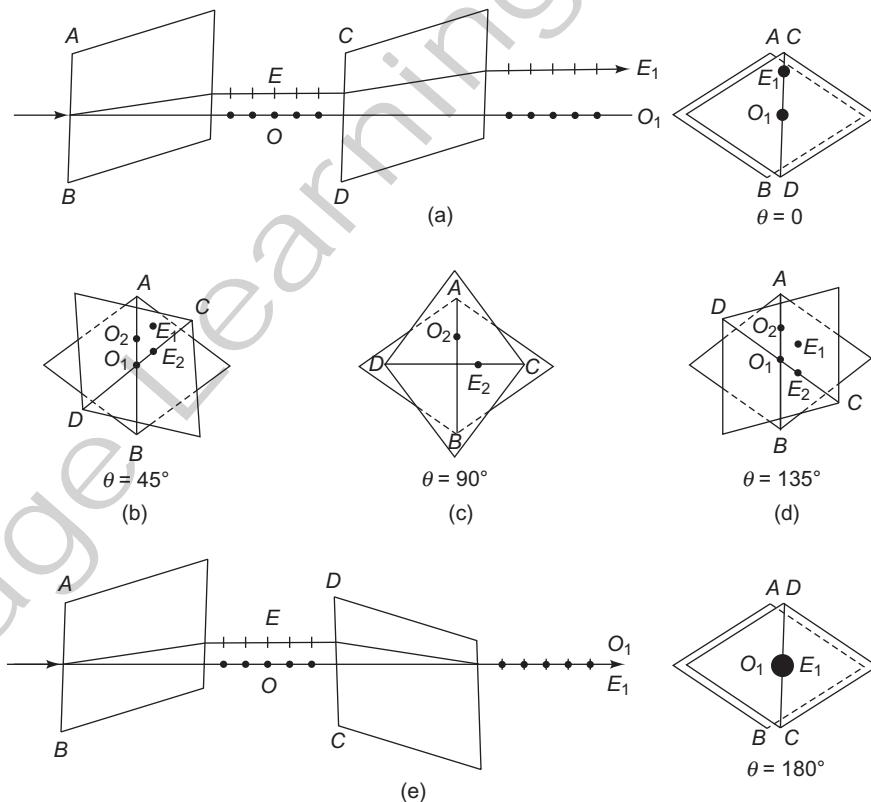


Fig. 14.10 Polarisation by double refraction

- (ii) When the second crystal is rotated, each of the rays O and E from the first crystal suffers double refraction in the second, giving rise to four images. Thus, besides O_1 and E_1 , two new images O_2 and E_2 appear. As the rotation is continued, O_1 and O_2 remain fixed while E_1 and E_2 rotate around O_1 and O_2 , respectively. Also the intensity of O_1 and E_1 goes on decreasing while that of O_2 and E_2 goes on increasing. When the principal section of the second crystal makes an angle of 45° with that of the first, all four images are of equal intensity [Fig. 14.10(b)].
- (iii) At the 90° position, original images O_1 and E_1 have faded to vanish and the new images O_2 and E_2 acquired maximum intensity [Fig. 14.10(c)].
- (iv) On further rotation, O_1 and E_1 reappear and start gaining intensity, while O_2 and E_2 start losing intensity. At $180^\circ - 45^\circ = 135^\circ$, the four images are once more equally intense [Fig. 14.10(d)].
- (v) At 180° , the principal sections of the two crystals are once more parallel but having optic axis oriented in opposite directions. The images O_2 and E_2 have vanished, while O_1 and E_1 have come together in the centre of the crystals and are of equal thickness [Fig. 14.10(e)].

14.6.4 Polarisation by Scattering

When a beam of white light passes through a medium of small particles of about the same size as the wavelength of light, then the light seen in the direction right angle to the incident beam appears bluish. This phenomenon is known as *scattering of light*. When the scattered blue light is seen through a rotating Nicol prism, a variation in intensity with minimum intensity zero is found. This shows that the light scattered in this direction is plane polarised.

14.7 DOUBLY REFRACTING CRYSTALS

There are certain crystals which split a ray of light incident upon them into two refracted rays. Such crystals are called *doubly refracting crystals*. These are of two types—uniaxial and biaxial. In uniaxial crystals, there is one direction called the optic axis along which the two refracted rays travel with the same velocity. The examples of such crystals are calcite, tourmaline, and quartz. In biaxial crystals, there are two optic axes. Topaz and aragonite are example of such crystals.

14.8 CALCITE CRYSTAL

The calcite crystal, also known as Iceland spar (CaCO_3), is a colourless crystal, transparent to visible as well as ultraviolet light. As shown in Fig. 14.11, each face of the crystal is a parallelogram having angles 102° and 78° . At the two diagonally opposite corners A and A' , three obtuse angles meet. These are called *blunt corners* of the crystal. At the rest of the six corners, one angle is obtuse and two are acute.

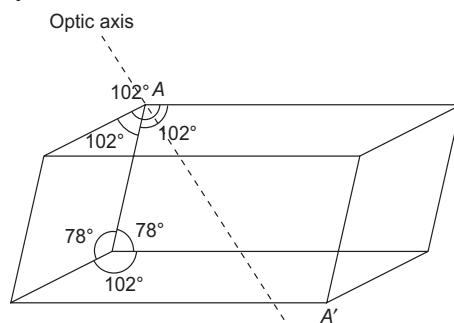


Fig. 14.11 Calcite crystal

14.8.1 Optic Axis of the Crystal

A line passing through any one of the blunt corners and making equal angles with the three faces which meet there is the direction of the optic axis of the crystal. Optic axis is a direction and not a line. Hence, any line parallel to the optic axis may be considered as optic axis.

14.8.2 Principal Section of the Crystal

A plane containing the optic axis and perpendicular to a pair of opposite faces of the crystal is called the principal section of the crystal for that pair of faces. Thus, there may be several principal sections passing through any point inside the crystal, one corresponding to each pair of opposite faces.

A principal section always cuts surfaces of the calcite crystal in a parallelogram having angles 71° and 109° as shown in Fig. 14.12(a). In Fig. 14.12(b), a face of the crystal is shown in which dotted line AB represents the end view of the principal section shown in Fig. 14.12(a).

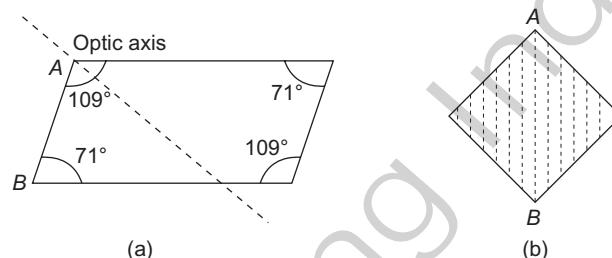


Fig. 14.12 (a) Principal section of calcite crystal and **(b)** face of the crystal

14.9 POLARISATION BY DOUBLE REFRACTION

When a ray of unpolarised light is incident on a calcite or quartz crystal, it splits up into two refracted rays. This phenomenon is called *double refraction*.

One of these two refracted rays is found to obey the laws of refraction, i.e., it always lies in the plane of incidence and its velocity in the crystal is the same in all directions. This ray is called the ordinary ray (*O-ray*). The other refracted ray does not obey the laws of refraction. It travels in the crystal with different speeds in different directions. Hence, it is called the extraordinary ray (*E-ray*). However, along the optic axis, both the rays have the same velocity and hence same refractive index.

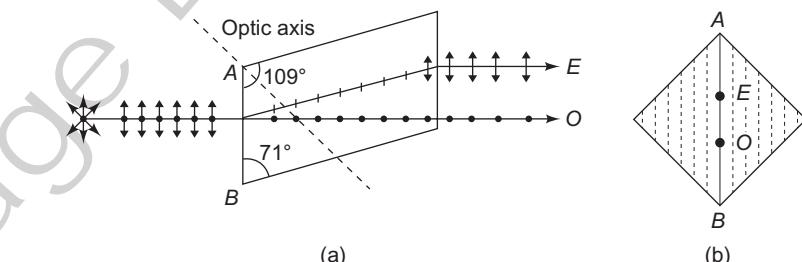


Fig. 14.13 (a) Double refraction of unpolarised light and **(b)** dots showing the images corresponding to *E*-and *O*-rays

Fig. 14.13(a) shows a ray of light incident normally on a crystal, which splits up into two rays *O* and *E*. The *O*-ray passes through the crystal undeviated, while the *E*-ray is refracted at some angle.

As the opposite faces of the crystal are parallel, the rays emerge parallel to the incident ray, but relatively displaced by a distance proportional to the thickness of the crystal.

As shown in Fig. 14.13(b), two images O and E corresponding to O - and E -rays are observed. The line joining them is in a principal section of the crystal and lies either along the shorter diagonal AB or along a line parallel to AB .

14.10 NICOL PRISM

A Nicol prism is an optical device made from a calcite crystal for producing and analysing plane polarised light with the help of the phenomenon of total internal reflection.

- Construction:** As shown in Fig. 14.14, to construct a Nicol prism, a calcite crystal $ABCD$ about three times as long as it is wide is taken. Its end faces AB and CD are ground such that the angles in principal section becomes 68° and 112° , instead of 71° and 109° . The crystal is then cut apart along the plane $A'D$ perpendicular to the principal section as well as to the end faces $A'B$ and CD . They are then cemented together by Canada balsam which is a transparent glue of refractive index 1.55 for sodium light. The crystal is then put inside a blackened tube.
- Working:** When a ray SM of unpolarised light nearly parallel to BD' is incident on the face $A'B$, it splits up into two refracted rays, O and E rays, both being plane polarised. The O -ray has vibrations perpendicular to the principal section of the crystal, while E -ray has vibration in the principal section.

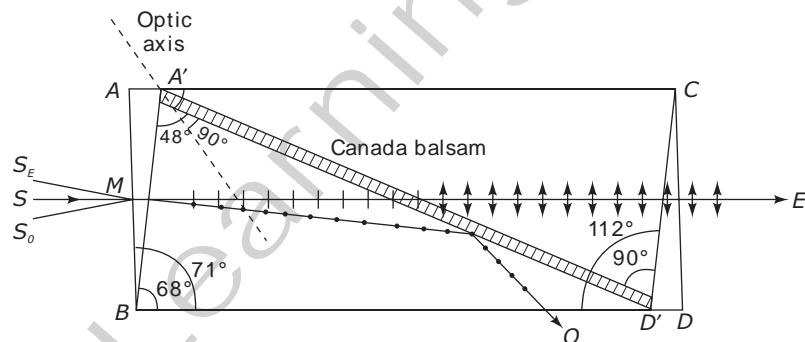


Fig. 14.14 Construction of Nicol prism

Now the refraction indices of Canada balsam, O -ray, and E -ray for calcite crystal are in order as $\mu_{CB} = 1.55$, $\mu_0 = 1.658$, and $\mu_E = 1.486$. Therefore, when the O -ray reaches the layer of Canada balsam, the total internal reflection at calcite–balsam surface takes place and this ray is absorbed by the tube containing the crystal. However, E -ray on reaching at calcite–balsam layer is transmitted because it passes from a rarer medium to a denser medium. Since the E -ray is plane polarised, the light emerging from the Nicol prism is plane polarised with vibrations parallel to the principal section.

14.10.1 Nicol Prism As a Polariser and As an Analyser

When an unpolarised ray of light is incident on a Nicol prism P , the ray emerging from P is plane polarised with vibrations in the principal section of P [Fig. 14.15(a)]. When this ray is allowed to fall on a second prism A kept in such a way that its principal section is parallel to P so that its vibrations lie in

15

Lasers and Holography

15.1 INTRODUCTION

The word *laser* is an acronym for “light amplification by stimulated emission of radiation”. Light has always fascinated human mind since early evolution. With the growth of civilisation, humans might have started thinking to create and control fire for their advantage, such as light in the night, safety from predators, and also for cooking those food items, hence, otherwise not suitable for raw eating. Since the early age of civilisation, light became an integral part of man’s daily life. But always using fire for light put some serious restriction on its use. Hence, the man started to search for some different mechanism of producing light. By knowing the specific properties of light such as amplitude, wavelength, frequency, and phases, it became easy to use it for specific purposes.

There are a variety of light sources emitting light of different types and quality, which can be broadly classified into general-purpose light and special-purpose light. The sources of ordinary light are fire and other combustion light sources such as lanterns and electric bulbs. These light sources are used for most common purposes such as seeing, reading, and for school-level science experiments. Special light is produced by light sources such as gas discharge lamp, sodium lamp, and mercury lamp. The light produced by these specific techniques has high-quality spectral light with monochromatic and polychromatic nature. Such light is used for specific purposes and researches. Laser light is also a special light produced by a specific technique. It is a source of light but it is different from many traditional sources; generally, it is not used for the purpose of illumination. Laser produces a highly directional and high-intensity light with a narrow frequency range. It is frequently used for special type of drill bit to drill holes in hard materials, as a cutter for thick metal sheets, as a phonograph needle for compact discs, as a knife during surgical operations, as a target designator for military weapons, etc.

In this chapter, we will describe spontaneous stimulated emission, and essentials of the laser action. We will also extend the discussion on three-level laser systems, four-level laser systems, ruby laser, He–Ne laser, CO₂ laser, and semiconductor laser with the basics of holography.

15.2 ATOMIC EXCITATION AND ENERGY STATES

According to the fundamentals of atomic structure, we know that in an atom, an electron in a ground state is stable and moves continuously in its orbit without radiating energy. When the electron receives an amount of energy equal to the difference of the energy of the ground state and one of the excited states,

it absorbs energy and jumps to the excited state. The electron can stay for a very short lifetime (10^{-8} s) in the excited state, but sometime may stay in the metastable state having relatively longer lifetime of 10^{-3} s. It is also important to note that an excited atom, while de-exciting, is not required to return to the same state from where it got excited. Let us consider in an atom, an electron absorbs an energy equal to the difference of the ground state (E_1) and the first excited state (E_2), i.e., $\Delta E = (E_2 - E_1)$. When this excited electron comes to the ground state by the process of de-excitation, then it releases a radiation of energy $\Delta E = E_2 - E_1$. The process of excitation and de-excitation of atom is shown in Fig. 15.1.

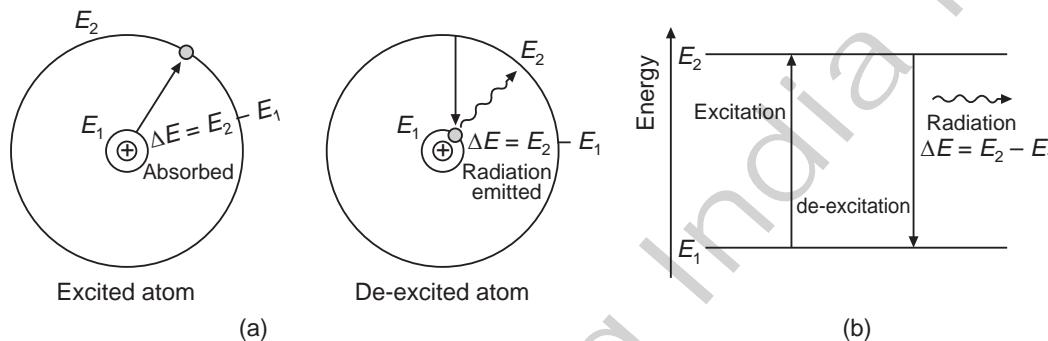


Fig. 15.1 (a) Presentation of excited and de-excited atom with the basic structure of atom and (b) demonstration of excitation and de-excitation with energy diagram

15.3 INTERACTION OF EXTERNAL ENERGY WITH THE ATOMIC ENERGY STATES

There are three kinds of interactions of the external energy with the atomic energy states. First is known as *absorption*, in which suitable amount of energy is absorbed by the atoms of the ground state to get excited to the higher energy states. Second is known as *spontaneous emission*, in which the excited atoms emit photon to come back in the lower energy state without any external impetus. Third is known as *stimulated emission*, in which atom in the excited state need not wait for the spontaneous emission to occur, but with the influence of suitable energy impetus, excited atom is triggered to the lower energy state, with the release of appropriate energy. These three important processes are discussed under the following headings.

15.3.1 Absorption

In order to describe the process of absorption, let us consider a sample in which there are two energy states— E_1 and E_2 , where E_1 is the lower energy state and E_2 is the higher energy state. The atom residing in lower energy level E_1 may absorb the incident photon to jump to the excited energy level E_2 as shown in Fig. 15.2. This process of transition is known as *induced or stimulated absorption*, or simply *absorption*. During each absorption transition, one photon disappears from the incident light. The energy of this photon is absorbed by an atom, whose position is raised to the excited energy level E_2 . If A is a particular atom which is absorbing the energy of a photon ($h\nu$) to become a member of the excited energy level E_2 , then this process can be represented as

$$A + h\nu \rightarrow A^* \quad (15.1)$$

where A^* is the excited atom.

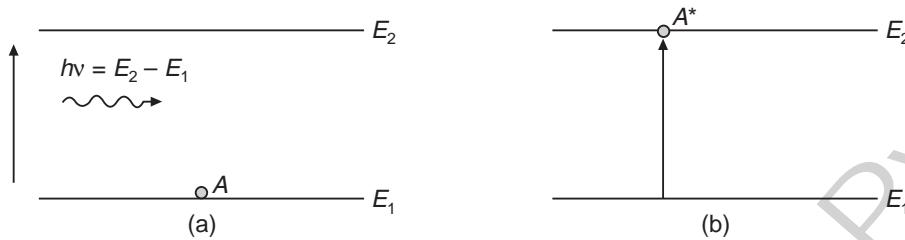


Fig. 15.2 Absorption process: (a) before absorption and (b) after absorption

The *rate of absorption transition* is defined as the number of atoms per unit volume per second which are raised from the lower energy level to the excited energy level. It is expressed as

$$R_{ab} = -\frac{dN_1}{dt}$$

where, $-dN_1/dt$ is the rate of decrease of population at the lower energy level E_1 . If dN_2/dt is the rate of increase of population at the higher energy level (E_2), then the absorption transition rate can also be expressed as

$$R_{ab} = \frac{dN_2}{dt}$$

From the above expression, we may conclude that

$$R_{ab} = -\frac{dN_1}{dt} = \frac{dN_2}{dt} \quad (15.2)$$

According to the above discussions, it is clear that the rate of absorption transition is proportional to the population of atoms in the lower energy state and the energy density of incidenting light [$(\rho(v))$]. Hence, the rate of absorption transition can be given as

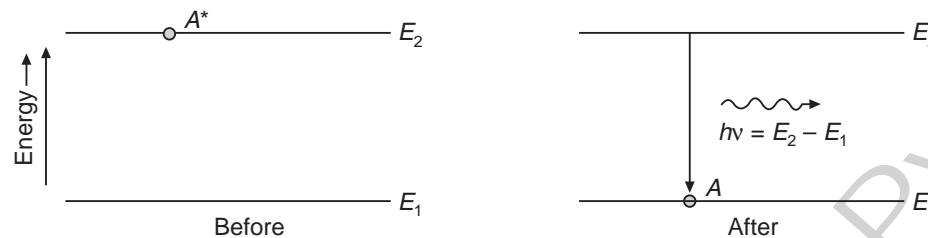
$$R_{ab} = B_{12} \rho(v) N_1 \quad (15.3)$$

where B_{12} is known as the *Einstein coefficient for induced absorption transition*, which represents the probability of induced transition from level $1 \rightarrow 2$.

15.3.2 Spontaneous Emission

The excited state of atom with higher energy is highly unstable. Without the influence of any external agent, the excited atom can stay in its state for only short average lifetime. If such atoms are not stimulated by external impetus, then the excited atom undergoes a transition to the lower energy level on its own. During the process of transition, the excess energy of excited atom comes out in the form of photon as shown in Fig. 15.3. Thus, the process in which an excited atom emits a photon itself without any external impetus is known as *spontaneous emission*. If A^* is the excited atom which is releasing the energy $h\nu$ to approach the lower energy state E_1 , then the process can be represented as



**Fig. 15.3** Spontaneous emission process

The rate of spontaneous transitions R_{sp} is given as

$$R_{sp} = \frac{-dN_2}{dt} = \frac{N_2}{T_{sp}} \quad (15.4)$$

where T_{sp} is the spontaneous transition lifetime.

Number of photons generated during the process of spontaneous transition will be proportional to the number of excited atoms only. It may be expressed as

$$R_{sp} = A_{21} N_2 \quad (15.5)$$

where A_{21} is known as *Einstein coefficient for spontaneous emission*. It gives the probability of spontaneous transition, from level $2 \rightarrow 1$.

Comparing Eqs. (15.4) and (15.5), we get

$$A_{21} = \frac{1}{T_{sp}} \quad (15.6)$$

Thus, the reciprocal of the coefficient A_{21} represents the lifetime of the spontaneous emission.

Some important features of spontaneous emission are as follows:

- (i) It is very difficult to control the process of spontaneous emission from outside. In other words, we can say that it is almost impossible.
- (ii) It is essentially probabilistic in nature.
- (iii) The emitted photons (radiation) during the process of spontaneous emission have different direction of propagation, initial phase, and plane of polarisation.
- (iv) The light emitted in the spontaneous emission is not monochromatic.
- (v) The emitted light has no preferred direction, thus, the light spreads in all the directions around the source. The intensity of light decreases rapidly with distance from the source.
- (vi) The emitted light during the process of spontaneous emission is incoherent.

15.3.3 Stimulated Emission

An excited atom need not to wait for spontaneous emission if a photon with suitable energy ($h\nu = E_2 - E_1$) interacts with it. This photon can trigger the excited atom for the transition from higher energy level (E_2) to lower energy level (E_1) by emitting another photon as shown in Fig. 15.4.

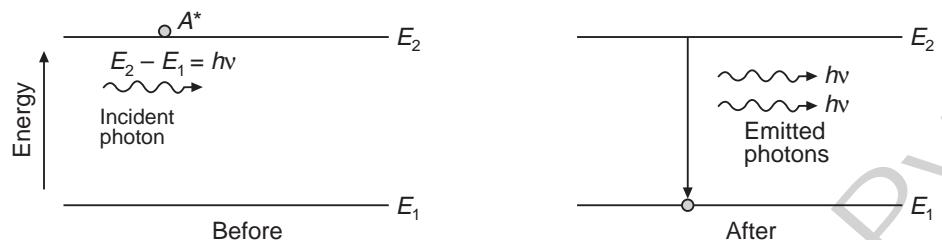


Fig. 15.4 Stimulated emission process

The process of emission of radiation from the excited atom under the influence of external impetus is known as *induced*, or *stimulated emission*. This process can be expressed as



The rate of stimulated emission of radiation can be given as

$$R_{st} = B_{21} \rho(v) N_2 \quad (15.7)$$

where \$B_{21}\$ is known as *Einstein coefficient for stimulated emission*, which indicates the stimulated emission from level \$2 \rightarrow 1\$.

Some important features of stimulated emission are as follows:

- (i) The process of stimulated emission can be controlled from outside.
- (ii) In the stimulated emission, the incident photon and emitted photon propagate in the same direction.
- (iii) The emitted photon and incident photon have same phase, frequency, and plane of polarisation.
- (iv) The light produced during stimulated emission is coherent and monochromatic.

15.4 EINSTEIN COEFFICIENTS

Einstein's \$A\$ and \$B\$ coefficients give the idea about the relation between spontaneous and stimulated emission probabilities. In order to obtain the relation between Einstein's \$A\$ and \$B\$ coefficients, let us consider that under thermal equilibrium, \$N_1\$ and \$N_2\$ are the mean population of atoms in the lower energy level \$E_1\$ and the upper energy level \$E_2\$, respectively. For the condition of equilibrium, the number of transitions from \$E_2\$ to \$E_1\$ must be equal to the transitions from \$E_1\$ and \$E_2\$. It means that the number of photons absorbed by atoms per unit time per unit volume must be equal to the number of photons emitted by atoms per unit time per unit volume.

If \$\rho(v)\$ is the photon density, then the number of atoms absorbing photons per unit time per unit volume is given as

$$B_{12} \rho(v) N_1$$

The number of atoms emitting photons per unit volume per unit time can be given as the sum of spontaneous emission and stimulated emission as

$$A_{21} N_2 + B_{21} \rho(v) N_2$$

At the condition of equilibrium, transitions from \$E_1\$ to \$E_2\$ must be equal to the transitions from \$E_2\$ to \$E_1\$. Thus, we can write

$$B_{12} \rho(v) N_1 = A_{21} N_2 + B_{21} \rho(v) N_2$$

or $\rho(v) [B_{12} N_1 - B_{21} N_2] = A_{21} N_2$

or $\rho(v) = \frac{A_{21} N_2}{(B_{12} N_1 - B_{21} N_2)}$

or $\rho(v) = \frac{A_{21} / B_{12}}{N_1 / N_2 - B_{21} / B_{12}}$ Dividing numerator and denominator of RHS by $N_2 B_{12}$ (15.8)

According to the Boltzmann equation, the population of atoms in the energy states E_1 and E_2 , respectively, can be given as

$$N_1 = e^{-E_1 / kT}$$

and $N_2 = e^{-E_2 / kT}$

The ratio of the populations in these two states, N_1/N_2 , is called the *relative population*, which is expressed as

$$\frac{N_2}{N_1} = e^{-(E_2 - E_1)/kT}$$

or $N_2 = N_1 e^{-\Delta E / kT}$ (15.9)

where $\Delta E = E_2 - E_1$, k is the the Boltzmann constant, and T is the absolute temperature.

Since $\Delta E = E_2 - E_1 = h\nu$, so Eq. (15.9) can be rewritten as

$$\frac{N_1}{N_2} = e^{h\nu / kT} \quad (15.10)$$

By putting the value of N_1/N_2 in Eq. (15.8), we get

$$\rho(v) = \frac{A_{21}}{B_{12}} \left[\frac{1}{e^{h\nu / kT} - (B_{21}/B_{12})} \right] \quad (15.11)$$

From the Planck's radiation law, we know that

$$\rho(v) = (8\pi h\nu^3/c^3) \frac{1}{e^{h\nu / kT} - 1} \quad (15.12)$$

where c is the velocity of light in free space.

Energy density $\rho(v)$ given by Eq. (15.11) will be consistent with Planck's law given by Eq. (15.12) only if

$$B_{21} = B_{12} \quad (15.13)$$

and $\frac{A_{21}}{B_{12}} = \frac{8\pi h\nu^3}{c^3}$ (15.14)

and hence, $B_{12} = B_{21} = \frac{c^3}{8\pi h\nu^3} A_{21}$ (15.15)

Equations (15.13) and (15.14) are known as *Einstein relations* and Eq. (15.15) gives the relationship between Einstein's A and B coefficients.

15.5 POPULATION INVERSION

Usually, at the state of thermal equilibrium, the number (population) of atoms in lower energy level (E_1) is more than the population of atoms at the higher energy level (E_2). But it is observed that for emission processes and for laser action, it is essential that the number of excited atoms must be more than the atoms in ground state. In other words, the number of atoms in higher energy level (E_3) must be greater than the number of atoms in lower energy level (E_1). The process by which this condition is achieved is known as the process of *population inversion*. This non-equilibrium state in which the population N_2 of the upper energy level exceeds to a large extent the population N_1 of the lower energy level is known as the *state of population inversion*.

In order to illustrate the phenomenon of population inversion, let us consider a system which has three energy states E_1 , E_2 , and E_3 in such a way that $E_3 > E_2 > E_1$. Let N_1 , N_2 , and N_3 be the number of atoms in the energy states E_1 , E_2 , and, E_3 , respectively, at the state of equilibrium. At the state of equilibrium, $N_1 > N_2 > N_3$, however, if the system is supplied with external energy such that $N_3 > N_2 > N_1$, we say that the system has reached at population inversion. The energy states and the population of atoms in different energy states are shown in Fig. 15.5.

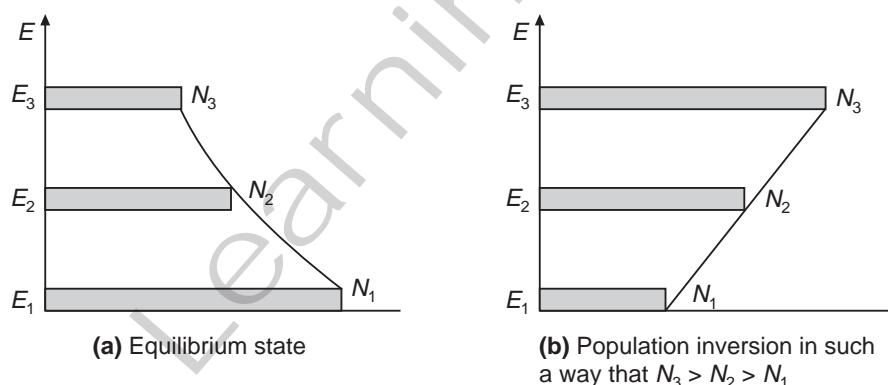


Fig. 15.5 Population inversion process

At the condition of population inversion, stimulated emission can produce a cascade of light. Initially, few random photons are emitted spontaneously; these photons trigger stimulated emission of more photons. These emitted photons induce still more stimulated emission and so on. This process of stimulated emissions remains continued as long as the excited state population is more than the lower level population, and consequently, light gets amplified as shown in Fig. 15.6. As soon as the population at the lower level becomes equal to or larger than that at the excited state, the population inversion ends, the stimulated emission diminishes, and the amplification of light ceases.

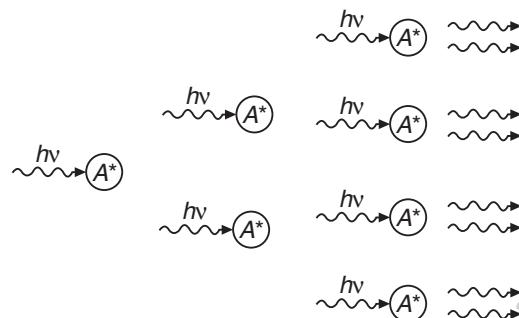


Fig. 15.6 Amplified light obtained by the multiplication of stimulated photons

15.6 PUMPING MECHANISM

The process by which we can realise and maintain the state of population inversion is known as *pumping*. In this process, it is necessary that atoms must be continuously promoted from the lower level to the excited level. The pumping energy is to be supplied somehow to the atoms to raise them from the lower level to the excited level and to ensure that the population at the excited level is at a value greater than that at the lower energy level. Thus, in simple words, we can say that the process by which atoms are raised from the lower energy level to the upper energy level is called *pumping*.

There are different methods by which we can supply the suitable value of energy to raise the atoms from the lower energy level to the higher energy level for achieving population inversion. These methods are described below.

15.6.1 Optical Pumping

In optical pumping, a light source (suitable photons) is used to supply luminous energy. Most often this energy is given in the form of short flashes of light. This technique was first used by Maiman in the ruby laser and is also widely used in solid-state lasers. In this method, the laser material is kept inside a helical xenon flash lamp similar to that which is used in photography.

15.6.2 Electric Discharge

It is another method of pumping in which direct electron excitation occurs through an electric discharge. This method is preferred in gaseous ion lasers, for example, argon-ion laser. In the case of gas laser, a high-voltage pulse initially ionises the gas so that it conducts electricity. An electric current flowing through the gas excites the atoms to the excited level from where they drop to the metastable upper laser level leading to population inversion.

15.6.3 Inelastic Collisions Between Atoms

In an important class of lasers, pumping by electrical discharge provides the initial excitation which raises one type of atoms to their excited states. These atoms collide inelastically with another atoms and provide them enough energy to excite them to the higher energy level and thus help in population inversion. This type of pumping occurs in helium-neon laser.

15.6.4 Direct Conversion

In light-emitting diodes (LEDs) and semiconductors, the electrons recombine with holes producing laser light. Thus, the direct conversion of electrical energy into radiation takes place.

15.6.5 Chemical Reaction

In chemical lasers, radiations come out of a chemical reaction, without any need of other energy source. For example, when hydrogen combines with fluorine, heat energy is generated as



This heat energy is enough to pump a CO_2 laser.

15.7 COMPONENTS OF LASER SYSTEM

In laser, we obtain the amplification of light using principles of stimulated emission of radiations. The stimulated emission produces completely coherent and intense radiations. There are three essential components of laser action, which are shown in Fig. 15.7 and described under the following sections.

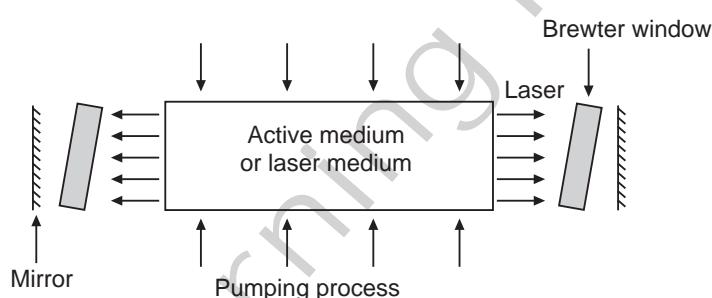


Fig. 15.7 Essentials of laser

15.7.1 Active Medium

It is the medium in which the laser action is made to take place. It may be in solid, liquid, or gaseous state where atoms/ions are lying in excited state (metastable state) to facilitate stimulated emission. Most of the lasers are named after the materials used as active or laser medium, e.g., ruby laser, He–Ne laser, CO_2 laser, Nd: YAG laser, and semiconductor lasers. The most important characteristic of laser medium is that it should be capable to obtain the population inversion in it.

15.7.2 Population Inversion

Under the normal conditions at thermal equilibrium, the lower state of energy is more populated than the higher states. According to the Boltzmann condition, if N_1 and N_2 are the number of atoms in the energy states E_1 and E_2 , respectively, then

$$\frac{N_2}{N_1} = e^{-hv/k_B T}$$

where

$$hv = E_2 - E_1$$

In order to facilitate the stimulated emission (laser action), it is must that the number of excited atoms should be greater than the number of atoms in the ground state. This condition can be achieved by pumping mechanism in which energy is supplied to the atoms by external impetus through different processes such as optical pumping, electric discharge, direct conversion, and chemical reactions.

15.7.3 The Optical Resonator

It consists of a pair of plane or spherical mirrors in which one is perfect and the other is a partial reflector having common principal axis. The reflection coefficient of one of the mirrors is very near to 1 and that of the other is kept somewhat less than one. Due to the stimulated emission, there are waves between the two mirrors propagating along both the directions which undergo interference to form a standing wave; but for a stable standing wave, the wavelength must satisfy the following condition:

If a wave is traveling to and fro, then the total phase change suffered by the wave in one complete round trip must be an integral multiple of 2π . Hence, if l is the length of the cavity, i.e., length of the laser tube, then

$$\left(\frac{2\pi}{\lambda}\right)2l = 2n\pi \quad \text{where } n = 1, 2, 3, \dots$$

or
$$l = \frac{n\lambda}{2} \quad (15.16)$$

15.8 THREE-LEVEL LASER

Laser action is highly dependent on the pumping scheme and the number of excited atoms ready for stimulated emission. The simplest method of pumping is two-level pumping scheme, but basically, a two-level pumping scheme is not suitable for attaining population inversion because in this scheme, the lifetime of spontaneous emission is very fast.

In order to achieve suitable population inversion for laser action, three-level pumping scheme is used. In this scheme, the excited atoms ready for stimulated emission have more lifetime than the atoms excited under two-level pumping scheme. A model of a three-level pumping scheme is shown in Fig. 15.8.

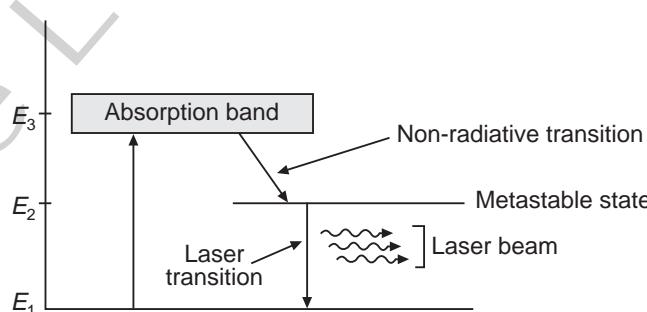


Fig. 15.8 Three-level pumping mechanism

In this scheme, the lower laser level is either the ground state or the level whose separation from the ground state is small compared to kT . When the atoms of the ground state are subjected to an intense radiation of suitable energy, i.e., $E_3 - E_1 = h\nu$, the atoms are excited to the higher energy level E_3 . Out of

16

Optical Fibres

16.1 INTRODUCTION

Optical fibres are mainly concerned with communication. *Communication* is defined as the transfer of information from one point to another. In the process of communication, information is sent from one place to another through carrier waves using the technique of modulation. Main constraints in the communication are transmission fidelity, data rate, distortions, and distance between relay stations. In spite of its limitations, the electrical communication is frequently used from the beginning of nineteenth century (with the invention of telegraph by Samuel F. B. Morse in 1838) to recent years. In order to meet out the demands of telecommunication companies worldwide, optical fibres are used as a dominant transmission system. This optical communication system consists of hair-thin glass fibres that guide light signals with minimum losses over long distances. Optical fibres are thin pipes of glass or plastic (in some cases) that conduct light in a better way as a copper wire conducts electricity or pipes conduct water. Optical fibres are widely used in that communication in which optical signals are used for the transfer of data from one place to another. Optical fibre communication permits the transmission of data over longer distances and at higher data rates than other form of wired and wireless communications. Fibres are frequently used in place of metal wires because signals propagating through fibres suffer less loss and are immune to electromagnetic interference.

In the process of communication, the amount of information carried by electromagnetic waves (carrier waves) having a frequency spectrum down to microwave region is proportional to the frequency. Due to this fact, the microwave communication and traditional communications have their limitations of reducing the heavy communication traffic. In order to reduce the existing problems of traditional communications, there was a search for better carrier waves. This important need of communication was catered by LASER and light emitting diodes. Actually, the invention of these suitable sources of light (carrier waves) gave birth to the optical communication. A light source acting as a carrier wave is capable of carrying far more information than radiowaves and microwaves. It has been observed that the light signals used instead of electric signals in the process of communication can transmit 45 million pulses per second. Fibre optics is the overlap of applied science and engineering concerned with the design and the application. Optical fibres have variety of applications such as in telecommunication industry, medical sciences, illumination technology, instrumentation, etc.

An optical fibre is a cylindrical waveguide system consisting of essentially three regions. The centre is known as the core, which has remarkable property of conducting an optical beam. The middle region is

known as cladding, which exhibits optical properties different from those of core. The refractive index of cladding is always lower than that of the core. The outer region is a protective sheath, which protects the cladding and the core from abrasions, contamination, and moisture.

This chapter will be dedicated to different structures, properties, characteristics, and the applications of optical fibres.

16.2 THE EVOLUTION OF FIBRE OPTIC SYSTEMS

Transmission of light by the method of refraction was first demonstrated by Daniel Colladon and Jacques Babinet in 1840s, with Irish inventor John Tyndall. In the real sense, optical fibre was developed in 1950s with the work of Hopkins and Narinder Singh Kapany in UK, and Van Heel in Holland. These workers did their experiments for the invention of optical fibre based on Tyndall's principle. Later on, more sophisticated optical fibres were developed, where the glass fibre coated with a transparent cladding was invented to offer a more suitable refractive index. After the invention of these optical fibres, scientists started thinking about fibre bundles. The first fibre optic semiflexible gastroscope was patented by Basil Hirschowitz, C. Wilbur Peters, and Lawrence Curtiss, researchers of the University of Michigan in 1956.

In the real sense, the actual evolution of optical fibre technology started in 1960s after the successful invention of suitable light sources such as semiconductor LASER and light emitting diodes (LEDs). The invention of the clad waveguide structure led to the first serious proposals by Charles K. Kao, George A. Hockhman, and Werts in 1966, to utilise optical fibres as a communication medium. In the beginning of their existence, these fibres had a loss of even more than 1000 dB/km. But the proposals led by the above researchers stimulated tremendous efforts to reduce the attenuation by purification of materials. These researchers pointed out that the optical fibre could be a practical medium for communication if the attenuation could be reduced below 20 dB/km. In 1970, Robert D. Maurer and co-workers reported a fibre with 17 dB attenuation per kilometre by doping silica glass with titanium. After a few years, they used germanium oxidized as core dopant to produce a fibre having loss less than 4 dB/km. Now, these days such optical fibre cables are available which have losses less than electrical copper cables. By using glass for both core and clad, Gerhard Bernsee in 1973 invented special optical fibres, which are still used frequently. Fibres fabricated with recently developed technology are characterised by extremely low losses (less than 0.2 dB/km) as a consequence of which, the distance between two successive repeaters could be as large as 250 km. Due to the low cost and better response, optical fibres are replacing the traditional copper cables. Optical fibres are also being extensively used in local area networks to wire up telephones, televisions, computers, or robots in offices and cities.

Photonic crystal fibre was developed in 1991 with the development of emerging field of photonic crystal. In such a crystal fibre, light is guided by means of diffraction rather than total internal reflection.

16.3 ADVANTAGES OF OPTICAL FIBRES

There are many advantages of optical fibres over conducting wires. Some of the important advantages are given below:

- (i) *Cheaper*: Silicon (Si) is the main component in the manufacturing of optical fibres. It is one of the most abundant materials on earth. The overall cost of optical fibre is lower than that of an equivalent cable used in communication.

