CONTENTS

Foreword *iii*
Preface *v*
A Note for the Teacher *x*

CHAPTER 1

Physical World

1.1 What is physics ? 1
1.2 Scope and excitement of physics 2
1.3 Physics, technology and society 5
1.4 Fundamental forces in nature 6
1.5 Nature of physical laws 10

CHAPTER 2

Units and Measurements

2.1 Introduction 16
2.2 The international system of units 16
2.3 Measurement of length 18
2.4 Measurement of mass 21
2.5 Measurement of time 22
2.6 Accuracy, precision of instruments and errors in measurement 22
2.7 Significant figures 27
2.8 Dimensions of physical quantities 31
2.9 Dimensional formulae and dimensional equations 31
2.10 Dimensional analysis and its applications 32

CHAPTER 3

Motion in a Straight Line

3.1 Introduction 39
3.2 Position, path length and displacement 39
3.3 Average velocity and average speed 42
3.4 Instantaneous velocity and speed 43
3.5 Acceleration 45
3.6 Kinematic equations for uniformly accelerated motion 47
3.7 Relative velocity 51

CHAPTER 4

Motion in a Plane

4.1 Introduction 65
4.2 Scalars and vectors 65
4.3 Multiplication of vectors by real numbers 67
4.4 Addition and subtraction of vectors – graphical method 67
4.5 Resolution of vectors 69

4.6 Vector addition – analytical method 71
4.7 Motion in a plane 72
4.8 Motion in a plane with constant acceleration 75
4.9 Relative velocity in two dimensions 76
4.10 Projectile motion 77
4.11 Uniform circular motion 79

CHAPTER 5

LAWS OF MOTION

5.1 Introduction 89
5.2 Aristotle's fallacy 90
5.3 The law of inertia 90
5.4 Newton's first law of motion 91
5.5 Newton's second law of motion 93
5.6 Newton's third law of motion 96
5.7 Conservation of momentum 98
5.8 Equilibrium of a particle 99
5.9 Common forces in mechanics 100
5.10 Circular motion 104
5.11 Solving problems in mechanics 105

CHAPTER 6

WORK, ENERGY AND POWER

6.1 Introduction 114
6.2 Notions of work and kinetic energy : The work-energy theorem 116
6.3 Work 116
6.4 Kinetic energy 117
6.5 Work done by a variable force 118
6.6 The work-energy theorem for a variable force 119
6.7 The concept of potential energy 120
6.8 The conservation of mechanical energy 121
6.9 The potential energy of a spring 123
6.10 Various forms of energy : the law of conservation of energy 126
6.11 Power 128
6.12 Collisions 129

CHAPTER 7

SYSTEM OF PARTICLES AND ROTATIONAL MOTION

7.1 Introduction 141
7.2 Centre of mass 144
7.3 Motion of centre of mass 148
7.4 Linear momentum of a system of particles 149
7.5 Vector product of two vectors 150
7.6 Angular velocity and its relation with linear velocity 152
7.7 Torque and angular momentum 154
7.8 Equilibrium of a rigid body 158
7.9 Moment of inertia 163
7.10 Theorems of perpendicular and parallel axes 164

7.11 Kinematics of rotational motion about a fixed axis 167
7.12 Dynamics of rotational motion about a fixed axis 169
7.13 Angular momentum in case of rotations about a fixed axis 171
7.14 Rolling motion 173

CHAPTER 8

GRAVITATION

8.1 Introduction 183
8.2 Kepler's laws 184
8.3 Universal law of gravitation 185
8.4 The gravitational constant 189
8.5 Acceleration due to gravity of the earth 189
8.6 Acceleration due to gravity below and above the surface of earth 190
8.7 Gravitational potential energy 191
8.8 Escape speed 193
8.9 Earth satellite 194
8.10 Energy of an orbiting satellite 195
8.11 Geostationary and polar satellites 196
8.12 Weightlessness 197

APPENDICES 203

ANSWERS 219

CHAPTER ONE

PHYSICAL WORLD

1.1 What is physics ?
1.2 Scope and excitement of physics
1.3 Physics, technology and society
1.4 Fundamental forces in nature
1.5 Nature of physical laws
Summary
Exercises

1.1 WHAT IS PHYSICS ?

Humans have always been curious about the world around them. The night sky with its bright celestial objects has fascinated humans since time immemorial. The regular repetitions of the day and night, the annual cycle of seasons, the eclipses, the tides, the volcanoes, the rainbow have always been a source of wonder. The world has an astonishing variety of materials and a bewildering diversity of life and behaviour. The inquiring and imaginative human mind has responded to the wonder and awe of nature in different ways. One kind of response from the earliest times has been to observe the physical environment carefully, look for any meaningful patterns and relations in natural phenomena, and build and use new tools to interact with nature. This human endeavour led, in course of time, to modern science and technology.

The word **Science** originates from the Latin verb *Scientia* meaning 'to know'. The Sanskrit word *Vijnan* and the Arabic word *Ilm* convey similar meaning, namely 'knowledge'. Science, in a broad sense, is as old as human species. The early civilisations of Egypt, India, China, Greece, Mesopotamia and many others made vital contributions to its progress. From the sixteenth century onwards, great strides were made in science in Europe. By the middle of the twentieth century, science had become a truly international enterprise, with many cultures and countries contributing to its rapid growth.

What is Science and what is the so-called **Scientific Method**? Science is a systematic attempt to understand natural phenomena in as much detail and depth as possible, and use the knowledge so gained to predict, modify and control phenomena. Science is exploring, experimenting and predicting from what we see around us. The curiosity to learn about the world, unravelling the secrets of nature is the first step towards the discovery of science. The scientific method involves several interconnected steps : Systematic observations, controlled experiments, qualitative and

quantitative reasoning, mathematical modelling, prediction and verification or falsification of theories. Speculation and conjecture also have a place in science; but ultimately, a scientific theory, to be acceptable, must be verified by relevant observations or experiments. There is much philosophical debate about the nature and method of science that we need not discuss here.

The interplay of theory and observation (or experiment) is basic to the progress of science. Science is ever dynamic. There is no 'final' theory in science and no unquestioned authority among scientists. As observations improve in detail and precision or experiments yield new results, theories must account for them, if necessary, by introducing modifications. Sometimes the modifications may not be drastic and may lie within the framework of existing theory. For example, when Johannes Kepler (1571-1630) examined the extensive data on planetary motion collected by Tycho Brahe (1546-1601), the planetary circular orbits in heliocentric theory (sun at the centre of the solar system) imagined by Nicolas Copernicus (1473–1543) had to be replaced by elliptical orbits to fit the data better. Occasionally, however, the existing theory is simply unable to explain new observations. This causes a major upheaval in science. In the beginning of the twentieth century, it was realised that Newtonian mechanics, till then a very successful theory, could not explain some of the most basic features of atomic phenomena. Similarly, the then accepted wave picture of light failed to explain the photoelectric effect properly. This led to the development of a radically new theory (*Quantum Mechanics*) to deal with atomic and molecular phenomena.

Just as a new experiment may suggest an alternative theoretical model, a theoretical advance may suggest what to look for in some experiments. The result of experiment of scattering of alpha particles by gold foil, in 1911 by Ernest Rutherford (1871–1937) established the nuclear model of the atom, which then became the basis of the quantum theory of hydrogen atom given in 1913 by Niels Bohr (1885–1962). On the other hand, the concept of antiparticle was first introduced theoretically by Paul Dirac (1902–1984) in 1930 and confirmed two years later by the experimental discovery of positron (antielectron) by Carl Anderson.

Physics is a basic discipline in the category of **Natural Sciences**, which also includes other disciplines like Chemistry and Biology. The word **Physics** comes from a Greek word meaning nature. Its Sanskrit equivalent is *Bhautiki* that is used to refer to the study of the physical world. A precise definition of this discipline is neither possible nor necessary. We can broadly describe physics as a study of the basic laws of nature and their manifestation in different natural phenomena. The scope of physics is described briefly in the next section. Here we remark on two principal thrusts in physics : **unification** and **reduction**.

In Physics, we attempt to explain diverse physical phenomena ***in terms of a few concepts and laws***. The effort is to see the physical world as manifestation of some universal laws in different domains and conditions. For example, the same law of gravitation (given by Newton) describes the fall of an apple to the ground, the motion of the moon around the earth and the motion of planets around the sun. Similarly, the basic laws of electromagnetism (Maxwell's equations) govern all electric and magnetic phenomena. The attempts to unify fundamental forces of nature (section 1.4) reflect this same quest for unification.

A related effort is to derive the properties of a bigger, more complex, system from the properties and interactions of its constituent simpler parts. This approach is called **reductionism** and is at the heart of physics. For example, the subject of thermodynamics, developed in the nineteenth century, deals with bulk systems in terms of macroscopic quantities such as temperature, internal energy, entropy, etc. Subsequently, the subjects of kinetic theory and statistical mechanics interpreted these quantities in terms of the properties of the molecular constituents of the bulk system. In particular, the temperature was seen to be related to the average kinetic energy of molecules of the system.

1.2 SCOPE AND EXCITEMENT OF PHYSICS

We can get some idea of the scope of physics by looking at its various sub-disciplines. Basically, there are two domains of interest : macroscopic and microscopic. The macroscopic domain includes phenomena at the laboratory, terrestrial and astronomical scales. The microscopic domain includes atomic, molecular and nuclear

phenomena*. **Classical Physics** deals mainly with macroscopic phenomena and includes subjects like **Mechanics**, **Electrodynamics, Optics** and **Thermodynamics**. Mechanics founded on Newton's laws of motion and the law of gravitation is concerned with the motion (or equilibrium) of particles, rigid and deformable bodies, and general systems of particles. The propulsion of a rocket by a jet of ejecting gases, propagation of water waves or sound waves in air, the equilibrium of a bent rod under a load, etc., are problems of mechanics. Electrodynamics deals with electric and magnetic phenomena associated with charged and magnetic bodies. Its basic laws were given by Coulomb, Oersted, Ampere and Faraday, and encapsulated by Maxwell in his famous set of equations. The motion of a current-carrying conductor in a magnetic field, the response of a circuit to an ac voltage (signal), the working of an antenna, the propagation of radio waves in the ionosphere, etc., are problems of electrodynamics. Optics deals with the phenomena involving light. The working of telescopes and microscopes, colours exhibited by thin films, etc., are topics in optics. Thermodynamics, in contrast to mechanics, does not deal with the motion of bodies as a whole. Rather, it deals with systems in macroscopic equilibrium and is concerned with changes in internal energy, temperature, entropy, etc., of the system through external work and transfer of heat. The efficiency of heat engines and refrigerators, the direction of a physical or chemical process, etc., are problems of interest in thermodynamics.

The microscopic domain of physics deals with the constitution and structure of matter at the minute scales of atoms and nuclei (and even lower scales of length) and their interaction with different probes such as electrons, photons and other elementary particles. Classical physics is inadequate to handle this domain and Quantum Theory is currently accepted as the proper framework for explaining microscopic phenomena. Overall, the edifice of physics is beautiful and imposing and you will appreciate it more as you pursue the subject.

Fig. 1.1 *Theory and experiment go hand in hand in physics and help each other's progress. The alpha scattering experiments of Rutherford gave the nuclear model of the atom.*

You can now see that the scope of physics is truly vast. It covers a tremendous range of magnitude of physical quantities like length, mass, time, energy, etc. At one end, it studies phenomena at the very small scale of length (10^{-14} m or even less) involving electrons, protons, etc.; at the other end, it deals with astronomical phenomena at the scale of galaxies or even the entire universe whose extent is of the order of 10^{26} m. The two length scales differ by a factor of 10^{40} or even more. The range of time scales can be obtained by dividing the length scales by the speed of light : 10^{-22} s to 10^{18} s. The range of masses goes from, say, 10^{-30} kg (mass of an electron) to 10^{55} kg (mass of known observable universe). Terrestrial phenomena lie somewhere in the middle of this range.

* *Recently, the domain intermediate between the macroscopic and the microscopic (the so-called mesoscopic physics), dealing with a few tens or hundreds of atoms, has emerged as an exciting field of research.*

Physics is exciting in many ways. To some people the excitement comes from the elegance and universality of its basic theories, from the fact that a few basic concepts and laws can explain phenomena covering a large range of magnitude of physical quantities. To some others, the challenge in carrying out imaginative new experiments to unlock the secrets of nature, to verify or refute theories, is thrilling. Applied physics is equally demanding. Application and exploitation of physical laws to make useful devices is the most interesting and exciting part and requires great ingenuity and persistence of effort.

What lies behind the phenomenal progress of physics in the last few centuries? Great progress usually accompanies changes in our basic perceptions. First, it was realised that for scientific progress, only qualitative thinking, though no doubt important, is not enough. Quantitative measurement is central to the growth of science, especially physics, because the laws of nature happen to be expressible in precise mathematical equations. The second most important insight was that the basic laws of physics are universal — the same laws apply in widely different contexts. Lastly, the strategy of approximation turned out to be very successful. Most observed phenomena in daily life are rather complicated manifestations of the basic laws. Scientists recognised the importance of extracting the essential features of a phenomenon from its less significant aspects. It is not practical to take into account all the complexities of a phenomenon in one go. A good strategy is to focus first on the essential features, discover the basic principles and then introduce corrections to build a more refined theory of the phenomenon. For example, a stone and a feather dropped from the same height do not reach the ground at the same time. The reason is that the essential aspect of the phenomenon, namely free fall under gravity, is complicated by the presence of air resistance. To get the law of free fall under gravity, it is better to create a situation wherein the air resistance is negligible. We can, for example, let the stone and the feather fall through a long evacuated tube. In that case, the two objects will fall almost at the same rate, giving the basic law that acceleration due to gravity is independent of the mass of the object. With the basic law thus found, we can go back to the feather, introduce corrections due to air resistance, modify the existing theory and try to build a more realistic

Hypothesis, axioms and models

One should not think that everything can be proved with physics and mathematics. All physics, and also mathematics, is based on assumptions, each of which is variously called a hypothesis or axiom or postulate, etc.

For example, the universal law of gravitation proposed by Newton is an assumption or hypothesis, which he proposed out of his ingenuity. Before him, there were several observations, experiments and data, on the motion of planets around the sun, motion of the moon around the earth, pendulums, bodies falling towards the earth etc. Each of these required a separate explanation, which was more or less qualitative. What the universal law of gravitation says is that, *if we assume* that any two bodies in the universe attract each other with a force proportional to the product of their masses and inversely proportional to the square of the distance between them, then we can explain all these observations in one stroke. It not only explains these phenomena, it also allows us to predict the results of future experiments.

A hypothesis is a supposition without assuming that it is true. It would not be fair to ask anybody to prove the universal law of gravitation, because it cannot be proved. It can be verified and substantiated by experiments and observations.

An axiom is a self-evident truth while a model is a theory proposed to explain observed phenomena. But you need not worry at this stage about the nuances in using these words. For example, next year you will learn about Bohr's model of hydrogen atom, in which Bohr assumed that an electron in the hydrogen atom follows certain rules (postutates). Why did he do that? There was a large amount of spectroscopic data before him which no other theory could explain. So Bohr said that if we assume that an atom behaves in such a manner, we can explain all these things at once.

Einstein's special theory of relativity is also based on two postulates, the constancy of the speed of electromagnetic radiation and the validity of physical laws in all inertial frame of reference. It would not be wise to ask somebody to prove that the speed of light in vacuum is constant, independent of the source or observer.

In mathematics too, we need axioms and hypotheses at every stage. Euclid's statement that parallel lines never meet, is a hypothesis. This means that if we assume this statement, we can explain several properties of straight lines and two or three dimensional figures made out of them. But if you don't assume it, you are free to use a different axiom and get a new geometry, as has indeed happened in the past few centuries and decades.

theory of objects falling to the earth under gravity.

1.3 PHYSICS, TECHNOLOGY AND SOCIETY

The connection between physics, technology and society can be seen in many examples. The discipline of thermodynamics arose from the need to understand and improve the working of heat engines. The steam engine, as we know, is inseparable from the Industrial Revolution in England in the eighteenth century, which had great impact on the course of human civilisation. Sometimes technology gives rise to new physics; at other times physics generates new technology. An example of the latter is the wireless communication technology that followed the discovery of the basic laws of electricity and magnetism in the nineteenth century. The applications of physics are not always easy to foresee. As late as 1933, the great physicist Ernest Rutherford had dismissed the possibility of tapping energy from atoms. But only a few years later, in 1938, Hahn and Meitner discovered the phenomenon of neutron-induced fission of uranium, which would serve as the basis of nuclear power reactors and nuclear weapons. Yet another important example of physics giving rise to technology is the silicon 'chip' that triggered the computer revolution in the last three decades of the twentieth century. A most significant area to which physics has and will contribute is the development of alternative energy resources. The fossil fuels of the planet are dwindling fast and there is an urgent need to discover new and affordable sources of energy. Considerable progress has already been made in this direction (for example, in conversion of solar energy, geothermal energy, etc., into electricity), but much more is still to be accomplished.

Table1.1 lists some of the great physicists, their major contribution and the country of origin. You will appreciate from this table the multi-cultural, international character of the scientific endeavour. Table 1.2 lists some important technologies and the principles of physics they are based on. Obviously, these tables are not exhaustive. We urge you to try to add many names and items to these tables with the help of your teachers, good books and websites on science. You will find that this exercise is very educative and also great fun. And, assuredly, it will never end. The progress of science is unstoppable!

Physics is the study of nature and natural phenomena. Physicists try to discover the rules that are operating in nature, on the basis of observations, experimentation and analysis. Physics deals with certain basic rules/laws governing the natural world. What is the nature

Table 1.1 Some physicists from different countries of the world and their major contributions

Name	Major contribution/discovery	Country of Origin
Archimedes	Principle of buoyancy; Principle of the lever	Greece
Galileo Galilei	Law of inertia	Italy
Christiaan Huygens	Wave theory of light	Holland
Isaac Newton	Universal law of gravitation; Laws of motion; Reflecting telescope	U.K.
Michael Faraday	Laws of electromagnetic induction	U.K.
James Clerk Maxwell	Electromagnetic theory; Light-an electromagnetic wave	U.K.
Heinrich Rudolf Hertz	Generation of electromagnetic waves	Germany
J.C. Bose	Ultra short radio waves	India
W.K. Roentgen	X-rays	Germany
J.J. Thomson	Electron	U.K.
Marie Sklodowska Curie	Discovery of radium and polonium; Studies on natural radioactivity	Poland
Albert Einstein	Explanation of photoelectric effect; Theory of relativity	Germany

Name	Major contribution/discovery	Country of Origin
Victor Francis Hess	Cosmic radiation	Austria
R.A. Millikan	Measurement of electronic charge	U.S.A.
Ernest Rutherford	Nuclear model of atom	New Zealand
Niels Bohr	Quantum model of hydrogen atom	Denmark
C.V. Raman	Inelastic scattering of light by molecules	India
Louis Victor de Borglie	Wave nature of matter	France
M.N. Saha	Thermal ionisation	India
S.N. Bose	Quantum statistics	India
Wolfgang Pauli	Exclusion principle	Austria
Enrico Fermi	Controlled nuclear fission	Italy
Werner Heisenberg	Quantum mechanics; Uncertainty principle	Germany
Paul Dirac	Relativistic theory of electron; Quantum statistics	U.K.
Edwin Hubble	Expanding universe	U.S.A.
Ernest Orlando Lawrence	Cyclotron	U.S.A.
James Chadwick	Neutron	U.K.
Hideki Yukawa	Theory of nuclear forces	Japan
Homi Jehangir Bhabha	Cascade process of cosmic radiation	India
Lev Davidovich Landau	Theory of condensed matter; Liquid helium	Russia
S. Chandrasekhar	Chandrasekhar limit, structure and evolution of stars	India
John Bardeen	Transistors; Theory of super conductivity	U.S.A.
C.H. Townes	Maser; Laser	U.S.A.
Abdus Salam	Unification of weak and electromagnetic interactions	Pakistan

of physical laws? We shall now discuss the nature of fundamental forces and the laws that govern the diverse phenomena of the physical world.

1.4 FUNDAMENTAL FORCES IN NATURE*

We all have an intuitive notion of force. In our experience, force is needed to push, carry or throw objects, deform or break them. We also experience the impact of forces on us, like when a moving object hits us or we are in a merry-go-round. Going from this intuitive notion to the proper scientific concept of force is not a trivial matter. Early thinkers like Aristotle had wrong ideas about it. The correct notion of force was arrived at by Isaac Newton in his famous laws of motion. He also gave an explicit form for the force for gravitational attraction between two bodies. We shall learn these matters in subsequent chapters.

In the macroscopic world, besides the gravitational force, we encounter several kinds of forces: muscular force, contact forces between bodies, friction (which is also a contact force parallel to the surfaces in contact), the forces exerted by compressed or elongated springs and taut strings and ropes (tension), the force of buoyancy and viscous force when solids are in

* *Sections 1.4 and 1.5 contain several ideas that you may not grasp fully in your first reading. However, we advise you to read them carefully to develop a feel for some basic aspects of physics. These are some of the areas which continue to occupy the physicists today.*

Table 1.2 Link between technology and physics

Technology	Scientific principle(s)
Steam engine	Laws of thermodynamics
Nuclear reactor	Controlled nuclear fission
Radio and Television	Generation, propagation and detection of electromagnetic waves
Computers	Digital logic
Lasers	Light amplification by stimulated emission of radiation
Production of ultra high magnetic fields	Superconductivity
Rocket propulsion	Newton's laws of motion
Electric generator	Faraday's laws of electromagnetic induction
Hydroelectric power	Conversion of gravitational potential energy into electrical energy
Aeroplane	Bernoulli's principle in fluid dynamics
Particle accelerators	Motion of charged particles in electromagnetic fields
Sonar	Reflection of ultrasonic waves
Optical fibres	Total internal reflection of light
Non-reflecting coatings	Thin film optical interference
Electron microscope	Wave nature of electrons
Photocell	Photoelectric effect
Fusion test reactor (Tokamak)	Magnetic confinement of plasma
Giant Metrewave Radio Telescope (GMRT)	Detection of cosmic radio waves
Bose-Einstein condensate	Trapping and cooling of atoms by laser beams and magnetic fields.

contact with fluids, the force due to pressure of a fluid, the force due to surface tension of a liquid, and so on. There are also forces involving charged and magnetic bodies. In the microscopic domain again, we have electric and magnetic forces, nuclear forces involving protons and neutrons, interatomic and intermolecular forces, etc. We shall get familiar with some of these forces in later parts of this course.

A great insight of the twentieth century physics is that these different forces occurring in different contexts actually arise from only a small number of fundamental forces in nature. For example, the elastic spring force arises due to the net attraction/repulsion between the neighbouring atoms of the spring when the spring is elongated/compressed. This net attraction/repulsion can be traced to the (unbalanced) sum of electric forces between the charged constituents of the atoms.

In principle, this means that the laws for 'derived' forces (such as spring force, friction) are not independent of the laws of fundamental forces in nature. The origin of these derived forces is, however, very complex.

At the present stage of our understanding, we know of four fundamental forces in nature, which are described in brief here :

Albert Einstein (1879-1955)

Albert Einstein, born in Ulm, Germany in 1879, is universally regarded as one of the greatest physicists of all time. His astonishing scientific career began with the publication of three path-breaking papers in 1905. In the first paper, he introduced the notion of light quanta (now called photons) and used it to explain the features of photoelectric effect that the classical wave theory of radiation could not account for. In the second paper, he developed a theory of Brownian motion that was confirmed experimentally a few years later and provided a convincing evidence of the atomic picture of matter. The third paper gave birth to the special theory of relativity that made Einstein a legend in his own life time. In the next decade, he explored the consequences of his new theory which included, among other things, the mass-energy equivalence enshrined in his famous equation $E = mc^2$. He also created the general version of relativity (The General Theory of Relativity), which is the modern theory of gravitation. Some of Einstein's most significant later contributions are: the notion of stimulated emission introduced in an alternative derivation of Planck's blackbody radiation law, static model of the universe which started modern cosmology, quantum statistics of a gas of massive bosons, and a critical analysis of the foundations of quantum mechanics. The year 2005 was declared as International Year of Physics, in recognition of Einstein's monumental contribution to physics, in year 1905, describing revolutionary scientific ideas that have since influenced all of modern physics.

1.4.1 Gravitational Force

The gravitational force is the force of mutual attraction between any two objects by virtue of their masses. It is a universal force. Every object experiences this force due to every other object in the universe. All objects on the earth, for example, experience the force of gravity due to the earth. In particular, gravity governs the motion of the moon and artificial satellites around the earth, motion of the earth and planets around the sun, and, of course, the motion of bodies falling to the earth. It plays a key role in the large-scale phenomena of the universe, such as formation and evolution of stars, galaxies and galactic clusters.

1.4.2 Electromagnetic Force

Electromagnetic force is the force between charged particles. In the simpler case when charges are at rest, the force is given by Coulomb's law : attractive for unlike charges and repulsive for like charges. Charges in motion produce magnetic effects and a magnetic field gives rise to a force on a moving charge. Electric and magnetic effects are, in general, inseparable – hence the name electromagnetic force. Like the gravitational force, electromagnetic force acts over large distances and does not need any intervening medium. It is enormously strong compared to gravity. The electric force between two protons, for example, is 10^{36} times the gravitational force between them, for any fixed distance.

Matter, as we know, consists of elementary charged constituents like electrons and protons. Since the electromagnetic force is so much stronger than the gravitational force, it dominates all phenomena at atomic and molecular scales. (The other two forces, as we shall see, operate only at nuclear scales.) Thus it is mainly the electromagnetic force that governs the structure of atoms and molecules, the dynamics of chemical reactions and the mechanical, thermal and other properties of materials. It underlies the macroscopic forces like 'tension', 'friction', 'normal force', 'spring force', etc.

Gravity is always attractive, while electromagnetic force can be attractive or repulsive. Another way of putting it is that mass comes only in one variety (there is no negative mass), but charge comes in two varieties : positive and negative charge. This is what makes all the difference. Matter is mostly electrically neutral (net charge is zero). Thus, electric force is largely zero and gravitational force dominates terrestrial phenomena. Electric force manifests itself in atmosphere where the atoms are ionised and that leads to lightning.

Satyendranath Bose (1894-1974)

Satyendranath Bose, born in Calcutta in 1894, is among the great Indian physicists who made a fundamental contribution to the advance of science in the twentieth century. An outstanding student throughout, Bose started his career in 1916 as a lecturer in physics in Calcutta University; five years later he joined Dacca University. Here in 1924, in a brilliant flash of insight, Bose gave a new derivation of Planck's law, treating radiation as a gas of photons and employing new statistical methods of counting of photon states. He wrote a short paper on the subject and sent it to Einstein who immediately recognised its great significance, translated it in German and forwarded it for publication. Einstein then applied the same method to a gas of molecules.

The key new conceptual ingredient in Bose's work was that the particles were regarded as indistinguishable, a radical departure from the assumption that underlies the classical Maxwell-Boltzmann statistics. It was soon realised that the new Bose-Einstein statistics was applicable to particles with integers spins, and a new quantum statistics (Fermi-Dirac statistics) was needed for particles with half integers spins satisfying Pauli's exclusion principle. Particles with integers spins are now known as bosons in honour of Bose.

An important consequence of Bose-Einstein statistics is that a gas of molecules below a certain temperature will undergo a phase transition to a state where a large fraction of atoms populate the same lowest energy state. Some seventy years were to pass before the pioneering ideas of Bose, developed further by Einstein, were dramatically confirmed in the observation of a new state of matter in a dilute gas of ultra cold alkali atoms - the Bose-Eintein condensate.

If we reflect a little, the enormous strength of the electromagnetic force compared to gravity is evident in our daily life. When we hold a book in our hand, we are balancing the gravitational force on the book due to the huge mass of the earth by the 'normal force' provided by our hand. The latter is nothing but the net electromagnetic force between the charged constituents of our hand and the book, at the surface in contact. If electromagnetic force were not intrinsically so much stronger than gravity, the hand of the strongest man would crumble under the weight of a feather ! Indeed, to be consistent, in that circumstance, we ourselves would crumble under our own weight !

1.4.3 Strong Nuclear Force

The strong nuclear force binds protons and neutrons in a nucleus. It is evident that without some attractive force, a nucleus will be unstable due to the electric repulsion between its protons. This attractive force cannot be gravitational since force of gravity is negligible compared to the electric force. A new basic force must, therefore, be invoked. The strong nuclear force is the strongest of all fundamental forces, about 100 times the electromagnetic force in strength. It is charge-independent and acts equally between a proton and a proton, a neutron and a neutron, and a proton and a neutron. Its range is, however, extremely small, of about nuclear dimensions (10^{-15}m). It is responsible for the stability of nuclei. The electron, it must be noted, does not experience this force.

Recent developments have, however, indicated that protons and neutrons are built out of still more elementary constituents called quarks.

1.4.4 Weak Nuclear Force

The weak nuclear force appears only in certain nuclear processes such as the β-decay of a nucleus. In β-decay, the nucleus emits an electron and an uncharged particle called neutrino. The weak nuclear force is not as weak as the gravitational force, but much weaker than the strong nuclear and electromagnetic forces. The range of weak nuclear force is exceedingly small, of the order of 10^{-16}m.

1.4.5 Towards Unification of Forces

We remarked in section 1.1 that unification is a basic quest in physics. Great advances in physics often amount to unification of different

Table 1.3 Fundamental forces of nature

Name	Relative strength	Range	Operates among
Gravitational force	10^{-39}	Infinite	All objects in the universe
Weak nuclear force	10^{-13}	Very short, Sub-nuclear size ($\sim 10^{-16}$m)	Some elementary particles, particularly electron and neutrino
Electromagnetic force	10^{-2}	Infinite	Charged particles
Strong nuclear force	1	Short, nuclear size ($\sim 10^{-15}$m)	Nucleons, heavier elementary particles

theories and domains. Newton unified terrestrial and celestial domains under a common law of gravitation. The experimental discoveries of Oersted and Faraday showed that electric and magnetic phenomena are in general inseparable. Maxwell unified electromagnetism and optics with the discovery that light is an electromagnetic wave. Einstein attempted to unify gravity and electromagnetism but could not succeed in this venture. But this did not deter physicists from zealously pursuing the goal of unification of forces.

Recent decades have seen much progress on this front. The electromagnetic and the weak nuclear force have now been unified and are seen as aspects of a single 'electro-weak' force. What this unification actually means cannot be explained here. Attempts have been (and are being) made to unify the electro-weak and the strong force and even to unify the gravitational force with the rest of the fundamental forces. Many of these ideas are still speculative and inconclusive. Table 1.4 summarises some of the milestones in the progress towards unification of forces in nature.

1.5 NATURE OF PHYSICAL LAWS

Physicists explore the universe. Their investigations, based on scientific processes, range from particles that are smaller than atoms in size to stars that are very far away. In addition to finding the facts by observation and experimentation, physicists attempt to discover the laws that summarise (often as mathematical equations) these facts.

In any physical phenomenon governed by different forces, several quantities may change with time. A remarkable fact is that some special physical quantities, however, remain constant in time. They are the conserved quantities of nature. Understanding these conservation principles is very important to describe the observed phenomena quantitatively.

For motion under an external conservative force, the total mechanical energy i.e. the sum of kinetic and potential energy of a body is a constant. The familiar example is the free fall of an object under gravity. Both the kinetic energy of the object and its potential energy change continuously with time, but the sum remains fixed. If the object is released from rest, the initial

Table 1.4 Progress in unification of different forces/domains in nature

Name of the physicist	Year	Achievement in unification
Isaac Newton	1687	Unified celestial and terrestrial mechanics; showed that the same laws of motion and the law of gravitation apply to both the domains.
Hans Christian Oersted Michael Faraday	1820 1830	Showed that electric and magnetic phenomena are inseparable aspects of a unified domain: electromagnetism.
James Clerk Maxwell	1873	Unified electricity, magnetism and optics; showed that light is an electromagnetic wave.
Sheldon Glashow, Abdus Salam, Steven Weinberg	1979	Showed that the 'weak' nuclear force and the electromagnetic force could be viewed as different aspects of a single electro-weak force.
Carlo Rubia, Simon Vander Meer	1984	Verified experimentally the predictions of the theory of electro-weak force.

potential energy is completely converted into the kinetic energy of the object just before it hits the ground. This law restricted for a conservative force should not be confused with the general law of conservation of energy of an isolated system (which is the basis of the First Law of Thermodynamics).

The concept of energy is central to physics and the expressions for energy can be written for every physical system. When all forms of energy e.g., heat, mechanical energy, electrical energy etc., are counted, it turns out that energy is conserved. The general law of conservation of energy is true for all forces and for any kind of transformation between different forms of energy. In the falling object example, if you include the effect of air resistance during the fall and see the situation after the object hits the ground and stays there, the total mechanical energy is obviously not conserved. The general law of energy conservation, however, is still applicable. The initial potential energy of the stone gets transformed into other forms of energy : heat and sound. (Ultimately, sound after it is absorbed becomes heat.) The total energy of the system (stone plus the surroundings) remains unchanged.

The law of conservation of energy is thought to be valid across all domains of nature, from the microscopic to the macroscopic. It is routinely applied in the analysis of atomic, nuclear and elementary particle processes. At the other end, all kinds of violent phenomena occur in the universe all the time. Yet the total energy of the universe (the most ideal isolated system possible!) is believed to remain unchanged.

Until the advent of Einstein's theory of relativity, the law of conservation of mass was regarded as another basic conservation law of nature, since matter was thought to be indestructible. It was (and still is) an important principle used, for example, in the analysis of chemical reactions. A chemical reaction is basically a rearrangement of atoms among different molecules. If the total binding energy of the reacting molecules is less than the total binding energy of the product molecules, the difference appears as heat and the reaction is exothermic. The opposite is true for energy absorbing (endothermic) reactions. However, since the atoms are merely rearranged but not destroyed, the total mass of the reactants is the same as the total mass of the products in a chemical reaction. The changes in the binding energy are too small to be measured as changes in mass.

According to Einstein's theory, mass m is equivalent to energy E given by the relation $E= mc^2$, where c is speed of light in vacuum.

In a nuclear process mass gets converted to energy (or vice-versa). This is the energy which is released in a nuclear power generation and nuclear explosions.

Sir C.V. Raman (1888-1970)

Chandrashekhara Venkata Raman was born on 07 Nov 1888 in Thiruvanaikkaval. He finished his schooling by the age of eleven. He graduated from Presidency College, Madras. After finishing his education he joined financial services of the Indian Government.

While in Kolkata, he started working on his area of interest at Indian Association for Cultivation of Science founded by Dr. Mahendra Lal Sirkar, during his evening hours. His area of interest included vibrations, variety of musical instruments, ultrasonics, diffraction and so on.

In 1917 he was offered Professorship at Calcutta University. In 1924 he was elected 'Fellow' of the Royal Society of London and received Nobel prize in Physics in 1930 for his discovery, now known as Raman Effect.

The **Raman Effect** deals with scattering of light by molecules of a medium when they are excited to vibrational energy levels. This work opened totally new avenues for research for years to come.

He spent his later years at Bangalore, first at Indian Institute of Science and then at Raman Research Institute. His work has inspired generation of young students.

Energy is a scalar quantity. But all conserved quantities are not necessarily scalars. The total linear momentum and the total angular momentum (both vectors) of an isolated system are also conserved quantities. These laws can be derived from Newton's laws of motion in mechanics. But their validity goes beyond mechanics. They are the basic conservation laws of nature in all domains, even in those where Newton's laws may not be valid.

Besides their great simplicity and generality, the conservation laws of nature are very useful in practice too. It often happens that we cannot solve the full dynamics of a complex problem involving different particles and forces. The conservation laws can still provide useful results. For example, we may not know the complicated forces that act during a collision of two automobiles; yet momentum conservation law enables us to bypass the complications and predict or rule out possible outcomes of the collision. In nuclear and elementary particle phenomena also, the conservation laws are important tools of analysis. Indeed, using the conservation laws of energy and momentum for β-decay, Wolfgang Pauli (1900-1958) correctly predicted in 1931 the existence of a new particle (now called neutrino) emitted in β-decay along with the electron.

Conservation laws have a deep connection with symmetries of nature that you will explore in more advanced courses in physics. For example, an important observation is that the laws of nature do not change with time! If you perform an experiment in your laboratory today and repeat the same experiment (on the same objects under identical conditions) after a year, the results are bound to be the same. It turns out that this symmetry of nature with respect to translation (i.e. displacement) in time is equivalent to the law of conservation of energy. Likewise, space is homogeneous and there is no (intrinsically) preferred location in the universe. To put it more clearly, the laws of nature are the same everywhere in the universe. (Caution : the phenomena may differ from place to place because of differing conditions at different locations. For example, the acceleration due to gravity at the moon is one-sixth that at the earth, but the **law of gravitation** is the same both on the moon and the earth.) This symmetry of the laws of nature with respect to translation in space gives rise to conservation of linear momentum. In the same way isotropy of space (no intrinsically preferred direction in space) underlies the law of conservation of angular momentum*. The conservation laws of charge and other attributes of elementary particles can also be related to certain abstract symmetries. Symmetries of space and time and other abstract symmetries play a central role in modern theories of fundamental forces in nature.

Conservation laws in physics

Conservation of energy, momentum, angular momentum, charge, etc are considered to be fundamental laws in physics. At this moment, there are many such conservation laws. Apart from the above four, there are others which mostly deal with quantities which have been introduced in nuclear and particle physics. Some of the conserved quantities are called spin, baryon number, strangeness, hypercharge, etc, but you need not worry about them.

A conservation law is a hypothesis, based on observations and experiments. It is important to remember that a conservation law cannot be proved. It can be verified, or disproved, by experiments. An experiment whose result is in conformity with the law verifies or substantiates the law; it does not prove the law. On the other hand, a single experiment whose result goes against the law is enough to disprove it.

It would be wrong to ask somebody to prove the law of conservation of energy. This law is an outcome of our experience over several centuries, and it has been found to be valid in all experiments, in mechanics, thermodynamics, electromagnetism, optics, atomic and nuclear physics, or any other area.

Some students feel that they can prove the conservation of mechanical energy from a body falling under gravity, by adding the kinetic and potential energies at a point and showing that it turns out to be constant. As pointed out above, this is only a verification of the law, not its proof.

* *See Chapter 7*

SUMMARY

1. Physics deals with the study of the basic laws of nature and their manifestation in different phenomena. The basic laws of physics are universal and apply in widely different contexts and conditions.
2. The scope of physics is wide, covering a tremendous range of magnitude of physical quantities.
3. Physics and technology are related to each other. Sometimes technology gives rise to new physics; at other times physics generates new technology. Both have direct impact on society.
4. There are four fundamental forces in nature that govern the diverse phenomena of the macroscopic and the microscopic world. These are the 'gravitational force', the 'electromagnetic force', the 'strong nuclear force', and the 'weak nuclear force'. Unification of different forces/domains in nature is a basic quest in physics.
5. The physical quantities that remain unchanged in a process are called conserved quantities. Some of the general conservation laws in nature include the laws of conservation of mass, energy, linear momentum, angular momentum, charge, parity, etc. Some conservation laws are true for one fundamental force but not for the other.
6. Conservation laws have a deep connection with symmetries of nature. Symmetries of space and time, and other types of symmetries play a central role in modern theories of fundamental forces in nature.

EXERCISES

Note for the student

The exercises given here are meant to enhance your awareness about the issues surrounding science, technology and society and to encourage you to think and formulate your views about them. The questions may not have clear-cut 'objective' answers.

Note for the teacher

The exercises given here are not for the purpose of a formal examination.

1.1 Some of the most profound statements on the nature of science have come from Albert Einstein, one of the greatest scientists of all time. What do you think did Einstein mean when he said : "The most incomprehensible thing about the world is that it is comprehensible"?

1.2 "Every great physical theory starts as a heresy and ends as a dogma". Give some examples from the history of science of the validity of this incisive remark.

1.3 "Politics is the art of the possible". Similarly, "Science is the art of the soluble". Explain this beautiful aphorism on the nature and practice of science.

1.4 Though India now has a large base in science and technology, which is fast expanding, it is still a long way from realising its potential of becoming a world leader in science. Name some important factors, which in your view have hindered the advancement of science in India.

1.5 No physicist has ever "seen" an electron. Yet, all physicists believe in the existence of electrons. An intelligent but superstitious man advances this analogy to argue that 'ghosts' exist even though no one has 'seen' one. How will you refute his argument?

1.6 The shells of crabs found around a particular coastal location in Japan seem mostly to resemble the legendary face of a Samurai. Given below are two explanations of this observed fact. Which of these strikes you as a scientific explanation ?

(a) A tragic sea accident several centuries ago drowned a young Samurai. As a tribute to his bravery, nature through its inscrutable ways immortalised his face by imprinting it on the crab shells in that area.

(b) After the sea tragedy, fishermen in that area, in a gesture of honour to their dead hero, let free any crab shell caught by them which accidentally had a shape resembling the face of a Samurai. Consequently, the particular shape of the crab shell survived longer and therefore in course of time the shape was genetically propagated. This is an example of evolution by artificial selection.

[Note : This interesting illustration taken from Carl Sagan's 'The Cosmos' highlights the fact that often strange and inexplicable facts which on the first sight appear 'supernatural' actually turn out to have simple scientific explanations. Try to think out other examples of this kind].

1.7 The industrial revolution in England and Western Europe more than two centuries ago was triggered by some key scientific and technological advances. What were these advances ?

1.8 It is often said that the world is witnessing now a second industrial revolution, which will transform the society as radically as did the first. List some key contemporary areas of science and technology, which are responsible for this revolution.

1.9 Write in about 1000 words a fiction piece based on your speculation on the science and technology of the twenty-second century.

1.10 Attempt to formulate your 'moral' views on the practice of science. Imagine yourself stumbling upon a discovery, which has great academic interest but is certain to have nothing but dangerous consequences for the human society. How, if at all, will you resolve your dilemma ?

1.11 Science, like any knowledge, can be put to good or bad use, depending on the user. Given below are some of the applications of science. Formulate your views on whether the particular application is good, bad or something that cannot be so clearly categorised :

(a) Mass vaccination against small pox to curb and finally eradicate this disease from the population. (This has already been successfully done in India).
(b) Television for eradication of illiteracy and for mass communication of news and ideas.
(c) Prenatal sex determination
(d) Computers for increase in work efficiency
(e) Putting artificial satellites into orbits around the Earth
(f) Development of nuclear weapons
(g) Development of new and powerful techniques of chemical and biological warfare).
(h) Purification of water for drinking
(i) Plastic surgery
(j) Cloning

1.12 India has had a long and unbroken tradition of great scholarship — in mathematics, astronomy, linguistics, logic and ethics. Yet, in parallel with this, several superstitious and obscurantistic attitudes and practices flourished in our society and unfortunately continue even today — among many educated people too. How will you use your knowledge of science to develop strategies to counter these attitudes ?

1.13 Though the law gives women equal status in India, many people hold unscientific views on a woman's innate nature, capacity and intelligence, and in practice give them a secondary status and role. Demolish this view using scientific arguments, and by quoting examples of great women in science and other spheres; and persuade yourself and others that, given equal opportunity, women are on par with men.

1.14 "It is more important to have beauty in the equations of physics than to have them agree with experiments". The great British physicist P. A. M. Dirac held this view. Criticize this statement. Look out for some equations and results in this book which strike you as beautiful.

1.15 Though the statement quoted above may be disputed, most physicists do have a feeling that the great laws of physics are at once simple and beautiful. Some of the notable physicists, besides Dirac, who have articulated this feeling, are : Einstein, Bohr, Heisenberg, Chandrasekhar and Feynman. You are urged to make special efforts to get

access to the general books and writings by these and other great masters of physics. (See the Bibliography at the end of this book.) Their writings are truly inspiring !

1.16 Textbooks on science may give you a wrong impression that studying science is dry and all too serious and that scientists are absent-minded introverts who never laugh or grin. This image of science and scientists is patently false. Scientists, like any other group of humans, have their share of humorists, and many have led their lives with a great sense of fun and adventure, even as they seriously pursued their scientific work. Two great physicists of this genre are Gamow and Feynman. You will enjoy reading their books listed in the Bibliography.

CHAPTER TWO

UNITS AND MEASUREMENT

2.1 Introduction
2.2 The international system of units
2.3 Measurement of length
2.4 Measurement of mass
2.5 Measurement of time
2.6 Accuracy, precision of instruments and errors in measurement
2.7 Significant figures
2.8 Dimensions of physical quantities
2.9 Dimensional formulae and dimensional equations
2.10 Dimensional analysis and its applications
Summary
Exercises
Additional exercises

2.1 INTRODUCTION

Measurement of any physical quantity involves comparison with a certain basic, arbitrarily chosen, internationally accepted reference standard called **unit**. The result of a measurement of a physical quantity is expressed by a number (or numerical measure) accompanied by a unit. Although the number of physical quantities appears to be very large, we need only a limited number of units for expressing all the physical quantities, since they are inter-related with one another. The units for the fundamental or base quantities are called **fundamental** or **base units**. The units of all other physical quantities can be expressed as combinations of the base units. Such units obtained for the derived quantities are called **derived units**. A complete set of these units, both the base units and derived units, is known as the **system of units**.

2.2 THE INTERNATIONAL SYSTEM OF UNITS

In earlier time scientists of different countries were using different systems of units for measurement. Three such systems, the CGS, the FPS (or British) system and the MKS system were in use extensively till recently.

The base units for length, mass and time in these systems were as follows :

- In CGS system they were centimetre, gram and second respectively.
- In FPS system they were foot, pound and second respectively.
- In MKS system they were metre, kilogram and second respectively.

The system of units which is at present internationally accepted for measurement is the *Système Internationale d' Unites* (French for International System of Units), abbreviated as SI. The SI, with standard scheme of symbols, units and abbreviations, was developed and recommended by General Conference on Weights and Measures in 1971 for

international usage in scientific, technical, industrial and commercial work. Because SI units used decimal system, conversions within the system are quite simple and convenient. We shall follow the SI units in this book.

In SI, there are seven base units as given in Table 2.1. Besides the seven base units, there are two more units that are defined for (a) plane angle $d\theta$ as the ratio of length of arc ds to the radius r and (b) solid angle $d\Omega$ as the ratio of the intercepted area dA of the spherical surface, described about the apex O as the centre, to the square of its radius r, as shown in Fig. 2.1(a) and (b) respectively. The unit for plane angle is radian with the symbol rad and the unit for the solid angle is steradian with the symbol sr. Both these are dimensionless quantities.

Fig. 2.1 Description of (a) plane angle $d\theta$ and (b) solid angle $d\Omega$.

Table 2.1 SI Base Quantities and Units*

Base quantity	SI Units		
	Name	Symbol	Definition
Length	metre	m	The metre is the length of the path travelled by light in vacuum during a time interval of 1/299,792,458 of a second. (1983)
Mass	kilogram	kg	The kilogram is equal to the mass of the international prototype of the kilogram (a platinum-iridium alloy cylinder) kept at international Bureau of Weights and Measures, at Sevres, near Paris, France. (1889)
Time	second	s	The second is the duration of 9,192,631,770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium-133 atom. (1967)
Electric current	ampere	A	The ampere is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross-section, and placed 1 metre apart in vacuum, would produce between these conductors a force equal to 2×10^{-7} newton per metre of length. (1948)
Thermo dynamic Temperature	kelvin	K	The kelvin, is the fraction 1/273.16 of the thermodynamic temperature of the triple point of water. (1967)
Amount of substance	mole	mol	The mole is the amount of substance of a system, which contains as many elementary entities as there are atoms in 0.012 kilogram of carbon - 12. (1971)
Luminous intensity	candela	cd	The candela is the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency 540×10^{12} hertz and that has a radiant intensity in that direction of 1/683 watt per steradian. (1979)

* *The values mentioned here need not be remembered or asked in a test. They are given here only to indicate the extent of accuracy to which they are measured. With progress in technology, the measuring techniques get improved leading to measurements with greater precision. The definitions of base units are revised to keep up with this progress.*

Table 2.2 Some units retained for general use (Though outside SI)

Name	Symbol	Value in SI Unit
minute	min	60 s
hour	h	60 min = 3600 s
day	d	24 h = 86400 s
year	y	365.25 d = 3.156×10^7 s
degree	°	$1° = (\pi/ 180)$ rad
litre	L	I $dm^3 = 10^{-3}$ m^3
tonne	t	10^3 kg
carat	c	200 mg
bar	bar	0.1 MPa = 10^5 Pa
curie	Ci	3.7×10^{10} s^{-1}
roentgen	R	2.58×10^{-4} C/kg
quintal	q	100 kg
barn	b	100 $fm^2 = 10^{-28}$ m^2
are	a	1 $dam^2 = 10^2$ m^2
hectare	ha	1 $hm^2 = 10^4$ m^2
standard atmospheric pressure	atm	101325 Pa = 1.013×10^5 Pa

Note that when mole is used, the elementary entities must be specified. These entities may be atoms, molecules, ions, electrons, other particles or specified groups of such particles.

We employ units for some physical quantities that can be derived from the seven base units (Appendix A 6). Some derived units in terms of the SI base units are given in (Appendix A 6.1). Some SI derived units are given special names (Appendix A 6.2) and some derived SI units make use of these units with special names and the seven base units (Appendix A 6.3). These are given in Appendix A 6.2 and A 6.3 for your ready reference. Other units retained for general use are given in Table 2.2.

Common SI prefixes and symbols for multiples and sub-multiples are given in Appendix A2. General guidelines for using symbols for physical quantities, chemical elements and nuclides are given in Appendix A7 and those for SI units and some other units are given in Appendix A8 for your guidance and ready reference.

2.3 MEASUREMENT OF LENGTH

You are already familiar with some direct methods for the measurement of length. For example, a metre scale is used for lengths from 10^{-3} m to 10^2 m. A vernier callipers is used for lengths to an accuracy of 10^{-4} m. A screw gauge and a spherometer can be used to measure lengths as less as to 10^{-5} m. To measure lengths beyond these ranges, we make use of some special indirect methods.

2.3.1 Measurement of Large Distances

Large distances such as the distance of a planet or a star from the earth cannot be measured directly with a metre scale. An important method in such cases is the **parallax method**.

When you hold a pencil in front of you against some specific point on the background (a wall) and look at the pencil first through your left eye A (closing the right eye) and then look at the pencil through your right eye B (closing the left eye), you would notice that the position of the pencil seems to change with respect to the point on the wall. This is called **parallax**. The distance between the two points of observation is called the **basis**. In this example, the basis is the distance between the eyes.

To measure the distance D of a far away planet S by the parallax method, we observe it from two different positions (observatories) A and B on the Earth, separated by distance AB = b at the same time as shown in Fig. 2.2. We measure the angle between the two directions along which the planet is viewed at these two points. The ∠ASB in Fig. 2.2 represented by symbol θ is called the **parallax angle** or **parallactic angle**.

As the planet is very far away, $\frac{b}{D} << 1$, and therefore, θ is very small. Then we approximately take AB as an arc of length b of a circle with centre at S and the distance D as

the radius AS = BS so that AB = $b = D\theta$ where θ is in radians.

$$D = \frac{b}{\theta} \tag{2.1}$$

Fig. 2.2 Parallax method.

Having determined D, we can employ a similar method to determine the size or angular diameter of the planet. If d is the diameter of the planet and α the angular size of the planet (the angle subtended by d at the earth), we have

$$\alpha = d/D \tag{2.2}$$

The angle α can be measured from the same location on the earth. It is the angle between the two directions when two diametrically opposite points of the planet are viewed through the telescope. Since D is known, the diameter d of the planet can be determined using Eq. (2.2).

Example 2.1 Calculate the angle of (a) 1^0 (degree) (b) $1'$ (minute of arc or arcmin) and (c) $1''$ (second of arc or arc second) in radians. Use $360^0 = 2\pi$ rad, $1^0 = 60'$ and $1' = 60''$

Answer (a) We have $360^0 = 2\pi$ rad

$1^0 = (\pi/180)$ rad $= 1.745\times10^{-2}$ rad

(b) $1^0 = 60' = 1.745\times10^{-2}$ rad

$1' = 2.908\times10^{-4}$ rad ; 2.91×10^{-4} rad

(c) $1' = 60'' = 2.908\times10^{-4}$ rad

$1'' = 4.847\times10^{-4}$ rad ; 4.85×10^{-6} rad

Example 2.2 A man wishes to estimate the distance of a nearby tower from him. He stands at a point A in front of the tower C and spots a very distant object O in line with AC. He then walks perpendicular to AC up to B, a distance of 100 m, and looks at O and C again. Since O is very distant, the direction BO is practically the same as AO; but he finds the line of sight of C shifted from the original line of sight by an angle $\theta = 40^0$ (θ is known as 'parallax') estimate the distance of the tower C from his original position A.

Fig. 2.3

Answer We have, parallax angle $\theta = 40^0$

From Fig. 2.3, AB = AC tan θ

AC = AB/tanθ = 100 m/tan 40^0

= 100 m/0.8391 = 119 m

Example 2.3 The moon is observed from two diametrically opposite points A and B on Earth. The angle θ subtended at the moon by the two directions of observation is $1°54'$. Given the diameter of the Earth to be about 1.276×10^7 m, compute the distance of the moon from the Earth.

Answer We have $\theta = 1°54' = 114'$

$$= (114\times60)'' \times (4.85\times10^{-6}) \text{ rad}$$

$$= 3.32\times10^{-2} \text{ rad},$$

since $1'' = 4.85\times10^{-6}$ *rad.*

Also $b = AB = 1.276\times10^7$ m

Hence from Eq. (2.1), we have the earth-moon distance,

$$D = b/\theta$$

$$= \frac{1.276\times10^7}{3.32\times10^{-2}}$$

$$= 3.84\times10^8 \text{ m}$$

Example 2.4 The Sun's angular diameter is measured to be $1920''$. The distance D of the Sun from the Earth is 1.496×10^{11} m. What is the diameter of the Sun ?

Answer Sun's angular diameter α

$$= 1920''$$

$$= 1920 \times 4.85 \times 10^{-6} \text{ rad}$$

$$= 9.31 \times 10^{-3} \text{ rad}$$

Sun's diameter

$$d = \alpha D$$

$$= \left(9.31 \times 10^{-3}\right) \times \left(1.496 \times 10^{11}\right) \text{m}$$

$$= 1.39 \times 10^{9} \text{ m}$$ ◀

2.3.2 Estimation of Very Small Distances: Size of a Molecule

To measure a very small size like that of a molecule (10^{-8} m to 10^{-10} m), we have to adopt special methods. We cannot use a screw gauge or similar instruments. Even a microscope has certain limitations. An optical microscope uses visible light to 'look' at the system under investigation. As light has wave like features, the resolution to which an optical microscope can be used is the wavelength of light (A detailed explanation can be found in the Class XII Physics textbook). For visible light the range of wavelengths is from about 4000 Å to 7000 Å (1 angstrom = 1 Å = 10^{-10} m). Hence an optical microscope cannot resolve particles with sizes smaller than this. Instead of visible light, we can use an electron beam. Electron beams can be focussed by properly designed electric and magnetic fields. The resolution of such an electron microscope is limited finally by the fact that electrons can also behave as waves ! (You will learn more about this in class XII). The wavelength of an electron can be as small as a fraction of an angstrom. Such electron microscopes with a resolution of 0.6 Å have been built. They can almost resolve atoms and molecules in a material. In recent times, tunnelling microscopy has been developed in which again the limit of resolution is better than an angstrom. It is possible to estimate the sizes of molecules.

A simple method for estimating the molecular size of oleic acid is given below. Oleic acid is a soapy liquid with large molecular size of the order of 10^{-9} m.

The idea is to first form mono-molecular layer of oleic acid on water surface.

We dissolve 1 cm^3 of oleic acid in alcohol to make a solution of 20 cm^3. Then we take 1 cm^3 of this solution and dilute it to 20 cm^3, using alcohol. So, the concentration of the solution is equal to $\left(\frac{1}{20 \times 20}\right)$cm^3 of oleic acid/cm^3 of solution. Next we lightly sprinkle some lycopodium powder on the surface of water in a large trough and we put one drop of this solution in the water. The oleic acid drop spreads into a thin, large and roughly circular film of molecular thickness on water surface. Then, we quickly measure the diameter of the thin film to get its area A. Suppose we have dropped n drops in the water. Initially, we determine the approximate volume of each drop (V cm^3).

Volume of n drops of solution

$$= nV \text{ cm}^3$$

Amount of oleic acid in this solution

$$= nV\left(\frac{1}{20 \times 20}\right)\text{cm}^3$$

This solution of oleic acid spreads very fast on the surface of water and forms a very thin layer of thickness t. If this spreads to form a film of area A cm^2, then the thickness of the film

$$t = \frac{\text{Volume of the film}}{\text{Area of the film}}$$

or,
$$t = \frac{nV}{20 \times 20\, A}\text{cm} \qquad (2.3)$$

If we assume that the film has mono-molecular thickness, then this becomes the size or diameter of a molecule of oleic acid. The value of this thickness comes out to be of the order of 10^{-9} m.

▶ ***Example 2.5*** If the size of a nucleus (in the range of 10^{-15} to 10^{-14} m) is scaled up to the tip of a sharp pin, what roughly is the size of an atom ? Assume tip of the pin to be in the range 10^{-5}m to 10^{-4}m.

Answer The size of a nucleus is in the range of 10^{-15} m and 10^{-14} m. The tip of a sharp pin is taken to be in the range of 10^{-5} m and 10^{-4} m. Thus we are scaling up by a factor of 10^{10}. An atom roughly of size 10^{-10} m will be scaled up to a size of 1 m. Thus a nucleus in an atom is as small in size as the tip of a sharp pin placed at the centre of a sphere of radius about a metre long. ◀

2.3.3 Range of Lengths

The sizes of the objects we come across in the universe vary over a very wide range. These may vary from the size of the order of 10^{-14} m of the tiny nucleus of an atom to the size of the order of 10^{26} m of the extent of the observable universe. Table 2.6 gives the range and order of lengths and sizes of some of these objects.

We also use certain special length units for short and large lengths. These are

1 fermi = 1 f = 10^{-15} m

1 angstrom = 1 Å = 10^{-10} m

1 astronomical unit = 1 AU (average distance of the Sun from the Earth) = 1.496×10^{11} m

1 light year = 1 ly = 9.46×10^{15} m (distance that light travels with velocity of 3×10^{8} m s^{-1} in 1 year)

1 parsec = 3.08×10^{16} m (Parsec is the distance at which average radius of earth's orbit subtends an angle of 1 arc second)

2.4 MEASUREMENT OF MASS

Mass is a basic property of matter. It does not depend on the temperature, pressure or location of the object in space. The SI unit of mass is kilogram (kg). The prototypes of the International standard kilogram supplied by the International Bureau of Weights and Measures (BIPM) are available in many other laboratories of different countries. In India, this is available at the National Physical Laboratory (NPL), New Delhi.

While dealing with atoms and molecules, the kilogram is an inconvenient unit. In this case, there is an important standard unit of mass, called the **unified atomic mass unit** (u), which has been established for expressing the mass of atoms as

1 unified atomic mass unit = 1u

= (1/12) of the mass of an atom of carbon-12 isotope $\left({}^{12}_{6}C\right)$ including the mass of electrons

= 1.66×10^{-27} kg

Mass of commonly available objects can be determined by a common balance like the one used in a grocery shop. Large masses in the universe like planets, stars, etc., based on Newton's law of gravitation can be measured by using gravitational method (See Chapter 8). For measurement of small masses of atomic/sub-atomic particles etc., we make use of mass spectrograph in which radius of the trajectory is proportional to the mass of a charged particle moving in uniform electric and magnetic field.

2.4.1 Range of Masses

The masses of the objects, we come across in the universe, vary over a very wide range. These may vary from tiny mass of the order of 10^{-30} kg of an electron to the huge mass of about 10^{55} kg of the known universe. Table 2.4 gives the range and order of the typical masses of various objects.

Table 2.3 Range and order of lengths

Size of object or distance	Length (m)
Size of a proton	10^{-15}
Size of atomic nucleus	10^{-14}
Size of hydrogen atom	10^{-10}
Length of typical virus	10^{-8}
Wavelength of light	10^{-7}
Size of red blood corpuscle	10^{-5}
Thickness of a paper	10^{-4}
Height of the Mount Everest above sea level	10^{4}
Radius of the Earth	10^{7}
Distance of moon from the Earth	10^{8}
Distance of the Sun from the Earth	10^{11}
Distance of Pluto from the Sun	10^{13}
Size of our galaxy	10^{21}
Distance to Andromeda galaxy	10^{22}
Distance to the boundary of observable universe	10^{26}

Table 2.4 Range and order of masses

Object	Mass (kg)
Electron	10^{-30}
Proton	10^{-27}
Uranium atom	10^{-25}
Red blood cell	10^{-13}
Dust particle	10^{-9}
Rain drop	10^{-6}
Mosquito	10^{-5}
Grape	10^{-3}
Human	10^{2}
Automobile	10^{3}
Boeing 747 aircraft	10^{8}
Moon	10^{23}
Earth	10^{25}
Sun	10^{30}
Milky way galaxy	10^{41}
Observable Universe	10^{55}

2.5 MEASUREMENT OF TIME

To measure any time interval we need a clock. We now use an **atomic standard of time**, which is based on the periodic vibrations produced in a cesium atom. This is the basis of the **cesium clock**, sometimes called **atomic clock**, used in the national standards. Such standards are available in many laboratories. In the cesium atomic clock, the second is taken as the time needed for 9,192,631,770 vibrations of the radiation corresponding to the transition between the two hyperfine levels of the ground state of cesium-133 atom. The vibrations of the cesium atom regulate the rate of this cesium atomic clock just as the vibrations of a balance wheel regulate an ordinary wristwatch or the vibrations of a small quartz crystal regulate a quartz wristwatch.

The cesium atomic clocks are very accurate. In principle they provide portable standard. The national standard of time interval 'second' as well as the frequency is maintained through four cesium atomic clocks. A cesium atomic clock is used at the National Physical Laboratory (NPL), New Delhi to maintain the Indian standard of time.

In our country, the NPL has the responsibility of maintenance and improvement of physical standards, including that of time, frequency, etc. Note that the Indian Standard Time (IST) is linked to this set of atomic clocks. The efficient cesium atomic clocks are so accurate that they impart the uncertainty in time realisation as $\pm 1 \times 10^{-13}$, i.e. 1 part in 10^{13}. This implies that the uncertainty gained over time by such a device is less than 1 part in 10^{13}; they lose or gain no more than 3 μs in one year. In view of the tremendous accuracy in time measurement, the SI unit of length has been expressed in terms the path length light travels in certain interval of time (1/299, 792, 458 of a second) (Table 2.1).

The time interval of events that we come across in the universe vary over a very wide range. Table 2.5 gives the range and order of some typical time intervals.

You may notice that there is an interesting coincidence between the numbers appearing in Tables 2.3 and 2.5. Note that the ratio of the longest and shortest lengths of objects in our universe is about 10^{41}. Interestingly enough, the ratio of the longest and shortest time intervals associated with the events and objects in our universe is also about 10^{41}. This number, 10^{41} comes up again in Table 2.4, which lists typical masses of objects. The ratio of the largest and smallest masses of the objects in our universe is about $(10^{41})^2$. Is this a curious coincidence between these large numbers purely accidental ?

2.6 ACCURACY, PRECISION OF INSTRUMENTS AND ERRORS IN MEASUREMENT

Measurement is the foundation of all experimental science and technology. The result of every measurement by any measuring instrument contains some uncertainty. This uncertainty is called **error**. Every calculated quantity which is based on measured values, also has an error. We shall distinguish between two terms: **accuracy** and **precision**. The accuracy of a measurement is a measure of how close the measured value is to the true value of the quantity. Precision tells us to what resolution or limit the quantity is measured.

The accuracy in measurement may depend on several factors, including the limit or the resolution of the measuring instrument. For example, suppose the true value of a certain length is near 3.678 cm. In one experiment, using a measuring instrument of resolution 0.1 cm, the measured value is found to be 3.5 cm, while in another experiment using a measuring device of greater resolution, say 0.01 cm, the length is determined to be 3.38 cm. The first measurement has more accuracy (because it is

Table 2.5 Range and order of time intervals

Event	Time interval(s)
Life-span of most unstable particle	10^{-24}
Time required for light to cross a nuclear distance	10^{-22}
Period of x-rays	10^{-19}
Period of atomic vibrations	10^{-15}
Period of light wave	10^{-15}
Life time of an excited state of an atom	10^{-8}
Period of radio wave	10^{-6}
Period of a sound wave	10^{-3}
Wink of eye	10^{-1}
Time between successive human heart beats	10^{0}
Travel time for light from moon to the Earth	10^{0}
Travel time for light from the Sun to the Earth	10^{2}
Time period of a satellite	10^{4}
Rotation period of the Earth	10^{5}
Rotation and revolution periods of the moon	10^{6}
Revolution period of the Earth	10^{7}
Travel time for light from nearest star	10^{8}
Average human life-span	10^{9}
Age of Egyptian pyramids	10^{11}
Time since dinosaurs became extinct	10^{15}
Age of the universe	10^{17}

closer to the true value) but less precision (its resolution is only 0.1 cm), while the second measurement is less accurate but more precise. Thus every measurement is approximate due to errors in measurement. In general, the errors in measurement can be broadly classified as (a) systematic errors and (b) random errors.

Systematic errors

The **systematic errors** are those errors that tend to be in one direction, either positive or negative. Some of the sources of systematic errors are :

(a) **Instrumental errors** that arise from the errors due to imperfect design or calibration of the measuring instrument, zero error in the instrument, etc. For example, the temperature graduations of a thermometer may be inadequately calibrated (it may read 104 °C at the boiling point of water at STP whereas it should read 100 °C); in a vernier callipers the zero mark of vernier scale may not coincide with the zero mark of the main scale, or simply an ordinary metre scale may be worn off at one end.

(b) **Imperfection in experimental technique or procedure** To determine the temperature of a human body, a thermometer placed under the armpit will always give a temperature lower than the actual value of the body temperature. Other external conditions (such as changes in temperature, humidity, wind velocity, etc.) during the experiment may systematically affect the measurement.

(c) **Personal errors** that arise due to an individual's bias, lack of proper setting of the apparatus or individual's carelessness in taking observations without observing proper precautions, etc. For example, if you, by habit, always hold your head a bit too far to the right while reading the position of a needle on the scale, you will introduce an error due to **parallax**.

Systematic errors can be minimised by improving experimental techniques, selecting better instruments and removing personal bias as far as possible. For a given set-up, these errors may be estimated to a certain extent and the necessary corrections may be applied to the readings.

Random errors

The **random errors** are those errors, which occur irregularly and hence are random with respect

to sign and size. These can arise due to random and unpredictable fluctuations in experimental conditions (e.g. unpredictable fluctuations in temperature, voltage supply, mechanical vibrations of experimental set-ups, etc), personal (unbiased) errors by the observer taking readings, etc. For example, when the same person repeats the same observation, it is very likely that he may get different readings everytime.

Least count error

The smallest value that can be measured by the measuring instrument is called its **least count**. All the readings or measured values are good only up to this value.

The **least count error** is the error associated with the resolution of the instrument. For example, a vernier callipers has the least count as 0.01 cm; a spherometer may have a least count of 0.001 cm. Least count error belongs to the category of random errors but within a limited size; it occurs with both systematic and random errors. If we use a metre scale for measurement of length, it may have graduations at 1 mm division scale spacing or interval.

Using instruments of higher precision, improving experimental techniques, etc., we can reduce the least count error. Repeating the observations several times and taking the arithmetic mean of all the observations, the mean value would be very close to the true value of the measured quantity.

2.6.1 Absolute Error, Relative Error and Percentage Error

(a) Suppose the values obtained in several measurements are $a_1, a_2, a_3 ...., a_n$. The arithmetic mean of these values is taken as the best possible value of the quantity under the given conditions of measurement as :

$$a_{mean} = (a_1 + a_2 + a_3 + ... + a_n) / n \quad (2.4)$$

or,

$$a_{mean} = \sum_{i=1}^{n} a_i / n \quad (2.5)$$

This is because, as explained earlier, it is reasonable to suppose that individual measurements are as likely to overestimate as to underestimate the true value of the quantity.

The magnitude of the difference between the true value of the quantity and the individual measurement value is called the absolute error of the measurement. This is denoted by $|\Delta a|$. In absence of any other method of knowing true value, we considered arithmatic mean as the true value. Then the errors in the individual measurement values are

$\Delta a_1 = a_{mean} - a_1$,
$\Delta a_2 = a_{mean} - a_2$,
....
....
$\Delta a_n = a_{mean} - a_n$

The Δa calculated above may be positive in certain cases and negative in some other cases. But absolute error $|\Delta a|$ will always be positive.

(b) The arithmetic mean of all the *absolute errors* is taken as the *final* or *mean absolute error* of the value of the physical quantity a. It is represented by Δa_{mean}.

Thus,

$$\Delta a_{mean} = (|\Delta a_1| + |\Delta a_2| + |\Delta a_3| + ... + |\Delta a_n|)/n \quad (2.6)$$

$$= \sum_{i=1}^{n} |\Delta a_i| / n \quad (2.7)$$

If we do a single measurement, the value we get may be in the range $a_{mean} \pm \Delta a_{mean}$

i.e. $a = a_{mean} \pm \Delta a_{mean}$

or,

$$a_{mean} - \Delta a_{mean} \le a \le a_{mean} + \Delta a_{mean} \quad (2.8)$$

This implies that any measurement of the physical quantity a is likely to lie between $(a_{mean} + \Delta a_{mean})$ and $(a_{mean} - \Delta a_{mean})$.

(c) Instead of the absolute error, we often use the **relative error** or the **percentage error** (δa). **The relative error is the ratio of the mean absolute error Δa_{mean} to the mean value a_{mean} of the quantity measured.**

Relative error = $\Delta a_{mean}/a_{mean}$ (2.9)

When the relative error is expressed in per cent, it is called the **percentage error** (δa).

Thus, Percentage error

$$\delta a = (\Delta a_{mean}/a_{mean}) \times 100\% \quad (2.10)$$

Let us now consider an example.

Example 2.6 Two clocks are being tested against a standard clock located in a national laboratory. At 12:00:00 noon by the standard clock, the readings of the two clocks are :

	Clock 1	Clock 2
Monday	12:00:05	10:15:06
Tuesday	12:01:15	10:14:59
Wednesday	11:59:08	10:15:18
Thursday	12:01:50	10:15:07
Friday	11:59:15	10:14:53
Saturday	12:01:30	10:15:24
Sunday	12:01:19	10:15:11

If you are doing an experiment that requires precision time interval measurements, which of the two clocks will you prefer ?

Answer The range of variation over the seven days of observations is 162 s for clock 1, and 31 s for clock 2. The average reading of clock 1 is much closer to the standard time than the average reading of clock 2. The important point is that a clock's zero error is not as significant for precision work as its variation, because a 'zero-error' can always be easily corrected. Hence clock 2 is to be preferred to clock 1. ◀

Example 2.7 We measure the period of oscillation of a simple pendulum. In successive measurements, the readings turn out to be 2.63 s, 2.56 s, 2.42 s, 2.71s and 2.80 s. Calculate the absolute errors, relative error or percentage error.

Answer The mean period of oscillation of the pendulum

$$T = \frac{(2.63 + 2.56 + 2.42 + 2.71 + 2.80)\text{s}}{5}$$

$$= \frac{13.12}{5}\ \text{s}$$

$$= 2.624\ \text{s}$$

$$= 2.62\ \text{s}$$

As the periods are measured to a resolution of 0.01 s, all times are to the second decimal; it is proper to put this mean period also to the second decimal.

The errors in the measurements are

2.63 s – 2.62 s = 0.01 s
2.56 s – 2.62 s = – 0.06 s
2.42 s – 2.62 s = – 0.20 s
2.71 s – 2.62 s = 0.09 s
2.80 s – 2.62 s = 0.18 s

Note that the errors have the same units as the quantity to be measured.

The arithmetic mean of all the absolute errors (for arithmetic mean, we take only the magnitudes) is

$$\Delta T_{mean} = [(0.01 + 0.06 + 0.20 + 0.09 + 0.18)\text{s}]/5$$
$$= 0.54\ \text{s}/5$$
$$= 0.11\ \text{s}$$

That means, the period of oscillation of the simple pendulum is (2.62 ± 0.11) s i.e. it lies between (2.62 + 0.11) s and (2.62 – 0.11) s or between 2.73 s and 2.51 s. As the arithmetic mean of all the absolute errors is 0.11 s, there is already an error in the tenth of a second. Hence there is no point in giving the period to a hundredth. A more correct way will be to write

$$T = 2.6 \pm 0.1\ \text{s}$$

Note that the last numeral 6 is unreliable, since it may be anything between 5 and 7. We indicate this by saying that the measurement has two significant figures. In this case, the two significant figures are 2, which is reliable and 6, which has an error associated with it. You will learn more about the significant figures in section 2.7.

For this example, the relative error or the percentage error is

$$\delta a = \frac{0.1}{2.6} \times 100 = 4\% \quad ◀$$

2.6.2 Combination of Errors

If we do an experiment involving several measurements, we must know how the errors in all the measurements combine. For example,

How will you measure the length of a line?

What a naïve question, at this stage, you might say! But what if it is not a straight line? Draw a zigzag line in your copy, or on the blackboard. Well, not too difficult again. You might take a thread, place it along the line, open up the thread, and measure its length.

Now imagine that you want to measure the length of a national highway, a river, the railway track between two stations, or the boundary between two states or two nations. If you take a string of length 1 metre or 100 metre, keep it along the line, shift its position every time, the arithmetic of man-hours of labour and expenses on the project is not commensurate with the outcome. Moreover, errors are bound to occur in this enormous task. There is an interesting fact about this. France and Belgium share a common international boundary, whose length mentioned in the official documents of the two countries differs substantially!

Go one step beyond and imagine the coastline where land meets sea. Roads and rivers have fairly mild bends as compared to a coastline. Even so, all documents, including our school books, contain information on the length of the coastline of Gujarat or Andhra Pradesh, or the common boundary between two states, etc. Railway tickets come with the distance between stations printed on them. We have 'milestones' all along the roads indicating the distances to various towns. So, how is it done?

One has to decide how much error one can tolerate and optimise cost-effectiveness. If you want smaller errors, it will involve high technology and high costs. Suffice it to say that it requires fairly advanced level of physics, mathematics, engineering and technology. It belongs to the areas of fractals, which has lately become popular in theoretical physics. Even then one doesn't know how much to rely on the figure that props up, as is clear from the story of France and Belgium. Incidentally, this story of the France-Belgium discrepancy appears on the first page of an advanced Physics book on the subject of fractals and chaos!

density is the ratio of the mass to the volume of the substance. If we have errors in the measurement of mass and of the sizes or dimensions, we must know what the error will be in the density of the substance. To make such estimates, we should learn how errors combine in various mathematical operations. For this, we use the following procedure.

(a) *Error of a sum or a difference*

Suppose two physical quantities A and B have measured values $A \pm \Delta A$, $B \pm \Delta B$ respectively where ΔA and ΔB are their absolute errors. We wish to find the error ΔZ in the sum

$$Z = A + B.$$

We have by addition, $Z \pm \Delta Z$

$$= (A \pm \Delta A) + (B \pm \Delta B).$$

The maximum possible error in Z

$$\Delta Z = \Delta A + \Delta B$$

For the difference $Z = A - B$, we have

$$Z \pm \Delta Z = (A \pm \Delta A) - (B \pm \Delta B)$$

$$= (A - B) \pm \Delta A \pm \Delta B$$

or, $\pm \Delta Z = \pm \Delta A \pm \Delta B$

The maximum value of the error ΔZ is again $\Delta A + \Delta B$.

Hence the rule : When two quantities are added or subtracted, the absolute error in the final result is the sum of the absolute errors in the individual quantities.

▶ ***Example 2.8*** The temperatures of two bodies measured by a thermometer are $t_1 = 20\ ^0C \pm 0.5\ ^0C$ and $t_2 = 50\ ^0C \pm 0.5\ ^0C$. Calculate the temperature difference and the error theirin.

Answer $t' = t_2 - t_1 = (50\ ^0C \pm 0.5\ ^0C) - (20^0C \pm 0.5\ ^0C)$

$t' = 30\ ^0C \pm 1\ ^0C$ ◀

(b) Error of a product or a quotient

Suppose $Z = AB$ and the measured values of A and B are $A \pm \Delta A$ and $B \pm \Delta B$. Then

$$Z \pm \Delta Z = (A \pm \Delta A)\ (B \pm \Delta B)$$

$$= AB \pm B\,\Delta A \pm A\,\Delta B \pm \Delta A\,\Delta B.$$

Dividing LHS by Z and RHS by AB we have,

$$1 \pm (\Delta Z/Z) = 1 \pm (\Delta A/A) \pm (\Delta B/B) \pm (\Delta A/A)(\Delta B/B).$$

Since ΔA and ΔB are small, we shall ignore their product.

Hence the maximum relative error

$$\Delta Z/\ Z = (\Delta A/A) + (\Delta B/B).$$

You can easily verify that this is true for division also.

Hence the rule : When two quantities are multiplied or divided, the relative error in the result is the sum of the relative errors in the multipliers.

Example 2.9 The resistance $R = V/I$ where $V = (100 \pm 5)$V and $I = (10 \pm 0.2)$A. Find the percentage error in R.

Answer The percentage error in V is 5% and in I it is 2%. The total error in R would therefore be 5% + 2% = 7%.

Example 2.10 Two resistors of resistances $R_1 = 100 \pm 3$ ohm and $R_2 = 200 \pm 4$ ohm are connected (a) in series, (b) in parallel. Find the equivalent resistance of the (a) series combination, (b) parallel combination. Use for (a) the relation $R = R_1 + R_2$ and for (b) $\frac{1}{R'} = \frac{1}{R_1} + \frac{1}{R_2}$ and $\frac{\Delta R'}{R'^2} = \frac{\Delta R_1}{R_1^2} + \frac{\Delta R_2}{R_2^2}$.

Answer (a) The equivalent resistance of series combination

$R = R_1 + R_2 = (100 \pm 3)$ ohm + (200 ± 4) ohm

$= 300 \pm 7$ ohm.

(b) The equivalent resistance of parallel combination

$$R' = \frac{R_1 R_2}{R_1 + R_2} = \frac{200}{3} = 66.7 \text{ ohm}$$

Then, from $\frac{1}{R'} = \frac{1}{R_1} + \frac{1}{R_2}$

we get,

$$\frac{\Delta R'}{R'^2} = \frac{\Delta R_1}{R_1^2} + \frac{\Delta R_2}{R_2^2}$$

$$\Delta R' = \left(R'^2\right)\frac{\Delta R_1}{R_1^2} + \left(R'^2\right)\frac{\Delta R_2}{R_2^2}$$

$$= \left(\frac{66.7}{100}\right)^2 3 + \left(\frac{66.7}{200}\right)^2 4$$

$= 1.8$

Then, $R' = 66.7 \pm 1.8$ ohm

(Here, ΔR is expresed as 1.8 instead of 2 to keep in confirmity with the rules of significant figures.)

(c) Error in case of a measured quantity raised to a power

Suppose $Z = A^2$,

Then,

$$\Delta Z/Z = (\Delta A/A) + (\Delta A/A) = 2\,(\Delta A/A).$$

Hence, the relative error in A^2 is two times the error in A.

In general, if $Z = A^p B^q / C^r$

Then,

$$\Delta Z/Z = p\,(\Delta A/A) + q\,(\Delta B/B) + r\,(\Delta C/C).$$

Hence the rule : The relative error in a physical quantity raised to the power k is the k times the relative error in the individual quantity.

Example 2.11 Find the relative error in Z, if $Z = A^4 B^{1/3}/CD^{3/2}$.

Answer The relative error in Z is $\Delta Z/Z = 4(\Delta A/A) + (1/3)(\Delta B/B) + (\Delta C/C) + (3/2)(\Delta D/D)$.

Example 2.12 The period of oscillation of a simple pendulum is $T = 2\pi\sqrt{L/g}$. Measured value of L is 20.0 cm known to 1 mm accuracy and time for 100 oscillations of the pendulum is found to be 90 s using a wrist watch of 1 s resolution. What is the accuracy in the determination of g ?

Answer $g = 4\pi^2 L/T^2$

Here, $T = \frac{t}{n}$ and $\Delta T = \frac{\Delta t}{n}$. Therefore, $\frac{\Delta T}{T} = \frac{\Delta t}{t}$.

The errors in both L and t are the least count errors. Therefore,

$(\Delta g/g) = (\Delta L/L) + 2(\Delta T/T)$

$$= \frac{0.1}{20.0} + 2\left(\frac{1}{90}\right) = 0.032$$

Thus, the percentage error in g is

$100\,(\Delta g/g) = 100(\Delta L/L) + 2 \times 100\,(\Delta T/T)$

$= 3\%$

2.7 SIGNIFICANT FIGURES

As discussed above, every measurement involves errors. Thus, the result of measurement should be reported in a way that indicates the precision of measurement. Normally, the reported result of measurement is a number that includes all digits in the number that are known reliably plus the first digit that is uncertain. The reliable digits plus

the first uncertain digit are known as **significant digits** or **significant figures**. If we say the period of oscillation of a simple pendulum is 1.62 s, the digits 1 and 6 are reliable and certain, while the digit 2 is uncertain. Thus, the measured value has three significant figures. The length of an object reported after measurement to be 287.5 cm has four significant figures, the digits 2, 8, 7 are certain while the digit 5 is uncertain. Clearly, reporting the result of measurement that includes more digits than the significant digits is superfluous and also misleading since it would give a wrong idea about the precision of measurement.

The rules for determining the number of significant figures can be understood from the following examples. Significant figures indicate, as already mentioned, the precision of measurement which depends on the least count of the measuring instrument. **A choice of change of different units does not change the number of significant digits or figures in a measurement.** This important remark makes most of the following observations clear:

(1) For example, the length 2.308 cm has four significant figures. But in different units, the same value can be written as 0.02308 m or 23.08 mm or 23080 μm.

All these numbers have the same number of significant figures (digits 2, 3, 0, 8), namely four. This shows that the location of decimal point is of no consequence in determining the number of significant figures.

The example gives the following rules :

- **All the non-zero digits are significant.**
- **All the zeros between two non-zero digits are significant, no matter where the decimal point is, if at all.**
- **If the number is less than 1, the zero(s) on the right of decimal point but to the left of the first non-zero digit are not significant.** [In 0.00 2308, the underlined zeroes are not significant].
- **The terminal or trailing zero(s) in a number without a decimal point are not significant.**
  [Thus 123 m = 12300 cm = 123000 mm has *three* significant figures, the trailing zero(s) being not significant.] However, you can also see the next observation.
- **The trailing zero(s) in a number with a decimal point are significant.**
  [The numbers 3.500 or 0.06900 have four significant figures each.]

(2) There can be some confusion regarding the trailing zero(s). Suppose a length is reported to be 4.700 m. It is evident that the zeroes here are meant to convey the precision of measurement and are, therefore, significant. [If these were not, it would be superfluous to write them explicitly, the reported measurement would have been simply 4.7 m]. Now suppose we change units, then

4.700 m = 470.0 cm = 4700 mm = 0.004700 km

Since the last number has trailing zero(s) in a number with no decimal, we would conclude erroneously from observation (1) above that the number has *two* significant figures, while in fact, it has four significant figures and a mere change of units cannot change the number of significant figures.

(3) **To remove such ambiguities in determining the number of significant figures, the best way is to report every measurement in scientific notation (in the power of 10).** In this notation, every number is expressed as $a \times 10^b$, where a is a number between 1 and 10, and b is any positive or negative exponent (or power) of 10. In order to get an approximate idea of the number, we may round off the number a to 1 (for $a \leq 5$) and to 10 (for $5 < a \leq 10$). Then the number can be expressed approximately as 10^b in which the exponent (or power) b of 10 is called **order of magnitude** of the physical quantity. When only an estimate is required, the quantity is of the order of 10^b. For example, the diameter of the earth (1.28×10^7m) is of the order of 10^7m with the order of magnitude 7. The diameter of hydrogen atom (1.06×10^{-10}m) is of the order of 10^{-10}m, with the order of magnitude −10. Thus, the diameter of the earth is 17 orders of magnitude larger than the hydrogen atom.

It is often customary to write the decimal after the first digit. Now the confusion mentioned in (a) above disappears :

$$4.700 \text{ m} = 4.700 \times 10^2 \text{ cm}$$
$$= 4.700 \times 10^3 \text{ mm} = 4.700 \times 10^{-3} \text{ km}$$

The power of 10 is irrelevant to the determination of significant figures. However, all

zeroes appearing in the base number in the scientific notation are significant. Each number in this case has *four* significant figures.

Thus, in the scientific notation, no confusion arises about the trailing zero(s) in the base number a. They are always significant.

(4) The scientific notation is ideal for reporting measurement. But if this is not adopted, we use the rules adopted in the preceding example :

- **For a number greater than 1, without any decimal, the trailing zero(s) are not significant.**
- **For a number with a decimal, the trailing zero(s) are significant.**

(5) The digit 0 conventionally put on the left of a decimal for a number less than 1 (like 0.1250) is never significant. However, the zeroes at the end of such number are significant in a measurement.

(6) The multiplying or dividing factors which are neither rounded numbers nor numbers representing measured values are exact and have infinite number of significant digits. For example in $r = \frac{d}{2}$ or $s = 2\pi r$, the factor 2 is an exact number and it can be written as 2.0, 2.00 or 2.0000 as required. Similarly, in $T = \frac{t}{n}$, n is an exact number.

2.7.1 Rules for Arithmetic Operations with Significant Figures

The result of a calculation involving approximate measured values of quantities (i.e. values with limited number of significant figures) must reflect the uncertainties in the original measured values. It cannot be more accurate than the original measured values themselves on which the result is based. In general, the final result should not have more significant figures than the original data from which it was obtained. Thus, if mass of an object is measured to be, say, 4.237 g (four significant figures) and its volume is measured to be 2.51 cm^3, then its density, by mere arithmetic division, is 1.68804780876 g/cm^3 upto 11 decimal places. It would be clearly absurd and irrelevant to record the calculated value of density to such a precision when the measurements on which the value is based, have much less precision. The following rules for arithmetic operations with significant figures ensure that the final result of a calculation is shown with the precision that is consistent with the precision of the input measured values :

(1) **In multiplication or division, the final result should retain as many significant figures as are there in the original number with the least significant figures.**

Thus, in the example above, density should be reported to *three* significant figures.

$$\text{Density} = \frac{4.237\text{g}}{2.51\ \text{cm}^3} = 1.69\ \text{g cm}^{-3}$$

Similarly, if the speed of light is given as 3.00×10^8 m s^{-1} (*three* significant figures) and one year (1y = 365.25 d) has 3.1557×10^7 s (*five* significant figures), the light year is 9.47×10^{15} m (*three* significant figures).

(2) **In addition or subtraction, the final result should retain as many decimal places as are there in the number with the least decimal places.**

For example, the sum of the numbers 436.32 g, 227.2 g and 0.301 g by mere arithmetic addition, is 663.821 g. But the least precise measurement (227.2 g) is correct to only one decimal place. The final result should, therefore, be rounded off to 663.8 g.

Similarly, the difference in length can be expressed as :

0.307 m – 0.304 m = 0.003 m = 3×10^{-3} m.

Note that we should not use the *rule* (1) applicable for multiplication and division and write 664 g as the result in the example of **addition** and 3.00×10^{-3} m in the example of **subtraction**. They do not convey the precision of measurement properly. For addition and subtraction, the rule is in terms of decimal places.

2.7.2 Rounding off the Uncertain Digits

The result of computation with approximate numbers, which contain more than one uncertain digit, should be rounded off. The rules for rounding off numbers to the appropriate significant figures are obvious in most cases. A number 2.74$\underline{6}$ rounded off to three significant figures is 2.75, while the number 2.743 would be 2.74. The *rule* by convention is that the **preceding digit is raised by 1 if the**

insignificant digit to be dropped (the underlined digit in this case) is more than 5, and is left unchanged if the latter is less than 5. But what if the number is 2.74$\underline{5}$ in which the insignificant digit is 5. Here, the convention is that **if the preceding digit is even, the insignificant digit is simply dropped and, if it is odd, the preceding digit is raised by** 1. Then, the number 2.74$\underline{5}$ rounded off to three significant figures becomes 2.74. On the other hand, the number 2.73$\underline{5}$ rounded off to three significant figures becomes 2.74 since the preceding digit is odd.

In any involved or complex multi-step calculation, you should retain, in intermediate steps, one digit more than the significant digits and round off to proper significant figures at the end of the calculation. Similarly, a number known to be within many significant figures, such as in 2.99792458×10^8 m/s for the speed of light in vacuum, is rounded off to an approximate value 3×10^8 m/s , which is often employed in computations. Finally, remember that exact numbers that appear in formulae like 2π in $T = 2\pi\sqrt{\frac{L}{g}}$, have a large (infinite) number of significant figures. The value of π = 3.1415926.... is known to a large number of significant figures. You may take the value as 3.142 or 3.14 for π, with limited number of significant figures as required in specific cases.

▶ ***Example 2.13*** Each side of a cube is measured to be 7.203 m. What are the total surface area and the volume of the cube to appropriate significant figures?

Answer The number of significant figures in the measured length is 4. The calculated area and the volume should therefore be rounded off to 4 significant figures.

Surface area of the cube $= 6(7.203)^2$ m^2

$= 311.299254$ m^2

$= 311.3$ m^2

Volume of the cube $= (7.203)^3$ m^3

$= 373.714754$ m^3

$= 373.7$ m^3 ◀

▶ ***Example 2.14*** 5.74 g of a substance occupies 1.2 cm^3. Express its density by keeping the significant figures in view.

Answer There are 3 significant figures in the measured mass whereas there are only 2 significant figures in the measured volume. Hence the density should be expressed to only 2 significant figures.

$$\text{Density} = \frac{5.74}{1.2}\ \text{g cm}^{-3}$$

$$= 4.8\ \text{g cm}^{-3}.$$ ◀

2.7.3 Rules for Determining the Uncertainty in the Results of Arithmatic Calculations

The rules for determining the uncertainty or error in the number/measured quantity in arithmetic operations can be understood from the following examples.

(1) If the length and breadth of a thin rectangular sheet are measured, using a metre scale as 16.2 cm and, 10.1 cm respectively, there are three significant figures in each measurement. It means that the length l may be written as

$$l = 16.2 \pm 0.1\ \text{cm}$$

$$= 16.2\ \text{cm} \pm 0.6\ \%.$$

Similarly, the breadth b may be written as

$$b = 10.1 \pm 0.1\ \text{cm}$$

$$= 10.1\ \text{cm} \pm 1\ \%$$

Then, the error of the product of two (or more) experimental values, using the combination of errors rule, will be

$$l\,b = 163.62\ \text{cm}^2 \pm 1.6\%$$

$$= 163.62 \pm 2.6\ \text{cm}^2$$

This leads us to quote the final result as

$$l\,b = 164 \pm 3\ \text{cm}^2$$

Here 3 cm^2 is the uncertainty or error in the estimation of area of rectangular sheet.

(2) **If a set of experimental data is specified to n significant figures, a result obtained by combining the data will also be valid to n significant figures.**

However, if data are subtracted, the number of significant figures can be reduced.

For example, 12.9 g – 7.06 g, both specified to three significant figures, cannot properly be evaluated as 5.84 g but only as 5.8 g, as uncertainties in subtraction or addition combine in a different fashion (smallest number of decimal places rather than the number of significant figures in any of the number added or subtracted).

(3) **The relative error of a value of number specified to significant figures depends not only on n but also on the number itself.**

For example, the accuracy in measurement of mass 1.02 g is ± 0.01 g whereas another measurement 9.89 g is also accurate to ± 0.01 g. The relative error in 1.02 g is

$$= (\pm 0.01/1.02) \times 100\ \%$$
$$= \pm 1\%$$

Similarly, the relative error in 9.89 g is

$$= (\pm 0.01/9.89) \times 100\ \%$$
$$= \pm 0.1\ \%$$

Finally, remember that **intermediate results in a multi-step computation should be calculated to one more significant figure in every measurement than the number of digits in the least precise measurement.** These should be justified by the data and then the arithmetic operations may be carried out; otherwise rounding errors can build up. For example, the reciprocal of 9.58, calculated (after rounding off) to the same number of significant figures (three) is 0.104, but the reciprocal of 0.104 calculated to three significant figures is 9.62. However, if we had written 1/9.58 = 0.1044 and then taken the reciprocal to three significant figures, we would have retrieved the original value of 9.58.

This example justifies the idea to retain one more extra digit (than the number of digits in the least precise measurement) in intermediate steps of the complex multi-step calculations in order to avoid additional errors in the process of rounding off the numbers.

2.8 DIMENSIONS OF PHYSICAL QUANTITIES

The nature of a physical quantity is described by its dimensions. All the physical quantities represented by derived units can be expressed in terms of some combination of seven fundamental or base quantities. We shall call these base quantities as the seven dimensions of the physical world, which are denoted with square brackets [ ]. Thus, length has the dimension [L], mass [M], time [T], electric current [A], thermodynamic temperature [K], luminous intensity [cd], and amount of substance [mol]. **The dimensions of a physical quantity are the powers (or exponents) to which the base quantities are raised to represent that quantity**. Note that using the square brackets [ ] round a quantity means that we are dealing with '**the dimensions of**' the quantity.

In mechanics, all the physical quantities can be written in terms of the dimensions [L], [M] and [T]. For example, the volume occupied by an object is expressed as the product of length, breadth and height, or three lengths. Hence the dimensions of volume are $[L] \times [L] \times [L] = [L]^3 = [L^3]$. As the volume is independent of mass and time, it is said to possess zero dimension in mass $[M^\circ]$, zero dimension in time $[T^\circ]$ and three dimensions in length.

Similarly, force, as the product of mass and acceleration, can be expressed as

$$\text{Force} = \text{mass} \times \text{acceleration}$$
$$= \text{mass} \times (\text{length})/(\text{time})^2$$

The dimensions of force are $[M]\ [L]/[T]^2 = [M\ L\ T^{-2}]$. Thus, the force has one dimension in mass, one dimension in length, and –2 dimensions in time. The dimensions in all other base quantities are zero.

Note that in this type of representation, the magnitudes are not considered. It is the quality of the type of the physical quantity that enters. Thus, a change in velocity, initial velocity, average velocity, final velocity, and speed are all equivalent in this context. Since all these quantities can be expressed as length/time, their dimensions are [L]/[T] or $[L\ T^{-1}]$.

2.9 DIMENSIONAL FORMULAE AND DIMENSIONAL EQUATIONS

The expression which shows how and which of the base quantities represent the dimensions of a physical quantity is called the *dimensional formula* of the given physical quantity. For example, the dimensional formula of the volume is $[M^\circ\ L^3\ T^\circ]$, and that of speed or velocity is $[M^\circ\ L\ T^{-1}]$. Similarly, $[M^\circ\ L\ T^{-2}]$ is the dimensional formula of acceleration and $[M\ L^{-3}\ T^\circ]$ that of mass density.

An equation obtained by equating a physical quantity with its dimensional formula is called the **dimensional equation** of the physical

quantity. Thus, the dimensional equations are the equations, which represent the dimensions of a physical quantity in terms of the base quantities. For example, the dimensional equations of volume [V], speed [υ], force [F] and mass density [ρ] may be expressed as

$[V] = [M^0 L^3 T^0]$
$[\upsilon] = [M^0 L T^{-1}]$
$[F] = [M L T^{-2}]$
$[\rho] = [M L^{-3} T^0]$

The dimensional equation can be obtained from the equation representing the relations between the physical quantities. The dimensional formulae of a large number and wide variety of physical quantities, derived from the equations representing the relationships among other physical quantities and expressed in terms of base quantities are given in Appendix 9 for your guidance and ready reference.

2.10 DIMENSIONAL ANALYSIS AND ITS APPLICATIONS

The recognition of concepts of dimensions, which guide the description of physical behaviour is of basic importance as only those physical quantities can be added or subtracted which have the same dimensions. A thorough understanding of dimensional analysis helps us in deducing certain relations among different physical quantities and checking the derivation, accuracy and dimensional consistency or homogeneity of various mathematical expressions. When magnitudes of two or more physical quantities are multiplied, their units should be treated in the same manner as ordinary algebraic symbols. We can cancel identical units in the numerator and denominator. The same is true for dimensions of a physical quantity. Similarly, physical quantities represented by symbols on both sides of a mathematical equation must have the same dimensions.

2.10.1 Checking the Dimensional Consistency of Equations

The magnitudes of physical quantities may be added together or subtracted from one another only if they have the same dimensions. In other words, we can add or subtract similar physical quantities. Thus, velocity cannot be added to force, or an electric current cannot be subtracted from the thermodynamic temperature. This simple principle called **the principle of homogeneity of dimensions** in an equation is extremely useful in checking the correctness of an equation. If the dimensions of all the terms are not same, the equation is wrong. Hence, if we derive an expression for the length (or distance) of an object, regardless of the symbols appearing in the original mathematical relation, when all the individual dimensions are simplified, the remaining dimension must be that of length. Similarly, if we derive an equation of speed, the dimensions on both the sides of equation, when simplified, must be of length/time, or [$L T^{-1}$].

Dimensions are customarily used as a preliminary test of the consistency of an equation, when there is some doubt about the correctness of the equation. However, the dimensional consistency does not guarantee correct equations. It is uncertain to the extent of dimensionless quantities or functions. The arguments of special functions, such as the trigonometric, logarithmic and exponential functions must be dimensionless. A pure number, ratio of similar physical quantities, such as angle as the ratio (length/length), refractive index as the ratio (speed of light in vacuum/speed of light in medium) etc., has no dimensions.

Now we can test the dimensional consistency or homogeneity of the equation

$$x = x_0 + \upsilon_0 t + (1/2)\, a\, t^2$$

for the distance x travelled by a particle or body in time t which starts from the position x_0 with an initial velocity υ_0 at time $t = 0$ and has uniform acceleration a along the direction of motion.

The dimensions of each term may be written as

$$[x] = [L]$$
$$[x_0] = [L]$$
$$[\upsilon_0 t] = [L T^{-1}]\ [T] = [L]$$
$$[(1/2)\, a\, t^2] = [L T^{-2}]\ [T^2] = [L]$$

As each term on the right hand side of this equation has the same dimension, namely that of length, which is same as the dimension of left hand side of the equation, hence this equation is a dimensionally correct equation.

It may be noted that a test of consistency of dimensions tells us no more and no less than a

test of consistency of units, but has the advantage that we need not commit ourselves to a particular choice of units, and we need not worry about conversions among multiples and sub-multiples of the units. It may be borne in mind that **if an equation fails this consistency test, it is proved wrong, but if it passes, it is not proved right. Thus, a dimensionally correct equation need not be actually an exact (correct) equation, but a dimensionally wrong (incorrect) or inconsistent equation must be wrong**.

Example 2.15 Let us consider an equation

$$\frac{1}{2} m v^2 = m g h$$

where m is the mass of the body, v its velocity, g is the acceleration due to gravity and h is the height. Check whether this equation is dimensionally correct.

Answer The dimensions of LHS are

$$[M]\ [L T^{-1}]^2 = [M]\ [L^2 T^{-2}]$$
$$= [M L^2 T^{-2}]$$

The dimensions of RHS are

$$[M][L T^{-2}]\ [L] = [M][L^2 T^{-2}]$$
$$= [M L^2 T^{-2}]$$

The dimensions of LHS and RHS are the same and hence the equation is dimensionally correct.

Example 2.16 The SI unit of energy is J = kg m^2 s^{-2}; that of speed v is m s^{-1} and of acceleration a is m s^{-2}. Which of the formulae for kinetic energy (K) given below can you rule out on the basis of dimensional arguments (m stands for the mass of the body) :

(a) $K = m^2 v^3$
(b) $K = (1/2)mv^2$
(c) $K = ma$
(d) $K = (3/16)mv^2$
(e) $K = (1/2)mv^2 + ma$

Answer Every correct formula or equation must have the same dimensions on both sides of the equation. Also, only quantities with the same physical dimensions can be added or subtracted. The dimensions of the quantity on the right side are $[M^2 L^3 T^{-3}]$ for (a); $[M L^2 T^{-2}]$ for (b) and (d); $[M L^{-1} T^{-2}]$ for (c). The quantity on the right side of (e) has no proper dimensions since two quantities of different dimensions have been added. Since the kinetic energy K has the dimensions of $[M L^2 T^{-2}]$, formulas (a), (c) and (e) are ruled out. Note that dimensional arguments cannot tell which of the two, (b) or (d), is the correct formula. For this, one must turn to the actual definition of kinetic energy (see Chapter 6). The correct formula for kinetic energy is given by (b).

2.10.2 Deducing Relation among the Physical Quantities

The method of dimensions can sometimes be used to deduce relation among the physical quantities. For this we should know the dependence of the physical quantity on other quantities (upto three physical quantities or linearly independent variables) and consider it as a product type of the dependence. Let us take an example.

Example 2.17 Consider a simple pendulum, having a bob attached to a string, that oscillates under the action of the force of gravity. Suppose that the period of oscillation of the simple pendulum depends on its length (l), mass of the bob (m) and acceleration due to gravity (g). Derive the expression for its time period using method of dimensions.

Answer The dependence of time period T on the quantities l, g and m as a product may be written as :

$$T = k\, l^x g^y m^z$$

where k is dimensionless constant and x, y and z are the exponents.

By considering dimensions on both sides, we have

$$[L^o M^o T^1] = [L^1]\,[L^1 T^{-2}]^y\,[M^1]^z$$
$$= L^{x+y} T^{-2y} M^z$$

On equating the dimensions on both sides, we have

$$x + y = 0;\ -2y = 1;\ \text{and } z = 0$$

So that $x = \frac{1}{2}, y = -\frac{1}{2}, z = 0$

Then, $T = k\, l^{1/2} g^{-1/2}$

or, $T = k\sqrt{\frac{l}{g}}$

Note that value of constant k can not be obtained by the method of dimensions. Here it does not matter if some number multiplies the right side of this formula, because that does not affect its dimensions.

Actually, $k = 2\pi$ so that $T = 2\pi\sqrt{\frac{l}{g}}$ ◀

Dimensional analysis is very useful in deducing relations among the interdependent physical quantities. However, dimensionless constants cannot be obtained by this method. The method of dimensions can only test the dimensional validity, but not the exact relationship between physical quantities in any equation. It does not distinguish between the physical quantities having same dimensions.

A number of exercises at the end of this chapter will help you develop skill in dimensional analysis.

SUMMARY

1. Physics is a quantitative science, based on measurement of physical quantities. Certain physical quantities have been chosen as fundamental or base quantities (such as length, mass, time, electric current, thermodynamic temperature, amount of substance, and luminous intensity).
2. Each base quantity is defined in terms of a certain basic, arbitrarily chosen but properly standardised reference standard called unit (such as metre, kilogram, second, ampere, kelvin, mole and candela). The units for the fundamental or base quantities are called fundamental or base units.
3. Other physical quantities, derived from the base quantities, can be expressed as a combination of the base units and are called derived units. A complete set of units, both fundamental and derived, is called a system of units.
4. The International System of Units (SI) based on seven base units is at present internationally accepted unit system and is widely used throughout the world.
5. The SI units are used in all physical measurements, for both the base quantities and the derived quantities obtained from them. Certain derived units are expressed by means of SI units with special names (such as joule, newton, watt, etc).
6. The SI units have well defined and internationally accepted unit symbols (such as m for metre, kg for kilogram, s for second, A for ampere, N for newton etc.).
7. Physical measurements are usually expressed for small and large quantities in scientific notation, with powers of 10. Scientific notation and the prefixes are used to simplify measurement notation and numerical computation, giving indication to the precision of the numbers.
8. Certain general rules and guidelines must be followed for using notations for physical quantities and standard symbols for SI units, some other units and SI prefixes for expressing properly the physical quantities and measurements.
9. In computing any physical quantity, the units for derived quantities involved in the relationship(s) are treated as though they were algebraic quantities till the desired units are obtained.
10. Direct and indirect methods can be used for the measurement of physical quantities. In measured quantities, while expressing the result, the accuracy and precision of measuring instruments along with errors in measurements should be taken into account.
11. In measured and computed quantities proper significant figures only should be retained. Rules for determining the number of significant figures, carrying out arithmetic operations with them, and 'rounding off ' the uncertain digits must be followed.
12. The dimensions of base quantities and combination of these dimensions describe the nature of physical quantities. Dimensional analysis can be used to check the dimensional consistency of equations, deducing relations among the physical quantities, etc. A dimensionally consistent equation need not be actually an exact (correct) equation, but a dimensionally wrong or inconsistent equation must be wrong.

EXERCISES

Note : In stating numerical answers, take care of significant figures.

2.1 Fill in the blanks
(a) The volume of a cube of side 1 cm is equal to .....m^3
(b) The surface area of a solid cylinder of radius 2.0 cm and height 10.0 cm is equal to ...$(mm)^2$
(c) A vehicle moving with a speed of 18 km h^{-1} covers....m in 1 s
(d) The relative density of lead is 11.3. Its density is ....g cm^{-3} or ....kg m^{-3}.

2.2 Fill in the blanks by suitable conversion of units
(a) 1 kg m^2 s^{-2} = ...g cm^2 s^{-2}
(b) 1 m = ... ly
(c) 3.0 m s^{-2} = ... km h^{-2}
(d) $G = 6.67 \times 10^{-11}$ N m^2 $(kg)^{-2}$ = ... $(cm)^3$ s^{-2} g^{-1}.

2.3 A calorie is a unit of heat or energy and it equals about 4.2 J where 1J = 1 kg m^2 s^{-2}. Suppose we employ a system of units in which the unit of mass equals α kg, the unit of length equals β m, the unit of time is γ s. Show that a calorie has a magnitude 4.2 $\alpha^{-1}\beta^{-2}\gamma^{2}$ in terms of the new units.

2.4 Explain this statement clearly :
"To call a dimensional quantity 'large' or 'small' is meaningless without specifying a standard for comparison". In view of this, reframe the following statements wherever necessary :
(a) atoms are very small objects
(b) a jet plane moves with great speed
(c) the mass of Jupiter is very large
(d) the air inside this room contains a large number of molecules
(e) a proton is much more massive than an electron
(f) the speed of sound is much smaller than the speed of light.

2.5 A new unit of length is chosen such that the speed of light in vacuum is unity. What is the distance between the Sun and the Earth in terms of the new unit if light takes 8 min and 20 s to cover this distance ?

2.6 Which of the following is the most precise device for measuring length :
(a) a vernier callipers with 20 divisions on the sliding scale
(b) a screw gauge of pitch 1 mm and 100 divisions on the circular scale
(c) an optical instrument that can measure length to within a wavelength of light ?

2.7 A student measures the thickness of a human hair by looking at it through a microscope of magnification 100. He makes 20 observations and finds that the average width of the hair in the field of view of the microscope is 3.5 mm. What is the estimate on the thickness of hair ?

2.8 Answer the following :
(a) You are given a thread and a metre scale. How will you estimate the diameter of the thread ?
(b) A screw gauge has a pitch of 1.0 mm and 200 divisions on the circular scale. Do you think it is possible to increase the accuracy of the screw gauge arbitrarily by increasing the number of divisions on the circular scale ?
(c) The mean diameter of a thin brass rod is to be measured by vernier callipers. Why is a set of 100 measurements of the diameter expected to yield a more reliable estimate than a set of 5 measurements only ?

2.9 The photograph of a house occupies an area of 1.75 cm^2 on a 35 mm slide. The slide is projected on to a screen, and the area of the house on the screen is 1.55 m^2. What is the linear magnification of the projector-screen arrangement.

2.10 State the number of significant figures in the following :
(a) 0.007 m^2
(b) 2.64×10^{24} kg
(c) 0.2370 g cm^{-3}

(d) 6.320 J
(e) 6.032 N m^{-2}
(f) 0.0006032 m^2

2.11 The length, breadth and thickness of a rectangular sheet of metal are 4.234 m, 1.005 m, and 2.01 cm respectively. Give the area and volume of the sheet to correct significant figures.

2.12 The mass of a box measured by a grocer's balance is 2.300 kg. Two gold pieces of masses 20.15 g and 20.17 g are added to the box. What is (a) the total mass of the box, (b) the difference in the masses of the pieces to correct significant figures ?

2.13 A physical quantity P is related to four observables a, b, c and d as follows :

$$P = a^3 b^2 / \left(\sqrt{c}\, d\right)$$

The percentage errors of measurement in a, b, c and d are 1%, 3%, 4% and 2%, respectively. What is the percentage error in the quantity P? If the value of P calculated using the above relation turns out to be 3.763, to what value should you round off the result ?

2.14 A book with many printing errors contains four different formulas for the displacement y of a particle undergoing a certain periodic motion :
(a) $y = a \sin 2\pi\, t/T$
(b) $y = a \sin vt$
(c) $y = (a/T) \sin t/a$
(d) $y = (a\sqrt{2})\ (\sin 2\pi t / T + \cos 2\pi t / T)$
(a = maximum displacement of the particle, v = speed of the particle. T = time-period of motion). Rule out the wrong formulas on dimensional grounds.

2.15 A famous relation in physics relates 'moving mass' m to the 'rest mass' m_o of a particle in terms of its speed v and the speed of light, c. (This relation first arose as a consequence of special relativity due to Albert Einstein). A boy recalls the relation almost correctly but forgets where to put the constant c. He writes :

$$m = \frac{m_0}{\left(1 - v^2\right)^{1/2}}.$$

Guess where to put the missing c.

2.16 The unit of length convenient on the atomic scale is known as an angstrom and is denoted by **Å**: 1 **Å** = 10^{-10} m. The size of a hydrogen atom is about 0.5 **Å**. What is the total atomic volume in m^3 of a mole of hydrogen atoms ?

2.17 One mole of an ideal gas at standard temperature and pressure occupies 22.4 L (molar volume). What is the ratio of molar volume to the atomic volume of a mole of hydrogen ? (Take the size of hydrogen molecule to be about 1 **Å**). Why is this ratio so large ?

2.18 Explain this common observation clearly : If you look out of the window of a fast moving train, the nearby trees, houses etc. seem to move rapidly in a direction opposite to the train's motion, but the distant objects (hill tops, the Moon, the stars etc.) seem to be stationary. (In fact, since you are aware that you are moving, these distant objects seem to move with you).

2.19 The principle of 'parallax' in section 2.3.1 is used in the determination of distances of very distant stars. The baseline AB is the line joining the Earth's two locations six months apart in its orbit around the Sun. That is, the baseline is about the diameter of the Earth's orbit $\approx 3 \times 10^{11}$m. However, even the nearest stars are so distant that with such a long baseline, they show parallax only of the order of 1" (second) of arc or so. A *parsec* is a convenient unit of length on the astronomical scale. It is the distance of an object that will show a parallax of 1" (second) of arc from opposite ends of a baseline equal to the distance from the Earth to the Sun. How much is a parsec in terms of metres ?

2.20 The nearest star to our solar system is 4.29 light years away. How much is this distance in terms of parsecs? How much parallax would this star (named Alpha Centauri) show when viewed from two locations of the Earth six months apart in its orbit around the Sun ?

2.21 Precise measurements of physical quantities are a *need* of science. For example, to ascertain the speed of an aircraft, one must have an accurate method to find its positions at closely separated instants of time. This was the actual motivation behind the discovery of radar in World War II. Think of different examples in modern science where precise measurements of length, time, mass etc. are needed. Also, wherever you can, give a quantitative idea of the precision needed.

2.22 Just as precise measurements are necessary in science, it is equally important to be able to make rough estimates of quantities using rudimentary ideas and common observations. Think of ways by which you can estimate the following (where an estimate is difficult to obtain, try to get an upper bound on the quantity) :

(a) the total mass of rain-bearing clouds over India during the Monsoon

(b) the mass of an elephant

(c) the wind speed during a storm

(d) the number of strands of hair on your head

(e) the number of air molecules in your classroom.

2.23 The Sun is a hot plasma (ionized matter) with its inner core at a temperature exceeding 10^7 K, and its outer surface at a temperature of about 6000 K. At these high temperatures, no substance remains in a solid or liquid phase. In what range do you expect the mass density of the Sun to be, in the range of densities of solids and liquids or gases ? Check if your guess is correct from the following data : mass of the Sun = 2.0×10^{30} kg, radius of the Sun = 7.0×10^8 m.

2.24 When the planet Jupiter is at a distance of 824.7 million kilometers from the Earth, its angular diameter is measured to be 35.72" of arc. Calculate the diameter of Jupiter.

Additional Exercises

2.25 A man walking briskly in rain with speed v must slant his umbrella forward making an angle θ with the vertical. A student derives the following relation between θ and v : $\tan\theta = v$ and checks that the relation has a correct limit: as $v \to 0$, $\theta \to 0$, as expected. (We are assuming there is no strong wind and that the rain falls vertically for a stationary man). Do you think this relation can be correct ? If not, guess the correct relation.

2.26 It is claimed that two cesium clocks, if allowed to run for 100 years, free from any disturbance, may differ by only about 0.02 s. What does this imply for the accuracy of the standard cesium clock in measuring a time-interval of 1 s ?

2.27 Estimate the average mass density of a sodium atom assuming its size to be about 2.5 **Å**. (Use the known values of Avogadro's number and the atomic mass of sodium). Compare it with the density of sodium in its crystalline phase : 970 kg m^{-3}. Are the two densities of the same order of magnitude ? If so, why ?

2.28 The unit of length convenient on the nuclear scale is a fermi : 1 f = 10^{-15} m. Nuclear sizes obey roughly the following empirical relation :

$$r = r_0 A^{1/3}$$

where r is the radius of the nucleus, A its mass number, and r_o is a constant equal to about, 1.2 f. Show that the rule implies that nuclear mass density is nearly constant for different nuclei. Estimate the mass density of sodium nucleus. Compare it with the average mass density of a sodium atom obtained in Exercise. 2.27.

2.29 A LASER is a source of very intense, monochromatic, and unidirectional beam of light. These properties of a laser light can be exploited to measure long distances. The distance of the Moon from the Earth has been already determined very precisely using a laser as a source of light. A laser light beamed at the Moon takes 2.56 s to

return after reflection at the Moon's surface. How much is the radius of the lunar orbit around the Earth ?

2.30 A SONAR (sound navigation and ranging) uses ultrasonic waves to detect and locate objects under water. In a submarine equipped with a SONAR the time delay between generation of a probe wave and the reception of its echo after reflection from an enemy submarine is found to be 77.0 s. What is the distance of the enemy submarine? (Speed of sound in water = 1450 m s^{-1}).

2.31 The farthest objects in our Universe discovered by modern astronomers are so distant that light emitted by them takes billions of years to reach the Earth. These objects (known as quasars) have many puzzling features, which have not yet been satisfactorily explained. What is the distance in km of a quasar from which light takes 3.0 billion years to reach us ?

2.32 It is a well known fact that during a total solar eclipse the disk of the moon almost completely covers the disk of the Sun. From this fact and from the information you can gather from examples 2.3 and 2.4, determine the approximate diameter of the moon.

2.33 A great physicist of this century (P.A.M. Dirac) loved playing with numerical values of Fundamental constants of nature. This led him to an interesting observation. Dirac found that from the basic constants of atomic physics (c, e, mass of electron, mass of proton) and the gravitational constant G, he could arrive at a number with the dimension of time. Further, it was a very large number, its magnitude being close to the present estimate on the age of the universe (~15 billion years). From the table of fundamental constants in this book, try to see if you too can construct this number (or any other interesting number you can think of). If its coincidence with the age of the universe were significant, what would this imply for the constancy of fundamental constants ?

CHAPTER THREE

MOTION IN A STRAIGHT LINE

3.1 Introduction
3.2 Position, path length and displacement
3.3 Average velocity and average speed
3.4 Instantaneous velocity and speed
3.5 Acceleration
3.6 Kinematic equations for uniformly accelerated motion
3.7 Relative velocity

Summary
Points to ponder
Exercises
Additional exercises
Appendix 3.1

3.1 INTRODUCTION

Motion is common to everything in the universe. We walk, run and ride a bicycle. Even when we are sleeping, air moves into and out of our lungs and blood flows in arteries and veins. We see leaves falling from trees and water flowing down a dam. Automobiles and planes carry people from one place to the other. The earth rotates once every twenty-four hours and revolves round the sun once in a year. The sun itself is in motion in the Milky Way, which is again moving within its local group of galaxies.

Motion is change in position of an object with time. How does the position change with time ? In this chapter, we shall learn how to describe motion. For this, we develop the concepts of velocity and acceleration. We shall confine ourselves to the study of motion of objects along a straight line, also known as **rectilinear motion**. For the case of rectilinear motion with uniform acceleration, a set of simple equations can be obtained. Finally, to understand the relative nature of motion, we introduce the concept of relative velocity.

In our discussions, we shall treat the objects in motion as point objects. This approximation is valid so far as the size of the object is much smaller than the distance it moves in a reasonable duration of time. In a good number of situations in real-life, the size of objects can be neglected and they can be considered as point-like objects without much error.

In **Kinematics**, we study ways to describe motion without going into the causes of motion. What causes motion described in this chapter and the next chapter forms the subject matter of Chapter 5.

3.2 POSITION, PATH LENGTH AND DISPLACEMENT

Earlier you learnt that motion is change in position of an object with time. In order to specify position, we need to use a reference point and a set of axes. It is convenient to choose

a rectangular coordinate system consisting of three mutually perpenducular axes, labelled X-, Y-, and Z- axes. The point of intersection of these three axes is called origin (O) and serves as the **reference point**. The coordinates ($x, y. z$) of an object describe the position of the object with respect to this coordinate system. To measure time, we position a clock in this system. This coordinate system along with a clock constitutes a **frame of reference**.

If one or more coordinates of an object change with time, we say that the object is in motion. Otherwise, the object is said to be at rest with respect to this frame of reference.

The choice of a set of axes in a frame of reference depends upon the situation. For example, for describing motion in one dimension, we need only one axis. To describe motion in two/three dimensions, we need a set of two/three axes.

Description of an event depends on the frame of reference chosen for the description. For example, when you say that a car is moving on a road, you are describing the car with respect to a frame of reference attached to you or to the ground. But with respect to a frame of reference attached with a person sitting in the car, the car is at rest.

To describe motion along a straight line, we can choose an axis, say X-axis, so that it coincides with the path of the object. We then measure the position of the object with reference to a conveniently chosen origin, say O, as shown in Fig. 3.1. Positions to the right of O are taken as positive and to the left of O, as negative. Following this convention, the position coordinates of point P and Q in Fig. 3.1 are +360 m and +240 m. Similarly, the position coordinate of point R is –120 m.

Path length

Consider the motion of a car along a straight line. We choose the x-axis such that it coincides with the path of the car's motion and origin of the axis as the point from where the car started moving, i.e. the car was at $x = 0$ at $t = 0$ (Fig. 3.1). Let P, Q and R represent the positions of the car at different instants of time. Consider two cases of motion. In the first case, the car moves from O to P. Then the distance moved by the car is OP = +360 m. **This distance is called the path length** traversed by the car**.** In the second case, the car moves from O to P and then moves back from P to Q. During this course of motion, the path length traversed is OP + PQ = + 360 m + (+120 m) = + 480 m. Path length is a scalar quantity — a quantity that has a magnitude only and no direction (see Chapter 4).

Displacement

It is useful to define another quantity displacement as the change in position. Let x_1 and x_2 be the positions of an object at time t_1 and t_2. Then its displacement, denoted by Δx, in time $\Delta t = (t_2 - t_1)$, is given by the difference between the final and initial positions :

$$\Delta x = x_2 - x_1$$

(We use the Greek letter delta (Δ) to denote a change in a quantity.)

If $x_2 > x_1$, Δx is positive; and if $x_2 < x_1$, Δx is negative.

Displacement has both magnitude and direction. Such quantities are represented by vectors. You will read about vectors in the next chapter. Presently, we are dealing with motion along a straight line (also called **rectilinear motion**) only. In one-dimensional motion, there are *only two* directions (backward and forward, upward and downward) in which an object can move, and these two directions can easily be specified by + and – signs. For example, displacement of the car in moving from O to P is :

$$\Delta x = x_2 - x_1 = (+360 \text{ m}) - 0 \text{ m} = +360 \text{ m}$$

The displacement has a magnitude of 360 m and is directed in the positive x direction as indicated by the + sign. Similarly, the displacement of the car from P to Q is 240 m – 360 m = – 120 m. The

Fig. 3.1 x-axis, origin and positions of a car at different times.

negative sign indicates the direction of displacement. Thus, it is not necessary to use vector notation for discussing motion of objects in one-dimension.

The magnitude of displacement may or may not be equal to the path length traversed by an object. For example, for motion of the car from O to P, the path length is +360 m and the displacement is +360 m. In this case, the magnitude of displacement (360 m) is equal to the path length (360 m). But consider the motion of the car from O to P and back to Q. In this case, the path length = (+360 m) + (+120 m) = + 480 m. However, the displacement = (+240 m) – (0 m) = + 240 m. Thus, the magnitude of displacement (240 m) is not equal to the path length (480 m).

The magnitude of the displacement for a course of motion may be zero but the corresponding path length is **not zero**. For example, if the car starts from O, goes to P and then returns to O, the final position coincides with the initial position and the displacement is zero. However, the path length of this journey is OP + PO = 360 m + 360 m = 720 m.

Motion of an object can be represented by a position-time graph as you have already learnt about it. Such a graph is a powerful tool to represent and analyse different aspects of motion of an object. For motion along a straight line, say *X*-axis, only *x*-coordinate varies with time and we have an *x*-*t* graph. Let us first consider the simple case in which an object is stationary, e.g. a car standing still at $x = 40$ m. The position-time graph is a straight line parallel to the time axis, as shown in Fig. 3.2(a).

If an object moving along the straight line covers equal distances in equal intervals of time, it is said to be in **uniform motion** along a straight line. Fig. 3.2(b) shows the position-time graph of such a motion.

Fig. 3.2 Position-time graph of (a) stationary object, and (b) an object in uniform motion.

Fig. 3.3 Position-time graph of a car.

Now, let us consider the motion of a car that starts from rest at time $t = 0$ s from the origin O and picks up speed till $t = 10$ s and thereafter moves with uniform speed till $t = 18$ s. Then the brakes are applied and the car stops at $t = 20$ s and $x = 296$ m. The position-time graph for this case is shown in Fig. 3.3. We shall refer to this graph in our discussion in the following sections.

3.3 AVERAGE VELOCITY AND AVERAGE SPEED

When an object is in motion, its position changes with time. But how fast is the position changing with time and in what direction? To describe this, we define the quantity **average velocity**. Average velocity is defined as the change in position or displacement (Δx) divided by the time intervals (Δt), in which the displacement occurs :

$$\bar{v} = \frac{x_2 - x_1}{t_2 - t_1} = \frac{\Delta x}{\Delta t} \qquad (3.1)$$

where x_2 and x_1 are the positions of the object at time t_2 and t_1, respectively. Here the bar over the symbol for velocity is a standard notation used to indicate an average quantity. The SI unit for velocity is m/s or m s^{-1}, although km h^{-1} is used in many everyday applications.

Like displacement, average velocity is also a vector quantity. But as explained earlier, for motion in a straight line, the directional aspect of the vector can be taken care of by + and – signs and we do not have to use the vector notation for velocity in this chapter.

Fig. 3.4 The average velocity is the slope of line P_1P_2.

Consider the motion of the car in Fig. 3.3. The portion of the x-t graph between $t = 0$ s and $t = 8$ s is blown up and shown in Fig. 3.4. As seen from the plot, the average velocity of the car between time $t = 5$ s and $t = 7$ s is :

$$\bar{v} = \frac{x_2 - x_1}{t_2 - t_1} = \frac{(27.4 - 10.0)\,\text{m}}{(7-5)\,\text{s}} = 8.7\ \text{m s}^{-1}$$

Geometrically, this is the slope of the straight line P_1P_2 connecting the initial position P_1 to the final position P_2 as shown in Fig. 3.4.

The average velocity can be positive or negative depending upon the sign of the displacement. It is zero if the displacement is zero. Fig. 3.5 shows the x-t graphs for an object, moving with positive velocity (Fig. 3.5a), moving with negative velocity (Fig. 3.5b) and at rest (Fig. 3.5c).

Fig. 3.5 Position-time graph for an object (a) moving with positive velocity, (b) moving with negative velocity, and (c) at rest.

Average velocity as defined above involves only the displacement of the object. We have seen earlier that the magnitude of displacement may be different from the actual path length. To describe the rate of motion over the actual path, we introduce another quantity called **average speed**.

Average speed is defined as the total path length travelled divided by the total time interval during which the motion has taken place :

$$\textit{Average speed} = \frac{\textit{Total path length}}{\textit{Total time interval}} \qquad (3.2)$$

Average speed has obviously the same unit (m s^{-1}) as that of velocity. But it does not tell us in what direction an object is moving. Thus, it is always positive (in contrast to the average velocity which can be positive or negative). If the motion of an object is along a straight line and in the **same direction**, the magnitude of displacement is equal to the total path length. In that case, the magnitude of average velocity

is equal to the average speed. This is not always the case, as you will see in the following example.

▶ ***Example 3.1*** A car is moving along a straight line, say OP in Fig. 3.1. It moves from O to P in 18 s and returns from P to Q in 6.0 s. What are the average velocity and average speed of the car in going (a) from O to P ? and (b) from O to P and back to Q ?

Answer (a)

$$\text{Average velocity} = \frac{\text{Displacement}}{\text{Time interval}}$$

$$\bar{v} = \frac{+360 \text{ m}}{18 \text{ s}} = +20 \text{ m s}^{-1}$$

$$\text{Average speed} = \frac{\text{Path length}}{\text{Time interval}}$$

$$= \frac{360 \text{ m}}{18 \text{ s}} = 20 \text{ m s}^{-1}$$

Thus, in this case the average speed is equal to the magnitude of the average velocity.

(b) In this case,

$$\text{Average velocity} = \frac{\text{Displacement}}{\text{Time interval}} = \frac{+240 \text{ m}}{(18+6.0) \text{ s}}$$

$$= +10 \text{ m s}^{-1}$$

$$\text{Average speed} = \frac{\text{Path length}}{\text{Time interval}} = \frac{\text{OP} + \text{PQ}}{\Delta t}$$

$$= \frac{(360+120) \text{ m}}{24 \text{ s}} = 20 \text{ m s}^{-1}$$

Thus, in this case the average speed is *not* equal to the magnitude of the average velocity. This happens because the motion here involves change in direction so that the path length is greater than the magnitude of displacement. This shows that **speed is, in general, greater than the magnitude of the velocity.** ◀

If the car in Example 3.1 moves from O to P and comes back to O in the same time interval, average speed is 20 m/s but the average velocity is zero !

3.4 INSTANTANEOUS VELOCITY AND SPEED

The average velocity tells us how fast an object has been moving over a given time interval but does not tell us how fast it moves at different instants of time during that interval. For this, we define **instantaneous velocity** or simply velocity v at an instant t.

The velocity at an instant is defined as the limit of the average velocity as the time interval Δt becomes infinitesimally small. In other words,

$$v = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} \tag{3.3a}$$

$$= \frac{dx}{dt} \tag{3.3b}$$

where the symbol $\lim_{\Delta t \to 0}$ stands for the operation of taking limit as $\Delta t \to 0$ of the quantity on its right. In the language of calculus, the quantity on the right hand side of Eq. (3.3a) is the differential coefficient of x with respect to t and is denoted by $\frac{dx}{dt}$ (see Appendix 3.1). It is the rate of change of position with respect to time, at that instant.

We can use Eq. (3.3a) for obtaining the value of velocity at an instant either **graphically** or **numerically**. Suppose that we want to obtain graphically the value of velocity at time $t = 4$ s (point P) for the motion of the car represented in Fig. 3.3. The figure has been redrawn in Fig. 3.6 choosing different scales to facilitate the

Fig. 3.6 Determining velocity from position-time graph. Velocity at t = 4 s is the slope of the tangent to the graph at that instant.

calculation. Let us take Δt = 2 s centred at t = 4 s. Then, by the definition of the average velocity, the slope of line P_1P_2 (Fig. 3.6) gives the value of average velocity over the interval 3 s to 5 s. Now, we decrease the value of Δt from 2 s to 1 s. Then line P_1P_2 becomes Q_1Q_2 and its slope gives the value of the average velocity over the interval 3.5 s to 4.5 s. In the limit $\Delta t \to 0$, the line P_1P_2 becomes tangent to the position-time curve at the point P and the velocity at t = 4 s is given by the slope of the tangent at that point. **It is difficult to show this process graphically. But if we use numerical method to obtain the value of the velocity, the meaning of the limiting process becomes clear**. For the graph shown in Fig. 3.6, $x = 0.08\ t^3$. Table 3.1 gives the value of $\Delta x/\Delta t$ calculated for Δt equal to 2.0 s, 1.0 s, 0.5 s, 0.1 s and 0.01 s centred at t = 4.0 s. The second and third columns give the value of $t_1 = \left(t - \frac{\Delta t}{2}\right)$ and $t_2 = \left(t + \frac{\Delta t}{2}\right)$ and the fourth and the fifth columns give the corresponding values of x, i.e. $x(t_1) = 0.08\ t_1^3$ and $x(t_2) = 0.08\ t_2^3$. The sixth column lists the difference $\Delta x = x(t_2) - x(t_1)$ and the last column gives the ratio of Δx and Δt, i.e. the average velocity corresponding to the value of Δt listed in the first column.

We see from Table 3.1 that as we decrease the value of Δt from 2.0 s to 0.010 s, the value of the average velocity approaches the limiting value 3.84 m s^{-1} which is the value of velocity at t = 4.0 s, i.e. the value of $\frac{dx}{dt}$ at t = 4.0 s. In this manner, we can calculate velocity at each instant for motion of the car shown in Fig. 3.3. For this case, the variation of velocity with time is found to be as shown in Fig. 3.7.

Fig. 3.7 Velocity–time graph corresponding to motion shown in Fig. 3.3.

The graphical method for the determination of the instantaneous velocity is always not a convenient method. For this, we must carefully plot the position–time graph and calculate the value of average velocity as Δt becomes smaller and smaller. It is easier to calculate the value of velocity at different instants if we have data of positions at different instants or exact expression for the position as a function of time. Then, we calculate $\Delta x/\Delta t$ from the data for decreasing the value of Δt and find the limiting value as we have done in Table 3.1 or use differential calculus for the given expression and calculate $\frac{dx}{dt}$ at different instants as done in the following example.

Table 3.1 Limiting value of $\frac{\Delta x}{\Delta t}$ at t = 4 s

Δt (s)	t_1 (s)	t_2 (s)	$x(t_1)$ (m)	$x(t_2)$ (m)	Δx (m)	$\Delta x / \Delta t$ (m s^{-1})
2.0	3.0	5.0	2.16	10.0	7.84	3.92
1.0	3.5	4.5	3.43	7.29	3.86	3.86
0.5	3.75	4.25	4.21875	6.14125	1.9225	3.845
0.1	3.95	4.05	4.93039	5.31441	0.38402	3.8402
0.01	3.995	4.005	5.100824	5.139224	0.0384	3.8400

Example 3.2 *The position of an object moving along x-axis is given by* $x = a + bt^2$ *where* a = 8.5 m, b = 2.5 m s^{-2} *and* t *is measured in seconds. What is its velocity at* t = 0 s *and* t = 2.0 s. *What is the average velocity between* t = 2.0 s *and* t = 4.0 s ?

Answer In notation of differential calculus, the velocity is

$$v = \frac{dx}{dt} = \frac{d}{dt}\left(a + bt^2\right) = 2b\,t = 5.0\ t \text{ m s}^{-1}$$

At t = 0 s, v = 0 m s^{-1} and at t = 2.0 s, v = 10 m s^{-1} .

$$\text{Average velocity} = \frac{x(4.0) - x(2.0)}{4.0 - 2.0}$$

$$= \frac{a + 16b - a - 4b}{2.0} = 6.0 \times b$$

$$= 6.0 \times 2.5 = 15 \text{ m s}^{-1}$$

From Fig. 3.7, we note that during the period t =10 s to 18 s the velocity is constant. Between period t =18 s to t = 20 s, it is uniformly decreasing and during the period t = 0 s to t = 10 s, it is increasing. **Note that for uniform motion, velocity is the same as the average velocity at all instants**.

Instantaneous speed or simply speed is the magnitude of velocity. For example, a velocity of + 24.0 m s^{-1} and a velocity of – 24.0 m s^{-1} — both have an associated speed of 24.0 m s^{-1}. It should be noted that though average speed over a finite interval of time is greater or equal to the magnitude of the average velocity, instantaneous speed at an instant is equal to the magnitude of the instantaneous velocity at that instant. Why so ?

3.5 ACCELERATION

The velocity of an object, in general, changes during its course of motion. How to describe this change? Should it be described as the rate of change in velocity **with distance** or **with time** ? This was a problem even in Galileo's time. It was first thought that this change could be described by the rate of change of velocity with distance. But, through his studies of motion of freely falling objects and motion of objects on an inclined plane, Galileo concluded that the rate of change of velocity with time is a constant of motion for all objects in free fall. On the other hand, the change in velocity with distance is not constant – it decreases with the increasing distance of fall. This led to the concept of acceleration as the rate of change of velocity with time.

The average acceleration $\bar{a}$ over a time interval is defined as the change of velocity divided by the time interval :

$$\bar{a} = \frac{v_2 - v_1}{t_2 - t_1} = \frac{\Delta v}{\Delta t} \qquad (3.4)$$

where v_2 and v_1 are the instantaneous velocities or simply velocities at time t_2 and t_1 . It is the average change of velocity per unit time. The SI unit of acceleration is m s^{-2} .

On a plot of velocity versus time, the average acceleration is the slope of the straight line connecting the points corresponding to (v_2, t_2) and (v_1, t_1). The average acceleration for velocity-time graph shown in Fig. 3.7 for different time intervals 0 s - 10 s, 10 s – 18 s, and 18 s – 20 s are :

0 s - 10 s $\quad \bar{a} = \dfrac{(24 - 0)\text{ m s}^{-1}}{(10 - 0)\text{ s}} = 2.4 \text{ m s}^{-2}$

10 s - 18 s $\quad \bar{a} = \dfrac{(24 - 24)\text{ m s}^{-1}}{(18 - 10)\text{ s}} = 0 \text{ m s}^{-2}$

18 s - 20 s $\quad \bar{a} = \dfrac{(0 - 24)\text{ m s}^{-1}}{(20 - 18)\text{ s}} = -12 \text{ m s}^{-2}$

Fig. 3.8 Acceleration as a function of time for motion represented in Fig. 3.3.

Instantaneous acceleration is defined in the same way as the instantaneous velocity :

$$a = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} = \frac{dv}{dt} \qquad (3.5)$$

The acceleration at an instant is the slope of the tangent to the v–t curve at that instant. For the v–t curve shown in Fig. 3.7, we can obtain acceleration at every instant of time. The resulting a – t curve is shown in Fig. 3.8. We see

that the acceleration is nonuniform over the period 0 s to 10 s. It is zero between 10 s and 18 s and is constant with value –12 m s^{-2} between 18 s and 20 s. When the acceleration is uniform, obviously, it equals the average acceleration over that period.

Since velocity is a quantity having both magnitude and direction, a change in velocity may involve either or both of these factors. Acceleration, therefore, may result from a change in speed (magnitude), a change in direction or changes in both. Like velocity, acceleration can also be positive, negative or zero. Position-time graphs for motion with positive, negative and zero acceleration are shown in Figs. 3.9 (a), (b) and (c), respectively. Note that the graph curves upward for positive acceleration; downward for negative acceleration and it is a straight line for zero acceleration. As an exercise, identify in Fig. 3.3, the regions of the curve that correspond to these three cases.

Although acceleration can vary with time, our study in this chapter will be restricted to motion with constant acceleration. In this case, the average acceleration equals the constant value of acceleration during the interval. If the velocity of an object is v_o at $t = 0$ and v at time t, we have

$$\bar{a} = \frac{v - v_0}{t - 0} \quad \text{or,} \quad v = v_0 + a\,t \qquad (3.6)$$

Fig. 3.9 Position-time graph for motion with (a) positive acceleration; (b) negative acceleration, and (c) zero acceleration.

Let us see how velocity-time graph looks like for some simple cases. Fig. 3.10 shows velocity-time graph for motion with constant acceleration for the following cases :

(a) An object is moving in a positive direction with a positive acceleration, for example the motion of the car in Fig. 3.3 between $t = 0$ s and $t = 10$ s.

(b) An object is moving in positive direction with a negative acceleration, for example, motion of the car in Fig 3.3 between $t = 18$ s and 20 s.

(c) An object is moving in negative direction with a negative acceleration, for example the motion of a car moving from O in Fig. 3.1 in negative x-direction with increasing speed.

(d) An object is moving in positive direction till time t_1, and then turns back with the same negative acceleration, for example the motion of a car from point O to point Q in Fig. 3.1 till time t_1 with decreasing speed and turning back and moving with the same negative acceleration.

An interesting feature of a velocity-time graph for any moving object is that **the area under the curve represents the displacement over a given time interval**. A general proof of this

Fig. 3.10 Velocity–time graph for motions with constant acceleration. (a) Motion in positive direction with positive acceleration, (b) Motion in positive direction with negative acceleration, (c) Motion in negative direction with negative acceleration, (d) Motion of an object with negative acceleration that changes direction at time t_1. Between times 0 to t_1, its moves in positive x - direction and between t_1 and t_2 it moves in the opposite direction.

statement requires use of calculus. We can, however, see that it is true for the simple case of an object moving with constant velocity u. Its velocity-time graph is as shown in Fig. 3.11.