- (ii) *Not hazardous*: In the wire communication, there is always a risk of accident. Due to short circuit and sparking, there is a risk of high damage. But there is no chance of such problems in the use of optical fibre cables.
- (iii) *Immune to EMI and RFI*: Optical fibres are frequently used in place of metal wires because signals propagating through fibres suffer less loss and are immune to electromagnetic interference and radio frequency interference. It is only due to this reason that the information is carried by photons in the optical fibre communication.
- (iv) *Small size, light weight, flexible, and strong*: The size of optical fibre is very small. It is of the order of few hundred microns. Its weight is very less. Optical fibres are flexible. They can be molded at any place with the help of suitable connectors and splices. An optical fibre has an outer jacket, which protects it from any outer damage and hence, makes it strong.
- (v) *No crosstalk*: There is no chance of crosstalk in the optical fibre communication because the information propagating through the optical fibre is trapped within the fibre and cannot leak out.
- (vi) *High information-carrying capacity*: A light source, acting as a carrier wave is capable of carrying far more information than radiowaves and microwaves. It has been observed that the light signals used instead of electric signals in the process of communication can transmit 45 million pulses per second.
- (vii) *Low loss*: Optical fibres fabricated with recently developed technology are characterised by extremely low losses (less than 0.2 dB/km) as a consequence of which, the distance between two successive repeaters can be as large as 250 km.
- (viii) *Higher data-rate transmission*: Optical fibre communication permits the transmission of data over longer distances and at higher data rates than other forms of wired and wireless communications.

16.4 FUNDAMENTALS OF OPTICAL FIBRES

Optical fibres are the dielectric waveguides which are fabricated from glass or plastic and are operated on optical frequencies. These are normally of cylindrical form. An optical fibre has three principal sections: (i) core, (ii) cladding, and (iii) jacket.

16.4.1 Core

It is the innermost region of the fibre (Fig. 16.1) which has specific property of conducting an optical beam. Core is usually made of glass or plastic. The core is the actual working structure of the fibre which is covered with another layer of glass or plastic having slightly different chemical composition known as cladding.

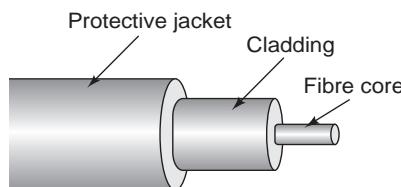


Fig. 16.1 An optical fibre waveguide showing core, cladding, and protective jacket

16.4.2 Cladding

It is the region just above the core region of the optical fibre. Usually, it has lower refractive index than the core region. The cladding has optical properties very different from those of the core. The optical fibre may have an abrupt boundary between the core and the cladding or there may be a gradual change in the material between the two.

16.4.3 Jacket

The outermost section of the optical fibre is known as *jacket*. It is made up of plastic or special kind of polymer and other materials usually opaque in nature. It protects the core from abrasion, interaction with environment, moisture, absorption, crushing, and other adversities of the terrestrial atmosphere and thus, enhances its tensile strength.

16.5 PROPAGATION OF LIGHT THROUGH OPTICAL FIBRE

In the optical fibre, the arrangement of core and cladding regions is done in such a way that the core acts like a continuous layer of two parallel mirrors. The message which has to be sent through fibre is first encoded into a light wave and then fed into the fibre where it is propagated as a result of multiple internal reflections.

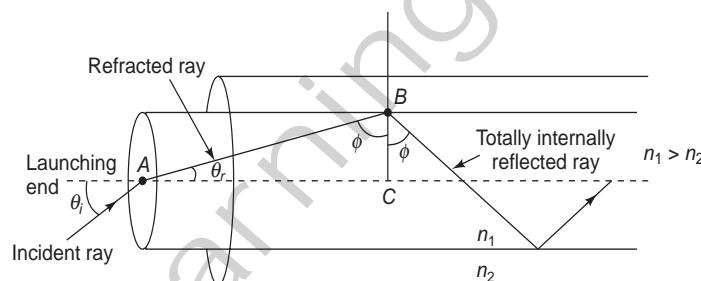


Fig. 16.2 Propagation of light in an optical fibre

The propagation mechanism of light through an optical fibre is shown in Fig. 16.2. The end at which the light enters the fibre is known as the launching end. Let the refractive index of the core be n_1 and the refractive index of cladding be n_2 ($n_2 < n_1$). Let the refractive index of outside medium from where the light is launched be n_0 . Let θ_i be the angle made by light with the axis of fibre, which is entering at launching end and θ_r be the angle made by the refracted ray with the axis which strikes the core-cladding interface at an angle ϕ .

If $\phi >$ critical angle (θ_c), then the ray undergoes total internal reflection at the interface. As long as the angle $\phi > \theta_c$, the light remains within the fibre.

Applying Snell's law at the launching face of the fibre, we get

$$\frac{\sin \theta_i}{\sin \theta_r} = \frac{n_1}{n_0} \quad (16.1)$$

Now, the largest value of θ_i will be at $\phi = \theta_c$.

From the right-angled triangle ABC , we have

$$\sin \theta_r = \sin (90^\circ - \phi) = \cos \phi$$

From Eq. (16.1), we know that

$$\sin \theta_i = \frac{n_1}{n_0} \sin \theta_r$$

By putting the value of $\sin \theta_r$, we get

$$\sin \theta_i = \frac{n_1}{n_0} \cos \phi$$

When $\phi = \theta_c$, $\theta_i = \theta_{\max}$

Now,

$$\sin \theta_{\max} = \frac{n_1}{n_0} \cos \theta_c \quad (16.2)$$

Using Snell's law at point B or cladding boundary,

$$n_1 \sin \theta_c = n_2 \sin 90^\circ$$

or $\sin \theta_c = \frac{n_2}{n_1}$ (Because for total internal reflection, refraction angle will be 90°)

So, $\cos \theta_c = \frac{\sqrt{n_1^2 - n_2^2}}{n_1}$ (16.3)

Using the value of $\cos \theta_c$ from Eq. (16.3) in Eq. (16.2), we get

$$\sin \theta_{\max} = \frac{\sqrt{n_1^2 - n_2^2}}{n_0} \quad (16.4)$$

For the conditions when $\sqrt{n_1^2 - n_2^2} \leq n_0$, for all values of angle of incidence, total internal reflection will occur. For special condition, when $n_0 = 1$, the maximum value of angle of incidence (θ_i) for the ray to be guided is given by

$$\begin{aligned} \sin \theta_m &= \sqrt{n_1^2 - n_2^2} \\ \text{or } \theta_m &= \sin^{-1} \left[\sqrt{n_1^2 - n_2^2} \right] \end{aligned} \quad (16.5)$$

In the above expression, θ_m is known as the acceptance angle of the fibre. Acceptance angle is defined as the maximum angle which incident light makes with the axis of fibre at which the ray is propagated (guided) through the fibre.

The light rays contained within the cone having a full angle $2\theta_m$ are accepted and transmitted along the fibre. This cone is known as acceptance cone.

16.6 FRACTIONAL REFRACTIVE INDEX CHANGE

Fractional refractive index change is defined as the ratio of the difference between the refractive indices of the core and the cladding to the refractive index of the core. It is denoted by Δ and is given as

$$\Delta = \frac{n_1 - n_2}{n_1} \quad (16.6)$$

The value of Δ is always positive and less than one because $n_1 > n_2$ (always), otherwise the phenomena of total internal reflection will not be fulfilled.

16.7 NUMERICAL APERTURE

Numerical aperture (NA) is a number, which defines the light acceptance or light propagating capacity of a fibre. Sometimes, it is also known as figure of merit. Numerical aperture is defined as the sine of maximum angle (acceptance angle) from the fibre axis at which the light may enter in the core of optical fibre and propagate through it by several internal reflections. It is expressed as

$$NA = \sin \theta_m = \frac{\sqrt{n_1^2 - n_2^2}}{n_0}$$

Usually n_0 is the refractive index of the air and is given as $n_0 = 1$. Hence, the value of NA can be given as

$$NA = \sin \theta_m = \sqrt{n_1^2 - n_2^2} \quad (16.7)$$

where n_1 and n_2 are the refractive indices of core and cladding, respectively.

Numerical aperture can also be expressed in terms of fractional refractive index change (Δ) as follows:

$$\begin{aligned} \Delta &= \frac{n_1 - n_2}{n_1} \\ &= \frac{(n_1 + n_2)(n_1 - n_2)}{(n_1 + n_2)n_1} \\ &= \frac{(n_1^2 - n_2^2)}{(n_1 + n_2)n_1} \end{aligned}$$

Since the difference in n_1 and n_2 is very small, so we can write $n_1 + n_2 \approx 2n_1$

$$\begin{aligned} \text{Hence, } \Delta &= \frac{n_1^2 - n_2^2}{2n_1 n_1} \\ &= \frac{n_1^2 - n_2^2}{2n_1^2} \\ &= \frac{(NA)^2}{2n_1^2} \end{aligned}$$

or $NA = n_1 \sqrt{2\Delta} \quad (16.8)$

Solved Examples**Example 16.1**

Compute the numerical aperture and hence, the acceptance angle for an optical fibre. Given that refractive indices of the core and the cladding are 1.42 and 1.40, respectively.

Solution

We know that the numerical aperture is given as

$$\text{NA} = \sqrt{n_1^2 - n_2^2}$$

Since, $n_1 = 1.42$ and $n_2 = 1.40$

$$\begin{aligned}\text{NA} &= \sqrt{(1.42)^2 - (1.40)^2} \\ &= \sqrt{0.0564} \\ &= 0.237\end{aligned}$$

Acceptance angle θ_m is given by

$$\begin{aligned}\theta_m &= \sin^{-1}[n_1^2 - n_2^2] \\ &= \sin^{-1}[\text{NA}] \\ &= \sin^{-1}[0.237] \\ &= 13^\circ 48'\end{aligned}$$

Example 16.2

An optical fibre has the core refractive index $n_1 = 1.36$ and the relative difference in index $\Delta = 0.025$. Find the

- (i) refractive index of cladding, i.e., n_2 ,
- (ii) numerical aperture, and
- (iii) acceptance angle.

Solution

(i) We know that

$$\Delta = \frac{n_1 - n_2}{n_1}$$

Given that $\Delta = 0.025$ and $n_1 = 1.36$. Therefore,

$$\begin{aligned}0.025 &= \frac{1.36 - n_2}{1.36} \\ \Rightarrow 0.034 &= 1.36 - n_2 \\ \Rightarrow n_2 &= 1.36 - 0.034 \\ &= 1.326\end{aligned}$$

(ii) We know that

$$\text{NA} = n_1 \sqrt{2\Delta}$$

Given that $n_1 = 1.36$ and $\Delta = 0.025$. Therefore,

$$\begin{aligned}\text{NA} &= 1.36 \sqrt{2 \times 0.025} \\ &= 1.36 \times 0.2236 \\ &= 0.304\end{aligned}$$

(iii) Acceptance angle α can be given as

$$\begin{aligned}\alpha &= \sin^{-1}(\text{NA}) \\ &= \sin^{-1}(0.304) \\ &= 17.7^\circ\end{aligned}$$

Example 16.3

An optical fibre core and its cladding have refractive indices of 1.545 and 1.495, respectively. Calculate the critical angle ϕ_c , the acceptance angle $\phi_{in(\max)}$, and the numerical aperture.

Solution

Using Snell's law at the boundary of core and cladding, we can write

$$n_1 \sin \phi_c = n_2 \sin 90^\circ \quad (\text{for total internal reflection } \theta = 90^\circ)$$

$$\text{Now, } \sin \phi_c = \frac{n_2}{n_1}$$

$$= \frac{1.495}{1.545}$$

$$\begin{aligned}\text{or } \phi_c &= \sin^{-1}(0.9676) \\ &= 83.75^\circ\end{aligned}$$

Acceptance angle

$$\begin{aligned}\phi_{in(\max)} &= \sin^{-1} \left[\sqrt{(1.545)^2 - (1.495)^2} \right] \\ &= \sin^{-1} \left[\sqrt{0.152} \right] \\ &= \sin^{-1}[0.3899] \\ &= 25.498^\circ\end{aligned}$$

Numerical aperture

$$\begin{aligned}\text{NA} &= \sqrt{n_1^2 - n_2^2} \\ &= \sqrt{(1.545)^2 - (1.495)^2} \\ &= \sqrt{0.152} \\ &= 0.3899\end{aligned}$$

Example 16.4

An optical fibre has an NA of 0.20 and a cladding refractive index of 1.59. Determine the acceptance angle for the fibre in water, which has refractive index of 1.33.

Solution

Numerical aperture

$$\begin{aligned} \text{NA} &= \sqrt{n_1^2 - n_2^2}, && (\text{where } n_0 = 1 \text{ for air}) \\ 0.20 &= \sqrt{n_1^2 - n_2^2} \\ \therefore n_1 &= \sqrt{(0.20)^2 + (1.59)^2} \\ &= 1.6025 \end{aligned}$$

In water,

$$\begin{aligned} n_\omega \text{ (refractive index in water)} &= 1.33 \\ \therefore \text{NA} &= \frac{\sqrt{n_1^2 - n_2^2}}{n_\omega} \\ &= \frac{\sqrt{(1.6025)^2 - (1.59)^2}}{1.33} \\ &= 0.15 \end{aligned}$$

∴ Acceptance angle

$$\begin{aligned} \theta_m &= \sin^{-1}(\text{NA}) \\ &= \sin^{-1}(0.15) = 8.6^\circ \end{aligned}$$

Example 16.5

An optical fibre has following data:

$$n_1 = 1.55$$

$$n_2 = 1.50$$

Core diameter = 50 μm

Calculate the

- (i) numerical aperture,
- (ii) acceptance angle, and
- (iii) how many reflections per metre are suffered by the guided ray at steepest angle with respect to the fibre axis.

Solution

Numerical aperture

$$\text{NA} = \sqrt{n_1^2 - n_2^2} = \sqrt{(1.55)^2 - (1.50)^2}$$

$$\begin{aligned}
 &= \sqrt{(1.55 + 1.50)(1.55 - 1.50)} \\
 &= \sqrt{3.05 \times 0.05} \\
 &= 0.3905
 \end{aligned}$$

(ii) Acceptance angle

$$\begin{aligned}
 \alpha &= \sin^{-1}(NA) \\
 &= \sin^{-1}(0.3905) \\
 &= 23^\circ
 \end{aligned}$$

(iii) The steepest angle of incidence at the core-cladding boundary exceeds the critical angle θ_c . Thus,

$$\begin{aligned}
 \theta_c &= \sin^{-1}\left(\frac{n_2}{n_1}\right) = \sin^{-1}\left(\frac{1.50}{1.55}\right) \\
 &= \sin^{-1}(0.9677) \\
 &= 75.4^\circ
 \end{aligned}$$

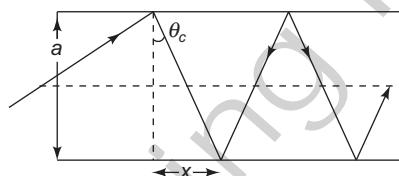


Fig. 1

If x is the axial distance traversed by the ray between two successive reflections suffered by the ray and a is the core diameter, then from the figure,

$$\begin{aligned}
 x &= a \tan \theta_c \\
 &= 50 \tan(75.4^\circ) \\
 &= 50 \times 3.8391 \\
 &= 191.95 \mu\text{m}
 \end{aligned}$$

Number of reflections per metre is

$$\begin{aligned}
 \frac{1\text{m}}{x} &= \frac{1\text{m}}{191.95 \times 10^{-6}} = 5209.6 \\
 &\approx 5209
 \end{aligned}$$

Example 16.6

A glass fibre has a core material of refractive index 1.45 and cladding material of refractive index 1.40. If it is surrounded by air, compute the critical angle

- (i) at the core-cladding boundary and
- (ii) cladding-air boundary.

Solution

(i) Given that $n_1 = 1.45$

17

Maxwell's Equations and Electromagnetic Waves

17.1 INTRODUCTION

The science of electricity has its roots in observation. It was known to Thales of Miletus in 600 B.C. that a rubbed piece of amber will attract bits of straw. The study of magnetism goes back to the observation that naturally occurring “stones” (i.e., magnetite) will attract iron. These two sciences developed quite separately until 1820, when Hans Christian Oersted (1777–1851) observed a connection between them, namely, that an electric current in a wire can affect a magnetic compass needle.

The new science of electromagnetism was developed further by many workers. One of the most important was Michael Faraday (1791–1867). Another worker, James Clerk Maxwell (1831–1879), put the laws of electromagnetism in the form in which we know them today. These laws are called *Maxwell's equations* and play the same role in electromagnetism that Newton's laws of motion do in mechanics.

Maxwell deduced that light is electromagnetic in nature and that its speed can be found by making purely electric and magnetic measurements. Thus, the science of optics was intimately connected with those of electricity and magnetism. The scope of Maxwell's equations is remarkable, as it includes the fundamental principles of large-scale electromagnetic and optical devices, such as motors, cyclotrons, electronic computers, radio, television, microwave radar, microscopes, and telescopes.

The development of classical electromagnetism did not end with Maxwell. Oliver Heaviside (1850–1925) and H.A. Lorentz (1853–1928) contributed substantially to the clarification of Maxwell's theory. Hertz (1857–1894) took a great step forward when, more than twenty years after Maxwell set up his theory, he produced in the laboratory electromagnetic “Maxwellian waves” of a kind that we now call *short radio waves*. It remained for Marconi and others to exploit this practical application of the electromagnetic waves of Maxwell and Hertz.

Present interest in electromagnetism takes two forms. At the level of engineering applications, Maxwell's equations are used constantly and universally in the solution of a wide variety of practical problems. At the level of the foundations of the theory, there is a continuing effort to extend its scope in such a way that electromagnetism is revealed as a special case of a more general theory.

17.1.1 Scalar and Vector Fields

A continuous function of the position of a point in a region of space is called a *point function*. The region of space in which it specifies a physical quantity is known as a *field*. These fields are classified into two groups:

- (i) *Scalar field*: A scalar field is defined as that region of space, whose each point is associated with a *scalar point function*, i.e., a continuous function which gives the value of a physical quantity as flux, potential, temperature, etc. In a scalar field, all the points having the same scalar physical quantity are connected by the means of surfaces called *equal or level surfaces*.
- (ii) *Vector field*: A vector field is specified by a continuous vector point function having magnitude and direction, both of which change from point to point, in the given region of field. The method of presentation of a vector field is called *vector lines*, or *lines of surfaces*. The tangent at a vector line gives the direction of the vector at the point.

17.1.2 Gradient, Divergence, and Curl

In vector calculus, we study about the rate of change of scalar and vector fields. For this purpose, a common operator called *del*, or *nabla*, is used, which is written as

$$\vec{\nabla} = i \frac{\partial}{\partial x} + j \frac{\partial}{\partial y} + k \frac{\partial}{\partial z}$$

If $\phi(x, y, z)$ is a differentiable scalar function, its gradient is defined as

$$\text{grad } \phi = \nabla \phi = \left(i \frac{\partial}{\partial x} + j \frac{\partial}{\partial y} + k \frac{\partial}{\partial z} \right) \phi$$

or
$$\vec{\nabla}\phi = i \frac{\partial\phi}{\partial x} + j \frac{\partial\phi}{\partial y} + k \frac{\partial\phi}{\partial z}$$

Physically, $\text{grad } \phi$ is a vector whose magnitude at any point is equal to the rate of change of ϕ at a point along the normal to the surface at that point.

If \vec{F} is a vector point function ($\vec{F} = F_1 i + F_2 j + F_3 k$), where F_1, F_2 , and F_3 are functions of x, y , and z , then its divergence written as $\text{div } F$, or $\vec{\nabla} \cdot \vec{F}$, is given by

$$\begin{aligned} \vec{\nabla} \cdot \vec{F} &= \left(i \frac{\partial}{\partial x} + j \frac{\partial}{\partial y} + k \frac{\partial}{\partial z} \right) \cdot (F_1 i + F_2 j + F_3 k) \\ &= \frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z} \quad \left[\begin{matrix} i \cdot i = j \cdot j = k \cdot k = 1 \\ i \cdot j = j \cdot k = k \cdot i = 0 \end{matrix} \right] \end{aligned}$$

Divergence of a vector point function, physically signifies the outward normal flux of vector field from a closed surface.

If the divergence of any vector function is zero, then the flux of vector function entering into a region must be equal to that leaving it. This vector function is called *solenoidal*.

If \vec{F} is a vector point function ($\vec{F} = F_1 \hat{i} + F_2 \hat{j} + F_3 \hat{k}$), where F_1 , F_2 , and F_3 are functions of x , y , and z , then its curl is defined as

$$\begin{aligned}\text{Curl } \vec{F} &= \begin{vmatrix} i & j & k \\ \partial / \partial x & \partial / \partial y & \partial / \partial z \\ F_1 & F_2 & F_3 \end{vmatrix} \\ &= i \left(\frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z} \right) - j \left(\frac{\partial F_1}{\partial z} - \frac{\partial F_3}{\partial x} \right) + k \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} \right)\end{aligned}$$

A vector field \vec{F} is called *irrotational* if $\text{curl } \vec{F} = 0$. Such fields are also known as *conservative fields*.

17.1.3 Gauss Divergence Theorem (Relation between Surface and Volume Integration)

This theorem states that the flux of a vector field \vec{F} , over any closed surface S , is equal to the volume integral of the divergence of that vector field over the volume V enclosed by the surface S .

$$\int_S \vec{F} \cdot d\vec{S} = \int_V \text{div } \vec{F} dV \quad (17.1)$$

17.1.4 Stokes Theorem (Relation between Line Integral and Surface Integration)

This theorem states that the surface integral of the curl of a vector field \vec{A} , taken over any surface S , is equal to the line integral of \vec{A} around the closed curve forming the periphery of the surface.

$$\begin{aligned}\iint_S (\text{curl } \vec{A}) \cdot d\vec{S} &= \oint_C \vec{A} \cdot d\vec{l} \\ \text{or} \quad \iint_S (\vec{\nabla} \times \vec{A}) \cdot d\vec{S} &= \oint_C \vec{A} \cdot d\vec{l}\end{aligned} \quad (17.2)$$

17.1.5 Poisson's and Laplace's Equations

Poisson's and Laplace's equations are very useful mathematical relations for the calculations of electric fields and potentials that cannot be computed by using Coulomb's and Gauss's law in electrostatic problems. These equations can be derived as follows:

Gauss law in electrostatics is given by

$$\text{div } E = \frac{\rho}{\epsilon_0}$$

Electric field and potential are related as

$$\vec{E} = -\text{grad } V = -\vec{\nabla} V$$

Thus, we obtain

$$\operatorname{div}(-\operatorname{grad} V) = \frac{\rho}{\epsilon_0}$$

or $\vec{\nabla} \cdot (-\vec{\nabla} V) = \frac{\rho}{\epsilon_0}$

or $\nabla^2 V = -\frac{\rho}{\epsilon_0}$ (17.3)

This equation is known as *Poisson's equation* for a homogeneous region. For a charge-free region, i.e., $\rho = 0$, Poisson's equation becomes

$$\nabla^2 V = 0$$

This is called *Laplace's equation*. This equation is applicable to those electrostatic problems, where the entire charge resides on the surface of the conductor or is concentrated in the form of point charges, line charges, or surface charges at a single position. It is also applicable in the cases, where the region between two conductors is filled with one or more homogeneous dielectrics.

17.2 FUNDAMENTAL LAWS OF ELECTRICITY AND MAGNETISM

To understand Maxwell's equation, we must go through the basic laws of electricity and magnetism.

(i) *Gauss's law in electrostatics:* $\oint \vec{E} \cdot \vec{dS} = q/\epsilon_0$ (17.4)

i.e., the electric flux from a closed surface is equal to $1/\epsilon_0$ times the charge enclosed by the surface.

(ii) *Gauss's law in magnetostatics:* $\oint \vec{B} \cdot \vec{dS} = 0$ (17.5)

i.e., the rate of change of magnetic flux through a closed surface is always equal to zero. This also signifies that monopole cannot exist.

(iii) *Faraday's law of electromagnetic induction:* This law states that the rate of change of magnetic flux in a closed circuit induces an emf which opposes the cause, i.e.,

$$e = -\frac{d\phi}{dt} \quad (17.6)$$

(iv) *Ampere's law:* $\oint \vec{B} \cdot \vec{dl} = \mu_0 I$ (17.7)

This law states that the line integral of magnetic flux is equal to μ_0 times the current enclosed by the current loop.

17.3 EQUATION OF CONTINUITY

Electric current is the rate of flow of charge. Therefore, we have

$$i = -\frac{dq}{dt} \quad (17.8)$$

If dq charge is enclosed in a volume element dV and is leaving a surface having area dS , we have

$$i = \int_s \vec{J} \cdot d\vec{S} \text{ and } q = \int_V \rho dV$$

where J is the current density and ρ is the volume charge density. Therefore, Eq. (17.8) becomes

$$\int_s \vec{J} \cdot d\vec{S} = -\frac{d}{dt} \int_V \rho dV$$

$$\text{or } \int_s \vec{J} \cdot d\vec{S} = \int_V \frac{\partial \rho}{\partial t} dV \quad (17.9)$$

Using Gauss divergence theorem on LHS of Eq. (17.9), we get

$$\int_s \vec{J} \cdot d\vec{S} = \int_V \operatorname{div} \vec{J} dV$$

Therefore, Eq. (17.9) becomes

$$\int_V \operatorname{div} \vec{J} dV = -\int_V \frac{\partial \rho}{\partial t} dV$$

$$\text{or } \int_V \left(\operatorname{div} \vec{J} + \frac{\partial \rho}{\partial t} \right) dV = 0$$

Therefore, for an arbitrary surface, we have

$$\operatorname{div} \vec{J} + \frac{\partial \rho}{\partial t} = 0 \quad (17.10)$$

This expression is called *continuity equation*.

17.4 DISPLACEMENT CURRENT

According to Maxwell, it is not only the current in a conductor that produces a magnetic field. A changing electric field in vacuum or in a dielectric also produces a magnetic field. This implies that a changing electric field is equivalent to a current, which flows till the electric field is changing. This equivalent current produces the same magnetic effects as a conventional current in a conductor. This equivalent current is known as *displacement current*.

17.4.1 Modified Ampere's Law

The concept of displacement current due to the discharge of a condenser leads to the modification in Ampere's law. Consider the process of charging of a parallel plate capacitor through a series circuit as shown in Fig. 17.1.

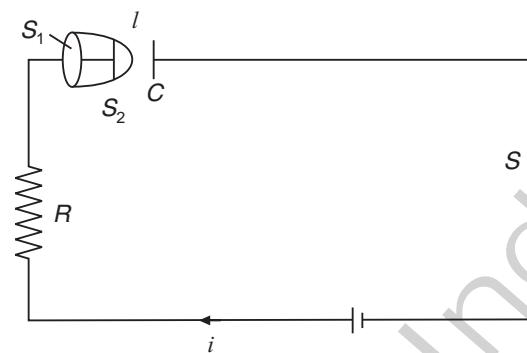


Fig. 17.1 Charging of a capacitor

Let us consider a plane surface S_1 and a hemispherical surface S_2 around the condenser plate as shown in Fig. 17.1. Let both surfaces be bounded by the same closed path l , and applying Ampere's law to the surface S_1 , we get

$$\oint_{S_1} \vec{B} \cdot d\vec{l} = \mu_0 i \quad (17.11)$$

Now, during the process of charging, current i has been flowing through the plane surface S_1 . If it is applied to the hemispherical surface S_2 , we get

$$\oint_{S_2} \vec{B} \cdot d\vec{l} = 0 \quad (17.12)$$

(because no current is enclosed by the surface S_2).

But Eqs. (17.11) and (17.12) show contradiction to each other. Hence, Maxwell introduced the idea that a changing electric field is a source of magnetic field in the gap between the capacitor plates (during charging) and is equivalent to the displacement current devalued by i_d . If ϕ_E is the electric flux, then from equation of continuity, i_d should be equal to $\epsilon_0 d\phi_E/dt$. Therefore, if along with an electric current, there exists a magnetic field, the modified Ampere's law becomes

$$\oint \vec{B} \cdot d\vec{l} = \mu_0 \left(i + \epsilon_0 \frac{d\phi_E}{dt} \right) = \mu_0 (i + i_d) \quad (17.13)$$

Now, the electric field for the charge q developed at the plates of a parallel plate capacitor, each having an area A is given by

$$E = \frac{q}{\epsilon_0 A}$$

or $\frac{dE}{dt} = \frac{1}{\epsilon_0 A} \frac{dq}{dt} = \frac{i}{\epsilon_0 A}$

or $i = \epsilon_0 A \frac{dE}{dt}$ (17.14)

$$= \epsilon_0 \frac{d(EA)}{dt} = \epsilon_0 \frac{d\phi_E}{dt} = i_d \quad (\phi_E = EA)$$

Thus, the displacement current in the gap is identical to the conduction current in the connecting wires.

From Eq. (17.14), we can write

$$i_d = A \frac{d(\epsilon_0 E)}{dt} = A \frac{dD}{dt} \quad (:\vec{D} = \epsilon_0 \vec{E})$$

or $\frac{i_d}{A} = \frac{dD}{dt}$

or $J_d = \frac{dD}{dt}$

Hence, modified Ampere's law becomes

$$\oint \vec{B} \cdot d\vec{l} = \mu_0 \left(i + \epsilon_0 \frac{d\phi_E}{dt} \right) \quad (17.15)$$

17.5 MAXWELL'S ELECTROMAGNETIC EQUATIONS

Maxwell's equations are based on the fundamental laws of physics, which we have already discussed in previous articles. With the help of these equations, one can analyse time-varying fields.

17.5.1 Maxwell's Equations in Differential Form

(i) $\vec{\nabla} \cdot \vec{D} = \rho$ or $\text{Div } \vec{D} = \rho$

(ii) $\vec{\nabla} \cdot \vec{B} = 0$ or $\text{Div } \vec{B} = 0$

(iii) $\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}$ or $\text{Curl } \vec{E} = -\frac{\partial \vec{B}}{\partial t}$

(iv) $\vec{\nabla} \times \vec{H} = \vec{J} + \frac{\partial \vec{D}}{\partial t}$ or $\text{Curl } \vec{H} = \vec{J} + \frac{\partial \vec{D}}{\partial t}$

where

\vec{D} = Electric displacement vector, \vec{B} = Magnetic flux density

\vec{E} = Electric field intensity, \vec{H} = Magnetic field intensity

\vec{J} = Current density (conventional)

ρ = Charge density

17.5.2 Maxwell's Equations in Integral Form

$$(i) \int_s \vec{D} \cdot d\vec{S} = \int_V \rho dV \text{ or } \oint_s \vec{E} \cdot d\vec{S} = q$$

$$(ii) \oint_s \vec{B} \cdot d\vec{S} = 0$$

$$(iii) \oint_s \vec{E} \cdot d\vec{l} = -\frac{\partial}{\partial t} \int_s \vec{B} \cdot d\vec{S}$$

$$(iv) \oint_s \vec{H} \cdot d\vec{l} = \int_s \left(\vec{J} + \frac{\partial \vec{D}}{\partial t} \right) \cdot d\vec{S}$$

Symbols used have the same meaning, as given in Section 17.5.1.

17.5.3 Derivation of Maxwell's Equations

1. Maxwell's first equation, $\operatorname{div} \vec{D} = \rho$ or $\vec{\nabla} \cdot \vec{D} = \rho$:

When a dielectric is placed in a uniform electric field, its molecules get polarised. Thus, a dielectric in an electric field contains two types of charges—free charges, which are embedded, and polarisation charges or bound charges. If ρ and ρ_p are the free and bound charge densities, respectively, at a point in a small volume element dV , then for such a medium, Gauss's law may be expressed as

$$\int_s \vec{E} \cdot d\vec{S} = \frac{1}{\epsilon_0} \int_V (\rho + \rho_p) dV \quad (17.16)$$

where ϵ_0 is the permittivity of the free space.

Now, the bound charge density

$$\rho_p = -\operatorname{div} \vec{P}, \text{ where } \vec{P} \text{ is electric polarisation.}$$

$$\text{Therefore, } \int_s \vec{E} \cdot d\vec{S} = \frac{1}{\epsilon_0} \int_V (\rho - \operatorname{div} \vec{P}) dV$$

Using Gauss divergence theorem on left-hand side of the above expression, we get

$$\int_s \vec{E} \cdot d\vec{S} = \int_v \operatorname{div} \vec{E} dV = \frac{1}{\epsilon_0} \int_v \rho dV - \frac{1}{\epsilon_0} \int_v \operatorname{div} \vec{P} dV$$

or $\int_v \epsilon_0 \operatorname{div} \vec{E} dV + \int_v \operatorname{div} \vec{P} dV = \int_v \rho dV$

$$\int_v \operatorname{div} (\epsilon_0 \vec{E}) dV + \int_v \operatorname{div} \vec{P} dV = \int_v \rho dV$$

$$\int_v \operatorname{div} (\epsilon_0 \vec{E} + \vec{P}) dV = \int_v \rho dV$$

But $\epsilon_0 \vec{E} + \vec{P} = \vec{D}$ is the electric displacement vector.

Thus, $\int_v \operatorname{div} \vec{D} dV = \int_v \rho dV$

or $\int_v (\operatorname{div} \vec{D} - \rho) dV = 0$

Therefore, for an arbitrary surface, we have

$$\operatorname{div} \vec{D} - \rho = 0$$

or $\operatorname{div} \vec{D} = \rho$

or $\vec{\nabla} \cdot \vec{D} = \rho$

This is the required Maxwell's first equation.

In free space, volume charge density $\rho = 0$.

Therefore, Maxwell's first equation in free space is reduced to

$$\operatorname{div} \vec{D} = 0 \text{ or } \vec{\nabla} \cdot \vec{D} = 0$$

or $\operatorname{div} \epsilon_0 \vec{E} = 0$

or $\epsilon_0 \operatorname{div} \vec{E} = 0$

or $\operatorname{div} \vec{E} = 0 \text{ or } \vec{\nabla} \cdot \vec{E} = 0$

2. Maxwell's second equation, $\operatorname{div} \vec{B} = 0$ or $\vec{\nabla} \cdot \vec{B} = 0$:

It has been experimentally observed that the number of magnetic lines of force entering any closed surface enclosing a volume is exactly the same as that leaving it, i.e., the net magnetic flux through any closed surface is always zero.

Hence,

$$\oint_s \vec{B} \cdot d\vec{S} = 0 \quad (17.17)$$

The above expression implies that a monopole or an isolated magnetic pole cannot exist to serve as a source or sink for the line of magnetic induction \vec{B} . This expression is also known as *Gauss's law in magnetostatics*.

Using Gauss divergence theorem in Eq. (17.6), we have

$$\oint_s \vec{B} \cdot d\vec{S} = \int_V \operatorname{div} \vec{B} dV = 0$$

where V is the volume enclosed by surface S .

Hence, for an arbitrary surface,

$$\operatorname{div} \vec{B} = 0$$

or $\vec{\nabla} \cdot \vec{B} = 0$

3. Maxwell's third equation (Faraday's law of electromagnetic induction):

According to Faraday's law of electromagnetic induction, the induced emf around a closed circuit is equal to the negative time rate of change of magnetic flux linked with the circuit, i.e.,

$$e = -\frac{d\phi_B}{dt} \quad (17.18)$$

If \vec{B} is the magnetic induction, then the magnetic flux linked with an area $d\vec{S}$ is

$$\phi_B = \int_s \vec{B} \cdot d\vec{S} \quad (17.19)$$

On combining Eqs. (17.18) and (17.19), we get

$$e = -\frac{d}{dt} \int_s (\vec{B} \cdot d\vec{S})$$

or $e = \int_s \frac{\partial}{\partial t} (\vec{B} \cdot d\vec{S}) \quad (17.20)$

According to definition, the induced emf is related to the corresponding electric field as

$$e = \int_c \vec{E} \cdot d\vec{l} \quad (17.21)$$

Equations (17.20) and (17.21) will give

$$\int_c \vec{E} \cdot d\vec{l} = -\int_s \frac{\partial}{\partial t} (\vec{B} \cdot d\vec{S})$$

18

Superconductivity

18.1 INTRODUCTION (A CHRONOLOGICAL VIEW)

Superconductivity was discovered by Heike Kamerlingh Onnes in 1911. He was studying the resistance of solid mercury at cryogenic temperatures using the recently discovered liquid helium as refrigerant. He observed that at 4.2 K, the resistance abruptly disappeared. In subsequent decades, superconductivity was found in several other materials. In 1913, lead (Pb) was found to behave as a superconductor at 7 K, and in 1941, it was found that niobium nitride became a superconductor at 16 K.

In 1933, Meissner and Oschenfeld discovered that superconductors repel applied magnetic field—a phenomenon known as the *Meissner effect*. In 1935, F. London and H. London showed that the Meissner effect was a consequence of the minimisation of the electromagnetic free energy carried by the superconducting current.

In 1950, the phenomenal Ginzburg–Landau theory of superconductivity was developed by Landau and Gingburg. This theory, which combined the Landau's theory of second-order phase transitions with a Schrödinger-type wave equation, was able to explain the macroscopic properties of superconductors. Abrikosov showed that Ginzburg–Landau theory predicts the division of superconductors into two categories, now referred to as *Type I* and *Type II*.

Also in 1950, Maxwell and Reynolds with other coworkers found that the critical temperature of a superconductor depends on the isotopic mass of the constituent elements. This important discovery pointed to the electron–phonon interaction as the microscopic mechanism responsible for superconductivity.

The complete microscopic theory of superconductivity was finally proposed in 1957 by Bardeen, Cooper, and Schrieffer through BCS theory. This BCS theory explained the superconducting current as a superfluid of Cooper pairs (pairs of electrons interacting through the exchange of phonons).

In 1962, the first commercial superconducting wire, a niobium–titanium wire, was developed, allowing the construction of first practical superconducting magnets. In the same year, Josephson made an important theoretical prediction that a supercurrent can flow between two pieces of a superconductor separated by a thin layer of an insulator. This phenomenon, now called the *Josephson effect*, is exploited by superconducting devices such as SQUIDS.

In 2008, Valeri Vinokur and Tatyana Baturina discovered that the same mechanism that produces superconductivity, can produce a superinsulating state in some materials with almost infinite electrical resistance.

18.2 SUPERCONDUCTIVITY

When certain materials and alloys are cooled to a very low temperature (near absolute zero), their electrical resistivity decreases in the usual way; but on reaching a temperature a few degrees above the absolute zero (0 K), their electrical resistivity falls sharply and becomes almost zero. They are then said to have entered a superconducting state. This new remarkable property of sudden disappearance of the electrical resistance at or below a finite and extremely low temperature (near absolute zero) is called *superconductivity* and the materials that exhibit this property are called *superconductors*.

The transition from a normal conducting state to the superconducting state occurs at a very specific temperature called *transition temperature*. Transition temperature is the characteristic of the material. It is defined as the temperature at or below which the electrical resistance of the specimen becomes zero.

18.3 TEMPERATURE DEPENDENCE IN SUPERCONDUCTING MATERIALS

It has been observed that the temperature dependence in superconducting materials can be classified in two ways: resistivity and external applied field. We will now discuss it in succeeding sections.

18.3.1 Temperature Dependence of Resistivity in Superconducting Materials

Generally, the electrical resistivity of a material decreases as the temperature is reduced. In fact, when the temperature is lowered, the thermal vibrations of the atoms decrease so that the conduction electrons are less frequently scattered. When the temperature reaches a few degrees just above the absolute zero, the material suddenly loses all its electrical resistance (Fig. 18.1) and a state called *superconducting state* is reached.

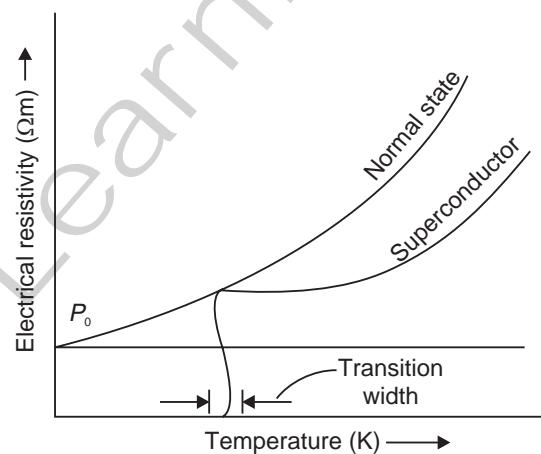


Fig. 18.1 Dependence of temperature on resistivity

It has been observed that the residual resistivity in a real specimen depends on the impurities the metal has, i.e., the larger the impurities, the larger will be its residual resistivity. However, the transformation to the superconducting state is independent of the impurities present in the specimen of superconducting materials. A small amount of impurities does not affect the transition temperature due to the fact that scattering of the electrons by the impurities is almost independent of the temperature. Though the magnetic impurities tend to decrease the transition temperature, the superconductivity of the pure metals,

which have very low transition temperature, may be vanished by doping with small amounts of magnetic impurities. Therefore, the elements like iridium and molybdenum show superconducting phenomenon only when they are in the intrinsic form. Not all pure metals can act as superconductors; copper, iron, and sodium do not exhibit superconducting properties even when cooled to temperatures close to absolute zero.

Figure 18.2 shows observations of H.K. Onnes. It is clear from the figure that the electrical resistance disappears below 4.2 K. It also shows that the width of the transition region in a particular specimen depends on the purity and metallurgical perfections, and can be as sharp as close to 10^{-3} degree or spread over several degrees, while the breadth of the transition region may increase if the sample is metallurgically imperfect. Figure 18.3 gives a comparison of the superconducting transition in a pure and an impure sample of tin.

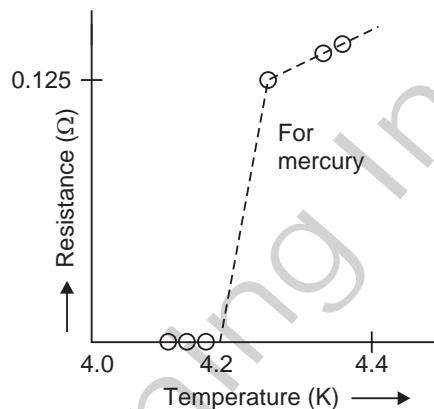


Fig. 18.2 H.K. Onnes' observation

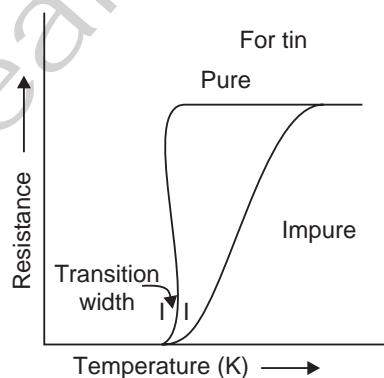


Fig. 18.3 Comparison between pure and impure samples of tin

18.3.2 Temperature Dependence of Critical Field (Effect of External Magnetic Field) on Superconductors

The existence of superconductivity in a metal depends on the intensity of the field in which it is placed and its temperature. Therefore, to maintain superconductivity in a metal, both the parameters, i.e., magnetic

field and temperature, should be less than their respective critical values. The state of superconductivity will vanish if the values of either of these parameters become greater than their critical values, i.e., $T > T_c$ or $H > H_c$. The minimum value of the applied magnetic field is when the material loses its superconductivity. The dependence of the temperature on the magnetic field can be represented by Eq. (18.1).

$$H_c = H_c(0) \left[1 - \left(\frac{T}{T_c} \right)^2 \right] \quad (18.1)$$

where H_c is the critical field strength at temperature T , $H_c(0)$ is the maximum field strength at absolute zero, and T_c is the critical temperature. Equation (18.1) defines a curve which separates the normal region from the superconducting region in the magnetic field versus temperature curve (Fig. 18.4). It is clear from Fig. 18.4 that at any temperature below the critical temperature (T_c), the material remains in the superconducting state till a corresponding critical magnetic field is applied. When a magnetic field is applied, which is higher than the critical magnetic field (H_c), the superconducting state is destroyed and the material comes to its normal state.

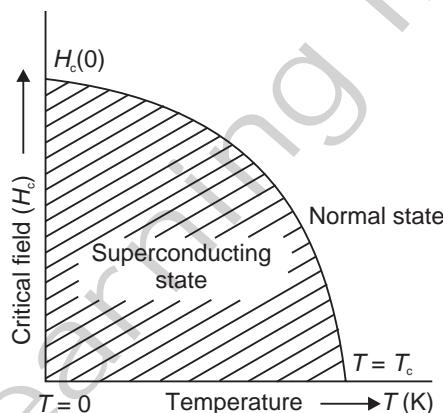


Fig. 18.4 Dependence on temperature

Solved Examples

Example 18.1

A superconducting tin has a critical temperature of 3.7 K in zero magnetic field and a critical field of 0.036 T at 0 K. Find the critical field at 2 K.

Solution

Given that $H_c(0) = 0.0306$ T, $T_c = 3.7$ K, and $T = 2$ K.

Now, the critical field at any temperature T kelvin is given as

$$H_c = H_c(0) \left[1 - \left(\frac{T}{T_c} \right)^2 \right]$$

$$\begin{aligned}
 &= 0.0306 \left[1 - \left(\frac{2}{3.7} \right)^2 \right] \\
 &= 0.0306 \times 0.708 \\
 &= 0.0216 \text{ T}
 \end{aligned}$$

Example 18.2

The transition temperature for lead (Pb) is 7.2 K. However, at 5 K, it loses the superconducting property when subjected to a magnetic field of 3.3×10^4 A/m. Find the value of the magnetic field that will allow the metal to retain its superconductivity at 0 K.

Solution

Given that $T_c = 7.2$ K, $T = 5$ K, and $H_c = 3.3 \times 10^4$ A/m.

$$\begin{aligned}
 H_c &= H_c(0) \left[1 - \left(\frac{T}{T_c} \right)^2 \right] \\
 H_c(0) &= H_c \left[1 - \left(\frac{T}{T_c} \right)^2 \right]^{-1} \\
 &= 3.3 \times 10^4 \left[1 - \left(\frac{5}{7.2} \right)^2 \right]^{-1} \\
 &= 6.37 \times 10^4 \text{ A/m}
 \end{aligned}$$

Example 18.3

At what temperature do we get $H_c = 0.1 H_c(0)$ for lead (Pb) having $T_c = 7.2$ K?

Solution

Given that $H_c = 0.1 H_c(0)$ and $T_c = 7.2$ K.

Now,

$$\begin{aligned}
 &= H_c(0) \left[1 - \left(\frac{T}{T_c} \right)^2 \right] \\
 0.1 H_c(0) &= H_c(0) \left[1 - \left(\frac{T}{7.2} \right)^2 \right] \\
 1 - \left(\frac{T}{7.2} \right)^2 &= 0.1 \\
 T^2 &= 0.9 \times 7.2 \times 7.2 \\
 T &= 6.83 \text{ K}
 \end{aligned}$$

Example 18.4

Calculate the critical field in lead (Pb) at $T = 4.2$ K when $H_c(0) = 0.0803$ Wb/m² for lead (Pb).

Solution

We know that the critical field and the temperature are related as

$$H_c = H_c(0) \left[1 - \left(\frac{T}{T_c} \right)^2 \right]$$

Now, for lead (Pb), we have

$$\begin{aligned} T_c &= 7.2 \text{ K} \\ \therefore H_c &= 0.0803 \left[1 - \left(\frac{4.2}{7.2} \right)^2 \right] \\ &= 0.0803 [1 - 0.34] \\ &= 0.0548 \text{ T} \end{aligned}$$

Example 18.5

A long thin superconducting wire of a metal produces a magnetic field 105×10^3 A/m on its surface due to the current passing through it at a certain temperature T kelvin. Calculate T . Given: $H_c(0) = 150 \times 10^3$ A/m and $T_c = 9.3$ K.

Solution

We know that

$$H_c = H_c(0) \left[1 - \left(\frac{T}{T_c} \right)^2 \right]$$

Rearrangement of the above equation will give

$$T = T_c \left[1 - \left(\frac{H_c}{H_c(0)} \right)^2 \right]^{1/2}$$

Substitution of values will give

$$\begin{aligned} T &= 9.2 \left[1 - \frac{105 \times 10^3}{150 \times 10^3} \right]^{1/2} \\ &= 9.2 \sqrt{\frac{3}{10}} \\ &= 5.03 \text{ K} \end{aligned}$$

Example 18.6

Calculate the transition temperature of niobium for which the critical field is 1×10^5 A/m at 8 K and 2×10^5 A/m at 0 K.

Solution

We know that

$$H_c = H_c(0) \left[1 - \left(\frac{T}{T_c} \right)^2 \right]$$

By suitable rearrangement, we get

$$\frac{T}{T_c} = \left[1 - \left(\frac{H_c(T)}{H_c(0)} \right) \right]^{1/2}$$

or

$$\begin{aligned} T_c &= \frac{T}{\left[1 - \left(\frac{H_c(T)}{H_c(0)} \right) \right]^{1/2}} \\ &= \frac{8}{\left[1 - \frac{1 \times 10^5}{2 \times 10^5} \right]^{1/2}} = 8\sqrt{2} \\ &= 11.3 \text{ K} \end{aligned}$$

Example 18.7

What observation can you make with following data for lead (Pb)?

Given: $T_c = 7.26$ K, $H_c(0) = 8 \times 10^5$ A/m, and $H_c(T) = 4 \times 10^4$ A/m.

Solution

We know that the variation of the magnetic field in the superconductors can be expressed as

$$H_c(T) = H_c(0) \left[1 - \left(\frac{T}{T_c} \right)^2 \right]$$

On rearrangement, this will give

$$T = T_c \left[1 - \left(\frac{H_c(T)}{H_c(0)} \right) \right]^{1/2}$$

On the basis of the given data, one obtains

$$T = 7.26 \left[1 - \frac{4 \times 10^4}{8 \times 10^5} \right]^{1/2}$$

$$= 7.08 \text{ K}$$

This shows that the temperature of the metal should be held below 7.08 K.

Example 18.8

For a specimen of V_3Ga , the critical fields are, respectively, 0.176 T and 0.528 T for 14 K and 13 K. Calculate the transition temperature and the critical fields at 0 K and 4.2 K.

Solution

The variation of the magnetic field is given by

$$H_c(T) = H_c(0) \left[1 - \left(\frac{T}{T_c} \right)^2 \right] \quad (1)$$

For the given set of values, the above equation becomes

$$0.176 = H_c(0) \left[1 - \left(\frac{14}{T_c} \right)^2 \right] \quad (2)$$

$$0.528 = H_c(0) \left[1 - \left(\frac{13}{T_c} \right)^2 \right] \quad (3)$$

On dividing Eq. (2) by Eq. (3), we get

$$\frac{1 - (13/T_c)^2}{(1 - 14/T_c)^2} = \frac{0.528}{0.176} = 3$$

$$1 - \frac{169}{T_c^2} = 3 - \frac{588}{T_c^2}$$

$$\frac{419}{T_c^2} = 2$$

$$T_c = \left(\frac{419}{2} \right)^{1/2} = 14.5 \text{ K}$$

Substituting the value of T_c in Eq. (2), we get

$$0.176 = H_c(0) \left[1 - \left(\frac{4.2}{14.5} \right)^2 \right]$$

$$0.176 = H_c(0) [1 - (0.9655)^2]$$

$$H_c(0) = 2.588 \text{ T}$$

and for $T = 4.2 \text{ K}$,

$$\begin{aligned}
 H_c(T) &= 2.588 \left[1 - \left(\frac{4.2}{14.5} \right)^2 \right] \\
 &= 2.588 [1 - 0.0839] \\
 &= 2.588 \times 0.9161 \\
 &= 2.37 \text{ T}
 \end{aligned}$$

18.4 MEISSNER EFFECT (EFFECT OF MAGNETIC FIELD)

When a material makes the transition from the normal to the superconducting state, it actively excludes magnetic fields from its interior. This phenomenon is called the *Meissner effect*.

Walther Meissner and Ochsenfeld (1953) concluded that if a superconductor is cooled in a longitudinal magnetic field, then above its transition temperature, the lines of magnetic induction pass through the specimen [Fig. 18.5(a)]. However, at or below the transition temperature, the magnetic flux is pushed out of the specimen and does not pass through it; instead it goes round the specimen as shown in Fig. 18.5(b). This indicates that below the transition temperature, the metal becomes perfectly diamagnetic. Thus, the phenomenon of exclusion of the magnetic flux from the interior of a superconductor below the transition temperature (T_c) is the *Meissner effect*.

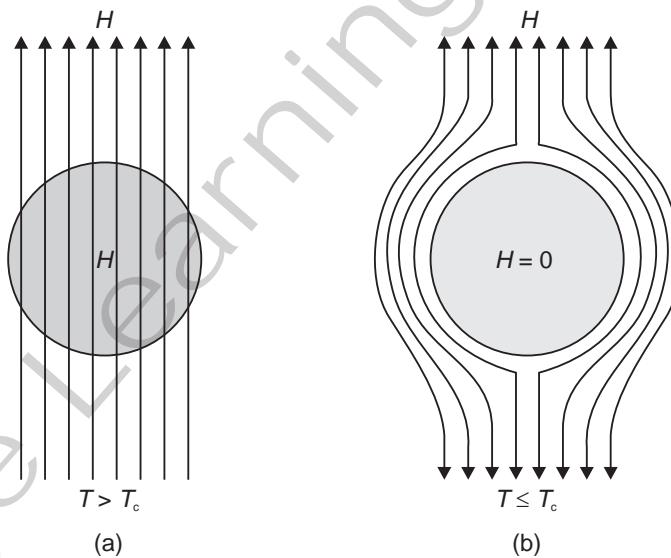


Fig. 18.5 Meissner effect: Magnetic field lines are excluded from a superconductor when it is below transition temperature

In a magnetic field, at or below the transition temperature, a persistent current generates on the surface and circulates so as to cancel the flux density inside the superconductor (like Lenz's law), in a way as a magnetic field is applied after the metal is cooled.

Thus, for a metal in a superconducting state, if the magnetic field (B) is zero inside the specimen, then we get

$$B = \mu_0 (H + M)$$

$$0 = \mu_0 (H + M)$$

or

$$M = -H$$

$$\text{or magnetic susceptibility, } \chi = \frac{M}{H} = -1$$

This is one of the characteristics of a perfect diamagnetic metal. Hence, the superconducting state is a perfect diamagnetic, provided the specimen is pure and strain-free.

Now, Ohm's law is given as

$$J = \sigma E$$

where σ is the conductivity.

In terms of the resistivity (ρ), it can also be written as

$$J = \left(\frac{1}{\rho} \cdot E \right)$$

or

$$E = \rho J \quad (18.2)$$

It is clear from Eq. (18.2) that if the resistivity (ρ) becomes zero, then the electric field (E) must be zero for finite current density (J), i.e.,

$$E = 0$$

Now, from the Maxwell's third equation, we have

$$\vec{\nabla} \times \vec{E} = - \frac{\partial \vec{B}}{\partial t}$$

We can have

$$\frac{\partial B}{\partial t} = 0 \quad (\text{by putting } E = 0)$$

or

$$B = \text{Constant}$$

This implies that the magnetic flux in the interior of the superconductor is independent of the transition temperature. This conclusion contradicts the Meissner effect. Hence, the resistivity and perfect diamagnetism are the two independent and essential properties of a superconductor.

18.5 CRITICAL CURRENTS

Since Biot-Savart's law, it is well known that a current-carrying conductor is associated with a magnetic field. Therefore, a superconductor in which a current is flowing is also associated with a magnetic field. Due to the condition of zero resistance, if the value of the current flowing in the superconductor is large enough so that it crosses the critical value H_c , then the superconductivity will disappear. Hence, a constraint, or limitation, is imposed on the value of the current that may flow in a superconductor. The minimum value of the current that can be passed in a sample of superconductor without destroying the superconductivity is known as *critical current* (I_c). Therefore, for a wire of radius r_1 having current I_c in it, the surface magnetic field (H_s) is given as

19

Nanomaterials and Their Applications

19.1 INTRODUCTION

Nanophysics refers to the science of very minute particles having their dimensions of the order of 10^{-9} m. In fact, the science of nanomaterials is interdisciplinary and it is expected that it will decide the shape and size of the things for their better applications in the days to come. Although the world of nanomaterials was suggested by great physicist Sir Richard Feynman during 1980s by delivering his famous statement “there are more rooms at the bottom.” but the story of real development of this field started from 1990s.

Nanoscience is the study of the fundamental phenomenon of structures and molecules having their dimensions in the range between 10 nm to 100 nm. It is an interdisciplinary regime where the expertise of chemists, physicists, material scientists, and engineers produce nanodevices. It has been found that many physical phenomenon such as statistical mechanical effects and quantum mechanical effects become more pronounced as the size of the system decreases. Furthermore, mechanical, electrical, and optical properties also change on moving from macroscopic system to nanoscale.

In this chapter, we will discuss the world of nanomaterials, important properties of nanomaterials, their synthesis using ball milling sol–gel techniques. We will also describe the synthesis and the properties of carbon nanotubes with important applications of nanomaterials.

19.2 NANOSCIENCE AND NANOTECHNOLOGY

Nanomaterial is the structure of material with improved properties through controlled synthesis and assembly of materials at nanoscale level. Nanoscience is the study of phenomena and manipulation of materials at atomic, molecular, and macromolecular scales. In nanoscience, the properties differ significantly from those at a larger scale. Nanotechnology deals with the design characterisation, production, and application of structures, devices, and systems. Above tasks are done by controlling the shape and size of materials at the nanometre scale. In short, nanoscience and nanotechnologies involve studying and working with matter on an ultra small scale.

19.3 SIGNIFICANCE OF THE NANOSCALE

It is observed that the physics is different on the nanoscale from those at the larger scale. The quantum, mechanical, and thermodynamical properties become important at nanolevel, which are not seen at macroscopic level. After studying an individual molecule's properties, we can put them together in very well defined ways to produce new materials with new and amazing properties. The science dealing with the materials of the nanoworld is an extension of the existing science into the nanoscale or a recasting of existing sciences using a newer, more modern terms. Nanoscience is based on the fact that the properties of materials change with the function of physical dimensions of the materials. The dimensions at which the changes in the properties of materials are observed depend on the specific material and the property in question, as well as on the three dimensions which are restricted in the real space. Actually, change in the properties of materials in these confined spaces are due to the changes in the electronic structure of the materials. The bulk properties of any material are merely the average of all the quantum forces affecting all the atoms. But on reducing the size of the material, we eventually reach a point where the averaging no longer holds. It is observed that the effect does not occur when we go from macro to micro dimensions. It becomes dominant only when the nano size range is reached. The properties of the materials are different at nanolevel due to two main reasons: increased surface area and quantum mechanical effect.

19.3.1 Surface Area

When a bulk material is reduced to the particles of smaller size, then the surface to volume ratio becomes very high. Thus, nanomaterials have a relatively large surface area when compared with the same volume or same mass of the material produced in a large form. For a sphere of radius r , the surface area can be given as

$$\text{Surface area} = 4\pi r^2$$

and its volume can be given as

$$\text{Volume} = \frac{4}{3} \pi r^3$$

$$\text{Now, } \quad \text{the surface to volume ratio} = \frac{3}{r}$$

Thus, we find that when the given volume is divided into smaller parts, surface area increases. Hence, nanoparticles have a greater surface area per given volume compared with larger particles. Since growth and catalytic chemical reactions occur at surfaces, this signifies that nanoparticles will be much more reactive than the same mass of material made up of larger particles. Hence, at nanoscale, physical processes such as friction and sticking become more pronounced than they are in a normal system. These specific factors related to nanoparticles will facilitate their use both inside and outside the system.

19.3.2 Quantum Confinement Effect

It is well known that isolated atoms have discrete energy levels. Naturally, solids have split energy levels known as energy bands. But nanomaterials have the intermediate stage. As far as the physics at nanolevel is concerned, it is the quantum effect that deals with the properties. At reduced dimensions, they are said to be either a quantum well, a quantum dot, or a quantum wire. The physics at these dimensions is entirely different. Actually, when the size of the grains is reduced to nanolevel, then overlapping of



wavefunction and quantum confinements occurs. If d is the diameter of the grain size, then the energy goes up by factor $1/d^2$.

From uncertainty principle, we know that

$$\Delta x \cdot \Delta p \approx h$$

where Δp is the uncertainty in momentum and h is Planck's constant. In case of nanoparticles,

$$\Delta x = d$$

$$\Delta p = \frac{h}{d}$$

If $p_{\max} = \Delta p$, then energy E can be given as

$$E = \frac{p^2}{2m} = \frac{p_{\max}^2}{2m}$$

Now, energy in terms of the size of the grain particle can be given as

$$E = \frac{h^2}{2md^2} \quad (19.1)$$

In Eq. (19.1) at right side, except d , rest are constant. Thus, energy at nanoscale is given as

$$E \propto \frac{1}{d^2} \quad (19.2)$$

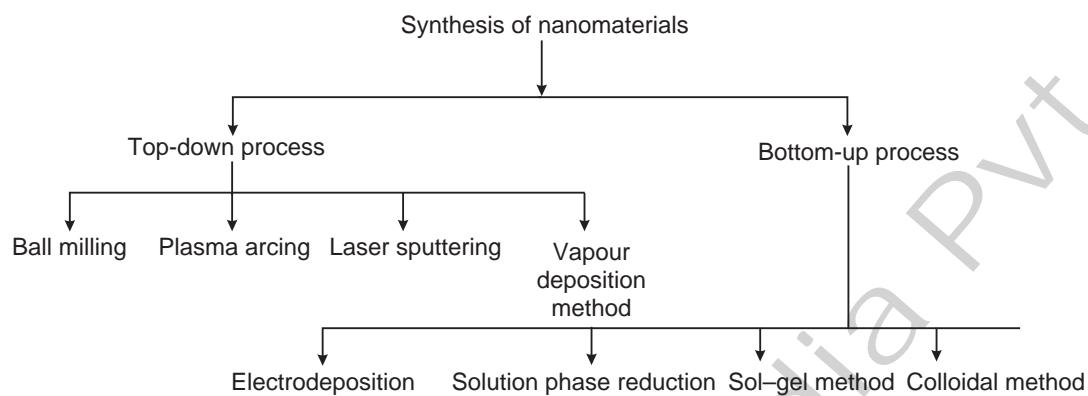
Hence, energy increases by a factor $1/d^2$.

This leads to excitons, and they are primarily responsible for enhanced optical and electrical properties at nanoscale. In case of metals, when their sizes are reduced, their properties are determined by plasmons.

Thus, we can conclude that when the dimensions of a potential well or a box concerned with a particle are reduced to the order of de Broglie wavelength of electron (within few tens of nanometres), then energy levels of electrons change. This effect is called *quantum confinement*. This can affect the optical, electrical, and magnetic behaviour of materials at nanolevels.

19.4 FABRICATION OF NANOMATERIALS

Nanomaterials can be fabricated in two ways, namely, top-down and bottom-up (Fig. 19.1). In top-down approach, nanomaterials are constructed by removing existing material from larger entities. In bottom-up approach, materials and devices are built atom by atom or molecule by molecule. This approach (bottom-up) is more time consuming. Therefore, a self assembly process is usually employed so that atoms or molecules spontaneously arrange themselves into the final product.

**Fig. 19.1** Flowchart showing processes for synthesis of nanomaterials

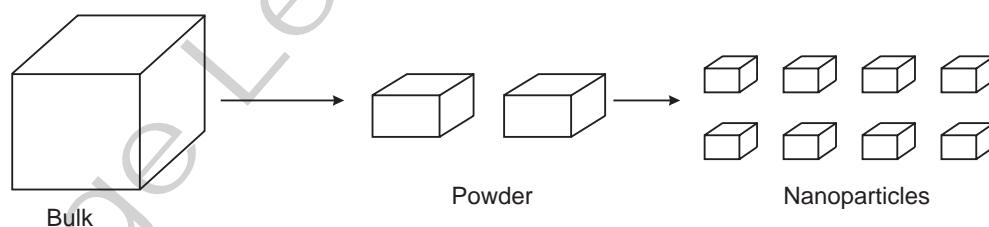
As discussed above, there are two main processes used for the synthesis of nanomaterials, which are described below.

19.4.1 Top-Down Process

Under this process of fabrication, bulk materials are broken into nano-sized particles. In this approach, there is no control over the size and the morphology of particles. There are many methods used in top-down approach to get nano-sized particles from bulk materials. Some common methods of top-down method are as follows.

- (i) Ball milling method
- (ii) Plasma arcing
- (iii) Laser sputtering
- (iv) Vapour deposition method

The schematic representation of the top-down approach for creating nanoparticles is shown in Fig. 19.2.

**Fig. 19.2** Schematic representation of top-down method

The main disadvantage of top-down approach is the imperfection in the surface structure and significant crystallographic damage to the processed patterns. These imperfections lead to extra challenges in the device design and fabrication. But this approach leads to the bulk production of nanomaterials. Out of many methods used for the fabrication of nanoparticle in the top-down process, ball milling is one of the simplest and frequently used method.

Ball Milling Method

Ball milling is a typical top-down method approach for the fabrication of nanoparticles. Basically, milling technique is used in this process of fabrication. In order to describe the various steps used for the preparation of nanoparticles by ball milling method, it is essential to have an idea about the principle and the working of ball milling method. Apart from knowing the operation, it is also necessary to explore the different types of attrition mills for various types of materials, with a variety of sizes of mills in order to up-scale the actual production from the laboratory level to the commercial production of nanoparticles. High-energy mills used for the fabrication of nanoparticles include attrition ball mill, planetary ball mill, vibrating ball mill, low-energy tumbling mill, and high-energy ball mill.

Principle: According to the basic principle of the ball milling method, small hard balls are allowed to rotate inside a container and then it is made to fall on a solid with high force to crush the solid into nanoparticles.

Construction and working: Hardened steel or tungsten carbide balls are put in a container along with the powder of particles ($50\text{ }\mu\text{m}$) of a desired material. The container is closed with tight lids (Fig. 19.3).

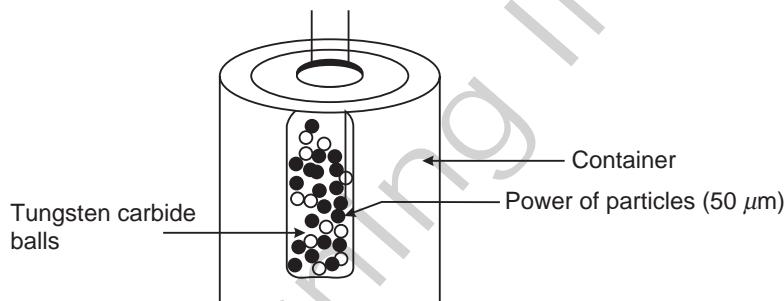


Fig. 19.3 Schematic diagram of the container of ball miller method

When the container is rotated around the central axis, the material is forced to press against the walls. The milling balls impart energy on collision and produce smaller grain size of nanoparticles. Ball milling is also known as mechanical alloying or crushing.

The main advantage of this method is that few milligrams to several kilograms of nanoparticles can be synthesised in a short period of time. This technique can be operated at large scale.

Applications: Some important applications of ball milling technique are as follows:

- (i) This method is useful in the preparation of elemental and metal oxide nanocrystals such as Co, Cr, AlFe, AgFe, and Fe.
- (ii) A variety of intermetallic compounds of Ni and Al can be formed.
- (iii) This method is useful in producing new types of building materials, fireproof materials, glass-ceramics, etc.

19.4.2 Bottom-Up Process

Bottom-up approach refers to the building up of a material from the bottom, i.e., atom by atom, molecule by molecule, or cluster by cluster. Colloidal dispersion is a good example of bottom-up approach in the synthesis of nanoparticles. There are different methods used for the synthesis of nanomaterials under bottom-up process. Some methods of this process are as follows:

- (i) Sol-gel method
- (ii) Colloidal method
- (iii) Electrodeposition
- (iv) Solution phase reductions

Schematic representation of bottom-up process is shown in Fig. 19.4. The most important feature of this process is that the size and morphology of fabricated nanoparticles are well controlled.

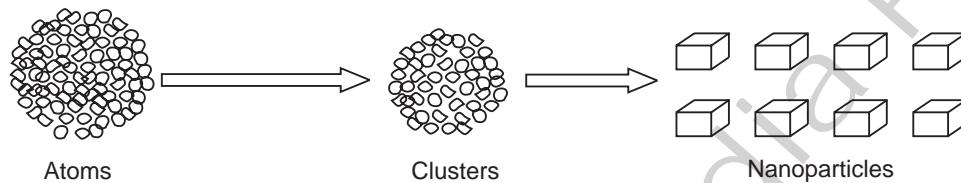


Fig. 19.4 Schematic representation of bottom-up process used for the synthesis of nanomaterials

Out of the many methods used for the synthesis of nanoparticles in bottom-up process, sol–gel is a versatile solution process used for the fabrication of nanomaterials.

Sol-Gel Method

The sol–gel method is an important technique which is frequently used not only for the synthesis of nanoparticles but also to fabricate ceramic or glass materials in a wide variety of forms. The different forms of materials produced by this method are ultra fine or spherical-shaped powders, thin film coatings, ceramic fibres, microporous inorganic membranes, monolithic ceramics and glasses, or extremely porous aerogel materials.

Before discussing different steps used in the sol–gel technique, it is very important to know about *sol* and *gel*. Actually, a sol is a dispersion of the solid particles (~ 0.1 to 1 mm) in a liquid, where only Brownian motions suspend the particles. A gel is a state where both liquids and solids are dispersed in each other, which presents a solid network containing liquid components.

Following four steps are used for the synthesis of nanomaterials using sol–gel method:

- (i) *Preparation of sol*: The starting materials used in the preparation of the sol are usually inorganic metal salts or metal organic compounds such as metal alkoxides.
 - (ii) *Preparation of gel*: In a typical sol–gel process, the precursor is subjected to a series of hydrolysis and polymerisation reaction to form a colloidal suspension known as gel.
 - (iii) *Drying and purification*: In this step, heat treatment is given to the gel emulsion to dry it up to the level necessary for further processing. The reaction mixture is refluxed and the washing is done by suitable agent several times before separating the powder by centrifuging system or spraying it by spin coating unit to the desired shape and size.
 - (iv) *Product formation*: Using different techniques, we can fabricate the particles or films of desired shape and size. For example, the centrifuge action is used to obtain the nanoparticles of zirconia while the thin films can be produced on a piece of substrate by spin-coating or dip-coating.
- Flowchart of the sol–gel method used for the fabrication of nanomaterials is given in Fig. 19.5.

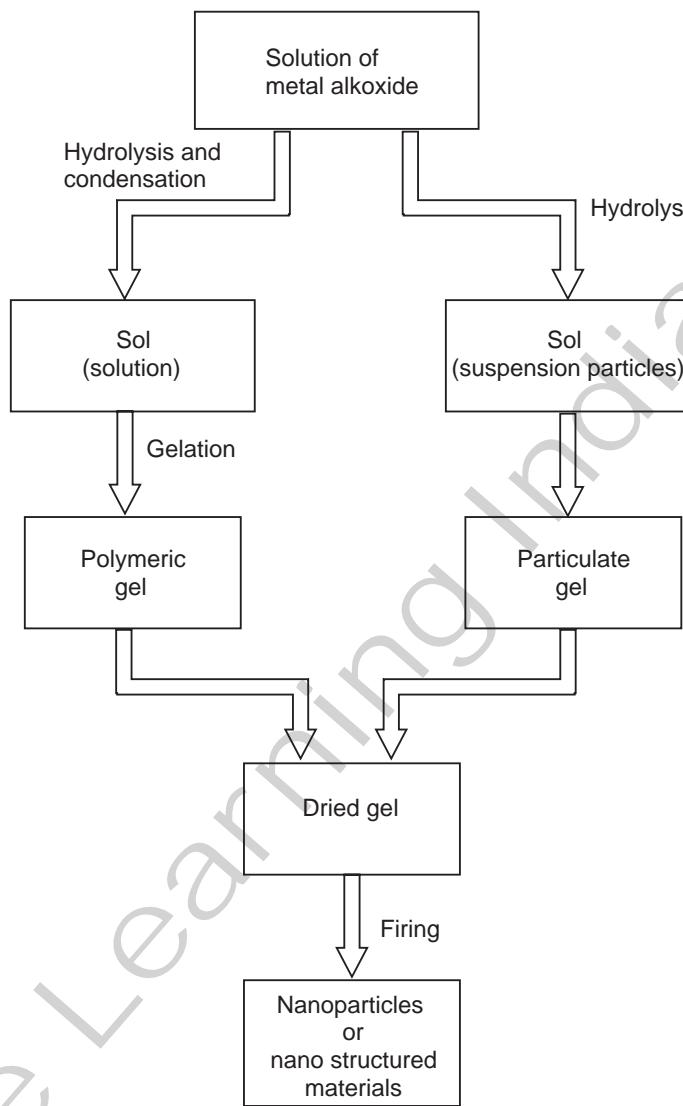


Fig. 19.5 Flowchart of sol-gel method to produce nanomaterials

19.5 QUANTUM SIZE EFFECT IN METAL OR SEMICONDUCTOR NANOPARTICLES

From the study of nanomaterials, it is observed that nanocrystalline particles represent a state of matter in the transition region between the bulk solid and a single molecule. It is also observed that their physical and chemical properties gradually change from solid state to molecular behaviour with decreasing particle size. There are two main reasons due to which there is a drastic change in the properties of materials at nanolevel. According to first, the nanoparticles have a relatively larger surface area as compared to the

same mass of materials produced in a larger form. This can make materials more chemically reactive and affect their electrical properties. Second is the quantum confinement. Actually, quantum effects dominate at nanolevel, which affects the optical, electrical, and the magnetic behaviour of the materials. Size quantisation effect and electronic state transition from bulk to atomic level are shown in Fig. 19.6.

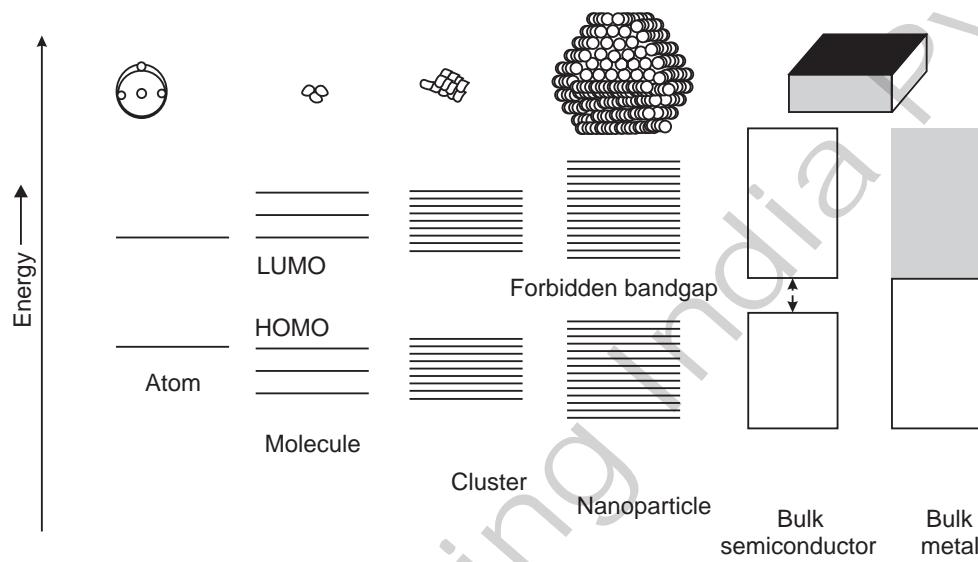


Fig. 19.6 Quantisation effect and electronic state transition from bulk to atomic level

19.6 DIFFERENT TYPES OF NANOSTRUCTURES: (CONFINEMENT DIMENSIONS 0-D, 1-D, 2-D, AND 3-D)

There are different classifications of nanostructures in nanotechnology. Nanostructures are usually classified according to their geometrical dimensions. They usually consist of nanocages, nanocrystallites, nanobelts, nanoneedles, nanocomposites, nanofabrics, nanofibres, nanoflakes, nanoflowers, nanofoams, nanomeshes, nanoparticles, nanopillars, nanopin films, nanorings, nanorods, nanoshells, nanopowders, nanoclusters, nanowires, nanotubes, quantum dots, and quantum ultrastructures. All these nanomaterials can be classified again according to the confinement dimensions or delocalisation dimensions. According to the above mentioned dimensions, there are four classes of quantum structures, namely, bulk conductors, quantum wells, quantum wires, and quantum dots.

A bulk conductor has all its three dimensions more than 100 nm. If one dimension is reduced to the nanorange while the other two dimensions remain large, then we get a structure known as *quantum well*. If two dimensions are reduced and one remains large, the resulting structure is referred to as the *quantum wire*. The extreme case of this process of size reduction in which all three dimensions reach at the nanorange is called *quantum dot*. The delocalisation dimensions of a bulk conductor, a quantum well, a quantum wire, and a quantum dot are $3(x, y, z)$, $2(x, y)$, $1(x)$, and 0 , respectively. But the confinement dimensions of a bulk conductor, a quantum well, a quantum wire, and a quantum dot are 0 , $1(x)$, $2(x, y)$, and $3(x, y, z)$, respectively. Localised and confinement dimensions of bulk, quantum well, quantum wire, and quantum dot are given in Table 19.1.

Table 19.1 Delocalisation and confinement dimensions of quantum nanostructures

S.No.	Quantum structure	Delocalisation dimensions	Confinement dimensions
1.	Bulk conductor	$3(x, y, z)$	0
2.	Quantum well	$2(x, y)$	$1(z)$
3.	Quantum wire	$1(z)$	$2(x, y)$
4.	Quantum dot	0	$3(x, y, z)$

Nanostructures of different dimensions are shown in Figs. 19.7 (a), (b), (c), and (d).

0-D (Delocalisation dimensions)

All dimensions (x, y, z) at nanometric scale.

Example: Nanoparticles

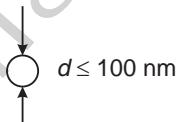


Fig. 19.7 (a) Quantum dot

1-D (Delocalisation dimensions)

Two dimensions (x, y) at nanometric scale; the other dimension is large.

Example: Nanorods, nanotubes

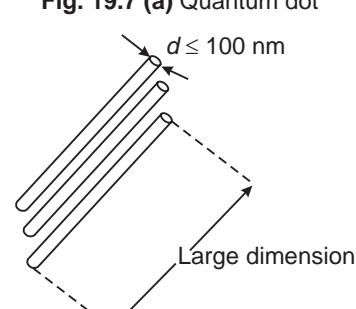


Fig. 19.7 (b) Quantum wire

2-D (Delocalisation dimensions)

One dimension (t measured along z -axis) at nanometric scale; two other dimensions (L_x, L_y) are large.

Example: Thin nanofilms

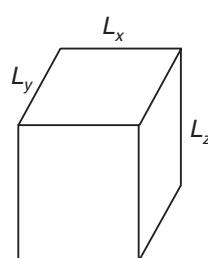


Fig. 19.7 (d) Bulk material

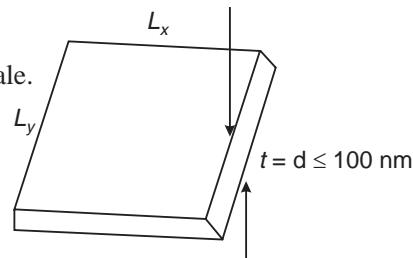


Fig. 19.7 (c) Quantum well

Fig. 19.7 Classification of nanostructures according to different dimensions

19.6.1 Quantum Wells

A potential well of nanometric size is known as quantum well. This *well* is like a cage in which the carrier particles (i.e., the excitons) can be trapped, like light can be trapped between the mirrors. We can create different layers of semiconductors having nanometric dimension to act as traps for excitons. These *trapped particles* can be considered to be in *quantum confinement*.

Due to this quantum confinement, the motion of carriers is restricted. In a quantum well, the excitons can move freely sideways in the plane of a thin layer, but they might like to move in the forward and backward directions as well. Due to the confinement of carriers, the structure quantum well has important applications in making different useful devices. Quantum wells are now widely used to make semiconductor lasers and other important devices. Quantum well is shown in Fig. 19.7 (c).

19.6.2 Quantum Wires

If a nanostructured material has one dimension that is outside the nanometric size range, then it is known as quantum wires. Actually, such a shape of nanomaterials has two dimensions in nanometric range and the third one is large. Nanostructures have a shape like a rod, and consist of nanotubes, nanorods, nanoneedles, and nanowires. The carriers trapped in such structures can be considered to be in 1-D quantum confinement. In this case, an exciton is only free to choose its trajectory along the wire. However, for each motion of its movement, the exciton can have various ways of being confined. This is a little bit tedious for the exciton as well as for us to explain, which creates a great technological challenge to make a defect-free good quantum wire. A simple structure of a quantum nanowire is shown in Fig. 19.7(b).

19.6.3 Quantum Dots

In the method of reducing the size of bulk material upto nanolevel, an extreme case occurs where all three dimensions of bulk are reduced to the low nanometric range. The resulting structure is known as quantum dot. As we have discussed in preceeding sections that a quantum well leads 2-D quantum confinement (delocalisation), and by going one step ahead, in quantum wire, there is 1-D quantum confinement. Now, we can go another step ahead, i.e., the final step, and put the exciton in 0-D quantum confinement in just a tiny dot of a given semiconductor material [Fig. 19.7(a)]. Hence, in this situation, the exciton only has confined states, i.e., there are no freely moving excitons. Now, it is clear that the quantum dots behave in very different and newer ways. Although a quantum dot has many thousands of atoms, but due to its peculiar properties, it is considered more like a single atom rather than many atoms. Due to its specific characteristics, quantum dot system may make the ideal laser material, which is very much focussed. It may be used as a basic building block in making a quantum computer.

19.7 UNUSUAL PROPERTIES OF NANOMATERIALS

It is very difficult to mention all unusual properties of nanomaterials because there is a dramatic change in the properties of individual bulk materials when they are reduced to the size below 100 nm. We can imagine the unusual properties of nanomaterials by taking the following examples into consideration.

Example 1: Melting point of gold decreases from 1200 K to 800 K when the particle size is reduced from 300 Å to 20 Å.