Fig. 3.11 Area under v–t curve equals displacement of the object over a given time interval.

The v-t curve is a straight line parallel to the time axis and the area under it between $t = 0$ and $t = T$ is the area of the rectangle of height u and base T. Therefore, area = $u \times T = uT$ which is the displacement in this time interval. How come in this case an area is equal to a distance? Think! Note the dimensions of quantities on the two coordinate axes, and you will arrive at the answer.

Note that the *x-t*, *v-t*, and *a-t* graphs shown in several figures in this chapter have sharp kinks at some points implying that the functions are not differentiable at these points. In any realistic situation, the functions will be differentiable at all points and the graphs will be smooth.

What this means physically is that acceleration and velocity cannot change values abruptly at an instant. Changes are always continuous.

3.6 KINEMATIC EQUATIONS FOR UNIFORMLY ACCELERATED MOTION

For uniformly accelerated motion, we can derive some simple equations that relate displacement (x), time taken (t), initial velocity (v_0), final velocity (v) and acceleration (a). Equation (3.6) already obtained gives a relation between final and initial velocities v and v_0 of an object moving with uniform acceleration a :

$$v = v_0 + at \tag{3.6}$$

This relation is graphically represented in Fig. 3.12. The area under this curve is :

Area between instants 0 and t = Area of triangle ABC + Area of rectangle OACD

$$= \frac{1}{2}(v - v_0)t + v_0 t$$

Fig. 3.12 Area under v-t curve for an object with uniform acceleration.

As explained in the previous section, the area under v-t curve represents the displacement. Therefore, the displacement x of the object is :

$$x = \frac{1}{2}(v - v_0)t + v_0 t \tag{3.7}$$

But $\quad v - v_0 = a\,t$

Therefore, $x = \frac{1}{2} a\, t^2 + v_0 t$

or, $$x = v_0 t + \frac{1}{2} a t^2 \tag{3.8}$$

Equation (3.7) can also be written as

$$x = \frac{v + v_0}{2} t = \bar{v}\, t \tag{3.9a}$$

where,

$$\bar{v} = \frac{v + v_0}{2} \quad \textbf{(constant acceleration only)} \tag{3.9b}$$

Equations (3.9a) and (3.9b) mean that the object has undergone displacement x with an average velocity equal to the arithmetic average of the initial and final velocities.

From Eq. (3.6), $t = (v - v_0)/a$. Substituting this in Eq. (3.9a), we get

$$x = \bar{v}\, t = \left(\frac{v + v_0}{2}\right)\left(\frac{v - v_0}{a}\right) = \frac{v^2 - v_0^2}{2a}$$

$$v^2 = v_0^2 + 2ax \tag{3.10}$$

This equation can also be obtained by substituting the value of t from Eq. (3.6) into Eq. (3.8). Thus, we have obtained three important equations :

$$v = v_0 + at$$

$$x = v_0 t + \frac{1}{2}at^2$$

$$v^2 = v_0^2 + 2ax \qquad (3.11a)$$

connecting five quantities v_0, v, a, t and x. These are kinematic equations of rectilinear motion for constant acceleration.

The set of Eq. (3.11a) were obtained by assuming that at t = 0, the position of the particle, x is 0. We can obtain a more general equation if we take the position coordinate at t = 0 as non-zero, say x_0. Then Eqs. (3.11a) are modified (replacing x by $x - x_0$) to :

$$v = v_0 + at$$

$$x = x_0 + v_0 t + \frac{1}{2}at^2 \qquad (3.11b)$$

$$v^2 = v_0^2 + 2a(x - x_0) \qquad (3.11c)$$

▶ ***Example 3.3*** Obtain equations of motion for constant acceleration using method of calculus.

Answer By definition

$$a = \frac{dv}{dt}$$

$$dv = a\,dt$$

Integrating both sides

$$\int_{v_0}^{v} dv = \int_0^t a\ dt$$

$$= a\int_0^t dt \qquad (a \text{ is constant})$$

$$v - v_0 = at$$

$$v = v_0 + at$$

Further, $$v = \frac{dx}{dt}$$

$$dx = v\,dt$$

Integrating both sides

$$\int_{x_0}^{x} dx = \int_0^t v\ dt$$

$$= \int_0^t (v_0 + at)\ dt$$

$$x - x_0 = v_0\ t + \frac{1}{2}a\ t^2$$

$$x = x_0 + v_0\ t + \frac{1}{2}\ a\ t^2$$

We can write

$$a = \frac{dv}{dt} = \frac{dv}{dx}\frac{dx}{dt} = v\frac{dv}{dx}$$

or, $v\,dv = a\,dx$

Integrating both sides,

$$\int_{v_0}^{v} v\,dv = \int_{x_0}^{x} a\,dx$$

$$\frac{v^2 - v_0^2}{2} = a(x - x_0)$$

$$v^2 = v_0^2 + 2a(x - x_0)$$

The advantage of this method is that it can be used for motion with non-uniform acceleration also.

Now, we shall use these equations to some important cases. ◀

▶ ***Example 3.4*** A ball is thrown vertically upwards with a velocity of 20 m s^{-1} from the top of a multistorey building. The height of the point from where the ball is thrown is 25.0 m from the ground. (a) How high will the ball rise ? and (b) how long will it be before the ball hits the ground? Take g = 10 m s^{-2}.

Answer (a) Let us take the y-axis in the vertically upward direction with zero at the ground, as shown in Fig. 3.13.

Now v_o = + 20 m s^{-1},

a = – g = –10 m s^{-2},

v = 0 m s^{-1}

If the ball rises to height y from the point of launch, then using the equation

$$v^2 = v_0^2 + 2\ a\ (y - y_0)$$

we get

$$0 = (20)^2 + 2(-10)(y - y_0)$$

Solving, we get, $(y - y_0)$ = 20 m.

(b) We can solve this part of the problem in two ways. **Note carefully the methods used**.

Fig. 3.13

FIRST METHOD : In the first method, we split the path in two parts : the upward motion (A to B) and the downward motion (B to C) and calculate the corresponding time taken t_1 and t_2. Since the velocity at B is zero, we have :

$$v = v_o + at$$

$$0 = 20 - 10t_1$$

Or, $t_1 = 2$ s

This is the time in going from *A* to *B*. From *B*, or the point of the maximum height, the ball falls freely under the acceleration due to gravity. The ball is moving in negative *y* direction. We use equation

$$y = y_0 + v_0 t + \frac{1}{2}at^2$$

We have, $y_0 = 45$ m, $y = 0$, $v_0 = 0$, $a = -g = -10$ m s^{-2}

$$0 = 45 + (½)(-10)\, t_2^2$$

Solving, we get $t_2 = 3$ s

Therefore, the total time taken by the ball before it hits the ground = $t_1 + t_2 = 2$ s + 3 s = 5 s.

SECOND METHOD : The total time taken can also be calculated by noting the coordinates of initial and final positions of the ball with respect to the origin chosen and using equation

$$y = y_0 + v_0 t + \frac{1}{2}at^2$$

Now $y_0 = 25$ m $\quad y = 0$ m

$v_o = 20$ m s^{-1}, $\quad a = -10$m s^{-2}, $\quad t = ?$

$$0 = 25 + 20\,t + (½)\,(-10)\,t^2$$

Or, $5t^2 - 20t - 25 = 0$

Solving this quadratic equation for *t*, we get

$t = 5$s

Note that the second method is better since we do not have to worry about the path of the motion as the motion is under constant acceleration.

◀

▶ ***Example 3.5*** **Free-fall :** Discuss the motion of an object under free fall. Neglect air resistance.

Answer An object released near the surface of the Earth is accelerated downward under the influence of the force of gravity. The magnitude of acceleration due to gravity is represented by ***g***. If air resistance is neglected, the object is said to be in **free fall**. If the height through which the object falls is small compared to the earth's radius, *g* can be taken to be constant, equal to 9.8 m s^{-2}. Free fall is thus a case of motion with uniform acceleration.

We assume that the motion is in *y*-direction, more correctly in –*y*-direction because we choose upward direction as positive. Since the acceleration due to gravity is always downward, it is in the negative direction and we have

$a = -g = -9.8$ m s^{-2}

The object is released from rest at $y = 0$. Therefore, $v_0 = 0$ and the equations of motion become:

$v = 0 - g\,t \quad = -9.8\,t$ m s^{-1}

$y = 0 - ½\,g\,t^2 \quad = -4.9\,t^2$ m

$v^2 = 0 - 2\,g\,y \quad = -19.6\,y$ m^2 s^{-2}

These equations give the velocity and the distance travelled as a function of time and also the variation of velocity with distance. The variation of acceleration, velocity, and distance, with time have been plotted in Fig. 3.14(a), (b) and (c).

(a)

(b)

(c)

Fig. 3.14 *Motion of an object under free fall.*
(a) Variation of acceleration with time.
(b) Variation of velocity with time.
(c) Variation of distance with time

Example 3.6 Galileo's law of odd numbers : *"The distances traversed, during equal intervals of time, by a body falling from rest, stand to one another in the same ratio as the odd numbers beginning with unity [namely, 1: 3: 5: 7.....]."* Prove it.

Answer Let us divide the time interval of motion of an object under free fall into many equal intervals τ and find out the distances traversed during successive intervals of time. Since initial velocity is zero, we have

$$y = -\frac{1}{2}gt^2$$

Using this equation, we can calculate the position of the object after different time intervals, 0, τ, 2τ, 3τ... which are given in second column of Table 3.2. If we take $(-1/2)\ g\tau^2$ as y_0 — the position coordinate after first time interval τ, then third column gives the positions in the unit of y_o. The fourth column gives the distances traversed in successive τs. We find that the distances are in the simple ratio 1: 3: 5: 7: 9: 11... as shown in the last column. This law was established by Galileo Galilei (1564-1642) who was the first to make quantitative studies of free fall.

Example 3.7 Stopping distance of vehicles : When brakes are applied to a moving vehicle, the distance it travels before stopping is called stopping distance. It is an important factor for road safety and depends on the initial velocity (v_0) and the braking capacity, or deceleration, $-a$ that is caused by the braking. Derive an expression for stopping distance of a vehicle in terms of v_o and a.

Answer Let the distance travelled by the vehicle before it stops be d_s. Then, using equation of motion $v^2 = v_o^2 + 2\,ax$, and noting that $v = 0$, we have the stopping distance

Thus, the stopping distance is proportional to the square of the initial velocity. Doubling the

Table 3.2

t	y	y in terms of y_o [=(- ½) g τ^2]	Distance traversed in successive intervals	Ratio of distances traversed
0	0	0		
τ	-(1/2) g τ^2	y_o	y_o	1
2 τ	-4(1/2) g τ^2	4 y_o	3 y_o	3
3 τ	-9(1/2) g τ^2	9 y_o	5 y_o	5
4 τ	-16(1/2) g τ^2	16 y_o	7 y_o	7
5 τ	-25(1/2) g τ^2	25 y_o	9 y_o	9
6 τ	-36(1/2) g τ^2	36 y_o	11 y_o	11

initial velocity increases the stopping distance by a factor of 4 (for the same deceleration).

For the car of a particular make, the braking distance was found to be 10 m, 20 m, 34 m and 50 m corresponding to velocities of 11, 15, 20 and 25 m/s which are nearly consistent with the above formula.

Stopping distance is an important factor considered in setting speed limits, for example, in school zones. ◀

▶ ***Example 3.8*** **Reaction time :** When a situation demands our immediate action, it takes some time before we really respond. Reaction time is the time a person takes to observe, think and act. For example, if a person is driving and suddenly a boy appears on the road, then the time elapsed before he slams the brakes of the car is the reaction time. Reaction time depends on complexity of the situation and on an individual.

You can measure your reaction time by a simple experiment. Take a ruler and ask your friend to drop it vertically through the gap between your thumb and forefinger (Fig. 3.15). After you catch it, find the distance d travelled by the ruler. In a particular case, d was found to be 21.0 cm. Estimate reaction time.

Fig. 3.15 *Measuring the reaction time.*

Answer The ruler drops under free fall. Therefore, $v_o = 0$, and $g = -9.8$ m s^{-2}. The distance travelled d and the reaction time t_r are related by

$$d = -\frac{1}{2} g t_r^2$$

Or, $$t_r = \sqrt{\frac{2d}{g}} \text{ s}$$

Given $d = 21.0$ cm and $g = 9.8$ m s^{-2} the reaction time is

$$t_r = \sqrt{\frac{2 \times 0.21}{9.8}} \text{ s} \cong 0.2 \text{ s.}$$ ◀

3.7 RELATIVE VELOCITY

You must be familiar with the experience of travelling in a train and being overtaken by another train moving in the same direction as you are. While that train must be travelling faster than you to be able to pass you, it does seem slower to you than it would be to someone standing on the ground and watching both the trains. In case both the trains have the same velocity with respect to the ground, then to you the other train would seem to be not moving at all. To understand such observations, we now introduce the concept of relative velocity.

Consider two objects A and B moving uniformly with average velocities v_A and v_B in one dimension, say along x-axis. (Unless otherwise specified, the velocities mentioned in this chapter are measured with reference to the ground). If $x_A(0)$ and $x_B(0)$ are positions of objects A and B, respectively at time $t = 0$, their positions $x_A(t)$ and $x_B(t)$ at time t are given by:

$$x_A(t) = x_A(0) + v_A t \quad (3.12a)$$
$$x_B(t) = x_B(0) + v_B t \quad (3.12b)$$

Then, the displacement from object A to object B is given by

$$x_{BA}(t) = x_B(t) - x_A(t)$$
$$= [\, x_B(0) - x_A(0) \,] + (v_B - v_A)\, t. \quad (3.13)$$

Equation (3.13) is easily interpreted. It tells us that as seen from object A, object B has a velocity $v_B - v_A$ because the displacement from A to B changes steadily by the amount $v_B - v_A$ in each unit of time. We say that the velocity of object B *relative to object* A is $v_B - v_A$:

$$v_{BA} = v_B - v_A \quad (3.14a)$$

Similarly, velocity of object A *relative to object* B is:

$$v_{AB} = v_A - v_B \quad (3.14b)$$

Fig. 3.16 Position-time graphs of two objects with equal velocities.

This shows: $v_{BA} = -v_{AB}$ (3.14c)

Now we consider some special cases :

(a) If $v_B = v_A$, $v_B - v_A = 0$. Then, from Eq. (3.13), $x_B(t) - x_A(t) = x_B(0) - x_A(0)$. Therefore, the two objects stay at a constant distance $(x_B(0) - x_A(0))$ apart, and their position–time graphs are straight lines parallel to each other as shown in Fig. 3.16. The relative velocity v_{AB} or v_{BA} is zero in this case.

(b) If $v_A > v_B$, $v_B - v_A$ is negative. One graph is steeper than the other and they meet at a common point. For example, suppose $v_A = 20$ m s^{-1} and $x_A(0) = 10$ m; and $v_B = 10$ m s^{-1}, $x_B(0) = 40$ m; then the time at which they meet is $t = 3$ s (Fig. 3.17). At this instant they are both at a position $x_A(t) = x_B(t) = 70$ m. Thus, object *A* overtakes object *B* at this time. In this case, $v_{BA} = 10$ m s^{-1} – 20 m s^{-1} = – 10 m s^{-1} = $-v_{AB}$.

(c) Suppose v_A and v_B are of opposite signs. For example, if in the above example object *A* is moving with 20 m s^{-1} starting at $x_A(0) = 10$ m and object *B* is moving with – 10 m s^{-1} starting at $x_B(0) = 40$ m, the two objects meet at $t = 1$ s (Fig. 3.18). The velocity of *B* relative to *A*, $v_{BA} = [-10 - (20)]$ m s^{-1} = –30 m s^{-1} = $-v_{AB}$. In this case, the magnitude of v_{BA} or v_{AB} (= 30 m s^{-1}) is greater than the magnitude of velocity of *A* or that of *B*. If the objects under consideration are two trains, then for a person sitting on either of the two, the other train seems to go very fast.

Note that Eq. (3.14) are valid even if v_A and v_B represent instantaneous velocities.

Fig. 3.17 Position-time graphs of two objects with unequal velocities, showing the time of meeting.

Fig. 3.18 Position-time graphs of two objects with velocities in opposite directions, showing the time of meeting.

▶ ***Example 3.9*** Two parallel rail tracks run north-south. Train A moves north with a speed of 54 km h^{-1}, and train B moves south with a speed of 90 km h^{-1}. What is the
(a) velocity of B with respect to A ?,
(b) velocity of ground with respect to B ?, and
(c) velocity of a monkey running on the roof of the train A against its motion (with a velocity of 18 km h^{-1} with respect to the train A) as observed by a man standing on the ground ?

Answer Choose the positive direction of *x*-axis to be from south to north. Then,

$v_A = +54 \text{ km h}^{-1} = 15 \text{ m s}^{-1}$
$v_B = -90 \text{ km h}^{-1} = -25 \text{ m s}^{-1}$

Relative velocity of B with respect to $A = v_B - v_A = -40 \text{ m s}^{-1}$, i.e. the train B appears to A to move with a speed of 40 m s^{-1} from north to south.

Relative velocity of ground with respect to $B = 0 - v_B = 25 \text{ m s}^{-1}$.

In (c), let the velocity of the monkey with respect to ground be v_M. Relative velocity of the monkey with respect to A,

$v_{MA} = v_M - v_A = -18 \text{ km h}^{-1} = -5 \text{ ms}^{-1}$. Therefore, $v_M = (15 - 5) \text{ m s}^{-1} = 10 \text{ m s}^{-1}$.

SUMMARY

1. An object is said to be in *motion* if its position changes with time. The position of the object can be specified with reference to a conveniently chosen origin. For motion in a straight line, position to the right of the origin is taken as positive and to the left as negative.
2. *Path length* is defined as the total length of the path traversed by an object.
3. *Displacement* is the change in position : $\Delta x = x_2 - x_1$. Path length is greater or equal to the magnitude of the displacement between the same points.
4. An object is said to be in *uniform motion* in a straight line if its displacement is equal in equal intervals of time. Otherwise, the motion is said to be *non-uniform.*
5. *Average velocity* is the displacement divided by the time interval in which the displacement occurs :

$$\bar{v} = \frac{\Delta x}{\Delta t}$$

On an x-t graph, the average velocity over a time interval is the slope of the line connecting the initial and final positions corresponding to that interval.

6. *Average Speed* is the ratio of total path length traversed and the corresponding time interval.

The average speed of an object is greater or equal to the magnitude of the average velocity over a given time interval.

7. *Instantaneous velocity* or simply velocity is defined as the limit of the average velocity as the time interval Δt becomes infinitesimally small :

$$v = \lim_{\Delta t \to 0} \bar{v} = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} = \frac{dx}{dt}$$

The velocity at a particular instant is equal to the slope of the tangent drawn on position-time graph at that instant.

8. *Average acceleration* is the change in velocity divided by the time interval during which the change occurs :

$$\bar{a} = \frac{\Delta v}{\Delta t}$$

9. *Instantaneous acceleration* is defined as the limit of the average acceleration as the time interval Δt goes to zero :

$$a = \lim_{\Delta t \to 0} \bar{a} = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} = \frac{dv}{dt}$$

The acceleration of an object at a particular time is the slope of the velocity-time graph at that instant of time. For uniform motion, acceleration is zero and the x-t graph is a straight line inclined to the time axis and the v-t graph is a straight line

parallel to the time axis. For motion with uniform acceleration, x-t graph is a parabola while the v-t graph is a straight line inclined to the time axis.

10. The area under the velocity-time curve between times t_1 and t_2 is equal to the displacement of the object during that interval of time.

11. For objects in uniformly accelerated rectilinear motion, the five quantities, displacement x, time taken t, initial velocity v_0, final velocity v and acceleration a are related by a set of simple equations called *kinematic equations of motion :*

$$v = v_0 + at$$

$$x = v_0 t + \frac{1}{2} at^2$$

$$v^2 = v_0^2 + 2ax$$

if the position of the object at time $t = 0$ is 0. If the particle starts at $x = x_0$, x in above equations is replaced by $(x - x_0)$.

Physical quantity	Symbol	Dimensions	Unit	Remarks
Path length		[L]	m	
Displacement	Δx	[L]	m	$= x_2 - x_1$ In one dimension, its sign indicates the direction.
Velocity		$[LT^{-1}]$	$m\ s^{-1}$	
(a) Average	$\bar{v}$			$= \frac{\Delta x}{\Delta t}$
(b) Instantaneous	v			$= \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} = \frac{dx}{dt}$ In one dimension, its sign indicates the direction.
Speed		$[LT^{-1}]$	$m\ s^{-1}$	
(a) Average				$= \frac{\text{Path length}}{\text{Time interval}}$
(b) Instantaneous				$= \frac{dx}{dt}$
Acceleration		$[LT^{-2}]$	$m\ s^{-2}$	
(a) Average	$\bar{a}$			$= \frac{\Delta v}{\Delta t}$
(b) Instantaneous	a			$= \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} = \frac{dv}{dt}$ In one dimension, its sign indicates the direction.

POINTS TO PONDER

1. The path length traversed by an object between two points is, in general, not the same as the magnitude of displacement. The displacement depends only on the end points; the path length (as the name implies) depends on the actual path. In one dimension, the two quantities are equal only if the object does not change its direction during the course of motion. In all other cases, the path length is greater than the magnitude of displacement.
2. In view of point 1 above, the average speed of an object is greater than or equal to the magnitude of the average velocity over a given time interval. The two are equal only if the path length is equal to the magnitude of displacement.
3. The origin and the positive direction of an axis are a matter of choice. You should first specify this choice before you assign signs to quantities like displacement, velocity and acceleration.
4. If a particle is speeding up, acceleration is in the direction of velocity; if its speed is decreasing, acceleration is in the direction opposite to that of the velocity. This statement is independent of the choice of the origin and the axis.
5. The sign of acceleration does not tell us whether the particle's speed is increasing or decreasing. The sign of acceleration (as mentioned in point 3) depends on the choice of the positive direction of the axis. For example, if the vertically upward direction is chosen to be the positive direction of the axis, the acceleration due to gravity is negative. If a particle is falling under gravity, this acceleration, though negative, results in increase in speed. For a particle thrown upward, the same negative acceleration (of gravity) results in decrease in speed.
6. The zero velocity of a particle at any instant does not necessarily imply zero acceleration at that instant. A particle may be momentarily at rest and yet have non-zero acceleration. For example, a particle thrown up has zero velocity at its uppermost point but the acceleration at that instant continues to be the acceleration due to gravity.
7. In the kinematic equations of motion [Eq. (3.11)], the various quantities are algebraic, i.e. they may be positive or negative. The equations are applicable in all situations (for one dimensional motion with constant acceleration) provided the values of different quantities are substituted in the equations with proper signs.
8. The definitions of instantaneous velocity and acceleration (Eqs. (3.3) and (3.5)) are exact and are always correct while the kinematic equations (Eq. (3.11)) are true only for motion in which the magnitude and the direction of acceleration are constant during the course of motion.

EXERCISES

3.1 In which of the following examples of motion, can the body be considered approximately a point object:
(a) a railway carriage moving without jerks between two stations.
(b) a monkey sitting on top of a man cycling smoothly on a circular track.
(c) a spinning cricket ball that turns sharply on hitting the ground.
(d) a tumbling beaker that has slipped off the edge of a table.

3.2 The position-time (x-t) graphs for two children A and B returning from their school O to their homes P and Q respectively are shown in Fig. 3.19. Choose the correct entries in the brackets below ;
(a) (A/B) lives closer to the school than (B/A)
(b) (A/B) starts from the school earlier than (B/A)
(c) (A/B) walks faster than (B/A)
(d) A and B reach home at the (same/different) time
(e) (A/B) overtakes (B/A) on the road (once/twice).

Fig. 3.19

3.3 A woman starts from her home at 9.00 am, walks with a speed of 5 km h^{-1} on a straight road up to her office 2.5 km away, stays at the office up to 5.00 pm, and returns home by an auto with a speed of 25 km h^{-1}. Choose suitable scales and plot the *x-t* graph of her motion.

3.4 A drunkard walking in a narrow lane takes 5 steps forward and 3 steps backward, followed again by 5 steps forward and 3 steps backward, and so on. Each step is 1 m long and requires 1 s. Plot the *x-t* graph of his motion. Determine graphically and otherwise how long the drunkard takes to fall in a pit 13 m away from the start.

3.5 A jet airplane travelling at the speed of 500 km h^{-1} ejects its products of combustion at the speed of 1500 km h^{-1} relative to the jet plane. What is the speed of the latter with respect to an observer on the ground ?

3.6 A car moving along a straight highway with speed of 126 km h^{-1} is brought to a stop within a distance of 200 m. What is the retardation of the car (assumed uniform), and how long does it take for the car to stop ?

3.7 Two trains A and B of length 400 m each are moving on two parallel tracks with a uniform speed of 72 km h^{-1} in the same direction, with A ahead of B. The driver of B decides to overtake A and accelerates by 1 m s^{-2}. If after 50 s, the guard of B just brushes past the driver of A, what was the original distance between them ?

3.8 On a two-lane road, car A is travelling with a speed of 36 km h^{-1}. Two cars B and C approach car A in opposite directions with a speed of 54 km h^{-1} each. At a certain instant, when the distance AB is equal to AC, both being 1 km, B decides to overtake A before C does. What minimum acceleration of car B is required to avoid an accident ?

3.9 Two towns A and B are connected by a regular bus service with a bus leaving in either direction every *T* minutes. A man cycling with a speed of 20 km h^{-1} in the direction A to B notices that a bus goes past him every 18 min in the direction of his motion, and every 6 min in the opposite direction. What is the period *T* of the bus service and with what speed (assumed constant) do the buses ply on the road?

3.10 A player throws a ball upwards with an initial speed of 29.4 m s^{-1}.

(a) What is the direction of acceleration during the upward motion of the ball ?

(b) What are the velocity and acceleration of the ball at the highest point of its motion ?

(c) Choose the $x = 0$ m and $t = 0$ s to be the location and time of the ball at its highest point, vertically downward direction to be the positive direction of *x*-axis, and give the signs of position, velocity and acceleration of the ball during its upward, and downward motion.

(d) To what height does the ball rise and after how long does the ball return to the player's hands ? (Take $g = 9.8$ m s^{-2} and neglect air resistance).

3.11 Read each statement below carefully and state with reasons and examples, if it is true or false ;

A particle in one-dimensional motion

(a) with zero speed at an instant may have non-zero acceleration at that instant

(b) with zero speed may have non-zero velocity,

(c) with constant speed must have zero acceleration,

(d) with positive value of acceleration *must* be speeding up.

3.12 A ball is dropped from a height of 90 m on a floor. At each collision with the floor, the ball loses one tenth of its speed. Plot the speed-time graph of its motion between t = 0 to 12 s.

3.13 Explain clearly, with examples, the distinction between :

(a) magnitude of displacement (sometimes called distance) over an interval of time, and the total length of path covered by a particle over the same interval;

(b) magnitude of average velocity over an interval of time, and the average speed over the same interval. [Average speed of a particle over an interval of time is defined as the total path length divided by the time interval]. Show in both (a) and (b) that the second quantity is either greater than or equal to the first. When is the equality sign true ? [For simplicity, consider one-dimensional motion only].

3.14 A man walks on a straight road from his home to a market 2.5 km away with a speed of 5 km h^{-1}. Finding the market closed, he instantly turns and walks back home with a speed of 7.5 km h^{-1}. What is the

(a) magnitude of average velocity, and

(b) average speed of the man over the interval of time (i) 0 to 30 min, (ii) 0 to 50 min, (iii) 0 to 40 min ? [Note: You will appreciate from this exercise why it is better to define average speed as total path length divided by time, and not as magnitude of average velocity. You would not like to tell the tired man on his return home that his average speed was zero !]

3.15 In Exercises 3.13 and 3.14, we have carefully distinguished between *average* speed and magnitude of *average* velocity. No such distinction is necessary when we consider instantaneous speed and magnitude of velocity. The instantaneous speed is always equal to the magnitude of instantaneous velocity. Why ?

3.16 Look at the graphs (a) to (d) (Fig. 3.20) carefully and state, with reasons, which of these *cannot* possibly represent one-dimensional motion of a particle.

Fig. 3.20

3.17 Figure 3.21 shows the *x-t* plot of one-dimensional motion of a particle. Is it correct to say from the graph that the particle moves in a straight line for $t < 0$ and on a parabolic path for $t > 0$? If not, suggest a suitable physical context for this graph.

Fig. 3.21

3.18 A police van moving on a highway with a speed of 30 km h^{-1} fires a bullet at a thief's car speeding away in the same direction with a speed of 192 km h^{-1}. If the muzzle speed of the bullet is 150 m s^{-1}, with what speed does the bullet hit the thief's car ? (Note: Obtain that speed which is relevant for damaging the thief's car).

3.19 Suggest a suitable physical situation for each of the following graphs (Fig 3.22):

Fig. 3.22

3.20 Figure 3.23 gives the *x-t* plot of a particle executing one-dimensional simple harmonic motion. (You will learn about this motion in more detail in Chapter14). Give the signs of position, velocity and acceleration variables of the particle at $t = 0.3$ s, 1.2 s, – 1.2 s.

Fig. 3.23

3.21 Figure 3.24 gives the *x-t* plot of a particle in one-dimensional motion. Three different equal intervals of time are shown. In which interval is the average speed greatest, and in which is it the least ? Give the sign of average velocity for each interval.

Fig. 3.24

3.22 Figure 3.25 gives a speed-time graph of a particle in motion along a constant direction. Three equal intervals of time are shown. In which interval is the average acceleration greatest in magnitude ? In which interval is the average speed greatest ? Choosing the positive direction as the constant direction of motion, give the signs of v and a in the three intervals. What are the accelerations at the points A, B, C and D ?

Fig. 3.25

Additional Exercises

3.23 A three-wheeler starts from rest, accelerates uniformly with 1 m s^{-2} on a straight road for 10 s, and then moves with uniform velocity. Plot the distance covered by the vehicle during the nth second (n = 1,2,3....) versus n. What do you expect this plot to be during accelerated motion : a straight line or a parabola ?

3.24 A boy standing on a stationary lift (open from above) throws a ball upwards with the maximum initial speed he can, equal to 49 m s^{-1}. How much time does the ball take to return to his hands? If the lift starts moving up with a uniform speed of 5 m s^{-1} and the boy again throws the ball up with the maximum speed he can, how long does the ball take to return to his hands ?

3.25 On a long horizontally moving belt (Fig. 3.26), a child runs to and fro with a speed 9 km h^{-1} (with respect to the belt) between his father and mother located 50 m apart on the moving belt. The belt moves with a speed of 4 km h^{-1}. For an observer on a stationary platform outside, what is the

(a) speed of the child running in the direction of motion of the belt ?.

(b) speed of the child running opposite to the direction of motion of the belt ?

(c) time taken by the child in (a) and (b) ?

Which of the answers alter if motion is viewed by one of the parents ?

Fig. 3.26

3.26 Two stones are thrown up simultaneously from the edge of a cliff 200 m high with initial speeds of 15 m s^{-1} and 30 m s^{-1}. Verify that the graph shown in Fig. 3.27 correctly represents the time variation of the relative position of the second stone with respect to the first. Neglect air resistance and assume that the stones do not rebound after hitting the ground. Take g = 10 m s^{-2}. Give the equations for the linear and curved parts of the plot.

Fig. 3.27

3.27 The speed-time graph of a particle moving along a fixed direction is shown in Fig. 3.28. Obtain the distance traversed by the particle between (a) t = 0 s to 10 s, (b) t = 2 s to 6 s.

Fig. 3.28

What is the average speed of the particle over the intervals in (a) and (b) ?

3.28 The velocity-time graph of a particle in one-dimensional motion is shown in Fig. 3.29 :

Fig. 3.29

Which of the following formulae are correct for describing the motion of the particle over the time-interval t_1 to t_2:

(a) $x(t_2) = x(t_1) + v(t_1)(t_2 - t_1) + (½)\, a\, (t_2 - t_1)^2$

(b) $v(t_2) = v(t_1) + a\,(t_2 - t_1)$

(c) $v_{average} = (x(t_2) - x(t_1))/(t_2 - t_1)$

(d) $a_{average} = (v(t_2) - v(t_1))/(t_2 - t_1)$

(e) $x(t_2) = x(t_1) + v_{average}(t_2 - t_1) + (½)\, a_{average}\, (t_2 - t_1)^2$

(f) $x(t_2) - x(t_1)$ = area under the v-t curve bounded by the t-axis and the dotted line shown.

APPENDIX 3.1 : ELEMENTS OF CALCULUS

Differential Calculus

Using the concept of 'differential coefficient' or 'derivative', we can easily define velocity and acceleration. Though you will learn in detail in mathematics about derivatives, we shall introduce this concept in brief in this Appendix so as to facilitate its use in describing physical quantities involved in motion.

Suppose we have a quantity y whose value depends upon a single variable x, and is expressed by an equation defining y as some specific function of x. This is represented as:

$$y = f(x) \tag{1}$$

This relationship can be visualised by drawing a graph of function $y = f(x)$ regarding y and x as Cartesian coordinates, as shown in Fig. 3.30 (a).

Fig. 3.30

Consider the point P on the curve $y = f(x)$ whose coordinates are (x, y) and another point Q where coordinates are $(x + \Delta x, y + \Delta y)$. The slope of the line joining P and Q is given by:

$$\tan\theta = \frac{\Delta y}{\Delta x} = \frac{(y + \Delta y) - y}{\Delta x} \tag{2}$$

Suppose now that the point Q moves along the curve towards P. In this process, Δy *and* Δx decrease and approach zero; though their ratio $\frac{\Delta y}{\Delta x}$ will not necessarily vanish. What happens to the line PQ as $\Delta y \to 0$, $\Delta x \to 0$. You can see that this line becomes a tangent to the curve at point P as shown in Fig. 3.30(b). This means that tan θ approaches the slope of the tangent at P, denoted by m:

$$m = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{(y + \Delta y) - y}{\Delta x} \tag{3}$$

The limit of the ratio $\Delta y/\Delta x$ as Δx approaches zero is called the derivative of y with respect to x and is written as dy/dx. It represents the slope of the tangent line to the curve $y = f(x)$ at the point (x, y).

Since $y = f(x)$ and $y + \Delta y = f(x + \Delta x)$, we can write the definition of the derivative as:

$$\frac{dy}{dx} = \frac{df(x)}{dx} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \left[\frac{f(x + \Delta x) - f(x)}{\Delta x} \right]$$

Given below are some elementary formulae for derivatives of functions. In these $u(x)$ and $v(x)$ represent arbitrary functions of x, and a and b denote constant quantities that are independent of x. Derivatives of some common functions are also listed .

$$\frac{d(a\,u)}{dx} = a\frac{du}{dx} \quad ; \quad \frac{du}{dt} = \frac{du}{dx}\cdot\frac{dx}{dt}$$

$$\frac{d(uv)}{dx} = u\frac{dv}{dx} + v\frac{du}{dx} \quad ; \quad \frac{d(u/v)}{dx} = \frac{1}{v^2}\frac{du}{dx} - u\frac{dv}{dx}$$

$$\frac{du}{dv} = \frac{du/dx}{dv/dx}$$

$$\frac{d}{dx}(\sin x) = \cos x \quad ; \quad \frac{d}{dx}(\cos x) = -\sin x$$

$$\frac{d}{dx}(\tan x) = \sec^2 x \quad ; \quad \frac{d}{dx}(\cot x) = -\operatorname{cosec}^2 x$$

$$\frac{d}{dx}(\sec x) = \tan x \sec x \quad ; \quad \frac{d}{dx}(\operatorname{cosec}^2 x) = -\cot x \operatorname{cosec} x$$

$$\frac{d}{dx}(u)^n = n\,u^{n-1}\frac{du}{dx} \quad ; \quad \frac{d}{du}(\ln u) = \frac{1}{u}$$

$$\frac{d}{du}(e^u) = e^u$$

In terms of derivatives, instantaneous velocity and acceleration are defined as

$$v = \lim_{\Delta t \to 0}\frac{\Delta x}{\Delta t} = \frac{dx}{dt}$$

$$a = \lim_{Dt \to 0}\frac{Dv}{Dt} = \frac{dv}{dt} = \frac{d^2x}{dt^2}$$

Integral Calculus

You are familiar with the notion of area. The formulae for areas of simple geometrical figures are also known to you. For example, the area of a rectangle is length times breadth and that of a triangle is half of the product of base and height. But how to deal with the problem of determination of area of an irregular figure? The mathematical notion of integral is necessary in connection with such problems.

Let us take a concrete example. Suppose a variable force $f(x)$ acts on a particle in its motion along x - axis from $x = a$ *to* $x = b$. The problem is to determine the work done (W) by the force on the particle during the motion. This problem is discussed in detail in Chapter 6.

Figure 3.31 shows the variation of $F(x)$ with x. If the force were constant, work would be simply the area $F(b\text{-}a)$ as shown in Fig. 3.31(i). But in the general case, force is varying .

Fig. 3.31

To calculate the area under this curve [Fig. 3.31 (ii)], let us employ the following trick. Divide the interval on *x-axis* from *a to b* into a large number (N) of small intervals: $x_0(=a)$ *to* x_1, x_1 *to* x_2; x_2 *to* x_3, ... x_{N-1} *to* $x_N(=b)$. The area under the curve is thus divided into N strips. Each strip is approximately a rectangle, since the variation of $F(x)$ over a strip is negligible. The area of the ith strip shown [Fig. 3.31(ii)] is then approximately

$$\Delta A_i = F(x_i)(x_i - x_{i-1}) = F(x_i)\Delta x$$

where Δx is the width of the strip which we have taken to be the same for all the strips. You may wonder whether we should put $F(x_{i-1})$ or the mean of $F(x_i)$ and $F(x_{i-1})$ in the above expression. If we take N to be very very large ($N \to \infty$), it does not really matter, since then the strip will be so thin that the difference between $F(x_i)$ and $F(x_{i-1})$ is vanishingly small. The total area under the curve then is:

$$A = \sum_{i=1}^{N} \Delta A_i = \sum_{i=1}^{N} F(x_i)\Delta x$$

The limit of this sum as $N \to \infty$ is known as the integral of $F(x)$ over x from a to b. It is given a special symbol as shown below:

$$A = \int_a^b F(x)dx$$

The integral sign $\int$ looks like an elongated S, reminding us that it basically is the limit of the sum of an infinite number of terms.

A most significant mathematical fact is that integration is, in a sense, an inverse of differentiation. Suppose we have a function $g(x)$ whose derivative is $f(x)$, i.e. $f(x) = \dfrac{dg(x)}{dx}$

The function g (x) is known as the indefinite integral of $f(x)$ and is denoted as:

$$g(x) = \int f(x)dx$$

An integral with lower and upper limits is known as a definite integral. It is a number. Indefinite integral has no limits; it is a function.

A fundamental theorem of mathematics states that

$$\int_a^b f(x)\,dx = g(x)\Big|_a^b \equiv g(b) - g(a)$$

As an example, suppose $f(x) = x^2$ and we wish to determine the value of the definite integral from $x = 1$ to $x = 2$. The function $g(x)$ whose derivative is x^2 is $x^3/3$. Therefore,

$$\int_1^2 x^2\,dx = \frac{x^3}{3}\Bigg|_1^2 = \frac{8}{3} - \frac{1}{3} = \frac{7}{3}$$

Clearly, to evaluate definite integrals, we need to know the corresponding indefinite integrals. Some common indefinite integrals are

$$\int x^n \mathrm{d}x = \frac{x^{n+1}}{n+1} \qquad (n \neq -1)$$

$$\int \left(\frac{1}{x}\right)\mathrm{d}x = \ln x \qquad (x > 0)$$

$$\int \sin x \ \mathrm{d}x = -\cos x \qquad \int \cos x \ \mathrm{d}x = \sin x$$

$$\int \mathrm{e}^x \mathrm{d}x = \mathrm{e}^x$$

This introduction to differential and integral calculus is not rigorous and is intended to convey to you the basic notions of calculus.

CHAPTER FOUR

MOTION IN A PLANE

4.1 Introduction
4.2 Scalars and vectors
4.3 Multiplication of vectors by real numbers
4.4 Addition and subtraction of vectors — graphical method
4.5 Resolution of vectors
4.6 Vector addition — analytical method
4.7 Motion in a plane
4.8 Motion in a plane with constant acceleration
4.9 Relative velocity in two dimensions
4.10 Projectile motion
4.11 Uniform circular motion
Summary
Points to ponder
Exercises
Additional exercises

4.1 INTRODUCTION

In the last chapter we developed the concepts of position, displacement, velocity and acceleration that are needed to describe the motion of an object along a straight line. We found that the directional aspect of these quantities can be taken care of by + and – signs, as in one dimension only two directions are possible. But in order to describe motion of an object in two dimensions (a plane) or three dimensions (space), we need to use vectors to describe the above-mentioned physical quantities. Therefore, it is first necessary to learn the language of vectors. What is a vector? How to add, subtract and multiply vectors ? What is the result of multiplying a vector by a real number ? We shall learn this to enable us to use vectors for defining velocity and acceleration in a plane. We then discuss motion of an object in a plane. As a simple case of motion in a plane, we shall discuss motion with constant acceleration and treat in detail the projectile motion. Circular motion is a familiar class of motion that has a special significance in daily-life situations. We shall discuss uniform circular motion in some detail.

The equations developed in this chapter for motion in a plane can be easily extended to the case of three dimensions.

4.2 SCALARS AND VECTORS

In physics, we can classify quantities as scalars or vectors. Basically, the difference is that a **direction** is associated with a vector but not with a scalar. A scalar quantity is a quantity with magnitude only. It is specified completely by a single number, along with the proper unit. Examples are : the distance between two points, mass of an object, the temperature of a body and the time at which a certain event happened. The rules for combining scalars are the rules of ordinary algebra. Scalars can be added, subtracted, multiplied and divided

just as the ordinary numbers*. For example, if the length and breadth of a rectangle are 1.0 m and 0.5 m respectively, then its perimeter is the sum of the lengths of the four sides, 1.0 m + 0.5 m +1.0 m + 0.5 m = 3.0 m. The length of each side is a scalar and the perimeter is also a scalar. Take another example: the maximum and minimum temperatures on a particular day are 35.6 °C and 24.2 °C respectively. Then, the difference between the two temperatures is 11.4 °C. Similarly, if a uniform solid cube of aluminium of side 10 cm has a mass of 2.7 kg, then its volume is 10^{-3} m^3 (a scalar) and its density is 2.7×10^3 kg m^{-3} (a scalar).

A **vector** quantity is a quantity that has both a magnitude and a direction and obeys the **triangle law of addition** or equivalently the **parallelogram law of addition**. So, a vector is specified by giving its magnitude by a number and its direction. Some physical quantities that are represented by vectors are displacement, velocity, acceleration and force.

To represent a vector, we use a bold face type in this book. Thus, a velocity vector can be represented by a symbol **v**. Since bold face is difficult to produce, when written by hand, a vector is often represented by an arrow placed over a letter, say $\vec{v}$. Thus, both **v** and $\vec{v}$ represent the velocity vector. The magnitude of a vector is often called its absolute value, indicated by $|\mathbf{v}| = v$. Thus, a vector is represented by a bold face, e.g. by **A, a, p, q, r, ... x, y**, with respective magnitudes denoted by light face *A, a, p, q, r, ... x, y*.

4.2.1 Position and Displacement Vectors

To describe the position of an object moving in a plane, we need to choose a convenient point, say O as origin. Let P and P′ be the positions of the object at time *t* and *t*′, respectively [Fig. 4.1(a)]. We join O and P by a straight line. Then, **OP** is the position vector of the object at time *t*. An arrow is marked at the head of this line. It is represented by a symbol **r**, i.e. **OP** = **r**. Point P′ is represented by another position vector, **OP**′ denoted by **r**′. The length of the vector **r** represents the magnitude of the vector and its direction is the direction in which P lies as seen from O. If the object moves from P to P′, the vector **PP**′ (with tail at P and tip at P′) is called the **displacement vector** corresponding to motion from point P (at time *t*) to point P′ (at time *t*′).

Fig. 4.1 (a) *Position and displacement vectors.* (b) *Displacement vector* **PQ** *and different courses of motion.*

It is important to note that displacement vector is the straight line joining the initial and final positions and does not depend on the actual path undertaken by the object between the two positions. For example, in Fig. 4.1b, given the initial and final positions as P and Q, the displacement vector is the same **PQ** for different paths of journey, say PABCQ, PDQ, and PBEFQ. Therefore, the **magnitude of displacement is either less or equal to the path length of an object between two points**. This fact was emphasised in the previous chapter also while discussing motion along a straight line.

4.2.2 Equality of Vectors

Two vectors **A** and **B** are said to be equal if, and only if, they have the same magnitude and the same direction.**

Figure 4.2(a) shows two equal vectors **A** and **B**. We can easily check their equality. Shift **B** parallel to itself until its tail Q coincides with that of *A*, i.e. Q coincides with O. Then, since their tips S and P also coincide, the two vectors are said to be equal. In general, equality is indicated

* *Addition and subtraction of scalars make sense only for quantities with same units. However, you can multiply and divide scalars of different units.*

** *In our study, vectors do not have fixed locations. So displacing a vector parallel to itself leaves the vector unchanged. Such vectors are called free vectors. However, in some physical applications, location or line of application of a vector is important. Such vectors are called localised vectors.*

Fig. 4.2 (a) *Two equal vectors* **A** *and* **B**. (b) *Two vectors* **A′** *and* **B′** *are unequal though they are of the same length.*

as **A** = **B**. Note that in Fig. 4.2(b), vectors **A′** and **B′** have the same magnitude but they are not equal because they have different directions. Even if we shift **B′** parallel to itself so that its tail Q′ coincides with the tail O′ of **A′**, the tip S′ of **B′** does not coincide with the tip P′ of **A′**.

4.3 MULTIPLICATION OF VECTORS BY REAL NUMBERS

Multiplying a vector **A** with a positive number λ gives a vector whose magnitude is changed by the factor λ but the direction is the same as that of **A** :

$$|\lambda \mathbf{A}| = \lambda\,|\mathbf{A}| \quad \text{if } \lambda > 0.$$

For example, if **A** is multiplied by 2, the resultant vector 2**A** is in the same direction as **A** and has a magnitude twice of |**A**| as shown in Fig. 4.3(a).

Multiplying a vector **A** by a negative number λ gives a vector λ**A** whose direction is opposite to the direction of **A** and whose magnitude is –λ times |**A**|.

Multiplying a given vector **A** by negative numbers, say –1 and –1.5, gives vectors as shown in Fig 4.3(b).

Fig. 4.3 (a) *Vector* **A** *and the resultant vector after multiplying* **A** *by a positive number 2.* (b) *Vector* **A** *and resultant vectors after multiplying it by a negative number –1 and –1.5.*

The factor λ by which a vector **A** is multiplied could be a scalar having its own physical dimension. Then, the dimension of λ **A** is the product of the dimensions of λ and **A**. For example, if we multiply a constant velocity vector by duration (of time), we get a displacement vector.

4.4 ADDITION AND SUBTRACTION OF VECTORS — GRAPHICAL METHOD

As mentioned in section 4.2, vectors, by definition, obey the triangle law or equivalently, the parallelogram law of addition. We shall now describe this law of addition using the graphical method. Let us consider two vectors **A** and **B** that lie in a plane as shown in Fig. 4.4(a). The lengths of the line segments representing these vectors are proportional to the magnitude of the vectors. To find the sum **A** + **B**, we place vector **B** so that its tail is at the head of the vector **A**, as in Fig. 4.4(b). Then, we join the tail of **A** to the head of **B**. This line OQ represents a vector **R,** that is, the sum of the vectors **A** and **B**. Since, in this procedure of vector addition, vectors are

Fig. 4.4 (a) *Vectors* **A** *and* **B**. (b) *Vectors* **A** *and* **B** *added graphically.* (c) *Vectors* **B** *and* **A** *added graphically.* (d) *Illustrating the associative law of vector addition.*

arranged head to tail, this graphical method is called the **head-to-tail method**. The two vectors and their resultant form three sides of a triangle, so this method is also known as **triangle method of vector addition**. If we find the resultant of **B** + **A** as in Fig. 4.4(c), the same vector **R** is obtained. Thus, vector addition is **commutative**:

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A} \qquad (4.1)$$

The addition of vectors also obeys the associative law as illustrated in Fig. 4.4(d). The result of adding vectors **A** and **B** first and then adding vector **C** is the same as the result of adding **B** and **C** first and then adding vector **A** :

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C}) \qquad (4.2)$$

What is the result of adding two equal and opposite vectors ? Consider two vectors **A** and –**A** shown in Fig. 4.3(b). Their sum is **A** + (–**A**). Since the magnitudes of the two vectors are the same, but the directions are opposite, the resultant vector has zero magnitude and is represented by **0** called a **null vector** or a **zero vector** :

$$\mathbf{A} - \mathbf{A} = \mathbf{0} \qquad |\mathbf{0}| = 0 \qquad (4.3)$$

Since the magnitude of a null vector is zero, its direction cannot be specified.

The null vector also results when we multiply a vector **A** by the number zero. The main properties of **0** are :

$$\mathbf{A} + \mathbf{0} = \mathbf{A}$$
$$\lambda\,\mathbf{0} = \mathbf{0}$$
$$0\,\mathbf{A} = \mathbf{0} \qquad (4.4)$$

What is the physical meaning of a zero vector? Consider the position and displacement vectors in a plane as shown in Fig. 4.1(a). Now suppose that an object which is at P at time *t*, moves to P′ and then comes back to P. Then, what is its displacement? Since the initial and final positions coincide, the displacement is a "null vector".

Subtraction of vectors can be defined in terms of addition of vectors. We define the difference of two vectors **A** and **B** as the sum of two vectors **A** and –**B** :

$$\mathbf{A} - \mathbf{B} = \mathbf{A} + (-\mathbf{B}) \qquad (4.5)$$

It is shown in Fig 4.5. The vector –**B** is added to vector **A** to get $\mathbf{R}_2 = (\mathbf{A} - \mathbf{B})$. The vector $\mathbf{R}_1 = \mathbf{A} + \mathbf{B}$ is also shown in the same figure for comparison. We can also use the **parallelogram method** to find the sum of two vectors. Suppose we have two vectors **A** and **B**. To add these vectors, we bring their tails to a common origin O as shown in Fig. 4.6(a). Then we draw a line from the head of **A** parallel to **B** and another line from the head of **B** parallel to **A** to complete a parallelogram OQSP. Now we join the point of the intersection of these two lines to the origin O. The resultant vector **R** is directed from the common origin O along the diagonal (OS) of the parallelogram [Fig. 4.6(b)]. In Fig.4.6(c), the triangle law is used to obtain the resultant of **A** and **B** and we see that the two methods yield the same result. Thus, the two methods are equivalent.

Fig. 4.5 (a) *Two vectors* **A** *and* **B**, – **B** *is also shown.* (b) *Subtracting vector* **B** *from vector* **A** – *the result is* $\mathbf{R}_2$. *For comparison, addition of vectors* **A** *and* **B**, *i.e.* $\mathbf{R}_1$ *is also shown.*

Fig. 4.6 (a) *Two vectors* **A** *and* **B** *with their tails brought to a common origin.* (b) *The sum* **A** + **B** *obtained using the parallelogram method.* (c) *The parallelogram method of vector addition is equivalent to the triangle method.*

▶ ***Example 4.1*** Rain is falling vertically with a speed of 35 m s^{-1}. Winds starts blowing after sometime with a speed of 12 m s^{-1} in east to west direction. In which direction should a boy waiting at a bus stop hold his umbrella ?

Fig. 4.7

Answer The velocity of the rain and the wind are represented by the vectors $\mathbf{v}_r$ and $\mathbf{v}_w$ in Fig. 4.7 and are in the direction specified by the problem. Using the rule of vector addition, we see that the resultant of $\mathbf{v}_r$ and $\mathbf{v}_w$ is **R** as shown in the figure. The magnitude of **R** is

$$R = \sqrt{v_r^2 + v_w^2} = \sqrt{35^2 + 12^2} \text{ m s}^{-1} = 37 \text{ m s}^{-1}$$

The direction θ that R makes with the vertical is given by

$$\tan\theta = \frac{v_w}{v_r} = \frac{12}{35} = 0.343$$

Or, $\theta = \tan^{-1}(0.343) = 19°$

Therefore, the boy should hold his umbrella in the vertical plane at an angle of about 19° with the vertical towards the east. ◀

4.5 RESOLUTION OF VECTORS

Let **a** and **b** be any two non-zero vectors in a plane with different directions and let **A** be another vector in the same plane(Fig. 4.8). **A** can be expressed as a sum of two vectors – one obtained by multiplying **a** by a real number and the other obtained by multiplying **b** by another real number. To see this, let O and P be the tail and head of the vector **A**. Then, through O, draw a straight line parallel to **a**, and through P, a straight line parallel to **b**. Let them intersect at Q. Then, we have

$$\mathbf{A} = \mathbf{OP} = \mathbf{OQ} + \mathbf{QP} \qquad (4.6)$$

But since **OQ** is parallel to **a**, and **QP** is parallel to **b**, we can write :

$$\mathbf{OQ} = \lambda\,\mathbf{a}, \text{ and } \mathbf{QP} = \mu\,\mathbf{b} \qquad (4.7)$$

where λ and μ are real numbers.

Therefore, $\mathbf{A} = \lambda\,\mathbf{a} + \mu\,\mathbf{b}$ (4.8)

Fig. 4.8 (a) *Two non-colinear vectors* **a** *and* **b**. (b) *Resolving a vector* **A** *in terms of vectors* **a** *and* **b**.

We say that **A** has been resolved into two component vectors λ **a** and μ **b** along **a** and **b** respectively. Using this method one can resolve

a given vector into two component vectors along a set of two vectors – all the three lie in the same plane. It is convenient to resolve a general vector along the axes of a rectangular coordinate system using vectors of unit magnitude. These are called unit vectors that we discuss now.

Unit vectors: A unit vector is a vector of unit magnitude and points in a particular direction. It has no dimension and unit. It is used to specify a direction only. Unit vectors along the *x*-, *y*- and *z*-axes of a rectangular coordinate system are denoted by $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$, respectively, as shown in Fig. 4.9(a).

Since these are unit vectors, we have

$$|\hat{\mathbf{i}}| = |\hat{\mathbf{j}}| = |\hat{\mathbf{k}}| = 1 \qquad (4.9)$$

These unit vectors are perpendicular to each other. In this text, they are printed in bold face with a cap (^) to distinguish them from other vectors. Since we are dealing with motion in two dimensions in this chapter, we require use of only two unit vectors. If we multiply a unit vector, say $\mathbf{n}$ by a scalar, the result is a vector $\boldsymbol{\lambda} = \lambda\,\mathbf{n}$. In general, a vector **A** can be written as

$$\mathbf{A} = |\mathbf{A}|\,\mathbf{n} \qquad (4.10)$$

where $\mathbf{n}$ is a unit vector along **A**.

We can now resolve a vector **A** in terms of component vectors that lie along unit vectors $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$. Consider a vector **A** that lies in *x*-*y* plane as shown in Fig. 4.9(b). We draw lines from the head of **A** perpendicular to the coordinate axes as in Fig. 4.9(b), and get vectors $\mathbf{A}_1$ and $\mathbf{A}_2$ such that $\mathbf{A}_1 + \mathbf{A}_2 = \mathbf{A}$. Since $\mathbf{A}_1$ is parallel to $\hat{\mathbf{i}}$ and $\mathbf{A}_2$ is parallel to $\hat{\mathbf{j}}$, we have :

$$\mathbf{A}_1 = A_x\,\hat{\mathbf{i}},\quad \mathbf{A}_2 = A_y\,\hat{\mathbf{j}} \qquad (4.11)$$

where A_x and A_y are real numbers.

Thus, $$\mathbf{A} = A_x\,\hat{\mathbf{i}} + A_y\,\hat{\mathbf{j}} \qquad (4.12)$$

This is represented in Fig. 4.9(c). The quantities A_x and A_y are called *x*-, and *y*- components of the vector **A**. Note that A_x is itself not a vector, but $A_x\,\hat{\mathbf{i}}$ is a vector, and so is $A_y\,\hat{\mathbf{j}}$. Using simple trigonometry, we can express A_x and A_y in terms of the magnitude of **A** and the angle θ it makes with the *x*-axis :

$$A_x = A\cos\theta$$
$$A_y = A\sin\theta \qquad (4.13)$$

As is clear from Eq. (4.13), a component of a vector can be positive, negative or zero depending on the value of θ.

Now, we have two ways to specify a vector **A** in a plane. It can be specified by :

(i) its magnitude *A* and the direction θ it makes with the *x*-axis; or

(ii) its components A_x and A_y

If A and θ are given, A_x and A_y can be obtained using Eq. (4.13). If A_x and A_y are given, *A* and θ can be obtained as follows :

$$A_x^2 + A_y^2 = A^2\cos^2\theta + A^2\sin^2\theta$$
$$= A^2$$

Or, $$A = \sqrt{A_x^2 + A_y^2} \qquad (4.14)$$

And $$\tan\theta = \frac{A_y}{A_x},\quad \theta = \tan^{-1}\frac{A_y}{A_x} \qquad (4.15)$$

Fig. 4.9 (a) *Unit vectors* $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$ *lie along the x-, y-, and z-axes.* (b) *A vector* **A** *is resolved into its components* $\mathbf{A}_1$ *and* $\mathbf{A}_2$ *along x-, and y- axes.* (c) $\mathbf{A}_1$ *and* $\mathbf{A}_2$ *expressed in terms of* $\hat{\mathbf{i}}$ *and* $\hat{\mathbf{j}}$.

So far we have considered a vector lying in an x-y plane. The same procedure can be used to resolve a general vector **A** into three components along x-, y-, and z-axes in three dimensions. If α, β, and γ are the angles* between **A** and the x-, y-, and z-axes, respectively Fig. 4.9(d), we have

Fig. 4.9 (d) *A vector* **A** *resolved into components along x-, y-, and z-axes*

$$A_x = A\cos\alpha,\ A_y = A\cos\beta,\ A_z = A\cos\gamma \quad (4.16a)$$

In general, we have

$$\mathbf{A} = A_x\hat{\mathbf{i}} + A_y\hat{\mathbf{j}} + A_z\hat{\mathbf{k}} \quad (4.16b)$$

The magnitude of vector **A** is

$$A = \sqrt{A_x^2 + A_y^2 + A_z^2} \quad (4.16c)$$

A position vector **r** can be expressed as

$$\mathbf{r} = x\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\hat{\mathbf{k}} \quad (4.17)$$

where x, y, and z are the components of **r** along x-, y-, z-axes, respectively.

4.6 VECTOR ADDITION – ANALYTICAL METHOD

Although the graphical method of adding vectors helps us in visualising the vectors and the resultant vector, it is sometimes tedious and has limited accuracy. It is much easier to add vectors by combining their respective components. Consider two vectors **A** and **B** in x-y plane with components A_x, A_y and B_x, B_y :

$$\mathbf{A} = A_x\hat{\mathbf{i}} + A_y\hat{\mathbf{j}} \quad (4.18)$$

$$\mathbf{B} = B_x\hat{\mathbf{i}} + B_y\hat{\mathbf{j}}$$

Let **R** be their sum. We have

$$\mathbf{R} = \mathbf{A} + \mathbf{B}$$

$$= \left(A_x\hat{\mathbf{i}} + A_y\hat{\mathbf{j}}\right) + \left(B_x\hat{\mathbf{i}} + B_y\hat{\mathbf{j}}\right) \quad (4.19a)$$

Since vectors obey the commutative and associative laws, we can arrange and regroup the vectors in Eq. (4.19a) as convenient to us :

$$\mathbf{R} = \left(A_x + B_x\right)\hat{\mathbf{i}} + \left(A_y + B_y\right)\hat{\mathbf{j}} \quad (4.19b)$$

Since $\mathbf{R} = R_x\hat{\mathbf{i}} + R_y\hat{\mathbf{j}}$ (4.20)

we have, $R_x = A_x + B_x,\ R_y = A_y + B_y$ (4.21)

Thus, each component of the resultant vector **R** is the sum of the corresponding components of **A** and **B**.

In three dimensions, we have

$$\mathbf{A} = A_x\hat{\mathbf{i}} + A_y\hat{\mathbf{j}} + A_z\hat{\mathbf{k}}$$

$$\mathbf{B} = B_x\hat{\mathbf{i}} + B_y\hat{\mathbf{j}} + B_z\hat{\mathbf{k}}$$

$$\mathbf{R} = \mathbf{A} + \mathbf{B} = R_x\hat{\mathbf{i}} + R_y\hat{\mathbf{j}} + R_z\hat{\mathbf{k}}$$

with $R_x = A_x + B_x$

$$R_y = A_y + B_y$$

$$R_z = A_z + B_z \quad (4.22)$$

This method can be extended to addition and subtraction of any number of vectors. For example, if vectors **a**, **b** and **c** are given as

$$\mathbf{a} = a_x\hat{\mathbf{i}} + a_y\hat{\mathbf{j}} + a_z\hat{\mathbf{k}}$$

$$\mathbf{b} = b_x\hat{\mathbf{i}} + b_y\hat{\mathbf{j}} + b_z\hat{\mathbf{k}}$$

$$\mathbf{c} = c_x\hat{\mathbf{i}} + c_y\hat{\mathbf{j}} + c_z\hat{\mathbf{k}} \quad (4.23a)$$

then, a vector **T** = **a** + **b** – **c** has components :

$$T_x = a_x + b_x - c_x$$

$$T_y = a_y + b_y - c_y \quad (4.23b)$$

$$T_z = a_z + b_z - c_z.$$

▶ ***Example 4.2*** Find the magnitude and direction of the resultant of two vectors **A** and **B** in terms of their magnitudes and angle θ between them.

* *Note that angles α, β, and γ are angles in space. They are between pairs of lines, which are not coplanar.*

Fig. 4.10

Answer Let **OP** and **OQ** represent the two vectors **A** and **B** making an angle θ (Fig. 4.10). Then, using the parallelogram method of vector addition, **OS** represents the resultant vector **R** :

$$\mathbf{R} = \mathbf{A} + \mathbf{B}$$

SN is normal to *OP* and *PM* is normal to *OS*.

From the geometry of the figure,

$$OS^2 = ON^2 + SN^2$$

but $ON = OP + PN = A + B\cos\theta$

$SN = B\sin\theta$

$$OS^2 = (A + B\cos\theta)^2 + (B\sin\theta)^2$$

or, $R^2 = A^2 + B^2 + 2AB\cos\theta$

$$R = \sqrt{A^2 + B^2 + 2AB\cos\theta} \qquad (4.24a)$$

In Δ OSN, $SN = OS\sin\alpha = R\sin\alpha$, and
in Δ PSN, $SN = PS\sin\theta = B\sin\theta$

Therefore, $R\sin\alpha = B\sin\theta$

or, $$\frac{R}{\sin\theta} = \frac{B}{\sin\alpha} \qquad (4.24b)$$

Similarly,

$$PM = A\sin\alpha = B\sin\beta$$

or, $$\frac{A}{\sin\beta} = \frac{B}{\sin\alpha} \qquad (4.24c)$$

Combining Eqs. (4.24b) and (4.24c), we get

$$\frac{R}{\sin\theta} = \frac{A}{\sin\beta} = \frac{B}{\sin\alpha} \qquad (4.24d)$$

Using Eq. (4.24d), we get:

$$\sin\alpha = \frac{B}{R}\sin\theta \qquad (4.24e)$$

where R is given by Eq. (4.24a).

or, $$\tan\alpha = \frac{SN}{OP + PN} = \frac{B\sin\theta}{A + B\cos\theta} \qquad (4.24f)$$

Equation (4.24a) gives the magnitude of the resultant and Eqs. (4.24e) and (4.24f) its direction. Equation (4.24a) is known as the **Law of cosines** and Eq. (4.24d) as the **Law of sines**. ◀

▶ ***Example 4.3*** A motorboat is racing towards north at 25 km/h and the water current in that region is 10 km/h in the direction of 60° east of south. Find the resultant velocity of the boat.

Answer The vector $\mathbf{v}_b$ representing the velocity of the motorboat and the vector $\mathbf{v}_c$ representing the water current are shown in Fig. 4.11 in directions specified by the problem. Using the parallelogram method of addition, the resultant **R** is obtained in the direction shown in the figure.

Fig. 4.11

We can obtain the magnitude of **R** using the Law of cosine :

$$R = \sqrt{v_b^2 + v_c^2 + 2v_b v_c \cos 120^\circ}$$

$$= \sqrt{25^2 + 10^2 + 2\times 25\times 10(-1/2)} \cong 22 \text{ km/h}$$

To obtain the direction, we apply the Law of sines

$$\frac{R}{\sin\theta} = \frac{v_c}{\sin\phi} \quad \text{or, } \sin\phi = \frac{v_c}{R}\sin\theta$$

$$= \frac{10\times\sin 120^\circ}{21.8} = \frac{10\sqrt{3}}{2\times 21.8} \cong 0.397$$

$$\phi \cong 23.4^\circ$$ ◀

4.7 MOTION IN A PLANE

In this section we shall see how to describe motion in two dimensions using vectors.

4.7.1 Position Vector and Displacement

The position vector **r** of a particle P located in a plane with reference to the origin of an *x-y* reference frame (Fig. 4.12) is given by

$$\mathbf{r} = x\,\hat{\mathbf{i}} + y\,\hat{\mathbf{j}}$$

where *x* and *y* are components of **r** along *x*-, and *y*- axes or simply they are the coordinates of the object.

Fig. 4.12 (a) *Position vector* **r**. (b) *Displacement* Δ**r** *and average velocity* **v** *of a particle.*

Suppose a particle moves along the curve shown by the thick line and is at P at time *t* and P′ at time *t*′ [Fig. 4.12(b)]. Then, the displacement is :

$$\Delta\mathbf{r} = \mathbf{r}' - \mathbf{r} \qquad (4.25)$$

and is directed from P to P′ .

We can write Eq. (4.25) in a component form:

$$\Delta\mathbf{r} = \left(x'\,\hat{\mathbf{i}} + y'\,\hat{\mathbf{j}}\right) - \left(x\,\hat{\mathbf{i}} + y\,\hat{\mathbf{j}}\right)$$

$$= \hat{\mathbf{i}}\Delta x + \hat{\mathbf{j}}\Delta y$$

where $\Delta x = x' - x,\ \Delta y = y' - y$ (4.26)

Velocity

The average velocity $(\bar{\mathbf{v}})$ of an object is the ratio of the displacement and the corresponding time interval :

$$\bar{\mathbf{v}} = \frac{\Delta\mathbf{r}}{\Delta t} = \frac{\Delta x\,\hat{\mathbf{i}} + \Delta y\,\hat{\mathbf{j}}}{\Delta t} = \hat{\mathbf{i}}\frac{\Delta x}{\Delta t} + \hat{\mathbf{j}}\frac{\Delta y}{\Delta t} \qquad (4.27)$$

Or, $\bar{\mathbf{v}} = \bar{v}_x\,\hat{\mathbf{i}} + \bar{v}_y\,\hat{\mathbf{j}}$

Since $\bar{\mathbf{v}} = \dfrac{\Delta\mathbf{r}}{\Delta t}$, the direction of the average velocity is the same as that of Δ**r** (Fig. 4.12). The **velocity** (instantaneous velocity) is given by the limiting value of the average velocity as the time interval approaches zero :

$$\mathbf{v} = \lim_{\Delta t \to 0} \frac{\Delta\mathbf{r}}{\Delta t} = \frac{d\mathbf{r}}{dt} \qquad (4.28)$$

The meaning of the limiting process can be easily understood with the help of Fig 4.13(a) to (d). In these figures, the thick line represents the path of an object, which is at P at time *t*. P_1, P_2 and P_3 represent the positions of the object after times Δt_1, Δt_2, and Δt_3. $\Delta\mathbf{r}_1$, $\Delta\mathbf{r}_2$, and $\Delta\mathbf{r}_3$ are the displacements of the object in times Δt_1, Δt_2, and

Fig. 4.13 *As the time interval* Δ*t approaches zero, the average velocity approaches the velocity* **v**. *The direction of* $\bar{\mathbf{v}}$ *is parallel to the line tangent to the path.*

Δt_3, respectively. The direction of the average velocity $\bar{\mathbf{v}}$ is shown in figures (a), (b) and (c) for three decreasing values of Δt, i.e. Δt_1, Δt_2, and Δt_3, ($\Delta t_1 > \Delta t_2 > \Delta t_3$). As $\Delta t \to 0$, $\Delta \mathbf{r} \to 0$ and is along the tangent to the path [Fig. 4.13(d)]. Therefore, **the direction of velocity at any point on the path of an object is tangential to the path at that point and is in the direction of motion**.

We can express **v** in a component form :

$$\mathbf{v} = \frac{d\mathbf{r}}{dt}$$

$$= \lim_{\Delta t \to 0}\left(\frac{\Delta x}{\Delta t}\hat{\mathbf{i}} + \frac{\Delta y}{\Delta t}\hat{\mathbf{j}}\right) \qquad (4.29)$$

$$= \hat{\mathbf{i}}\lim_{\Delta t \to 0}\frac{\Delta x}{\Delta t} + \hat{\mathbf{j}}\lim_{\Delta t \to 0}\frac{\Delta y}{\Delta t}$$

Or, $$\mathbf{v} = \hat{\mathbf{i}}\frac{dx}{dt} + \hat{\mathbf{j}}\frac{dy}{dt} = v_x\hat{\mathbf{i}} + v_y\hat{\mathbf{j}}.$$

where $$v_x = \frac{dx}{dt}, v_y = \frac{dy}{dt} \qquad (4.30a)$$

So, if the expressions for the coordinates x and y are known as functions of time, we can use these equations to find v_x and v_y.

The magnitude of **v** is then

$$v = \sqrt{v_x^2 + v_y^2} \qquad (4.30b)$$

and the direction of **v** is given by the angle θ :

$$\tan\theta = \frac{v_y}{v_x}, \quad \theta = \tan^{-1}\left(\frac{v_y}{v_x}\right) \qquad (4.30c)$$

v_x, v_y and angle θ are shown in Fig. 4.14 for a velocity vector **v**.

*Fig. 4.14 The components v_x and v_y of velocity **v** and the angle θ it makes with x-axis. Note that $v_x = v\cos\theta$, $v_y = v\sin\theta$.*

Acceleration

The **average acceleration** **a** of an object for a time interval Δt moving in x-y plane is the change in velocity divided by the time interval :

$$\bar{\mathbf{a}} = \frac{\Delta\mathbf{v}}{\Delta t} = \frac{\Delta\left(v_x\hat{\mathbf{i}} + v_y\hat{\mathbf{j}}\right)}{\Delta t} = \frac{\Delta v_x}{\Delta t}\hat{\mathbf{i}} + \frac{\Delta v_y}{\Delta t}\hat{\mathbf{j}} \qquad (4.31a)$$

Or, $$\bar{\mathbf{a}} = a_x\hat{\mathbf{i}} + a_y\hat{\mathbf{j}}. \qquad (4.31b)$$

The **acceleration** (instantaneous acceleration) is the limiting value of the average acceleration as the time interval approaches zero :

$$\mathbf{a} = \lim_{\Delta t \to 0}\frac{\Delta\mathbf{v}}{\Delta t} \qquad (4.32a)$$

Since $\Delta\mathbf{v} = \Delta v_x\hat{\mathbf{i}} + \Delta v_y\hat{\mathbf{j}}$, we have

$$\mathbf{a} = \hat{\mathbf{i}}\lim_{\Delta t \to 0}\frac{\Delta v_x}{\Delta t} + \hat{\mathbf{j}}\lim_{\Delta t \to 0}\frac{\Delta v_y}{\Delta t}$$

Or, $$\mathbf{a} = a_x\hat{\mathbf{i}} + a_y\hat{\mathbf{j}} \qquad (4.32b)$$

where, $$a_x = \frac{dv_x}{dt}, \; a_y = \frac{dv_y}{dt} \qquad (4.32c)^*$$

As in the case of velocity, we can understand graphically the limiting process used in defining acceleration on a graph showing the path of the object's motion. This is shown in Figs. 4.15(a) to (d). P represents the position of the object at time t and P_1, P_2, P_3 positions after time Δt_1, Δt_2, Δt_3, respectively ($\Delta t_1 > \Delta t_2 > \Delta t_3$). The velocity vectors at points P, P_1, P_2, P_3 are also shown in Figs. 4.15 (a), (b) and (c). In each case of Δt, $\Delta\mathbf{v}$ is obtained using the triangle law of vector addition. By definition, the direction of average acceleration is the same as that of $\Delta\mathbf{v}$. We see that as Δt decreases, the direction of $\Delta\mathbf{v}$ changes and consequently, the direction of the acceleration changes. Finally, in the limit $\Delta t \to 0$

* *In terms of x and y, a_x and a_y can be expressed as*

$$a_x = \frac{d}{dt}\left(\frac{dx}{dt}\right) = \frac{d^2x}{dt^2}, \quad a_y = \frac{d}{dt}\left(\frac{dy}{dt}\right) = \frac{d^2y}{dt^2}$$

Fig. 4.15 The average acceleration for three time intervals (a) Δt_1, (b) Δt_2, *and* (c) Δt_3, ($\Delta t_1 > \Delta t_2 > \Delta t_3$). (d) *In the limit* $\Delta t \to 0$, *the average acceleration becomes the acceleration.*

Fig. 4.15(d), the average acceleration becomes the instantaneous acceleration and has the direction as shown.

Note that in one dimension, the velocity and the acceleration of an object are always along the same straight line (either in the same direction or in the opposite direction). However, for motion in two or three dimensions, velocity and acceleration vectors may have any angle between 0° and 180° between them.

▶ ***Example 4.4*** The position of a particle is given by

$$\mathbf{r} = 3.0t\,\hat{\mathbf{i}} + 2.0t^2\hat{\mathbf{j}} + 5.0\,\hat{\mathbf{k}}$$

where t is in seconds and the coefficients have the proper units for $\mathbf{r}$ to be in metres. (a) Find $\mathbf{v}(t)$ and $\mathbf{a}(t)$ of the particle. (b) Find the magnitude and direction of $\mathbf{v}(t)$ at $t = 1.0$ s.

Answer

$$\mathbf{v}(t) = \frac{d\mathbf{r}}{dt} = \frac{d}{dt}\left(3.0\,t\,\hat{\mathbf{i}} + 2.0t^2\,\hat{\mathbf{j}} + 5.0\,\hat{\mathbf{k}}\right)$$

$$= 3.0\hat{\mathbf{i}} + 4.0t\hat{\mathbf{j}}$$

$$\mathbf{a}\,(t) = \frac{d\mathbf{v}}{dt} = +4.0\hat{\mathbf{j}}$$

$a = 4.0$ m s^{-2} along y- direction

At $t = 1.0$ s, $\mathbf{v} = 3.0\hat{\mathbf{i}} + 4.0\hat{\mathbf{j}}$

It's magnitude is $v = \sqrt{3^2 + 4^2} = 5.0$ m s^{-1}

and direction is

$$\theta = \tan^{-1}\left(\frac{v_y}{v_x}\right) = \tan^{-1}\left(\frac{4}{3}\right) \cong 53^\circ \text{ with } x\text{-axis.}$$

◀

4.8 MOTION IN A PLANE WITH CONSTANT ACCELERATION

Suppose that an object is moving in x-y plane and its acceleration $\mathbf{a}$ is constant. Over an interval of time, the average acceleration will equal this constant value. Now, let the velocity of the object be $\mathbf{v}_0$ at time $t = 0$ and $\mathbf{v}$ at time t. Then, by definition

$$\mathbf{a} = \frac{\mathbf{v} - \mathbf{v_0}}{t - 0} = \frac{\mathbf{v} - \mathbf{v_0}}{t}$$

Or, $\mathbf{v} = \mathbf{v_0} + \mathbf{a}t$ (4.33a)

In terms of components :

$$v_x = v_{ox} + a_x t$$

$$v_y = v_{oy} + a_y t \qquad (4.33b)$$

Let us now find how the position $\mathbf{r}$ changes with time. We follow the method used in the one-dimensional case. Let $\mathbf{r}_o$ and $\mathbf{r}$ be the position vectors of the particle at time 0 and t and let the velocities at these instants be $\mathbf{v}_o$ and $\mathbf{v}$. Then, over this time interval t, the average velocity is $(\mathbf{v}_o + \mathbf{v})/2$. The displacement is the average velocity multiplied by the time interval :

$$\mathbf{r} - \mathbf{r_0} = \left(\frac{\mathbf{v} + \mathbf{v_0}}{2}\right)t = \left(\frac{(\mathbf{v_0} + \mathbf{a}t) + \mathbf{v_0}}{2}\right)t$$

$$= \mathbf{v_0}t + \frac{1}{2}\mathbf{a}t^2$$

Or, $$\mathbf{r} = \mathbf{r_0} + \mathbf{v_0}t + \frac{1}{2}\mathbf{a}t^2 \qquad (4.34a)$$

It can be easily verified that the derivative of Eq. (4.34a), i.e. $\frac{d\mathbf{r}}{dt}$ gives Eq.(4.33a) and it also satisfies the condition that at t=0, $\mathbf{r} = \mathbf{r_o}$. Equation (4.34a) can be written in component form as

$$x = x_0 + v_{ox}t + \frac{1}{2}a_x t^2$$

$$y = y_0 + v_{oy}t + \frac{1}{2}a_y t^2 \qquad (4.34b)$$

One immediate interpretation of Eq.(4.34b) is that the motions in x- and y-directions can be treated independently of each other. That is, **motion in a plane (two-dimensions) can be treated as two separate simultaneous one-dimensional motions with constant acceleration along two perpendicular directions**. This is an important result and is useful in analysing motion of objects in two dimensions. A similar result holds for three dimensions. The choice of perpendicular directions is convenient in many physical situations, as we shall see in section 4.10 for projectile motion.

▶ ***Example 4.5*** A particle starts from origin at $t = 0$ with a velocity $5.0\,\hat{\mathbf{i}}$ m/s and moves in x-y plane under action of a force which produces a constant acceleration of $(3.0\hat{\mathbf{i}} + 2.0\hat{\mathbf{j}})$ m/s². (a) What is the y-coordinate of the particle at the instant its x-coordinate is 84 m ? (b) What is the speed of the particle at this time ?

Answer The position of the particle is given by

$$\mathbf{r}(t) = \mathbf{v_0}t + \frac{1}{2}\mathbf{a}t^2$$

$$= 5.0\hat{\mathbf{i}}\,t + (1/2)(3.0\hat{\mathbf{i}} + 2.0\hat{\mathbf{j}})\,t^2$$

$$= (5.0t + 1.5t^2)\hat{\mathbf{i}} + 1.0t^2\hat{\mathbf{j}}$$

Therefore, $$x(t) = 5.0t + 1.5t^2$$

$$y(t) = +1.0t^2$$

Given $x(t)$ = 84 m, t = ?

$$5.0\,t + 1.5\,t^2 = 84 \Rightarrow t = 6 \text{ s}$$

At t = 6 s, y = 1.0 (6)² = 36.0 m

Now, the velocity $\mathbf{v} = \frac{d\mathbf{r}}{dt} = (5.0 + 3.0t)\hat{\mathbf{i}} + 2.0t\,\hat{\mathbf{j}}$

At t = 6 s, $\mathbf{v} = 23.0\hat{\mathbf{i}} + 12.0\hat{\mathbf{j}}$

speed $= |\mathbf{v}| = \sqrt{23^2 + 12^2} \cong 26 \text{ m s}^{-1}$. ◀

4.9 RELATIVE VELOCITY IN TWO DIMENSIONS

The concept of relative velocity, introduced in section 3.7 for motion along a straight line, can be easily extended to include motion in a plane or in three dimensions. Suppose that two objects A and B are moving with velocities $\mathbf{v_A}$ and $\mathbf{v_B}$ (each with respect to some common frame of reference, say ground.). Then, velocity of object A **relative to that of B is** :

$$\mathbf{v}_{AB} = \mathbf{v}_A - \mathbf{v}_B \qquad (4.35a)$$

and similarly, the velocity of object B *relative to that of* A is :

$$\mathbf{v}_{BA} = \mathbf{v}_B - \mathbf{v}_A$$

Therefore, $$\mathbf{v}_{AB} = -\mathbf{v}_{BA} \qquad (4.35b)$$

and, $$|\mathbf{v}_{AB}| = |\mathbf{v}_{BA}| \qquad (4.35c)$$

▶ ***Example 4.6*** Rain is falling vertically with a speed of 35 m s⁻¹. A woman rides a bicycle with a speed of 12 m s⁻¹ in east to west direction. What is the direction in which she should hold her umbrella ?

Answer In Fig. 4.16 $\mathbf{v}_r$ represents the velocity of rain and $\mathbf{v}_b$, the velocity of the bicycle, the woman is riding. Both these velocities are with respect to the ground. Since the woman is riding a bicycle, the velocity of rain as experienced by

Fig. 4.16

her is the velocity of rain relative to the velocity of the bicycle she is riding. That is $\mathbf{v}_{rb} = \mathbf{v}_r - \mathbf{v}_b$

This relative velocity vector as shown in Fig. 4.16 makes an angle θ with the vertical. It is given by

$$\tan\theta = \frac{v_b}{v_r} = \frac{12}{35} = 0.343$$

Or, $\theta \cong 19°$

Therefore, the woman should hold her umbrella at an angle of about 19° with the vertical towards the west.

Note carefully the difference between this Example and the Example 4.1. In Example 4.1, the boy experiences the resultant (vector sum) of two velocities while in this example, the woman experiences the velocity of rain relative to the bicycle (the vector difference of the two velocities). ◀

4.10 PROJECTILE MOTION

As an application of the ideas developed in the previous sections, we consider the motion of a projectile. An object that is in flight after being thrown or projected is called a **projectile**. Such a projectile might be a football, a cricket ball, a baseball or any other object. The motion of a projectile may be thought of as the result of two separate, simultaneously occurring components of motions. One component is along a horizontal direction without any acceleration and the other along the vertical direction with constant acceleration due to the force of gravity. It was Galileo who first stated this independency of the horizontal and the vertical components of projectile motion in his **Dialogue on the great world systems** (1632).

In our discussion, we shall assume that the air resistance has negligible effect on the motion of the projectile. Suppose that the projectile is launched with velocity $\mathbf{v}_o$ that makes an angle θ_o with the x-axis as shown in Fig. 4.17.

After the object has been projected, the acceleration acting on it is that due to gravity which is directed vertically downward:

$$\mathbf{a} = -g\,\hat{\mathbf{j}}$$

Or, $a_x = 0,\ a_y = -g$ (4.36)

The components of initial velocity $\mathbf{v}_o$ are :

$$v_{ox} = v_o \cos\theta_o$$
$$v_{oy} = v_o \sin\theta_o \qquad (4.37)$$

Fig 4.17 Motion of an object projected with velocity $\mathbf{v}_o$ at angle θ_0.

If we take the initial position to be the origin of the reference frame as shown in Fig. 4.17, we have :

$$x_o = 0,\ y_o = 0$$

Then, Eq.(4.47b) becomes :

$$x = v_{ox}\,t = (v_o \cos\theta_o)\,t$$

and $y = (v_o \sin\theta_o)\,t - (\tfrac{1}{2})\,g\,t^2$ (4.38)

The components of velocity at time t can be obtained using Eq.(4.33b) :

$$v_x = v_{ox} = v_o \cos\theta_o$$
$$v_y = v_o \sin\theta_o - g\,t \qquad (4.39)$$

Equation (4.38) gives the x-, and y-coordinates of the position of a projectile at time t in terms of two parameters — initial speed v_o and projection angle θ_o. Notice that the choice of mutually perpendicular x-, and y-directions for the analysis of the projectile motion has resulted in a simplification. One of the components of velocity, i.e. x-component remains constant throughout the motion and only the y- component changes, like an object in free fall in vertical direction. This is shown graphically at few instants in Fig. 4.18. Note that at the point of maximum height, $v_y = 0$ and therefore,

$$\theta = \tan^{-1}\frac{v_y}{v_x} = 0$$

Equation of path of a projectile

What is the shape of the path followed by the projectile? This can be seen by eliminating the time between the expressions for x and y as given in Eq. (4.38). We obtain:

$$y = (\tan\theta_o)x - \frac{g}{2\,(v_o\cos\theta_o)^2}x^2 \qquad (4.40)$$

Now, since g, θ_o and v_o are constants, Eq. (4.40) is of the form $y = a\,x + b\,x^2$, in which a and b are constants. This is the equation of a parabola, i.e. the path of the projectile is a parabola (Fig. 4.18).

Fig. 4.18 The path of a projectile is a parabola.

Time of maximum height

How much time does the projectile take to reach the maximum height ? Let this time be denoted by t_m. Since at this point, $v_y = 0$, we have from Eq. (4.39):

$$v_y = v_o \sin\theta_o - g\,t_m = 0$$

Or, $$t_m = v_o \sin\theta_o / g \qquad (4.41a)$$

The total time T_f during which the projectile is in flight can be obtained by putting $y = 0$ in Eq. (4.38). We get :

$$T_f = 2\,(v_o \sin\theta_o)/g \qquad (4.41b)$$

T_f is known as the **time of flight** of the projectile. We note that $T_f = 2\,t_m$, which is expected because of the symmetry of the parabolic path.

Maximum height of a projectile

The maximum height h_m reached by the projectile can be calculated by substituting $t = t_m$ in Eq. (4.38) :

$$y = h_m = (v_0\sin\theta_0)\left(\frac{v_0\sin\theta_0}{g}\right) - \frac{g}{2}\left(\frac{v_0\sin\theta_0}{g}\right)^2$$

Or, $$h_m = \frac{(v_0\sin\theta_0)^2}{2g} \qquad (4.42)$$

Horizontal range of a projectile

The horizontal distance travelled by a projectile from its initial position ($x = y = 0$) to the position where it passes $y = 0$ during its fall is called the **horizontal range**, R. It is the distance travelled during the time of flight T_f . Therefore, the range R is

$$R = (v_o \cos\theta_o)\,(T_f)$$
$$= (v_o \cos\theta_o)\,(2\,v_o \sin\theta_o)/g$$

Or, $$R = \frac{v_0^2 \sin 2\theta_0}{g} \qquad (4.43a)$$

Equation (4.43a) shows that for a given projection velocity v_o , R is maximum when sin $2\theta_0$ is maximum, i.e., when $\theta_0 = 45^0$.
The maximum horizontal range is, therefore,

$$R_m = \frac{v_0^2}{g} \qquad (4.43b)$$

▶ ***Example 4.7*** Galileo, in his book **Two new sciences,** stated that "for elevations which exceed or fall short of 45° by equal amounts, the ranges are equal". Prove this statement.

Answer For a projectile launched with velocity $\mathbf{v_o}$ at an angle θ_o , the range is given by

$$R = \frac{v_0^2 \sin 2\theta_0}{g}$$

Now, for angles, (45° + α) and (45° – α), $2\theta_o$ is (90° + 2α) and (90° – 2α) , respectively. The values of sin (90° + 2α) and sin (90° – 2α) are the same, equal to that of cos 2α. Therefore, ranges are equal for elevations which exceed or fall short of 45° by equal amounts α. ◀

▶ ***Example 4.8*** A hiker stands on the edge of a cliff 490 m above the ground and throws a stone horizontally with an initial speed of 15 m s^{-1}. Neglecting air resistance, find the time taken by the stone to reach the ground, and the speed with which it hits the ground. (Take g = 9.8 m s^{-2}).

Answer We choose the origin of the x-,and y-axis at the edge of the cliff and t = 0 s at the instant the stone is thrown. Choose the positive direction of x-axis to be along the initial velocity and the positive direction of y-axis to be the vertically upward direction. The x-, and y-components of the motion can be treated independently. The equations of motion are :

$$x(t) = x_o + v_{ox}\, t$$
$$y(t) = y_o + v_{oy}\, t + (1/2)\, a_y\, t^2$$

Here, $x_o = y_o = 0$, $v_{oy} = 0$, $a_y = -g = -9.8$ m s^{-2}, $v_{ox} = 15$ m s^{-1}.

The stone hits the ground when $y(t) = -490$ m.

$$-490 \text{ m} = -(1/2)(9.8)\, t^2.$$

This gives $t = 10$ s.

The velocity components are $v_x = v_{ox}$ and $v_y = v_{oy} - g\, t$

so that when the stone hits the ground :

$$v_{ox} = 15 \text{ m s}^{-1}$$
$$v_{oy} = 0 - 9.8 \times 10 = -98 \text{ m s}^{-1}$$

Therefore, the speed of the stone is

$$\sqrt{v_x^2 + v_y^2} = \sqrt{15^2 + 98^2} = 99 \text{ m s}^{-1}$$

◀

▶ ***Example 4.9*** A cricket ball is thrown at a speed of 28 m s^{-1} in a direction 30° above the horizontal. Calculate (a) the maximum height, (b) the time taken by the ball to return to the same level, and (c) the distance from the thrower to the point where the ball returns to the same level.

Answer (a) The maximum height is given by

$$h_m = \frac{(v_0 \sin\theta_o)^2}{2g} = \frac{(28 \sin 30°)^2}{2\ (9.8)} \text{ m}$$

$$= \frac{14 \times 14}{2 \times 9.8} = 10.0 \text{ m}$$

(b) The time taken to return to the same level is

$$T_f = (2\, v_o \sin\theta_o)/g = (2 \times 28 \times \sin 30°)/9.8$$
$$= 28/9.8 \text{ s} = 2.9 \text{ s}$$

(c) The distance from the thrower to the point where the ball returns to the same level is

$$R = \frac{(v_o^2 \sin 2\theta_o)}{g} = \frac{28 \times 28 \times \sin 60°}{9.8} = 69 \text{ m}$$

◀

Neglecting air resistance - what does the assumption really mean?

While treating the topic of projectile motion, we have stated that we assume that the air resistance has no effect on the motion of the projectile. You must understand what the statement really means. Friction, force due to viscosity, air resistance are all dissipative forces. In the presence of any of such forces opposing motion, any object will lose some part of its initial energy and consequently, momentum too. Thus, a projectile that traverses a parabolic path would certainly show deviation from its idealised trajectory in the presence of air resistance. It will not hit the ground with the same speed with which it was projected from it. In the absence of air resistance, the x-component of the velocity remains constant and it is only the y-component that undergoes a continuous change. However, in the presence of air resistance, both of these would get affected. That would mean that the range would be less than the one given by Eq. (4.43). Maximum height attained would also be less than that predicted by Eq. (4.42). Can you then, anticipate the change in the time of flight?

In order to avoid air resistance, we will have to perform the experiment in vacuum or under low pressure, which is not easy. When we use a phrase like 'neglect air resistance', we imply that the change in parameters such as range, height etc. is much smaller than their values without air resistance. The calculation without air resistance is much simpler than that with air resistance.

4.11 UNIFORM CIRCULAR MOTION

When an object follows a circular path at a constant speed, the motion of the object is called **uniform circular motion**. The word "uniform" refers to the speed, which is uniform (constant) throughout the motion. Suppose an object is moving with uniform speed v in a circle of radius R as shown in Fig. 4.19. Since the velocity of the object is changing continuously in direction, the object undergoes acceleration. Let us find the magnitude and the direction of this acceleration.

Fig. 4.19 Velocity and acceleration of an object in uniform circular motion. The time interval Δt decreases from (a) to (c) where it is zero. The acceleration is directed, at each point of the path, towards the centre of the circle.

Let **r** and **r′** be the position vectors and **v** and **v′** the velocities of the object when it is at point *P* and *P′* as shown in Fig. 4.19(a). By definition, velocity at a point is along the tangent at that point in the direction of motion. The velocity vectors **v** and **v′** are as shown in Fig. 4.19(a1). Δ**v** is obtained in Fig. 4.19 (a2) using the triangle law of vector addition. Since the path is circular, **v** is perpendicular to **r** and so is **v′** to **r′**. Therefore, Δ**v** is perpendicular to Δ**r**. Since average acceleration is along Δ**v** $\left(\bar{\mathbf{a}} = \frac{\Delta \mathbf{v}}{\Delta t}\right)$, the average acceleration $\bar{\mathbf{a}}$ is perpendicular to Δ**r.** If we place Δ**v** on the line that bisects the angle between **r** and **r′**, we see that it is directed towards the centre of the circle. Figure 4.19(b) shows the same quantities for smaller time interval. Δ**v** and hence $\bar{\mathbf{a}}$ is again directed towards the centre. In Fig. 4.19(c), $\Delta t \to 0$ and the average acceleration becomes the instantaneous acceleration. It is directed towards the centre*. Thus, we find that the acceleration of an object in uniform circular motion is always directed towards the centre of the circle. Let us now find the magnitude of the acceleration.

The magnitude of **a** is, by definition, given by

$$|\mathbf{a}| = \lim_{\Delta t \to 0} \frac{|\Delta \mathbf{v}|}{\Delta t}$$

Let the angle between position vectors **r** and **r′** be Δθ. Since the velocity vectors **v** and **v′** are always perpendicular to the position vectors, the angle between them is also Δθ . Therefore, the triangle CPP′ formed by the position vectors and the triangle GHI formed by the velocity vectors **v**, **v′** and Δ**v** are similar (Fig. 4.19a). Therefore, the ratio of the base-length to side-length for one of the triangles is equal to that of the other triangle. That is :

$$\frac{|\Delta \mathbf{v}|}{v} = \frac{|\Delta \mathbf{r}|}{R}$$

Or, $$|\Delta \mathbf{v}| = v\frac{|\Delta \mathbf{r}|}{R}$$

Therefore,

$$|\mathbf{a}| = \lim_{\Delta t \to 0} \frac{|\Delta \mathbf{v}|}{\Delta t} = \lim_{\Delta t \to 0} \frac{v|\Delta \mathbf{r}|}{R\Delta t} = \frac{v}{R} \lim_{\Delta t \to 0} \frac{|\Delta \mathbf{r}|}{\Delta t}$$

If Δt is small, Δθ will also be small and then arc *PP′* can be approximately taken to be | Δ**r** | :

$$|\Delta \mathbf{r}| \cong v\Delta t$$

$$\frac{|\Delta \mathbf{r}|}{\Delta t} \cong v$$

Or, $$\lim_{\Delta t \to 0} \frac{|\Delta \mathbf{r}|}{\Delta t} = v$$

Therefore, the centripetal acceleration a_c is :

* *In the limit Δt→0, Δ**r** becomes perpendicular to **r**. In this limit Δ**v**→**0** and is consequently also perpendicular to **V**. Therefore, the acceleration is directed towards the centre, at each point of the circular path.*

$$a_c = \left(\frac{v}{R}\right) v = v^2/R \qquad (4.44)$$

Thus, the acceleration of an object moving with speed v in a circle of radius R has a magnitude v^2/R and is always **directed towards the centre**. This is why this acceleration is called **centripetal acceleration** (a term proposed by Newton). A thorough analysis of centripetal acceleration was first published in 1673 by the Dutch scientist Christiaan Huygens (1629-1695) but it was probably known to Newton also some years earlier. "Centripetal" comes from a Greek term which means 'centre-seeking'. Since v and R are constant, the magnitude of the centripetal acceleration is also constant. However, the direction changes — pointing always towards the centre. Therefore, a centripetal acceleration is not a constant vector.

We have another way of describing the velocity and the acceleration of an object in uniform circular motion. As the object moves from P to P′ in time Δt (= $t' - t$), the line CP (Fig. 4.19) turns through an angle $\Delta\theta$ as shown in the figure. $\Delta\theta$ is called angular distance. We define the angular speed ω(Greek letter omega) as the time rate of change of angular displacement :

$$\omega = \frac{\Delta\theta}{\Delta t} \qquad (4.45)$$

Now, if the distance travelled by the object during the time Δt is Δs, i.e. PP' is Δs, then :

$$v = \frac{\Delta s}{\Delta t}$$

but $\Delta s = R\,\Delta\theta$. Therefore :

$$v = R\frac{\Delta\theta}{\Delta t} = R\,\omega$$

$$v = R\,\omega \qquad (4.46)$$

We can express centripetal acceleration a_c in terms of angular speed :

$$a_c = \frac{v^2}{R} = \frac{\omega^2 R^2}{R} = \omega^2 R$$

$$a_c = \omega^2 R \qquad (4.47)$$

The time taken by an object to make one revolution is known as its time period T and the number of revolution made in one second is called its frequency ν (=1/T). However, during this time the distance moved by the object is $s = 2\pi R$.

Therefore, $v = 2\pi R/T = 2\pi R\nu$ (4.48)

In terms of frequency ν, we have

$$\omega = 2\pi\nu$$

$$v = 2\pi R\nu$$

$$a_c = 4\pi^2 \nu^2 R \qquad (4.49)$$

▶ ***Example 4.10*** An insect trapped in a circular groove of radius 12 cm moves along the groove steadily and completes 7 revolutions in 100 s. (a) What is the angular speed, and the linear speed of the motion? (b) Is the acceleration vector a constant vector ? What is its magnitude ?

Answer This is an example of uniform circular motion. Here R = 12 cm. The angular speed ω is given by

$$\omega = 2\pi/T = 2\pi \times 7/100 = 0.44 \text{ rad/s}$$

The linear speed v is :

$$v = \omega R = 0.44 \text{ s}^{-1} \times 12 \text{ cm} = 5.3 \text{ cm s}^{-1}$$

The direction of velocity $\boldsymbol{v}$ is along the tangent to the circle at every point. The acceleration is directed towards the centre of the circle. Since this direction changes continuously, acceleration here is *not* a constant vector. However, the magnitude of acceleration is constant:

$$a = \omega^2 R = (0.44 \text{ s}^{-1})^2 (12 \text{ cm})$$

$$= 2.3 \text{ cm s}^{-2}$$ ◀

SUMMARY

1. *Scalar quantities* are quantities with magnitudes only. Examples are distance, speed, mass and temperature.
2. *Vector quantities* are quantities with magnitude and direction both. Examples are displacement, velocity and acceleration. They obey special rules of vector algebra.
3. A vector **A** multiplied by a real number λ is also a vector, whose magnitude is λ times the magnitude of the vector **A** and whose direction is the same or opposite depending upon whether λ is positive or negative.
4. Two vectors **A** and **B** may be *added graphically* using *head-to-tail method* or *parallelogram method*.
5. Vector addition is *commutative* :
$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$$
It also obeys the *associative law* :
$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$$
6. A *null* or *zero vector* is a vector with zero magnitude. Since the magnitude is zero, we don't have to specify its direction. It has the properties :
$$\mathbf{A} + \mathbf{0} = \mathbf{A}$$
$$\lambda\mathbf{0} = \mathbf{0}$$
$$0\,\mathbf{A} = \mathbf{0}$$
7. The *subtraction* of vector **B** from **A** is defined as the sum of **A** and **–B** :
$$\mathbf{A} - \mathbf{B} = \mathbf{A} + (-\mathbf{B})$$
8. A vector **A** can be *resolved* into component along two given vectors **a** and **b** lying in the same plane :
$$\mathbf{A} = \lambda\,\mathbf{a} + \mu\,\mathbf{b}$$
where λ and μ are real numbers.
9. A *unit vector* associated with a vector **A** has magnitude one and is along the vector **A**:
$$\hat{\mathbf{n}} = \frac{\mathbf{A}}{|\mathbf{A}|}$$
The unit vectors $\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}$ are vectors of unit magnitude and point in the direction of the x-, y-, and z-axes, respectively in a right-handed coordinate system.
10. A vector **A** can be expressed as
$$\mathbf{A} = A_x\hat{\mathbf{i}} + A_y\hat{\mathbf{j}}$$
where A_x, A_y are its components along x-, and y -axes. If vector **A** makes an angle θ with the x-axis, then $A_x = A\cos\theta$, $A_y = A\sin\theta$ and $A = |\mathbf{A}| = \sqrt{A_x^2 + A_y^2}$, $\tan\theta = \dfrac{A_y}{A_x}$.
11. Vectors can be conveniently added using *analytical method*. If sum of two vectors **A** and **B**, that lie in x-y plane, is **R**, then :
$\mathbf{R} = R_x\hat{\mathbf{i}} + R_y\hat{\mathbf{j}}$, where, $R_x = A_x + B_x$, and $R_y = A_y + B_y$
12. The *position vector* of an object in x-y plane is given by $\mathbf{r} = x\,\hat{\mathbf{i}} + y\,\hat{\mathbf{j}}$ and the *displacement* from position **r** to position **r'** is given by
$$\Delta\mathbf{r} = \mathbf{r}' - \mathbf{r}$$
$$= (x' - x)\,\hat{\mathbf{i}} + (y' - y)\,\hat{\mathbf{j}}$$
$$= \Delta x\,\hat{\mathbf{i}} + \Delta y\,\hat{\mathbf{j}}$$
13. If an object undergoes a displacement $\Delta\mathbf{r}$ in time Δt, its *average velocity* is given by $\mathbf{v} = \dfrac{\Delta\mathbf{r}}{\Delta t}$. The *velocity* of an object at time t is the limiting value of the average velocity

as Δt tends to zero :

$\mathbf{v} = \lim_{\Delta t \to 0} \frac{\Delta \mathbf{r}}{\Delta t} = \frac{d\mathbf{r}}{dt}$. It can be written in unit vector notation as :

$\mathbf{v} = v_x\hat{\mathbf{i}} + v_y\hat{\mathbf{j}} + v_z\hat{\mathbf{k}}$ where $v_x = \frac{dx}{dt}, v_y = \frac{dy}{dt}, v_z = \frac{dz}{dt}$

When position of an object is plotted on a coordinate system, $\mathbf{v}$ is always tangent to the curve representing the path of the object.

14. If the velocity of an object changes from $\mathbf{v}$ to $\mathbf{v}'$ in time Δt, then its *average acceleration* is given by: $\bar{\mathbf{a}} = \frac{\mathbf{v} - \mathbf{v}'}{\Delta t} = \frac{\Delta \mathbf{v}}{\Delta t}$

The *acceleration* $\mathbf{a}$ at any time t is the limiting value of $\bar{\mathbf{a}}$ as $\Delta t \to 0$:

$$\mathbf{a} = \lim_{\Delta t \to 0} \frac{\Delta \mathbf{v}}{\Delta t} = \frac{d\mathbf{v}}{dt}$$

In component form, we have : $\mathbf{a} = a_x\hat{\mathbf{i}} + a_y\hat{\mathbf{j}} + a_z\hat{\mathbf{k}}$

where, $a_x = \frac{dv_x}{dt}$, $a_y = \frac{dv_y}{dt}$, $a_z = \frac{dv_z}{dt}$

15. If an object is moving in a plane with constant acceleration $a = |\mathbf{a}| = \sqrt{a_x^2 + a_y^2}$ and its position vector at time $t = 0$ is $\mathbf{r}_o$, then at any other time t, it will be at a point given by:

$$\mathbf{r} = \mathbf{r}_o + \mathbf{v}_o t + \frac{1}{2}\mathbf{a}t^2$$

and its velocity is given by :

$$\mathbf{v} = \mathbf{v}_o + \mathbf{a}\, t$$

where $\mathbf{v}_o$ is the velocity at time $t = 0$

In component form :

$$x = x_o + v_{ox}t + \frac{1}{2}a_x t^2$$

$$y = y_o + v_{oy}\, t + \frac{1}{2}a_y t^2$$

$$v_x = v_{ox} + a_x t$$

$$v_y = v_{oy} + a_y t$$

Motion in a plane can be treated as superposition of two separate simultaneous one-dimensional motions along two perpendicular directions

16. An object that is in flight after being projected is called a *projectile*. If an object is projected with initial velocity $\mathbf{v}_o$ making an angle θ_o with x-axis and if we assume its initial position to coincide with the origin of the coordinate system, then the position and velocity of the projectile at time t are given by :

$$x = (v_o \cos\theta_o)\, t$$

$$y = (v_o \sin\theta_o)\, t - (1/2)\, g\, t^2$$

$$v_x = v_{ox} = v_o \cos\theta_o$$

$$v_y = v_o \sin\theta_o - g\, t$$

The path of a projectile is *parabolic* and is given by :

$$y = (\tan\theta_0)x - \frac{gx^2}{2(v_o \cos\theta_o)^2}$$

The *maximum height* that a projectile attains is :

$$h_m = \frac{(v_o \sin q_o)^2}{2g}$$

The *time* taken to reach this height is :

$$t_m = \frac{v_o \sin\theta_o}{g}$$

The horizontal distance travelled by a projectile from its initial position to the position it passes $y = 0$ during its fall is called the *range*, R of the projectile. It is :

$$R = \frac{v_o^2}{g} \sin 2\theta_o$$

17. When an object follows a circular path at constant speed, the motion of the object is called *uniform circular motion*. The magnitude of its acceleration is $a_c = v^2/R$. The direction of a_c is always towards the centre of the circle.
The angular speed ω is the rate of change of angular distance. It is related to velocity v by $v = \omega R$. The acceleration is $a_c = \omega^2 R$.
If T is the time period of revolution of the object in circular motion and ν is its frequency, we have $\omega = 2\pi\nu$, $v = 2\pi\nu R$, $a_c = 4\pi^2\nu^2 R$

Physical Quantity	Symbol	Dimensions	Unit	Remark
Position vector	$\mathbf{r}$	[L]	m	Vector. It may be denoted by any other symbol as well.
Displacement	$\Delta\mathbf{r}$	[L]	m	- do -
Velocity		$[LT^{-1}]$	$m\ s^{-1}$	
(a) Average	$\bar{\mathbf{v}}$			$= \frac{\Delta\mathbf{r}}{\Delta t}$, vector
(b) Instantaneous	$\mathbf{v}$			$= \frac{d\mathbf{r}}{dt}$, vector
Acceleration		$[LT^{-2}]$	$m\ s^{-2}$	
(a) Average	$\bar{\mathbf{a}}$			$= \frac{\Delta\mathbf{v}}{\Delta t}$, vector
(b) Instantaneous	$\mathbf{a}$			$= \frac{d\mathbf{v}}{dt}$, vector
Projectile motion				
(a) Time of max. height	t_m	[T]	s	$= \frac{v_0 \sin\theta_0}{g}$
(b) Max. height	h_m	[L]	m	$= \frac{(v_0 \sin\theta_0)^2}{2g}$
(c) Horizontal range	R	[L]	m	$= \frac{v_0^2 \sin 2\theta_0}{g}$
Circular motion				
(a) Angular speed	ω	$[T^{-1}]$	rad/s	$= \frac{\Delta\theta}{\Delta t} = \frac{v}{r}$
(b) Centripetal acceleration	a_c	$[LT^{-2}]$	$m\ s^{-2}$	$= \frac{v^2}{r}$

POINTS TO PONDER

1. The path length traversed by an object between two points is, in general, not the same as the magnitude of displacement. The displacement depends only on the end points; the path length (as the name implies) depends on the actual path. The two quantities are equal only if the object does not change its direction during the course of motion. In all other cases, the path length is greater than the magnitude of displacement.
2. In view of point 1 above, the average speed of an object is greater than or equal to the magnitude of the average velocity over a given time interval. The two are equal only if the path length is equal to the magnitude of displacement.
3. The vector equations (4.33a) and (4.34a) do not involve any choice of axes. Of course, you can always resolve them along any two independent axes.
4. The kinematic equations for uniform acceleration do not apply to the case of uniform circular motion since in this case the magnitude of acceleration is constant but its direction is changing.
5. An object subjected to two velocities $\mathbf{v}_1$ and $\mathbf{v}_2$ has a resultant velocity $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$. Take care to distinguish it from velocity of object 1 relative to velocity of object 2 : $\mathbf{v}_{12} = \mathbf{v}_1 - \mathbf{v}_2$. Here $\mathbf{v}_1$ and $\mathbf{v}_2$ are velocities with reference to some common reference frame.
6. The resultant acceleration of an object in circular motion is towards the centre only if the speed is constant.
7. The shape of the trajectory of the motion of an object is not determined by the acceleration alone but also depends on the initial conditions of motion (initial position and initial velocity). For example, the trajectory of an object moving under the same acceleration due to gravity can be a straight line or a parabola depending on the initial conditions.

EXERCISES

4.1 State, for each of the following physical quantities, if it is a scalar or a vector : volume, mass, speed, acceleration, density, number of moles, velocity, angular frequency, displacement, angular velocity.

4.2 Pick out the two scalar quantities in the following list :
force, angular momentum, work, current, linear momentum, electric field, average velocity, magnetic moment, relative velocity.

4.3 Pick out the only vector quantity in the following list :
Temperature, pressure, impulse, time, power, total path length, energy, gravitational potential, coefficient of friction, charge.

4.4 State with reasons, whether the following algebraic operations with scalar and vector physical quantities are meaningful :
(a) adding any two scalars, (b) adding a scalar to a vector of the same dimensions , (c) multiplying any vector by any scalar, (d) multiplying any two scalars, (e) adding any two vectors, (f) adding a component of a vector to the same vector.

4.5 Read each statement below carefully and state with reasons, if it is true or false :
(a) The magnitude of a vector is always a scalar, (b) each component of a vector is always a scalar, (c) the total path length is always equal to the magnitude of the displacement vector of a particle. (d) the average speed of a particle (defined as total path length divided by the time taken to cover the path) is either greater or equal to the magnitude of average velocity of the particle over the same interval of time, (e) Three vectors not lying in a plane can never add up to give a null vector.

4.6 Establish the following vector inequalities geometrically or otherwise :

(a) $|\mathbf{a}+\mathbf{b}| \leq |\mathbf{a}| + |\mathbf{b}|$

(b) $|\mathbf{a}+\mathbf{b}| \geq ||\mathbf{a}| - |\mathbf{b}||$

(c) $|\mathbf{a}-\mathbf{b}| \leq |\mathbf{a}| + |\mathbf{b}|$

(d) $|\mathbf{a}-\mathbf{b}| \geq ||\mathbf{a}| - |\mathbf{b}||$

When does the equality sign above apply?

4.7 Given $\mathbf{a} + \mathbf{b} + \mathbf{c} + \mathbf{d} = \mathbf{0}$, which of the following statements are correct :

(a) **a**, **b**, **c**, and **d** must each be a null vector,

(b) The magnitude of $(\mathbf{a} + \mathbf{c})$ equals the magnitude of $(\mathbf{b} + \mathbf{d})$,

(c) The magnitude of **a** can never be greater than the sum of the magnitudes of **b**, **c**, and **d**,

(d) **b** + **c** must lie in the plane of **a** and **d** if **a** and **d** are not collinear, and in the line of **a** and **d**, if they are collinear ?

4.8 Three girls skating on a circular ice ground of radius 200 m start from a point *P* on the edge of the ground and reach a point **Q** diametrically opposite to *P* following different paths as shown in Fig. 4.20. What is the magnitude of the displacement vector for each ? For which girl is this equal to the actual length of path skate ?

Fig. 4.20

4.9 A cyclist starts from the centre *O* of a circular park of radius 1 km, reaches the edge *P* of the park, then cycles along the circumference, and returns to the centre along **QO** as shown in Fig. 4.21. If the round trip takes 10 min, what is the (a) net displacement, (b) average velocity, and (c) average speed of the cyclist ?

Fig. 4.21

4.10 On an open ground, a motorist follows a track that turns to his left by an angle of 60^0 after every 500 m. Starting from a given turn, specify the displacement of the motorist at the third, sixth and eighth turn. Compare the magnitude of the displacement with the total path length covered by the motorist in each case.

4.11 A passenger arriving in a new town wishes to go from the station to a hotel located 10 km away on a straight road from the station. A dishonest cabman takes him along a circuitous path 23 km long and reaches the hotel in 28 min. What is (a) the average speed of the taxi, (b) the magnitude of average velocity ? Are the two equal ?

4.12 Rain is falling vertically with a speed of 30 m s^{-1}. A woman rides a bicycle with a speed of 10 m s^{-1} in the north to south direction. What is the direction in which she should hold her umbrella ?

4.13 A man can swim with a speed of 4.0 km/h in still water. How long does he take to cross a river 1.0 km wide if the river flows steadily at 3.0 km/h and he makes his

strokes normal to the river current? How far down the river does he go when he reaches the other bank ?

4.14 In a harbour, wind is blowing at the speed of 72 km/h and the flag on the mast of a boat anchored in the harbour flutters along the N-E direction. If the boat starts moving at a speed of 51 km/h to the north, what is the direction of the flag on the mast of the boat ?

4.15 The ceiling of a long hall is 25 m high. What is the maximum horizontal distance that a ball thrown with a speed of 40 m s^{-1} can go without hitting the ceiling of the hall ?

4.16 A cricketer can throw a ball to a maximum horizontal distance of 100 m. How much high above the ground can the cricketer throw the same ball ?

4.17 A stone tied to the end of a string 80 cm long is whirled in a horizontal circle with a constant speed. If the stone makes 14 revolutions in 25 s, what is the magnitude and direction of acceleration of the stone ?

4.18 An aircraft executes a horizontal loop of radius 1.00 km with a steady speed of 900 km/h. Compare its centripetal acceleration with the acceleration due to gravity.

4.19 Read each statement below carefully and state, with reasons, if it is true or false :

(a) The net acceleration of a particle in circular motion is *always* along the radius of the circle towards the centre

(b) The velocity vector of a particle at a point is *always* along the tangent to the path of the particle at that point

(c) The acceleration vector of a particle in *uniform* circular motion averaged over one cycle is a null vector

4.20 The position of a particle is given by

$$\mathbf{r} = 3.0t\,\hat{\mathbf{i}} - 2.0t^2\,\hat{\mathbf{j}} + 4.0\,\hat{\mathbf{k}}\ \text{m}$$

where t is in seconds and the coefficients have the proper units for $\mathbf{r}$ to be in metres.

(a) Find the $\mathbf{v}$ and $\mathbf{a}$ of the particle? (b) What is the magnitude and direction of velocity of the particle at t = 2.0 s ?

4.21 A particle starts from the origin at t = 0 s with a velocity of $10.0\,\hat{\mathbf{j}}$ m/s and moves in the x-y plane with a constant acceleration of $\left(8.0\hat{\mathbf{i}} + 2.0\hat{\mathbf{j}}\right)$ m s^{-2}. (a) At what time is the x- coordinate of the particle 16 m? What is the y-coordinate of the particle at that time? (b) What is the speed of the particle at the time ?

4.22 $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ are unit vectors along x- and y- axis respectively. What is the magnitude and direction of the vectors $\hat{\mathbf{i}} + \hat{\mathbf{j}}$, and $\hat{\mathbf{i}} - \hat{\mathbf{j}}$? What are the components of a vector $\mathbf{A} = 2\,\hat{\mathbf{i}} + 3\hat{\mathbf{j}}$ along the directions of $\hat{\mathbf{i}} + \hat{\mathbf{j}}$ and $\hat{\mathbf{i}} - \hat{\mathbf{j}}$? [You may use graphical method]

4.23 For any arbitrary motion in space, which of the following relations are true :

(a) $\mathbf{v}_{average} = (1/2)\,(\mathbf{v}\,(t_1) + \mathbf{v}\,(t_2))$

(b) $\mathbf{v}_{average} = [\mathbf{r}(t_2) - \mathbf{r}(t_1)\,]\,/(t_2 - t_1)$

(c) $\mathbf{v}\,(t) = \mathbf{v}\,(0) + \mathbf{a}\,t$

(d) $\mathbf{r}\,(t) = \mathbf{r}\,(0) + \mathbf{v}\,(0)\,t + (1/2)\,\mathbf{a}\,t^2$

(e) $\mathbf{a}_{average} = [\,\mathbf{v}\,(t_2) - \mathbf{v}\,(t_1\,)]\,/(\,t_2 - t_1)$

(The 'average' stands for average of the quantity over the time interval t_1 to t_2)

4.24 Read each statement below carefully and state, with reasons and examples, if it is true or false :

A scalar quantity is one that

(a) is conserved in a process

(b) can never take negative values

(c) must be dimensionless

(d) does not vary from one point to another in space

(e) has the same value for observers with different orientations of axes.

4.25 An aircraft is flying at a height of 3400 m above the ground. If the angle subtended at a ground observation point by the aircraft positions 10.0 s apart is 30°, what is the speed of the aircraft ?

Additional Exercises

4.26 A vector has magnitude and direction. Does it have a location in space ? Can it vary with time ? Will two equal vectors **a** and **b** at different locations in space necessarily have identical physical effects ? Give examples in support of your answer.

4.27 A vector has both magnitude and direction. Does it mean that anything that has magnitude and direction is necessarily a vector ? The rotation of a body can be specified by the direction of the axis of rotation, and the angle of rotation about the axis. Does that make any rotation a vector ?

4.28 Can you associate vectors with (a) the length of a wire bent into a loop, (b) a plane area, (c) a sphere ? Explain.

4.29 A bullet fired at an angle of 30° with the horizontal hits the ground 3.0 km away. By adjusting its angle of projection, can one hope to hit a target 5.0 km away ? Assume the *muzzle* speed to the fixed, and neglect air resistance.

4.30 A fighter plane flying horizontally at an altitude of 1.5 km with speed 720 km/h passes directly overhead an anti-aircraft gun. At what angle from the vertical should the gun be fired for the shell with muzzle speed 600 m s^{-1} to hit the plane ? At what minimum altitude should the pilot fly the plane to avoid being hit ? (Take g = 10 m s^{-2}).

4.31 A cyclist is riding with a speed of 27 km/h. As he approaches a circular turn on the road of radius 80 m, he applies brakes and reduces his speed at the constant rate of 0.50 m/s every second. What is the magnitude and direction of the net acceleration of the cyclist on the circular turn ?

4.32 (a) Show that for a projectile the angle between the velocity and the x-axis as a function of time is given by

$$\theta(t) = \tan^{-1}\left(\frac{v_{0y} - gt}{v_{ox}}\right)$$

(b) Shows that the projection angle θ_0 for a projectile launched from the origin is given by

$$\theta_0 = \tan^{-1}\left(\frac{4h_m}{R}\right)$$

where the symbols have their usual meaning.

CHAPTER FIVE

LAWS OF MOTION

5.1 Introduction
5.2 Aristotle's fallacy
5.3 The law of inertia
5.4 Newton's first law of motion
5.5 Newton's second law of motion
5.6 Newton's third law of motion
5.7 Conservation of momentum
5.8 Equilibrium of a particle
5.9 Common forces in mechanics
5.10 Circular motion
5.11 Solving problems in mechanics
Summary
Points to ponder
Exercises
Additional exercises

5.1 INTRODUCTION

In the preceding Chapter, our concern was to describe the motion of a particle in space quantitatively. We saw that uniform motion needs the concept of velocity alone whereas non-uniform motion requires the concept of acceleration in addition. So far, we have not asked the question as to what governs the motion of bodies. In this chapter, we turn to this basic question.

Let us first guess the answer based on our common experience. To move a football at rest, someone must kick it. To throw a stone upwards, one has to give it an upward push. A breeze causes the branches of a tree to swing; a strong wind can even move heavy objects. A boat moves in a flowing river without anyone rowing it. Clearly, some external agency is needed to provide force to move a body from rest. Likewise, an external force is needed also to retard or stop motion. You can stop a ball rolling down an inclined plane by applying a force against the direction of its motion.

In these examples, the external agency of force (hands, wind, stream, etc) is in contact with the object. This is not always necessary. A stone released from the top of a building accelerates downward due to the gravitational pull of the earth. A bar magnet can attract an iron nail from a distance. **This shows that external agencies (e.g. gravitational and magnetic forces) can exert force on a body even from a distance.**

In short, a force is required to put a stationary body in motion or stop a moving body, and some external agency is needed to provide this force. The external agency may or may not be in contact with the body.

So far so good. But what if a body is moving uniformly (e.g. a skater moving straight with constant speed on a horizontal ice slab) ? **Is an external force required to keep a body in uniform motion?**

5.2 ARISTOTLE'S FALLACY

The question posed above appears to be simple. However, it took ages to answer it. Indeed, the correct answer to this question given by Galileo in the seventeenth century was the foundation of Newtonian mechanics, which signalled the birth of modern science.

The Greek thinker, Aristotle (384 B.C– 322 B.C.), held the view that if a body is moving, something external is required to keep it moving. According to this view, for example, an arrow shot from a bow keeps flying since the air behind the arrow keeps pushing it. The view was part of an elaborate framework of ideas developed by Aristotle on the motion of bodies in the universe. Most of the Aristotelian ideas on motion are now known to be wrong and need not concern us. For our purpose here, the Aristotelian law of motion may be phrased thus: **An external force is required to keep a body in motion**.

Aristotelian law of motion is flawed, as we shall see. However, it is a natural view that anyone would hold from common experience. Even a small child playing with a simple (non-electric) toy-car on a floor knows intuitively that it needs to constantly drag the string attached to the toy-car with some force to keep it going. If it releases the string, it comes to rest. This experience is common to most terrestrial motion. External forces seem to be needed to keep bodies in motion. Left to themselves, all bodies eventually come to rest.

What is the flaw in Aristotle's argument? The answer is: a moving toy car comes to rest because the external force of friction on the car by the floor opposes its motion. To counter this force, the child has to apply an external force on the car in the direction of motion. When the car is in uniform motion, there is no net external force acting on it: the force by the child cancels the force (friction) by the floor. The corollary is: if there were no friction, the child would not be required to apply any force to keep the toy car in uniform motion.

The opposing forces such as friction (solids) and viscous forces (for fluids) are always present in the natural world. This explains why forces by external agencies are necessary to overcome the frictional forces to keep bodies in uniform motion. Now we understand where Aristotle went wrong. He coded this practical experience in the form of a basic argument. To get at the true law of nature for forces and motion, one has to imagine a world in which uniform motion is possible with no frictional forces opposing. This is what Galileo did.

5.3 THE LAW OF INERTIA

Galileo studied motion of objects on an inclined plane. Objects (i) moving down an inclined plane accelerate, while those (ii) moving up retard. (iii) Motion on a horizontal plane is an intermediate situation. Galileo concluded that an object moving on a frictionless horizontal plane must neither have acceleration nor retardation, i.e. it should move with constant velocity (Fig. 5.1(a)).

(i) (ii) (iii)

Fig. 5.1(a)

Another experiment by Galileo leading to the same conclusion involves a double inclined plane. A ball released from rest on one of the planes rolls down and climbs up the other. If the planes are smooth, the final height of the ball is nearly the same as the initial height (a little less but never greater). In the ideal situation, when friction is absent, the final height of the ball is the same as its initial height.

If the slope of the second plane is decreased and the experiment repeated, the ball will still reach the same height, but in doing so, it will travel a longer distance. In the limiting case, when the slope of the second plane is zero (i.e. is a horizontal) the ball travels an infinite distance. In other words, its motion never ceases. This is, of course, an idealised situation (Fig. 5.1(b)).

Fig. 5.1(b) The law of inertia was inferred by Galileo from observations of motion of a ball on a double inclined plane.

In practice, the ball does come to a stop after moving a finite distance on the horizontal plane, because of the opposing force of friction which can never be totally eliminated. However, if there were no friction, the ball would continue to move with a constant velocity on the horizontal plane.

Galileo thus, arrived at a new insight on motion that had eluded Aristotle and those who followed him. The state of rest and the state of uniform linear motion (motion with constant velocity) are equivalent. In both cases, there is no net force acting on the body. It is incorrect to assume that a net force is needed to keep a body in uniform motion. To maintain a body in uniform motion, we need to apply an external force to ecounter the frictional force, so that the two forces sum up to zero net external force.

To summarise, if the net external force is zero, a body at rest continues to remain at rest and a body in motion continues to move with a uniform velocity. This property of the body is called inertia. Inertia means **'resistance to change'**. A body does not change its state of rest or uniform motion, unless an external force compels it to change that state.

5.4 NEWTON'S FIRST LAW OF MOTION

Galileo's simple, but revolutionary ideas dethroned Aristotelian mechanics. A new mechanics had to be developed. This task was accomplished almost single-handedly by Isaac Newton, one of the greatest scientists of all times.

Newton built on Galileo's ideas and laid the foundation of mechanics in terms of three laws of motion that go by his name. Galileo's law of inertia was his starting point which he formulated as the **First Law of motion**:

Every body continues to be in its state of rest or of uniform motion in a straight line unless compelled by some external force to act otherwise.

Ideas on Motion in Ancient Indian Science

Ancient Indian thinkers had arrived at an elaborate system of ideas on motion. Force, the cause of motion, was thought to be of different kinds : force due to continuous pressure (nodan), as the force of wind on a sailing vessel; impact (abhighat), as when a potter's rod strikes the wheel; persistent tendency (sanskara) to move in a straight line(vega) or restoration of shape in an elastic body; transmitted force by a string, rod, etc. The notion of (vega) in the Vaisesika theory of motion perhaps comes closest to the concept of inertia. Vega, the tendency to move in a straight line, was thought to be opposed by contact with objects including atmosphere, a parallel to the ideas of friction and air resistance. It was correctly summarised that the different kinds of motion (translational, rotational and vibrational) of an extended body arise from only the translational motion of its constituent particles. A falling leaf in the wind may have downward motion as a whole (patan) and also rotational and vibrational motion (bhraman, spandan), but each particle of the leaf at an instant only has a definite (small) displacement. There was considerable focus in Indian thought on measurement of motion and units of length and time. It was known that the position of a particle in space can be indicated by distance measured along three axes. Bhaskara (1150 A.D.) had introduced the concept of 'instantaneous motion' (*tatkaliki gati*), which anticipated the modern notion of instantaneous velocity using Differential Calculus. The difference between a wave and a current (of water) was clearly understood; a current is a motion of particles of water under gravity and fluidity while a wave results from the transmission of vibrations of water particles.

The state of rest or uniform linear motion both imply zero acceleration. The first law of motion can, therefore, be simply expressed as:

If the net external force on a body is zero, its acceleration is zero. Acceleration can be non zero only if there is a net external force on the body.

Two kinds of situations are encountered in the application of this law in practice. In some examples, we know that the net external force on the object is zero. In that case we can conclude that the acceleration of the object is zero. For example, a spaceship out in interstellar space, far from all other objects and with all its rockets turned off, has no net external force acting on it. Its acceleration, according to the First Law, must be zero. If it is in motion, it must continue to move with a uniform velocity.

Galileo Galilei (1564 - 1642)

Galileo Galilei, born in Pisa, Italy in 1564 was a key figure in the scientific revolution in Europe about four centuries ago. Galileo proposed the concept of acceleration. From experiments on motion of bodies on inclined planes or falling freely, he contradicted the Aristotelian notion that a force was required to keep a body in motion, and that heavier bodies fall faster than lighter bodies under gravity. He thus arrived at the law of inertia that was the starting point of the subsequent epochal work of Isaac Newton.

Galileo's discoveries in astronomy were equally revolutionary. In 1609, he designed his own telescope (invented earlier in Holland) and used it to make a number of startling observations : mountains and depressions on the surface of the moon; dark spots on the sun; the moons of Jupiter and the phases of Venus. He concluded that the Milky Way derived its luminosity because of a large number of stars not visible to the naked eye. In his masterpiece of scientific reasoning : Dialogue on the Two Chief World Systems, Galileo advocated the heliocentric theory of the solar system proposed by Copernicus, which eventually got universal acceptance.

With Galileo came a turning point in the very method of scientific inquiry. Science was no longer merely observations of nature and inferences from them. Science meant devising and doing experiments to verify or refute theories. Science meant measurement of quantities and a search for mathematical relations between them. Not undeservedly, many regard Galileo as the father of modern science.

More often, however, we do not know all the forces to begin with. In that case, if we know that an object is unaccelerated (i.e. it is either at rest or in uniform linear motion), we can infer from the first law that the net external force on the object must be zero. Gravity is everywhere. For terrestrial phenomena, in particular, every object experiences gravitational force due to the earth. Also objects in motion generally experience friction, viscous drag, etc. If then, on earth, an object is at rest or in uniform linear motion, it is not because there are no forces acting on it, but because the various external forces cancel out i.e. add up to zero net external force.

Consider a book at rest on a horizontal surface Fig. (5.2(a)). It is subject to two external forces : the force due to gravity (i.e. its weight W) acting downward and the upward force on the book by the table, the normal force R . R is a self-adjusting force. This is an example of the kind of situation mentioned above. The forces are not quite known fully but the state of motion is known. We observe the book to be at rest. Therefore, we conclude from the first law that the magnitude of R equals that of W. A statement often encountered is : "Since $W = R$, forces cancel and, therefore, the book is at rest". This is incorrect reasoning. The correct statement is : "Since the book is observed to be at rest, the net external force on it must be zero, according to the first law. This implies that the normal force R must be equal and opposite to the weight W ".

Fig. 5.2 (a) *a book at rest on the table, and* (b) *a car moving with uniform velocity. The net force is zero in each case.*

Consider the motion of a car starting from rest, picking up speed and then moving on a smooth straight road with uniform speed (Fig. (5.2(b)). When the car is stationary, there is no net force acting on it. During pick-up, it accelerates. This must happen due to a net external force. Note, it has to be an external force. The acceleration of the car cannot be accounted for by any internal force. This might sound surprising, but it is true. The only conceivable external force along the road is the force of friction. It is the frictional force that accelerates the car as a whole. (You will learn about friction in section 5.9). When the car moves with constant velocity, there is no net external force.

The property of inertia contained in the First law is evident in many situations. Suppose we are standing in a stationary bus and the driver starts the bus suddenly. We get thrown backward with a jerk. Why ? Our feet are in touch with the floor. If there were no friction, we would remain where we were, while the floor of the bus would simply slip forward under our feet and the back of the bus would hit us. However, fortunately, there is some friction between the feet and the floor. If the start is not too sudden, i.e. if the acceleration is moderate, the frictional force would be enough to accelerate our feet along with the bus. But our body is not strictly a rigid body. It is deformable, i.e. it allows some relative displacement between different parts. What this means is that while our feet go with the bus, the rest of the body remains where it is due to inertia. Relative to the bus, therefore, we are thrown backward. As soon as that happens, however, the muscular forces on the rest of the body (by the feet) come into play to move the body along with the bus. A similar thing happens when the bus suddenly stops. Our feet stop due to the friction which does not allow relative motion between the feet and the floor of the bus. But the rest of the body continues to move forward due to inertia. We are thrown forward. The restoring muscular forces again come into play and bring the body to rest.

▶ ***Example 5.1*** An astronaut accidentally gets separated out of his small spaceship accelerating in inter stellar space at a constant rate of 100 m s^{-2}. What is the acceleration of the astronaut the instant after he is outside the spaceship ? (Assume that there are no nearby stars to exert gravitational force on him.)

Answer Since there are no nearby stars to exert gravitational force on him and the small spaceship exerts negligible gravitational attraction on him, the net force acting on the astronaut, once he is out of the spaceship, is zero. By the first law of motion the acceleration of the astronaut is zero. ◀

5.5 NEWTON'S SECOND LAW OF MOTION

The first law refers to the simple case when the net external force on a body is zero. The second law of motion refers to the general situation when there is a net external force acting on the body. It relates the net external force to the acceleration of the body.

Momentum

Momentum, **P** of a body is defined to be the product of its mass m and velocity **v**, and is denoted by **p**:

$$\mathbf{p} = m\mathbf{v} \tag{5.1}$$

Momentum is clearly a vector quantity. The following common experiences indicate the importance of this quantity for considering the effect of force on motion.

- Suppose a light-weight vehicle (say a small car) and a heavy weight vehicle (say a loaded truck) are parked on a horizontal road. We all know that a much greater force is needed to push the truck than the car to bring them to the same speed in same time. Similarly, a greater opposing force is needed to stop a heavy body than a light body in the same time, if they are moving with the same speed.
- If two stones, one light and the other heavy, are dropped from the top of a building, a person on the ground will find it easier to catch the light stone than the heavy stone. The mass of a body is thus an important parameter that determines the effect of force on its motion.
- Speed is another important parameter to consider. A bullet fired by a gun can easily pierce human tissue before it stops, resulting in casualty. The same bullet fired with moderate speed will not cause much damage. Thus for a given mass, the greater the speed, the greater is the opposing force needed to stop the body in a certain time. Taken together, the product of mass and velocity, that is momentum, is evidently a relevant variable of motion. The greater the change in the momentum in a given time, the greater is the force that needs to be applied.
- A seasoned cricketer catches a cricket ball coming in with great speed far more easily than a novice, who can hurt his hands in the act. One reason is that the cricketer allows a longer time for his hands to stop the ball. As you may have noticed, he draws in the hands backward in the act of catching the ball (Fig. 5.3). The novice, on the other hand, keeps his hands fixed and tries to catch the ball almost instantly. He needs to provide a much greater force to stop the ball instantly, and

this hurts.The conclusion is clear: force not only depends on the change in momentum, but also on how fast the change is brought about. The same change in momentum brought about in a shorter time needs a greater applied force. In short, the greater the rate of change of momentum, the greater is the force.

Fig. 5.3 *Force not only depends on the change in momentum but also on how fast the change is brought about. A seasoned cricketer draws in his hands during a catch, allowing greater time for the ball to stop and hence requires a smaller force.*

- Observations confirm that the product of mass and velocity (i.e. momentum) is basic to the effect of force on motion. Suppose a fixed force is applied for a certain interval of time on two bodies of different masses, initially at rest, the lighter body picks up a greater speed than the heavier body. However, at the end of the time interval, observations show that each body acquires the same momentum. **Thus the same force for the same time causes the same change in momentum for different bodies**. This is a crucial clue to the second law of motion.
- In the preceding observations, the vector character of momentum has not been evident. In the examples so far, momentum and change in momentum both have the same direction. But this is not always the case. Suppose a stone is rotated with uniform speed in a horizontal plane by means of a string, the magnitude of momentum is fixed, but its direction changes (Fig. 5.4). A force is needed to cause this change in momentum vector. This force is provided by our hand through the string. Experience suggests that our hand needs to exert a greater force if the stone is rotated at greater speed or in a circle of smaller radius, or both. This corresponds to greater acceleration or equivalently a greater rate of change in momentum vector. This suggests that the greater the rate of change in momentum vector the greater is the force applied.

Fig. 5.4 *Force is necessary for changing the direction of momentum, even if its magnitude is constant. We can feel this while rotating a stone in a horizontal circle with uniform speed by means of a string.*

These qualitative observations lead to the **Second Law of motion** expressed by Newton as follows :

The rate of change of momentum of a body is directly proportional to the applied force and takes place in the direction in which the force acts.

Thus, if under the action of a force **F** for time interval Δt, the velocity of a body of mass m changes from $\mathbf{v}$ to $\mathbf{v} + \Delta\mathbf{v}$ i.e. its initial momentum $\boldsymbol{p} = m\,\mathbf{v}$ changes by $\Delta\boldsymbol{p} = m\Delta\boldsymbol{v}$. According to the Second Law,

$$\mathbf{F} \propto \frac{\Delta\mathbf{p}}{\Delta t} \quad \text{or} \quad \mathbf{F} = k\,\frac{\Delta\mathbf{p}}{\Delta t}$$

where k is a constant of proportionality. Taking the limit $\Delta t \to 0$, the term $\frac{\Delta\mathbf{p}}{\Delta t}$ becomes the derivative or differential co-efficient of $\mathbf{p}$ with respect to t, denoted by $\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t}$. Thus

$$\mathbf{F} = k\frac{d\mathbf{p}}{dt} \quad (5.2)$$

For a body of fixed mass m,

$$\frac{d\mathbf{p}}{dt} = \frac{d}{dt}(m\,\mathbf{v}) = m\frac{d\mathbf{v}}{dt} = m\,\mathbf{a} \quad (5.3)$$

i.e the Second Law can also be written as

$$\mathbf{F} = k\,m\,\mathbf{a} \quad (5.4)$$

which shows that force is proportional to the product of mass m and acceleration **a**.

The unit of force has not been defined so far. In fact, we use Eq. (5.4) to define the unit of force. We, therefore, have the liberty to choose any constant value for k. For simplicity, we choose $k = 1$. The Second Law then is

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} = m\,\mathbf{a} \quad (5.5)$$

In SI unit force is one that causes an acceleration of 1 m s^{-2} to a mass of 1 kg. This unit is known as **newton** : 1 N = 1 kg m s^{-2}.

Let us note at this stage some important points about the second law :

1. In the second law, **F** = 0 implies **a** = 0. The second Law is obviously consistent with the first law.
2. The second law of motion is a vector law. It is equivalent to three equations, one for each component of the vectors :

$$F_x = \frac{dp_x}{dt} = ma_x$$

$$F_y = \frac{dp_y}{dt} = ma_y$$

$$F_z = \frac{dp_z}{dt} = m\,a_z \quad (5.6)$$

This means that if a force is not parallel to the velocity of the body, but makes some angle with it, it changes only the component of velocity along the direction of force. The component of velocity normal to the force remains unchanged. For example, in the motion of a projectile under the vertical gravitational force, the horizontal component of velocity remains unchanged (Fig. 5.5).

3. The second law of motion given by Eq. (5.5) is applicable to a single point particle. The force **F** in the law stands for the net external force on the particle and **a** stands for acceleration of the particle. It turns out, however, that the law in the same form applies to a rigid body or, even more generally, to a system of particles. In that case, **F** refers to the total external force on the system and **a** refers to the acceleration of the system as a whole. More precisely, **a** is the acceleration of the centre of mass of the system. **Any internal forces in the system are not to be included in F**.