Example 2: In metals, if the size is reduced from the bulk, the electronic band becomes narrower and the value of ionisation potential is raised.

20

New Engineering Materials

20.1 INTRODUCTION

The human quest for material started with the beginning of civilisation. Transportation, housing, clothing, communication, recreation, and food production, virtually almost every segment of our everyday lives is highly influenced by the materials. With the development and advancement of societies, the scientists and technocrats have made consistent effort to produce and manipulate the materials. The search for new and advanced materials has been the major preoccupation of materials scientists during the past several decades. Recent investigations have focussed on the improvement of the properties and performance of existing materials or synthesising and developing completely new materials to cater the need of society. Significant improvements have been achieved in the mechanical, chemical, and physical properties of materials by the addition of alloying elements, by microstructural modification, and by subjecting the materials to thermal, mechanical, or thermo-mechanical processing methods. Now, in these days completely new materials, unheard of earlier, have also been synthesised. Some of them are metallic glasses, quasicrystals, shape memory alloys, nanocrystalline, active, and smart materials, high-temperature superconductors, etc.

In this chapter we will describe metallic glasses, types, properties, preparation, and their applications; shape memory alloys, types, shape memory effect, pseudo elasticity, properties, and their applications; and biomaterials, general information, biomedical compatibility of Ti-Al-Nb alloys, and their applications.

20.2 METALLIC GLASSES

Most of the metallic metals and alloys are usually crystalline in nature possessing translational symmetry, i.e., their constituent atoms are arranged in a regular and periodic manner in three dimensions. The regular arrangement of atoms in the crystalline materials can be directly viewed using a transmission electron microscope (TEM). Many of the important properties of engineering alloys can be explained in terms of their crystalline order or in many cases, in terms of the crystal defects in the crystal structure.

Metallic glasses, in contrast, are alloys that are noncrystalline or amorphous. In these materials, there is no long-range atomic order and atoms are more or less randomly arranged. The real world of metallic glasses came into existence with the revolution in the concept of metals during 1960, after the successful synthesis of an Au–25 at. % Si alloy in the glassy state by rapid solidification of the liquid at the rates approaching a million degrees per second. Investigations on metallic glasses and other alloys rapidly

solidified from the liquid state at the rates of about 105–106 K/s have unambiguously confirmed that these alloys are truly amorphous and lack crystallinity. The amorphous nature of the quenched material was confirmed by transmission electron microscopy (TEM) and electron diffraction techniques, in addition to the XRD methods. Besides these methods, several other advanced techniques have also brought the same conclusion later. Since these amorphous materials are based on metals, these were referred to as glassy metals or *metallic glasses*.

20.3 TYPES OF METALLIC GLASSES

Since the discovery of a metallic glass in 1960, hundreds and thousands of alloys of different compositions have been prepared as metallic glasses. Due to their unique properties, a large variety of metallic glasses have been developed during the last nearly 50 years. These can be broadly classified into metal–metalloid or metal–metal type, even though other classifications also exist.

20.3.1 Metal–Metalloid Type Metallic Glasses

In a typical metal–metalloid type metallic glass, the metal atoms constitute about 80% and the metalloid atoms (typically B, C, P, and Si) about 20%. In the constituent of these bulk metallic glasses (BMGs), the metal atoms may be of one type or a combination of different metals, but the total amount of metal atoms is about 80%. Similarly, the metalloid atoms may be of one type or a combination of different metalloid atoms, again the total amount of metalloid atoms is about 20%. Some well investigated examples of this category include $Pd_{80}Si_{20}$, $Pd_{77}Cu_6Si_{17}$, $Fe_{80}B_{20}$, $Fe_{40}Ni_{40}B_{20}$, $Ni_{75}Si_8B_{17}$, $Fe_{40}Ni_{40}P_{14}B_6$, $Fe_{70}Cr_{10}P_{13}C_7$, $Ni_{49}Fe_{29}B_6P_{14}Si_2$, and some exotic compositions such as $W_{35}Mo_{20}Cr_{15}Fe_5Ni_5P_6B_6C_5Si_3$ (the subscripts denote the atomic percentages of the elements in the alloy).

20.3.2 Metal–Metal Type Metallic Glasses

In the metal–metal type metallic glasses, only metal type of atoms are involved. Some of the alloys that have been well investigated in this category include $Ni_{60}Nb_{40}$, $Cu_{57}Zr_{43}$, $Mg_{70}Zn_{30}$, $La_{80}Au_{20}$, and $Fe_{90}Zr_{10}$. Anyone can immediately notice the difference between these two classes of metallic glasses only by observing their compositions. In the metal–metalloid type metallic glasses, the metalloid content is usually around 20 at. % and rest are metals, whereas in metal–metal type metallic glasses, there is no such compositional restriction. In the case of metal–metal type metallic glasses, the content of second metal component can be as small as 9–10 at. % or as large as nearly 50 at. %.

20.4 PREPARATION OF METALLIC GLASSES

The revolution in the concept of metals got new dimension with the successful synthesis of the glassy metals. The first metallic glass was synthesised by Pol Duwez at the California Institute of Technology in Pasadena, California, during 1960. He synthesised an Au–25 at. % Si alloy in the glassy state by rapid solidification of the liquid at the rates approaching a million degrees per second.

These high solidification rates were achieved by a special technique in which the small droplets of the liquid alloys are propelled tangentially onto a highly conducting substrate such as copper or silver to enable the liquid to spread out in the form of a thin foil on the substrate surface. The good thermal contact with the substrate ensures that the heat is extracted rapidly by the large substrate through the small thickness of the foil. This technique is known as the “*gun technique*” of liquid quenching, in which

the solidification rates were estimated to vary from as low as 10^4 K/s to as high as 10^{10} K/s depending on the thickness of the foil, the nature of the substrate, the type of material solidified, and how good the thermal contact was between the foil and the substrate. A typical solidification rate for a foil of $50\text{ }\mu\text{m}$ thickness is about 10^6 K/s.

20.5 METHODS TO SYNTHESISE METALLIC GLASSES

Metallic glasses and noncrystalline materials can be prepared by a variety of methods. Some important methods such as vapour-state processes, liquid-state processes, or solid-state processes are briefly described in the following subsections.

20.5.1 Vapour-State Processes

Vapour-state process is the earliest method of formation of an amorphous phase in a metal. Physicists produced amorphous metals by the vapour deposition route to study their superconducting properties. According to the available reports, the German physicist Kramer synthesised amorphous Sb by first evaporating the Sb metal and then allowing it to condense on a cold substrate. After the successful synthesis of amorphous Sb, different scientists synthesised thin films of Bi, Ga, and Sn–Bi alloys by the same process. Even though they believed that their films were ultrafine-grained from the *in situ* electron diffraction experiments, it became clear later that the films were actually amorphous. There is still debate in the literature whether the vapour-deposited pure metals Sb and Bi were truly amorphous or microcrystalline. Irrespective of whether they were amorphous or not, this technique is presently used to synthesise amorphous materials in small quantities.

20.5.2 Liquid-State Processes

Metallic glasses were usually prepared by rapidly solidifying their metallic melts at solidification rates of about 10^6 K/s, which can be achieved using the techniques such as melt spinning and its variants. These processes of rapid solidification processing (RSP) result the products in the form of thin ribbons, even though wires, powders, and other shapes have also been produced. In order to form the glassy structure in metallic melts, it is essential to cool the melt with very high solidification rates to prevent the formation of crystal nuclei. Consequently, the minimum cooling rate required to form the glass (critical cooling rate) is about 10^5 – 10^6 K/s. Such high solidification rate can be attained only when the heat is removed from the melt rapidly, and as a result, the section thickness is limited to a few tens of micrometres.

With the proper choice of the alloy system and its composition, it is possible to reduce the critical cooling rate for glass formation. Consequently, the section thickness may increase from a few micrometres (20 – $50\text{ }\mu\text{m}$) in the RSP alloys to a few millimetres or even a few centimetres. The metallic glasses as described above are referred to as BMGs.

20.5.3 Solid-State Processes

Solid-state process is another popular technique frequently used for the preparation of metallic glasses. A number of different methods in this major category are available, and the most commonly used method amongst these is mechanical alloying or milling. *Mechanical alloying* refers to the process which involves putting the required amounts of the elemental or pre-alloyed powders in a hardened steel container along with hardened steel balls and agitating the whole mixture for the required period of time. Alloying occurs between the metal powders and can result in the formation of a supersaturated solid solution, an inter-

metallic phase, a quasicrystalline phase, or an amorphous alloy, depending on the powder composition and milling conditions. The choice of the phase formed depends on the thermodynamic stability of the different competing phases.

20.6 PROPERTIES OF METALLIC GLASSES

Due to their unique structures, metallic glasses have a variety of applications in engineering. Since the structure of a material determines its properties, one might expect that a material with an unusual structure might have interesting properties. This is certainly true for the metallic glasses. It is observed that metallic glasses including BMGs have unique properties. They are typically much stronger than crystalline metal counterparts by a factor of 2 to 3. They are quite tough (more than ceramics) and they have very high strain limits for Hook's law having materials as shown in Fig. 20.1.

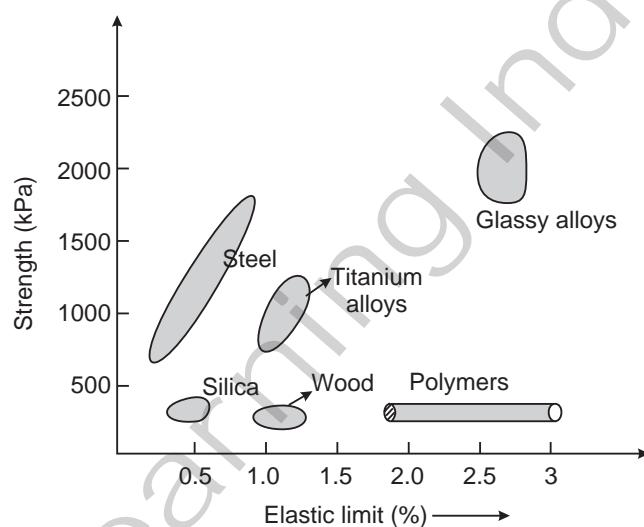


Fig. 20.1 Typical strengths and elastic limits for various materials

From Fig. 20.1, it is clear that glassy alloys have unique strengths and elastic limits due to which they are used in specific engineering applications. Besides this, metallic glasses possess several other superior properties as compared to crystalline metals. Some important properties of metallic glasses are as follows:

- (i) They have high strength; they are 3 times stronger than steel.
- (ii) They bear high springing nature—about 10 times that of best-quality industrial steel.
- (iii) They have high strength to weight ratio.
- (iv) They are very hard.
- (v) They are highly temperature-sensitive and become soft on heating, and therefore are easily malleable.
- (vi) They bear extreme springiness and rebound characteristics.
- (vii) They have good acoustic damping properties.
- (viii) They are highly elastic.
- (ix) They are very tough.

20.7 APPLICATIONS OF METALLIC GLASSES

Different material systems of metallic glasses and their specialties are given in Table 20.1.

Table 20.1 Different metallic glasses with their constituents and specialities

S. No.	Material systems	Specialities
1.	Cu–Zr	Binary alloy
2.	Cu–Zr–Al	Ternary alloy
3.	Cu–Zr–Al–Y	Quaternary alloy
4.	Ta–Cu–Ni–Al	Plastic strain= 4.5%
5.	Pt–Cu–Ni–P	Plastic strain= 20%
6.	Pd ₄₀ Ni ₄₀ P ₂₀	Palladium–nickel based alloy
7.	Fe ₄₀ Ni ₄₀ P ₁₄ B ₆	Iron–nickel based alloy
8.	Zr _{41.2} T _{13.8} Cu _{12.5} Ni ₁₀ Be _{22.5}	Vetreloy used for golf-club heads
9.	Al ₉₀ Fe ₂₅ Ce ₅	Aluminium–iron–cesium based alloys
10.	La ₅₅ Al ₂₅ Ni ₂₀	Lanthanide based alloy

These materials have potential applications in science, technology, industry, frontier areas of researches, and material sciences. Some of their important applications are as follows:

1. In transformers
2. In magnetic head recorders
3. In metallic glass ribbon for different uses
4. In resistant coatings
5. In magnetic shielding
6. In electric motors
7. In the engine parts
8. In military applications such as armor-piercing projectiles
9. In electronics

20.8 BULK METALLIC GLASSES

Bulk metallic glasses are those noncrystalline solids obtained by continuous cooling with slow solidification rate from the liquid state which have a section thickness of at least a few millimetres. In general, the metallic glasses with at least a diameter or section thickness of 1 mm are considered “bulk.” But in these days the minimum size limit of BMGs has been extended to 10 mm. BMGs have the following four important characteristics:

- (i) The alloy systems have a minimum of three components due to which they are frequently referred as multicomponent alloy systems. The binary BMGs are also reported but the maximum diameter of the rod which can be obtained in a fully glassy condition is usually reported to be 1 or 2 mm.
- (ii) Bulk metallic glasses can be produced by slow solidification rates, typically 10³ K/s or less. The lowest solidification rate at which BMGs have been obtained is reported as 0.067 K/s, that is, 4 K/min.

- (iii) BMGs exhibit large section thicknesses or diameters, a minimum of about 1 mm. The largest diameter of a bulk metallic glass rod produced till date is 72 mm in a $Pd_{40}Cu_{30}Ni_{10}P_{20}$ alloy.
- (iv) BMGs exhibit a large supercooled liquid region. The difference between the glass transition temperature and the crystallisation temperature is large, usually a few tens of degrees, and the highest reported value so far is 131 K in case of $Pd_{43}Ni_{10}Cu_{27}P_{20}$ alloy.

20.9 IMPORTANT APPLICATIONS OF BULK METALLIC GLASSES

Due to the attractive physical and chemical properties with the advantages of a large section thickness, BMGs have limitless applications in the fields of industries, medicals, and researches. Some important applications of BMGs are as follows:

(i) Chemical applications

The superior corrosion resistance of BMGs over their crystalline counterparts plays a major role in this group of applications. BMGs are also the most appropriate materials for fuel-cell separators.

(ii) Magnetic applications

The magnetic properties of melt-spun glassy ribbons have been exploited for a variety of applications. The outstanding soft magnetic properties of the Fe-based melt-spun ribbons have found applications in power distribution transformers and several other applications. The present-day electronic devices are going down in size, and miniaturisation is reaching micrometre and even nanometre levels. Micro-electromechanical systems (MEMS) and nano-electromechanical systems (NEMS) are the operative words of today.

(iii) Miscellaneous applications

The high strength and wear resistance of BMGs along with their biocompatibility, smooth surface finish, and aesthetic appearance helps in determining their appropriate applications.

(iv) Jewellery applications

Jewellery has fascinated humankind for several hundreds of years. In a survey, it has been reported that in more than two-third jewelleries, gold, silver, platinum, and diamonds are used. Due to the special attributes of BMGs based on Au and Pt, it is possible to develop these alloys for jewellery applications.

(v) Biomedical applications

Many metallic materials are used as implants in the human body. Because, these materials are biocompatible and inert (having high corrosion resistance), and have mechanical properties that are very similar to the part that is being replaced.

(vi) Medical devices

As mentioned in the above applications, BMGs are corrosion resistant and biocompatible. The above features are taken into consideration while using BMGs to develop products for reconstructive devices, fractured fixations, spinal implants, ophthalmic surgery, and cataract surgery. Vitreloys (trade name for Zr-based BMG alloys manufactured by Liquidmetal Technologies) are being used for materials in knee-replacement devices, pacemaker casings, etc.

20.10 DISTINCTION BETWEEN CRYSTALS AND GLASSES

Solids (i.e., metals and alloys) are traditionally considered crystalline in nature, in which constituent atoms are arranged in a periodic manner in three dimensions (thus, the periodicity is the same in the three directions). In other words, if we know the coordinates of one atom in the crystal, and the size and shape of the unit cell (i.e., periodicity in the three directions), then it is possible to predict the position of other atoms. Further, the concept of space lattice requires that every atom in the crystal have identical surroundings, i.e., the nearest neighbor distances and the coordination number (CN).

The solid in which there is no regular arrangement of atoms, that is, periodicity, is considered “non-crystalline” in character. It means, these materials do not possess any crystallinity. These solids are also known as “amorphous” or “glassy” materials. Since the atomic arrangement is random (i.e., there is no periodicity) in these noncrystalline materials, it will be difficult to define either the nearest neighbor distances or the coordination number. Each atom in the noncrystalline solid will have different nearest neighbor distances and CNs. But it can be safely stated that the nearest neighbor distances are longer and the CNs smaller in a noncrystalline solid in comparison to its crystal counterpart, assuming that a solid can exist both in the crystalline and the noncrystalline states.

20.11 DIFFERENCES BETWEEN AMORPHOUS ALLOYS AND METALLIC GLASSES

The terms noncrystalline, amorphous, or glassy refer to similar (random) atomic arrangements in solid materials, and therefore, these terms have been used interchangeably (quite understandably so) in the literature. Sometimes, the use of these terms leads to some confusion. In the literature it is observed that some researchers have been preferably using the term “amorphous”, while others have been using “glassy”, and a few others “noncrystalline”. Additionally, some researchers refer to the thin ribbon glassy materials as amorphous and the bulk glassy alloys only as glasses. Thus, several terms have been used to describe these noncrystalline materials. To avoid the ambiguity and confusion, we would like to clarify each of these terms where it should be used. It has been generally agreed upon that noncrystalline is a generic term used to describe any solid material that does not possess crystallinity. A noncrystalline solid formed by continuous cooling from the liquid state is known as a “glass”. While a noncrystalline material obtained by any other process such as vapour deposition or solid-state processing method is referred as “amorphous” material. But it is worth to mention that both amorphous materials and glasses are noncrystalline.

20.12 SHAPE MEMORY ALLOYS

Most of the solids, especially metals, are characterised by their physical properties such as tensile strength, malleability, and conductivity. In the case of shape memory alloys (SMAs), we can add the anthropomorphic qualities of memory and trainability. The shape memory alloys exhibit the shape memory effect. If these alloys are plastically deformed at one temperature, they will completely recover their original shape when their temperature is raised to a higher limit. It is observed that during the recovery of their shape, the alloys can produce a displacement or a force as a function of temperature. In many alloys, a combination of both is possible. Metals can change shape, change position, pull, compress, expand, bend, or turn, with heat as the only activator. The special features of products that possess this shape memory property are as follows:

- (i) A high force during shape change
- (ii) Large movement with small temperature change

- (iii) A high permanent strength
- (iv) Simple application, because no special tools are required; many possible shapes and configurations
- (v) Easy to use

Due to these specific properties, shape memory alloys are helping to solve a wide variety of problems. SMAs are also used in a wide range of medical and dental applications.

SMAs are novel and special materials; they can remember their shape when heated above a certain temperature. They are also known as muscle wires.

For the first time the shape memory behaviour was reported by Olander in his study of “rubber like effect” in samples of gold–cadmium during 1932. Again it was confirmed by Greninger and Mooradian in their study on brass alloys (copper–zinc) in 1938. Later in 1951, Chang and Read first reported the term “*shape recovery*”. They were also working on gold–cadmium alloys. In 1962, William J. Buehler and his co-workers at the Naval Ordnance Laboratory discovered shape memory effect in an alloy of nickel and titanium. They named it NiTiNOL (for nickel–titanium Naval Ordnance Laboratory).

20.13 PHASE CHANGE IN SHAPE MEMORY ALLOYS

Shape memory alloys are a special class of metal alloys that can remember their shape when heated above a certain temperature. SMAs change shape based on a solid-sate phase transformation, which are atomic level changes (rearrangement of atoms). The change in shape occurs at specific temperature (shape memory effect). Shape memory alloys have two stable phases: the high-temperature phase, called *austenite*, and the low-temperature phase, called *martensite*.

The main characteristic of all shape memory alloys is the occurrence of a martensitic phase transformation, which is a phase change between two solid phases through the rearrangement of atoms within the crystal lattice. The martensitic transformation is related to the inelastic deformation of the crystal lattice where no diffusive process is involved. This phase transformation is the result of collective motion of atoms on distances smaller than the lattice parameters. The growing speeds of martensite plates can approach the sound velocity in the metal (up to 1100 m/s). Martensitic transformation can occur at low temperatures where atomic mobility may be very small, resulting in the absence of diffusion. Due to the absence of diffusion, the martensitic phase transformation is almost instantaneous. This transformation is also known as the first-order transition. When a shape memory alloy undergoes a martensitic phase transformation, it transforms from its high-symmetry (usually cubic) austenitic phase to a low-symmetry martensitic phase (highly twinned monoclinic structure).

The number of equivalent shear directions through which the martensite can form within a region of parent-phase is the main feature of all martensitic transformations. This results in the formation of martensite variants within the microstructure of a transformed alloy. Two crystallographically equivalent martensite variants (twinned and detwinned martensite) created by different atomic shears from the parent phase have been shown in Fig. 20.2. The macroscopic shape of the crystal block is maintained by two opposite shears. The microstructure in which the shear of one variant is accommodated or “cancelled” by the other is known as a self-accommodated structure.

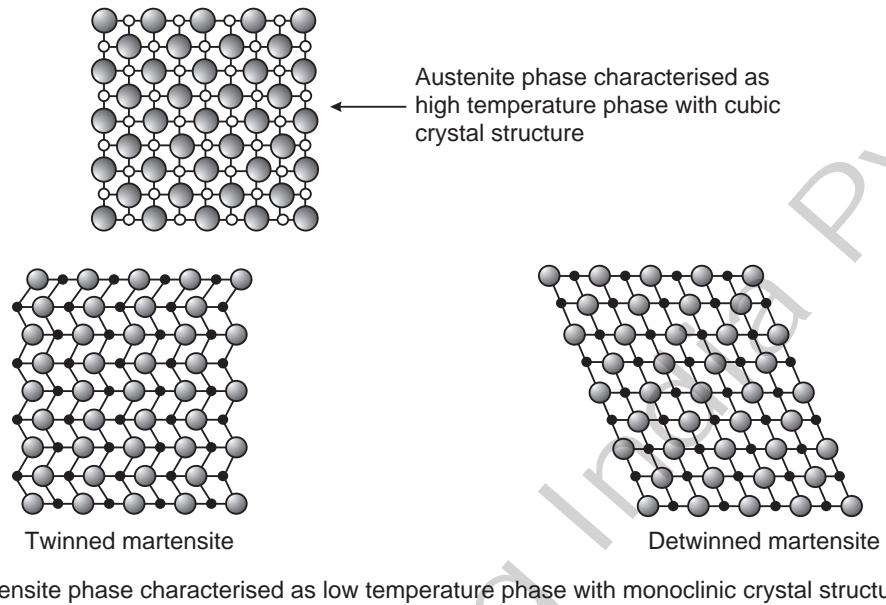


Fig. 20.2 Different phases of a shape memory alloy

The process of temperature-induced phase transformation without the mechanical loading of shape memory alloy has been shown in Fig. 20.3. Upon cooling without the application of mechanical loading,

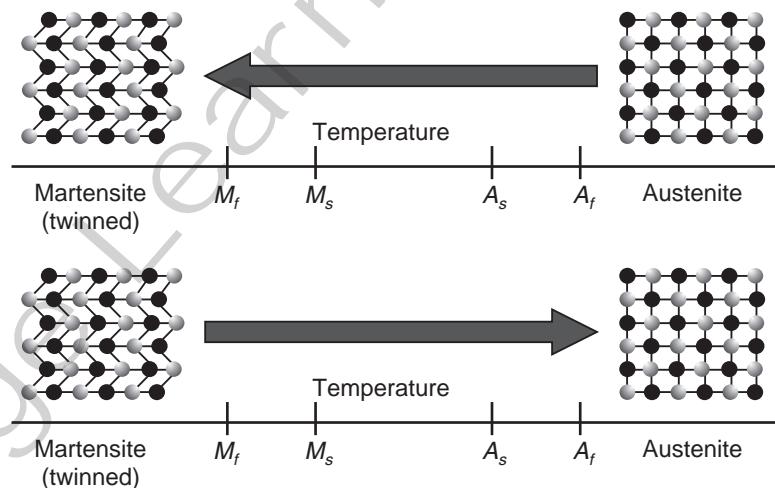


Fig. 20.3 Phase transformation of a shape memory alloy, induced by temperature without mechanical loading

the material transforms from austenite into twinned martensite. With the same condition of mechanical loading when twinned martensite is heated, a reverse martensitic transformation takes place and as a

result the material transforms into austenite. In this process there are four critical temperatures which are as follows:

- (i) M_s (martensitic start temperature): It is the temperature at which the material starts transforming from austenite to martensite.
 - (ii) M_f (martensitic finish temperature): It is the temperature at which the transformation is complete and the material is fully in the martensitic phase.
- Similar temperatures are defined for reversible transformation.
- (i) A_s (austenite start temperature): It is the temperature at which the material starts transforming from martensite to austenite.
 - (ii) A_f (austenite finish temperature): It is the temperature at which the transformation is complete and the material is fully in the austenite phase.

The hysteresis curve obtained for a thermoelastic martensitic transformation has been shown in Fig. 20.4. The overall hysteresis between forward and reverse transformation pathways in shape memory alloys is small, and typically lies between 10°C and 50°C. The behaviour of shape memory alloys is based on the thermoelastic martensitic transformations. In Fig. 20.4, ξ_0 and T_0 are the prescribed initial conditions for a given shape memory alloy. They represent the condition that the material contains some martensite (ξ_0) and some austenite ($1 - \xi_0$) at a temperature T_0 . This can be repeated definitely as long as high temperature excursions are avoided.

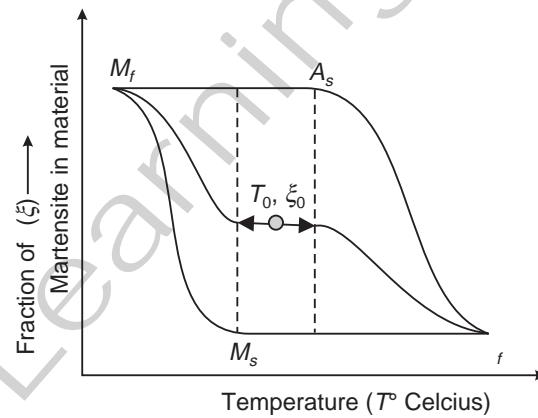


Fig. 20.4 Temperature-induced phase transformation of a shape memory alloy without mechanical loading

20.14 ONE-WAY MEMORY EFFECT

One way memory effect can be discussed at macroscopic and microscopic scales, which are described below.

20.14.1 Macroscopic Scale

If a self-accommodated martensite alloy deformed by applying mechanical load and then unloaded, remains deformed, but when reheated to a temperature above the austenite finish temperature, recovers the original macroscopic shape, then the alloy is known as *one-way shape memory alloy* and this effect

21

Elementary Modern Physics

21.1 INTRODUCTION

New era in the development of modern physics (i.e., wave mechanics) started with the revolutionary idea of Max Planck, which states that electromagnetic radiation may be considered as a stream of small energy packets called photons. Wave mechanics deals with the microscopic properties of materials. Before wave mechanics, phenomena such as photoelectric effect, Compton effect, blackbody radiation, atomic spectra, and so on could not be explained. Wave mechanics was able to correctly explain these phenomena. Macroscopic phenomena such as interference, diffraction, and polarisation firmly confirm the wave character of electromagnetic radiation which may have particle nature as well. In other words, electromagnetic radiation has dual character; in certain situation it exhibits properties of wave, while in others it acts as a particle.

In this chapter, we will discuss blackbody radiation spectrum, Planck's hypothesis and radiation law, Compton's effect, and photoelectric effect in context of old (classical) and modern (quantum) theory. In the same chapter, we will also study the wave–particle duality along with the concept of phase velocity and group velocity, and matter-waves.

21.2 BLACKBODY

A blackbody is a body which absorbs radiations of all wavelengths incidenting on it. Blackbody neither reflects nor transmits any of the incident radiation, and therefore, appears black. In actual practice, there is no body as a perfect blackbody, but lamp black may be regarded as blackbody because it is able to absorb nearly 99% of the incident radiations. Scientist Ferry suggested a blackbody which consisted of a hollow thick-walled sphere, painted lamp black internally and provided a small circular opening to enter the radiations. In order to prevent the direct reflection from the inner surface, there is the provision of projection in front of the opening of the blackbody. When any radiation enters the blackbody region through the opening, it suffers multiple reflection inside the sphere and is finally absorbed. When the walls of such a cavity are heated to a temperature T , it emits the radiations which fill the cavity and it comes out through the opening of cavity. These radiations are known as *blackbody radiations*, which are the characteristic of its temperature. Here, it is worth to mention that the thermal radiations and the light radiations are identical, except the difference in their wavelengths.

21.2.1 Blackbody Radiation Spectrum

Among the scientists, it was a matter of great interest that how the energy was distributed in the different wavelengths in the spectrum of blackbody radiation. In this regard, the first important affect was made by Stefan and Boltzmann during 1884. According to them, the energy of radiation in unit volume of space due to all the wavelengths in the spectrum is proportional to the fourth power of the absolute temperature of the blackbody. Although this law has been verified experimentally, but it is unable to give any idea about the energy distribution among the different wavelengths of the blackbody radiation.

Lummer and Pringsheim investigated the distribution of energy amongst the different wavelengths in the thermal spectrum of a blackbody radiation. The results obtained by Lummer and Pringsheim are shown in Fig. 21.1. Here the graph is drawn between emissive power and wavelength for various temperatures.

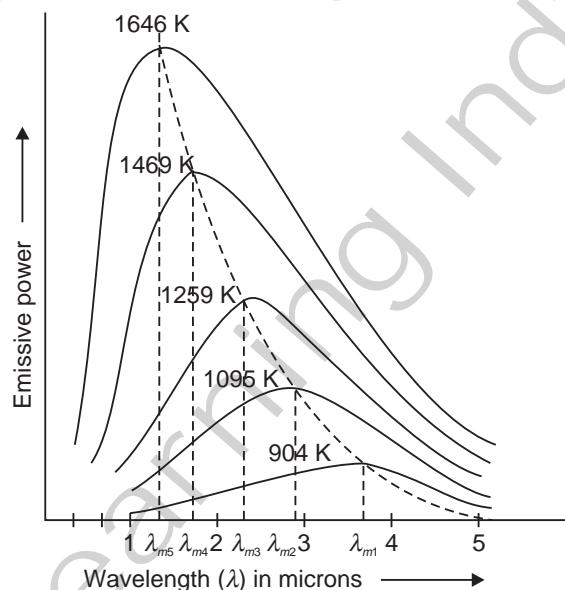


Fig. 21.1 Spectrum of blackbody radiation according to Lummer and Pringsheim

Following are the important findings of Lummer and Pringsheim:

- At a given temperature, the energy is not uniformly distributed in the radiation spectrum of the blackbody.
- At a given temperature, the intensity of radiation increases with increase in wavelength and at a particular wavelength, its value is maximum. With further increase in wavelength, the intensity of radiation decreases.
- With the increase in temperature, λ_m decreases, where λ_m is the wavelength at which the maximum emission of energy takes place. The points on the dotted line represent λ_m at various temperatures.
- There is increase in energy emission with the increase in temperature corresponding to all the wavelengths.
- The area under each curve gives the total energy emitted for the complete spectrum at a particular temperature. With the increase in temperature, this area increases. It is also observed that the area is directly proportional to the fourth power of the temperature of the blackbody, i.e., $E \propto T^4$.

21.3 WIEN'S DISPLACEMENT LAW

According to Wien's displacement law, the product of the wavelength corresponding to maximum energy and the absolute temperature is constant, i.e.,

$$\lambda_m \cdot T = \text{constant} = 0.2892 \text{ cm K} \quad (21.1)$$

This is known as Wien's displacement law. It shows that with the increase in temperature, λ_m decreases.

Wien also suggests that the energy E_{\max} is directly proportional to the fifth power of the absolute temperature

$$\begin{aligned} \text{i.e.,} \quad E_m &\propto T^5 \\ \text{or} \quad E_m &= \text{constant} \times T^5 \end{aligned} \quad (21.2)$$

Using certain assumptions, Wien deduced the radiation law for the energy density in the wavelength range λ and $\lambda + d\lambda$ at a temperature T , which is given as

$$E_\lambda d\lambda = \frac{c_1 \lambda^{-5}}{e^{c_2/\lambda T}} d\lambda \quad (21.3)$$

where c_1 and c_2 are the constants.

Equation (21.3) gives the Wien's law of distribution of energy.

This law is applicable only for short wavelengths of radiation.

21.4 RAYLEIGH-JEANS LAW

According to Rayleigh, the distribution of energy in radiation spectrum is given by the formula

$$E_\lambda d\lambda = \frac{8\pi kT}{\lambda^4} d\lambda \quad (21.4)$$

where k is Boltzmann's constant.

It is observed that Wien's formula agreed with the experimental results of Lummer and Pringsheim, in the region of shorter wavelength, while the Rayleigh–Jeans formula agreed with long wavelengths. As none of these formulae could account for the entire shape of the radiation curve, therefore, Paschen suggested that the fundamental assumptions of the classical theory were at fault. This exactly occurred to Planck who proposed a new revolutionary hypothesis in 1901 known as theory of quanta. In order to explain the distribution of energy in the radiation spectrum, Planck's theory of quanta is used successfully.

21.5 PLANCK'S HYPOTHESIS AND RADIATION LAW

In order to derive the theoretical expression for the energy distribution in the spectrum of blackbody, Planck proposed the following hypothesis in 1901:

- A chamber containing blackbody radiations also contains simple harmonic oscillators of molecular dimensions which can vibrate with all possible frequencies.
- The frequency of radiation emitted by an oscillator is the same as the frequency of its vibration.
- The frequency of radiation emitted by an oscillator is not continuous but it is only in the multiples of a small unit called quantum (photon). As the energy of photon is $h\nu$, the energy emitted is equal to $h\nu, 2h\nu, 3h\nu, \dots$, so on.

- (iv) The oscillators can emit or absorb radiation energy in packets of $h\nu$. It means that the exchange of energy between the radiation and the matter is not continuous but is limited to a discrete set of values: $0, h\nu, 2h\nu, 3h\nu, \dots, nh\nu$, i.e., $E_n = nh\nu$, where n is an integer and h is Plank's constant (6.625×10^{-27} erg s)

On the basis of the above assumptions, Planck deduced the formula

$$E_\lambda d\lambda = \frac{8\pi hc}{\lambda^5 \left[e^{\frac{hc}{\lambda kT}} - 1 \right]} d\lambda \quad (21.5)$$

where c is the velocity of light and h is Planck's constant. Equation (21.5) agrees with all the experimental results (for short as well as for long wavelengths) obtained by Lummer and Pringsheim.

Deductions: Using suitable conditions, Wien's law and Rayleigh-Jeans law can be deduced from Eq. (21.5).

- (i) For shorter wavelengths,

$$e^{\frac{hc}{\lambda kT}} \gg 1$$

From Eq. (21.5), we know that

$$\begin{aligned} E_\lambda d\lambda &= 8\pi hc \lambda^{-5} e^{-\frac{hc}{\lambda kT}} d\lambda \\ \text{or } E_\lambda d\lambda &= c_1 \lambda^{-5} e^{-\frac{c_2}{\lambda T}} d\lambda \\ \text{where } c_1 &= 8\pi hc \text{ and } c_2 = \frac{hc}{k} \end{aligned} \quad (21.6)$$

Equation (21.6) represents Wien's radiation law.

- (ii) Expanding $\frac{hc}{kT}$ and neglecting higher powers in Eq. (21.5), we get

$$\begin{aligned} E_\lambda d\lambda &= \frac{8\pi hc}{\lambda^5 \left[1 + \frac{hc}{kT} - 1 \right]} d\lambda \\ &= \frac{8\pi hc \lambda kT}{\lambda^5 \cdot hc} d\lambda \\ \text{or } E_\lambda d\lambda &= \frac{8\pi kT}{\lambda^4} d\lambda \end{aligned} \quad (21.7)$$

Equation (21.7) represents Rayleigh-Jeans law.

Hence, Planck's formula for the energy distribution in a radiation spectrum is applicable for all wavelengths.

21.6 PHOTOELECTRIC EFFECT

The emission of electrons from the metal surface when illuminated by light or any other radiation of suitable wavelength (or frequency) is known as *photoelectric effect*. The emitted electrons are known as *photoelectrons* and the current constituted by these electrons is known as *photocurrent*.

21.6.1 Experimental Study

Simple arrangement used for the study of photoelectric effect is shown in Fig. 21.2.

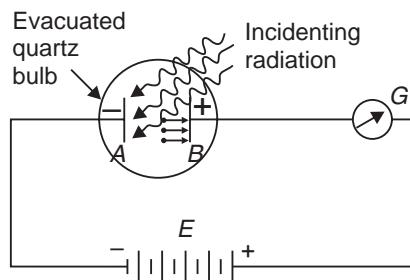


Fig. 21.2 Experimental arrangement for the study of photoelectric effect

In this arrangement, A and B are two photosensitive surfaces, which are enclosed in the evacuated quartz bulb. Plate A is connected to the negative terminal of the battery while plate B is connected to the positive terminal through a galvanometer G . In the absence of any light, there is no deflection in the galvanometer. But when a beam of radiation is allowed to incident on the plate A , a current starts to flow in the circuit due to which the galvanometer shows deflection. The flow of current (photocurrent) ensures the fact that when radiations of suitable wavelength are allowed to incident on a photosensitive surface, electrons are ejected. The number of electrons emitted and their kinetic energy depend upon the following factors:

- The potential difference between plates A and B .
- The intensity of incident radiations.
- The frequency of incident radiations.
- The photosensitive metals or plates used.

21.6.2 Fundamental Laws of Photoelectric Emission

On the basis of the potential difference between plates A and B , the intensity of incident radiation, its frequency, and the effect of photometals, following fundamental laws can be deducted:

- The number of electrons emitted per second, i.e., photoelectric current is proportional to the intensity of the incident light.
- For the ejection of photoelectrons from the surface of a given material, certain minimum frequency ($h\nu$) of the incident light (radiation) is required. If the frequency is less than the minimum frequency, no electron can be emitted from the metal surface however intense the incident light may be. The minimum value of the frequency is known as *threshold frequency* and the corresponding wavelength is known as *threshold wavelength*.
- The kinetic energy of the photoelectrons depends on the frequency of the incident radiation and not on the intensity. The kinetic energy of the photoelectrons is proportional to the frequency of the incident radiation.
- The phenomenon of photoelectric effect is different from thermionic emission because the rate at which the electrons are emitted from the photo-cathode is independent of its temperature.

- (v) The time-lag between the incidenting radiation and the ejected photoelectron from the photosensitive surface is very small (less than 10^{-8} s). This instantaneous emission continues as long as the frequency of the incident radiation is greater than the threshold frequency.
- (vi) The stopping potential (the negative potential of B at which the photoelectric current is zero) is proportional to the frequency of the incidenting radiation but is independent of its intensity.

21.6.3 Einstein's Photoelectric Equation

On the basis of Planck's revolutionary idea of photon theory, Einstein suggested a successful explanation of the photoelectric effect during 1905. According to Einstein's explanation, in the phenomenon of photoelectric effect, one photon is completely absorbed by one electron of the metal (photosensitive), thus, the electron gains the quantum of energy and may be emitted from the metal surface. Hence, according to this explanation, the photon's energy is utilised in the following two ways:

- (i) A part of its energy is used to free the electron from the atom and to move it from the metal surface. The energy used in this task is known as *photoelectric work function* of the metal. It is denoted by W .
- (ii) Rest (other part) of the energy is used to give the kinetic energy ($1/2 m v^2$) to the ejected photoelectrons. Thus, the total energy ($h\nu$) of the incidenting photon is given as

$$h\nu = W_0 + \frac{1}{2} m v^2 \quad (21.8)$$

where v is the velocity of the ejected photoelectron. Equation (21.8) is known as *Einstein's photoelectric equation*.

If the energy of the photon is just sufficient to liberate the electron from the metal, then the kinetic energy of the electron will be zero. At this condition, Eq. (21.8) reduces to

$$h\nu_0 = W_0 \quad (21.9)$$

where ν_0 is the threshold frequency. It is defined as the minimum value of frequency which can cause photoelectric emission. The wavelength corresponding to the threshold frequency represents the upper limit of wavelength (λ_0) for photoelectric effect. Its physical significance suggests that only those radiations are capable to eject the electrons which have their wavelengths less than the wavelength λ_0 (i.e., $\lambda < \lambda_0$).

In terms of work function, λ_0 is given as

$$\lambda_0 = \frac{c}{\nu_0} = \frac{ch}{W_0} \quad (\because W_0 = h\nu_0) \quad (21.10)$$

or
$$\lambda_0 = \frac{19.875 \times 10^{-26}}{W_0} \text{ m}$$

In the above expression, W_0 is expressed in joules. If it is expressed in electron volt, then

$$\lambda_0 = \frac{19.875 \times 10^{-26}}{1.602 \times 10^{-19} W_0} = \frac{12.4 \times 10^{-7}}{W_0} \text{ m}$$

or
$$\lambda_0 = \frac{12400}{W_0} \text{ Å} \quad (21.11)$$

Now, using Eq. (21.9) in Eq. (21.8), we get

$$\begin{aligned} h\nu &= h\nu_0 + \frac{1}{2}mv^2 \\ \text{or } \frac{1}{2}mv^2 &= (h\nu - h\nu_0) = h(v - v_0) \end{aligned} \quad (21.12)$$

This is also known as *Einstein's photoelectric equation*. From Eq. (21.12), it is clear that

$$v^2 \propto v \quad (21.13)$$

(because h , v_0 , and m are constant)

Equation (21.13) indicates that the velocity of the ejected electrons increases with the increase in frequency of the incident radiation, provided that the intensity of the incident light is constant.