Fig. 5.5 *Acceleration at an instant is determined by the force at that instant. The moment after a stone is dropped out of an accelerated train, it has no horizontal acceleration or force, if air resistance is neglected. The stone carries no memory of its acceleration with the train a moment ago.*

4. The second law of motion is a local relation which means that force **F** at a point in space (location of the particle) at a certain instant of time is related to **a** at that point at that instant. Acceleration here and now is determined by the force here and now, **not by any history of the motion of the particle (See Fig. 5.5)**.

▶ ***Example 5.2*** A bullet of mass 0.04 kg moving with a speed of 90 m s^{-1} enters a heavy wooden block and is stopped after a distance of 60 cm. What is the average resistive force exerted by the block on the bullet?

Answer The retardation 'a' of the bullet (assumed constant) is given by

$$a = \frac{-u^2}{2s} = \frac{-90 \times 90}{2 \times 0.6}\,\text{m s}^{-2} = -6750\,\text{m s}^{-2}$$

The retarding force, by the Second Law of motion, is

$$= 0.04 \text{ kg} \times 6750 \text{ m s}^{-2} = 270 \text{ N}$$

The actual resistive force, and therefore, retardation of the bullet may not be uniform. The answer therefore, only indicates the average resistive force. ◀

▶ ***Example 5.3*** The motion of a particle of mass m is described by $y = ut + \frac{1}{2}gt^2$. Find the force acting on the particle.

Answer We know

$$y = ut + \frac{1}{2}gt^2$$

Now,

$$v = \frac{dy}{dt} = u + gt$$

acceleration, $a = \frac{dv}{dt} = g$

Then the force is given by Eq. (5.5)

$$F = ma = mg$$ ◀

Thus the given equation describes the motion of a particle under acceleration due to gravity and y is the position coordinate in the direction of g.

Impulse

We sometimes encounter examples where a large force acts for a very short duration producing a finite change in momentum of the body. For example, when a ball hits a wall and bounces back, the force on the ball by the wall acts for a very short time when the two are in contact, yet the force is large enough to reverse the momentum of the ball. Often, in these situations, the force and the time duration are difficult to ascertain separately. However, the product of force and time, which is the change in momentum of the body remains a measurable quantity. This product is called impulse:

$$\text{Impulse} = \text{Force} \times \text{time duration} = \text{Change in momentum} \qquad (5.7)$$

A large force acting for a short time to produce a finite change in momentum is called an *impulsive force*. In the history of science, impulsive forces were put in a conceptually different category from ordinary forces. Newtonian mechanics has no such distinction. Impulsive force is like any other force – except that it is large and acts for a short time.

▶ ***Example 5.4*** A batsman hits back a ball straight in the direction of the bowler without changing its initial speed of 12 m s^{-1}. If the mass of the ball is 0.15 kg, determine the impulse imparted to the ball. (Assume linear motion of the ball)

Answer Change in momentum

$$= 0.15 \times 12 - (-0.15 \times 12)$$

$$= 3.6 \text{ N s},$$

Impulse = 3.6 N s,
in the direction from the batsman to the bowler.

This is an example where the force on the ball by the batsman and the time of contact of the ball and the bat are difficult to know, but the impulse is readily calculated. ◀

5.6 NEWTON'S THIRD LAW OF MOTION

The second law relates the external force on a body to its acceleration. What is the origin of the external force on the body ? What agency provides the external force ? The simple answer in Newtonian mechanics is that the external force on a body always arises due to some other body. Consider a pair of bodies *A* and *B*. *B* gives rise to an external force on *A*. A natural question is: Does *A* in turn give rise to an external force on *B* ? In some examples, the answer seems clear. If you press a coiled spring, the spring is compressed by the force of your hand. The compressed spring in turn exerts a force on your hand and you can feel it. But what if the bodies are not in contact ? The earth pulls a stone downwards due to gravity. Does the stone exert a force on the earth ? The answer is not obvious since we hardly see the effect of the stone on the earth. The answer according to Newton is: Yes, the stone does exert an equal and opposite force on the earth. We do not notice it since the earth is very massive and the effect of a small force on its motion is negligible.

Thus, according to Newtonian mechanics, force never occurs singly in nature. Force is the mutual interaction between two bodies. Forces

always occur in pairs. Further, the mutual forces between two bodies are always equal and opposite. This idea was expressed by Newton in the form of the **third law of motion**.

To every action, there is always an equal and opposite reaction.

Newton's wording of the third law is so crisp and beautiful that it has become a part of common language. For the same reason perhaps, misconceptions about the third law abound. Let us note some important points about the third law, particularly in regard to the usage of the terms : action and reaction.

1. The terms action and reaction in the third law mean nothing else but 'force'. Using different terms for the same physical concept can sometimes be confusing. A simple and clear way of stating the third law is as follows :

   Forces always occur in pairs. Force on a body A by B is equal and opposite to the force on the body B by A.

2. The terms action and reaction in the third law may give a wrong impression that action comes before reaction i.e action is the cause and reaction the effect. **There is no cause-effect relation implied in the third law. The force on *A* by *B* and the force on *B* by *A* act at the same instant.** By the same reasoning, any one of them may be called action and the other reaction.
3. Action and reaction forces act on different bodies, not on the same body. Consider a pair of bodies *A* and *B*. According to the third law,

$$\mathbf{F}_{AB} = -\mathbf{F}_{BA} \tag{5.8}$$

(force on *A* by *B*) = – (force on *B* by *A*)

Thus if we are considering the motion of any one body (*A* or *B*), only one of the two forces is relevant. It is an error to add up the two forces and claim that the net force is zero.

However, if you are considering the system of two bodies as a whole, $\mathbf{F}_{AB}$ and $\mathbf{F}_{BA}$ are internal forces of the system (*A* + *B*). They add up to give a null force. Internal forces in a body or a system of particles thus cancel away in pairs. This is an important fact that enables the second law to be applicable to a body or a system of particles (See Chapter 7).

Isaac Newton (1642 – 1727)

Isaac Newton was born in Woolsthorpe, England in 1642, the year Galileo died. His extraordinary mathematical ability and mechanical aptitude remained hidden from others in his school life. In 1662, he went to Cambridge for undergraduate studies. A plague epidemic in 1665 forced the university town to close and Newton had to return to his mother's farm. There in two years of solitude, his dormant creativity blossomed in a deluge of fundamental discoveries in mathematics and physics : binomial theorem for negative and fractional exponents, the beginning of calculus, the inverse square law of gravitation, the spectrum of white light, and so on. Returning to Cambridge, he pursued his investigations in optics and devised a reflecting telescope.

In 1684, encouraged by his friend Edmund Halley, Newton embarked on writing what was to be one of the greatest scientific works ever published : The Principia Mathematica. In it, he enunciated the three laws of motion and the universal law of gravitation, which explained all the three Kepler's laws of planetary motion. The book was packed with a host of path-breaking achievements : basic principles of fluid mechanics, mathematics of wave motion, calculation of masses of the earth, the sun and other planets, explanation of the precession of equinoxes, theory of tides, etc. In 1704, Newton brought out another masterpiece Opticks that summarized his work on light and colour.

The scientific revolution triggered by Copernicus and steered vigorously ahead by Kepler and Galileo was brought to a grand completion by Newton. Newtonian mechanics unified terrestrial and celestial phenomena. The same mathematical equation governed the fall of an apple to the ground and the motion of the moon around the earth. The age of reason had dawned.

Example 5.5 Two identical billiard balls strike a rigid wall with the same speed but at different angles, and get reflected without any change in speed, as shown in Fig. 5.6. What is (i) the direction of the force on the wall due to each ball? (ii) the ratio of the magnitudes of impulses imparted to the balls by the wall ?

Fig. 5.6

Answer An instinctive answer to (i) might be that the force on the wall in case (a) is normal to the wall, while that in case (b) is inclined at 30° to the normal. This answer is wrong. The force on the wall is normal to the wall in both cases.

How to find the force on the wall? The trick is to consider the force (or impulse) on the ball due to the wall using the second law, and then use the third law to answer (i). Let u be the speed of each ball before and after collision with the wall, and m the mass of each ball. Choose the x and y axes as shown in the figure, and consider the change in momentum of the ball in each case :

Case (a)

$$(p_x)_{\text{initial}} = mu \qquad (p_y)_{\text{initial}} = 0$$

$$(p_x)_{\text{final}} = -mu \qquad (p_y)_{\text{final}} = 0$$

Impulse is the change in momentum vector. Therefore,

x-component of impulse $= -2\,m\,u$
y-component of impulse $= 0$

Impulse and force are in the same direction. Clearly, from above, the force on the ball due to the wall is normal to the wall, along the negative x-direction. Using Newton's third law of motion, the force on the wall due to the ball is normal to the wall along the positive x-direction. The magnitude of force cannot be ascertained since the small time taken for the collision has not been specified in the problem.

Case (b)

$$(p_x)_{initial} = m\,u\cos 30^\circ,\ (p_y)_{initial} = -\,m\,u\sin 30^\circ$$

$$(p_x)_{final} = -\,m\,u\cos 30^\circ,\ (p_y)_{final} = -\,m\,u\sin 30^\circ$$

Note, while p_x changes sign after collision, p_y does not. Therefore,

x-component of impulse $= -2\,m\,u\cos 30^\circ$
y-component of impulse $= 0$

The direction of impulse (and force) is the same as in (a) and is normal to the wall along the negative x direction. As before, using Newton's third law, the force on the wall due to the ball is normal to the wall along the positive x direction.

The ratio of the magnitudes of the impulses imparted to the balls in (a) and (b) is

$$2\,m\,u/\left(2\,m\,u\cos 30^\circ\right) = \frac{2}{\sqrt{3}} \approx 1.2$$

5.7 CONSERVATION OF MOMENTUM

The second and the third laws of motion lead to an important consequence: the law of conservation of momentum. Take a familiar example. A bullet is fired from a gun. If the force on the bullet by the gun is **F**, the force on the gun by the bullet is $-\mathbf{F}$, according to the third law. The two forces act for a common interval of time Δt. According to the second law, $\mathbf{F}\,\Delta t$ is the change in momentum of the bullet and $-\mathbf{F}\,\Delta t$ is the change in momentum of the gun. Since initially, both are at rest, the change in momentum equals the final momentum for each. Thus if $\mathbf{p}_b$ is the momentum of the bullet after firing and $\mathbf{p}_g$ is the recoil momentum of the gun, $\mathbf{p}_g = -\mathbf{p}_b$ *i.e.* $\mathbf{p}_b + \mathbf{p}_g = 0$. That is, the total momentum of the (bullet + gun) system is conserved.

Thus in an isolated system (i.e. a system with no external force), mutual forces between pairs of particles in the system can cause momentum change in individual particles, but since the mutual forces for each pair are equal and opposite, the momentum changes cancel in pairs and the total momentum remains unchanged. This fact is known as the **law of conservation of momentum** :

The total momentum of an isolated system of interacting particles is conserved.

An important example of the application of the law of conservation of momentum is the collision of two bodies. Consider two bodies A and B, with initial momenta $\mathbf{p}_A$ and $\mathbf{p}_B$. The bodies collide, get apart, with final momenta $\mathbf{p}'_A$ and $\mathbf{p}'_B$ respectively. By the Second Law

$$\mathbf{F}_{AB}\Delta t = \mathbf{p}'_A - \mathbf{p}_A \quad \text{and}$$

$$\mathbf{F}_{BA}\Delta t = \mathbf{p}'_B - \mathbf{p}_B$$

(where we have taken a common interval of time for both forces i.e. the time for which the two bodies are in contact.)

Since $\mathbf{F}_{AB} = -\mathbf{F}_{BA}$ by the third law,

$$\mathbf{p}'_A - \mathbf{p}_A = -(\mathbf{p}'_B - \mathbf{p}_B)$$

i.e. $$\mathbf{p}'_A + \mathbf{p}'_B = \mathbf{p}_A + \mathbf{p}_B \qquad (5.9)$$

which shows that the total final momentum of the isolated system equals its initial momentum. Notice that this is true whether the collision is elastic or inelastic. In elastic collisions, there is a second condition that the total initial kinetic energy of the system equals the total final kinetic energy (See Chapter 6).

5.8 EQUILIBRIUM OF A PARTICLE

Equilibrium of a particle in mechanics refers to the situation when the net external force on the particle is zero.* According to the first law, this means that, the particle is either at rest or in uniform motion.

If two forces $\mathbf{F}_1$ and $\mathbf{F}_2$, act on a particle, equilibrium requires

$$\mathbf{F}_1 = -\mathbf{F}_2 \qquad (5.10)$$

i.e. the two forces on the particle must be equal and opposite. Equilibrium under three concurrent forces $\mathbf{F}_1$, $\mathbf{F}_2$ and $\mathbf{F}_3$ requires that the vector sum of the three forces is zero.

$$\mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3 = 0 \qquad (5.11)$$

Fig. 5.7 Equilibrium under concurrent forces.

In other words, the resultant of any two forces say $\mathbf{F}_1$ and $\mathbf{F}_2$, obtained by the parallelogram law of forces must be equal and opposite to the third force, $\mathbf{F}_3$. As seen in Fig. 5.7, the three forces in equilibrium can be represented by the sides of a triangle with the vector arrows taken in the same sense. The result can be generalised to any number of forces. A particle is in equilibrium under the action of forces $\mathbf{F}_1$, $\mathbf{F}_2$,... $\mathbf{F}_n$ if they can be represented by the sides of a closed n-sided polygon with arrows directed in the same sense.

Equation (5.11) implies that

$$F_{1x} + F_{2x} + F_{3x} = 0$$
$$F_{1y} + F_{2y} + F_{3y} = 0$$
$$F_{1z} + F_{2z} + F_{3z} = 0 \qquad (5.12)$$

where F_{1x}, F_{1y} and F_{1z} are the components of F_1 along x, y and z directions respectively.

▶ ***Example 5.6*** See Fig. 5.8 A mass of 6 kg is suspended by a rope of length 2 m from the ceiling. A force of 50 N in the horizontal direction is applied at the mid-point P of the rope, as shown. What is the angle the rope makes with the vertical in equilibrium ? (Take g = 10 m s^{-2}). Neglect the mass of the rope.

Fig. 5.8

* *Equilibrium of a body requires not only translational equilibrium (zero net external force) but also rotational equilibrium (zero net external torque), as we shall see in Chapter 7.*

Answer Figures 5.8(b) and 5.8(c) are known as free-body diagrams. Figure 5.8(b) is the free-body diagram of W and Fig. 5.8(c) is the free-body diagram of point P.

Consider the equilibrium of the weight W. Clearly, $T_2 = 6 \times 10 = 60$ N.

Consider the equilibrium of the point P under the action of three forces - the tensions T_1 and T_2, and the horizontal force 50 N. The horizontal and vertical components of the resultant force must vanish separately :

$$T_1 \cos\theta = T_2 = 60 \text{ N}$$

$$T_1 \sin\theta = 50 \text{ N}$$

which gives that

$$\tan\theta = \frac{5}{6} \text{ or } \theta = \tan^{-1}\left(\frac{5}{6}\right) = 40^\circ$$

Note the answer does not depend on the length of the rope (assumed massless) nor on the point at which the horizontal force is applied. ◀

5.9 COMMON FORCES IN MECHANICS

In mechanics, we encounter several kinds of forces. The gravitational force is, of course, all pervasive. Every object on the earth experiences the force of gravity due to the earth. Gravity also governs the motion of celestial bodies. The gravitational force can act at a distance without the need of any intervening medium.

All the other forces common in mechanics are contact forces.* As the name suggests, a contact force on an object arises due to contact with some other object: solid or fluid. When bodies are in contact (e.g. a book resting on a table, a system of rigid bodies connected by rods, hinges and other types of supports), there are mutual contact forces (for each pair of bodies) satisfying the third law. The component of contact force normal to the surfaces in contact is called normal reaction. The component parallel to the surfaces in contact is called friction. Contact forces arise also when solids are in contact with fluids. For example, for a solid immersed in a fluid, there is an upward bouyant force equal to the weight of the fluid displaced. The viscous force, air resistance, etc are also examples of contact forces (Fig. 5.9).

Two other common forces are tension in a string and the force due to spring. When a spring is compressed or extended by an external force, a restoring force is generated. This force is usually proportional to the compression or elongation (for small displacements). The spring force F is written as $F = -k\,x$ where x is the displacement and k is the force constant. The negative sign denotes that the force is opposite to the displacement from the unstretched state. For an inextensible string, the force constant is very high. The restoring force in a string is called tension. It is customary to use a constant tension T throughout the string. This assumption is true for a string of negligible mass.

In Chapter 1, we learnt that there are four fundamental forces in nature. Of these, the weak and strong forces appear in domains that do not concern us here. Only the gravitational and electrical forces are relevant in the context of mechanics. The different contact forces of mechanics mentioned above fundamentally arise from electrical forces. This may seem surprising

Fig. 5.9 Some examples of contact forces in mechanics.

* *We are not considering, for simplicity, charged and magnetic bodies. For these, besides gravity, there are electrical and magnetic non-contact forces.*

since we are talking of uncharged and non-magnetic bodies in mechanics. At the microscopic level, all bodies are made of charged constituents (nuclei and electrons) and the various contact forces arising due to elasticity of bodies, molecular collisions and impacts, etc. can ultimately be traced to the electrical forces between the charged constituents of different bodies. The detailed microscopic origin of these forces is, however, complex and not useful for handling problems in mechanics at the macroscopic scale. This is why they are treated as different types of forces with their characteristic properties determined empirically.

5.9.1 Friction

Let us return to the example of a body of mass m at rest on a horizontal table. The force of gravity (mg) is cancelled by the normal reaction force (N) of the table. Now suppose a force F is applied horizontally to the body. We know from experience that a small applied force may not be enough to move the body. But if the applied force F were the only external force on the body, it must move with acceleration F/m, however small. Clearly, the body remains at rest because some other force comes into play in the horizontal direction and opposes the applied force F, resulting in zero net force on the body. This force f_s parallel to the surface of the body in contact with the table is known as frictional force, or simply friction (Fig. 5.10(a)). The subscript stands for static friction to distinguish it from kinetic friction f_k that we consider later (Fig. 5.10(b)). Note that static friction does not exist by itself. When there is no applied force, there is no static friction. It comes into play the moment there is an applied force. As the applied force F increases, f_s also increases, remaining equal and opposite to the applied force (up to a certain limit), keeping the body at rest. Hence, it is called **static friction**. Static friction opposes **impending motion**. The term impending motion means motion that would take place (but does not actually take place) under the applied force, if friction were absent.

Fig. 5.10 Static and sliding friction: (a) Impending motion of the body is opposed by static friction. When external force exceeds the maximum limit of static friction, the body begins to move. (b) Once the body is in motion, it is subject to sliding or kinetic friction which opposes relative motion between the two surfaces in contact. Kinetic friction is usually less than the maximum value of static friction.

We know from experience that as the applied force exceeds a certain limit, the body begins to move. It is found experimentally that the limiting value of static friction $(f_s)_{max}$ is independent of the area of contact and varies with the normal force (N) approximately as :

$$(f_s)_{max} = \mu_s N \qquad (5.13)$$

where μ_s is a constant of proportionality depending only on the nature of the surfaces in contact. The constant μ_s is called the coefficient of static friction. The law of static friction may thus be written as

$$f_s \le \mu_s \text{ N} \qquad (5.14)$$

If the applied force F exceeds $(f_s)_{max}$ the body begins to slide on the surface. It is found experimentally that when relative motion has started, the frictional force decreases from the static maximum value $(f_s)_{max}$. Frictional force that opposes relative motion between surfaces in contact is called kinetic or sliding friction and is denoted by f_k. Kinetic friction, like static friction, is found to be independent of the area of contact. Further, it is nearly independent of the velocity. It satisfies a law similar to that for static friction:

$$f_k = \mu_k \mathbf{N} \qquad (5.15)$$

where μ_k, the coefficient of kinetic friction, depends only on the surfaces in contact. As mentioned above, experiments show that μ_k is less than μ_s. When relative motion has begun, the acceleration of the body according to the Second Law is $(F - f_k)/m$. For a body moving with constant velocity, $F = f_k$. If the applied force on the body is removed, its acceleration is $-f_k/m$ and it eventually comes to a stop.

The laws of friction given above do not have the status of fundamental laws like those for gravitational, electric and magnetic forces. They are empirical relations that are only

approximately true. Yet they are very useful in practical calculations in mechanics.

Thus, when two bodies are in contact, each experiences a contact force by the other. Friction, by definition, is the component of the contact force parallel to the surfaces in contact, which opposes impending or actual relative motion between the two surfaces. Note that it is not motion, but **relative motion** that the frictional force opposes. Consider a box lying in the compartment of a train that is accelerating. If the box is stationary relative to the train, it is in fact accelerating along with the train. What forces cause the acceleration of the box? Clearly, the only conceivable force in the horizontal direction is the force of friction. If there were no friction, the floor of the train would slip by and the box would remain at its initial position due to inertia (and hit the back side of the train). This impending relative motion is opposed by the static friction f_s. Static friction provides the same acceleration to the box as that of the train, keeping it stationary relative to the train.

▶ ***Example 5.7*** Determine the maximum acceleration of the train in which a box lying on its floor will remain stationary, given that the co-efficient of static friction between the box and the train's floor is 0.15.

Answer Since the acceleration of the box is due to the static friction,

$$ma = f_s \le \mu_s N = \mu_s\, m g$$

i.e. $a \le \mu_s g$

$\therefore a_{max} = \mu_s g = 0.15 \times 10 \text{ m s}^{-2}$

$= 1.5 \text{ m s}^{-2}$ ◀

▶ ***Example 5.8*** See Fig. 5.11. A mass of 4 kg rests on a horizontal plane. The plane is gradually inclined until at an angle $\theta = 15°$ with the horizontal, the mass just begins to slide. What is the coefficient of static friction between the block and the surface ?

Fig. 5.11

Answer The forces acting on a block of mass m at rest on an inclined plane are (i) the weight mg acting vertically downwards (ii) the normal force N of the plane on the block, and (iii) the static frictional force f_s opposing the impending motion. In equilibrium, the resultant of these forces must be zero. Resolving the weight mg along the two directions shown, we have

$$m g \sin\theta = f_s \;, \quad m g \cos\theta = N$$

As θ increases, the self-adjusting frictional force f_s increases until at $\theta = \theta_{max}$, f_s achieves its maximum value, $(f_s)_{max} = \mu_s N$.

Therefore,

$$\tan\theta_{max} = \mu_s \;\; or \;\; \theta_{max} = \tan^{-1}\mu_s$$

When θ becomes just a little more than θ_{max}, there is a small net force on the block and it begins to slide. Note that θ_{max} depends only on μ_s and is independent of the mass of the block.

For $\theta_{max} = 15°$,

$\mu_s = \tan 15°$

$= 0.27$ ◀

▶ ***Example 5.9*** What is the acceleration of the block and trolley system shown in a Fig. 5.12(a), if the coefficient of kinetic friction between the trolley and the surface is 0.04? What is the tension in the string? (Take g = 10 m s^{-2}). Neglect the mass of the string.

Fig. 5.12

Answer As the string is inextensible, and the pully is smooth, the 3 kg block and the 20 kg trolley both have same magnitude of acceleration. Applying second law to motion of the block (Fig. 5.12(b)),

$$30 - T = 3a$$

Apply the second law to motion of the trolley (Fig. 5.12(c)),

$$T - f_k = 20\,a.$$

Now $f_k = \mu_k N$,

Here $\mu_k = 0.04$,

$N = 20 \times 10$

$= 200$ N.

Thus the equation for the motion of the trolley is

$$T - 0.04 \times 200 = 20\,a \text{ Or } T - 8 = 20a.$$

These equations give $a = \frac{22}{23}$ m s^{-2} = 0.96 m s^{-2} and T = 27.1 N. ◀

Rolling friction

A body like a ring or a sphere rolling without slipping over a horizontal plane will suffer no friction, in principle. At every instant, there is just one point of contact between the body and the plane and this point has no motion relative to the plane. In this ideal situation, kinetic or static friction is zero and the body should continue to roll with constant velocity. We know, in practice, this will not happen and some resistance to motion (rolling friction) does occur, i.e. to keep the body rolling, some applied force is needed. For the same weight, rolling friction is much smaller (even by 2 or 3 orders of magnitude) than static or sliding friction. This is the reason why discovery of the wheel has been a major milestone in human history.

Rolling friction again has a complex origin, though somewhat different from that of static and sliding friction. During rolling, the surfaces in contact get momentarily deformed a little, and this results in a finite area (not a point) of the body being in contact with the surface. The net effect is that the component of the contact force parallel to the surface opposes motion.

We often regard friction as something undesirable. In many situations, like in a machine with different moving parts, friction does have a negative role. It opposes relative motion and thereby dissipates power in the form of heat, etc. Lubricants are a way of reducing kinetic friction in a machine. Another way is to use ball bearings between two moving parts of a machine. (Fig. 5.13(a)) Since the rolling friction between ball bearings and the surfaces in contact is very small, power dissipation is reduced. A thin cushion of air maintained between solid surfaces in relative motion is another effective way of reducing friction (Fig. 5.13(a)).

In many practical situations, however, friction is critically needed. Kinetic friction that dissipates power is nevertheless important for quickly stopping relative motion. It is made use of by brakes in machines and automobiles. Similarly, static friction is important in daily life. We are able to walk because of friction. It is impossible for a car to move on a very slippery road. On an ordinary road, the friction between the tyres and the road provides the necessary external force to accelerate the car.

Fig. 5.13 *Some ways of reducing friction. (a) Ball bearings placed between moving parts of a machine. (b) Compressed cushion of air between surfaces in relative motion.*

5.10 CIRCULAR MOTION

We have seen in Chapter 4 that acceleration of a body moving in a circle of radius R with uniform speed v is v^2/R directed towards the centre. According to the second law, the force f providing this acceleration is :

$$f = \frac{mv^2}{R} \qquad (5.16)$$

where m is the mass of the body. This force directed forwards the centre is called the centripetal force. For a stone rotated in a circle by a string, the centripetal force is provided by the tension in the string. The centripetal force for motion of a planet around the sun is the gravitational force on the planet due to the sun. For a car taking a circular turn on a horizontal road, the centripetal force is the force of friction.

The circular motion of a car on a flat and banked road give interesting application of the laws of motion.

Motion of a car on a level road

Three forces act on the car. (Fig. 5.14(a)

(i) The weight of the car, mg

(ii) Normal reaction, N

(iii) Frictional force, f

As there is no acceleration in the vertical direction

$N - mg = 0$

$$N = mg \qquad (5.17)$$

The centripetal force required for circular motion is along the surface of the road, and is provided by the component of the contact force between road and the car tyres along the surface. This by definition is the frictional force. Note that it is the static friction that provides the centripetal acceleration. Static friction opposes the impending motion of the car moving away from the circle. Using equation (5.14) & (5.16) we get the result

$$f \le \mu_s N = \frac{mv^2}{R}$$

$$v^2 \le \frac{\mu_s RN}{m} = \mu_s Rg \qquad [\text{Q } N = mg]$$

which is independent of the mass of the car. This shows that for a given value of μ_s and R, there is a maximum speed of circular motion of the car possible, namely

$$v_{\max} = \sqrt{\mu_s Rg} \qquad (5.18)$$

Fig. 5.14 Circular motion of a car on (a) a level road, (b) a banked road.

Motion of a car on a banked road

We can reduce the contribution of friction to the circular motion of the car if the road is banked (Fig. 5.14(b)). Since there is no acceleration along the vertical direction, the net force along this direction must be zero. Hence,

$$N\cos\theta = mg + f\sin\theta \qquad (5.19a)$$

The centripetal force is provided by the horizontal components of N and f.

$$N\sin\theta + f\cos\theta = \frac{mv^2}{R} \qquad (5.20a)$$

But $f \le \mu_s N$

Thus to obtain v_{max} we put

$f = \mu_s N$.

Then Eqs. (5.19) & (5.20) become

$$N\cos\theta = mg + \mu_s N\sin\theta \qquad (5.19b)$$

$$N \sin\theta + \mu_s N \cos\theta = mv^2/R \qquad (5.20b)$$

We obtain

$$N = \frac{mg}{\cos\theta - \mu_s \sin\theta}$$

Substituting value of N in Eq. (5.20b), we get

$$\frac{mg(\sin\theta - \mu_s \cos\theta)}{\cos\theta - \mu_s \sin\theta} = \frac{mv_{max}^2}{R}$$

$$\text{or } v_{max} = \left(Rg\frac{\mu_s + \tan\theta}{1 - \mu_s \tan\theta}\right)^{\frac{1}{2}} \qquad (5.21)$$

Comparing this with Eq. (5.18) we see that maximum possible speed of a car on a banked road is greater than that on a flat road.

For $\mu_s = 0$ in Eq. (5.21),

$$v_o = (R g \tan\theta)^{1/2} \qquad (5.22)$$

At this speed, frictional force is not needed at all to provide the necessary centripetal force. Driving at this speed on a banked road will cause little wear and tear of the tyres. The same equation also tells you that for $v < v_o$, frictional force will be up the slope and that a car can be parked only if $\tan\theta \le \mu_s$.

▶ ***Example 5.10*** A cyclist speeding at 18 km/h on a level road takes a sharp circular turn of radius 3 m without reducing the speed. The co-efficient of static friction between the tyres and the road is 0.1. Will the cyclist slip while taking the turn ?

Answer On an unbanked road, frictional force alone can provide the centripetal force needed to keep the cyclist moving on a circular turn without slipping. If the speed is too large, or if the turn is too sharp (i.e. of too small a radius) or both, the frictional force is not sufficient to provide the necessary centripetal force, and the cyclist slips. The condition for the cyclist not to slip is given by Eq. (5.18) :

$$v^2 \le \mu_s R g$$

Now, $R = 3$ m, $g = 9.8$ m s^{-2}, $\mu_s = 0.1$. That is, $\mu_s R g = 2.94$ m^2 s^{-2}. $v = 18$ km/h $= 5$ m s^{-1}; i.e., $v^2 = 25$ m^2 s^{-2}. The condition is not obeyed. The cyclist will slip while taking the circular turn. ◀

▶ ***Example 5.11*** A circular racetrack of radius 300 m is banked at an angle of 15°. If the coefficient of friction between the wheels of a race-car and the road is 0.2, what is the (a) optimum speed of the race-car to avoid wear and tear on its tyres, and (b) maximum permissible speed to avoid slipping ?

Answer On a banked road, the horizontal component of the normal force and the frictional force contribute to provide centripetal force to keep the car moving on a circular turn without slipping. At the optimum speed, the normal reaction's component is enough to provide the needed centripetal force, and the frictional force is not needed. The optimum speed v_o is given by Eq. (5.22):

$$v_O = (R g \tan\theta)^{1/2}$$

Here $R = 300$ m, $\theta = 15°$, $g = 9.8$ m s^{-2}; we have

$$v_O = 28.1 \text{ m s}^{-1}.$$

The maximum permissible speed v_{max} is given by Eq. (5.21):

$$v_{max} = \left(Rg\frac{\mu_s + \tan\theta}{1 - \mu_s \tan\theta}\right)^{1/2} = 38.1 \text{ m s}^{-1}$$ ◀

5.11 SOLVING PROBLEMS IN MECHANICS

The three laws of motion that you have learnt in this chapter are the foundation of mechanics. You should now be able to handle a large variety of problems in mechanics. A typical problem in mechanics usually does not merely involve a single body under the action of given forces. More often, we will need to consider an assembly of different bodies exerting forces on each other. Besides, each body in the assembly experiences the force of gravity. When trying to solve a problem of this type, it is useful to remember the fact that we can choose any part of the assembly and apply the laws of motion to that part provided we include all forces on the chosen part due to the remaining parts of the assembly. We may call the chosen part of the assembly as the system and the remaining part of the assembly (plus any other agencies of forces) as the environment. We have followed the same

method in solved examples. To handle a typical problem in mechanics systematically, one should use the following steps :

(i) Draw a diagram showing schematically the various parts of the assembly of bodies, the links, supports, etc.

(ii) Choose a convenient part of the assembly as one system.

(iii) Draw a separate diagram which shows this system and all the forces on the system by the remaining part of the assembly. Include also the forces on the system by other agencies. **Do not include the forces on the environment by the system**. A diagram of this type is known as 'a free-body diagram'. (Note this does not imply that the system under consideration is without a net force).

(iv) In a free-body diagram, include information about forces (their magnitudes and directions) that are either given or you are sure of (e.g., the direction of tension in a string along its length). The rest should be treated as unknowns to be determined using laws of motion.

(v) If necessary, follow the same procedure for another choice of the system. In doing so, employ Newton's third law. That is, if in the free-body diagram of *A*, the force on *A* due to *B* is shown as **F,** then in the free-body diagram of *B*, the force on *B* due to *A* should be shown as **–F.**

The following example illustrates the above procedure :

▶ ***Example 5.12*** See (Fig. 5.15) A wooden block of mass 2 kg rests on a soft horizontal floor. When an iron cylinder of mass 25 kg is placed on top of the block, the floor yields steadily and the block and the cylinder together go down with an acceleration of 0.1 m s^{-2}. What is the action of the block on the floor (a) before and (b) after the floor yields ? Take g = 10 m s^{-2}. Identify the action-reaction pairs in the problem.

Answer

(a) The block is at rest on the floor. Its free-body diagram shows two forces on the block, the force of gravitational attraction by the earth equal to 2 × 10 = 20 N; and the normal force *R* of the floor on the block. By the First Law, the net force on the block must be zero i.e., *R* = 20 N. Using third law the action of the block (i.e. the force exerted on the floor by the block) is equal to 20 N and directed vertically downwards.

(b) The system (block + cylinder) accelerates downwards with 0.1 m s^{-2}. The free-body diagram of the system shows two forces on the system : the force of gravity due to the earth (270 N); and the normal force *R′* by the floor. Note, the free-body diagram of the system does not show the internal forces between the block and the cylinder. Applying the second law to the system,

$$270 - R' = 27 \times 0.1\text{N}$$
$$\text{ie. } R' = 267.3 \text{ N}$$

Fig. 5.15

By the third law, the action of the system on the floor is equal to 267.3 N vertically downward.

Action-reaction pairs

For (a): (i) the force of gravity (20 N) on the block by the earth (say, action); the force of gravity on the earth by the block (reaction) equal to 20 N directed upwards (not shown in the figure).
(ii) the force on the floor by the block (action); the force on the block by the floor (reaction).

For (b): (i) the force of gravity (270 N) on the system by the earth (say, action); the force of gravity on the earth by the system (reaction), equal to 270 N,

directed upwards (not shown in the figure).

(ii) the force on the floor by the system (action); the force on the system by the floor (reaction). In addition, for (b), the force on the block by the cylinder and the force on the cylinder by the block also constitute an action-reaction pair.

The important thing to remember is that an action-reaction pair consists of mutual forces which are always equal and opposite between two bodies. Two forces on the same body which happen to be equal and opposite can never constitute an action-reaction pair. The force of gravity on the mass in (a) or (b) and the normal force on the mass by the floor are not action-reaction pairs. These forces happen to be equal and opposite for (a) since the mass is at rest. They are not so for case (b), as seen already. The weight of the system is 270 N, while the normal force R' is 267.3 N. ◀

The practice of drawing free-body diagrams is of great help in solving problems in mechanics. It allows you to clearly define your system and consider all forces on the system due to objects that are not part of the system itself. A number of exercises in this and subsequent chapters will help you cultivate this practice.

SUMMARY

1. Aristotle's view that a force is necessary to keep a body in uniform motion is wrong. A force is necessary in practice to counter the opposing force of friction.
2. Galileo extrapolated simple observations on motion of bodies on inclined planes, and arrived at the law of inertia. Newton's first law of motion is the same law rephrased thus: *"Everybody continues to be in its state of rest or of uniform motion in a straight line, unless compelled by some external force to act otherwise"*. In simple terms, the First Law is **"If external force on a body is zero, its acceleration is zero"**.
3. Momentum ($\mathbf{p}$) of a body is the product of its mass *(m)* and velocity *(**v**)* :
$$\mathbf{p} = m\,\mathbf{v}$$
4. Newton's second law of motion :
The rate of change of momentum of a body is proportional to the applied force and takes place in the direction in which the force acts. Thus
$$\mathbf{F} = k\frac{d\mathbf{p}}{dt} = k\,m\,\mathbf{a}$$
where $\mathbf{F}$ is the net external force on the body and $\mathbf{a}$ its acceleration. We set the constant of proportionality $k = 1$ in S.I. Then
$$\mathbf{F} = \frac{d\mathbf{p}}{dt} = m\mathbf{a}$$
The SI unit of force is newton : 1 N = 1 kg m s^{-2}.
   (a) The second law is consistent with the First Law ($\mathbf{F} = 0$ implies $\mathbf{a} = 0$)
   (b) It is a vector equation
   (c) It is applicable to a particle, and also to a body or a system of particles, provided $\mathbf{F}$ is the total external force on the system and $\mathbf{a}$ is the acceleration of the system as a whole.
   (d) $\mathbf{F}$ at a point at a certain instant determines $\mathbf{a}$ at the same point at that instant. That is the Second Law is a local law; $\mathbf{a}$ at an instant does not depend on the history of motion.
5. Impulse is the product of force and time which equals change in momentum.
The notion of impulse is useful when a large force acts for a short time to produce a measurable change in momentum. Since the time of action of the force is very short, one can assume that there is no appreciable change in the position of the body during the action of the impulsive force.
6. Newton's third law of motion:
To every action, there is always an equal and opposite reaction

In simple terms, the law can be stated thus :

Forces in nature always occur between pairs of bodies. Force on a body A by body B is equal and opposite to the force on the body B by A.

Action and reaction forces are simultaneous forces. There is no cause-effect relation between action and reaction. Any of the two mutual forces can be called action and the other reaction. Action and reaction act on different bodies and so they cannot be cancelled out. The internal action and reaction forces between different parts of a body do, however, sum to zero.

7. *Law of Conservation of Momentum*

The total momentum of an isolated system of particles is conserved. The law follows from the second and third law of motion.

8. *Friction*

Frictional force opposes (impending or actual) relative motion between two surfaces in contact. It is the component of the contact force along the common tangent to the surface in contact. Static friction f_s opposes impending relative motion; kinetic friction f_k opposes actual relative motion. They are independent of the area of contact and satisfy the following approximate laws :

$$f_s \le (f_s)_{max} = \mu_s R$$

$$f_k = \mu_k R$$

μ_s (co-efficient of static friction) and μ_k (co-efficient of kinetic friction) are constants characteristic of the pair of surfaces in contact. It is found experimentally that μ_k is less than μ_s .

Quantity	Symbol	Units	Dimensions	Remarks
Momentum	p	kg m s^{-1} or N s	$[MLT^{-1}]$	Vector
Force	F	N	$[MLT^{-2}]$	F = ma Second Law
Impulse		kg m s^{-1} or N s	$[M\ LT^{-1}]$	Impulse = force x time = change in momentum
Static friction	f_s	N	$[MLT^{-2}]$	$f_s \le \mu_s$ N
Kinetic friction	f_k	N	$[MLT^{-2}]$	$f_k = \mu_k$ N

POINTS TO PONDER

1. Force is not always in the direction of motion. Depending on the situation, **F** may be along **v**, opposite to **v**, normal to **v** or may make some other angle with **v**. In every case, it is parallel to acceleration.
2. If **v** = 0 at an instant, i.e. if a body is momentarily at rest, it does not mean that force or acceleration are necessarily zero at that instant. For example, when a ball thrown upward reaches its maximum height, **v** = 0 but the force continues to be its weight *mg* and the acceleration is not zero but *g*.
3. Force on a body at a given time is determined by the situation at the location of the body at that time. Force is not 'carried' by the body from its earlier history of motion. The moment after a stone is released out of an accelerated train, there is no horizontal force (or acceleration) on the stone, if the effects of the surrounding air are neglected. The stone then has only the vertical force of gravity.
4. In the second law of motion **F** = *m* **a**, **F** stands for the net force due to all material agencies external to the body. **a** is the effect of the force. *m***a** should

not be regarded as yet another force, besides **F**.

5. The centripetal force should not be regarded as yet another kind of force. It is simply a name given to the force that provides inward radial acceleration to a body in circular motion. We should always look for some material force like tension, gravitational force, electrical force, friction, etc as the centripetal force in any circular motion.
6. Static friction is a self-adjusting force up to its limit $\mu_s N$ ($f_s \leq \mu_s N$). Do not put $f_s = \mu_s N$ without being sure that the maximum value of static friction is coming into play.
7. The familiar equation $mg = R$ for a body on a table is true only if the body is in equilibrium. The two forces mg and R can be different (e.g. a body in an accelerated lift). The equality of mg and R has no connection with the third law.
8. The terms 'action' and 'reaction' in the third Law of Motion simply stand for simultaneous mutual forces between a pair of bodies. Unlike their meaning in ordinary language, action does not precede or cause reaction. Action and reaction act on different bodies.
9. The different terms like 'friction', 'normal reaction' 'tension', 'air resistance', 'viscous drag', 'thrust', 'buoyancy' 'weight' 'centripetal force' all stand for 'force' in different contexts. For clarity, every force and its equivalent terms encountered in mechanics should be reduced to the phrase 'force on *A* by *B*'.
10. For applying the second law of motion, there is no conceptual distinction between inanimate and animate objects. An animate object such as a human also requires an external force to accelerate. For example, without the external force of friction, we cannot walk on the ground.
11. The objective concept of force in physics should not be confused with the subjective concept of the 'feeling of force'. On a merry-go-around, all parts of our body are subject to an inward force, but we have a feeling of being pushed outward – the direction of impending motion.

EXERCISES

(For simplicity in numerical calculations, take g = 10 m s^{-2})

5.1 Give the magnitude and direction of the net force acting on
(a) a drop of rain falling down with a constant speed,
(b) a cork of mass 10 g floating on water,
(c) a kite skillfully held stationary in the sky,
(d) a car moving with a constant velocity of 30 km/h on a rough road,
(e) a high-speed electron in space far from all material objects, and free of electric and magnetic fields.

5.2 A pebble of mass 0.05 kg is thrown vertically upwards. Give the direction and magnitude of the net force on the pebble,
(a) during its upward motion,
(b) during its downward motion,
(c) at the highest point where it is momentarily at rest. Do your answers change if the pebble was thrown at an angle of 45° with the horizontal direction?

Ignore air resistance.

5.3 Give the magnitude and direction of the net force acting on a stone of mass 0.1 kg,
(a) just after it is dropped from the window of a stationary train,
(b) just after it is dropped from the window of a train running at a constant velocity of 36 km/h,
(c) just after it is dropped from the window of a train accelerating with 1 m s^{-2},
(d) lying on the floor of a train which is accelerating with 1 m s^{-2}, the stone being at rest relative to the train.

Neglect air resistance throughout.

5.4 One end of a string of length l is connected to a particle of mass m and the other to a small peg on a smooth horizontal table. If the particle moves in a circle with speed v the net force on the particle (directed towards the centre) is :

(i) T, (ii) $T - \frac{mv^2}{l}$, (iii) $T + \frac{mv^2}{l}$, (iv) 0

T is the tension in the string. [Choose the correct alternative].

5.5 A constant retarding force of 50 N is applied to a body of mass 20 kg moving initially with a speed of 15 m s^{-1}. How long does the body take to stop ?

5.6 A constant force acting on a body of mass 3.0 kg changes its speed from 2.0 m s^{-1} to 3.5 m s^{-1} in 25 s. The direction of the motion of the body remains unchanged. What is the magnitude and direction of the force ?

5.7 A body of mass 5 kg is acted upon by two perpendicular forces 8 N and 6 N. Give the magnitude and direction of the acceleration of the body.

5.8 The driver of a three-wheeler moving with a speed of 36 km/h sees a child standing in the middle of the road and brings his vehicle to rest in 4.0 s just in time to save the child. What is the average retarding force on the vehicle ? The mass of the three-wheeler is 400 kg and the mass of the driver is 65 kg.

5.9 A rocket with a lift-off mass 20,000 kg is blasted upwards with an initial acceleration of 5.0 m s^{-2}. Calculate the initial thrust (force) of the blast.

5.10 A body of mass 0.40 kg moving initially with a constant speed of 10 m s^{-1} to the north is subject to a constant force of 8.0 N directed towards the south for 30 s. Take the instant the force is applied to be $t = 0$, the position of the body at that time to be $x = 0$, and predict its position at t = –5 s, 25 s, 100 s.

5.11 A truck starts from rest and accelerates uniformly at 2.0 m s^{-2}. At $t = 10$ s, a stone is dropped by a person standing on the top of the truck (6 m high from the ground). What are the (a) velocity, and (b) acceleration of the stone at t = 11s ? (Neglect air resistance.)

5.12 A bob of mass 0.1 kg hung from the ceiling of a room by a string 2 m long is set into oscillation. The speed of the bob at its mean position is 1 m s^{-1}. What is the trajectory of the bob if the string is cut when the bob is (a) at one of its extreme positions, (b) at its mean position.

5.13 A man of mass 70 kg stands on a weighing scale in a lift which is moving
(a) upwards with a uniform speed of 10 m s^{-1},
(b) downwards with a uniform acceleration of 5 m s^{-2},
(c) upwards with a uniform acceleration of 5 m s^{-2}.
What would be the readings on the scale in each case?
(d) What would be the reading if the lift mechanism failed and it hurtled down freely under gravity ?

5.14 Figure 5.16 shows the position-time graph of a particle of mass 4 kg. What is the (a) force on the particle for $t < 0$, $t > 4$ s, $0 < t < 4$ s? (b) impulse at $t = 0$ and $t = 4$ s ? (Consider one-dimensional motion only).

Fig. 5.16

5.15 Two bodies of masses 10 kg and 20 kg respectively kept on a smooth, horizontal surface are tied to the ends of a light string. a horizontal force F = 600 N is applied to (i) A, (ii) B along the direction of string. What is the tension in the

string in each case?

5.16 Two masses 8 kg and 12 kg are connected at the two ends of a light inextensible string that goes over a frictionless pulley. Find the acceleration of the masses, and the tension in the string when the masses are released.

5.17 A nucleus is at rest in the laboratory frame of reference. Show that if it disintegrates into two smaller nuclei the products must move in opposite directions.

5.18 Two billiard balls each of mass 0.05 kg moving in opposite directions with speed 6 m s^{-1} collide and rebound with the same speed. What is the impulse imparted to each ball due to the other ?

5.19 A shell of mass 0.020 kg is fired by a gun of mass 100 kg. If the muzzle speed of the shell is 80 m s^{-1}, what is the recoil speed of the gun ?

5.20 A batsman deflects a ball by an angle of 45° without changing its initial speed which is equal to 54 km/h. What is the impulse imparted to the ball ? (Mass of the ball is 0.15 kg.)

5.21 A stone of mass 0.25 kg tied to the end of a string is whirled round in a circle of radius 1.5 m with a speed of 40 rev./min in a horizontal plane. What is the tension in the string ? What is the maximum speed with which the stone can be whirled around if the string can withstand a maximum tension of 200 N ?

5.22 If, in Exercise 5.21, the speed of the stone is increased beyond the maximum permissible value, and the string breaks suddenly, which of the following correctly describes the trajectory of the stone after the string breaks :
(a) the stone moves radially outwards,
(b) the stone flies off tangentially from the instant the string breaks,
(c) the stone flies off at an angle with the tangent whose magnitude depends on the speed of the particle ?

5.23 Explain why
(a) a horse cannot pull a cart and run in empty space,
(b) passengers are thrown forward from their seats when a speeding bus stops suddenly,
(c) it is easier to pull a lawn mower than to push it,
(d) a cricketer moves his hands backwards while holding a catch.

Additional Exercises

5.24 Figure 5.17 shows the position-time graph of a body of mass 0.04 kg. Suggest a suitable physical context for this motion. What is the time between two consecutive impulses received by the body ? What is the magnitude of each impulse ?

Fig. 5.17

5.25 Figure 5.18 shows a man standing stationary with respect to a horizontal conveyor belt that is accelerating with 1 m s^{-2}. What is the net force on the man? If the coefficient of static friction between the man's shoes and the belt is 0.2, up to what acceleration of the belt can the man continue to be stationary relative to the belt ? (Mass of the man = 65 kg.)

Fig. 5.18

5.26 A stone of mass m tied to the end of a string revolves in a vertical circle of radius R. The net forces at the lowest and highest points of the circle directed vertically downwards are : [Choose the correct alternative]

	Lowest Point	Highest Point
(a)	$mg - T_1$	$mg + T_2$
(b)	$mg + T_1$	$mg - T_2$
(c)	$mg + T_1 - (m v_1^2) / R$	$mg - T_2 + (m v_1^2) / R$
(d)	$mg - T_1 - (m v_1^2) / R$	$mg + T_2 + (m v_1^2) / R$

T_1 and v_1 denote the tension and speed at the lowest point. T_2 and v_2 denote corresponding values at the highest point.

5.27 A helicopter of mass 1000 kg rises with a vertical acceleration of 15 m s^{-2}. The crew and the passengers weigh 300 kg. Give the magnitude and direction of the
(a) force on the floor by the crew and passengers,
(b) action of the rotor of the helicopter on the surrounding air,
(c) force on the helicopter due to the surrounding air.

5.28 A stream of water flowing horizontally with a speed of 15 m s^{-1} gushes out of a tube of cross-sectional area 10^{-2} m^2, and hits a vertical wall nearby. What is the force exerted on the wall by the impact of water, assuming it does not rebound ?

5.29 Ten one-rupee coins are put on top of each other on a table. Each coin has a mass m. Give the magnitude and direction of
(a) the force on the 7th coin (counted from the bottom) due to all the coins on its top,
(b) the force on the 7th coin by the eighth coin,
(c) the reaction of the 6th coin on the 7th coin.

5.30 An aircraft executes a horizontal loop at a speed of 720 km/h with its wings banked at 15°. What is the radius of the loop ?

5.31 A train runs along an unbanked circular track of radius 30 m at a speed of 54 km/h. The mass of the train is 10^6 kg. What provides the centripetal force required for this purpose — The engine or the rails ? What is the angle of banking required to prevent wearing out of the rail ?

5.32 A block of mass 25 kg is raised by a 50 kg man in two different ways as shown in Fig. 5.19. What is the action on the floor by the man in the two cases ? If the floor yields to a normal force of 700 N, which mode should the man adopt to lift the block without the floor yielding ?

Fig. 5.19

5.33 A monkey of mass 40 kg climbs on a rope (Fig. 5.20) which can stand a maximum tension of 600 N. In which of the following cases will the rope break: the monkey
(a) climbs up with an acceleration of 6 m s^{-2}
(b) climbs down with an acceleration of 4 m s^{-2}
(c) climbs up with a uniform speed of 5 m s^{-1}
(d) falls down the rope nearly freely under gravity?
(Ignore the mass of the rope).

Fig. 5.20

5.34 Two bodies *A* and *B* of masses 5 kg and 10 kg in contact with each other rest on a table against a rigid wall (Fig. 5.21). The coefficient of friction between the bodies and the table is 0.15. A force of 200 N is applied horizontally to *A*. What are (a) the reaction of the partition (b) the action-reaction forces between *A* and *B* ? What happens when the wall is removed? Does the answer to (b) change, when the bodies are in motion? Ignore the difference between μ_s and μ_k.

Fig. 5.21

5.35 A block of mass 15 kg is placed on a long trolley. The coefficient of static friction between the block and the trolley is 0.18. The trolley accelerates from rest with 0.5 m s^{-2} for 20 s and then moves with uniform velocity. Discuss the motion of the block as viewed by (a) a stationary observer on the ground, (b) an observer moving with the trolley.

5.36 The rear side of a truck is open and a box of 40 kg mass is placed 5 m away from the open end as shown in Fig. 5.22. The coefficient of friction between the box and the surface below it is 0.15. On a straight road, the truck starts from rest and accelerates with 2 m s^{-2}. At what distance from the starting point does the box fall off the truck? (Ignore the size of the box).

Fig. 5.22

5.37 A disc revolves with a speed of $33\frac{1}{3}$ rev/min, and has a radius of 15 cm. Two coins are placed at 4 cm and 14 cm away from the centre of the record. If the co-efficient of friction between the coins and the record is 0.15, which of the coins will revolve with the record ?

5.38 You may have seen in a circus a motorcyclist driving in vertical loops inside a 'death-well' (a hollow spherical chamber with holes, so the spectators can watch from outside). Explain clearly why the motorcyclist does not drop down when he is at the uppermost point, with no support from below. What is the minimum speed required at the uppermost position to perform a vertical loop if the radius of the chamber is 25 m ?

5.39 A 70 kg man stands in contact against the inner wall of a hollow cylindrical drum of radius 3 m rotating about its vertical axis with 200 rev/min. The coefficient of friction between the wall and his clothing is 0.15. What is the minimum rotational speed of the cylinder to enable the man to remain stuck to the wall (without falling) when the floor is suddenly removed ?

5.40 A thin circular loop of radius *R* rotates about its vertical diameter with an angular frequency ω Show that a small bead on the wire loop remains at its lowermost point for $\omega \le \sqrt{g/R}$. What is the angle made by the radius vector joining the centre to the bead with the vertical downward direction for $\omega = \sqrt{2g/R}$? Neglect friction.

CHAPTER SIX

WORK, ENERGY AND POWER

6.1 Introduction
6.2 Notions of work and kinetic energy : The work-energy theorem
6.3 Work
6.4 Kinetic energy
6.5 Work done by a variable force
6.6 The work-energy theorem for a variable force
6.7 The concept of potential energy
6.8 The conservation of mechanical energy
6.9 The potential energy of a spring
6.10 Various forms of energy : the law of conservation of energy
6.11 Power
6.12 Collisions

Summary
Points to ponder
Exercises
Additional exercises
Appendix 6.1

6.1 INTRODUCTION

The terms 'work', 'energy' and 'power' are frequently used in everyday language. A farmer ploughing the field, a construction worker carrying bricks, a student studying for a competitive examination, an artist painting a beautiful landscape, all are said to be working. In physics, however, the word 'Work' covers a definite and precise meaning. Somebody who has the capacity to work for 14-16 hours a day is said to have a large stamina or energy. We admire a long distance runner for her stamina or energy. Energy is thus our capacity to do work. In Physics too, the term 'energy' is related to work in this sense, but as said above the term 'work' itself is defined much more precisely. The word 'power' is used in everyday life with different shades of meaning. In karate or boxing we talk of 'powerful' punches. These are delivered at a great speed. This shade of meaning is close to the meaning of the word 'power' used in physics. We shall find that there is at best a loose correlation between the physical definitions and the physiological pictures these terms generate in our minds. The aim of this chapter is to develop an understanding of these three physical quantities. Before we proceed to this task, we need to develop a mathematical prerequisite, namely the scalar product of two vectors.

6.1.1 The Scalar Product

We have learnt about vectors and their use in Chapter 4. Physical quantities like displacement, velocity, acceleration, force etc. are vectors. We have also learnt how vectors are added or subtracted. We now need to know how vectors are multiplied. There are two ways of multiplying vectors which we shall come across : one way known as the scalar product gives a scalar from two vectors and the other known as the vector product produces a new vector from two vectors. We shall look at the vector product in Chapter 7. Here we take up the scalar product of two vectors. The scalar product or dot product of any two vectors **A** and **B**, denoted as **A·B** (read

A dot **B**) is defined as

$$\mathbf{A}\cdot\mathbf{B} = A\,B\cos\theta \qquad (6.1a)$$

where θ is the angle between the two vectors as shown in Fig. 6.1a. Since A, B and $\cos\theta$ are scalars, the dot product of **A** and **B** is a scalar quantity. Each vector, **A** and **B**, has a direction but their scalar product does not have a direction.

From Eq. (6.1a), we have

$$\mathbf{A}\cdot\mathbf{B} = A\,(B\cos\theta)$$
$$= B\,(A\cos\theta)$$

Geometrically, $B\cos\theta$ is the projection of **B** onto **A** in Fig.6.1 (b) and $A\cos\theta$ is the projection of **A** onto **B** in Fig. 6.1 (c). So, **A·B** is the product of the magnitude of **A** and the component of **B** along **A**. Alternatively, it is the product of the magnitude of **B** and the component of **A** along **B**.

Equation (6.1a) shows that the scalar product follows the commutative law :

$$\mathbf{A}\cdot\mathbf{B} = \mathbf{B}\cdot\mathbf{A}$$

Scalar product obeys the **distributive law**:

$$\mathbf{A}\cdot(\mathbf{B}+\mathbf{C}) = \mathbf{A}\cdot\mathbf{B} + \mathbf{A}\cdot\mathbf{C}$$

Further, $\mathbf{A}\cdot(\lambda\,\mathbf{B}) = \lambda\,(\mathbf{A}\cdot\mathbf{B})$

where λ is a real number.

The proofs of the above equations are left to you as an exercise.

For unit vectors $\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}$ we have

$$\hat{\mathbf{i}}\cdot\hat{\mathbf{i}} = \hat{\mathbf{j}}\cdot\hat{\mathbf{j}} = \hat{\mathbf{k}}\cdot\hat{\mathbf{k}} = 1$$
$$\hat{\mathbf{i}}\cdot\hat{\mathbf{j}} = \hat{\mathbf{j}}\cdot\hat{\mathbf{k}} = \hat{\mathbf{k}}\cdot\hat{\mathbf{i}} = 0$$

Given two vectors

$$\mathbf{A} = A_x\hat{\mathbf{i}} + A_y\hat{\mathbf{j}} + A_z\hat{\mathbf{k}}$$
$$\mathbf{B} = B_x\hat{\mathbf{i}} + B_y\hat{\mathbf{j}} + B_z\hat{\mathbf{k}}$$

their scalar product is

$$\mathbf{A}\cdot\mathbf{B} = \left(A_x\hat{\mathbf{i}} + A_y\hat{\mathbf{j}} + A_z\hat{\mathbf{k}}\right)\cdot\left(B_x\hat{\mathbf{i}} + B_y\hat{\mathbf{j}} + B_z\hat{\mathbf{k}}\right)$$
$$= A_xB_x + A_yB_y + A_zB_z \qquad (6.1b)$$

From the definition of scalar product and, (Eq. 6.1b) we have :

(i) $\mathbf{A}\cdot\mathbf{A} = A_xA_x + A_yA_y + A_zA_z$

Or, $$A^2 = A_x^2 + A_y^2 + A_z^2 \qquad (6.1c)$$

since $\mathbf{A}\cdot\mathbf{A} = |\mathbf{A}|\,|\mathbf{A}|\cos 0 = A^2$.

(ii) $\mathbf{A}\cdot\mathbf{B} = 0$, if **A** and **B** are perpendicular.

▶ ***Example 6.1*** Find the angle between force $\mathbf{F} = (3\,\hat{i} + 4\,\hat{j}\quad 5\,\hat{k})$ unit and displacement $\mathbf{d} = (5\,\hat{i} + 4\,\hat{j} + 3\,\hat{k})$ unit. Also find the projection of **F** on **d**.

Answer $\mathbf{F}\cdot\mathbf{d} = F_xd_x + F_yd_y + F_zd_z$
$= 3\,(5) + 4\,(4) + (-3)\,(3)$
$= 16$ unit

Hence $\mathbf{F}\cdot\mathbf{d} = F\,d\,\cos\theta = 16$ unit

Now $\mathbf{F}\cdot\mathbf{F} = F^2 = F_x^2 + F_y^2 + F_z^2$
$= 9 + 16 + 25$
$= 50$ unit

and $\mathbf{d}\cdot\mathbf{d} = d^2 = d_x^2 + d_y^2 + d_z^2$
$= 25 + 16 + 9$
$= 50$ unit

$$\therefore \cos\theta = \frac{16}{\sqrt{50}\sqrt{50}} = \frac{16}{50} = 0.32,$$
$$\theta = \cos^{-1} 0.32$$ ◀

Fig. 6.1 (a) *The scalar product of two vectors* **A** *and* **B** *is a scalar :* $\mathbf{A}\cdot\mathbf{B} = A\,B\cos\theta$. (b) $B\cos\theta$ *is the projection of* **B** *onto* **A**. (c) $A\cos\theta$ *is the projection of* **A** *onto* **B**.

6.2 NOTIONS OF WORK AND KINETIC ENERGY: THE WORK-ENERGY THEOREM

The following relation for rectilinear motion under constant acceleration a has been encountered in Chapter 3,

$$v^2 - u^2 = 2\,as$$

where u and v are the initial and final speeds and s the distance traversed. Multiplying both sides by $m/2$, we have

$$\frac{1}{2}mv^2 - \frac{1}{2}mu^2 = mas = Fs \qquad (6.2a)$$

where the last step follows from Newton's Second Law. We can generalise Eq. (6.1) to three dimensions by employing vectors

$$v^2 - u^2 = 2\,\mathbf{a.d}$$

Once again multiplying both sides by $m/2$, we obtain

$$\frac{1}{2}mv^2 - \frac{1}{2}mu^2 = m\,a.d = F.d \qquad (6.2b)$$

The above equation provides a motivation for the definitions of work and kinetic energy. The left side of the equation is the difference in the quantity 'half the mass times the square of the speed' from its initial value to its final value. We call each of these quantities the 'kinetic energy', denoted by K. The right side is a product of the displacement and the component of the force along the displacement. This quantity is called 'work' and is denoted by W. Eq. (6.2) is then

$$K_f - K_i = W \qquad (6.3)$$

where K_i and K_f are respectively the initial and final kinetic energies of the object. Work refers to the force and the displacement over which it acts. **Work is done by a force on the body over a certain displacement**.

Equation (6.2) is also a special case of the work-energy (WE) theorem : **The change in kinetic energy of a particle is equal to the work done on it by the net force**. We shall generalise the above derivation to a varying force in a later section.

▶ ***Example 6.2*** It is well known that a raindrop falls under the influence of the downward gravitational force and the opposing resistive force. The latter is known to be proportional to the speed of the drop but is otherwise undetermined. Consider a drop of mass 1.00 g falling from a height 1.00 km. It hits the ground with a speed of 50.0 m s^{-1}. (a) What is the work done by the gravitational force ? What is the work done by the unknown resistive force?

Answer (a) The change in kinetic energy of the drop is

$$\Delta K = \frac{1}{2}m\,v^2 - 0$$

$$= \frac{1}{2}\times 10^{-3}\times 50\times 50$$

$$= 1.25 \text{ J}$$

where we have assumed that the drop is initially at rest.

Assuming that g is a constant with a value 10 m/s^2, the work done by the gravitational force is,

$$W_g = mgh$$
$$= 10^{-3}\times 10\times 10^3$$
$$= 10.0 \text{ J}$$

(b) From the work-energy theorem

$$\Delta K = W_g + W_r$$

where W_r is the work done by the resistive force on the raindrop. Thus

$$W_r = \Delta K - W_g$$
$$= 1.25 - 10$$
$$= -8.75 \text{ J}$$

is negative. ◀

6.3 WORK

As seen earlier, work is related to force and the displacement over which it acts. Consider a constant force **F** acting on an object of mass m. The object undergoes a displacement **d** in the positive x-direction as shown in Fig. 6.2.

Fig. 6.2 An object undergoes a displacement **d** *under the influence of the force* **F**.

The work done by the force is defined to be the product of component of the force in the direction of the displacement and the magnitude of this displacement. Thus

$$W = (F\cos\theta)d = \mathbf{F.d} \qquad (6.4)$$

We see that if there is no displacement, there is no work done even if the force is large. Thus, when you push hard against a rigid brick wall, the force you exert on the wall does no work. Yet your muscles are alternatively contracting and relaxing and internal energy is being used up and you do get tired. Thus, the meaning of work in physics is different from its usage in everyday language.

No work is done if :

(i) the displacement is zero as seen in the example above. A weightlifter holding a 150 kg mass steadily on his shoulder for 30 s does no work on the load during this time.

(ii) the force is zero. A block moving on a smooth horizontal table is not acted upon by a horizontal force (since there is no friction), but may undergo a large displacement.

(iii) the force and displacement are mutually perpendicular. This is so since, for $\theta = \pi/2$ rad (= 90°), cos ($\pi/2$) = 0. For the block moving on a smooth horizontal table, the gravitational force mg does no work since it acts at right angles to the displacement. If we assume that the moon's orbits around the earth is perfectly circular then the earth's gravitational force does no work. The moon's instantaneous displacement is tangential while the earth's force is radially inwards and $\theta = \pi/2$.

Work can be both positive and negative. If θ is between 0° and 90°, cos θ in Eq. (6.4) is positive. If θ is between 90° and 180°, cos θ is negative. In many examples the frictional force opposes displacement and θ = 180°. Then the work done by friction is negative (cos 180° = –1).

From Eq. (6.4) it is clear that work and energy have the same dimensions, $[ML^2T^{-2}]$. The SI unit of these is joule (J), named after the famous British physicist James Prescott Joule (1811-1869). Since work and energy are so widely used as physical concepts, alternative units abound and some of these are listed in Table 6.1.

Table 6.1 Alternative Units of Work/Energy in J

erg	10^{-7} J
electron volt (eV)	1.6×10^{-19} J
calorie (cal)	4.186 J
kilowatt hour (kWh)	3.6×10^{6} J

▶ ***Example 6.3*** A cyclist comes to a skidding stop in 10 m. During this process, the force on the cycle due to the road is 200 N and is directly opposed to the motion. (a) How much work does the road do on the cycle ? (b) How much work does the cycle do on the road ?

Answer Work done on the cycle by the road is the work done by the stopping (frictional) force on the cycle due to the road.

(a) The stopping force and the displacement make an angle of 180° (π rad) with each other. Thus, work done by the road,

$$W_r = Fd\cos\theta$$
$$= 200 \times 10 \times \cos\pi$$
$$= -2000\ \text{J}$$

It is this negative work that brings the cycle to a halt in accordance with WE theorem.

(b) From Newton's Third Law an equal and opposite force acts on the road due to the cycle. Its magnitude is 200 N. However, the road undergoes no displacement. Thus, work done by cycle on the road is zero. ◀

The lesson of this example is that though the force on a body A exerted by the body B is always equal and opposite to that on B by A (Newton's Third Law); the work done on A by B is not necessarily equal and opposite to the work done on B by A.

6.4 KINETIC ENERGY

As noted earlier, if an object of mass m has velocity $\mathbf{v}$, its kinetic energy K is

$$K = \frac{1}{2}m\,\mathbf{v.v} = \frac{1}{2}mv^2 \qquad (6.5)$$

Kinetic energy is a scalar quantity. The kinetic energy of an object is a measure of the work an

Table 6.2 Typical kinetic energies (K)

Object	Mass (kg)	Speed (m s^{-1})	K (J)
Car	2000	25	6.3×10^{5}
Running athlete	70	10	3.5×10^{3}
Bullet	5×10^{-2}	200	10^{3}
Stone dropped from 10 m	1	14	10^{2}
Rain drop at terminal speed	3.5×10^{-5}	9	1.4×10^{-3}
Air molecule	$\simeq 10^{-26}$	500	$\simeq 10^{-21}$

object can do by the virtue of its motion. This notion has been intuitively known for a long time. The kinetic energy of a fast flowing stream has been used to grind corn. Sailing ships employ the kinetic energy of the wind. Table 6.2 lists the kinetic energies for various objects.

▶ ***Example 6.4*** In a ballistics demonstration a police officer fires a bullet of mass 50.0 g with speed 200 m s^{-1} (see Table 6.2) on soft plywood of thickness 2.00 cm. The bullet emerges with only 10% of its initial kinetic energy. What is the emergent speed of the bullet ?

Answer The initial kinetic energy of the bullet is $mv^2/2$ = 1000 J. It has a final kinetic energy of 0.1×1000 = 100 J. If v_f is the emergent speed of the bullet,

$$\frac{1}{2}mv_f^2 = 100 \text{ J}$$

$$v_f = \sqrt{\frac{2\times100 \text{ J}}{0.05 \text{ kg}}}$$

$$= 63.2 \text{ m s}^{-1}$$

The speed is reduced by approximately 68% (not 90%). ◀

6.5 WORK DONE BY A VARIABLE FORCE

A constant force is rare. It is the variable force, which is more commonly encountered. Fig. 6.2 is a plot of a varying force in one dimension.

If the displacement Δx is small, we can take the force $F(x)$ as approximately constant and the work done is then

$$\Delta W = F(x)\,\Delta x$$

This is illustrated in Fig. 6.3(a). Adding successive rectangular areas in Fig. 6.3(a) we get the total work done as

$$W \cong \sum_{x_i}^{x_f} F(x)\Delta x \qquad (6.6)$$

where the summation is from the initial position x_i to the final position x_f.

If the displacements are allowed to approach zero, then the number of terms in the sum increases without limit, but the sum approaches a definite value equal to the area under the curve in Fig. 6.3(b). Then the work done is

$$W = \lim_{\Delta x \to 0} \sum_{x_i}^{x_f} F(x)\Delta x$$

$$= \int_{x_i}^{x_f} F(x)\,dx \qquad (6.7)$$

where '*lim*' stands for the limit of the sum when Δx tends to zero. Thus, for a varying force the work done can be expressed as a definite integral of force over displacement (see also Appendix 3.1).

Fig. 6.3(a)

Fig. 6.3 (a) The shaded rectangle represents the work done by the varying force F(x), over the small displacement Δx, ΔW = F(x)Δx. (b) adding the areas of all the rectangles we find that for Δx →0, the area under the curve is exactly equal to the work done by F(x).

▶ ***Example 6.5*** A woman pushes a trunk on a railway platform which has a rough surface. She applies a force of 100 N over a distance of 10 m. Thereafter, she gets progressively tired and her applied force reduces linearly with distance to 50 N. The total distance through which the trunk has been moved is 20 m. Plot the force applied by the woman and the frictional force, which is 50 N. Calculate the work done by the two forces over 20 m.

Answer

Fig. 6.4 Plot of the force F applied by the woman and the opposing frictional force f.

The plot of the applied force is shown in Fig. 6.4. At x = 20 m, F = 50 N (≠ 0). We are given that the frictional force f is $|\mathbf{f}|$ = 50 N. It opposes motion and acts in a direction opposite to **F**. It is therefore, shown on the negative side of the force axis.

The work done by the woman is

W_F →area of the rectangle ABCD + area of the trapezium CEID

$$W_F = 100\times10 + \frac{1}{2}(100+50)\times10$$
$$= 1000 + 750$$
$$= 1750 \text{ J}$$

The work done by the frictional force is

W_F →area of the rectangle AGHI

$$W_f = (-50)\times 20$$
$$= -1000 \text{ J}$$

The area on the negative side of the force axis has a negative sign. ◀

6.6 THE WORK-ENERGY THEOREM FOR A VARIABLE FORCE

We are now familiar with the concepts of work and kinetic energy to prove the work-energy theorem for a variable force. We confine ourselves to one dimension. The time rate of change of kinetic energy is

$$\frac{dK}{dt} = \frac{d}{dt}\left(\frac{1}{2}m\,v^2\right)$$
$$= m\frac{dv}{dt}v$$
$$= F\,v \text{ (from Newton's Second Law)}$$
$$= F\frac{dx}{dt}$$

Thus

$$dK = F dx$$

Integrating from the initial position (x_i) to final position (x_f), we have

$$\int_{K_i}^{K_f} dK = \int_{x_i}^{x_f} F dx$$

where, K_i and K_f are the initial and final kinetic energies corresponding to x_i and x_f.

or $$K_f - K_i = \int_{x_i}^{x_f} F dx \qquad (6.8a)$$

From Eq. (6.7), it follows that

$$K_f - K_i = W \qquad (6.8b)$$

Thus, the WE theorem is proved for a variable force.

While the WE theorem is useful in a variety of problems, it does not, in general, incorporate the complete dynamical information of Newton's Second Law. It is an integral form of Newton's second law. Newton's second law is a relation between acceleration and force at any instant of time. Work-energy theorem involves an integral over an interval of time. In this sense, the temporal (time) information contained in the statement of Newton's second law is 'integrated over' and is

not available explicitly. Another observation is that Newton's second law for two or three dimensions is in vector form whereas the work-energy theorem is in scalar form. In the scalar form, information with respect to directions contained in Newton's second law is not present.

▶ ***Example 6.6*** A block of mass m = 1 kg, moving on a horizontal surface with speed v_i = 2 ms^{-1} enters a rough patch ranging from x = 0.10 m to x = 2.01 m. The retarding force F_r on the block in this range is inversely proportional to x over this range,

$$F_r = \frac{-k}{x} \text{ for } 0.1 < x < 2.01 \text{ m}$$

$$= 0 \text{ for } x < 0.1\text{m and } x > 2.01 \text{ m}$$

where k = 0.5 J. What is the final kinetic energy and speed v_f of the block as it crosses this patch ?

Answer From Eq. (6.8a)

$$K_f = K_i + \int_{0.1}^{2.01} \frac{(-k)}{x} \mathrm{d}x$$

$$= \frac{1}{2} m v_i^2 - k \ln(x)\Big|_{0.1}^{2.01}$$

$$= \frac{1}{2} m v_i^2 - k \ln(2.01/0.1)$$

$$= 2 - 0.5 \ln(20.1)$$

$$= 2 - 1.5 = 0.5 \text{ J}$$

$$v_f = \sqrt{2K_f/m} = 1 \text{ m s}^{-1}$$

Here, note that ln is a symbol for the natural logarithm to the base e and not the logarithm to the base 10 [$\ln X = \log_e X = 2.303 \log_{10} X$]. ◀

6.7 THE CONCEPT OF POTENTIAL ENERGY

The word potential suggests possibility or capacity for action. The term potential energy brings to one's mind 'stored' energy. A stretched bow-string possesses potential energy. When it is released, the arrow flies off at a great speed. The earth's crust is not uniform, but has discontinuities and dislocations that are called fault lines. These fault lines in the earth's crust are like 'compressed springs'. They possess a large amount of potential energy. An earthquake results when these fault lines readjust. Thus, potential energy is the 'stored energy' by virtue of the position or configuration of a body. The body left to itself releases this stored energy in the form of kinetic energy. Let us make our notion of potential energy more concrete.

The gravitational force on a ball of mass m is mg. g may be treated as a constant near the earth surface. By 'near' we imply that the height h of the ball above the earth's surface is very small compared to the earth's radius R_E ($h << R_E$) so that we can ignore the variation of g near the earth's surface*. In what follows we have taken the upward direction to be positive. Let us raise the ball up to a height h. The work done by the external agency against the gravitational force is mgh. This work gets stored as potential energy. Gravitational potential energy of an object, as a function of the height h, is denoted by $V(h)$ and it is the negative of work done by the gravitational force in raising the object to that height.

$$V(h) = mgh$$

If h is taken as a variable, it is easily seen that the gravitational force F equals the negative of the derivative of $V(h)$ with respect to h. Thus,

$$F = -\frac{\mathrm{d}}{\mathrm{d}h} V(h) = -m\,g$$

The negative sign indicates that the gravitational force is downward. When released, the ball comes down with an increasing speed. Just before it hits the ground, its speed is given by the kinematic relation,

$$v^2 = 2gh$$

This equation can be written as

$$\frac{1}{2} m\, v^2 = m\, g\, h$$

which shows that the gravitational potential energy of the object at height h, when the object is released, manifests itself as kinetic energy of the object on reaching the ground.

Physically, the notion of potential energy is applicable only to the class of forces where work done against the force gets 'stored up' as energy. When external constraints are removed, it manifests itself as kinetic energy. Mathematically, (for simplicity, in one dimension) the potential

* *The variation of g with height is discussed in Chapter 8 on Gravitation.*

energy $V(x)$ is defined if the force $F(x)$ can be written as

$$F(x) = -\frac{dV}{dx}$$

This implies that

$$\int_{x_i}^{x_f} F(x)dx = -\int_{V_i}^{V_f} dV = V_i - V_f$$

The work done by a conservative force such as gravity depends on the initial and final positions only. In the previous chapter we have worked on examples dealing with inclined planes. If an object of mass m is released from rest, from the top of a smooth (frictionless) inclined plane of height h, its speed at the bottom is $\sqrt{2gh}$ irrespective of the angle of inclination. Thus, at the bottom of the inclined plane it acquires a kinetic energy, mgh. If the work done or the kinetic energy did depend on other factors such as the velocity or the particular path taken by the object, the force would be called non-conservative.

The dimensions of potential energy are $[ML^2T^{-2}]$ and the unit is joule (J), the same as kinetic energy or work. To reiterate, the change in potential energy, for a conservative force, ΔV is equal to the negative of the work done by the force

$$\Delta V = -F(x)\ \Delta x \qquad (6.9)$$

In the example of the falling ball considered in this section we saw how potential energy was converted to kinetic energy. This hints at an important principle of conservation in mechanics, which we now proceed to examine.

6.8 THE CONSERVATION OF MECHANICAL ENERGY

For simplicity we demonstrate this important principle for one-dimensional motion. Suppose that a body undergoes displacement Δx under the action of a conservative force F. Then from the WE theorem we have,

$$\Delta K = F(x)\ \Delta x$$

If the force is conservative, the potential energy function $V(x)$ can be defined such that

$$-\Delta V = F(x)\ \Delta x$$

The above equations imply that

$$\Delta K + \Delta V = 0$$

$$\Delta(K + V) = 0 \qquad (6.10)$$

which means that $K + V$, the sum of the kinetic and potential energies of the body is a constant. Over the whole path, x_i to x_f, this means that

$$K_i + V(x_i) = K_f + V(x_f) \qquad (6.11)$$

The quantity $K + V(x)$, is called the total mechanical energy of the system. Individually the kinetic energy K and the potential energy $V(x)$ may vary from point to point, but the sum is a constant. The aptness of the term 'conservative force' is now clear.

Let us consider some of the definitions of a conservative force.

- A force $F(x)$ is conservative if it can be derived from a scalar quantity $V(x)$ by the relation given by Eq. (6.9). The three-dimensional generalisation requires the use of a vector derivative, which is outside the scope of this book.
- The work done by the conservative force depends only on the end points. This can be seen from the relation,
  $W = K_f - K_i = V(x_i) - V(x_f)$
  which depends on the end points.
- A third definition states that the work done by this force in a closed path is zero. This is once again apparent from Eq. (6.11) since $x_i = x_f$.

Thus, the principle of conservation of total mechanical energy can be stated as

The total mechanical energy of a system is conserved if the forces, doing work on it, are conservative.

The above discussion can be made more concrete by considering the example of the gravitational force once again and that of the spring force in the next section. Fig. 6.5 depicts a ball of mass m being dropped from a cliff of height H.

Fig. 6.5 The conversion of potential energy to kinetic energy for a ball of mass m dropped from a height H.

The total mechanical energies E_0, E_h, and E_H of the ball at the indicated heights zero (ground level), h and H, are

$$E_H = mgH \quad (6.11\text{ a})$$

$$E_h = mgh + \frac{1}{2}mv_h^2 \quad (6.11\text{ b})$$

$$E_0 = (1/2)\, mv^2 \quad (6.11\text{ c})$$

The constant force is a special case of a spatially dependent force $F(x)$. Hence, the mechanical energy is conserved. Thus

$$E_H = E_0$$

or, $$mgH = \frac{1}{2}mv_f^2$$

$$v_f = \sqrt{2gH}$$

a result that was obtained in section 3.7 for a freely falling body.

Further,

$$E_H = E_h$$

which implies,

$$v_h^2 = 2g(H - h) \quad (6.11\text{ d})$$

and is a familiar result from kinematics.

At the height H, the energy is purely potential. It is partially converted to kinetic at height h and is fully kinetic at ground level. This illustrates the conservation of mechanical energy.

▶ ***Example 6.7*** A bob of mass m is suspended by a light string of length L. It is imparted a horizontal velocity v_o at the lowest point A such that it completes a semi-circular trajectory in the vertical plane with the string becoming slack only on reaching the topmost point, C. This is shown in Fig. 6.6. Obtain an expression for (i) v_o; (ii) the speeds at points B and C; (iii) the ratio of the kinetic energies (K_B/K_C) at B and C. Comment on the nature of the trajectory of the bob after it reaches the point C.

Fig. 6.6

Answer (i) There are two external forces on the bob : gravity and the tension (T) in the string. The latter does no work since the displacement of the bob is always normal to the string. The potential energy of the bob is thus associated with the gravitational force only. The total mechanical energy E of the system is conserved. We take the potential energy of the system to be zero at the lowest point A. Thus, at A :

$$E = \frac{1}{2}mv_0^2 \quad (6.12)$$

$$T_A - mg = \frac{mv_0^2}{L} \quad \text{[Newton's Second Law]}$$

where T_A is the tension in the string at A. At the highest point C, the string slackens, as the tension in the string (T_C) becomes zero.

Thus, at C

$$E = \frac{1}{2}mv_c^2 + 2mgL \quad (6.13)$$

$$mg = \frac{mv_c^2}{L} \quad \text{[Newton's Second Law]} \quad (6.14)$$

where v_C is the speed at C. From Eqs. (6.13) and (6.14)

$$E = \frac{5}{2}mgL$$

Equating this to the energy at A

$$\frac{5}{2}mgL = \frac{m}{2}v_0^2$$

or, $$v_0 = \sqrt{5gL}$$

(ii) It is clear from Eq. (6.14)

$$v_C = \sqrt{gL}$$

At B, the energy is

$$E = \frac{1}{2}mv_B^2 + mgL$$

Equating this to the energy at A and employing the result from (i), namely $v_0^2 = 5gL$,

$$\frac{1}{2}mv_B^2 + mgL = \frac{1}{2}mv_0^2$$

$$= \frac{5}{2}m\, g\, L$$

$$\therefore v_B = \sqrt{3gL}$$

(iii) The ratio of the kinetic energies at B and C is :

$$\frac{K_B}{K_C} = \frac{\frac{1}{2}mv_B^2}{\frac{1}{2}mv_C^2} = \frac{3}{1}$$

At point C, the string becomes slack and the velocity of the bob is horizontal and to the left. If the connecting string is cut at this instant, the bob will execute a projectile motion with horizontal projection akin to a rock kicked horizontally from the edge of a cliff. Otherwise the bob will continue on its circular path and complete the revolution. ◀

6.9 THE POTENTIAL ENERGY OF A SPRING

The spring force is an example of a variable force which is conservative. Fig. 6.7 shows a block attached to a spring and resting on a smooth horizontal surface. The other end of the spring is attached to a rigid wall. The spring is light and may be treated as massless. In an ideal spring, the spring force F_s is proportional to x where x is the displacement of the block from the equilibrium position. The displacement could be either positive [Fig. 6.7(b)] or negative [Fig. 6.7(c)]. This force law for the spring is called Hooke's law and is mathematically stated as

$$F_s = -kx$$

The constant k is called the spring constant. Its unit is N m^{-1}. The spring is said to be stiff if k is large and soft if k is small.

Suppose that we pull the block outwards as in Fig. 6.7(b). If the extension is x_m, the work done by the spring force is

$$W_s = \int_0^{x_m} F_s \, dx \quad = -\int_0^{x_m} kx \, dx$$

$$= -\frac{k x_m^2}{2} \qquad (6.15)$$

This expression may also be obtained by considering the area of the triangle as in Fig. 6.7(d). Note that the work done by the external pulling force F is positive since it overcomes the spring force.

$$W = +\frac{k x_m^2}{2} \qquad (6.16)$$

Fig. 6.7 Illustration of the spring force with a block attached to the free end of the spring. (a) The spring force F_s is zero when the displacement x from the equilibrium position is zero. (b) For the stretched spring $x > 0$ and $F_s < 0$ (c) For the compressed spring $x < 0$ and $F_s > 0$. (d) The plot of F_s versus x. The area of the shaded triangle represents the work done by the spring force. Due to the opposing signs of F_s and x, this work done is negative, $W_s = -kx_m^2/2$.

The same is true when the spring is compressed with a displacement x_c (< 0). The spring force does work $W_s = -kx_c^2/2$ while the

external force F does work $+ kx_c^2 / 2$. If the block is moved from an initial displacement x_i to a final displacement x_f, the work done by the spring force W_s is

$$W_s = -\int_{x_i}^{x_f} k\,x\,dx \quad = \frac{k\,x_i^2}{2} - \frac{k\,x_f^2}{2} \qquad (6.17)$$

Thus the work done by the spring force depends only on the end points. Specifically, if the block is pulled from x_i and allowed to return to x_i;

$$W_s = -\int_{x_i}^{x_i} k\,x\,dx \quad = \frac{k\,x_i^2}{2} - \frac{k\,x_i^2}{2}$$

$$= 0 \qquad (6.18)$$

The work done by the spring force in a cyclic process is zero. We have explicitly demonstrated that the spring force (i) is position dependent only as first stated by Hooke, ($F_s = -kx$); (ii) does work which only depends on the initial and final positions, e.g. Eq. (6.17). Thus, the spring force is a **conservative force**.

We define the potential energy $V(x)$ of the spring to be zero when block and spring system is in the equilibrium position. For an extension (or compression) x the above analysis suggests that

$$V(x) = \frac{kx^2}{2} \qquad (6.19)$$

You may easily verify that $-dV/dx = -kx$, the spring force. If the block of mass m in Fig. 6.7 is extended to x_m and released from rest, then its total mechanical energy at any arbitrary point x, where x lies between $-x_m$ and $+x_m$, will be given by

$$\frac{1}{2}k\,x_m^2 = \frac{1}{2}k\,x^2 + \frac{1}{2}m\,v^2$$

where we have invoked the conservation of mechanical energy. This suggests that the speed and the kinetic energy will be maximum at the equilibrium position, $x = 0$, i.e.,

$$\frac{1}{2}m\,v_m^2 = \frac{1}{2}k\,x_m^2$$

where v_m is the maximum speed.

or $$v_m = \sqrt{\frac{k}{m}}\,x_m$$

Note that k/m has the dimensions of $[T^{-2}]$ and our equation is dimensionally correct. The kinetic energy gets converted to potential energy and vice versa, however, the total mechanical energy remains constant. This is graphically depicted in Fig. 6.8.

Example 6.8 To simulate car accidents, auto manufacturers study the collisions of moving cars with mounted springs of different spring constants. Consider a typical simulation with a car of mass 1000 kg moving with a speed 18.0 km/h on a smooth road and colliding with a horizontally mounted spring of spring constant 6.25×10^3 N m^{-1}. What is the maximum compression of the spring ?

Fig. 6.8 Parabolic plots of the potential energy V and kinetic energy K of a block attached to a spring obeying Hooke's law. The two plots are complementary, one decreasing as the other increases. The total mechanical energy E = K + V remains constant.

Answer At maximum compression the kinetic energy of the car is converted entirely into the potential energy of the spring.

The kinetic energy of the moving car is

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

$$= \frac{1}{2} \times 10^3 \times 5 \times 5$$

$$K = 1.25 \times 10^4 \text{ J}$$

where we have converted 18 km h^{-1} to 5 m s^{-1} **[It is useful to remember that 36 km h^{-1} = 10 m s^{-1}]**. At maximum compression x_m, the potential energy V of the spring is equal to the kinetic energy K of the moving car from the principle of conservation of mechanical energy.

$$V = \frac{1}{2}k\,x_m^2$$

$$= 1.25 \times 10^4 \text{ J}$$

We obtain

$$x_m = 2.00 \text{ m}$$

We note that we have idealised the situation. The spring is considered to be massless. The surface has been considered to possess negligible friction. ◀

We conclude this section by making a few remarks on conservative forces.

(i) Information on time is absent from the above discussions. In the example considered above, we can calculate the compression, but not the time over which the compression occurs. A solution of Newton's Second Law for this system is required for temporal information.

(ii) Not all forces are conservative. Friction, for example, is a non-conservative force. The principle of conservation of energy will have to be modified in this case. This is illustrated in Example 6.9.

(iii) The zero of the potential energy is arbitrary. It is set according to convenience. For the spring force we took $V(x) = 0$, at $x = 0$, i.e. the unstretched spring had zero potential energy. For the constant gravitational force mg, we took $V = 0$ on the earth's surface. In a later chapter we shall see that for the force due to the universal law of gravitation, the zero is best defined at an infinite distance from the gravitational source. However, once the zero of the potential energy is fixed in a given discussion, it must be consistently adhered to throughout the discussion. You cannot change horses in midstream !

▶ ***Example 6.9*** Consider Example 6.7 taking the coefficient of friction, μ, to be 0.5 and calculate the maximum compression of the spring.

Answer In presence of friction, both the spring force and the frictional force act so as to oppose the compression of the spring as shown in Fig. 6.9.

We invoke the work-energy theorem, rather than the conservation of mechanical energy.

The change in kinetic energy is

Fig. 6.9 The forces acting on the car.

$$\Delta K = K_f - K_i = 0 - \frac{1}{2} m v^2$$

The work done by the net force is

$$W = -\frac{1}{2} k x_m^2 - \mu m g x_m$$

Equating we have

$$\frac{1}{2} m v^2 = \frac{1}{2} k x_m^2 + \mu m g x_m$$

Now $\mu mg = 0.5 \times 10^3 \times 10 = 5 \times 10^3$ N (taking $g = 10.0$ m s^{-2}). After rearranging the above equation we obtain the following quadratic equation in the unknown x_m.

$$k x_m^2 + 2\mu m g x_m - m v^2 = 0$$

$$x_m = \frac{-\mu m g + \left[\mu^2 m^2 g^2 + m k v^2\right]^{1/2}}{k}$$

where we take the positive square root since x_m is positive. Putting in numerical values we obtain

$$x_m = 1.35 \text{ m}$$

which, as expected, is less than the result in Example 6.8.

If the two forces on the body consist of a conservative force F_c and a non-conservative force F_{nc}, the conservation of mechanical energy formula will have to be modified. By the WE theorem

$$(F_c + F_{nc})\, \Delta x = \Delta K$$

But $\quad F_c\, \Delta x = -\Delta V$

Hence, $\quad \Delta(K + V) = F_{nc} \Delta x$

$$\Delta E = F_{nc} \Delta x$$

where E is the total mechanical energy. Over the path this assumes the form

$$E_f - E_i = W_{nc}$$

Where W_{nc} is the total work done by the non-conservative forces over the path. Note that

unlike the conservative force, W_{nc} depends on the particular path i to f. ◀

6.10 VARIOUS FORMS OF ENERGY : THE LAW OF CONSERVATION OF ENERGY

In the previous section we have discussed mechanical energy. We have seen that it can be classified into two distinct categories : one based on motion, namely kinetic energy; the other on configuration (position), namely potential energy. Energy comes in many a forms which transform into one another in ways which may not often be clear to us.

6.10.1 Heat

We have seen that the frictional force is excluded from the category of conservative forces. However, work is associated with the force of friction. A block of mass m sliding on a rough horizontal surface with speed v_0 comes to a halt over a distance x_0. The work done by the force of kinetic friction f over x_0 is $-f\,x_0$. By the work-energy theorem $m\,v_0^2/2 = f\,x_0$. If we confine our scope to mechanics, we would say that the kinetic energy of the block is 'lost' due to the frictional force. On examination of the block and the table we would detect a slight increase in their temperatures. The work done by friction is not 'lost', but is transferred as heat energy. This raises the internal energy of the block and the table. In winter, in order to feel warm, we generate heat by vigorously rubbing our palms together. We shall see later that the internal energy is associated with the ceaseless, often random, motion of molecules. A quantitative idea of the transfer of heat energy is obtained by noting that 1 kg of water releases about 42000 J of energy when it cools by10 °C.

6.10.2 Chemical Energy

One of the greatest technical achievements of humankind occurred when we discovered how to ignite and control fire. We learnt to rub two flint stones together (mechanical energy), got them to heat up and to ignite a heap of dry leaves (chemical energy), which then provided sustained warmth. A matchstick ignites into a bright flame when struck against a specially prepared chemical surface. The lighted matchstick, when applied to a firecracker, results in a spectacular display of sound and light.

Chemical energy arises from the fact that the molecules participating in the chemical reaction have different binding energies. A stable chemical compound has less energy than the separated parts. A chemical reaction is basically a rearrangement of atoms. If the total energy of the reactants is more than the products of the reaction, heat is released and the reaction is said to be an **exothermic** reaction. If the reverse is true, heat is absorbed and the reaction is **endothermic**. Coal consists of carbon and a kilogram of it when burnt releases 3×10^7 J of energy.

Chemical energy is associated with the forces that give rise to the stability of substances. These forces bind atoms into molecules, molecules into polymeric chains, etc. The chemical energy arising from the combustion of coal, cooking gas, wood and petroleum is indispensable to our daily existence.

6.10.3 Electrical Energy

The flow of electrical current causes bulbs to glow, fans to rotate and bells to ring. There are laws governing the attraction and repulsion of charges and currents, which we shall learn later. Energy is associated with an electric current. An urban Indian household consumes about 200 J of energy per second on an average.

6.10.4 The Equivalence of Mass and Energy

Till the end of the nineteenth century, physicists believed that in every physical and chemical process, the mass of an isolated system is conserved. Matter might change its phase, e.g. glacial ice could melt into a gushing stream, but matter is neither created nor destroyed; Albert Einstein (1879-1955) however, showed that mass and energy are equivalent and are related by the relation

$$E = m\,c^2 \qquad (6.20)$$

where c, the speed of light in vacuum is approximately 3×10^8 m s^{-1}. Thus, a staggering amount of energy is associated with a mere kilogram of matter

$$E = 1\times (3\times 10^8)^2 \text{ J} = 9\times 10^{16} \text{ J}.$$

This is equivalent to the annual electrical output of a large (3000 MW) power generating station.

6.10.5 Nuclear Energy

The most destructive weapons made by man, the fission and fusion bombs are manifestations of

Table 6.3 Approximate energy associated with various phenomena

Description	Energy (J)
Big Bang	10^{68}
Radio energy emitted by the galaxy during its lifetime	10^{55}
Rotational energy of the Milky Way	10^{52}
Energy released in a supernova explosion	10^{44}
Ocean's hydrogen in fusion	10^{34}
Rotational energy of the earth	10^{29}
Annual solar energy incident on the earth	5×10^{24}
Annual wind energy dissipated near earth's surface	10^{22}
Annual global energy usage by human	3×10^{20}
Annual energy dissipated by the tides	10^{20}
Energy release of 15-megaton fusion bomb	10^{17}
Annual electrical output of large generating plant	10^{16}
Thunderstorm	10^{15}
Energy released in burning 1000 kg of coal	3×10^{10}
Kinetic energy of a large jet aircraft	10^{9}
Energy released in burning 1 litre of gasoline	3×10^{7}
Daily food intake of a human adult	10^{7}
Work done by a human heart per beat	0.5
Turning this page	10^{-3}
Flea hop	10^{-7}
Discharge of a single neuron	10^{-10}
Typical energy of a proton in a nucleus	10^{-13}
Typical energy of an electron in an atom	10^{-18}
Energy to break one bond in DNA	10^{-20}

the above equivalence of mass and energy [Eq. (6.20)]. On the other hand the explanation of the life-nourishing energy output of the sun is also based on the above equation. In this case effectively four light hydrogen nuclei fuse to form a helium nucleus whose mass is less than the sum of the masses of the reactants. This mass difference, called the mass defect Δm is the source of energy $(\Delta m)c^2$. In fission, a heavy nucleus like uranium $^{235}_{92}U$, is split by a neutron into lighter nuclei. Once again the final mass is less than the initial mass and the mass difference translates into energy, which can be tapped to provide electrical energy as in nuclear power plants (controlled nuclear fission) or can be employed in making nuclear weapons (uncontrolled nuclear fission). Strictly, the energy ΔE released in a chemical reaction can also be related to the mass defect $\Delta m = \Delta E/c^2$. However, for a chemical reaction, this mass defect is much smaller than for a nuclear reaction. Table 6.3 lists the total energies for a variety of events and phenomena.

▶ ***Example 6.10*** Examine Tables 6.1-6.3 and express (a) The energy required to break one bond in DNA in eV; (b) The kinetic energy of an air molecule (10^{-21} J) in eV; (c) The daily intake of a human adult in kilocalories.

Answer (a) Energy required to break one bond of DNA is

$$\frac{10^{-20}\,\text{J}}{1.6\times10^{-19}\,\text{J/eV}} \simeq 0.06\ \text{eV}$$

where the symbol '$\simeq$' stands for approximate. Note 0.1 eV = 100 meV (100 millielectron volt).

(b) The kinetic energy of an air molecule is

$$\frac{10^{-21}\,\text{J}}{1.6\times10^{-19}\,\text{J/eV}} \simeq 0.0062\ \text{eV}$$

This is the same as 6.2 meV.

(c) The average human consumption in a day is

$$\frac{10^{7}\,\text{J}}{4.2\times10^{3}\ \text{J/kcal}} \simeq 2400\ \text{kcal}$$

We point out a common misconception created by newspapers and magazines. They mention food values in calories and urge us to restrict diet intake to below 2400 calories. What they should be saying is kilocalories (kcal) and not calories. A person consuming 2400 calories a day will soon starve to death! 1 food calorie is 1 kcal. ◀

6.10.6 The Principle of Conservation of Energy

We have seen that the total mechanical energy of the system is conserved if the forces doing work on it are conservative. If some of the forces involved are non-conservative, part of the mechanical energy may get transformed into other forms such as heat, light and sound. However, the total energy of an isolated system does not change, as long as one accounts for all forms of energy. Energy may be transformed from one form to another but the total energy of an isolated system remains constant. Energy can neither be created, nor destroyed.

Since the universe as a whole may be viewed as an isolated system, the total energy of the universe is constant. If one part of the universe loses energy, another part must gain an equal amount of energy.

The principle of conservation of energy cannot be proved. However, no violation of this principle has been observed. The concept of conservation and transformation of energy into various forms links together various branches of physics, chemistry and life sciences. It provides a unifying, enduring element in our scientific pursuits. From engineering point of view all electronic, communication and mechanical devices rely on some forms of energy transformation.

6.11 POWER

Often it is interesting to know not only the work done on an object, but also the rate at which this work is done. We say a person is physically fit if he not only climbs four floors of a building but climbs them fast. **Power** is defined as the time rate at which work is done or energy is transferred.

The average power of a force is defined as the ratio of the work, W, to the total time t taken

$$P_{av} = \frac{W}{t}$$

The instantaneous power is defined as the limiting value of the average power as time interval approaches zero,

$$P = \frac{\mathrm{d}W}{\mathrm{d}t} \tag{6.21}$$

The work dW done by a force F for a displacement $d\mathbf{r}$ is $dW = \mathbf{F}.d\mathbf{r}$. The instantaneous power can also be expressed as

$$P = \mathbf{F}.\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t}$$

$$= \mathbf{F}.\mathbf{v} \tag{6.22}$$

where $\mathbf{v}$ is the instantaneous velocity when the force is $\mathbf{F}$.

Power, like work and energy, is a scalar quantity. Its dimensions are $[ML^2T^{-3}]$. In the SI, its unit is called a watt (W). The watt is 1 J s^{-1}. The unit of power is named after James Watt, one of the innovators of the steam engine in the eighteenth century.

There is another unit of power, namely the horse-power (hp)

1 hp = 746 W

This unit is still used to describe the output of automobiles, motorbikes, etc.

We encounter the unit watt when we buy electrical goods such as bulbs, heaters and refrigerators. A 100 watt bulb which is on for 10 hours uses 1 kilowatt hour (kWh) of energy.

100 (watt) × 10 (hour)
= 1000 watt hour
= 1 kilowatt hour (kWh)
= 10^3 (W) × 3600 (s)
= 3.6×10^6 J

Our electricity bills carry the energy consumption in units of kWh. Note that kWh is a unit of energy and not of power.

▶ ***Example 6.11*** An elevator can carry a maximum load of 1800 kg (elevator + passengers) is moving up with a constant speed of 2 m s^{-1}. The frictional force opposing the motion is 4000 N. Determine the minimum power delivered by the motor to the elevator in watts as well as in horse power.

Answer The downward force on the elevator is

$$F = m\,g + F_f = (1800 \times 10) + 4000 = 22000 \text{ N}$$

The motor must supply enough power to balance this force. Hence,

$$P = \mathbf{F}.\,\mathbf{v} = 22000 \times 2 = 44000 \text{ W} = 59 \text{ hp}$$ ◀

6.12 COLLISIONS

In physics we study motion (change in position). At the same time, we try to discover physical quantities, which do not change in a physical process. The laws of momentum and energy conservation are typical examples. In this section we shall apply these laws to a commonly encountered phenomena, namely collisions. Several games such as billiards, marbles or carrom involve collisions.We shall study the collision of two masses in an idealised form.

Consider two masses m_1 and m_2. The particle m_1 is moving with speed v_{1i}, the subscript 'i' implying initial. We can cosider m_2 to be at rest. No loss of generality is involved in making such a selection. In this situation the mass m_1 collides with the stationary mass m_2 and this is depicted in Fig. 6.10.

Fig. 6.10 Collision of mass m_1, with a stationary mass m_2.

The masses m_1 and m_2 fly-off in different directions. We shall see that there are relationships, which connect the masses, the velocities and the angles.

6.12.1 Elastic and Inelastic Collisions

In all collisions the total linear momentum is conserved; the initial momentum of the system is equal to the final momentum of the system. One can argue this as follows. When two objects collide, the mutual impulsive forces acting over the collision time Δt cause a change in their respective momenta :

$$\Delta\mathbf{p}_1 = \mathbf{F}_{12}\,\Delta t$$
$$\Delta\mathbf{p}_2 = \mathbf{F}_{21}\,\Delta t$$

where $\mathbf{F}_{12}$ is the force exerted on the first particle by the second particle. $\mathbf{F}_{21}$ is likewise the force exerted on the second particle by the first particle. Now from Newton's Third Law, $\mathbf{F}_{12} = -\mathbf{F}_{21}$. This implies

$$\Delta\mathbf{p}_1 + \Delta\mathbf{p}_2 = \mathbf{0}$$

The above conclusion is true even though the forces vary in a complex fashion during the collision time Δt. Since the third law is true at every instant, the total impulse on the first object is equal and opposite to that on the second.

On the other hand, the total kinetic energy of the system is not necessarily conserved. The impact and deformation during collision may generate heat and sound. Part of the initial kinetic energy is transformed into other forms of energy. A useful way to visualise the deformation during collision is in terms of a 'compressed spring'. If the 'spring' connecting the two masses regains its original shape without loss in energy, then the initial kinetic energy is equal to the final kinetic energy but the kinetic energy during the collision time Δt is not constant. Such a collision is called an **elastic collision**. On the other hand the deformation may not be relieved and the two bodies could move together after the collision. A collision in which the two particles move together after the collision is called a **completely inelastic collision**. The intermediate case where the deformation is partly relieved and some of the initial kinetic energy is lost is more common and is appropriately called an **inelastic collision**.

6.12.2 Collisions in One Dimension

Consider first a **completely inelastic collision** in one dimension. Then, in Fig. 6.10,

$$\theta_1 = \theta_2 = 0$$

$$m_1 v_{1i} = (m_1 + m_2) v_f \quad \text{(momentum conservation)}$$

$$v_f = \frac{m_1}{m_1 + m_2} v_{1i} \qquad (6.23)$$

The loss in kinetic energy on collision is

$$\Delta K = \frac{1}{2} m_1 v_{1i}^2 - \frac{1}{2}(m_1 + m_2) v_f^2$$

$$= \frac{1}{2} m_1 v_{1i}^2 - \frac{1}{2}\frac{m_1^2}{m_1 + m_2} v_{1i}^2 \quad \text{[using Eq. (6.23)]}$$

$$= \frac{1}{2} m_1 v_{1i}^2 \left[1 - \frac{m_1}{m_1 + m_2}\right]$$

An experiment on head-on collision

In performing an experiment on collision on a horizontal surface, we face three difficulties. One, there will be friction and bodies will not travel with uniform velocities. Two, if two bodies of different sizes collide on a table, it would be difficult to arrange them for a head-on collision unless their centres of mass are at the same height above the surface. Three, it will be fairly difficult to measure velocities of the two bodies just before and just after collision.

By performing this experiment in a vertical direction, all the three difficulties vanish. Take two balls, one of which is heavier (basketball/football/volleyball) and the other lighter (tennis ball/rubber ball/table tennis ball). First take only the heavier ball and drop it vertically from some height, say 1 m. Note to which it rises. This gives the velocities near the floor or ground, just before and just after the bounce (by using $v^2 = 2gh$). Hence you will get the coefficient of restitution.

Now take the big ball and a small ball and hold them in your hands one over the other, with the heavier ball below the lighter one, as shown here. Drop them together, taking care that they remain together while falling, and see what happens. You will find that the heavier ball rises less than when it was dropped alone, while the lighter one shoots up to about 3 m. With practice, you will be able to hold the ball properly so that the lighter ball rises vertically up and does not fly sideways. This is head-on collision.

You can try to find the best combination of balls which gives you the best effect. You can measure the masses on a standard balance. We leave it to you to think how you can determine the initial and final velocities of the balls.

$$= \frac{1}{2}\frac{m_1 m_2}{m_1 + m_2} v_{1i}^2$$

which is a positive quantity as expected.

Consider next an elastic collision. Using the above nomenclature with $\theta_1 = \theta_2 = 0$, the momentum and kinetic energy conservation equations are

$$m_1 v_{1i} = m_1 v_{1f} + m_2 v_{2f} \quad (6.24)$$

$$m_1 v_{1i}^2 = m_1 v_{1f}^2 + m_2 v_{2f}^2 \quad (6.25)$$

From Eqs. (6.24) and (6.25) it follows that,

$$m_1 v_{1i}(v_{2f} - v_{1i}) = m_1 v_{1f}(v_{2f} - v_{1f})$$

or, $$v_{2f}(v_{1i} - v_{1f}) = v_{1i}^2 - v_{1f}^2$$

$$= (v_{1i} - v_{1f})(v_{1i} + v_{1f})$$

Hence, $$\therefore v_{2f} = v_{1i} + v_{1f} \quad (6.26)$$

Substituting this in Eq. (6.24), we obtain

$$v_{1f} = \frac{(m_1 - m_2)}{m_1 + m_2} v_{1i} \quad (6.27)$$

and $$v_{2f} = \frac{2m_1 v_{1i}}{m_1 + m_2} \quad (6.28)$$

Thus, the 'unknowns' $\{v_{1f}, v_{2f}\}$ are obtained in terms of the 'knowns' $\{m_1, m_2, v_{1i}\}$. Special cases of our analysis are interesting.

Case I : If the two masses are equal

$$v_{1f} = 0$$
$$v_{2f} = v_{1i}$$

The first mass comes to rest and pushes off the second mass with its initial speed on collision.

Case II : If one mass dominates, e.g. $m_2 >> m_1$

$$v_{1f} \simeq -v_{1i} \qquad v_{2f} \simeq 0$$

The heavier mass is undisturbed while the lighter mass reverses its velocity.

▶ ***Example 6.12*** **Slowing down of neutrons**: In a nuclear reactor a neutron of high speed (typically 10^7 m s^{-1}) must be slowed

to 10^3 m s^{-1} so that it can have a high probability of interacting with isotope $^{235}_{92}U$ and causing it to fission. Show that a neutron can lose most of its kinetic energy in an elastic collision with a light nuclei like deuterium or carbon which has a mass of only a few times the neutron mass. The material making up the light nuclei, usually heavy water (D_2O) or graphite, is called a moderator.

Answer The initial kinetic energy of the neutron is

$$K_{1i} = \frac{1}{2} m_1 v_{1i}^2$$

while its final kinetic energy from Eq. (6.27)

$$K_{1f} = \frac{1}{2} m_1 v_{1f}^2 = \frac{1}{2} m_1 \left(\frac{m_1 - m_2}{m_1 + m_2}\right)^2 v_{1i}^2$$

The fractional kinetic energy lost is

$$f_1 = \frac{K_{1f}}{K_{1i}} = \left(\frac{m_1 - m_2}{m_1 + m_2}\right)^2$$

while the fractional kinetic energy gained by the moderating nuclei K_{2f}/K_{1i} is

$$f_2 = 1 - f_1 \text{ (elastic collision)}$$

$$= \frac{4 m_1 m_2}{(m_1 + m_2)^2}$$

One can also verify this result by substituting from Eq. (6.28).

For deuterium $m_2 = 2m_1$ and we obtain $f_1 = 1/9$ while $f_2 = 8/9$. Almost 90% of the neutron's energy is transferred to deuterium. For carbon $f_1 = 71.6\%$ and $f_2 = 28.4\%$. In practice, however, this number is smaller since head-on collisions are rare. ◀

If the initial velocities and final velocities of both the bodies are along the same straight line, then it is called a one-dimensional collision, or **head-on collision.** In the case of small spherical bodies, this is possible if the direction of travel of body 1 passes through the centre of body 2 which is at rest. In general, the collision is two-dimensional, where the initial velocities and the final velocities lie in a plane.

6.12.3 Collisions in Two Dimensions

Fig. 6.10 also depicts the collision of a moving mass m_1 with the stationary mass m_2. Linear momentum is conserved in such a collision. Since momentum is a vector this implies three equations for the three directions {*x, y, z*}. Consider the plane determined by the final velocity directions of m_1 and m_2 and choose it to be the *x-y* plane. The conservation of the *z*-component of the linear momentum implies that the entire collision is in the *x-y* plane. The *x*- and *y*-component equations are

$$m_1 v_{1i} = m_1 v_{1f} \cos\theta_1 + m_2 v_{2f} \cos\theta_2 \quad (6.29)$$

$$0 = m_1 v_{1f} \sin\theta_1 - m_2 v_{2f} \sin\theta_2 \quad (6.30)$$

One knows $\{m_1, m_2, v_{1i}\}$ in most situations. There are thus four unknowns $\{v_{1f}, v_{2f}, \theta_1 \text{ and } \theta_2\}$, and only two equations. If $\theta_1 = \theta_2 = 0$, we regain Eq. (6.24) for one dimensional collision.

If, further the collision is elastic,

$$\frac{1}{2} m_1 v_{1i}^2 = \frac{1}{2} m_1 v_{1f}^2 + \frac{1}{2} m_2 v_{2f}^2 \quad (6.31)$$

We obtain an additional equation. That still leaves us one equation short. At least one of the four unknowns, say θ_1, must be made known for the problem to be solvable. For example, θ_1 can be determined by moving a detector in an angular fashion from the *x* to the *y* axis. Given $\{m_1, m_2, v_{1i}, \theta_1\}$ we can determine $\{v_{1f}, v_{2f}, \theta_2\}$ from Eqs. (6.29)-(6.31).

▶ ***Example 6.13*** Consider the collision depicted in Fig. 6.10 to be between two billiard balls with equal masses $m_1 = m_2$. The first ball is called the cue while the second ball is called the target. The billiard player wants to 'sink' the target ball in a corner pocket, which is at an angle $\theta_2 = 37°$. Assume that the collision is elastic and that friction and rotational motion are not important. Obtain θ_1.

Answer From momentum conservation, since the masses are equal

$$\mathbf{v}_{1i} = \mathbf{v}_{1f} + \mathbf{v}_{2f}$$

or

$$v_{1i}^2 = (\mathbf{v}_{1f} + \mathbf{v}_{2f}) \cdot (\mathbf{v}_{1f} + \mathbf{v}_{2f})$$

$$= v_{1f}^2 + v_{2f}^2 + 2\mathbf{v}_{1f} \cdot \mathbf{v}_{2f}$$

$$= \quad v_{1f}^{\,2} + v_{2f}^{\,2} + 2v_{1f}v_{2f}\cos\left(\theta_1 + 37°\right)\Big\} \qquad (6.32)$$

Since the collision is elastic and $m_1 = m_2$ it follows from conservation of kinetic energy that

$$v_{1i}^{\,2} = v_{1f}^{\,2} + v_{2f}^{\,2} \qquad (6.33)$$

Comparing Eqs. (6.32) and (6.33), we get

$$\cos(\theta_1 + 37°) = 0$$

or $\quad \theta_1 + 37° = 90°$

Thus, $\quad \theta_1 = 53°$

This proves the following result : when two equal masses undergo a glancing elastic collision with one of them at rest, after the collision, they will move at right angles to each other. ◀

The matter simplifies greatly if we consider spherical masses with smooth surfaces, and assume that collision takes place only when the bodies touch each other. This is what happens in the games of marbles, carrom and billiards.

In our everyday world, collisions take place only when two bodies touch each other. But consider a comet coming from far distances to the sun, or alpha particle coming towards a nucleus and going away in some direction. Here we have to deal with forces involving action at a distance. Such an event is called scattering. The velocities and directions in which the two particles go away depend on their initial velocities as well as the type of interaction between them, their masses, shapes and sizes.

SUMMARY

1. The *work-energy theorem* states that the change in kinetic energy of a body is the work done by the net force on the body.
$$K_f - K_i = W_{net}$$
2. A force is *conservative* if (i) work done by it on an object is path independent and depends only on the end points $\{x_i, x_j\}$, or (ii) the work done by the force is zero for an arbitrary closed path taken by the object such that it returns to its initial position.
3. For a conservative force in one dimension, we may define a *potential energy* function $V(x)$ such that
$$F(x) = -\frac{\mathrm{d}V(x)}{\mathrm{d}x}$$
or
$$V_i - V_f = \int_{x_i}^{x_f} F(x)\,\mathrm{d}x$$
4. The principle of conservation of mechanical energy states that the total mechanical energy of a body remains constant if the only forces that act on the body are conservative.
5. The *gravitational potential energy* of a particle of mass m at a height x about the earth's surface is
$$V(x) = m\,g\,x$$
where the variation of g with height is ignored.
6. The elastic potential energy of a spring of force constant k and extension x is
$$V(x) = \frac{1}{2}k\,x^2$$
7. The scalar or dot product of two vectors **A** and **B** is written as **A.B** and is a scalar quantity given by : $\mathbf{A}\cdot\mathbf{B} = AB\cos\theta$, where θ is the angle between **A** and **B**. It can be positive, negative or zero depending upon the value of θ. The scalar product of two vectors can be interpreted as the product of magnitude of one vector and component of the other vector along the first vector. For unit vectors :
$$\hat{\mathbf{i}}\cdot\hat{\mathbf{i}} = \hat{\mathbf{j}}\cdot\hat{\mathbf{j}} = \hat{\mathbf{k}}\cdot\hat{\mathbf{k}} = 1 \text{ and } \hat{\mathbf{i}}\cdot\hat{\mathbf{j}} = \hat{\mathbf{j}}\cdot\hat{\mathbf{k}} = \hat{\mathbf{k}}\cdot\hat{\mathbf{i}} = 0$$
Scalar products obey the commutative and the distributive laws.

Physical Quantity	Symbol	Dimensions	Units	Remarks
Work	W	$[ML^2T^{-2}]$	J	W = **F.d**
Kinetic energy	K	$[ML^2T^{-2}]$	J	$K = \frac{1}{2}mv^2$
Potential energy	$V(x)$	$[ML^2T^{-2}]$	J	$F(x) = -\frac{dV(x)}{dx}$
Mechanical energy	E	$[ML^2T^{-2}]$	J	$E = K + V$
Spring constant	k	$[MT^{-2}]$	N m^{-1}	$F = -kx$ $V(x) = \frac{1}{2}kx^2$
Power	P	$[ML^2T^{-3}]$	W	P = **F.v** $P = \frac{dW}{dt}$

POINTS TO PONDER

1. The phrase 'calculate the work done' is incomplete. We should refer (or imply clearly by context) to the work done by a specific force or a group of forces on a given body over a certain displacement.
2. Work done is a scalar quantity. It can be positive or negative unlike mass and kinetic energy which are positive scalar quantities. The work done by the friction or viscous force on a moving body is negative.
3. For two bodies, the sum of the mutual forces exerted between them is zero from Newton's Third Law,
$$\mathbf{F}_{12} + \mathbf{F}_{21} = 0$$
But the sum of the work done by the two forces need not *always cancel, i.e.*
$$W_{12} + W_{21} \neq 0$$
However, it *may sometimes be true.*
4. The work done by a force can be calculated sometimes even if the exact nature of the force is not known. This is clear from Example 6.1 where the WE theorem is used in such a situation.
5. The WE theorem is not independent of Newton's Second Law. The WE theorem may be viewed as a scalar form of the Second Law. The principle of conservation of mechanical energy may be viewed as a consequence of the WE theorem for conservative forces.
6. The WE theorem holds in all inertial frames. It can also be extended to non-inertial frames provided we include the pseudoforces in the calculation of the net force acting on the body under consideration.
7. The potential energy of a body subjected to a conservative force is always undetermined upto a constant. For example, the point where the potential energy is zero is a matter of choice. For the gravitational potential energy mgh, the zero of the potential energy is chosen to be the ground. For the spring potential energy $kx^2/2$, the zero of the potential energy is the equilibrium position of the oscillating mass.
8. Every force encountered in mechanics does not have an associated potential energy. For example, work done by friction over a closed path is not zero and no potential energy can be associated with friction.
9. During a collision : (a) the total linear momentum is conserved at each instant of the collision ; (b) the kinetic energy conservation (even if the collision is elastic) applies after the collision is over and does not hold at every instant of the collision. In fact the two colliding objects are deformed and may be momentarily at rest with respect to each other.

EXERCISES

6.1 The sign of work done by a force on a body is important to understand. State carefully if the following quantities are positive or negative:

(a) work done by a man in lifting a bucket out of a well by means of a rope tied to the bucket.

(b) work done by gravitational force in the above case,

(c) work done by friction on a body sliding down an inclined plane,

(d) work done by an applied force on a body moving on a rough horizontal plane with uniform velocity,

(e) work done by the resistive force of air on a vibrating pendulum in bringing it to rest.

6.2 A body of mass 2 kg initially at rest moves under the action of an applied horizontal force of 7 N on a table with coefficient of kinetic friction = 0.1. Compute the

(a) work done by the applied force in 10 s,

(b) work done by friction in 10 s,

(c) work done by the net force on the body in 10 s,

(d) change in kinetic energy of the body in 10 s,

and interpret your results.

6.3 Given in Fig. 6.11 are examples of some potential energy functions in one dimension. The total energy of the particle is indicated by a cross on the ordinate axis. In each case, specify the regions, if any, in which the particle cannot be found for the given energy. Also, indicate the minimum total energy the particle must have in each case. Think of simple physical contexts for which these potential energy shapes are relevant.

Fig. 6.11

6.4 The potential energy function for a particle executing linear simple harmonic motion is given by $V(x) = kx^2/2$, where k is the force constant of the oscillator. For $k = 0.5$ N m^{-1}, the graph of $V(x)$ versus x is shown in Fig. 6.12. Show that a particle of total energy 1 J moving under this potential must 'turn back' when it reaches $x = \pm 2$ m.

Fig. 6.12

6.5 Answer the following :

(a) The casing of a rocket in flight burns up due to friction. At whose expense is the heat energy required for burning obtained? The rocket or the atmosphere?

(b) Comets move around the sun in highly elliptical orbits. The gravitational force on the comet due to the sun is not normal to the comet's velocity in general. Yet the work done by the gravitational force over every complete orbit of the comet is zero. Why ?

Fig. 6.13

(c) An artificial satellite orbiting the earth in very thin atmosphere loses its energy gradually due to dissipation against atmospheric resistance, however small. Why then does its speed increase progressively as it comes closer and closer to the earth ?

(d) In Fig. 6.13(i) the man walks 2 m carrying a mass of 15 kg on his hands. In Fig. 6.13(ii), he walks the same distance pulling the rope behind him. The rope goes over a pulley, and a mass of 15 kg hangs at its other end. In which case is the work done greater ?

6.6 Underline the correct alternative :

(a) When a conservative force does positive work on a body, the potential energy of the body increases/decreases/remains unaltered.

(b) Work done by a body against friction always results in a loss of its kinetic/potential energy.

(c) The rate of change of total momentum of a many-particle system is proportional to the external force/sum of the internal forces on the system.

(d) In an inelastic collision of two bodies, the quantities which do not change after the collision are the total kinetic energy/total linear momentum/total energy of the system of two bodies.

6.7 State if each of the following statements is true or false. Give reasons for your answer.

(a) In an elastic collision of two bodies, the momentum and energy of each body is conserved.

(b) Total energy of a system is always conserved, no matter what internal and external forces on the body are present.

(c) Work done in the motion of a body over a closed loop is zero for every force in nature.

(d) In an inelastic collision, the final kinetic energy is always less than the initial kinetic energy of the system.

6.8 Answer carefully, with reasons :

(a) In an elastic collision of two billiard balls, is the total kinetic energy conserved during the short time of collision of the balls (i.e. when they are in contact) ?

(b) Is the total linear momentum conserved during the short time of an elastic collision of two balls ?

(c) What are the answers to (a) and (b) for an inelastic collision ?

(d) If the potential energy of two billiard balls depends only on the separation distance between their centres, is the collision elastic or inelastic ? (Note, we are talking here of potential energy corresponding to the force during collision, not gravitational potential energy).

6.9 A body is initially at rest. It undergoes one-dimensional motion with constant acceleration. The power delivered to it at time t is proportional to

(i) $t^{1/2}$ (ii) t (iii) $t^{3/2}$ (iv) t^2

6.10 A body is moving unidirectionally under the influence of a source of constant power. Its displacement in time t is proportional to

(i) $t^{1/2}$ (ii) t (iii) $t^{3/2}$ (iv) t^2

6.11 A body constrained to move along the z-axis of a coordinate system is subject to a constant force $\mathbf{F}$ given by

$$\mathbf{F} = -\hat{\mathbf{i}} + 2\,\hat{\mathbf{j}} + 3\,\hat{\mathbf{k}}\ \text{N}$$

where $\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}$ are unit vectors along the x-, y- and z-axis of the system respectively. What is the work done by this force in moving the body a distance of 4 m along the z-axis ?

6.12 An electron and a proton are detected in a cosmic ray experiment, the first with kinetic energy 10 keV, and the second with 100 keV. Which is faster, the electron or the proton ? Obtain the ratio of their speeds. (electron mass = 9.11×10^{-31} kg, proton mass = 1.67×10^{-27} kg, 1 eV = 1.60×10^{-19} J).

6.13 A rain drop of radius 2 mm falls from a height of 500 m above the ground. It falls with decreasing acceleration (due to viscous resistance of the air) until at half its original height, it attains its maximum (terminal) speed, and moves with uniform speed thereafter. What is the work done by the gravitational force on the drop in the first and second half of its journey ? What is the work done by the resistive force in the entire journey if its speed on reaching the ground is 10 m s^{-1} ?

6.14 A molecule in a gas container hits a horizontal wall with speed 200 m s^{-1} and angle 30° with the normal, and rebounds with the same speed. Is momentum conserved in the collision ? Is the collision elastic or inelastic ?

6.15 A pump on the ground floor of a building can pump up water to fill a tank of volume 30 m^3 in 15 min. If the tank is 40 m above the ground, and the efficiency of the pump is 30%, how much electric power is consumed by the pump ?

6.16 Two identical ball bearings in contact with each other and resting on a frictionless table are hit head-on by another ball bearing of the same mass moving initially with a speed V. If the collision is elastic, which of the following (Fig. 6.14) is a possible result after collision ?

Fig. 6.14

6.17 The bob A of a pendulum released from 30° to the vertical hits another bob B of the same mass at rest on a table as shown in Fig. 6.15. How high does the bob A rise after the collision ? Neglect the size of the bobs and assume the collision to be elastic.

Fig. 6.15

6.18 The bob of a pendulum is released from a horizontal position. If the length of the pendulum is 1.5 m, what is the speed with which the bob arrives at the lowermost point, given that it dissipated 5% of its initial energy against air resistance ?

6.19 A trolley of mass 300 kg carrying a sandbag of 25 kg is moving uniformly with a speed of 27 km/h on a frictionless track. After a while, sand starts leaking out of a hole on the floor of the trolley at the rate of 0.05 kg s^{-1}. What is the speed of the trolley after the entire sand bag is empty ?

6.20 A body of mass 0.5 kg travels in a straight line with velocity $v = a\,x^{3/2}$ where a = 5 $m^{-1/2}$ s^{-1}. What is the work done by the net force during its displacement from x = 0 to x = 2 m ?

6.21 The blades of a windmill sweep out a circle of area A. (a) If the wind flows at a velocity v perpendicular to the circle, what is the mass of the air passing through it in time t ? (b) What is the kinetic energy of the air ? (c) Assume that the windmill converts 25% of the wind's energy into electrical energy, and that A = 30 m^2, v = 36 km/h and the density of air is 1.2 kg m^{-3}. What is the electrical power produced ?

6.22 A person trying to lose weight (dieter) lifts a 10 kg mass, one thousand times, to a height of 0.5 m each time. Assume that the potential energy lost each time she lowers the mass is dissipated. (a) How much work does she do against the gravitational force ? (b) Fat supplies 3.8 × 10^7J of energy per kilogram which is converted to mechanical energy with a 20% efficiency rate. How much fat will the dieter use up?

6.23 A family uses 8 kW of power. (a) Direct solar energy is incident on the horizontal surface at an average rate of 200 W per square meter. If 20% of this energy can be converted to useful electrical energy, how large an area is needed to supply 8 kW? (b) Compare this area to that of the roof of a typical house.

Additional Exercises

6.24 A bullet of mass 0.012 kg and horizontal speed 70 m s^{-1} strikes a block of wood of mass 0.4 kg and instantly comes to rest with respect to the block. The block is suspended from the ceiling by means of thin wires. Calculate the height to which the block rises. Also, estimate the amount of heat produced in the block.

6.25 Two inclined frictionless tracks, one gradual and the other steep meet at A from where two stones are allowed to slide down from rest, one on each track (Fig. 6.16). Will the stones reach the bottom at the same time? Will they reach there with the same speed? Explain. Given $\theta_1 = 30^0$, $\theta_2 = 60^0$, and h = 10 m, what are the speeds and times taken by the two stones ?

Fig. 6.16

6.26 A 1 kg block situated on a rough incline is connected to a spring of spring constant 100 N m^{-1} as shown in Fig. 6.17. The block is released from rest with the spring in the unstretched position. The block moves 10 cm down the incline before coming to rest. Find the coefficient of friction between the block and the incline. Assume that the spring has a negligible mass and the pulley is frictionless.

Fig. 6.17

6.27 A bolt of mass 0.3 kg falls from the ceiling of an elevator moving down with an uniform speed of 7 m s^{-1}. It hits the floor of the elevator (length of the elevator = 3 m) and does not rebound. What is the heat produced by the impact ? Would your answer be different if the elevator were stationary ?

6.28 A trolley of mass 200 kg moves with a uniform speed of 36 km/h on a frictionless track. A child of mass 20 kg runs on the trolley from one end to the other (10 m away) with a speed of 4 m s^{-1} relative to the trolley in a direction opposite to the its motion, and jumps out of the trolley. What is the final speed of the trolley ? How much has the trolley moved from the time the child begins to run ?

6.29 Which of the following potential energy curves in Fig. 6.18 cannot possibly describe the elastic collision of two billiard balls ? Here r is the distance between centres of the balls.

Fig. 6.18

6.30 Consider the decay of a free neutron at rest : $n \rightarrow p + e^-$

Show that the two-body decay of this type must necessarily give an electron of fixed energy and, therefore, cannot account for the observed continuous energy distribution in the β-decay of a neutron or a nucleus (Fig. 6.19).

Fig. 6.19

[Note: The simple result of this exercise was one among the several arguments advanced by W. Pauli to predict the existence of a third particle in the decay products of β-decay. This particle is known as neutrino. We now know that it is a particle of intrinsic spin ½ (like e^-, p or n), but is neutral, and either massless or having an extremely small mass (compared to the mass of electron) and which interacts very weakly with matter. The correct decay process of neutron is : $n \rightarrow p + e^- + \nu$]

APPENDIX 6.1 : POWER CONSUMPTION IN WALKING

The table below lists the approximate power expended by an adult human of mass 60 kg.

Table 6.4 Approximate power consumption

Activity	Power (W)
Sleeping	75
Slow Walking	200
Bicycling	500
Heart beat	1.2

Mechanical work must not be confused with the everyday usage of the term work. A woman standing with a very heavy load on her head may get very tired. But no mechanical work is involved. That is not to say that mechanical work cannot be estimated in ordinary human activity.

Consider a person walking with constant speed v_0. The mechanical work he does may be estimated simply with the help of the work-energy theorem. Assume :

(a) The major work done in walking is due to the acceleration and deceleration of the legs with each stride (See Fig. 6.20).
(b) Neglect air resistance.
(c) Neglect the small work done in lifting the legs against gravity.
(d) Neglect the swinging of hands etc. as is common in walking.

As we can see in Fig. 6.20, in each stride the leg is brought from rest to a speed, approximately equal to the speed of walking, and then brought to rest again.

Fig. 6.20 An illustration of a single stride in walking. While the first leg is maximally off the round, the second leg is on the ground and vice-versa

The work done by one leg in each stride is $m_l\, v_0^2$ by the work-energy theorem. Here m_l is the mass of the leg. Note $m_l\, v_0^2/2$ energy is expended by one set of leg muscles to bring the foot from rest to speed v_0 while an additional $m_l\, v_0^2/2$ is expended by a complementary set of leg muscles to bring the foot to rest from speed v_0. Hence work done by both legs in one stride is (study Fig. 6.20 carefully)

$$W_s = 2m_l\, v_0^2 \tag{6.34}$$

Assuming m_l = 10 kg and slow running of a nine-minute mile which translates to 3 m s^{-1} in SI units, we obtain

$$W_s = 180 \text{ J / stride}$$

If we take a stride to be 2 m long, the person covers 1.5 strides per second at his speed of 3 m s^{-1}. Thus the power expended

$$P = 180 \frac{\text{J}}{\text{stride}} \times 1.5 \frac{\text{stride}}{\text{second}}$$
$$= 270 \text{ W}$$

We must bear in mind that this is a lower estimate since several avenues of power loss (e.g. swinging of hands, air resistance etc.) have been ignored. The interesting point is that we did not worry about the forces involved. The forces, mainly friction and those exerted on the leg by the muscles of the rest of the body, are hard to estimate. Static friction does no work and we bypassed the impossible task of estimating the work done by the muscles by taking recourse to the work-energy theorem. We can also see the advantage of a wheel. The wheel permits smooth locomotion without the continual starting and stopping in mammalian locomotion.

CHAPTER SEVEN

SYSTEMS OF PARTICLES AND ROTATIONAL MOTION

7.1 Introduction
7.2 Centre of mass
7.3 Motion of centre of mass
7.4 Linear momentum of a system of particles
7.5 Vector product of two vectors
7.6 Angular velocity and its relation with linear velocity
7.7 Torque and angular momentum
7.8 Equilibrium of a rigid body
7.9 Moment of inertia
7.10 Theorems of perpendicular and parallel axes
7.11 Kinematics of rotational motion about a fixed axis
7.12 Dynamics of rotational motion about a fixed axis
7.13 Angular momentum in case of rotation about a fixed axis
7.14 Rolling motion
Summary
Points to Ponder
Exercises
Additional exercises

7.1 INTRODUCTION

In the earlier chapters we primarily considered the motion of a single particle (A particle is represented as a point mass. It has practically no size). We applied the results of our study even to the motion of bodies of finite size, assuming that motion of such bodies can be described in terms of the motion of a particle.

Any real body which we encounter in daily life has a finite size. In dealing with the motion of extended bodies (bodies of finite size) often the idealised model of a particle is inadequate. In this chapter we shall try to go beyond this inadequacy. We shall attempt to build an understanding of the motion of extended bodies. An extended body, in the first place, is a system of particles. We shall begin with the consideration of motion of the system as a whole. The centre of mass of a system of particles will be a key concept here. We shall discuss the motion of the centre of mass of a system of particles and usefulness of this concept in understanding the motion of extended bodies.

A large class of problems with extended bodies can be solved by considering them to be rigid bodies. **Ideally a rigid body is a body with a perfectly definite and unchanging shape. The distances between different pairs of such a body do not change.** It is evident from this definition of a rigid body that no real body is truly rigid, since real bodies deform under the influence of forces. But in many situations the deformations are negligible. Thus, in a number of situations involving bodies such as wheels, tops, steel beams, molecules and planets on the other hand, we can ignore that they warp, bend or vibrate and treat them as rigid.

7.1.1 What kind of motion can a rigid body have?

Let us try to explore this question by taking some examples of the motion of rigid bodies. Let us begin with a rectangular block sliding down an inclined plane without any sidewise

Fig 7.1 Translational (sliding) motion of a block down an inclined plane.
(Any point like P_1 or P_2 of the block moves with the same velocity at any instant of time.)

movement. The block is a rigid body. Its motion down the plane is such that all the particles of the body are moving together, i.e. they have the same velocity at any instant of time. The rigid body here is in pure translational motion (Fig. 7.1).

In pure translational motion at any instant of time every particle of the body has the same velocity.

Consider now the rolling motion of a solid metallic or wooden cylinder down the same inclined plane (Fig. 7.2). The rigid body in this problem, namely the cylinder, shifts from the top to the bottom of the inclined plane, and thus, has translational motion. But as Fig. 7.2 shows, all its particles are not moving with the same velocity at any instant. The body therefore, is not in pure translation. Its motion is translation plus 'something else.'

Fig. 7.2 Rolling motion of a cylinder It is not pure translational motion. Points P_1, P_2, P_3 and P_4 have different velocities (shown by arrows) at any instant of time. In fact, the velocity of the point of contact P_3 is zero at any instant, if the cylinder rolls without slipping.

In order to understand what this 'something else' is, let us take a rigid body so constrained that it cannot have translational motion. The most common way to constrain a rigid body so that it does not have translational motion is to fix it along a straight line. The only possible motion of such a rigid body is **rotation**. The line along which the body is fixed is termed as its **axis of rotation**. If you look around, you will come across many examples of rotation about an axis, a ceiling fan, a potter's wheel, a giant wheel in a fair, a merry-go-round and so on (Fig 7.3(a) and (b)).

Fig. 7.3 Rotation about a fixed axis
(a) A ceiling fan
(b) A potter's wheel.

Let us try to understand what rotation is, what characterises rotation. You may notice that **in rotation of a rigid body about a fixed**

Fig. 7.4 *A rigid body rotation about the z-axis (Each point of the body such as P_1 or P_2 describes a circle with its centre (C_1 or C_2) on the axis. The radius of the circle (r_1 or r_2) is the perpendicular distance of the point (P_1 or P_2) from the axis. A point on the axis like P_3 remains stationary).*

axis, every particle of the body moves in a circle, which lies in a plane perpendicular to the axis and has its centre on the axis. Fig. 7.4 shows the rotational motion of a rigid body about a fixed axis (the *z*-axis of the frame of reference). Let P_1 be a particle of the rigid body, arbitrarily chosen and at a distance r_1 from fixed axis. The particle P_1 describes a circle of radius r_1 with its centre C_1 on the fixed axis. The circle lies in a plane perpendicular to the axis. The figure also shows another particle P_2 of the rigid body, P_2 is at a distance r_2 from the fixed axis. The particle P_2 moves in a circle of radius r_2 and with centre C_2 on the axis. This circle, too, lies in a plane perpendicular to the axis. Note that the circles described by P_1 and P_2 may lie in different planes; both these planes, however, are perpendicular to the fixed axis. For any particle on the axis like P_3, $r = 0$. Any such particle remains stationary while the body rotates. This is expected since the axis is fixed.

Fig. 7.5 (a) *A spinning top (The point of contact of the top with the ground, its tip O, is fixed.)*

Fig. 7.5 (b) *An oscillating table fan. The pivot of the fan, point O, is fixed.*

In some examples of rotation, however, the axis may not be fixed. A prominent example of this kind of rotation is a top spinning in place [Fig. 7.5(a)]. (We assume that the top does not slip from place to place and so does not have translational motion.) We know from experience that the axis of such a spinning top moves around the vertical through its point of contact with the ground, sweeping out a cone as shown in Fig. 7.5(a). (This movement of the axis of the top around the vertical is termed **precession**.) Note, the **point of contact of the top with ground is fixed**. The axis of rotation of the top at any instant passes through the point of contact. Another simple example of this kind of rotation is the oscillating table fan or a pedestal fan. You may have observed that the axis of

rotation of such a fan has an oscillating (sidewise) movement in a horizontal plane about the vertical through the point at which the axis is pivoted (point O in Fig. 7.5(b)).

While the fan rotates and its axis moves sidewise, this point is fixed. Thus, in more general cases of rotation, such as the rotation of a top or a pedestal fan, **one point and not one line,** of the rigid body is fixed. In this case the axis is not fixed, though it always passes through the fixed point. In our study, however, we mostly deal with the simpler and special case of rotation in which one line (i.e. the axis) is fixed. Thus, for us rotation will be about a fixed axis only unless stated otherwise.

Fig. 7.6(a) *Motion of a rigid body which is pure translation.*

Fig. 7.6(b) *Motion of a rigid body which is a combination of translation and rotation.*

Fig 7.6 (a) and 7.6 (b) illustrate different motions of the same body. Note P is an arbitrary point of the body; O is the centre of mass of the body, which is defined in the next section. Suffice to say here that the trajectories of O are the translational trajectories Tr_1 and Tr_2 of the body. The positions O and P at three different instants of time are shown by O_1, O_2, and O_3, and P_1, P_2 and P_3 respectively in Fig. 7.6 (a) and (b) both. As seen from Fig. 7.6(a), at any instant the velocities of any particles like O and P of the body are the same in pure translation. Notice, in this case the orientation of OP, i.e. the angle OP makes with a fixed direction, say the horizontal, remains the same, i.e. $\alpha_1 = \alpha_2 = \alpha_3$. Fig. 7.6 (b) illustrates a case of combination of translation and rotation. In this case, at any instants the velocities of O and P differ. Also, α_1, α_2 and α_3 may all be different.

The rolling motion of a cylinder down an inclined plane is a combination of rotation about a fixed axis and translation. Thus, the 'something else' in the case of rolling motion which we referred to earlier is rotational motion. You will find Fig. 7.6(a) and (b) instructive from this point of view. Both these figures show motion of the same body along identical translational trajectory. In one case, Fig. 7.6(a), the motion is a pure translation; in the other case [Fig. 7.6(b)] it is a combination of translation and rotation. (You may try to reproduce the two types of motion shown using a rigid object like a heavy book.)