The increase in the intensity of the incident radiation is equivalent to the increase in the number of photons falling on the emitting surface. If these photons have their frequency greater than the threshold frequency (i.e., $v > v_0$), then the photoelectrons will be ejected. Thus, the number of ejected electrons is directly proportional to the intensity of the incident radiation, provided that the energy of the individual photons is greater than certain maximum energy, i.e., greater than $h\nu_0$.

If V_0 is the stopping potential, then

$$\begin{aligned} eV_0 &= h\nu - h\nu_0 \\ \text{or } V_0 &= \frac{h\nu}{e} - \frac{h\nu_0}{e} \\ \text{or } V_0 &= \frac{h}{e}[v - v_0] \end{aligned} \quad (21.14)$$

Solved Examples

Example 21.1

Find the average energy of an oscillator of frequency 5×10^{12} /s at $T = 300$ K, treating it as (i) classical oscillator and (ii) Planck's oscillator.

Solution

(i) The average energy of the classical oscillator is given as

$$\begin{aligned} \bar{E} &= kT = 1.380 \times 10^{-23} \times 300 \\ &= 4.14 \times 10^{-21} \text{ J} \end{aligned}$$

(ii) The average energy \bar{E} of Planck's oscillator is given as

$$\bar{E} = \frac{hv}{\left[\exp\left(\frac{hv}{kT}\right) - 1\right]}$$

$$\begin{aligned}
 &= \frac{6.626 \times 10^{-34} \times 5.0 \times 10^{12}}{\left[\exp\left(\frac{6.626 \times 10^{-34} \times 5.0 \times 10^{12}}{1.380 \times 10^{-23} \times 300} \right) - 1 \right]} \\
 &= \frac{3.313 \times 10^{-21}}{\left(\exp \frac{3.313 \times 10^{-21}}{4.14 \times 10^{-21}} \right) - 1} \\
 &= \frac{3.313 \times 10^{-21}}{\left[e^{0.80} - 1 \right]} \\
 &= \frac{3.313 \times 10^{-21}}{2.23 - 1} = 2.6935 \times 10^{-21} \text{ J}
 \end{aligned}$$

Example 21.2

Electrons are emitted with zero velocity from a certain metal surface when it is exposed to radiations of $\lambda = 7000 \text{ \AA}$. Calculate the threshold frequency and the work function of the metal.

Solution

Given wavelength 7000 \AA corresponds to the threshold frequency v_0 .

$$\begin{aligned}
 \text{Now, } c &= v_0 \lambda_0 \\
 \text{or } v_0 &= \frac{c}{\lambda_0} = \frac{3 \times 10^8}{7000 \times 10^{-10}} \\
 &= 4.286 \times 10^{14} \text{ Hz}
 \end{aligned}$$

Work function is given as

$$\begin{aligned}
 W_0 &= hV_0 = \frac{hc}{\lambda_0} = \frac{6.6 \times 10^{-34} \times 3 \times 10^8}{7 \times 10^{-7}} \\
 &= \frac{19.8 \times 10^{-26}}{7 \times 10^{-7}} \\
 &= 2.8 \times 10^{-19} \text{ J} \\
 &= \frac{2.8 \times 10^{-19}}{1.6 \times 10^{-19}} = 1.77 \text{ eV}
 \end{aligned}$$

Example 21.3

A radiation of wavelength $3 \times 10^{-7} \text{ m}$ is capable of emitting photoelectrons from gold cathode. Calculate the stopping potential for the photoelectrons, if the work function of gold is 3 eV.

Solution

Stopping potential is given as

$$\begin{aligned} eV_0 &= h\nu - W_0 \\ &= \frac{hc}{\lambda} - W_0 \\ \text{or } V_0 &= \frac{hc}{\lambda e} - \frac{W_0}{e} \\ &= \left[\frac{6.626 \times 10^{-34} \times 3 \times 10^8}{3 \times 10^{-7} \times 1.6 \times 10^{-19}} - 3 \right] \\ &= \left[\frac{19.878 \times 10^{-26}}{4.8 \times 10^{-26}} - 3 \right] \\ &= [4.1413 - 3] \\ &= 1.141 \text{ V} \end{aligned}$$

Example 21.4

If the work function of a target material is 1.24 eV and the wavelength of the incident light is 4.4×10^{-7} m then calculate

- (i) the velocity of the photoelectron and
- (ii) the retarding potential necessary to stop the emission of these electrons.

Solution

- (i) According to Einstein's photoelectric equation, we know that

$$\begin{aligned} \frac{1}{2} mv^2 &= h\nu - W_0 \\ \text{or } v^2 &= \frac{2}{m} [h\nu - W_0] = \frac{2}{m} \left[\frac{hc}{\lambda} - W_0 \right] \\ &= \frac{2 \left[\frac{6.6 \times 10^{-34} \times 3 \times 10^8}{4.4 \times 10^{-7}} - 1.24 \times 1.6 \times 10^{-19} \right]}{9.1 \times 10^{-31}} \\ &= \frac{2 \left[4.5 \times 10^{-19} - 1.98 \times 10^{-19} \right]}{9.1 \times 10^{-31}} \\ &= \frac{2 \times 2.52 \times 10^{-19}}{9.1 \times 10^{-31}} \\ &= \frac{5.04 \times 10^{-19}}{9.1 \times 10^{-31}} \end{aligned}$$

$$= 0.5538 \times 10^{12}$$

$$v = 7.44 \times 10^5 \text{ m/s}$$

(ii) If V_0 is the potential required to stop the emission, then

$$eV_0 = \frac{1}{2} mv^2$$

$$V_0 = \frac{\frac{1}{2} \times 9.1 \times 10^{-31} \times 0.5538 \times 10^{12}}{1.6 \times 10^{-19}}$$

$$= \frac{2.52 \times 10^{-19}}{1.6 \times 10^{-19}}$$

$$V_0 = 1.575 \text{ V}$$

21.7 FORMULAE AND HIGHLIGHTS

1. A blackbody neither reflects nor transmits any of the incident radiations and, therefore, appears black.
2. In the spectrum of blackbody radiation, λ_m decreases with the increase in temperature, where λ_m is the wavelength at which maximum emission of energy takes place.
3. According to Wien's displacement law,

$$\lambda_m \cdot T = \text{constant} = 0.2892 \text{ cm K}$$

4. Wien's law of distribution of energy density in the wavelength range λ and $\lambda + d\lambda$ is given as

$$E_\lambda = c_1 \lambda^{-5} e^{c_2/\lambda T} d\lambda$$

5. Rayleigh-Jeans law of distribution of energy in radiation spectrum is given as

$$E_\lambda = \frac{8\pi kT}{\lambda^4} d\lambda$$

6. According to Planck's hypothesis, the frequency of radiations emitted by an oscillator is not continuous but it is only in the multiples of a small unit called quantum (photon).
7. Planck's formula for the distribution of energy in radiation spectrum is given as

$$E_\lambda = \frac{8\pi hc}{\lambda^5 [e^{hc/kT} - 1]} d\lambda$$

8. The emission of electrons from the metal surface when illuminated by light or any other radiation of suitable wavelength is known as photoelectric effect.
9. The number of photoelectrons emitted per second is proportional to the intensity of the incident radiations, while their kinetic energy depends on the energy of the individual photons responsible for the emission of electrons.

22

de Broglie Matter-Waves and Wave–Particle Duality

22.1 INTRODUCTION

The development of wave mechanics started with the revolutionary idea of Max Planck, which states that the electromagnetic radiation may be considered as a stream of small energy packets called *photons*. Wave mechanics deals with the microscopic properties of materials. Before wave mechanics, phenomena such as photoelectric effect, Compton effect, black body radiation, atomic spectra, and so on could not be explained. Wave mechanics was able to correctly explain these phenomena. Macroscopic phenomena such as interference, diffraction, and polarisation firmly confirm the wave character of electromagnetic radiation, but Planck's idea of photons (quanta of energy) suggested that radiation may have particle nature as well. In other words, electromagnetic radiation has dual character; in certain situations it exhibits properties of wave, while in others it acts as a particle.

In this chapter, we will study the wave–particle duality along with the concept of group velocity, phase velocity, and matter-waves.

22.2 WAVE–PARTICLE DUALITY

On the basis of existing experimental facts like photoelectric effect, emission and absorption spectra, and black body radiation, it is clear that electromagnetic radiation consists of discrete indivisible packets of energy ($h\nu$), i.e., photons. This establishes the particle character of radiation. Similarly, major optical phenomena like interference, diffraction, and polarisation establish its wave character. Hence, we can conclude that electromagnetic radiation has dual character, i.e., in certain situations it exhibits characteristics of wave, while in others it shows its particle characteristics.

In order to understand the wave–particle duality, it is essential to know the clear distinctions between particle and wave.

It is easy to realise the concept of a particle by taking the following points into consideration:

- (i) *Mass*: A particle must have definite mass.
- (ii) *Velocity*: A particle can move from one place to another with a certain velocity.
- (iii) *Position*: A particle may be located at some definite place or point.

- (iv) *Momentum*: A particle having mass and velocity possesses momentum during its motion.
- (v) *Energy*: A particle has energy in different forms in different situations such as potential energy, kinetic energy, rest-mass energy, and so on.

However, the concept of a wave is a little bit more complicated than that of a particle. The following points may be taken into account for a wave:

- (i) *Lack of position*: A wave is always realised as a disturbance; it is spread out over a relatively large region of space. It cannot be located at some definite place or point.
- (ii) *Mass*: It is very difficult to think about the mass being associated with a wave.
- (iii) *Frequency and wavelength*: A wave or disturbance, which advances in a medium, has a certain frequency and wavelength.
- (iv) *Phase of wave velocity*: Phase gives an idea about the instantaneous position and direction of a wave.
- (v) *Amplitude*: The amplitude of a wave gives an idea of the intensity of disturbance in the medium.

In view of the above points, it may seem very difficult to accept the wave–particle duality. For example, if we take the superposition of two waves, we will observe that these two waves are lying together, but it is not so for two particles because two particles cannot appear at the same position simultaneously.

Although it is quite difficult to accept that a radiation will have both wave and particle characters, the existing experimental facts, particularly interference, diffraction, and polarisation force us to consider radiation as a wave. At the same time, the phenomena like photoelectric effect, Compton effect, black body radiation, and emission and absorption spectra suggest us that radiation has particle nature as well. Hence, we can conclude that radiation sometimes behaves as a wave and at other times as a particle, i.e., it has wave–particle duality. Here it is important to remember that *radiation cannot exhibit its particle and wave properties simultaneously*.

22.3 de BROGLIE HYPOTHESIS OF MATTER-WAVES

Issac Newton explained the phenomena of reflection and refraction on the basis of corpuscular theory. Corpuscular theory assumes that both matter and radiation consist of particles. But the discovery of phenomena like interference, diffraction, and polarisation proved that light is a kind of wave motion as well. In the beginning of twentieth century, with the discovery of phenomena like photoelectric effect, Compton effect, and black body radiation, another aspect of light came into existence. These phenomena could not be explained on the basis of the wave nature of light. For the satisfactory explanation of these phenomena, corpuscular or quantum nature of light was taken into account. According to this theory, light may be considered as a stream of photons (particles) having mass ($h v/c^2$), energy ($h v$), velocity (c), and momentum ($h v/c$). However, quantum theory (photon theory) could not explain the phenomena like interference, diffraction, and polarisation. It indicates that the various phenomena of light (radiation) can be made only on the basis of the dual nature of light (radiation). Thus, light has dual nature, i.e., it possesses both particle and wave nature. It is important to remember that wave and particle nature can never appear together.

Extending the idea of wave–particle duality of radiation (light), Louis de Broglie in 1924 suggested that this duality is true not only for radiation but it is also true for all the moving material particles of the universe. It means that like radiation, matter also have wave–particle duality.

de Broglie proposed that there is a close connection between waves and particles not only in the case of radiation but also in the case of matter. A wave is always associated with the motion of a particle and a particle is controlled by a wave in the same way as photons are controlled by wave (radiation).

According to de Broglie's concept, a moving particle always has a wave associated with it and the motion of the particle is guided by that wave in the same way as a photon is controlled by a wave. The wavelength of the matter-wave is given by

$$\lambda = \frac{h}{mv} = \frac{h}{p}$$

where m is the mass of the particle, v is its velocity, and p is its momentum.

22.4 EXPRESSION FOR WAVELENGTH OF MATTER-WAVE (de BROGLIE WAVE-LENGTH)

The mathematical expression for the wavelength of the matter-wave can be derived by taking photon into consideration in radiation. According to the photon consideration, an electromagnetic wave (radiation) is associated with photon, and photon is controlled by wave.

A light wave of frequency ν is associated with a photon of energy E . By applying Planck's relation, we get $E = h\nu$.

According to Einstein's theory of relativity, a particle of mass m is equivalent to energy mc^2 , i.e.,

$$E = mc^2$$

If a photon attains a mass m during its motion with velocity c , then we can write

$$h\nu = mc^2$$

Since the momentum of photon is $p = mc$, the above equation can be written as

$$\begin{aligned} h\nu &= pc \\ &= p\nu\lambda \quad (\text{because } c = \nu\lambda) \\ \text{or } \lambda &= \frac{h}{p} \end{aligned} \tag{22.1}$$

Equation (22.1) gives the de Broglie wavelength for photon. Both the material particle and its counterpart (matter-wave) may have the same expression. Here, the momentum is determined by the mass and velocity of the moving particle (approaching the velocity of light). Thus, the momentum will be $P = mv$, and the expression for the wavelength of wave associated with the particle can be given as

$$\lambda = \frac{h}{p} = \frac{h}{mv} \tag{22.2}$$

The expression for the de Broglie wavelength may be expressed in different ways in different situations and circumstances. Some cases are discussed in the following points:

- (i) If the particle is accelerated through kinetic energy E_k , then

$$E_k = \frac{1}{2}mv^2 = \frac{1}{2}\frac{(mv)^2}{m} = \frac{p^2}{2m}$$

or $p = \sqrt{2mE_k}$

Now, the expression for the de Broglie wavelength can be given as

$$\lambda = \frac{h}{\sqrt{2mE_k}} \quad (22.3)$$

- (ii) If the particle is accelerated with the help of electrostatic energy, then the kinetic energy can be given as

$$E_k = qV$$

where V is the accelerating potential.

Now, the momentum of the particle can be given as

$$p = \sqrt{2mqV}$$

Hence, the de Broglie wavelength associated with the particle can be given as

$$\lambda = \frac{h}{\sqrt{2mqV}} \quad (22.4)$$

- (iii) When the particle is accelerated with the help of thermal energy, then its kinetic energy can be given as

$$E_k = \frac{3}{2} kT$$

where k is Boltzmann's constant (1.38×10^{-23} J/K) and T is the absolute temperature. The momentum of the particle will be given as

$$p = \sqrt{2m\left(\frac{3}{2}kT\right)} = \sqrt{3mkT}$$

Therefore, the expression for the wavelength can be given as

$$\lambda = \frac{h}{\sqrt{3mkT}} \quad (22.5)$$

- (iv) When the particle is moving with a velocity comparable to that of light, then the mass of the particle no longer remains constant. It will vary according to the following equation:

$$m = \frac{m_0}{\sqrt{1 - v/c^2}}$$

where m is the mass of the particle during its motion, m_0 is the rest mass of the particle, v is the velocity of the particle, and c is the velocity of light.

Hence, the expression for de Broglie wavelength will be slightly modified, as shown in the following expression:

$$\lambda = \frac{h}{mv} = \frac{h\sqrt{[1 - (v^2/c^2)]}}{m_0v} \quad (22.6)$$



- (v) de Broglie wavelength is associated with electron. Let us consider an electron of rest mass M_0 and charge e accelerated by the potential difference V . The kinetic energy of the electron can be given as

$$\frac{1}{2} m_0 v^2 = eV$$

Hence, the velocity of the electron will be given as

$$v = \sqrt{\frac{2eV}{m_0}}$$

Thus, the de Broglie wavelength associated with the electron can be given as

$$\lambda = \frac{h}{m_0 v} = \frac{h\sqrt{m_0}}{m_0 \sqrt{2eV}}$$

or $\lambda = \frac{h}{\sqrt{2eVm_0}}$

where h is Planck's constant (6.625×10^{-34} Js). Now, by putting the values of constants, we get

$$\lambda = \frac{6.625 \times 10^{-34}}{\sqrt{[2 \times 1.632 \times 10^{-19} \times V \times 9.1 \times 10^{-31}]}}$$

or $\lambda = \frac{12.26}{\sqrt{V}} \text{ Å}$ (22.7)

If we take $V = 100$ V, then the value of λ can be given as

$$\lambda = 1.226 \text{ Å}$$

Therefore, the wavelength associated with an electron accelerated by the potential difference of 100 V is 1.226 Å.

22.5 PROPERTIES OF MATTER-WAVES

de Broglie matter-waves have the following properties:

- (i) The lighter particles have greater wavelength than the heavier particles.
- (ii) The smaller the velocity of the particle, the greater is the wavelength associated with it.
- (iii) From the expression of the de Broglie wavelength, i.e., $\lambda = (h/mv)$, if $v = 0$, then $\lambda = \infty$, whereas if $v = \infty$, then $\lambda = 0$. This shows that the matter-waves are generated only when the particle is in motion.
- (iv) The matter-waves are independent of the charge. Thus, they are produced by both charged and uncharged particles. This shows that the matter-waves are not electromagnetic waves; they are entirely different waves.
- (v) The velocity of the matter-waves is not constant. It depends on the velocity of the particle, while the velocity of the electromagnetic waves is constant.

- (vi) The velocity of the matter-waves may be greater than the velocity of light. It can be proved as follows:

During the motion of a matter-wave, the particle and the wave associated with it are treated as complementary to each other. Hence, there will be two different velocities; one referring to the mechanical motion of the particle represented by v and the second related to the propagation of the wave represented by ω .

We know that

$$E = hv$$

Also,

$$E = mc^2$$

or $hv = mc^2$

or $v = \frac{mc^2}{h}$

The wave velocity (ω) is given by

$$\omega = v \times \lambda$$

$$= \frac{mc^2}{h} \frac{h}{mv} = \frac{c^2}{v}$$

Since a particle cannot travel with the velocity more than the velocity of light, thus, ω will be greater than the velocity of light. The satisfactory explanation of such unexpected result can be explained with the help of wave velocity and group velocity.

- (vii) The wave and particle aspects of matter never appear simultaneously in the same experiment. These aspects are complementary to each other and are exhibited only in mutually exclusive experimental arrangements. In some phenomena, particle nature predominates while in others wave nature predominates.
- (viii) Wave velocity of a matter-wave is given in terms of group velocity as

$$v_{\text{phase}} = \frac{c^2}{v_{\text{group}}}$$

where v_{group} is equal to the particle velocity (see Section 22.15). Since a particle cannot travel with the velocity more than the velocity of light, the velocity of matter-wave will be greater than c .

- (ix) Wave-particle duality (wave nature of matter) introduces the concept of uncertainty. This concept suggests that if the particle nature of matter becomes certain, the wave nature will be uncertain and vice versa.

22.6 DETECTION OF MATTER-WAVES

A wave associated with the motion of a material particle is known as a matter-wave. The matter-waves are generated by both microscopic and macroscopic bodies. However, the wavelengths of matter-waves associated with the macroscopic bodies are so small that they cannot be detected by any instrument. Examples 22.1 to 22.3 will be enough to explain the fact why the wave nature of matter is not observed in our daily life.

Solved Examples**Example 22.1**

Calculate the de Broglie wavelength (λ) of the earth, taking the mass of the earth to be 6×10^{24} kg. Orbital velocity of the earth is 3×10^4 m/s.

Solution

We know that

$$\lambda = \frac{h}{mv}$$

On substituting the values of the constants, we get

$$\begin{aligned}\lambda &= \frac{6.625 \times 10^{-34}}{6 \times 10^{24} \times 3 \times 10^4} \\ &= 3.68 \times 10^{-63} \text{ m}\end{aligned}$$

Thus, the value of the de Broglie wavelength of the earth, i.e., 3.68×10^{-63} m, is so small that it cannot be detected by any measuring instrument.

Example 22.2

A body of mass 1 kg is moving with the velocity of 10 m/s. What will be the de Broglie wavelength associated with this body?

Solution

We know that

$$\lambda = \frac{h}{mv}$$

On substituting the values of the constants, we get

$$\lambda = \frac{6.625 \times 10^{-34}}{1 \times 10} = 6.625 \times 10^{-35} \text{ m}$$

This wavelength is also very small and cannot be detected by any instrument.

Example 22.3

An electron of mass of order 10^{-30} kg is moving with the velocity of 10^5 m/s. What will be its wavelength? Can it be detected?

Solution

The wavelength of the wave associated with the electron can be given as

$$\lambda = \frac{6.625 \times 10^{-34}}{10^{-30} \times 10^5} = 6.625 \times 10^{-9} \text{ m}$$

This value of λ is within the limits of measurement. It is of the order of the wavelength of X-rays which can be detected.

22.7 NATURE OF ELECTRON

Before 1924, the electron was exclusively regarded as a particle but after the suggestion of de Broglie, the electron was given the wave-particle duality. There are many experimental evidences which prove that the electron is a particle, as it has a definite mass, charge, energy, and momentum. Additionally, the impact of the electron on the screen of zinc sulphide proves its identity as a particle.

At the same time, there are some experimental evidences which prove that the electron has a wave nature as well, i.e., a wave is associated with the electron during its motion. The following are some experiments which support the wave nature of electrons:

- (i) Davisson and Germer diffraction experiment
- (ii) G.P. Thomson experiment

We will now discuss Davisson and Germer experiment in detail.

22.8 DAVISSON AND GERMER EXPERIMENT FOR MATTER-WAVES

Initially, de Broglie hypothesis of matter-waves was accepted provisionally, as it had not been proved experimentally. This important task was accomplished in 1927 by C.J. Davisson and L.H. Germer in the United States. They proposed that if the de Broglie hypothesis of the matter-waves is correct, then the particles such as electrons and protons should exhibit diffraction patterns similar to those of electromagnetic waves.

The detail discussion of Davisson and Germer experiment is as follows:

- (i) *Principle:* If an electron is accelerated through a potential difference of V volts, then the electron acquires the kinetic energy equivalent to qV joule, where q is the charge on the electron (in coulomb). The total energy E of the electron becomes

$$E = m_0 c^2 + qV \quad (22.8)$$

where $m_0 c^2$ is the rest-mass energy of the electron.

But from the special theory of relativity, we have

$$E = \sqrt{m_0^2 c^4 + p^2 c^2} \quad (22.9)$$

From Eqs. (22.8) and (22.9), we get

$$m_0^2 c^4 + p^2 c^2 = (m_0 c^2 + qV)^2$$

or

$$p^2 c^2 = 2m_0 c^2 qV + (qV)^2$$

or

$$p = \sqrt{\left\{ 2m_0 qV \left(1 + \frac{qV}{2m_0 c^2} \right) \right\}} \quad (22.10)$$

The de Broglie wavelength corresponding to such an electron is

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{\{2m_0qV(1 + qV/2m_0c^2)\}}} \quad (22.11)$$

Putting the values of various constants in Eq. (22.11), we obtain

$$\lambda = \frac{12.27 \times 10^{-10}}{\sqrt{\{V(1 + 9.8 \times 10^{-7}V)\}}} \text{ m} \quad (22.12)$$

If V is not very large, the second term in the denominator of Eq. (22.12) can be neglected, and hence, we get

$$\lambda = \frac{12.27}{\sqrt{V}} \text{ Å} \quad (22.13)$$

The theoretical calculation of the de Broglie wavelength associated with an electron gives the value of λ according to Eq. (22.13). If we calculate the wavelength of a wave associated with the electron, which is accelerated by the potential difference of 100 V, then the value of the wavelength will be 1.277 Å. For experimental verification, if a stream of such electron beam is passed through an aperture of the order of 1 Å, then a diffraction pattern should be produced. Since this order of the wavelength is the same as the distances between atomic planes in crystals, so these (crystals) can be used as grating (or aperture) for the diffraction of the electron beam.

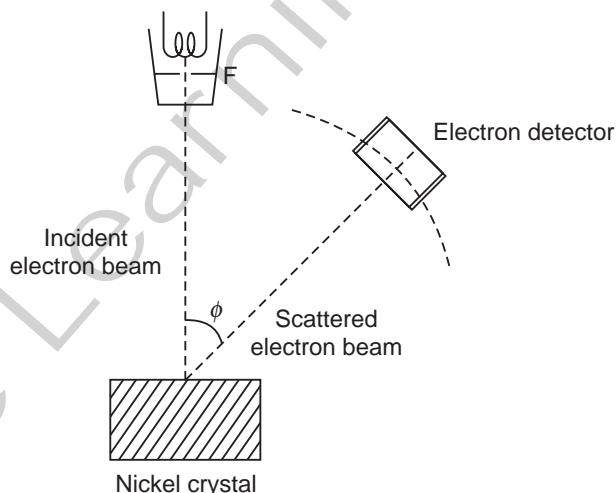


Fig. 22.1 Davisson-Germer experimental setup

- (ii) *Experimental arrangement:* The experimental setup used by Davisson and Germer is given in Fig. 22.1. Filament F was used to emit electrons at different potentials. Well-collimated beams of electrons were incident normally on a nickel crystal. The electrons were scattered in all the directions by the atoms in the crystal. The intensity of the electrons scattered in a given direction was measured by receiving the electrons at electron detector, which can be rotated through any

angular position with respect to the crystal. In their experiment, Davisson and Germer measured the intensity of the scattered electrons as a function of the scattering angle ϕ .

- (iii) *Results and discussion:* The experimental results are discussed on the basis of different curves obtained between scattering angle ϕ and the intensity of scattered beam of electrons, corresponding to different accelerating voltages V .

These graphs are plotted in polar coordinates. In these plots, the length of the radius vector is proportional to the intensity of the scattered beam, and the angle between the vertical axis and the radius vector is the angle of scattering. A set of curves has been shown in Fig. 22.2. By observing closely, we find that in the plot obtained for the electrons accelerated through 40 V, there is a hump at about 60° . We draw a vertical line in the first graph and show a small hump at 60° . Then the fifth and the sixth plots are distorted and the position of the hump gets shifted downward.

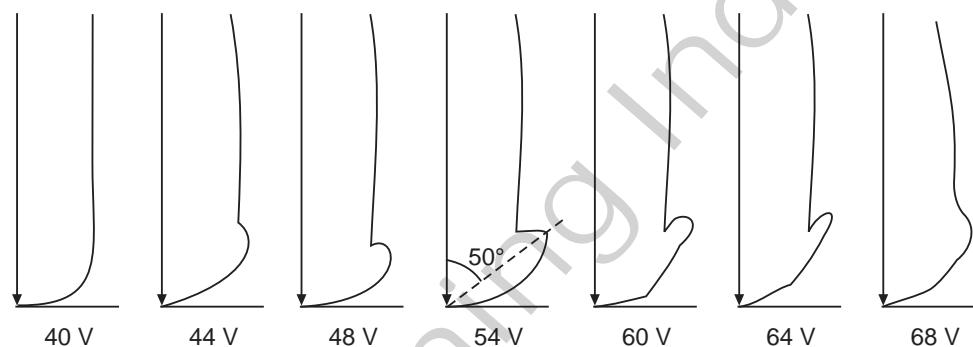


Fig. 22.2 Curves showing the variation of intensity with scattering angle at different accelerating potentials

It is also observed that at higher values of the accelerating potential, the hump moves upward and becomes deeper. At accelerating potential of 54 V, the hump is converted into a spur. In this plot, it has been observed that the length of the spur increases gradually and becomes most distinct when the scattering angle (ϕ) becomes 50° at the accelerating potential 54 V. At higher values of the accelerating potential, the length of the spur decreases and gradually disappears. Since the length of the radius vector gives the intensity, the existence of the spur of the maximum length corresponds to the maximum intensity of scattered electrons at $\phi = 50^\circ$ and $V = 54$ V. This maximum intensity can be explained in terms of the constructive interference between electron waves from the periodically arranged atoms of nickel crystal. Bragg's atomic planes, which are drawn through the regularly spaced atoms of the nickel crystal, have been shown in Fig. 22.3.

According to Fig. 22.3, a beam of the incident electrons makes an angle θ with Bragg's atomic planes. This angle is known as *glancing angle*. For maximum intensity of the scattered electron beam (wave), it has been observed at scattering angle (ϕ) = 50° . Corresponding to this value of the scattering angle, the glancing angle (θ) can be calculated as follows:

23

Uncertainty Principle and Quantum Mechanics

23.1 INTRODUCTION

In classical mechanics, we can simultaneously determine the position and the momentum of a particle. However, this is not the case in quantum mechanics. From de Broglie hypothesis of matter waves, we know that a moving particle may be considered as a group of waves and the particle may be positioned anywhere within the wave packet. This indicates that the position of the particle is uncertain within the limits of a wave packet. This uncertainty is only for microscopic particles; it has no importance for macroscopic particles. In atomic systems, the classical mechanics fails to explain the microscopic system of particles due to the above said reasons. The uncertainty principle states that the position and the momentum of a microscopic particle cannot be simultaneously measured accurately. The measurement of one quantity introduces a measure of uncertainty into the other. Hence, the classical mechanics, for which position and momentum have definite values at all instants, is not valid for atomic systems.

From the concept of de Broglie, we know that a wave is associated with a material particle. Clearly, a mathematical wave equation is required to deal with such waves. Since classical equations (mathematical formulations) were not valid at microscopic level, new equations known as *wave mechanics* or *quantum mechanics* were developed by Schrödinger in 1926. In his wave mechanics, Schrödinger assumed wave function $\psi(x, y, z, t)$ as the amplitude of matter waves. Wave function $\psi(x, y, z, t)$ is a complex quantity and is the function of position and time. It gives the idea for the probability of finding a particle in a particular region of space.

In this chapter, we will discuss uncertainty principle and Schrödinger wave equations with their applications.

23.2 BASIS FOR UNCERTAINTY PRINCIPLE

Although in the beginning scientists were reluctant to accept this principle, but the strong evidences forced them to accept the uncertainty principle. The following are some important observations about the uncertainty principle:

- (i) The material particle exhibits particle nature as well as exhibits wave nature, but it does not simultaneously possess both the natures.
- (ii) Instead of being contradictory, the wave and particle natures are complementary.

- (iii) Bohr's principle of complementarity is the consequence of de Broglie hypothesis.
- (iv) Under the de Broglie hypothesis, particles may be represented as wave packets. The particle may be anywhere inside the wave packet. Hence, there will be uncertainty in the measurement of position of the particle.

23.3 HEISENBERG'S UNCERTAINTY PRINCIPLE

Due to the dual nature of material particles, it is very difficult to locate the exact position and the momentum of the particle simultaneously. This uncertainty was explained by Werner Heisenberg in 1927 through his *uncertainty principle*. To understand the uncertainty in the measurement of position and momentum of microscopic particles, let us take the examples of narrow and wide wave packets.

In a narrow wave packet [Fig. 23.1(a)], the position of the particle can be precisely determined, but not the wavelength. As a result, the particle's momentum cannot be measured accurately as there are not enough waves to exactly measure the wavelength ($\lambda = h/mv$). On the other hand, in a wider wave packet

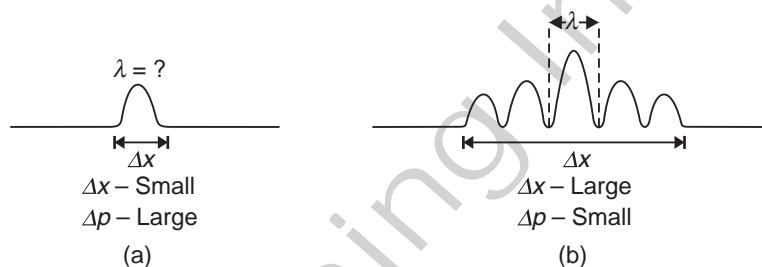


Fig. 23.1 Uncertainty in position and time

[Fig. 23.1(b)], the wavelength can be determined exactly but the position of the particle will be uncertain due to the large width of the wave packet. Hence, it can be concluded that it is impossible to simultaneously determine the exact position and the exact momentum of a particle.

23.3.1 Statement of the Uncertainty Principle

The Heisenberg's uncertainty principle states that it is not possible to simultaneously measure the position and the momentum of a particle to any desired degree of accuracy. In other words, the product of uncertainty in the measurement of position (Δx) and uncertainty in the measurement of momentum (Δp) is always constant, and it is at least equal to Planck's constant (h), i.e.,

$$\Delta p \cdot \Delta x = h \quad (23.1)$$

Equation (23.1) implies that if uncertainty in the measurement of position (Δx) increases, then uncertainty in the measurement of momentum (Δp) will decrease and vice versa.

Similar to Eq. (23.1), we can write

$$\Delta E \cdot \Delta t = h \quad (23.2)$$

and $\Delta J \cdot \Delta \theta = h \quad (23.3)$

where ΔE and Δt are the uncertainties in determining energy and time, respectively. Similarly, ΔJ and $\Delta \theta$ are the uncertainties in the measurement of angular momentum and angle, respectively.

It has been experimentally observed that the product of uncertainties in position and momentum is equal to or greater than $h/4\pi$, i.e.,

$$\Delta p \cdot \Delta x \geq \frac{h}{4\pi} \quad (23.4)$$

where h is Planck's constant.

23.4 DERIVATION OF UNCERTAINTY PRINCIPLE

Uncertainty principle is the direct consequence of the dual nature of the material particle. The principle of dual nature holds that a moving particle is associated with group(s) of waves and these waves travel with the velocity equal to that of the particle.

Let us consider two simple harmonic plane waves of same amplitude A having nearly equal frequencies ω_1 and ω_2 with propagation vectors k_1 and k_2 , respectively. These plane waves can be written as

$$y_1 = A \sin(\omega_1 t - k_1 x) \quad (23.5)$$

$$y_2 = A \sin(\omega_2 t - k_2 x) \quad (23.6)$$

where y_1 and y_2 are the displacements of these plane waves. Due to superposition of these waves, we will get the wave group as shown below:

$$\begin{aligned} y &= y_1 + y_2 \\ &= A \sin(\omega_1 t - k_1 x) + A \sin(\omega_2 t - k_2 x) \\ &= 2A \sin\left[\left(\frac{\omega_1 + \omega_2}{2}\right)t - \left(\frac{k_1 + k_2}{2}\right)x\right] \cos\left[\left(\frac{\omega_1 - \omega_2}{2}\right)t - \left(\frac{k_1 - k_2}{2}\right)x\right] \\ &= 2A \sin(\omega t - kx) \cos\left(\frac{\Delta\omega}{2}t - \frac{\Delta k}{2}x\right) \end{aligned} \quad (23.7)$$

where $\omega = \frac{\omega_1 + \omega_2}{2}$, $k = \frac{k_1 + k_2}{2}$, $\Delta\omega = \omega_1 - \omega_2$, and $\Delta k = k_1 - k_2$.

Equation (23.7) represents the wave group, which is the result of Eqs. (23.5) and (23.6). In Eq. (23.7), the second term is responsible for the modulation and forms a wave packet as shown in Fig. 23.2.



Fig. 23.2 Group of waves

In wave packets, the position of the particle remains uncertain between successive nodes (or extreme points of loop). From Eq. (23.7), the condition for node formation is given as:

$$\cos\left(\frac{\Delta\omega}{2}t - \frac{\Delta k}{2}x\right) = 0$$

$$\text{or } \left(\frac{\Delta\omega}{2}t - \frac{\Delta k}{2}x \right) = \frac{(2n+1)\pi}{2} \quad (23.8)$$

where $n = 0, 1, 2$, and so on.

If x_n and x_{n+1} are positions of n th and $(n+1)$ th nodes, then from Eq. (23.8), we get

$$\frac{\Delta\omega}{2}t - \frac{\Delta k}{2}x_n = \frac{(2n+1)\pi}{2} \quad (23.9)$$

$$\text{and } \frac{\Delta\omega}{2}t - \frac{\Delta k}{2}x_{n+1} = \frac{(2n+3)\pi}{2} \quad (23.10)$$

On subtracting Eq. (23.9) from Eq. (23.10), we get

$$\frac{\Delta k}{2}(x_{n+1} - x_n) = \pi$$

$$\text{Now, } X_{n+1} - X_n = \Delta x = \frac{2\pi}{\Delta k} \quad (23.11)$$

From the definition of propagation vector, we know that

$$\begin{aligned} k &= \frac{2\pi}{\lambda} \\ &= \frac{2\pi}{h/p} \\ &= \frac{2\pi p}{h} \\ \text{or } \Delta k &= \frac{2\pi \Delta p}{h} \\ \text{or } \Delta p &= \frac{\Delta k \cdot h}{2\pi} \end{aligned} \quad (23.12)$$

Now, from Eqs. (23.11) and (23.12), we get

$$\begin{aligned} \Delta x \cdot \Delta p &= \frac{\cancel{2\pi}}{\cancel{\Delta k}} \cdot \frac{\cancel{\Delta k} \cdot h}{\cancel{2\pi}} \\ \Delta x \cdot \Delta p &= h \end{aligned} \quad (23.13)$$

Equation (23.13) represents the original statement of the uncertainty principle of Heisenberg.

23.5 TIME-ENERGY UNCERTAINTY PRINCIPLE

We can derive the expression for time-energy uncertainty with the help of position and momentum uncertainties. Let us consider a particle of rest mass m_0 moving with velocity v_x in the X -direction. The kinetic energy of the particle can be given as

$$E = \frac{1}{2} m_0 v_x^2 = \frac{p_x^2}{2m_0} \quad (23.14)$$

Differential form of Eq. (23.14) can be written as

$$\Delta E = \frac{2 p_x \cdot \Delta p_x}{2m_0}$$

If ΔE and Δp_x are the uncertainties in energy and momentum, respectively, then

$$p_x \cdot \Delta p_x = m_0 \Delta E$$

$$\text{or } \Delta p_x = \frac{m_0}{p_x} \cdot \Delta E = \frac{1}{v_x} \cdot \Delta E \quad (23.15)$$

Let the uncertainty in measuring the time interval at point x be Δt . Then, uncertainty in position Δx can be given as

$$\Delta = v_x \cdot \Delta t \quad (23.16)$$

From Eqs. (23.15) and (23.16), we get

$$\begin{aligned} \Delta p_x \cdot \Delta x &= \frac{1}{v_x} \Delta E \cdot v_x \cdot \Delta t \\ &= \Delta E \cdot \Delta t \end{aligned}$$

$$\text{or } \Delta E \cdot \Delta t = h \quad [\text{from Eq. (23.1), } \Delta x \cdot \Delta p_x = h]$$

$$\text{or } \Delta E \cdot \Delta t \geq h/4\pi \quad (23.17)$$

Note: For calculations, the uncertainty principle expressions $\Delta x \cdot \Delta p \geq \hbar$ and $\Delta x \cdot \Delta p \geq \hbar/2$ are equally good. One can use any one of them for the solution of numerical problems. It is equally applicable for other uncertainty equations.

23.6 EXPERIMENTAL EXAMPLES OF UNCERTAINTY PRINCIPLE

There are many examples confirming the uncertainty principle. Some of them are given below.

23.6.1 Determination of the Position of a Particle by γ -ray Microscope

To measure the exact position and the momentum of an electron along the X-axis in the field of view of an ideally high resolving power microscope, let us consider a photon being incident on an electron in the field of view of microscope as shown in Fig. 23.3.

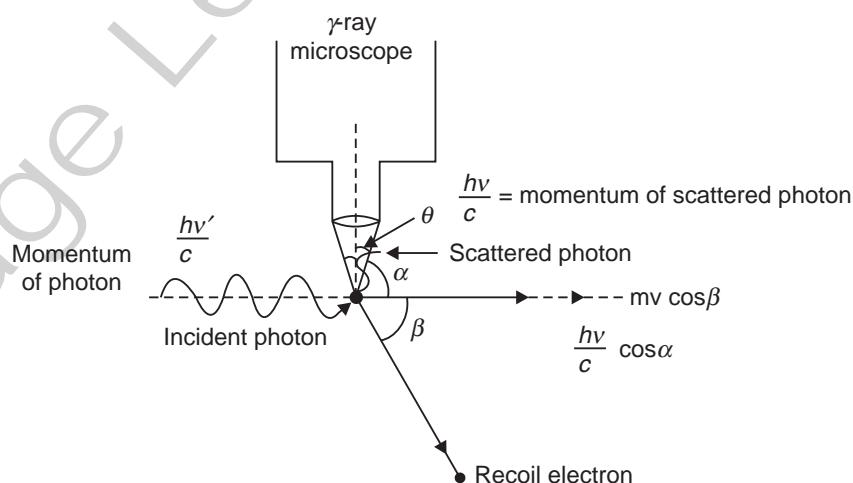


Fig. 23.3 Determination of position and momentum of electron by γ -ray microscope

The resolving power of a microscope can be given as

$$\Delta x = \frac{\lambda}{2 \sin \theta} \quad (23.18)$$

where Δx is the minimum distance in the field of view that can just be resolved by the microscope (i.e., uncertainty in the measurement of the position of the electron), λ is the wavelength of the photon received at microscope after scattering, and θ is the semi-vertical angle of the cone of the light rays entering the objective lens of the microscope.