We now recapitulate the most important observations of the present section: **The motion of a rigid body which is not pivoted or fixed in some way is either a pure translation or a combination of translation and rotation. The motion of a rigid body which is pivoted or fixed in some way is rotation.** The rotation may be about an axis that is fixed (e.g. a ceiling fan) or moving (e.g. an oscillating table fan). We shall, in the present chapter, consider rotational motion about a fixed axis only.

7.2 CENTRE OF MASS

We shall first see what the centre of mass of a system of particles is and then discuss its significance. For simplicity we shall start with a two particle system. We shall take the line joining the two particles to be the *x*- axis.

Fig. 7.7

Let the distances of the two particles be x_1 and x_2 respectively from some origin O. Let m_1 and m_2 be respectively the masses of the two

particles. The centre of mass of the system is that point C which is at a distance X from O, where X is given by

$$X = \frac{m_1 x_1 + m_2 x_2}{m_1 + m_2} \tag{7.1}$$

In Eq. (7.1), X can be regarded as the mass-weighted mean of x_1 and x_2. If the two particles have the same mass $m_1 = m_2 = m$ then

$$X = \frac{m x_1 + m x_2}{2m} = \frac{x_1 + x_2}{2}$$

Thus, for two particles of equal mass the centre of mass lies exactly midway between them.

If we have n particles of masses m_1, m_2, ...m_n respectively, along a straight line taken as the x- axis, then by definition the position of the centre of the mass of the system of particles is given by

$$X = \frac{m_1 x_1 + m_2 x_2 + ... + m_n x_n}{m_1 + m_2 + ... + m_n} = \frac{\sum m_i x_i}{\sum m_i} \tag{7.2}$$

where x_1, x_2,...x_n are the distances of the particles from the origin; X is also measured from the same origin. The symbol $\sum$ (the Greek letter sigma) denotes summation, in this case over n particles. The sum

$$\sum m_i = M$$

is the total mass of the system.

Suppose that we have three particles, not lying in a straight line. We may define x and y-axes in the plane in which the particles lie and represent the positions of the three particles by coordinates (x_1, y_1), (x_2, y_2) and (x_3, y_3) respectively. Let the masses of the three particles be m_1, m_2 and m_3 respectively. The centre of mass C of the system of the three particles is defined and located by the coordinates (X, Y) given by

$$X = \frac{m_1 x_1 + m_2 x_2 + m_3 x_3}{m_1 + m_2 + m_3} \tag{7.3a}$$

$$Y = \frac{m_1 y_1 + m_2 y_2 + m_3 y_3}{m_1 + m_2 + m_3} \tag{7.3b}$$

For the particles of equal mass $m = m_1 = m_2 = m_3$,

$$X = \frac{m(x_1 + x_2 + x_3)}{3m} = \frac{x_1 + x_2 + x_3}{3}$$

$$Y = \frac{m(y_1 + y_2 + y_3)}{3m} = \frac{y_1 + y_2 + y_3}{3}$$

Thus, for three particles of equal mass, the centre of mass coincides with the centroid of the triangle formed by the particles.

Results of Eqs. (7.3a) and (7.3b) are generalised easily to a system of n particles, not necessarily lying in a plane, but distributed in space. The centre of mass of such a system is at (X, Y, Z), where

$$X = \frac{\sum m_i x_i}{M} \tag{7.4a}$$

$$Y = \frac{\sum m_i y_i}{M} \tag{7.4b}$$

and $$Z = \frac{\sum m_i z_i}{M} \tag{7.4c}$$

Here $M = \sum m_i$ is the total mass of the system. The index i runs from 1 to n; m_i is the mass of the ith particle and the position of the ith particle is given by (x_i, y_i, z_i).

Eqs. (7.4a), (7.4b) and (7.4c) can be combined into one equation using the notation of position vectors. Let $\mathbf{r}_i$ be the position vector of the ith particle and $\mathbf{R}$ be the position vector of the centre of mass:

$$\mathbf{r}_i = x_i \hat{\mathbf{i}} + y_i \hat{\mathbf{j}} + z_i \hat{\mathbf{k}}$$

and $\mathbf{R} = X\hat{\mathbf{i}} + Y\hat{\mathbf{j}} + Z\hat{\mathbf{k}}$

Then $$\mathbf{R} = \frac{\sum m_i \mathbf{r}_i}{M} \tag{7.4d}$$

The sum on the right hand side is a vector sum.

Note the economy of expressions we achieve by use of vectors. If the origin of the frame of reference (the coordinate system) is chosen to be the centre of mass then $\sum m_i \mathbf{r}_i = 0$ for the given system of particles.

A rigid body, such as a metre stick or a flywheel, is a system of closely packed particles; Eqs. (7.4a), (7.4b), (7.4c) and (7.4d) are therefore, applicable to a rigid body. The number of particles (atoms or molecules) in such a body is so large that it is impossible to carry out the summations over individual particles in these equations. Since the spacing of the particles is

small, we can treat the body as a continuous distribution of mass. We subdivide the body into n small elements of mass; $\Delta m_1, \Delta m_2 \ldots \Delta m_n$; the ith element Δm_i is taken to be located about the point (x_i, y_i, z_i). The coordinates of the centre of mass are then approximately given by

$$X = \frac{\sum(\Delta m_i)x_i}{\sum \Delta m_i}, Y = \frac{\sum(\Delta m_i)y_i}{\sum \Delta m_i}, Z = \frac{\sum(\Delta m_i)z_i}{\sum \Delta m_i}$$

As we make n bigger and bigger and each Δm_i smaller and smaller, these expressions become exact. In that case, we denote the sums over i by integrals. Thus,

$$\sum \Delta m_i \rightarrow \int \mathrm{d}m = M,$$

$$\sum(\Delta m_i)x_i \rightarrow \int x\,\mathrm{d}m,$$

$$\sum(\Delta m_i)y_i \rightarrow \int y\,\mathrm{d}m,$$

and $$\sum(\Delta m_i)z_i \rightarrow \int z\,\mathrm{d}m$$

Here M is the total mass of the body. The coordinates of the centre of mass now are

$$X = \frac{1}{M}\int x\,\mathrm{d}m,\ Y = \frac{1}{M}\int y\,\mathrm{d}m \text{ and } Z = \frac{1}{M}\int z\,\mathrm{d}m \quad (7.5a)$$

The vector expression equivalent to these three scalar expressions is

$$\mathbf{R} = \frac{1}{M}\int \mathbf{r}\,\mathrm{d}m \quad (7.5b)$$

If we choose, the centre of mass as the origin of our coordinate system,

$$\mathbf{R}\,(x, y, z) = \mathbf{0}$$

i.e., $\int \mathbf{r}\,\mathrm{d}m = \mathbf{0}$

or $$\int x\,\mathrm{d}m = \int y\,\mathrm{d}m = \int z\,\mathrm{d}m = 0 \quad (7.6)$$

Often we have to calculate the centre of mass of homogeneous bodies of regular shapes like rings, discs, spheres, rods etc. (By a homogeneous body we mean a body with uniformly distributed mass.) By using symmetry consideration, we can easily show that the centres of mass of these bodies lie at their geometric centres.

Fig. 7.8 Determining the CM of a thin rod.

Let us consider a thin rod, whose width and breath (in case the cross section of the rod is rectangular) or radius (in case the cross section of the rod is cylindrical) is much smaller than its length. Taking the origin to be at the geometric centre of the rod and x-axis to be along the length of the rod, we can say that on account of reflection symmetry, for every element dm of the rod at x, there is an element of the same mass dm located at $-x$ (Fig. 7.8).

The net contribution of every such pair to the integral and hence the integral $\int x\,\mathrm{d}m$ itself is zero. From Eq. (7.6), the point for which the integral itself is zero, is the centre of mass. Thus, the centre of mass of a homogenous thin rod coincides with its geometric centre. This can be understood on the basis of reflection symmetry.

The same symmetry argument will apply to homogeneous rings, discs, spheres, or even thick rods of circular or rectangular cross section. For all such bodies you will realise that for every element dm at a point (x,y,z) one can always take an element of the same mass at the point $(-x,-y,-z)$. (In other words, the origin is a point of reflection symmetry for these bodies.) As a result, the integrals in Eq. (7.5 a) all are zero. This means that for all the above bodies, their centre of mass coincides with their geometric centre.

▶ ***Example 7.1*** Find the centre of mass of three particles at the vertices of an equilateral triangle. The masses of the particles are 100g, 150g, and 200g respectively. Each side of the equilateral triangle is 0.5m long.

Answer

Fig. 7.9

With the X and Y axes chosen as shown in Fig. 7.9, the coordinates of points O, A and B forming the equilateral triangle are respectively (0,0), (0.5,0), (0.25,0.25$\sqrt{3}$). Let the masses 100 g, 150g and 200g be located at O, A and B be respectively. Then,

$$X = \frac{m_1x_1 + m_2x_2 + m_3x_3}{m_1 + m_2 + m_3}$$

$$= \frac{[100(0) + 150(0.5) + 200(0.25)]\ \text{g m}}{(100 + 150 + 200)\ \text{g}}$$

$$= \frac{75 + 50}{450}\text{m} = \frac{125}{450}\text{m} = \frac{5}{18}\text{m}$$

$$Y = \frac{[100(0) + 150(0) + 200(0.25\sqrt{3})]\ \text{g m}}{450\ \text{g}}$$

$$= \frac{50\sqrt{3}}{450}\text{m} = \frac{\sqrt{3}}{9}\text{m} = \frac{1}{3\sqrt{3}}\text{m}$$

The centre of mass C is shown in the figure. Note that it is not the geometric centre of the triangle OAB. Why?

Example 7.2 Find the centre of mass of a triangular lamina.

Answer The lamina (ΔLMN) may be subdivided into narrow strips each parallel to the base *(MN)* as shown in Fig. 7.10

Fig. 7.10

By symmetry each strip has its centre of mass at its midpoint. If we join the midpoint of all the strips we get the median LP. The centre of mass of the triangle as a whole therefore, has to lie on the median LP. Similarly, we can argue that it lies on the median MQ and NR. This means the centre of mass lies on the point of concurrence of the medians, i.e. on the centroid G of the triangle.

Example 7.3 Find the centre of mass of a uniform L-shaped lamina (a thin flat plate) with dimensions as shown. The mass of the lamina is 3 kg.

Answer Choosing the X and Y axes as shown in Fig. 7.11 we have the coordinates of the vertices of the L-shaped lamina as given in the figure. We can think of the L-shape to consist of 3 squares each of length 1m. The mass of each square is 1kg, since the lamina is uniform. The centres of mass C_1, C_2 and C_3 of the squares are, by symmetry, their geometric centres and have coordinates (1/2,1/2), (3/2,1/2), (1/2,3/2) respectively. We take the masses of the squares to be concentrated at these points. The centre of mass of the whole L shape (X, Y) is the centre of mass of these mass points.

Fig. 7.11

Hence

$$X = \frac{[1(1/2) + 1(3/2) + 1(1/2)]\text{kg m}}{(1 + 1 + 1)\text{kg}} = \frac{5}{6}\text{m}$$

$$Y = \frac{[[1(1/2) + 1(1/2) + 1(3/2)]]\text{kg m}}{(1 + 1 + 1)\text{kg}} = \frac{5}{6}\text{m}$$

The centre of mass of the L-shape lies on the line OD. We could have guessed this without calculations. Can you tell why? Suppose, the three squares that make up the L shaped lamina

of Fig. 7.11 had different masses. How will you then determine the centre of mass of the lamina? ◀

7.3 MOTION OF CENTRE OF MASS

Equipped with the definition of the centre of mass, we are now in a position to discuss its physical importance for a system of particles. We may rewrite Eq.(7.4d) as

$$M\mathbf{R} = \sum m_i \mathbf{r}_i = m_1\mathbf{r}_1 + m_2\mathbf{r}_2 + \ldots + m_n\mathbf{r}_n \qquad (7.7)$$

Differentiating the two sides of the equation with respect to time we get

$$M\frac{\mathrm{d}\mathbf{R}}{\mathrm{d}t} = m_1\frac{\mathrm{d}\mathbf{r}_1}{\mathrm{d}t} + m_2\frac{\mathrm{d}\mathbf{r}_2}{\mathrm{d}t} + \ldots + m_n\frac{\mathrm{d}\mathbf{r}_n}{dt}$$

or

$$M\,\mathbf{V} = m_1\mathbf{v}_1 + m_2\mathbf{v}_2 + \ldots + m_n\mathbf{v}_n \qquad (7.8)$$

where $\mathbf{v}_1\,(= \mathrm{d}\mathbf{r}_1/\mathrm{d}t)$ is the velocity of the first particle $\mathbf{v}_2\,(= d\mathbf{r}_2/dt)$ is the velocity of the second particle etc. and $\mathbf{V} = \mathrm{d}\mathbf{R}/\mathrm{d}t$ is the velocity of the centre of mass. Note that we assumed the masses m_1, m_2, ... etc. do not change in time. We have therefore, treated them as constants in differentiating the equations with respect to time.

Differentiating Eq.(7.8) with respect to time, we obtain

$$M\frac{\mathrm{d}\mathbf{V}}{\mathrm{d}t} = m_1\frac{\mathrm{d}\mathbf{v}_1}{\mathrm{d}t} + m_2\frac{\mathrm{d}\mathbf{v}_2}{\mathrm{d}t} + \ldots + m_n\frac{\mathrm{d}\mathbf{v}_n}{\mathrm{d}t}$$

or

$$M\mathbf{A} = m_1\mathbf{a}_1 + m_2\mathbf{a}_2 + \ldots + m_n\mathbf{a}_n \qquad (7.9)$$

where $\mathbf{a}_1\,(= \mathrm{d}\mathbf{v}_1/\mathrm{d}t)$ is the acceleration of the first particle, $\mathbf{a}_2\,(= \mathrm{d}\mathbf{v}_2/\mathrm{d}t)$ is the acceleration of the second particle etc. and $\mathbf{A}\,(= \mathrm{d}\mathbf{V}/\mathrm{d}t)$ is the acceleration of the centre of mass of the system of particles.

Now, from Newton's second law, the force acting on the first particle is given by $\mathbf{F}_1 = m_1\mathbf{a}_1$. The force acting on the second particle is given by $\mathbf{F}_2 = m_2\mathbf{a}_2$ and so on. Eq. (7.9) may be written as

$$M\mathbf{A} = \mathbf{F}_1 + \mathbf{F}_2 + \ldots + \mathbf{F}_n \qquad (7.10)$$

Thus, the total mass of a system of particles times the acceleration of its centre of mass is the vector sum of all the forces acting on the system of particles.

Note when we talk of the force $\mathbf{F}_1$ on the first particle, it is not a single force, but the vector sum of all the forces on the first particle; likewise for the second particle etc. Among these forces on each particle there will be **external** forces exerted by bodies outside the system and also **internal** forces exerted by the particles on one another. We know from Newton's third law that these internal forces occur in equal and opposite pairs and in the sum of forces of Eq. (7.10), their contribution is zero. Only the external forces contribute to the equation. We can then rewrite Eq. (7.10) as

$$M\mathbf{A} = \mathbf{F}_{ext} \qquad (7.11)$$

where $\mathbf{F}_{ext}$ represents the sum of all external forces acting on the particles of the system.

Eq. (7.11) states that **the centre of mass of a system of particles moves as if all the mass of the system was concentrated at the centre of mass and all the external forces were applied at that point.**

Notice, to determine the motion of the centre of mass no knowledge of internal forces of the system of particles is required; for this purpose we need to know only the external forces.

To obtain Eq. (7.11) we did not need to specify the nature of the system of particles. The system may be a collection of particles in which there may be all kinds of internal motions, or it may be a rigid body which has either pure translational motion or a combination of translational and rotational motion. Whatever is the system and the motion of its individual particles, the centre of mass moves according to Eq. (7.11).

Instead of treating extended bodies as single particles as we have done in earlier chapters, we can now treat them as systems of particles. We can obtain the translational component of their motion, i.e. the motion centre of mass of the system, by taking the mass of the whole system to be concentrated at the centre of mass and all the external forces on the system to be acting at the centre of mass.

This is the procedure that we followed earlier in analysing forces on bodies and solving

problems without explicitly outlining and justifying the procedure. We now realise that in earlier studies we assumed, without saying so, that rotational motion and/or internal motion of the particles were either absent or negligible. We no longer need to do this. We have not only found the justification of the procedure we followed earlier; but we also have found how to describe and separate the translational motion of (1) a rigid body which may be rotating as well, or (2) a system of particles with all kinds of internal motion.

Fig. 7.12 The centre of mass of the fragments of the projectile continues along the same parabolic path which it would have followed if there were no explosion.

Figure 7.12 is a good illustration of Eq. (7.11). A projectile, following the usual parabolic trajectory, explodes into fragments midway in air. The forces leading to the explosion are internal forces. They contribute nothing to the motion of the centre of mass. The total external force, namely, the force of gravity acting on the body, is the same before and after the explosion. The centre of mass under the influence of the external force continues, therefore, along the same parabolic trajectory as it would have followed if there were no explosion.

7.4 LINEAR MOMENTUM OF A SYSTEM OF PARTICLES

Let us recall that the linear momentum of a particle is defined as

$$\mathbf{p} = m\,\mathbf{v} \tag{7.12}$$

Let us also recall that Newton's second law written in symbolic form for a single particle is

$$\mathbf{F} = \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} \tag{7.13}$$

where $\mathbf{F}$ is the force on the particle. Let us consider a system of n particles with masses m_1, $m_2, \ldots m_n$ respectively and velocities $\mathbf{v}_1, \mathbf{v}_2, \ldots\ldots \mathbf{v}_n$ respectively. The particles may be interacting and have external forces acting on them. The linear momentum of the first particle is $m_1\mathbf{v}_1$, of the second particle is $m_2\mathbf{v}_2$ and so on.

For the system of n particles, the linear momentum of the system is defined to be the vector sum of all individual particles of the system,

$$\mathbf{P} = \mathbf{p}_1 + \mathbf{p}_2 + \ldots + \mathbf{p}_n$$

$$= m_1\mathbf{v}_1 + m_2\mathbf{v}_2 + \ldots + m_n\mathbf{v}_n \tag{7.14}$$

Comparing this with Eq. (7.8)

$$\mathbf{P} = M\,\mathbf{V} \tag{7.15}$$

Thus, **the total momentum of a system of particles is equal to the product of the total mass of the system and the velocity of its centre of mass**. Differentiating Eq. (7.15) with respect to time,

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}t} = M\frac{\mathrm{d}\mathbf{V}}{\mathrm{d}t} = M\mathbf{A} \tag{7.16}$$

Comparing Eq.(7.16) and Eq. (7.11),

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}t} = \mathbf{F}_{ext} \tag{7.17}$$

This is the statement of **Newton's second law extended to a system of particles.**

Suppose now, that the sum of external forces acting on a system of particles is zero. Then from Eq.(7.17)

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}t} = 0 \quad \textbf{or} \quad \mathbf{P} = \text{Constant} \tag{7.18a}$$

Thus, when the total external force acting on a system of particles is zero, the total linear momentum of the system is constant. This is the law of conservation of the total linear momentum of a system of particles. Because of Eq. (7.15), this also means that when the total external force on the system is zero the velocity of the centre of mass remains constant. (We assume throughout the discussion on systems of particles in this chapter that the total mass of the system remains constant.)

Note that on account of the internal forces, i.e. the forces exerted by the particles on one another, the individual particles may have

complicated trajectories. Yet, if the total external force acting on the system is zero, the centre of mass moves with *a* constant velocity, i.e., moves uniformly in a straight line like a free particle.

The vector Eq. (7.18a) is equivalent to three scalar equations,

$$P_x = c_1,\ P_y = c_2 \text{ and } P_z = c_3 \qquad (7.18\text{ b})$$

Here P_x, P_y and P_z are the components of the total linear momentum vector **P** along the *x*, *y* and *z* axes respectively; c_1, c_2 and c_3 are constants.

Fig. 7.13 (a) A heavy nucleus (Ra) splits into a lighter nucleus (Rn) and an alpha particle (He). The CM of the system is in uniform motion.
(b) The same spliting of the heavy nucleus (Ra) with the centre of mass at rest. The two product particles fly back to back.

As an example, let us consider the radioactive decay of a moving unstable particle, like the nucleus of radium. A radium nucleus disintegrates into a nucleus of radon and an alpha particle. The forces leading to the decay are internal to the system and the external forces on the system are negligible. So the total linear momentum of the system is the same before and after decay. The two particles produced in the decay, the radon nucleus and the alpha particle, move in different directions in such a way that their centre of mass moves along the same path along which the original decaying radium nucleus was moving [Fig. 7.13(a)].

If we observe the decay from the frame of reference in which the centre of mass is at rest, the motion of the particles involved in the decay looks particularly simple; the product particles

Fig. 7.14 (a) Trajectories of two stars, S_1 (dotted line) and S_2 (solid line) forming a binary system with their centre of mass C in uniform motion.
(b) The same binary system, with the centre of mass C at rest.

move back to back with their centre of mass remaining at rest as shown in Fig.7.13 (b).

In many problems on the system of particles as in the above radioactive decay problem, it is convenient to work in the centre of mass frame rather than in the laboratory frame of reference.

In astronomy, binary (double) stars is a common occurrence. If there are no external forces, the centre of mass of a double star moves like a free particle, as shown in Fig.7.14 (a). The trajectories of the two stars of equal mass are also shown in the figure; they look complicated. If we go to the centre of mass frame, then we find that there the two stars are moving in a circle, about the centre of mass, which is at rest. Note that the position of the stars have to be diametrically opposite to each other [Fig. 7.14(b)]. Thus in our frame of reference, the trajectories of the stars are a combination of (i) uniform motion in a straight line of the centre of mass and (ii) circular orbits of the stars about the centre of mass.

As can be seen from the two examples, **separating** the motion of different parts of a system into motion **of the centre of mass and motion about the centre of mass** is a very useful technique that helps in understanding the motion of the system.

7.5 VECTOR PRODUCT OF TWO VECTORS

We are already familiar with vectors and their use in physics. In chapter 6 (Work, Energy, Power) we defined the scalar product of two vectors. An important physical quantity, work, is defined as a scalar product of two vector quantities, force and displacement.

We shall now define another product of two vectors. This product is a vector. Two important quantities in the study of rotational motion, namely, moment of a force and angular momentum, are defined as vector products.

Definition of Vector Product

A vector product of two vectors **a** and **b** is a vector **c** such that

(i) magnitude of **c** = $c = ab\sin\theta$ where a and b are magnitudes of **a** and **b** and θ is the angle between the two vectors.

(ii) **c** is perpendicular to the plane containing **a** and **b**.

(iii) if we take a right handed screw with its head lying in the plane of **a** and **b** and the screw perpendicular to this plane, and if we turn the head in the direction from **a** to **b**, then the tip of the screw advances in the direction of **c**. This right handed screw rule is illustrated in Fig. 7.15a.

Alternately, if one curls up the fingers of right hand around a line perpendicular to the plane of the vectors **a** and **b** and if the fingers are curled up in the direction from **a** to **b**, then the stretched thumb points in the direction of **c**, as shown in Fig. 7.15b.

Fig. 7.15 (a) Rule of the right handed screw for defining the direction of the vector product of two vectors.
(b) Rule of the right hand for defining the direction of the vector product.

A simpler version of the right hand rule is the following : Open up your right hand palm and curl the fingers pointing from **a** to **b**. Your stretched thumb points in the direction of **c**.

It should be remembered that there are two angles between any two vectors **a** and **b** . In Fig. 7.15 (a) or (b) they correspond to θ (as shown) and ($360^0 - \theta$). While applying either of the above rules, the rotation should be taken through the smaller angle ($<180^0$) between **a** and **b**. It is θ here.

Because of the cross used to denote the vector product, it is also referred to as cross product.

- Note that scalar product of two vectors is commutative as said earlier, **a.b** = **b.a**

The vector product, however, is not commutative, i.e. $\mathbf{a} \times \mathbf{b} \neq \mathbf{b} \times \mathbf{a}$

The magnitude of both $\mathbf{a} \times \mathbf{b}$ and $\mathbf{b} \times \mathbf{a}$ is the same ($ab\sin\theta$); also, both of them are perpendicular to the plane of **a** and **b**. But the rotation of the right-handed screw in case of $\mathbf{a} \times \mathbf{b}$ is from **a** to **b**, whereas in case of $\mathbf{b} \times \mathbf{a}$ it is from **b** to **a**. This means the two vectors are in opposite directions. We have

$$\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$$

- Another interesting property of a vector product is its behaviour under reflection. Under reflection (i.e. on taking the mirror image) we have $x \to -x, y \to -y$ **and** $z \to -z$. As a result all the components of a vector change sign and thus $a \to -a,\ b \to -b$. What happens to $\mathbf{a} \times \mathbf{b}$ under reflection?

$$\mathbf{a} \times \mathbf{b} \to (-\mathbf{a}) \times (-\mathbf{b}) = \mathbf{a} \times \mathbf{b}$$

Thus, $\mathbf{a} \times \mathbf{b}$ does not change sign under reflection.

- Both scalar and vector products are distributive with respect to vector addition. Thus,

$$\mathbf{a}.(\mathbf{b} + \mathbf{c}) = \mathbf{a}.\mathbf{b} + \mathbf{a}.\mathbf{c}$$

$$\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}$$

- We may write **c** = **a** × **b** in the component form. For this we first need to obtain some elementary cross products:

(i) $\mathbf{a} \times \mathbf{a} = \mathbf{0}$ (**0** is a null vector, i.e. a vector with zero magnitude)

This follows since magnitude of $\mathbf{a} \times \mathbf{a}$ is $a^2 \sin 0° = 0$ ·

From this follow the results

$\hat{\mathbf{i}} \times \hat{\mathbf{i}} = \mathbf{0},\ \hat{\mathbf{j}} \times \hat{\mathbf{j}} = \mathbf{0},\ \hat{\mathbf{k}} \times \hat{\mathbf{k}} = \mathbf{0}$

(ii) $\hat{\mathbf{i}} \times \hat{\mathbf{j}} = \hat{\mathbf{k}}$

Note that the magnitude of $\hat{\mathbf{i}} \times \hat{\mathbf{j}}$ is $\sin 90^0$ or 1, since $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ both have unit magnitude and the angle between them is 90^0. Thus, $\hat{\mathbf{i}} \times \hat{\mathbf{j}}$ is a unit vector. A unit vector perpendicular to the plane of $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ and related to them by the right hand screw rule is $\hat{\mathbf{k}}$. Hence, the above result. You may verify similarly,

$\hat{\mathbf{j}} \times \hat{\mathbf{k}} = \hat{\mathbf{i}}$ and $\hat{\mathbf{k}} \times \hat{\mathbf{i}} = \hat{\mathbf{j}}$

From the rule for commutation of the cross product, it follows:

$\hat{\mathbf{j}} \times \hat{\mathbf{i}} = -\hat{\mathbf{k}},\ \ \hat{\mathbf{k}} \times \hat{\mathbf{j}} = -\hat{\mathbf{i}},\ \ \hat{\mathbf{i}} \times \hat{\mathbf{k}} = -\hat{\mathbf{j}}$

Note if $\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}$ occur cyclically in the above vector product relation, the vector product is positive. If $\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}$ do not occur in cyclic order, the vector product is negative.

Now,

$$\mathbf{a} \times \mathbf{b} = (a_x\hat{\mathbf{i}} + a_y\hat{\mathbf{j}} + a_z\hat{\mathbf{k}}) \times (b_x\hat{\mathbf{i}} + b_y\hat{\mathbf{j}} + b_z\hat{\mathbf{k}})$$

$$= a_x b_y \hat{\mathbf{k}} - a_x b_z \hat{\mathbf{j}} - a_y b_x \hat{\mathbf{k}} + a_y b_z \hat{\mathbf{i}} + a_z b_x \hat{\mathbf{j}} - a_z b_y \hat{\mathbf{i}}$$

$$= (a_y b_z - a_z b_x)\hat{\mathbf{i}} + (a_z b_x - a_x b_z)\hat{\mathbf{j}} + (a_x b_y - a_y b_x)\hat{\mathbf{k}}$$

We have used the elementary cross products in obtaining the above relation. The expression for $\mathbf{a} \times \mathbf{b}$ can be put in a determinant form which is easy to remember.

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{vmatrix}$$

▶ ***Example 7.4*** *Find the scalar and vector products of two vectors.* $\mathbf{a} = (3\hat{\mathbf{i}} - 4\hat{\mathbf{j}} + 5\hat{\mathbf{k}})$ *and* $\mathbf{b} = (-2\hat{\mathbf{i}} + \hat{\mathbf{j}} - 3\hat{\mathbf{k}})$

Answer

$$\mathbf{a} \cdot \mathbf{b} = (3\hat{\mathbf{i}} - 4\hat{\mathbf{j}} + 5\hat{\mathbf{k}}) \cdot (-2\hat{\mathbf{i}} + \hat{\mathbf{j}} - 3\hat{\mathbf{k}})$$
$$= -6 - 4 - 15$$
$$= -25$$

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ 3 & -4 & 5 \\ -2 & 1 & -3 \end{vmatrix} = 7\hat{\mathbf{i}} - \hat{\mathbf{j}} - 5\hat{\mathbf{k}}$$

Note $\mathbf{b} \times \mathbf{a} = -7\hat{\mathbf{i}} + \hat{\mathbf{j}} + 5\hat{\mathbf{k}}$ ◀

7.6 ANGULAR VELOCITY AND ITS RELATION WITH LINEAR VELOCITY

In this section we shall study what is angular velocity and its role in rotational motion. We have seen that every particle of a rotating body moves in a circle. The linear velocity of the particle is related to the angular velocity. The relation between these two quantities involves a vector product which we learnt about in the last section.

Let us go back to Fig. 7.4. As said above, in rotational motion of a rigid body about a fixed axis, every particle of the body moves in a circle,

Fig. 7.16 *Rotation about a fixed axis. (A particle (P) of the rigid body rotating about the fixed (z-) axis moves in a circle with centre (C) on the axis.)*

which lies in a plane perpendicular to the axis and has its centre on the axis. In Fig. 7.16 we redraw Fig. 7.4, showing a typical particle (at a point P) of the rigid body rotating about a fixed axis (taken as the *z*-axis). The particle describes

a circle with a centre C on the axis. The radius of the circle is r, the perpendicular distance of the point P from the axis. We also show the linear velocity vector **v** of the particle at P. It is along the tangent at P to the circle.

Let P′ be the position of the particle after an interval of time Δt (Fig. 7.16). The angle PCP′ describes the angular displacement $\Delta\theta$ of the particle in time Δt. The average angular velocity of the particle over the interval Δt is $\Delta\theta/\Delta t$. As Δt tends to zero (i.e. takes smaller and smaller values), the ratio $\Delta\theta/\Delta t$ approaches a limit which is the instantaneous angular velocity $d\theta/dt$ of the particle at the position P. We denote the **instantaneous angular velocity** by ω (the Greek letter omega). We know from our study of circular motion that the magnitude of linear velocity v of a particle moving in a circle is related to the angular velocity of the particle ω by the simple relation $v = \omega r$, where r is the radius of the circle.

We observe that at any given instant the relation $v = \omega r$ applies to all particles of the rigid body. Thus for a particle at a perpendicular distance r_i from the fixed axis, the linear velocity at a given instant v_i is given by

$$\omega = d\theta/dt \qquad v_i = \omega r_i \tag{7.19}$$

The index i runs from 1 to n, where n is the total number of particles of the body.

For particles on the axis, $\boldsymbol{r} = \mathbf{0}$, and hence $v = \omega r = 0$. Thus, particles on the axis are stationary. This verifies that the axis is *fixed*.

Note that we use the same angular velocity ω for all the particles. **We therefore, refer to ω as the angular velocity of the whole body**.

We have characterised pure translation of a body by all parts of the body having the same velocity at any instant of time. Similarly, we may characterise pure rotation by all parts of the body having the same angular velocity at any instant of time. Note that this characterisation of the rotation of a rigid body about a fixed axis is **just another way** of saying as in Sec. 7.1 that each particle of the body moves in a circle, which lies in a plane perpendicular to the axis and has the centre on the axis.

In our discussion so far the angular velocity appears to be a scalar. In fact, it is a vector. We shall not justify this fact, but we shall accept it. For rotation about a fixed axis, the angular velocity vector lies along the axis of rotation, and points out in the direction in which a right handed screw would advance, if the head of the screw is rotated with the body. (See Fig. 7.17a).

The magnitude of this vector is referred as above.

Fig. 7.17 (a) If the head of a right handed screw rotates with the body, the screw advances in the direction of the angular velocity ω. If the sense (clockwise or anticlockwise) of rotation of the body changes, so does the direction of ω

Fig. 7.17 (b) The angular velocity vector ω is directed along the fixed axis as shown. The linear velocity of the particle at P is $\boldsymbol{v} = \boldsymbol{\omega} \times \boldsymbol{r}$. *It is perpendicular to both* $\boldsymbol{\omega}$ *and* $\boldsymbol{r}$ *and is directed along the tangent to the circle described by the particle.*

We shall now look at what the vector product $\boldsymbol{\omega} \times \mathbf{r}$ corresponds to. Refer to Fig. 7.17(b) which is a part of Fig. 7.16 reproduced to show the path of the particle P. The figure shows the vector ω directed along the fixed (z-) axis and also the position vector $\mathbf{r} = \mathbf{OP}$ of the particle at P of the rigid body with respect to the origin O. Note that the origin is chosen to be on the axis of rotation.

Now $\boldsymbol{\omega} \times \mathbf{r} = \boldsymbol{\omega} \times \mathbf{OP} = \boldsymbol{\omega} \times (\mathbf{OC} + \mathbf{CP})$

But $\boldsymbol{\omega} \times \mathbf{OC} = \mathbf{0}$ *as* $\boldsymbol{\omega}$ is along **OC**

Hence $\boldsymbol{\omega} \times \mathbf{r} = \boldsymbol{\omega} \times \mathbf{CP}$

The vector $\boldsymbol{\omega} \times \mathbf{CP}$ is perpendicular to ω, i.e. to **CP**, the *z*-axis and also to the radius of the circle described by the particle at P. It is therefore, along the tangent to the circle at P. Also, the magnitude of $\omega \times \mathbf{CP}$ is ω(CP) since ω and **CP** are perpendicular to each other. We shall denote **CP** by $\mathbf{r}_\perp$ and not by **r**, as we did earlier.

Thus, $\omega \times \mathbf{r}$ is a vector of magnitude $\omega r_\perp$ and is along the tangent to the circle described by the particle at P. The linear velocity vector **v** at P has the same magnitude and direction. Thus,

$$\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r} \tag{7.20}$$

In fact, the relation, Eq. (7.20), holds good even for rotation of a rigid body with one point fixed, such as the rotation of the top [Fig. 7.6(a)]. In this case **r** represents the position vector of the particle with respect to the fixed point taken as the origin.

We note that **for rotation about a fixed axis, the direction of the vector ω does not change with time. Its magnitude may, however, change from instant to instant. For the more general rotation, both the magnitude and the direction of ω may change from instant to instant.**

7.6.1 Angular acceleration

You may have noticed that we are developing the study of rotational motion along the lines of the study of translational motion with which we are already familiar. Analogous to the kinetic variables of linear displacement and velocity (**v**) in translational motion, we have angular displacement and angular velocity ($\boldsymbol{\omega}$) in rotational motion. It is then natural to define in rotational motion the concept of angular acceleration in analogy with linear acceleration defined as the time rate of change of velocity in translational motion. We define angular acceleration **α** as the time rate of change of angular velocity; Thus,

$$\boldsymbol{\alpha} = \frac{\mathrm{d}\boldsymbol{\omega}}{\mathrm{d}t} \tag{7.21}$$

If the axis of rotation is fixed, the direction of ω and hence, that of α is fixed. In this case the vector equation reduces to a scalar equation

$$\alpha = \frac{\mathrm{d}\omega}{\mathrm{d}t} \tag{7.22}$$

7.7 TORQUE AND ANGULAR MOMENTUM

In this section, we shall acquaint ourselves with two physical quantities which are defined as vector products of two vectors. These as we shall see, are especially important in the discussion of motion of systems of particles, particularly rigid bodies.

7.7.1 Moment of force (Torque)

We have learnt that the motion of a rigid body in general is a combination of rotation and translation. If the body is fixed at a point or along a line, it has only rotational motion. We know that force is needed to change the translational state of a body, i.e. to produce linear acceleration. We may then ask, what is the analogue of force in the case of rotational motion? To look into the question in a concrete situation let us take the example of opening or closing of a door. A door is a rigid body which can rotate about a fixed vertical axis passing through the hinges. What makes the door rotate? It is clear that unless a force is applied the door does not rotate. But any force does not do the job. A force applied to the hinge line cannot produce any rotation at all, whereas a force of given magnitude applied at right angles to the door at its outer edge is most effective in producing rotation. It is not the force alone, but how and where the force is applied is important in rotational motion.

The rotational analogue of force is **moment of force**. It is also referred to as **torque**. (We shall use the words moment of force and torque interchangeably.) We shall first define the moment of force for the special case of a single particle. Later on we shall extend the concept to systems of particles including rigid bodies. We shall also relate it to a change in the state of rotational motion, i.e. is angular acceleration of a rigid body.

Fig. 7.18 $\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F}$, τ *is perpendicular to the plane containing* **r** *and* **F**, *and its direction is given by the right handed screw rule.*

If a force acts on a single particle at a point P whose position with respect to the origin O is given by the position vector **r** (Fig. 7.18), the moment of the force acting on the particle with respect to the origin O is defined as the vector product

$l = r\,p_{\perp}$

$$\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F} \tag{7.23}$$

The moment of force (or torque) is a vector quantity. The symbol $\boldsymbol{\tau}$ stands for the Greek letter **tau**. The magnitude of $\boldsymbol{\tau}$ is

$$\tau = rF\sin\theta \tag{7.24a}$$

where r is the magnitude of the position vector **r**, i.e. the length OP, F is the magnitude of force **F** and θ is the angle between **r** and **F** as shown.

Moment of force has dimensions $M\,L^2\,T^{-2}$. Its dimensions are the same as those of work or energy. It is, however, a very different physical quantity than work. Moment of a force is a vector, while work is a scalar. The SI unit of moment of force is Newton-metre (Nm). The magnitude of the moment of force may be written

$$\tau = (r\sin\theta)F = r_{\perp}F \tag{7.24b}$$

$$\text{or } \tau = rF\sin\theta = rF_{\perp} \tag{7.24c}$$

where $r_{\perp} = r\sin\theta$ is the perpendicular distance of the line of action of **F** form the origin and $F_{\perp}(= F\sin\theta)$ is the component of **F** in the direction perpendicular to **r**. Note that $\tau = 0$ if $r = 0$, $F = 0$ or $\theta = 0^0$ or 180^0 . Thus, the moment of a force vanishes if either the magnitude of the force is zero, or if the line of action of the force passes through the origin.

One may note that since $\mathbf{r} \times \mathbf{F}$ is a vector product, properties of a vector product of two vectors apply to it. If the direction of **F** is reversed, the direction of the moment of force is reversed. If directions of both **r** and **F** are reversed, the direction of the moment of force remains the same.

7.7.2 Angular momentum of a particle

Just as the moment of a force is the rotational analogue of force, the quantity angular momentum is the rotational analogue of linear momentum. We shall first define angular momentum for the special case of a single particle and look at its usefulness in the context of single particle motion. We shall then extend the definition of angular momentum to systems of particles including rigid bodies.

Like moment of a force, angular momentum is also a vector product. It could also be referred to as moment of (linear) momentum. From this term one could guess how angular momentum is defined.

Consider a particle of mass m and linear momentum **p** at a position **r** relative to the origin O. The angular momentum **l** of the particle with respect to the origin O is defined to be

$$\mathbf{l} = \mathbf{r} \times \mathbf{p} \tag{7.25a}$$

The magnitude of the angular momentum vector is

$$l = r\,p\sin\theta \tag{7.26a}$$

where p is the magnitude of **p** and θ is the angle between **r** and **p**. We may write

$$\text{or } r_{\perp}p \tag{7.26b}$$

where $r_{\perp}$ (= $r\sin\theta$) is the perpendicular distance of the directional line of **p** from the origin and $p_{\perp}(= p\sin\theta)$ is the component of **p** in a direction perpendicular to **r**. We expect the angular momentum to be zero (l = 0), if the linear momentum vanishes (p = 0), if the particle is at the origin (r = 0), or if the directional line of **p** passes through the origin $\theta = 0^0$ or 180^0.

The physical quantities, moment of a force and angular momentum, have an important relation between them. It is the rotational analogue of the relation between force and linear momentum. For deriving the relation in the context of a single particle, we differentiate $\mathbf{l} = \mathbf{r} \times \mathbf{p}$ with respect to time,

$$\frac{d\mathbf{l}}{dt} = \frac{d}{dt}(\mathbf{r} \times \mathbf{p})$$

Applying the product rule for differentiation to the right hand side,

$$\frac{d}{dt}(\mathbf{r} \times \mathbf{p}) = \frac{d\mathbf{r}}{dt} \times \mathbf{p} + \mathbf{r} \times \frac{d\mathbf{p}}{dt}$$

Now, the velocity of the particle is $\mathbf{v} = d\mathbf{r}/dt$ and $\mathbf{p} = m\,\mathbf{v}$

Because of this $\frac{d\mathbf{r}}{dt} \times \mathbf{p} = \mathbf{v} \times m\,\mathbf{v} = 0$,

as the vector product of two parallel vectors vanishes. Further, since $d\mathbf{p} / dt = \mathbf{F}$,

$$\mathbf{r} \times \frac{d\mathbf{p}}{dt} = \mathbf{r} \times \mathbf{F} = \boldsymbol{\tau}$$

Hence $\frac{d}{dt}(\mathbf{r} \times \mathbf{p}) = \boldsymbol{\tau}$

or $$\frac{d\mathbf{l}}{dt} = \boldsymbol{\tau} \qquad (7.27)$$

Thus, the time rate of change of the angular momentum of a particle is equal to the torque acting on it. This is the rotational analogue of the equation $\mathbf{F} = d\mathbf{p}/dt$, which expresses Newton's second law for the translational motion of a single particle.

An experiment with the bicycle rim

Take a bicycle rim and extend its axle on both sides. Tie two strings at both ends A and B, as shown in the adjoining figure. Hold both the strings together in one hand such that the rim is vertical. If you leave one string, the rim will tilt. Now keeping the rim in vertical position with both the strings in one hand, put the wheel in fast rotation around the axle with the other hand. Then leave one string, say B, from your hand, and observe what happens.

The rim keeps rotating in a vertical plane and the plane of rotation turns around the string A which you are holding. We say that the axis of rotation of the rim or equivalently its angular momentum precesses about the string A.

The rotating rim gives rise to an angular momentum. Determine the direction of this angular momentum. When you are holding the rotating rim with string A, a torque is generated. (We leave it to you to find out how the torque is generated and what its direction is.) The effect of the torque on the angular momentum is to make it precess around an axis perpendicular to both the angular momentum and the torque. Verify all these statements.

Torque and angular momentum for a system of particles

To get the total angular momentum of a system of particles about a given point we need to add vectorially the angular momenta of individual particles. Thus, for a system of n particles,

$$\mathbf{L} = \mathbf{l}_1 + \mathbf{l}_2 + \ldots + \mathbf{l}_n = \sum_{i=1}^{n} \mathbf{l}_i$$

The angular momentum of the ith particle is given by

$\mathbf{l}_i = \mathbf{r}_i \times \mathbf{p}_i$

where $\mathbf{r}_i$ is the position vector of the ith particle with respect to a given origin and $\mathbf{p} = (m_i\mathbf{v}_i)$ is the linear momentum of the particle. (The particle has mass m_i and velocity $\mathbf{v}_i$) We may write the total angular momentum of a system of particles as

$$\mathbf{L} = \sum \mathbf{l}_i = \sum_i \mathbf{r}_i \times \mathbf{p}_i \qquad (7.25b)$$

This is a generalisation of the definition of angular momentum (Eq. 7.25a) for a single particle to a system of particles.

Using Eqs. (7.23) and (7.25b), we get

$$\frac{d\mathbf{L}}{dt} = \frac{d}{dt}\left(\sum \mathbf{l}_i\right) = \sum_i \frac{d\mathbf{l}_i}{dt} = \sum_i \boldsymbol{\tau}_i \qquad (7.28a)$$

where $\boldsymbol{\tau}_i$ is the torque acting on the ith particle:

$$\boldsymbol{\tau}_i = \mathbf{r}_i \times \mathbf{F}_i$$

The force $\mathbf{F}_i$ on the ith particle is the vector sum of external forces $\mathbf{F}_i^{ext}$ acting on the particle and the internal forces $\mathbf{F}_i^{int}$ exerted on it by the other particles of the system. We may therefore separate the contribution of the external and the internal forces to the total torque $\boldsymbol{\tau} = \sum_i \boldsymbol{\tau}_i = \sum_i \mathbf{r}_i \times \mathbf{F}_i$ as

$$\boldsymbol{\tau} = \boldsymbol{\tau}_{ext} + \boldsymbol{\tau}_{int},$$

where $\boldsymbol{\tau}_{ext} = \sum_i \mathbf{r}_i \times \mathbf{F}_i^{ext}$

and $\boldsymbol{\tau}_{int} = \sum_i \mathbf{r}_i \times \mathbf{F}_i^{int}$

We shall assume not only Newton's third law, i.e. the forces between any two particles of the system are equal and opposite, but also that these forces are directed along the line joining the two particles. In this case the contribution of the internal forces to the total torque on the system is zero, since the torque resulting from each action-reaction pair of forces is zero. We thus have, $\boldsymbol{\tau}_{int} = \mathbf{0}$ and therefore $\boldsymbol{\tau} = \boldsymbol{\tau}_{ext}$.

Since $\boldsymbol{\tau} = \sum \boldsymbol{\tau}_i$, it follows from Eq. (7.28a) that

$$\frac{d\mathbf{L}}{dt} = \boldsymbol{\tau}_{ext} \qquad (7.28\text{ b})$$

Thus, the time rate of the total angular momentum of the system is equal to the total torque (or the external torques) acting on the system taken about the same point. Eq. (7.28 b) is the generalisation of the single particle case of Eq. (7.23) to a system of particles. Note that when we have only one particle, there are no internal forces or torques. Eq.(7.28 b) is the rotational analogue of

$$\frac{d\mathbf{P}}{dt} = \mathbf{F}_{ext} \qquad (7.17)$$

Note that like Eq.(7.17), Eq.(7.28b) holds good for any system of particles, whether it is a rigid body or its individual particles have all kinds of internal motion.

Conservation of angular momentum

If $\boldsymbol{\tau}_{ext} = \mathbf{0}$, Eq. (7.28b) reduces to

$$\frac{d\mathbf{L}}{dt} = 0$$

or $\mathbf{L} =$ constant. (7.29a)

Thus, if the total external torque on a system of particles is zero, then the total angular momentum of the system is conserved, i.e. remains constant. Eq. (7.29a) is equivalent to three scalar equations,

$$L_x = K_1,\ L_y = K_2 \text{ and } L_z = K_3 \qquad (7.29\text{ b})$$

Here K_1, K_2 and K_3 are constants; L_x, L_y and L_z are the components of the total angular momentum vector $\mathbf{L}$ along the x, y and z axes respectively. The statement that the total angular momentum is conserved means that each of these three components is conserved.

Eq. (7.29a) is the rotational analogue of Eq. (7.18a), i.e. the conservation law of the total linear momentum for a system of particles. Like Eq. (7.18a), it has applications in many practical situations. We shall look at a few of the interesting applications later on in this chapter.

▶ ***Example 7.5*** Find the torque of a force $7\hat{\mathbf{i}} + 3\hat{\mathbf{j}} - 5\hat{\mathbf{k}}$ about the origin. The force acts on a particle whose position vector is $\hat{\mathbf{i}} - \hat{\mathbf{j}} + \hat{\mathbf{k}}$.

Answer Here $\mathbf{r} = \hat{\mathbf{i}} - \hat{\mathbf{j}} + \hat{\mathbf{k}}$

and $\mathbf{F} = 7\hat{\mathbf{i}} + 3\hat{\mathbf{j}} - 5\hat{\mathbf{k}}$.

$$\boldsymbol{\tau} = \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ 1 & -1 & 1 \\ 7 & 3 & -5 \end{vmatrix} = (5-3)\hat{\mathbf{i}} - (-5-7)\hat{\mathbf{j}} + (3-(-7))\hat{\mathbf{k}}$$

or $\boldsymbol{\tau} = 2\hat{\mathbf{i}} + 12\hat{\mathbf{j}} + 10\hat{\mathbf{k}}$ ◀

▶ ***Example 7.6*** Show that the angular momentum about any point of a single particle moving with constant velocity remains constant throughout the motion.

Answer Let the particle with velocity **v** be at point P at some instant *t*. We want to calculate the angular momentum of the particle about an arbitrary point O.

Fig 7.19

The angular momentum is $\mathbf{l} = \mathbf{r} \times m\mathbf{v}$. Its magnitude is $mvr \sin\theta$, where θ is the angle between **r** and **v** as shown in Fig. 7.19. Although the particle changes position with time, the line of direction of **v** remains the same and hence OM = $r \sin \theta$. is a constant.

Further, the direction of **l** is perpendicular to the plane of **r** and **v**. It is into the page of the figure.This direction does not change with time.

Thus, **l** remains the same in magnitude and direction and is therefore conserved. Is there any external torque on the particle? ◀

7.8 EQUILIBRIUM OF A RIGID BODY

We are now going to concentrate on the motion of rigid bodies rather than on the motion of general systems of particles.

We shall recapitulate what effect the external forces have on a rigid body. (Henceforth we shall omit the adjective 'external' because unless stated otherwise, we shall deal with only external forces and torques.) The forces change the translational state of the motion of the rigid body, i.e. they change its total linear momentum in accordance with Eq. (7.17). But this is not the only effect the forces have. The total torque on the body may not vanish. Such a torque changes the rotational state of motion of the rigid body, i.e. it changes the total angular momentum of the body in accordance with Eq. (7.28 b).

A rigid body is said to be in mechanical equilibrium, if both its linear momentum and angular momentum are not changing with time, or equivalently, the body has neither linear acceleration nor angular acceleration. This means

(1) the total force, i.e. the vector sum of the forces, on the rigid body is zero;

$$\mathbf{F}_1 + \mathbf{F}_2 + \ldots + \mathbf{F}_n = \sum_{i=1}^{n} \mathbf{F}_i = \mathbf{0} \qquad (7.30a)$$

If the total force on the body is zero, then the total linear momentum of the body does not change with time. Eq. (7.30a) gives the condition for the translational equilibrium of the body.

(2) The total torque, i.e. the vector sum of the torques on the rigid body is zero,

$$\boldsymbol{\tau}_1 + \boldsymbol{\tau}_2 + \ldots + \boldsymbol{\tau}_n = \sum_{i=1}^{n} \boldsymbol{\tau}_i = \mathbf{0} \qquad (7.30b)$$

If the total torque on the rigid body is zero, the total angular momentum of the body does not change with time. Eq. (7.30 b) gives the condition for the rotational equilibrium of the body.

One may raise a question, whether the rotational equilibrium condition [Eq. 7.30(b)] remains valid, if the origin with respect to which the torques are taken is shifted. One can show that if the translational equilibrium condition [Eq. 7.30(a)] holds for a rigid body, then such a shift of origin does not matter, i.e. the rotational equilibrium condition is independent of the location of the origin about which the torques are taken. Example 7.7 gives a proof of this result in a special case of a couple, i.e. two forces acting on a rigid body in translational equilibrium. The generalisation of this result to *n* forces is left as an exercise.

Eq. (7.30a) and Eq. (7.30b), both, are vector equations. They are equivalent to three scalar equations each. Eq. (7.30a) corresponds to

$$\sum_{i=1}^{n} F_{ix} = 0, \quad \sum_{i=1}^{n} F_{iy} = 0 \quad \text{and} \quad \sum_{i=1}^{n} F_{iz} = 0 \qquad (7.31a)$$

where F_{ix}, F_{iy} and F_{iz} are respectively the *x*, *y* and *z* components of the forces $\mathbf{F}_i$. Similarly, Eq. (7.30b) is equivalent to three scalar equations

$$\sum_{i=1}^{n} \tau_{ix} = 0, \quad \sum_{i=1}^{n} \tau_{iy} = 0 \quad \text{and} \quad \sum_{i=1}^{n} \tau_{iz} = 0 \qquad (7.31b)$$

where τ_{ix}, τ_{iy} and τ_{iz} are respectively the *x*, *y* and *z* components of the torque τ_i.

Eq. (7.31a) and (7.31b) give six independent conditions to be satisfied for mechanical equilibrium of a rigid body. In a number of problems all the forces acting on the body are coplanar. Then we need only three conditions to be satisfied for mechanical equilibrium. Two of these conditions correspond to translational equilibrium; the sum of the components of the forces along any two perpendicular axes in the plane must be zero. The third condition corresponds to rotational equilibrium. The sum of the components of the torques along any axis perpendicular to the plane of the forces must be zero.

The conditions of equilibrium of a rigid body may be compared with those for a particle, which we considered in earlier chapters. Since consideration of rotational motion does not apply to a particle, only the conditions for translational equilibrium (Eq. 7.30 a) apply to a particle. Thus, for equilibrium of a particle the vector sum of all the forces on it must be zero. Since all these forces act on the single particle, they must be concurrent. Equilibrium under concurrent forces was discussed in the earlier chapters.

A body may be in partial equilibrium, i.e., it may be in translational equilibrium and not in rotational equilibrium, or it may be in rotational equilibrium and not in translational equilibrium.

Consider a light (i.e. of negligible mass) rod (AB), at the two ends (A and B) of which two parallel forces both equal in magnitude are applied perpendicular to the rod as shown in Fig. 7.20(a).

Fig. 7.20 (a)

Let C be the midpoint of AB, CA = CB = a. the moment of the forces at A and B will both be equal in magnitude (aF), but opposite in sense as shown. The net moment on the rod will be zero. The system will be in rotational equilibrium, but it will not be in translational equilibrium; $\sum \mathbf{F} \neq \mathbf{0}$

Fig. 7.20 (b)

The force at B in Fig. 7.20(a) is reversed in Fig. 7.20(b). Thus, we have the same rod with two equal and opposite forces applied perpendicular to the rod, one at end A and the other at end B. Here the moments of both the forces are equal, but they are not opposite; they act in the same sense and cause anticlockwise rotation of the rod. The total force on the body is zero; so the body is in translational equilibrium; but it is not in rotational equilibrium. Although the rod is not fixed in any way, it undergoes pure rotation (i.e. rotation without translation).

A pair of equal and opposite forces with different lines of action is known as a **couple**. A couple produces rotation without translation.

When we open the lid of a bottle by turning it, our fingers are applying a couple to the lid [Fig. 7.21(a)]. Another known example is a compass needle in the earth's magnetic field as shown in the Fig. 7.21(b). The earth's magnetic field exerts equal forces on the north and south poles. The force on the North Pole is towards the north, and the force on the South Pole is toward the south. Except when the needle points in the north-south direction; the two forces do not have the same line of action. Thus there is a **couple** acting on the needle due to the earth's magnetic field.

Fig. 7.21(a) Our fingers apply a couple to turn the lid.

Fig. 7.21(b) The Earth's magnetic field exerts equal and opposite forces on the poles of a compass needle. These two forces form a couple.

▶ ***Example 7.7*** Show that moment of a couple does not depend on the point about which you take the moments.

Answer

Fig. 7.22

Consider a couple as shown in Fig. 7.22 acting on a rigid body. The forces **F** and -**F** act respectively at points B and A. These points have position vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ with respect to origin O. Let us take the moments of the forces about the origin.

The moment of the couple = sum of the moments of the two forces making the couple

$= \mathbf{r}_1 \times (-\mathbf{F}) + \mathbf{r}_2 \times \mathbf{F}$

$= \mathbf{r}_2 \times \mathbf{F} - \mathbf{r}_1 \times \mathbf{F}$

$= (\mathbf{r}_2 - \mathbf{r}_1) \times \mathbf{F}$

But $\mathbf{r}_1 + \mathbf{AB} = \mathbf{r}_2$, and hence $\mathbf{AB} = \mathbf{r}_2 - \mathbf{r}_1$.

The moment of the couple, therefore, is $\mathbf{AB} \times \mathbf{F}$.

Clearly this is independent of the origin, the point about which we took the moments of the forces. ◀

7.8.1 Principle of moments

An ideal lever is essentially a light (i.e. of negligible mass) rod pivoted at a point along its length. This point is called the fulcrum. A see-saw on the children's playground is a typical example of a lever. Two forces F_1 and F_2, parallel to each other and usually perpendicular to the lever, as shown here, act on the lever at distances d_1 and d_2 respectively from the fulcrum as shown in Fig. 7.23.

Fig. 7.23

The lever is a system in mechanical equilibrium. Let **R** be the reaction of the support at the fulcrum; **R** is directed opposite to the forces F_1 and F_2. For translational equilibrium,

$$R - F_1 - F_2 = 0 \qquad \text{(i)}$$

For considering rotational equilibrium we take the moments about the fulcrum; the sum of moments must be zero,

$$d_1F_1 - d_2F_2 = 0 \qquad \text{(ii)}$$

Normally the anticlockwise (clockwise) moments are taken to be positive (negative). Note R acts at the fulcrum itself and has zero moment about the fulcrum.

In the case of the lever force F_1 is usually some weight to be lifted. It is called the *load* and its distance from the fulcrum d_1 is called the *load arm*. Force F_2 is the *effort* applied to lift the load; distance d_2 of the effort from the fulcrum is the *effort arm*.

Eq. (ii) can be written as

$$d_1F_1 = d_2F_2 \qquad \text{(7.32a)}$$

or load arm × load = effort arm × effort

The above equation expresses the principle of moments for a lever. Incidentally the ratio F_1/F_2 is called the Mechanical Advantage (M.A.);

$$\text{M.A.} = \frac{F_1}{F_2} = \frac{d_2}{d_1} \qquad \text{(7.32b)}$$

If the effort arm d_2 is larger than the load arm, the mechanical advantage is greater than one. Mechanical advantage greater than one means that a small effort can be used to lift a large load. There are several examples of a lever around you besides the see-saw. The beam of a balance is a lever. Try to find more such

examples and identify the fulcrum, the effort and effort arm, and the load and the load arm of the lever in each case.

You may easily show that the principle of moment holds even when the parallel forces F_1 and F_2 are not perpendicular, but act at some angle, to the lever.

7.8.2 Centre of gravity

Many of you may have the experience of balancing your notebook on the tip of a finger. Figure 7.24 illustrates a similar experiment that you can easily perform. Take an irregular-shaped cardboard and a narrow tipped object like a pencil. You can locate by trial and error a point G on the cardboard where it can be balanced on the tip of the pencil. (The cardboard remains horizontal in this position.) This point of balance is the centre of gravity (CG) of the cardboard. The tip of the pencil provides a vertically upward force due to which the cardboard is in mechanical equilibrium. As shown in the Fig. 7.24, the reaction of the tip is equal and opposite to $M\mathbf{g}$, the total weight of (i.e., the force of gravity on) the cardboard and hence the cardboard is in translational equilibrium. It is also in rotational equilibrium; if it were not so, due to the unbalanced torque it would tilt and fall. There are torques on the card board due to the forces of gravity like $m_1\mathbf{g}$, $m_2\mathbf{g}$ .... etc, acting on the individual particles that make up the cardboard.

Fig. 7.24 Balancing a cardboard on the tip of a pencil. The point of support, G, is the centre of gravity.

The CG of the cardboard is so located that the total torque on it due to the forces $m_1\mathbf{g}$, $m_2\mathbf{g}$ .... etc. is zero.

If $\mathbf{r}_i$ is the position vector of the ith particle of an extended body with respect to its CG, then the torque about the CG, due to the force of gravity on the particle is $\boldsymbol{\tau}_i = \mathbf{r}_i \times m_i\mathbf{g}$. The total gravitational torque about the CG is zero, i.e.

$$\boldsymbol{\tau}_g = \sum \boldsymbol{\tau}_i = \sum \mathbf{r}_i \times m_i\mathbf{g} = \mathbf{0} \qquad (7.33)$$

We may therefore, define the CG of a body as that point where the total gravitational torque on the body is zero.

We notice that in Eq. (7.33), $\mathbf{g}$ is the same for all particles, and hence it comes out of the summation. This gives, since $\mathbf{g}$ is nonzero, $\sum m_i\mathbf{r}_i = \mathbf{0}$. Remember that the position vectors ($\mathbf{r}_i$) are taken with respect to the CG. Now, in accordance with the reasoning given below Eq. (7.4a) in Sec. 7.2, if the sum is zero, the origin must be the centre of mass of the body. Thus, the centre of gravity of the body coincides with the centre of mass. We note that this is true because the body being small, $\mathbf{g}$ does not

Fig. 7.25 Determining the centre of gravity of a body of irregular shape. The centre of gravity G lies on the vertical AA_1 through the point of suspension of the body A.

vary from one point of the body to the other. If the body is so extended that **g** varies from part to part of the body, then the centre of gravity and centre of mass will not coincide. Basically, the two are different concepts. The centre of mass has nothing to do with gravity. It depends only on the distribution of mass of the body.

In Sec. 7.2 we found out the position of the centre of mass of several regular, homogeneous objects. Obviously the method used there gives us also the centre of gravity of these bodies, if they are small enough.

Figure 7.25 illustrates another way of determining the CG of an regular shaped body like a cardboard. If you suspend the body from some point like A, the vertical line through A passes through the CG. We mark the vertical AA_1. We then suspend the body through other points like B and C. The intersection of the verticals gives the CG. Explain why the method works. Since the body is small enough, the method allows us to determine also its centre of mass.

▶ ***Example 7.8*** A metal bar 70 cm long and 4.00 kg in mass supported on two knife-edges placed 10 cm from each end. A 6.00 kg weight is suspended at 30 cm from one end. Find the reactions at the knife-edges. (Assume the bar to be of uniform cross section and homogeneous.)

Answer

Fig. 7.26

Figure 7.26 shows the rod AB, the positions of the knife edges K_1 and K_2 , the centre of gravity of the rod at G and the suspended weight at P.

Note the weight of the rod W acts at its centre of gravity G. The rod is uniform in cross section and homogeneous; hence G is at the centre of the rod; AB = 70 cm. AG = 35 cm, AP = 30 cm, PG = 5 cm, $AK_1 = BK_2 = 10$ cm and $K_1G = K_2G = 25$ cm. Also, W= weight of the rod = 4.00 kg and W_1= suspended weight = 6.00 kg; R_1 and R_2 are the normal reactions of the support at the knife edges.

For translational equilibrium of the rod,

$R_1 + R_2 - W_1 - W = 0$ (i)

Note W_1 and W act vertically down and R_1 and R_2 act vertically up.

For considering rotational equilibrium, we take moments of the forces. A convenient point to take moments about is G. The moments of R_2 and W_1 are anticlockwise (+ve), whereas the moment of R_1 is clockwise (-ve).

For rotational equilibrium,

$-R_1 (K_1G) + W_1 (PG) + R_2 (K_2G) = 0$ (ii)

It is given that $W = 4.00g$ N and $W_1 = 6.00g$ N, where g = acceleration due to gravity. We take $g = 9.8$ m/s^2.

With numerical values inserted, from (i)

$R_1 + R_2 - 4.00g - 6.00g = 0$

or $R_1 + R_2 = 10.00g$ N (iii)

$= 98.00$ N

From (ii) $-0.25\, R_1 + 0.05\, W_1 + 0.25\, R_2 = 0$

or $R_2 - R_1 = 1.2g$ N = 11.76 N (iv)

From (iii) and (iv), $R_1 = 54.88$ N,

$R_2 = 43.12$ N

Thus the reactions of the support are about 55 N at K_1 and 43 N at K_2. ◀

▶ ***Example 7.9*** A 3m long ladder weighing 20 kg leans on a frictionless wall. Its feet rest on the floor 1 m from the wall as shown in Fig.7.27. Find the reaction forces of the wall and the floor.

Answer

Fig. 7.27

The ladder AB is 3 m long, its foot A is at distance AC = 1 m from the wall. From

Pythagoras theorem, BC = $2\sqrt{2}$ m. The forces on the ladder are its weight W acting at its centre of gravity D, reaction forces F_1 and F_2 of the wall and the floor respectively. Force F_1 is perpendicular to the wall, since the wall is frictionless. Force F_2 is resolved into two components, the normal reaction N and the force of friction F. Note that F prevents the ladder from sliding away from the wall and is therefore directed toward the wall.

For translational equilibrium, taking the forces in the vertical direction,

$$N - W = 0 \quad \text{(i)}$$

Taking the forces in the horizontal direction,

$$F - F_1 = 0 \quad \text{(ii)}$$

For rotational equilibrium, taking the moments of the forces about A,

$$2\sqrt{2}\,F_1 - (1/2)\,W = 0 \quad \text{(iii)}$$

Now $W = 20$ g $= 20 \times 9.8$ N $= 196.0$ N

From (i) $N = 196.0$

From (iii) $F_1 = W/4\sqrt{2} = 196.0/4\sqrt{2} = 34.6$ N

From (ii) $F = F_1 = 34.6$ N

$$F_2 = \sqrt{F^2 + N^2} = 199.0 \text{ N}$$

The force F_2 makes an angle α with the horizontal,

$$\tan\alpha = N/F = 4\sqrt{2}, \quad \alpha = \tan^{-1}(4\sqrt{2}) \approx 80^\circ$$ ◀

7.9 MOMENT OF INERTIA

We have already mentioned that we are developing the study of rotational motion parallel to the study of translational motion with which we are familiar. We have yet to answer one major question in this connection. **What is the analogue of mass in rotational motion?** We shall attempt to answer this question in the present section. To keep the discussion simple, we shall consider rotation about a fixed axis only. Let us try to get an expression for *the* **kinetic energy of a rotating body**. We know that for a body rotating about a fixed axis, each particle of the body moves in a circle with linear velocity given by Eq. (7.19). (Refer to Fig. 7.16). For a particle at a distance from the axis, the linear velocity is $v_i = r_i\omega$. The kinetic energy of motion of this particle is

$$k_i = \frac{1}{2}m_i v_i^2 = \frac{1}{2}m_i r_i^2 \omega^2$$

where m_i is the mass of the particle. The total kinetic energy K of the body is then given by the sum of the kinetic energies of individual particles,

$$K = \sum_{i=1}^{n} k_i = \frac{1}{2}\sum_{i=1}^{n}(m_i r_i^2 \omega^2)$$

Here n is the number of particles in the body. Note ω is the same for all particles. Hence, taking ω out of the sum,

$$K = \frac{1}{2}\omega^2\left(\sum_{i=1}^{n} m_i r_i^2\right)$$

We define a new parameter characterising the rigid body, called the moment of inertia I, given by

$$I = \sum_{i=1}^{n} m_i r_i^2 \quad (7.34)$$

With this definition,

$$K = \frac{1}{2}I\omega^2 \quad (7.35)$$

Note that the parameter I is independent of the magnitude of the angular velocity. It is a characteristic of the rigid body and the axis about which it rotates.

Compare Eq. (7.35) for the kinetic energy of a rotating body with the expression for the kinetic energy of a body in linear (translational) motion,

$$K = \frac{1}{2}m\,v^2$$

Here m is the mass of the body and v is its velocity. We have already noted the analogy between angular velocity ω (in respect of rotational motion about a fixed axis) and linear velocity v (in respect of linear motion). It is then evident that the parameter, moment of inertia I, is the desired rotational analogue of mass. In rotation (about a fixed axis), the moment of inertia plays a similar role as mass does in linear motion.

We now apply the definition Eq. (7.34), to calculate the moment of inertia in two simple cases.

(a) Consider a thin ring of radius R and mass M, rotating in its own plane around its

centre with angular velocity ω Each mass element of the ring is at a distance R from the axis, and moves with a speed $R\omega$ The kinetic energy is therefore,

$$K = \frac{1}{2}Mv^2 = \frac{1}{2}MR^2\omega^2$$

Comparing with Eq. (7.35) we get $I = MR^2$ for the ring.

Fig. 7.28 A light rod of length l with a pair of masses rotating about an axis through the centre of mass of the system and perpendicular to the rod. The total mass of the system is M.

(b) Next, take a rigid massless rod of length l with a pair of small masses, rotating about an axis through the centre of mass perpendicular to the rod (Fig. 7.28). Each mass $M/2$ is at a distance $l/2$ from the axis. The moment of inertia of the masses is therefore given by

$$(M/2)\,(l/2)^2 + (M/2)(l/2)^2$$

Thus, for the pair of masses, rotating about the axis through the centre of mass perpendicular to the rod

$I = Ml^2 / 4$

Table 7.1 gives the moment of inertia of various familiar regular shaped solids about specific axes.

As the mass of a body resists a change in its state of linear motion, it is a measure of its inertia in linear motion. Similarly, as the moment of inertia about a given axis of rotation resists a change in its rotational motion, it can be regarded as a measure of rotational inertia of the body; it is a measure of the way in which different parts of the body are distributed at different distances from the axis. Unlike the mass of a body, the moment of inertia is not a fixed quantity but depends on the orientation and position of the axis of rotation with respect to the body as a whole. As a measure of the way in which the mass of a rotating rigid body is distributed with respect to the axis of rotation, we can define a new parameter, the **radius of gyration**. It is related to the moment of inertia and the total mass of the body.

Notice from the Table 7.1 that in all cases, we can write $I = Mk^2$, where k has the dimension of length. For a rod, about the perpendicular axis at its midpoint, $k^2 = L^2/12$, **i.e.** $k^2 = L/\sqrt{12}$. Similarly, $k = R/2$ for the circular disc about its diameter. The length k is a geometric property of the body and axis of rotation. It is called the **radius of gyration**. **The radius of gyration of a body about an axis** may be defined as the distance from the axis of a mass point whose mass is equal to the mass of the whole body and whose moment of inertia is equal to the moment of inertia of the body about the axis.

Thus, the moment of inertia of a rigid body depends on the mass of the body, its shape and size; distribution of mass about the axis of rotation, and the position and orientation of the axis of rotation.

From the definition, Eq. (7.34), we can infer that the dimensions of moments of inertia we ML^2 and its SI units are kg m^2.

The property of this extremely important quantity I as a measure of rotational inertia of the body has been put to a great practical use. The machines, such as steam engine and the automobile engine, etc., that produce rotational motion have a disc with a large moment of inertia, called a **flywheel**. Because of its large moment of inertia, the flywheel resists the sudden increase or decrease of the speed of the vehicle. It allows a gradual change in the speed and prevents jerky motions, thereby ensuring a smooth ride for the passengers on the vehicle.

7.10 THEOREMS OF PERPENDICULAR AND PARALLEL AXES

These are two useful theorems relating to moment of inertia. We shall first discuss the theorem of perpendicular axes and its simple yet instructive application in working out the moments of inertia of some regular-shaped bodies.

Table 7.1 Moments of Inertia of some regular shaped bodies about specific axes

Z	Body	Axis	Figure	*I*
(1)	Thin circular ring, radius *R*	Perpendicular to plane, at centre		$M R^2$
(2)	Thin circular ring, radius *R*	Diameter		$M R^2/2$
(3)	Thin rod, length L	Perpendicular to rod, at mid point		$M L^2/12$
(4)	Circular disc, radius *R*	Perpendicular to disc at centre		$M R^2/2$
(5)	Circular disc, radius *R*	Diameter		$M R^2/4$
(6)	Hollow cylinder, radius *R*	Axis of cylinder		$M R^2$
(7)	Solid cylinder, radius *R*	Axis of cylinder		$M R^2/2$
(8)	Solid sphere, radius *R*	Diameter		$2 M R^2/5$

Theorem of perpendicular axes

This theorem is applicable to bodies which are planar. In practice this means the theorem applies to flat bodies whose thickness is very small compared to their other dimensions (e.g. length, breadth or radius). Fig. 7.29 illustrates the theorem. It states that **the moment of inertia of a planar body (lamina) about an axis perpendicular to its plane is equal to the sum of its moments of inertia about two perpendicular axes concurrent with perpendicular axis and lying in the plane of the body.**

Fig. 7.29 *Theorem of perpendicular axes applicable to a planar body; x and y axes are two perpendicular axes in the plane and the z-axis is perpendicular to the plane.*

The figure shows a planar body. An axis perpendicular to the body through a point O is taken as the z-axis. Two mutually perpendicular axes lying in the plane of the body and concurrent with z-axis, i.e. passing through O, are taken as the x and y-axes. The theorem states that

$$I_z = I_x + I_y \tag{7.36}$$

Let us look at the usefulness of the theorem through an example.

Example 7.10 What is the moment of inertia of a disc about one of its diameters?

Fig. 7.30 *M.I. of a disc about a diameter, given its M.I. about the perpendicular axis through its centre.*

Answer We assume the moment of inertia of the disc about an axis perpendicular to it and through its centre to be known; it is $MR^2/2$, where M is the mass of the disc and R is its radius (Table 7.1)

The disc can be considered to be a planar body. Hence the theorem of perpendicular axes is applicable to it. As shown in Fig. 7.30, we take three concurrent axes through the centre of the disc, O as the x,y,z axes; x and y-axes lie in the plane of the disc and z is perpendicular to it. By the theorem of perpendicular axes,

$$I_z = I_x + I_y$$

Now, x and y axes are along two diameters of the disc, and by symmetry the moment of inertia of the disc is the same about any diameter. Hence

$$I_x = I_y$$

and $\quad I_z = 2I_x$

But $\quad I_z = MR^2/2$

So finally, $\quad I_x = I_z/2 = MR^2/4$

Thus the moment of inertia of a disc about any of its diameter is $MR^2/4$.

Find similarly the moment of inertia of a ring about any of its diameter. Will the theorem be applicable to a solid cylinder?

Fig.7.31 *The theorem of parallel axes The z and z′ axes are two parallel axes separated by a distance a; O is the centre of mass of the body, **OO′** = a.*

7.10.1 Theorem of parallel axes

This theorem is applicable to a body of any shape. It allows to find the moment of inertia of a body about any axis, given the moment of inertia of the body about a parallel axis through the centre of mass of the body. We shall only state this theorem and not give its proof. We shall, however, apply it to a few simple situations which will be enough to convince us about the usefulness of the theorem. The theorem may be stated as follows:

The moment of inertia of a body about any axis is equal to the sum of the moment of inertia of the body about a parallel axis passing through its centre of mass and the product of its mass and the square of the distance between the two parallel axes. As shown in the Fig. 7.31, z and z' are two parallel axes separated by a distance a. The z-axis passes through the centre of mass O of the rigid body. Then according to the theorem of parallel axes

$$I_{z'} = I_z + Ma^2 \qquad (7.37)$$

where I_z and $I_{z'}$ are the moments of inertia of the body about the z and z' axes respectively, M is the total mass of the body and a is the perpendicular distance between the two parallel axes.

▶ ***Example 7.11*** What is the moment of inertia of a rod of mass M, length l about an axis perpendicular to it through one end?

Answer For the rod of mass M and length l, $I = Ml^2/12$. Using the parallel axes theorem, $I' = I + Ma^2$ with $a = l/2$ we get,

$$I' = M\frac{l^2}{12} + M\left(\frac{l}{2}\right)^2 = \frac{Ml^2}{3}$$

We can check this independently since I is half the moment of inertia of a rod of mass $2M$ and length $2l$ about its midpoint,

$$I' = 2M.\frac{4l^2}{12} \times \frac{1}{2} = \frac{Ml^2}{3}$$ ◀

▶ ***Example 7.12*** What is the moment of inertia of a ring about a tangent to the circle of the ring?

Answer

The tangent to the ring in the plane of the ring is parallel to one of the diameters of the ring. The distance between these two parallel axes is R, the radius of the ring. Using the parallel axes theorem,

Fig. 7.32

$$I_{\text{tangent}} = I_{dia} + MR^2 = \frac{MR^2}{2} + MR^2 = \frac{3}{2}MR^2 .$$ ◀

7.11 KINEMATICS OF ROTATIONAL MOTION ABOUT A FIXED AXIS

We have already indicated the analogy between rotational motion and translational motion. For example, the angular velocity ω plays the same role in rotation as the linear velocity $\mathbf{v}$ in translation. We wish to take this analogy further. In doing so we shall restrict the discussion only to rotation about fixed axis. This case of motion involves only one degree of freedom, i.e., needs only one independent variable to describe the motion. This in translation corresponds to linear motion. This section is limited only to kinematics. We shall turn to dynamics in later sections.

We recall that for specifying the angular displacement of the rotating body we take any particle like P (Fig.7.33) of the body. Its angular displacement θ in the plane it moves is the angular displacement of the whole body; θ is measured from a fixed direction in the plane of motion of P, which we take to be the x' - axis, chosen parallel to the x-axis. Note, as shown, the axis of rotation is the z – axis and the plane of the motion of the particle is the x - y plane. Fig. 7.33 also shows θ_0, the angular displacement at $t = 0$.

We also recall that the angular velocity is the time rate of change of angular displacement, $\omega = d\theta/dt$. Note since the axis of rotation is

fixed, there is no need to treat angular velocity as a vector. Further, the angular acceleration, $\alpha = d\omega/dt$.

The kinematical quantities in rotational motion, angular displacement (θ), angular velocity (ω) and angular acceleration (α) respectively correspond to kinematic quantities in linear motion, displacement (x), velocity (v) and acceleration (a). We know the kinematical equations of linear motion with uniform (i.e. constant) acceleration:

$$v = v_0 + at \quad \text{(a)}$$

$$x = x_0 + v_0 t + \frac{1}{2}at^2 \quad \text{(b)}$$

$$v^2 = v_0^2 + 2ax \quad \text{(c)}$$

where x_0 = initial displacement and v_0= initial velocity. The word 'initial' refers to values of the quantities at $t = 0$

The corresponding kinematic equations for rotational motion with uniform angular acceleration are:

$$\omega = \omega_0 + \alpha t \quad (7.38)$$

$$\theta = \theta_0 + \omega_0 t + \frac{1}{2}\alpha t^2 \quad (7.39)$$

$$\text{and } \omega^2 = \omega_0^2 + 2\alpha(\theta - \theta_0) \quad (7.40)$$

where θ_0= initial angular displacement of the rotating body, and ω_0 = initial angular velocity of the body.

Fig.7.33 *Specifying the angular position of a rigid body.*

Example 7.13 Obtain Eq. (7.38) from first principles.

Answer The angular acceleration is uniform, hence

$$\frac{d\omega}{dt} = \alpha = constant \quad \text{(i)}$$

Integrating this equation,

$$\omega = \int \alpha\, dt + c$$

$$= \alpha t + c \quad \text{(as } \alpha \text{ is constant)}$$

At $t = 0$, $\omega = \omega_0$ (given)

From (i) we get at $t = 0$, $\omega = c = \omega_0$

Thus, $\omega = \alpha t + \omega_0$ as required.

With the definition of $\omega = d\theta/dt$ we may integrate Eq. (7.38) to get Eq. (7.39). This derivation and the derivation of Eq. (7.40) is left as an exercise.

Example 7.14 The angular speed of a motor wheel is increased from 1200 rpm to 3120 rpm in 16 seconds. (i) What is its angular acceleration, assuming the acceleration to be uniform? (ii) How many revolutions does the engine make during this time?

Answer

(i) We shall use $\omega = \omega_0 + \alpha t$

ω_0 = initial angular speed in rad/s

$= 2\pi \times$ angular speed in rev/s

$$= \frac{2\pi \times \text{angular speed in rev/min}}{60 \text{ s/min}}$$

$$= \frac{2\pi \times 1200}{60} \text{ rad/s}$$

$= 40\pi$ rad/s

Similarly ω= final angular speed in rad/s

$$= \frac{2\pi \times 3120}{60} \text{ rad/s}$$

$= 2\pi \times 52$ rad/s

$= 104\,\pi$ rad/s

$\therefore$ Angular acceleration

$$\alpha = \frac{\omega - \omega_0}{t} = 4\,\pi \text{ rad/s}^2$$

The angular acceleration of the engine = 4π rad/s^2

(ii) The angular displacement in time t is given by

$$\theta = \omega_0 t + \frac{1}{2}\alpha t^2$$

$$= (40\pi \times 16 + \frac{1}{2} \times 4\pi \times 16^2) \text{ rad}$$

$$= (640\pi + 512\pi) \text{ rad}$$

$$= 1152\pi \text{ rad}$$

Number of revolutions = $\frac{1152\pi}{2\pi} = 576$ ◀

7.12 DYNAMICS OF ROTATIONAL MOTION ABOUT A FIXED AXIS

Table 7.2 lists quantities associated with linear motion and their analogues in rotational motion. We have already compared kinematics of the two motions. Also, we know that in rotational motion moment of inertia and torque play the same role as mass and force respectively in linear motion. Given this we should be able to guess what the other analogues indicated in the table are. For example, we know that in linear motion, work done is given by $F\,dx$, in rotational motion about a fixed axis it should be $\tau d\theta$, since we already know the correspondence $dx \to d\theta$ and $F \to \tau$. It is, however, necessary that these correspondences are established on sound dynamical considerations. This is what we now turn to.

Before we begin, we note *a* **simplification that arises in the case of rotational motion about a fixed axis**. Since the axis is fixed, only those components of torques, which are along the direction of the fixed axis need to be considered in our discussion. Only these components can cause the body to rotate about the axis. A component of the torque perpendicular to the axis of rotation will tend to turn the axis from its position. We specifically assume that there will arise necessary forces of constraint to cancel the effect of the perpendicular components of the (external) torques, so that the fixed position of the axis will be maintained. The perpendicular components of the torques, therefore need not be taken into account. This means that for our calculation of torques on a rigid body:

(1) We need to consider only those forces that lie in planes perpendicular to the axis. Forces which are parallel to the axis will give torques perpendicular to the axis and need not be taken into account.

(2) We need to consider only those components of the position vectors which are perpendicular to the axis. Components of position vectors along the axis will result in torques perpendicular to the axis and need not be taken into account.

Work done by a torque

Fig. 7.34 Work done by a force $\mathbf{F}_1$ *acting on a particle of a body rotating about a fixed axis; the particle describes a circular path with centre C on the axis; arc* $P_1P'_1$ *(*ds_1*) gives the displacement of the particle.*

Figure 7.34 shows a cross-section of a rigid body rotating about a fixed axis, which is taken as the *z*-axis (perpendicular to the plane of the page; see Fig. 7.33). As said above we need to consider only those forces which lie in planes perpendicular to the axis. Let $\mathbf{F}_1$ be one such typical force acting as shown on a particle of the body at point P_1 with its line of action in a plane perpendicular to the axis. For convenience we call this to be the $x'-y'$ plane (coincident with the plane of the page). The particle at P_1 describes a circular path of radius r_1 with centre C on the axis; $CP_1 = r_1$.

In time Δt, the point moves to the position P_1'. The displacement of the particle $d\mathbf{s}_1$, therefore, has magnitude $ds_1 = r_1 d\theta$ and direction tangential at P_1 to the circular path as shown. Here $d\theta$ is the angular displacement of the particle, $d\theta = \angle P_1CP_1'$. The work done by the force on the particle is

$$dW_1 = \mathbf{F}_1 . d\mathbf{s}_1 = F_1 ds_1 \cos\phi_1 = F_1 (r_1\, d\theta) \sin\alpha_1$$

where ϕ_1 is the angle between $\mathbf{F}_1$ and the tangent

Table 7.2 Comparison of Translational and Rotational Motion

	Linear Motion	Rotational Motion about a Fixed Axis
1	Displacement x	Angular displacement θ
2	Velocity $v = dx/dt$	Angular velocity $\omega = d\theta/dt$
3	Acceleration $a = dv/dt$	Angular acceleration $\alpha = d\omega/dt$
4	Mass M	Moment of inertia I
5	Force $F = Ma$	Torque $\tau = I\alpha$
6	Work $dW = F\,ds$	Work $W = \tau\, d\theta$
7	Kinetic energy $K = Mv^2/2$	Kinetic energy $K = I\omega^2/2$
8	Power $P = F\,v$	Power $P = \tau\omega$
9	Linear momentum $p = Mv$	Angular momentum $L = I\omega$

at P_1, and α_1 is the angle between $\mathbf{F}_1$ and the radius vector $\mathbf{OP}_1$; $\phi_1 + \alpha_1 = 90°$.

The torque due to $\mathbf{F}_1$ about the origin is $\mathbf{OP}_1 \times \mathbf{F}_1$. Now $\mathbf{OP}_1 = \mathbf{OC} + \mathbf{OP}_1$. [Refer to Fig. 7.17(b).] Since **OC** is along the axis, the torque resulting from it is excluded from our consideration. The effective torque due to $\mathbf{F}_1$ is $\boldsymbol{\tau}_1 = \mathbf{CP} \times \mathbf{F}_1$; it is directed along the axis of rotation and has a magnitude $\tau_1 = r_1 F_1 \sin\alpha$, Therefore,

$$dW_1 = \tau_1 d\theta$$

If there are more than one forces acting on the body, the work done by all of them can be added to give the total work done on the body. Denoting the magnitudes of the torques due to the different forces as τ_1, τ_2, ... etc,

$$dW = (\tau_1 + \tau_2 + ...)d\theta$$

Remember, the forces giving rise to the torques act on different particles, but the angular displacement $d\theta$ is the same for all particles. Since all the torques considered are parallel to the fixed axis, the magnitude τ of the total torque is just the algebraic sum of the magnitudes of the torques, i.e., $\tau = \tau_1 + \tau_2 + ...$ We, therefore, have

$$dW = \tau d\theta \qquad (7.41)$$

This expression gives the work done by the total (external) torque τ which acts on the body rotating about a fixed axis. Its similarity with the corresponding expression

$$dW = F\,ds$$

for linear (translational) motion is obvious.

Dividing both sides of Eq. (7.41) by dt gives

$$P = \frac{dW}{dt} = \tau\frac{d\theta}{dt} = \tau\omega$$

or $P = \tau\omega$ (7.42)

This is the instantaneous power. Compare this expression for power in the case of rotational motion about a fixed axis with the expression for power in the case of linear motion,

$$P = Fv$$

In a perfectly rigid body there is no internal motion. The work done by external torques is therefore, not dissipated and goes on to increase the kinetic energy of the body. The rate at which work is done on the body is given by Eq. (7.42). This is to be equated to the rate at which kinetic energy increases. The rate of increase of kinetic energy is

$$\frac{d}{dt}\left(\frac{I\omega^2}{2}\right) = I\frac{(2\omega)}{2}\frac{d\omega}{dt}$$

We assume that the moment of inertia does not change with time. This means that the mass of the body does not change, the body remains rigid and also the axis does not change its position with respect to the body.

Since $\alpha = d\omega/dt$, we get

$$\frac{d}{dt}\left(\frac{I\omega^2}{2}\right) = I\,\omega\,\alpha$$

Equating rates of work done and of increase in kinetic energy,

$$\tau\omega = I\,\omega\,\alpha$$

$$\tau = I\alpha \qquad (7.43)$$

Eq. (7.43) is similar to Newton's second law for linear motion expressed symbolically as

$F = ma$

Just as force produces acceleration, torque produces angular acceleration in a body. The angular acceleration is directly proportional to the applied torque and is inversely proportional to the moment of inertia of the body. Eq.(7.43) can be called Newton's second law for rotation about a fixed axis.

Example 7.15 A cord of negligible mass is wound round the rim of a fly wheel of mass 20 kg and radius 20 cm. A steady pull of 25 N is applied on the cord as shown in Fig. 7.35. The flywheel is mounted on a horizontal axle with frictionless bearings.

(a) Compute the angular acceleration of the wheel.

(b) Find the work done by the pull, when 2m of the cord is unwound.

(c) Find also the kinetic energy of the wheel at this point. Assume that the wheel starts from rest.

(d) Compare answers to parts (b) and (c).

Answer

Fig. 7.35

(a) We use $I\alpha = \tau$

the torque $\tau = FR$

$= 25 \times 0.20$ Nm (as $R = 0.20$m)

$= 5.0$ Nm

I = M. I. of flywheel about its axis $= \dfrac{MR^2}{2}$

$= \dfrac{20.0 \times (0.2)^2}{2} = 0.4$ kg m^2

α = angular acceleration

$= 5.0$ N m/0.4 kg m^2 = 12.35 s^{-2}

(b) Work done by the pull unwinding 2m of the cord

= 25 N × 2m = 50 J

(c) Let ω be the final angular velocity. The kinetic energy gained = $\dfrac{1}{2}I\omega^2$,

since the wheel starts from rest. Now,

$\omega^2 = \omega_0^2 + 2\alpha\theta, \quad \omega_0 = 0$

The angular displacement θ = length of unwound string / radius of wheel

= 2m/0.2 m = 10 rad

$\omega^2 = 2 \times 12.5 \times 10.0 = 250\,(\text{rad/s})^2$

$\therefore$ *K.E.* gained $= \dfrac{1}{2} \times 0.4 \times 250 = 50\,\mathbf{J}$

(d) The answers are the same, i.e. the kinetic energy gained by the wheel = work done by the force. There is no loss of energy due to friction.

7.13 ANGULAR MOMENTUM IN CASE OF ROTATION ABOUT A FIXED AXIS

We have studied in section 7.7, the angular momentum of a system of particles. We already know from there that the time rate of total angular momentum of a system of particles about a point is equal to the total external torque on the system taken about the same point. When the total external torque is zero, the total angular momentum of the system is conserved.

We now wish to study the angular momentum in the special case of rotation about a fixed axis. The general expression for the total angular momentum of the system is

$$\mathbf{L} = \sum_{i=1}^{N} \mathbf{r}_i \times \mathbf{p}_i \qquad (7.25b)$$

We first consider the angular momentum of a typical particle of the rotating rigid body. We then sum up the contributions of individual particles to get **L** of the whole body.

For a typical particle $\mathbf{l} = \mathbf{r} \times \mathbf{p}$. As seen in the last section $\mathbf{r} = \mathbf{OP} = \mathbf{OC} + \mathbf{CP}$ [Fig. 7.17(b)].

With $\mathbf{p} = m\,\mathbf{v}$,

$$\mathbf{l} = (\mathbf{OC} \times m\mathbf{v}) + (\mathbf{CP} \times m\mathbf{v})$$

The magnitude of the linear velocity $\mathbf{v}$ of the particle at P is given by $v = \omega r_\perp$ where $r_\perp$ is the length of CP or the perpendicular distance of P from the axis of rotation. Further, $\mathbf{v}$ is tangential at P to the circle which the particle describes. Using the right-hand rule one can check that $\mathbf{CP} \times \mathbf{v}$ is parallel to the fixed axis. The unit vector along the fixed axis (chosen as the *z*-axis) is $\mathbf{k}$. Hence

$$\mathbf{CP} \times m\mathbf{v} = r_\perp (mv)\mathbf{k}$$

$$= mr_\perp^2 \omega \mathbf{k} \quad (\text{since } v = wr_\perp)$$

Similarly, we can check that $\mathbf{OC} \times \mathbf{v}$ is perpendicular to the fixed axis. Let us denote the part of $\mathbf{l}$ along the fixed axis (i.e. the *z*-axis) by $\mathbf{l}_z$, then

$$\mathbf{l}_z = \mathbf{CP} \times m\mathbf{v} = mr_\perp^2 \omega \mathbf{k}$$

and $\mathbf{l} = \mathbf{l}_z + \mathbf{OC} \times m\mathbf{v}$

We note that $\mathbf{l}_z$ is parallel to the fixed axis, but $\mathbf{l}$ is not. In general, for a particle, the angular momentum $\mathbf{l}$ is not along the axis of rotation, i.e. for a particle, $\mathbf{l}$ and ω are not necessarily parallel. Compare this with the corresponding fact in translation. For a particle, $\mathbf{p}$ and $\mathbf{v}$ are always parallel to each other.

For computing the total angular momentum of the whole rigid body, we add up the contribution of each particle of the body.

Thus $\quad \mathbf{L} = \sum \mathbf{l}_i = \sum l_{iz} + \sum \mathbf{OC}_i \times m_i \mathbf{v}_i$

We denote by $\mathbf{L}_\perp$ and $\mathbf{L}_z$ the components of $\mathbf{L}$ respectively perpendicular to the *z*-axis and along the *z*-axis;

$$\mathbf{L}_\perp = \sum \mathbf{OC}_i \times m_i \mathbf{v}_i \qquad (7.44a)$$

where m_i and $\mathbf{v}_i$ are respectively the mass and the velocity of the ith particle and C_i is the centre of the circle described by the particle;

and $\quad \mathbf{L}_z = \sum \mathbf{l}_{iz} = \left(\sum_i m_i r_i^2 \right) \omega \mathbf{k}$

or $\quad \mathbf{L}_z = I\omega \mathbf{k} \qquad (7.44b)$

The last step follows since the perpendicular distance of the ith particle from the axis is r_i; and by definition the moment of inertia of the body about the axis of rotation is $I = \sum m_i r_i^2$.

Note $\quad \mathbf{L} = \mathbf{L}_z + \mathbf{L}_\perp \qquad (7.44c)$

The rigid bodies which we have mainly considered in this chapter are symmetric about the axis of rotation, i.e. the axis of rotation is one of their symmetry axes. For such bodies, for a given $\mathbf{OC}_i$, for every particle which has a velocity $\mathbf{v}_i$, there is another particle of velocity $-\mathbf{v}_i$ located diametrically opposite on the circle with centre C_i described by the particle. Together such pairs will contribute zero to $\mathbf{L}_\perp$ and as a result for symmetric bodies $\mathbf{L}_\perp$ is zero, and hence

$$\mathbf{L} = \mathbf{L}_z = I\omega \mathbf{k} \qquad (7.44d)$$

For bodies, which are not symmetric about the axis of rotation, $\mathbf{L}$ is not equal to $\mathbf{L}_z$ and hence $\mathbf{L}$ does not lie along the axis of rotation.

Referring to table 7.1, can you tell in which cases $\mathbf{L} = \mathbf{L}_z$ will not apply?

Let us differentiate Eq. (7.44b). Since $\mathbf{k}$ is a fixed (constant) vector, we get

$$\frac{d}{dt}(\mathbf{L}_z) = \left(\frac{d}{dt}(I\omega) \right) \mathbf{k}$$

Now, Eq. (7.28b) states

$$\frac{d\mathbf{L}}{dt} = \boldsymbol{\tau}$$

As we have seen in the last section, only those components of the external torques which are along the axis of rotation, need to be taken into account, when we discuss rotation about a fixed axis. This means we can take $\boldsymbol{\tau} = \tau \mathbf{k}$. Since $\mathbf{L} = \mathbf{L}_z + \mathbf{L}_\perp$ and the direction of $\mathbf{L}_z$ (vector $\mathbf{k}$) is fixed, it follows that for rotation about a fixed axis,

$$\frac{d\mathbf{L}_z}{dt} = \tau \mathbf{k} \qquad (7.45a)$$

and $\quad \dfrac{d\mathbf{L}_\perp}{dt} = 0 \qquad (7.45b)$

Thus, for rotation about a fixed axis, the component of angular momentum perpendicular to the fixed axis is constant. As $\mathbf{L}_z = I\omega \mathbf{k}$, we get from Eq. (7.45a),

$$\frac{d}{dt}(I\omega) = \tau \qquad (7.45c)$$

If the moment of inertia I does not change with time,

$$\frac{d}{dt}(I\omega) = I\frac{d\omega}{dt} = I\alpha$$

and we get from Eq. (7.45c),

$$\tau = I\alpha \qquad (7.43)$$

We have already derived this equation using the work - kinetic energy route.

7.13.1 Conservation of angular momentum

We are now in a position to revisit the principle of conservation of angular momentum in the context of rotation about a fixed axis. From Eq. (7.45c), if the external torque is zero,

$$L_z = I\omega = \text{constant} \qquad (7.46)$$

For symmetric bodies, from Eq. (7.44d), L_z may be replaced by L .(L and L_z are respectively the magnitudes of $\mathbf{L}$ and $\mathbf{L}_z$.)

This then is the required form, for fixed axis rotation, of Eq. (7.29a), which expresses the general law of conservation of angular momentum of a system of particles. Eq. (7.46) applies to many situations that we come across in daily life. You may do this experiment with your friend. Sit on a swivel chair with your arms folded and feet not resting on, i.e., away from, the ground. Ask your friend to rotate the chair rapidly. While the chair is rotating with considerable angular speed stretch your arms horizontally. What happens? Your angular speed is reduced. If you bring back your arms closer to your body, the angular speed increases again. This is a situation where the principle of conservation of angular momentum is applicable. If friction in the rotational mechanism is neglected, there is no external torque about the axis of rotation of the chair and hence $I\omega$ is constant. Stretching the arms increases I about the axis of rotation, resulting in decreasing the angular speed ω. Bringing the arms closer to the body has the opposite effect.

A circus acrobat and a diver take advantage of this principle. Also, skaters and classical, Indian or western, dancers performing a pirouette on the toes of one foot display 'mastery' over this principle. Can you explain?

Fig 7.36 (a) A demonstration of conservation of angular momentum. A girl sits on a swivel chair and stretches her arms/ brings her arms closer to the body.

7.14 ROLLING MOTION

One of the most common motions observed in daily life is the rolling motion. All wheels used in transportation have rolling motion. For specificness we shall begin with the case of a disc, but the result will apply to any rolling body rolling on a level surface. We shall assume that the disc rolls without slipping. This means that at any instant of time the bottom of the disc which is in contact with the surface is at rest on the surface.

Fig 7.36 (b) An acrobat employing the principle of conservation of angular momentum in her performance.

We have remarked earlier that rolling motion is a combination of rotation and translation. We know that the translational motion of a system of particles is the motion of its centre of mass.

Fig. 7.37 The rolling motion (without slipping) of a disc on a level surface. Note at any instant, the point of contact P_0 of the disc with the surface is at rest; the centre of mass of the disc moves with velocity, v_{cm}. The disc rotates with angular velocity about its axis which passes through C; $v_{cm} = R\omega$, where R is the radius of the disc.

Let $\mathbf{v}_{cm}$ be the velocity of the centre of mass and therefore the translational velocity of the disc. Since the centre of mass of the rolling disc is at its geometric centre C (Fig. 7. 37), $\mathbf{v}_{cm}$ is the velocity of C. It is parallel to the level surface. The rotational motion of the disc is about its symmetry axis, which passes through C. Thus, the velocity of any point of the disc, like P_0, P_1 or P_2, consists of two parts, one is the translational velocity $\mathbf{v}_{cm}$ and the other is the linear velocity $\mathbf{v}_r$ on account of rotation. The magnitude of $\mathbf{v}_r$ is $v_r = r\omega$, where ω is the angular velocity of the rotation of the disc about the axis and r is the distance of the point from the axis (i.e. from C). The velocity $\mathbf{v}_r$ is directed perpendicular to the radius vector of the given point with respect to C. In Fig. 7.37, the velocity of the point P_2 ($\mathbf{v}_2$) and its components $\mathbf{v}_r$ and $\mathbf{v}_{cm}$ are shown; $\mathbf{v}_r$ here is perpendicular to CP_2. It is easy to show that $\mathbf{v}_z$ is perpendicular to the line P_OP_2. Therefore the line passing through P_O and parallel to ω is called the instantaneous axis of rotation.

At P_o, the linear velocity, $\mathbf{v}_r$, due to rotation is directed exactly opposite to the translational velocity $\mathbf{v}_{cm}$. Further the magnitude of $\mathbf{v}_r$ here is $R\omega$, where R is the radius of the disc. The condition that P_o is instantaneously at rest requires $v_{cm} = R\omega$ Thus for the disc the condition for rolling without slipping is

(7.47)

Incidentally, this means that the velocity of point P_1 at the top of the disc ($\mathbf{v}_1$) has a magnitude $v_{cm} + R\omega$ or $2\ v_{cm}$ and is directed parallel to the level surface. The condition (7.47) applies to all rolling bodies.

7.14.1 Kinetic Energy of Rolling Motion

Our next task will be to obtain an expression for the kinetic energy of a rolling body. The kinetic energy of a rolling body can be separated into kinetic energy of translation and kinetic energy of rotation. This is a special case of a general result for a system of particles, according to which the kinetic energy of a system of particles (K) can be separated into the kinetic energy of motion of the centre of mass (translation) ($MV^2/2$) and kinetic energy of rotational motion about the centre of mass of the system of particles (K'). Thus,

$$K = K' + MV^2/2 \qquad (7.48)$$

We assume this general result (see Exercise 7.31), and apply it to the case of rolling motion. In our notation, the kinetic energy of the centre of mass, i.e., the kinetic energy of translation, of the rolling body is $mv^2_{cm}/2$, where m is the mass of the body and v_{cm} is the centre of the mass velocity. Since the motion of the rolling body about the centre of mass is rotation, K' represents the kinetic energy of rotation of the body; $K' = I\omega^2/2$, where I is the moment of inertia about the appropriate axis, which is the symmetry axis of the rolling body. The kinetic energy of a rolling body, therefore, is given by

$$K = \frac{1}{2} I\omega^2 + \frac{1}{2} m v_{cm}^2 \qquad (7.49a)$$

Substituting $I = mk^2$ where k = the corresponding radius of gyration of the body and $v_{cm} = R\,\omega$, we get

$$K = \frac{1}{2}\frac{mk^2 v_{cm}^2}{R^2} + \frac{1}{2} m v_{cm}^2$$

$$\text{or} \quad K = \frac{1}{2} m v_{cm}^2 \left(1 + \frac{k^2}{R^2}\right) \qquad (7.49b)$$

Equation (7.49b) applies to any rolling body: a disc, a cylinder, a ring or a sphere.

▶ ***Example 7.16*** Three bodies, a ring, a solid cylinder and a solid sphere roll down the same inclined plane without slipping. They start from rest. The radii of the bodies are identical. Which of the bodies reaches the ground with maximum velocity?

Answer We assume conservation of energy of the rolling body, i.e. there is no loss of energy due to friction etc. The potential energy lost by the body in rolling down the inclined plane (= mgh) must, therefore, be equal to kinetic energy gained. (See Fig.7.38) Since the bodies start from rest the kinetic energy gained is equal to the final kinetic energy of the bodies. From Eq. (7.49b), $K = \frac{1}{2}mv^2\left(1+\frac{k^2}{R^2}\right)$, where v is the final velocity of (the centre of mass of) the body. Equating K and mgh,

Fig.7.38

$$mgh = \frac{1}{2}mv^2\left(1+\frac{k^2}{R^2}\right)$$

or $$v^2 = \left(\frac{2gh}{1+k^2/R^2}\right)$$

Note is independent of the mass of the rolling body;

For a ring, $k^2 = R^2$

$$v_{ring} = \sqrt{\frac{2gh}{1+1}},$$

$$= \sqrt{gh}$$

For a solid cylinder $k^2 = R^2/2$

$$v_{disc} = \sqrt{\frac{2gh}{1+1/2}}$$

$$= \sqrt{\frac{4gh}{3}}$$

For a solid sphere $k^2 = 2R^2/5$

$$v_{sphere} = \sqrt{\frac{2gh}{1+2/5}}$$

$$= \sqrt{\frac{10gh}{7}}$$

From the results obtained it is clear that among the three bodies the sphere has the greatest and the ring has the least velocity of the centre of mass at the bottom of the inclined plane.

Suppose the bodies have the same mass. Which body has the greatest rotational kinetic energy while reaching the bottom of the inclined plane? ◀

SUMMARY

1. Ideally, a rigid body is one for which the distances between different particles of the body do not change, even though there are forces on them.
2. A rigid body fixed at one point or along a line can have only rotational motion. A rigid body not fixed in some way can have either pure translation or a combination of translation and rotation.
3. In rotation about a fixed axis, every particle of the rigid body moves in a circle which lies in a plane perpendicular to the axis and has its centre on the axis. Every Point in the rotating rigid body has the same angular velocity at any instant of time.
4. In pure translation, every particle of the body moves with the same velocity at any instant of time.
5. Angular velocity is a vector. Its magnitude is $\omega = d\theta/dt$ and it is directed along the axis of rotation. For rotation about a fixed axis, this vector ω has a fixed direction.

6. The vector or cross product of two vector **a** and **b** is a vector written as $\mathbf{a} \times \mathbf{b}$. The magnitude of this vector is $ab\sin\theta$ and its direction is given by the right handed screw or the right hand rule.
7. The linear velocity of a particle of a rigid body rotating about a fixed axis is given by $\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}$, where **r** is the position vector of the particle with respect to an origin along the fixed axis. The relation applies even to more general rotation of a rigid body with one point fixed. In that case **r** is the position vector of the particle with respect to the fixed point taken as the origin.
8. The centre of mass of a system of particles is defined as the point whose position vector is

$$\mathbf{R} = \frac{\sum m_i \mathbf{r}_i}{M}$$

9. Velocity of the centre of mass of a system of particles is given by $\mathbf{V} = \mathbf{P}/M$, where **P** is the linear momentum of the system. The centre of mass moves as if all the mass of the system is concentrated at this point and all the external forces act at it. If the total external force on the system is zero, then the total linear momentum of the system is constant.

10. The angular momentum of a system of n particles about the origin is

$$\mathbf{L} = \sum_{i=1}^{n} \mathbf{r}_i \times \mathbf{p}_i$$

The torque or moment of force on a system of n particles about the origin is

$$\boldsymbol{\tau} = \sum_{1} \mathbf{r}_i \times \mathbf{F}_i$$

The force $\mathbf{F}_i$ acting on the ith particle includes the external as well as internal forces. Assuming Newton's third law and that forces between any two particles act along the line joining the particles, we can show $\boldsymbol{\tau}_{int} = \mathbf{0}$ and

$$\frac{d\mathbf{L}}{dt} = \boldsymbol{\tau}_{ext}$$

11. A rigid body is in mechanical equilibrium if
    (1) it is in translational equilibrium, i.e., the total external force on it is zero : $\sum \mathbf{F}_i = \mathbf{0}$, and
    (2) it is in rotational equilibrium, i.e. the total external torque on it is zero : $\sum \boldsymbol{\tau}_i = \sum \mathbf{r}_i \times \mathbf{F}_i = \mathbf{0}$.
12. The centre of gravity of an extended body is that point where the total gravitational torque on the body is zero.
13. The moment of intertia of a rigid body about an axis is defined by the formula $I = \sum m_i r_i^2$ where r_i is the perpendicular distance of the ith point of the body from the axis. The kinetic energy of rotation is $K = \frac{1}{2} I\omega^2$.
14. The theorem of parallel axes: $I'_z = I_z + Ma^2$, allows us to determine the moment of intertia of a rigid body about an axis as the sum of the moment of inertia of the body about a parallel axis through its centre of mass and the product of mass and square of the perpendicular distance between these two axes.

15. Rotation about a fixed axis is directly analogous to linear motion in respect of kinematics and dynamics.
16. The angular acceleration of a rigid body rotating about a fixed axis is given by $I\alpha = \tau$. If the external torque τ is zero, the component of angular momentum about the fixed axis $I\omega$ of such a rotating body is constant.
17. For rolling motion without slipping $v_{cm} = R\omega$ where v_{cm} is the velocity of translation (i.e. of the centre of mass), R is the radius and m is the mass of the body. The kinetic energy of such a rolling body is the sum of kinetic energies of translation and rotation:

$$K = \frac{1}{2} m\, v_{cm}^2 + \frac{1}{2} I\omega^2 .$$

Quantity	Symbols	Dimensions	Units	Remarks
Angular velocity	ω	$[T^{-1}]$	rad s^{-1}	$v = \omega \times r$
Angular momentum	L	$[ML^2 T^{-1}]$	J s	$L = r \times p$
Torque	τ	$[ML^2 T^{-2}]$	N m	$\tau = r \times F$
Moment of inertia	I	$[ML^2]$	kg m^2	$I = \sum m_i r_i^2$

POINTS TO PONDER

1. To determine the motion of the centre of mass of a system no knowledge of internal forces of the system is required. For this purpose we need to know only the external forces on the body.
2. Separating the motion of a system of particles as, i.e. the motion of the centre of mass transnational motion of the system and motion about (i.e. relative to) the centre of mass of the system is a useful technique in dynamics of a system of particles. One example of this technique is separating the kinetic energy of a system of particles K as the kinetic energy of the system about its centre of mass K′ and the kinetic energy of the centre of mass $MV^2/2$,

$$K = K' + MV^2/2$$

3. Newton's Second Law for finite sized bodies (or systems of particles) is based in Newton's Second Law and also Newton's Third Law for particles.
4. To establish that the time rate of change of the total angular momentum of a system of particles is the total external torque in the system, we need not only Newton's second law for particles, but also Newton's third law with the provision that the forces between any two particle act along the line joining the particles.
5. The vanishing of the total external force and the vanishing of the total external torque are independent conditions. We can have one without the other. In a couple, total external force is zero but total torque is non-zero.
6. The total torque on a system is independent of the origin if the total external force is zero.
7. The centre of gravity of a body coincides with its centre of mass only if the gravitational field does not vary from one part of the body to the other.
8. The angular momentum **L** and the angular velocity $\boldsymbol{\omega}$ are not necessarily parallel vectors. However, for the simpler situations discussed in this chapter when rotation is about a fixed axis which is an axis of symmetry of the rigid body, the relation $\mathbf{L} = I\omega$ holds good, where I is the moment of the inertia of the body about the rotation axis.

EXERCISES

7.1 Give the location of the centre of mass of a (i) sphere, (ii) cylinder, (iii) ring, and (iv) cube, each of uniform mass density. Does the centre of mass of a body necessarily lie inside the body ?

7.2 In the HC1 molecule, the separation between the nuclei of the two atoms is about 1.27 Å (1 Å = 10^{-10} m). Find the approximate location of the CM of the molecule, given that a chlorine atom is about 35.5 times as massive as a hydrogen atom and nearly all the mass of an atom is concentrated in its nucleus.

7.3 A child sits stationary at one end of a long trolley moving uniformly with a speed V on a smooth horizontal floor. If the child gets up and runs about on the trolley in any manner, what is the speed of the CM of the (trolley + child) system ?

7.4 Show that the area of the triangle contained between the vectors **a** and **b** is one half of the magnitude of **a** × **b**.

7.5 Show that **a**·(**b** × **c**) is equal in magnitude to the volume of the parallelepiped formed on the three vectors , **a**, **b** and **c**.

7.6 Find the components along the x, y, z axes of the angular momentum **l** of a particle, whose position vector is **r** with components x, y, z and momentum is **p** with components p_x, p_y and p_z. Show that if the particle moves only in the x-y plane the angular momentum has only a z-component.

7.7 Two particles, each of mass m and speed v, travel in opposite directions along parallel lines separated by a distance d. Show that the vector angular momentum of the two particle system is the same whatever be the point about which the angular momentum is taken.

7.8 A non-uniform bar of weight W is suspended at rest by two strings of negligible weight as shown in Fig.7.39. The angles made by the strings with the vertical are 36.9° and 53.1° respectively. The bar is 2 m long. Calculate the distance d of the centre of gravity of the bar from its left end.

Fig. 7.39

7.9 A car weighs 1800 kg. The distance between its front and back axles is 1.8 m. Its centre of gravity is 1.05 m behind the front axle. Determine the force exerted by the level ground on each front wheel and each back wheel.

7.10 (a) Find the moment of inertia of a sphere about a tangent to the sphere, given the moment of inertia of the sphere about any of its diameters to be $2MR^2/5$, where M is the mass of the sphere and R is the radius of the sphere.

(b) Given the moment of inertia of a disc of mass M and radius R about any of its diameters to be $MR^2/4$, find its moment of inertia about an axis normal to the disc and passing through a point on its edge.

7.11 Torques of equal magnitude are applied to a hollow cylinder and a solid sphere, both having the same mass and radius. The cylinder is free to rotate about its standard axis of symmetry, and the sphere is free to rotate about an axis passing through its centre. Which of the two will acquire a greater angular speed after a given time.

7.12 A solid cylinder of mass 20 kg rotates about its axis with angular speed 100 rad s^{-1}. The radius of the cylinder is 0.25 m. What is the kinetic energy associated with the rotation of the cylinder? What is the magnitude of angular momentum of the cylinder about its axis?

7.13 (a) A child stands at the centre of a turntable with his two arms outstretched. The turntable is set rotating with an angular speed of 40 rev/min. How much is the angular speed of the child if he folds his hands back and thereby reduces his moment of inertia to 2/5 times the initial value ? Assume that the turntable rotates without friction.

(b) Show that the child's new kinetic energy of rotation is more than the initial kinetic energy of rotation. How do you account for this increase in kinetic energy?

7.14 A rope of negligible mass is wound round a hollow cylinder of mass 3 kg and radius 40 cm. What is the angular acceleration of the cylinder if the rope is pulled with a force of 30 N ? What is the linear acceleration of the rope ? Assume that there is no slipping.

7.15 To maintain a rotor at a uniform angular speed or 200 rad s^{-1}, an engine needs to transmit a torque of 180 N m. What is the power required by the engine ? (Note: uniform angular velocity in the absence of friction implies zero torque. In practice, applied torque is needed to counter frictional torque). Assume that the engine is 100% efficient.

7.16 From a uniform disk of radius R, a circular hole of radius $R/2$ is cut out. The centre of the hole is at $R/2$ from the centre of the original disc. Locate the centre of gravity of the resulting flat body.

7.17 A metre stick is balanced on a knife edge at its centre. When two coins, each of mass 5 g are put one on top of the other at the 12.0 cm mark, the stick is found to be balanced at 45.0 cm. What is the mass of the metre stick?

7.18 A solid sphere rolls down two different inclined planes of the same heights but different angles of inclination. (a) Will it reach the bottom with the same speed in each case? (b) Will it take longer to roll down one plane than the other? (c) If so, which one and why?

7.19 A hoop of radius 2 m weighs 100 kg. It rolls along a horizontal floor so that its centre of mass has a speed of 20 cm/s. How much work has to be done to stop it?

7.20 The oxygen molecule has a mass of 5.30×10^{-26} kg and a moment of inertia of 1.94×10^{-46} kg m^2 about an axis through its centre perpendicular to the lines joining the two atoms. Suppose the mean speed of such a molecule in a gas is 500 m/s and that its kinetic energy of rotation is two thirds of its kinetic energy of translation. Find the average angular velocity of the molecule.

7.21 A solid cylinder rolls up an inclined plane of angle of inclination 30°. At the bottom of the inclined plane the centre of mass of the cylinder has a speed of 5 m/s.
(a) How far will the cylinder go up the plane?
(b) How long will it take to return to the bottom?

Additional Exercises

7.22 As shown in Fig.7.40, the two sides of a step ladder BA and CA are 1.6 m long and hinged at A. A rope DE, 0.5 m is tied half way up. A weight 40 kg is suspended from a point F, 1.2 m from B along the ladder BA. Assuming the floor to be frictionless and neglecting the weight of the ladder, find the tension in the rope and forces exerted by the floor on the ladder. (Take g = 9.8 m/s^2)
(Hint: Consider the equilibrium of each side of the ladder separately.)

Fig.7.40

7.23 A man stands on a rotating platform, with his arms stretched horizontally holding a 5 kg weight in each hand. The angular speed of the platform is 30 revolutions per minute. The man then brings his arms close to his body with the distance of each weight from the axis changing from 90cm to 20cm. The moment of inertia of the man together with the platform may be taken to be constant and equal to 7.6 kg m^2.
(a) What is his new angular speed? (Neglect friction.)
(b) Is kinetic energy conserved in the process? If not, from where does the change come about?

7.24 A bullet of mass 10 g and speed 500 m/s is fired into a door and gets embedded exactly at the centre of the door. The door is 1.0 m wide and weighs 12 kg. It is hinged at one end and rotates about a vertical axis practically without friction. Find the angular speed of the door just after the bullet embeds into it.
(Hint: The moment of inertia of the door about the vertical axis at one end is $ML^2/3$.)

7.25 Two discs of moments of inertia I_1 and I_2 about their respective axes (normal to the disc and passing through the centre), and rotating with angular speeds ω_1 and ω_2 are brought into contact face to face with their axes of rotation coincident. (a) What is the angular speed of the two-disc system? (b) Show that the kinetic energy of the combined system is less than the sum of the initial kinetic energies of the two discs. How do you account for this loss in energy? Take $\omega_1 \neq \omega_2$.

7.26 (a) Prove the theorem of perpendicular axes.

(Hint : Square of the distance of a point (x, y) in the x–y plane from an axis through the origin perpendicular to the plane is x^2+y^2).

(b) Prove the theorem of parallel axes.

(Hint : If the centre of mass is chosen to be the origin $\sum m_i \mathbf{r}_i = 0$).

7.27 Prove the result that the velocity v of translation of a rolling body (like a ring, disc, cylinder or sphere) at the bottom of an inclined plane of a height h is given by

$$v^2 = \frac{2gh}{\left(1 + k^2 / R^2\right)}$$

using dynamical consideration (i.e. by consideration of forces and torques). Note k is the radius of gyration of the body about its symmetry axis, and R is the radius of the body. The body starts from rest at the top of the plane.

7.28 A disc rotating about its axis with angular speed ω_0 is placed lightly (without any translational push) on a perfectly frictionless table. The radius of the disc is R. What are the linear velocities of the points A, B and C on the disc shown in Fig. 7.41? Will the disc roll in the direction indicated ?

Fig. 7.41

7.29 Explain why friction is necessary to make the disc in Fig. 7.41 roll in the direction indicated.

(a) Give the direction of frictional force at B, and the sense of frictional torque, before perfect rolling begins.

(b) What is the force of friction after perfect rolling begins ?

7.30 A solid disc and a ring, both of radius 10 cm are placed on a horizontal table simultaneously, with initial angular speed equal to 10 π rad s^{-1}. Which of the two will start to roll earlier ? The co-efficient of kinetic friction is μ_k = 0.2.

7.31 A cylinder of mass 10 kg and radius 15 cm is rolling perfectly on a plane of inclination 30°. The co-efficient of static friction μ_S = 0.25.

(a) How much is the force of friction acting on the cylinder ?

(b) What is the work done against friction during rolling ?

(c) If the inclination θ of the plane is increased, at what value of θ does the cylinder begin to skid, and not roll perfectly ?

$\mathbf{L} = \mathbf{L}' + \mathbf{R} \times M\mathbf{V}$

7.32 Read each statement below carefully, and state, with reasons, if it is true or false;

(a) During rolling, the force of friction acts in the same direction as the direction of motion of the CM of the body.

(b) The instantaneous speed of the point of contact during rolling is zero.

(c) The instantaneous acceleration of the point of contact during rolling is zero.

(d) For perfect rolling motion, work done against friction is zero.

(e) A wheel moving down a perfectly *frictionless* inclined plane will undergo slipping (not rolling) motion.

7.33 Separation of Motion of a system of particles into motion of the centre of mass and motion about the centre of mass :

(a) Show $\mathbf{p} = \mathbf{p}'_i + m_i\mathbf{V}$

where $\mathbf{p}_i$ is the momentum of the *i*th particle (of mass m_i) and $\mathbf{p}'_i = m_i\mathbf{v}'_i$. Note $\mathbf{v}'_i$ is the velocity of the *i*th particle relative to the centre of mass.

Also, prove using the definition of the centre of mass $\sum \mathbf{p}'_i = 0$

(b) Show $K = K' + ½MV^2$

where K is the total kinetic energy of the system of particles, K' is the total kinetic energy of the system when the particle velocities are taken with respect to the centre of mass and $MV^2/2$ is the kinetic energy of the translation of the system as a whole (i.e. of the centre of mass motion of the system). The result has been used in Sec. 7.14.

(c) Show

where $\mathbf{L}' = \sum \mathbf{r}'_i \times \mathbf{p}'_i$ is the angular momentum of the system about the centre of mass with velocities taken relative to the centre of mass. Remember $\mathbf{r}'_i = \mathbf{r}_i - \mathbf{R}$; rest of the notation is the

standard notation used in the chapter. Note $\mathbf{L}'$ and $M\mathbf{R}\times\mathbf{V}$ can be said to be angular momenta, respectively, about and of the centre of mass of the system of particles.

(d) Show $\frac{d\mathbf{L}'}{dt}=\sum \mathbf{r}_i'\times\frac{d\mathbf{p}'}{dt}$

Further, show that

$$\frac{d\mathbf{L}'}{dt}=\boldsymbol{\tau}_{ext}'$$

where $\boldsymbol{\tau}_{ext}'$ is the sum of all external torques acting on the system about the centre of mass.

(Hint : Use the definition of centre of mass and Newton's Third Law. Assume the internal forces between any two particles act along the line joining the particles.)

CHAPTER EIGHT

GRAVITATION

8.1 Introduction
8.2 Kepler's laws
8.3 Universal law of gravitation
8.4 The gravitational constant
8.5 Acceleration due to gravity of the earth
8.6 Acceleration due to gravity below and above the surface of earth
8.7 Gravitational potential energy
8.8 Escape speed
8.9 Earth satellites
8.10 Energy of an orbiting satellite
8.11 Geostationary and polar satellites
8.12 Weightlessness
Summary
Points to ponder
Exercises
Additional exercises

8.1 INTRODUCTION

Early in our lives, we become aware of the tendency of all material objects to be attracted towards the earth. Anything thrown up falls down towards the earth, going uphill is lot more tiring than going downhill, raindrops from the clouds above fall towards the earth and there are many other such phenomena. Historically it was the Italian Physicist Galileo (1564-1642) who recognised the fact that all bodies, irrespective of their masses, are accelerated towards the earth with a constant acceleration. It is said that he made a public demonstration of this fact. To find the truth, he certainly did experiments with bodies rolling down inclined planes and arrived at a value of the acceleration due to gravity which is close to the more accurate value obtained later.

A seemingly unrelated phenomenon, observation of stars, planets and their motion has been the subject of attention in many countries since the earliest of times. Observations since early times recognised stars which appeared in the sky with positions unchanged year after year. The more interesting objects are the planets which seem to have regular motions against the background of stars. The earliest recorded model for planetary motions proposed by Ptolemy about 2000 years ago was a 'geocentric' model in which all celestial objects, stars, the sun and the planets, all revolved around the earth. The only motion that was thought to be possible for celestial objects was motion in a circle. Complicated schemes of motion were put forward by Ptolemy in order to describe the observed motion of the planets. The planets were described as moving in circles with the center of the circles themselves moving in larger circles. Similar theories were also advanced by Indian astronomers some 400 years later. However a more elegant model in which the Sun was the center around which the planets revolved – the 'heliocentric' model – was already mentioned by Aryabhatta (5th century A.D.) in his treatise. A thousand years later, a Polish monk named Nicolas

Copernicus (1473-1543) proposed a definitive model in which the planets moved in circles around a fixed central sun. His theory was discredited by the church, but notable amongst its supporters was Galileo who had to face prosecution from the state for his beliefs.

It was around the same time as Galileo, a nobleman called Tycho Brahe (1546-1601) hailing from Denmark, spent his entire lifetime recording observations of the planets with the naked eye. His compiled data were analysed later by his assistant Johannes Kepler (1571-1640). He could extract from the data three elegant laws that now go by the name of Kepler's laws. These laws were known to Newton and enabled him to make a great scientific leap in proposing his universal law of gravitation.

8.2 KEPLER'S LAWS

The three laws of Kepler can be stated as follows:

1. Law of orbits : All planets move in elliptical orbits with the Sun situated at one of the foci of the ellipse (Fig. 8.1a). This law was a deviation from the Copernican model which allowed only circular orbits. The ellipse, of which the circle is a special case, is a closed curve which can be drawn very simply as follows.

Fig. 8.1(a) An ellipse traced out by a planet around the sun. The closest point is P and the farthest point is A, P is called the perihelion and A the aphelion. The semimajor axis is half the distance AP.

Fig. 8.1(b) Drawing an ellipse. A string has its ends fixed at F_1 and F_2. The tip of a pencil holds the string taut and is moved around.

Select two points F_1 and F_2. Take a length of a string and fix its ends at F_1 and F_2 by pins. With the tip of a pencil stretch the string taut and then draw a curve by moving the pencil keeping the string taut throughout.(Fig. 8.1(b)) The closed curve you get is called an ellipse. Clearly for any point T on the ellipse, the sum of the distances from F_1 and F_2 is a constant. F_1, F_2 are called the focii. Join the points F_1 and F_2 and extend the line to intersect the ellipse at points P and A as shown in Fig. 8.1(b). The midpoint of the line PA is the centre of the ellipse O and the length PO = AO is called the semi-major axis of the ellipse. For a circle, the two focii merge into one and the semi-major axis becomes the radius of the circle.

2. Law of areas : The line that joins any planet to the sun sweeps equal areas in equal intervals of time (Fig. 8.2). This law comes from the observations that planets appear to move slower when they are farther from the sun than when they are nearer.

Fig. 8.2 The planet P moves around the sun in an elliptical orbit. The shaded area is the area ΔA swept out in a small interval of time Δt.

3. Law of periods : The square of the time period of revolution of a planet is proportional to the cube of the semi-major axis of the ellipse traced out by the planet.

The table below gives the approximate time periods of revolution of nine planets around the sun along with values of their semi-major axes.

Table 1 **Data from measurement of planetary motions given below confirm Kepler's Law of Periods**

a ≡ Semi-major axis in units of 10^{10} m.
T ≡ Time period of revolution of the planet in years(y).
Q ≡ The quotient (T^2/a^3) in units of 10^{-34} y^2 m^{-3}.

Planet	a	T	Q
Mercury	5.79	0.24	2.95
Venus	10.8	0.615	3.00
Earth	15.0	1	2.96
Mars	22.8	1.88	2.98
Jupiter	77.8	11.9	3.01
Saturn	143	29.5	2.98
Uranus	287	84	2.98
Neptune	450	165	2.99
Pluto	590	248	2.99

The law of areas can be understood as a consequence of conservation of angular momentum whch is valid for any central force . A central force is such that the force on the planet is along the vector joining the sun and the planet. Let the sun be at the origin and let the position and momentum of the planet be denoted by **r** and **p** respectively. Then the area swept out by the planet of mass m in time interval Δt is (Fig. 8.2) $\Delta\mathbf{A}$ given by

$$\Delta\mathbf{A} = \tfrac{1}{2}\ (\mathbf{r} \times \mathbf{v}\Delta t) \qquad (8.1)$$

Hence

$$\Delta\mathbf{A}/\Delta t = \tfrac{1}{2}\ (\mathbf{r} \times \mathbf{p})/m, \text{ (since } \mathbf{v} = \mathbf{p}/m)$$
$$= \mathbf{L} / (2\ m) \qquad (8.2)$$

where **v** is the velocity, **L** is the angular momentum equal to (**r** × **p**). For a central force, which is directed along **r**, **L** is a constant as the planet goes around. Hence, $\Delta\mathbf{A}/\Delta t$ is a constant according to the last equation. This is the law of areas. Gravitation is a central force and hence the law of areas follows.

Johannes Kepler (1571–1630) was a scientist of German origin. He formulated the three laws of planetary motion based on the painstaking observations of Tycho Brahe and coworkers. Kepler himself was an assistant to Brahe and it took him sixteen long years to arrive at the three planetary laws. He is also known as the founder of geometrical optics, being the first to describe what happens to light after it enters a telescope.

▶ ***Example 8.1*** Let the speed of the planet at the perihelion *P* in Fig. 8.1(a) be v_P and the Sun-planet distance SP be r_P. Relate $\{r_P, v_P\}$ to the corresponding quantities at the aphelion $\{r_A\ v_A\}$. Will the planet take equal times to traverse *BAC* and *CPB* ?

Answer The magnitude of the angular momentum at *P* is $L_p = m_p r_p v_p$, since inspection tells us that $\mathbf{r}_p$ and $\mathbf{v}_p$ are mutually perpendicular. Similarly, $L_A = m_p r_A v_A$. From angular momentum conservation

$$m_p r_p v_p = m_p r_A v_A$$

or $$\frac{v_p}{v_A} = \frac{r_A}{r_p}$$ ◀

Since $r_A > r_p$, $v_p > v_A$.

The area *SBAC* bounded by the ellipse and the radius vectors *SB* and *SC* is larger than SBPC in Fig. 8.1. From Kepler's second law, equal areas are swept in equal times. Hence the planet will take a longer time to traverse *BAC* than *CPB*.

8.3 UNIVERSAL LAW OF GRAVITATION

Legend has it that observing an apple falling from a tree, Newton was inspired to arrive at an universal law of gravitation that led to an explanation of terrestrial gravitation as well as of Kepler's laws. Newton's reasoning was that the moon revolving in an orbit of radius R_m was subject to a centripetal acceleration due to earth's gravity of magnitude

$$a_m = \frac{V^2}{R_m} = \frac{4\pi^2 R_m}{T^2} \qquad (8.3)$$

where *V* is the speed of the moon related to the time period *T* by the relation $V = 2\pi R_m / T$. The time period *T* is about 27.3 days and R_m was already known then to be about 3.84×10^8m. If we substitute these numbers in equation (8.3), we get a value of a_m much smaller than the value of acceleration due to gravity g on the surface of the earth, arising also due to earth's gravitational attraction.

Central Forces

We know the time rate of change of the angular momentum of a single particle about the origin is

$$\frac{d\mathbf{l}}{dt} = \mathbf{r} \times \mathbf{F}$$

The angular momentum of the particle is conserved, if the torque $\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F}$ due to the force **F** on it vanishes. This happens either when **F** is zero or when **F** is along **r**. We are interested in forces which satisfy the latter condition. Central forces satisfy this condition.

A 'central' force is always directed towards or away from a fixed point, i.e., along the position vector of the point of application of the force with respect to the fixed point. (See Figure below.) Further, the magnitude of a central force F depends on r, the distance of the point of application of the force from the fixed point; $F = F(r)$.

In the motion under a central force the angular momentum is always conserved. Two important results follow from this:

(1) The motion of a particle under the central force is always confined to a plane.

(2) The position vector of the particle with respect to the centre of the force (i.e. the fixed point) has a constant areal velocity. In other words the position vector sweeps out equal areas in equal times as the particle moves under the influence of the central force.

Try to prove both these results. You may need to know that the areal velocity is given by : $dA/dt = ½\, r\, \upsilon \sin \alpha$.

An immediate application of the above discussion can be made to the motion of a planet under the gravitational force of the sun. For convenience the sun may be taken to be so heavy that it is at rest. The gravitational force of the sun on the planet is directed towards the sun. This force also satisfies the requirement $F = F(r)$, since $F = G\, m_1 m_2/r^2$ where m_1 and m_2 are respectively the masses of the planet and the sun and G is the universal constant of gravitation. The two results (1) and (2) described above, therefore, apply to the motion of the planet. In fact, the result (2) is the well-known second law of Kepler.

Tr is the trejectory of the particle under the central force. At a position P, the force is directed along **OP**, *O is the centre of the force taken as the origin. In time Δt, the particle moves from* P *to P′ , arc PP′ = Δs = υ Δt. The tangent PQ at P to the trajectory gives the direction of the velocity at P. The area swept in Δt is the area of sector POP′ ≈* $(r \sin \alpha)$ *PP′ / 2 = (r υ sin a) Δt/2.)*

This clearly shows that the force due to earth's gravity decreases with distance. If one assumes that the gravitational force due to the earth decreases in proportion to the inverse square of the distance from the center of the earth, we will have $a_m \alpha R_m^{-2}$; $g \alpha R_E^{-2}$ and we get

$$\frac{g}{a_m} = \frac{R_m^2}{R_E^2} \; ; \; 3600 \qquad (8.4)$$

in agreement with a value of g ; 9.8 m s^{-2} and the value of a_m from Eq. (8.3). These observations led Newton to propose the following Universal Law of Gravitation :

Every body in the universe attracts every other body with a force which is directly proportional to the product of their masses and inversely proportional to the square of the distance between them.

The quotation is essentially from Newton's famous treatise called 'Mathematical Principles of Natural Philosophy' (Principia for short).

Stated Mathematically, Newton's gravitation law reads : The force **F** on a point mass m_2 due to another point mass m_1 has the magnitude

$$|\mathbf{F}| = G \frac{m_1 \; m_2}{r^2} \qquad (8.5)$$

Equation (8.5) can be expressed in vector form as

$$\mathbf{F} = G \frac{m_1 \; m_2}{r^2} \left(-\hat{\mathbf{r}}\right) = -G \frac{m_1 \; m_2}{r^2} \hat{\mathbf{r}}$$

$$= -G \frac{m_1 \; m_2}{|\mathbf{r}|^3} \hat{\mathbf{r}}$$

where G is the universal gravitational constant, $\hat{\mathbf{r}}$ is the unit vector from m_1 to m_2 and $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$ as shown in Fig. 8.3.

Fig. 8.3 Gravitational force on m_1 due to m_2 is along **r** *where the vector* **r** *is* ($\mathbf{r}_2 - \mathbf{r}_1$).

The gravitational force is attractive, i.e., the force **F** is along – **r**. The force on point mass m_1 due to m_2 is of course – **F** by Newton's third law. Thus, the gravitational force $\mathbf{F}_{12}$ on the body 1 due to 2 and $\mathbf{F}_{21}$ on the body 2 due to 1 are related as $\mathbf{F}_{12} = -\mathbf{F}_{21}$.

Before we can apply Eq. (8.5) to objects under consideration, we have to be careful since the law refers to **point** masses whereas we deal with extended objects which have finite size. . If we have a collection of point masses,the force on any one of them is the vector sum of the gravitational forces exerted by the other point masses as shown in Fig 8.4.

Fig. 8.4 Gravitational force on point mass m_1 is the vector sum of the gravitational forces exerted by m_2, m_3 and m_4.

The total force on m_1 is

$$\mathbf{F}_1 = \frac{Gm_2 \, m_1}{r_{21}^2} \hat{\mathbf{r}}_{21} + \frac{Gm_3 \, m_1}{r_{31}^2} \hat{\mathbf{r}}_{31} + \frac{Gm_4 \, m_1}{r_{41}^2} \hat{\mathbf{r}}_{41}$$

Example 8.2 Three equal masses of m kg each are fixed at the vertices of an equilateral triangle ABC.
(a) What is the force acting on a mass 2*m* placed at the centroid G of the triangle?
(b) What is the force if the mass at the vertex A is doubled ?
Take AG = BG = CG = 1m (see Fig. 8.5)

Fig. 8.5 Three equal masses are placed at the three vertices of the Δ ABC. A mass 2m is placed at the centroid G.

Answer (a) The angle between GC and the positive x-axis is 30° and so is the angle between GB and the negative x-axis. The individual forces in vector notation are

$$\mathbf{F}_{GA} = \frac{Gm(2m)}{1}\hat{\mathbf{j}}$$

$$\mathbf{F}_{GB} = \frac{Gm(2m)}{1}\left(-\hat{\mathbf{i}}\cos 30^o - \hat{\mathbf{j}}\sin 30^o\right)$$

$$\mathbf{F}_{GC} = \frac{Gm(2m)}{1}\left(+\hat{\mathbf{i}}\cos 30^o - \hat{\mathbf{j}}\sin 30^o\right)$$

From the principle of superposition and the law of vector addition, the resultant gravitational force $\mathbf{F}_R$ on $(2m)$ is

$$\mathbf{F}_R = \mathbf{F}_{GA} + \mathbf{F}_{GB} + \mathbf{F}_{GC}$$

$$\mathbf{F}_R = 2Gm^2\,\hat{\mathbf{j}} + 2Gm^2\left(-\hat{\mathbf{i}}\cos 30^o - \hat{\mathbf{j}}\sin 30^o\right)$$

$$+ 2Gm^2\left(\hat{\mathbf{i}}\cos 30^o - \hat{\mathbf{j}}\sin 30^o\right) = 0$$

Alternatively, one expects on the basis of symmetry that the resultant force ought to be zero.

(b) By symmetry the x-component of the force cancels out. The y-component survives.

$$\mathbf{F}_R = 4Gm^2\hat{\mathbf{j}} - 2Gm^2\hat{\mathbf{j}} = 2Gm^2\hat{\mathbf{j}}$$ ◀

For the gravitational force between an extended object (like the earth) and a point mass, Eq. (8.5) is not directly applicable. Each point mass in the extended object will exert a force on the given point mass and these force will not all be in the same direction. We have to add up these forces vectorially for all the point masses in the extended object to get the total force. This is easily done using calculus. For two special cases, a simple law results when you do that :

(1) **The force of attraction between a hollow spherical shell of uniform density and a point mass situated outside is just as if the entire mass of the shell is concentrated at the centre of the shell.** Qualitatively this can be understood as follows: Gravitational forces caused by the various regions of the shell have components along the line joining the point mass to the centre as well as along a direction prependicular to this line. The components prependicular to this line cancel out when summing over all regions of the shell leaving only a resultant force along the line joining the point to the centre. The magnitude of this force works out to be as stated above.

Newton's Principia

Kepler had formulated his third law by 1619. The announcement of the underlying universal law of gravitation came about seventy years later with the publication in 1687 of Newton's masterpiece **Philosophiae Naturalis Principia Mathematica**, often simply called the **Principia**.

Around 1685, Edmund Halley (after whom the famous Halley's comet is named), came to visit Newton at Cambridge and asked him about the nature of the trajectory of a body moving under the influence of an inverse square law. Without hesitation Newton replied that it had to be an ellipse, and further that he had worked it out long ago around 1665 when he was forced to retire to his farm house from Cambridge on account of a plague outbreak. Unfortunately, Newton had lost his papers. Halley prevailed upon Newton to produce his work in book form and agreed to bear the cost of publication. Newton accomplished this feat in eighteen months of superhuman effort. The **Principia** is a singular scientific masterpiece and in the words of Lagrange it is "the greatest production of the human mind." The Indian born astrophysicist and Nobel laureate S. Chandrasekhar spent ten years writing a treatise on the **Principia**. His book, Newton's **Principia for the Common Reader** brings into sharp focus the beauty, clarity and breath taking economy of Newton's methods.

(2) **The force of attraction due to a hollow spherical shell of uniform density, on a point mass situated inside it is zero.** Qualitatively, we can again understand this result. Various regions of the spherical shell attract the point mass inside it in various directions. These forces cancel each other completely.

8.4 THE GRAVITATIONAL CONSTANT

The value of the gravitational constant G entering the Universal law of gravitation can be determined experimentally and this was first done by English scientist Henry Cavendish in 1798. The apparatus used by him is schematically shown in figure.8.6

Fig. 8.6 Schematic drawing of Cavendish's experiment. S_1 and S_2 are large spheres which are kept on either side (shown shades) of the masses at A and B. When the big spheres are taken to the other side of the masses (shown by dotted circles), the bar AB rotates a little since the torque reverses direction. The angle of rotation can be measured experimentally.

The bar AB has two small lead spheres attached at its ends. The bar is suspended from a rigid support by a fine wire. Two large lead spheres are brought close to the small ones but on opposite sides as shown. The big spheres attract the nearby small ones by equal and opposite force as shown. There is no net force on the bar but only a torque which is clearly equal to F times the length of the bar,where F is the force of attraction between a big sphere and its neighbouring small sphere. Due to this torque, the suspended wire gets twisted till such time as the restoring torque of the wire equals the gravitational torque . If θ is the angle of twist of the suspended wire, the restoring torque is proportional to θ, equal to $\tau\theta$. Where τ is the restoring couple per unit angle of twist. τ can be measured independently e.g. by applying a known torque and measuring the angle of twist. The gravitational force between the spherical balls is the same as if their masses are concentrated at their centers. Thus if d is the separation between the centers of the big and its neighbouring small ball, M and m their masses, the gravitational force between the big sphere and its neighouring small ball is.

$$F = G\frac{Mm}{d^2} \tag{8.6}$$

If L is the length of the bar AB , then the torque arising out of F is F multiplied by L. At equilibrium, this is equal to the restoring torque and hence

$$G\frac{Mm}{d^2}L = \tau\,\theta \tag{8.7}$$

Observation of θ thus enables one to calculate G from this equation.

Since Cavendish's experiment, the measurement of G has been refined and the currently accepted value is

$$G = 6.67\times10^{-11}\ \text{N m}^2/\text{kg}^2 \tag{8.8}$$

8.5 ACCELERATION DUE TO GRAVITY OF THE EARTH

The earth can be imagined to be a sphere made of a large number of concentric spherical shells with the smallest one at the centre and the largest one at its surface. A point outside the earth is obviously outside all the shells. Thus, all the shells exert a gravitational force at the point outside just as if their masses are concentrated at their common centre according to the result stated in the last section. The total mass of all the shells combined is just the mass of the earth. Hence, at a point outside the earth, the gravitational force is just as if its entire mass of the earth is concentrated at its center.

For a point inside the earth, the situation is different. This is illustrated in Fig. 8.7.

Fig. 8.7 The mass m is in a mine located at a depth d below the surface of the Earth of mass M_E and radius R_E. We treat the Earth to be spherically symmetric.

Again consider the earth to be made up of concentric shells as before and a point mass m situated at a distance r from the centre. The point P lies outside the sphere of radius r. For the shells of radius greater than r, the point P lies inside. Hence according to result stated in the last section, they exert no gravitational force on mass m kept at P. The shells with radius $\leq$ r make up a sphere of radius r for which the point P lies on the surface. This smaller sphere therefore exerts a force on a mass m at P as if its mass m_r is concentrated at the center. Thus the force on the mass m at P has a magnitude

$$F = \frac{Gm\,(m_r)}{r^2} \tag{8.9}$$

We assume that the entire earth is of uniform density and hence its mass is $M_E = \frac{4\pi}{3} R_E^3 \rho$ where M_E is the mass of the earth R_E is its radius and ρ is the density. On the other hand the mass of the sphere M_r of radius r is $\frac{4\pi}{3}\rho r^3$ and hence

$$F = Gm\left(\frac{4\pi}{3}\rho\right)\frac{r^3}{r^2} = Gm\left(\frac{M_E}{R_E^3}\right)\frac{r^3}{r^2}$$

$$= \frac{G\,m\,M_E}{R_E{}^3}\,r \tag{8.10}$$

If the mass m is situated on the surface of earth, then r = R_E and the gravitational force on it is, from Eq. (8.10)

$$F = G\,\frac{M_E m}{R_E^2} \tag{8.11}$$

The acceleration experienced by the mass m, which is usually denoted by the symbol g is related to F by Newton's 2nd law by relation F = mg. Thus

$$g = \frac{F}{m} = \frac{GM_E}{R_E^2} \tag{8.12}$$

Acceleration g is readily measurable. R_E is a known quantity. The measurement of G by Cavendish's experiment (or otherwise), combined with knowledge of g and R_E enables one to estimate M_E from Eq. (8.12). This is the reason why there is a popular statement regarding Cavendish : "Cavendish weighed the earth".

8.6 ACCELERATION DUE TO GRAVITY BELOW AND ABOVE THE SURFACE OF EARTH

Consider a point mass m at a height h above the surface of the earth as shown in Fig. 8.8(a). The radius of the earth is denoted by R_E. Since this point is outside the earth,

(a)

Fig. 8.8 (a) g at a height h above the surface of the earth.

its distance from the centre of the earth is $(R_E + h)$. If $F(h)$ denoted the magnitude of the force on the point mass m, we get from Eq. (8.5) :

$$F(h) = \frac{GM_E m}{(R_E + h)^2} \tag{8.13}$$

The acceleration experienced by the point mass is $F(h)/m \equiv g(h)$ and we get

$$g(h) = \frac{F(h)}{m} = \frac{GM_E}{(R_E + h)^2} \quad . \qquad (8.14)$$

This is clearly less than the value of g on the surface of earth : $g = \frac{GM_E}{R_E^2}$. For $h << R_E$, we can expand the RHS of Eq. (8.14) :

$$g(h) = \frac{GM}{R_E^2(1 + h/R_E)^2} = g(1 + h/R_E)^{-2}$$

For $\frac{h}{R_E} << 1$, using binomial expression,

$$g(h) \cong g\left(1 - \frac{2h}{R_E}\right) \quad . \qquad (8.15)$$

Equation (8.15) thus tells us that for small heights h above the value of g decreases by a factor $(1 - 2h/R_E)$.

Now, consider a point mass m at a depth d below the surface of the earth (Fig. 8.8(b)), so that its distance from the center of the earth is $(R_E - d)$ as shown in the figure. The earth can be thought of as being composed of a smaller sphere of radius $(R_E - d)$ and a spherical shell of thickness d. The force on m due to the outer shell of thickness d is zero because the result quoted in the previous section. As far as the smaller sphere of radius $(R_E - d)$ is concerned, the point mass is outside it and hence according to the result quoted earlier, the force due to this smaller sphere is just as if the entire mass of the smaller sphere is concentrated at the centre. If M_s is the mass of the smaller sphere, then,

$$M_s / M_E = (R_E - d)^3 / R_E^3 \qquad (8.16)$$

Since mass of a sphere is proportional to be cube of its radius.

(b)

Fig. 8.8 (b) g at a depth d. In this case only the smaller sphere of radius $(R_E - d)$ contributes to g.

Thus the force on the point mass is

$$F(d) = G M_s m / (R_E - d)^2 \qquad (8.17)$$

Substituting for M_s from above , we get

$$F(d) = G M_E m (R_E - d) / R_E^3 \qquad (8.18)$$

and hence the acceleration due to gravity at a depth d,

$$g(d) = \frac{F(d)}{m} \text{ is}$$

$$g(d) = \frac{F(d)}{m} = \frac{GM_E}{R_E^3}(R_E - d)$$

$$= g\frac{R_E - d}{R_E} = g(1 - d/R_E) \qquad (8.19)$$

Thus, as we go down below earth's surface, the acceleration due gravity decreases by a factor $(1 - d/R_E)$. The remarkable thing about acceleration due to earth's gravity is that it is maximum on its surface decreasing whether you go up or down.

8.7 GRAVITATIONAL POTENTIAL ENERGY

We had discussed earlier the notion of potential energy as being the energy stored in the body at its given position. If the position of the particle changes on account of forces acting on it, then the change in its potential energy is just the amount of work done on the body by the force. As we had discussed earlier, forces for which the work done is independent of the path are the conservative forces.

The force of gravity is a conservative force and we can calculate the potential energy of a body arising out of this force, called the gravitational potential energy. Consider points close to the surface of earth, at distances from the surface much smaller than the radius of the earth. In such cases, the force of gravity is practically a constant equal to mg, directed towards the center of the earth. If we consider a point at a height h_1 from the surface of the earth and another point vertically above it at a height h_2 from the surface, the work done in lifting the particle of mass m from the first to the second position is denoted by W_{12}

$$W_{12} = \text{Force} \times \text{displacement}$$
$$= mg(h_2 - h_1) \qquad (8.20)$$

If we associate a potential energy $W(h)$ at a point at a height h above the surface such that

$$W(h) = mgh + W_o \quad (8.21)$$

(where W_o = constant) ;
then it is clear that

$$W_{12} = W(h_2) - W(h_1) \quad (8.22)$$

The work done in moving the particle is just the difference of potential energy between its final and initial positions.Observe that the constant W_o cancels out in Eq. (8.22). Setting $h = 0$ in the last equation, we get $W(h = 0) = W_o$. $h = 0$ means points on the surface of the earth. Thus, W_o is the potential energy on the surface of the earth.

If we consider points at arbitrary distance from the surface of the earth, the result just derived is not valid since the assumption that the gravitational force mg is a constant is no longer valid. However, from our discussion we know that a point outside the earth, the force of gravitation on a particle directed towards the center of the earth is

$$F = \frac{G M_E m}{r^2} \quad (8.23)$$

where M_E = mass of earth, m = mass of the particle and r its distance from the center of the earth. If we now calculate the work done in lifting a particle from $r = r_1$ to $r = r_2$ ($r_2 > r_1$) along a vertical path, we get instead of Eq. (8.20)

$$W_{12} = \int_{r_1}^{r_2} \frac{G M m}{r^2} dr$$

$$= -G M_E m \left(\frac{1}{r_2} - \frac{1}{r_1}\right) \quad (8.24)$$

In place of Eq. (8.21), we can thus associate a potential energy $W(r)$ at a distance r, such that

$$W(r) = -\frac{G M_E m}{r} + W_1, \quad (8.25)$$

valid for $r > R$,

so that once again $W_{12} = W(r_2) - W(r_1)$. Setting r = infinity in the last equation, we get W (r = infinity) = W_1. Thus, W_1 is the potential energy at infinity. One should note that only the difference of potential energy between two points has a definite meaning from Eqs. (8.22) and (8.24). One conventionally sets W_1 equal to zero, so that the potential energy at a point is just the amount of work done in displacing the particle from infinity to that point.

We have calculated the potential energy at a point of a particle due to gravitational forces on it due to the earth and it is proportional to the mass of the particle. The gravitational potential due to the gravitational force of the earth is defined as the potential energy of a particle of unit mass at that point. From the earlier discussion, we learn that the gravitational potential energy associated with two particles of masses m_1 and m_2 separated by distance by a distance r is given by

$$V = -\frac{G m_1 m_2}{r} \text{ (if we choose } V = 0 \text{ as } r \to \infty)$$

It should be noted that an isolated system of particles will have the total potential energy that equals the sum of energies (given by the above equation) for all possible pairs of its constituent particles. This is an example of the application of the superposition principle.

Example 8.3 Find the potential energy of a system of four particles placed at the vertices of a square of side l. Also obtain the potential at the centre of the square.

Answer Consider four masses each of mass m at the corners of a square of side l; See Fig. 8.9. We have four mass pairs at distance l and two diagonal pairs at distance $\sqrt{2}\, l$

Hence,

$$W(r) = -4\frac{G m^2}{l} - 2\frac{G m^2}{\sqrt{2}\, l}$$

Fig. 8.9

$$= -\frac{2\,G\,m^2}{l}\left(2+\frac{1}{\sqrt{2}}\right) = -5.41\frac{G\,m^2}{l}$$

The gravitational potential at the centre of the square $\left(r=\sqrt{2}\,l/2\right)$ is

$$U(r) = -4\sqrt{2}\,\frac{G\,m}{l}.$$

◀

8.8 ESCAPE SPEED

If a stone is thrown by hand, we see it falls back to the earth. Of course using machines we can shoot an object with much greater speeds and with greater and greater initial speed, the object scales higher and higher heights. A natural query that arises in our mind is the following: 'can we throw an object with such high initial speeds that it does not fall back to the earth?'

The principle of conservation of energy helps us to answer this question. Suppose the object did reach infinity and that its speed there was V_f. The energy of an object is the sum of potential and kinetic energy. As before W_1 denotes that gravitational potential energy of the object at infinity. The total energy of the projectile at infinity then is

$$E\,(\infty) = W_1 + \frac{mV_f^2}{2} \tag{8.26}$$

If the object was thrown initially with a speed V_i from a point at a distance $(h+R_E)$ from the center of the earth (R_E = radius of the earth), its energy initially was

$$E\,(h+R_E) = \frac{1}{2}mV_i^2 - \frac{GmM_E}{(h+R_E)} + W_1 \tag{8.27}$$

By the principle of energy conservation Eqs. (8.26) and (8.27) must be equal. Hence

$$\frac{mV_i^2}{2} - \frac{GmM_E}{(h+R_E)} = \frac{mV_f^2}{2} \tag{8.28}$$

The R.H.S. is a positive quantity with a minimum value zero hence so must be the L.H.S. Thus, an object can reach infinity as long as V_i is such that

$$\frac{mV_i^2}{2} - \frac{GmM_E}{(h+R_E)} \geq 0 \tag{8.29}$$

The minimum value of V_i corresponds to the case when the L.H.S. of Eq. (8.29) equals zero. Thus, the minimum speed required for an object to reach infinity (i.e. escape from the earth) corresponds to

$$\frac{1}{2}m\left(V_i^2\right)_{\min} = \frac{GmM_E}{h+R_E} \tag{8.30}$$

If the object is thrown from the surface of the earth, h=0, and we get

$$\left(V_i\right)_{\min} = \frac{\sqrt{2GM_E}}{R_E} \tag{8.31}$$

Using the relation $g = GM_E/R_E^2$, we get

$$\left(V_i\right)_{\min} = \sqrt{2gR_E} \tag{8.32}$$

Using the value of g and R_E, numerically $(V_i)_{min} \approx$ 11.2 km/s. This is called the escape speed, sometimes loosely called the escape velocity.

Equation (8.32) applies equally well to an object thrown from the surface of the moon with g replaced by the acceleration due to Moon's gravity on its surface and r_E replaced by the radius of the moon. Both are smaller than their values on earth and the escape speed for the moon turns out to be 2.3 km/s, about five times smaller. This is the reason that moon has no atmosphere. Gas molecules if formed on the surface of the moon having velocities larger than this will escape the gravitational pull of the moon.

▶ ***Example 8.4*** Two uniform solid spheres of equal radii R, but mass M and 4 M have a center to centre separation 6 R, as shown in Fig. 8.10. The two spheres are held fixed. A projectile of mass m is projected from the surface of the sphere of mass M directly towards the centre of the second sphere. Obtain an expression for the minimum speed v of the projectile so that it reaches the surface of the second sphere.

Fig. 8.10

Answer The projectile is acted upon by two mutually opposing gravitational forces of the two

spheres. The neutral point N (see Fig. 8.10) is defined as the position where the two forces cancel each other exactly. If ON = r, we have

$$\frac{G\,M\,m}{r^2} = \frac{4\,G\,M\,m}{(6R-r)^2}$$

$$(6R - r)^2 = 4r^2$$

$$6R - r = \pm 2r$$

$$r = 2R \quad \text{or} - 6R.$$

The neutral point $r = -6R$ does not concern us in this example. Thus ON = r = $2R$. It is sufficient to project the particle with a speed which would enable it to reach N. Thereafter, the greater gravitational pull of $4M$ would suffice. The mechanical energy at the surface of M is

$$E_i = \frac{1}{2} m\, v^2 - \frac{G\,M\,m}{R} - \frac{4\,G\,M\,m}{5\,R}.$$

At the neutral point N, the speed approaches zero. The mechanical energy at N is purely potential.

$$E_N = -\frac{G\,M\,m}{2\,R} - \frac{4\,G\,M\,m}{4\,R}.$$

From the principle of conservation of mechanical energy

$$\frac{1}{2} v^2 - \frac{GM}{R} - \frac{4GM}{5R} = -\frac{GM}{2R} - \frac{GM}{R}$$

or

$$v^2 = \frac{2\,G\,M}{R}\left(\frac{4}{5} - \frac{1}{2}\right)$$

$$v = \left(\frac{3\,G\,M}{5\,R}\right)^{1/2}$$

A point to note is that the speed of the projectile is zero at N, but is nonzero when it strikes the heavier sphere 4 M. The calculation of this speed is left as an exercise to the students.

8.9 EARTH SATELLITES

Earth satellites are objects which revolve around the earth. Their motion is very similar to the motion of planets around the Sun and hence Kepler's laws of planetary motion are equally applicable to them. In particular, their orbits around the earth are circular or elliptic. Moon is the only natural satellite of the earth with a near circular orbit with a time period of approximately 27.3 days which is also roughly equal to the rotational period of the moon about its own axis. Since, 1957, advances in technology have enabled many countries including India to launch artificial earth satellites for practical use in fields like telecommunication, geophysics and meteorology.

We will consider a satellite in a circular orbit of a distance $(R_E + h)$ from the centre of the earth, where R_E = radius of the earth. If m is the mass of the satellite and V its speed, the centripetal force required for this orbit is

$$\text{F(centripetal)} = \frac{mV^2}{(R_E + h)} \qquad (8.33)$$

directed towards the center. This centripetal force is provided by the gravitational force, which is

$$\text{F(gravitation)} = \frac{G\,m\,M_E}{(R_E + h)^2} \qquad (8.34)$$

Where M_E is the mass of the earth.

Equating R.H.S of Eqs. (8.33) and (8.34) and cancelling out m, we get

$$V^2 = \frac{G\,M_E}{(R_E + h)} \qquad (8.35)$$

Thus V decreases as h increases. From equation (8.35), the speed V for $h = 0$ is

$$V^2 \quad (h = 0) = GM / R_E = gR_E \qquad (8.36)$$

where we have used the relation $g = GM / R_E^2$. In every orbit, the satellite traverses a distance $2\pi(R_E + h)$ with speed V. It's time period T therefore is

$$T = \frac{2\pi(R_E + h)}{V} = \frac{2\pi(R_E + h)^{3/2}}{\sqrt{G\,M_E}} \qquad (8.37)$$

on substitution of value of V from Eq. (8.35). Squaring both sides of Eq. (8.37), we get

$$T^2 = k\ (R_E + h)^3 \quad (\text{where } k = 4\,\pi^2 / GM_E) \qquad (8.38)$$

which is Kepler's law of periods, as applied to motion of satellites around the earth. For a satellite very close to the surface of earth h can be neglected in comparison to R_E in Eq. (8.38). Hence, for such satellites, T is T_o, where

$$T_0 = 2\pi\sqrt{R_E / g} \qquad (8.39)$$

If we substitute the numerical values g ; 9.8 m s^{-2} and R_E = 6400 km., we get

$$T_0 = 2\pi\sqrt{\frac{6.4 \times 10^6}{9.8}} \ \text{s}$$

Which is approximately 85 minutes.

Example 8.5 The planet Mars has two moons, phobos and delmos. (i) phobos has a period 7 hours, 39 minutes and an orbital radius of 9.4×10^3 km. Calculate the mass of mars. (ii) Assume that earth and mars move in circular orbits around the sun, with the martian orbit being 1.52 times the orbital radius of the earth. What is the length of the martian year in days ?

Answer (i) We employ Eq. (8.38) with the sun's mass replaced by the martian mass M_m

$$T^2 = \frac{4\pi^2}{GM_m} R^3$$

$$M_m = \frac{4\pi^2}{G} \frac{R^3}{T^2}$$

$$= \frac{4 \times (3.14)^2 \times (9.4)^3 \times 10^{18}}{6.67 \times 10^{-11} \times (459 \times 60)^2}$$

$$M_m = \frac{4 \times (3.14)^2 \times (9.4)^3 \times 10^{18}}{6.67 \times (4.59 \times 6)^2 \times 10^{-5}}$$

$= 6.48 \times 10^{23}$ kg.

(ii) Once again Kepler's third law comes to our aid,

$$\frac{T_M^2}{T_E^2} = \frac{R_{MS}^3}{R_{ES}^3}$$

where R_{MS} is the mars -sun distance and R_{ES} is the earth-sun distance.

$\therefore T_M = (1.52)^{3/2} \times 365$

$= 684$ days

We note that the orbits of all planets except Mercury, Mars and Pluto are very close to being circular. For example, the ratio of the semi-minor to semi-major axis for our Earth is, b/a = 0.99986.

Example 8.6 **Weighing the Earth** : You are given the following data: $g = 9.81$ ms^{-2}, $R_E = 6.37 \times 10^6$ m, the distance to the moon $R = 3.84 \times 10^8$ m and the time period of the moon's revolution is 27.3 days. Obtain the mass of the Earth M_E in two different ways.

Answer From Eq. (8.12) we have

$$M_E = \frac{g R_E^2}{G}$$

$$= \frac{9.81 \times (6.37 \times 10^6)^2}{6.67 \times 10^{-11}}$$

$= 5.97 \times 10^{24}$ kg.

The moon is a satellite of the Earth. From the derivation of Kepler's third law [see Eq. (8.38)]

$$T^2 = \frac{4\pi^2 R^3}{G M_E}$$

$$M_E = \frac{4\pi^2 R^3}{G T^2}$$

$$= \frac{4 \times 3.14 \times 3.14 \times (3.84)^3 \times 10^{24}}{6.67 \times 10^{-11} \times (27.3 \times 24 \times 60 \times 60)^2}$$

$= 6.02 \times 10^{24}$ kg

Both methods yield almost the same answer, the difference between them being less than 1%.

Example 8.7 Express the constant k of Eq. (8.38) in days and kilometres. Given k = 10^{-13} s^2 m^{-3}. The moon is at a distance of 3.84×10^5km from the earth. Obtain its time-period of revolution in days.

Answer Given

$k = 10^{-13}$ s^2 m^{-3}

$$= 10^{-13} \left[\frac{1}{(24 \times 60 \times 60)^2} \text{d}^2 \right] \left[\frac{1}{(1/1000)^3 \text{ km}^3} \right]$$

$= 1.33 \times 10^{-14}$ d^2 km^{-3}

Using Eq. (8.38) and the given value of k, the time period of the moon is

$T^2 = (1.33 \times 10^{-14})(3.84 \times 10^5)^3$

$T = 27.3$ d

Note that Eq. (8.38) also holds for elliptical orbits if we replace $(R_E + h)$ by the semi-major axis of the ellipse. The earth will then be at one of the foci of this ellipse.

8.10 ENERGY OF AN ORBITING SATELLITE

Using Eq. (8.35), the kinetic energy of the satellite in a circular orbit with speed v is

$$KgE = \frac{1}{2} m v^2$$

$$= \frac{G m M_E}{2(R_E + h)}, \qquad (8.40)$$

Considering gravitational potential energy at infinity to be zero, the potential energy at distance (R_e+h) from the center of the earth is

$$P.E = -\frac{G m M_E}{(R_E + h)} \quad (8.41)$$

The K.E is positive whereas the P.E is negative. However, in magnitude the K.E. is half the P.E, so that the total E is

$$E = K.E + P.E = -\frac{G m M_E}{2(R_E + h)} \quad (8.42)$$

The total energy of an circularly orbiting satellite is thus negative, with the potential energy being negative but twice is magnitude of the positive kinetic energy.

When the orbit of a satellite becomes elliptic, both the *K.E.* and *P.E.* vary from point to point. The total energy which remains constant is negative as in the circular orbit case. This is what we expect, since as we have discussed before if the total energy is positive or zero, the object escapes to infinity. Satellites are always at finite distance from the earth and hence their energies cannot be positive or zero.

Example 8.8 A 400 kg satellite is in a circular orbit of radius $2R_E$ about the Earth. How much energy is required to transfer it to a circular orbit of radius $4R_E$? What are the changes in the kinetic and potential energies ?

Answer Initially,

$$E_i = -\frac{G M_E m}{4 R_E}$$

While finally

$$E_f = -\frac{G M_E m}{8 R_E}$$

The change in the total energy is

$$\Delta E = E_f - E_i$$

$$= \frac{G M_E m}{8 R_E} = \left(\frac{G M_E}{R_E^2}\right)\frac{m R_E}{8}$$

$$\Delta E = \frac{g m R_E}{8} = \frac{9.81 \times 400 \times 6.37 \times 10^6}{8} = 3.13 \times 10^9 \text{ J}$$

The kinetic energy is reduced and it mimics ΔE, namely, $\Delta K = K_f - K_i = -3.13 \times 10^9$ J.

The change in potential energy is twice the change in the total energy, namely

$\Delta V = V_f - V_i = -6.25 \times 10^9$ J

8.11 GEOSTATIONARY AND POLAR SATELLITES

An interesting phenomenon arises if in we arrange the value of (R_E+ h) such that T in Eq. (8.37) becomes equal to 24 hours. If the circular orbit is in the equatorial plane of the earth, such a satellite, having the same period as the period of rotation of the earth about its own axis would appear stationery viewed from a point on earth. The (R_E + h) for this purpose works out to be large as compared to R_E :

$$R_E + h = \left(\frac{T^2 G M_E}{4\pi^2}\right)^{1/3} \quad (8.43)$$

and for T = 24 hours, h works out to be 35800 km. which is much larger than R_E. Satellites in a circular orbits around the earth in the

India's Leap into Space

India entered the space age with the launching of the low orbit satellite Aryabhatta in 1975. In the first few years of its programme the launch vehicles were provided by the erstwhile Soviet Union. Indigenous launch vehicles were employed in the early 1980's to send the Rohini series of satellites into space. The programme to send polar satellites into space began in late 1980's. A series of satellites labelled IRS (Indian Remote Sensing Satellites) have been launched and this programme is expected to continue in future. The satellites have been employed for surveying, weather prediction and for carrying out experiments in space. The INSAT (Indian National Satellite) series of satellites were designed and made operational for communications and weather prediction purposes beginning in 1982. European launch vehicles have been employed in the INSAT series. India tested its geostationary launch capability in 2001 when it sent an experimental communications satellite (GSAT-1) into space. In 1984 Rakesh Sharma became the first Indian astronaut. The Indian Space Research Organisation (ISRO) is the umbrella organisation that runs a number of centre. Its main lauch centre at Sriharikota (SHAR) is 100 km north of Chennai. The National Remote Sensing Agency (NRSA) is near Hyderabad. Its national centre for research in space and allied sciences is the Physical Research Laboratory (PRL) at Ahmedabad.

equatorial plane with T = 24 hours are called Geostationery Satellites. Clearly, since the earth rotates with the same period, the satellite would appear fixed from any point on earth. It takes very powerful rockets to throw up a satellite to such large heights above the earth but this has been done in view of the several benefits of many practical applications.

It is known that electromagnetic waves above a certain frequency are not reflected from ionosphere. Radio waves used for radio broadcast which are in the frequency range 2 MHz to 10 MHz, are below the critical frequency. They are therefore reflected by the ionosphere. Thus radio waves broadcast from an antenna can be received at points far away where the direct wave fail to reach on account of the curvature of the earth. Waves used in television broadcast or other forms of communication have much higher frequencies and thus cannot be received beyond the line of sight. A Geostationery satellite, appearing fixed above the broadcasting station can however receive these signals and broadcast them back to a wide area on earth. The INSAT group of satellites sent up by India are one such group of Geostationary satellites widely used for telecommunications in India.

Fig. 8.11 A Polar satellite. A strip on earth's surface (shown shaded) is visible from the satellite during one cycle. For the next revolution of the satellite, the earth has rotated a little on its axis so that an adjacent strip becomes visible.

Another class of satellites are called the Polar satellites (Fig. 8.11). These are low altitude ($h \approx$ 500 to 800 km) satellites, but they go around the poles of the earth in a north-south direction whereas the earth rotates around its axis in an east-west direction. Since its time period is around 100 minutes it crosses any altitude many times a day. However, since its height h above the earth is about 500-800 km, a camera fixed on it can view only small strips of the earth in one orbit. Adjacent strips are viewed in the next orbit, so that in effect the whole earth can be viewed strip by strip during the entire day. These satellites can view polar and equatorial regions at close distances with good resolution. Information gathered from such satellites is extremely useful for remote sensing, meterology as well as for environmental studies of the earth.

8.12 WEIGHTLESSNESS

Weight of an object is the force with which the earth attracts it. We are conscious of our own weight when we stand on a surface, since the surface exerts a force opposite to our weight to keep us at rest. The same principle holds good when we measure the weight of an object by a spring balance hung from a fixed point e.g. the ceiling. The object would fall down unless it is subject to a force opposite to gravity. This is exactly what the spring exerts on the object. This is because the spring is pulled down a little by the gravitational pull of the object and in turn the spring exerts a force on the object vertically upwards.

Now, imagine that the top end of the balance is no longer held fixed to the top ceiling of the room. Both ends of the spring as well as the object move with identical acceleration g. The spring is not stretched and does not exert any upward force on the object which is moving down with acceleration g due to gravity. The reading recorded in the spring balance is zero since the spring is not stretched at all. If the object were a human being, he or she will not feel his weight since there is no upward force on him. Thus, when an object is in free fall, it is weightless and this phenomenon is usually called the phenomenon of weightlessness.

In a satellite around the earth, every part and parcel of the satellite has an acceleration towards the center of the earth which is exactly

the value of earth's acceleration due to gravity at that position. Thus in the satellite everything inside it is in a state of free fall. This is just as if we were falling towards the earth from a height. Thus, in a manned satellite, people inside experience no gravity. Gravity for us defines the vertical direction and thus for them there are no horizontal or vertical directions, all directions are the same. Pictures of astronauts floating in a satellite reflect show this fact.

SUMMARY

1. Newton's law of universal gravitation states that the gravitational force of attraction between any two particles of masses m_1 and m_2 separated by a distance r has the magnitude

$$F = G\frac{m_1 m_2}{r^2}$$

where G is the universal gravitational constant, which has the value 6.672×10^{-11} N m^2 kg^{-2}.

2. If we have to find the resultant gravitational force acting on the particle m due to a number of masses $M_1, M_2, \ldots M_n$ etc. we use the principle of superposition. Let $F_1, F_2, \ldots F_n$ be the individual forces due to $M_1, M_2, \ldots M_n$ each given by the law of gravitation. From the principle of superposition each force acts independently and uninfluenced by the other bodies. The resultant force F_R is then found by vector addition

$$F_R = F_1 + F_2 + \ldots\ldots + F_n = \sum_{i=1}^{n} \boldsymbol{F}_i$$

where the symbol 'Σ' stands for summation.

3. Kepler's laws of planetary motion state that
   (a) All planets move in elliptical orbits with the Sun at one of the focal points
   (b) The radius vector drawn from the sun to a planet sweeps out equal areas in equal time intervals. This follows from the fact that the force of gravitation on the planet is central and hence angular momentum is conserved.
   (c) The square of the orbital period of a planet is proportional to the cube of the semi-major axis of the elliptical orbit of the planet

   The period T and radius R of the circular orbit of a planet about the Sun are related by

$$T^2 = \left(\frac{4\pi^2}{G\,M_s}\right)R^3$$

where M_s is the mass of the Sun. Most planets have nearly circular orbits about the Sun. For elliptical orbits, the above equation is valid if R is replaced by the semi-major axis, a.

4. The acceleration due to gravity.
   (a) at a height h above the Earth's surface

$$g(h) = \frac{G\,M_E}{(R_E + h)^2}$$

$$\approx \frac{G\,M_E}{R_E^2}\left(1 - \frac{2h}{R_E}\right) \quad \text{for } h << R_E$$

$$g(h) = g(0)\left(1 - \frac{2h}{R_E}\right) \quad \text{where } g(0) = \frac{G\,M_E}{R_E^2}$$

(b) at depth d below the Earth's surface is

$$g(d) = \frac{G\,M_E}{R_E^2}\left(1 - \frac{d}{R_E}\right) = g(0)\left(1 - \frac{d}{R_E}\right)$$

5. The gravitational force is a conservative force, and therefore a potential energy function can be defined. The *gravitational potential energy* associated with two particles separated by a distance r is given by

$$V = -\frac{G\,m_1\,m_2}{r}$$

where V is taken to be zero at $r \to \infty$. The total potential energy for a system of particles is the sum of energies for all pairs of particles, with each pair represented by a term of the form given by above equation. This prescription follows from the principle of superposition.

6. If an isolated system consists of a particle of mass m moving with a speed v in the vicinity of a massive body of mass M, the total mechanical energy of the particle is given by

$$E = \frac{1}{2}\,m\,v^2 - \frac{G\,M\,m}{r}$$

That is, the total mechanical energy is the sum of the kinetic and potential energies. The total energy is a constant of motion.

7. If m moves in a circular orbit of radius a about M, where $M >> m$, the total energy of the system is

$$E = -\frac{G\,M\,m}{2a}$$

with the choice of the arbitrary constant in the potential energy given in the point 5., above. The total energy is negative for any bound system, that is, one in which the orbit is closed, such as an elliptical orbit. The kinetic and potential energies are

$$K = \frac{G\,M\,m}{2a}$$

$$V = -\frac{G\,M\,m}{a}$$

8. The escape speed from the surface of the Earth is

$$v_e = \sqrt{\frac{2\,G\,M_E}{R_E}} = \sqrt{2gR_E}$$

and has a value of 11.2 km s^{-1}.

9. If a particle is outside a uniform spherical shell or solid sphere with a spherically symmetric internal mass distribution, the sphere attracts the particle as though the mass of the sphere or shell were concentrated at the centre of the sphere.

10. If a particle is inside a uniform spherical shell, the gravitational force on the particle is zero. If a particle is inside a homogeneous solid sphere, the force on the particle acts toward the centre of the sphere. This force is exerted by the spherical mass interior to the particle.

11. A geostationary (geosynchronous communication) satellite moves in a circular orbit in the equatorial plane at a approximate distance of 4.22×10^4 km from the Earth's centre.

Physical Quantity	Symbol	Dimensions	Unit	Remarks
Gravitational Constant	G	$[M^{-1} L^3 T^{-2}]$	$N\ m^2\ kg^{-2}$	6.67×10^{-11}
Gravitational Potential Energy	$V(r)$	$[M\ L^2 T^{-2}]$	J	$-\frac{GMm}{r}$ (scalar)
Gravitational Potential	$U(r)$	$[L^2T^{-2}]$	$J\ kg^{-1}$	$-\frac{GM}{r}$ (scalar)
Gravitational Intensity	$\mathbf{E}$ or $\mathbf{g}$	$[LT^{-2}]$	$m\ s^{-2}$	$\frac{GM}{r^2}\hat{r}$ (vector)

POINTS TO PONDER

1. In considering motion of an object under the gravitational influence of another object the following quantities are conserved:
   (a) Angular momentum
   (b) Total mechanical energy
   Linear momentum is **not** conserved
2. Angular momentum conservation leads to Kepler's second law. However, it is not special to the inverse square law of gravitation. It holds for any central force.
3. In Kepler's third law (see Eq. (8.1) and $T^2 = K_S R^3$. The constant K_S is the same for all planets in circular orbits. This applies to satellites orbiting the Earth [(Eq. (8.38)].
4. An astronaut experiences weightlessness in a space satellite. This is not because the gravitational force is small at that location in space. It is because both the astronaut and the satellite are in "free fall" towards the Earth.
5. The *gravitational potential energy* associated with two particles separated by a distance r is given by

$$V = -\frac{G m_1 m_2}{r} + \text{constant}$$

   The constant can be given any value. The simplest choice is to take it to be zero. With this choice

$$V = -\frac{G m_1 m_2}{r}$$

   This choice implies that $V \to 0$ as $r \to \infty$. Choosing location of zero of the gravitational energy is the same as choosing the arbitrary constant in the potential energy. Note that the gravitational force is not altered by the choice of this constant.
6. The total mechanical energy of an object is the sum of its kinetic energy (which is always positive) and the potential energy. Relative to infinity (i.e. if we presume that the potential energy of the object at infinity is zero), the gravitational potential energy of an object is negative. The total energy of a satellite is negative.
7. The commonly encountered expression mgh for the potential energy is actually an approximation to the difference in the gravitational potential energy discussed in the point 6, above.
8. Although the gravitational force between two particles is central, the force between two finite rigid bodies is not necessarily along the line joining their centre of mass. For a spherically symmetric body however the force on a particle external to the body is as if the mass is concentrated at the centre and this force is therefore central.
9. The gravitational force on a particle inside a spherical shell is zero. However, (unlike a metallic shell which shields electrical forces) the shell does not shield other bodies outside it from exerting gravitational forces on a particle inside. *Gravitational shielding is not possible.*

EXERCISES

8.1 Answer the following :

(a) You can shield a charge from electrical forces by putting it inside a hollow conductor. Can you shield a body from the gravitational influence of nearby matter by putting it inside a hollow sphere or by some other means ?

(b) An astronaut inside a small space ship orbiting around the earth cannot detect gravity. If the space station orbiting around the earth has a large size, can he hope to detect gravity ?

(c) If you compare the gravitational force on the earth due to the sun to that due to the moon, you would find that the Sun's pull is greater than the moon's pull. (you can check this yourself using the data available in the succeeding exercises). However, the tidal effect of the moon's pull is greater than the tidal effect of sun. Why ?

8.2 Choose the correct alternative :

(a) Acceleration due to gravity increases/decreases with increasing altitude.

(b) Acceleration due to gravity increases/decreases with increasing depth (assume the earth to be a sphere of uniform density).

(c) Acceleration due to gravity is independent of mass of the earth/mass of the body.

(d) The formula $-G\,Mm(1/r_2 - 1/r_1)$ is more/less accurate than the formula $mg(r_2 - r_1)$ for the difference of potential energy between two points r_2 and r_1 distance away from the centre of the earth.

8.3 Suppose there existed a planet that went around the sun twice as fast as the earth. What would be its orbital size as compared to that of the earth ?

8.4 Io, one of the satellites of Jupiter, has an orbital period of 1.769 days and the radius of the orbit is 4.22×10^8 m. Show that the mass of Jupiter is about one-thousandth that of the sun.

8.5 Let us assume that our galaxy consists of 2.5×10^{11} stars each of one solar mass. How long will a star at a distance of 50,000 ly from the galactic centre take to complete one revolution ? Take the diameter of the Milky Way to be 10^5 ly.

8.6 Choose the correct alternative:

(a) If the zero of potential energy is at infinity, the total energy of an orbiting satellite is negative of its kinetic/potential energy.

(b) The energy required to launch an orbiting satellite out of earth's gravitational influence is more/less than the energy required to project a stationary object at the same height (as the satellite) out of earth's influence.

8.7 Does the escape speed of a body from the earth depend on (a) the mass of the body, (b) the location from where it is projected, (c) the direction of projection, (d) the height of the location from where the body is launched?

8.8 A comet orbits the sun in a highly elliptical orbit. Does the comet have a constant (a) linear speed, (b) angular speed, (c) angular momentum, (d) kinetic energy, (e) potential energy, (f) total energy throughout its orbit? Neglect any mass loss of the comet when it comes very close to the Sun.

8.9 Which of the following symptoms is likely to afflict an astronaut in space (a) swollen feet, (b) swollen face, (c) headache, (d) orientational problem.

In the following two exercises, choose the correct answer from among the given ones:

8.10 The gravitational intensity at the centre of a hemispherical shell of uniform mass density has the direction indicated by the arrow (see Fig 8.12) (i) a, (ii) b, (iii) c, (iv) 0.

Fig. 8.12

8.11 For the above problem, the direction of the gravitational intensity at an arbitrary point P is indicated by the arrow (i) d, (ii) e, (iii) f, (iv) g.

8.12 A rocket is fired from the earth towards the sun. At what distance from the earth's centre is the gravitational force on the rocket zero ? Mass of the sun = 2×10^{30} kg, mass of the earth = 6×10^{24} kg. Neglect the effect of other planets etc. (orbital radius = 1.5×10^{11} m).

8.13 How will you 'weigh the sun', that is estimate its mass? The mean orbital radius of the earth around the sun is 1.5×10^{8} km.

8.14 A saturn year is 29.5 times the earth year. How far is the saturn from the sun if the earth is 1.50×10^{8} km away from the sun ?

8.15 A body weighs 63 N on the surface of the earth. What is the gravitational force on it due to the earth at a height equal to half the radius of the earth ?

8.16 Assuming the earth to be a sphere of uniform mass density, how much would a body weigh half way down to the centre of the earth if it weighed 250 N on the surface ?

8.17 A rocket is fired vertically with a speed of 5 km s^{-1} from the earth's surface. How far from the earth does the rocket go before returning to the earth ? Mass of the earth = 6.0×10^{24} kg; mean radius of the earth = 6.4×10^{6} m; $G = 6.67 \times 10^{-11}$ N m^2 kg^{-2}.

8.18 The escape speed of a projectile on the earth's surface is 11.2 km s^{-1}. A body is projected out with thrice this speed. What is the speed of the body far away from the earth? Ignore the presence of the sun and other planets.

8.19 A satellite orbits the earth at a height of 400 km above the surface. How much energy must be expended to rocket the satellite out of the earth's gravitational influence? Mass of the satellite = 200 kg; mass of the earth = 6.0×10^{24} kg; radius of the earth = 6.4×10^{6} m; $G = 6.67 \times 10^{-11}$ N m^2 kg^{-2}.

8.20 Two stars each of one solar mass (= 2×10^{30} kg) are approaching each other for a head on collision. When they are a distance 10^{9} km, their speeds are negligible. What is the speed with which they collide ? The radius of each star is 10^{4} km. Assume the stars to remain undistorted until they collide. (Use the known value of *G*).

8.21 Two heavy spheres each of mass 100 kg and radius 0.10 m are placed 1.0 m apart on a horizontal table. What is the gravitational force and potential at the mid point of the line joining the centres of the spheres ? Is an object placed at that point in equilibrium? If so, is the equilibrium stable or unstable ?

Additional Exercises

8.22 As you have learnt in the text, a geostationary satellite orbits the earth at a height of nearly 36,000 km from the surface of the earth. What is the potential due to earth's gravity at the site of this satellite ? (Take the potential energy at infinity to be zero). Mass of the earth = 6.0×10^{24} kg, radius = 6400 km.

8.23 A star 2.5 times the mass of the sun and collapsed to a size of 12 km rotates with a speed of 1.2 rev. per second. (Extremely compact stars of this kind are known as neutron stars. Certain stellar objects called pulsars belong to this category). Will an object placed on its equator remain stuck to its surface due to gravity ? (mass of the sun = 2×10^{30} kg).

8.24 A spaceship is stationed on Mars. How much energy must be expended on the spaceship to launch it out of the solar system ? Mass of the space ship = 1000 kg; mass of the sun = 2×10^{30} kg; mass of mars = 6.4×10^{23} kg; radius of mars = 3395 km; radius of the orbit of mars = 2.28×10^{8} km; $G = 6.67\times 10^{-11}$ N m^2 kg^{-2}.

8.25 A rocket is fired 'vertically' from the surface of mars with a speed of 2 km s^{-1}. If 20% of its initial energy is lost due to martian atmospheric resistance, how far will the rocket go from the surface of mars before returning to it ? Mass of mars = 6.4×10^{23} kg; radius of mars = 3395 km; $G = 6.67\times 10^{-11}$ N m^2 kg^{-2}.

APPENDICES

APPENDIX A 1
THE GREEK ALPHABET

Alpha	Α	α	Iota	Ι	ι	Rho	Ρ	ρ
Beta	Β	β	Kappa	Κ	κ	Sigma	Σ	σ
Gamma	Γ	γ	Lambda	Λ	λ	Tau	Τ	τ
Delta	Δ	δ	Mu	Μ	μ	Upsilon	Υ	υ
Epsilon	Ε	ε	Nu	Ν	ν	Phi	Φ	ϕ, φ
Zeta	Ζ	ς	Xi	Ξ	ξ	Chi	Χ	χ
Eta	Η	η	Omicron	Ο	ο	Psi	Ψ	ψ
Theta	Θ	θ	Pi	Π	π	Omega	Ω	ω

APPENDIX A 2
COMMON SI PREFIXES AND SYMBOLS FOR MULTIPLES AND SUB-MULTIPLES

Multiple			Sub-Multiple		
Factor	**Prefix**	**Symbol**	**Factor**	**Prefix**	**symbol**
10^{18}	Exa	E	10^{-18}	atto	a
10^{15}	Peta	P	10^{-15}	femto	f
10^{12}	Tera	T	10^{-12}	pico	p
10^{9}	Giga	G	10^{-9}	nano	n
10^{6}	Mega	M	10^{-6}	micro	μ
10^{3}	kilo	k	10^{-3}	milli	m
10^{2}	Hecto	h	10^{-2}	centi	c
10^{1}	Deca	da	10^{-1}	deci	d

APPENDIX A 3

SOME IMPORTANT CONSTANTS

Name	Symbol	Value
Speed of light in vacuum	c	2.9979×10^{8} m s^{-1}
Charge of electron	e	1.602×10^{-19} C
Gravitational constant	G	6.673×10^{-11} N m^{2} kg^{-2}
Planck constant	h	6.626×10^{-34} J s
Boltzmann constant	k	1.381×10^{-23} J K^{-1}
Avogadro number	N_A	6.022×10^{23}mol^{-1}
Universal gas constant	R	8.314 J mol^{-1} K^{-1}
Mass of electron	m_e	9.110×10^{-31} kg
Mass of neutron	m_n	1.675×10^{-27} kg
Mass of proton	m_p	1.673×10^{-27} kg
Electron-charge to mass ratio	e/m_e	1.759×10^{11} C/kg
Faraday constant	F	9.648×10^{4} C/mol
Rydberg constant	R	1.097×10^{7} m^{-1}
Bohr radius	a_0	5.292×10^{-11} m
Stefan-Boltzmann constant	σ	5.670×10^{-8} W m^{-2} K^{-4}
Wien's Constant	b	2.898×10^{-3} m K
Permittivity of free space	ε_0 $1/4\pi\,\varepsilon_0$	8.854×10^{-12} C^{2} N^{-1}m^{-2} 8.987×10^{9} N m^{2} C^{-2}
Permeability of free space	μ_0	$4\pi \times 10^{-7}$ T m A^{-1} $\cong 1.257 \times 10^{-6}$ Wb A^{-1} m^{-1}

Other useful constants

Name	Symbol	Value
Mechanical equivalent of heat	J	4.186 J cal^{-1}
Standard atmospheric pressure	1 atm	1.013×10^{5}Pa
Absolute zero	0 K	–273.15 °C
Electron volt	1 eV	1.602×10^{-19}J
Unified Atomic mass unit	1 u	1.661×10^{-27} kg
Electron rest energy	m c^{2}	0.511 MeV
Energy equivalent of 1 u	1 u c^{2}	931.5 MeV
Volume of ideal gas(0 °C and 1atm)	V	22.4 L mol^{-1}
Acceleration due to gravity (sea level, at equator)	g	9.78049 m s^{-2}

APPENDIX A 4
CONVERSION FACTORS

Conversion factors are written as equations for simplicity.

Length

1 km = 0.6215 mi

1mi = 1.609 km

1m = 1.0936 yd = 3.281 ft = 39.37 in

1 in = 2.54 cm

1 ft = 12 in = 30.48 cm

1 yd = 3ft = 91.44 cm

1 lightyear = 1 ly = 9.461 x 10^{15}m

1 Å= 0.1nm

Area

1 m^2 = 10^4 cm^2

1km^2 = 0.3861 mi^2 = 247.1 acres

1 in^2= 6.4516 cm^2

1ft^2= 9.29 x $10^{-2}m^2$

1 m^2= 10.76 ft^2

1 acre = 43,560 ft^2

1 mi^2= 460 acres = 2.590 km^2

Volume

1m^3= 10^6cm^3

1 L = 1000 cm^3 = 10^{-3} m^3

1 gal = 3.786 L

1 gal = 4 qt = 8 pt = 128 oz = 231 in^3

1 in^3 = 16.39 cm^3

1ft^3 = 1728 in^3 = 28.32 L = 2.832 × 10^4 cm^3

Speed

1 km h^{-1} = 0.2778 m s^{-1} = 0.6215 mi h^{-1}

1mi h^{-1} = 0.4470 m s^{-1} = 1.609 km h^{-1}

1mi h^{-1} = 1.467 ft s^{-1}

Magnetic Field

1 G = 10^{-4} T

1 T = 1 Wb m^{-2} = 10^4 G

Angle and Angular Speed

π rad = 180°

1 rad = 57.30°

1° = 1.745 × 10^{-2} rad

1 rev min^{-1} = 0.1047 rad s^{-1}

1 rad s^{-1} = 9.549 rev min^{-1}

Mass

1 kg = 1000 g

1 tonne = 1000 kg = 1 Mg

1 u = 1.6606 × 10^{-27} kg

1 kg = 6.022 × 10^{26} u

1 slug = 14.59 kg

1 kg = 6.852 × 10^{-2} slug

1 u = 931.50 MeV/c^2

Density

1 g cm^{-3} = 1000 kg m^{-3} = 1 kg L^{-1}

Force

1 N = 0.2248 lbf = 10^5 dyn

1 lbf = 4.4482 N

1 kgf = 2.2046 lbf

Time

1 h = 60 min = 3.6 ks

1 d = 24 h = 1440 min = 86.4 ks

1y = 365.24 d = 31.56 Ms

Pressure

1 Pa = 1 N m^{-2}

1 bar = 100 kPa

1 atm = 101.325 kPa = 1.01325 bar

1atm = 14.7 lbf/in^2 = 760 mm Hg

= 29.9 in Hg = 33.8 ft H_2O

1 lbf in^{-2} = 6.895 kPa

1 torr = 1mm Hg = 133.32 Pa

Energy

1 kW h = 3.6 MJ

1 cal = 4.186 J

1ft lbf = 1.356 J = 1.286×10^{-3} Btu

1 L atm = 101.325 J

1 L atm = 24.217 cal

1 Btu = 778 ft lb = 252 cal = 1054.35 J

1 eV = 1.602×10^{-19}J

1 u c^2 = 931.50 MeV

1 erg = 10^{-7}J

Power

1 horsepower (hp) = 550 ft lbf/s

= 745.7 W

1 Btu min^{-1} = 17.58 W

1 W = 1.341×10^{-3} hp

= 0.7376 ft lbf/s

Thermal Conductivity

1 W m^{-1} K^{-1} = 6.938 Btu in/hft^2 °F

1 Btu in/hft^2 °F = 0.1441 W/m K

APPENDIX A 5
MATHEMATICAL FORMULAE

Geometry

Circle of radius r: circumference = $2\pi r$;
area = πr^2

Sphere of radius r: area = $4\pi r^2$;

volume = $\frac{4}{3}\pi r^3$

Right circular cylinder of radius r and height h: area = $2\pi r^2 + 2\pi r h$;

volume = $\pi r^2 h$;

Triangle of base a and altitude h.

area = $\frac{1}{2} a h$

Quadratic Formula

If $ax^2 + bx + c = 0$,

then $$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2}$$

Trigonometric Functions of Angle θ

Fig. A 5.1

$$\sin\theta = \frac{y}{r} \qquad \cos\theta = \frac{x}{r}$$

$$\tan\theta = \frac{y}{x} \qquad \cot\theta = \frac{x}{y}$$

$$\sec\theta = \frac{r}{x} \qquad \csc\theta = \frac{r}{y}$$

Pythagorean Theorem

In this right triangle, $a^2 + b^2 = c^2$

Fig. A 5.2

Triangles

Angles are A, B, C

Opposite sides are a, b, c

Angles $A + B + C = 180^0$

$$\frac{\sin A}{a} = \frac{\sin B}{b} = \frac{\sin C}{c}$$

$c^2 = a^2 + b^2 - 2ab \cos C$

Exterior angle $D = A + C$

Fig. A 5.3

Mathematical Signs and Symbols

= equals
≅ equals approximately
~ is the order of magnitude of
≠ is not equal to
≡ is identical to, is defined as
> is greater than (>> is much greater than)
< is less than (<< is much less than)
≥ is greater than or equal to (or, is no less than)
≤ is less than or equal to (or, is no more than)
± plus or minus
∝ is proportional to
∑ the sum of
$\bar{x}$ or $<x>$ or x_{av} the average value of x

Trigonometric Identities

$\sin(90^0 - \theta) = \cos\theta$

$\cos(90^0 - \theta) = \sin\theta$

$\sin\theta / \cos\theta = \tan\theta$

$\sin^2\theta + \cos^2\theta = 1$

$\sec^2\theta - \tan^2\theta = 1$

$\csc^2\theta - \cot^2\theta = 1$

$\sin 2\theta = 2\sin\theta\cos\theta$

$\cos 2\theta = \cos^2\theta - \sin^2\theta = 2\cos^2\theta - 1$
$= 1 - 2\sin^2\theta$

$\sin(\alpha \pm \beta) = \sin\alpha\cos\beta \pm \cos\alpha\sin\beta$

$\cos(\alpha \pm \beta) = \cos\alpha\cos\beta \mp \sin\alpha\sin\beta$

$\tan(\alpha \pm \beta) = \dfrac{\tan\alpha \pm \tan\beta}{1 \mp \tan\alpha\tan\beta}$

$\sin\alpha \pm \sin\beta = 2\sin\frac{1}{2}(a \pm b)\cos\frac{1}{2}(a \mp b)$

$\cos\alpha + \cos\beta$

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

$\cos\alpha - \cos\beta$

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

Binomial Theorem

$$(1 \pm x)^n = 1 \pm \frac{nx}{1!} + \frac{n(n-1)x^2}{2!} + \ldots\ldots (x^2 < 1)$$

$$(1 \pm x)^{-n} = 1 \mp \frac{nx}{1!} + \frac{n(n+1)x^2}{2!} + \ldots\ldots (x^2 < 1)$$

Exponential Expansion

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots\ldots$$

Logarithmic Expansion

$$\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \ldots\ldots \quad (|x| < 1)$$

Trigonometric Expansion
(θ in radians)

$$\sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \ldots\ldots$$

$$\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \ldots\ldots$$

$$\tan\theta = \theta + \frac{\theta^3}{3} + \frac{2\theta^5}{15} - \ldots\ldots$$

Products of Vectors

Let $\hat{\mathbf{i}}, \hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$ be unit vectors in the x, y and z directions. Then

$$\hat{\mathbf{i}} \cdot \hat{\mathbf{i}} = \hat{\mathbf{j}} \cdot \hat{\mathbf{j}} = \hat{\mathbf{k}} \cdot \hat{\mathbf{k}} = 1, \quad \hat{\mathbf{i}} \cdot \hat{\mathbf{j}} = \hat{\mathbf{j}} \cdot \hat{\mathbf{k}} = \hat{\mathbf{k}} \cdot \hat{\mathbf{i}} = 0$$

$$\hat{\mathbf{i}} \times \hat{\mathbf{i}} = \hat{\mathbf{j}} \times \hat{\mathbf{j}} = \hat{\mathbf{k}} \times \hat{\mathbf{k}} = 0, \quad \hat{\mathbf{i}} \times \hat{\mathbf{j}} = \hat{\mathbf{k}}, \hat{\mathbf{j}} \times \hat{\mathbf{k}} = \hat{\mathbf{i}}, \hat{\mathbf{k}} \times \hat{\mathbf{i}} = \hat{\mathbf{j}}$$

Any vector **a** with components a_x, a_y, and a_z along the x, y, and z axes can be written,

$$\mathbf{a} = a_x\hat{\mathbf{i}} + a_y\hat{\mathbf{j}} + a_z\hat{\mathbf{k}}$$

Let **a**, **b** and **c** be arbitary vectors with magnitudes a, b and c. Then

$$\mathbf{a}\times(\mathbf{b}+\mathbf{c}) = (\mathbf{a}\times\mathbf{b}) + (\mathbf{a}\times\mathbf{c})$$

$$(s\mathbf{a})\times\mathbf{b} = \mathbf{a}\times(s\mathbf{b}) = s(\mathbf{a}\times\mathbf{b}) \quad \text{(s is a scalar)}$$

Let θ be the smaller of the two angles between **a** and **b**. Then

$$\mathbf{a}\cdot\mathbf{b} = \mathbf{b}\cdot\mathbf{a} = a_x b_x + a_y b_y + a_z b_z = ab\cos\theta$$

$$|\mathbf{a}\times\mathbf{b}| = ab\sin\theta$$

$$\mathbf{a}\times\mathbf{b} = -\mathbf{b}\times\mathbf{a} = \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{vmatrix}$$

$$= (a_y b_z - b_y a_z)\,\hat{\mathbf{i}} + (a_z b_x - b_z a_x)\,\hat{\mathbf{j}} + (a_x b_y - b_x a_y)\,\hat{\mathbf{k}}$$

$$\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = \mathbf{b}\cdot(\mathbf{c}\times\mathbf{a}) = \mathbf{c}\cdot(\mathbf{a}\times\mathbf{b})$$

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = (\mathbf{a}\cdot\mathbf{c})\,\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\,\mathbf{c}$$

APPENDIX A 6

SI DERIVED UNITS

A 6.1 Some SI Derived Units expressed in SI Base Units

Physical quantity	SI Unit	
	Name	Symbol
Area	square metre	m^2
Volume	cubic metre	m^3
Speed, velocity	metre per second	m/s or m s^{-1}
Angular velocity	radian per second	rad/s or rad s^{-1}
Acceleration	metre per second square	m/s^2 or m s^{-2}
Angular acceleration	radian per second square	rad/s^2 or rad s^{-2}
Wave number	per metre	m^{-1}
Density, mass density	kilogram per cubic metre	kg/m^3 or kg m^{-3}
Current density	ampere per square metre	A/m^2 or A m^{-2}
Magnetic field strength, magnetic intensity, magnetic moment density	ampere per metre	A/m or A m^{-1}
Concentration (of amount of substance)	mole per cubic metre	mol/m^3 or mol m^{-3}
Specific volume	cubic metre per kilogram	m^3/kg or m^3 kg^{-1}
Luminance, intensity of illumination	candela per square metre	cd/m^2 or cd m^{-2}
Kinematic viscosity	square metre per second	m^2/s or m^2 s^{-1}
Momentum	kilogram metre per second	kg m s^{-1}
Moment of inertia	kilogram square metre	kg m^2
Radius of gyration	metre	m
Linear/superficial/volume expansivities	per kelvin	K^{-1}
Flow rate	cubic metre per second	m^3 s^{-1}

A 6.2 SI Derived Units with special names

Physical quantity	SI Unit			
	Name	Symbol	Expression in terms of other units	Expression in terms of SI base Units
Frequency	hertz	Hz	-	s^{-1}
Force	newton	N	-	kg m s^{-2} or kg m/s^2
Pressure, stress	Pascal	Pa	N/m^2 or N m^{-2}	kg m^{-1} s^{-2} or kg /s^2 m
Energy, work, quantity of heat	joule	J	N m	kg m^2 s^{-2} or kg m^2/s^2
Power, radiant flux	watt	W	J/s or J s^{-1}	kg m^2 s^{-3}or kg m^2/s^3
Quantity of electricity, electric charge	coulomb	C	-	A s
Electric potential, potential difference, electromotive force	volt	V	W/A or W A^{-1}	kg $m^2$$s^{-3}$ A^{-1} or kg m^2/s^3 A
Capacitance	farad	F	C/V	A^2 s^4 kg^{-1} m^{-2}
Electric resistance	ohm	Ω	V/A	kg m^2 s $^{-3}$ A^{-2}
Conductance	siemens	S	A/V	m^{-2} kg^{-1} s^3 A^2
Magnetic flux	weber	Wb	V s or J/A	kg m^2 s^{-2} A^{-1}
Magnetic field, magnetic flux density, magnetic induction	tesla	T	Wb/m^2	kg s^{-2} A^{-1}
Inductance	henry	H	Wb/A	kg m^2 s^{-2} A^{-2}
Luminous flux, luminous power	lumen	lm	-	cd /sr
Illuminance	lux	lx	lm/m^2	m^{-2} cd sr^{-1}
Activity (of a radio nuclide/radioactive source)	becquerel	Bq	-	s^{-1}
Absorbed dose, absorbed dose index	gray	Gy	J/kg	m^2/s^2 or m^2 s^{-2}

A 6.3 Some SI Derived Units expressed by means of SI Units with special names

Physical quantity	SI Unit		
	Name	Symbol	Expression in terms of SI base units
Magnetic moment	joule per tesla	J T^{-1}	m^2 A
Dipole moment	coulomb metre	C m	s A m
Dynamic viscosity	poiseiulles or pascal second or newton second per square metre	Pl or Pa s or N s m^{-2}	m^{-1} kg s^{-1}
Torque, couple, moment of force	newton metre	N m	m^2 kg s^{-2}
Surface tension	newton per metre	N/m	kg s^{-2}
Power density, irradiance, heat flux density	watt per square metre	W/m^2	kg s^{-3}

Heat capacity, entropy	joule per kelvin	J/K	m^2 kg s^{-2} K^{-1}
Specific heat capacity, specific entropy	joule per kilogram kelvin	J/kg K	m^2 s^{-2} K^{-1}
Specific energy, latent heat	joule per kilogram	J/kg	m^2 s^{-2}
Radiant intensity	watt per steradian	W sr^{-1}	kg m^2 s^{-3} sr^{-1}
Thermal conductivity	watt per metre kelvin	W m^{-1} K^{-1}	m kg s^{-3} K^{-1}
Energy density	joule per cubic metre	J/m^3	kg m^{-1} s^{-2}
Electric field strength	volt per metre	V/m	m kg s^3 A^{-1}
Electric charge density	coulomb per cubic metre	C/m^3	m^{-3} A s
Electric flux density	coulomb per square metre	C/m^2	m^{-2} A s
Permittivity	farad per metre	F/m	m^{-3} kg^{-1} s^4 A^2
Permeability	henry per metre	H/m	m kg s^{-2} A^{-2}
Molar energy	joule per mole	J/mol	m^2 kg $s^{-2}$$mol^{-1}$
Angular momentum, Planck's constant	joule second	J s	kg m^2 s^{-1}
Molar entropy, molar heat capacity	joule per mole kelvin	J/mol K	m^2 kg s^{-2} K^{-1} mol^{-1}
Exposure (x-rays and γ-rays)	coulomb per kilogram	C/kg	kg^{-1} s A
Absorbed dose rate	gray per second	Gy/s	m^2 s^{-3}
Compressibility	per pascal	Pa^{-1}	m kg^{-1} s^2
Elastic moduli	newton per square metre	N/m^2 or N m^{-2}	kg m^{-1} s^{-2}
Pressure gradient	pascal per metre	Pa/m or N m^{-3}	kg m^{-2} s^{-2}
Surface potential	joule per kilogram	J/kg or N m/kg	m^2 s^{-2}
Pressure energy	pascal cubic metre	Pa m^3 or N m	kg m^2 s^{-2}
Impulse	newton second	N s	kg m s^{-1}
Angular impulse	newton metre second	N m s	kg m^2 s^{-1}
Specific resistance	ohm metre	Ωm	kg m^3 s^{-3} A^{-2}
Surface energy	joule per square metre	J/m^2 or N/m	kg s^{-2}

APPENDIX A 7
GENERAL GUIDELINES FOR USING SYMBOLS FOR PHYSICAL QUANTITIES, CHEMICAL ELEMENTS AND NUCLIDES

- Symbols for physical quantities are normally single letters and printed in italic (or sloping) type. However, in case of the two letter symbols, appearing as a factor in a product, some spacing is necessary to separate this symbol from other symbols.
- Abbreviations, i.e., shortened forms of names or expressions, such as p.e. for potential energy, are not used in physical equations. These abbreviations in the text are written in ordinary normal/roman (upright) type.
- Vectors are printed in bold and normal/roman (upright) type. However, in class room situations, vectors may be indicated by an arrow on the top of the symbol.
- Multiplication or product of two physical quantities is written with some spacing between them. Division of one physical quantity by another may be indicated with a horizontal bar or with

solidus, a slash or a short oblique stroke mark (/) or by writing it as a product of the numerator and the inverse first power of the denominator, using brackets at appropriate places to clearly distinguish between the numerator and the denominator.

- Symbols for chemical elements are written in normal/roman (upright) type. The symbol is not followed by a full stop.
  For example, Ca, C, H, He, U, etc.
- The attached numerals specifying a nuclide are placed as a left subscript (atomic number) and superscript (mass number).
  For example, a U-235 nuclide is expressed as $^{235}_{92}\mathbf{U}$ (with 235 expressing the mass number and 92 as the atomic number of uranium with chemical symbol U).
- The right superscript position is used, if required, for indicating a state of ionisation (in case of ions).
  For example, Ca^{2+}, PO_4^{3-}

APPENDIX A 8
GENERAL GUIDELINES FOR USING SYMBOLS FOR SI UNITS, SOME OTHER UNITS, AND SI PREFIXES

- Symbols for units of physical quantities are printed/written in Normal/Roman (upright) type.
- Standard and recommended symbols for units are written in lower case roman (upright) type, starting with small letters. The shorter designations for units such as kg, m, s, cd, etc., are symbols and not the abbreviations. The unit names are never capitalised. However, the unit symbols are capitalised only if the symbol for a unit is derived from a proper name of scientist, beginning with a capital, normal/roman letter.
  For example, m for the unit 'metre', d for the unit 'day', atm for the unit 'atmospheric pressure', Hz for the unit 'hertz', Wb for the unit 'weber', J for the unit 'joule', A for the unit 'ampere', V for the unit 'volt', etc. The single exception is L, which is the symbol for the unit 'litre'. This exception is made to avoid confusion of the lower case letter l with the Arabic numeral l.
- Symbols for units do not contain any final full stop at the end of recommended letter and remain unaltered in the plural, using only singular form of the unit.
  For example, for a length of 25 centimetres the unit symbol is written as 25 cm and not 25 cms or 25 cm. or 25 cms., etc.
- Use of solidus (/) is recommended only for indicating a division of one letter unit symbol by another unit symbol. Not more than one solidus is used.
  For example :
  m/s^2 or $m\ s^{-2}$ (with a spacing between m and s^{-2}) but not m/s/s;
  1 Pl =1 N s m^{-2} = 1 N s/m^2 = 1 kg/s m=1 kg m^{-1} s^{-1}, but not 1 kg/m/s;
  J/K mol or J K^{-1} mol^{-1}, but not J/K/mol; etc.
- Prefix symbols are printed in normal/roman (upright) type without spacing between the prefix symbol and the unit symbol. Thus certain approved prefixes written very close to the unit symbol are used to indicate decimal fractions or multiples of a SI unit, when it is inconveniently small or large.
  For example :

megawatt (1MW = 10^6 W);	nanosecond (1 ns = 10^{-9} s);
centimetre (1 cm = 10^{-2} m);	picofarad (1 pF = 10^{-12} F);.
kilometre (1 km = 10^3 m);	microsecond (1μs = 10^{-6} s);
millivolt (1 mV= 10^{-3} V);	gigahertz (1GHz = 10^9 Hz);

kilowatt-hour (1 kW h = 10^3 W h = 3.6 MJ = 3.6×10^6 J);
microampere (1μ A = 10^{-6} A); micron (1μm = 10^{-6} m);
angstrom (1 Å =0.1 nm = 10^{-10} m); etc.

The unit 'micron' which equals 10^{-6} m, i.e. a micrometre, is simply the name given to convenient sub-multiple of the metre. In the same spirit, the unit 'fermi', equal to a femtometre or 10^{-15} m has been used as the convenient length unit in nuclear studies. Similarly, the unit 'barn', equal to 10^{-28} m^2, is a convenient measure of cross-sectional areas in sub-atomic particle collisions. However, the unit 'micron' is preferred over the unit 'micrometre' to avoid confusion of the 'micrometre' with the length measuring instrument called 'micrometer'. These newly formed multiples or sub-multiples (cm, km, μm, μs, ns) of SI units, metre and second, constitute a new composite inseparable symbol for units.

- When a prefix is placed before the symbol of a unit, the combination of prefix and symbol is considered as a new symbol, for the unit, which can be raised to a positive or negative power without using brackets. These can be combined with other unit symbols to form compound unit. Rules for binding-in indices are not those of ordinary algebra.
  For example :
  cm^3 means always $(cm)^3 = (0.01\ m)^3 = (10^{-2}\ m)^3 = 10^{-6}\ m^3$, but never 0.01 m^3 or $10^{-2}\ m^3$ or 1cm^3 (prefix c with a spacing with m^3 is meaningless as prefix c is to be attached to a unit symbol and it has no physical significance or independent existence without attachment with a unit symbol).
  Similarly, mA^2 means always $(mA)^2 = (0.001A)^2 = (10^{-3}\ A)^2 = 10^{-6}\ A^2$, but never 0.001 A^2 or $10^{-3}\ A^2$ or m A^2;
  1 $cm^{-1} = (10^{-2}m)^{-1} = 10^2\ m^{-1}$, but not 1c m^{-1} or $10^{-2}\ m^{-1}$;
  1μs^{-1} means always $(10^{-6}s)^{-1} = 10^6\ s^{-1}$, but not $1 \times 10^{-6}\ s^{-1}$;
  1 km^2 means always $(km)^2 = (10^3\ m)^2 = 10^6\ m^2$, but not $10^3\ m^2$;
  1mm^2 means always $(mm)^2 = (10^{-3}\ m)^2 = 10^{-6}\ m^2$, but not $10^{-3}\ m^2$.
- A prefix is never used alone. It is always attached to a unit symbol and written or fixed before (pre-fix) the unit symbol.
  For example :
  $10^3/m^3$ means $1000/m^3$ or 1000 m^{-3}, but not k/m^3 or k m^{-3}.
  $10^6/m^3$ means $10,00,000/m^3$ or 10,00,000 m^{-3}, but not M/m^3 or M m^{-3}
- Prefix symbol is written very close to the unit symbol without spacing between them, while unit symbols are written separately with spacing when units are multiplied together.
  For example :
  m s^{-1} (symbols m and s^{-1}, in lower case, small letter m and s, are separate and independent unit symbols for metre and second respectively, with spacing between them) means 'metre per second', but not 'milli per second'.
  Similarly, ms^{-1} [symbol m and s are written very close to each other, with prefix symbol m (for prefix milli) and unit symbol s, in lower case, small letter (for unit 'second') without any spacing between them and making ms as a new composite unit] means 'per millisecond', but never 'metre per second'.
  mS^{-1}[symbol m and S are written very close to each other, with prefix symbol m (for prefix milli) and unit symbol S, in capital roman letter S (for unit 'siemens') without any spacing between them, and making mS as a new composite unit] means 'per millisiemens', but never 'per millisecond'.
  C m [symbol C and m are written separately, representing unit symbols C (for unit 'coulomb') and m (for unit 'metre'), with spacing between them] means 'coulomb metre', but never 'centimetre', etc.
- The use of double prefixes is avoided when single prefixes are available.
  For example :

10^{-9} m = 1nm (nanometre), but not 1mμm (millimicrometre),
10^{-6} m= 1μm (micron), but not 1mmm(millimillimetre),
10^{-12} F= 1 pF (picofarad), but not 1μμF (micromicrofarad),
10^{9} W=1 GW (giga watt), but not 1 kMW (kilomegawatt), etc.

- The use of a combination of unit and the symbols for units is avoided when the physical quantity is expressed by combining two or more units.
  For example :
  joule per mole kelvin is written as J/mol K or J mol^{-1} K^{-1}, but not joule/mole K or J/ mol kelvin or J/mole K, etc.
  joule per tesla is written as J/T or J T^{-1}, but not joule /T or J per tesla or J/tesla, etc.
  newton metre second is written as N m s, but not Newton m second or N m second or N metre s or newton metre s, etc.
  joule per kilogram kelvin is written as J/kg K or J kg^{-1} K^{-1}, but not J/kilog K or joule/kg K or J/ kg kelvin or J/kilogram K, etc.
- To simplify calculations, the prefix symbol is attached to the unit symbol in the numerator and not to the denominator.
  For example :
  10^{6} N/m^{2} is written more conveniently as MN/m^{2}, in preference to N/mm^{2}.
  A preference has been expressed for multiples or sub-multiples involving the factor 1000, $10^{\pm 3n}$ where n is the integer.
- Proper care is needed when same symbols are used for physical quantities and units of physical quantities.
  For example :
  The physical quantity weight (*W*) expressed as a product of mass (*m*) and acceleration due to gravity (*g*) may be written in terms of symbols *W*, *m* and *g* printed in italic (or sloping) type as *W* = *m g*, preferably with a spacing between *m* and *g*. It should not be confused with the unit symbols for the units watt (W), metre (m) and gram (g). However, in the equation *W*=*m g*, the symbol *W* expresses the weight with a unit symbol J, *m* as the mass with a unit symbol kg and *g* as the acceleration due to gravity with a unit symbol m/s^{2}. Similarly, in equation *F* = *m a*, the symbol *F* expresses the force with a unit symbol N, *m* as the mass with a unit symbol kg, and *a* as the acceleration with a unit symbol m/s^{2}. These symbols for physical quantities should not be confused with the unit symbols for the units 'farad' (F), 'metre'(m) and 'are' (a).
  Proper distinction must be made while using the symbols h (prefix hecto, and unit hour), c (prefix centi, and unit carat), d (prefix deci and unit day), T (prefix tera, and unit tesla), a (prefix atto, and unit are), da (prefix deca, and unit deciare), etc.
- SI base unit 'kilogram' for mass is formed by attaching SI prefix (a multiple equal to 10^{3}) 'kilo' to a cgs (centimetre, gram, second) unit 'gram' and this may seem to result in an anomaly. Thus, while a thousandth part of unit of length (metre) is called a millimetre (mm), a thousandth part of the unit of mass (kg) is not called a millikilogram, but just a gram. This appears to give the impression that the unit of mass is a gram (g) which is not true. Such a situation has arisen because we are unable to replace the name 'kilogram' by any other suitable unit. Therefore, as an exception, name of the multiples and sub-multiples of the unit of mass are formed by attaching prefixes to the word 'gram' and not to the word 'kilogram'.
  For example :
  10^{3} kg =1 megagram (1Mg), but not 1 kilo kilogram (1 kkg);
  10^{-6} kg = 1 milligram (1 mg), but not 1 microkilogram (1μkg);
  10^{-3} kg = 1 gram (1g), but not 1 millikilogram (1 mkg), etc.

It may be emphasised again that you should use the internationally approved and recommended symbols only. Continual practice of following general rules and guidelines in unit symbol writing would make you learn mastering the correct use of SI units, prefixes and related symbols for physical quantities in a proper perspective.

APPENDIX A 9
DIMENSIONAL FORMULAE OF PHYSICAL QUANTITIES

S.No	Physical quantity	Relationship with other physical quantities	Dimensions	Dimensional formula
1.	Area	Length × breadth	$[L^2]$	$[M^0 L^2 T^0]$
2.	Volume	Length × breadth × height	$[L^3]$	$[M^0 L^3 T^0]$
3.	Mass density	Mass/volume	$[M]/[L^3]$ or $[M L^{-3}]$	$[M L^{-3} T^0]$
4.	Frequency	1/time period	$1/[T]$	$[M^0 L^0 T^{-1}]$
5.	Velocity, speed	Displacement/time	$[L]/[T]$	$[M^0 LT^{-1}]$
6.	Acceleration	Velocity /time	$[LT^{-1}]/[T]$	$[M^0LT^{-2}]$
7.	Force	Mass × acceleration	$[M][LT^{-2}]$	$[M LT^{-2}]$
8.	Impulse	Force × time	$[M LT^{-2}][T]$	$[M LT^{-1}]$
9.	Work, Energy	Force × distance	$[MLT^{-2}] [L]$	$[M L^2 T^{-2}]$
10.	Power	Work/time	$[ML^2 T^{-2}]/ [T]$	$[M L^2 T^{-3}]$
11.	Momentum	Mass × velocity	$[M] [LT^{-1}]$	$[M LT^{-1}]$
12.	Pressure, stress	Force/area	$[M LT^{-2}]/[L^2]$	$[ML^{-1}T^{-2}]$
13.	Strain	$\frac{\text{Change in dimension}}{\text{Oringinal dimension}}$	$[L] / [L]$ or $[L^3] / [L^3]$	$[M^oL^oT^o]$
14.	Modulus of elasticity	Stress/strain	$\frac{[ML^{-1}T^{-2}]}{[M^0L^0T^0]}$	$[M L^{-1} T^{-2}]$
15	Surface tension	Force/length	$[MLT^{-2}]/[L]$	$[ML^0 T^{-2}]$
16.	Surface energy	Energy/area	$[ML^2 T^{-2}]/[L^2]$	$[ML^0T^{-2}]$
17.	Velocity gradient	Velocity/distance	$[LT^{-1}]/[L]$	$[M^0L^0T^{-1}]$
18.	Pressure gradient	Pressure/distance	$[ML^{-1}T^{-2}]/[L]$	$[ML^{-2}T^{-2}]$
19.	Pressure energy	Pressure × volume	$[ML^{-1}T^{-2}] [L^3]$	$[ML^2 T^{-2}]$
20.	Coefficient of viscosity	Force/area × velocity gradient	$\frac{[MLT^{-2}]}{[L^2][LT^{-1} / L]}$	$[ML^{-1} T^{-1}]$
21.	Angle, Angular displacement	Arc/radius	$[L]/[L]$	$[M^0L^0T^0]$
22.	Trigonometric ratio ($\sin\theta$, $\cos\theta$, $\tan\theta$, etc.)	Length/length	$[L]/[L]$	$[M^0L^0T^0]$
23.	Angular velocity	Angle/time	$[L^0]/[T]$	$[M^0L^0T^{-1}]$

24.	Angular acceleration	Angular velocity/time	$[T^{-1}]/[T]$	$[M^0L^0T^{-2}]$
25.	Radius of gyration	Distance	$[L]$	$[M^0LT^0]$
26.	Moment of inertia	Mass × (radius of gyration)^2	$[M]\,[L^2]$	$[ML^2\,T^0]$
27.	Angular momentum	Moment of inertia × angular velocity	$[ML^2]\,[T^{-1}]$	$[ML^2\,T^{-1}]$
28.	Moment of force, moment of couple	Force x distance	$[MLT^{-2}]\,[L]$	$[ML^2\,T^{-2}]$
29.	Torque	Angular momentum/time, Or Force × distance	$[ML^2\,T^{-1}]\,/\,[T]$ or $[MLT^{-2}]\,[L]$	$[ML^2\,T^{-2}]$
30.	Angular frequency	2π × Frequency	$[T^{-1}]$	$[M^0L^0T^{-1}]$
31.	Wavelength	Distance	$[L]$	$[M^0LT^0]$
32.	Hubble constant	Recession speed/distance	$[LT^{-1}]/[L]$	$[M^0L^0T^{-1}]$
33.	Intensity of wave	(Energy/time)/area	$[ML^2\,T^{-2}/T]/[L^2]$	$[ML^0T^{-3}]$
34.	Radiation pressure	$\frac{\text{Intensity of wave}}{\text{Speed of light}}$	$[MT^{-3}]/[LT^{-1}]$	$[ML^{-1}\,T^{-2}]$
35.	Energy density	Energy/volume	$[ML^2\,T^{-2}]/\,[L^3]$	$[ML^{-1}\,T^{-2}]$
36.	Critical velocity	$\frac{\text{Reynold's number} \times \text{coefficient of viscocity}}{\text{Mass density} \times \text{radius}}$	$\frac{[M^0L^0T^0][ML^{-1}\,T^{-1}]}{[ML^{-3}][L]}$	$[M^0LT^{-1}]$
37.	Escape velocity	(2 × acceleration due to gravity × earth's radius)$^{1/2}$	$[LT^{-2}]^{1/2}$ x $[L]^{1/2}$	$[M^0LT^{-1}]$
38.	Heat energy, internal energy	Work (= Force × distance)	$[MLT^{-2}]\,[L]$	$[ML^2\,T^{-2}]$
39.	Kinetic energy	(1/2) mass × (velocity)^2	$[M]\,[LT^{-1}]^2$	$[ML^2T^{-2}]$
40	Potential energy	Mass × acceleration due to gravity × height	$[M]\,[LT^{-2}]\,[L]$	$[ML^2\,T^{-2}]$
41.	Rotational kinetic energy	½ × moment of inertia × (angular velocity)^2	$[M^0L^0T^0]\,[ML^2]x[T^{-1}]^2$	$[M\,L^2\,T^{-2}]$
42.	Efficiency	$\frac{\text{Output work or energy}}{\text{Input work or energy}}$	$\frac{[ML^2\,T^{-2}]}{[ML^2\,T^{-2}]}$	$[M^0L^0T^0]$
43.	Angular impulse	Torque × time	$[ML^2\,T^{-2}]\,[T]$	$[M\,L^2\,T^{-1}]$
44.	Gravitational constant	$\frac{\text{Force} \times (\text{distance})^2}{\text{mass} \times \text{mass}}$	$\frac{[MLT^{-2}][L^2]}{[M]\,[M]}$	$[M^{-1}L^3T^{-2}]$
45.	Planck constant	Energy/frequency	$[ML^2\,T^{-2}]\,/[T^{-1}]$	$[ML^2\,T^{-1}]$

46.	Heat capacity, entropy	Heat energy / temperature	$[ML^2 T^{-2}]/[K]$	$[ML^2T^{-2}K^{-1}]$
47.	Specific heat capacity	$\frac{\text{Heat Energy}}{\text{Mass} \times \text{temperature}}$	$[ML^2 T^{-2}]/[M]\,[K]$	$[M^0L^2 T^{-2} K^{-1}]$
48.	Latent heat	Heat energy/mass	$[ML^2 T^{-2}]/[M]$	$[M^0L^2 T^{-2}]$
49.	Thermal expansion coefficient or Thermal expansivity	$\frac{\text{Change in dimension}}{\text{Original dimension} \times \text{temperature}}$	$[L]/[L][K]$	$[M^0L^0K^{-1}]$
50.	Thermal conductivity	$\frac{\text{Heat energy} \times \text{thickness}}{\text{Area} \times \text{temperature} \times \text{time}}$	$\frac{[ML^2T^{-2}][L]}{[L^2]\,[K]\,[T]}$	$[MLT^{-3} K^{-1}]$
51.	Bulk modulus or (compressibility)$^{-1}$	$\frac{\text{Volume} \times \text{(change in pressure)}}{\text{(change in volume)}}$	$\frac{[L^3]\,[ML^{-1}T^{-2}]}{[L^3]}$	$[ML^{-1} T^{-2}]$
52.	Centripetal acceleration	$(\text{Velocity})^2/\text{radius}$	$[LT^{-1}]^2/[L]$	$[M^0 LT^{-2}]$
53.	Stefan constant	$\frac{(\text{Energy / area} \times \text{time})}{(\text{Temperature})^4}$	$\frac{[ML^2 T^{-2}]}{[L^2]\,[T]\,[K]^4}$	$[ML^0 T^{-3}K^{-4}]$
54.	Wien constant	Wavelength × temperature	$[L]\,[K]$	$[M^0 LT^0K]$
55.	Boltzmann constant	Energy/temperature	$[ML^2 T^{-2}]/[K]$	$[ML^2 T^{-2} K^{-1}]$
56.	Universal gas constant	$\frac{\text{Pressure} \times \text{volume}}{\text{mole} \times \text{temperature}}$	$\frac{[ML^{-1} T^{-2}][L^3]}{[mol]\,[K]}$	$[ML^2 T^{-2} K^{-1} mol^{-1}]$
57.	Charge	Current × time	$[A]\,[T]$	$[M^0 L^0TA]$
58.	Current density	Current /area	$[A]/[L^2]$	$[M^0L^{-2} T^0A]$
59.	Voltage, electric potential, electromotive force	Work/charge	$[ML^2T^{-2}]/[AT]$	$[ML^2 T^{-3} A^{-1}]$
60.	Resistance	$\frac{\text{Potential difference}}{\text{Current}}$	$\frac{[ML^2 T^{-3} A^{-1}]}{[A]}$	$[ML^2 T^{-3} A^{-2}]$
61.	Capacitance	Charge/potential difference	$\frac{[AT]}{[ML^2 T^{-3} A^{-1}]}$	$[M^{-1}L^{-2} T^4 A^2]$
62.	Electrical resistivity or (electrical conductivity)$^{-1}$	$\frac{\text{Resistance} \times \text{area}}{\text{length}}$	$[ML^2 T^{-3} A^{-2}]$ $[L^2]/[L]$	$[ML^3 T^{-3} A^{-2}]$
63.	Electric field	Electrical force/charge	$[MLT^{-2}]/[AT]$	$[MLT^{-3} A^{-1}]$
64.	Electric flux	Electric field × area	$[MLT^{-3}A^{-1}][L^2]$	$[ML^3 T^{-3} A^{-1}]$

65.	Electric dipole moment	Torque/electric field	$\frac{[ML^2\ T^{-2}]}{[MLT^{-3}\ A^{-1}]}$	$[M^0\ LTA]$
66.	Electric field strength or electric intensity	$\frac{\text{Potential difference}}{\text{distance}}$	$\frac{[ML^2\ T^{-3}\ A^{-1}]}{[L]}$	$[MLT^{-3}\ A^{-1}]$
67.	Magnetic field, magnetic flux density, magnetic induction	$\frac{\text{Force}}{\text{Current} \times \text{length}}$	$[MLT^{-2}]/[A]\ [L]$	$[ML^0\ T^{-2}\ A^{-1}]$
68.	Magnetic flux	Magnetic field × area	$[MT^{-2}\ A^{-2}]\ [L^2]$	$[ML^2\ T^{-2}\ A^{-1}]$
69.	Inductance	$\frac{\text{Magnetic flux}}{\text{Current}}$	$\frac{[ML^2\ T^{-2}\ A^{-1}]}{[A]}$	$[ML^2\ T^{-2}\ A^{-2}]$
70.	Magnetic dipole moment	Torque/magnetic field or current × area	$[ML^2\ T^{-2}]\ /\ [MT^{-2}\ A^{-1}]$ or $[A]\ [L^2]$	$[M^0L^2T^0A]$
71.	Magnetic field strength, magnetic intensity or magnetic moment density	$\frac{\text{Magnetic moment}}{\text{Volume}}$	$\frac{[L^2A]}{[L^3]}$	$[M^0L^{-1}T^0A]$
72	Permittivity constant (of free space)	$\frac{\text{Charge} \times \text{charge}}{4\pi \times \text{electric force} \times (\text{distance})^2}$	$\frac{[AT][AT]}{[MLT^{-2}][L]^2}$	$[M^{-1}L^{-3}T^4\ A^2]$
73.	Permeability constant (of free space)	$\frac{2\pi \times \text{force} \times \text{distance}}{\text{current} \times \text{current} \times \text{length}}$	$\frac{[M^0\ L^0\ T^0][M\ L\ T^{-2}][L]}{[A][A][L]}$	$[MLT^{-2}\ A^{-2}]$
74.	Refractive index	$\frac{\text{Speed of light in vacuum}}{\text{Speed of light in medium}}$	$[LT^{-1}]/LT^{-1}]$	$[M^0L^0T^0]$
75.	Faraday constant	Avogadro constant × elementary charge	$[AT]/[mol]$	$[M^0L^0TA\ mol^{-1}]$
76.	Wave number	2π/wavelength	$[M^0L^0T^0]\ /\ [L]$	$[M^0L^{-1}\ T^0]$
77.	Radiant flux, Radiant power	Energy emitted/time	$[ML^2T^{-2}]/[T]$	$[ML^2T^{-3}]$
78.	Luminosity of radiant flux or radiant intensity	$\frac{\text{Radiant power or radiant flus of source}}{\text{Solid angle}}$	$[ML^2T^{-3}]\ /\ [M^0L^0T^0]$	$[ML^2\ T^{-3}]$
79.	Luminous power or luminous flux of source	$\frac{\text{Luminous energy emitted}}{\text{time}}$	$[ML^2\ T^{-2}]/[T]$	$[ML^2\ T^{-3}]$

80.	Luminous intensity or illuminating power of source	$\frac{\text{Luminous flux}}{\text{Soild angle}}$	$\frac{[ML^2T^{-3}]}{[M^0L^0T^0]}$	$[ML^2T^{-3}]$
81.	Intensity of illumination or luminance	$\frac{\text{Luminous intensity}}{(\text{distance})^2}$	$[ML^2T^{-3}]/[L^2]$	$[ML^0T^{-3}]$
82.	Relative luminosity	$\frac{\text{Luminous flux of a source of given wavelength}}{\text{luminous flux of peak sensitivity wavelength (555 nm) source of same power}}$	$\frac{[ML^2T^{-1}]}{[ML^2T^{-3}]}$	$[M^0L^0T^0]$
83.	Luminous efficiency	$\frac{\text{Total luminous flux}}{\text{Total radiant flux}}$	$[ML^2T^{-3}] / [ML^2T^{-3}]$	$[M^0L^0T^0]$
84.	Illuminance or illumination	$\frac{\text{Luminous flux incident}}{\text{area}}$	$[ML^2T^{-3}]/[L^2]$	$[ML^0T^{-3}]$
85.	Mass defect	(sum of masses of nucleons)-(mass of the nucleus)	$[M]$	$[ML^0T^0]$
86.	Binding energy of nucleus	Mass defect × (speed of light in vacuum)^2	$[M]\,[LT^{-1}]^2$	$[ML^2T^{-2}]$
87.	Decay constant	0.693/half life	$[T^{-1}]$	$[M^0L^0T^{-1}]$
88.	Resonant frequency	$(\text{Inductance} \times \text{capacitance})^{-\frac{1}{2}}$	$[ML^2T^{-2}A^{-2}]^{-\frac{1}{2}}$ x $[M^{-1}L^{-2}T^4A^2]^{-\frac{1}{2}}$	$[M^0L^0A^0T^{-1}]$
89.	Quality factor or Q-factor of coil	$\frac{\text{Resonant frequency} \times \text{inductance}}{\text{Resistance}}$	$\frac{[T^{-1}][ML^2T^{-2}A^{-2}]}{[ML^2T^{-3}A^{-2}]}$	$[M^0L^0T^0]$
90.	Power of lens	$(\text{Focal length})^{-1}$	$[L^{-1}]$	$[M^0L^{-1}T^0]$
91.	Magnification	$\frac{\text{Image distance}}{\text{Object distance}}$	$[L]/[L]$	$[M^0L^0T^0]$
92.	Fluid flow rate	$\frac{(\pi/8)(\text{pressure})\times(\text{radius})^4}{(\text{viscosity coefficient})\times(\text{length})}$	$\frac{[ML^{-1}T^{-2}]\,[L^4]}{[ML^{-1}T^{-1}]\,[L]}$	$[M^0L^3T^{-1}]$
93	Capacitive reactance	$(\text{Angular frequency} \times \text{capacitance})^{-1}$	$[T^{-1}]^{-1}[M^{-1}L^{-2}T^4A^2]^{-1}$	$[ML^2T^{-3}A^{-2}]$
94.	Inductive reactance	(Angular frequency × inductance)	$[T^{-1}][ML^2T^{-2}A^{-2}]$	$[ML^2T^{-3}A^{-2}]$

ANSWERS

Chapter 2

2.1 (a) 10^{-6} ; (b) 1.5×10^{4} ; (c) 5 ; (d) 11.3, 1.13×10^{4}.

2.2 (a) 10^{7} ; (b) 10^{-16} ; (c) 3.9×10^{4} ; (d) 6.67×10^{-8}.

2.5 500

2.6 (c)

2.7 0.035 mm

2.9 94.1

2.10 (a) 1 ; (b) 3 ; (c) 4 ; (d) 4 ; (e) 4 ; (f) 4.

2.11 8.72 m^2; 0.0855 m^3

2.12 (a) 2.3 kg ; (b) 0.02 g

2.13 13%; 3.8

2.14 (b) and (c) are wrong on dimensional grounds. Hint: The argument of a trigonometric function must always be dimensionless.

2.15 The correct formula is $m = m_0 (1 - v^2/c^2)^{-½}$

2.16 $\cong 3 \times 10^{-7}$ m^3

2.17 $\cong 10^{4}$; intermolecular separation in a gas is much larger than the size of a molecule.

2.18 Near objects make greater angle than distant (far off) objects at the eye of the observer. When you are moving, the angular change is less for distant objects than nearer objects. So, these distant objects seem to move along with you, but the nearer objects in opposite direction.

2.19 $\cong 3 \times 10^{16}$ m; as a unit of length 1 parsec is defined to be equal to 3.084×10^{16} m.

2.20 1.32 parsec; 2.64″ (second of arc)

2.23 1.4×10^{3} kg m^{-3}; the mass density of the Sun is in the range of densities of liquids / solids and *not* gases. This high density arises due to inward gravitational attraction on outer layers due to inner layers of the Sun.

2.24 1.429×10^{5} km

2.25 Hint: tan θ must be dimensionless. The correct formula is tan $\theta = v/v'$ where v' is the speed of rainfall.

2.26 Accuracy of 1 part in 10^{11} to 10^{12}

2.27 $\cong 0.7 \times 10^3$ kg m^{-3}. In the solid phase atoms are tightly packed, so the atomic mass density is close to the mass density of the solid.

2.28 $\cong 0.3 \times 10^{18}$ m^{-3} – Nuclear density is typically 10^{15} times atomic density of matter.

2.29 3.84×10^8 m

2.30 55.8 km

2.31 2.8×10^{22} km

2.32 3,581 km

2.33 Hint: the quantity $e^4/(16\pi^2 \epsilon_0^2 m_p m_e^2 c^3 G)$ has the dimension of time.

Chapter 3

3.1 (a), (b)

3.2 (a) A....B, (b) A....B, (c) B....A, (d) Same, (e) B....A....once.

3.4 37 s

3.5 1000 km/h

3.6 3.06 m s^{-2}; 11.4 s

3.7 1250 m (Hint: view the motion of B relative to A)

3.8 1 m s^{-2} (Hint: view the motion of B and C relative to A)

3.9 T = 9 min, speed = 40 km/h. Hint: $vT/(v-20) = 18$; $vT/(v+20) = 6$

3.10 (a) Vertically downwards; (b) zero velocity, acceleration of 9.8 m s^{-2} downwards; (c) $x > 0$ (upward and downward motion); $v < 0$ (upward), $v > 0$ (downward), $a > 0$ throughout; (d) 44.1 m, 6 s.

3.11 (a) True;, (b) False; (c) True (if the particle rebounds instantly with the same speed, it implies infinite acceleration which is unphysical); (d) False (true only when the chosen positive direction is along the direction of motion)

3.14 (a) 5 km h^{-1}, 5 km h^{-1}; (b) 0, 6 km h^{-1}; (c) $\frac{15}{8}$ km h^{-1}, $\frac{45}{8}$ km h^{-1}

3.15 Because, for an arbitrarily small interval of time, the magnitude of displacement is equal to the length of the path.

3.16 All the four graphs are impossible. (a) a particle cannot have two different positions at the same time; (b) a particle cannot have velocity in opposite directions at the same time; (c) speed is always non-negative; (d) total path length of a particle can never decrease with time. (Note, the arrows on the graphs are meaningless).

3.17 No, wrong. x-t plot does not show the trajectory of a particle. Context: A body is dropped from a tower ($x = 0$) at $t = 0$.

3.18 105 m s^{-1}

3.19 (a) A ball at rest on a smooth floor is kicked, it rebounds from a wall with reduced speed and moves to the opposite wall which stops it; (b) A ball thrown up with some initial velocity rebounding from the floor with reduced speed after each hit; (c) A uniformly moving cricket ball turned back by hitting it with a bat for a very short time-interval.

3.20 $x < 0$, $v < 0$, $a > 0$; $x > 0$, $v > 0$, $a < 0$; $x < 0$, $v > 0$, $a > 0$.

3.21 Greatest in 3, least in 2; $v > 0$ in 1 and 2, $v < 0$ in 3.

3.22 Acceleration magnitude greatest in 2; speed greatest in 3; $v > 0$ in 1, 2 and 3; $a > 0$ in 1 and 3, $a < 0$ in 2; $a = 0$ at A, B, C, D.

3.23 A straight line inclined with the time-axis for uniformly accelerated motion; parallel to the time- axis for uniform motion.

3.24 10 s, 10 s

3.25 (a) 13 km h^{-1} ; (b) 5 km h^{-1}; (c) 20 s in either direction, viewed by any one of the parents, the speed of the child is 9 km h^{-1} in either direction; answer to (c) is unaltered.

3.26 $x_2 - x_1 = 15\ t$ (linear part); $x_2 - x_1 = 200 + 30\ t - 5\ t^2$ (curved part).

3.27 (a) 60 m, 6 m s^{-1} ; (b) 36 m, 9 m s^{-1}

3.28 (c), (d), (f)

Chapter 4

4.1 Volume, mass, speed, density, number of moles, angular frequency are scalars; the rest are vectors.

4.2 Work, current

4.3 Impulse

4.4 Only (c) and (d) are permissible

4.5 (a) T, (b) F, (c) F, (d) T, (e) T

4.6 Hint: The sum (difference) of any two sides of a triangle is never less (greater) than the third side. Equality holds for collinear vectors.

4.7 All statements except (a) are correct

4.8 400 m for each; B

4.9 (a) O; (b) O; (c) 21.4 km h^{-1}

4.10 Displacement of magnitude 1 km and direction 60° with the initial direction; total path length = 1.5 km (third turn); null displacement vector; path length = 3 km (sixth turn); 866 m, 30°, 4 km (eighth turn)

4.11 (a) 49.3 km h^{-1} ; (b) 21.4 km h^{-1}. No, the average speed equals average velocity magnitude only for a straight path.

4.12 About 18° with the vertical, towards the south.

4.13 15 min, 750 m

4.14 East (approximately)

4.15 150.5 m

4.16 50 m

4.17 9.9 m s^{-2}, along the radius at every point towards the centre.

4.18 6.4 g

4.19 (a) False (true only for uniform circular motion)

(b) True, (c) True.

4.20 (a) $\mathbf{v}(t) = (3.0\,\hat{\mathbf{i}} - 4.0t\,\hat{\mathbf{j}})$ $\hat{\mathbf{a}}(t) = -4.0\hat{\mathbf{j}}$

(b) 8.54 m s^{-1}, 70° with x-axis.

4.21 (a) 2 s, 24 m, 21.26 m s^{-1}

4.22 $\sqrt{2}$, 45° with the x-axis; $\sqrt{2}$, – 45° with the x- axis, $\left(5/\sqrt{2},\ -1/\sqrt{2}\right)$.

4.23 (b) and (e)

4.24 Only (e) is true

4.25 182 m s^{-1}

4.27 No. Rotations in *general* cannot be associated with vectors

4.28 A vector can be associated with a plane area

4.29 No

4.30 At an angle of $\sin^{-1}(1/3) = 19.5°$ with the vertical; 16 km.

4.31 0.86 m s^{-2}, 54.5° with the direction of velocity

Chapter 5

5.1 (a) to (d) No net force according to the First Law
(e) No force, since it is far away from all material agencies producing electromagnetic and gravitational forces.

5.2 The only force in each case is the force of gravity, (neglecting effects of air) equal to 0.5 N vertically downward. The answers do not change, even if the motion of the pebble is not along the vertical. The pebble is not at rest at the highest point. It has a constant horizontal component of velocity throughout its motion.

5.3 (a) 1 N vertically downwards (b) same as in (a)
(c) same as in (a); force at an instant depends on the situation at that instant, not on history.
(d) 0.1 N in the direction of motion of the train.

5.4 (i) T

5.5 $a = -2.5$ m s^{-2}. Using $v = u + at$, $0 = 15 - 2.5\,t$ i.e., $t = 6.0$ s

5.6 $a = 1.5/25 = 0.06$ m s^{-2}
$F = 3 \times 0.06 = 0.18$ N in the direction of motion.

5.7 Resultant force = 10 N at an angle of $\tan^{-1}(3/4) = 37°$ with the direction of 8 N force. Acceleration = 2 m s^{-2} in the direction of the resultant force.

5.8 $a = -2.5$ m s^{-2}, Retarding force = $465 \times 2.5 = 1.2 \times 10^3$ N

5.9 $F - 20{,}000 \times 10 = 20000 \times 5.0$, i.e., $F = 3.0 \times 10^5$ N

5.10 $a = -20$ m s^{-2} $0 \le t \le 30$ s

$t = -5$ s : $x = u\,t = -10 \times 5 = -50$ m

$t = 25$ s : $x = u\,t + (½)\ a\,t^2 = (10 \times 25 - 10 \times 625)$m $= -6$ km

$t = 100$ s : First consider motion up to 30 s

$x_1 = 10 \times 30 - 10 \times 900 = -8700$ m

At $t = 30$ s, $v = 10 - 20 \times 30 = -590$ m s^{-1}

For motion from 30 s to 100 s : $x_2 = -590 \times 70 = -41300$ m

$x = x_1 + x_2 = -50$ km

5.11 (a) Velocity of car (at $t = 10$ s) $= 0 + 2 \times 10 = 20$ m s^{-1}

By the First Law, the horizontal component of velocity is 20 m s^{-1} throughout.
Vertical component of velocity (at $t = 11$s) $= 0 + 10 \times 1 = 10$ m s^{-1}

Velocity of stone (at $t = 11$s) $= \sqrt{20^2 + 10^2} = \sqrt{500} = 22.4$ m s^{-1} at an angle of tan^{-1} (½) with the horizontal.
(b) 10 m s^{-2} vertically downwards.

5.12 (a) At the extreme position, the speed of the bob is zero. If the string is cut, it will fall vertically downwards.
(b) At the mean position, the bob has a horizontal velocity. If the string is cut, it will fall along a parabolic path.

5.13 The reading on the scale is a measure of the force on the floor by the man. By the Third Law, this is equal and opposite to the normal force N on the man by the floor.

(a) $N = 70 \times 10 = 700$ N ; Reading is 70 kg
(b) $70 \times 10 - N = 70 \times 5$; Reading is 35 kg
(c) $N - 70 \times 10 = 70 \times 5$; Reading is 105 kg
(d) $70 \times 10 - N = 70 \times 10$; Reading would be zero; the scale would read zero.

5.14 (a) In all the three intervals, acceleration and, therefore, force are zero.
(b) 3 kg m s^{-1} at $t = 0$; (c) -3 kg m s^{-1} at $t = 4$ s.

5.15 If the 20 kg mass is pulled,

$600 - T = 20\,a$, $T = 10\,a$

$a = 20$ m s^{-2}, $T = 200$ N

If the 10 kg mass is pulled, a = 20 m s^{-2}, $T = 400$ N

5.16 $T - 8 \times 10 = 8\,a, 12 \times 10 - T = 12a$
i.e. $a = 2$ m s^{-2}, $T = 96$ N

5.17 By momentum conservation principle, total final momentum is zero. Two momentum vectors cannot sum to a null momentum unless they are equal and opposite.

5.18 Impulse on each ball = 0.05 ×12 = 0.6 kg m s^{-1} in magnitude. The two impulses are opposite in direction.

5.19 Use momentum conservation : $100\,v = 0.02 \times 80$
$v = 0.016$ m s^{-1} = 1.6 cm s^{-1}

5.20 Impulse is directed along the bisector of the initial and final directions. Its magnitude is $0.15 \times 2 \times 15 \times \cos 22.5° = 4.2$ kg m s^{-1}

5.21 $v \quad 2 \quad 1.5 \quad \frac{40}{60} \quad 2$ m s^{-1}

$T \quad \frac{mv^2}{R} \quad \frac{0.25 \quad 4 \quad ^2}{1.5} \quad 6.6$ N

$200 \quad \frac{mv_{max}^2}{R}$, which gives $v_{max} \quad 35 \text{ m s}^{-1}$

5.22 Alternative (b) is correct, according to the First Law

5.23 (a) The horse-cart system has no external force in empty space. The mutual forces between the horse and the cart cancel (Third Law). On the ground, the contact force between the system and the ground (friction) causes their motion from rest.

(b) Due to inertia of the body not directly in contact with the seat.

(c) A lawn mower is pulled or pushed by applying force at an angle. When you push, the normal force (N) must be more than its weight, for equilibrium in the vertical direction. This results in greater friction f ($f \propto N$) and, therefore, a greater applied force to move. Just the opposite happens while pulling.

(d) To reduce the rate of change of momentum and hence to reduce the force necessary to stop the ball.

5.24 A body with a constant speed of 1 cm s^{-1} receives impulse of magnitude 0.04 kg × 0.02 m s^{-1} = 8×10^{-4} kg m s^{-1} after every 2 s from the walls at $x = 0$ and $x = 2$ cm.

5.25 Net force = 65 kg × 1 m s^{-2} = 65 N

$a_{max} = \mu_s g = 2 \text{ m s}^{-2}$

5.26 Alternative (a) is correct. Note $mg + T_2 = m\boldsymbol{v}_2^2/R$; $T_1 - mg = m\boldsymbol{v}_1^2/R$

The moral is : do not confuse the actual material forces on a body (tension, gravitational force, etc) with the effects they produce : centripetal acceleration $\boldsymbol{v}_2^2/R$ or $\boldsymbol{v}_1^2/R$ in this example.

5.27 (a) 'Free body' : crew and passengers

Force on the system by the floor = F upwards; weight of system = mg downwards;

$\therefore F - mg = ma$

$F - 300 \times 10 = 300 \times 15$

$F = 7.5 \times 10^3$ N upward

By the Third Law, force on the floor by the crew and passengers = 7.5×10^3N downwards.

(b) 'Free body' : helicopter plus the crew and passengers

Force by air on the system = R upwards; weight of system = mg downwards

$\therefore R - mg = ma$

$R - 1300 \times 10 = 1300 \times 15$

$R = 3.25 \times 10^4$ N upwards

By the Third Law, force (action) on the air by the helicopter = 3.25×10^4 N downwards.

(c) 3.25×10^4 N upwards

5.28 Mass of water hitting the wall per second

$= 10^3 \text{ kg m}^{-3} \times 10^{-2} \text{ m}^2 \times 15 \text{ m s}^{-1} = 150 \text{ kg s}^{-1}$

Force by the wall = momentum loss of water per second = 150 kg s^{-1} × 15 m s^{-1} = 2.25 × 10^3 N

5.29 (a) 3 mg (down) (b) 3 mg (down) (c) 4 mg (up)

5.30 If N is the normal force on the wings,

$N\cos\theta = mg, \qquad N\sin\theta = \dfrac{mv^2}{R}$

which give $R = \dfrac{v^2}{g \tan\theta} = \dfrac{200 \times 200}{10 \times \tan 15°} = 15\text{km}$

5.31 The centripetal force is provided by the lateral thrust by the rail on the flanges of the wheels. By the Third Law, the train exerts an equal and opposite thrust on the rail causing its wear and tear.

Angle of banking = $\tan^{-1}\left(\dfrac{v^2}{R\,g}\right) = \tan^{-1}\left(\dfrac{15 \times 15}{30 \times 10}\right) \square\ 37°$

5.32 Consider the forces on the man in equilibrium : his weight, force due to the rope and normal force due to the floor.

(a) 750 N (b) 250 N; mode (b) should be adopted.

5.33 (a) $T - 400 = 240$, $T = 640$ N

(b) $400 - T = 160$, $T = 240$ N

(c) $T = 400$ N

(*d*) $T = 0$

The rope will break in case (a).

5.34 We assume perfect contact between bodies A and B and the rigid partition. In that case, the self-adjusting normal force on B by the partition (reaction) equals 200 N. There is no impending motion and no friction. The action-reaction forces between A and B are also 200 N. When the partition is removed, kinetic friction comes into play.

Acceleration of A + B = [200 – (150 × 0.15)] / 15 =11.8 m s^{-2}

Friction on A = 0.15 × 50 = 7.5 N

200 – 7.5 – F_{AB} = 5 × 11.8

$F_{AB} = 1.3 \times 10^2$ N; opposite to motion .

$F_{BA} = 1.3 \times 10^2$ N; in the direction of motion.

5.35 (a) Maximum frictional force possible for opposing impending relative motion between the block and the trolley = 150 × 0.18 = 27 N, which is more than the frictional force of 15 × 0.5 = 7.5 N needed to accelerate the box with the trolley. When the trolley moves with uniform velocity, there is no force of friction acting on the block.

(b) For the accelerated (non-inertial) observer, frictional force is opposed by the pseudo-force of the same magnitude, keeping the box at rest relative to the observer. When the trolley moves with uniform velocity there is no pseudo-force for the moving (inertial) observer and no friction.

5.36 Acceleration of the box due to friction = μg = 0.15 × 10 = 1.5 m s^{-2}. But the acceleration of the truck is greater. The acceleration of the box relative to the truck is 0.5 m s^{-2} towards the rear end. The time taken for the box to fall off the truck = $\sqrt{\dfrac{2\ \ 5}{0.5}}\ \ \sqrt{20}$ s .

During this time, the truck covers a distance = ½ × 2 × 20 = 20 m.

5.37 For the coin to revolve with the disc, the force of friction should be enough to provide the necessary centripetal force, i.e $\frac{mv^2}{r} \le \mu m g$. Now $v = r\omega$, where $\omega = \frac{2\pi}{T}$ is the angular frequency of the disc. For a given μ and ω, the condition is $r \le \mu g / \omega^2$. The condition is satisfied by the nearer coin (4 cm from the centre).

5.38 At the uppermost point, $N + mg = \frac{mv^2}{R}$, where N is the normal force (downwards) on the motorcyclist by the ceiling of the chamber. The minimum possible speed at the uppermost point corresponds to $N = 0$.

i.e. $v_{min} \quad \sqrt{Rg} \quad \sqrt{25 \quad 10} \quad 16\,\mathrm{m\,s^{-1}}$

5.39 The horizontal force N by the wall on the man provides the needed centripetal force : $N = m R \omega^2$. The frictional force f (vertically upwards) opposes the weight mg. The man remains stuck to the wall after the floor is removed if $mg = f < \mu N$ i.e. $mg < \mu m R \omega^2$. The minimum angular speed of rotation of the cylinder is $\omega_{min} = \sqrt{g / \mu R} = 5\ \mathrm{s}^{-1}$

5.40 Consider the free-body diagram of the bead when the radius vector joining the centre of the wire makes an angle θ with the vertical downward direction. We have $mg = N\cos\theta$ and $m R \sin\theta\ \omega^2 = N \sin\theta$. These equations give $\cos\theta = g/R\omega^2$. Since $\cos\theta \le 1$, the bead remains at its lowermost point for $\omega \le \sqrt{\frac{g}{R}}$.

For $\boldsymbol{\omega} = \sqrt{\frac{2g}{R}}$, $\cos\theta = \frac{1}{2}$ i.e. $\theta = 60^0$.

Chapter 6

6.1 (a) +ve (b) –ve (c) –ve (d) + ve (e) – ve

6.2 (a) 882 J ; (b) –247 J; (c) 635 J ; (d) 635 J;
Work done by the net force on a body equals change in its kinetic energy.

6.3 (a) $x > a$; 0 (c) $x < a$, $x > b$; $-V_1$
(b) $-\infty < x < \infty$; V_1 (d) $-b/2 < x < -a/2$, $a/2 < x < b/2$; $-V_1$

6.5 (a) rocket; (b) For a conservative force work done over a path is minus of change in potential energy. Over a complete orbit, there is no change in potential energy; (c) K.E. increases, but P.E. decreases, and the sum decreases due to dissipation against friction; (d) in the second case.

6.6 (a) decrease; (b) kinetic energy; (c) external force; (d) total linear momentum, and also total energy (if the system of two bodies is isolated).

6.7 (a) F ; (b) F ; (c) F ; (d) F (true usually but not always, why?)

6.8 (a) No
(b) Yes
(c) Linear momentum is conserved during an inelastic collision, kinetic energy is, of course, not conserved even after the collision is over.
(d) elastic.

6.9 (b) t

6.10 (c)

6.11 12 J

6.12 The electron is faster, $v_e / v_p = 13.5$

6.13 0.082 J in each half ; – 0.163 J

6.14 Yes, momentum of the molecule + wall system is conserved. The wall has a recoil momentum such that the momentum of the wall + momentum of the outgoing molecule equals momentum of the incoming molecule, assuming the wall to be stationary initially. However, the recoil momentum produces negligible velocity because of the large mass of the wall. Since kinetic energy is also conserved, the collision is elastic.

6.15 43.6 kW

6.16 (b)

6.17 It transfers its entire momentum to the ball on the table, and does not rise at all.

6.18 5.3 m s^{-1}

6.19 27 km h^{-1} (no change in speed)

6.20 50 J

6.21 (a) $\dot{m}$ $Av t$ (b) $\dot{K}$ $Av^3 t/2$ (c) P = 4.5 kW

6.22 (a) 49,000 J (b) 6.45 10^{-3} kg

6.23 (a) 200 m^2(b) comparable to the roof of a large house of dimension 14m × 14m.

6.24 21.2 cm, 28.5 J

6.25 No, the stone on the steep plane reaches the bottom earlier; yes, they reach with the same speed v, [since $mgh = (1/2)\, m v^2$]

$v_B = v_C = 14.1$ m s^{-1} , $t_B = 2\sqrt{2}$ s , $t_C = 2\sqrt{2}$ s

6.26 0.125

6.27 8.82 J for both cases.

6.28 The child gives an impulse to the trolley at the start and then runs with a constant relative velocity of 4 m s^{-1} with respect to the trolley's new velocity. Apply momentum conservation for an observer outside. 10.36 m s^{-1}, 25.9 m.

6.29 All except (V) are impossible.

Chapter 7

7.1 The geometrical centre of each. No, the CM may lie outside the body, as in case of a ring, a hollow sphere, a hollow cylinder, a hollow cube etc.

7.2 Located on the line joining H and C1 nuclei at a distance of 1.24 Å from the H end.

7.3 The speed of the CM of the (trolley + child) system remains unchanged (equal to v) because no external force acts on the system. The forces involved in running on the trolley are internal to this system.

7.6 $l_z = xp_y - yp_x$, $l_x = yp_z - zp_y$, $l_y = zp_x - xp_z$

7.8 72 cm

7.9 3675 N on each front wheel, 5145 N on each back wheel.

7.10 (a) 7/5 MR^2 (b) 3/2 MR^2

7.11 Sphere

7.12 Kinetic Energy = 3125 J; Angular Momentum = 62.5 J s

7.13 (a) 100 rev/min (use angular momentum conservation).

(b) The new kinetic energy is 2.5 times the initial kinetic energy of rotation. The child uses his internal energy to increase his rotational kinetic energy.

7.14 25 s^{-2}; 10 m s^{-2}

7.15 36 kW

7.16 at R/6 from the center of original disc opposite to the center of cut portion.

7.17 66.0 g

7.18 (a) Yes; (b) Yes, (c) the plane with smaller inclination ($\because a \alpha \sin\theta$)

7.19 4J

7.20 6.75×10^{12} rad s^{-1}

7.21 (a) 3.8 m (b) 3.0 s

7.22 Tension = 98 N, N_B = 245 N, N_C = 147 N.

7.23 (a) 59 rev/min, (b) No, the K.E. is increased and it comes from work done by man in the process.

7.24 0.625 rad s^{-1}

7.27 (a) By angular momentum conservation, the common angular speed

$$\omega = (I_1 \omega_1 + I_2 \omega_2) / (I_1 + I_2)$$

(b) The loss is due to energy dissipation in frictional contact which brings the two discs to a common angular speed ω. However, since frictional torques are internal to the system, angular momentum is unaltered.

7.28 Velocity of A = $\omega_o R$ in the same direction as the arrow; velocity of B = $\omega_o R$ in the opposite direction to the arrow; velocity of C = $\omega_o R/2$ in the same direction as the arrow. The disc will not roll on a frictionless plane.

7.29 (a) Frictional force at B opposes velocity of B. Therefore, frictional force is in the same direction as the arrow. The sense of frictional torque is such as to oppose angular motion. ω_o and τ are both normal to the paper, the first into the paper, and the second coming out of the paper.

(b) Frictional force decreases the velocity of the point of contact B. Perfect rolling ensues when this velocity is zero. Once this is so, the force of friction is zero.

7.30 Frictional force causes the CM to accelerate from its initial zero velocity. Frictional torque causes retardation in the initial angular speed ω_o. The equations of motion are : $\mu_k m g = m a$ and $\mu_k m g R = -I\alpha$, which yield $v = \mu_k g t$, $\omega = \omega_o - \mu_k m g R t / I$. Rolling begins when $v = R\omega$. For a ring, $I = m R^2$, and rolling begins at $t = \omega_o R/2 \mu_k g$. For a disc, $I = ½ m R^2$ and rolling starts at break line $t = R\omega_o/3 \mu_k g$. Thus, the disc begins to roll earlier than the ring, for the same R and ω_o. The actual times can be obtained for R = 10 cm, $\omega_o = 10\pi$ rad s^{-1}, $\mu_k = 0.2$

7.31 (a) 16.4 N

(b) Zero

(c) 37° approx.

Chapter 8

8.1 (a) No.
(b) Yes, if the size of the space ship is large enough for him to detect the variation in g.
(c) Tidal effect depends inversely on the cube of the distance unlike force, which depends inversely on the square of the distance.

8.2 (a) decreases; (b) decreases; (c) mass of the body; (d) more.

8.3 Smaller by a factor of 0.63.

8.5 3.54×10^8 years.

8.6 (a) Kinetic energy, (b) less,

8.7 (a) No, (b) No, (c) No, (d) Yes

[The escape velocity is independent of mass of the body and the direction of projection. It depends upon the gravitational potential at the point from where the body is launched. Since this potential depends (slightly) on the latitude and height of the point, the escape velocity (speed) depends (slightly) on these factors.]

8.8 All quantities vary over an orbit except angular momentum and total energy.

8.9 (b), (c) and (d)

8.10 and **8.11** For these two problems, complete the hemisphere to sphere. At both P, and C, potential is constant and hence intensity = 0. Therefore, for the hemisphere, (c) and (e) are correct.

8.12 2.6×10^8 m

8.13 2.0×10^{30} kg

8.14 1.43×10^{12} m

8.15 28 N

8.16 125 N

8.17 8.0×10^6 m from the earth's centre

8.18 31.7 km/s

8.19 5.9×10^9 J

8.20 2.6×10^6 m/s

8.21 0, 2.7×10^{-8} J/kg; an object placed at the mid point is in an unstable equilibrium

8.22 -9.4×10^6 J/kg

8.23 $GM / R^2 = 2.3 \times 10^{12}$ m s^{-2} , $\omega^2 R = 1.1 \times 10^6$ m s^{-2}; here ω is the angular speed of rotation. Thus in the rotating frame of the star, the inward force is much greater than the outward centrifugal force at its equator. The object will remain stuck (and not fly off due to centrifugal force). Note, if angular speed of rotation increases say by a factor of 2000, the object will fly off.

8.24 3×10^{11} J

8.25 495 km

Physics

Part II

Textbook for Class XI

Physics Part II Textbook for Class XI

CONTENTS

Foreword iii
Preface vii
A Note for the Teachers x

CHAPTER 9

Mechanical Properties of Solids

9.1 Introduction 231
9.2 Elastic behaviour of solids 232
9.3 Stress and strain 232
9.4 Hooke's law 234
9.5 Stress-strain curve 234
9.6 Elastic moduli 235
9.7 Applications of elastic behaviour of materials 240

CHAPTER 10

Mechanical Properties of Fluids

10.1 Introduction 246
10.2 Pressure 246
10.3 Streamline flow 253
10.4 Bernoulli's principle 254
10.5 Viscosity 258
10.6 Reynolds number 260
10.7 Surface tension 261

CHAPTER 11

Thermal Properties of Matter

11.1 Introduction 274
11.2 Temperature and heat 274
11.3 Measurement of temperature 275
11.4 Ideal-gas equation and absolute temperature 275
11.5 Thermal expansion 276
11.6 Specific heat capacity 280
11.7 Calorimetry 281
11.8 Change of state 282
11.9 Heat transfer 286
11.10 Newton's law of cooling 290

CHAPTER 12

Thermodynamics

12.1 Introduction 298
12.2 Thermal equilibrium 299

12.3 Zeroth law of thermodynamics 300
12.4 Heat, internal energy and work 300
12.5 First law of thermodynamics 302
12.6 Specific heat capacity 303
12.7 Thermodynamic state variables and equation of state 304
12.8 Thermodynamic processes 305
12.9 Heat engines 308
12.10 Refrigerators and heat pumps 308
12.11 Second law of thermodynamics 309
12.12 Reversible and irreversible processes 310
12.13 Carnot engine 311

CHAPTER 13

KINETIC THEORY

13.1 Introduction 318
13.2 Molecular nature of matter 318
13.3 Behaviour of gases 320
13.4 Kinetic theory of an ideal gas 323
13.5 Law of equipartition of energy 327
13.6 Specific heat capacity 328
13.7 Mean free path 330

CHAPTER 14

OSCILLATIONS

14.1 Introduction 336
14.2 Periodic and oscilatory motions 337
14.3 Simple harmonic motion 339
14.4 Simple harmonic motion and uniform circular motion 341
14.5 Velocity and acceleration in simple harmonic motion 343
14.6 Force law for simple harmonic motion 345
14.7 Energy in simple harmonic motion 346
14.8 Some systems executing SHM 347
14.9 Damped simple harmonic motion 351
14.10 Forced oscillations and resonance 353

CHAPTER 15

WAVES

15.1 Introduction 363
15.2 Transverse and longitudinal waves 365
15.3 Displacement relation in a progressive wave 367
15.4 The speed of a travelling wave 369
15.5 The principle of superposition of waves 373
15.6 Reflection of waves 374

15.7 Beats 379
15.8 Doppler effect 381

ANSWERS 391

BIBLIOGRAPHY 401

INDEX 403

CHAPTER TEN

MECHANICAL PROPERTIES OF FLUIDS

10.1 Introduction
10.2 Pressure
10.3 Streamline flow
10.4 Bernoulli's principle
10.5 Viscosity
10.6 Reynolds number
10.7 Surface tension
Summary
Points to ponder
Exercises
Additional exercises
Appendix

10.1 INTRODUCTION

In this chapter, we shall study some common physical properties of liquids and gases. Liquids and gases can flow and are therefore, called fluids. It is this property that distinguishes liquids and gases from solids in a basic way.

Fluids are everywhere around us. Earth has an envelop of air and two-thirds of its surface is covered with water. Water is not only necessary for our existence; every mammalian body constitute mostly of water. All the processes occurring in living beings including plants are mediated by fluids. Thus understanding the behaviour and properties of fluids is important.

How are fluids different from solids? What is common in liquids and gases? Unlike a solid, a fluid has no definite shape of its own. Solids and liquids have a fixed volume, whereas a gas fills the entire volume of its container. We have learnt in the previous chapter that the volume of solids can be changed by stress. The volume of solid, liquid or gas depends on the stress or pressure acting on it. When we talk about fixed volume of solid or liquid, we mean its volume under atmospheric pressure. The difference between gases and solids or liquids is that for solids or liquids the change in volume due to change of external pressure is rather small. In other words solids and liquids have much lower compressibility as compared to gases.

Shear stress can change the shape of a solid keeping its volume fixed. The key property of fluids is that they offer very little resistance to shear stress; their shape changes by application of very small shear stress. The shearing stress of fluids is about million times smaller than that of solids.

10.2 PRESSURE

A sharp needle when pressed against our skin pierces it. Our skin, however, remains intact when a blunt object with a wider contact area (say the back of a spoon) is pressed against it with the same force. If an elephant were to step on a man's chest, his ribs would crack. A circus performer across whose

chest a large, light but strong wooden plank is placed first, is saved from this accident. Such everyday experiences convince us that both the force and its coverage area are important. Smaller the area on which the force acts, greater is the impact. This concept is known as pressure.

When an object is submerged in a fluid at rest, the fluid exerts a force on its surface. This force is always normal to the object's surface. This is so because if there were a component of force parallel to the surface, the object will also exert a force on the fluid parallel to it; as a consequence of Newton's third law. This force will cause the fluid to flow parallel to the surface. Since the fluid is at rest, this cannot happen. Hence, the force exerted by the fluid at rest has to be perpendicular to the surface in contact with it. This is shown in Fig.10.1(a).

The normal force exerted by the fluid at a point may be measured. An idealised form of one such pressure-measuring device is shown in Fig. 10.1(b). It consists of an evacuated chamber with a spring that is calibrated to measure the force acting on the piston. This device is placed at a point inside the fluid. The inward force exerted by the fluid on the piston is balanced by the outward spring force and is thereby measured.

Fig. 10.1 (a) The force exerted by the liquid in the beaker on the submerged object or on the walls is normal (perpendicular) to the surface at all points.
(b) An idealised device for measuring pressure.

If F is the magnitude of this normal force on the piston of area A then the **average pressure** P_{av} is defined as the normal force acting per unit area.

$$P_{av} = \frac{F}{A} \tag{10.1}$$

In principle, the piston area can be made arbitrarily small. The pressure is then defined in a limiting sense as

$$P = \lim_{\Delta A \to 0} \frac{F}{A} \tag{10.2}$$

Pressure is a scalar quantity. We remind the reader that it is the component of the force normal to the area under consideration and not the (vector) force that appears in the numerator in Eqs. (10.1) and (10.2). Its dimensions are $[ML^{-1}T^{-2}]$. The SI unit of pressure is N m^{-2}. It has been named as pascal (Pa) in honour of the French scientist Blaise Pascal (1623-1662) who carried out pioneering studies on fluid pressure. A common unit of pressure is the atmosphere (atm), i.e. the pressure exerted by the atmosphere at sea level (1 atm = 1.013×10^5 Pa).

Another quantity, that is indispensable in describing fluids, is the density ρ. For a fluid of mass m occupying volume V,

$$\rho = \frac{m}{V} \tag{10.3}$$

The dimensions of density are $[ML^{-3}]$. Its SI unit is kg m^{-3}. It is a positive scalar quantity. A liquid is largely incompressible and its density is therefore, nearly constant at all pressures. Gases, on the other hand exhibit a large variation in densities with pressure.

The density of water at 4°C (277 K) is 1.0×10^3 kg m^{-3}. The relative density of a substance is the ratio of its density to the density of water at 4°C. It is a dimensionless positive scalar quantity. For example the relative density of aluminium is 2.7. Its density is 2.7×10^3 kg m^{-3}. The densities of some common fluids are displayed in Table 10.1.

Table 10.1 Densities of some common fluids at STP*

Fluid	ρ (kg m^{-3})
Water	1.00×10^3
Sea water	1.03×10^3
Mercury	13.6×10^3
Ethyl alcohol	0.806×10^3
Whole blood	1.06×10^3
Air	1.29
Oxygen	1.43
Hydrogen	9.0×10^{-2}
Interstellar space	$\approx 10^{-20}$

* *STP means standard temperature (0°C) and 1 atm pressure.*

Example 10.1 The two thigh bones (femurs), each of cross-sectional area 10 cm^2 support the upper part of a human body of mass 40 kg. Estimate the average pressure sustained by the femurs.

Answer Total cross-sectional area of the femurs is $A = 2 \times 10\ cm^2 = 20 \times 10^{-4}\ m^2$. The force acting on them is F = 40 kg wt = 400 N (taking g = 10 m s^{-2}). This force is acting vertically down and hence, normally on the femurs. Thus, the average pressure is

$$P_{av} = \frac{F}{A} = 2 \times 10^5\ \text{N m}^{-2}$$

10.2.1 Pascal's Law

The French scientist Blaise Pascal observed that the pressure in a fluid at rest is the same at all points if they are at the same height. This fact may be demonstrated in a simple way.

Fig. 10.2 Proof of Pascal's law. ABC-DEF is an element of the interior of a fluid at rest. This element is in the form of a right-angled prism. The element is small so that the effect of gravity can be ignored, but it has been enlarged for the sake of clarity.

Fig. 10.2 shows an element in the interior of a fluid at rest. This element ABC-DEF is in the form of a right-angled prism. In principle, this prismatic element is very small so that every part of it can be considered at the same depth from the liquid surface and therefore, the effect of the gravity is the same at all these points. But for clarity we have enlarged this element. The forces on this element are those exerted by the rest of the fluid and they must be normal to the surfaces of the element as discussed above. Thus, the fluid exerts pressures P_a, P_b and P_c on this element of area corresponding to the normal forces F_a, F_b and F_c as shown in Fig. 10.2 on the faces BEFC, ADFC and ADEB denoted by A_a, A_b and A_c respectively. Then

$F_b \sin\theta = F_c$, $\quad F_b \cos\theta = F_a$ (by equilibrium)

$A_b \sin\theta = A_c$, $\quad A_b \cos\theta = A_a$ (by geometry)

Thus,

$$\frac{F_b}{A_b} = \frac{F_c}{A_c} = \frac{F_a}{A_a}; \qquad P_b = P_c = P_a \tag{10.4}$$

Hence, pressure exerted is same in all directions in a fluid at rest. It again reminds us that like other types of stress, pressure is not a vector quantity. No direction can be assigned to it. The force against any area within (or bounding) a fluid at rest and under pressure is normal to the area, regardless of the orientation of the area.

Now consider a fluid element in the form of a horizontal bar of uniform cross-section. The bar is in equilibrium. The horizontal forces exerted at its two ends must be balanced or the pressure at the two ends should be equal. This proves that for a liquid in equilibrium the pressure is same at all points in a horizontal plane. Suppose the pressure were not equal in different parts of the fluid, then there would be a flow as the fluid will have some net force acting on it. Hence in the absence of flow the pressure in the fluid must be same everywhere. Wind is flow of air due to pressure differences.

10.2.2 Variation of Pressure with Depth

Consider a fluid at rest in a container. In Fig. 10.3 point 1 is at height h above a point 2. The pressures at points 1 and 2 are P_1 and P_2 respectively. Consider a cylindrical element of fluid having area of base A and height h. As the fluid is at rest the resultant horizontal forces should be zero and the resultant vertical forces should balance the weight of the element. The forces acting in the vertical direction are due to the fluid pressure at the top (P_1A) acting downward, at the bottom (P_2A) acting upward. If mg is weight of the fluid in the cylinder we have

$$(P_2 - P_1)\ A = mg \tag{10.5}$$

Now, if ρ is the mass density of the fluid, we have the mass of fluid to be $m = \rho V = \rho h A$ so that

$$P_2 - P_1 = \rho g h \tag{10.6}$$

Fig.10.3 Fluid under gravity. The effect of gravity is illustrated through pressure on a vertical cylindrical column.

Pressure difference depends on the vertical distance h between the points (1 and 2), mass density of the fluid ρ and acceleration due to gravity g. If the point 1 under discussion is shifted to the top of the fluid (say water), which is open to the atmosphere, P_1 may be replaced by atmospheric pressure (P_a) and we replace P_2 by P. Then Eq. (10.6) gives

$$P = P_a + \rho gh \qquad (10.7)$$

Thus, the pressure P, at depth below the surface of a liquid open to the atmosphere is greater than atmospheric pressure by an amount ρgh. The excess of pressure, $P - P_a$, at depth h is called a **gauge pressure** at that point.

The area of the cylinder is not appearing in the expression of absolute pressure in Eq. (10.7). Thus, the height of the fluid column is important and not cross sectional or base area or the shape of the container. The liquid pressure is the same at all points at the same horizontal level (same depth). The result is appreciated through the example of **hydrostatic paradox**. Consider three vessels A, B and C [Fig.10.4] of different shapes. They are connected at the bottom by a horizontal pipe. On filling with water the level in the three vessels is the same though they hold different amounts of water. This is so, because water at the bottom has the same pressure below each section of the vessel.

Fig 10.4 Illustration of hydrostatic paradox. The three vessels A, B and C contain different amounts of liquids, all upto the same height.

Example 10.2 What is the pressure on a swimmer 10 m below the surface of a lake?

Answer Here

h = 10 m *and* ρ = 1000 kg m^{-3}. Take g = 10 m s^{-2}

From Eq. (10.7)

$P = P_a + \rho gh$

$= 1.01 \times 10^5$ Pa + 1000 kg m$^{-3} \times$ 10 m s$^{-2} \times$ 10 m

$= 2.01 \times 10^5$ Pa

$\approx$ 2 atm

This is a 100% increase in pressure from surface level. At a depth of 1 km the increase in pressure is 100 atm! Submarines are designed to withstand such enormous pressures.

10.2.3 Atmospheric Pressure and Gauge Pressure

The pressure of the atmosphere at any point is equal to the weight of a column of air of unit cross sectional area extending from that point to the top of the atmosphere. At sea level it is 1.013×10^5 Pa (1 atm). Italian scientist Evangelista Torricelli (1608-1647) devised for the first time, a method for measuring atmospheric pressure. A long glass tube closed at one end and filled with mercury is inverted into a trough of mercury as shown in Fig.10.5 (a). This device is known as mercury barometer. The space above the mercury column in the tube contains only mercury vapour whose pressure P is so small that it may be neglected. The pressure inside the column at point A must equal the pressure at point B, which is at the same level. Pressure at B = atmospheric pressure = P_a

$$P_a = \rho gh \qquad (10.8)$$

where ρ is the density of mercury and h is the height of the mercury column in the tube.

In the experiment it is found that the mercury column in the barometer has a height of about 76 cm at sea level equivalent to one atmosphere (1 atm). This can also be obtained using the value of ρ in Eq. (10.8). A common way of stating pressure is in terms of cm or mm of mercury (Hg). A pressure equivalent of 1 mm is called a torr (after Torricelli).

1 torr = 133 Pa.

The mm of Hg and torr are used in medicine and physiology. In meteorology, a common unit is the bar and millibar.

1 bar = 10^5 Pa

An open-tube manometer is a useful instrument for measuring pressure differences. It consists of a U-tube containing a suitable liquid i.e. a low density liquid (such as oil) for measuring small pressure differences and a high density liquid (such as mercury) for large pressure differences. One end of the tube is open to the atmosphere and other end is connected to the system whose pressure we want to measure [see Fig. 10.5 (b)]. The pressure P at A is equal to pressure at point B. What we normally measure is the gauge pressure, which is $P - P_a$, given by Eq. (10.8) and is proportional to manometer height h.

Fig 10.5 (a) The mercury barometer.

(b) the open tube manometer

Fig 10.5 Two pressure measuring devices.

Pressure is same at the same level on both sides of the U-tube containing a fluid. For liquids the density varies very little over wide ranges in pressure and temperature and we can treat it safely as a constant for our present purposes. Gases on the other hand, exhibits large variations of densities with changes in pressure and temperature. Unlike gases, liquids are therefore, largely treated as incompressible.

Example 10.3 The density of the atmosphere at sea level is 1.29 kg/m^3. Assume that it does not change with altitude. Then how high would the atmosphere extend?

Answer We use Eq. (10.7)

$$\rho g h = 1.29\ \text{kg m}^{-3} \times 9.8\ \text{m s}^{2} \times h\ \text{m} = 1.01 \times 10^5\ \text{Pa}$$

$$\therefore h = 7989\ \text{m} \approx 8\ \text{km}$$

In reality the density of air decreases with height. So does the value of g. The atmospheric cover extends with decreasing pressure over 100 km. We should also note that the sea level atmospheric pressure is not always 760 mm of Hg. A drop in the Hg level by 10 mm or more is a sign of an approaching storm.

Example 10.4 At a depth of 1000 m in an ocean (a) what is the absolute pressure? (b) What is the gauge pressure? (c) Find the force acting on the window of area 20 cm × 20 cm of a submarine at this

depth, the interior of which is maintained at sea-level atmospheric pressure. (The density of sea water is 1.03×10^3 kg m^{-3}, g = 10m s^{-2}.)

Answer Here h = 1000 m and $\rho = 1.03 \times 10^3$ kg m^{-3}.

(a) From Eq. (10.6), absolute pressure

$P = P_a + \rho gh$

$= 1.01 \times 10^5$ Pa

$+ 1.03 \times 10^3$ kg m^{-3} $\times$ 10 m s^{-2} $\times$ 1000 m

$= 104.01 \times 10^5$ Pa

$\approx$ 104 atm

(b) Gauge pressure is $P - P_a = \rho gh = P_g$

$P_g = 1.03 \times 10^3$ kg m^{-3} $\times$ 10 ms^2 $\times$ 1000 m

$= 103 \times 10^5$ Pa

$\approx$ 103 atm

(c) The pressure outside the submarine is $P = P_a + \rho gh$ and the pressure inside it is P_a. Hence, the net pressure acting on the window is gauge pressure, $P_g = \rho gh$. Since the area of the window is A = 0.04 m^2, the force acting on it is

$F = P_g A = 103 \times 10^5 \text{Pa} \times 0.04 \text{ m}^2 = 4.12 \times 10^5$ N

t

10.2.4 Hydraulic Machines

Let us now consider what happens when we change the pressure on a fluid contained in a vessel. Consider a horizontal cylinder with a piston and three vertical tubes at different points. The pressure in the horizontal cylinder is indicated by the height of liquid column in the vertical tubes.It is necessarily the same in all. If we push the piston, the fluid level rises in all the tubes, again reaching the same level in each one of them.