If α and β are the angles made by the scattered photon and the recoiled electron from the +ve X-axis, respectively, the law of conservation of momentum along the X-axis can be given as

$$\frac{hv'}{c} = \frac{hv}{c} \cos \alpha + mv \cos \beta \quad (23.19)$$

where $mv \cos \beta$ is the component of momentum of the recoiled electron along the +ve X-axis and α is the scattering angle of photon.

Now, from Eq. (23.19), we can write

$$mv \cos \beta = p_x = \frac{hv'}{c} - \frac{hv}{c} \cos \alpha$$

Here, α may vary from $(90 - \theta)$ to $(90 + \theta)$. Hence, we get

$$\begin{aligned} \frac{h}{c} (v' - v \cos (90^\circ - \theta)) &\leq p_x \leq \frac{h}{c} (v' - v \cos (90^\circ + \theta)) \\ \text{or } \Delta p_x &= \frac{h}{c} [v' - v \cos (90^\circ + \theta) - v' + v \cos (90^\circ - \theta)] \\ &= \frac{h}{c} (v \sin \theta + v \sin \theta) \\ &= \frac{2hv}{c} \sin \theta \\ \Delta p_x &= \frac{2h \sin \theta}{\lambda} \end{aligned} \quad (23.20)$$

From Eqs. (23.18) and (23.20), we get

$$\Delta x \cdot \Delta p_x = h \quad (23.21)$$

From Eq. (23.21), it is clear that the product of uncertainties in determining the position and the momentum is constant. This suggests that any attempt to increase the accuracy of the measurement of position of the electron will increase the inaccuracy (uncertainty) of the measurement of momentum of the electron.

Regarding uncertainty measurements, it has been observed that the product of uncertainties in the measurements of position and momentum is equal to $h/4\pi$. Hence, the exact statement of uncertainty principle can be mathematically expressed as

$$\Delta x \cdot \Delta p \geq h/4\pi \quad (23.22)$$

23.6.2 Diffraction of a Beam of Electrons by a Slit

Let a beam of electrons is transmitted through a slit of width Δy (comparable with λ) and received on a photographic plate P , kept at some distance from the slit (shown in Fig. 23.4).

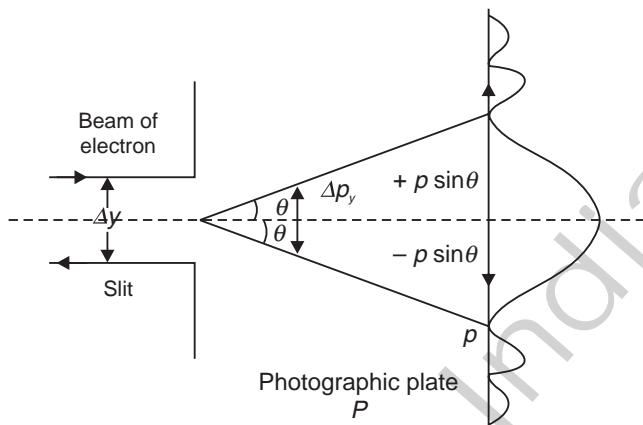


Fig. 23.4 Diffraction of an electron beam

For the electrons of the beam, we can only say that these are passing through the slit, but we cannot specify their exact location in the slit. Hence, the position of any electron recorded on the photographic plate is uncertain by the amount equal to the width of the slit (Δy). Let λ be the wavelength of the wave associated with electrons and θ be the angle of deviation corresponding to the first minimum. In Fraunhofer's diffraction at a single slit, the direction of the first minimum is given by $d \sin \theta = \lambda$, where d is the width of the slit. Thus, according to Fig. 23.4,

$$\begin{aligned} \Delta y \sin \theta &= \lambda \\ \text{or } \Delta y &= \frac{\lambda}{\sin \theta} \end{aligned} \quad (23.23)$$

Equation (23.23) gives the uncertainty in determining the position of electrons along the Y -axis.

Initially, the electrons are moving along the X -axis and therefore, they have no component of momentum along the Y -axis. As the electrons are deviated at the slit from their initial path to form the pattern, they acquire an additional component of momentum along the Y -axis. If p is the momentum of an electron after emerging from the slit, the component of momentum of the electron along the Y -axis is $p \sin \theta$. As the electron may be anywhere between $-p \sin \theta$ and $+p \sin \theta$, the uncertainty in its Y -component of momentum can be given as

$$\begin{aligned} \Delta p_y &= 2p \sin \theta = \frac{2h}{\lambda} \sin \theta \\ \Delta y \cdot \Delta p_y &\geq \frac{\lambda}{\sin \theta} \cdot \frac{2h}{\lambda} \sin \theta = 2h \\ \Delta y \cdot \Delta p_y &\geq h \end{aligned} \quad (23.24)$$

Equation (23.24) shows the uncertainty principle, which states that the product of uncertainties in the measurements of position and momentum is of the order of Planck's constant.

23.6.3 Consequences of Uncertainty Principle

The most important consequence of uncertainty principle is the dual nature of matter. In the dual nature, it is not possible to determine the wave and particle properties exactly at the same time. The complementarity principle states that the wave and particle aspects of matter are complementary, instead of being contradictory. This principle suggests that the consideration of particle and light natures is necessary to have a complete picture of the same system.

23.7 APPLICATIONS OF UNCERTAINTY PRINCIPLE

Uncertainty principle is used for explaining many facts of quantum mechanics, which cannot be explained by classical mechanics. Some of the applications of uncertainty principle are as follows.

23.7.1 Non-Existence of Electrons in the Nucleus

To prove the non-existence of an electron in the nucleus, let us first assume that the electron is present in nucleus. Since the diameter of nucleus is of the order of 10^{-14} m, the maximum uncertainty in the measurement of position of the electron in the nucleus will be of the order of $\Delta x = 10^{-14}$ m. Using Heisenberg's uncertainty relation, the uncertainty in the measurement of momentum of the electron is given as

$$\begin{aligned}\Delta p_x &\geq \frac{h}{4\pi \cdot \Delta x} \\ &= \frac{6.63 \times 10^{-34} \text{ Js}}{4 \times 3.14 \times (10^{-14} \text{ m})} \quad (\Delta x \approx 10^{-14} \text{ m}) \\ &= 0.527 \times 10^{-20} \text{ kgm/s}\end{aligned}$$

or $\Delta p_x \geq 0.527 \times 10^{-20} \text{ kgm/s}$

If uncertainty in momentum of the electron is of the order of 0.527×10^{-20} kgm/s, then the momentum of the electron must be at least of the same order. Since the mass of the electron is 9.1×10^{-31} kg, the electron in nucleus should have a velocity comparable to that of light. Using the relativistic formula for energy E of the electron, we have

$$E^2 = p^2 c^2 + m_0^2 c^4$$

Since the rest mass energy ($m_0 c^2$) is negligible (0.511 MeV) in comparison to $p^2 c^2$, the expression of energy can be given as

$$\begin{aligned}E^2 &= p^2 c^2 \\ \text{or } E &= pc \\ &\approx (0.527 \times 10^{-20}) \times (3 \times 10^8) \\ &= \frac{0.527 \times 10^{-20} \times 3 \times 10^8}{1.6 \times 10^{-19}} \text{ eV} \\ E &= 9.88 \text{ MeV}\end{aligned}$$

The above calculation shows that an electron can exist in the nucleus if its energy is of the order of 9.88 MeV. But we know that the electrons emitted by radioactive nuclei during β -decay have energies of the order of 3 MeV to 4 MeV only. Hence, electrons cannot exist in the nucleus.



23.7.2 Existence of Protons, Neutrons, and α -particles in the Nucleus

To prove the existence of protons, neutrons, and α -particles in the nucleus, let us start with the maximum uncertainty in the measurements of positions of these particles in the nucleus. This uncertainty will be equal to the order of the diameter of nucleus, i.e., $\Delta x = 10^{-14}$ m. Using the uncertainty principle, the uncertainty in the momentum of above said particles can be given as

$$\begin{aligned}\Delta p_x &\geq \frac{h}{4\pi \cdot \Delta x} \\ &= \frac{6.63 \times 10^{-34} \text{ Js}}{4 \times 3.14 \times (10^{-14} \text{ m})} \\ &= 0.527 \times 10^{-20} \text{ kgm/s}\end{aligned}\quad (\Delta x = 10^{-14} \text{ m})$$

For protons and neutrons, $m_0 = 1.67 \times 10^{-27}$ kg. This is a non-relativistic problem as for these particles, $v = p/m_0 = 3 \times 10^6$ m/s. The kinetic energy in this case can be given as

$$\begin{aligned}E_k &= \frac{p^2}{2m} = \frac{(0.527 \times 10^{-20})^2}{2 \times 1.67 \times 10^{-27}} \text{ J} \\ &= \frac{(0.527 \times 10^{-20})^2}{2 \times 1.67 \times 10^{-27} \times 1.6 \times 10^{-19}} \text{ eV} = E_k = 52 \text{ keV}\end{aligned}$$

This value of this kinetic energy is smaller than the energies of the particles emitted by the nucleus. Hence, the particles such as protons and neutrons, or the particles heavier than these, can exist inside the nucleus.

23.7.3 Binding Energy of an Electron in an Atom

The electrons in an atom revolve in different orbits under the influence of electrostatic attraction of positively charged nucleus. If an electron is revolving in an orbit of radius R , then the maximum uncertainty in the measurement of the position of the electron will be equal to the diameter of that orbit, i.e., $2R$. Now, using the uncertainty principle, the uncertainty in the momentum can be given as

$$\Delta p_x \geq \frac{h}{4\pi \cdot 2R}$$

The above expression suggests that the momentum of an electron in an atomic orbit should be at least of the order of Δp_x . For $R = 10^{-10}$ m, the momentum of an electron can be given as

$$p_x \geq \frac{h}{4\pi \cdot 2R} = 0.527 \times 10^{-24} \text{ kgm/s}$$

This shows that p_x is non-relativistic. Now, the kinetic energy of an electron is given as

$$\begin{aligned}E_k &= \frac{p^2}{2m_0} \\ &= \left(\frac{h}{4\pi R} \right)^2 \cdot \frac{1}{2m_0} \\ &= \frac{h^2}{32\pi^2 m_0 R^2}\end{aligned}$$

The electrostatic potential energy of an electron under the influence of a nucleus of atomic number Z is given as

$$V = \frac{-Ze^2}{4\pi^2 \epsilon_0 R}$$

Now, the total energy of an electron in its orbit is

$$\begin{aligned} E &= E_k + V \\ &= \frac{h^2}{32\pi^2 m_0 R^2} - \frac{Ze^2}{4\pi^2 \epsilon_0 R} \\ &= \frac{(1.055 \times 10^{-34})^2}{8 \times 9.1 \times 10^{-31} \times R^2} - \frac{Z(1.6 \times 10^{-19})^2}{4 \times 3.14 \times 8.85 \times 10^{-12} \times R} \\ &= \left(\frac{10^{-20}}{R^2} - \frac{15 \times 10^{-19} Z}{R} \right) \text{eV} \end{aligned}$$

For $R = 10^{-10}$ m, we have

$$E = (1 - 15Z) \text{ eV}$$

Since first term in the bracket of the above expression is negligible compared to the second, we can write

$$E \approx (-15Z) \text{ eV}$$

It is well known that the binding energies of outermost electrons in H ($Z = 1$) and He ($Z = 2$) are -13.6 eV and -24.6 eV, respectively. From the above calculation, it is clear that the calculated value of the binding energy is comparable with the known magnitudes. Hence, the binding energy of an electron in an atom can be calculated.

23.7.4 Zero-Point Energy of a Harmonic Oscillator

From quantum mechanics, we know that the lowest energy of a simple harmonic oscillator is not zero; instead it is equal to $1/2 \hbar \omega$ (where $\hbar = h/2\pi$) and is known as *zero-point energy*. This zero-point energy of the oscillator can be obtained with the help of uncertainty principle.

Let Δx and Δp_x be the uncertainties in the simultaneous measurements of the position and the momentum of a particle of mass m executing simple harmonic motion along the X -axis. Now, from the uncertainty principle, we can write

$$\Delta x \cdot \Delta p_x \geq \frac{\hbar}{2}$$

or
$$\Delta p_x = \frac{\hbar}{2\Delta x}$$

Total energy of the particle of mass m can be given as

$$E = E_k + V$$

24

Thermodynamics

24.1 INTRODUCTION

Thermodynamics is the study of heat in motion and its relation with the other forms of energy. It is based on the principles which have been formulated by generalizing experimental observations made on the variety of physical systems. Basically it is related to the macroscopic theory of natural phenomena, the aim of which is to establish relations between directly observed macroscopic quantities, viz., pressure, volume, temperature, canonization, electric field, magnetic field, etc., which define the state of system under investigation. The principles of thermodynamics are widely used in the design of heat engines, nuclear power stations, rocket propulsion cryogenics, biological phenomena, etc.

In this chapter we shall discuss the basic laws of thermodynamics, different thermodynamic processes, variables, potentials, such as internal energy, enthalpy, Gibb's free energy, and entropy. We shall also consider the *T-S* diagram, Carnot cycle, and Maxwell's relations.

24.2 LAWS OF THERMODYNAMICS

In order to describe the relations between heat energy and mechanical work, as well as other forms of energy, different principles of thermodynamics are suggested. These principles are known as the *laws of thermodynamics*.

24.2.1 Zeroth Law of Thermodynamics

When two bodies are in thermal contact with each other and no exchange of heat takes place between them, they are said to be in *thermal equilibrium*.

"It is observed that when two bodies are in thermal equilibrium with a third body, then these two bodies are also in thermal equilibrium with each other." This statement is known as the *zeroth law of thermodynamics*. Regarding this law the term "zero" is used because of the fact that this law logically provides the basis to the first and second laws of thermodynamics. For better illustration of this law, let us consider two systems *A* and *B* separated by an adiabatic wall and the third system *C* separated from *A* and *B* both by a diathermic wall as shown in Fig. 24.1. In this arrangement, systems *A* and *C* will be in thermal equilibrium and so will be systems *B* and *C*. Now if the adiabatic wall between *A* and *B* is replaced by a diathermic wall, then *A* and *B* will also be in thermal equilibrium. From the above discussion it is clear that if system *A* is in thermal equilibrium with *B* and *B* is in thermal equilibrium with *C*, then *A* and *C* will also be in thermal equilibrium.

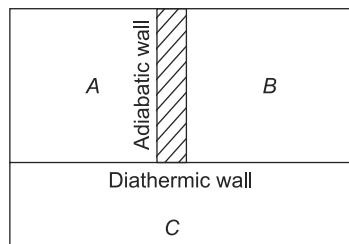


Fig. 24.1 Systems A and B are separated by an adiabatic wall and system C is separated from A and B through a diathermic wall

24.2.2 First Law of Thermodynamics

According to the first law of thermodynamics, “The mechanical work and heat are equivalent to each other”, i.e., whenever work is converted into heat or heat into work, the quantity of work is mechanically equivalent to the quantity of heat. If W is the work corresponding to heat H , then

$$W \propto H$$

$$\text{or} \quad \frac{W}{H} = J \text{ (constant)} \quad (24.1)$$

where J is known as the mechanical equivalent of heat and has value 4.2×10^7 ergs/cal in C.G.S. units and 4.2 J/cal in M.K.S. units. From the first law of thermodynamics it is clear that the law of conservation of energy in mechanics can be generalized to include heat energy also.

Many attempts have often been made to construct engines which may continue to operate without any expenditure of fuel and may thus create energy out of nothing. None of these attempts have proved to be successful. The first law of thermodynamics makes a formal assertion that such attempts are futile and can never be successful.

For infinitesimal change in a system, if dU and dH are the changes in internal energy and absorbed heat with dW work done by the system, then according to the first law of thermodynamics

$$dU = dH - dW \quad (24.2)$$

$$\text{or} \quad dH = dU + dW \quad (24.3a)$$

For P, V, T system, $dW = PdV$. Thus Eq. (24.2) can be given as

$$dH = dU + PdV \quad (24.3b)$$

Equations (24.2) and (24.3) are the mathematical forms of the first law of thermodynamics, according to which the supplied thermal energy to a system is partially used for increasing its internal energy and the rest appears as the work done on the system.

24.2.3 Second Law of Thermodynamics

The first law of thermodynamics tells us that heat and mechanical work both are interconvertible. But it does not tell us the limitation and condition for this conversion, i.e., how much heat is converted into work and whether the transformation itself can take place or not. Actually the law specifying the condition of transformation of heat into work is called the second law of thermodynamics.

In order to explain the conditions and limitations for the conversion of heat energy into work and vice versa, the second law of thermodynamics was put forward which may be stated in the following two ways.

Kelvin Statement

According to Kelvin, the second law of thermodynamics can be stated as:

"It is impossible to get a continuous supply of work from a body by cooling it to a temperature lower than that of its surroundings."

For better understanding of the above statement, let us consider the case of a heat engine in which the working substance does some work and rejects the remaining heat to the sink (heat reservoir having infinite thermal capacity). For such a heat engine the temperature of the source must be higher than that of the surroundings and the engine will not work when the temperatures of the source and sink are the same.

Clausius Statement

According to the clausius statement of second law of thermodynamics,

"It is impossible to make heat flow from a body at a lower temperature to a body at a higher temperature without doing external work on the working substance."

This statement is applicable to ice plants and refrigerators. Heat itself cannot flow from a body at a lower temperature to a body at a higher temperature. But this is possible at the cost of some external work done on the working substance.

Thus, the second law of thermodynamics plays an important role in practical devices, e.g., heat engines and refrigerators. The first law of thermodynamics gives only the relation between the work done and the heat produced. But the second law of thermodynamics gives the condition under which heat can be converted into work.

Solved Examples

Example 24.1

When a system is taken from state A to state B, along the path ACB, 80 J of heat flows into the system and the system does 20 J of work (Fig. 24.2).

- How much heat flows into the system along the path ADB, if the work done is 20 J?
- The system is returned from state B to state A along the curved path. The work done on the system is 10 J. Does the system absorb or liberate heat? If yes, how much?
- If $U_A = 0$, $U_D = 40$ J, find the heat absorbed in the processes AD and DB.

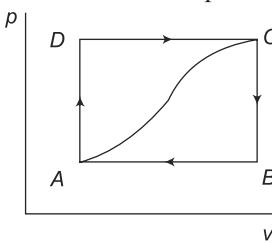


Fig. 24.2 P-V diagram of the system in Example 24.1

Solution

According to the first law of thermodynamic, along the path ACB

$$H_{ACB} = U_B - U_A + W$$

$$H = +80 \text{ J}$$

$$W = +20 \text{ J}$$

$$\begin{aligned}\text{Now } U_B - U_A &= H - W \\ &= 80 - 20 \\ &= 60 \text{ J}\end{aligned}$$

(i) Along the path ADB

$$W = 20 \text{ J}$$

$$H_{ABD} = U_B - U_A + W$$

$$\begin{aligned}\text{or } H &= 60 + 20 \\ &= 80 \text{ J}\end{aligned}$$

(ii) From the curved path B to A

$$W = -10 \text{ J}$$

$$\begin{aligned}H &= (U_A - U_B) + W \\ &= -60 - 10 = -70 \text{ J}\end{aligned}$$

(Negative sign shows that heat is liberated by the system)

$$\begin{aligned}\text{(iii) } U_A &= 0, U_D = 40 \text{ J} \\ U_B - U_A &= 60 \text{ J} \\ \therefore U_B &= 60 \text{ J}\end{aligned}$$

In the process $A \xrightarrow{DB} B$, 20 J of work is done. Work done from A to D is 20 J and D to B is zero.

$$\begin{aligned}\text{For } AD \quad H_{AD} &= U_D - U_A + W \\ &= 40 + 20 = 60 \text{ J}\end{aligned}$$

$$\begin{aligned}\text{For } DB \quad H_{DB} &= U_B - U_D + W \\ &= 60 - 40 + 0 = 20 \text{ J.}\end{aligned}$$

Example 24.2

The temperature of 5 gm of air is raised from 0°C to 1°C . What is the increase in the internal energy of air?

Given that $C_V = 0.172 \text{ cal g}^{-1} \text{ }^\circ\text{C}^{-1}$ and $J = 4.18 \times 10^7 \text{ erg/cal}$.

Solution

According to the first law of thermodynamics,

$$dH = dU + dW$$

If volume is constant, $dW = 0$, because $dW = PdV$

$$\text{Now } dH = dU$$

It means the entire heat given to the system is spent in increasing the internal energy.

$$\begin{aligned} dH &= dU = m C_V dT \\ &= 5 \times 0.172 \times 1 \text{ cal} \\ &= 5 \times 0.172 \times 4.18 \times 10 \text{ erg} \\ &= 3.595 \times 10^7 \text{ erg} \end{aligned}$$

Example 24.3

At atmospheric pressure 10 kg water is converted into steam which occupies the volume of 20 m^3 . Calculate the work done on the system and increase in the internal energy. Given that the latent heat of steam (L) = $2.263 \times 10^6 \text{ J/kg}$.

Solution

$$\text{Initial volume } V_i = \frac{m}{\rho} = \frac{10}{10^3} = 10^{-2} \text{ m}^3$$

$$\text{Final volume} = 20 \text{ m}^3$$

Now the work done in an isobaric process

$$\begin{aligned} W &= P(V_f - V_i) \\ &= 1.01 \times 10^5 (20 - .01) \\ W &= 20.1899 \times 10^5 \text{ J} \end{aligned}$$

Heat given to the system

$$\begin{aligned} H &= mL \\ &= 10 \times 2.26 \times 10^6 \end{aligned}$$

$$\text{or } H = 22.6 \times 10^6 \text{ J}$$

According to the first law of thermodynamics,

$$dH = dU + dW$$

$$\text{or } H = dU + W$$

$$\begin{aligned} \text{Now } dU &= H - W = 22.6 \times 10^6 - 20.1899 \times 10^5 \\ &= 20.58 \times 10^6 \text{ J} \end{aligned}$$

Example 24.4

A quantity of dry air at 27°C is compressed (i) slowly (ii) suddenly to 1/3 of its volume. Find the change in temperature in each case. (Given that $\gamma(\text{air}) = 1.4$)

Solution

- (i) Slow compression is isothermal and hence there is no change in temperature.
- (ii) Sudden compression is an adiabatic change in which we use the equation

$$T_1 V_1^{\gamma-1} = T_2 V_2^{\gamma-1} \quad (\text{i})$$

According this problem, $V_1 = V$, $V_2 = V/3$

$$T_1 = 300 \text{ K}, T_2 = ?$$

From Eq. (i), we have

$$\begin{aligned} T_2 &= \left(\frac{V_1}{V_2} \right)^{\gamma-1} \times T_1 \\ &= (3)^{0.4} \times 300 \end{aligned}$$

$$\text{or} \quad T_2 = 192.5 \text{ }^{\circ}\text{C}$$

Now increase in temperature

$$= 192.5 \text{ }^{\circ}\text{C} - 27$$

$$= 165.5 \text{ }^{\circ}\text{C}$$

Example 24.5

Air is compressed adiabatically to half its volume. Calculate the change in its temperature. For air $\gamma = 1.4$.

Solution

Let us consider $T_1 \text{ K}$ and $T_2 \text{ K}$ are the initial and final temperatures of the air, respectively. If V_1 and V_2 are its initial and final volumes, then

$$V_2 = \frac{V_1}{2}$$

During adiabatic process

$$T_1 V_1^{\gamma-1} = T_2 V_2^{\gamma-1}$$

$$\text{or} \quad T_2 = T_1 \left[\frac{V_1}{V_2} \right]^{\gamma-1}$$

$$T_2 = T_1 [2]^{1.4-1}$$

$$= T_1 [2]^{1.4-1}$$

$$\text{or} \quad = T_1 [2]^{0.4}$$

$$T_2 = 1.319 T_1$$

Now change in temperature

$$\begin{aligned} dT &= T_2 - T_1 \\ &= 1.319 T_1 - T_1 \\ &= 0.319 T_1 \end{aligned}$$

Example 24.6

One mole of a monoatomic gas ($\gamma = 5/3$) is initially at 27°C. It is compressed adiabatically from an initial pressure of 1 atm to 27 atm. Calculate the final temperature of the gas.

Solution

For an adiabatic process, we can write

$$\begin{aligned} \frac{T_1^\gamma}{P_1^{\gamma-1}} &= \frac{T_2^\gamma}{P_2^{\gamma-1}} \\ T_2^\gamma &= \left(\frac{P_2}{P_1} \right)^{\gamma-1} T_1^\gamma \\ (T_2)^{5/3} &= (27)^{2/3} (300)^{5/3} \end{aligned}$$

$$\text{or } \frac{5}{3} \log T_2 = \log 9 + \frac{5}{2} \log 300$$

$$\text{Now } T_2 = 1121.16 \text{ K}$$

24.2.4 Third Law of Thermodynamics or Nernst Heat Theorem

This principle governs the behaviour of systems as their temperature approaches absolute zero. In usual condition, in all heat engines, there is always a loss of heat in the form of conduction, radiation, and friction. Therefore, in actual heat engines entropy H_1/T_1 is not equal to H_2/T_2 , where H_1 is the heat absorbed by the working substance at the temperature T_1 and H_2 is the heat rejected to the sink at the temperature T_2 .

According to the above discussion, $H_1/T_1 - H_2/T_2$ is not zero but it is a positive quantity. It means that when the different cycles of a reversible process are repeated, entropy of the system increases and tends to reach a maximum value. When the system has attained the maximum value, a stage of stagnancy will be reached and no work can be done by the engine at this stage.

Since entropy is the measure of the disorder of a system, so, in substance, an increase in entropy will increase the disorder of its molecules. And with a decrease in entropy, the disorder decreases.

Using quantum statistics, Nernst proved that

$$\lim_{T \rightarrow 0} dS = 0$$

where dS is the change in entropy in an infinitesimal isothermal process.

$$\text{or} \quad \lim_{T \rightarrow 0} (S_1 - S_2) = 0 \quad (24.4)$$

This condition is known as the *Nernst heat theorem*, according to which, "In the neighbourhood of absolute zero, all reactions in a liquid or solid in internal equilibrium take place with no change in entropy."

During 1910 Planck went a step further in that not only the entropy difference vanishes at $T \rightarrow 0$ but entropy of every condensed phase (i.e. solid or liquid) vanishes individually at absolute zero temperature, i.e.,

$$\lim_{T \rightarrow 0} S = 0$$

This is known as the third law of thermodynamics.

We can conclude some interesting results from the third law of thermodynamic as follows:

- (i) Coefficient of volume expansion γ vanishes at absolute zero temperature.
- (ii) Heat capacity vanishes at absolute zero temperature.
- (iii) Absolute zero is unattainable by a finite change in parameters. This is also known as the *principle of unattainability of absolute zero*, according to which, "It is impossible by any procedure, no matter how idealized, to reduce any system to absolute zero temperature in a finite number of operations."

24.3 THERMODYNAMIC PROCESSES

The thermodynamical state of a system can be defined with the help of the thermodynamical coordinates of the system. Thermodynamical state of a system can be changed with the alteration in its thermodynamical coordinates. Change in the state of a system with the alteration in its thermodynamical coordinates is known as a *thermodynamic process*. During any process performed on or with a system, its physical as well as chemical condition may change and this is generally associated with a change in the form of some of its energy content due to the transformation of heat energy into work. If the system does not interact with the surroundings, it is classed as a *closed system*, and its energy remains constant throughout. If the system returns to its initial state after undergoing a series of operations, it is said to perform a *cyclic process*.

In order to explain thermodynamic processes further, let us consider two states of a system, say, state *A* and state *B*. Change of state from *A* to *B* or vice versa is a process and the direction of the process will depend upon a new thermodynamical coordinate called entropy. It is observed that all process are not possible in universe. For better understanding, following thermodynamical process can be considered:

- (i) Let us consider two blocks *A* and *B* at temperatures T_1 and T_2 , respectively ($T_1 > T_2$). These two blocks are kept in contact but the system as a whole is insulated from the surrounding. Now the conduction of heat takes place between the blocks, due to which the temperature of *A* falls and the temperature of *B* rises and thermodynamical equilibrium will be reached.
- (ii) Let us consider the case of a flywheel rotating with an angular velocity ω . Initially, this flywheel has kinetic energy — . If no external force is applied to the flywheel, then after some time the wheel comes to rest and its kinetic energy is utilized in overcoming friction at the bearings.

The temperature of the wheel and the bearings rises and the increase in their internal energy is equal to the original kinetic energy of the flywheel.

- (iii). Let us consider two flasks *A* and *B* which are connected with a glass tube provided with a stopcock. Out of these two flasks, *A* contains air at high pressure and *B* is evacuated. This entire system is isolated from the surroundings. Now if the stopcock is opened, air rushes from *A* to *B*. The pressure in *A* decreases and the volume of air increases.

Although the above three examples are different, they refer to the thermodynamical processes occurring due to the change in thermodynamical coordinates. In these processes the law of conservation of energy is not violated because the total energy of the system is conserved, which satisfies the first law of thermodynamics. To determine the direction of the process, a new thermodynamical coordinate has been devised by Clausius, named as entropy of the system. Similar to energy, entropy is also a function of the state of a system. For any possible process, the entropy of an isolated system should increase or remain constant. The process in which there is a possibility of decrease in entropy cannot take place.

In order to study the transformation of heat energy into work, all the thermodynamic processes can be classified into two groups:

- (i) Reversible process
- (ii) Irreversible process

24.3.1 Reversible Process

A *reversible process* is that which can be retraced in the reverse direction, i.e., if a minute change is made in the reverse direction, the process is reversed completely and the working substance undergoes exactly the same conditions as it does in the direct process. This process should take place with extremely slow rate. During this process, there should not be any loss of heat due to friction or radiation.

In order to explain this process, let us consider an example of a cylinder containing a gas at a certain pressure and temperature. The cylinder is fitted with a frictionless piston. If the pressure is decreased, the gas expands slowly and maintains a constant temperature (isothermal process). The energy required to maintain the temperature of the gas due to expansion is drawn continuously from the source (surroundings). If the pressure on the piston is increased, the gas contracts slowly and maintains a constant temperature (isothermal process). The energy liberated during compression is given to the sink (surroundings). This is also true for an adiabatic process provided the process takes place infinitely slowly.

The following conditions are required for a reversible process or heat engine:

- (i) The temperature and pressure of the working substance must not differ appreciably from those of the surroundings at any stage of the cycle of operation.
- (ii) All the processes occurring in the cycle of operation must be infinitely slow.
- (iii) The working part of the engine must be completely free from friction.
- (iv) There should not be any loss of energy due to conduction or radiation during the cycle of operation.

Here it is important to mention that the complete reversible process or cycle of operation is only an ideal case. In an actual process, there is always a loss of heat due to friction, radiation, and conduction, etc. It is also observed that the temperature and pressure of the working substance differ appreciably from those of the surroundings.

24.3.2 Irreversible Process

A thermodynamic process which cannot be retraced is called an *irreversible thermodynamical process*. In other words, if the whole system (including the surroundings) cannot be restored to its initial condition everywhere, even with the use of all possible physical means, the process is said to be *irreversible*.

In the case of thermodynamical process, it is observed that “the processes in which the entropy of an isolated system decreases do not take place, or for all processes taking place in an isolated system, the entropy of the system should increase or remain constant.” It means that a process is irreversible if entropy decreases when the direction of the process is reversed. During this process a part of heat energy is always utilized to overcome friction. Energy is also dissipated in the form of radiation and conduction. This loss of energy always takes place whether the engine works in one direction or the reverse direction. Such energy cannot be regained. In actual practice all the engines are irreversible. Some examples of the irreversible process are as follows:

- (i) Work done against friction is irreversible because even if the direction of motion is reversed, work has to be done against friction.
- (ii) If electric current is passed through a wire, heat is produced. If the direction of the current is reversed, heat is again produced.
- (iii) Almost all the chemical reactions and natural processes are irreversible.

24.4 THERMODYNAMIC VARIABLES

The behaviour of a system and its introduction to the surroundings or both can be studied through two points of view:

- (i) Macroscopic point of view
- (ii) Microscopic point of view

24.4.1 Macroscopic Quantities

A macroscopic description of a system involves the specification of a few fundamental properties of the system without enquiring the internal structure of the matter or system. *Macroscopic quantities* refer to gross characteristics or large-scale properties of system and provide a macroscopic description. Therefore, these are known as *macroscopic coordinates* or *variables*. A macroscopic description of a system involves the specification of a few fundamental measurable properties of the system, such as volume, pressure, and temperature, etc.

24.4.2 Microscopic Quantities

Microscopic quantities are concerned with the internal structure of the matter or system. These quantities postulate the existence of molecules, their motion, their energy states, their interactions, etc. These quantities refer to the small-scale properties of the system and provide a microscopic description. They are therefore called *microscopic coordinates* or *variables*.

Although the above two points of view to study a system or some matter are entirely different and incompatible, when these two are applied to the same system, they must lead to the same conclusion. In fact the macroscopic properties are really average of time of a large number of microscopic characteristics over a period of time. For example, the macroscopic quantity, pressure, is the average rate of change of momentum due to all the molecular collisions made on a unit area of the container of a gas.

25

Semiconductor Physics

25.1 INTRODUCTION

Among all the classes of solids semiconductors are the most interesting and useful substances. These materials exhibit a wide spectrum of phenomena, covering the entire range from the strictly metallic to the insulator, having extremely versatile in terms of applications. The variety of physical devices employing semiconductors is truly impressive. Although semiconductors have been studied for a long time since the 1920's but they become more popular after the invention of transistor during 1940's by Shockley, Bardeen and Brattain. With the invention of transistors and development of other related devices, semiconductors have become the most actively studied substances in solid state physics. During these studies we have learned much about the basic properties of these solids, and how to utilize them in designing still newer and more efficient devices.

This chapter is devoted to the basic physical properties of semiconducting materials such as qualitative discussion of band theory of solids with special reference to the band structure of pure semiconductors, classification of solids on the basis of band theory, charge density in semiconductor, effect of temperature and doping on the band structure and carrier concentration, extrinsic semiconductors, idea of effective mass, density of state expression of density of states in valence band and conduction band, fermi distribution function and importance of fermi level, carrier concentration in valence and conduction band of intrinsic and extrinsic semiconductors, position of fermi level in intrinsic and extrinsic semiconductors, mobility of electron and holes in valence band conduction band, expression for conductivity, resistivity, expression for conductivity of intrinsic semiconductors and experimental determination of band gap in intrinsic semiconductors.

25.2 ENERGY BANDS IN SOLIDS

In an isolated atom, there are discrete energy levels as $1s$, $2s$, $3s$, ... which are occupied by the electrons of the atom, as shown in Fig. 25.1(a). If all the atoms of a solid are assumed to be isolated from one another, then they have completely identical electronic schemes of their energy levels. When these isolated atoms are brought together to form a solid, various interactions occur between the neighbouring atoms. During this interaction, the higher energy levels are affected considerably, and consequently, the energy levels of the outer shells are slightly altered without violating Pauli's exclusion principle. Hence, the single energy levels of atoms split into large number of energy levels as shown in Fig. 25.1(b). Since, there are

many atoms in a solid, so the separation between the split energy levels is very small. Hence, the split energy levels are almost continuous and are said to form an energy band.

There are energy bands in a solid corresponding to the energy levels in an atom. An electron in a solid can have only those discrete energies that lie within these energy bands. These energy bands are known as the allowed energy bands, which are generally separated by some energy gap known as the forbidden energy bands or not allowed energy bands. Energy band occupied by the valence electrons is known as the valence band, and the energy band which is empty or occupied by the conduction electrons is known as the conduction band.

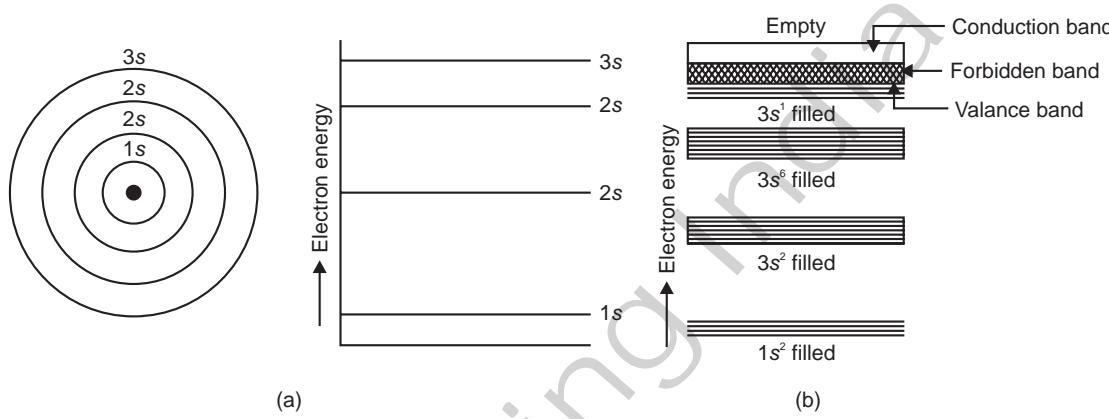


Fig. 25.1 (a) Electrons of an isolated atom in discrete energy levels and **(b)** energy bands formed due to the interaction between different sodium atoms in a solid

In order to clarify the concept of valence band and conduction band. Let us take an example of sodium crystal. This crystal has many closely packed sodium atoms, which is usually referred to as crystal lattice. Each sodium atom has eleven electrons with electronic configuration $1s^2, 2s^2, 2p^6$, and $3s^1$. In this electronic configuration, all the subshells are filled except $3s^1$. The outermost electron $3s^1$ of this atom is called valence electron. When more number of atoms are brought together to form a solid, then the outer energy levels are split to form an energy band (i.e., large number of split energy levels closely spaced). In this band, there are two regions: one region is occupied by the valence electrons known as, valence band, whereas the other region is either empty or occupied by the conduction electrons (called conduction band) as shown in Fig. 25.1(b). Valence band is also known as the highest occupied band, whereas the conduction band is the lowest unoccupied band. The forbidden energy gap between the conduction band and the valence band is known as the energy band gap.

25.3 CLASSIFICATION OF SOLIDS ON THE BASIS OF ENERGY BAND DIAGRAM

On the basis of the forbidden energy gap, solid are broadly classified into following three categories:

- (i) conductors (metals), (ii) insulators, (iii) semiconductors,

The electrical properties of the materials can be explained by the concept of energy bands. Depending on the nature of band occupation by electrons and the width of forbidden bands, some important characteristics of above said class of solids are discussed below.

25.3.1 Conductors (Metals)

In case of conductors, there is no forbidden gap between the valence band and the conduction band. It is observed that the valence band overlaps with the conduction band in metals as shown in Fig. 25.2. There are sufficient number of free electrons, available for electrical conduction. Since, conduction band and valence band overlap in case of metals, so the electrons from valence band freely enter the conduction band. Most important point in conductors is that due to the absence of forbidden gap between the conduction band and the valence band, there is no chance of the presence of holes. Hence, the total current in conductors is simply due to the flow of electrons only.

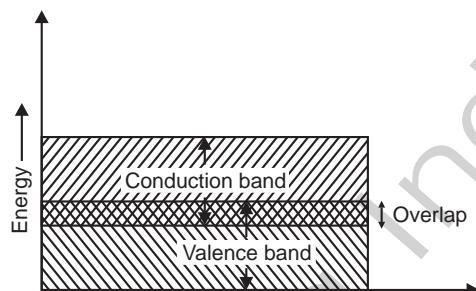


Fig. 25.2 Conduction band and valence band conductors

For further discussion, let us consider the example of sodium, which has electronic configuration as $1s^2, 2s^2, 2p^6, 3s^1$. From this configuration, it is clear that $1s, 2s, 2p$ bands in sodium are completely filled, while the outermost ($3s$) band has only one electron and can accommodate one more electron. Thus, the outer valence band is half-filled. Under the influence of external electric field, electrons of valence band acquire sufficient kinetic energy to move towards the conduction band, where they can move freely to participate in the process of conduction.

Thus, we can conclude that in conductors, there is overlapping between valence and conduction bands. Hence, valence band itself behaves as the conduction band. Due to this reason, metals in general have high electrical conductivity.

25.3.2 Insulators

A class of solids behaves as insulator if it satisfies the following two conditions.

- (i) it has even number of valence electrons per atom and
- (ii) the valence band and the conduction band are separated by a large energy gap compared with kT .

The band diagram of insulators is given in Fig. 25.3(a). Actually, due to the large band gap, electrons cannot jump from the valence band to the conduction band. In insulators, the valence electrons are bound tightly to their parent atoms. For example, in case of glass, the valence band is completely filled at 0 K and energy band gap between the valence band and the conduction band is of the order of 10 eV. An external electric field cannot provide this amount of energy to an electron in solid. Therefore, such solids behave as insulators because no free electrons are available to conduct electricity. Hence, we can conclude that the insulators are those solids which have full valence band and empty conduction band with a large energy band gap between the valence band and the conduction band.