This indicates that when the pressure on the cylinder was increased, it was distributed uniformly throughout. We can say **whenever external pressure is applied on any part of a fluid contained in a vessel, it is transmitted undiminished and equally in all directions. This is the Pascal's law for transmission of fluid pressure and has many applications in daily life.**

A number of devices such as **hydraulic lift** and **hydraulic brakes** are based on the Pascal's law. In these devices fluids are used for transmitting pressure. In a hydraulic lift as shown in Fig. 10.6 two pistons are separated by the space filled with a liquid. A piston of small cross section A_1 is used to exert a force F_1 directly on the liquid. The pressure $P = \frac{F_1}{A_1}$ is transmitted throughout the liquid to the larger cylinder attached with a larger piston of area A_2, which results in an upward force of $P \times A_2$. Therefore, the piston is capable of supporting a large force (large weight of, say a car, or a truck, placed on the platform) $F_2 = PA_2 = \frac{F_1 A_2}{A_1}$. By changing the force at A_1, the platform can be

Archemedes' Principle

Fluid appears to provide partial support to the objects placed in it. When a body is wholly or partially immersed in a fluid at rest, the fluid exerts pressure on the surface of the body in contact with the fluid. The pressure is greater on lower surfaces of the body than on the upper surfaces as pressure in a fluid increases with depth. The resultant of all the forces is an upward force called buoyant force. Suppose that a cylindrical body is immersed in the fluid. The upward force on the bottom of the body is more than the downward force on its top. The fluid exerts a resultant upward force or buoyant force on the body equal to $(P_2 - P_1)A$. We have seen in equation 10.4 that $(P_2 - P_1)A = \rho ghA$. Now hA is the volume of the solid and ρhA is the weight of an equivaliant volume of the fluid. $(P_2 - P_1)A = mg$. Thus the upward force exerted is equal to the weight of the displaced fluid.

The result holds true irrespective of the shape of the object and here cylindrical object is considered only for convenience. This is Archimedes' principle. For totally immersed objects the volume of the fluid displaced by the object is equal to its own volume. If the density of the immersed object is more than that of the fluid, the object will sink as the weight of the body is more than the upward thrust. If the density of the object is less than that of the fluid, it floats in the fluid partially submerged. To calculate the volume submerged. Suppose the total volume of the object is V_s and a part V_p of it is submerged in the fluid. Then the upward force which is the weight of the displaced fluid is $\rho_f g V_p$, which must equal the weight of the body; $\rho_s g V_s = \rho_f g V_p$ or $\rho_s / \rho_f = V_p / V_s$ The apparent weight of the floating body is zero.

This principle can be summarised as; 'the loss of weight of a body submerged (partially or fully) in a fluid is equal to the weight of the fluid displaced'.

moved up or down. Thus, the applied force has been increased by a factor of $\frac{A_2}{A_1}$ and this factor is the mechanical advantage of the device. The example below clarifies it.

Fig 10.6 Schematic diagram illustrating the principle behind the hydraulic lift, a device used to lift heavy loads.

Example 10.5 Two syringes of different cross sections (without needles) filled with water are connected with a tightly fitted rubber tube filled with water. Diameters of the smaller piston and larger piston are 1.0 cm and 3.0 cm respectively. (a) Find the force exerted on the larger piston when a force of 10 N is applied to the smaller piston. (b) If the smaller piston is pushed in through 6.0 cm, how much does the larger piston move out?

Answer (a) Since pressure is transmitted undiminished throughout the fluid,

$$F_2 = \frac{A_2}{A_1} F_1 = \frac{3/2 \times 10^{-2}\,\text{m}^2}{1/2 \times 10^{-2}\,\text{m}^2} \times 10\,\text{N}$$

$$= 90\ \text{N}$$

(b) Water is considered to be perfectly incompressible. Volume covered by the movement of smaller piston inwards is equal to volume moved outwards due to the larger piston.

$$L_1 A_1 = L_2 A_2$$

$$\simeq 0.67 \times 10^{-2}\,\text{m} = 0.67\ \text{cm}$$

Note, atmospheric pressure is common to both pistons and has been ignored.

Example 10.6 In a car lift compressed air exerts a force F_1 on a small piston having a radius of 5.0 cm. This pressure is transmitted to a second piston of radius 15 cm (Fig 10.7). If the mass of the car to be lifted is 1350 kg, calculate F_1. What is the pressure necessary to accomplish this task? (g = 9.8 ms^{-2}).

Answer Since pressure is transmitted undiminished throughout the fluid,

$$F_1 = \frac{A_1}{A_2} F_2 = \frac{5 \times 10^{-2}\,\text{m}^2}{15 \times 10^{-2}\,\text{m}^2} \times 1350\ \text{N} \times 9.8\ \text{m s}^{-2}$$

$$= 1470\ \text{N}$$

$$\approx 1.5 \times 10^3\ \text{N}$$

The air pressure that will produce this force is

$$P = \frac{F_1}{A_1} = \frac{1.5 \times 10^3\,\text{N}}{5 \times 10^{-2}\,\text{m}^2} = 1.9 \times 10^5\,\text{Pa}$$

This is almost double the atmospheric pressure.

Hydraulic brakes in automobiles also work on the same principle. When we apply a little

Archimedes (287 – 212 B.C.)

Archimedes was a Greek philosopher, mathematician, scientist and engineer. He invented the catapult and devised a system of pulleys and levers to handle heavy loads. The king of his native city Syracuse, Hiero II asked him to determine if his gold crown was alloyed with some cheaper metal such as silver without damaging the crown. The partial loss of weight he experienced while lying in his bathtub suggested a solution to him. According to legend, he ran naked through the streets of Syracuse exclaiming "Eureka, eureka!", which means "I have found it, I have found it!"

force on the pedal with our foot the master piston moves inside the master cylinder, and the pressure caused is transmitted through the brake oil to act on a piston of larger area. A large force acts on the piston and is pushed down expanding the brake shoes against brake lining. In this way a small force on the pedal produces a large retarding force on the wheel. An important advantage of the system is that the pressure set up by pressing pedal is transmitted equally to all cylinders attached to the four wheels so that the braking effort is equal on all wheels.

10.3 STREAMLINE FLOW

So far we have studied fluids at rest. The study of the fluids in motion is known as fluid dynamics. When a water-tap is turned on slowly, the water flow is smooth initially, but loses its smoothness when the speed of the outflow is increased. In studying the motion of fluids we focus our attention on what is happening to various fluid particles at a particular point in space at a particular time. The flow of the fluid is said to be **steady** if at any given point, the velocity of each passing fluid particle remains constant in time. This does not mean that the velocity at different points in space is same. The velocity of a particular particle may change as it moves from one point to another. That is, at some other point the particle may have a different velocity, but every other particle which passes the second point behaves exactly as the previous particle that has just passed that point. Each particle follows a smooth path, and the paths of the particles do not cross each other.

Fig. 10.7 The meaning of streamlines. (a) A typical trajectory of a fluid particle. (b) A region of streamline flow.

The path taken by a fluid particle under a steady flow is a **streamline**. It is defined as a curve whose tangent at any point is in the direction of the fluid velocity at that point. Consider the path of a particle as shown in Fig.10.7 (a), the curve describes how a fluid particle moves with time. The curve PQ is like a permanent map of fluid flow, indicating how the fluid streams. No two streamlines can cross, for if they do, an oncoming fluid particle can go either one way or the other and the flow would not be steady. Hence, in steady flow, the map of flow is stationary in time. How do we draw closely spaced streamlines ? If we intend to show streamline of every flowing particle, we would end up with a continuum of lines. Consider planes perpendicular to the direction of fluid flow e.g., at three points P, R and Q in Fig.10.7 (b). The plane pieces are so chosen that their boundaries be determined by the same set of streamlines. This means that number of fluid particles crossing the surfaces as indicated at P, R and Q is the same. If area of cross-sections at these points are A_P,A_R and A_Q and speeds of fluid particles are v_P, v_R and v_Q, then mass of fluid Δm_P crossing at A_P in a small interval of time Δt is $\rho_P A_P v_P \Delta t$. Similarly mass of fluid Δm_R flowing or crossing at A_R in a small interval of time Δt is $\rho_R A_R v_R \Delta t$ and mass of fluid Δm_Q is $\rho_Q A_Q v_Q \Delta t$ crossing at A_Q. The mass of liquid flowing out equals the mass flowing in, holds in all cases. Therefore,

$$\rho_P A_P v_P \Delta t = \rho_R A_R v_R \Delta t = \rho_Q A_Q v_Q \Delta t \qquad (10.9)$$

For flow of incompressible fluids

$$\rho_P = \rho_R = \rho_Q$$

Equation (10.9) reduces to

$$A_P v_P = A_R v_R = A_Q v_Q \qquad (10.10)$$

which is called the **equation of continuity** and it is a statement of conservation of mass in flow of incompressible fluids. In general

$$Av = \text{constant} \qquad (10.11)$$

Av gives the volume flux or flow rate and remains constant throughout the pipe of flow. Thus, at narrower portions where the streamlines are closely spaced, velocity increases and its vice versa. From (Fig 10.7b) it is clear that $A_R > A_Q$ or $v_R < v_Q$, the fluid is accelerated while passing from R to Q. This is associated with a change in pressure in fluid flow in horizontal pipes.

Steady flow is achieved at low flow speeds. Beyond a limiting value, called critical speed, this flow loses steadiness and becomes **turbulent**. One sees this when a fast flowing

stream encounters rocks, small foamy whirlpool-like regions called 'white water rapids are formed.

Figure 10.8 displays streamlines for some typical flows. For example, Fig. 10.8(a) describes a laminar flow where the velocities at different points in the fluid may have different magnitudes but their directions are parallel. Figure 10.8 (b) gives a sketch of turbulent flow.

(a) (b)

Fig. 10.8 (a) Some streamlines for fluid flow. (b) A jet of air striking a flat plate placed perpendicular to it. This is an example of turbulent flow.

10.4 BERNOULLI'S PRINCIPLE

Fluid flow is a complex phenomenon. But we can obtain some useful properties for steady or streamline flows using the conservation of energy.

Consider a fluid moving in a pipe of varying cross-sectional area. Let the pipe be at varying heights as shown in Fig. 10.9. We now suppose that an incompressible fluid is flowing through the pipe in a steady flow. Its velocity must change as a consequence of equation of continuity. A force is required to produce this acceleration, which is caused by the fluid surrounding it, the pressure must be different in different regions. Bernoulli's equation is a general expression that relates the pressure difference between two points in a pipe to both velocity changes (kinetic energy change) and elevation (height) changes (potential energy change). The Swiss Physicist Daniel Bernoulli developed this relationship in 1738.

Consider the flow at two regions 1 (i.e. BC) and 2 (i.e. DE). Consider the fluid initially lying between B and D. In an infinitesimal time interval Δt, this fluid would have moved. Suppose v_1 is the speed at B and v_2 at D, then fluid initially at B has moved a distance $v_1\Delta t$ to C ($v_1\Delta t$ is small enough to assume constant cross-section along BC). In the same interval Δt the fluid initially at D moves to E, a distance equal to $v_2\Delta t$. Pressures P_1 and P_2 act as shown on the plane faces of areas A_1 and A_2 binding the two regions. The work done on the fluid at left end (BC) is $W_1 = P_1A_1(v_1\Delta t) = P_1\Delta V$. Since the same volume ΔV passes through both the regions (from the equation of continuity) the work done by the fluid at the other end (DE) is $W_2 = P_2A_2(v_2\Delta t) = P_2\Delta V$ or, the work done on the fluid is $-P_2\Delta V$. So the total work done on the fluid is

$$W_1 - W_2 = (P_1 - P_2)\,\Delta V$$

Part of this work goes into changing the kinetic energy of the fluid, and part goes into changing the gravitational potential energy. If the density of the fluid is ρ and $\Delta m = \rho A_1 v_1 \Delta t = \rho\Delta V$ is the mass passing through the pipe in time Δt, then change in gravitational potential energy is

$$\Delta U = \rho g \Delta V\,(h_2 - h_1)$$

The change in its kinetic energy is

$$\Delta K = \left(\frac{1}{2}\right)\rho\,\Delta V\,(v_2^2 - v_1^2)$$

We can employ the work – energy theorem (Chapter 6) to this volume of the fluid and this yields

$$(P_1 - P_2)\,\Delta V = \left(\frac{1}{2}\right)\rho\,\Delta V\,(v_2^2 - v_1^2) + \rho g \Delta V\,(h_2 - h_1)$$

We now divide each term by ΔV to obtain

$$(P_1 - P_2) = \left(\frac{1}{2}\right)\rho\,(v_2^2 - v_1^2) + \rho g\,(h_2 - h_1)$$

Daniel Bernoulli (1700-1782)

Daniel Bernoulli was a Swiss scientist and mathematician who along with Leonard Euler had the distinction of winning the French Academy prize for mathematics ten times. He also studied medicine and served as a professor of anatomy and botany for a while at Basle, Switzerland. His most well known work was in hydrodynamics, a subject he developed from a single principle: the conservation of energy. His work included calculus, probability, the theory of vibrating strings, and applied mathematics. He has been called the founder of mathematical physics.

We can rearrange the above terms to obtain

$$P_1 + \left(\frac{1}{2}\right)\rho v_1^2 + \rho g h_1 = P_2 + \left(\frac{1}{2}\right)\rho v_2^2 + \rho g h_2 \quad (10.12)$$

This is **Bernoulli's equation**. Since 1 and 2 refer to any two locations along the pipeline, we may write the expression in general as

$$P + \left(\frac{1}{2}\right)\rho v^2 + \rho g h = \text{constant} \quad (10.13)$$

Fig. 10.9 The flow of an ideal fluid in a pipe of varying cross section. The fluid in a section of length $v_1\Delta t$ moves to the section of length $v_2\Delta t$ in time Δt.

In words, the Bernoulli's relation may be stated as follows: As we move along a streamline the sum of the pressure (P), the kinetic energy per unit volume $\frac{v^2}{2}$ and the potential energy per unit volume ($\rho g h$) remains a constant.

Note that in applying the energy conservation principle, there is an assumption that no energy is lost due to friction. But in fact, when fluids flow, some energy does get lost due to internal friction. This arises due to the fact that in a fluid flow, the different layers of the fluid flow with different velocities. These layers exert frictional forces on each other resulting in a loss of energy. This property of the fluid is called viscosity and is discussed in more detail in a later section. The lost kinetic energy of the fluid gets converted into heat energy. Thus, Bernoulli's equation ideally applies to fluids with zero viscosity or non-viscous fluids. Another restriction on application of Bernoulli theorem is that the fluids must be incompressible, as the elastic energy of the fluid is also not taken into consideration. In practice, it has a large number of useful applications and can help explain a wide variety of phenomena for low viscosity incompressible fluids. Bernoulli's equation also does not hold for non-steady or turbulent flows, because in that situation velocity and pressure are constantly fluctuating in time.

When a fluid is at rest i.e. its velocity is zero everywhere, Bernoulli's equation becomes

$$P_1 + \rho g h_1 = P_2 + \rho g h_2$$

$$(P_1 - P_2) = \rho g\,(h_2 - h_1)$$

which is same as Eq. (10.6).

10.4.1 Speed of Efflux: Torricelli's Law

The word efflux means fluid outflow. Torricelli discovered that the speed of efflux from an open tank is given by a formula identical to that of a freely falling body. Consider a tank containing a liquid of density ρ with a small hole in its side at a height y_1 from the bottom (see Fig. 10.10). The air above the liquid, whose surface is at height y_2, is at pressure P. From the equation of continuity [Eq. (10.10)] we have

$$v_1 A_1 = v_2 A_2$$

$$v_2 = \frac{A_1}{A_2} v_1$$

Fig. 10.10 Torricelli's law. The speed of efflux, v_1, from the side of the container is given by the application of Bernoulli's equation. If the container is open at the top to the atmosphere then $v_1 \quad \sqrt{2gh}$.

If the cross sectional area of the tank A_2 is much larger than that of the hole ($A_2 >> A_1$), then we may take the fluid to be approximately at rest at the top, i.e. $v_2 = 0$. Now applying the Bernoulli equation at points 1 and 2 and noting that at the hole $P_1 = P_a$, the atmospheric pressure, we have from Eq. (10.12)

$$P_a \quad \frac{1}{2} \quad v_1^2 \qquad g\,y_1 \quad P \qquad g\,y_2$$

Taking $y_2 - y_1 = h$ we have

$$v_1 \quad \sqrt{2g\,h \quad \frac{2\ P \quad P_a}{}} \qquad (10.14)$$

When $P >> P_a$ and $2\,g\,h$ may be ignored, the speed of efflux is determined by the container pressure. Such a situation occurs in rocket propulsion. On the other hand if the tank is open to the atmosphere, then $P = P_a$ and

$$v_1 = \sqrt{2gh} \qquad (10.15)$$

This is the speed of a freely falling body. Equation (10.15) is known as **Torricelli's law**.

10.4.2 Venturi-meter

The Venturi-meter is a device to measure the flow speed of incompressible fluid. It consists of a tube with a broad diameter and a small constriction at the middle as shown in Fig. (10.11). A manometer in the form of a U-tube is also attached to it, with one arm at the broad neck point of the tube and the other at constriction as shown in Fig. (10.11). The manometer contains a liquid of density ρ_m. The speed v_1 of the liquid flowing through the tube at the broad neck area A is to be measured from equation of continuity Eq. (10.10) the speed at the constriction becomes $v_2 = \frac{A}{a} v_1$.

Then using Bernoulli's equation, we get

$$P_1 + \frac{1}{2} \rho v_1^2 = P_2 + \frac{1}{2} \rho v_1^2 (A/a)^2$$

So that

$$P_1 - P_2 = \frac{1}{2} \rho v_1^2 \left[\left(\frac{A}{a}\right)^2 - 1\right] \qquad (10.16)$$

This pressure difference causes the fluid in the U tube connected at the narrow neck to rise in comparison to the other arm. The difference in height h measure the pressure difference.

Fig. 10.11 A schematic diagram of Venturi-meter.

$$P_1 - P_2 = \rho_m g h = \frac{1}{2} \rho v_1^2 \left[\left(\frac{A}{a}\right)^2 - 1\right]$$

So that the speed of fluid at wide neck is

$$v_1 = \sqrt{\left(\frac{2\rho_m g h}{\rho}\right)} \left(\left(\frac{A}{a}\right)^2 - 1\right)^{-1/2} \qquad (10.17)$$

The principle behind this meter has many applications. The carburetor of automobile has a Venturi channel (nozzle) through which air flows with a large speed. The pressure is then lowered at the narrow neck and the petrol (gasoline) is sucked up in the chamber to provide the correct mixture of air to fuel necessary for combustion. Filter pumps or aspirators, Bunsen burner, atomisers and sprayers [See Fig. 10.12] used for perfumes or to spray insecticides work on the same principle.

Fig. 10.12 The spray gun. Piston forces air at high speeds causing a lowering of pressure at the neck of the container.

Example 10.7 **Blood velocity:** The flow of blood in a large artery of an anesthetised dog is diverted through a Venturi meter. The wider part of the meter has a cross-sectional area equal to that of the artery. A = 8 mm^2. The narrower part has an area a = 4 mm^2. The pressure drop in the artery is 24 Pa. What is the speed of the blood in the artery?

Answer We take the density of blood from Table 10.1 to be 1.06×10^3 kg m^{-3}. The ratio of the areas is $\left(\frac{A}{a}\right) = 2$. Using Eq. (10.17) we obtain

$$v_1 = \sqrt{\frac{2 \times 24 Pa}{1060\,\text{kg m}^{-3} \times \left(2^2 - 1\right)}} = 0.125\,\text{m s}^{-1}$$

10.4.3 Blood Flow and Heart Attack

Bernoulli's principle helps in explaining blood flow in artery. The artery may get constricted due to the accumulation of plaque on its inner walls. In order to drive the blood through this constriction a greater demand is placed on the activity of the heart. The speed of the flow of the blood in this region is raised which lowers the pressure inside and the artery may collapse due to the external pressure. The heart exerts further pressure to open this artery and forces the blood through. As the blood rushes through the opening, the internal pressure once again drops due to same reasons leading to a repeat collapse. This may result in heart attack.

10.4.4 Dynamic Lift

Dynamic lift is the force that acts on a body, such as airplane wing, a hydrofoil or a spinning ball, by virtue of its motion through a fluid. In many games such as cricket, tennis, baseball, or golf, we notice that a spinning ball deviates from its parabolic trajectory as it moves through air. This deviation can be partly explained on the basis of Bernoulli's principle.

(i) **Ball moving without spin**: Fig. 10.13(a) shows the streamlines around a non-spinning ball moving relative to a fluid. From the symmetry of streamlines it is clear that the velocity of fluid (air) above and below the ball at corresponding points is the same resulting in zero pressure difference. The air therefore, exerts no upward or downward force on the ball.

(ii) **Ball moving with spin**: A ball which is spinning drags air along with it. If the surface is rough more air will be dragged. Fig 10.13(b) shows the streamlines of air for a ball which is moving and spinning at the same time. The ball is moving forward and relative to it the air is moving backwards. Therefore, the velocity of air above the ball relative to it is larger and below it is smaller. The stream lines thus get crowded above and rarified below.

This difference in the velocities of air results in the pressure difference between the lower and upper faces and their is a net upward force on the ball. This dynamic lift due to spining is called **Magnus effect**.

Fig 10.13 *(a) Fluid streaming past a static sphere. (b) Streamlines for a fluid around a sphere spinning clockwise. (c) Air flowing past an aerofoil.*

Aerofoil or lift on aircraft wing: Figure 10.13 (c) shows an aerofoil, which is a solid piece shaped to provide an upward dynamic lift when it moves horizontally through air. The cross-section of the wings of an aeroplane looks somewhat like the aerofoil shown in Fig. 10.13 (c) with streamlines around it. When the aerofoil moves against the wind, the orientation of the wing relative to flow direction causes the streamlines to crowd together above the wing more than those below it. The flow speed on top is higher than that below it. There is an upward force resulting in a dynamic lift of the wings and this balances the weight of the plane. The following example illustrates this.

Example 10.8 A fully loaded Boeing aircraft has a mass of 3.3×10^5 kg. Its total wing area is 500 m^2. It is in level flight with a speed of 960 km/h. (a) Estimate the pressure difference between the lower and upper surfaces of the wings (b) Estimate the fractional increase in the speed of the air on the upper surface of the wing relative to the lower surface. [The density of air is ρ = 1.2 kg m^{-3}]

Answer (a) The weight of the Boeing aircraft is balanced by the upward force due to the pressure difference

$\Delta P \times A = 3.3 \times 10^5 \text{ kg} \times 9.8$

$\Delta P = (3.3 \times 10^5 \text{ kg} \times 9.8 \text{ m s}^{-2}) / 500 \text{ m}^2$

$= 6.5 \times 10^3 \text{ Nm}^{-2}$

(b) We ignore the small height difference between the top and bottom sides in Eq. (10.12). The pressure difference between them is then

where v_2 is the speed of air over the upper surface and v_1 is the speed under the bottom surface.

$$v_2 - v_1 \quad \frac{2 \quad P}{v_2 \quad v_1}$$

Taking the average speed

$v_{av} = (v_2 + v_1)/2$ = 960 km/h = 267 m s^{-1},

we have

$$v_2 - v_1 \quad / v_{av} \quad \frac{P}{v_{av}^2} \approx 0.08$$

The speed above the wing needs to be only 8 % higher than that below.

10.5 VISCOSITY

Most of the fluids are not ideal ones and offer some resistance to motion. This resistance to fluid motion is like an internal friction analogous to friction when a solid moves on a surface. It is called viscosity. This force exists when there is relative motion between layers of the liquid. Suppose we consider a fluid like oil enclosed between two glass plates as shown in Fig. 10.15 (a). The bottom plate is fixed while the top plate is moved with a constant velocity **v** relative to the fixed plate. If oil is replaced by honey, a greater force is required to move the plate with the same velocity. Hence we say that honey is more viscous than oil. The fluid in contact with a surface has the same velocity as that of the surfaces. Hence, the layer of the liquid in contact with top surface moves with a velocity **v** and the layer of the liquid in contact with the fixed surface is stationary. The velocities of layers increase uniformly from bottom (zero velocity) to the top layer (velocity **v**). For any layer of liquid, its upper layer pulls it forward while lower layer pulls it backward. This results in force between the layers. This type of flow is known as laminar. The layers of liquid slide over one another as the pages of a book do when it is placed flat on a table and a horizontal force is applied to the top cover. When a fluid is flowing in a pipe or a tube, then velocity of the liquid layer along the axis of the tube is maximum and decreases gradually as we move towards the walls where it becomes zero, Fig. 10.14 (b). The velocity on a cylindrical surface in a tube is constant.

On account of this motion, a portion of liquid, which at some instant has the shape ABCD, take the shape of AEFD after short interval of time (Δt). During this time interval the liquid has undergone a shear strain of $\Delta x/l$. Since, the strain in a flowing fluid increases with time continuously. Unlike a solid, here the stress is found experimentally to depend on 'rate of change of strain' or 'strain rate' i.e. $\Delta x/(l\,\Delta t)$ or v/l instead of strain itself. The coefficient of viscosity (pronounced 'eta') for a fluid is defined as the ratio of shearing stress to the strain rate.

$$\frac{F/A}{v/l} \quad \frac{Fl}{vA} \tag{10.18}$$

The SI unit of viscosity is poiseiulle (Pl). Its other units are N s m^{-2} or Pa s. The dimensions

Fig 10.14 *(a) A layer of liquid sandwiched between two parallel glass plates in which the lower plate is fixed and the upper one is moving to the right with velocity* **v** *(b) velocity distribution for viscous flow in a pipe.*

of viscosity are $[ML^{-1}T^{-1}]$. Generally thin liquids like water, alcohol etc. are less viscous than thick liquids like coal tar, blood, glycerin etc. The coefficients of viscosity for some common fluids are listed in Table 10.2. We point out two facts about blood and water that you may find interesting. As Table 10.2 indicates, blood is 'thicker' (more viscous) than water. Further the relative viscosity (η/η_{water}) of blood remains constant between 0°C and 37°C.

The viscosity of liquids decreases with temperature while it increases in the case of gases.

Example 10.9 A metal block of area 0.10 m² is connected to a 0.010 kg mass via a string that passes over an ideal pulley (considered massless and frictionless), as in Fig. 10.15. A liquid with a film thickness of 0.30 mm is placed between the block and the table. When released the block moves to the right with a constant speed of 0.085 m s⁻¹. Find the coefficient of viscosity of the liquid.

Fig. 10.15 *Measurement of the coefficient of viscosity of a liquid.*

Answer The metal block moves to the right because of the tension in the string. The tension T is equal in magnitude to the weight of the suspended mass m. Thus the shear force F is $F = T = mg = 0.010\ \text{kg} \times 9.8\ \text{m s}^{-2} = 9.8 \times 10^{-2}\ \text{N}$

$$\text{Shear stress on the fluid} = F/A = \frac{9.8 \quad 10^{-2}}{0.10}$$

$$\text{Strain rate} = \frac{\upsilon}{l} \quad \frac{0.085}{0.030}$$

$$\eta = \frac{stress}{strain\ rate}$$

$$\frac{9.8 \quad 10^{-2} N \quad 0.30 \quad 10^{-3} m}{0.085 m\ s^{-1} \quad 0.10 m^2}$$

$$= 3.45 \times 10^{-3}\ \text{Pa s}$$

Table 10.2 The viscosities of some fluids

Fluid	T(°C)	Viscosity (mPl)
Water	20	1.0
	100	0.3
Blood	37	2.7
Machine Oil	16	113
	38	34
Glycerine	20	830
Honey		200
Air	0	0.017
	40	0.019

10.5.1 Stokes' Law

When a body falls through a fluid it drags the layer of the fluid in contact with it. A relative

motion between the different layers of the fluid is set and as a result the body experiences a retarding force. Falling of a raindrop and swinging of a pendulum bob are some common examples of such motion. It is seen that the viscous force is proportional to the velocity of the object and is opposite to the direction of motion. The other quantities on which the force F depends on viscosity η of the fluid and radius a of the sphere. Sir George G. Stokes (1819-1903), an English scientist enunciated clearly the viscous drag force F as

$$F \quad 6 \quad av \qquad (10.19)$$

This is known as Stokes' law.We shall not derive Stokes' law.

This law is an interesting example of retarding force which is proportional to velocity. We can study its consequences on an object falling through a viscous medium. We consider a raindrop in air. It accelerates initially due to gravity. As the velocity increases, the retarding force also increases. Finally when viscous force plus buoyant force becomes equal to force due to gravity, the net force becomes zero and so does the acceleration. The sphere (raindrop) then descends with a constant velocity. Thus in equilibrium, this terminal velocity v_t is given by

$$6\pi\eta a v_t = (4\pi/3)\ a^3\ (\rho - \sigma) g$$

where ρ and σ are mass densities of sphere and the fluid respectively. We obtain

$$v_t = 2a^2\ (\rho - \sigma) g\ /\ (9\eta) \qquad (10.20)$$

So the terminal velocity v_t depends on the square of the radius of the sphere and inversely on the viscosity of the medium.

You may like to refer back to Example 6.2 in this context.

Example 10.10 The terminal velocity of a copper ball of radius 2.0 mm falling through a tank of oil at 20°C is 6.5 cm s^{-1}. Compute the viscosity of the oil at 20°C. Density of oil is 1.5×10^3 kg m^{-3}, density of copper is 8.9×10^3 kg m^{-3}.

Answer We have $v_t = 6.5 \times 10^{-2}$ ms^{-1}, $a = 2 \times 10^{-3}$ m, $g = 9.8$ ms^{-2}, $\rho = 8.9 \times 10^3$ kg m^{-3}, $\sigma = 1.5 \times 10^3$ kg m^{-3}. From Eq. (10.20)

$$\frac{2}{9}\ \frac{2 \quad 10^{-3} \quad \text{m}^2 \quad 9.8\,\text{m s}^{-2}}{6.5 \quad 10^{-2}\ \text{m s}^{-1}} \quad 7.4 \quad 10^3\ \text{kg m}^{-3}$$

$$= 9.9 \times 10^{-1}\ \text{kg m}^{-1}\ \text{s}^{-1}$$

10.6 REYNOLDS NUMBER

When the rate of flow of a fluid is large, the flow no longer remain laminar, but becomes turbulent. In a turbulent flow the velocity of the fluids at any point in space varies rapidly and randomly with time. Some circular motions called eddies are also generated. An obstacle placed in the path of a fast moving fluid causes turbulence [Fig. 10.8 (b)]. The smoke rising from a burning stack of wood, oceanic currents are turbulent. Twinkling of stars is the result of atmospheric turbulence. The wakes in the water and in the air left by cars, aeroplanes and boats are also turbulent.

Osborne Reynolds (1842-1912) observed that turbulent flow is less likely for viscous fluid flowing at low rates. He defined a dimensionless number, whose value gives one an approximate idea whether the flow would be turbulent . This number is called the Reynolds R_e.

$$R_e = \rho v d / \eta \qquad (10.21)$$

where ρ is the density of the fluid flowing with a speed v, d stands for the dimension of the pipe, and η is the viscosity of the fluid. R_e is a dimensionless number and therefore, it remains same in any system of units. It is found that flow is streamline or laminar for R_e less than 1000. The flow is turbulent for $R_e > 2000$. The flow becomes unsteady for R_e between 1000 and 2000. The critical value of R_e (known as critical Reynolds number), at which turbulence sets, is found to be the same for the geometrically similar flows. For example when oil and water with their different densities and viscosities, flow in pipes of same shapes and sizes, turbulence sets in at almost the same value of R_e. Using this fact a small scale laboratory model can be set up to study the character of fluid flow. They are useful in designing of ships, submarines, racing cars and aeroplanes.

R_e can also be written as

$$R_e = \rho v^2\ /\ (\eta v/d) = \rho A v^2\ /\ (\eta A v/d) \qquad (10.22)$$

= inertial force/force of viscosity.

Thus R_e represents the ratio of inertial force (force due to inertia i.e. mass of moving fluid or due to inertia of obstacle in its path) to viscous force.

Turbulence dissipates kinetic energy usually in the form of heat. Racing cars and planes are engineered to precision in order to minimise turbulence. The design of such vehicles involves

experimentation and trial and error. On the other hand turbulence (like friction) is sometimes desirable. Turbulence promotes mixing and increases the rates of transfer of mass, momentum and energy. The blades of a kitchen mixer induce turbulent flow and provide thick milk shakes as well as beat eggs into a uniform texture.

Example 10.11 The flow rate of water from a tap of diameter 1.25 cm is 0.48 L/min. The coefficient of viscosity of water is 10^{-3} Pa s. After sometime the flow rate is increased to 3 L/min. Characterise the flow for both the flow rates.

Answer Let the speed of the flow be v and the diameter of the tap be d = 1.25 cm. The volume of the water flowing out per second is

$Q = v \times \pi d^2 / 4$

$v = 4 Q / d^2$

We then estimate the Reynolds number to be

$R_e = 4 \rho Q / \pi d \eta$

$= 4 \times 10^3 \text{kg m}^{-3} \times Q / (3.14 \times 1.25 \times 10^{-2} \text{m} \times 10^{-3} \text{Pa s})$

$= 1.019 \times 10^8 \text{ m}^{-3} \text{ s } Q$

Since initially

$Q = 0.48 \text{ L / min} = 8 \text{ cm}^3 / \text{s} = 8 \times 10^{-6} \text{ m}^3 \text{ s}^{-1}$,

we obtain,

$R_e = 815$

Since this is below 1000, the flow is steady. After some time when

$Q = 3 \text{ L / min} = 50 \text{ cm}^3 / \text{s} = 5 \times 10^{-5} \text{ m}^3 \text{ s}^{-1}$,

we obtain,

$R_e = 5095$

The flow will be turbulent. You may carry out an experiment in your washbasin to determine the transition from laminar to turbulent flow.

10.7 SURFACE TENSION

You must have noticed that, oil and water do not mix; water wets you and me but not ducks; mercury does not wet glass but water sticks to it, oil rises up a cotton wick, inspite of gravity, Sap and water rise up to the top of the leaves of the tree, hairs of a paint brush do not cling together when dry and even when dipped in water but form a fine tip when taken out of it. All these and many more such experiences are related with the free surfaces of liquids. As liquids have no definite shape but have a definite volume, they acquire a free surface when poured in a container. These surfaces possess some additional energy. This phenomenon is known as surface tension and it is concerned with only liquid as gases do not have free surfaces. Let us now understand this phenomena.

10.7.1 Surface Energy

A liquid stays together because of attraction between molecules. Consider a molecule well inside a liquid. The intermolecular distances are such that it is attracted to all the surrounding molecules [Fig. 10.16(a)]. This attraction results in a negative potential energy for the molecule, which depends on the number and distribution of molecules around the chosen one. But the average potential energy of all the molecules is the same. This is supported by the fact that to take a collection of such molecules (the liquid) and to disperse them far away from each other in order to evaporate or vaporise, the heat of evaporation required is quite large. For water it is of the order of 40 kJ/mol.

Let us consider a molecule near the surface Fig. 10.16(b). Only lower half side of it is surrounded by liquid molecules. There is some negative potential energy due to these, but obviously it is less than that of a molecule in bulk, i.e., the one fully inside. Approximately it is half of the latter. Thus, molecules on a liquid surface have some extra energy in comparison to molecules in the interior. A liquid thus tends to have the least surface area which external conditions permit. Increasing surface area requires energy. Most surface phenomenon can be understood in terms of this fact. What is the energy required for having a molecule at the surface? As mentioned above, roughly it is half the energy required to remove it entirely from the liquid i.e., half the heat of evaporation.

Finally, what is a surface? Since a liquid consists of molecules moving about, there cannot be a perfectly sharp surface. The density of the liquid molecules drops rapidly to zero around $z = 0$ as we move along the direction indicated Fig 10.16 (c) in a distance of the order of a few molecular sizes.

Fig. 10.16 *Schematic picture of molecules in a liquid, at the surface and balance of forces. (a) Molecule inside a liquid. Forces on a molecule due to others are shown. Direction of arrows indicates attraction of repulsion. (b) Same, for a molecule at a surface. (c) Balance of attractive (A) and repulsive (R) forces.*

10.7.2 Surface Energy and Surface Tension

As we have discussed that an extra energy is associated with surface of liquids, the creation of more surface (spreading of surface) keeping other things like volume fixed requires additional energy. To appreciate this, consider a horizontal liquid film ending in bar free to slide over parallel guides Fig (10.17).

Fig. 10.17 *Stretching a film. (a) A film in equilibrium; (b) The film stretched an extra distance.*

Suppose that we move the bar by a small distance d as shown. Since the area of the surface increases, the system now has more energy, this means that some work has been done against an internal force. Let this internal force be **F**, the work done by the applied force is $\mathbf{F}\cdot\mathbf{d} = Fd$. From conservation of energy, this is stored as additional energy in the film. If the surface energy of the film is S per unit area, the extra area is $2dl$. A film has two sides and the liquid in between, so there are two surfaces and the extra energy is

$$S\,(2dl) = Fd \qquad (10.23)$$

$$\text{Or, } S = Fd/2dl = F/2l \qquad (10.24)$$

This quantity S is the magnitude of surface tension. It is equal to the surface energy per unit area of the liquid interface and is also equal to the force per unit length exerted by the fluid on the movable bar.

So far we have talked about the surface of one liquid. More generally, we need to consider fluid surface in contact with other fluids or solid surfaces. The surface energy in that case depends on the materials on both sides of the surface. For example, if the molecules of the materials attract each other, surface energy is reduced while if they repel each other the surface energy is increased. Thus, more appropriately, the surface energy is the energy of the interface between two materials and depends on both of them.

We make the following observations from above:

(i) Surface tension is a force per unit length (or surface energy per unit area) acting in the plane of the interface between the plane of the liquid and any other substance; it also is the extra energy that the molecules at the interface have as compared to molecules in the interior.

(ii) At any point on the interface besides the boundary, we can draw a line and imagine equal and opposite surface tension forces S per unit length of the line acting perpendicular to the line, in the plane of the interface. The line is in equilibrium. To be more specific, imagine a line of atoms or molecules at the surface. The atoms to the left pull the line towards them; those to the right pull it towards them! This line of atoms is in equilibrium under tension. If the line really marks the end of the interface, as in

Figure 10.16 (a) and (b) there is only the force S per unit length acting inwards.

Table 10.3 gives the surface tension of various liquids. The value of surface tension depends on temperature. Like viscosity, the surface tension of a liquid usually falls with temperature.

Table 10.3 Surface tension of some liquids at the temperatures indicated with the heats of the vaporisation

Liquid	Temp (°C)	Surface Tension (N/m)	Heat of vaporisation (kJ/mol)
Helium	–270	0.000239	0.115
Oxygen	–183	0.0132	7.1
Ethanol	20	0.0227	40.6
Water	20	0.0727	44.16
Mercury	20	0.4355	63.2

A fluid will stick to a solid surface if the surface energy between fluid and the solid is smaller than the sum of surface energies between solid-air, and fluid-air. Now there is cohesion between the solid surface and the liquid. It can be directly measured experimentaly as schematically shown in Fig. 10.18. A flat vertical glass plate, below which a vessel of some liquid is kept, forms one arm of the balance. The plate is balanced by weights on the other side, with its horizontal edge just over water. The vessel is raised slightly till the liquid just touches the glass plate and pulls it down a little because of surface tension. Weights are added till the plate just clears water.

Fig. 10.18 Measuring Surface Tension.

Suppose the additional weight required is W. Then from Eq. 10.24 and the discussion given there, the surface tension of the liquid-air interface is

$$S_{la} = (W/2l) = (mg/2l) \qquad (10.25)$$

where m is the extra mass and l is the length of the plate edge. The subscript (la) emphasises the fact that the liquid-air interface tension is involved.

10.7.3 Angle of Contact

The surface of liquid near the plane of contact, with another medium is in general curved. The angle between tangent to the liquid surface at the point of contact and solid surface inside the liquid is termed as angle of contact. It is denoted by θ. It is different at interfaces of different pairs of liquids and solids. The value of θ determines whether a liquid will spread on the surface of a solid or it will form droplets on it. For example, water forms droplets on lotus leaf as shown in Fig. 10.19 (a) while spreads over a clean plastic plate as shown in Fig. 10.19(b).

Fig. 10.19 Different shapes of water drops with interfacial tensions (a) on a lotus leaf (b) on a clean plastic plate.

We consider the three interfacial tensions at all the three interfaces, liquid-air, solid-air and solid-liquid denoted by S_{la}, S_{sa} & S_{sl} respectively as given in Fig. 10.19 (a) and (b). At the line of contact, the surface forces between the three media must be in equilibrium. From the Fig. 10.19(b) the following relation is easily derived.

$$S_{la} \cos \theta + S_{sl} = S_{sa} \quad (10.26)$$

The angle of contact is an obtuse angle if $S_{sl} > S_{la}$ as in the case of water-leaf interface while it is an acute angle if $S_{sl} < S_{la}$ as in the case of water-plastic interface. When θ is an obtuse angle then molecules of liquids are attracted strongly to themselves and weakly to those of solid, it costs a lot of energy to create a liquid-solid surface, and liquid then does not wet the solid. This is what happens with water on a waxy or oily surface, and with mercury on any surface. On the other hand, if the molecules of the liquid are strongly attracted to those of the solid, this will reduce S_{sl} and therefore, cos θ may increase or θ may decrease. In this case θ is an acute angle. This is what happens for water on glass or on plastic and for kerosene oil on virtually anything (it just spreads). Soaps, detergents and dying substances are wetting agents. When they are added the angle of contact becomes small so that these may penetrate well and become effective. Water proofing agents on the other hand are added to create a large angle of contact between the water and fibres.

10.7.4 Drops and Bubbles

One consequence of surface tension is that free liquid drops and bubbles are spherical if effects of gravity can be neglected. You must have seen this especially clearly in small drops just formed in a high-speed spray or jet, and in soap bubbles blown by most of us in childhood. Why are drops and bubbles spherical? What keeps soap bubbles stable?

As we have been saying repeatedly, a liquid-air interface has energy, so for a given volume the surface with minimum energy is the one with the least area. The sphere has this property. Though it is out of the scope of this book, but you can check that a sphere is better than at least a cube in this respect! So, if gravity and other forces (e.g. air resistance) were ineffective, liquid drops would be spherical.

Another interesting consequence of surface tension is that the pressure inside a spherical drop Fig. 10.20(a) is more than the pressure outside. Suppose a spherical drop of radius r is in equilibrium. If its radius increase by Δr. The extra surface energy is

$$[4\pi(r + \Delta r)^2 - 4\pi r^2]\, S_{la} = 8\pi r\, \Delta r\, S_{la} \quad (10.27)$$

If the drop is in equilibrium this energy cost is balanced by the energy gain due to expansion under the pressure difference $(P_i - P_o)$ between the inside of the bubble and the outside. The work done is

$$W = (P_i - P_o)\, 4\pi r^2 \Delta r \quad (10.28)$$

so that

$$(P_i - P_o) = (2\, S_{la} / r) \quad (10.29)$$

In general, for a liquid-gas interface, the convex side has a higher pressure than the concave side. For example, an air bubble in a liquid, would have higher pressure inside it. See Fig 10.20 (b).

Fig. 10.20 Drop, cavity and bubble of radius r.

A bubble Fig 10.20 (c) differs from a drop and a cavity; in this it has two interfaces. Applying the above argument we have for a bubble

$$(P_i - P_o) = (4\, S_{la} / r) \quad (10.30)$$

This is probably why you have to blow hard, but not too hard, to form a soap bubble. A little extra air pressure is needed inside!

10.7.5 Capillary Rise

One consequence of the pressure difference across a curved liquid-air interface is the well-known effect that water rises up in a narrow tube in spite of gravity. The word capilla means

Fig. 10.21 Capillary rise, (a) Schematic picture of a narrow tube immersed water. (b) Enlarged picture near interface.

hair in Latin; if the tube were hair thin, the rise would be very large. To see this, consider a vertical capillary tube of circular cross section (radius a) inserted into an open vessel of water (Fig. 10.21). The contact angle between water and glass is acute. Thus the surface of water in **the capillary is concave. This means that there is a pressure difference between the two sides of the top surface. This is given by**

$$(P_i - P_o) = (2S/r) = 2S/(a \sec \theta)$$
$$= (2S/a) \cos \theta \qquad (10.31)$$

Thus the pressure of the water inside the tube, just at the meniscus (air-water interface) is less than the atmospheric pressure. Consider the two points A and B in Fig. 10.21(a). They must be at the same pressure, namely

$$P_0 + h \rho g = P_i = P_A \qquad (10.32)$$

where ρ is the density of water and h is called the capillary rise [Fig. 10.21(a)]. Using Eq. (10.31) and (10.32) we have

$$h \rho g = (P_i - P_0) = (2S \cos \theta)/a \qquad (10.33)$$

The discussion here, and the Eqs. (10.28) and (10.29) make it clear that the capillary rise is due to surface tension. It is larger, for a smaller a. Typically it is of the order of a few cm for fine capillaries. For example, if a = 0.05 cm, using the value of surface tension for water (Table 10.3), we find that

$$h = 2S/(\rho g a)$$

$$\frac{2 \quad 0.073 \text{ Nm}^{-1}}{10^3 \text{ kg m}^{-3} \quad 9.8 \text{ m s}^{-2} \quad 5 \quad 10^{-4} \text{ m}}$$

$$= 2.98 \times 10^{-2} \text{ m} = 2.98 \text{ cm}$$

Notice that if the liquid meniscus is convex, as for mercury, i.e., if cos θ is negative then from Eq. (10.32) for example, it is clear that the liquid will be lower in the capillary !

10.7.6 Detergents and Surface Tension

We clean dirty clothes containing grease and oil stains sticking to cotton or other fabrics by adding detergents or soap to water, soaking clothes in it and shaking. Let us understand this process better.

Washing with water does not remove grease stains. This is because water does not wet greasy dirt; i.e., there is very little area of contact between them. If water could wet grease, the flow of water could carry some grease away. Something of this sort is achieved through detergents. The molecules of detergents are hairpin shaped, with one end attracted to water and the other to molecules of grease, oil or wax, thus tending to form water-oil interfaces. The result is shown in Fig. 10.22 as a sequence of figures.

In our language, we would say that addition of detergents, whose molecules attract at one end and say, oil on the other, reduces drastically the surface tension S (water-oil). It may even become energetically favourable to form such interfaces, i.e., globs of dirt surrounded by detergents and then by water. This kind of process using surface active detergents or surfactants is important not only for cleaning, but also in recovering oil, mineral ores etc.

Fig. 10.22 Detergent action in terms of what detergent molecules do.

Example 10.12 The lower end of a capillary tube of diameter 2.00 mm is dipped 8.00 cm below the surface of water in a beaker. What is the pressure required in the tube in order to blow a hemispherical bubble at its end in water? The surface tension of water at temperature of the experiments is 7.30×10^{-2} Nm^{-1}. 1 atmospheric pressure = 1.01×10^5 Pa, density of water = 1000 kg/m^3, g = 9.80 m s^{-2}. Also calculate the excess pressure.

Answer The excess pressure in a bubble of gas in a liquid is given by $2S/r$, where S is the surface tension of the liquid-gas interface. You should note there is only one liquid surface in this case. (For a bubble of liquid in a gas, there are two liquid surfaces, so the formula for excess pressure in that case is $4S/r$.) The radius of the bubble is r. Now the pressure outside the bubble P_o equals atmospheric pressure plus the pressure due to 8.00 cm of water column. That is

$$P_o = (1.01 \times 10^5 \text{ Pa} + 0.08 \text{ m} \times 1000 \text{ kg m}^{-3} \times 9.80 \text{ m s}^{-2})$$
$$= 1.01784 \times 10^5 \text{ Pa}$$

Therefore, the pressure inside the bubble is

$$P_i = P_o + 2S/r$$
$$= 1.01784 \times 10^5 \text{ Pa} + (2 \times 7.3 \times 10^{-2} \text{ Pa m}/10^{-3} \text{ m})$$
$$= (1.01784 + 0.00146) \times 10^5 \text{ Pa}$$
$$= 1.02 \times 10^5 \text{ Pa}$$

where the radius of the bubble is taken to be equal to the radius of the capillary tube, since the bubble is hemispherical ! (The answer has been rounded off to three significant figures.) The excess pressure in the bubble is 146 Pa.

SUMMARY

1. The basic property of a fluid is that it can flow. The fluid does not have any resistance to change of its shape. Thus, the shape of a fluid is governed by the shape of its container.
2. A liquid is incompressible and has a free surface of its own. A gas is compressible and it expands to occupy all the space available to it.
3. If F is the normal force exerted by a fluid on an area A then the average pressure P_{av} is defined as the ratio of the force to area
$$P_{av} = \frac{F}{A}$$
4. The unit of the pressure is the pascal (Pa). It is the same as N m^{-2}. Other common units of pressure are
1 atm = 1.01×10^5 Pa
1 bar = 10^5 Pa
1 torr = 133 Pa = 0.133 kPa
1 mm of Hg = 1 torr = 133 Pa
5. *Pascal's law* states that: Pressure in a fluid at rest is same at all points which are at the same height. A change in pressure applied to an enclosed fluid is transmitted undiminished to every point of the fluid and the walls of the containing vessel.
6. The pressure in a fluid varies with depth h according to the expression
$P = P_a + \rho g h$
where ρ is the density of the fluid, assumed uniform.
7. The volume of an incompressible fluid passing any point every second in a pipe of non uniform crossection is the same in the steady flow.
$v A$ = constant (v is the velocity and A is the area of crossection)
The equation is due to mass conservation in incompressible fluid flow.
8. *Bernoulli's principle* states that as we move along a streamline, the sum of the pressure (P), the kinetic energy per unit volume ($\rho v^2/2$) and the potential energy per unit volume ($\rho g y$) remains a constant.
$P + \rho v^2/2 + \rho g y$ = constant

The equation is basically the conservation of energy applied to non viscuss fluid motion in steady state. There is no fluid which have zero viscosity, so the above statement is true only approximately. The viscosity is like friction and converts the kinetic energy to heat energy.

9. Though shear strain in a fluid does not require shear stress, when a shear stress is applied to a fluid, the motion is generated which causes a shear strain growing with time. The ratio of the shear stress to the time rate of shearing strain is known as coefficient of viscosity, η
where symbols have their usual meaning and are defined in the text.
10. *Stokes' law* states that the viscous drag force **F** on a sphere of radius a moving with velocity **v** through a fluid of viscosity is, $\mathbf{F} = -6\pi\eta a\mathbf{v}$.
11. The onset of turbulence in a fluid is determined by a dimensionless parameter is called the *Reynolds number* given by
$R_e = \rho v d/\eta$
Where d is a typical geometrical length associated with the fluid flow and the other symbols have their usual meaning.
12. Surface tension is a force per unit length (or surface energy per unit area) acting in the plane of interface between the liquid and the bounding surface. It is the extra energy that the molecules at the interface have as compared to the interior.

POINTS TO PONDER

1. Pressure is a *scalar quantity*. The definition of the pressure as "force per unit area" may give one false impression that pressure is a vector. The "force" in the numerator of the definition is the component of the force normal to the area upon which it is impressed. While describing fluids as a conceptual shift from particle and rigid body mechanics is required. We are concerned with properties that vary from point to point in the fluid.
2. One should not think of pressure of a fluid as being exerted only on a solid like the walls of a container or a piece of solid matter immersed in the fluid. Pressure exists at all points in a fluid. An element of a fluid (such as the one shown in Fig. 10.2) is in equilibrium because the pressures exerted on the various faces are equal.
3. The expression for pressure
$P = P_a + \rho g h$
holds true if fluid is incompressible. Practically speaking it holds for liquids, which are largely incompressible and hence is a constant with height.
4. The gauge pressure is the difference of the actual pressure and the atmospheric pressure.
$P - P_a = P_g$
Many pressure-measuring devices measure the gauge pressure. These include the tyre pressure gauge and the blood pressure gauge (sphygmomanometer).
5. A streamline is a map of fluid flow. In a steady flow two streamlines do not intersect as it means that the fluid particle will have two possible velocities at the point.
6. Bernoulli's principal does not hold in presence of viscous drag on the fluid. The work done by this dissipative viscous force must be taken into account in this case, and P_2 [Fig. 10.9] will be lower than the value given by Eq. (10.12).
7. As the temperature rises the atoms of the liquid become more mobile and the coefficient of viscosity, η, falls. In a gas the temperature rise increases the random motion of atoms and η increases.
8. The critical Reynolds number for the onset of turbulence is in the range 1000 to 10000, depending on the geometry of the flow. For most cases $R_e < 1000$ signifies laminar flow; $1000 < R_e < 2000$ is unsteady flow and $R_e > 2000$ implies turbulent flow.
9. Surface tension arises due to excess potential energy of the molecules on the surface in comparison to their potential energy in the interior. Such a surface energy is present at the interface separating two substances at least one of which is a fluid. It is not the property of a single fluid alone.

Physical Quantity	Symbol	Dimensions	Unit	Remarks
Pressure	P	$[M\ L^{-1} T^{-2}]$	pascal (Pa)	1 atm = 1.013×10^5 Pa, Scalar
Density	ρ	$[M\ L^{-3}]$	kg m^{-3}	Scalar
Specific Gravity		No	No	$\frac{\rho_{substance}}{\rho_{water}}$, Scalar
Co-efficient of viscosity	η	$[M\ L^{-1}\ T^{-1}]$	Pa s or poiseiulles (Pl)	Scalar
Reynold's Number	R_e	No	No	$R_e = \frac{\rho v d}{\eta}$ scalar
Surface Tension	S	$[M\ T^{-2}]$	N m^{-1}	Scalar

EXERCISES

10.1 Explain why

(a) The blood pressure in humans is greater at the feet than at the brain

(b) Atmospheric pressure at a height of about 6 km decreases to nearly half of its value at the sea level, though the height of the atmosphere is more than 100 km

(c) Hydrostatic pressure is a scalar quantity even though pressure is force divided by area.

10.2 Explain why

(a) The angle of contact of mercury with glass is obtuse, while that of water with glass is acute.

(b) Water on a clean glass surface tends to spread out while mercury on the same surface tends to form drops. (Put differently, water wets glass while mercury does not.)

(c) Surface tension of a liquid is independent of the area of the surface

(d) Water with detergent disolved in it should have small angles of contact.

(e) A drop of liquid under no external forces is always spherical in shape

10.3 Fill in the blanks using the word(s) from the list appended with each statement:

(a) Surface tension of liquids generally ... with temperatures (increases / decreases)

(b) Viscosity of gases ... with temperature, whereas viscosity of liquids ... with temperature (increases / decreases)

(c) For solids with elastic modulus of rigidity, the shearing force is proportional to ... , while for fluids it is proportional to ... (shear strain / rate of shear strain)

(d) For a fluid in a steady flow, the increase in flow speed at a constriction follows (conservation of mass / Bernoulli's principle)

(e) For the model of a plane in a wind tunnel, turbulence occurs at a ... speed for turbulence for an actual plane (greater / smaller)

10.4 Explain why

(a) To keep a piece of paper horizontal, you should blow over, not under, it

(b) When we try to close a water tap with our fingers, fast jets of water gush through the openings between our fingers

(c) The size of the needle of a syringe controls flow rate better than the thumb pressure exerted by a doctor while administering an injection

(d) A fluid flowing out of a small hole in a vessel results in a backward thrust on the vessel

(e) A spinning cricket ball in air does not follow a parabolic trajectory

10.5 A 50 kg girl wearing high heel shoes balances on a single heel. The heel is circular with a diameter 1.0 cm. What is the pressure exerted by the heel on the horizontal floor ?

10.6 Toricelli's barometer used mercury. Pascal duplicated it using French wine of density 984 kg m^{-3}. Determine the height of the wine column for normal atmospheric pressure.

10.7 A vertical off-shore structure is built to withstand a maximum stress of 10^9 Pa. Is the structure suitable for putting up on top of an oil well in the ocean ? Take the depth of the ocean to be roughly 3 km, and ignore ocean currents.

10.8 A hydraulic automobile lift is designed to lift cars with a maximum mass of 3000 kg. The area of cross-section of the piston carrying the load is 425 cm^2. What maximum pressure would the smaller piston have to bear ?

10.9 A U-tube contains water and methylated spirit separated by mercury. The mercury columns in the two arms are in level with 10.0 cm of water in one arm and 12.5 cm of spirit in the other. What is the specific gravity of spirit ?

10.10 In the previous problem, if 15.0 cm of water and spirit each are further poured into the respective arms of the tube, what is the difference in the levels of mercury in the two arms ? (Specific gravity of mercury = 13.6)

10.11 Can Bernoulli's equation be used to describe the flow of water through a rapid in a river ? Explain.

10.12 Does it matter if one uses gauge instead of absolute pressures in applying Bernoulli's equation ? Explain.

10.13 Glycerine flows steadily through a horizontal tube of length 1.5 m and radius 1.0 cm. If the amount of glycerine collected per second at one end is 4.0×10^{-3} kg s^{-1}, what is the pressure difference between the two ends of the tube ? (Density of glycerine = 1.3×10^3 kg m^{-3} and viscosity of glycerine = 0.83 Pa s). [You may also like to check if the assumption of laminar flow in the tube is correct].

10.14 In a test experiment on a model aeroplane in a wind tunnel, the flow speeds on the upper and lower surfaces of the wing are 70 m s^{-1} and 63 m s^{-1} respectively. What is the lift on the wing if its area is 2.5 m^2 ? Take the density of air to be 1.3 kg m^{-3}.

10.15 Figures 10.23(a) and (b) refer to the steady flow of a (non-viscous) liquid. Which of the two figures is incorrect ? Why ?

Fig. 10.23

10.16 The cylindrical tube of a spray pump has a cross-section of 8.0 cm^2 one end of which has 40 fine holes each of diameter 1.0 mm. If the liquid flow inside the tube is 1.5 m min^{-1}, what is the speed of ejection of the liquid through the holes ?

10.17 A U-shaped wire is dipped in a soap solution, and removed. The thin soap film formed between the wire and the light slider supports a weight of 1.5×10^{-2} N (which includes the small weight of the slider). The length of the slider is 30 cm. What is the surface tension of the film ?

10.18 Figure 10.24 (a) shows a thin liquid film supporting a small weight = 4.5×10^{-2} N. What is the weight supported by a film of the same liquid at the same temperature in Fig. (b) and (c) ? Explain your answer physically.

Fig. 10.24

10.19 What is the pressure inside the drop of mercury of radius 3.00 mm at room temperature ? Surface tension of mercury at that temperature (20 °C) is 4.65×10^{-1} N m^{-1}. The atmospheric pressure is 1.01×10^5 Pa. Also give the excess pressure inside the drop.

10.20 What is the excess pressure inside a bubble of soap solution of radius 5.00 mm, given that the surface tension of soap solution at the temperature (20 °C) is 2.50×10^{-2} N m^{-1}? If an air bubble of the same dimension were formed at depth of 40.0 cm inside a container containing the soap solution (of relative density 1.20), what would be the pressure inside the bubble ? (1 atmospheric pressure is 1.01×10^5 Pa).

Additional Exercises

10.21 A tank with a square base of area 1.0 m^2 is divided by a vertical partition in the middle. The bottom of the partition has a small-hinged door of area 20 cm^2. The tank is filled with water in one compartment, and an acid (of relative density 1.7) in the other, both to a height of 4.0 m. compute the force necessary to keep the door close.

10.22 A manometer reads the pressure of a gas in an enclosure as shown in Fig. 10.25 (a) When a pump removes some of the gas, the manometer reads as in Fig. 10.25 (b) The liquid used in the manometers is mercury and the atmospheric pressure is 76 cm of mercury.

(a) Give the absolute and gauge pressure of the gas in the enclosure for cases (a) and (b), in units of cm of mercury.

(b) How would the levels change in case (b) if 13.6 cm of water (immiscible with mercury) are poured into the right limb of the manometer ? (Ignore the small change in the volume of the gas).

Fig. 10.25

10.23 Two vessels have the same base area but different shapes. The first vessel takes twice the volume of water that the second vessel requires to fill upto a particular common height. Is the force exerted by the water on the base of the vessel the same in the two cases ? If so, why do the vessels filled with water to that same height give different readings on a weighing scale ?

10.24 During blood transfusion the needle is inserted in a vein where the gauge pressure is 2000 Pa. At what height must the blood container be placed so that blood may just enter the vein ? [Use the density of whole blood from Table 10.1].

10.25 In deriving Bernoulli's equation, we equated the work done on the fluid in the tube to its change in the potential and kinetic energy. (a) What is the largest average velocity of blood flow in an artery of diameter 2×10^{-3} m if the flow must remain laminar ? (b) Do the dissipative forces become more important as the fluid velocity increases ? Discuss qualitatively.

10.26 (a) What is the largest average velocity of blood flow in an artery of radius 2×10^{-3}m if the flow must remain lanimar? (b) What is the corresponding flow rate ? (Take viscosity of blood to be 2.084×10^{-3} Pa s).

10.27 A plane is in level flight at constant speed and each of its two wings has an area of 25 m^2. If the speed of the air is 180 km/h over the lower wing and 234 km/h over the upper wing surface, determine the plane's mass. (Take air density to be 1 kg m^{-3}).

10.28 In Millikan's oil drop experiment, what is the terminal speed of an uncharged drop of radius 2.0×10^{-5} m and density 1.2×10^3 kg m^{-3}. Take the viscosity of air at the temperature of the experiment to be 1.8×10^{-5} Pa s. How much is the viscous force on the drop at that speed ? Neglect buoyancy of the drop due to air.

10.29 Mercury has an angle of contact equal to 140° with soda lime glass. A narrow tube of radius 1.00 mm made of this glass is dipped in a trough containing mercury. By what amount does the mercury dip down in the tube relative to the liquid surface outside ? Surface tension of mercury at the temperature of the experiment is 0.465 N m^{-1}. Density of mercury = 13.6×10^3 kg m^{-3}.

10.30 Two narrow bores of diameters 3.0 mm and 6.0 mm are joined together to form a U-tube open at both ends. If the U-tube contains water, what is the difference in its levels in the two limbs of the tube ? Surface tension of water at the temperature of the experiment is 7.3×10^{-2} N m^{-1}. Take the angle of contact to be zero and density of water to be 1.0×10^3 kg m^{-3} (g = 9.8 m s^{-2}) .

Calculator/Computer – Based Problem

10.31 (a) It is known that density ρ of air decreases with height y as

$$_0e^{\ y/y_o}$$

where ρ_0 = 1.25 kg m^{-3} is the density at sea level, and y_0 is a constant. This density variation is called the law of atmospheres. Obtain this law assuming that the temperature of atmosphere remains a constant (isothermal conditions). Also assume that the value of g remains constant.

(b) A large He balloon of volume 1425 m^3 is used to lift a payload of 400 kg. Assume that the balloon maintains constant radius as it rises. How high does it rise ?

[Take y_0 = 8000 m and ρ_{He} = 0.18 kg m^{-3}].

APPENDIX 10.1 : WHAT IS BLOOD PRESSURE ?

In evolutionary history there occurred a time when animals started spending a significant amount of time in the upright position. This placed a number of demands on the circulatory system. The venous system that returns blood from the lower extremities to the heart underwent changes. You will recall that veins are blood vessels through which blood returns to the heart. Humans and animals such as the giraffe have adapted to the problem of moving blood upward against gravity. But animals such as snakes, rats and rabbits will die if held upwards, since the blood remains in the lower extremities and the venous system is unable to move it towards the heart.

Fig. 10.26 Schematic view of the gauge pressures in the arteries in various parts of the human body while standing or lying down. The pressures shown are averaged over a heart cycle.

Figure 10.26 shows the average pressures observed in the arteries at various points in the human body. Since viscous effects are small, we can use Bernoulli's equation, Eq. (10.13),

$$P \quad \frac{1}{2} \quad \upsilon^2 \qquad gy \quad \text{Constant}$$

to understand these pressure values. The kinetic energy term ($\rho\ \upsilon^2/2$) can be ignored since the velocities in the three arteries are small ($\approx$ 0.1 m s^{-1}) and almost constant. Hence the gauge pressures at the brain P_B, the heart P_H, and the foot P_F are related by

$$P_F = P_H + \rho\ g\, h_H = P_B + \rho\ g\, h_B \tag{10.34}$$

where ρ is the density of blood.

Typical values of the heights to the heart and the brain are h_H = 1.3 m and h_B = 1.7 m. Taking $\rho = 1.06 \times 10^3$ kg m^{-3} we obtain that P_F = 26.8 kPa (kilopascals) and P_B = 9.3 kPa given that P_H = 13.3 kPa. Thus the pressures in the lower and upper parts of the body are so different when a person is standing, but are almost equal when he is lying down. As mentioned in the text the units for pressure more commonly employed in medicine and physiology are torr and mm of Hg. 1 mm of Hg = 1 torr = 0.133 kPa. Thus the average pressure at the heart is P_H = 13.3 kPa = 100 mm of Hg.

The human body is a marvel of nature. The veins in the lower extremities are equipped with valves, which open when blood flows towards the heart and close if it tends to drain down. Also, blood is returned at least partially by the pumping action associated with breathing and by the flexing of the skeletal muscles during walking. This explains why a soldier who is required to stand at attention may faint because of insufficient return of the blood to the heart. Once he is made to lie down, the pressures become equalized and he regains consciousness.

An instrument called the sphygmomanometer usually measures the blood pressure of humans. It is a fast, painless and non-invasive technique and gives the doctor a reliable idea about the patient's health. The measurement process is shown in Fig. 10.27. There are two reasons why the upper arm is used. First, it is at the same level as the heart and measurements here give values close to that at the heart. Secondly, the upper arm contains a single bone and makes the artery there (called the brachial artery) easy to compress. We have all measured pulse rates by placing our fingers over the wrist. Each pulse takes a little less than a second. During each pulse the pressure in the heart and the circulatory system goes through a

maximum as the blood is pumped by the heart (**systolic pressure**) and a minimum as the heart relaxes **(diastolic pressure)**. The sphygmomanometer is a device, which measures these extreme pressures. It works on the principle that **blood flow in the brachial (upper arm) artery can be made to go from laminar to turbulent by suitable compression. Turbulent flow is dissipative, and its sound can be picked up on the stethoscope**.

The gauge pressure in an air sack wrapped around the upper arm is measured using a manometer or a dial pressure gauge (Fig. 10.27). The pressure in the sack is first increased till the brachial artery is closed. The pressure in the sack is then slowly reduced while a stethoscope placed just below the sack is used to listen to noises arising in the brachial artery. When the pressure is just below the **systolic** (peak) pressure, the artery opens briefly. During this brief period, the blood velocity in the highly constricted artery is high and turbulent and hence noisy. The resulting noise is heard as a **tapping sound** on the stethoscope. When the pressure in the sack is lowered further, the artery remains open for a longer portion of the heart cycle. Nevertheless, it remains closed during the **diastolic** (minimum pressure) phase of the heartbeat. Thus the duration of the tapping sound is longer. When the pressure in the sack reaches the diastolic pressure the artery is open during the entire heart cycle. The flow is however, still turbulent and noisy. But instead of a tapping sound we hear a steady, continuous roar on the stethoscope.

Fig. 10.27 Blood pressure measurement using the sphygmomanometer and stethoscope.

The blood pressure of a patient is presented as the ratio of systolic/diastolic pressures. For a resting healthy adult it is typically 120/80 mm of Hg (120/80 torr). Pressures above 140/90 require medical attention and advice. High blood pressures may seriously damage the heart, kidney and other organs and must be controlled.

CHAPTER ELEVEN

THERMAL PROPERTIES OF MATTER

11.1 Introduction
11.2 Temperature and heat
11.3 Measurement of temperature
11.4 Ideal-gas equation and absolute temperature
11.5 Thermal expansion
11.6 Specific heat capacity
11.7 Calorimetry
11.8 Change of state
11.9 Heat transfer
11.10 Newton's law of cooling
Summary
Points to ponder
Exercises

11.1 INTRODUCTION

We all have common-sense notions of heat and temperature. Temperature is a measure of 'hotness' of a body. A kettle with boiling water is hotter than a box containing ice. In physics, we need to define the notion of heat, temperature, etc., more carefully. In this chapter, you will learn what heat is and how it is measured, and study the various proceses by which heat flows from one body to another. Along the way, you will find out why blacksmiths heat the iron ring before fitting on the rim of a wooden wheel of a bullock cart and why the wind at the beach often reverses direction after the sun goes down. You will also learn what happens when water boils or freezes, and its temperature does not change during these processes even though a great deal of heat is flowing into or out of it.

11.2 TEMPERATURE AND HEAT

We can begin studying thermal properties of matter with definitions of temperature and heat. Temperature is a relative measure, or indication of hotness or coldness. A hot utensil is said to have a high temperature, and ice cube to have a low temperature. An object that has a higher temperature than another object is said to be hotter. Note that hot and cold are relative terms, like tall and short. We can perceive temperature by touch. However, this temperature sense is somewhat unreliable and its range is too limited to be useful for scientific purposes.

We know from experience that a glass of ice-cold water left on a table on a hot summer day eventually warms up whereas a cup of hot tea on the same table cools down. It means that when the temperature of body, ice-cold water or hot tea in this case, and its surrounding medium are different, heat transfer takes place between the system and the surrounding medium, until the body and the surrounding medium are at the same temperature. We also know that in the case of glass tumbler of ice cold water, heat flows from the environment to

the glass tumbler, whereas in the case of hot tea, it flows from the cup of hot tea to the environment. So, we can say that **heat is the form of energy transferred between two (or more) systems or a system and its surroundings by virtue of temperature difference.** The SI unit of heat energy transferred is expressed in joule (J) while SI unit of temperature is kelvin (K), and °C is a commonly used unit of temperature. When an object is heated, many changes may take place. Its temperature may rise, it may expand or change state. We will study the effect of heat on different bodies in later sections.

11.3 MEASUREMENT OF TEMPERATURE

A measure of temperature is obtained using a thermometer. Many physical properties of materials change sufficiently with temperature to be used as the basis for constructing thermometers. The commonly used property is variation of the volume of a liquid with temperature. For example, a common thermometer (the liquid-in-glass type) with which you are familiar. Mercury and alcohol are the liquids used in most liquid-in-glass thermometers.

Thermometers are calibrated so that a numerical value may be assigned to a given temperature. For the definition of any standard scale, two fixed reference points are needed. Since all substances change dimensions with temperature, an absolute reference for expansion is not available. However, the necessary fixed points may be correlated to physical phenomena that always occur at the same temperature. The ice point and the steam point of water are two convenient fixed points and are known as the freezing and boiling points. These two points are the temperatures at which pure water freezes and boils under standard pressure. The two familiar temperature scales are the Fahrenheit temperature scale and the Celsius temperature scale. The ice and steam point have values 32 °F and 212 °F respectively, on the Fahrenheit scale and 0 °C and 100 °C on the Celsius scale. On the Fahrenheit scale, there are 180 equal intervals between two reference points, and on the celsius scale, there are 100.

Fig. 11.1 A plot of Fahrenheit temperature (t_F) versus Celsius temperature (t_c).

A relationship for converting between the two scales may be obtained from a graph of Fahrenheit temperature (t_F) versus celsius temperature (t_C) in a straight line (Fig. 11.1), whose equation is

$$\frac{t_F - 32}{180} = \frac{t_C}{100} \tag{11.1}$$

11.4 IDEAL-GAS EQUATION AND ABSOLUTE TEMPERATURE

Liquid-in-glass thermometers show different readings for temperatures other than the fixed points because of differing expansion properties. A thermometer that uses a gas, however, gives the same readings regardless of which gas is used. Experiments show that all gases at low densities exhibit same expansion behaviour. The variables that describe the behaviour of a given quantity (mass) of gas are pressure, volume, and temperature (P, V, and T)(where $T = t + 273.15$; t is the temperature in °C). When temperature is held constant, the pressure and volume of a quantity of gas are related as PV = constant. This relationship is known as Boyle's law, after Robert Boyle (1627-1691) the English Chemist who discovered it. When the pressure is held constant, the volume of a quantity of the gas is related to the temperature as V/T = constant. This relationship is known as Charles' law, after the French scientist Jacques Charles (1747-1823). Low density gases obey these laws, which may be combined into a single relationship.

Fig. 11.2 Pressure versus temperature of a low density gas kept at constant volume.

Fig. 11.3 A plot of pressure versus temperature and extrapolation of lines for low density gases indicates the same absolute zero temperature.

Notice that since PV = constant and V/T = constant for a given quantity of gas, then PV/T should also be a constant. This relationship is known as ideal gas law. It can be written in a more general form that applies not just to a given quantity of a single gas but to any quantity of any dilute gas and is known as **ideal-gas equation**:

$$\frac{PV}{T} = \mu R$$

or $PV = \mu RT$ (11.2)

where, μ is the number of moles in the sample of gas and R is called universal gas constant:

$R = 8.31$ J mol^{-1} K^{-1}

In Eq. 11.2, we have learnt that the pressure and volume are directly proportional to temperature : $PV \propto T$. This relationship allows a gas to be used to measure temperature in a constant volume gas thermometer. Holding the volume of a gas constant, it gives $P \propto T$. Thus, with a constant-volume gas thermometer, temperature is read in terms of pressure. A plot of pressure versus temperature gives a straight line in this case, as shown in Fig. 11.2.

However, measurements on real gases deviate from the values predicted by the ideal gas law at low temperature. But the relationship is linear over a large temperature range, and it looks as though the pressure might reach zero with decreasing temperature if the gas continued to be a gas. The absolute minimum temperature for an ideal gas, therefore, inferred by extrapolating the straight line to the axis, as in Fig. 11.3. This temperature is found to be – 273.15 °C and is designated as **absolute zero**. Absolute zero is the foundation of the Kelvin temperature scale or absolute scale temperature named after the British scientist Lord Kelvin. On this scale, – 273.15 °C is taken as the zero point, that is 0 K (Fig. 11.4).

Fig. 11.4 Comparision of the Kelvin, Celsius and Fahrenheit temperature scales.

The size of the unit for Kelvin temperature is the same celsius degree, so temperature on these scales are related by

$$T = t_C + 273.15 \quad (11.3)$$

11.5 THERMAL EXPANSION

You may have observed that sometimes sealed bottles with metallic lids are so tightly screwed that one has to put the lid in hot water for sometime to open the lid. This would allow the metallic cover to expand, thereby loosening it to unscrew easily. In case of liquids, you may have observed that mercury in a thermometer rises, when the thermometer is put in a slightly warm water. If we take out the thermometer from the

warm water the level of mercury falls again. Similarly, in the case of gases, a balloon partially inflated in a cool room may expand to full size when placed in warm water. On the other hand, a fully inflated balloon when immersed in cold water would start shrinking due to contraction of the air inside.

It is our common experience that most substances expand on heating and contract on cooling. A change in the temperature of a body causes change in its dimensions. The increase in the dimensions of a body due to the increase in its temperature is called thermal expansion. The expansion in length is called **linear expansion**. The expansion in area is called **area expansion**. The expansion in volume is called **volume expansion** (Fig. 11.5).

Fig. 11.5 Thermal Expansion.

If the substance is in the form of a long rod, then for small change in temperature, ΔT, the fractional change in length, $\Delta l/l$, is directly proportional to ΔT.

$$\frac{\Delta l}{l} = \alpha_1 \Delta T \qquad (11.4)$$

where α_1 is known as the **coefficient of linear expansion** and is characteristic of the material of the rod. In Table 11.1 are given typical average values of the coefficient of linear expansion for some materials in the temperature range 0 °C to 100 °C. From this Table, compare the value of α_1 for glass and copper. We find that copper expands about five times more than glass for the same rise in temperature. Normally, metals expand more and have relatively high values of α_1.

Table 11.1 Values of coefficient of linear expansion for some materials

Materials	α_1 (10^{-5} K^{-1})
Aluminium	2.5
Brass	1.8
Iron	1.2
Copper	1.7
Silver	1.9
Gold	1.4
Glass (pyrex)	0.32
Lead	0.29

Similarly, we consider the fractional change in volume, $\frac{\Delta V}{V}$, of a substance for temperature change ΔT and define the **coefficient of volume expansion**, α_V as

$$\alpha_V = \left(\frac{\Delta V}{V}\right)\frac{1}{\Delta T} \qquad (11.5)$$

Here α_V is also a characteristic of the substance but is not strictly a constant. It depends in general on temperature (Fig 11.6). It is seen that α_V becomes constant only at a high temperature.

Fig. 11.6 Coefficient of volume expansion of copper as a function of temperature.

Table 11.2 gives the values of co-efficient of volume expansion of some common substances in the temperature range 0 –100 °C. You can see that thermal expansion of these substances (solids and liquids) is rather small, with

materials like pyrex glass and invar (a special iron-nickel alloy) having particularly low values of α_V. From this Table we find that the value of α_v for alcohol (ethyl) is more than mercury and expands more than mercury for the same rise in temperature.

Table 11.2 Values of coefficient of volume expansion for some substances

Materials	α_v (K^{-1})
Aluminium	7×10^{-5}
Brass	6×10^{-5}
Iron	3.55×10^{-5}
Paraffin	58.8×10^{-5}
Glass (ordinary)	2.5×10^{-5}
Glass (pyrex)	1×10^{-5}
Hard rubber	2.4×10^{-4}
Invar	2×10^{-6}
Mercurry	18.2×10^{-5}
Water	20.7×10^{-5}
Alcohol (ethyl)	110×10^{-5}

Water exhibits an anomalous behavour; it contracts on heating between 0 °C and 4 °C. The volume of a given amount of water decreases as it is cooled from room temperature, until its temperature reaches 4 °C, [Fig. 11.7(a)]. Below 4 °C, the volume increases, and therefore the density decreases [Fig. 11.7(b)].

This means that water has a maximum density at 4 °C. This property has an important environmental effect: Bodies of water, such as lakes and ponds, freeze at the top first. As a lake cools toward 4 °C, water near the surface loses energy to the atmosphere, becomes denser, and sinks; the warmer, less dense water near the bottom rises. However, once the colder water on top reaches temperature below 4 °C, it becomes less dense and remains at the surface, where it freezes. If water did not have this property, lakes and ponds would freeze from the bottom up, which would destroy much of their animal and plant life.

Gases at ordinary temperature expand more than solids and liquids. For liquids, the coefficient of volume expansion is relatively independent of the temperature. However, for gases it is dependent on temperature. For an ideal gas, the coefficient of volume expansion at constant pressure can be found from the ideal gas equation :

$PV = \mu RT$

At constant pressure

$P\Delta V = \mu R\,\Delta T$

$$\frac{\Delta V}{V} = \frac{\Delta T}{T}$$

i.e. $\alpha_v = \dfrac{1}{T}$ for ideal gas (11.6)

At 0 °C, $\alpha_v = 3.7 \times 10^{-3}\ K^{-1}$, which is much larger than that for solids and liquids. Equation (11.6) shows the temperature dependence of α_v; it decreases with increasing temperature. For a gas at room temperature and constant pressure α_v is about $3300 \times 10^{-6} K^{-1}$, as

Fig. 11.7 Thermal expansion of water.

much as order(s) of magnitude larger than the coefficient of volume expansion of typical liquids.

There is a simple relation between the coefficient of volume expansion (α_v) and coefficient of linear expansion (α_l). Imagine a cube of length, l, that expands equally in all directions, when its temperature increases by ΔT. We have

$\Delta l = \alpha_l \, l \, \Delta T$

so, $\Delta V = (l+\Delta l)^3 - l^3 \; ; \quad 3l^2 \, \Delta l$ (11.7)

In equation (11.7), terms in $(\Delta l)^2$ and $(\Delta l)^3$ have been neglected since Δl is small compared to l. So

$$\Delta V = \frac{3V\,\Delta l}{l} = 3V\alpha_l \; \Delta T \qquad (11.8)$$

which gives

$$\alpha_v = 3\alpha_l \qquad (11.9)$$

What happens by preventing the thermal expansion of a rod by fixing its ends rigidly? Clearly, the rod acquires a compressive strain due to the external forces provided by the rigid support at the ends. The corresponding stress set up in the rod is called **thermal stress**. For example, consider a steel rail of length 5 m and area of cross section 40 cm^2 that is prevented from expanding while the temperature rises by 10 °C. The coefficient of linear expansion of steel is $\alpha_{l(\text{steel})} = 1.2 \times 10^{-5}$ K^{-1}. Thus, the compressive strain is $\frac{\Delta l}{l} = \alpha_{l(\text{steel})}\Delta T = 1.2 \times 10^{-5} \times 10 = 1.2 \times 10^{-4}$. Youngs modulus of steel is $Y_{(\text{steel})} = 2 \times 10^{11}$ N m^{-2}. Therefore, the thermal stress developed is $\frac{\Delta F}{A} = Y_{steel}\left(\frac{\Delta l}{l}\right) = 2.4 \times 10^7$ N m^{-2}, which corresponds to an external force of

$$\Delta F = AY_{\text{steel}}\left(\frac{\Delta l}{l}\right) = 2.4 \times 10^7 \times 40 \times 10^{-4} \simeq 10^5 \text{N}.$$

If two such steel rails, fixed at their outer ends, are in contact at their inner ends, a force of this magnitude can easily bend the rails.

Example 11.1 Show that the coefficient of area expansions, $(\Delta A/A)/\Delta T$, of a rectangular sheet of the solid is twice its linear expansivity, α_l.

Answer

Fig. 11.8

Consider a rectangular sheet of the solid material of length a and breadth b (Fig. 11.8). When the temperature increases by ΔT, a increases by $\Delta a = \alpha_l a \Delta T$ and b increases by $\Delta b = \alpha_l b \, \Delta T$. From Fig. 11.8, the increase in area

$$\Delta A = \Delta A_1 + \Delta A_2 + \Delta A_3$$
$$\Delta A = a\,\Delta b + b\,\Delta a + (\Delta a)\;(\Delta b)$$
$$= a\,\alpha_l b\,\Delta T + b\,\alpha_l\,a\,\Delta T + (\alpha_l)^2\,ab\,(\Delta T)^2$$
$$= \alpha_l\,ab\,\Delta T\,(2 + \alpha_l\,\Delta T) = \alpha_l\,A\,\Delta T\,(2 + \alpha_l\,\Delta T)$$

Since α_l ; 10^{-5} K^{-1}, from Table 11.1, the product $\alpha_l \Delta T$ for fractional temperature is small in comparision with 2 and may be neglected. Hence,

$$\left(\frac{\Delta A}{A}\right)\frac{1}{\Delta T} \simeq 2\alpha_l$$

Example 11.2 A blacksmith fixes iron ring on the rim of the wooden wheel of a bullock cart. The diameter of the rim and the iron ring are 5.243 m and 5.231 m respectively at 27 °C. To what temperature should the ring be heated so as to fit the rim of the wheel?

Given, $T_1 = 27$ °C

$L_{T1} = 5.231$ m

$L_{T2} = 5.243$ m

So,

$$L_{T2} = L_{T1}\;[1+\alpha_l\,(T_2 - T_1)]$$

$$5.243 \text{ m} = 5.231 \text{ m}\,[1 + 1.20\times10^{-5}\text{ K}^{-1}\,(T_2 - 27\text{ °C})]$$

or $T_2 = 218$ °C.

11.6 SPECIFIC HEAT CAPACITY

Take some water in a vessel and start heating it on a burner. Soon you will notice that bubbles begin to move upward. As the temperature is raised the motion of water particles increases till it becomes turbulent as water starts boiling. What are the factors on which the quantity of heat required to raise the temperature of a substance depend? In order to answer this question in the first step, heat a given quantity of water to raise its temperature by, say 20 °C and note the time taken. Again take the same amount of water and raise its temperature by 40 °C using the same source of heat. Note the time taken by using a stopwatch. You will find it takes about twice the time and therefore, double the quantity of heat required raising twice the temperature of same amount of water.

In the second step, now suppose you take double the amount of water and heat it, using the same heating arrangement, to raise the temperature by 20 °C, you will find the time taken is again twice that required in the first step.

In the third step, in place of water, now heat the same quantity of some oil, say mustard oil, and raise the temperature again by 20 °C. Now note the time by the same stopwatch. You will find the time taken will be shorter and therefore, the quantity of heat required would be less than that required by the same amount of water for the same rise in temperature.

The above observations show that the quantity of heat required to warm a given substance depends on its mass, m, the change in temperature, ΔT and the nature of substance. The change in temperature of a substance, when a given quantity of heat is absorbed or rejected by it, is characterised by a quantity called the **heat capacity** of that substance. We define heat capacity, S of a substance as

$$S = \frac{\Delta Q}{\Delta T} \tag{11.10}$$

where ΔQ is the amount of heat supplied to the substance to change its temperature from T to $T + \Delta T$.

You have observed that if equal amount of heat is added to equal masses of different substances, the resulting temperature changes will not be the same. It implies that every substance has a unique value for the amount of heat absorbed or rejected to change the temperature of unit mass of it by one unit. This quantity is referred to as the **specific heat capacity** of the substance.

If ΔQ stands for the amount of heat absorbed or rejected by a substance of mass m when it undergoes a temperature change ΔT, then the specific heat capacity, of that substance is given by

$$s = \frac{S}{m} = \frac{1}{m}\frac{\Delta Q}{\Delta T} \tag{11.11}$$

The **specific heat capacity** is the property of the substance which determines the change in the temperature of the substance (undergoing no phase change) when a given quantity of heat is absorbed (or rejected) by it. It is defined as the amount of heat per unit mass absorbed or rejected by the substance to change its temperature by one unit. It depends on the nature of the substance and its temperature. The SI unit of specific heat capacity is J kg^{-1} K^{-1}.

If the amount of substance is specified in terms of moles μ, instead of mass m in kg, we can define heat capacity per mole of the substance by

$$C = \frac{S}{\mu} = \frac{1}{\mu}\frac{\Delta Q}{\Delta T} \tag{11.12}$$

where C is known as **molar specific heat capacity** of the substance. Like S, C also depends on the nature of the substance and its temperature. The SI unit of molar specific heat capacity is J mol^{-1} K^{-1}.

However, in connection with specific heat capacity of gases, additional conditions may be needed to define C. In this case, heat transfer can be achieved by keeping either pressure or volume constant. If the gas is held under constant pressure during the heat transfer, then it is called the **molar specific heat capacity at constant pressure** and is denoted by C_p. On the other hand, if the volume of the gas is maintained during the heat transfer, then the corresponding molar specific heat capacity is called **molar specific heat capacity at constant volume** and is denoted by C_v. For details see Chapter 12. Table 11.3 lists measured specific heat capacity of some substances at atmospheric pressure and ordinary temperature while Table 11.4 lists molar specific heat capacities of some gases. From Table 11.3 you can note that water

Table 11.3 Specific heat capacity of some substances at room temperature and atmospheric pressure

Substance	Specific heat capacity ($J\ kg^{-1}\ K^{-1}$)	Substance	Specific heat capacity ($J\ kg^{-1}\ K^{-1}$)
Aluminium	900.0	Ice	2060
Carbon	506.5	Glass	840
Copper	386.4	Iron	450
Lead	127.7	Kerosene	2118
Silver	236.1	Edible oil	1965
Tungesten	134.4	Mercury	140
Water	4186.0		

has the highest specific heat capacity compared to other substances. For this reason water is used as a coolant in automobile radiators as well as a heater in hot water bags. Owing to its high specific heat capacity, the water warms up much more slowly than the land during summer and consequently wind from the sea has a cooling effect. Now, you can tell why in desert areas, the earth surface warms up quickly during the day and cools quickly at night.

Table 11.4 Molar specific heat capacities of some gases

Gas	C_p ($J\ mol^{-1}K^{-1}$)	C_v($J\ mol^{-1}K^{-1}$)
He	20.8	12.5
H_2	28.8	20.4
N_2	29.1	20.8
O_2	29.4	21.1
CO_2	37.0	28.5

11.7 CALORIMETRY

A system is said to be isolated if no exchange or transfer of heat occurs between the system and its surroundings. When different parts of an isolated system are at different temperature, a quantity of heat transfers from the part at higher temperature to the part at lower temperature. The heat lost by the part at higher temperature is equal to the heat gained by the part at lower temperature.

Calorimetry means measurement of heat. When a body at higher temperature is brought in contact with another body at lower temperature, the heat lost by the hot body is equal to the heat gained by the colder body, provided no heat is allowed to escape to the surroundings. A device in which heat measurement can be made is called a **calorimeter**. It consists a metallic vessel and stirrer of the same material like copper or alumiunium. The vessel is kept inside a wooden jacket which contains heat insulating materials like glass wool etc. The outer jacket acts as a heat shield and reduces the heat loss from the inner vessel. There is an opening in the outer jacket through which a mercury thermometer can be inserted into the calorimeter. The following example provides a method by which the specific heat capacity of a given solid can be determinated by using the principle, heat gained is equal to the heat lost.

Example 11.3 A sphere of aluminium of 0.047 kg placed for sufficient time in a vessel containing boiling water, so that the sphere is at 100 °C. It is then immediately transfered to 0.14 kg copper calorimeter containing 0.25 kg of water at 20 °C. The temperature of water rises and attains a steady state at 23 °C. Calculate the specific heat capacity of aluminium.

Answer In solving this example we shall use the fact that at a steady state, heat given by an aluminium sphere will be equal to the heat absorbed by the water and calorimeter.

Mass of aluminium sphere (m_1) = 0.047 kg
Initial temp. of aluminium sphere = 100 °C
Final temp. = 23 °C
Change in temp (ΔT) = (100 °C - 23 °C) = 77 °C
Let specific heat capacity of aluminium be s_{Al}.

The amount of heat lost by the aluminium sphere = $m_1 s_{Al} \Delta T = 0.047 \text{kg} \times s_{Al} \times 77\ °\text{C}$

Mass of water (m_2) = 0.25 kg

Mass of calorimeter (m_3) = 0.14 kg

Initial temp. of water and calorimeter = 20 °C

Final temp. of the mixture = 23 °C

Change in temp. (ΔT_2) = 23 °C – 20 °C = 3 °C

Specific heat capacity of water (s_w)

$= 4.18 \times 10^3\ \text{J kg}^{-1}\ \text{K}^{-1}$

Specific heat capacity of copper calorimeter

$= 0.386 \times 10^3\ \text{J kg}^{-1}\ \text{K}^{-1}$

The amount of heat gained by water and calorimeter = $m_2\ s_w\ \Delta T_2 + m_3 s_{cu} \Delta T_2$

$= (m_2 s_w + m_3 s_{cu})\ (\Delta T_2)$

$= 0.25\ \text{kg} \times 4.18 \times 10^3\ \text{J kg}^{-1}\ \text{K}^{-1} + 0.14\ \text{kg} \times 0.386 \times 10^3\ \text{J kg}^{-1}\ \text{K}^{-1})\ (23\ °\text{C} - 20\ °\text{C})$

In the steady state heat lost by the aluminium sphere = heat gained by water + heat gained by calorimeter.

So, $0.047\ \text{kg} \times s_{Al} \times 77\ °\text{C}$

$= (0.25\ \text{kg} \times 4.18 \times 10^3\ \text{J kg}^{-1}\ \text{K}^{-1} + 0.14\ \text{kg} \times 0.386 \times 10^3\ \text{J kg}^{-1}\ \text{K}^{-1})(3\ °\text{C})$

$s_{Al} = 0.911\ \text{kJ kg}^{-1}\ \text{K}^{-1}$ ◀

11.8 CHANGE OF STATE

Matter normally exists in three states: solid, liquid, and gas. A transition from one of these states to another is called a change of state. Two common changes of states are solid to liquid and liquid to gas (and vice versa). These changes can occur when the exchange of heat takes place between the substance and its surroundings. To study the change of state on heating or cooling, let us perform the following activity.

Take some cubes of ice in a beaker. Note the temperature of ice (0 °C). Start heating it slowly on a constant heat source. Note the temperature after every minute. Continuously stir the mixture of water and ice. Draw a graph between temperature and time (Fig. 11.9). You will observe no change in the temperature so long as there is ice in the beaker. In the above process, the temperature of the system does not change even though heat is being continuously supplied. The heat supplied is being utilised in changing the state from solid (ice) to liquid (water).

Fig. 11.9 A plot of temperature versus time showing the changes in the state of ice on heating (not to scale).

The change of state from solid to liquid is called **melting** and from liquid to solid is called **fusion**. It is observed that the temperature remains constant until the entire amount of the solid substance melts. That is, **both the solid and liquid states of the substance coexist in thermal equilibrium during the change of states from solid to liquid.** The temperature at which the solid and the liquid states of the substance in thermal equilibrium with each other is called its **melting point**. It is characteristic of the substance. It also depends on pressure. The melting point of a substance at standard atomspheric pressure is called its **normal melting point**. Let us do the following activity to understand the process of melting of ice.

Take a slab of ice. Take a metallic wire and fix two blocks, say 5 kg each, at its ends. Put the wire over the slab as shown in Fig. 11.10. You will observe that the wire passes through the ice slab. This happens due to the fact that just below the wire, ice melts at lower temperature due to increase in pressure. When the wire has passed, water above the wire freezes again. Thus the wire passes through the slab and the slab does not split. This phenomenon of refreezing is called **regelation**. Skating is possible on snow due to the formation of water below the skates. Water is formed due to the increase of pressure and it acts as a lubricant.

Fig. 11.10

After the whole of ice gets converted into water and as we continue further heating, we shall see that temperature begins to rise. The temperature keeps on rising till it reaches nearly 100 °C when it again becomes steady. The heat supplied is now being utilised to change water from liquid state to vapour or gaseous state.

The change of state from liquid to vapour (or gas) is called **vaporisation**. It is observed that the temperature remains constant until the entire amount of the liquid is converted into vapour. That is, both the liquid and vapour states of the substance coexist in thermal equilibrium, during the change of state from liquid to vapour. The temperature at which the liquid and the vapour states of the substance coexist is called its **boiling point**. Let us do the following activity to understand the process of boiling of water.

Take a round-bottom flask, more than half filled with water. Keep it over a burner and fix a

Triple Point

The temperature of a substance remains constant during its change of state (phase change). A graph between the temperature *T* and the Pressure *P* of the substance is called a phase diagram or *P* – *T* diagram. The following figure shows the phase diagram of water and CO_2. Such a phase diagram divides the *P* – *T* plane into a solid-region, the vapour-region and the liquid-region. The regions are separated by the curves such as sublimation curve (BO), **fusion curve** (AO) and **vaporisation curve** (CO). The points on **sublimation curve** represent states in which solid and vapour phases coexist. The point on the sublimation curve BO represent states in which the solid and vapour phases co-exist. Points on the fusion curve AO represent states in which solid and liquid phase coexist. Points on the vapourisation curve CO represent states in which the liquid and vapour phases coexist. The temperature and pressure at which the fusion curve, the vaporisation curve and the sublimation curve meet and all the three phases of a substance coexist is called the **triple point** of the substance. For example the triple point of water is represented by the temperature 273.16 K and pressure 6.11×10^{-3} *Pa*.

Pressure-temperature phase diagrams for (a) water and (b) CO_2 (not to the scale).

thermometer and steam outlet through the cork of the flask (Fig. 11.11). As water gets heated in the flask, note first that the air, which was dissolved in the water, will come out as small bubbles. Later, bubbles of steam will form at the bottom but as they rise to the cooler water near the top, they condense and disappear. Finally, as the temperature of the entire mass of the water reaches 100 °C, bubbles of steam reach the surface and boiling is said to occur. The steam in the flask may not be visible but as it comes out of the flask, it condenses as tiny droplets of water, giving a foggy appearance.

Fig. 11.11 Boiling process.

If now the steam outlet is closed for a few seconds to increase the pressure in the flask, you will notice that boiling stops. More heat would be required to raise the temperature (depending on the increase in pressure) before boiling begins again. Thus boiling point increases with increase in pressure.

Let us now remove the burner. Allow water to cool to about 80 °C. Remove the thermometer and steam outlet. Close the flask with the airtight cork. Keep the flask turned upside down on the stand. Pour ice-cold water on the flask. Water vapours in the flask condense reducing the pressure on the water surface inside the flask. Water begins to boil again, now at a lower temperature. Thus boiling point decreases with decrease in pressure.

This explains why cooking is difficult on hills. At high altitudes, atmospheric pressure is lower, reducing the boiling point of water as compared to that at sea level. On the other hand, boiling point is increased inside a pressure cooker by increasing the pressure. Hence cooking is faster. The boiling point of a substance at standard atmospheric pressure is called its **normal boiling point**.

However, all substances do not pass through the three states: solid-liquid-gas. There are certain substances which normally pass from the solid to the vapour state directly and vice versa. The change from solid state to vapour state without passing through the liquid state is called **sublimation**, and the substance is said to sublime. Dry ice (solid CO_2) sublimes, so also iodine. During the sublimation process both the solid and vapour states of a substance coexist in thermal equilibrium.

11.8.1 Latent Heat

In Section 11.8, we have learnt that certain amount of heat energy is transferred between a substance and its surroundings when it undergoes a change of state. The amount of heat per unit mass transferred during change of state of the substance is called latent heat of the substance for the process. For example, if heat is added to a given quantity of ice at –10 °C, the temperature of ice increases until it reaches its melting point (0 °C). At this temperature, the addition of more heat does not increase the temperature but causes the ice to melt, or changes its state. Once the entire ice melts, adding more heat will cause the temperature of the water to rise. A similar situation occurs during liquid gas change of state at the boiling point. Adding more heat to boiling water causes vaporisation, without increase in temperature.

Table 11.5 **Temperatures of the change of state and latent heats for various substances at 1 atm pressure**

Substance	Melting Point (°C)	L_f (10^5J kg^{-1})	Boiling Point (°C)	L_v (10^5J kg^{-1})
Ethyl alcohol	–114	1.0	78	8.5
Gold	1063	0.645	2660	15.8
Lead	328	0.25	1744	8.67
Mercury	–39	0.12	357	2.7
Nitrogen	–210	0.26	–196	2.0
Oxygen	–219	0.14	–183	2.1
Water	0	3.33	100	22.6

The heat required during a change of state depends upon the heat of transformation and the mass of the substance undergoing a change of state. Thus, if mass m of a substance undergoes a change from one state to the other, then the quantity of heat required is given by

$$Q = m\,L$$

or $$L = Q/m \qquad (11.13)$$

where L is known as latent heat and is a characteristic of the substance. Its SI unit is J kg^{-1}. The value of L also depends on the pressure. Its value is usually quoted at standard atmospheric pressure. The latent heat for a solid-liquid state change is called the **latent heat of fusion** (L_f), and that for a liquid-gas state change is called the **latent heat of vaporisation** (L_v). These are often referred to as the heat of fusion and the heat of vaporisation. A plot of temperature versus heat energy for a quantity of water is shown in Fig. 11.12. The latent heats of some substances, their freezing and boiling points, are given in Table 11.5.

Fig. 11.12 Temperature versus heat for water at 1 atm pressure (not to scale).

Note that when heat is added (or removed) during a change of state, the temperature remains constant. Note in Fig. 11.12 that the slopes of the phase lines are not all the same, which indicates that specific heats of the various states are not equal. For water, the latent heat of fusion and vaporisation are $L_f = 3.33 \times 10^5$ J kg^{-1} and $L_v = 22.6 \times 10^5$ J kg^{-1} respectively. That is 3.33×10^5 J of heat are needed to melt 1 kg of ice at 0 °C, and 22.6×10^5 J of heat are needed to convert 1 kg of water to steam at 100 °C. So, steam at 100 °C carries 22.6×10^5 J kg^{-1} more heat than water at 100 °C. This is why burns from steam are usually more serious than those from boiling water.

Example 11.4 When 0.15 kg of ice of 0 °C mixed with 0.30 kg of water at 50 °C in a container, the resulting temperature is 6.7 °C. Calculate the heat of fusion of ice. (s_{water} = 4186 J kg^{-1} K^{-1})

Answer

Heat lost by water = $ms_w\,(\theta_f - \theta_i)_w$

= (0.30 kg) (4186 J kg^{-1} K^{-1}) (50.0 °C – 6.7 °C)

= 54376.14 J

Heat required to melt ice = $m_2 L_f$ = (0.15 kg) L_f

Heat required to raise temperature of ice water to final temperature = $m_1 s_w\,(\theta_f - \theta_i)_1$

= (0.15 kg) (4186 J kg^{-1} K^{-1}) (6.7 °C – 0 °C)

= 4206.93 J

Heat lost = heat gained

54376.14 J = (0.15 kg) L_f + 4206.93 J

$L_f = 3.34 \times 10^5$ J kg^{-1}.

Example 11.5 Calculate the heat required to convert 3 kg of ice at –12 °C kept in a calorimeter to steam at 100 °C at atmospheric pressure. Given specific heat capacity of ice = 2100 J kg^{-1} K^{-1}, specific heat capacity of water = 4186 J kg^{-1} K^{-1}, latent heat of fusion of ice = 3.35×10^5 J kg^{-1} and latent heat of steam = 2.256×10^6 J kg^{-1}.

Answer We have

Mass of the ice, $m = 3$ kg

specific heat capacity of ice, s_{ice} = 2100 J kg^{-1} K^{-1}

specific heat capacity of water, s_{water} = 4186 J kg^{-1} K^{-1}

latent heat of fusion of ice, $L_{f\,ice}$ = 3.35×10^5 J kg^{-1}

latent heat of steam, L_{steam} = 2.256×10^6 J kg^{-1}

Now, Q = heat required to convert 3 kg of ice at –12 °C to steam at 100 °C,

Q_1 = heat required to convert ice at –12 °C to ice at 0 °C.

= $m\, s_{ice}\, \Delta T_1$ = (3 kg) (2100 J kg^{-1} K^{-1}) [0–(–12)]°C = 75600 J

Q_2 = heat required to melt ice at 0 °C to water at 0 °C

= $m L_{f\,ice}$ = (3 kg) (3.35×10^5 J kg^{-1})

= 1005000 J

Q_3 = heat required to convert water at 0 °C to water at 100 °C.

= $m s_w\, \Delta T_2$ = (3kg) (4186J kg^{-1} K^{-1}) (100 °C)

= 1255800 J

Q_4 = heat required to convert water at 100 °C to steam at 100 °C.

= $m\, L_{steam}$ = (3 kg) (2.256×10^6 J kg^{-1})

= 6768000 J

So, $Q = Q_1 + Q_2 + Q_3 + Q_4$

= 75600J + 1005000 J + 1255800 J + 6768000 J

= 9.1×10^6 J

11.9 HEAT TRANSFER

We have seen that heat is energy transfer from one system to another or from one part of a system to another part, arising due to temperature difference. What are the different ways by which this energy transfer takes place? There are three distinct modes of heat transfer : conduction, convection and radiation (Fig. 11.13).

Fig. 11.13 Heating by conduction, convection and radiation.

11.9.1 Conduction

Conduction is the mechanism of transfer of heat between two adjacent parts of a body because of their temperature difference. Suppose one end of a metallic rod is put in a flame, the other end of the rod will soon be so hot that you cannot hold it by your bare hands. Here heat transfer takes place by conduction from the hot end of the rod through its different parts to the other end. Gases are poor thermal conductors while liquids have conductivities intermediate between solids and gases.

Heat conduction may be described quantitatively as the time rate of heat flow in a material for a given temperature difference. Consider a metallic bar of length L and uniform cross section A with its two ends maintained at different temperatures. This can be done, for example, by putting the ends in thermal contact with large reservoirs at temperatures, say, T_C and T_D respectively (Fig. 11.14). Let us assume the ideal condition that the sides of the bar are fully insulated so that no heat is exchanged between the sides and the surroundings.

After sometime, a steady state is reached; the temperature of the bar decreases uniformly with distance from T_C to T_D; ($T_C > T_D$). The reservoir at C supplies heat at a constant rate, which transfers through the bar and is given out at the same rate to the reservoir at D. It is found

Fig. 11.14 Steady state heat flow by conduction in a bar with its two ends maintained at temperatures T_C and T_D; ($T_C > T_D$).

experimentally that in this steady state, the rate of flow of heat (or heat current) H is proportional to the temperature difference ($T_C - T_D$) and the area of cross section A and is inversely proportional to the length L :

$$H = KA \frac{T_C - T_D}{L} \tag{11.14}$$

The constant of proportionality K is called the **thermal conductivity** of the material. The greater the value of K for a material, the more rapidly will it conduct heat. The SI unit of K is J S^{-1} m^{-1} K^{-1} or W m^{-1} K^{-1}. The thermal conductivities of various substances are listed in Table 11.5. These values vary slightly with temperature, but can be considered to be constant over a normal temperature range.

Compare the relatively large thermal conductivities of the good thermal conductors, the metals, with the relatively small thermal conductivities of some good thermal insulators, such as wood and glass wool. You may have noticed that some cooking pots have copper coating on the bottom. Being a good conductor of heat, copper promotes the distribution of heat over the bottom of a pot for uniform cooking. Plastic foams, on the other hand, are good insulators, mainly because they contain pockets of air. Recall that gases are poor conductors, and note the low thermal conductivity of air in the Table 11.5. Heat retention and transfer are important in many other applications. Houses made of concrete roofs get very hot during summer days, because thermal conductivity of concrete (though much smaller than that of a metal) is still not small enough. Therefore, people usually prefer to give a layer of earth or foam insulation on the ceiling so that heat transfer is prohibited and keeps the room cooler. In some situations, heat transfer is critical. In a nuclear reactor, for example, elaborate heat transfer systems need to be installed so that the enormous energy produced by nuclear fission in the core transits out sufficiently fast, thus preventing the core from overheating.

Table 11.6 Thermal conductivities of some materials

Materials	Thermal conductivity (J s^{-1} m^{-1} K^{-1})
Metals	
Silver	406
Copper	385
Aluminium	205
Brass	109
Steel	50.2
Lead	34.7
Mercury	8.3
Non-metals	
Insulating brick	0.15
Concrete	0.8
Body fat	0.20
Felt	0.04
Glass	0.8
Ice	1.6
Glass wool	0.04
Wood	0.12
Water	0.8
Gases	
Air	0.024
Argon	0.016
Hydrogen	0.14

Example 11.6 What is the temperature of the steel-copper junction in the steady state of the system shown in Fig. 11.15. Length of the steel rod = 15.0 cm, length of the copper rod = 10.0 cm, temperature of the furnace = 300 °C, temperature of the other end = 0 °C. The area of cross section of the steel rod is twice that of the copper rod. (Thermal conductivity of steel = 50.2 J s^{-1} $m^{-1}K^{-1}$; and of copper = 385 J $s^{-1}m^{-1}K^{-1}$).

Fig. 11.15

Answer The insulating material around the rods reduces heat loss from the sides of the rods. Therefore, heat flows only along the length of the rods. Consider any cross section of the rod. In the steady state, heat flowing into the element must equal the heat flowing out of it; otherwise there would be a net gain or loss of heat by the element and its temperature would not be steady. Thus in the steady state, rate of heat flowing across a cross section of the rod is the same at every point along the length of the combined steel-copper rod. Let T be the temperature of the steel-copper junction in the steady state. Then,

$$\frac{K_1 A_1 \quad 300 \quad T}{L_1} = \frac{K_2 A_2 \quad T - 0}{L_2}$$

where 1 and 2 refer to the steel and copper rod respectively. For $A_1 = 2 A_2$, $L_1 = 15.0$ cm, $L_2 = 10.0$ cm, $K_1 = 50.2$ J s^{-1} m^{-1} K^{-1}, $K_2 = 385$ J s^{-1} m^{-1} K^{-1}, we have

$$\frac{50.2 \times 2\,(300 - T)}{15} = \frac{385T}{10}$$

which gives $T = 44.4$ °C ◀

▶ ***Example 11.7*** An iron bar ($L_1 = 0.1$ m, $A_1 = 0.02$ m^2, $K_1 = 79$ W m^{-1} K^{-1}) and a brass bar ($L_2 = 0.1$ m, $A_2 = 0.02$ m^2, $K_2 = 109$ W m^{-1}K^{-1}) are soldered end to end as shown in Fig. 11.16. The free ends of the iron bar and brass bar are maintained at 373 K and 273 K respectively. Obtain expressions for and hence compute (i) the temperature of the junction of the two bars, (ii) the equivalent thermal conductivity of the compound bar, and (iii) the heat current through the compound bar.

Answer

Fig 11.16

Given, $L_1 = L_2 = L = 0.1$ m, $A_1 = A_2 = A = 0.02$ m^2 $K_1 = 79$ W m^{-1} K^{-1}, $K_2 = 109$ W m^{-1} K^{-1}, $T_1 = 373$ K, and $T_2 = 273$ K.

Under steady state condition, the heat current (H_1) through iron bar is equal to the heat current (H_2) through brass bar.

So, $H = H_1 = H_2$

$$= \frac{K_1 A_1 \quad T_1 - T_0}{L_1} \quad \frac{K_2 A_2 (T_0 - T_2)}{L_2}$$

For $A_1 = A_2 = A$ and $L_1 = L_2 = L$, this equation leads to

$K_1 (T_1 - T_0) = K_2 (T_0 - T_2)$

Thus the junction temperature T_0 of the two bars is

$$T_0 = \frac{K_1 T_1 \quad K_2 T_2}{K_1 \quad K_2}$$

Using this equation, the heat current H through either bar is

$$H = \frac{K_1 A \quad T_1 - T_0}{L} \quad \frac{K_2 A (T_0 - T_2)}{L}$$

$$\frac{K_1 K_2}{K_1 \quad K_2} \quad \frac{A \quad T_1 - T_0}{L} \quad \frac{A \quad T_1 - T_2}{L \quad \frac{1}{K_1} \quad \frac{1}{K_2}}$$

Using these equations, the heat current H' through the compound bar of length $L_1 + L_2 = 2L$ and the equivalent thermal conductivity K', of the compound bar are given by

$$H \quad \frac{K \quad A \quad T_1 - T_2}{2\,L} \quad H$$

$$K \quad \frac{2\,K_1 \quad K_2}{K_1 \quad K_2}$$

(i) $$T_0 \quad \frac{K_1 T_1 \quad K_2 T_2}{K_1 \quad K_2}$$

$$\frac{79\,W\ \text{m}^{-1}\text{K}^{-1} \quad 373\,\text{K} \quad 109\ \text{W m}^{-1}\text{K}^{-1} \quad 273\,\text{K}}{79\ \text{W m}^{-1}\text{K}^{-1} \quad 109\ \text{W m}^{-1}\text{K}^{-1}}$$

$= 315$ K

(ii) $$K = \frac{2K_1 \quad K_2}{K_1 \quad K_2}$$

$$= \frac{2 \times (79\ \text{W m}^{-1}\ \text{K}^{-1}) \times (109\ \text{W m}^{-1}\ \text{K}^{-1})}{79\ \text{W m}^{-1}\ \text{K}^{-1} + 109\ \text{W m}^{-1}\ \text{K}^{-1}}$$

$= 91.6$ W m^{-1} K^{-1}

(iii) $H \quad H \quad \dfrac{K \;\; A \;\; T_1 - T_2}{2 \;\; L}$

$$\frac{91.6\ \text{W m}^{-1}\ \text{K}^{-1} \times 0.02\ \text{m}^2 \times 373\,\text{K} - 273\,\text{K}}{2 \times 0.1\ \text{m}}$$

$= 916.1$ W ◀

11.9.2 Convection

Convection is a mode of heat transfer by actual motion of matter. It is possible only in fluids. Convection can be natural or forced. In natural convection, gravity plays an important part. When a fluid is heated from below, the hot part expands and, therefore, becomes less dense. Because of buoyancy, it rises and the upper colder part replaces it. This again gets heated, rises up and is replaced by the colder part of the fluid. The process goes on. This mode of heat transfer is evidently different from conduction. Convection involves bulk transport of different parts of the fluid. In forced convection, material is forced to move by a pump or by some other physical means. The common examples of forced convection systems are forced-air heating systems in home, the human circulatory system, and the cooling system of an automobile engine. In the human body, the heart acts as the pump that circulates blood through different parts of the body, transferring heat by forced convection and maintaining it at a uniform temperature.

Natural convection is responsible for many familiar phenomena. During the day, the ground heats up more quickly than large bodies of water do. This occurs both because the water has a greater specific heat and because mixing currents disperse the absorbed heat throughout the great volume of water. The air in contact with the warm ground is heated by conduction. It expands, becoming less dense than the surrounding cooler air. As a result, the warm air rises (air currents) and other air moves (winds) to fill the space-creating a sea breeze near a large body of water. Cooler air descends, and a thermal convection cycle is set up, which transfers heat away from the land. At night, the ground loses its heat more quickly, and the water surface is warmer than the land. As a result, the cycle is reveresed (Fig. 11.17).

The other example of natural convection is the steady surface wind on the earth blowing in from north-east towards the equator, the so called trade wind. A resonable explanation is as follows : the equatorial and polar regions of the earth receive unequal solar heat. Air at the earth's surface near the equator is hot while the air in the upper atmosphere of the poles is cool. In the absence of any other factor, a convection current would be set up, with the air at the equatorial surface rising and moving out towards the poles, descending and streaming in towards the equator. The rotation of the earth, however, modifies this convection current. Because of this, air close to the equator has an eastward speed of 1600 km/h, while it is zero close to the poles. As a result, the air descends not at the poles but at 30° N (North) latitude and returns to the equator. This is called **trade wind**.

Fig. 11.17 Convection cycles.

11.9.3 Radiation

Conduction and convection require some material as a transport medium. These modes of heat transfer cannot operate between bodies separated by a distance in vacuum. But the earth does receive heat from the sun across a huge distance and we quickly feel the warmth of the fire nearby even though air conducts poorly and before convection can set in. The third mechanism for heat transfer needs no medium; it is called radiation and the energy so radiated by electromagnetic waves is called radiant energy. In an electromagnetic wave electric and magnetic fields oscillate in space and time. Like any wave, electromagnetic waves can have different wavelengths and can travel in vacuum with the same speed, namely the speed of light i.e., 3×10^8 m s^{-1}. You will learn these matters in more details later, but you now know why heat transfer by radiation does not need any medium and why it is so fast. This is how heat is transfered to the earth from the sun through empty space. All bodies emit radiant energy, whether they are solid, liquid or gases. The electromagnetic radiation emitted by a body by virtue of its temperature like the radiation by a red hot iron or light from a filament lamp is called thermal radiation.

When this thermal radiation falls on other bodies, it is partly reflected and partly absorbed. The amount of heat that a body can absorb by radiation depends on the colour of the body.

We find that black bodies absorb and emit radiant energy better than bodies of lighter colours. This fact finds many applications in our daily life. We wear white or light coloured clothes in summer so that they absorb the least heat from the sun. However, during winter, we use dark coloured clothes which absorb heat from the sun and keep our body warm. The bottoms of the utensils for cooking food are blackened so that they absorb maximum heat from the fire and give it to the vegetables to be cooked.

Similarly, a Dewar flask or thermos bottle is a device to minimise heat transfer between the contents of the bottle and outside. It consists of a double-walled glass vessel with the inner and outer walls coated with silver. Radiation from the inner wall is reflected back into the contents of the bottle. The outer wall similarly reflects back any incoming radiation. The space between the walls is evacuted to reduce conduction and convection losses and the flask is supported on an insulator like cork. The device is, therefore, useful for preventing hot contents (like milk) from getting cold, or alternatively to store cold contents (like ice).

11.10 NEWTON'S LAW OF COOLING

We all know that hot water or milk when left on a table begins to cool gradually. Ultimately it attains the temperature of the surroundings. To study how a given body can cool on exchanging heat with its surroundings, let us perform the following activity.

Take some water, say 300 ml, in a calorimeter with a stirrer and cover it with two holed lid. Fix a thermometer through a hole in the lid and make sure that the bulb of thermometer is immersed in the water. Note the reading of the thermometer. This reading T_1 is the temperature of the surroundings. Heat the water kept in the calorimeter till it attains a temperature, say, 40 °C above room temperature (i.e., temperature of the surroundings). Then stop heating the water by removing the heat source. Start the stop-watch and note the reading of the thermometer after fixed interval of time, say after every one minute of stirring gently with the stirrer. Continue to note the temperature (T_2) of water till it attains a temperature about 5 °C above that of the surroundings. Then plot a graph by taking each value of temperature $\Delta T = T_2 - T_1$ along y axis and the coresponding value of t along x-axis (Fig. 11.18).

Fig. 11.18 Curve showing cooling of hot water with time.

From the graph you will infer how the cooling of hot water depends on the difference of its temperature from that of the surroundings. You will also notice that initially the rate of cooling is higher and decreases as the temperature of the body falls.

The above activity shows that a hot body loses heat to its surroundings in the form of heat radiation. The rate of loss of heat depends on the difference in temperature between the body and its surroundings. Newton was the first to study, in a systematic manner, the relation between the heat lost by a body in a given enclosure and its temperature.

According to Newton's law of cooling, the rate of loss of heat, $-\mathrm{d}Q/\mathrm{d}t$ of the body is directly proportional to the difference of temperature $\Delta T = (T_2 - T_1)$ of the body and the surroundings. The law holds good only for small difference of temperature. Also, the loss of heat by radiation depends upon the nature of the surface of the body and the area of the exposed surface. We can write

$$-\frac{dQ}{dt} = k(T_2 \quad T_1) \tag{11.15}$$

where k is a positive constant depending upon the area and nature of the surface of the body. Suppose a body of mass m and specific heat capacity s is at temperature T_2. Let T_1 be the temperature of the surroundings. If the temperature falls by a small amount $\mathrm{d}T_2$ in time $\mathrm{d}t$, then the amount of heat lost is

$\mathrm{d}Q = ms\,\mathrm{d}T_2$

$\therefore$Rate of loss of heat is given by

$$\frac{dQ}{dt} = ms\frac{dT_2}{dt} \tag{11.16}$$

From Eqs. (11.15) and (11.16) we have

$$-ms\frac{dT_2}{dt} = k(T_2 - T_1)$$

$$\frac{dT_2}{T_2 - T_1} = -\frac{k}{ms}dt = -K\,dt \tag{11.17}$$

where $K = k/m\,s$

On integrating,

$$\log_e (T_2 - T_1) = -K\,t + c \tag{11.18}$$

$$\text{or } T_2 = T_1 + C'\,e^{-Kt};\ \text{where } C' = e^c \tag{11.19}$$

Equation (11.19) enables you to calculate the time of cooling of a body through a particular range of temperature.

For small temperature differences, the rate of cooling, due to conduction, convection, and radiation combined, is proportional to the difference in temperature. It is a valid approximation in the transfer of heat from a radiator to a room, the loss of heat through the wall of a room, or the cooling of a cup of tea on the table.

Fig. 11.19 Verification of Newton's Law of cooling.

Newton's law of cooling can be verified with the help of the experimental set-up shown in Fig. 11.19(a). The set-up consists of a double walled vessel (V) containing water in between the two walls. A copper calorimeter (C) containing hot water is placed inside the double walled vessel. Two thermometers through the corks are used to note the temperatures T_2 of water in calorimeter and T_1 of hot water in between the double walls respectively. Temperature of hot water in the calorimeter is noted after equal intervals of time. A graph is plotted between $\log_e (T_2 - T_1)$ and time (t). The nature of the graph is observed to be a straight line having a negative slope as shown in Fig. 11.19(b). This is in support of Eq. (11.18).

Example 11.8 A pan filled with hot food cools from 94 °C to 86 °C in 2 minutes when the room temperature is at 20 °C. How long will it take to cool from 71 °C to 69 °C?

Answer The average temperature of 94 °C and 86 °C is 90 °C, which is 70 °C above the room temperature. Under these conditions the pan cools 8 °C in 2 minutes.

Using Eq. (11.17), we have

$$\frac{\text{Change in temperature}}{\text{Time}} \quad K \quad T$$

$$\frac{8°C}{2\ \text{min}} = K(70°C)$$

The average of 69 °C and 71 °C is 70 °C, which is 50 °C above room temperature. K is the same for this situation as for the original.

$$\frac{2°C}{\text{Time}} = K\ (50\ °C)$$

When we divide above two equations, we have

$$\frac{8\ \ C/2\ \text{min}}{2\ \ C/\text{time}} = \frac{K\ (70\ \ C)}{K\ (50\ \ C)}$$

Time = 0.7 min

= 42 s

SUMMARY

1. Heat is a form of energy that flows between a body and its surrounding medium by virtue of temperature difference between them. The degree of hotness of the body is quantitatively represented by temperature.
2. A temperature-measuring device (thermometer) makes use of some measurable property (called thermometric property) that changes with temperature. Different thermometers lead to different temperature scales. To construct a temperature scale, two fixed points are chosen and assigned some arbitrary values of temperature. The two numbers fix the origin of the scale and the size of its unit.
3. The Celsius temperature (t_C) and the Farenheit temperare (t_F)are related by

$$t_F = (9/5)\ t_C + 32$$

4. The ideal gas equation connecting pressure (P), volume (V) and absolute temperature (T) is :

$$PV = \mu RT$$

where μ is the number of moles and R is the universal gas constant.

5. In the absolute temperature scale, the zero of the scale is the absolute zero of temperature – the temperature where every substance in nature has the least possible molecular activity. The Kelvin absolute temperature scale (T) has the same unit size as the Celsius scale (T_c), but differs in the origin :

$$T_C = T - 273.15$$

6. The coefficient of linear expansion (α_l) and volume expansion (α_v) are defined by the relations :

$$\frac{l}{l} \quad {}_l \quad T$$

$$\frac{V}{V} \quad {}_V \quad T$$

where Δl and ΔV denote the change in length l and volume V for a change of temperature ΔT. The relation between them is :

$$\alpha_v = 3\ \alpha_l$$

7. The specific heat capacity of a substance is defined by

$$s = \frac{1}{m}\frac{\Delta Q}{\Delta T}$$

where m is the mass of the substance and ΔQ is the heat required to change its temperature by ΔT. The molar specific heat capacity of a substance is defined by

$$C = \frac{1}{\mu}\frac{\Delta Q}{\Delta T}$$

where μ is the number of moles of the substance.

8. The latent heat of fusion (L_f) is the heat per unit mass required to change a substance from solid into liquid at the same temperature and pressure. The latent heat of vaporisation (L_v) is the heat per unit mass required to change a substance from liquid to the vapour state without change in the temperature and pressure.
9. The three modes of heat transfer are conduction, convection and radiation.
10. In conduction, heat is transferred between neighbouring parts of a body through molecular collisions, without any flow of matter. For a bar of length L and uniform cross section A with its ends maintained at temperatures T_C and T_D, the rate of flow of heat H is :

$$H = K\,A\,\frac{T_C - T_D}{L}$$

where K is the thermal conductivity of the material of the bar.

11. Newton's Law of Cooling says that the rate of cooling of a body is proportional to the excess temperature of the body over the surroundings :

$$\frac{dQ}{dt} = -k\,(T_2 - T_1)$$

Where T_1 is the temperature of the surrounding medium and T_2 is the temperature of the body.

Quantity	Symbol	Dimensions	Unit	Remark
Amount of substance	μ	[mol]	mol	
Celsius temperature	t_c	[K]	°C	
Kelvin absolute temperature	T	[K]	K	$t_c = T - 273.15$
Co-efficient of linear expansion	α_l	$[K^{-1}]$	K^{-1}	
Co-efficient of volume expansion	α_v	$[K^{-1}]$	K^{-1}	$\alpha_v = 3\,\alpha_l$
Heat supplied to a system	ΔQ	$[ML^2T^{-2}]$	J	Q is not a state variable
Specific heat	s	$[L^2T^{-2}K^{-1}]$	$J\,kg^{-1}\,K^{-1}$	
Thermal Conductivity	K	$[M\,LT^{-3}K^{-1}]$	$J\,s^{-1}K^{-1}$	$H = -KA\,\frac{dT}{dx}$

POINTS TO PONDER

1. The relation connecting Kelvin temperature (T) and the Celsius temperature t_c

$$T = t_c + 273.15$$

and the assignment T = 273.16 K for the triple point of water are exact relations (by choice). With this choice, the Celsius temperature of the melting point of water and boiling point of water (both at 1 atm pressure) are very close to, but not exactly equal to 0 °C and 100 °C respectively. In the original Celsius scale, these latter fixed points were exactly at 0 °C and 100 °C (by choice), but now the triple point of water is the preferred choice for the fixed point, because it has a unique temperature.

2. A liquid in equilibrium with vapour has the same pressure and temperature throughout the system; the two phases in equilibrium differ in their molar volume (i.e. density). This is true for a system with any number of phases in equilibrium.

3. Heat transfer always involves temperature difference between two systems or two parts of the same system. Any energy transfer that does not involve temperature difference in some way is not heat.

4. Convection involves flow of matter *within a fluid* due to unequal temperatures of its parts. A hot bar placed under a running tap loses heat by conduction between the surface of the bar and water and not by convection within water.

EXERCISES

11.1 The triple points of neon and carbon dioxide are 24.57 K and 216.55 K respectively. Express these temperatures on the Celsius and Fahrenheit scales.

11.2 Two absolute scales A and B have triple points of water defined to be 200 A and 350 B. What is the relation between T_A and T_B ?

11.3 The electrical resistance in ohms of a certain thermometer varies with temperature according to the approximate law :

$$R = R_o [1 + \alpha (T - T_o)]$$

The resistance is 101.6 Ω at the triple-point of water 273.16 K, and 165.5 Ω at the normal melting point of lead (600.5 K). What is the temperature when the resistance is 123.4 Ω ?

11.4 Answer the following :

(a) The triple-point of water is a standard fixed point in modern thermometry. Why ? What is wrong in taking the melting point of ice and the boiling point of water as standard fixed points (as was originally done in the Celsius scale) ?

(b) There were two fixed points in the original Celsius scale as mentioned above which were assigned the number 0 °C and 100 °C respectively. On the absolute scale, one of the fixed points is the triple-point of water, which on the Kelvin absolute scale is assigned the number 273.16 K. What is the other fixed point on this (Kelvin) scale ?

(c) The absolute temperature (Kelvin scale) T is related to the temperature t_c on the Celsius scale by

$$t_c = T - 273.15$$

Why do we have 273.15 in this relation, and not 273.16 ?

(d) What is the temperature of the triple-point of water on an absolute scale whose unit interval size is equal to that of the Fahrenheit scale ?

11.5 Two ideal gas thermometers *A* and *B* use oxygen and hydrogen respectively. The following observations are made :

Temperature	**Pressure thermometer A**	**Pressure thermometer B**
Triple-point of water	1.250×10^5 Pa	0.200×10^5 Pa
Normal melting point of sulphur	1.797×10^5 Pa	0.287×10^5 Pa

(a) What is the absolute temperature of normal melting point of sulphur as read by thermometers *A* and *B* ?

(b) What do you think is the reason behind the slight difference in answers of thermometers *A* and *B* ? (The thermometers are not faulty). What further procedure is needed in the experiment to reduce the discrepancy between the two readings ?

11.6 A steel tape 1m long is correctly calibrated for a temperature of 27.0 °C. The length of a steel rod measured by this tape is found to be 63.0 cm on a hot day when the temperature is 45.0 °C. What is the actual length of the steel rod on that day ? What is the length of the same steel rod on a day when the temperature is 27.0 °C ? Coefficient of linear expansion of steel = 1.20×10^{-5} K^{-1} .

11.7 A large steel wheel is to be fitted on to a shaft of the same material. At 27 °C, the outer diameter of the shaft is 8.70 cm and the diameter of the central hole in the wheel is 8.69 cm. The shaft is cooled using 'dry ice'. At what temperature of the shaft does the wheel slip on the shaft? Assume coefficient of linear expansion of the steel to be constant over the required temperature range :
$\alpha_{steel} = 1.20 \times 10^{-5}$ K^{-1}.

11.8 A hole is drilled in a copper sheet. The diameter of the hole is 4.24 cm at 27.0 °C. What is the change in the diameter of the hole when the sheet is heated to 227 °C? Coefficient of linear expansion of copper = 1.70×10^{-5} K^{-1}.

11.9 A brass wire 1.8 m long at 27 °C is held taut with little tension between two rigid supports. If the wire is cooled to a temperature of –39 °C, what is the tension developed in the wire, if its diameter is 2.0 mm ? Co-efficient of linear expansion of brass = 2.0×10^{-5} K^{-1}; Young's modulus of brass = 0.91×10^{11} Pa.

11.10 A brass rod of length 50 cm and diameter 3.0 mm is joined to a steel rod of the same length and diameter. What is the change in length of the combined rod at 250 °C, if the original lengths are at 40.0 °C? Is there a 'thermal stress' developed at the junction ? The ends of the rod are free to expand (Co-efficient of linear expansion of brass = 2.0×10^{-5} K^{-1}, steel = 1.2×10^{-5} K^{-1}).

11.11 The coefficient of volume expansion of glycerin is 49×10^{-5} K^{-1}. What is the fractional change in its density for a 30 °C rise in temperature ?

11.12 A 10 kW drilling machine is used to drill a bore in a small aluminium block of mass 8.0 kg. How much is the rise in temperature of the block in 2.5 minutes, assuming 50% of power is used up in heating the machine itself or lost to the surroundings. Specific heat of aluminium = 0.91 J g^{-1} K^{-1}.

11.13 A copper block of mass 2.5 kg is heated in a furnace to a temperature of 500 °C and then placed on a large ice block. What is the maximum amount of ice that can melt? (Specific heat of copper = 0.39 J g^{-1} K^{-1}; heat of fusion of water = 335 J g^{-1}).

11.14 In an experiment on the specific heat of a metal, a 0.20 kg block of the metal at 150 °C is dropped in a copper calorimeter (of water equivalent 0.025 kg) containing 150 cm^3 of water at 27 °C. The final temperature is 40 °C. Compute the specific

heat of the metal. If heat losses to the surroundings are not negligible, is your answer greater or smaller than the actual value for specific heat of the metal ?

11.15 Given below are observations on molar specific heats at room temperature of some common gases.

Gas	**Molar specific heat (C_v) (cal mol^{-1} K^{-1})**
Hydrogen	4.87
Nitrogen	4.97
Oxygen	5.02
Nitric oxide	4.99
Carbon monoxide	5.01
Chlorine	6.17

The measured molar specific heats of these gases are markedly different from those for monatomic gases. Typically, molar specific heat of a monatomic gas is 2.92 cal/mol K. Explain this difference. What can you infer from the somewhat larger (than the rest) value for chlorine ?

11.16 Answer the following questions based on the *P-T* phase diagram of carbon dioxide:

(a) At what temperature and pressure can the solid, liquid and vapour phases of CO_2 co-exist in equilibrium ?

(b) What is the effect of decrease of pressure on the fusion and boiling point of CO_2 ?

(c) What are the critical temperature and pressure for CO_2 ? What is their significance ?

(d) Is CO_2 solid, liquid or gas at (a) –70 °C under 1 atm, (b) –60 °C under 10 atm, (c) 15 °C under 56 atm ?

11.17 Answer the following questions based on the *P – T* phase diagram of CO_2:

(a) CO_2 at 1 atm pressure and temperature – 60 °C is compressed isothermally. Does it go through a liquid phase ?

(b) What happens when CO_2 at 4 atm pressure is cooled from room temperature at constant pressure ?

(c) Describe qualitatively the changes in a given mass of solid CO_2 at 10 atm pressure and temperature –65 °C as it is heated up to room temperature at constant pressure.

(d) CO_2 is heated to a temperature 70 °C and compressed isothermally. What changes in its properties do you expect to observe ?

11.18 A child running a temperature of 101°F is given an antipyrin (i.e. a medicine that lowers fever) which causes an increase in the rate of evaporation of sweat from his body. If the fever is brought down to 98 °F in 20 min, what is the average rate of extra evaporation caused, by the drug. Assume the evaporation mechanism to be the only way by which heat is lost. The mass of the child is 30 kg. The specific heat of human body is approximately the same as that of water, and latent heat of evaporation of water at that temperature is about 580 cal g^{-1}.

11.19 A 'thermacole' icebox is a cheap and efficient method for storing small quantities of cooked food in summer in particular. A cubical icebox of side 30 cm has a thickness of 5.0 cm. If 4.0 kg of ice is put in the box, estimate the amount of ice remaining after 6 h. The outside temperature is 45 °C, and co-efficient of thermal conductivity of thermacole is 0.01 J s^{-1} m^{-1} K^{-1}. [Heat of fusion of water = 335 × 10^3 J kg^{-1}]

11.20 A brass boiler has a base area of 0.15 m^2 and thickness 1.0 cm. It boils water at the rate of 6.0 kg/min when placed on a gas stove. Estimate the temperature of the part

of the flame in contact with the boiler. Thermal conductivity of brass = 109 J s^{-1} m^{-1} K^{-1}; Heat of vaporisation of water = 2256×10^3 J kg^{-1}.

11.21 Explain why :

(a) a body with large reflectivity is a poor emitter

(b) a brass tumbler feels much colder than a wooden tray on a chilly day

(c) an optical pyrometer (for measuring high temperatures) calibrated for an ideal black body radiation gives too low a value for the temperature of a red hot iron piece in the open, but gives a correct value for the temperature when the same piece is in the furnace

(d) the earth without its atmosphere would be inhospitably cold

(e) heating systems based on circulation of steam are more efficient in warming a building than those based on circulation of hot water

11.22 A body cools from 80 °C to 50 °C in 5 minutes. Calculate the time it takes to cool from 60 °C to 30 °C. The temperature of the surroundings is 20 °C.

CHAPTER TWELVE

THERMODYNAMICS

12.1 Introduction
12.2 Thermal equilibrium
12.3 Zeroth law of Thermodynamics
12.4 Heat, internal energy and work
12.5 First law of thermodynamics
12.6 Specific heat capacity
12.7 Thermodynamic state variables and equation of state
12.8 Thermodynamic processes
12.9 Heat engines
12.10 Refrigerators and heat pumps
12.11 Second law of thermodynamics
12.12 Reversible and irreversible processes
12.13 Carnot engine
Summary
Points to ponder
Exercises

12.1 INTRODUCTION

In previous chapter we have studied thermal properties of matter. In this chapter we shall study laws that govern thermal energy. We shall study the processes where work is converted into heat and vice versa. In winter, when we rub our palms together, we feel warmer; here work done in rubbing produces the 'heat'. Conversely, in a steam engine, the 'heat' of the steam is used to do useful work in moving the pistons, which in turn rotate the wheels of the train.

In physics, we need to define the notions of heat, temperature, work, etc. more carefully. Historically, it took a long time to arrive at the proper concept of 'heat'. Before the modern picture, heat was regarded as a fine invisible fluid filling in the pores of a substance. On contact between a hot body and a cold body, the fluid (called caloric) flowed from the colder to the hotter body ! This is similar to what happens when a horizontal pipe connects two tanks containing water up to different heights. The flow continues until the levels of water in the two tanks are the same. Likewise, in the 'caloric' picture of heat, heat flows until the 'caloric levels' (i.e., the temperatures) equalise.

In time, the picture of heat as a fluid was discarded in favour of the modern concept of heat as a form of energy. An important experiment in this connection was due to Benjamin Thomson (also known as Count Rumford) in 1798. He observed that boring of a brass cannon generated a lot of heat, indeed enough to boil water. More significantly, the amount of heat produced depended on the work done (by the horses employed for turning the drill) but not on the sharpness of the drill. In the caloric picture, a sharper drill would scoop out more heat fluid from the pores; but this was not observed. A most natural explanation of the observations was that heat was a form of energy and the experiment demonstrated conversion of energy from one form to another–from work to heat.

Thermodynamics is the branch of physics that deals with the concepts of heat and temperature and the inter-conversion of heat and other forms of energy. Thermodynamics is a macroscopic science. It deals with bulk systems and does not go into the molecular constitution of matter. In fact, its concepts and laws were formulated in the nineteenth century before the molecular picture of matter was firmly established. Thermodynamic description involves relatively few macroscopic variables of the system, which are suggested by common sense and can be usually measured directly. A microscopic description of a gas, for example, would involve specifying the co-ordinates and velocities of the huge number of molecules constituting the gas. The description in kinetic theory of gases is not so detailed but it does involve molecular distribution of velocities. Thermodynamic description of a gas, on the other hand, avoids the molecular description altogether. Instead, the state of a gas in thermodynamics is specified by macroscopic variables such as pressure, volume, temperature, mass and composition that are felt by our sense perceptions and are measurable*.

The distinction between mechanics and thermodynamics is worth bearing in mind. In mechanics, our interest is in the motion of particles or bodies under the action of forces and torques. Thermodynamics is not concerned with the motion of the system as a whole. It is concerned with the internal macroscopic state of the body. When a bullet is fired from a gun, what changes is the mechanical state of the bullet (its kinetic energy, in particular), not its temperature. When the bullet pierces a wood and stops, the kinetic energy of the bullet gets converted into heat, changing the temperature of the bullet and the surrounding layers of wood. Temperature is related to the energy of the internal (disordered) motion of the bullet, not to the motion of the bullet as a whole.

12.2 THERMAL EQUILIBRIUM

Equilibrium in mechanics means that the net external force and torque on a system are zero. The term 'equilibrium' in thermodynamics appears in a different context : we say the state of a system is an equilibrium state if the macroscopic variables that characterise the system do not change in time. For example, a gas inside a closed rigid container, completely insulated from its surroundings, with fixed values of pressure, volume, temperature, mass and composition that do not change with time, is in a state of thermodynamic equilibrium.