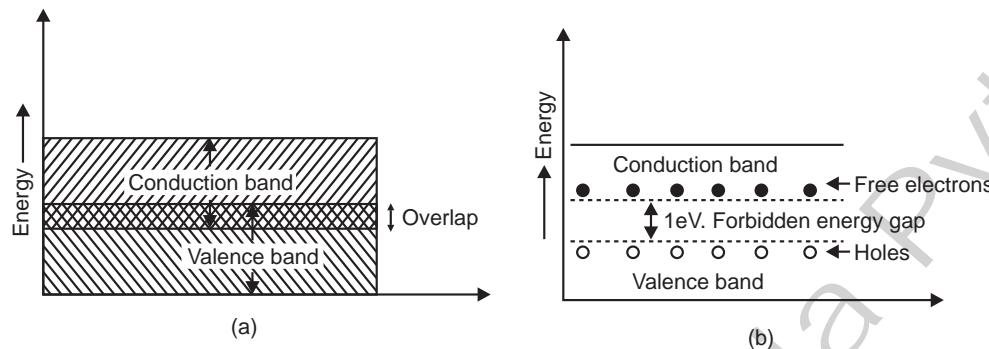


Fig. 25.3 Conduction band and valence band in insulator and semiconductors

25.3.3 Semiconductors

In case of semiconductors, the energy band gap (forbidden gap) between the filled valence band and the empty conduction band is small as compared to the insulators and more as compared to the conductors. Usually, this energy band gap is of the order of 1 eV. Energy band structure of semiconductors is shown to Fig. 25.3(b). Germanium and silicon are the common examples of semiconductors. In germanium the forbidden energy gap is 0.7 eV, whereas in case of silicon, it is 1.1 eV. Due to the small energy band gap between the valence band and the conduction band, the electrons of valence band can be thermally excited to the conduction band, where they are able to conduct small current on application of electric field to the crystal. At 0 K temperature, a semiconductor behaves as an insulator, but at room temperature a semiconductor has the following peculiar properties:

- A partially filled conduction band
- A partially filled valence band
- A very narrow energy band gap (\approx eV)

25.4 THEORY OF BAND FORMATION IN SOLIDS

According to the Bohr's quantum mechanical theory for isolated atom electron have only discrete energy levels considering the simplest atom (hydrogen atom) the energy of n th orbit of an isolated atom can be given as

$$E_n = -\frac{q^4 m_0}{8 \epsilon_0^2 h^2 n^2} \quad (25.1)$$

where

m_0 = mass of free electron

q = charge of electron

ϵ_0 = Permittivity of free space

h = Plank's constant

n = Principal quantum number

Energy expressed in equation (25.1) is usually measured in electron volt which is abbreviated as eV.

Putting the values of different parameters in equation (25.1) we get

$$E_n = -\frac{13.6}{n^2} \quad (25.2)$$

where $n = 1, 2, 3 \dots$ and negative sign indicate that electrons are bound in their respective energy levels.

According to the detailed study it is observed that for higher principal quantum number ($n \geq 2$) energy levels split according to their angular momentum quantum number ($l = 0, 1, 2, \dots, n-1$).

When two identical atoms are brought closer, the doubly degenerate energy level split into two closely spaced energy levels, as shown in Fig. 25.4. The splitting of energy level is the consequence of interaction of wave functions of electrons of individual energy level.

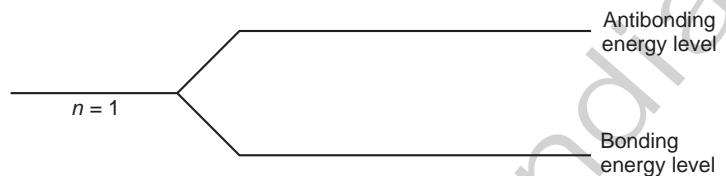


Fig. 25.4 Splitting of single energy level into doubly degenerate energy level.

Similar to above discussion when N isolated atoms are brought together to form a solid, the orbits of the outer electrons of different atoms overlap and interact with other. This interaction including those forces of attraction and repulsion between atoms, causes a shift in the energy levels. Due to this interaction instead of two levels, N separate but closely levels are formed. For the large value of N result is an essentially continuous band of energy. Depending on the interatomic spacing for the crystal the band of N levels can extend over a few eV. The splitting of degenerate state into a band of allowed energies is shown in Fig. 25.5, where a is the equilibrium interatomic distance.

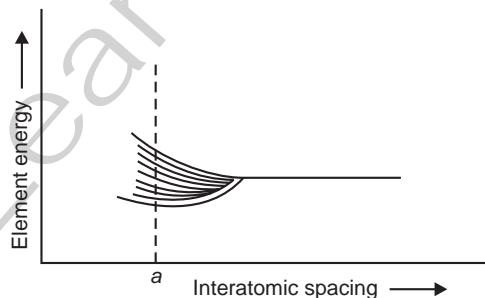


Fig. 25.5 Splitting of degenerate state into a band of allowed energies.

25.5 BAND FORMATION IN PURE SEMICONDUCTOR (SILICON)

In case of semiconducting materials the actual band splitting is much more complicated. In order to illustrate the formation of band in pure semiconductor let us consider the case of silicon. Isolated atom of silicon have 14 electrons which has been shown in Fig. 25.6. Out of these 14 electrons, 10 occupy deep lying energy levels whose orbital radius is much smaller than the interatomic separation in the crystal. The four remaining valence electrons of outer energy level ($n = 3$) are loosely bound and can be involved in the chemical reactions or in the process of conduction under the influence of external

impetus. Thus we need to consider only the outer shell (for $n = 3$) for the valence electrons, because rest two inner shells are completely full and tightly bound to the nucleus. The $3s$ subshell (i.e. for $n = 3$, and $l = 0$) has two allowed quantum states per atom. This subshell contains two valence electrons at $T = 0$ K. The $3p$ subshell (for $n = 3$ and $l = 1$) has six allowed quantum state per atom and contains remaining two valence electrons.

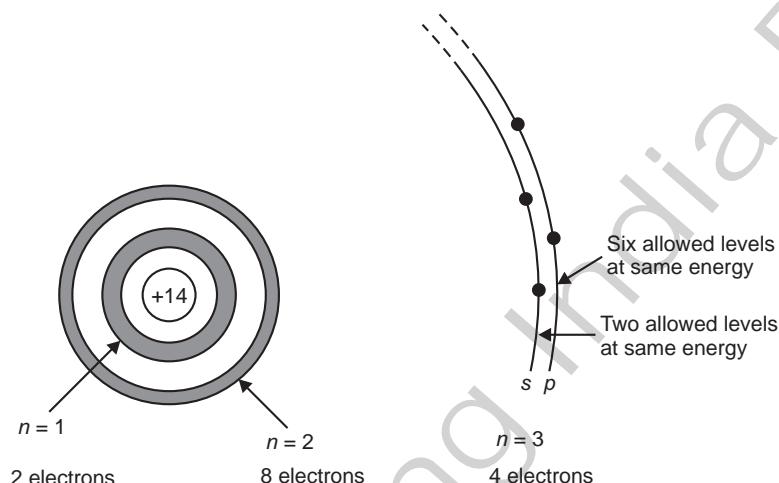


Fig. 25.6 Isolated atom of silicon with 14 electrons.

The schematic diagram of the formation of band in a silicon crystal with N isolated silicon atoms is shown in Fig. 25.7.

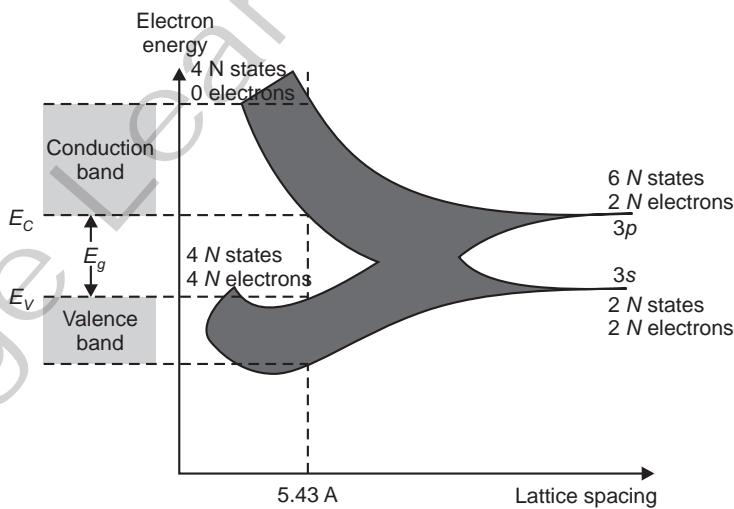


Fig. 25.7 Schematic diagram of the formation of bands in silicon crystal with N atoms.

As the interaction distance decreases, the $3s$ and $3p$ subshell of N silicon interact and overlap to form bands. When $3s$ and $2p$ bands grow, they merge into a single band and this band (band of $3s$ - $3p$ levels) contains $8N$

available states. At the equilibrium interatomic spacing of silicon, the band splits into two bands separated by an energy gap E_g . The upper band contains $4N$ states called conduction band. The lower band called valence band, contains rest $4N$ states as it is shown in Fig. 25.7. At the temperature $T = 0$ K, the electron occupy the lowest energy states, so that all the states in the lower band (the valence band) will be full and all states in the upper band (conduction band) will be empty. The lowest energy level of conduction band (bottom of conduction band) is called E_C , and highest energy level (top of the valence band) is known as E_V .

The energy band gap between the bottom of the conduction band top of the valence band ($E_C - E_V$) is known as E_g as it is indicated in the left of Fig. 25.7. Physically, E_g is the energy required to break a band in the semiconductor to free the electron to the conduction band which leave a hole in the valence bond.

25.6 IDEA OF EFFECTIVE MASS OF ELECTRON

It is observed that a moving electron in the crystal lattice of a solid material faces interaction with the atoms which affect its movement. Due to this reason the movement of electron is not described by Newton's law of motion. In order to describe the motion of electron according to the Newton's law of motion the concept of effective mass is introduced. Effective mass of electron may be different due to the different circumstances.

Let us consider on electron inside a crystal in a state k with the electronic wave function $\exp(ik)$. For a wave packet the group velocity can be given as

$$v = \frac{dw}{dk} = \frac{1}{\hbar} = \frac{d(\hbar w)}{dk} = \frac{1}{\hbar} \left(\frac{dE}{dk} \right) \quad (25.3)$$

Under the influence of external electric field, the force experienced by the electron is

$$F = \frac{dp}{dt} = m \frac{dv}{dt} \quad (25.4)$$

where p is the momentum of electron.

Now putting the value of v in equation (25.4) from equation (25.3) and using the relation $p = \hbar k$ we get,

$$\begin{aligned} \frac{\hbar}{m} \frac{dk}{dt} &= \frac{d}{dt} \left[\frac{1}{\hbar} \frac{dE}{dk} \right] \\ \text{or } \frac{\hbar}{m} \frac{dk}{dt} &= \frac{1}{\hbar} \frac{d}{dt} \left[\frac{dE}{dk} \right] \\ &= \frac{1}{\hbar} \frac{d}{dk} \left(\frac{dE}{dk} \right) \frac{dk}{dt} \\ \text{or } \frac{1}{m} &= \frac{1}{\hbar^2} \left(\frac{d^2 E}{dk^2} \right) \end{aligned}$$

Now replacing m by m^* in above equation we get

$$m^* = \frac{\hbar^2}{\frac{d^2 E}{dk^2}} \quad (25.5)$$

Equation (25.5) gives the idea of effective mass of an electron moving in the crystal.

Effective mass of electron has a special importance in solid state electronics because the results of free electron theory can be applied in band theory of solids by replacing the rest mass m of electron by effective

mass of electron. It is used in transport calculations, such as transport of electrons under the influence of fields or carrier gradients in different semiconductors. Also used to calculate the density of states.

25.7 BAND STRUCTURE IN SEMICONDUCTING MATERIALS WITH THE CONCEPT OF EFFECTIVE MASS

Semiconductors are the solids in which the highest occupied energy band, the valence band, is completely full at $T = 0^\circ\text{K}$, but in which the gap above this band is also small, so that electrons may be excited thermally at room temperature from the valence band to the next-higher band, which is known as the conduction band. Usually the number of excited electrons is appreciable (at room temperature) whenever the energy gap E_g is less than 2 eV. The material may then be classified as a semiconductor.

When electrons are excited across the gap, they reach to the bottom of conduction band (CB), thus the population of electron increase, and the top of valence band (VB) populated by holes. Due to above process, both bands are now only partially filled, and would carry a current if an electric field is applied. Although the conductivity of semiconductor is very small but it is sufficiently large for many practical purposes.

Only the conduction band and valence bands are of much interest to us here, because only these two bands contribute to the current. Other bands below valence band are completely filled and those above conduction band are completely empty, thus these group of bands contribute no current; hence they may be ignored so far as semiconducting properties are concerned.

Simple band structure of a semiconductor is shown in Fig. 25.8. The energy of the conduction band can be given as

$$E_C(k) = \frac{\hbar^2 k^2}{2m_e^*} + E_g \quad (25.6)$$

where k is the wave vector and m_e^* is the effective mass of the electron. E_g is the energy band gap. The zero energy level lie at the top of valence band.

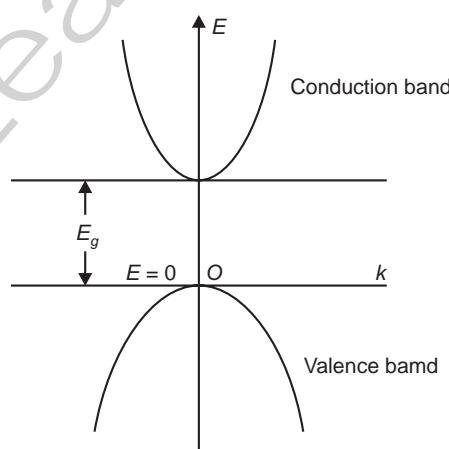


Fig. 25.8 Band structure in semiconductor

The energy of the valence band can be given as

$$E_V(k) = \frac{\hbar^2 k^2}{2m_h^*} \quad (25.7)$$

where m_h^* is the effective mass of the hole. As it is clear from above equations the primary band structure parameters are the electron and hole masses m_e and m_h and the band gap E_g . These parameters for some semiconductors are given in table 25.1

Table 25.1 Parameters for band structure of semiconductors at room temperature

Group	Crystal	E_g , eV	Effective mass, m/m_o	
			Electrons	Holes
IV	C	5.3	$m_l = 0.97, m_t = 0.91$ $m_l = 1.6, m_t = 0.08$	0.5, 0.16
	Si	1.1		0.3, 0.04
	Ge	0.7		
	Sn	0.08		
III-V	Ga As	1.4	0.07	0.09
	Ga P	2.3	0.12	0.50
	GaSb	0.7	0.20	0.39
	In As	0.4	0.03	0.02
	In P	1.3	0.07	0.69
	In Sb	0.2	0.01	0.18

Note

m_l and m_t are the longitudinal and transverse masses, respectively, of ellipsoidal energy.

The conduction and valence bands in semiconductors are related to the atomic states. In order to understand it, let us take the example of hydrogen molecule, when two hydrogen atoms are brought together to form a molecule the atomic $1s$ state splits into two states: a low-energy bonding state and a high-energy antibonding state. In solid hydrogen, these states broaden into bonding and antibonding energy bands, respectively. Similarly, the valence and conduction bands in semiconductors respectively, the bonding and antibonding bands of the corresponding atomic valence states. Thus the valence band and conduction band in Si, for example, result from the bonding and antibonding states of the hybrid $3s^1 3p^3$. Similar explanations can be given to the bands in Ge, C and other semiconductors.

25.8 DENSITY OF STATES

Density of states is very important for the better understanding of electronic processes, particularly in transport phenomena. The density of states for electrons in a band yields the number of states in a certain energy range. It is denoted by $D(E)$ and defined as

$$D(E) dE = \text{The number of electron states per unit volume in the energy range } (E, E + dE) \quad (25.8)$$

In order to evaluate $D(E)$ we use the definition given in equation (25.8). Let us draw a shell in k -space which inner and outer surfaces are determined by the energy contours $E(k) = E$ and $E(k) = E + dE$, respectively, as shown in Fig. 25.9.

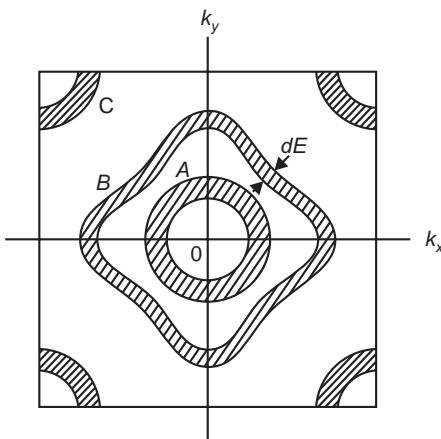


Fig. 25.9 Shells in k -space used to evaluate $D(E)$

The number of k values lying inside this shell then gives the number of states. When this number of states divided by the thickness of the shell dE , yields the desired function $D(E)$.

From the above discussions it is clear that $D(E)$ is closely related to the shape of the energy contours, and hence the band structure. The complexities of this structure are reflected in terms of the form taken by $D(E)$.

In order to evaluate $D(E)$ first we consider the case in which the dispersion relation for electron energy has the standard form,

$$E = \frac{\hbar^2 k^2}{2m^*} \quad (25.9)$$

As we know that the dispersion relation depicted in equation (25.9) are true for those states lying close to the bottom of the band near the origin of the Brillouin zone. The energy contours corresponding to equation (25.9) are clearly concentric spheres surrounding the origin. Thus the resulting density of states will be spherical in shape as illustrated by shell A in Fig. (25.9). The volume of this spherical shell can be given as $4\pi k^2 dk$, where k is the radius and dk is the thickness of the shell. The number of allowed k values per unit volume of k -space is $1/(2\pi)^3$, thus the number of states lying in the shell in the energy range E and $E + dE$ can be given as.

$$\text{Number of states} = \frac{1}{(2\pi)^3} 4\pi k^2 dk \quad (25.10)$$

The right hand side of equation (25.10) can be written in terms of energy rather than in terms of k , using equation (25.9). Now equation (25.10) can be expressed as.

$$\text{Number of states} = \frac{1}{4\pi^2} \left(\frac{2m^*}{\hbar^2} \right)^{3/2} E^{1/2} dE \quad (25.11)$$

Comparing equation (25.11) with the definition of $D(E)$ expressed in equation (25.8). We get

$$D(E) = \frac{1}{4\pi^2} \left(\frac{2m^*}{\hbar^2} \right)^{3/2} E^{1/2} \quad (25.12)$$

26

Mechanical Properties of Matter

26.1 INTRODUCTION

When a system of forces or couples in equilibrium act on a body, there is in general either a change of shape or size of the body. When a body regains completely its original shape and size after the removal of deforming force, it is said to be *perfectly elastic*; while if it completely retains its modified form, it is said to be perfectly plastic, property by virtue of which it resists and recovers from the change of the shape and size is called *elasticity* of the body. There is no perfectly elastic or plastic body. Actually, bodies lie between these two extremes, but all bodies are nearly perfectly elastic if the deformation is small. The nearest approach to a perfectly elastic body is a quartz fiber, and to a perfectly plastic body is putty. It is important to mention here that the deformation in the shape or size of a body is the result of equal and opposite forces or couples in equilibrium.

In this chapter the elastic properties of matter will be discussed. We will describe stress strain-Hooke's law, elastic constants and their relations. Special attention will also be given to the torsional pendulum cantilever, stress-strain behavior of the matters and bending of beam.

26.2 STRESS

When a body is deformed by the application of external forces, displacements of its various particles take place. For such deformation, the body offers resistance, and internal forces of reaction are developed inside the body. These internal forces, so developed are equal in magnitude and opposite to the deforming forces, and equilibrium is established. In fact, these forces of internal reaction tend to restore the body to its original form. Under the action of these restoring forces, the body recovers its shape and size when deforming forces are removed.

The restoring force per unit area set up inside the body, is known as stress and it is measured by the magnitude of the deforming force acting on unit area of the body within the elastic limit. If a force F is acting on an area A then,

$$\text{Stress} = \frac{F}{A}$$

Stress acting normal to the surface is called normal stress. In case of a change in length of the wire or change in the volume of a body, the stress is always normal. Normal stress is of two types: compressive and expansive (i.e. tensile), according to a decrease or increase in volume takes place. The tangential

stress at a surface is also known as shearing stress. Since stress is defined as force per unit area, so its dimensional formula can be given as $ML^{-1}T^{-2}$.

26.3 STRAIN

If a system of forces or couples acting upon a body causes relative displacement of its various parts, a change in length, volume or shape is produced. In this condition body is said to be strained.

“The change of a shape or the fractional change of size of a body by a given set of forces or couples is called strain, and its character evidently depends upon the nature of the forces producing it”.

“Strain is defined as the fractional deformation produced in a body when it is subjected to a set of deforming forces”. Since strain is a ratio so it has no unit depending on the nature of deforming forces and change produced in the shape and size of a body. Strain can be observed as three types as discussed in the below sections.

26.3.1 Longitudinal or tensile strain

Usually, longitudinal strain is produced when forces act along the length of body. In this case a change in the length of body is produced, without any change in its shape. Longitudinal or tensile strain is defined as the ratio of change in length to the original length of the body. For example, if L is the original length of a body and ΔL is the change in the length produced by the force F along its length, as shown in Fig. 26.1, then longitudinal strain is given as

$$\text{Longitudinal strain} = \frac{\Delta L}{L}$$

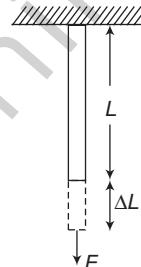


Fig. 26.1 Change in the length of body under action of force F .

26.3.2 Shear or shearing strain

Under the action of tangential force applied on the surface of a body, its shape changes but volume remains same. In this situation body is said to be in sheared state, and occurred strain is known as shearing strain. Thus, the angular deformation produced under the action of tangential force is known as shearing strain or shear.

26.3.3 Volumetric strain

Volumetric strain is defined as the change in volume per unit original volume of a body without any change in its shape.

Let us consider a cube has its initial volume V deformed under the action of normal forces acting on each side of it, as shown in Fig. 26.2. If the change in volume is ΔV without change in its shape then

$$\text{Volumetric strain} = \frac{\Delta V}{V}$$

When the body is deformed, the internal forces which play per unit area to restore it to its original state is called stress. Since the body is in the state of equilibrium, the stress must be equal and opposite to the deforming forces per unit area. Thus, stress is measured as the deforming force or load applied per unit area on the body.

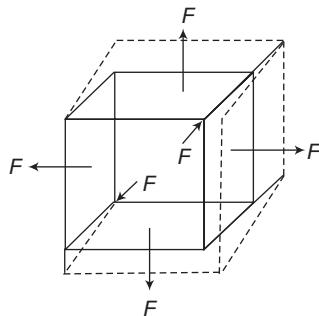


Fig. 26.2 Equal force is acting on each wall of cube due to which volume has been changed

26.4 BEHAVIOUR OF A WIRE UNDER INCREASING LOAD

Let a wire be clamped at one end and loaded at the other gradually until the wire breaks down. It means we are increasing the stress applied on the wire, which leads a change in the strain of the wire. To study the stress-strain behaviour of the wire, let us plot stress on y -axis and strain on x -axis; we get the plot as shown in Fig. 26.3. Explanation of the plot can be divided in different segments as given below

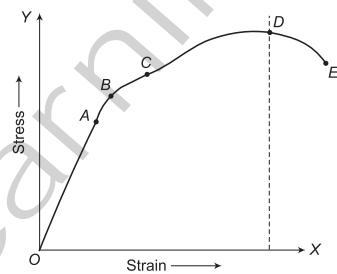


Fig. 26.3 Stress-strain behaviour of a wire

OA region This region of the plot is straight line which shows that stress is proportional to the strain, i.e., Hooke's law is obeyed up to A .

AB region Point B is the elastic limit of the material. This point lies near the point A and up to this point the wire takes back its original length, when the load is removed. In this region stress must not necessarily be proportional to strain.

BC region If the stress is further increased beyond the elastic limit, the stress-strain curve takes a bend. In this region, at any point if wire is unloaded it does not regain its original length and get permanent stretch, which is known as permanent check this word for corrections.

C-D region If load is further increased beyond the point C , the extension increases rapidly without an increase in the load, i.e. the material of the wire flows beyond C , which is known as plastic flow. In this region the wire becomes thin, thus the stress (load per unit area) becomes considerably greater.

D-E region Beyond point D when load is further increased, the thinning of wire no longer remains uniform and diameter of a section decreases considerably, i.e. the specimen wire exhibits a local constriction, causing “necking” in the wire’s behavior and further a breaking point E is reached.

26.5 HOOKE'S LAW

From the close observation on the stress-strain behavior of a material, Robert Hook established, basic law of elasticity in 1678. According to the Hooke's law

“Within the elastic limit, stress is proportional to the corresponding strain”, i.e.

$$\begin{aligned} & \text{Stress} \propto \text{strain} \\ \text{Or} \quad & \text{Stress} = \text{strain} \\ \text{Or} \quad & E = \frac{\text{Stress}}{\text{Strain}} \end{aligned} \quad (26.1)$$

where E is proportionality constant known as modulus of elasticity. Its value is independent to the magnitude of stress and strain, it depends on the nature of material and its other conditions, the dimension and unit of elasticity is same as of stress i.e. N/m^2 and $\text{ML}^{-1}\text{T}^{-2}$. There are three types of modulus of elasticity corresponding to three of strain:

1. Corresponding to longitudinal strain there is *Young's modulus*.
2. Corresponding volume strain there is *Bulk Modulus*.
3. Corresponding to shearing strain there is *Modulus of rigidity*.

26.6 YOUNG'S MODULUS (Y)

Within the elastic limit to which Hooke's law is applicable, the ratio of longitudinal stress to the corresponding strain produced is called Young's modulus for a material of a body and is denoted by ' Y '.

Let us consider a body of original length L and cross sectional area A is elongated by l with the application of a force F . In the state of equilibrium.

$$\begin{aligned} \text{Stress} &= \frac{F}{A} \text{ and strain} = \frac{l}{L} \\ \text{Now Young's Modulus } Y &= \frac{F/A}{l/L} = \frac{FL}{Al} \end{aligned} \quad (26.2)$$

This relation holds good for compressive as well as extensive stresses. Since strain is a ratio which has no unit, so Young's Modulus has same unit as stress i.e. N/m^2

26.7 BULK MODULUS (K)

When a uniform pressure P is applied over the whole surface of a body it produces a uniform compression such that each elementary cube of the body is compressed in the same proportion. If the pressure is not too great. The compression is proportional to the pressure and the ratio of the pressure to the volume strain is known as *bulk modulus* or volume elasticity of the material of the body. If a pressure P causes the volume of the body to reduce from V to $(V - v)$, the compression or volume strain is v/V and hence the bulk modulus is given by

$$K = \frac{P}{-v/V} = \frac{-PV}{v} \quad (26.3)$$

If the compression is not proportional to the pressure we can still define it as the ratio of an infinitesimal increment of pressure from P to $(P + dP)$ to the corresponding volume strain $-\frac{dv}{V}$, hence in general case

$$K = \frac{\frac{dP}{-dV}}{V} = -V \frac{dP}{dV} \quad (26.4)$$

The negative sign indicates here that by the increase of pressure a decrease in volume occurs.

The reciprocal of bulk modulus of a substance is called its compressibility, i.e.

$$\text{Compressibility} = \frac{1}{K} = -\frac{v/V}{P} \quad (26.5)$$

The only stress which a liquid or a gas can permanently sustain is a hydrostatic pressure. Hence, the bulk modulus is the only elastic constant for fluids and it is often called as the elasticity of the liquid or the gas. It has same unit as pressure, i.e., N/m².

26.8 MODULUS OF RIGIDITY OR SHEAR MODULUS

Within the elastic limit, the ratio of tangential stress to the shearing strain is known as the modulus of rigidity of the material of a body and is denoted by a symbol ' η '.

Under the sheared condition the shape of a body changes but its volume remains unchanged.

Let us consider a cube of side L , which is fixed at its lower edge CD . A tangential force F is applied in the direction shown over the upper face of area A . A reaction of same magnitude F acts on the lower fixed face CD in the opposite direction as shown in Fig. 26.4. These two equal and opposite forces constitute a couple. Under the action of this force, couple face AD of the cube is displaced to the position $A'D$ and force BC to the position $B'C$ through an angle θ known as the angle of shear or shearing strain. The rotation of cube under above mentioned couple will be prevented by an equal and opposite moment of force couple provided by the constraints.

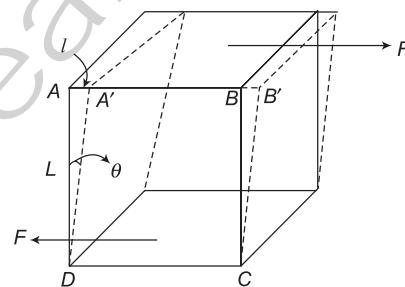


Fig. 26.4 A sheared body under the action of tangential force F .

Now shearing strain $= \theta = \tan \theta = \frac{l}{L}$ for same value of θ .

When $AA' = BB' = l$ and $AD = BC = L$

$$\text{Shearing stress} = \frac{\text{Tangential force}}{\text{Area of the face}} = \frac{F}{A}$$

$$\text{Now the modulus of rigidity } \eta = \frac{F/A}{\theta} = \frac{F}{A\theta}$$

$$\text{Or} \quad \eta = \frac{FL}{Al} = \frac{F}{A\theta} \quad (26.6)$$

Above relation is similar to the expression for Young's modulus (Y), but here in place of linear stress a tangential stress is applied and l , a displacement at right angles L occurs instead of along it. Modulus of rigidity has same unit as Young's modulus (Y) and Bulk modulus (K), i.e., N/m².

26.9 POISSON'S RATIO

In the case of a wire elongated by the application of two equal and opposite forces acting along its length we have assumed that the strain is only a change in length and no change in the diameter of the cross section occurs.

In practice, however, the strain of the wire is not so simple. Poisson's found that in every case of such an elongation there is a contraction of the lateral dimension.

"Within elastic limit there is a complete proportionality between the lateral strain and longitudinal strain, the ratio of lateral strain and longitudinal stain is constant for a material of a body which is called *Poisson's ratio*".

Let us consider a rod has its original diameter and length as D and L , respectively, and if increase in length due to a given longitudinal force is l and reduction of diameter is d , then Poisson's ratio is given as

$$\sigma = \frac{d / D}{l / L} = \frac{Ld}{lD} \quad (26.7)$$

$$\text{Or} \quad \frac{L}{D} \frac{dD}{dL} \quad (26.8)$$

where dL and dD are the change in the length and diameter of rod.

Since both the lateral and longitudinal strains are pure numbers so Poisson's ratio has no unit. The value of Poisson's ratio depends only on the nature of the material, and for homogeneous isotropic materials its value lies between +1/2 and -1.

26.10 RELATION BETWEEN ELASTIC CONSTANTS

The four elastic constants (Y , K , η , and σ) discussed in the preceding sections for a material of a body are not independent to one another but are connected by two equations, and hence, if two of these constants are given any arbitrary values the other two can be determined using the following equations. The two equations connecting the four elastic constants are

$$Y = \frac{9\eta K}{3K + \eta} \quad (26.9)$$

$$\text{And} \quad \sigma = \frac{3K - 2\eta}{6K + 2\eta} \quad (26.10)$$

26.11 SHEARING STRESS EQUIVALENT TO AN EQUAL LINEAR TENSILE STRESS AND EQUAL COMPRESSIVE STRESS AT RIGHT ANGLES TO EACH OTHER

Let us consider a front view of a cube $ABCD$ with its side L as shown in Fig. 26.5. DC is the base of this face which is fixed and force F is applied on the upper face AB .

Now the shearing stress can be given as F/L^2 .

If F is the only force acting on the face AB . Then cube should move as a whole in the direction of applied force. But the cube is fixed at lower face DC . These two forces, therefore, constitute a couple, tendency of which is to rotate cube in clockwise direction, hence equal and opposite force acts on the face DC . Actually, cube does not rotate means that the constituted couple must have been balanced by an equal and opposite couple exerted by the fixed face DC . This couple is exerting an equal and opposite force (F' say) on the faces AD and CB as shown in Fig. 26.5.

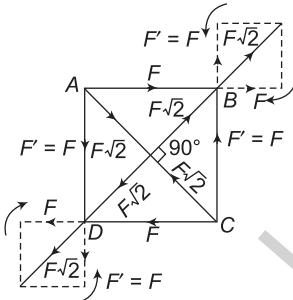


Fig. 26.5 Couple of forces acting on the front face of a cube ABCD

The moment of clockwise couple $= F \times AD$
and moment of anticlockwise couple $= F' \times AB$
since the cube is in the state of equilibrium so

$$F \times AD = F' \times AB$$

$$\Rightarrow F = F' \text{ because } AD = AB = L$$

Thus, the tangential force F applied at face AB is exerting equal tangential forces acting along all faces of cube in suitable direction as shown in Fig. 26.5.

The force acting along the faces AB and CB gives a resultant force $F\sqrt{2}$ along DB and the forces along AD and CD gives a resultant force $F\sqrt{2}$ along BD . Thus, an outward pull is exerted along the diagonal BD at its corners B and D . Under the action of this diagonal pull an extension is produced.

Similarly there will be resultant force $F\sqrt{2}$ acting in inward direction along the diagonal AC acting at the corners A and C , respectively. Under the influence of inward forces acting at A and C there will be contraction along the diagonal AC .

Now the area of the plane drawn perpendicular to the plane of paper through the diagonals AC and BD in the cube can be given as

$$= L \times L\sqrt{2} = L^2\sqrt{2}$$

$$\text{Hence, linear tensile stress along } BD = \frac{F\sqrt{2}}{L^2\sqrt{2}} = \frac{F}{L^2}$$

$$\text{Similarly, linear compressive stress along } AC = \frac{F\sqrt{2}}{L^2\sqrt{2}} = \frac{F}{L^2}$$

From the above discussion it is obvious that F/L^2 is the shearing stress over the face AB of the cube. Hence a shearing stress over is equivalent to an equal linear tensile stress and an equal compressive stress at right angles to each other.

26.12 RELATION BETWEEN Y, K AND σ

In order to obtain a relation connecting elastic constant Y , K and σ , let us consider a cube $ABCDEFGH$ of unit side. Let a force F be applied normally and uniformly on its all six faces acting in the outward direction, since the side of cube is unity so all those forces are supposed to be a force acting on unit area, i.e. stresses.

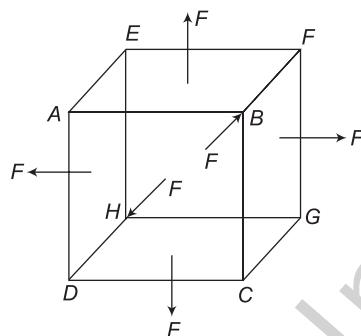


Fig. 26.6 A cube of unit side on each faces uniform normal force F is applied.

If α is the increase per unit length per unit tension along the direction of force, then the elongation produced along each edges, i.e., AB , BF and BC will be $F\alpha$.

If β is the contraction produced per unit stress perpendicular to the edges then contraction produced perpendicular to AB , BF and BC will be $F\beta$. Thus, each edge there is elongation of $F\alpha$ and simultaneously a contraction of $(F\beta + F\alpha)$.

Thus, the sides of cube can be given as

$$AB = 1 + F\alpha + (-F\beta - F\beta) = 1 + F(\alpha - 2\beta)$$

$$BF = 1 + F\alpha + (-F\beta - F\beta) = 1 + F(\alpha - 2\beta)$$

$$BC = 1 + F\alpha + (-F\beta - F\beta) = 1 + F(\alpha - 2\beta)$$

Now the new volume of the cube is

$$AB \times BF \times BC = (1 + F(\alpha - 2\beta))^3$$

$$= 1 + 3F(\alpha - 2\beta) \text{ (using Binomial expansion)}$$

Since α and β are small so higher terms in Binomial expansion have been neglected

Now change in volume = New volume – Initial volume

$$= 1 + 3F(\alpha - 2\beta) - 1$$

$$= 3F(\alpha - 2\beta)$$

$$\text{Volume strain} = \frac{3F(\alpha - 2\beta)}{1} = 3F(\alpha - 2\beta)$$

$$\text{Now Bulk modulus } K = \frac{\text{Normal stress}}{\text{Volumetric strain}} = \frac{F}{3F(\alpha - 2\beta)}$$

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

Since α is defined as increase per unit length per unit stress so linear strain = $\frac{\alpha}{1} = \alpha$ {Because side of cube is unity}

$$\text{Thus } Y = \frac{\text{Linear stress}}{\text{Linear strain}} = \frac{1}{\alpha}$$

And $\beta/\alpha = \sigma$ = poisons ratio

$$\text{Thus } K = \frac{Y}{3(1 - 2\sigma)} \quad (26.11)$$

26.13 RELATION BETWEEN Y , η AND α

In order to obtain relation connecting elastic constants Y , η and σ , let us consider $ABCD$ as the front view of a cube face having its side L . A tangential force F is applied on its upper face AB and the bottom face DC is fixed. As a result of this force the cube is sheared to $A'B'C'D$ through an angle θ as shown in Figure 26.7.

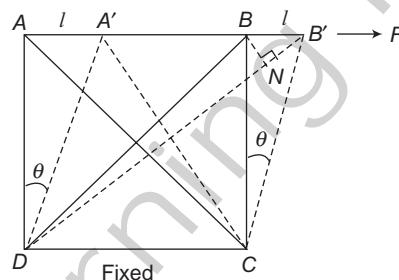


Fig. 26.7 Front view of a face of a cube $ABCD$ under the action of tangential force F .

Shearing strain under the action of tangential force F can be given as

$$\theta = \frac{AA'}{AD} = \frac{BB'}{BC} = \frac{l}{L}$$

Where AA' and BB' are equal displacement produced in the face of a cube.

Shearing stress is given as

$$T = \frac{F}{\text{Area of the upper face of the cube}}$$

$$\text{Or } T = \frac{F}{L^2}$$

Now the modulus of rigidity can be given as

$$\eta = T/\theta$$

But from section 26.10 we know that shearing stress along AB is equivalent to a tensile stress along DB and an equal compressive stress along AC at right angles to each other.

If α and β are the longitudinal and lateral strains per unit stress then the extension along diagonal DB due to the tensile stress is given as $DB.T.\alpha$

And extension along DB due to compression stress along $AC = DB \cdot T \cdot \beta$

Now total extension along $DB' = NB' = DB \cdot T \cdot (\alpha + \beta) = L\sqrt{2}T(\alpha + \beta)$

Increase in the length of DB is NB' where BN is the perpendicular on DB' from B .

Since θ is very small so $\angle AB'C$ is nearly 90° and $\angle BB'N = 45^\circ$

$$\text{Thus, } NB' = BB' \cos 45^\circ = \frac{BB'}{\sqrt{2}} = \frac{l}{\sqrt{2}} \quad (26.13)$$

From equation (26.12) and (26.13) we have

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

$$\text{or } T \frac{L}{l} = \frac{l}{2(\alpha + \beta)}$$

$$\text{But } T \frac{L}{l} = \frac{T}{l/T} = \frac{T}{\theta} = \eta$$

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

$$\text{But } \beta/\alpha = \sigma \text{ and } \gamma = \frac{\text{Stress}}{\text{Longitudinal strain}} = \frac{1}{\alpha}$$

$$\text{Therefore } \eta = \frac{Y}{2(1 + \sigma)} \quad (26.14)$$

26.14 RELATION BETWEEN Y , K AND η

In order to obtain the relation connecting elastic constant Y , K and η , let us use equation (26.11) and (26.14) from which we can write

$$1 - 2\sigma = Y/3K$$

$$2 + 2\sigma = Y/\eta$$

adding above equations we get

$$3 = \frac{Y}{3K} + \frac{Y}{\eta} = \frac{(Y(3K + \eta))}{3K\eta}$$

$$\text{or } \gamma = \frac{9\eta K}{3K + \eta} \quad (26.15)$$

This relation can be rewritten as,

$$\frac{9}{Y} = \frac{3K + \eta}{\eta K} \quad (26.16)$$

$$\text{or } \frac{9}{Y} = \frac{3}{\eta} + \frac{1}{K} \quad (26.17)$$

Appendix

Topics:

- A.1 Gradient, Divergence, and Curl
- A.2 Oscillation
- A.3 Coupled Oscillations
- A.4 Ferromagnetism
- A.5 Weiss Law

A.1 Gradient, Divergence, and Curl

GRADIENT OF A SCALAR FIELD

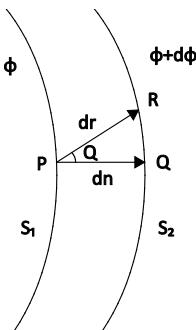
For a continuously differentiable scalar function $\phi(x, y, z)$, gradient is defined as

$$\begin{aligned}\text{grad } \phi &= i \frac{\partial \phi}{\partial x} + j \frac{\partial \phi}{\partial y} + k \frac{\partial \phi}{\partial z} \\ \left(i \frac{\partial}{\partial x} + j \frac{\partial}{\partial y} + k \frac{\partial}{\partial z} \right) \phi &= \nabla \phi\end{aligned}\dots(1)$$

where $\nabla = i \frac{\partial}{\partial x} + j \frac{\partial}{\partial y} + k \frac{\partial}{\partial z}$ is called *del* or *nabla*.