In general, whether or not a system is in a state of equilibrium depends on the surroundings and the nature of the wall that separates the system from the surroundings. Consider two gases A and B occupying two different containers. We know experimentally that pressure and volume of a given mass of gas can be chosen to be its two independent variables. Let the pressure and volume of the gases be (P_A, V_A) and (P_B, V_B) respectively. Suppose first that the two systems are put in proximity but are separated by an **adiabatic wall** – an insulating wall (can be movable) that does not allow flow of energy (heat) from one to another. The systems are insulated from the rest of the surroundings also by similar adiabatic walls. The situation is shown schematically in Fig. 12.1 (a). In this case, it is found that any possible pair of values (P_A, V_A) will be in equilibrium with any possible pair of values (P_B, V_B). Next, suppose that the adiabatic wall is replaced by a **diathermic wall** – a conducting wall that allows energy flow (heat) from one to another. It is then found that the macroscopic variables of the systems A and B change spontaneously until both the systems attain equilibrium states. After that there is no change in their states. The situation is shown in Fig. 12.1(b). The pressure and volume variables of the two gases change to (P_B', V_B') and (P_A', V_A') such that the new states of A and B are in equilibrium with each other**. There is no more energy flow from one to another. We then say that the system A is in thermal equilibrium with the system B.

What characterises the situation of thermal equilibrium between two systems ? You can guess the answer from your experience. In thermal equilibrium, the temperatures of the two systems

* *Thermodynamics may also involve other variables that are not so obvious to our senses e.g. entropy, enthalpy, etc., and they are all macroscopic variables.*

** *Both the variables need not change. It depends on the constraints. For instance, if the gases are in containers of fixed volume, only the pressures of the gases would change to achieve thermal equilibrium.*

are equal. We shall see how does one arrive at the concept of temperature in thermodynamics? The Zeroth law of thermodynamics provides the clue.

12.3 ZEROTH LAW OF THERMODYNAMICS

Imagine two systems *A* and *B*, separated by an adiabatic wall, while each is in contact with a third system *C*, via a conducting wall [Fig. 12.2(a)]. The states of the systems (i.e., their macroscopic variables) will change until both *A* and *B* come to thermal equilibrium with *C*. After this is achieved, suppose that the adiabatic wall between *A* and *B* is replaced by a conducting wall and *C* is insulated from *A* and *B* by an adiabatic wall [Fig. 12.2(b)]. It is found that the states of *A* and *B* change no further i.e. they are found **to be in thermal equilibrium with each other**. This observation forms the basis of the **Zeroth Law of Thermodynamics**, which states that **'two systems in thermal equilibrium with a third system separately are in thermal equilibrium with each other'**. R.H. Fowler formulated this law in 1931 long after the first and second Laws of thermodynamics were stated and so numbered.

Fig. 12.1 *(a) Systems A and B (two gases) separated by an adiabatic wall – an insulating wall that does not allow flow of heat. (b) The same systems A and B separated by a diathermic wall – a conducting wall that allows heat to flow from one to another. In this case, thermal equilibrium is attained in due course.*

The Zeroth Law clearly suggests that when two systems *A* and *B*, are in thermal equilibrium, there must be a physical quantity that has the same value for both. This thermodynamic variable whose value is equal for two systems in thermal equilibrium is called temperature (T). Thus, if *A* and *B* are separately in equilibrium with *C*, $T_A = T_C$ and $T_B = T_C$. This implies that $T_A = T_B$ i.e. the systems *A* and *B* are also in thermal equilibrium.

We have arrived at the concept of temperature formally via the Zeroth Law. The next question is : how to assign numerical values to temperatures of different bodies ? In other words, how do we construct a scale of temperature ? Thermometry deals with this basic question to which we turn in the next section.

Fig. 12.2 *(a) Systems A and B are separated by an adiabatic wall, while each is in contact with a third system C via a conducting wall. (b) The adiabatic wall between A and B is replaced by a conducting wall, while C is insulated from A and B by an adiabatic wall.*

12.4 HEAT, INTERNAL ENERGY AND WORK

The Zeroth Law of Thermodynamics led us to the concept of temperature that agrees with our commonsense notion. Temperature is a marker

of the 'hotness' of a body. It determines the direction of flow of heat when two bodies are placed in thermal contact. Heat flows from the body at a higher temperature to the one at lower temperature. The flow stops when the temperatures equalise; the two bodies are then in thermal equilibrium. We saw in some detail how to construct temperature scales to assign temperatures to different bodies. We now describe the concepts of heat and other relevant quantities like internal energy and work.

The concept of internal energy of a system is not difficult to understand. We know that every bulk system consists of a large number of molecules. Internal energy is simply the sum of the kinetic energies and potential energies of these molecules. We remarked earlier that in thermodynamics, the kinetic energy of the system, as a whole, is not relevant. Internal energy is thus, the sum of molecular kinetic and potential energies in the frame of reference relative to which the centre of mass of the system is at rest. Thus, it includes only the (disordered) energy associated with the random motion of molecules of the system. We denote the internal energy of a system by U.

Though we have invoked the molecular picture to understand the meaning of internal energy, as far as thermodynamics is concerned, U is simply a macroscopic variable of the system. The important thing about internal energy is that it depends only on the state of the system, not on how that state was achieved. Internal energy U of a system is an example of a thermodynamic 'state variable' – its value depends only on the given state of the system, not on history i.e. not on the 'path' taken to arrive at that state. Thus, the internal energy of a given mass of gas depends on its state described by specific values of pressure, volume and temperature. It does not depend on how this state of the gas came about. Pressure, volume, temperature, and internal energy are thermodynamic state variables of the system (gas) (see section 12.7). If we neglect the small intermolecular forces in a gas, the internal energy of a gas is just the sum of kinetic energies associated with various random motions of its molecules. We will see in the next chapter that in a gas this motion is not only translational (i.e. motion from one point to another in the volume of the container); it also includes rotational and vibrational motion of the molecules (Fig. 12.3).

What are the ways of changing internal energy of a system ? Consider again, for simplicity, the system to be a certain mass of gas contained in a cylinder with a movable piston as shown in Fig. 12.4. Experience shows there are two ways of changing the state of the gas (and hence its internal energy). One way is to put the cylinder in contact with a body at a higher temperature than that of the gas. The temperature difference will cause a flow of energy (heat) from the hotter body to the gas, thus increasing the internal energy of the gas. The other way is to push the piston down i.e. to do work on the system, which again results in increasing the internal energy of the gas. Of course, both these things could happen in the reverse direction. With surroundings at a lower temperature, heat would flow from the gas to the surroundings. Likewise, the gas could push the piston up and do work on the surroundings. In short, heat and work are two different modes of altering the state of a thermodynamic system and changing its internal energy.

The notion of heat should be carefully distinguished from the notion of internal energy. Heat is certainly energy, but it is the energy in transit. This is not just a play of words. The distinction is of basic significance. The state of a thermodynamic system is characterised by its internal energy, not heat. A statement like **'a gas in a given state has a certain amount of heat'** is as meaningless as the statement that **'a gas in a given state has a certain amount of work'**. In contrast, **'a gas in a given state has a certain amount of internal energy'** is a perfectly meaningful statement. Similarly, the statements **'a certain amount of heat is supplied to the system' or 'a certain amount of work was done by the system'** are perfectly meaningful.

To summarise, heat and work in thermodynamics are not state variables. They are modes of energy transfer to a system resulting in change in its internal energy, which, as already mentioned, is a state variable.

In ordinary language, we often confuse heat with internal energy. The distinction between them is sometimes ignored in elementary

physics books. For proper understanding of thermodynamics, however, the distinction is crucial.

Fig. 12.3 *(a) Internal energy U of a gas is the sum of the kinetic and potential energies of its molecules when the box is at rest. Kinetic energy due to various types of motion (translational, rotational, vibrational) is to be included in U. (b) If the same box is moving as a whole with some velocity, the kinetic energy of the box is not to be included in U.*

Fig. 12.4 *Heat and work are two distinct modes of energy transfer to a system that results in change in its internal energy. (a) Heat is energy transfer due to temperature difference between the system and the surroundings. (b) Work is energy transfer brought about by means (e.g. moving the piston by raising or lowering some weight connected to it) that do not involve such a temperature difference.*

12.5 FIRST LAW OF THERMODYNAMICS

We have seen that the internal energy U of a system can change through two modes of energy transfer : heat and work. Let

ΔQ = Heat supplied *to* the system *by* the surroundings

ΔW = Work done *by* the system *on* the surroundings

ΔU = Change in internal energy of the system

The general principle of conservation of energy then implies that

$$\Delta Q = \Delta U + \Delta W \qquad (12.1)$$

i.e. the energy (ΔQ) supplied to the system goes in partly to increase the internal energy of the system (ΔU) and the rest in work on the environment (ΔW). Equation (12.1) is known as the **First Law of Thermodynamics**. It is simply the general law of conservation of energy applied to any system in which the energy transfer from or to the surroundings is taken into account.

Let us put Eq. (12.1) in the alternative form

$$\Delta Q - \Delta W = \Delta U \qquad (12.2)$$

Now, the system may go from an initial state to the final state in a number of ways. For example, to change the state of a gas from (P_1, V_1) to (P_2, V_2), we can first change the volume of the gas from V_1 to V_2, keeping its pressure constant i.e. we can first go the state (P_1, V_2) and then change the pressure of the gas from P_1 to P_2, keeping volume constant, to take the gas to (P_2, V_2). Alternatively, we can first keep the volume constant and then keep the pressure constant. Since U is a state variable, ΔU depends only on the initial and final states and not on the path taken by the gas to go from one to the other. However, ΔQ and ΔW will, in general, depend on the path taken to go from the initial to final states. From the First Law of Thermodynamics, Eq. (12.2), it is clear that the combination $\Delta Q - \Delta W$, is however, path independent. This shows that if a system is taken through a process in which $\Delta U = 0$ (for example, isothermal expansion of an ideal gas, see section 12.8),

$$\Delta Q = \Delta W$$

i.e., heat supplied to the system is used up entirely by the system in doing work on the environment.

If the system is a gas in a cylinder with a movable piston, the gas in moving the piston

does work. Since force is pressure times area, and area times displacement is volume, work done by the system against a constant pressure P is

$$\Delta W = P\,\Delta V$$

where ΔV is the change in volume of the gas. Thus, for this case, Eq. (12.1) gives

$$\Delta Q = \Delta U + P\,\Delta V \qquad (12.3)$$

As an application of Eq. (12.3), consider the change in internal energy for 1 g of water when we go from its liquid to vapour phase. The measured latent heat of water is 2256 J/g. i.e., for 1 g of water ΔQ = 2256 J. At atmospheric pressure, 1 g of water has a volume 1 cm^3 in liquid phase and 1671 cm^3 in vapour phase.

Therefore,

$$\Delta W = P(V_g - V_l) = 1.013 \times 10^5 \times (1670) \times 10^{-6} = 169.2\ \text{J}$$

Equation (12.3) then gives

$$\Delta U = 2256 - 169.2 = 2086.8\ \text{J}$$

We see that most of the heat goes to increase the internal energy of water in transition from the liquid to the vapour phase.

12.6 SPECIFIC HEAT CAPACITY

Suppose an amount of heat ΔQ supplied to a substance changes its temperature from T to $T + \Delta T$. We define heat capacity of a substance (see Chapter 11) to be

$$S = \frac{\Delta Q}{\Delta T} \qquad (12.4)$$

We expect ΔQ and, therefore, heat capacity S to be proportional to the mass of the substance. Further, it could also depend on the temperature, i.e., a different amount of heat may be needed for a unit rise in temperature at different temperatures. To define a constant characteristic of the substance and independent of its amount, we divide S by the mass of the substance m in kg :

$$s = \frac{S}{m} = \frac{1}{m}\frac{Q}{T} \qquad (12.5)$$

s is known as the specific heat capacity of the substance. It depends on the nature of the substance and its temperature. The unit of specific heat capacity is J kg^{-1} K^{-1}.

If the amount of substance is specified in terms of moles μ (instead of mass m in kg), we can define heat capacity per mole of the substance by

$$C = \frac{S}{} = \frac{1}{}\frac{Q}{T} \qquad (12.6)$$

C is known as molar specific heat capacity of the substance. Like s, C is independent of the amount of substance. C depends on the nature of the substance, its temperature and the conditions under which heat is supplied. The unit of C is J mol^{-1} K^{-1}. As we shall see later (in connection with specific heat capacity of gases), additional conditions may be needed to define C or s. The idea in defining C is that simple predictions can be made in regard to molar specific heat capacities.

Table 12.1 lists measured specific and molar heat capacities of solids at atmospheric pressure and ordinary room temperature.

We will see in Chapter 13 that predictions of specific heats of gases generally agree with experiment. We can use the same law of equipartition of energy that we use there to predict molar specific heat capacities of solids. Consider a solid of N atoms, each vibrating about its mean position. An oscillator in one dimension has average energy of $2 \times \frac{1}{2}\, k_B T = k_B T$. In three dimensions, the average energy is $3\, k_B T$. For a mole of a solid, the total energy is

$$U = 3\, k_B T \times N_A = 3\, RT$$

Now, at constant pressure, $\Delta Q = \Delta U + P\,\Delta V \cong \Delta U$, since for a solid ΔV is negligible. Therefore,

$$C = \frac{\Delta Q}{\Delta T} = \frac{\Delta U}{\Delta T} = 3R \qquad (12.7)$$

Table 12.1 Specific and molar heat capacities of some solids at room temperature and atmospheric pressure

Substance	Specific-^{v} heat (J kg^{-1}K^{-1})	Molar specific Heat(J mol^{-1} K^{-1})
Aluminium	900.0	24.4
Carbon	506.5	6.1
Copper	386.4	24.5
Lead	127.7	26.5
Silver	236.1	25.5
Tungsten	134.4	24.9

As Table 12.1 shows, the experimentally measured values which generally agrees with

predicted value 3R at ordinary temperatures. (Carbon is an exception.) The agreement is known to break down at low temperatures.

Specific heat capacity of water

The old unit of heat was calorie. One calorie was earlier defined to be the amount of heat required to raise the temperature of 1g of water by 1°C. With more precise measurements, it was found that the specific heat of water varies slightly with temperature. Figure 12.5 shows this variation in the temperature range 0 to 100 °C.

Fig. 12.5 *Variation of specific heat capacity of water with temperature.*

For a precise definition of calorie, it was, therefore, necessary to specify the unit temperature interval. One calorie is defined to be the amount of heat required to raise the temperature of 1g of water from 14.5 °C to 15.5 °C. Since heat is just a form of energy, it is preferable to use the unit joule, J. In SI units, the specific heat capacity of water is 4186 J kg^{-1} K^{-1} i.e. 4.186 J g^{-1} K^{-1}. The so called mechanical equivalent of heat defined as the amount of work needed to produce 1 cal of heat is in fact just a conversion factor between two different units of energy : calorie to joule. Since in SI units, we use the unit joule for heat, work or any other form of energy, the term mechanical equivalent is now superfluous and need not be used.

As already remarked, the specific heat capacity depends on the process or the conditions under which heat capacity transfer takes place. For gases, for example, we can define two specific heats : **specific heat capacity at constant volume and specific heat capacity at constant pressure**. For an ideal gas, we have a simple relation.

$$C_p - C_v = R \qquad (12.8)$$

where C_p and C_v are molar specific heat capacities of an ideal gas at constant pressure and volume respectively and R is the universal gas constant. To prove the relation, we begin with Eq. (12.3) for 1 mole of the gas :

$$\Delta Q = \Delta U + P\,\Delta V$$

If ΔQ is absorbed at constant volume, $\Delta V = 0$

$$C_v \quad \frac{Q}{T}_v \quad \frac{U}{T}_v \quad \frac{U}{T} \qquad (12.9)$$

where the subscript v is dropped in the last step, since U of an ideal gas depends only on temperature. (The subscript denotes the quantity kept fixed.) If, on the other hand, ΔQ is absorbed at constant pressure,

$$C_p \quad \frac{Q}{T}_p \quad \frac{U}{T}_p \quad P\,\frac{V}{T}_p \qquad (12.10)$$

The subscript p can be dropped from the first term since U of an ideal gas depends only on T. Now, for a mole of an ideal gas

$$PV = RT$$

which gives

$$P\,\frac{V}{T}_p \quad R \qquad (12.11)$$

Equations (12.9) to (12.11) give the desired relation, Eq. (12.8).

12.7 THERMODYNAMIC STATE VARIABLES AND EQUATION OF STATE

Every **equilibrium state** of a thermodynamic system is completely described by specific values of some macroscopic variables, also called state variables. For example, an equilibrium state of a gas is completely specified by the values of pressure, volume, temperature, and mass (and composition if there is a mixture of gases). A thermodynamic system is not always in equilibrium. For example, a gas allowed to expand freely against vacuum is not an equilibrium state [Fig. 12.6(a)]. During the rapid expansion, pressure of the gas may

not be uniform throughout. Similarly, a mixture of gases undergoing an explosive chemical reaction (e.g. a mixture of petrol vapour and air when ignited by a spark) is not an equilibrium state; again its temperature and pressure are not uniform [Fig. 12.6(b)]. Eventually, the gas attains a uniform temperature and pressure and comes to thermal and mechanical equilibrium with its surroundings.

Fig. 12.6 (a) The partition in the box is suddenly removed leading to free expansion of the gas. (b) A mixture of gases undergoing an explosive chemical reaction. In both situations, the gas is not in equilibrium and cannot be described by state variables.

In short, thermodynamic state variables describe equilibrium states of systems. The various state variables are not necessarily independent. The connection between the state variables is called the equation of state. For example, for an ideal gas, the equation of state is the ideal gas relation

$$P\,V = \mu\,R\,T$$

For a fixed amount of the gas i.e. given μ, there are thus, only two independent variables, say P and V or T and V. The pressure-volume curve for a fixed temperature is called an **isotherm**. Real gases may have more complicated equations of state.

The thermodynamic state variables are of two kinds: **extensive** and **intensive**. Extensive variables indicate the 'size' of the system. Intensive variables such as pressure and temperature do not. To decide which variable is extensive and which intensive, think of a relevant system in equilibrium, and imagine that it is divided into two equal parts. The variables that remain unchanged for each part are intensive. The variables whose values get halved in each part are extensive. It is easily seen, for example, that internal energy U, volume V, total mass M are extensive variables. Pressure P, temperature T, and density ρ are intensive variables. It is a good practice to check the consistency of thermodynamic equations using this classification of variables. For example, in the equation

$$\Delta Q = \Delta U + P\,\Delta V$$

quantities on both sides are extensive*. (The product of an intensive variable like P and an extensive quantity ΔV is extensive.)

12.8 THERMODYNAMIC PROCESSES

12.8.1 Quasi-static process

Consider a gas in thermal and mechanical equilibrium with its surroundings. The pressure of the gas in that case equals the external pressure and its temperature is the same as that of its surroundings. Suppose that the external pressure is suddenly reduced (say by lifting the weight on the movable piston in the container). The piston will accelerate outward. During the process, the gas passes through states that are not equilibrium states. The non-equilibrium states do not have well-defined pressure and temperature. In the same way, if a finite temperature difference exists between the gas and its surroundings, there will be a rapid exchange of heat during which the gas will pass through non-equilibrium states. In due course, the gas will settle to an equilibrium state with well-defined temperature and pressure equal to those of the surroundings. The free expansion of a gas in vacuum and a mixture of gases undergoing an explosive chemical reaction, mentioned in section 12.7 are also examples where the system goes through non-equilibrium states.

Non-equilibrium states of a system are difficult to deal with. It is, therefore, convenient to imagine an idealised process in which at every stage the system is an equilibrium state. Such a

* *As emphasised earlier, Q is not a state variable. However, ΔQ is clearly proportional to the total mass of system and hence is extensive.*

process is, in principle, infinitely slow-hence the name quasi-static (meaning nearly static). The system changes its variables (P, T, V) so slowly that it remains in thermal and mechanical equilibrium with its surroundings throughout. In a quasi-static process, at every stage, the difference in the pressure of the system and the external pressure is infinitesimally small. The same is true of the temperature difference between the system and its surroundings. To take a gas from the state (P, T) to another state (P', T') via a quasi-static process, we change the external pressure by a very small amount, allow the system to equalise its pressure with that of the surroundings and continue the process infinitely slowly until the system achieves the pressure P'. Similarly, to change the temperature, we introduce an infinitesimal temperature difference between the system and the surrounding reservoirs and by choosing reservoirs of progressively different temperatures T to T', the system achieves the temperature T'.

Fig. 12.7 *In a quasi-static process, the temperature of the surrounding reservoir and the external pressure differ only infinitesimally from the temperature and pressure of the system.*

A quasi-static process is obviously a hypothetical construct. In practice, processes that are sufficiently slow and do not involve accelerated motion of the piston, large temperature gradient, etc. are reasonably approximation to an ideal quasi-static process. We shall from now on deal with quasi-static processes only, except when stated otherwise.

A process in which the temperature of the system is kept fixed throughout is called an **isothermal process**. The expansion of a gas in a metallic cylinder placed in a large reservoir of fixed temperature is an example of an isothermal process. (Heat transferred from the reservoir to the system does not materially affect the temperature of the reservoir, because of its very large heat capacity.) In **isobaric processes** the pressure is constant while in **isochoric processes** the volume is constant. Finally, if the system is insulated from the surroundings and no heat flows between the system and the surroundings, the process is **adiabatic**. The definitions of these special processes are summarised in Table. 12.2

Table 12.2 **Some special thermodynamic processes**

Type of processes	Feature
Isothermal	Temperature constant
Isobaric	Pressure constant
Isochoric	Volume constant
Adiabatic	No heat flow between the system and the surroundings ($\Delta Q = 0$)

We now consider these processes in some detail :

Isothermal process

For an isothermal process (T fixed), the ideal gas equation gives

$$PV = \text{constant}$$

i.e., pressure of a given mass of gas varies inversely as its volume. This is nothing but Boyle's Law.

Suppose an ideal gas goes isothermally (at temperature T) from its initial state (P_1, V_1) to the final state (P_2, V_2). At any intermediate stage with pressure P and volume change from V to $V + \Delta V$ (ΔV small)

$$\Delta W = P \Delta V$$

Taking ($\Delta V \to 0$) and summing the quantity ΔW over the entire process,

$$W = \int_{V_1}^{V_2} P \, dV$$

$$= RT \int_{V_1}^{V_2} \frac{dV}{V} \qquad RT \ \text{In} \frac{V_2}{V_1} \qquad (12.12)$$

where in the second step we have made use of the ideal gas equation $PV = \mu RT$ and taken the constants out of the integral. For an ideal gas, internal energy depends only on temperature. Thus, there is no change in the internal energy of an ideal gas in an isothermal process. The First Law of Thermodynamics then implies that heat supplied to the gas equals the work done by the gas : $Q = W$. Note from Eq. (12.12) that for $V_2 > V_1$, $W > 0$; and for $V_2 < V_1$, $W < 0$. That is, in an isothermal expansion, the gas absorbs heat and does work while in an isothermal compression, work is done on the gas by the environment and heat is released.

Adiabatic process

In an adiabatic process, the system is insulated from the surroundings and heat absorbed or released is zero. From Eq. (12.1), we see that work done by the gas results in decrease in its internal energy (and hence its temperature for an ideal gas). We quote without proof (the result that you will learn in higher courses) that for an adiabatic process of an ideal gas.

$$P\,V^{\gamma} = \text{const} \qquad (12.13)$$

where γ is the ratio of specific heats (ordinary or molar) at constant pressure and at constant volume.

$$\gamma = \frac{C_p}{C_v}$$

Thus if an ideal gas undergoes a change in its state adiabatically from (P_1, V_1) to (P_2, V_2) :

$$P_1\,V_1^{\gamma} = P_2\,V_2^{\gamma} \qquad (12.14)$$

Figure 12.8 shows the *P-V* curves of an ideal gas for two adiabatic processes connecting two isotherms.

Fig. 12.8 *P-V curves for isothermal and adiabatic processes of an ideal gas.*

We can calculate, as before, the work done in an adiabatic change of an ideal gas from the state (P_1, V_1, T_1) to the state (P_2, V_2, T_2).

$$W = \int_{V_1}^{V_2} P\,dV$$

$$= constant \int_{V_1}^{V_2} \frac{dV}{V} \quad constant \left. \frac{V^{-+1}}{1} \right|_{V_1}^{V_2}$$

$$= \frac{constant}{(1-\gamma)} \times \left[\frac{1}{V_2^{\gamma-1}} - \frac{1}{V_1^{\gamma-1}} \right] \qquad (12.15)$$

From Eq. (12.34), the constant is $P_1V_1^{\gamma}$ or $P_2V_2^{\gamma}$

$$W = \frac{1}{1-} \quad \frac{P_2V_2}{V_2^{\ 1}} \quad \frac{P_1V_1}{V_1^{\ 1}}$$

$$= \frac{1}{1-} \quad P_2V_2 \quad P_1V_1 \quad \frac{R(T_1 - T_2)}{-} \qquad (12.16)$$

As expected, if work is done by the gas in an adiabatic process ($W > 0$), from Eq. (12.16), $T_2 < T_1$. On the other hand, if work is done on the gas ($W < 0$), we get $T_2 > T_1$ i.e., the temperature of the gas rises.

Isochoric process

In an isochoric process, V is constant. No work is done on or by the gas. From Eq. (12.1), the heat absorbed by the gas goes entirely to change its internal energy and its temperature. The change in temperature for a given amount of heat is determined by the specific heat of the gas at constant volume.

Isobaric process

In an isobaric process, P is fixed. Work done by the gas is

$$W = P\,(V_2 - V_1) = \mu\,R\,(T_2 - T_1) \qquad (12.17)$$

Since temperature changes, so does internal energy. The heat absorbed goes partly to increase internal energy and partly to do work. The change in temperature for a given amount of heat is determined by the specific heat of the gas at constant pressure.

Cyclic process

In a cyclic process, the system returns to its initial state. Since internal energy is a state variable, $\Delta U = 0$ for a cyclic process. From

Eq. (12.1), the total heat absorbed equals the work done by the system.

12.9 HEAT ENGINES

Heat engine is a device by which a system is made to undergo a cyclic process that results in conversion of heat to work.

(1) It consists of a **working substance**–the system. For example, a mixture of fuel vapour and air in a gasoline or diesel engine or steam in a steam engine are the working substances.

(2) The working substance goes through a cycle consisting of several processes. In some of these processes, it absorbs a total amount of heat Q_1 from an external reservoir at some high temperature T_1.

(3) In some other processes of the cycle, the working substance releases a total amount of heat Q_2 to an external reservoir at some lower temperature T_2.

(4) The work done (W) by the system in a cycle is transferred to the environment via some arrangement (e.g. the working substance may be in a cylinder with a moving piston that transfers mechanical energy to the wheels of a vehicle via a shaft).

The basic features of a heat engine are schematically represented in Fig. 12.9.

Fig. 12.9 Schematic representation of a heat engine. The engine takes heat Q_1 from a hot reservoir at temperature T_1, releases heat Q_2 to a cold reservoir at temperature T_2 and delivers work W to the surroundings.

The cycle is repeated again and again to get useful work for some purpose. The discipline of thermodynamics has its roots in the study of heat engines. A basic question relates to the efficiency of a heat engine. The efficiency (η) of a heat engine is defined by

$$\eta = \frac{W}{Q_1} \tag{12.18}$$

where Q_1 is the heat input i.e., the heat absorbed by the system in one complete cycle and W is the work done on the environment in a cycle. In a cycle, a certain amount of heat (Q_2) may also be rejected to the environment. Then, according to the First Law of Thermodynamics, over one complete cycle,

$$W = Q_1 - Q_2 \tag{12.19}$$

i.e.,

$$\eta = 1 - \frac{Q_2}{Q_1} \tag{12.20}$$

For $Q_2 = 0$, $\eta = 1$, i.e., the engine will have 100% efficiency in converting heat into work. Note that the First Law of Thermodynamics i.e., the energy conservation law does not rule out such an engine. But experience shows that such an ideal engine with $\eta = 1$ is never possible, even if we can eliminate various kinds of losses associated with actual heat engines. It turns out that there is a fundamental limit on the efficiency of a heat engine set by an independent principle of nature, called the Second Law of Thermodynamics (section 12.11).

The mechanism of conversion of heat into work varies for different heat engines. Basically, there are two ways : the system (say a gas or a mixture of gases) is heated by an external furnace, as in a steam engine; or it is heated internally by an exothermic chemical reaction as in an internal combustion engine. The various steps involved in a cycle also differ from one engine to another. For the purpose of general analysis, it is useful to conceptualise a heat engine as having the following essential ingredients.

12.10 REFRIGERATORS AND HEAT PUMPS

A refrigerator is the reverse of a heat engine. Here the working substance extracts heat Q_2 from the cold reservoir at temperature T_2, some external work W is done on it and heat Q_1 is released to the hot reservoir at temperature T_1 (Fig. 12.10).

Fig. 12.10 Schematic representation of a refrigerator or a heat pump, the reverse of a heat engine.

Pioneers of Thermodynamics

Lord Kelvin (William Thomson) (1824-1907), born in Belfast, Ireland, is among the foremost British scientists of the nineteenth century. Thomson played a key role in the development of the law of conservation of energy suggested by the work of James Joule (1818-1889), Julius Mayer (1814-1878) and Hermann Helmholtz (1821-1894). He collaborated with Joule on the so-called Joule-Thomson effect : cooling of a gas when it expands into vacuum. He introduced the notion of the absolute zero of temperature and proposed the absolute temperature scale, now called the Kelvin scale in his honour. From the work of Sadi Carnot (1796-1832), Thomson arrived at a form of the Second Law of Thermodynamics. Thomson was a versatile physicist, with notable contributions to electromagnetic theory and hydrodynamics.

Rudolf Clausius (1822-1888), born in Poland, is generally regarded as the discoverer of the Second Law of Thermodynamics. Based on the work of Carnot and Thomson, Clausius arrived at the important notion of entropy that led him to a fundamental version of the Second Law of Thermodynamics that states that the entropy of an isolated system can never decrease. Clausius also worked on the kinetic theory of gases and obtained the first reliable estimates of molecular size, speed, mean free path, etc.

A heat pump is the same as a refrigerator. What term we use depends on the purpose of the device. If the purpose is to cool a portion of space, like the inside of a chamber, and higher temperature reservoir is surrounding, we call the device a refrigerator; if the idea is to pump heat into a portion of space (the room in a building when the outside environment is cold), the device is called a heat pump.

In a refrigerator the working substance (usually, in gaseous form) goes through the following steps : (a) sudden expansion of the gas from high to low pressure which cools it and converts it into a vapour-liquid mixture, (b) absorption by the cold fluid of heat from the region to be cooled converting it into vapour, (c) heating up of the vapour due to external work done on the system, and (d) release of heat by the vapour to the surroundings, bringing it to the initial state and completing the cycle.

The coefficient of performance (α) of a refrigerator is given by

$$\frac{Q_2}{W} \qquad (12.21)$$

where Q_2 is the heat extracted from the cold reservoir and W is the work done on the system–the refrigerant. (α for heat pump is defined as Q_1/W) Note that while η by definition can never exceed 1, α can be greater than 1. By energy conservation, the heat released to the hot reservoir is

$$Q_1 = W + Q_2$$

i.e.,
$$\frac{Q_2}{Q_1 - Q_2} \qquad (12.22)$$

In a heat engine, heat cannot be fully converted to work; likewise a refrigerator cannot work without some external work done on the system, i.e., the coefficient of performance in Eq. (12.21) cannot be infinite.

12.11 SECOND LAW OF THERMODYNAMICS

The First Law of Thermodynamics is the principle of conservation of energy. Common experience shows that there are many conceivable processes that are perfectly allowed by the First Law and yet are never observed. For example, nobody has ever seen a book lying on a table jumping to a height by itself. But such a thing

would be possible if the principle of conservation of energy were the only restriction. The table could cool spontaneously, converting some of its internal energy into an equal amount of mechanical energy of the book, which would then hop to a height with potential energy equal to the mechanical energy it acquired. But this never happens. Clearly, some additional basic principle of nature forbids the above, even though it satisfies the energy conservation principle. This principle, which disallows many phenomena consistent with the First Law of Thermodynamics is known as the Second Law of Thermodynamics.

The Second Law of Thermodynamics gives a fundamental limitation to the efficiency of a heat engine and the co-efficient of performance of a refrigerator. In simple terms, it says that efficiency of a heat engine can never be unity. According to Eq. (12.20), this implies that heat released to the cold reservoir can never be made zero. For a refrigerator, the Second Law says that the co-efficient of performance can never be infinite. According to Eq. (12.21), this implies that external work (W) can never be zero. The following two statements, one due to Kelvin and Planck denying the possibility of a perfect heat engine, and another due to Clausius denying the possibility of a perfect refrigerator or heat pump, are a concise summary of these observations.

Second Law of Thermodynamics

Kelvin-Planck statement

No process is possible whose sole result is the absorption of heat from a reservoir and the complete conversion of the heat into work.

Clausius statement

No process is possible whose sole result is the transfer of heat from a colder object to a hotter object.

It can be proved that the two statements above are completely equivalent.

12.12 REVERSIBLE AND IRREVERSIBLE PROCESSES

Imagine some process in which a thermodynamic system goes from an initial state i to a final state f. During the process the system absorbs heat Q from the surroundings and performs work W on it. Can we reverse this process and bring both the system and surroundings to their initial states with no other effect anywhere? Experience suggests that for most processes in nature this is not possible. The spontaneous processes of nature are irreversible. Several examples can be cited. The base of a vessel on an oven is hotter than its other parts. When the vessel is removed, heat is transferred from the base to the other parts, bringing the vessel to a uniform temperature (which in due course cools to the temperature of the surroundings). The process cannot be reversed; a part of the vessel will not get cooler spontaneously and warm up the base. It will violate the Second Law of Thermodynamics, if it did. The free expansion of a gas is irreversible. The combustion reaction of a mixture of petrol and air ignited by a spark cannot be reversed. Cooking gas leaking from a gas cylinder in the kitchen diffuses to the entire room. The diffusion process will not spontaneously reverse and bring the gas back to the cylinder. The stirring of a liquid in thermal contact with a reservoir will convert the work done into heat, increasing the internal energy of the reservoir. The process cannot be reversed exactly; otherwise it would amount to conversion of heat entirely into work, violating the Second Law of Thermodynamics. Irreversibility is a rule rather an exception in nature.

Irreversibility arises mainly from two causes: one, many processes (like a free expansion, or an explosive chemical reaction) take the system to non-equilibrium states; two, most processes involve friction, viscosity and other dissipative effects (e.g., a moving body coming to a stop and losing its mechanical energy as heat to the floor and the body; a rotating blade in a liquid coming to a stop due to viscosity and losing its mechanical energy with corresponding gain in the internal energy of the liquid). Since dissipative effects are present everywhere and can be minimised but not fully eliminated, most processes that we deal with are irreversible.

A thermodynamic process (state $i \rightarrow$ state f) is reversible if the process can be turned back such that both the system and the surroundings return to their original states, with no other change anywhere else in the universe. From the preceding discussion, a reversible process is an idealised notion. A process is reversible only if it is quasi-static (system in equilibrium with the

surroundings at every stage) and there are no dissipative effects. For example, a quasi-static isothermal expansion of an ideal gas in a cylinder fitted with a frictionless movable piston is a reversible process.

Why is reversibility such a basic concept in thermodynamics ? As we have seen, one of the concerns of thermodynamics is the efficiency with which heat can be converted into work. The Second Law of Thermodynamics rules out the possibility of a perfect heat engine with 100% efficiency. But what is the highest efficiency possible for a heat engine working between two reservoirs at temperatures T_1 and T_2? It turns out that a heat engine based on idealised reversible processes achieves the highest efficiency possible. All other engines involving irreversibility in any way (as would be the case for practical engines) have lower than this limiting efficiency.

12.13 CARNOT ENGINE

Suppose we have a hot reservoir at temperature T_1 and a cold reservoir at temperature T_2. What is the maximum efficiency possible for a heat engine operating between the two reservoirs and what cycle of processes should be adopted to achieve the maximum efficiency ? Sadi Carnot, a French engineer, first considered this question in 1824. Interestingly, Carnot arrived at the correct answer, even though the basic concepts of heat and thermodynamics had yet to be firmly established.

We expect the ideal engine operating between two temperatures to be a reversible engine. Irreversibility is associated with dissipative effects, as remarked in the preceding section, and lowers efficiency. A process is reversible if it is quasi-static and non-dissipative. We have seen that a process is not quasi-static if it involves finite temperature difference between the system and the reservoir. This implies that in a reversible heat engine operating between two temperatures, heat should be absorbed (from the hot reservoir) isothermally and released (to the cold reservoir) isothermally. We thus have identified two steps of the reversible heat engine : isothermal process at temperature T_1 absorbing heat Q_1 from the hot reservoir, and another isothermal process at temperature T_2 releasing heat Q_2 to the cold reservoir. To complete a cycle, we need to take the system from temperature T_1 to T_2 and then back from temperature T_2 to T_1. Which processes should we employ for this purpose that are reversible? A little reflection shows that we can only adopt reversible adiabatic processes for these purposes, which involve no heat flow from any reservoir. If we employ any other process that is not adiabatic, say an isochoric process, to take the system from one temperature to another, we shall need a series of reservoirs in the temperature range T_2 to T_1 to ensure that at each stage the process is quasi-static. (Remember again that for a process to be quasi-static and reversible, there should be no finite temperature difference between the system and the reservoir.) But we are considering a reversible engine that operates between only two temperatures. Thus adiabatic processes must bring about the temperature change in the system from T_1 to T_2 and T_2 to T_1 in this engine.

Fig. 12.11 Carnot cycle for a heat engine with an ideal gas as the working substance.

A reversible heat engine operating between two temperatures is called a Carnot engine. We have just argued that such an engine must have the following sequence of steps constituting one cycle, called the Carnot cycle, shown in Fig. 12.11. We have taken the working substance of the Carnot engine to be an ideal gas.

(a) *Step* 1 $\rightarrow$2 Isothermal expansion of the gas taking its state from (P_1, V_1, T_1) to (P_2, V_2, T_1).

The heat absorbed by the gas (Q_1) from the reservoir at temperature T_1 is given by

Eq. (12.12). This is also the work done ($W_{1\to 2}$) by the gas on the environment.

$$W_{1\to 2} = Q_1 = \mu R T_1 \ln\left(\frac{V_2}{V_1}\right) \qquad (12.23)$$

(b) *Step* 2 $\to$3 Adiabatic expansion of the gas from (P_2, V_2, T_1) to (P_3, V_3, T_2) Work done by the gas, using Eq. (12.16), is

$$W_{2\ \ 3} \quad \frac{R\ T_1 \quad T_2}{1} \qquad (12.24)$$

(c) *Step* 3 $\to$4 Isothermal compression of the gas from (P_3, V_3, T_2) to (P_4, V_4, T_2).

Heat released (Q_2) by the gas to the reservoir at temperature T_2 is given by Eq. (12.12). This is also the work done ($W_{3\to 4}$) on the gas by the environment.

$$W_{3\ \ 4} \quad Q_2 \quad RT_2 \ln \frac{V_3}{V_4} \qquad (12.25)$$

(d) *Step* 4 $\to$1 Adiabatic compression of the gas from (P_4, V_4, T_2) to (P_1,V_1, T_1).

Work done on the gas, [using Eq.(12.16)], is

$$W_{4\ \ 1} \quad R\ \frac{T_1 \quad T_2}{-1} \qquad (12.26)$$

From Eqs. (12.23) to (12.26) total work done by the gas in one complete cycle is

$$W = W_{1\to 2} + W_{2\to 3} - W_{3\ \to 4} - W_{4\to 1}$$

$$= \mu RT_1 \ln\left(\frac{V_2}{V_1}\right) - \mu RT_2 \ln\left(\frac{V_3}{V_4}\right) \qquad (12.27)$$

The efficiency $\eta$of the Carnot engine is

$$\frac{W}{Q_1} \quad 1 \quad \frac{Q_2}{Q_1}$$

$$1 \quad \frac{T_2}{T_1} \quad \frac{\text{In}\ \frac{V_3}{V_4}}{\text{In}\ \frac{V_2}{V_1}} \qquad (12.28)$$

Now since step 2 $\to$3 is an adiabatic process,

$$T_1 V_2^{\ \ 1} \quad T_2 V_3^{\ \ 1}$$

$$\text{i.e.} \quad \frac{V_2}{V_3} = \frac{T_2}{T_1}^{\ 1/(\ \ 1)} \qquad (12.29)$$

Similarly, since step 4 $\to$1 is an adiabatic process

$$T_2 V_4^{\ \ 1} \quad T_1 V_1^{\ \ 1}$$

$$\text{i.e.} \quad \frac{V_1}{V_4} = \frac{T_2}{T_1}^{\ 1/\ \ 1} \qquad (12.30)$$

From Eqs. (12.29) and (12.30),

$$\frac{V_3}{V_4} = \frac{V_2}{V_1} \qquad (12.31)$$

Using Eq. (12.31) in Eq. (12.28), we get

$$1 - \frac{T_2}{T_1} \quad \text{(Carnot engine)} \qquad (12.32)$$

We have already seen that a Carnot engine is a reversible engine. Indeed it is the only reversible engine possible that works between two reservoirs at different temperatures. Each step of the Carnot cycle given in Fig. 12.11 can be reversed. This will amount to taking heat Q_2 from the cold reservoir at T_2, doing work W on the system, and transferring heat Q_1 to the hot reservoir. This will be a reversible refrigerator.

We next establish the important result (sometimes called Carnot's theorem) that (a) working between two given temperatures T_1 and T_2 of the hot and cold reservoirs respectively, no engine can have efficiency more than that of the Carnot engine and (b) the efficiency of the Carnot engine is independent of the nature of the working substance.

To prove the result (a), imagine a reversible (Carnot) engine R and an irreversible engine I working between the same source (hot reservoir) and sink (cold reservoir). Let us couple the engines, I and R, in such a way so that I acts like a heat engine and R acts as a refrigerator. Let I absorb heat Q_1 from the source, deliver work W' and release the heat $Q_1 - W'$ to the sink. We arrange so that R returns the same heat Q_1 to the source, taking heat Q_2 from the sink and requiring work $W = Q_1 - Q_2$ to be done on it.

Now suppose $\eta_R < \eta_I$ i.e. if R were to act as an engine it would give less work output than that of I i.e. $W < W'$ for a given Q_1. With R acting like a refrigerator, this would mean $Q_2 = Q_1 - W > Q_1 - W'$. Thus on the whole, the coupled I-R system extracts heat $(Q_1 - W) - (Q_1 - W') = (W' - W)$ from the cold reservoir and delivers the same amount of work in one cycle, without any change in the source or anywhere else. This is clearly against the Kelvin-Planck statement of the Second Law of Thermodynamics. Hence the assertion $\eta_I > \eta_R$ is wrong. No engine can have efficiency greater than that of the Carnot engine. A similar argument can be constructed to show that a reversible engine with one particular substance cannot be more efficient than the one using another substance. The maximum efficiency of a Carnot engine given by Eq. (12.32) is independent of the nature of the system performing the Carnot cycle of operations. Thus we are justified in using an ideal gas as a system in the calculation of efficiency η of a Carnot engine. The ideal gas has a simple equation of state, which allows us to readily calculate η, but the final result for η [Eq. (12.32)], is true for any Carnot engine.

Fig. 12.12 An irreversible engine (I) coupled to a reversible refrigerator (R). If $W' > W$, this would amount to extraction of heat $W' - W$ from the sink and its full conversion to work, in contradiction with the Second Law of Thermodynamics.

This final remark shows that in a Carnot cycle,

$$\frac{Q_1}{Q_2} = \frac{T_1}{T_2} \tag{12.33}$$

is a universal relation independent of the nature of the system. Here Q_1 and Q_2 are respectively, the heat absorbed and released isothermally (from the hot and to the cold reservoirs) in a Carnot engine. Equation (12.33), can, therefore, be used as a relation to define a truly universal thermodynamic temperature scale that is independent of any particular properties of the system used in the Carnot cycle. Of course, for an ideal gas as a working substance, this universal temperature is the same as the ideal gas temperature introduced in section 12.11.

SUMMARY

1. The zeroth law of thermodynamics states that '*two systems in thermal equilibrium with a third system are in thermal equilibrium with each other*'. The Zeroth Law leads to the concept of temperature.
2. Internal energy of a system is the sum of kinetic energies and potential energies of the molecular constituents of the system. It does not include the over-all kinetic energy of the system. Heat and work are two modes of energy transfer to the system. Heat is the energy transfer arising due to temperature difference between the system and the surroundings. Work is energy transfer brought about by other means, such as moving the piston of a cylinder containing the gas, by raising or lowering some weight connected to it.
3. The first law of thermodynamics is the general law of conservation of energy applied to any system in which energy transfer from or to the surroundings (through heat and work) is taken into account. It states that

$$\Delta Q = \Delta U + \Delta W$$

where ΔQ is the heat supplied to the system, ΔW is the work done by the system and ΔU is the change in internal energy of the system.

4. The specific heat capacity of a substance is defined by

$$s = \frac{1}{m}\frac{\Delta Q}{\Delta T}$$

where m is the mass of the substance and ΔQ is the heat required to change its temperature by ΔT. The molar specific heat capacity of a substance is defined by

$$C \quad \frac{1}{\quad}\frac{Q}{T}$$

where μ is the number of moles of the substance. For a solid, the law of equipartition of energy gives

$$C = 3R$$

which generally agrees with experiment at ordinary temperatures.

Calorie is the old unit of heat. 1 calorie is the amount of heat required to raise the temperature of 1 g of water from 14.5 °C to 15.5 °C. 1 cal = 4.186 J.

5. For an ideal gas, the molar specific heat capacities at constant pressure and volume satisfy the relation

$$C_p - C_v = R$$

where R is the universal gas constant.

6. Equilibrium states of a thermodynamic system are described by state variables. The value of a state variable depends only on the particular state, not on the path used to arrive at that state. Examples of state variables are pressure (P), volume (V), temperature (T), and mass (m). Heat and work are not state variables. An Equation of State (like the ideal gas equation $PV = \mu RT$) is a relation connecting different state variables.

7. A quasi-static process is an infinitely slow process such that the system remains in thermal and mechanical equilibrium with the surroundings throughout. In a quasi-static process, the pressure and temperature of the environment can differ from those of the system only infinitesimally.

8. In an isothermal expansion of an ideal gas from volume V_1 to V_2 at temperature T the heat absorbed (Q) equals the work done (W) by the gas, each given by

$$Q = W = \mu RT \ \ln\left(\frac{V_2}{V_1}\right)$$

9. In an adiabatic process of an ideal gas

$$PV^{\gamma} = \text{constant}$$

where $\dfrac{C_p}{C_v}$

Work done by an ideal gas in an adiabatic change of state from (P_1, V_1, T_1) to (P_2, V_2, T_2) is

$$W \quad \frac{R \ T_1 \quad T_2}{-1}$$

10. Heat engine is a device in which a system undergoes a cyclic process resulting in conversion of heat into work. If Q_1 is the heat absorbed from the source, Q_2 is the heat released to the sink, and the work output in one cycle is W, the efficiency η of the engine is:

$$\frac{W}{Q_1} \quad 1 \quad \frac{Q_2}{Q_1}$$

11. In a refrigerator or a heat pump, the system extracts heat Q_2 from the cold reservoir and releases Q_1 amount of heat to the hot reservoir, with work W done on the system. The co-efficient of performance of a refrigerator is given by

$$\alpha = \frac{Q_2}{W} = \frac{Q_2}{Q_1 - Q_2}$$

12. The second law of thermodynamics disallows some processes consistent with the First Law of Thermodynamics. It states

    Kelvin-Planck statement

    No process is possible whose sole result is the absorption of heat from a reservoir and complete conversion of the heat into work.

    Clausius statement

    No process is possible whose sole result is the transfer of heat from a colder object to a hotter object.

    Put simply, the Second Law implies that no heat engine can have efficiency η equal to 1 or no refrigerator can have co-efficient of performance α equal to infinity.

13. A process is reversible if it can be reversed such that both the system and the surroundings return to their original states, with no other change anywhere else in the universe. Spontaneous processes of nature are irreversible. The idealised reversible process is a quasi-static process with no dissipative factors such as friction, viscosity, etc.

14. Carnot engine is a reversible engine operating between two temperatures T_1 (source) and T_2 (sink). The Carnot cycle consists of two isothermal processes connected by two adiabatic processes. The efficiency of a Carnot engine is given by

$$\eta = 1 - \frac{T_2}{T_1} \qquad \text{(Carnot engine)}$$

    No engine operating between two temperatures can have efficiency greater than that of the Carnot engine.

15. If $Q > 0$, heat is added to the system

    If $Q < 0$, heat is removed to the system

    If $W > 0$, Work is done by the system

    If $W < 0$, Work is done on the system

Quantity	Symbol	Dimensions	Unit	Remark
Co-efficienty of volume expansion	α_v	$[K^{-1}]$	K^{-1}	$\alpha_v = 3\,\alpha_l$
Heat supplied to a system	ΔQ	$[ML^2 T^{-2}]$	J	Q is not a state variable
Specific heat	s	$[L^2 T^{-2} K^{-1}]$	$J\ kg^{-1}\ K^{-1}$	
Thermal Conductivity	K	$[MLT^{-3} K^{-1}]$	$J\ s^{-1}\ K^{-1}$	$H = -KA\dfrac{dt}{dx}$

POINTS TO PONDER

1. Temperature of a body is related to its average internal energy, not to the kinetic energy of motion of its centre of mass. A bullet fired from a gun is not at a higher temperature because of its high speed.
2. Equilibrium in thermodynamics refers to the situation when macroscopic variables describing the thermodynamic state of a system do not depend on time. Equilibrium of a system in mechanics means the net external force and torque on the system are zero.
3. In a state of thermodynamic equilibrium, the microscopic constituents of a system are not in equilibrium (in the sense of mechanics).
4. Heat capacity, in general, depends on the process the system goes through when heat is supplied.
5. In isothermal quasi-static processes, heat is absorbed or given out by the system even though at every stage the gas has the same temperature as that of the surrounding reservoir. This is possible because of the infinitesimal difference in temperature between the system and the reservoir.

EXERCISES

12.1 A geyser heats water flowing at the rate of 3.0 litres per minute from 27 °C to 77 °C. If the geyser operates on a gas burner, what is the rate of consumption of the fuel if its heat of combustion is 4.0×10^4 J/g ?

12.2 What amount of heat must be supplied to 2.0×10^{-2} kg of nitrogen (at room temperature) to raise its temperature by 45 °C at constant pressure ? (Molecular mass of N_2 = 28; R = 8.3 J mol^{-1} K^{-1}.)

12.3 Explain why

(a) Two bodies at different temperatures T_1 and T_2 if brought in thermal contact do not necessarily settle to the mean temperature $(T_1 + T_2)/2$.

(b) The coolant in a chemical or a nuclear plant (i.e., the liquid used to prevent the different parts of a plant from getting too hot) should have high specific heat.

(c) Air pressure in a car tyre increases during driving.

(d) The climate of a harbour town is more temperate than that of a town in a desert at the same latitude.

12.4 A cylinder with a movable piston contains 3 moles of hydrogen at standard temperature and pressure. The walls of the cylinder are made of a heat insulator, and the piston is insulated by having a pile of sand on it. By what factor does the pressure of the gas increase if the gas is compressed to half its original volume ?

12.5 In changing the state of a gas adiabatically from an equilibrium state A to another equilibrium state B, an amount of work equal to 22.3 J is done on the system. If the gas is taken from state A to B via a process in which the net heat absorbed by the system is 9.35 cal, how much is the net work done by the system in the latter case ? (Take 1 cal = 4.19 J)

12.6 Two cylinders A and B of equal capacity are connected to each other via a stopcock. A contains a gas at standard temperature and pressure. B is completely evacuated. The entire system is thermally insulated. The stopcock is suddenly opened. Answer the following :

(a) What is the final pressure of the gas in A and B ?

(b) What is the change in internal energy of the gas ?

(c) What is the change in the temperature of the gas ?

(d) Do the intermediate states of the system (before settling to the final equilibrium state) lie on its P-V-T surface ?

12.7 A steam engine delivers 5.4×10^8J of work per minute and services 3.6×10^9J of heat per minute from its boiler. What is the efficiency of the engine? How much heat is wasted per minute?

12.8 An electric heater supplies heat to a system at a rate of 100W. If system performs work at a rate of 75 joules per second. At what rate is the internal energy increasing?

12.9 A thermodynamic system is taken from an original state to an intermediate state by the linear process shown in Fig. (12.13)

Fig. 12.13

Its volume is then reduced to the original value from E to F by an isobaric process. Calculate the total work done by the gas from D to E to F

12.10 A refrigerator is to maintain eatables kept inside at 9^0C. If room temperature is 36^0C, calculate the coefficient of performance.

CHAPTER THIRTEEN

KINETIC THEORY

13.1 Introduction
13.2 Molecular nature of matter
13.3 Behaviour of gases
13.4 Kinetic theory of an ideal gas
13.5 Law of equipartition of energy
13.6 Specific heat capacity
13.7 Mean free path
Summary
Points to ponder
Exercises
Additional exercises

13.1 INTRODUCTION

Boyle discovered the law named after him in 1661. Boyle, Newton and several others tried to explain the behaviour of gases by considering that gases are made up of tiny atomic particles. The actual atomic theory got established more than 150 years later. Kinetic theory explains the behaviour of gases based on the idea that the gas consists of rapidly moving atoms or molecules. This is possible as the inter-atomic forces, which are short range forces that are important for solids and liquids, can be neglected for gases. The kinetic theory was developed in the nineteenth century by Maxwell, Boltzmann and others. It has been remarkably successful. It gives a molecular interpretation of pressure and temperature of a gas, and is consistent with gas laws and Avogadro's hypothesis. It correctly explains specific heat capacities of many gases. It also relates measurable properties of gases such as viscosity, conduction and diffusion with molecular parameters, yielding estimates of molecular sizes and masses. This chapter gives an introduction to kinetic theory.

13.2 MOLECULAR NATURE OF MATTER

Richard Feynman, one of the great physicists of 20th century considers the discovery that "Matter is made up of atoms" to be a very significant one. Humanity may suffer annihilation (due to nuclear catastrophe) or extinction (due to environmental disasters) if we do not act wisely. If that happens, and all of scientific knowledge were to be destroyed then Feynman would like the 'Atomic Hypothesis' to be communicated to the next generation of creatures in the universe. Atomic Hypothesis: All things are made of atoms - little particles that move around in perpetual motion, attracting each other when they are a little distance apart, but repelling upon being squeezed into one another.

Speculation that matter may not be continuous, existed in many places and cultures. Kanada in India and Democritus

Atomic Hypothesis in Ancient India and Greece

Though John Dalton is credited with the introduction of atomic viewpoint in modern science, scholars in ancient India and Greece conjectured long before the existence of atoms and molecules. In the Vaiseshika school of thought in India founded by Kanada (Sixth century B.C.) the atomic picture was developed in considerable detail. Atoms were thought to be eternal, indivisible, infinitesimal and ultimate parts of matter. It was argued that if matter could be subdivided without an end, there would be no difference between a mustard seed and the Meru mountain. The four kinds of atoms (**Paramanu** — Sanskrit word for the smallest particle) postulated were Bhoomi (Earth), Ap (water), Tejas (fire) and Vayu (air) that have characteristic mass and other attributes, were propounded. Akasa (space) was thought to have no atomic structure and was continuous and inert. Atoms combine to form different molecules (e.g. two atoms combine to form a diatomic molecule dvyanuka, three atoms form a tryanuka or a triatomic molecule), their properties depending upon the nature and ratio of the constituent atoms. The size of the atoms was also estimated, by conjecture or by methods that are not known to us. The estimates vary. In Lalitavistara, a famous biography of the Buddha written mainly in the second century B.C., the estimate is close to the modern estimate of atomic size, of the order of 10^{-10} m.

In ancient Greece, Democritus (Fourth century B.C.) is best known for his atomic hypothesis. The word 'atom' means 'indivisible' in Greek. According to him, atoms differ from each other physically, in shape, size and other properties and this resulted in the different properties of the substances formed by their combination. The atoms of water were smooth and round and unable to 'hook' on to each other, which is why liquid /water flows easily. The atoms of earth were rough and jagged, so they held together to form hard substances. The atoms of fire were thorny which is why it caused painful burns. These fascinating ideas, despite their ingenuity, could not evolve much further, perhaps because they were intuitive conjectures and speculations not tested and modified by quantitative experiments - the hallmark of modern science.

in Greece had suggested that matter may consist of indivisible constituents. The scientific 'Atomic Theory' is usually credited to John Dalton. He proposed the atomic theory to explain the laws of definite and multiple proportions obeyed by elements when they combine into compounds. The first law says that any given compound has, a fixed proportion by mass of its constituents. The second law says that when two elements form more than one compound, for a fixed mass of one element, the masses of the other elements are in ratio of small integers.

To explain the laws Dalton suggested, about 200 years ago, that the smallest constituents of an element are atoms. Atoms of one element are identical but differ from those of other elements. A small number of atoms of each element combine to form a molecule of the compound. Gay Lussac's law, also given in early 19th century, states: When gases combine chemically to yield another gas, their volumes are in the ratios of small integers. Avogadro's law (or hypothesis) says: Equal volumes of all gases at equal temperature and pressure have the same number of molecules. Avogadro's law, when combined with Dalton's theory explains Gay Lussac's law. Since the elements are often in the form of molecules, Dalton's atomic theory can also be referred to as the molecular theory of matter. The theory is now well accepted by scientists. However even at the end of the nineteenth century there were famous scientists who did not believe in atomic theory !

From many observations, in recent times we now know that molecules (made up of one or more atoms) constitute matter. Electron microscopes and scanning tunnelling microscopes enable us to even see them. The size of an atom is about an angstrom (10^{-10} m). In solids, which are tightly packed, atoms are spaced about a few angstroms (2 Å) apart. In liquids the separation between atoms is also about the same. In liquids the atoms are not as rigidly fixed as in solids, and can move around. This enables a liquid to flow. In gases the interatomic distances are in tens of angstroms. The average distance a molecule can travel without colliding is called the **mean free path**. The mean free path, in gases, is of the order of thousands of angstroms. The atoms are much freer in gases and can travel long distances without colliding. If they are not enclosed, gases disperse away. In solids and liquids the closeness makes the interatomic force important. The force has a long range attraction and a short range repulsion. The atoms attract when they are at a few angstroms but repel when they come closer. The static appearance of a gas

is misleading. The gas is full of activity and the equilibrium is a dynamic one. In dynamic equilibrium, molecules collide and change their speeds during the collision. Only the average properties are constant.

Atomic theory is not the end of our quest, but the beginning. We now know that atoms are not indivisible or elementary. They consist of a nucleus and electrons. The nucleus itself is made up of protons and neutrons. The protons and neutrons are again made up of quarks. Even quarks may not be the end of the story. There may be string like elementary entities. Nature always has surprises for us, but the search for truth is often enjoyable and the discoveries beautiful. In this chapter, we shall limit ourselves to understanding the behaviour of gases (and a little bit of solids), as a collection of moving molecules in incessant motion.

13.3 BEHAVIOUR OF GASES

Properties of gases are easier to understand than those of solids and liquids. This is mainly because in a gas, molecules are far from each other and their mutual interactions are negligible except when two molecules collide. Gases at low pressures and high temperatures much above that at which they liquefy (or solidify) approximately satisfy a simple relation between their pressure, temperature and volume given by (see Ch. 11)

$$PV = KT \tag{13.1}$$

for a given sample of the gas. Here T is the temperature in kelvin or (absolute) scale. K is a constant for the given sample but varies with the volume of the gas. If we now bring in the idea of atoms or molecules then K is proportional to the number of molecules, (say) N in the sample. We can write $K = N k$. Observation tells us that this k is same for all gases. It is called Boltzmann constant and is denoted by k_B.

$$\text{As } \frac{P_1V_1}{N_1T_1} \quad \frac{P_2V_2}{N_2T_2} = \text{constant} = k_B \tag{13.2}$$

if P, V and T are same, then N is also same for all gases. This is Avogadro's hypothesis, that the number of molecules per unit volume is same for all gases at a fixed temperature and pressure. The number in 22.4 litres of any gas is 6.02×10^{23}. This is known as Avogadro number and is denoted by N_A. The mass of 22.4 litres of any gas is equal to its molecular weight in grams at S.T.P (standard temperature 273 K and pressure 1 atm). This amount of substance is called a mole (see Chapter 2 for a more precise definition). Avogadro had guessed the equality of numbers in equal volumes of gas at a fixed temperature and pressure from chemical reactions. Kinetic theory justifies this hypothesis.

The perfect gas equation can be written as

$$PV = \mu RT \tag{13.3}$$

where μ is the number of moles and $R = N_A k_B$ is a universal constant. The temperature T is absolute temperature. Choosing kelvin scale for

John Dalton (1766- 1844)

He was an English chemist. When different types of atoms combine, they obey certain simple laws. Dalton's atomic theory explains these laws in a simple way. He also gave a theory of colour blindness.

Amedeo Avogadro (1776 – 1856)

He made a brilliant guess that equal volumes of gases have equal number of molecules at the same temperature and pressure. This helped in understanding the combination of different gases in a very simple way. It is now called Avogadro's hypothesis (or law). He also suggested that the smallest constituent of gases like hydrogen, oxygen and nitrogen are not atoms but diatomic molecules.

absolute temperature, R = 8.314 J mol^{-1}K^{-1}. Here

$$\frac{M}{M_0} \quad \frac{N}{N_A} \tag{13.4}$$

where M is the mass of the gas containing N molecules, M_0 is the molar mass and N_A the Avogadro's number. Using Eqs. (13.4) and (13.3) can also be written as

$PV = k_B NT$ or $P = k_B nT$

Fig.13.1 *Real gases approach ideal gas behaviour at low pressures and high temperatures.*

where n is the number density, i.e. number of molecules per unit volume. k_B is the Boltzmann constant introduced above. Its value in SI units is 1.38 × 10^{-23} J K^{-1}.

Another useful form of Eq. (13.3) is

$$P \quad \frac{RT}{M_0} \tag{13.5}$$

where ρ is the mass density of the gas.

A gas that satisfies Eq. (13.3) exactly at all pressures and temperatures is defined to be an **ideal gas**. An ideal gas is a simple theoretical model of a gas. No real gas is truly ideal. Fig. 13.1 shows departures from ideal gas behaviour for a real gas at three different temperatures. Notice that all curves approach the ideal gas behaviour for low pressures and high temperatures.

At low pressures or high temperatures the molecules are far apart and molecular interactions are negligible. Without interactions the gas behaves like an ideal one.

If we fix μ and T in Eq. (13.3), we get

$$PV = \text{constant} \tag{13.6}$$

i.e., keeping temperature constant, pressure of a given mass of gas varies inversely with volume. This is the famous **Boyle's law**. Fig. 13.2 shows comparison between experimental P-V curves and the theoretical curves predicted by Boyle's law. Once again you see that the agreement is good at high temperatures and low pressures. Next, if you fix P, Eq. (13.1) shows that $V \propto T$ i.e., for a fixed pressure, the volume of a gas is proportional to its absolute temperature T **(Charles' law)**. See Fig. 13.3.

Fig.13.2 *Experimental P-V curves (solid lines) for steam at three temperatures compared with Boyle's law (dotted lines). P is in units of 22 atm and V in units of 0.09 litres.*

Finally, consider a mixture of non-interacting ideal gases: μ_1 moles of gas 1, μ_2 moles of gas 2, etc. in a vessel of volume V at temperature T and pressure P. It is then found that the equation of state of the mixture is :

$$PV = (\mu_1 + \mu_2 + \ldots)\, RT \tag{13.7}$$

$$\text{i.e. } P \quad {}_1\frac{RT}{V} \quad {}_2\frac{RT}{V} \quad \ldots \tag{13.8}$$

$$= P_1 + P_2 + \ldots \tag{13.9}$$

Clearly $P_1 = \mu_1 RT/V$ is the pressure gas 1 would exert at the same conditions of volume and temperature if no other gases were present. This is called the partial pressure of the gas. Thus, the total pressure of a mixture of ideal gases is the sum of partial pressures. This is Dalton's law of partial pressures.

Fig. 13.3 Experimental T-V curves (solid lines) for CO_2 at three pressures compared with Charles' law (dotted lines). T is in units of 300 K and V in units of 0.13 litres.

We next consider some examples which give us information about the volume occupied by the molecules and the volume of a single molecule.

Example 13.1 The density of water is 1000 kg m^{-3}. The density of water vapour at 100 °C and 1 atm pressure is 0.6 kg m^{-3}. The volume of a molecule multiplied by the total number gives ,what is called, molecular volume. Estimate the ratio (or fraction) of the molecular volume to the total volume occupied by the water vapour under the above conditions of temperature and pressure.

Answer For a given mass of water molecules, the density is less if volume is large. So the volume of the vapour is 1000/0.6 = /(6 ×10 $^{-4}$) times larger. If densities of bulk water and water molecules are same, then the fraction of molecular volume to the total volume in liquid state is 1. As volume in vapour state has increased, the fractional volume is less by the same amount, i.e. 6×10^{-4}. ◀

Example 13.2 Estimate the volume of a water molecule using the data in Example 13.1.

Answer In the liquid (or solid) phase, the molecules of water are quite closely packed. The density of water molecule may therefore, be regarded as roughly equal to the density of bulk water = 1000 kg m^{-3}. To estimate the volume of a water molecule, we need to know the mass of a single water molecule. We know that 1 mole of water has a mass approximately equal to

(2 + 16)g = 18 g = 0.018 kg.

Since 1 mole contains about 6×10^{23} molecules (Avogadro's number), the mass of a molecule of water is $(0.018)/(6 \times 10^{23})$ kg = 3×10^{-26} kg. Therefore, a rough estimate of the volume of a water molecule is as follows :

Volume of a water molecule

= $(3 \times 10^{-26}$ kg)/ (1000 kg m^{-3})

= 3×10^{-29} m^3

= (4/3) π (Radius)^3

Hence, Radius ≈ 2×10^{-10} m = 2 Å ◀

Example 13.3 What is the average distance between atoms (interatomic distance) in water? Use the data given in Examples 13.1 and 13.2.

Answer : A given mass of water in vapour state has 1.67×10^3 times the volume of the same mass of water in liquid state (Ex. 13.1). This is also the increase in the amount of volume available for each molecule of water. When volume increases by 10^3 times the radius increases by $V^{1/3}$ or 10 times, i.e., 10 × 2 Å = 20 Å. So the average distance is 2 × 20 = 40 Å. ◀

Example 13.4 A vessel contains two non-reactive gases : neon (monatomic) and oxygen (diatomic). The ratio of their partial pressures is 3:2. Estimate the ratio of (i) number of molecules and (ii) mass density of neon and oxygen in the vessel. Atomic mass of Ne = 20.2 u, molecular mass of O_2 = 32.0 u.

Answer Partial pressure of a gas in a mixture is the pressure it would have for the same volume and temperature if it alone occupied the vessel. (The total pressure of a mixture of non-reactive gases is the sum of partial pressures due to its constituent gases.) Each gas (assumed ideal) obeys the gas law. Since V and T are common to the two gases, we have $P_1V = \mu_1 RT$ and $P_2V = \mu_2 RT$, i.e. $(P_1/P_2) = (\mu_1 / \mu_2)$. Here 1 and 2 refer to neon and oxygen respectively. Since (P_1/P_2) = (3/2) (given), $(\mu_1/ \mu_2) = 3/2$.

(i) By definition $\mu_1 = (N_1/N_A)$ and $\mu_2 = (N_2/N_A)$ where N_1 and N_2 are the number of molecules of 1 and 2, and N_A is the Avogadro's number. Therefore, $(N_1/N_2) = (\mu_1 / \mu_2) = 3/2$.

(ii) We can also write $\mu_1 = (m_1/M_1)$ and $\mu_2 = (m_2/M_2)$ where m_1 and m_2 are the masses of 1 and 2; and M_1 and M_2 are their molecular masses. (Both m_1 and M_1; as well as m_2 and M_2 should be expressed in the same units). If ρ_1 and ρ_2 are the mass densities of 1 and 2 respectively, we have

$$\frac{1}{2}\ \frac{m_1/V}{m_2/V}\ \frac{m_1}{m_2}\ \frac{1}{2}\ \frac{M_1}{M_2}$$

$$\frac{3}{2}\ \frac{20.2}{32.0}\ 0.947$$ ◀

13.4 KINETIC THEORY OF AN IDEAL GAS

Kinetic theory of gases is based on the molecular picture of matter. A given amount of gas is a collection of a large number of molecules (typically of the order of Avogadro's number) that are in incessant random motion. At ordinary pressure and temperature, the average distance between molecules is a factor of 10 or more than the typical size of a molecule (2 Å). Thus the interaction between the molecules is negligible and we can assume that they move freely in straight lines according to Newton's first law. However, occasionally, they come close to each other, experience intermolecular forces and their velocities change. These interactions are called collisions. The molecules collide incessantly against each other or with the walls and change their velocities. The collisions are considered to be elastic. We can derive an expression for the pressure of a gas based on the kinetic theory.

We begin with the idea that molecules of a gas are in incessant random motion, colliding against one another and with the walls of the container. All collisions between molecules among themselves or between molecules and the walls are elastic. This implies that total kinetic energy is conserved. The total momentum is conserved as usual.

13.4.1 Pressure of an Ideal Gas

Consider a gas enclosed in a cube of side l. Take the axes to be parallel to the sides of the cube, as shown in Fig. 13.4. A molecule with velocity

Fig. 13.4 Elastic collision of a gas molecule with the wall of the container.

(v_x, v_y, v_z) hits the planar wall parallel to yz-plane of area A (= l^2). Since the collision is elastic, the molecule rebounds with the same velocity; its y and z components of velocity do not change in the collision but the x-component reverses sign. That is, the velocity after collision is $(-v_x, v_y, v_z)$. The change in momentum of the molecule is : $-mv_x - (mv_x) = -2mv_x$. By the principle of conservation of momentum, the momentum imparted to the wall in the collision $= 2mv_x$.

To calculate the force (and pressure) on the wall, we need to calculate momentum imparted to the wall per unit time. In a small time interval Δt, a molecule with x-component of velocity v_x will hit the wall if it is within the distance $v_x \Delta t$ from the wall. That is, all molecules within the volume $Av_x \Delta t$ only can hit the wall in time Δt. But, on the average, half of these are moving towards the wall and the other half away from the wall. Thus the number of molecules with velocity (v_x, v_y, v_z) hitting the wall in time Δt is $\frac{1}{2} A\, v_x\, \Delta t\, n$ where n is the number of molecules per unit volume. The total momentum transferred to the wall by these molecules in time Δt is :

$$Q = (2mv_x)\,(\tfrac{1}{2}\, n\, A\, v_x\, \Delta t) \qquad (13.10)$$

The force on the wall is the rate of momentum transfer $Q/\Delta t$ and pressure is force per unit area :

$$P = Q/(A\,\Delta t) = n\, m\, v_x^2 \qquad (3.11)$$

Actually, all molecules in a gas do not have the same velocity; there is a distribution in velocities. The above equation therefore, stands for pressure due to the group of molecules with speed v_x in the x-direction and n stands for the number density of that group of molecules. The

total pressure is obtained by summing over the contribution due to all groups:

$$P = n\, m\, \overline{v_x^2} \qquad (13.12)$$

where $\overline{v_x^2}$ is the average of v_x^2. Now the gas is isotropic, i.e. there is no preferred direction of velocity of the molecules in the vessel. Therefore, by symmetry,

$$\overline{v_x^2} = \overline{v_y^2} = \overline{v_z^2}$$

$$= (1/3)\,[\overline{v_x^2} + \overline{v_y^2} + \overline{v_z^2}] = (1/3)\,\overline{v^2} \qquad (13.13)$$

where v is the speed and $\overline{v^2}$ denotes the mean of the squared speed. Thus

$$P = (1/3)\, n\, m\, \overline{v^2} \qquad (13.14)$$

Some remarks on this derivation. First, though we choose the container to be a cube, the shape of the vessel really is immaterial. For a vessel of arbitrary shape, we can always choose a small infinitesimal (planar) area and carry through the steps above. Notice that both A and Δt do not appear in the final result. By Pascal's law, given in Ch. 10, pressure in one portion of the gas in equilibrium is the same as anywhere else. Second, we have ignored any collisions in the derivation. Though this assumption is difficult to justify rigorously, we can qualitatively see that it will not lead to erroneous results. The number of molecules hitting the wall in time Δt was found to be $\frac{1}{2}\, n A v_x \Delta t$. Now the collisions are random and the gas is in a steady state. Thus, if a molecule with velocity (v_x, v_y, v_z) acquires a different velocity due to collision with some molecule, there will always be some other molecule with a different initial velocity which after a collision acquires the velocity (v_x, v_y, v_z). If this were not so, the distribution of velocities would not remain steady. In any case we are finding $\overline{v_x^2}$. Thus, on the whole, molecular collisions (if they are not too frequent and the time spent in a collision is negligible compared to time between collisions) will not affect the calculation above.

13.4.2 Kinetic Interpretation of Temperature

Equation (13.14) can be written as

$$PV = (1/3)\, nV\, m\, \overline{v^2} \qquad (13.15a)$$

Founders of Kinetic Theory of Gases

James Clerk Maxwell (1831 – 1879), born in Edinburgh, Scotland, was among the greatest physicists of the nineteenth century. He derived the thermal velocity distribution of molecules in a gas and was among the first to obtain reliable estimates of molecular parameters from measurable quantities like viscosity, etc. Maxwell's greatest achievement was the unification of the laws of electricity and magnetism (discovered by Coulomb, Oersted, Ampere and Faraday) into a consistent set of equations now called Maxwell's equations. From these he arrived at the most important conclusion that light is an electromagnetic wave. Interestingly, Maxwell did not agree with the idea (strongly suggested by the Faraday's laws of electrolysis) that electricity was particulate in nature.

Ludwig Boltzmann (1844 – 1906) born in Vienna, Austria, worked on the kinetic theory of gases independently of Maxwell. A firm advocate of atomism, that is basic to kinetic theory, Boltzmann provided a statistical interpretation of the Second Law of thermodynamics and the concept of entropy. He is regarded as one of the founders of classical statistical mechanics. The proportionality constant connecting energy and temperature in kinetic theory is known as Boltzmann's constant in his honour.

$$PV = (2/3)\, N \times \tfrac{1}{2}\, m\, \overline{v^2} \qquad (13.15b)$$

where N (= nV) is the number of molecules in the sample.

The quantity in the bracket is the average translational kinetic energy of the molecules in the gas. Since the internal energy E of an ideal gas is purely kinetic*,

$$E = N \times (1/2)\, m\, \overline{v^2} \qquad (13.16)$$

Equation (13.15) then gives :

$$PV = (2/3)\, E \qquad (13.17)$$

We are now ready for a kinetic interpretation of temperature. Combining Eq. (13.17) with the ideal gas Eq. (13.3), we get

$$E = (3/2)\, k_B\, NT \qquad (13.18)$$

$$\text{or} \quad E/N = \tfrac{1}{2}\, m\, \overline{v^2} = (3/2)\, k_B T \qquad (13.19)$$

i.e., the average kinetic energy of a molecule is proportional to the absolute temperature of the gas; it is independent of pressure, volume or the nature of the ideal gas. This is a fundamental result relating temperature, a macroscopic measurable parameter of a gas (a thermodynamic variable as it is called) to a molecular quantity, namely the average kinetic energy of a molecule. The two domains are connected by the Boltzmann constant. We note in passing that Eq. (13.18) tells us that internal energy of an ideal gas depends only on temperature, not on pressure or volume. With this interpretation of temperature, kinetic theory of an ideal gas is completely consistent with the ideal gas equation and the various gas laws based on it.

For a mixture of non-reactive ideal gases, the total pressure gets contribution from each gas in the mixture. Equation (13.14) becomes

$$P = (1/3)\, [n_1 m_1 \overline{v_1^2} + n_2 m_2 \overline{v_2^2} + \ldots] \qquad (13.20)$$

In equilibrium, the average kinetic energy of the molecules of different gases will be equal. That is,

$$\tfrac{1}{2}\, m_1\, \overline{v_1^2} = \tfrac{1}{2}\, m_2\, \overline{v_2^2} = (3/2)\, k_B\, T$$

so that

$$P = (n_1 + n_2 + \ldots)\, k_B\, T \qquad (13.21)$$

which is Dalton's law of partial pressures.

From Eq. (13.19), we can get an idea of the typical speed of molecules in a gas. At a temperature T = 300 K, the mean square speed of a molecule in nitrogen gas is :

$$m = \frac{M_{N_2}}{N_A} = \frac{28}{6.02 \times 10^{26}} = 4.65 \times 10^{-26}\ \text{kg}.$$

$$\overline{v^2} = 3\, k_B\, T / m = (516)^2\ \text{m}^2\text{s}^{-2}$$

The square root of $\overline{v^2}$ is known as root mean square (rms) speed and is denoted by v_{rms},

(We can also write $\overline{v^2}$ as $\langle v^2 \rangle$.)

$$v_{rms} = 516\ \text{m s}^{-1}$$

The speed is of the order of the speed of sound in air. It follows from Eq. (13.19) that at the same temperature, lighter molecules have greater rms speed.

Example 13.5 A flask contains argon and chlorine in the ratio of 2:1 by mass. The temperature of the mixture is 27 °C. Obtain the ratio of (i) average kinetic energy per molecule, and (ii) root mean square speed v_{rms} of the molecules of the two gases. Atomic mass of argon = 39.9 u; Molecular mass of chlorine = 70.9 u.

Answer The important point to remember is that the average kinetic energy (per molecule) of any (ideal) gas (be it monatomic like argon, diatomic like chlorine or polyatomic) is always equal to (3/2) $k_B T$. It depends only on temperature, and is independent of the nature of the gas.

(i) Since argon and chlorine both have the same temperature in the flask, the ratio of average kinetic energy (per molecule) of the two gases is 1:1.

(ii) Now $\tfrac{1}{2}\, m\, v_{rms}^2$ = average kinetic energy per molecule = (3/2)) $k_B T$ where m is the mass of a molecule of the gas. Therefore,

$$\frac{\mathbf{v}_{rms\ Ar}^2}{\mathbf{v}_{rms\ Cl}^2} = \frac{m_{Cl}}{m_{Ar}} = \frac{M_{Cl}}{M_{Ar}} = \frac{70.9}{39.9} = 1.77$$

where M denotes the molecular mass of the gas. (For argon, a molecule is just an atom of argon.) Taking square root of both sides,

$$\frac{\mathbf{v}_{rms\ Ar}}{\mathbf{v}_{rms\ Cl}} = 1.33$$

You should note that the composition of the mixture by mass is quite irrelevant to the above

* *E denotes the translational part of the internal energy U that may include energies due to other degrees of freedom also. See section 13.5.*

Maxwell Distribution Function

In a given mass of gas, the velocities of all molecules are not the same, even when bulk parameters like pressure, volume and temperature are fixed. Collisions change the direction and the speed of molecules. However in a state of equilibrium, the distribution of speeds is constant or fixed.

Distributions are very important and useful when dealing with systems containing large number of objects. As an example consider the ages of different persons in a city. It is not feasible to deal with the age of each individual. We can divide the people into groups: children up to age 20 years, adults between ages of 20 and 60, old people above 60. If we want more detailed information we can choose smaller intervals, 0-1, 1-2,..., 99-100 of age groups. When the size of the interval becomes smaller, say half year, the number of persons in the interval will also reduce, roughly half the original number in the one year interval. The number of persons $dN(x)$ in the age interval x and $x+dx$ is proportional to dx or $dN(x) = n_x\, dx$. We have used n_x to denote the number of persons at the value of x.

Maxwell distribution of molecular speeds

In a similar way the molecular speed distribution gives the number of molecules between the speeds v and $v+dv$. $dN(v) = 4p\, N\, a^3 e^{-bv^2}\, v^2\, dv = n_v dv$. This is called Maxwell distribution. The plot of n_v against v is shown in the figure. The fraction of the molecules with speeds v and $v+dv$ is equal to the area of the strip shown. The average of any quantity like v^2 is defined by the integral $\langle v^2 \rangle = (1/N) \int v^2\, dN(v) = \text{Å}(3k_B T/m)$ which agrees with the result derived from more elementary considerations.

calculation. Any other proportion by mass of argon and chlorine would give the same answers to (i) and (ii), provided the temperature remains unaltered. ◀

▶ ***Example 13.6*** Uranium has two isotopes of masses 235 and 238 units. If both are present in Uranium hexafluoride gas which would have the larger average speed ? If atomic mass of fluorine is 19 units, estimate the percentage difference in speeds at any temperature.

Answer At a fixed temperature the average energy $= ½\, m \langle v^2 \rangle$ is constant. So smaller the mass of the molecule, faster will be the speed. The ratio of speeds is inversely proportional to the square root of the ratio of the masses. The masses are 349 and 352 units. So

$$v_{349} / v_{352} = (352/349)^{1/2} = 1.0044 .$$

Hence difference $\frac{V}{V} = 0.44\ \%$.

[^{235}U is the isotope needed for nuclear fission. To separate it from the more abundant isotope ^{238}U, the mixture is surrounded by a porous cylinder. The porous cylinder must be thick and narrow, so that the molecule wanders through individually, colliding with the walls of the long pore. The faster molecule will leak out more than

the slower one and so there is more of the lighter molecule (enrichment) outside the porous cylinder (Fig. 13.5). The method is not very efficient and has to be repeated several times for sufficient enrichment.]. ◀

When gases diffuse, their rate of diffusion is inversely proportional to square root of the masses (see Exercise 13.12). Can you guess the explanation from the above answer?

Fig. 13.5 Molecules going through a porous wall.

▶ ***Example 13.7*** (a) When a molecule (or an elastic ball) hits a (massive) wall, it rebounds with the same speed. When a ball hits a massive bat held firmly, the same thing happens. However, when the bat is moving towards the ball, the ball rebounds with a different speed. Does the ball move faster or slower? (Ch.6 will refresh your memory on elastic collisions.)

(b) When gas in a cylinder is compressed by pushing in a piston, its temperature rises. Guess at an explanation of this in terms of kinetic theory using (a) above.

(c) What happens when a compressed gas pushes a piston out and expands. What would you observe ?

(d) Sachin Tendulkar uses a heavy cricket bat while playing. Does it help him in anyway ?

Answer (a) Let the speed of the ball be u relative to the wicket behind the bat. If the bat is moving towards the ball with a speed V relative to the wicket, then the relative speed of the ball to bat is $V + u$ towards the bat. When the ball rebounds (after hitting the massive bat) its speed, relative to bat, is $V + u$ moving away from the bat. So relative to the wicket the speed of the rebounding ball is $V + (V + u) = 2V + u$, moving away from the wicket. So the ball speeds up after the collision with the bat. The rebound speed will be less than u if the bat is not massive. For a molecule this would imply an increase in temperature.

You should be able to answer (b) (c) and (d) based on the answer to (a).

(Hint: Note the correspondence, piston→ bat, cylinder → wicket, molecule → ball.) ◀

13.5 LAW OF EQUIPARTITION OF ENERGY

The kinetic energy of a single molecule is

$$_t \quad \frac{1}{2}mv_x^2 \quad \frac{1}{2}mv_y^2 \quad \frac{1}{2}mv_z^2 \qquad (13.22)$$

For a gas in thermal equilibrium at temperature T the average value of energy denoted by $< \ _t >$ is

$$\langle \ _t \rangle \quad \left\langle \frac{1}{2}mv_x^2 \right\rangle \quad \left\langle \frac{1}{2}mv_y^2 \right\rangle \quad \left\langle \frac{1}{2}mv_z^2 \right\rangle \quad \frac{3}{2}k_BT \qquad (13.23)$$

Since there is no preferred direction, Eq. (13.23) implies

$$\left\langle \frac{1}{2}\ mv_x^2 \right\rangle \quad \frac{1}{2}\ k_BT\ , \left\langle \frac{1}{2}\ mv_y^2 \right\rangle \quad \frac{1}{2}\ k_BT\ ,$$

$$\left\langle \frac{1}{2}\ mv_z^2 \right\rangle \quad \frac{1}{2}\ k_BT \qquad (13.24)$$

A molecule free to move in space needs three coordinates to specify its location. If it is constrained to move in a plane it needs two;and if constrained to move along a line, it needs just one coordinate to locate it. This can also be expressed in another way. We say that it has one degree of freedom for motion in a line, two for motion in a plane and three for motion in space. Motion of a body as a whole from one point to another is called translation. Thus, a molecule free to move in space has three translational degrees of freedom. Each translational degree of freedom contributes a term that contains square of some variable of motion, e.g., ½ mv_x^2 and similar terms in v_y and v_z. In, Eq. (13.24) we see that in thermal equilibrium, the average of each such term is ½ k_BT .

Molecules of a monatomic gas like argon have only translational degrees of freedom. But what about a diatomic gas such as O_2 or N_2? A molecule of O_2 has three translational degrees of freedom. But in addition it can also rotate about its centre of mass. Figure 13.6 shows the two independent axes of rotation 1 and 2, normal to the axis joining the two oxygen atoms about which the molecule can rotate*. The molecule thus has two rotational degrees of freedom, each of which contributes a term to the total energy consisting of translational energy $_t$ and rotational energy $_r$.

$$_t \quad _t \quad \frac{1}{2}mv_x^2 \quad \frac{1}{2}mv_y^2 \quad \frac{1}{2}mv_z^2 \quad \frac{1}{2}I_1 \; _1^2 \quad \frac{1}{2}I_2 \; _2^2 \qquad (13.25)$$

Fig. 13.6 The two independent axes of rotation of a diatomic molecule

where ω_1 and ω_2 are the angular speeds about the axes 1 and 2 and I_1, I_2 are the corresponding moments of inertia. Note that each rotational degree of freedom contributes a term to the energy that contains square of a rotational variable of motion.

We have assumed above that the O_2 molecule is a 'rigid rotator', i.e. the molecule does not vibrate. This assumption, though found to be true (at moderate temperatures) for O_2, is not always valid. Molecules like CO even at moderate temperatures have a mode of vibration, i.e. its atoms oscillate along the interatomic axis like a one-dimensional oscillator, and contribute a vibrational energy term ε_v to the total energy:

$$_v \quad \frac{1}{2}m \left(\frac{dy}{dt}\right)^2 \quad \frac{1}{2}ky^2$$

$$_t \quad _r \quad _v \qquad (13.26)$$

where k is the force constant of the oscillator and y the vibrational co-ordinate.

Once again the vibrational energy terms in Eq. (13.26) contain squared terms of vibrational variables of motion y and dy/dt.

At this point, notice an important feature in Eq.(13.26). While each translational and rotational degree of freedom has contributed only one 'squared term' in Eq.(13.26), one vibrational mode contributes two 'squared terms' : kinetic and potential energies.

Each quadratic term occurring in the expression for energy is a mode of absorption of energy by the molecule. We have seen that in thermal equilibrium at absolute temperature T, for each translational mode of motion, the average energy is ½ k_BT. A most elegant principle of classical statistical mechanics (first proved by Maxwell) states that this is so for each mode of energy: translational, rotational and vibrational. That is, in equilibrium, the total energy is equally distributed in all possible energy modes, with each mode having an average energy equal to ½ k_BT. This is known as the **law of equipartition of energy**. Accordingly, each translational and rotational degree of freedom of a molecule contributes ½ k_BT to the energy while each vibrational frequency contributes $2 \times ½\, k_BT = k_BT$, since a vibrational mode has both kinetic and potential energy modes.

The proof of the law of equipartition of energy is beyond the scope of this book. Here we shall apply the law to predict the specific heats of gases theoretically. Later we shall also discuss briefly, the application to specific heat of solids.

13.6 SPECIFIC HEAT CAPACITY

13.6.1 Monatomic Gases

The molecule of a monatomic gas has only three translational degrees of freedom. Thus, the average energy of a molecule at temperature T is $(3/2)k_BT$. The total internal energy of a mole of such a gas is

* *Rotation along the line joining the atoms has very small moment of inertia and does not come into play for quantum mechanical reasons. See end of section 13.6.*

$$U \quad \frac{3}{2}k_B T \quad N_A \quad \frac{3}{2}RT \qquad (13.27)$$

The molar specific heat at constant volume, C_v, is

$$C_v \text{ (monatomic gas)} = \frac{dU}{dT} = \frac{3}{2}RT \qquad (13.28)$$

For an ideal gas,

$$C_p - C_v = R \qquad (13.29)$$

where C_p is the molar specific heat at constant pressure. Thus,

$$C_p = \frac{5}{2}R \qquad (13.30)$$

The ratio of specific heats $\frac{C_p}{C_v} \quad \frac{5}{3}$ (13.31)

13.6.2 Diatomic Gases

As explained earlier, a diatomic molecule treated as a rigid rotator like a dumbbell has 5 degrees of freedom : 3 translational and 2 rotational. Using the law of equipartition of energy, the total internal energy of a mole of such a gas is

$$U \quad \frac{5}{2}k_B T \quad N_A \quad \frac{5}{2}RT \qquad (13.32)$$

The molar specific heats are then given by

$$C_v \text{ (rigid diatomic)} = \frac{5}{2}R, \; C_p = \frac{7}{2}R \qquad (13.33)$$

$$\gamma \text{ (rigid diatomic)} = \frac{7}{5} \qquad (13.34)$$

If the diatomic molecule is not rigid but has in addition a vibrational mode

$$U = \left(\frac{5}{2}k_B T + k_B T\right)N_A = \frac{7}{2}RT$$

$$C_v \quad \frac{7}{2}R, \; C_p \quad \frac{9}{2}R, \quad \frac{9}{7}R \qquad (13.35)$$

13.6.3 Polyatomic Gases

In general a polyatomic molecule has 3 translational, 3 rotational degrees of freedom and a certain number (f) of vibrational modes. According to the law of equipartition of energy, it is easily seen that one mole of such a gas has

$$U = \left(\frac{3}{2}k_B T + \frac{3}{2}k_B T + f k_B T\right) N_A$$

i.e. $C_v = (3 + f)\,R$, $C_p = (4 + f)\,R$,

$$\frac{f}{f} \qquad (13.36)$$

Note that $C_p - C_v = R$ is true for any ideal gas, whether mono, di or polyatomic.

Table 13.1 summarises the theoretical predictions for specific heats of gases ignoring any vibrational modes of motion. The values are in good agreement with experimental values of specific heats of several gases given in Table 13.2. Of course, there are discrepancies between predicted and actual values of specific heats of several other gases (not shown in the table), such as Cl_2, C_2H_6 and many other polyatomic gases. Usually, the experimental values for specific heats of these gases are greater than the predicted values given in Table 13.1 suggesting that the agreement can be improved by including vibrational modes of motion in the calculation. The law of equipartition of energy is thus well

Table 13.1 Predicted values of specific heat capacities of gases (ignoring vibrational modes),

Nature of Gas	C_v (J mol^{-1} K^{-1})	C_p (J mol^{-1} K^{-1})	$C_p - C_v$ (J mol^{-1} K^{-1})	γ
Monatomic	12.5	20.8	8.31	1.67
Diatomic	20.8	29.1	8.31	1.40
Triatomic	24.93	33.24	8.31	1.33

Table13.2 Measured values of specific heat capacities of some gases

Nature of gas	Gas	C_v (J mol^{-1} K^{-1})	C_p (J mol^{-1} K^{-1})	$C_p - C_v$ (J mol^{-1} K^{-1})	γ
Monatomic	He	12.5	20.8	8.30	1.66
Monatomic	Ne	12.7	20.8	8.12	1.64
Monatomic	Ar	12.5	20.8	8.30	1.67
Diatomic	H_2	20.4	28.8	8.45	1.41
Diatomic	O_2	21.0	29.3	8.32	1.40
Diatomic	N_2	20.8	29.1	8.32	1.40
Triatomic	H_2O	27.0	35.4	8.35	1.31
Polyatomic	CH_4	27.1	35.4	8.36	1.31

verified experimentally at ordinary temperatures.

Example 13.8 A cylinder of fixed capacity 44.8 litres contains helium gas at standard temperature and pressure. What is the amount of heat needed to raise the temperature of the gas in the cylinder by 15.0 °C ? (R = 8.31 J mol^{-1} K^{-1}).

Answer Using the gas law $PV = \mu RT$, you can easily show that 1 mol of any (ideal) gas at standard temperature (273 K) and pressure (1 atm = 1.01 × 10^5 Pa) occupies a volume of 22.4 litres. This universal volume is called molar volume. Thus the cylinder in this example contains 2 mol of helium. Further, since helium is monatomic, its predicted (and observed) molar specific heat at constant volume, $C_v = (3/2)\ R$, and molar specific heat at constant pressure, $C_p = (3/2)\ R + R = (5/2)\ R$. Since the volume of the cylinder is fixed, the heat required is determined by C_v. Therefore,

Heat required = no. of moles × molar specific heat × rise in temperature

$= 2 \times 1.5\ R \times 15.0 = 45\ R$

$= 45 \times 8.31 = 374$ J.

13.6.4 Specific Heat Capacity of Solids

We can use the law of equipartition of energy to determine specific heats of solids. Consider a solid of N atoms, each vibrating about its mean position. An oscillation in one dimension has average energy of $2 \times ½\ k_BT = k_BT$. In three dimensions, the average energy is $3\ k_BT$. For a mole of solid, $N = N_A$, and the total energy is

$$U = 3\ k_BT \times N_A = 3\ RT$$

Now at constant pressure $\Delta Q = \Delta U + P\Delta V = \Delta U$, since for a solid ΔV is negligible. Hence,

$$C \quad \frac{Q}{T} \quad \frac{U}{T} \quad 3R \qquad (13.37)$$

Table 13.3 Specific Heat Capacity of some solids at room temperature and atmospheric pressure

Substance	Specific heat (J kg^{-1} K^{-1})	Molar specific Heat(J mol^{-1} K^{-1})
Aluminium	900.0	24.4
Carbon	506.5	6.1
Copper	386.4	24.5
Lead	127.7	26.5
Silver	236.1	25.5
Tungsten	134.4	24.9

As Table 13.3 shows the prediction generally agrees with experimental values at ordinary temperature (Carbon is an exception).

13.6.5 Specific Heat Capacity of Water

We treat water like a solid. For each atom average energy is $3k_BT$. Water molecule has three atoms, two hydrogen and one oxygen. So it has

$$U = 3 \times 3\ k_BT \times N_A = 9\ RT$$

and $C = \Delta Q/\ \Delta T = \Delta\ U\ /\ \Delta T = 9R$.

This is the value observed and the agreement is very good. In the calorie, gram, degree units, water is defined to have unit specific heat. As 1 calorie = 4.179 joules and one mole of water is 18 grams, the heat capacity per mole is ~ 75 J mol^{-1} K^{-1} ~ 9R. However with more complex molecules like alcohol or acetone the arguments, based on degrees of freedom, become more complicated.

Lastly, we should note an important aspect of the predictions of specific heats, based on the classical law of equipartition of energy. The predicted specific heats are independent of temperature. As we go to low temperatures, however, there is a marked departure from this prediction. Specific heats of all substances approach zero as $T \rightarrow 0$. This is related to the fact that degrees of freedom get frozen and ineffective at low temperatures. According to classical physics degrees of freedom must remain unchanged at all times. The behaviour of specific heats at low temperatures shows the inadequacy of classical physics and can be explained only by invoking quantum considerations, as was first shown by Einstein. Quantum mechanics requires a minimum, nonzero amount of energy before a degree of freedom comes into play. This is also the reason why vibrational degrees of freedom come into play only in some cases.

13.7 MEAN FREE PATH

Molecules in a gas have rather large speeds of the order of the speed of sound. Yet a gas leaking from a cylinder in a kitchen takes considerable time to diffuse to the other corners of the room. The top of a cloud of smoke holds together for hours. This happens because molecules in a gas have a finite though small size, so they are bound to undergo collisions. As a result, they cannot

Seeing is Believing

Can one see atoms rushing about. Almost but not quite. One can see pollen grains of a flower being pushed around by molecules of water. The size of the grain is ~ 10^{-5} m. In 1827, a Scottish botanist Robert Brown, while examining, under a microscope, pollen grains of a flower suspended in water noticed that they continuously moved about in a zigzag, random fashion.

Kinetic theory provides a simple explanation of the phenomenon. Any object suspended in water is continuously bombarded from all sides by the water molecules. Since the motion of molecules is random, the number of molecules hitting the object in any direction is about the same as the number hitting in the opposite direction. The small difference between these molecular hits is negligible compared to the total number of hits for an object of ordinary size, and we do not notice any movement of the object.

When the object is sufficiently small but still visible under a microscope, the difference in molecular hits from different directions is not altogether negligible, i.e. the impulses and the torques given to the suspended object through continuous bombardment by the molecules of the medium (water or some other fluid) do not exactly sum to zero. There is a net impulse and torque in this or that direction. The suspended object thus, moves about in a zigzag manner and tumbles about randomly. This motion called now 'Brownian motion' is a visible proof of molecular activity. In the last 50 years or so molecules have been seen by scanning tunneling and other special microscopes.

In 1987 Ahmed Zewail, an Egyptian scientist working in USA was able to observe not only the molecules but also their detailed interactions. He did this by illuminating them with flashes of laser light for very short durations, of the order of tens of femtoseconds and photographing them. (1 femto-second = 10^{-15} s). One could study even the formation and breaking of chemical bonds. That is really seeing !

move straight unhindered; their paths keep getting incessantly deflected.

Fig. 13.7 The volume swept by a molecule in time Δt in which any molecule will collide with it.

Suppose the molecules of a gas are spheres of diameter d. Focus on a single molecule with the average speed $\langle v \rangle$. It will suffer collision with any molecule that comes within a distance d between the centres. In time Δt, it sweeps a volume $\pi d^2 \langle v \rangle \Delta t$ wherein any other molecule will collide with it (see Fig. 13.7). If n is the number of molecules per unit volume, the molecule suffers $n\pi d^2 \langle v \rangle \Delta t$ collisions in time Δt. Thus the rate of collisions is $n\pi d^2 \langle v \rangle$ or the time between two successive collisions is on the average,

$$\tau = 1/(n\pi \langle v \rangle d^2) \qquad (13.38)$$

The average distance between two successive collisions, called the mean free path l, is :

$$l = \langle v \rangle \tau = 1/(n\pi d^2) \qquad (13.39)$$

In this derivation, we imagined the other molecules to be at rest. But actually all molecules are moving and the collision rate is determined by the average relative velocity of the molecules. Thus we need to replace $\langle v \rangle$ by $\langle v_r \rangle$ in Eq. (13.38). A more exact treatment gives

$$l \quad 1/\sqrt{2}\, n \; d^2 \qquad (13.40)$$

Let us estimate l and τ for air molecules with average speeds $\langle v \rangle$ = (485m/s). At STP

$$n = \frac{0.02 \quad 10^{23}}{22.4 \quad 10^{-3}}$$

$= 2.7 \times 10^{25}$ m $^{-3}$.

Taking, $d = 2 \times 10^{-10}$ m,

$\tau = 6.1 \times 10^{-10}$ s

and $l = 2.9 \times 10^{-7}$ m $\approx$ 1500d (13.41)

As expected, the mean free path given by Eq. (13.40) depends inversely on the number density and the size of the molecules. In a highly evacuated tube n is rather small and the mean free path can be as large as the length of the tube.

> ***Example 13.9*** Estimate the mean free path for a water molecule in water vapour at 373 K. Use information from Exercises 13.1 and Eq. (13.41) above.

Answer The d for water vapour is same as that of air. The number density is inversely proportional to absolute temperature.

So $n \quad 2.7 \quad 10^{25} \quad \frac{273}{373} \quad 2 \quad 10^{25}\,\text{m}^{-3}$

Hence, mean free path $l \quad 4 \quad 10^{-7}\,\text{m}$ ◀

Note that the mean free path is 100 times the interatomic distance ~ 40 Å = 4×10^{-9} m calculated earlier. It is this large value of mean free path that leads to the typical gaseous behaviour. Gases can not be confined without a container.

Using, the kinetic theory of gases, the bulk measurable properties like viscosity, heat conductivity and diffusion can be related to the microscopic parameters like molecular size. It is through such relations that the molecular sizes were first estimated.

SUMMARY

1. The ideal gas equation connecting pressure (P), volume (V) and absolute temperature (T) is

$$PV = \mu RT \quad = k_B NT$$

where μ is the number of moles and N is the number of molecules. R and k_B are universal constants.

$$R = 8.314 \text{ J mol}^{-1}\text{ K}^{-1}, \quad k_B = \frac{R}{N_A} = 1.38 \times 10^{-23} \text{ J K}^{-1}$$

Real gases satisfy the ideal gas equation only approximately, more so at low pressures and high temperatures.

2. Kinetic theory of an ideal gas gives the relation

$$P \quad \frac{1}{3} n\, m\, \overline{v^2}$$

where n is number density of molecules, m the mass of the molecule and $\overline{v^2}$ is the mean of squared speed. Combined with the ideal gas equation it yields a kinetic interpretation of temperature.

$$\frac{1}{2} m\, \overline{v^2} \quad \frac{3}{2} k_B\, T, \quad v_{rms} \quad \overline{v^2}^{\,1/2} \quad \sqrt{\frac{3k_B T}{m}}$$

This tells us that the temperature of a gas is a measure of the average kinetic energy of a molecule, *independent of the nature of the gas or molecule.* In a mixture of gases at a fixed temperature the heavier molecule has the lower average speed.

3. The translational kinetic energy

$$E = \frac{3}{2} k_B NT.$$

This leads to a relation

$$PV = \frac{2}{3} E$$

4. The law of equipartition of energy states that if a system is in equilibrium at absolute temperature T, the total energy is distributed equally in different energy modes of

absorption, the energy in each mode being equal to ½ $k_B T$. Each translational and rotational degree of freedom corresponds to one energy mode of absorption and has energy ½ $k_B T$. Each vibrational frequency has two modes of energy (kinetic and potential) with corresponding energy equal to
$2 \times ½\ k_B T = k_B T$.

5. Using the law of equipartition of energy, the molar specific heats of gases can be determined and the values are in agreement with the experimental values of specific heats of several gases. The agreement can be improved by including vibrational modes of motion.

6. The mean free path l is the average distance covered by a molecule between two successive collisions :

$$\bar{l} \quad \frac{1}{\sqrt{2}\, n \quad d^2}$$

where n is the number density and d the diameter of the molecule.

POINTS TO PONDER

1. Pressure of a fluid is not only exerted on the wall. Pressure exists everywhere in a fluid. Any layer of gas inside the volume of a container is in equilibrium because the pressure is the same on both sides of the layer.

2. We should not have an exaggerated idea of the intermolecular distance in a gas. At ordinary pressures and temperatures, this is only 10 times or so the interatomic distance in solids and liquids. What is different is the mean free path which in a gas is 100 times the interatomic distance and 1000 times the size of the molecule.

3. The law of equipartition of energy is stated thus: the energy for each degree of freedom in thermal equilibrium is ½ $k_B T$. Each quadratic term in the total energy expression of a molecule is to be counted as a degree of freedom. Thus, each vibrational mode gives 2 (not 1) degrees of freedom (kinetic and potential energy modes), corresponding to the energy $2 \times ½\ k_B T = k_B T$.

4. Molecules of air in a room do not all fall and settle on the ground (due to gravity) because of their high speeds and incessant collisions. In equilibrium, there is a very slight increase in density at lower heights (like in the atmosphere). The effect is small since the potential energy (mgh) for ordinary heights is much less than the average kinetic energy ½ mv^2 of the molecules.
5. $< v^2 >$ is not always equal to $(< v >)^2$. The average of a squared quantity is not necessarily the square of the average. Can you find examples for this statement.

EXERCISES

13.1 Estimate the fraction of molecular volume to the actual volume occupied by oxygen gas at STP. Take the diameter of an oxygen molecule to be 3 Å.

13.2 Molar volume is the volume occupied by 1 mol of any (ideal) gas at standard temperature and pressure (STP : 1 atmospheric pressure, 0 °C). Show that it is 22.4 litres.

13.3 Figure 13.8 shows plot of PV/T versus P for 1.00×10^{-3} kg of oxygen gas at two different temperatures.

Fig. 13.8

(a) What does the dotted plot signify?
(b) Which is true: $T_1 > T_2$ or $T_1 < T_2$?
(c) What is the value of PV/T where the curves meet on the y-axis?
(d) If we obtained similar plots for 1.00×10^{-3} kg of hydrogen, would we get the same value of PV/T at the point where the curves meet on the y-axis? If not, what mass of hydrogen yields the same value of PV/T (for low pressurehigh temperature region of the plot) ? (Molecular mass of H_2 = 2.02 u, of O_2 = 32.0 u, R = 8.31 J mol^{-1} K^{-1}.)

13.4 An oxygen cylinder of volume 30 litres has an initial gauge pressure of 15 atm and a temperature of 27 °C. After some oxygen is withdrawn from the cylinder, the gauge pressure drops to 11 atm and its temperature drops to 17 °C. Estimate the mass of oxygen taken out of the cylinder (R = 8.31 J mol^{-1} K^{-1}, molecular mass of O_2 = 32 u).

13.5 An air bubble of volume 1.0 cm^3 rises from the bottom of a lake 40 m deep at a temperature of 12 °C. To what volume does it grow when it reaches the surface, which is at a temperature of 35 °C ?

13.6 Estimate the total number of air molecules (inclusive of oxygen, nitrogen, water vapour and other constituents) in a room of capacity 25.0 m^3 at a temperature of 27 °C and 1 atm pressure.

13.7 Estimate the average thermal energy of a helium atom at (i) room temperature (27 °C), (ii) the temperature on the surface of the Sun (6000 K), (iii) the temperature of 10 million kelvin (the typical core temperature in the case of a star).

13.8 Three vessels of equal capacity have gases at the same temperature and pressure. The first vessel contains neon (monatomic), the second contains chlorine (diatomic), and the third contains uranium hexafluoride (polyatomic). Do the vessels contain equal number of respective molecules ? Is the root mean square speed of molecules the same in the three cases? If not, in which case is v_{rms} the largest ?

13.9 At what temperature is the root mean square speed of an atom in an argon gas cylinder equal to the rms speed of a helium gas atom at – 20 °C ? (atomic mass of Ar = 39.9 u, of He = 4.0 u).

13.10 Estimate the mean free path and collision frequency of a nitrogen molecule in a cylinder containing nitrogen at 2.0 atm and temperature 17 ^{0}C. Take the radius of a nitrogen molecule to be roughly 1.0 Å. Compare the collision time with the time the molecule moves freely between two successive collisions (Molecular mass of N_2 = 28.0 u).

Additional Exercises

13.11 A metre long narrow bore held horizontally (and closed at one end) contains a 76 cm long mercury thread, which traps a 15 cm column of air. What happens if the tube is held vertically with the open end at the bottom ?

13.12 From a certain apparatus, the diffusion rate of hydrogen has an average value of 28.7 $cm^3 s^{-1}$. The diffusion of another gas under the same conditions is measured to have an average rate of 7.2 $cm^3 s^{-1}$. Identify the gas.
[**Hint** : Use Graham's law of diffusion: $R_1/R_2 = (M_2/M_1)^{1/2}$, where R_1, R_2 are diffusion rates of gases 1 and 2, and M_1 and M_2 their respective molecular masses. The law is a simple consequence of kinetic theory.]

13.13 A gas in equilibrium has uniform density and pressure throughout its volume. This is strictly true only if there are no external influences. A gas column under gravity, for example, does not have uniform density (and pressure). As you might expect, its density decreases with height. The precise dependence is given by the so-called law of atmospheres

$$n_2 = n_1 \exp[-mg(h_2 - h_1)/k_B T]$$

where n_2, n_1 refer to number density at heights h_2 and h_1 respectively. Use this relation to derive the equation for sedimentation equilibrium of a suspension in a liquid column:

$$n_2 = n_1 \exp[-mg N_A (\rho - P')(h_2 - h_1)/(\rho RT)]$$

where ρ is the density of the suspended particle, and ρ' that of surrounding medium. [N_A is Avogadro's number, and R the universal gas constant.] [**Hint** : Use Archimedes principle to find the apparent weight of the suspended particle.]

13.14 Given below are densities of some solids and liquids. Give rough estimates of the size of their atoms :

Substance	Atomic Mass (u)	Density (10^3 Kg m^{-3})
Carbon (diamond)	12.01	2.22
Gold	197.00	19.32
Nitrogen (liquid)	14.01	1.00
Lithium	6.94	0.53
Fluorine (liquid)	19.00	1.14

[**Hint** : Assume the atoms to be 'tightly packed' in a solid or liquid phase, and use the known value of Avogadro's number. You should, however, not take the actual numbers you obtain for various atomic sizes too literally. Because of the crudeness of the tight packing approximation, the results only indicate that atomic sizes are in the range of a few Å].

CHAPTER FOURTEEN

OSCILLATIONS

14.1 Introduction
14.2 Periodic and oscilatory motions
14.3 Simple harmonic motion
14.4 Simple harmonic motion and uniform circular motion
14.5 Velocity and acceleration in simple harmonic motion
14.6 Force law for simple harmonic motion
14.7 Energy in simple harmonic motion
14.8 Some systems executing SHM
14.9 Damped simple harmonic motion
14.10 Forced oscillations and resonance

Summary
Points to ponder
Exercises
Additional Exercises
Appendix

14.1 INTRODUCTION

In our daily life we come across various kinds of motions. You have already learnt about some of them, e.g. rectilinear motion and motion of a projectile. Both these motions are non-repetitive. We have also learnt about uniform circular motion and orbital motion of planets in the solar system. In these cases, the motion is repeated after a certain interval of time, that is, it is periodic. In your childhood you must have enjoyed rocking in a cradle or swinging on a swing. Both these motions are repetitive in nature but different from the periodic motion of a planet. Here, the object moves to and fro about a mean position. The pendulum of a wall clock executes a similar motion. There are leaves and branches of a tree oscillating in breeze, boats bobbing at anchor and the surging pistons in the engines of cars. All these objects execute a periodic to and fro motion. Such a motion is termed as oscillatory motion. In this chapter we study this motion.

The study of oscillatory motion is basic to physics; its concepts are required for the understanding of many physical phenomena. In musical instruments like the sitar, the guitar or the violin, we come across vibrating strings that produce pleasing sounds. The membranes in drums and diaphragms in telephone and speaker systems vibrate to and fro about their mean positions. The vibrations of air molecules make the propagation of sound possible. Similarly, the atoms in a solid oscillate about their mean positions and convey the sensation of temperature. The oscillations of electrons in the antennas of radio, TV and satellite transmitters convey information.

The description of a periodic motion in general, and oscillatory motion in particular, requires some fundamental concepts like period, frequency, displacement, amplitude and phase. These concepts are developed in the next section.

14.2 PERIODIC AND OSCILLATORY MOTIONS

Fig 14.1 shows some periodic motions. Suppose an insect climbs up a ramp and falls down it comes back to the initial point and repeats the process identically. If you draw a graph of its height above the ground versus time, it would look something like Fig. 14.1 (a). If a child climbs up a step, comes down, and repeats the process, its height above the ground would look like that in Fig 14.1 (b). When you play the game of bouncing a ball off the ground, between your palm and the ground, its height versus time graph would look like the one in Fig 14.1 (c). Note that both the curved parts in Fig 14.1 (c) are sections of a parabola given by the Newton's equation of motion (see section 3.6),

$h = ut + \frac{1}{2}gt^2$ for downward motion, and

$h = ut - \frac{1}{2}gt^2$ for upward motion,

with different values of u in each case. These are examples of periodic motion. Thus, a motion that repeats itself at regular intervals of time is called **periodic motion.**

Fig 14.1 Examples of periodic motion. The period T is shown in each case.

Very often the body undergoing periodic motion has an equilibrium position somewhere inside its path. When the body is at this position no net external force acts on it. Therefore, if it is left there at rest, it remains there forever. If the body is given a small displacement from the position, a force comes into play which tries to bring the body back to the equilibrium point, giving rise to **oscillations** or **vibrations**. For example, a ball placed in a bowl will be in equilibrium at the bottom. If displaced a little from the point, it will perform oscillations in the bowl. Every oscillatory motion is periodic, but every periodic motion need not be oscillatory. Circular motion is a periodic motion, but it is not oscillatory.

There is no significant difference between oscillations and vibrations. It seems that when the frequency is small, we call it oscillation (like the oscillation of a branch of a tree), while when the frequency is high, we call it vibration (like the vibration of a string of a musical instrument).

Simple harmonic motion is the simplest form of oscillatory motion. This motion arises when the force on the oscillating body is directly proportional to its displacement from the mean position, which is also the equilibrium position. Further, at any point in its oscillation, this force is directed towards the mean position.

In practice, oscillating bodies eventually come to rest at their equilibrium positions, because of the damping due to friction and other dissipative causes. However, they can be forced to remain oscillating by means of some external periodic agency. We discuss the phenomena of damped and forced oscillations later in the chapter.

Any material medium can be pictured as a collection of a large number of coupled oscillators. The collective oscillations of the constituents of a medium manifest themselves as waves. Examples of waves include water waves, seismic waves, electromagnetic waves. We shall study the wave phenomenon in the next chapter.

14.2.1 Period and frequency

We have seen that any motion that repeats itself at regular intervals of time is called **periodic motion**. **The smallest interval of time after which the motion is repeated is called its period**. Let us denote the period by the symbol T. Its SI unit is second. For periodic motions,

which are either too fast or too slow on the scale of seconds, other convenient units of time are used. The period of vibrations of a quartz crystal is expressed in units of microseconds (10^{-6} s) abbreviated as μs. On the other hand, the orbital period of the planet Mercury is 88 earth days. The Halley's comet appears after every 76 years.

The reciprocal of T gives the number of repetitions that occur per unit time. This quantity is called the **frequency of the periodic motion**. It is represented by the symbol ν . The relation between ν and T is

$$\nu = 1/T \qquad (14.1)$$

The unit of ν is thus s^{-1}. After the discoverer of radio waves, Heinrich Rudolph Hertz (1857-1894), a special name has been given to the unit of frequency. It is called hertz (abbreviated as Hz). Thus,

$$1 \text{ hertz} = 1 \text{ Hz} = 1 \text{ oscillation per second} = 1 s^{-1} \qquad (14.2)$$

Note, that the frequency, ν, is not necessarily an integer.

Example 14.1 On an average a human heart is found to beat 75 times in a minute. Calculate its frequency and period.

Answer The beat frequency of heart = 75/(1 min)
= 75/(60 s)
= 1.25 s^{-1}
= 1.25 Hz
The time period T = 1/(1.25 s^{-1})
= 0.8 s

14.2.2 Displacement

In section 4.2, we defined displacement of a particle as the change in its position vector. In this chapter, we use the term displacement in a more general sense. It refers to change with time of any physical property under consideration. For example, in case of rectilinear motion of a steel ball on a surface, the distance from the starting point as a function of time is its position displacement. The choice of origin is a matter of convenience. Consider a block attached to a spring, the other end of which is fixed to a rigid wall [see Fig.14.2(a)]. Generally it is convenient to measure displacement of the body from its equilibrium position. For an oscillating simple pendulum, the angle from the vertical as a function of time may be regarded as a displacement variable [see Fig.14.2(b)]. The term displacement is not always to be referred in the context of position only. There can be many other kinds of displacement variables. The voltage across a capacitor, changing with time in an a.c. circuit, is also a displacement variable. In the same way, pressure variations in time in the propagation of sound wave, the changing electric and magnetic fields in a light wave are examples of displacement in different contexts. The displacement variable may take both positive and negative values. In experiments on oscillations, the displacement is measured for different times.

Fig. 14.2(a) A block attached to a spring, the other end of which is fixed to a rigid wall. The block moves on a frictionless surface. The motion of the block can be described in terms of its distance or displacement x from the wall.

Fig.14.2(b) An oscillating simple pendulum; its motion can be described in terms of angular displacement θ from the vertical.

The displacement can be represented by a mathematical function of time. In case of periodic motion, this function is periodic in time. One of the simplest periodic functions is given by

$$f(t) = A \cos \omega t \qquad (14.3a)$$

If the argument of this function, ωt, is increased by an integral multiple of 2π radians,

the value of the function remains the same. The function $f(t)$ is then periodic and its period, T, is given by

$$T = \frac{2\pi}{\omega} \qquad (14.3b)$$

Thus, the function $f(t)$ is periodic with period T,

$$f(t) = f(t+T)$$

The same result is obviously correct if we consider a sine function, $f(t) = A \sin \omega t$. Further, a linear combination of sine and cosine functions like,

$$f(t) = A \sin \omega t + B \cos \omega t \qquad (14.3c)$$

is also a periodic function with the same period T. Taking,

$$A = D \cos \phi \text{ and } B = D \sin \phi$$

Eq. (14.3c) can be written as,

$$f(t) = D \sin (\omega t + \phi) , \qquad (14.3d)$$

Here D and ϕ are constant given by

$$D = \sqrt{A^2 + B^2} \text{ and } \tan^{-1} \frac{B}{A}$$

The great importance of periodic sine and cosine functions is due to a remarkable result proved by the French mathematician, Jean Baptiste Joseph Fourier (1768-1830): **Any periodic function can be expressed as a superposition of sine and cosine functions of different time periods with suitable coefficients**.

Example 14.2 Which of the following functions of time represent (a) periodic and (b) non-periodic motion? Give the period for each case of periodic motion [ω is any positive constant].

(i) $\sin \omega t + \cos \omega t$

(ii) $\sin \omega t + \cos 2\omega t + \sin 4\omega t$

(iii) $e^{-\omega t}$

(iv) $\log (\omega t)$

Answer

(i) $\sin \omega t + \cos \omega t$ is a periodic function, it can also be written as $\sqrt{2} \sin (\omega t + \pi/4)$.

Now $\sqrt{2} \sin (\omega t + \pi/4) = \sqrt{2} \sin (\omega t + \pi/4 + 2\pi)$

$= \sqrt{2} \sin [\omega(t + 2\pi/\omega) + \pi/4]$

The periodic time of the function is $2\pi/\omega$

(ii) This is an example of a periodic motion. It can be noted that each term represents a periodic function with a different angular frequency. Since period is the least interval of time after which a function repeats its value, $\sin \omega t$ has a period $T_0 = 2\pi/\omega$; $\cos 2\omega t$ has a period $\pi/\omega = T_0/2$; and $\sin 4\omega t$ has a period $2\pi/4\omega = T_0/4$. The period of the first term is a multiple of the periods of the last two terms. Therefore, the smallest interval of time after which the sum of the three terms repeats is T_0, and thus the sum is a periodic function with a period $2\pi/\omega$

(iii) The function $e^{-\omega t}$ is not periodic, it decreases monotonically with increasing time and tends to zero as $t \to \infty$ and thus, never repeats its value.

(iv) The function $\log(\omega t)$ increases monotonically with time t. It, therefore, never repeats its value and is a non-periodic function. It may be noted that as $t \to \infty$, $\log(\omega t)$ diverges to ∞. It, therefore, cannot represent any kind of physical displacement. ◀

14.3 SIMPLE HARMONIC MOTION

Let us consider a particle vibrating back and forth about the origin of an x-axis between the limits $+A$ and $-A$ as shown in Fig. 14.3. In between these extreme positions the particle

Fig. 14.3 A particle vibrating back and forth about the origin of x-axis, between the limits +A and –A.

moves in such a manner that its speed is maximum when it is at the origin and zero when it is at $\pm A$. The time t is chosen to be zero when the particle is at $+A$ and it returns to $+A$ at $t = T$. In this section we will describe this motion. Later, we shall discuss how to achieve it. To study the motion of this particle, we record its positions as a function of time

by taking 'snapshots' at regular intervals of time. A set of such snapshots is shown in Fig. 14.4. The position of the particle with reference to the origin gives its displacement at any instant of time. For such a motion the displacement $x(t)$ of the particle from a certain chosen origin is found to vary with time as,

Fig. 14.4 *A sequence of 'snapshots' (taken at equal intervals of time) showing the position of a particle as it oscillates back and forth about the origin along an x-axis, between the limits +A and –A. The length of the vector arrows is scaled to indicate the speed of the particle. The speed is maximum when the particle is at the origin and zero when it is at ± A. If the time t is chosen to be zero when the particle is at +A, then the particle returns to +A at t = T, where T is the period of the motion. The motion is then repeated. It is represented by Eq. (14.4) for $\phi = 0$.*

Fig. 14.5 *A graph of x as a function of time for the motion represented by Eq. (14.4).*

$$x(t) = A\cos(\omega t + \phi) \qquad (14.4)$$

in which A, ω, and ϕ are constants.

$$\underset{\text{Displacement}}{x(t)} = \underset{\text{Amplitude}}{A} \cos(\overbrace{\underset{\text{Angular frequency}}{\omega t} + \underset{\text{Phase constant}}{\phi}}^{\text{Phase}})$$

Fig. 14.6 *A reference of the quantities in Eq. (14.4).*

The motion represented by Eq. (14.4) is called **simple harmonic motion** (SHM); a term that means the periodic motion is a sinusoidal function of time. Equation (14.4), in which the sinusoidal function is a cosine function, is plotted in Fig. 14.5. The quantities that determine

Fig. 14.7 (a) *A plot of displacement as a function of time as obtained from Eq. (14.4) with $\phi = 0$. The curves 1 and 2 are for two different amplitudes A and B.*

the shape of the graph are displayed in Fig. 14.6 along with their names.

We shall now define these quantities.

The quantity A is called the **amplitude** of the motion. It is a positive constant which represents the magnitude of the maximum displacement of the particle in either direction. The cosine function in Eq. (14.4) varies between the limits ±1, so the displacement $x(t)$ varies between the limits ± A. In Fig. 14.7 (a), the curves 1 and 2 are plots of Eq. (14.4) for two different amplitudes A and B. The difference between these curves illustrates the significance of amplitude.

The time varying quantity, $(\omega t + \phi)$, in Eq. (14.4) is called the **phase** of the motion. It describes the state of motion at a given time. The constant ϕ is called the **phase constant** (or **phase angle**). The value of ϕ depends on the displacement and velocity of the particle at $t = 0$. This can be understood better by considering Fig. 14.7(b). In this figure, the curves 3 and 4 represent plots of Eq. (14.4) for two values of the phase constant ϕ.

It can be seen that the phase constant signifies the initial conditions.

The constant ω, called the angular frequency of the motion, is related to the period T. To get

Fig. 14.7 (b) A plot obtained from Eq. 14.4. The curves 3 and 4 are for ϕ = 0 and -π/4 respectively. The amplitude A is same for both the plots.

their relationship, let us consider Eq. (14.4) with ϕ = 0; it then reduces to,

$$x(t) = A \cos \omega t \qquad (14.5)$$

Now since the motion is periodic with a period T, the displacement $x(t)$ must return to its initial value after one period of the motion; that is, **$x(t)$ must be equal to $x(t+T)$** for all t. Applying this condition to Eq. (14.5) leads to,

$$A \cos \omega t = A \cos \omega(t + T) \qquad (14.6)$$

As the cosine function first repeats itself when its argument (the phase) has increased by 2π, Eq. (14.6) gives,

$$\omega(t + T) = \omega t + 2\pi$$

or $$\omega T = 2\pi$$

Thus, the angular frequency is,

$$\omega = 2\pi / T \qquad (14.7)$$

The SI unit of angular frequency is radians per second. To illustrate the significance of period T, sinusoidal functions with two different periods are plotted in Fig. 14.8.

Fig. 14.8 Plots of Eq. (14.4) for ϕ = 0 for two different periods.

In this plot the SHM represented by curve a, has a period T and that represented by curve b, has a period $T' = T/2$.

We have had an introduction to simple harmonic motion. In the next section we will discuss the simplest example of simple harmonic motion. It will be shown that the projection of uniform circular motion on a diameter of the circle executes simple harmonic motion.

Example 14.3 Which of the following functions of time represent (a) simple harmonic motion and (b) periodic but not simple harmonic? Give the period for each case.
(1) $\sin \omega t - \cos \omega t$
(2) $\sin^2 \omega t$

Answer

(a) $\sin \omega t - \cos \omega t$

$= \sin \omega t - \sin (\pi/2 - \omega t)$

$= 2 \cos (\pi/4) \sin (\omega t - \pi/4)$

$= \sqrt{2} \sin (\omega t - \pi/4)$

This function represents a simple harmonic motion having a period $T = 2\pi/\omega$ and a phase angle $(-\pi/4)$ or $(7\pi/4)$

(b) $\sin^2 \omega t$

$= ½ - ½ \cos 2\omega t$

The function is periodic having a period $T = \pi/\omega$ It also represents a harmonic motion with the point of equilibrium occurring at ½ instead of zero.

14.4 SIMPLE HARMONIC MOTION AND UNIFORM CIRCULAR MOTION

In 1610, Galileo discovered four principal moons of the planet Jupiter. To him, each moon seemed to move back and forth relative to the planet in a simple harmonic motion; the disc of the planet forming the mid point of the motion. The record of his observations, written in his own hand, is still available. Based on his data, the position of the moon Callisto relative to Jupiter is plotted in Fig. 14.9. In this figure, the circles represent Galileo's data points and the curve drawn is a best fit to the data. The curve obeys Eq. (14.4), which is the displacement function for SHM. It gives a period of about 16.8 days.

It is now well known that Callisto moves with essentially a constant speed in an almost circular orbit around Jupiter. Its true motion is uniform circular motion. What Galileo saw and what we can also see, with a good pair of binoculars, is the projection of this uniform circular motion on a line in the plane of motion. This can easily be visualised by performing a

Fig. 14.9 The angle between Jupiter and its moon Callisto as seen from earth. The circles are based on Galileo's measurements of 1610. The curve is a best fit suggesting a simple harmonic motion. At Jupiter's mean distance, 10 minutes of arc corresponds to about 2×10^6 km.

simple experiment. Tie a ball to the end of a string and make it move in a horizontal plane about a fixed point with a constant angular speed. The ball would then perform a uniform circular motion in the horizontal plane. Observe the ball sideways or from the front, fixing your attention in the plane of motion. The ball will appear to execute to and fro motion along a horizontal line with the point of rotation as the midpoint. You could alternatively observe the shadow of the ball on a wall which is perpendicular to the plane of the circle. In this process what we are observing is the motion of the ball on a diameter of the circle normal to the direction of viewing. This experiment provides an analogy to Galileo's observation.

In Fig. 14.10, we show the motion of a **reference particle** P executing a uniform circular motion with (constant) angular speed ω in a **reference circle**. The radius A of the circle is the magnitude of the particle's position vector. At any time t, the angular position of the particle

Fig. 14.10 The motion of a reference particle P executing a uniform circular motion with (constant) angular speed ω in a reference circle of radius A.

is $\omega t + \phi$, where ϕ is its angular position at $t = 0$. The projection of particle P on the x-axis is a point P′, which we can take as a second particle. The projection of the position vector of particle P on the x-axis gives the location $x(t)$ of P′. Thus we have,

$$x(t) = A \cos(\omega t + \phi)$$

which is the same as Eq. (14.4). This shows that if the reference particle P moves in a uniform circular motion, its projection particle P′ executes a simple harmonic motion along a diameter of the circle.

From Galileo's observation and the above considerations, we are led to the conclusion that circular motion viewed edge-on is simple harmonic motion. In a more formal language we can say that : **Simple harmonic motion is the projection of uniform circular motion on a diameter of the circle in which the latter motion takes place**.

Example 14.4 Fig. 14.11 depicts two circular motions. The radius of the circle, the period of revolution, the initial position and the sense of revolution are indicated on the figures. Obtain the simple harmonic motions of the x-projection of the radius vector of the rotating particle P in each case.

Fig. 14.11

Answer

(a) At $t = 0$, OP makes an angle of $45° = \pi/4$ rad with the (positive direction of) x-axis. After time t, it covers an angle $\frac{2\pi}{T}t$ in the anticlockwise sense, and makes an angle of $\frac{2\pi}{T}t + \frac{\pi}{4}$ with the x-axis.

The projection of OP on the x-axis at time t is given by,

$$x(t) = A\cos\left(\frac{2}{T}t + \frac{}{4}\right)$$

For $T = 4$ s,

$$x(t) = A\cos\ \frac{2}{4}t + \frac{}{4}$$

which is a SHM of amplitude A, period 4 s, and an initial phase* $= \frac{}{4}$.

(b) In this case at $t = 0$, OP makes an angle of $90^\circ = \frac{}{2}$ with the x-axis. After a time t, it covers an angle of $\frac{2}{T}t$ in the clockwise sense and makes an angle of $\frac{}{2} - \frac{2}{T}t$ with the x-axis. The projection of OP on the x-axis at time t is given by

$$x(t) = B\cos\ \frac{}{2} - \frac{2}{T}t$$

$$= B\sin\left(\frac{2}{T}t\right)$$

For $T = 30$ s,

$$x(t) = B\sin\ \frac{}{15}t$$

Writing this as $x(t) = B\cos\ \frac{}{15}t - \frac{}{2}$, and comparing with Eq. (14.4). We find that this represents a SHM of amplitude B, period 30 s, and an initial phase of $\frac{}{2}$. ◀

14.5 VELOCITY AND ACCELERATION IN SIMPLE HARMONIC MOTION

It can be seen easily that the magnitude of velocity, **v**, with which the reference particle P (Fig. 14.10) is moving in a circle is related to its angular speed, ω, as

$$v = \omega A \tag{14.8}$$

where A is the radius of the circle described by the particle P. The magnitude of the velocity vector **v** of the projection particle is ωA; its projection on the x-axis at any time t, as shown in Fig. 14.12, is

$$v(t) = -\omega A\sin(\omega t + \phi) \tag{14.9}$$

Fig. 14.12 *The velocity, v (t), of the particle P′ is the projection of the velocity v of the reference particle, P.*

The negative sign appears because the velocity component of P is directed towards the left, in the negative direction of x. Equation (14.9) expresses the instantaneous velocity of the particle P′ (projection of P). Therefore, **it expresses the instantaneous velocity of a particle executing SHM.** Equation (14.9) can also be obtained by differentiating Eq. (14.4) with respect to time as,

$$v(t) = \frac{d}{dt}x(t) \tag{14.10}$$

* *The natural unit of angle is radian, defined through the ratio of arc to radius. Angle is a dimensionless quantity. Therefore it is not always necessary to mention the unit 'radian' when we use π, its multiples or submultiples. The conversion between radian and degree is not similar to that between metre and centimetre or mile. If the argument of a trigonometric function is stated without units, it is understood that the unit is radian. On the other hand, if degree is to be used as the unit of angle, then it must be shown explicitly. For example, sin(15⁰) means sine of 15 degree, but sin(15) means sine of 15 radians. Hereafter, we will often drop 'rad' as the unit, and it should be understood that whenever angle is mentioned as a numerical value, without units, it is to be taken as radians.*

Fig. 14.13 *The acceleration, a(t), of the particle P′ is the projection of the acceleration* **a** *of the reference particle P.*

We have seen that a particle executing a uniform circular motion is subjected to a radial acceleration **a** directed towards the centre. Figure 14.13 shows such a radial acceleration, **a**, of the reference particle P executing uniform circular motion. The magnitude of the radial acceleration of P is $\omega^2 A$. Its projection on the x-axis at any time t is,

$$a(t) = -\omega^2 A \cos(\omega t + \phi)$$
$$= -\omega^2 x(t) \qquad (14.11)$$

which is the acceleration of the particle P′ (the projection of particle P). Equation (14.11), therefore, represents the instantaneous acceleration of the particle P′, which is executing **SHM**. Thus Eq. (14.11) **expresses the acceleration of a particle executing SHM**. It is an important result for *SHM*. It shows that in **SHM**, **the acceleration is proportional to the displacement and is always directed towards the mean position**. Eq. (14.11) can also be obtained by differentiating Eq. (14.9) with respect to time as,

$$a(t) = \frac{d}{dt} v(t) \qquad (14.12)$$

The inter-relationship between the displacement of a particle executing simple harmonic motion, its velocity and acceleration can be seen in Fig. 14.14. In this figure (a) is a plot of Eq. (14.4) with $\phi = 0$ and (b) depicts Eq. (14.9) also with $\phi = 0$. Similar to the amplitude A in Eq. (14.4), the positive quantity ωA in Eq. (14.9) is called the **velocity amplitude** v_m. In Fig. 14.14(b), it can be seen that the velocity of the

Fig. 14.14 *The particle displacement, velocity and acceleration in a simple harmonic motion. (a) The displacement x (t) of a particle executing SHM with phase angle ϕ equal to zero. (b) The velocity v (t) of the particle. (c) The acceleration a (t) of the particle.*

oscillating particle varies between the limits $\pm v_m = \pm \omega A$. Note that the curve of $v(t)$ is shifted (to the left) from the curve of $x(t)$ by one quarter period and thus the particle velocity lags behind the displacement by a phase angle of $\pi/2$; when the magnitude of displacement is the greatest, the magnitude of the velocity is the least. When the magnitude of displacement is the least, the velocity is the greatest. Figure14.14(c) depicts the variation of the particle acceleration $a(t)$. It is seen that when the displacement has its greatest positive value, the acceleration has its greatest negative value and vice versa. When the displacement is zero, the acceleration is also zero.

Example 14.5 A body oscillates with SHM according to the equation (in SI units),

$$x = 5 \cos [2\pi t + \pi/4].$$

At $t = 1.5$ s, calculate the (a) displacement, (b) speed and (c) acceleration of the body.

Answer The angular frequency ω of the body $= 2\pi\ \text{s}^{-1}$ and its time period $T = 1$ s.

At $t = 1.5$ s

(a) displacement = (5.0 m) cos $[(2\pi\ \text{s}^{-1}) \times 1.5\ \text{s} + \pi/4]$

$= (5.0 \text{ m}) \cos [(3\pi + \pi/4)]$
$= -5.0 \times 0.707 \text{ m}$
$= -3.535 \text{ m}$

(b) Using Eq. (14.9), the speed of the body
$= -(5.0 \text{ m})(2\pi \text{ s}^{-1}) \sin [(2\pi \text{ s}^{-1}) \times 1.5 \text{ s} + \pi/4]$
$= -(5.0 \text{ m})(2\pi \text{ s}^{-1}) \sin [(3\pi + \pi/4)]$
$= 10\pi \times 0.707 \text{ m s}^{-1}$
$= 22 \text{ m s}^{-1}$

(c) Using Eq.(14.10), the acceleration of the body
$= -(2\pi \text{ s}^{-1})^2 \times \text{displacement}$
$= -(2\pi \text{ s}^{-1})^2 \times (-3.535 \text{ m})$
$= 140 \text{ m s}^{-2}$ ◀

14.6 FORCE LAW FOR SIMPLE HARMONIC MOTION

In Section14.3, we described the simple harmonic motion. Now we discuss how it can be generated. Newton's second law of motion relates the force acting on a system and the corresponding acceleration produced. Therefore, if we know how the acceleration of a particle varies with time, this law can be used to learn about the force, which must act on the particle to give it that acceleration. If we combine Newton's second law and Eq. (14.11), we find that for simple harmonic motion,

$$F(t) = ma$$
$$= -m\omega^2 x(t)$$

or $$F(t) = -k\,x(t) \qquad (14.13)$$

where $$k = m\omega^2 \qquad (14.14a)$$

or

$$\omega = \sqrt{\frac{k}{m}} \qquad (14.14b)$$

Equation (14.13) gives the force acting on the particle. It is proportional to the displacement and directed in an opposite direction. Therefore, it is a restoring force. Note that unlike the centripetal force for uniform circular motion that is constant in magnitude, the restoring force for SHM is time dependent. The force law expressed by Eq. (14.13) can be taken as an alternative definition of simple harmonic motion. It states : **Simple harmonic motion is the motion executed by a particle subject to a force, which is proportional to the displacement of the particle and is directed towards the mean position.**

Since the force F is proportional to x rather than to some other power of x, such a system is also referred to as a **linear harmonic oscillator**. Systems in which the restoring force is a non-linear function of x are termed as non-linear harmonic or anharmonic oscillators.

▶ ***Example 14.6*** Two identical springs of spring constant k are attached to a block of mass m and to fixed supports as shown in Fig. 14.15. Show that when the mass is displaced from its equilibrium position on either side, it executes a simple harmonic motion. Find the period of oscillations.

Fig. 14.15

Answer Let the mass be displaced by a small distance x to the right side of the equilibrium position, as shown in Fig. 14.16. Under this situation the spring on the left side gets

Fig. 14.16

elongated by a length equal to x and that on the right side gets compressed by the same length. The forces acting on the mass are then,

$F_1 = -k\,x$ (force exerted by the spring on the left side, trying to pull the mass towards the mean position)

$F_2 = -k\,x$ (force exerted by the spring on the right side, trying to push the mass towards the mean position)

The net force, F, acting on the mass is then given by,

$$F = -2kx$$

Hence the force acting on the mass is proportional to the displacement and is directed towards the mean position; therefore, the motion executed by the mass is simple harmonic. The time period of oscillations is,

$$T = 2\pi \sqrt{\frac{m}{2k}}$$

14.7 ENERGY IN SIMPLE HARMONIC MOTION

A particle executing simple harmonic motion has kinetic and potential energies, both varying between the limits, zero and maximum.

In section14.5 we have seen that the velocity of a particle executing SHM, is a periodic function of time. It is zero at the extreme positions of displacement. Therefore, the kinetic energy (K) of such a particle, which is defined as

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

$$= \frac{1}{2} m\, \omega^2 A^2 \sin^2(\omega t + \phi)$$

$$= \frac{1}{2} k\, A^2 \sin^2(\omega t + \phi) \qquad (14.15)$$

is also a periodic function of time, being zero when the displacement is maximum and maximum when the particle is at the mean position. Note, since the sign of v is immaterial in K, the period of K is $T/2$.

What is the potential energy (PE) of a particle executing simple harmonic motion ? In Chapter 6, we have seen that the concept of potential energy is possible only for conservative forces. The spring force $F = -kx$ is a conservative force, with associated potential energy

$$U = \frac{1}{2} k\, x^2 \qquad 14.16)$$

Hence the potential energy of a particle executing simple harmonic motion is,

$$U(x) = \frac{1}{2} k\, x^2$$

$$= \frac{1}{2} k\, A^2 \cos^2(\omega t + \phi) \qquad (14.17)$$

Thus the potential energy of a particle executing simple harmonic motion is also periodic, with period $T/2$, being zero at the mean position and maximum at the extreme displacements.