Physical significance: In the scalar field, let there are two surfaces S_1 and S_2 very close to each other characterized by scalar functions ϕ and $\phi + d\phi$ respectively. Consider two points P and R on the level surfaces S_1 and S_2 respectively. Let r and $r + dr$ be the position vectors of P and R respectively relative to any arbitrary origin, then $\overline{PR} = dr$.

The coordinates of P and R are (x, y, z) and $(x + dx, y + dy + z + dz)$ respectively, then



$$dr = idx + jdy + kdz \quad \dots(2)$$

as the value of scalar function at $P(x, y, z)$ and $R(x+dx, y+dy, z+dz)$ are ϕ and $\phi+d\phi$. We may write

$$d\phi = \frac{\partial \phi}{\partial x} dx + \frac{\partial \phi}{\partial y} dy + \frac{\partial \phi}{\partial z} dz \quad \dots(3)$$

$$\begin{aligned} &= \left(i \frac{\partial \phi}{\partial x} + j \frac{\partial \phi}{\partial y} + k \frac{\partial \phi}{\partial z} \right) \cdot (idx + jdy + kdz) \\ &= (\nabla \phi) \cdot dr \end{aligned} \quad \dots(4)$$

In particular, if we consider that the point R (i.e., dr) lies on the level surface S_1 , then $d\phi = 0$. so that

$$(\nabla \phi) \cdot dr = 0 \quad \dots(5)$$

thereby showing that the vector $\nabla \phi$ is normal to the level surface S_1 .

If dn represents the distance along the normal from P to the surface S_2 , then

$$dn = PQ = dr \cos \theta = \hat{n} \cdot dr \quad \dots(6)$$

where \hat{n} is a unit vector normal to the surface S_1 at P . If we promote from P to Q , the value of scalar function ϕ increases by an amount $d\phi$; therefore, we may write

$$d\phi = \frac{\partial \phi}{\partial n} \cdot dn = \frac{\partial \phi}{\partial n} \hat{n} \cdot dr \quad [\text{using (6)}] \quad \dots(7)$$

Comparing (4) and (7), we get

$$\text{grad } \phi = \nabla \phi = \frac{\partial \phi}{\partial n} \hat{n} \quad \dots(8)$$

Thus, the gradient of a scalar function ϕ at any point is a vector whose magnitude is equal to the rate of change of scalar function ϕ along the normal to the level surface at the point.

As $\frac{\partial \phi}{\partial n} \cdot \hat{n}$ gives the greatest rate of increase of ϕ with respect to space variables, $\text{grad } \phi$ may be defined as follows:

The gradient of a scalar function ϕ is a vector whose magnitude is equal to maximum rate of change of scalar function ϕ with respect to space variable and whose direction is along that change.

DIVERGENCE OF A VECTOR FUNCTION

In a certain region of space for a vector field A , divergence at a certain point $P(x, y, z)$ is defined as the outward flux of the vector field A per unit volume enclosed through an infinitesimal closed surface surrounding the point P .

If τ is the infinitesimal volume enclosed by an infinitely small closed surface s surrounding the point $P(x, y, z)$, then the divergence of the vector field A at point P is defined as

$$\operatorname{div} A = \lim_{\tau \rightarrow 0} \frac{\iint_s A \cdot ds}{\tau} \quad \dots(9)$$

The divergence of a vector function is a scalar quantity. It is important to mention that divergence itself is simply an operator and has no physical meaning in itself. After operating on suitable physical vector functions, it represents various significant physical scalar quantities. To understand the physical meaning of the divergence of a vector field, we consider the following two examples:

- (i) The divergence of fluid velocity v at a point represents the quantity of fluid flowing out per second per unit volume enclosed by infinitesimal closed surface surrounding that point. If $\operatorname{div} v$ at a point is negative, it means that fluid is constantly flowing towards that point and thus there exists a sink for the fluid. If $\operatorname{div} v$ at a point is positive, it indicates the existence of the sources of fluid at that point.
- (ii) The divergence of current density J at a point is numerically equal to the quantity of electric charge flowing out per second per unit volume through an infinitesimal closed surface surrounding that point. Taking the law of conservation of charge into consideration, $\operatorname{div} J$ at any point is equal to the rate of decrease of charge density in the neighbourhood of that point, thus

$$\operatorname{div} J = -\frac{\partial \rho}{\partial t}$$

or

$$\operatorname{div} + \frac{\partial \rho}{\partial t} = 0 \quad \dots(10)$$

This equation is known as *equation of continuity* and represents the law of conservation of charge.

If the divergence of any vector function in a region is zero (i.e., $\operatorname{div} A = 0$), it means that the flux of the vector function entering any element of this region is equal to that leaving it. Any vector A satisfying the condition $\operatorname{div} A = 0$ is called a *solenoidal vector*.

The divergence of a vector field A in Cartesian coordinates, $\operatorname{div} A$ may also be written as

$$\operatorname{div} A = \left(i \frac{\partial}{\partial x} + j \frac{\partial}{\partial y} + k \frac{\partial}{\partial z} \right) \cdot (i \hat{A}_x + j \hat{A}_y + k \hat{A}_z)$$

or

$$\operatorname{div} A = \vec{\nabla} \cdot \vec{A} \quad \dots(11)$$

where $\vec{\nabla}$ is a vector operator defined as

$$\vec{\nabla} = i \frac{\partial}{\partial x} + j \frac{\partial}{\partial y} + k \frac{\partial}{\partial z}$$

CURL OF A VECTOR

The curl of a vector field at any point is defined as a vector quantity whose magnitude is equal to the maximum line integral per unit area along the boundary of an infinitesimal test area at that point and whose direction is perpendicular to the plane of the test area.

In order to obtain mathematical expression of curl of a vector $A = iA_x + jA_y + kA_z$ at a point P in a given infinitesimal rectangular area $EFGH$ (as shown in Fig. 2), we will use the above definition of curl of a

A.2 Oscillation

DAMPED HARMONIC OSCILLATION

A body executing simple harmonic oscillations has constant amplitude only under ideal condition. It is possible only when there is no dissipative or damping force acting on the body thrown to execute simple harmonic motion. In all practical cases, it is observed that the vibrating system experiences a resisting force which causes decay of its amplitude.

If a bob of a pendulum is displaced slightly from its mean position, it starts oscillating. The oscillation continues with its natural frequency but the amplitude gradually dies away and finally comes to rest. Decay in the amplitude of oscillation is due to the various resisting forces opposing the motion of the body, which diminishes the amplitude or damps the oscillation. Such vibrations with decreasing amplitude are called *damped oscillations*. The energy given to the body initially to begin the oscillations is used up in doing work against damping forces. Thus, the energy given to the system decrease with time, and finally the system reaches to rest.

Wave Equation of Damped Vibration: If y is the instantaneous displacement from the equilibrium position of a body of mass M , performing damped simple harmonic oscillation, then the forces acting on the body are as follows:

- (i) **Restoring force:** The tendency of this force is that it always tries to bring the body (executing oscillation) back to its mean position and is proportional to the displacement. Thus, it may be written as

$$\begin{aligned} F_R &\propto -y \\ \text{or} \quad F_R &= -ky \end{aligned} \quad \dots(1)$$

- (ii) **Damping force:** It is the retarding force proportional to the velocity acting in the opposite direction of motion.

Thus, it may be written as

$$\begin{aligned} F_D &\propto -\frac{dy}{dt} \\ \text{or} \quad F_D &= -b \frac{dy}{dt} \end{aligned} \quad \dots(2)$$

- (iii) **Force of inertia:** Each body ready to set itself for oscillatory motion has inertia, which can be given as

$$F_I \propto \frac{d^2y}{dt^2} \quad \text{or} \quad F_I = m \frac{d^2y}{dt^2} \quad \dots(3)$$

At equilibrium, the force of inertia is balanced by restoring force and damping force. Hence, the equation of motion of a body executing damped harmonic vibration is given as

$$\begin{aligned} m \frac{d^2y}{dt^2} &= -ky - b \frac{dy}{dt} \\ \text{or} \quad \frac{d^2y}{dt^2} + \frac{b}{m} \frac{dy}{dt} + \frac{k}{m} y &= 0 \\ \text{or} \quad \frac{d^2y}{dt^2} + 2\beta \frac{dy}{dt} + w_o^2 y &= 0 \end{aligned} \quad \dots(4)$$

where $\beta = \frac{b}{2m}$ is the damping coefficient and $w_o^2 = \frac{k}{m}$, and equation (4) is called the differential equation for damped oscillator. This is a linear, homogeneous, second order differential equation with constant coefficients, 2β and w_o^2 .

The general solution of this equation for $\beta \neq w_o$ is

$$y(t) = e^{-\beta t} [A_1 e^{\sqrt{\beta^2 - w_o^2} t} + A_2 e^{-\sqrt{\beta^2 - w_o^2} t}] \quad \dots(5)$$

where the constants A_1 and A_2 depend on the initial (at $t = 0$) position and velocity of the oscillator.

The behaviour of a damped oscillator depends on the relative values of the restoring force and the damping force which regulate its motion. These two forces, in turn, determine the constants w_o and β , respectively.

The quantity $(\beta^2 - w_o^2)$ occurring in equation (5) may be positive, negative, or zero. Accordingly, the solution given in equation (5) will lead to three different situations:

1. Underdamping (oscillatory): $w_o^2 > \beta^2$
2. Overdamping (non-oscillatory): $w_o^2 < \beta^2$
3. Critical damping (non-oscillatory): $w_o^2 = \beta^2$

Case-I: Underdamping ($w_o^2 > \beta^2$)

Under the weak damping, $\beta^2 < w_o^2$; thus, the quantity $(\beta^2 - w_o^2)$ is negative.

We can write

$$\beta^2 - w_o^2 = -w_1^2 \text{ so that } \sqrt{\beta^2 - w_o^2} = iw_1 \quad \dots(6)$$

where w_1 is a real quantity.

Now, the general solution (5) can be written as

$$y(t) = e^{-\beta t} [A_1 e^{iw_1 t} + A_2 e^{-iw_1 t}] \quad \dots(7)$$

By setting $A_1 = \frac{A}{2} e^{i\theta}$ and $A_2 = \frac{A}{2} e^{-i\theta}$, we get

$$y(t) = e^{-\beta t} \frac{A}{2} (e^{i(w_1 t + \theta)} + e^{-j(w_1 t + \theta)}) \quad \dots(8)$$

or $y(t) = Ae^{-\beta t} \cos(w_1 t + \theta)$

The constants A and θ are determined from initial position and velocity of the oscillator.

Cosine term in equation (8) represents oscillatory motion with time period.

$T_1 = \frac{2\pi}{w_1} = \frac{2\pi}{\sqrt{w_o^2 - \beta^2}}$ which is greater than the time period $T = \frac{2\pi}{-w_o}$ of undamped oscillator. Thus,

the damped oscillator has low frequency than that of undamped oscillator.

The amplitude of $Ae^{-\beta t}$ of the damped oscillator is no more constant and decreases exponentially with time, till the oscillation dies out as shown in Fig. 1. The total mechanical energy of the oscillator also decreases with time. Since energy is proportional to square of the amplitude ($e^{-2\beta t}$), thus energy also decreases exponentially with time, i.e., as $e^{-2\beta t}$.

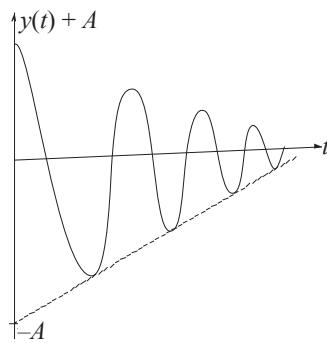


Fig. 1 Variation of amplitude during damped oscillation

Decrement: The ratio between amplitudes of two successive maxima i.e.

$$\frac{Ae^{-\beta t}}{Ae^{-\beta(t+T_1)}} = e^{\beta T_1}$$

is known as the decrement of the oscillator. The natural logarithm of the decrement

$$\delta = \log e^{\beta T_1} = \beta T_1 = \frac{2\pi\beta}{\sqrt{w_o^2 - \beta^2}} \quad \dots(9)$$

is known as the logarithmic decrement of the oscillator.

Case-II: Overdamping ($w_o^2 < \beta^2$)

For large damping, $\beta^2 > w_o^2$, so that $(\beta^2 - w_o^2)$ is a positive. If we write $\sqrt{\beta^2 - w_o^2} = \alpha$, then equation (5) takes the form

$$\begin{aligned} y(t) &= e^{-\beta t}(A_1 e^{\alpha t} + A_2 e^{-\alpha t}) \\ y(t) &= A_1 e^{-(\beta-\alpha)t} + A_2 e^{-(\beta+\alpha)t} \end{aligned} \quad \dots(10)$$

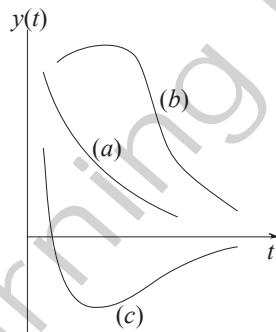


Fig. 2 Oscillation under overdamping with different initial conditions.

under the relative values of α , β , A_1 and A_2 with initial position and velocity, the oscillator comes back to the equilibrium position in one of the following ways:

- (a) The displacement decreases to zero exponentially in a continuous manner.
- (b) The displacement first increases, attains a maximum, and then decreases to zero.
- (c) The displacement changes sign, i.e., the oscillator overshoots to the other side of the equilibrium position, attains a maximum, and then continuously comes back to equilibrium position.

Motion of simple pendulum in highly viscous liquid medium is an example of overdamped motion.

Case-III: Critical damping ($w_o^2 = \beta^2$)

The general solution of differential equation of damped harmonic motion given in equation (4) is

$$y(t) = (At + B)e^{-\beta t} \quad \dots(11)$$

where A and B are constants of integration. In this case, motion is not oscillatory; it approaches to zero asymptotically. If the damping is weaker, the motion is oscillatory; if damping is stronger, the motion is overdamped. For given values of initial position and velocity, a critically damped oscillator approaches the equilibrium position faster than either the underdamped oscillator or overdamped oscillator, see Fig. 3.

A.3 Coupled Oscillations

COUPLED VIBRATIONS

A system whose motion and position are characterized by two or more independent quantities is known as system having two or more degrees of freedom. The vibrations of such system which involve more than one vibration system are known as *coupled vibrations*. The nature has many fascinating examples of such systems, such as molecules and elementary particles. In order to derive the differential equation for coupled oscillations for its analytical treatment, let us consider the example of coupled pendulum. The detail analysis of this system is discussed in the following section.

COUPLED PENDULUM

This system consists of two identical simple pendulums whose bobs are connected by an elastic spring of stiffness constant K . Both pendulums having equal mass are suspended on a light rod of length l as shown in Fig. 1.

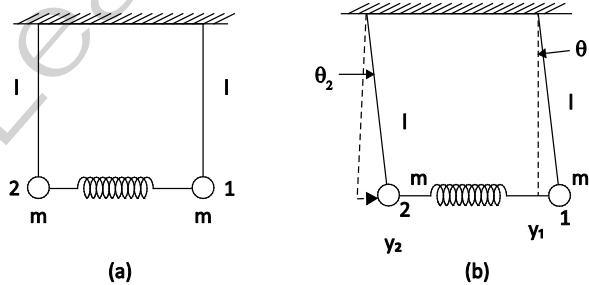


Fig. 1 Coupled pendulum

Suppose we draw 1 aside, holding 2 fixed and then release both from equilibrium position Fig 1(b). Pendulum 1 oscillates but its amplitude continuously decreases. Pendulum 2, initially undisplaced, also begins to oscillate with increasing amplitude very soon the amplitude of oscillations of pendulum 1 and 2 becomes equal. The amplitude of 1 continues to decrease becomes zero, while that of 2 continues to increase and becomes equal to the amplitude initially given to 1. Thus, the starting condition is reversed. Again the motion of 2 is transferred back to 1 and so on. The energy initially given to 1 thus shuttles back and forth between 1 and 2 through the coupling spring.

Let the displacement of the bobs 1 and 2 at any instant be y_1 and y_2 , respectively, such that $y_1 > y_2$. The spring is stretched by an amount $y_1 - y_2$ and pulls 1 and 2 with an elastic force. Thus, the net restoring force (gravity + elasticity) on 1 is

$$\begin{aligned} & -mg \sin \theta_1 - k(y_1 - y_2) \\ & = -mg \frac{y_1}{l_1} - k(y_1 - y_2) \end{aligned}$$

Similarly, the net restoring force on 2 is

$$\begin{aligned} & -mg \sin \theta_2 + k(y_1 - y_2) \\ & = -mg \frac{y_2}{l} + k(y_1 - y_2) \end{aligned}$$

The elastic forces are negative and positive for pendulum 1 and 2 as it opposes the motion of 1 and helps that of 2.

Therefore, the equations of motion of the two bobs are

$$m \frac{d^2 y_1}{dt^2} = -mg \frac{y_1}{l} - k(y_1 - y_2)$$

and

$$m \frac{d^2 y_2}{dt^2} = -\frac{mg}{l} y_2 + k(y_1 - y_2)$$

We can write the above equations as

$$\frac{d^2 y_1}{dt^2} = -\omega_0^2 y_1 - \omega_c^2 (y_1 - y_2) \quad \dots(1)$$

$$\frac{d^2 y_2}{dt^2} = -\omega_0^2 y_2 + \omega_c^2 (y_1 - y_2) \quad \dots(2)$$

where $\omega_0 = g/l$ is the natural frequency of each pendulum and $\omega_c^2 = k/m$ is coupling frequency of the spring.

NORMAL FREQUENCY

In order to calculate the normal frequencies, let us suppose that the system is allowed to vibrate in one of its normal modes with frequency w . The solution of equations (1) and (2) may be assumed to be

$$y_1 = A_1 e^{iwt} \quad \text{and} \quad y_2 = A_2 e^{iwt}, \text{ where } A_1 \text{ and } A_2 \text{ being the amplitudes of oscillations.}$$

Thus,

$$\frac{d^2 y_1}{dt^2} = -\omega^2 A_1 e^{i\omega t} = -\omega^2 y_1$$

and

$$\frac{d^2 y_2}{dt^2} = -\omega^2 A_2 e^{i\omega t} = -\omega^2 y_2$$

Substituting these values in equations (1) and (2), we get

$$-\omega^2 y_1 = -\omega_0^2 y_1 - \omega_c^2 (y_1 - y_2)$$

or $(-\omega^2 + \omega_0^2 + \omega_c^2) y_1 = \omega_c^2 y_2 \quad \dots(3)$

and $-\omega^2 y_2 = -\omega_0^2 y_2 + \omega_c^2 (y_1 - y_2)$

or $(-\omega^2 + \omega_0^2 + \omega_c^2) y_2 = \omega_c^2 y_1 \quad \dots(4)$

From equation (3), we can write

$$\frac{y_1}{y_2} = \frac{\omega_c^2}{(-\omega^2 + \omega_0^2 + \omega_c^2)} \quad \dots(5)$$

and from equation (4), we get

$$\frac{y_1}{y_2} = \frac{(-\omega^2 + \omega_0^2 + \omega_c^2)}{\omega_c^2} \quad \dots(6)$$

From equations (5) and (6), we get

$$\frac{\omega_c^2}{-\omega^2 + \omega_0^2 + \omega_c^2} = -\frac{(-\omega^2 + \omega_0^2 + \omega_c^2)}{\omega_c^2}$$

or

$$(-\omega^2 + \omega_0^2 + \omega_c^2)^2 = (\omega_c^2)^2$$

or

$$\omega^2 + \omega_0^2 + \omega_c^2 = \pm \omega_c^2 \quad \dots(7)$$

from equation (7), we can write

$$\omega^2 = \omega_0^2 \quad \dots(8)$$

or

$$\omega^2 = \omega_0^2 + 2\omega_c^2 \quad \dots(9)$$

or

$$\omega^2 = \omega_0^2 + \frac{2k}{m} \quad \dots(10)$$

Thus, the normal frequencies would be either

$$\omega_0 \text{ or } \left(\omega_0^2 + \frac{2k}{m}\right)^{\frac{1}{2}}$$

ω_0 is called lower mode frequency and $\left(\omega_0^2 + \frac{2k}{m}\right)^{\frac{1}{2}}$ is higher mode frequency.

The ratio of the displacement of two bobs in different modes determine the configuration of the system.

From equation (5), we have $\frac{y_1}{y_2} = \frac{\omega_c^2}{(-\omega^2 + \omega_0^2 + \omega_c^2)}$

In the lower mode $\frac{y_1}{y_2} = 1$, i.e., $y_1 = y_2$

In this mode, both the bobs have equal displacements in the same direction. It means the spring remains unstretched, i.e., stiffness of the coupling has no effect.

In the higher mode, when $\omega^2 = \omega_0^2 + 2 k/m$

$$\frac{y_1}{y_2} = -1, \text{ i.e., } y_1 = -y_2$$

In this mode, the bobs have equal but opposite displacements. The spring is stretched and compressed alternately, which causes the greater restoring force.

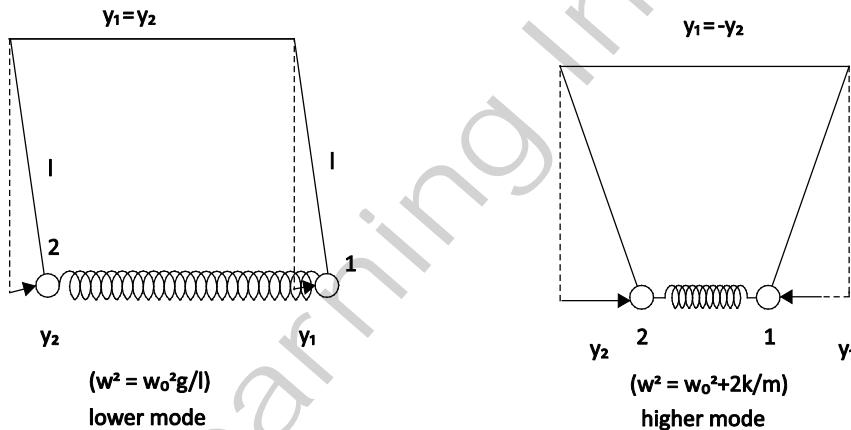


Fig. 2 Lower and higher modes of vibration of coupled pendulum

NORMAL COORDINATES

The oscillation of coupled system can be described in terms of normal coordinates which have been discussed below.

The equation of motion of two identical pendulum equation (1) and (2) are coupled equations so they cannot be solved independently. They are solved simultaneously.

Adding equations (1) and (2), we get

$$\frac{d^2}{dt^2}(y_1 + y_2) + \omega_0^2(y_1 + y_2) = 0 \quad \dots(11)$$

On subtracting (1) and (2), we get

$$\frac{d^2}{dt^2}(y_1 - y_2) + (\omega_0^2 + 2\omega_c^2)(y_1 - y_2) = 0 \quad \dots(12)$$

Let us now choose two new coordinates, Y_1 and Y_2 , such that

A.4 Ferromagnetism

Ferromagnetism is characterized by the presence of parallel alignment of permanent magnetic dipole moments in a single direction. Fe, Co, Ni, Gd, and Dy are some examples of ferromagnetic materials. Usually, the materials showing ferromagnetic behaviour have partly-filled inner electron shells. In such materials, magnetization arises spontaneously, over large region of material. It is important to mention that all this happens in the absence of an external magnetic field. In the successive sections, we will discuss the temperature dependence of ferromagnetism. It is observed in case of ferromagnetic materials that the spontaneous magnetization can occur only below a certain critical temperature T_c , called *Ferromagnetic Curie temperature*. Well above the Curie temperature, such materials behave like a paramagnetic material and have a well-defined susceptibility given by Curie–Weiss law as follows:

$$\chi_m = \frac{c}{T - \theta} \quad \dots(1)$$

where c is the Curie constant and θ is called the paramagnetic Curie temperature. θ is slightly higher than T_c . A typical curve relative χ_m to temperature above the Curie temperature for nickel is shown in the following Fig. 1.

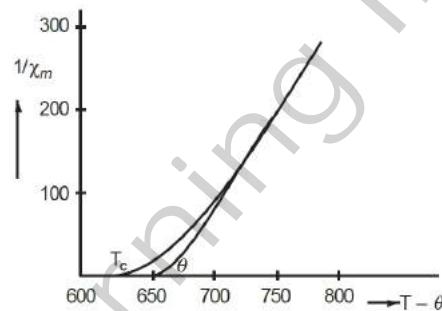


Fig. 1 $1/\chi_m$ vs temperature of nickel.

The above curve shows that ferromagnetic behaviour ($T > \theta$) and paramagnetic behaviour are similar; the only difference is that for most of the paramagnetic materials, $\theta = 0$.

FERROMAGNETIC BEHAVIOUR BELOW CURIE TEMPERATURE T_c

Below the ferromagnetic Curie temperature, ferromagnetic materials show a marked increase in susceptibility and exhibit the well-known B–H characteristics, which displays the similar hysteresis effect. A schematic representation of this behaviour is shown in Fig. 2. We can explain the behaviour in the following way: starting with a virgin specimen ($B = H = 0$), if H is increased, B at first increases (oa).

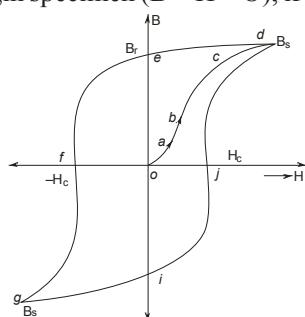


Fig. 2 Hysteresis loop for a ferromagnetic specimen

As the field H is increased further, B starts increasing rapidly until point c on the curve, where saturation sets in when position d is reached; further increase in H causes negligible change in B , and at that point $B = B_{\text{sat}}$, called the saturation field. At saturation, if now, H is decreased the curve de results and a remnant flux density B_r is observed with $H = 0$. Since there is no longer an external excitation, the specimen has become spontaneously magnetized. The magnetization corresponding to B_r is equal

$$\text{to } M = \frac{B_r}{\mu_0} \text{ as } H = 0.$$

A field $-H_c$ required to reduce the flux density to zero is called the *coercive force*. Further, decrease of H creates curve fg , with reverse saturation of B field occurring at g . Another reversal of H traces out $gijd$. The portion of the curve ef is called the *demagnetizing curve*.

From the above observations, regarding magnetizing and demagnetizing of ferromagnetic materials, it is clear that the magnetization of ferromagnetic materials does not get traced back when magnetizing field is reduced. Rather, hysteresis is observed. Depending on the amount of coercive and remnant forces, materials are characterized as soft or hard magnetic materials.

SPONTANEOUS MAGNETIZATION AND WEISS THEORY OF FERROMAGNETISM

The properties of ferromagnetic materials are quite different above and below the Curie temperature. Below the Curie temperature, the ferromagnetic materials show spontaneous magnetization, and above this temperature they obey the well-known Curie–Weiss law. Here, we shall discuss the atomic interpretation of spontaneous magnetization and of the Curie–Weiss law.

As discussed in earlier text, the remnant flux density B_r of a typical ferromagnetic materials used for permanent magnet is about 1 weber/m². Since $H = 0$ at this point on the hysteresis loop, as shown in

$$\text{Fig. 2, we have spontaneous magnetization } M_r = \frac{B_r}{\mu_0} \approx 10^6 \text{ A/m.}$$

If each atom in the material has a magnetic moment of the order of one Bohr magneton, then the atomic density is $N = \frac{M_r}{p_B} \approx 10^{29} \text{ atoms/m}^3$, where p_B is the magnetic moment per spin. Since the number

of atoms in a solid is approximately 10^{29} per m³, it is clear that such a high remnant flux density can be explained only by assuming a spontaneous magnetization in which all the atoms of the material are so oriented that their magnetic moments are parallel. From this example, we can conclude that even when $H = 0$, there exists some force of field in a ferromagnetic material which tends to orient the magnetic dipole moments in parallel.

On the basis of the interaction of elementary moments with one another, a French scientist, Weiss in 1907 suggested the idea to understand the behaviour of ferromagnetic materials. He suggested that interaction could be expressed in terms of fictitious internal field, which he called the molecular field H_m , and which acted in addition to the applied field H . The molecular field was thought to be in some way caused by the magnetization of the surrounding material.

Weiss assumed that the intensity of the molecular field was directly proportional to the magnetization such that

$$M_H = \gamma M \quad \dots(2)$$

where γ is called the molecular field constant. Therefore, the total internal field which acts in the material is

$$\begin{aligned} H_i &= H + H_M \\ H_i &= H + \gamma M \end{aligned} \quad \dots(3)$$

γ is also referred as internal field constant and it determines the strength of the interaction between the magnetic dipole moments in a material.

We shall now do the analysis of ferromagnetic materials, assuming the internal field as given by equation (3). It is desirable to consider separately two temperature regions.

FERROMAGNETIC BEHAVIOUR AT HIGH TEMPERATURE ($T > T_c$)

At very high temperature ($T > T_c$), the thermal agitation in case of ferromagnetic materials is so great that internal field is not sufficient to maintain alignment of magnetic dipole moments. The behaviour becomes analogous to a paramagnetic material except that the field acting on the dipole is the internal field given by equation (3).

As a model, let us consider a system of N spins per m^3 , each giving rise to a magnetic moment of 1 Bohr magneton, p_B , either parallel or antiparallel to the external field. The magnetization of such a system may be obtained from expression of magnetization for paramagnetic materials given by

$$M = Np_B \tanh(\mu_0 p_B H / kT) \quad \dots(4)$$

Replacing H by H_i in equation (4), we get

$$M = Np_B \tanh\left[\frac{\mu_0 p_B}{kT}(H + \gamma M)\right] \quad \dots(5)$$

at sufficiently high temperatures, $\mu_0 p_B (H + \gamma M) \ll kT$. Then, since $\tanh \alpha \approx \alpha$ for $\alpha \ll 1$, we may write equation (5) as

$$\begin{aligned} M &= \left(N\mu_0 p_B^2 / kT\right)(H + \gamma M) \\ \text{or} \quad M &\left(1 - \frac{\gamma N\mu_0 p_B^2}{kT}\right) = \frac{N\mu_0 p_B^2 H}{kT} \end{aligned} \quad \dots(6)$$

From this expression, we have magnetic susceptibility as

$$\chi_m = \frac{M}{H} = \frac{N\mu_0 p_B^2 / k}{T - (\gamma N\mu_0 p_B^2 / k)} = \frac{c}{T - \theta} \quad \dots(7)$$

where $c = N\mu_0 p_B^2 / k$ and $\theta = \gamma c$

For the model, we have a relation which is identical in form with the Curie–Weiss law. For a ferromagnetic material, c and θ can be determined from measurements of susceptibility as a function of temperature. The value of γ for a ferromagnetic material is of the order of 10^3 .

It is important to note that the molecular field is in no sense a real field; it is rather a force which tends to align or dis-align the atomic or magnetic moments, and the strength of the force depends on the amount of alignment already attained because the molecular field is proportional to magnetization. This force is also called *exchange force*.

FERROMAGNETIC MATERIALS BELOW CURIE POINT ($T < T_c$)—SPONTANEOUS MAGNETIZATION

Curie: Weiss law, given in equation (7), cannot field $T \leq \theta$ in this case as magnetic susceptibility becomes negative and it would pass through pole; at $T = \theta$, the susceptibility would become infinite. This fact suggests that at $T = \theta$, spontaneous magnetization may occur (i.e., when $H = 0$, M is non-vanishing). This will be clear from the following analysis.

From the model discussed earlier, magnetization field can be given as

$$M = Np_B \tanh \left[\frac{\mu_0 p_B}{kT} (H + \gamma M) \right] \quad \dots(8)$$

When $H = 0$, let us see whether we get non-vanishing value of M or not. From equation (8), when $H = 0$

$$M = Np_B \tanh \frac{\mu_0 p_B \gamma M}{kT}$$

and $\frac{M}{Np_B} = \frac{M}{M_{sat}} = \tanh \alpha$... (9)

where $\alpha = \frac{\mu_0 p_B \gamma M}{kT}$... (10)

where $M_{sat} = Np_B$ represents the saturation value of magnetization, i.e., maximum magnetization when all the dipoles are parallel. M/M_{sat} as a function of α is plotted in the following Fig. 3.

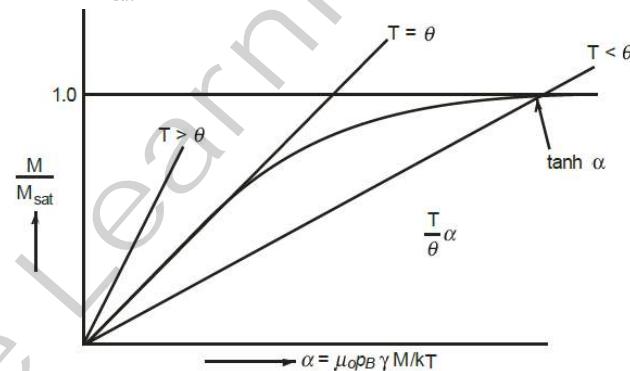


Fig. 3 Intersection of two curves $\frac{M}{M_{sat}}$ ($= \tan \alpha$) vs α and $\frac{M}{M_{sat}}$ ($= \frac{T}{\theta} \cdot \alpha$) vs α

Using equation (10), putting the value of M in equation (9), we get

$$\frac{M}{M_{sat}} = \frac{M}{Np_B} = \frac{kT}{N\gamma\mu_0 p_B^2} \cdot \alpha = \frac{T}{\theta} \cdot \alpha \quad \dots(11)$$

where $\theta = \frac{N\gamma\mu_0 p_B^2}{k}$.

A.5 Weiss Law

WEISS LAW (MOLECULAR FIELD THEORY OF PARAMAGNETISM)

Actually, Langevin's theory could not explain a more complicated dependence of susceptibility upon temperature exhibited by several paramagnetics such as highly compressed and cooled gases, very concentrated solutions of salts, solid salts, and crystals. This theory could not explain the intimate relation observed between para and ferromagnetism.

In order to explain those points, Weiss, in 1907, improved the Langevin's treatment, introducing a new concept of internal molecular field. Weiss observed that in a real gas, molecules are mutually influenced by their magnetic moment. He proposed similarly within the gas at any point, a molecular field is produced by all the molecules in the neighbourhood, proportional to and acting in the same sense as the magnetization vector \vec{M} . Thus, we have

$$\begin{aligned}\text{Molecular field} &\propto \text{Magnetization } \vec{M} \\ &= \gamma \vec{M}\end{aligned}$$

where γ is the proportionality constant known as molecular field coefficient. Then the effective field may be regarded as the vector sum of the externally applied field \vec{B} and the internal molecular field, i.e.,

$$\vec{B}_{\text{eff}} = \vec{B} + \gamma \vec{M} \quad \dots(1)$$

According to the Langevin's theory for paramagnetism, the mean effective magnetic moment can be given as

$$M = \frac{m^2 B_{\text{eff}}}{3KT} \quad \dots(2)$$

where m is the magnetic moment of an electron orbitting in a circle, K is Boltzmann constant, and T is the absolute temperature. From equations (1) and (2), we get

$$M = \frac{m^2 (B + \gamma M)}{3KT}$$

Now, the magnetic moment in a gram molecule,

$$M = \frac{N_m m^2 (B + \gamma M)}{3KT}$$

where N_m is the number of molecules in a gram molecule. Rearranging the terms

$$M \left(1 - \frac{N_m m^2 \gamma}{3KT} \right) = \left(\frac{N_m m^2 B}{3KT} \right)$$

$$M = \frac{N_m m^2 B}{3KT \left(1 - \frac{N_m m^2 \gamma}{3KT} \right)}$$

$$M = \frac{N_m m^2 B}{3K \left(T - \frac{N_m m^2 \gamma}{3K} \right)}$$

But $B = \mu H$

$$\text{Therefore, } M = \frac{N_m m^2 \mu H}{3K \left(T - \frac{N_m m^2 \gamma}{3K} \right)}$$

Now, the molecular susceptibility

$$\chi_m = \frac{M}{H} = \frac{N_m m^2 \mu}{3K \left(T - \frac{N_m m^2 \gamma}{3K} \right)}$$

$$\text{We know that } K = \frac{R}{N_m}$$

Thus,

$$\chi_m = \frac{M}{H} = \frac{N_m^2 m^2 \mu}{3R \left(T - \frac{N_m^2 m^2 \gamma}{3R} \right)} \quad \dots(3)$$

Let $C_m = \frac{N_m^2 m^2 \mu}{3R}$ is Curie constant and $\theta = \frac{N_m^2 m^2 \gamma}{3R}$ is Curie point.

Now, using C_m and θ in equation (3), we get

$$\chi_m = \frac{C_m}{T - \theta} \quad \dots(4)$$

Equation (4) is known as Curie–Weiss law derived on the basis of Weiss molecular field theory of paramagnetism. According to this law, the susceptibility of a paramagnetic substance having molecular field varies inversely as the excess of temperature not from a certain critical temperature, the Curie point. The Curie point is able to explain the intimate relation observed between paramagnetic and ferromagnetic substances. Below the Curie point θ , the ferromagnetism exists; but above the Curie point, the ferromagnetic properties disappear, and the substance becomes paramagnetic.

Solved Examples

Example 1: The magnetic susceptibility of medium is 948. Calculate the permeability (or absolute permeability) and relative permeability.

Solution: Given that

$$\chi_m = 948$$

Absolute permeability:

$$\begin{aligned} \mu &= \mu_0 (1 + \chi_m) \\ &= 4\pi \times 10^{-7} (1 + 948) \text{ henry/m} \\ &= 4\pi \times 10^{-7} \times (949) \text{ henry/m} \\ &= 2 \times 5.96 \times 10^{-4} \text{ henry/m} \\ &= 1.192 \times 10^{-3} \text{ henry/m} \end{aligned}$$

$$\text{Relative permeability: } \mu_r = \frac{\mu}{\mu_0} = \frac{1.192 \times 10^{-3}}{4\pi \times 10^{-7}} = 949$$

Example 2: An iron rod of 0.2 cm^2 cross-sectional area is subjected to a magnetizing field of 1200 Am^{-1} . The susceptibility of iron is 599. Find the permeability and the magnetic flux produced.

Solution: The permeability μ is related to the susceptibility as follows

$$\mu = \mu_0 (1 + \chi_m)$$

$$\text{Now, } \mu_0 = 4\pi \times 10^{-7} \text{ TmA}^{-1}$$

Therefore, $\mu = 4\pi \times 10^{-7} (1+599) = 7.536 \times 10^{-4} \text{ TmA}^{-1}$

The magnetizing field H is 1200 AM^{-1} . Therefore, the magnetic induction B is given by

$$\begin{aligned} B &= \mu H = (7.536 \times 10^{-4} \text{ TmA}^{-1}) \times (1200 \text{ Am}^{-1}) \\ &= 0.904 \text{ T} \end{aligned}$$

The magnetic flux produced is given by

$$\begin{aligned} \phi &= BA = 0.094T \times (0.2 \times 10^{-4}) \text{ m}^2 \\ &= 1.81 \times 10^{-5} \text{ Tm}^2 \end{aligned}$$

Since $1 \text{ T} = 1 \text{ wb/m}^2$, so $\phi = 1.81 \times 10^{-5} \text{ wb}$

Example 3: A paramagnetic gas has atoms each of dipole moment $1.5 \times 10^{-23} \text{ Am}^2$. The temperature of the gas is 27 and it has $2.0 \times 10^{26} \text{ atoms/m}^3$. Find the maximum possible magnetization of gas under an external magnetizing field. Can the gas reach close to maximum magnetization in a magnetizing field of 3.0 T? The Boltzmann's constant is $1.38 \times 10^{-23} \text{ J/K}$.

Solution: The maximum possible magnetization (magnetic moment per unit volume) corresponding to all the atoms aligned along the magnetizing field is given by

$$\begin{aligned} I &= (1.5 \times 10^{-23} \text{ Am}^2) \times (2.0 \times 10^{26} \text{ m}^{-3}) \\ &= 3 \times 10^3 \text{ Am}^{-1} \end{aligned}$$

The kinetic energy of translation (thermal energy) of an atomic dipole is

$$\begin{aligned} \frac{3}{2}KT &= \frac{3}{2} \times (1.38 \times 10^{-23} \text{ JK}^{-1}) \times (300 \text{ K}) \\ &= 6.2 \times 10^{-21} \text{ J} \end{aligned}$$

The minimum potential energy of the dipole in an external magnetic field B ($= 3.0 \text{ T}$) is

$$\begin{aligned} MB &= (1.5 \times 10^{-23} \text{ Am}^2) \times (3.0 \text{ NA}^{-1} \text{ m}^{-1}) \\ &= 4.5 \times 10^{-23} \text{ Nm} = 4.5 \times 10^{-23} \text{ J} \end{aligned}$$

$$\text{Thus, } \frac{\frac{3}{2}KT}{MB} = \frac{6.2 \times 10^{-21} \text{ J}}{4.5 \times 10^{-23} \text{ J}} = 138$$