It follows from Eqs. (14.15) and (14.17) that the total energy, E, of the system is,

$$E = U + K$$

$$= \frac{1}{2} k\, A^2 \cos^2(\omega t + \phi) + \frac{1}{2} k\, A^2 \sin^2(\omega t + \phi)$$

$$= \frac{1}{2} k\, A^2 \left[\cos^2(\omega t + \phi) + \sin^2(\omega t + \phi)\right]$$

The quantity within the square brackets above is unity and we have,

$$E = \frac{1}{2} k\, A^2 \qquad (14.18)$$

The total mechanical energy of a harmonic oscillator is thus independent of time as expected for motion under any conservative force. The time and displacement dependence of the potential and kinetic energies of a linear simple harmonic oscillator are shown in Fig. 14.17.

It is observed that in a linear harmonic oscillator, all energies are positive and peak twice during every period. For $x = 0$, the energy is all kinetic and for $x = \pm A$ it is all potential.

In between these extreme positions, the potential energy increases at the expense of kinetic energy. This behaviour of a linear harmonic oscillator suggests that it possesses an element of springiness and an element of inertia. The former stores its potential energy and the latter stores its kinetic energy.

Fig. 14.17 *(a) Potential energy U(t), kinetic energy K(t) and the total energy E as functions of time t for a linear harmonic oscillator. All energies are positive and the potential and kinetic energies peak twice in every period of the oscillator. (b) Potential energy U(x), kinetic energy K(x) and the total energy E as functions of position x for a linear harmonic oscillator with amplitude A. For x = 0, the energy is all kinetic and for x = ± A it is all potential.*

Example 14.7 A block whose mass is 1 kg is fastened to a spring. The spring has a spring constant of 50 N m^{-1}. The block is pulled to a distance x = 10 cm from its equilibrium position at x= 0 on a frictionless surface from rest at t = 0. Calculate the kinetic, potential and total energies of the block when it is 5 cm away from the mean position.

Answer The block executes SHM, its angular frequency, as given by Eq. (14.14b), is

$$\omega = \sqrt{\frac{k}{m}}$$

$$= \sqrt{\frac{50 \text{ N m}^{-1}}{1\text{kg}}}$$

$$= 7.07 \text{ rad s}^{-1}$$

Its displacement at any time t is then given by,

$$x(t) = 0.1 \cos (7.07t)$$

Therefore, when the particle is 5 cm away from the mean position, we have

$$0.05 = 0.1 \cos (7.07t)$$

Or $\cos (7.07t) = 0.5$ and hence

$$\sin (7.07t) = \frac{\sqrt{3}}{2} = 0.866,$$

Then the velocity of the block at x = 5 cm is

$$= 0.1 \times 7.07 \times 0.866 \text{ m s}^{-1}$$

$$= 0.61 \text{ m s}^{-1}$$

Hence the K.E. of the block,

$$= \frac{1}{2} m v^2$$

$$= ½[1\text{kg} \times (0.6123 \text{ m s}^{-1})^2]$$

$$= 0.19 \text{ J}$$

The P.E. of the block,

$$= \frac{1}{2} k x^2$$

$$= ½(50 \text{ N m}^{-1} \times 0.05 \text{ m} \times 0.05 \text{ m})$$

$$= 0.0625 \text{ J}$$

The total energy of the block at x = 5 cm,

$$= \text{K.E.} + \text{P.E.}$$

$$= 0.25 \text{ J}$$

we also know that at maximum displacement, K.E. is zero and hence the total energy of the system is equal to the P.E. Therefore, the total energy of the system,

$$= ½(50 \text{ N m}^{-1} \times 0.1 \text{ m} \times 0.1 \text{ m})$$

$$= 0.25 \text{ J}$$

which is same as the sum of the two energies at a displacement of 5 cm. This is in conformity with the principle of conservation of energy.

14.8 SOME SYSTEMS EXECUTING SIMPLE HARMONIC MOTION

There are no physical examples of absolutely pure **simple harmonic motion**. In practice we come across systems that execute simple harmonic motion approximately under certain conditions. In the subsequent part of this section, we discuss the motion executed by some such systems.

Fig. 14.18 A linear simple harmonic oscillator consisting of a block of mass m attached to a spring. The block moves over a frictionless surface. Once pulled to the side and released, it executes simple harmonic motion.

14.8.1 Oscillations due to a Spring

The simplest observable example of simple harmonic motion is the small oscillations of a block of mass m fixed to a spring, which in turn is fixed to a rigid wall as shown in Fig. 14.18. The block is placed on a frictionless horizontal surface. If the block is pulled on one side and is released, it then executes a to and fro motion about a mean position. Let $x = 0$, indicate the position of the centre of the block when the spring is in equilibrium. The positions marked as $-A$ and $+A$ indicate the maximum displacements to the left and the right of the mean position. We have already learnt that springs have special properties, which were first discovered by the English physicist Robert Hooke. He had shown that such a system when deformed, is subject to a restoring force, the magnitude of which is proportional to the deformation or the displacement and acts in opposite direction. This is known as Hooke's law (Chapter 9). It holds good for displacements small in comparison to the length of the spring. At any time t, if the displacement of the block from its mean position is x, the restoring force F acting on the block is,

$$F(x) = -k\,x \qquad (14.19)$$

The constant of proportionality, k, is called the spring constant, its value is governed by the elastic properties of the spring. A stiff spring has large k and a soft spring has small k. Equation (14.19) is same as the force law for SHM and therefore the system executes a simple harmonic motion. From Eq. (14.14) we have,

$$\omega = \sqrt{\frac{k}{m}} \qquad (14.20)$$

and the period, T, of the oscillator is given by,

$$T = 2\sqrt{\frac{m}{k}} \qquad (14.21)$$

Equations (14.20) and (14.21) tell us that a large angular frequency and hence a small period is associated with a stiff spring (high k) and a light block (small m).

Example 14.8 A 5 kg collar is attached to a spring of spring constant 500 N m^{-1}. It slides without friction over a horizontal rod. The collar is displaced from its equilibrium position by 10.0 cm and released. Calculate
(a) the period of oscillation,
(b) the maximum speed and
(c) maximum acceleration of the collar.

Answer (a) The period of oscillation as given by Eq. (14.21) is,

$$T = 2\sqrt{\frac{m}{k}}$$

$$= 2\pi\sqrt{\frac{5.0\,\text{kg}}{500\,\text{N m}^{-1}}}$$

$$= (2\pi/10)\ \text{s}$$

$$= 0.63\ \text{s}$$

(b) The velocity of the collar executing SHM is given by,

$$v(t) = -A\omega\sin(\omega t + \phi)$$

The maximum speed is given by,

$$v_m = A\omega$$

$$= 0.1 \times \sqrt{\frac{k}{m}}$$

$$= 0.1 \times \sqrt{\frac{500\,\text{N m}^{-1}}{5\,\text{kg}}}$$

$$= 1\ \text{m s}^{-1}$$

and it occurs at $x = 0$

(c) The acceleration of the collar at the displacement $x(t)$ from the equilibrium is given by,

$$a(t) = -\omega^2 x(t)$$

$$= -\frac{k}{m}\,x(t)$$

Therefore the maximum acceleration is,

$$a_{max} = \omega^2 A$$

$$= \frac{500 \text{ N m}^{-1}}{5 \text{ kg}} \times 0.1 \text{ m}$$

$$= 10 \text{ m s}^{-2}$$

and it occurs at the extremities. ◀

14.8.2 The Simple Pendulum

It is said that Galileo measured the periods of a swinging chandelier in a church by his pulse beats. He observed that the motion of the chandelier was periodic. The system is a kind of pendulum. You can also make your own pendulum by tying a piece of stone to a long unstretchable thread, approximately 100 cm long. Suspend your pendulum from a suitable support so that it is free to oscillate. Displace the stone to one side by a small distance and let it go. The stone executes a to and fro motion, it is periodic with a period of about two seconds. Is this motion simple harmonic ? To answer this question, we consider a **simple pendulum**, which consists of a particle of mass m (called the bob of the pendulum) suspended from one end of an unstretchable, massless string of length L fixed at the other end as shown in Fig. 14.19(a). The bob is free to swing to and fro in the plane of the page, to the left and right of a vertical line through the pivot point.

The forces acting on the bob are the force **T**, tension in the string and the gravitational force $\mathbf{F}_g$ (= $m\,\mathbf{g}$), as shown in Fig. 14.19(b). The string makes an angle θ with the vertical. We resolve the force $\mathbf{F}_g$ into a radial component $F_g \cos\theta$ and a tangential component $F_g \sin\theta$. The radial component is cancelled by the tension, since there is no motion along the length of the string. The tangential component produces a restoring torque about the pendulum's pivot point. This

(a)

(b)

Fig. 14.19 (a) A simple pendulum. (b) The forces acting on the bob are the force due to gravity, F_g (= m g), and the tension T in the string. (b) The tangential component F_g of the gravitational force is a restoring force that tends to bring the pendulum back to the central position.

torque always acts opposite to the displacement of the bob so as to bring it back towards its central location. The central location is called the **equilibrium position** ($\theta = 0$), because at this position the pendulum would be at rest if it were not swinging.

The restoring torque τ is given by,

$$\tau = -L\,(F_g \sin\theta) \qquad (14.22)$$

where the negative sign indicates that the torque acts to reduce θ, and L is the length of the moment arm of the force $F_g \sin\theta$ about the pivot point. For rotational motion we have,

$$\tau = I\,\alpha \qquad (14.23)$$

where I is the pendulum's rotational inertia about the pivot point and α is its angular acceleration about that point. From Eqs. (14.22) and (14.23) we have,

$$-L\,(F_g \sin\theta) = I\,\alpha \qquad (14.24)$$

Substituting the magnitude of F_g, i.e. mg, we have,

$$-L\,m\,g \sin\theta = I\,\alpha$$

Or,

$$\alpha = \frac{m\,g\,L}{I} \sin \qquad (14.25)$$

We can simplify Eq. (14.25) if we assume that the displacement θ is small. We know that $\sin\theta$ can be expressed as,

$$\sin \qquad \frac{}{3!} \quad \frac{}{5!} \pm \ldots \qquad (14.26)$$

where θ is in radians.

Now if θ is small, sin θ can be approximated by θ and Eq. (14.25) can then be written as,

$$\alpha = -\frac{mgL}{I}\theta \qquad (14.27)$$

In Table 14.1, we have listed the angle θ in degrees, its equivalent in radians, and the value of the function sin θ. From this table it can be seen that for θ as large as 20 degrees, sin θ is nearly the same as θ **expressed in radians**.

Table 14.1 sin θ as a function of angle θ

θ (degrees)	θ (radians)	sin θ
0	0	0
5	0.087	0.087
10	0.174	0.174
15	0.262	0.256
20	0.349	0.342

Equation (14.27) is the angular analogue of Eq. (14.11) and tells us that the angular acceleration of the pendulum is proportional to the angular displacement θ but opposite in sign. Thus as the pendulum moves to the right, its pull to the left increases until it stops and begins to return to the left. Similarly, when it moves towards left, its acceleration to the right tends to return it to the right and so on, as it swings to and fro in SHM. Thus the motion of a **simple pendulum swinging through small angles is approximately SHM**.

Comparing Eq. (14.27) with Eq. (14.11), we see that the angular frequency of the pendulum is,

$$\omega = \sqrt{\frac{mgL}{I}}$$

and the period of the pendulum, T, is given by,

$$T \quad \sqrt{\frac{I}{mgL}} \qquad (14.28)$$

All the mass of a simple pendulum is centred in the mass m of the bob, which is at a radius of L from the pivot point. Therefore, for this system, we can write $I = m L^2$ and substituting this in Eq. (14.28) we get,

SHM - how small should the amplitude be?

When you perform the experiment to determine the time period of a simple pendulum, your teacher tells you to keep the amplitude small. But have you ever asked how small is small? Should the amplitude to 5^0, 2^0, 1^0, or 0.5^0? Or could it be 10^0, 20^0, or 30^0?

To appreciate this, it would be better to measure the time period for different amplitudes, up to large amplitudes. Of course, for large oscillations, you will have to take care that the pendulum oscillates in a vertical plane. Let us denote the time period for small-amplitude oscillations as $T(0)$ and write the time period for amplitude θ_0 as $T(\theta_0) = cT(0)$, where c is the multiplying factor. If you plot a graph of c versus θ_0, you will get values somewhat like this:

θ_0	:	20^0	45^0	50^0	70^0	90^0
c	:	1.02	1.04	1.05	1.10	1.18

This means that the error in the time period is about 2% at an amplitude of 20^0, 5% at an amplitude of 50^0, and 10% at an amplitude of 70^0 and 18% at an amplitude of 90^0.

In the experiment, you will never be able to measure $T(0)$ because this means there are no oscillations. Even theoretically, sin θ is exactly equal to θ only for $\theta = 0$. There will be some inaccuracy for all other values of θ. The difference increases with increasing θ. Therefore we have to decide how much error we can tolerate. No measurement is ever perfectly accurate. You must also consider questions like these: What is the accuracy of the stopwatch? What is your own accuracy in starting and stopping the stopwatch? You will realise that the accuracy in your measurements at this level is never better than 5% or 10%. Since the above table shows that the time period of the pendulum increases hardly by 5% at an amplitude of 50^0 over its low amplitude value, you could very well keep the amplitude to be 50° in your experiments.

$$T \quad \sqrt{\frac{L}{g}} \tag{14.29}$$

Equation (14.29) represents a simple expression for the time period of a simple pendulum.

Example 14.9 What is the length of a simple pendulum, which ticks seconds ?

Answer From Eq. (14.29), the time period of a simple pendulum is given by,

$$T \quad \sqrt{\frac{L}{g}}$$

From this relation one gets,

$$L \quad \frac{gT^2}{4 \quad ^2}$$

The time period of a simple pendulum, which ticks seconds, is 2 s. Therefore, for $g = 9.8$ m s^{-2} and $T = 2$ s, L is

$$\frac{9.8(\text{m s}^{-2}) \quad 4(\text{s}^2)}{4 \quad ^2}$$

$$= 1 \text{ m}$$

14.9 DAMPED SIMPLE HARMONIC MOTION

We know that the motion of a simple pendulum, swinging in air, dies out eventually. Why does it happen ? This is because the air drag and the friction at the support oppose the motion of the pendulum and dissipate its energy gradually. The pendulum is said to execute **damped oscillations**. In damped oscillations, although the energy of the system is continuously dissipated, the oscillations remain apparently periodic. The dissipating forces are generally the frictional forces. To understand the effect of such external forces on the motion of an oscillator, let us consider a system as shown in Fig. 14.20. Here a block of mass m oscillates vertically on a spring with spring constant k. The block is connected to a vane through a rod (the vane and the rod are considered to be massless). The vane is submerged in a liquid. As the block oscillates up and down, the vane also moves along with it in the liquid. The up and down motion of the vane displaces the liquid, which in turn, exerts an inhibiting drag force (viscous drag) on it and thus on the entire oscillating system. With time, the mechanical energy of the block-spring system decreases, as energy is transferred to the thermal energy of the liquid and vane.

Let the damping force exerted by the liquid on the system be* $\mathbf{F}_d$. Its magnitude is proportional to the velocity $\mathbf{v}$ of the vane or the

Fig. 14.20 *A damped simple harmonic oscillator. The vane immersed in a liquid exerts a damping force on the block as it oscillates up and down.*

block. The force acts in a direction opposite to the direction of $\mathbf{v}$. This assumption is valid only when the vane moves slowly. Then for the motion along the x-axis (vertical direction as shown in Fig. 14.20), we have

$$\mathbf{F}_d = -b\mathbf{v} \tag{14.30}$$

where b is a **damping constant** that depends on the characteristics of the liquid and the vane. The negative sign makes it clear that the force is opposite to the velocity at every moment.

When the mass m is attached to the spring and released, the spring will elongate a little and the mass will settle at some height. This position, shown by O in Fig 14.20, is the equilibrium position of the mass. If the mass is pulled down or pushed up a little, the restoring force on the block due to the spring is $\mathbf{F}_S = -k\mathbf{x}$, where $\mathbf{x}$ is the displacement of the mass from its equilibrium position. Thus the total force acting

* *Under gravity, the block will be at a certain equilibrium position* O *on the spring; x here represents the displacement from that position.*

on the mass at any time t is $\mathbf{F} = -k\,\mathbf{x} - b\,\mathbf{v}$. If $\mathbf{a}(t)$ is the acceleration of the mass at time t, then by Newton's second law of motion for force components along the x-axis, we have

$$m\,a(t) = -k\,x(t) - b\,v(t) \tag{14.31}$$

Here we have dropped the vector notation because we are discussing one-dimensional motion. Substituting $\mathrm{d}x/\mathrm{d}t$ for $v(t)$ and $\mathrm{d}^2x/\mathrm{d}t^2$ for the acceleration $a(t)$ and rearranging gives us the differential equation,

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} \quad b\frac{\mathrm{d}x}{\mathrm{d}t} \quad k\,x \quad 0 \tag{14.32}$$

The solution of Eq. (14.32) describes the motion of the block under the influence of a damping force which is proportional to velocity. The solution is found to be of the form

$$x(t) \;=\; A\,e^{-b\,t/2m}\cos(\omega' t + \phi) \tag{14.33}$$

where a is the amplitude and ω' is the angular frequency of the damped oscillator given by,

$$\omega' = \sqrt{\frac{k}{m} - \frac{b^2}{4m^2}} \tag{14.34}$$

In this function, the cosine function has a period $2\pi/\omega'$ but the function $x(t)$ is not strictly periodic because of the factor $e^{-b\,t/2m}$ which decreases continuously with time. However, if the decrease is small in one time period T, the motion represented by Eq. (14.33) is approximately periodic.

The solution, Eq. (14.33), can be graphically represented as shown in Fig. 14.21. We can regard it as a cosine function whose amplitude, which is $Ae^{-b\,t/2m}$, gradually decreases with time.

Fig. 14.21 Displacement as a function of time in damped harmonic oscillations. Damping goes on increasing successively from curve **a** *to* **d**.

If $b = 0$ (there is no damping), then Eqs. (14.33) and (14.34) reduce to Eqs. (14.4) and (14.14b), expressions for the displacement and angular frequency of an undamped oscillator. We have seen that the mechanical energy of an undamped oscillator is constant and is given by Eq. (14.18) ($E = 1/2\ k\,A^2$). If the oscillator is damped, the mechanical energy is not constant but decreases with time. If the damping is small, we can find $E(t)$ by replacing A in Eq. (14.18) by $Ae^{-bt/2m}$, the amplitude of the damped oscillations. Thus we find,

$$E(t) = \frac{1}{2}k\,A^2\,e^{-b\,t/m} \tag{14.35}$$

Equation (14.35) shows that the total energy of the system decreases exponentially with time. Note that small damping means that the dimensionless ratio $\left(\frac{b}{\sqrt{k\,m}}\right)$ is much less than 1.

Example 14.10 For the damped oscillator shown in Fig. 14.20, the mass m of the block is 200 g, k = 90 N m^{-1} and the damping constant b is 40 g s^{-1}. Calculate (a) the period of oscillation, (b) time taken for its amplitude of vibrations to drop to half of its initial value and (c) the time taken for its mechanical energy to drop to half its initial value.

Answer (a) We see that km = 90×0.2 = 18 kg N m^{-1} = kg^2 s^{-2}; therefore $\sqrt{km}$ = 4.243 kg s^{-1}, and b = 0.04 kg s^{-1}. Therefore b is much less than $\sqrt{km}$. Hence the time period T from Eq. (14.34) is given by

$$T \quad 2\sqrt{\frac{m}{k}}$$

$$2\sqrt{\frac{0.2\ kg}{90\ \mathrm{N\ m^{\pm 1}}}}$$

$$= 0.3\ \mathrm{s}$$

(b) Now, from Eq. (14.33), the time, $T_{1/2}$, for the amplitude to drop to half of its initial value is given by,

$$T_{1/2} = \frac{\ln(1/2)}{\boldsymbol{b}/2\boldsymbol{m}}$$

$$\frac{0.693}{40} \quad 2 \quad 200 \text{ s}$$

$$= 6.93 \text{ s}$$

(c) For calculating the time, $t_{1/2}$, for its mechanical energy to drop to half its initial value we make use of Eq. (14.35). From this equation we have,

$$E(t_{1/2})/E(0) = \exp(-bt_{1/2}/m)$$

Or $$\tfrac{1}{2} = \exp(-bt_{1/2}/m)$$

$$\ln(1/2) = -(bt_{1/2}/m)$$

Or $$t_{1/2} \quad \frac{0.693}{40 \text{ g s}^{-1}} \quad 200 \text{ g}$$

$$= 3.46 \text{ s}$$

This is just half of the decay period for amplitude. This is not suprising, because, according to Eqs. (14.33) and (14.35), energy depends on the square of the amplitude. Notice that there is a factor of 2 in the exponents of the two exponentials. ◀

14.10 FORCED OSCILLATIONS AND RESONANCE

A person swinging in a swing without anyone pushing it or a simple pendulum, displaced and released, are examples of free oscillations. In both the cases, the amplitude of swing will gradually decrease and the system would, ultimately, come to a halt. Because of the ever-present dissipative forces, the free oscillations cannot be sustained in practice. They get damped as seen in section 14.9. However, while swinging in a swing if you apply a push periodically by pressing your feet against the ground, you find that not only the oscillations can now be maintained but the amplitude can also be increased. Under this condition the swing has **forced,** or **driven, oscillations**. In case of a system executing driven oscillations under the action of a harmonic force, two angular frequencies are important : (1) **the natural** angular frequency ω of the system, which is the angular frequency at which it will oscillate if it were displaced from equilibrium position and then left to oscillate freely, and (2) the angular frequency ω_d of the external force causing the driven oscillations.

Suppose an external force $F(t)$ of amplitude F_0 that varies periodically with time is applied to a damped oscillator. Such a force can be represented as,

$$F(t) = F_o \cos \omega_d t \tag{14.36}$$

The motion of a particle under the combined action of a linear restoring force, damping force and a time dependent driving force represented by Eq. (14.36) is given by,

$$m\,a(t) = -k\,x(t) - bv(t) + F_o \cos \omega_d t \tag{14.37a}$$

Substituting d^2x/dt^2 for acceleration in Eq. (14.37a) and rearranging it, we get

$$m\frac{d^2x}{dt^2} \quad b\frac{dx}{dt} \quad kx \quad F_o \cos \omega_d t \tag{14.37b}$$

This is the equation of an oscillator of mass m on which a periodic force of (angular) frequency ω_d is applied. The oscillator initially oscillates with its natural frequency ω When we apply the external periodic force, the oscillations with the natural frequency die out, and then the body oscillates with the (angular) frequency of the external periodic force. Its displacement, after the natural oscillations die out, is given by

$$x(t) = A \cos(\omega_d t + \phi) \tag{14.38}$$

where t is the time measured from the moment when we apply the periodic force.

The amplitude A is a function of the forced frequency ω_d and the natural frequency ω Analysis shows that it is given by

$$A \quad \frac{F}{m \quad {}_d^{\,2} \quad {}_d b^{\,1/2}} \tag{14.39a}$$

and $$\tan\phi = \frac{-v_o}{\omega_d x_o} \tag{14.39b}$$

where m is the mass of the particle and v_0 and x_0 are the velocity and the displacement of the particle at time $t = 0$, which is the moment when we apply the periodic force. Equation (14.39) shows that the amplitude of the forced oscillator depends on the (angular) frequency of the driving force. We can see a different behaviour of the oscillator when ω_d is far from $\omega$and when it is close to ω We consider these two cases.

(a) **Small Damping, Driving Frequency far from Natural Frequency :** In this case, $\omega_d b$ will be much smaller than $m(\omega^2 - \omega_d^2)$, and we can neglect that term. Then Eq. (14.39) reduces to

$$A \quad \frac{F}{m \quad {}^{2}_{d}} \tag{14.40}$$

Figure 14.22 shows the dependence of the displacement amplitude of an oscillator on the angular frequency of the driving force for different amounts of damping present in the system. It may be noted that in all the cases the amplitude is greatest when $\omega_d / \omega = 1$. The curves in this figure show that smaller the damping, the taller and narrower is the resonance peak.

If we go on changing the driving frequency, the amplitude tends to infinity when it equals the natural frequency. But this is the ideal case of zero damping, a case which never arises in a real system as the damping is never perfectly zero. You must have experienced in a swing that when the timing of your push exactly matches with the time period of the swing, your swing gets the maximum amplitude. This amplitude is large, but not infinity, because there is always some damping in your swing. This will become clear in the (b).

(b) **Driving Frequency Close to Natural Frequency :** If ω_d is very close to ω, $m\,(\omega^2 - {}^{2}_{d})$

Fig. 14.22 The amplitude of a forced oscillator as a function of the angular frequency of the driving force. The amplitude is greatest near $\omega_d / \omega = 1$. The five curves correspond to different extents of damping present in the system. Curve **a** *corresponds to the least damping, and damping goes on increasing successively in curves* ***b, c, d, e****. Notice that the peak shifts to the left with increasing* ***b****.*

would be much less than $\omega_d b$, for any reasonable value of b, then Eq. (14.39) reduces to

$$A = \frac{F_o}{\omega_d b} \tag{14.41}$$

This makes it clear that the maximum possible amplitude for a given driving frequency is governed by the driving frequency and the damping, and is never infinity. The phenomenon of increase in amplitude when the driving force is close to the natural frequency of the oscillator is called **resonance**.

In our daily life we encounter phenomena which involve resonance. Your experience with swings is a good example of resonance. You might have realised that the skill in swinging to greater heights lies in the synchronisation of the rhythm of pushing against the ground with the natural frequency of the swing.

To illustrate this point further, let us consider a set of five simple pendulums of assorted lengths suspended from a common rope as shown in Fig. 14.23. The pendulums 1 and 4 have the same lengths and the others have different lengths. Now let us set pendulum 1 into motion. The energy from this pendulum gets transferred to other pendulums through the connecting rope and they start oscillating. The driving force is provided through the connecting rope. The frequency of this force is the frequency with which pendulum 1 oscillates. If we observe the response of pendulums 2, 3 and 5, they first start oscillating with their natural frequencies

Fig. 14.23 A system of five simple pendulums suspended from a common rope.

of oscillations and different amplitudes, but this motion is gradually damped and not sustained. Their frequencies of oscillation gradually change and ultimately they oscillate with the frequency of pendulum 1, i.e. the frequency of the driving force but with different amplitudes. They oscillate with small amplitudes. The response of pendulum 4 is in contrast to this set of pendulums. It oscillates with the same frequency as that of pendulum 1 and its amplitude gradually picks up and becomes very large. A resonance-like response is seen. This happens because in this the condition for resonance is satisfied, i.e. the natural frequency of the system coincides with that of the driving force.

All mechanical structures have one or more natural frequencies, and if a structure is subjected to a strong external periodic driving force that matches one of these frequencies, the resulting oscillations of the structure may rupture it. The Tacoma Narrows Bridge at Puget Sound, Washington, USA was opened on July 1, 1940. Four months later winds produced a pulsating resultant force in resonance with the natural frequency of the structure. This caused a steady increase in the amplitude of oscillations until the bridge collapsed. It is for the same reason the marching soldiers break steps while crossing a bridge. Aircraft designers make sure that none of the natural frequencies at which a wing can oscillate match the frequency of the engines in flight. Earthquakes cause vast devastation. It is interesting to note that sometimes, in an earthquake, short and tall structures remain unaffected while the medium height structures fall down. This happens because the natural frequencies of the short structures happen to be higher and those of taller structures lower than the frequency of the seismic waves.

SUMMARY

1. The motions which repeat themselves are called *periodic motions.*
2. The *period* T is the time required for one complete oscillation, or cycle. It is related to the frequency ν by,

$$T = \frac{1}{\nu}$$

The *frequency* ν of periodic or oscillatory motion is the number of oscillations per unit time. In the SI, it is measured in hertz :

$$1 \text{ hertz} = 1 \text{ Hz} = 1 \text{ oscillation per second} = 1\text{s}^{-1}$$

3. In *simple harmonic motion* (SHM), the displacement $x(t)$ of a particle from its equilibrium position is given by,

$$x(t) = A\cos(\omega t + \phi) \qquad \text{(displacement)},$$

in which A is the *amplitude* of the displacement, the quantity $(\omega t + \phi)$ is the phase of the motion, and ϕ is the *phase constant*. The *angular frequency* ω is related to the period and frequency of the motion by,

$$\frac{2}{T} \quad 2 \qquad \text{(angular frequency)}.$$

4. Simple harmonic motion is the projection of uniform circular motion on the diameter of the circle in which the latter motion occurs.
5. The particle velocity and acceleration during SHM as functions of time are given by,

$$v(t) = -\omega A\sin(\omega t + \phi) \qquad \text{(velocity)},$$

$$a(t) = -\omega^2 A\cos(\omega t + \phi)$$

$$= -\omega^2 x(t) \qquad \text{(acceleration)},$$

Thus we see that both velocity and acceleration of a body executing simple harmonic motion are periodic functions, having the velocity *amplitude* $v_m = \omega A$ and *acceleration amplitude* $a_m = \omega^2 A$, respectively.

6. The force acting simple harmonic motion is proportional to the displacement and is always directed towards the centre of motion.

7. A particle executing simple harmonic motion has, at any time, kinetic energy $K = \frac{1}{2} mv^2$ and potential energy $U = \frac{1}{2} kx^2$. If no friction is present the mechanical energy of the system, $E = K + U$ always remains constant even though K and U change with time

8. A particle of mass m oscillating under the influence of a Hooke's law restoring force given by $F = -kx$ exhibits simple harmonic motion with

$$\omega = \sqrt{\frac{k}{m}} \qquad \text{(angular frequency)}$$

$$T \quad 2\sqrt{\frac{m}{k}} \qquad \text{(period)}$$

Such a system is also called a linear oscillator.

9. The motion of a simple pendulum swinging through small angles is approximately simple harmonic. The period of oscillation is given by,

$$T \quad 2\sqrt{\frac{L}{g}}$$

10. The mechanical energy in a real oscillating system decreases during oscillations because external forces, such as drag, inhibit the oscillations and transfer mechanical energy to thermal energy. The real oscillator and its motion are then said to be *damped*. If the *damping force* is given by $F_d = -bv$, where v is the velocity of the oscillator and b is a *damping constant*, then the displacement of the oscillator is given by,

$$x(t) = A e^{-bt/2m} \cos(\omega' t + \phi)$$

where ω', the angular frequency of the damped oscillator, is given by

$$\sqrt{\frac{k}{m} \quad \frac{b^2}{4m^2}}$$

If the damping constant is small then $\omega' \approx \omega$ where ω is the angular frequency of the undamped oscillator. The mechanical energy E of the damped oscillator is given by

$$E(t) \quad \frac{1}{2} kA^2 e^{\ bt/m}$$

11. If an external force with angular frequency ω_d acts on an oscillating system with natural angular frequency ω the system oscillates with angular frequency ω_d. The amplitude of oscillations is the greatest when

$$\omega_d = \omega$$

a condition called *resonance*.

Physical quantity	Symbol	Dimensions	Unit	Remarks
Period	T	[T]	s	The least time for motion to repeat itself
Frequency	ν (or f)	$[T^{-1}]$	s^{-1}	$\nu = \frac{1}{T}$
Angular frequency	ω	$[T^{-1}]$	s^{-1}	$\omega = 2\pi\nu$
Phase constant	ϕ	Dimensionless	rad	Initial value of phase of displacement in SHM
Force constant	k	$[MT^{-2}]$	$N\ m^{-1}$	Simple harmonic motion $F = -kx$

POINTS TO PONDER

1. The period T is the *least time* after which motion repeats itself. Thus, motion repeats itself after nT where n is an integer.
2. Every periodic motion is not simple harmonic motion. Only that periodic motion governed by the force law $F = -kx$ is simple harmonic.
3. Circular motion can arise due to an inverse-square law force (as in planetary motion) as well as due to simple harmonic force in two dimensions equal to: $-m\omega^2 r$. In the latter case, the phases of motion, in two perpendicular directions (x and y) must differ by $\omega/2$. Thus, a particle subject to a force $-m\omega^2 r$ with initial position (o, A) and velocity (ωA, o) will move uniformly in a circle of radius A.
4. For linear simple harmonic motion with a given ω two arbitrary initial conditions are necessary and sufficient to determine the motion completely. The initial condition may be (i) initial position and initial velocity or (ii) amplitude and phase or (iii) energy and phase.
5. From point 4 above, given amplitude or energy, phase of motion is determined by the initial position or initial velocity.
6. A combination of two simple harmonic motions with arbitrary amplitudes and phases is not necessarily periodic. It is periodic only if frequency of one motion is an integral multiple of the other's frequency. However, a periodic motion can always be expressed as a sum of infinite number of harmonic motions with appropriate amplitudes.
7. The period of SHM does not depend on amplitude or energy or the phase constant. Contrast this with the periods of planetary orbits under gravitation (Kepler's third law).
8. The motion of a simple pendulum is simple harmonic for small angular displacement.
9. For motion of a particle to be simple harmonic, its displacement x must be expressible in either of the following forms :

   $$x = A\cos\omega t + B\sin\omega t$$

   $$x = A\cos(\omega t + \alpha),\ x = B\sin(\omega t + \beta)$$

   The three forms are completely equivalent (any one can be expressed in terms of any other two forms).

   Thus, damped simple harmonic motion [Eq. (14.31)] is not strictly simple harmonic. It is approximately so only for time intervals much less than $2m/b$ where b is the damping constant.
10. In forced oscillations, the steady state motion of the particle (after the force oscillations die out) is simple harmonic motion whose frequency is the frequency of the driving frequency ω_d, not the natural frequency ω of the particle.

11. In the ideal case of zero damping, the amplitude of simple harmonic motion at resonance is infinite. This is no problem since all real systems have some damping, however, small.
12. Under forced oscillation, the phase of harmonic motion of the particle differs from the phase of the driving force.

Exercises

14.1 Which of the following examples represent periodic motion?
(a) A swimmer completing one (return) trip from one bank of a river to the other and back.
(b) A freely suspended bar magnet displaced from its N-S direction and released.
(c) A hydrogen molecule rotating about its center of mass.
(d) An arrow released from a bow.

14.2 Which of the following examples represent (nearly) simple harmonic motion and which represent periodic but not simple harmonic motion?
(a) the rotation of earth about its axis.
(b) motion of an oscillating mercury column in a U-tube.
(c) motion of a ball bearing inside a smooth curved bowl, when released from a point slightly above the lower most point.
(d) general vibrations of a polyatomic molecule about its equilibrium position.

14.3 Figure 14.27 depicts four *x-t* plots for linear motion of a particle. Which of the plots represent periodic motion? What is the period of motion (in case of periodic motion) ?

Fig. 14.27

14.4 Which of the following functions of time represent (a) simple harmonic, (b) periodic but not simple harmonic, and (c) non-periodic motion? Give period for each case of periodic motion (ω is any positive constant):

(a) $\sin \omega t - \cos \omega t$
(b) $\sin^3 \omega t$
(c) $3 \cos (\pi/4 - 2\omega t)$
(d) $\cos \omega t + \cos 3\omega t + \cos 5\omega t$
(e) $\exp (-\omega^2 t^2)$
(f) $1 + \omega t + \omega^2 t^2$

14.5 A particle is in linear simple harmonic motion between two points, A and B, 10 cm apart. Take the direction from A to B as the positive direction and give the signs of velocity, acceleration and force on the particle when it is

(a) at the end A,
(b) at the end B,
(c) at the mid-point of AB going towards A,
(d) at 2 cm away from B going towards A,
(e) at 3 cm away from A going towards B, and
(f) at 4 cm away from B going towards A.

14.6 Which of the following relationships between the acceleration a and the displacement x of a particle involve simple harmonic motion?

(a) $a = 0.7x$
(b) $a = -200x^2$
(c) $a = -10x$
(d) $a = 100x^3$

14.7 The motion of a particle executing simple harmonic motion is described by the displacement function,

$$x(t) = A \cos (\omega t + \phi).$$

If the initial ($t = 0$) position of the particle is 1 cm and its initial velocity is ωcm/s, what are its amplitude and initial phase angle ? The angular frequency of the particle is $\pi\ s^{-1}$. If instead of the cosine function, we choose the sine function to describe the SHM : $x = B \sin (\omega t + \alpha)$, what are the amplitude and initial phase of the particle with the above initial conditions.

14.8 A spring balance has a scale that reads from 0 to 50 kg. The length of the scale is 20 cm. A body suspended from this balance, when displaced and released, oscillates with a period of 0.6 s. What is the weight of the body ?

14.9 A spring having with a spring constant 1200 N m^{-1} is mounted on a horizontal table as shown in Fig. 14.28. A mass of 3 kg is attached to the free end of the spring. The mass is then pulled sideways to a distance of 2.0 cm and released.

Fig. 14.28

Determine (i) the frequency of oscillations, (ii) maximum acceleration of the mass, and (iii) the maximum speed of the mass.

14.10 In Exercise 14.9, let us take the position of mass when the spring is unstreched as $x = 0$, and the direction from left to right as the positive direction of x-axis. Give x as a function of time t for the oscillating mass if at the moment we start the stopwatch ($t = 0$), the mass is

(a) at the mean position,
(b) at the maximum stretched position, and
(c) at the maximum compressed position.

In what way do these functions for SHM differ from each other, in frequency, in amplitude or the initial phase?

14.11 Figures 14.29 correspond to two circular motions. The radius of the circle, the period of revolution, the initial position, and the sense of revolution (i.e. clockwise or anti-clockwise) are indicated on each figure.

Fig. 14.29

Obtain the corresponding simple harmonic motions of the x-projection of the radius vector of the revolving particle P, in each case.

14.12 Plot the corresponding reference circle for each of the following simple harmonic motions. Indicate the initial (t =0) position of the particle, the radius of the circle, and the angular speed of the rotating particle. For simplicity, the sense of rotation may be fixed to be anticlockwise in every case: (x is in cm and t is in s).

(a) $x = -2 \sin (3t + \pi/3)$

(b) $x = \cos (\pi/6 - t)$

(c) $x = 3 \sin (2\pi t + \pi/4)$

(d) $x = 2 \cos \pi t$

14.13 Figure 14.30 (a) shows a spring of force constant k clamped rigidly at one end and a mass m attached to its free end. A force **F** applied at the free end stretches the spring. Figure 14.30 (b) shows the same spring with both ends free and attached to a mass m at either end. Each end of the spring in Fig. 14.30(b) is stretched by the same force **F**.

Fig. 14.30

(a) What is the maximum extension of the spring in the two cases ?
(b) If the mass in Fig. (a) and the two masses in Fig. (b) are released, what is the period of oscillation in each case ?

14.14 The piston in the cylinder head of a locomotive has a stroke (twice the amplitude) of 1.0 m. If the piston moves with simple harmonic motion with an angular frequency of 200 rad/min, what is its maximum speed ?

14.15 The acceleration due to gravity on the surface of moon is 1.7 m s^{-2}. What is the time period of a simple pendulum on the surface of moon if its time period on the surface of earth is 3.5 s ? (g on the surface of earth is 9.8 m s^{-2})

14.16 Answer the following questions :

(a) Time period of a particle in SHM depends on the force constant k and mass m of the particle:

$T = 2\pi\sqrt{\frac{m}{k}}$. A simple pendulum executes SHM approximately. Why then is the time period of a pendulum independent of the mass of the pendulum?

(b) The motion of a simple pendulum is approximately simple harmonic for small angle oscillations. For larger angles of oscillation, a more involved analysis shows that T is greater than $2\pi\sqrt{\frac{l}{g}}$. Think of a qualitative argument to appreciate this result.

(c) A man with a wristwatch on his hand falls from the top of a tower. Does the watch give correct time during the free fall ?

(d) What is the frequency of oscillation of a simple pendulum mounted in a cabin that is freely falling under gravity ?

14.17 A simple pendulum of length l and having a bob of mass M is suspended in a car. The car is moving on a circular track of radius R with a uniform speed v. If the pendulum makes small oscillations in a radial direction about its equilibrium position, what will be its time period ?

14.18 A cylindrical piece of cork of density of base area A and height h floats in a liquid of density ρ_l. The cork is depressed slightly and then released. Show that the cork oscillates up and down simple harmonically with a period

$$T = 2\pi\sqrt{\frac{h\rho}{\rho_l g}}$$

where ρ is the density of cork. (Ignore damping due to viscosity of the liquid).

14.19 One end of a U-tube containing mercury is connected to a suction pump and the other end to atmosphere. A small pressure difference is maintained between the two columns. Show that, when the suction pump is removed, the column of mercury in the U-tube executes simple harmonic motion.

Additional Exercises

14.20 An air chamber of volume V has a neck area of cross section a into which a ball of mass m just fits and can move up and down without any friction (Fig.14.33). Show that when the ball is pressed down a little and released , it executes SHM. Obtain an expression for the time period of oscillations assuming pressure-volume variations of air to be isothermal [see Fig. 14.33].

Fig.14.33

14.21 You are riding in an automobile of mass 3000 kg. Assuming that you are examining the oscillation characteristics of its suspension system. The suspension sags 15 cm when the entire automobile is placed on it. Also, the amplitude of oscillation decreases by 50% during one complete oscillation. Estimate the values of (a) the spring constant k and (b) the damping constant b for the spring and shock absorber system of one wheel, assuming that each wheel supports 750 kg.

14.22 Show that for a particle in linear SHM the average kinetic energy over a period of oscillation equals the average potential energy over the same period.

14.23 A circular disc of mass 10 kg is suspended by a wire attached to its centre. The wire is twisted by rotating the disc and released. The period of torsional oscillations is found to be 1.5 s. The radius of the disc is 15 cm. Determine the torsional spring constant of the wire. (Torsional spring constant α is defined by the relation $J = -\alpha\,\theta$, where J is the restoring couple and θ the angle of twist).

14.24 A body describes simple harmonic motion with an amplitude of 5 cm and a period of 0.2 s. Find the acceleration and velocity of the body when the displacement is (a) 5 cm, (b) 3 cm, (c) 0 cm.

14.25 A mass attached to a spring is free to oscillate, with angular velocity ω, in a horizontal plane without friction or damping. It is pulled to a distance x_0 and pushed towards the centre with a velocity v_0 at time $t = 0$. Determine the amplitude of the resulting oscillations in terms of the parameters ω, x_0 and v_0. [Hint : Start with the equation $x = a \cos(\omega t+\theta)$ and note that the initial velocity is negative.]

CHAPTER FIFTEEN

WAVES

15.1 Introduction
15.2 Transverse and longitudinal waves
15.3 Displacement relation in a progressive wave
15.4 The speed of a travelling wave
15.5 The principle of superposition of waves
15.6 Reflection of waves
15.7 Beats
15.8 Doppler effect
Summary
Points to ponder
Exercises
Additional exercises

15.1 INTRODUCTION

In the previous Chapter, we studied the motion of objects oscillating in isolation. What happens in a system, which is a collection of such objects ? A material medium provides such an example. Here, elastic forces bind the constituents to each other and, therefore, the motion of one affects that of the other. If you drop a little pebble in a pond of still water, the water surface gets disturbed. The disturbance does not remain confined to one place, but propagates outward along a circle. If you continue dropping pebbles in the pond, you see circles rapidly moving outward from the point where the water surface is disturbed. It gives a feeling as if the water is moving outward from the point of disturbance. If you put some cork pieces on the disturbed surface, it is seen that the cork pieces move up and down but do not move away from the centre of disturbance. This shows that the water mass does not flow outward with the circles, but rather a moving disturbance is created. Similarly, when we speak, the sound moves outward from us, without any flow of air from one part of the medium to another. The disturbances produced in air are much less obvious and only our ears or a microphone can detect them. These patterns, which move without the actual physical transfer or flow of matter as a whole, are called **waves**. In this Chapter, we will study such waves.

In a wave, information and energy, in the form of signals, propagate from one point to another but no material object makes the journey. All our communications depend on the transmission of signals through waves. When we make a telephone call to a friend at a distant place, a sound wave carries the message from our vocal cords to the telephone. There, an electrical signal is generated which propagates along the copper wire. If the distance is too large, the electrical signal generated may be transformed into a light signal or

electromagnetic waves and transmitted through optical cables or the atmosphere, possibly by way of a communication satellite. At the receiving end, the electrical or light signal or the electromagnetic waves are transformed back into sound waves travelling from the telephone to the ear.

Not all waves require a medium for their propagation. We know that light waves can travel through vacuum. The light emitted by stars, which are hundreds of light years away, reaches us through inter-stellar space, which is practically a vacuum.

The waves we come across are mainly of three types: (a) mechanical waves, (b) electromagnetic waves and (c) matter waves. Mechanical waves are most familiar because we encounter them constantly; common examples include water waves, sound waves, seismic waves, etc. All these waves have certain central features : They are governed by Newton's laws, and can exist only within a material medium, such as water, air, and rock. The common examples of electromagnetic waves are visible and ultra-violet light, radio waves, microwaves, x-rays etc. All electromagnetic waves travel through vacuum at the same speed c, given by

$$c = 299,792,458 \text{ m s}^{-1} \quad \text{(speed of light)} \quad (15.1)$$

Unlike the mechanical waves, the electromagnetic waves do not require any medium for their propagation. You would learn more about these waves later.

Matter waves are associated with moving electrons, protons, neutrons and other fundamental particles, and even atoms and molecules. Because we commonly think of these as constituting matter, such waves are called **matter waves**. They arise in quantum mechanical description of nature that you will learn in your later studies. Though conceptually more abstract than mechanical or electromagnetic waves, they have already found applications in several devices basic to modern technology; matter waves associated with electrons are employed in electron microscopes.

In this chapter we will study mechanical waves, which require a material medium for their propagation.

The aesthetic influence of waves on art and literature is seen from very early times; yet the first scientific analysis of wave motion dates back to the seventeenth century. Some of the famous scientists associated with the physics of wave motion are Christiaan Huygens (1629-1695), Robert Hooke and Isaac Newton. The understanding of physics of waves followed the physics of oscillations of masses tied to springs and physics of the simple pendulum. Waves in elastic media are intimately connected with harmonic oscillations. (Stretched strings, coiled springs, air, etc., are examples of elastic media.) We shall illustrate this connection through simple examples.

Consider a collection of springs connected to one another as shown in Fig. 15.1. If the spring at one end is pulled suddenly and released, the disturbance travels to the other end. What has happened ? The first spring is disturbed from its equilibrium length. Since the second spring is connected to the first, it is also stretched or compressed, and so on. The disturbance moves

Fig. 15.1 A collection of springs connected to each other. The end A is pulled suddenly generating a disturbance, which then propagates to the other end.

from one end to the other; but each spring only executes small oscillations about its equilibrium position. As a practical example of this situation, consider a stationary train at a railway station. Different bogies of the train are coupled to each other through a spring coupling. When an engine is attached at one end, it gives a push to the bogie next to it; this push is transmitted from one bogie to another without the entire train being bodily displaced.

Now let us consider the propagation of sound waves in air. As the wave passes through air, it compresses or expands a small region of air. This causes a change in the density of that region, say $\delta\rho$, this change induces a change in pressure, δp, in that region. Pressure is force per unit area, so there is a **restoring force proportional** to the disturbance, just like in a spring. In this case, the quantity similar to extension or compression of the spring is the change in density. If a region is compressed, the molecules in that region are packed together, and they tend to move out to the adjoining region, thereby increasing the density or creating compression in the adjoining region. Consequently, the air

in the first region undergoes rarefaction. If a region is comparatively rarefied the surrounding air will rush in making the rarefaction move to the adjoining region. Thus, the compression or rarefaction moves from one region to another, making the propagation of a disturbance possible in air.

In solids, similar arguments can be made. In a crystalline solid, atoms or group of atoms are arranged in a periodic lattice. In these, each atom or group of atoms is in equilibrium, due to forces from the surrounding atoms. Displacing one atom, keeping the others fixed, leads to restoring forces, exactly as in a spring. So we can think of atoms in a lattice as end points, with springs between pairs of them.

In the subsequent sections of this chapter we are going to discuss various characteristic properties of waves.

15.2 TRANSVERSE AND LONGITUDINAL WAVES

Mechanical waves can be transverse or longitudinal depending on the relationship between the directions of disturbance or displacement in the medium and that of the propagation of wave. To differentiate between them let us consider the response of a stretched string fixed at one end. If you give a single up-and-down jerk to the free end of this string, as shown in Fig. 15.2, a wave in the form of a single **pulse** travels along the string. We assume that the string is very long as compared to the size of the pulse, so that the pulse dissipates out by the time it reaches the other end and, therefore, its reflection from the other end may be ignored. The formation and propagation of this pulse is possible because the string is under tension. When you pull your end of the string upwards it begins to pull upwards on the adjacent section of the string, because of the tension between the two sections. As the adjacent section begins to move upwards, it begins to pull the next section upwards, and so on. In the meanwhile you have pulled down your end of the string. As each section moves upwards in turn, it begins to be pulled back downwards by neighbouring sections that are already on the way down. The net result is that a distortion in the shape of the string (the pulse) moves along the string with a certain velocity **v**.

Fig. 15.2 A single pulse is sent along a stretched string. A typical element of the string (such as that marked with a dot) moves up and then down as the pulse passes through. The element's motion is perpendicular to the direction in which the wave travels.

If you move your end up and down in a continuous manner, a continuous wave travels along the string with a velocity **v**. However, if the motion of your hand is a sinusoidal function of time, at any given instant of time the wave will have a sinusoidal shape as shown in Fig. 15.3. The wave has the shape of a sine or cosine curve.

The waves shown in Fig. 15.3 can be studied in two ways. Oneway is to monitor the waveforms as they move to the right, i.e. take a 'snapshot' of the string at a given instant of time. Alternatively, we fix our attention to a particular position on the string and monitor the motion of an element at that point as it oscillates up and down while a wave passes through it. We would find that the displacement of every such oscillating string element is transverse (i.e. perpendicular) to the direction of travel of the wave as indicated in Fig. 15.3. Such a wave is said to be a **transverse wave**.

Fig. 15.3 A sinusoidal wave is sent along the string. A typical element of the string moves up and down continuously as the wave passes. It is a transverse wave.

Now, let us consider the production of waves in a long air-filled pipe by the movement of a piston as shown in Fig. 15.4. If you suddenly move the piston to the right and then to the left, you are sending a pulse of pressure along the pipe. The motion of the piston to the right pushes

Fig. 15.4 A sound wave is set up in an air filled pipe by moving a piston back and forth. As the oscillations of an element of air are parallel to the direction in which the wave travels, the wave is a longitudinal wave.

the elements of air next towards the right, changing the air pressure there. The increased pressure in this region then pushes on the elements of air somewhat farther along the pipe. Moving the piston to the left reduces the air pressure next to it. This causes the elements of air next to it move back to the left and then the farther elements follow. Thus, the motion of the air and the change in air pressure travel towards the right along the pipe as a pulse.

If you push and pull on the piston in a simple harmonic manner, a sinusoidal wave travels along the pipe. It may be noted that the motion of the elements of air is parallel to the direction of propagation of the wave. This motion is said to be **longitudinal** and the wave produced is, therefore, called a **longitudinal wave**. The sound waves produced in air are such pressure waves and are therefore of longitudinal character.

In short, **in transverse waves, the constituents of the medium oscillate perpendicular to the direction of wave propagation and in longitudinal waves they oscillate along the direction of wave propagation**.

A wave, transverse or longitudinal, is said to be **travelling or progressive** if it travels from one point of the medium to another. A progressive wave is to be distinguished from a standing or stationary wave (see Section 15.7). In Fig. 15.3 transverse waves travel from one end of the string to the other end while the longitudinal waves in Fig. 15.4 travel from one end of the pipe to its other end. We note again that in both the cases, it is the wave or the disturbance that moves from end to end, not the material through which the waves propagate.

In transverse waves, the particle motion is normal to the direction of propagation of the wave. Therefore, as the wave propagates, each element of the medium undergoes a shearing strain. Transverse waves can, therefore, be propagated only in those media which can sustain shearing stress, such as solids and strings, and not in fluids. Fluids as well as solids can sustain compressive strain; therefore, longitudinal waves can propagate in all elastic media. For example, in medium like a steel bar, both transverse and longitudinal waves can propagate while air can sustain only longitudinal waves. The waves on the surface of water are of two kinds: **capillary waves** and **gravity waves**. The former are ripples of fairly short wavelength—no more than a few centimetres—and the restoring force that produces them is the surface tension of water. Gravity waves have wavelengths typically ranging from several metres to several hundred metres. The restoring force that produces these waves is the pull of gravity, which tends to keep the water surface at its lowest level. The oscillations of the particles in these waves are not confined to the surface only, but extend with diminishing amplitude to the very bottom. The particle motion in the water waves involves a complicated motion; they not only move up and down but also back and forth. The waves in an ocean are a combination of both longitudinal and transverse waves.

It is found that generally transverse and longitudinal waves travel with different speeds in the same medium.

▶ ***Example 15.1*** Given below are some examples of wave motion. State in each case if the wave motion is transverse, longitudinal or a combination of both:

(a) Motion of a kink in a longitudinal spring produced by displacing one end of the spring sideways.

(b) Waves produced in a cylinder containing a liquid by moving its piston back and forth.

(c) Waves produced by a motorboat sailing in water.

(d) Ultrasonic waves in air produced by a vibrating quartz crystal.

Answer

(a) Transverse and longitudinal
(b) Longitudinal
(c) Transverse and longitudinal
(d) Longitudinal ◀

15.3 DISPLACEMENT RELATION IN A PROGRESSIVE WAVE

To describe the propagation of a wave in a medium (and the motion of any constituent of the medium), we need a function that completely gives the shape of the wave at every instant of time. For example, to completely describe the wave on a string (and the motion of any element along its length) we need a relation which describes the displacement of an element at a particular position as a function of time and also describes the state of vibration of various elements of the string along its length at a given instant of time. For a sinusoidal wave, as shown in Fig. 15.3, this function should be periodic in space as well as in time. Let $y(x, t)$ denote the transverse displacement of the element at position x at time t. As the wave sweeps through succeeding elements of the string, the elements oscillate parallel to the y-axis. At any time t, the displacement y of the element located at position x is given by

$$y(x, t) = a \sin(kx - \omega t + \phi) \qquad (15.2)$$

One can as well choose a cosine function or a linear combination of sine and cosine functions such as,

$$y(x, t) = A \sin(kx - \omega t) + B \cos(kx - \omega t), \qquad (15.3)$$

then in Eq. (15.2),

$$a = \sqrt{A^2 + B^2} \text{ and } \phi = \tan^{-1}\left(\frac{B}{A}\right)$$

The function represented in Eq. (15.2) is periodic in position coordinate x and time t. It represents a transverse wave moving along the x-axis. At any time t, it gives the displacement of the elements of the string as a function of their position. It can tell us the shape of the wave at any given time and show how the wave progresses. Functions, such as that given in Eq. (15.2), represent a progressive wave travelling along the positive direction of the x-axis. On the other hand a function,

$$y(x, t) = a \sin(kx + \omega t + \phi), \qquad (15.4)$$

represents a wave travelling in the negative direction of x-axis (see Section 15.4). The set of four parameters a, ϕ, k, and ω in Eq. (15.2) completely describe a harmonic wave. The names of these parameters are displayed in Fig. 15.5 and are defined later.

$$\underbrace{y(x, t)}_{\text{Displacement}} = \underbrace{a}_{\text{Amplitude}} \sin(\overbrace{\underset{\substack{\uparrow \\ \text{Angular} \\ \text{Wave} \\ \text{Number}}}{kx} - \underset{\substack{\uparrow \\ \text{Angular} \\ \text{Frequency}}}{\omega t} + \underset{\substack{\uparrow \\ \text{Initial} \\ \text{Phase} \\ \text{Angle}}}{\phi}}^{\text{Phase}})$$

Fig. 15.5 *The names of the quantities in Eq. (15.2) for a progressive wave.*

To understand the definition of the quantities in Eq. (15.2), let us consider the graphs shown in Fig. 15.6. These graphs represent plots of Eq. (15.2) for five different values of time t as the wave travels in positive direction of x-axis. A point of maximum positive displacement in a wave, shown by the arrow, is called *crest*, and a point of maximum negative displacement is called **trough**. The progress of the wave is indicated by the progress of the short arrow pointing to a crest of the wave towards the right. As we move from one plot to another, the short arrow moves to the right with the wave shape,

Fig. 15.6 *Plots of Eq. (15.2) for a wave travelling in the positive direction of an x-axis at five different values of time t.*

but the string moves only parallel to y-axis. It can be seen that as we go from plot (a) to (e), a particular element of the string has undergone one complete cycle of changes or completed one full oscillation. During this course of time the short arrow head or the wave has moved by a characteristic distance along the x-axis.

In the context of the above five plots, we will now define various quantities in Eq. (15.2) and shown in Fig.15.5.

15.3.1 Amplitude and Phase

The **amplitude** a of a wave such as that in Figs.15.5 and 15.6 is the magnitude of the maximum displacement of the elements from their equilibrium positions as the wave passes through them. It is depicted in Fig. 15.6 (a). Since a is a magnitude, it is a positive quantity, even if the displacement is negative.

The **phase** of the wave is the argument $(kx - \omega t + \phi)$ of the oscillatory term $\sin(kx - \omega t + \phi)$ in Eq. (15.2). It describes the state of motion as the wave sweeps through a string element at a particular position x. It changes linearly with time t. The sine function also changes with time, oscillating between +1 and –1. Its extreme positive value +1 corresponds to a peak of the wave moving through the element; then the value of y at position x is a. Its extreme negative value –1 corresponds to a valley of the wave moving through the element, then the value of y at position x is $-a$. Thus, the sine function and the time dependent phase of a wave correspond to the oscillation of a string element, and the amplitude of the wave determines the extremes of the element's displacement. The constant ϕ is called the **initial phase angle**. The value of ϕ is determined by the initial (t = 0) displacement and velocity of the element (say, at x = 0).

It is always possible to choose origin (x = 0) and the initial instant (t = 0) such that ϕ = 0. There is no loss of generality in working with Eq. (15.2) with ϕ = 0.

15.3.2 Wavelength and Angular Wave Number

The **wavelength** λ of a wave is the distance (parallel to the direction of wave propagation) between the consecutive repetitions of the shape of the wave. It is the minimum distance between two consecutive troughs or crests or two consecutive points in the same phase of wave motion. A typical wavelength is marked in Fig.15.6(a), which is a plot of Eq. (15.2) for $t = 0$ and $\phi = 0$. At this time Eq. (15.2) reduces to

$$y(x, 0) = a \sin kx \qquad (15.5)$$

By definition, the displacement y is same at both ends of this wavelength, that is at $x = x_1$ and at $x = x_1 + \lambda$. Thus, by Eq. (15.2),

$$a \sin k x_1 = a \sin k (x_1 + \lambda)$$
$$= a \sin (k x_1 + k \lambda)$$

This condition can be satisfied only when,

$$k \lambda = 2\pi n$$

where $n = 1, 2, 3...$ Since λ is defined as the least distance between points with the same phase, n =1 and

$$k \quad \frac{2}{} \qquad (15.6)$$

k is called the **propagation constant** or the **angular wave number** ; its SI unit is radian per metre or rad m^{-1}.*

It may be noted that in Fig. 15.6, as we move from one plot to another, the wave moves to the right by a distance equal to ¼ λ. Thus, by the fifth plot, it has moved to the right by a distance equal to λ.

15.3.3 Period, Angular Frequency and Frequency

Figure15.7 shows a graph of the displacement y, of Eq. (15.2), versus time t at a certain position along the string, taken to be x = 0. If you were to monitor the string, you would see that the

Fig. 15.7 A graph of the displacement of the string element at x = 0 as a function of time, as the sinusoidal wave of Fig. 14.6 passes through it. The amplitude a is indicated. A typical period T, measured from an arbitrary time t_1 is also indicated.

* *Here again, 'radian' could be dropped and the units could be written merely as m^{-1}. Thus, k represents 2π times the number of waves (or the total phase difference) that can be accommodated per unit length, with SI units m^{-1}.*

element of the string at that position moves up and down in simple harmonic motion given by Eq. (15.2) with $x = 0$,

$$y\,(0,t) = a \sin(-\omega t)$$
$$= -a \sin \omega t$$

Figure 15.7 is a graph of this equation; it does not show the shape of the wave.

The **period** of oscillation T of a wave is defined as the time any string element takes to move through one complete oscillation. A typical period is marked on the Fig. 15.7. Applying Eq. (15.2) on both ends of this time interval, we get

$$-a \sin \omega t_1 = -a \sin \omega(t_1 + T)$$
$$= -a \sin(\omega t_1 + \omega T)$$

This can be true only if the least value of ωT is 2π, or if

$$\omega = 2\pi/T \qquad (15.7)$$

ω is called the **angular frequency** of the wave, its SI unit is rad s^{-1}.

Look back at the five plots of a travelling wave in Fig. 15.6. The time between two consecutive plots is $T/4$. Thus, by the fifth plot, every string element has made one full oscillation.

The **frequency** ν of a wave is defined as $1/T$ and is related to the angular frequency ω by

$$\nu = \frac{1}{T} = \frac{\omega}{2\pi} \qquad (15.8)$$

It is the number of oscillations per unit time made by a string element as the wave passes through it. It is usually measured in hertz.

In the discussion above, reference has always been made to a wave travelling along a string or a transverse wave. In a longitudinal wave, the displacement of an element of the medium is parallel to the direction of propagation of the wave. In Eq. (15.2), the displacement function for a longitudinal wave is written as,

$$s(x, t) = a \sin(kx - \omega t + \phi) \qquad (15.9)$$

where $s(x, t)$ is the displacement of an element of the medium in the direction of propagation of the wave at position x and time t. In Eq. (15.9), a is the displacement amplitude; other quantities have the same meaning as in case of a transverse wave except that the displacement function $y(x, t)$ is to be replaced by the function $s(x, t)$.

Example 15.2 A wave travelling along a string is described by,

$y(x, t) = 0.005 \sin(80.0\,x - 3.0\,t)$,

in which the numerical constants are in SI units (0.005 m, 80.0 rad m^{-1}, and 3.0 rad s^{-1}). Calculate (a) the amplitude, (b) the wavelength, and (c) the period and frequency of the wave. Also, calculate the displacement y of the wave at a distance $x = 30.0$ cm and time $t = 20$ s ?

Answer On comparing this displacement equation with Eq. (15.2),

$$y\,(x, t) = a \sin(kx - \omega t),$$

we find

(a) the amplitude of the wave is 0.005 m = 5 mm.

(b) the angular wave number k and angular frequency ω are

$k = 80.0\ m^{-1}$ and $\omega = 3.0\ s^{-1}$

We then relate the wavelength λ to k through Eq. (15.6),

$$\lambda = 2\pi/k$$
$$= \frac{2\pi}{80.0\ m^{-1}}$$
$$= 7.85\ \text{cm}$$

(c) Now we relate T to ω by the relation

$$T = 2\pi/\omega$$
$$= \frac{2\pi}{3.0\ s^{-1}}$$
$$= 2.09\ \text{s}$$

and frequency, $\nu = 1/T = 0.48$ Hz

The displacement y at $x = 30.0$ cm and time $t = 20$ s is given by

$$y = (0.005\ \text{m}) \sin(80.0 \times 0.3 - 3.0 \times 20)$$
$$= (0.005\ \text{m}) \sin(-36 + 12\pi)$$
$$= (0.005\ \text{m}) \sin(1.699)$$
$$= (0.005\ \text{m}) \sin(97^0) \simeq 5\ \text{mm}$$

15.4 THE SPEED OF A TRAVELLING WAVE

Let us monitor the propagation of a travelling wave represented by Eq. (15.2) along a string. The wave is travelling in the positive direction of x. We find that an element of string at a particular position x moves up and down as a function of time but the waveform advances to

the right. The displacement of various elements of the string at two different instants of time t differing by a small time interval Δt is depicted in Fig. 15.8 (the phase angle ϕ has been taken to be zero). It is observed that during this interval of time the entire wave pattern moves by a distance Δx in the positive direction of x. Thus the wave is travelling to the right, in the positive direction of x. The ratio $\Delta x/\Delta t$ is the wave speed v.

Fig. 15.8 The plots of Eq.(15.2) at two instants of time differing by an interval Δt, at $t = 0$ and then at $t = \Delta t$. As the wave moves to the right at velocity **v**, *the entire curve shifts a distance Δx during Δt. The point A rides the waveform but the string element moves only up and down.*

As the wave moves (see Fig. 15.8), each point of the moving waveform represents a particular phase of the wave and retains its displacement y. It may be noted that the points on the string do not retain their displacement, but the points on the waveform do. Let us consider a point like A marked on a peak of the waveform. If a point like A on the waveform retains its displacement as it moves, it follows from Eq. (15.2) that this is possible only when the argument is constant. It, therefore, follows that

$$kx - \omega t = \text{constant} \qquad (15.10)$$

Note that in the argument both x and t are changing; therefore, to keep the argument constant, if t increases, x must also increase. This is possible only when the wave is moving in the positive direction of x.

To find the wave speed v, let us differentiate Eq. (15.10) with respect to time ;

$$\frac{d}{dt}\ kx \quad t \quad 0$$

or $$k\frac{dx}{dt} \quad 0$$

or $$\frac{dx}{dt} \quad \frac{}{k} \quad v \qquad (15.11)$$

Making use of Eqs. (15.6)-(15.8), we can write,

$$v = \frac{\omega}{k} = \frac{\lambda}{T} = \lambda\nu \qquad (15.12)$$

Equation (15.11) is a general relation valid for all progressive waves. It merely states that the wave moves a distance of one wavelength in one period of oscillation. The speed of a wave is related to its wavelength and frequency by the Eq. (15.12), **but it is determined by the properties of the medium.** If a wave is to travel in a medium like air, water, steel, or a stretched string, it must cause the particles of that medium to oscillate as it passes through it. For this to happen, the medium must possess mass and elasticity. Therefore the linear mass density (or mass per unit length, in case of linear systems like a stretched string) and the elastic properties determine how fast the wave can travel in the medium. Conversely, it should be possible to calculate the speed of the wave through the medium in terms of these properties. In subsequent sub-sections of this chapter, we will obtain specific expressions for the speed of mechanical waves in some media.

15.4.1 Speed of a Transverse Wave on Stretched String

The speed of transverse waves on a string is determined by two factors, (i) the linear mass density or mass per unit length, μ, and (ii) the tension T. The mass is required so that there is kinetic energy and without tension no disturbance can be propagated in the string. The exact derivation of the relationship between the speed of wave in a stretched string and the two parameters mentioned above is outside the scope of this book. However, we take recourse to a simpler procedure. In dimensional analysis, we have already learnt (Chapter 2) how to get a relationship between different quantities which are interrelated. Such a relation is, however, uncertain to the extent of a constant factor.

The linear mass density, μ, of a string is the mass m of the string divided by its length l. Therefore its dimension is $[ML^{-1}]$. The tension T

has the dimension of force — namely, $[M\,L\,T^{-2}]$. Our goal is to combine μ and T in such a way as to generate v [dimension $(L\,T^{-1})$]. If we examine the dimensions of these quantities, it can be seen that the ratio T/μ has the dimension

$$\frac{\left[MLT^{-2}\right]}{\left[ML^{-1}\right]} = \left[L^2\ T^{-2}\right]$$

Therefore, if v depends only on T and μ, the relation between them must be

$$v = C\sqrt{\frac{T}{\mu}} \qquad (15.13)$$

Here C is a dimensionless constant that cannot be determined by dimensional analysis. By adopting a more rigorous procedure it can be shown that the constant C is indeed equal to unity. The speed of transverse waves on a stretched string is, therefore, given by

$$v = \sqrt{\frac{T}{\mu}} \qquad (15.14)$$

Equation (15.14) tells us :

The speed of a wave along a stretched ideal string depends only on the tension and the linear mass density of the string and does not depend on the frequency of the wave.

The **frequency** of the wave is determined by the source that generates the wave. The **wavelength** is then fixed by Eq. (15.12) in the form,

$$\lambda = \frac{v}{\nu} \qquad (15.15)$$

Example 15.3 A steel wire 0.72 m long has a mass of 5.0×10^{-3} kg. If the wire is under a tension of 60 N, what is the speed of transverse waves on the wire ?

Answer Mass per unit length of the wire,

$$\mu = \frac{5.0\times10^{-3}\ \text{kg}}{0.72\ \text{m}}$$

$$= 6.9 \times 10^{-3}\ \text{kg m}^{-1}$$

Tension, $T = 60$ N

The speed of wave on the wire is given by

$$v = \sqrt{\frac{T}{\mu}} = \sqrt{\frac{60\,\text{N}}{6.9\times10^{-3}\,\text{kg m}^{-1}}} = 93\ \text{m s}^{-1}$$

15.4.2 Speed of a Longitudinal Wave Speed of Sound

In a longitudinal wave the constituents of the medium oscillate forward and backward in the direction of propagation of the wave. We have already seen that the sound waves travel in the form of compressions and rarefactions of small

Propagation of a pulse on a rope

You can easily see the motion of a pulse on a rope. You can also see its reflection from a rigid boundary and measure its velocity of travel. You will need a rope of diameter 1 to 3 cm, two hooks and some weights. You can perform this experiment in your classroom or laboratory.

Take a long rope or thick string of diameter 1 to 3 cm, and tie it to hooks on opposite walls in a hall or laboratory. Let one end pass on a hook and hang some weight (about 1 to 5 kg) to it. The walls may be about 3 to 5 m apart.

Take a stick or a rod and strike the rope hard at a point near one end. This creates a pulse on the rope which now travels on it. You can see it reaching the end and reflecting back from it. You can check the phase relation between the incident pulse and reflected pulse. You can easily watch two or three reflections before the pulse dies out. You can take a stopwatch and find the time for the pulse to travel the distance between the walls, and thus measure its velocity. Compare it with that obtained from Eq. (15.14).

This is also what happens with a thin metallic string of a musical instrument. The major difference is that the velocity on a string is fairly high because of low mass per unit length, as compared to that on a thick rope. The low velocity on a rope allows us to watch the motion and make measurements beautifully.

volume elements of air. The property that determines the extent to which the volume of an element of a medium changes when the pressure on it changes, is the **bulk modulus** B, defined (see Chapter 9) as,

$$B = -\frac{\Delta P}{\Delta V/V} \tag{15.16}$$

Here $\Delta V/V$ is the fractional change in volume produced by a change in pressure ΔP. The SI unit for pressure is N m^{-2} or pascal (Pa). Now since the longitudinal waves in a medium travel in the form of compressions and rarefactions or changes in density, the inertial property of the medium, which could be involved in the process, is the density ρ. The dimension of density is $[ML^{-3}]$. Thus, the dimension of the ratio B/ρ is,

$$\frac{\left[M\,L^{-1}\,T^{-2}\right]}{\left[M\,L^{-3}\right]} = \left[L^{2}\;\;T^{-2}\right] \tag{15.17}$$

Therefore, on the basis of dimensional analysis the most appropriate expression for the speed of longitudinal waves in a medium is

$$v = C\sqrt{\frac{B}{\rho}} \tag{15.18}$$

where C is a dimensionless constant and can be shown to be unity. Thus the speed of longitudinal waves in a medium is given by,

$$v = \sqrt{\frac{B}{\rho}} \tag{15.19}$$

The speed of propagation of a longitudinal wave in a fluid therefore depends only on the bulk modulus and the density of the medium.

When a solid bar is struck a blow at one end, the situation is somewhat different from that of a fluid confined in a tube or cylinder of constant cross section. For this case, the relevant modulus of elasticity is the Young's modulus, since the sideway expansion of the bar is negligible and only longitudinal strain needs to be considered. It can be shown that the speed of a longitudinal wave in the bar is given by,

$$v = \sqrt{\frac{Y}{\rho}} \tag{15.20}$$

where Y is the Young's modulus of the material of the bar.

Table 15.1 gives the speed of sound in various media.

Table 15.1 Speed of Sound in some Media

Medium	Speed (m s^{-1})
Gases	
Air (0 °C)	331
Air (20 °C)	343
Helium	965
Hydrogen	1284
Liquids	
Water (0 °C)	1402
Water (20 °C)	1482
Seawater	1522
Solids	
Aluminium	6420
Copper	3560
Steel	5941
Granite	6000
Vulcanised Rubber	54

It may be noted that although the densities of liquids and solids are much higher than those of the gases, the speed of sound in them is higher. It is because liquids and solids are less compressible than gases, i.e. have much greater bulk modulus.

In the case of an ideal gas, the relation between pressure P and volume V is given by (see Chapter 11)

$$PV = Nk_BT \tag{15.21}$$

where N is the number of molecules in volume V, k_B is the Boltzmann constant and T the temperature of the gas (in Kelvin). Therefore, for an isothermal change it follows from Eq.(15.21) that

$$V\Delta P + P\Delta V = 0$$

or $$-\frac{\Delta P}{\Delta V/V} = P$$

Hence, substituting in Eq. (15.16), we have

$$B = P$$

Therefore, from Eq. (15.19) the speed of a longitudinal wave in an ideal gas is given by,

$$v = \sqrt{\frac{P}{\rho}} \tag{15.22}$$

This relation was first given by Newton and is known as Newton's formula.

Example 15.4 Estimate the speed of sound in air at standard temperature and pressure. The mass of 1 mole of air is 29.0×10^{-3} kg.

Answer We know that 1 mole of any gas occupies 22.4 litres at STP. Therefore, density of air at STP is :

ρ_o = (mass of one mole of air)/ (volume of one mole of air at STP)

$$\frac{29.0 \quad 10^{3} \text{ kg}}{22.4 \quad 10^{3} \text{ m}^3}$$

$$= 1.29 \text{ kg m}^{-3}$$

According to Newton's formula for the speed of sound in a medium, we get for the speed of sound in air at STP,

$$v = \left[\frac{1.01 \times 10^5 \text{ N m}^{-2}}{1.29 \text{ kg m}^{-3}}\right]^{1/2} = 280 \text{ m s}^{-1} \quad (15.23)$$

The result shown in Eq.(15.23) is about 15% smaller as compared to the experimental value of 331 m s^{-1} as given in Table 15.1. Where did we go wrong ? If we examine the basic assumption made by Newton that the pressure variations in a medium during propagation of sound are isothermal, we find that this is not correct. It was pointed out by Laplace that the pressure variations in the propagation of sound waves are so fast that there is little time for the heat flow to maintain constant temperature. These variations, therefore, are adiabatic and not isothermal. For adiabatic processes the ideal gas satisfies the relation,

$$PV^{\gamma} = \text{constant}$$

i.e. $\Delta(PV^{\gamma}) = 0$

or $P\gamma V^{\gamma-1}\Delta V + V^{\gamma}\Delta P = 0$

Thus for an ideal gas the adiabatic bulk modulus is given by,

$$B_{ad} = -\frac{\Delta P}{\Delta V/V}$$
$$= \gamma P$$

where γ is the ratio of two specific heats, C_p/C_v. The speed of sound is, therefore, given by,

$$v = \sqrt{\frac{\gamma P}{\rho}} \quad (15.24)$$

This modification of Newton's formula is referred to as the **Laplace correction**. For air $\gamma = 7/5$. Now using Eq. (15.24) to estimate the speed of sound in air at STP, we get a value 331.3 m s^{-1}, which agrees with the measured speed.

15.5 THE PRINCIPLE OF SUPERPOSITION OF WAVES

Let us consider that two waves are travelling simultaneously along the same stretched string in opposite directions. The sequence of pictures shown in Fig. 15.9 depicts the state of displacement of various elements of the string at different time instant. Each picture depicts the resultant waveform in the string at a given instant of time. It is observed that the **net displacement of any element of the string at a given time is the algebraic sum of the displacements due to each wave**. This way of

Fig. 15.9 A sequence of pictures depicting two pulses travelling in opposite directions along a stretched string. They meet and pass through each other and move on independently as shown by the sequence of time snapshots (a) through (d). The total disturbance is the algebraic sum of the displacements due to each pulse. When the two disturbances overlap they give a complicated pattern as shown in (c). In region (d) they have passed each other and proceed unchanged.

addition of individual waveforms to determine the net waveform is called the **principle of superposition**. To put this rule in a mathematical form, let $y_1(x, t)$ and $y_2(x, t)$ be the displacements that any element of the string would experience if each wave travelled alone. The displacement $y(x,t)$ of an element of the string when the waves overlap is then given by,

$$y(x, t) = y_1(x, t) + y_2(x, t) \quad (15.25)$$

The principle of superposition can also be expressed by stating that **overlapping waves algebraically add to produce a resultant wave (or a net wave)**. The principle implies that the overlapping waves do not, in any way, alter the travel of each other.

If we have two or more waves moving in the medium the resultant waveform is the sum of wave functions of individual waves. That is, if the wave functions of the moving waves are

$$y_1 = f_1(x-vt),$$

$$y_2 = f_2(x-vt),$$

.....

.....

$$y_n = f_n(x-vt)$$

then the wave function describing the disturbance in the medium is

$$y = f_1(x - vt) + f_2(x - vt) + \ldots + f_n(x - vt)$$

$$= \sum_{i=1}^{n} f_i(x - vt) \quad (15.26)$$

As illustrative examples of this principle we shall study the phenomena of interference and reflection of waves.

Let a wave travelling along a stretched string be given by,

$$y_1(x, t) = a \sin(kx - \omega t) \quad (15.27)$$

and another wave, shifted from the first by a phase ϕ,

$$y_2(x, t) = a \sin(kx - \omega t + \phi) \quad (15.28)$$

Both the waves have the same angular frequency, same angular wave number k (same wavelength) and the same amplitude a. They travel in the positive direction of x-axis, with the same speed. Their phases at a given distance and time differ by a constant angle ϕ. These waves are said to be out of phase by ϕ or have a phase difference ϕ.

Now, applying the superposition principle, the resultant wave is the algebraic sum of the two constituent waves and has displacement

$$y(x, t) = a \sin(kx - \omega t) + a \sin(kx - \omega t + \phi) \quad (15.29)$$

We now use the trigonometric relation

$$\sin\alpha + \sin\beta = 2 \sin\frac{1}{2}(\alpha + \beta) \cos\frac{1}{2}(\alpha - \beta) \quad (15.30)$$

Applying this relation to Eq. (15.29) we have

$$y(x, t) = [2a \cos \frac{1}{2}\phi] \sin(kx - \omega t + \frac{1}{2}\phi) \quad (15.31)$$

Equation (15.31) shows that the resultant wave is also a sinusoidal wave, travelling in the positive direction of x-axis.

The resultant wave differs from the constituent waves in two respects: (1) its phase angle is $(½)\phi$ and (2) its amplitude is the quantity in brackets in Eq. (15.31) viz.,

$$A(\phi) = 2a \cos(½)\phi \quad (15.32)$$

If $\phi = 0$, i.e. the two waves are in phase, Eq. (15.31) reduces to

$$A(0) = 2a \sin(kx - \omega t) \quad (15.33)$$

The amplitude of the resultant wave is $2a$, which is the largest possible value of $A(\phi)$.

If $\phi = \pi$, the two waves are completely out of phase, the amplitude of the resultant wave given by Eq. (15.32) reduces to zero. We then have for all x and t,

$$y(x, t) = 0 \quad (15.34)$$

These cases are shown in Fig. 15.10.

15.6 REFLECTION OF WAVES

In previous sections we have discussed wave propagation in unbounded media. What happens when a pulse or a travelling wave encounters a rigid boundary ? It is a common experience that under such a situation the pulse or the wave gets reflected. An everyday example of the reflection of sound waves from a rigid boundary is the phenomenon of echo. If the boundary is not completely rigid or is an interface between two different elastic media, the effect of boundary conditions on

Fig. 15.10 Two identical sinusoidal waves, $y_1(x, t)$ and $y_2(x, t)$, travel along a stretched string in the positive direction of x-axis. They give rise to a resultant wave y (x, t). The phase difference between the two waves is (a) 0 and (b) π or 180°. The corresponding resultant waves are shown in (c) and (d).

an incident pulse or a wave is somewhat complicated. A part of the wave is reflected and a part is transmitted into the second medium. If a wave is incident obliquely on the boundary between two different media the transmitted wave is called the **refracted wave**. The incident and refracted waves obey Snell's law of refraction, and the incident and reflected waves obey the usual laws of reflection.

To illustrate the reflection of waves at a boundary, we consider two situations. First, a string is fixed to a rigid wall at its left end, as shown in Fig. 15.11(a). Second, the left end of the string is tied to a ring, which slides up and down without any friction on a rod, as shown in Fig. 15.11(b). A pulse is allowed to propagate in both these strings, the pulse, on reaching the left end, gets reflected; the state of disturbance in the string at various times is shown in Fig. 15.11.

In Fig. 15.11(a), the string is fixed to the wall at its left end. When the pulse arrives at that end, it exerts an upward force on the wall. By Newton's third law, the wall exerts an opposite force of equal magnitude on the string. This second force generates a pulse at the support (the wall), which travels back along the string in the direction opposite to that of the incident pulse. In a reflection of this kind, there must be no displacement at the support as the string is fixed there. The reflected and incident pulses must have opposite signs, so as to cancel each other at that point. Thus, in case of a travelling wave, the reflection at a rigid boundary will take place with a phase reversal or with a phase difference of π or 180°.

In Fig. 15.11(b), the string is fastened to a ring, which slides without friction on a rod. In this case, when the pulse arrives at the left end, the ring moves up the rod. As the ring moves, it pulls on the string, stretching the string and producing a reflected pulse with the same sign and amplitude as the incident pulse. Thus, in such a reflection, the incident and reflected pulses reinforce each other, creating the

Fig. 15.11 (a) A pulse incident from the right is reflected at the left end of the string, which is tied to a wall. Note that the reflected pulse is inverted from the incident pulse. (b) Here the left end is tied to a ring that can slide up and down without friction on the rod. Now the reflected pulse is not inverted by reflection.

maximum displacement at the end of the string: the maximum displacement of the ring is twice the amplitude of either of the pulses. Thus, the reflection is without any additional phase shift. In case of a travelling wave the reflection at an open boundary, such as the open end of an organ pipe, the reflection takes place without any phase change.

We can thus, summarise the reflection of waves at a boundary or interface between two media as follows:

A travelling wave, at a rigid boundary or a closed end, is reflected with a phase reversal but the reflection at an open boundary takes place without any phase change.

To express the above statement mathematically, let the incident wave be represented by

$$y_i(x, t) = a \sin (kx - \omega t),$$

then, for reflection at a rigid boundary the reflected wave is represented by,

$$y_r(x, t) = a \sin (kx + \omega t + \pi).$$

$$= - a \sin (kx + \omega t) \qquad (15.35)$$

For reflection at an open boundary, the reflected wave is represented by

$$y_r (x, t) = a \sin (kx + \omega t). \qquad (15.36)$$

15.6.1 Standing Waves and Normal Modes

In the previous section we have considered a system which is bounded at one end. Let us now consider a system which is bounded at both the ends such as a stretched string fixed at the ends or an air column of finite length. In such a system suppose that we send a continuous sinusoidal wave of a certain frequency, say, toward the right. When the wave reaches the right end, it gets reflected and begins to travel back. The left-going wave then overlaps the wave, travelling to the right. When the left-going wave reaches the left end, it gets reflected again and the newly reflected wave begins to travel to the right, overlapping the left-going wave. This process will continue and, therefore, very soon we have many overlapping waves, which interfere with one another. In such a system, at any point x and at any time t, there are always two waves, one moving to the left and another to the right. We, therefore, have

$y_1(x, t) = a \sin (kx - \omega t)$ (wave travelling in the positive direction of x-axis)

and $y_2(x, t) = a \sin (kx + \omega t)$ (wave travelling in the negative direction of x-axis).

The principle of superposition gives, for the combined wave

$$y (x, t) = y_1(x, t) + y_2(x, t)$$

$$= a \sin (kx - \omega t) + a \sin (kx + \omega t)$$

$$= (2a \sin kx) \cos \omega t \qquad (15.37)$$

The wave represented by Eq. (15.37) does not describe a travelling wave, as the waveform or the disturbance does not move to either side. Here, the quantity $2a \sin kx$ within the brackets is the amplitude of oscillation of the element of the string located at the position x. In a travelling wave, in contrast, the amplitude of the wave is the same for all elements. Equation (15.37), therefore, represents a **standing wave**, a wave in which the waveform does not move. The formation of such waves is illustrated in Fig. 15.12.

It is seen that the points of maximum or minimum amplitude stay at one position.

The amplitude is zero for values of kx that give $\sin kx = 0$. Those values are given by

$$kx = n\pi, \text{ for } n = 0, 1, 2, 3, \ldots$$

Substituting $k = 2\pi/\lambda$ in this equation, we get

$$x = n \frac{\lambda}{2}, \text{ for } n = 0, 1, 2, 3, \ldots \qquad (15.38)$$

The positions of zero amplitude are called **nodes**. Note that a distance of $\frac{\lambda}{2}$ or half a wavelength separates two consecutive nodes.

The amplitude has a maximum value of $2a$, which occurs for the values of kx that give $|\sin kx| = 1$. Those values are

$$kx = (n + ½) \pi \text{ for } n = 0, 1, 2, 3, \ldots$$

Substituting $k = 2\pi/\lambda$ in this equation, we get

$$x = (n + ½) \frac{\lambda}{2} \text{ for } n = 0, 1, 2, 3, \ldots \qquad (15.39)$$

as the positions of maximum amplitude. These are called the **antinodes**. The antinodes are separated by $\lambda/2$ and are located half way between pairs of nodes.

For a stretched string of length L, fixed at both ends, the two ends of the string have to be nodes.

Fig. 15.12 The formation of a standing wave in a stretched string. Two sinusoidal waves of same amplitude travel along the string in opposite directions. The set of pictures represent the state of displacements at four different times. The displacement at positions marked as N is zero at all times. These positions are called nodes.

If one of the ends is chosen as position $x = 0$, then the other end is $x = L$. In order that this end is a node; the length L must satisfy the condition

$$L = n\,\frac{\lambda}{2}, \text{ for } n = 1, 2, 3, \ldots \qquad (15.40)$$

This condition shows that standing waves on a string of length L have restricted wavelength given by

$$\lambda = \frac{2L}{n}, \text{ for } n = 1, 2, 3, \ldots \text{ etc.} \qquad (15.41)$$

The frequencies corresponding to these wavelengths follow from Eq. (15.12) as

$$\nu = n\,\frac{v}{2L}, \text{ for } n = 1, 2, 3, \ldots \text{ etc.} \qquad (15.42)$$

where v is the speed of travelling waves on the string. The set of frequencies given by Eq. (15.42) are called the natural frequencies or **modes** of oscillation of the system. This equation tells us that the natural frequencies of a string are integral multiples of the lowest frequency $\nu = \frac{v}{2L}$, which corresponds to $n = 1$. The oscillation mode with that lowest frequency is called the **fundamental mode** or the **first harmonic**. The **second harmonic** is the oscillation mode with $n = 2$. The third harmonic corresponds to $n = 3$ and so on. The frequencies associated with these modes are often labelled as ν_1, ν_2, ν_3 and so on. The collection of all possible modes is called the **harmonic series** and n is called the harmonic number.

Some of the harmonics of a stretched string fixed at both the ends are shown in Fig. 15.13. According to the principle of superposition, a stretched string tied at both ends can vibrate simultaneously in more than one modes. Which mode is strongly excited depends on where the

string is plucked or bowed. Musical instruments like sitar and violin are designed on this principle.

We now study the modes of vibration of a system closed at one end, with the other end being free. Air columns such as glass tubes partially filled with water provide examples of such systems. In these, the length of the air column can be adjusted by changing the water level in the tube. In such systems, the end of the air column in touch with the water suffers no displacement as the reflected and incident waves are exactly out of phase. For this reason the pressure changes here are the largest, since when the compressional part is reflected the pressure increase is doubled, and when the rarefaction is reflected the decrease in pressure is doubled. On the other hand, at the open end, there is maximum displacement and minimum pressure change. The two waves travelling in opposite directions are in phase here, so there are no pressure changes. Now if the length of the air column is L, then the open end, $x = L$, is an antinode and therefore, it follows from Eq. (15.39) that

Fig. 15.13 Stationary waves in a stretched string fixed at both ends. Various modes of vibration are shown.

$$L = \left(n + \frac{1}{2}\right)\frac{\lambda}{2}, \text{ for } n = 0, 1, 2, 3, \ldots$$

The modes, which satisfy the condition

$$\lambda = \frac{2L}{(n + 1/2)}, \text{ for } n = 0, 1, 2, 3, \ldots \qquad (15.43)$$

are sustained in such an air column. The corresponding frequencies of various modes of such an air column are given by,

$$\nu = \left(n + \frac{1}{2}\right)\frac{v}{2L}, \text{ for } n = 0, 1, 2, 3, \ldots \qquad (15.44)$$

Some of the normal modes in an air column with the open end are shown in Fig. 15.14. The fundamental frequency is $\frac{v}{4L}$ and the higher frequencies are **odd harmonics** of the fundamental frequency, i.e. $3\frac{v}{4L}$, $5\frac{v}{4L}$, etc.

In the case of a pipe open at both ends, there will be antinodes at both ends, and **all harmonics** will be generated.

Normal modes of a circular membrane rigidly clamped to the circumference as in a tabla are determined by the boundary condition that no

point on the circumference of the membrane vibrates. Estimation of the frequencies of normal modes of this system is more complex. This problem involves wave propagation in two dimensions. However, the underlying physics is the same.

We have seen above that in a string, fixed at both ends, standing waves are produced only at certain frequencies as given by Eq. (15.42) or the system **resonates** at these frequencies. Similarly an air column open at one end resonates at frequencies given by Eq. (15.44).

> ***Example 15.5*** A pipe, 30.0 cm long, is open at both ends. Which harmonic mode of the pipe rasonates a 1.1 kHz source? Will resonance with the same source be observed if one end of the pipe is closed ? Take the speed of sound in air as 330 m s^{-1}.

Answer The first harmonic frequency is given by

$$\nu_1 = \frac{v}{\lambda_1} = \frac{v}{2L} \qquad \text{(open pipe)}$$

where L is the length of the pipe. The frequency of its nth harmonic is:

$$\nu_n = \frac{nv}{2L}, \text{ for } n = 1, 2, 3, \ldots \text{ (open pipe)}$$

First few modes of an open pipe are shown in Fig. 15.14.

Fig. 15.14 *Some of the normal modes of vibration of an air column open at one end.*

For L = 30.0 cm, v = 330 m s^{-1},

$$\nu_n = \frac{n \times 330 \text{ (m s}^{-1})}{0.6 \text{ (m)}} = 550\, n \text{ s}^{-1}$$

Clearly, a source of frequency 1.1 kHz will resonate at ν_2, i.e. the **second harmonic**.

Now if one end of the pipe is closed (Fig. 15.15), it follows from Eq. (14.50) that the fundamental frequency is

$$\nu_1 = \frac{v}{\lambda_1} = \frac{v}{4L} \text{ (pipe closed at one end)}$$

and only the odd numbered harmonics are present :

$$\nu_3 = \frac{3v}{4L}, \ \nu_5 = \frac{5v}{4L}, \text{ and so on.}$$

For L = 30 cm and v = 330 m s^{-1}, the fundamental frequency of the pipe closed at one end is 275 Hz and the source frequency corresponds to its fourth harmonic. Since this harmonic is not a possible mode, no resonance will be observed with the source, the moment one end is closed.

15.7 BEATS

If we listen, a few minutes apart, two sounds of very close frequencies, say 256 Hz and 260 Hz, we will not be able to discriminate between

Fig. 15.15 Standing waves in an open pipe, first four harmonics are depicted.

them. However, if both these sounds reach our ears simultaneously, what we hear is a sound of frequency 258 Hz, the **average** of the two combining frequencies. In addition we hear a striking variation in the intensity of sound — it increases and decreases in slow, wavering **beats** that repeat at a frequency of 4 Hz, the **difference** between the frequencies of two incoming sounds. The phenomenon of wavering of sound intensity when two waves of nearly same frequencies and amplitudes travelling in the same direction, are superimposed on each other is called **beats**.

Let us find out what happens when two waves having slightly different frequencies are superposed on each other. Let the time dependent variations of the displacements due to two sound waves at a particular location be

$$s_1 = a \cos \omega_1 t \quad \text{and} \quad s_2 = a \cos \omega_2 t \qquad (15.45)$$

where $\omega_1 > \omega_2$. We have assumed, for simplicity, that the waves have same amplitude and phase. According to the superposition principle, the resultant displacement is

$$s = s_1 + s_2 = a\,(\cos \omega_1 t + \cos \omega_2 t)$$

$$= 2\,a \cos \frac{(\omega_1 - \omega_2)t}{2} \cos \frac{(\omega_1 + \omega_2)t}{2} \qquad (15.46)$$

If we write $\omega_b = \dfrac{(\omega_1 - \omega_2)}{2}$ and $\omega_a = \dfrac{(\omega_1 + \omega_2)}{2}$

then Eq. (15.46) can be written as

$$s = [2\,a \cos \omega_b t] \cos \omega_a t \qquad (15.47)$$

If $|\omega_1 - \omega_2| << \omega_1, \omega_2, \omega_a >> \omega_b$, then in Eq. (15.47) the main time dependence arises from cosine function whose angular frequency is ω_a. The quantity in the brackets can be regarded as the amplitude of this function (which is not a constant but, has a small variation of angular frequency ω_b). It becomes maximum whenever $\cos \omega_b t$ has the value +1 or –1, which happens twice in each repetition of cosine function. Since ω_1 and ω_2 are very close, ω_a cannot be differentiated easily from either of them. Thus, the result of superposition of two waves having nearly the same

Musical Pillars

Temples often have some pillars portraying human figures playing musical instruments, but seldom do these pillars themselves produce music. At the Nellaiappar temple in Tamil Nadu, gentle taps on a cluster of pillars carved out of a single piece of rock produce the basic notes of Indian classical music, viz. Sa, Re, Ga, Ma, Pa, Dha, Ni, Sa. Vibrations of these pillars depend on elasticity of the stone used, its density and shape.

Musical pillars are categorised into three types: The first is called the **Shruti Pillar**, as it can produce the basic notes — the "swaras". The second type is the **Gana Thoongal**, which generates the basic tunes that make up the "ragas". The third variety is the **Laya Thoongal** pillars that produce "taal" (beats) when tapped. The pillars at the Nellaiappar temple are a combination of the Shruti and Laya types.

Archaeologists date the Nelliappar temple to the 7th century and claim it was built by successive rulers of the Pandyan dynasty.

The musical pillars of Nelliappar and several other temples in southern India like those at Hampi (picture), Kanyakumari, and Thiruvananthapuram are unique to the country and have no parallel in any other part of the world.

Reflection of sound in an open pipe

When a high pressure pulse of air traveling down an open pipe reaches the other end, its momentum drags the air out into the open, where pressure falls rapidly to the atmospheric pressure. As a result the air following after it in the tube is pushed out. The low pressure at the end of the tube draws air from further up the tube. The air gets drawn towards the open end forcing the low pressure region to move upwards. As a result a pulse of *high* pressure air travelling *down* the tube turns into a pulse of *low* pressure air travelling *up* the tube. We say a pressure wave has been reflected at the open end with a change in phase of 180^0. Standing waves in an open pipe organ like the flute is a result of this phenomenon.

Compare this with what happens when a pulse of high pressure air arrives at a closed end: it collides and as a result pushes the air back in the opposite direction. Here, we say that the pressure wave is reflected, with no change in phase.

frequencies is a wave with nearly same angular frequency but its amplitude is not constant. Thus the intensity of resultant sound varies with an angular frequency $\omega_{beat} = 2\,\omega_b = \omega_1 - \omega_2$. Now using the relation,

$$\omega = 2\pi\nu$$

the beat frequency, ν_{beat}, is given by

$$\nu_{beat} = \nu_1 - \nu_2 \qquad (15.48)$$

Thus we hear a waxing and waning of sound with a frequency equal to the difference between the frequencies of the superposing waves. The time-displacement graphs of two waves of frequency 11 Hz and 9 Hz is shown in Figs. 15.16(a) and 15.16(b). The result of their 'superposition' is shown in Fig. 15.16(c).

Fig. 15.16 *(a) Plot of a harmonic wave of frequency 11 Hz. (b) Plot of a harmonic wave of frequency 9 Hz. (c) Superposition of (a) and (b), showing clearly the beats in the slow (2 Hz) of the total disturbance.*

Musicians use the beat phenomenon in tuning their instruments. If an instrument is sounded against a standard frequency and tuned until the beat disappears, then the instrument is in tune with that standard.

▶ ***Example 15.6*** Two sitar strings A and B playing the note '*Dha*' are slightly out of tune and produce beats of frequency 5 Hz. The tension of the string B is slightly increased and the beat frequency is found to decrease to 3 Hz. What is the original frequency of B if the frequency of A is 427 Hz ?

Answer Increase in the tension of a string increases its frequency. If the original frequency of B (ν_B) were greater than that of A (ν_A), further increase in ν_B should have resulted in an increase in the beat frequency. But the beat frequency is found to decrease. This shows that $\nu_B < \nu_A$. Since $\nu_A - \nu_B = 5$ Hz, and $\nu_A = 427$ Hz, we get $\nu_B = 422$ Hz. ◀

15.8 DOPPLER EFFECT

It is an everyday experience that the pitch (or frequency) of the whistle of a fast moving train

decreases as it recedes away. When we approach a stationary source of sound with high speed, the pitch of the sound heard appears to be higher than that of the source. As the observer recedes away from the source, the observed **pitch** (or frequency) becomes lower than that of the source. This motion-related frequency change is called **Doppler effect**. The Austrian physicist Johann Christian Doppler first proposed the effect in 1842. Buys Ballot in Holland tested it experimentally in 1845. Doppler effect is a wave phenomenon, it holds not only for sound waves but also for electromagnetic waves. However, here we shall consider only sound waves.

We shall analyse changes in frequency under three different situations: (1) observer is stationary but the source is moving, (2) observer is moving but the source is stationary, and (3) both the observer and the source are moving. The situations (1) and (2) differ from each other because of the absence or presence of relative motion between the observer and the medium. Most waves require a medium for their propagation; however, electromagnetic waves do not require any medium for propagation. If there is no medium present, the Doppler shifts are same irrespective of whether the source moves or the observer moves, since there is no way of distinction between the two situations.

15.8.1 Source Moving ; Observer Stationary

Let us choose the convention to take the direction from the observer to the source as the positive direction of velocity. Consider a source S moving with velocity v_s and an observer who is stationary in a frame in which the medium is also at rest. Let the speed of a wave of angular frequency ω and period T_o, both measured by an observer at rest with respect to the medium, be v. We assume that the observer has a detector that counts every time a wave crest reaches it. As shown in Fig. 15.17, at time $t = 0$ the source is at point S_1, located at a distance L from the observer, and emits a crest. This reaches the observer at time $t_1 = L/v$. At time $t = T_o$ the source has moved a distance v_sT_o and is at point S_2, located at a distance $(L + v_sT_o)$ from the observer. At S_2, the source emits a second crest. This reaches the observer at

$$t_2 = T_0 + \frac{(L + v_s T_0)}{v}$$

Fig. 15.17 A source moving with velocity v_s emits a wave crest at the point S_1. It emits the next wave crest at S_2 after moving a distance v_sT_o.

At time $n\,T_o$, the source emits its $(n+1)^{th}$ crest and this reaches the observer at time

$$t_{n+1} = n\,T_0 + \frac{L + n\,v_s T_0}{v}$$

Hence, in a time interval

$$nT_0 + \frac{L + nv_s T_0}{v} - \frac{L}{v}$$

the observer's detector counts n crests and the observer records the period of the wave as T given by

$$T = \left[nT_0 + \frac{L + nv_s T_0}{v} - \frac{L}{v}\right]/n$$

$$= T_0 + \frac{v_s T_0}{v}$$

$$= T_0\left(1 + \frac{v_s}{v}\right) \qquad (15.49)$$

Equation (15.49) may be rewritten in terms of the frequency ν_o that would be measured if the source and observer were stationary, and the frequency ν observed when the source is moving, as

$$\nu = \nu_0\left(1 + \frac{v_s}{v}\right)^{-1} \qquad (15.50)$$

If v_s is small compared with the wave speed v, taking binomial expansion to terms in first order in v_s/v and neglecting higher power, Eq. (15.50) may be approximated, giving

$$\nu = \nu_0\left(1 - \frac{v_s}{v}\right) \qquad (15.51)$$

For a source approaching the observer, we replace v_s by $-v_s$ to get

$$\nu = \nu_0\left(1+\frac{v_s}{v}\right) \qquad (15.52)$$

The observer thus measures a lower frequency when the source recedes from him than he does when it is at rest. He measures a higher frequency when the source approaches him.

15.8.2 Observer Moving; Source Stationary

Now to derive the Doppler shift when the observer is moving with velocity v_o towards the source and the source is at rest, we have to proceed in a different manner. We work in the reference frame of the moving observer. In this reference frame the source and medium are approaching at speed v_o and the speed with which the wave approaches is $v_o + v$. Following a similar procedure as in the previous case, we find that the time interval between the arrival of the first and the $(n+1)$ th crests is

$$t_{n+1} - t_1 = n\,T_0 - \frac{nv_0T_0}{v_0+v}$$

The observer thus, measures the period of the wave to be

$$= T_0\left(1-\frac{v_0}{v_0+v}\right)$$

$$T_0 \quad 1 \quad \frac{v_0}{v}^{-1}$$

giving

$$\nu = {}_0\left(1+\frac{v_0}{v}\right) \qquad (15.53)$$

If $\frac{v_0}{v}$ is small, the Doppler shift is almost same whether it is the observer or the source moving since Eq. (15.53) and the approximate relation Eq. (15.51) are the same.

15.8.3 Both Source and Observer Moving

We will now derive a general expression for Doppler shift when both the source and the observer are moving. As before, let us take the direction from the observer to the source as the positive direction. Let the source and the observer be moving with velocities v_s and v_o respectively as shown in Fig.15.18. Suppose at time $t = 0$, the observer is at O_1 and the source is at S_1, O_1 being to the left of S_1. The source emits a wave of velocity v, of frequency ν and period T_0 all measured by an observer at rest with respect to the medium. Let L be the distance between O_1 and S_1 at $t = 0$, when the source emits the first crest. Now, since the observer is moving, the velocity of the wave relative to the observer is $v+v_0$. Therfore the first crest reaches the observer at time $t_1 = L/(v+v_0)$. At time $t = T_0$, both the observer and the source have moved to their new positions O_2 and S_2 respectively. The new distance between the observer and the source, $O_2\,S_2$, would be $L+(v_s-v_0)\,T_0]$. At S_2, the source emits a second crest. This reaches the observer at time.

$$t_2 = T_o + [L + (v_s - v_o)T_o)]\ /(v + v_o)$$

At time nT_o the source emits its $(n+1)$ th crest and this reaches the observer at time

$$t_{n+1} = nT_o + [L + n\,(v_s - v_o)T_o)]\ /(v + v_o)$$

Hence, in a time interval $t_{n+1} - t_1$, i.e.,

$$nT_o + [L + n\,(v_s - v_o)T_o)]\ /(v + v_o) - L\ /(v + v_o),$$

the observer counts n crests and the observer records the period of the wave as equal to T given by

Application of Doppler effect

The change in frequency caused by a moving object due to Doppler effect is used to measure their velocities in diverse areas such as military, medical science, astrophysics, etc. It is also used by police to check over-speeding of vehicles.

A sound wave or electromagnetic wave of known frequency is sent towards a moving object. Some part of the wave is reflected from the object and its frequency is detected by the monitoring station. This change in frequency is called **Doppler shift**.

It is used at airports to guide aircraft, and in the military to detect enemy aircraft. Astrophysicists use it to measure the velocities of stars.

Doctors use it to study heart beats and blood flow in different part of the body. Here they use ulltrasonic waves, and in common practice, it is called **sonography**. Ultrasonic waves enter the body of the person, some of them are reflected back, and give information about motion of blood and pulsation of heart valves, as well as pulsation of the heart of the foetus. In the case of heart, the picture generated is called **echocardiogram**.

$$T = T_0\left(1+\frac{\upsilon_s - \upsilon_o}{\upsilon + \upsilon_0}\right) = T_0\left(\frac{\upsilon+\upsilon_s}{\upsilon+\upsilon_0}\right) \quad (15.54)$$

The frequency ν observed by the observer is given by

$$\nu = \nu_0\left(\frac{\upsilon+\upsilon_0}{\upsilon+\upsilon_s}\right) \quad (15.55)$$

Consider a passenger sitting in a train moving on a straight track. Suppose she hears a whistle sounded by the driver of the train. What frequency will she measure or hear? Here both the observer and the source are moving with the same velocity, so there will be no shift in frequency and the passanger will note the natural frequency. But an observer outside who is stationary with respect to the track will note a higher frequency if the train is approaching him and a lower frequency when it recedes from him.

Note that we have defined the direction from the observer to the source as the positive direction. Therefore, if the observer is moving towards the source, υ_0 has a positive (numerical) value whereas if O is moving away from S, υ_0 has a negative value. On the other hand, if S is moving away from O, υ_s has a positive value whereas if it is moving towards O, υ_s has a negative value. The sound emitted by the source travels in all directions. It is that part of sound coming towards the observer which the observer receives and detects. Therefore the relative velocity of sound with respect to the observer is $\upsilon + \upsilon_0$ in all cases.

Fig. 15.18 *The observer O and the source S, both moving respectively with velocities υ_0 and υ_s. They are at position O_1 and S_1 at time t = 0, when the source emits the first crest of a sound, whose velocity is υ with respect to the medium. After one period, $t = T_0$, they have moved to O_2 and S_2, respectively through distances $\upsilon_0 T_0$ and $\upsilon_s T_0$, when the source emits the next crest.*

Example 15.7 A rocket is moving at a speed of 200 m s^{-1} towards a stationary target. While moving, it emits a wave of frequency 1000 Hz. Some of the sound reaching the target gets reflected back to the rocket as an echo. Calculate (1) the frequency of the sound as detected by the target and (2) the frequency of the echo as detected by the rocket.

Answer (1) The observer is at rest and the source is moving with a speed of 200 m s^{-1}. Since this is comparable with the velocity of sound, 330 m s^{-1}, we must use Eq. (15.50) and not the approximate Eq. (15.51). Since the source is approaching a stationary target, $\upsilon_o = 0$, and υ_s must be replaced by $-\upsilon_s$. Thus, we have

$$\nu = \nu_0\left(1-\frac{\upsilon_s}{\upsilon}\right)^{-1}$$

$$\nu = 1000 \text{ Hz} \times [1 - 200 \text{ m s}^{-1}/330 \text{ m s}^{-1}]^{-1}$$

$$\simeq 2540 \text{ Hz}$$

(2) The target is now the source (because it is the source of echo) and the rocket's detector is now the detector or observer (because it detects echo). Thus, $\upsilon_s = 0$ and υ_o has a positive value. The frequency of the sound emitted by the source (the target) is ν, the frequency intercepted by the target and not ν_o. Therefore, the frequency as registered by the rocket is

$$\nu' = \nu\left(\frac{\upsilon+\upsilon_0}{\upsilon}\right)$$

$$= 2540 \text{ Hz} \times \frac{200 \text{ m s}^{-1} + 330 \text{ m s}^{-1}}{330 \text{ m s}^{-1}}$$

$$\simeq 4080 \text{ Hz}$$

SUMMARY

1. *Mechanical waves* can exist in material media and are governed by Newton's Laws.
2. *Transverse waves* are waves in which the particles of the medium oscillate perpendicular to the direction of wave propagation.
3. *Longitudinal waves* are waves in which the particles of the medium oscillate along the direction of wave propagation.
4. *Progressive wave* is a wave that moves from one point of medium to another.
5. *The displacement* in a sinusoidal wave propagating in the positive x direction is given by

$$y\,(x, t) = a \sin\,(kx - \omega t + \phi)$$

where a *is the* amplitude of the wave, k is the angular wave number, ω *is the* angular frequency, $(kx - \omega t + \phi)$ *is the* phase, *and* ϕ *is the* phase constant or phase angle.

6. *Wavelength* λ of a progressive wave is the distance between two consecutive points of the same phase at a given time. In a stationary wave, it is twice the distance between two consecutive nodes or anti nodes.
7. *Period* T of oscillation of a wave is defined as the time any element of the medium takes to move through one complete oscillation. It is related to the *angular frequency* ω through the relation

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

8. *Frequency* ν of a wave is defined as $1/T$ and is related to angular frequency by

$$f = \frac{\omega}{2\pi}$$

9. *Speed* of a progressive wave is given by $\upsilon \quad \overline{\mathrm{k}} \quad \overline{\mathrm{T}}$
10. *The speed of a transverse wave* on a stretched string is set by the properties of the string. The speed on a string with tension T and linear mass density μ is

$$\upsilon = \sqrt{\frac{T}{\mu}}$$

11. *Sound waves* are longitudinal mechanical waves that can travel through solids, liquids, or gases. The speed υ of sound wave in a fluid having *bulk modulus* B and density ρ is

$$\upsilon = \sqrt{\frac{B}{\rho}}$$

The speed of longitudinal waves in a metallic bar is

$$\upsilon = \sqrt{\frac{Y}{\rho}}$$

For gases, since $B = \gamma P$, the speed of sound is

$$\upsilon = \sqrt{\frac{\gamma P}{\rho}}$$

12. When two or more waves traverse the same medium, the displacement of any element of the medium is the algebraic sum of the displacements due to each wave. This is known as the *principle of superposition* of waves

$$y = \sum_{i=1}^{n} f_i(x - vt)$$

13. Two sinusoidal waves on the same string exhibit *interference*, adding or cancelling according to the principle of superposition. If the two are travelling in the same direction and have the same amplitude a and frequency but differ in phase by a *phase constant* ϕ, the result is a single wave with the same frequency ω:

$$y\ (x,\ t) \ = \ 2a\cos\frac{1}{2} \quad \sin\ kx \quad t \quad \frac{1}{2}$$

If $\phi = 0$ or an integral multiple of 2π, the waves are exactly in phase and the interference is constructive; if $\phi = \pi$, they are exactly out of phase and the interference is destructive.

14. A travelling wave, at a rigid boundary or a closed end, is reflected with a phase reversal but the reflection at an open boundary takes place without any phase change.

For an incident wave

$$y_i\ (x,\ t) = a \sin\ (kx - \omega t\ )$$

the reflected wave at a rigid boundary is

$$y_r\ (x,\ t) = -\ a \sin\ (kx + \omega t\ )$$

For reflection at an open boundary

$$y_r\ (x,t\ ) = a \sin\ (kx + \omega t)$$

15. The interference of two identical waves moving in opposite directions produces *standing waves*. For a string with fixed ends, the standing wave is given by

$$y\ (x,\ t) = [2a \sin\ kx\ ] \cos\ \omega t$$

Standing waves are characterised by fixed locations of zero displacement called *nodes* and fixed locations of maximum displacements called *antinodes*. The separation between two consecutive nodes or antinodes is $\lambda/2$.

A stretched string of length L fixed at both the ends vibrates with frequencies given by

$$\nu \ = \frac{1}{2}\ \frac{v}{2L}, \qquad n = 1,\ 2,\ 3,\ \ldots$$

The set of frequencies given by the above relation are called the *normal modes* of oscillation of the system. The oscillation mode with lowest frequency is called the *fundamental mode* or the *first harmonic*. The *second harmonic* is the oscillation mode with $n = 2$ and so on.

A pipe of length L with one end closed and other end open (such as air columns) vibrates with frequencies given by

$$\nu \quad n \quad ½ \quad \frac{v}{2L}, \qquad n = 0,\ 1,\ 2,\ 3,\ \ldots$$

The set of frequencies represented by the above relation are the *normal modes* of oscillation of such a system. The lowest frequency given by $v/4L$ is the fundamental mode or the first harmonic.

16. A string of length L fixed at both ends or an air column closed at one end and open at the other end, vibrates with frequencies called its normal modes. Each of these frequencies is a *resonant frequency* of the system.

17. *Beats* arise when two waves having slightly different frequencies, ν_1 and ν_2 and comparable amplitudes, are superposed. The beat frequency is

$$\nu_{beat} = \nu_1 \sim \nu_2$$

18. The *Doppler effect* is a change in the observed frequency of a wave when the source and the observer O moves relative to the medium. For sound the observed frequency ν is given in terms of the source frequency ν_o by

$$\nu = \nu_o \left(\frac{\upsilon + \upsilon_0}{\upsilon + \upsilon_s} \right)$$

here υ is the speed of sound through the medium, υ_o is the velocity of observer relative to the medium, and υ_s is the source velocity relative to the medium. In using this formula, velocities in the direction OS should be treated as positive and those opposite to it should be taken to be negative.

Physical quantity	Symbol	Dimensions	Unit	Remarks
Wavelength	λ	[L]	m	Distance between two consecutive points with the same phase.
Propagation constant	k	$[L^{-1}]$	m^{-1}	$k = \frac{2\pi}{\lambda}$
Wave speed	υ	$[LT^{-1}]$	$m\ s^{-1}$	$\nu = \upsilon\lambda$
Beat frequency	ν_{beat}	$[T^{-1}]$	s^{-1}	Difference of two close frequencies of superposing waves.

POINTS TO PONDER

1. A wave is not motion of matter as a whole in a medium. A wind is different from the sound wave in air. The former involves motion of air from one place to the other. The latter involves compressions and rarefactions of layers of air.
2. In a wave, energy and *not the matter* is transferred from one point to the other.
3. Energy transfer takes place because of the coupling through elastic forces between neighbouring oscillating parts of the medium.
4. Transverse waves can propagate only in medium with shear modulus of elasticity, Longitudinal waves need bulk modulus of elasticity and are therefore, possible in all media, solids, liquids and gases.
5. In a harmonic progressive wave of a given frequency all particles have the same amplitude but different phases at a given instant of time. In a stationary wave, all particles between two nodes have the same phase at a given instant but have different amplitudes.
6. Relative to an observer at rest in a medium the speed of a mechanical wave in that medium (υ) depends only on elastic and other properties (such as mass density) of the medium. It does not depend on the velocity of the source.
7. For an observer moving with velocity υ_o relative to the medium, the speed of a wave is obviously different from υ and is given by $\upsilon \pm \upsilon_o$.

EXERCISES

15.1 A string of mass 2.50 kg is under a tension of 200 N. The length of the stretched string is 20.0 m. If the transverse jerk is struck at one end of the string, how long does the disturbance take to reach the other end?

15.2 A stone dropped from the top of a tower of height 300 m high splashes into the water of a pond near the base of the tower. When is the splash heard at the top given that the speed of sound in air is 340 m s^{-1}? (g = 9.8 m s^{-2})

15.3 A steel wire has a length of 12.0 m and a mass of 2.10 kg. What should be the tension in the wire so that speed of a transverse wave on the wire equals the speed of sound in dry air at 20°C = 343 m s^{-1}.

15.4 Use the formula $v = \sqrt{\frac{\gamma P}{\rho}}$ to explain why the speed of sound in air

(a) is independent of pressure,
(b) increases with temperature,
(c) increases with humidity.

15.5 You have learnt that a travelling wave in one dimension is represented by a function $y = f(x, t)$ where x and t must appear in the combination $x - v\,t$ or $x + v\,t$, i.e. $y = f(x \pm v\,t)$. Is the converse true? Examine if the following functions for y can possibly represent a travelling wave :

(a) $(x - vt)^2$
(b) $\log[(x + vt)/x_0]$
(c) $1/(x + vt)$

15.6 A bat emits ultrasonic sound of frequency 1000 kHz in air. If the sound meets a water surface, what is the wavelength of (a) the reflected sound, (b) the transmitted sound? Speed of sound in air is 340 m s^{-1} and in water 1486 m s^{-1}.

15.7 A hospital uses an ultrasonic scanner to locate tumours in a tissue. What is the wavelength of sound in the tissue in which the speed of sound is 1.7 km s^{-1}? The operating frequency of the scanner is 4.2 MHz.

15.8 A transverse harmonic wave on a string is described by

$$y(x, t) = 3.0 \sin(36\ t + 0.018\ x + \pi/4)$$

where x and y are in cm and t in s. The positive direction of x is from left to right.

(a) Is this a travelling wave or a stationary wave ?
If it is travelling, what are the speed and direction of its propagation ?
(b) What are its amplitude and frequency ?
(c) What is the initial phase at the origin ?
(d) What is the least distance between two successive crests in the wave ?

15.9 For the wave described in Exercise 15.8, plot the displacement (y) versus (t) graphs for x = 0, 2 and 4 cm. What are the shapes of these graphs? In which aspects does the oscillatory motion in travelling wave differ from one point to another: amplitude, frequency or phase ?

15.10 For the travelling harmonic wave

$$y(x, t) = 2.0 \cos 2\pi(10t - 0.0080\ x + 0.35)$$

where x and y are in cm and t in s. Calculate the phase difference between oscillatory motion of two points separated by a distance of

(a) 4 m,
(b) 0.5 m,
(c) $\lambda/2$,
(d) $3\lambda/4$

15.11 The transverse displacement of a string (clamped at its both ends) is given by

$$y(x, t) = 0.06 \sin \frac{2}{3}x \cos (120 \pi t)$$

where x and y are in m and t in s. The length of the string is 1.5 m and its mass is 3.0×10^{-2} kg.

Answer the following :

(a) Does the function represent a travelling wave or a stationary wave?

(b) Interpret the wave as a superposition of two waves travelling in opposite directions. What is the wavelength, frequency, and speed of each wave ?

(c) Determine the tension in the string.

15.12 (i) For the wave on a string described in Exercise 15.11, do all the points on the string oscillate with the same (a) frequency, (b) phase, (c) amplitude? Explain your answers. (ii) What is the amplitude of a point 0.375 m away from one end?

15.13 Given below are some functions of x and t to represent the displacement (transverse or longitudinal) of an elastic wave. State which of these represent (i) a travelling wave, (ii) a stationary wave or (iii) none at all:

(a) $y = 2 \cos (3x) \sin (10t)$

(b) $y = 2\sqrt{x - vt}$

(c) $y = 3 \sin (5x - 0.5t) + 4 \cos (5x - 0.5t)$

(d) $y = \cos x \sin t + \cos 2x \sin 2t$

15.14 A wire stretched between two rigid supports vibrates in its fundamental mode with a frequency of 45 Hz. The mass of the wire is 3.5×10^{-2} kg and its linear mass density is 4.0×10^{-2} kg m^{-1}. What is (a) the speed of a transverse wave on the string, and (b) the tension in the string?

15.15 A metre-long tube open at one end, with a movable piston at the other end, shows resonance with a fixed frequency source (a tuning fork of frequency 340 Hz) when the tube length is 25.5 cm or 79.3 cm. Estimate the speed of sound in air at the temperature of the experiment. The edge effects may be neglected.

15.16 A steel rod 100 cm long is clamped at its middle. The fundamental frequency of longitudinal vibrations of the rod are given to be 2.53 kHz. What is the speed of sound in steel?

15.17 A pipe 20 cm long is closed at one end. Which harmonic mode of the pipe is resonantly excited by a 430 Hz source ? Will the same source be in resonance with the pipe if both ends are open? (speed of sound in air is 340 m s^{-1}).

15.18 Two sitar strings A and B playing the note '*Ga*' are slightly out of tune and produce beats of frequency 6 Hz. The tension in the string A is slightly reduced and the beat frequency is found to reduce to 3 Hz. If the original frequency of A is 324 Hz, what is the frequency of B?

15.19 Explain why (or how):

(a) in a sound wave, a displacement node is a pressure antinode and vice versa,

(b) bats can ascertain distances, directions, nature, and sizes of the obstacles without any "eyes",

(c) a violin note and sitar note may have the same frequency, yet we can distinguish between the two notes,

(d) solids can support both longitudinal and transverse waves, but only longitudinal waves can propagate in gases, and

(e) the shape of a pulse gets distorted during propagation in a dispersive medium.

15.20 A train, standing at the outer signal of a railway station blows a whistle of frequency 400 Hz in still air. (i) What is the frequency of the whistle for a platform observer when the train (a) approaches the platform with a speed of 10 m s^{-1}, (b) recedes from the platform with a speed of 10 m s^{-1}? (ii) What is the speed of sound in each case ? The speed of sound in still air can be taken as 340 m s^{-1}.

15.21 A train, standing in a station-yard, blows a whistle of frequency 400 Hz in still air. The wind starts blowing in the direction from the yard to the station with at a speed of 10 m s^{-1}. What are the frequency, wavelength, and speed of sound for an observer standing on the station's platform? Is the situation exactly identical to the case when the air is still and the observer runs towards the yard at a speed of 10 m s^{-1}? The speed of sound in still air can be taken as 340 m s^{-1}

Additional Exercises

15.22 A travelling harmonic wave on a string is described by

$$y(x, t) = 7.5 \sin (0.0050x + 12t + \pi/4)$$

(a) what are the displacement and velocity of oscillation of a point at $x = 1$ cm, and $t = 1$ s? Is this velocity equal to the velocity of wave propagation?

(b) Locate the points of the string which have the same transverse displacements and velocity as the $x = 1$ cm point at $t = 2$ s, 5 s and 11 s.

15.23 A narrow sound pulse (for example, a short pip by a whistle) is sent across a medium. (a) Does the pulse have a definite (i) frequency, (ii) wavelength, (iii) speed of propagation? (b) If the pulse rate is 1 after every 20 s, (that is the whistle is blown for a split of second after every 20 s), is the frequency of the note produced by the whistle equal to 1/20 or 0.05 Hz ?

15.24 One end of a long string of linear mass density 8.0×10^{-3} kg m^{-1} is connected to an electrically driven tuning fork of frequency 256 Hz. The other end passes over a pulley and is tied to a pan containing a mass of 90 kg. The pulley end absorbs all the incoming energy so that reflected waves at this end have negligible amplitude. At $t = 0$, the left end (fork end) of the string $x = 0$ has zero transverse displacement ($y = 0$) and is moving along positive y-direction. The amplitude of the wave is 5.0 cm. Write down the transverse displacement y as function of x and t that describes the wave on the string.

15.25 A SONAR system fixed in a submarine operates at a frequency 40.0 kHz. An enemy submarine moves towards the SONAR with a speed of 360 km h^{-1}. What is the frequency of sound reflected by the submarine ? Take the speed of sound in water to be 1450 m s^{-1}.

15.26 Earthquakes generate sound waves inside the earth. Unlike a gas, the earth can experience both transverse (S) and longitudinal (P) sound waves. Typically the speed of S wave is about 4.0 km s^{-1}, and that of P wave is 8.0 km s^{-1}. A seismograph records P and S waves from an earthquake. The first P wave arrives 4 min before the first S wave. Assuming the waves travel in straight line, at what distance does the earthquake occur ?

15.27 A bat is flitting about in a cave, navigating via ultrasonic beeps. Assume that the sound emission frequency of the bat is 40 kHz. During one fast swoop directly toward a flat wall surface, the bat is moving at 0.03 times the speed of sound in air. What frequency does the bat hear reflected off the wall ?

CHAPTER NINE

MECHANICAL PROPERTIES OF SOLIDS

9.1 Introduction
9.2 Elastic behaviour of solids
9.3 Stress and strain
9.4 Hooke's law
9.5 Stress-strain curve
9.6 Elastic moduli
9.7 Applications of elastic behaviour of materials
Summary
Points to ponder
Exercises
Additional exercises

9.1 INTRODUCTION

In Chapter 7, we studied the rotation of the bodies and then realised that the motion of a body depends on how mass is distributed within the body. We restricted ourselves to simpler situations of rigid bodies. A rigid body generally means a hard solid object having a definite shape and size. But in reality, bodies can be stretched, compressed and bent. Even the appreciably rigid steel bar can be deformed when a sufficiently large external force is applied on it. This means that solid bodies are not perfectly rigid.

A solid has definite shape and size. In order to change (or deform) the shape or size of a body, a force is required. If you stretch a helical spring by gently pulling its ends, the length of the spring increases slightly. When you leave the ends of the spring, it regains its original size and shape. The property of a body, by virtue of which it tends to regain its original size and shape when the applied force is removed, is known as **elasticity** and the deformation caused is known as **elastic** deformation. However, if you apply force to a lump of putty or mud, they have no gross tendency to regain their previous shape, and they get permanently deformed. Such substances are called **plastic** and this property is called **plasticity.** Putty and mud are close to ideal plastics.

The elastic behaviour of materials plays an important role in engineering design. For example, while designing a building, knowledge of elastic properties of materials like steel, concrete etc. is essential. The same is true in the design of bridges, automobiles, ropeways etc. One could also ask — Can we design an aeroplane which is very light but sufficiently strong? Can we design an artificial limb which is lighter but stronger? Why does a railway track have a particular shape like **I**? Why is glass brittle while brass is not? Answers to such questions begin with the study of how relatively simple kinds of loads or forces act to deform different solids bodies. In this chapter, we shall study the

elastic behaviour and mechanical properties of solids which would answer many such questions.

9.2 ELASTIC BEHAVIOUR OF SOLIDS

We know that in a solid, each atom or molecule is surrounded by neighbouring atoms or molecules. These are bonded together by interatomic or intermolecular forces and stay in a stable equilibrium position. When a solid is deformed, the atoms or molecules are displaced from their equilibrium positions causing a change in the interatomic (or intermolecular) distances. When the deforming force is removed, the interatomic forces tend to drive them back to their original positions. Thus the body regains its original shape and size. The restoring mechanism can be visualised by taking a model of spring-ball system shown in the Fig. 9.1. Here the balls represent atoms and springs represent interatomic forces.

Fig. 9.1 Spring-ball model for the illustration of elastic behaviour of solids.

If you try to displace any ball from its equilibrium position, the spring system tries to restore the ball back to its original position. Thus elastic behaviour of solids can be explained in terms of microscopic nature of the solid. Robert Hooke, an English physicist (1635 - 1703 A.D) performed experiments on springs and found that the elongation (change in the length) produced in a body is proportional to the applied force or load. In 1676, he presented his law of elasticity, now called Hooke's law. We shall study about it in Section 9.4. This law, like Boyle's law, is one of the earliest quantitative relationships in science. It is very important to know the behaviour of the materials under various kinds of load from the context of engineering design.

9.3 STRESS AND STRAIN

When a force is applied on body, it is deformed to a small or large extent depending upon the nature of the material of the body and the magnitude of the deforming force. The deformation may not be noticeable visually in many materials but it is there. When a body is subjected to a deforming force, a restoring force is developed in the body. This restoring force is equal in magnitude but opposite in direction to the applied force. The restoring force per unit area is known as **stress.** If F is the force applied and A is the area of cross section of the body,

Magnitude of the stress = F/A (9.1)

The SI unit of stress is N m^{-2} or pascal (Pa) and its dimensional formula is [$ML^{-1}T^{-2}$].

There are three ways in which a solid may change its dimensions when an external force acts on it. These are shown in Fig. 9.2. In Fig.9.2(a), a cylinder is stretched by two equal forces applied normal to its cross-sectional area. The restoring force per unit area in this case is called **tensile stress**. If the cylinder is compressed under the action of applied forces, the restoring force per unit area is known as **compressive stress**. Tensile or compressive stress can also be termed as longitudinal stress.

In both the cases, there is a change in the length of the cylinder. The change in the length ΔL to the original length L of the body (cylinder in this case) is known as **longitudinal strain**.

$$\text{Longitudinal strain} \quad \frac{L}{L} \qquad (9.2)$$

However, if two equal and opposite deforming forces are applied parallel to the cross-sectional area of the cylinder, as shown in Fig. 9.2(b), there is relative displacement between the opposite faces of the cylinder. The restoring force per unit area developed due to the applied tangential force is known as **tangential** or **shearing stress**.

Robert Hooke
(1635 – 1703 A.D.)

Robert Hooke was born on July 18, 1635 in Freshwater, Isle of Wight. He was one of the most brilliant and versatile seventeenth century English scientists. He attended Oxford University but never graduated. Yet he was an extremely talented inventor, instrument-maker and building designer. He assisted Robert Boyle in the construction of Boylean air pump. In 1662, he was appointed as Curator of Experiments to the newly founded Royal Society. In 1665, he became Professor of Geometry in Gresham College where he carried out his astronomical observations. He built a Gregorian reflecting telescope; discovered the fifth star in the trapezium and an asterism in the constellation Orion; suggested that Jupiter rotates on its axis; plotted detailed sketches of Mars which were later used in the 19th century to determine the planet's rate of rotation; stated the inverse square law to describe planetary motion, which Newton modified later etc. He was elected Fellow of Royal Society and also served as the Society's Secretary from 1667 to 1682. In his series of observations presented in Micrographia, he suggested wave theory of light and first used the word 'cell' in a biological context as a result of his studies of cork.

Robert Hooke is best known to physicists for his discovery of law of elasticity: **Ut tensio, sic vis** (This is a Latin expression and it means as the distortion, so the force). This law laid the basis for studies of stress and strain and for understanding the elastic materials.

As a result of applied tangential force, there is a relative displacement Δx between opposite faces of the cylinder as shown in the Fig. 9.2(b). The strain so produced is known as **shearing strain** and it is defined as the ratio of relative displacement of the faces Δx to the length of the cylinder L.

$$\text{Shearing strain} \quad \frac{x}{L} = \tan\theta \qquad (9.3)$$

where θ is the angular displacement of the cylinder from the vertical (original position of the cylinder). Usually θ is very small, $\tan\theta$ is nearly equal to angle θ, (if $\theta = 10°$, for example, there is only 1% difference between θ and $\tan\theta$).

It can also be visualised, when a book is pressed with the hand and pushed horizontally, as shown in Fig. 9.2 (c).

Thus, shearing strain $= \tan\theta \approx \theta$ (9.4)

In Fig. 9.2 (d), a solid sphere placed in the fluid under high pressure is compressed uniformly on all sides. The force applied by the fluid acts in perpendicular direction at each point of the surface and the body is said to be under hydraulic compression. This leads to decrease in its volume without any change of its geometrical shape.

The body develops internal restoring forces that are equal and opposite to the forces applied by the fluid (the body restores its original shape and size when taken out from the fluid). The internal restoring force per unit area in this case

Fig. 9.2 *(a) Cylinder subjected to tensile stress stretches it by an amount ΔL. (b) A cylinder subjected to shearing (tangential) stress deforms by an angle θ .(c) A book subjected to a shearing stress (d) A solid sphere subjected to a uniform hydraulic stress shrinks in volume by an amount ΔV.*

is known as **hydraulic stress** and in magnitude is equal to the hydraulic pressure (applied force per unit area).

The strain produced by a hydraulic pressure is called **volume strain** and is defined as the ratio of change in volume (ΔV) to the original volume (V).

$$\text{Volume strain} \quad \frac{V}{V} \qquad (9.5)$$

Since the strain is a ratio of change in dimension to the original dimension, it has no units or dimensional formula.

9.4 HOOKE'S LAW

Stress and strain take different forms in the situations depicted in the Fig. (9.2). For small deformations the stress and strain are proportional to each other. This is known as Hooke's law.

Thus,

$$\text{stress} \propto \text{strain}$$

$$\text{stress} = k \times \text{strain} \qquad (9.6)$$

where k is the proportionality constant and is known as modulus of elasticity.

Hooke's law is an empirical law and is found to be valid for most materials. However, there are some materials which do not exhibits this linear relationship.

9.5 STRESS-STRAIN CURVE

The relation between the stress and the strain for a given material under tensile stress can be found experimentally. In a standard test of tensile properties, a test cylinder or a wire is stretched by an applied force. The fractional change in length (the strain) and the applied force needed to cause the strain are recorded. The applied force is gradually increased in steps and the change in length is noted. A graph is plotted between the stress (which is equal in magnitude to the applied force per unit area) and the strain produced. A typical graph for a metal is shown in Fig. 9.3. Analogous graphs for compression and shear stress may also be obtained. The stress-strain curves vary from material to material. These curves help us to understand how a given material deforms with increasing loads. From the graph, we can see that in the region between O to A, the curve is linear. In this region, Hooke's law is obeyed. The body regains its original dimensions when the applied force is removed. In this region, the solid behaves as an elastic body.

Fig. 9.3 A typical stress-strain curve for a metal.

In the region from A to B, stress and strain are not proportional. Nevertheless, the body still returns to its original dimension when the load is removed. The point B in the curve is known as **yield point** (also known as **elastic limit**) and the corresponding stress is known as **yield strength** (S_y) of the material.

If the load is increased further, the stress developed exceeds the yield strength and strain increases rapidly even for a small change in the stress. The portion of the curve between B and D shows this. When the load is removed, say at some point C between B and D, the body does not regain its original dimension. In this case, even when the stress is zero, the strain is not zero. The material is said to have a **permanent set**. The deformation is said to be **plastic deformation**. The point D on the graph is the ultimate **tensile strength** (S_u) of the material. Beyond this point, additional strain is produced even by a reduced applied force and fracture occurs at point E. If the ultimate strength and fracture points D and E are close, the material is said to be brittle. If they are far apart, the material is said to be ductile.

As stated earlier, the stress-strain behaviour varies from material to material. For example, rubber can be pulled to several times its original length and still returns to its original shape. Fig. 9.4 shows stress-strain curve for the elastic tissue of aorta, present in the heart. Note that although elastic region is very large, the material

Fig. 9.4 *Stress-strain curve for the elastic tissue of Aorta, the large tube (vessel) carrying blood from the heart.*

does not obey Hooke's law over most of the region. Secondly, there is no well defined plastic region. Substances like tissue of aorta, rubber etc. which can be stretched to cause large strains are called **elastomers**.

9.6 ELASTIC MODULI

The proportional region within the elastic limit of the stress-strain curve (region OA in Fig. 9.3) is of great importance for structural and manufacturing engineering designs. The ratio of stress and strain, called **modulus of elasticity**, is found to be a characteristic of the material.

9.6.1 Young's Modulus

Experimental observation show that for a given material, the magnitude of the strain produced is same whether the stress is tensile or compressive. The ratio of tensile (or compressive) stress (σ) to the longitudinal strain (ε) is defined as **Young's modulus** and is denoted by the symbol Y.

$$Y = \frac{\quad}{\quad} \tag{9.7}$$

From Eqs. (9.1) and (9.2), we have

$$Y = (F/A)/(\Delta L/L)$$
$$= (F \times L)/(A \times \Delta L) \tag{9.8}$$

Since strain is a dimensionless quantity, the unit of Young's modulus is the same as that of stress *i.e.*, N m^{-2} or Pascal (Pa). Table 9.1 gives the values of Young's moduli and yield strengths of some materials.

From the data given in Table 9.1, it is noticed that for metals Young's moduli are large. Therefore, these materials require a large force to produce small change in length. To increase the length of a thin steel wire of 0.1 cm^2 cross-sectional area by 0.1%, a force of 2000 N is required. The force required to produce the same strain in aluminium, brass and copper wires having the same cross-sectional area are 690 N, 900 N and 1100 N respectively. It means that steel is more elastic than copper, brass and aluminium. It is for this reason that steel is

Table 9.1 Young's moduli and yield strengths of some materials.

Substance	Density ρ (kg m^{-3})	Young's modulus Y (10^9 N m^{-2})	Ultimate strength, S_u (10^6 N m^{-2})	Yield strength S_y (10^6 N m^{-2})
Aluminium	2710	70	110	95
Copper	8890	110	400	200
Iron (wrought)	7800-7900	190	330	170
Steel	7860	200	400	250
Glass#	2190	65	50	—
Concrete	2320	30	40	—
Wood#	525	13	50	—
Bone#	1900	9	170	—
Polystyrene	1050	3	48	—

 # substance tested under compression

preferred in heavy-duty machines and in structural designs. Wood, bone, concrete and glass have rather small Young's moduli.

Example 9.1 A structural steel rod has a radius of 10 mm and a length of 1.0 m. A 100 kN force stretches it along its length. Calculate (a) stress, (b) elongation, and (c) strain on the rod. Young's modulus, of structural steel is 2.0×10^{11} N m^{-2}.

Answer We assume that the rod is held by a clamp at one end, and the force F is applied at the other end, parallel to the length of the rod. Then the stress on the rod is given by

$$Stress = \frac{F}{A} = \frac{F}{\pi r^2}$$

$$\frac{100 \quad 10^3 \text{ N}}{3.14 \quad 10^{-2} \text{ m}^2}$$

$$= 3.18 \times 10^8 \text{ N m}^{-2}$$

The elongation,

$$\Delta L = \frac{(F/A)L}{Y}$$

$$= \frac{3.18 \quad 10^8 \text{ N m}^{-2} \quad 1\text{m}}{2 \quad 10^{11} \text{ N m}^{-2}}$$

$$= 1.59 \times 10^{-3} \text{ m}$$

$$= 1.59 \text{ mm}$$

The strain is given by

$$\text{Strain} = \Delta L/L$$

$$= (1.59 \times 10^{-3} \text{ m})/(1\text{m})$$

$$= 1.59 \times 10^{-3}$$

$$= 0.16 \ \%$$

Example 9.2 A copper wire of length 2.2 m and a steel wire of length 1.6 m, both of diameter 3.0 mm, are connected end to end. When stretched by a load, the net elongation is found to be 0.70 mm. Obtain the load applied.

Answer The copper and steel wires are under a tensile stress because they have the same tension (equal to the load W) and the same area of cross-section A. From Eq. (9.7) we have stress = strain × Young's modulus. Therefore

$$W/A = Y_c \times (\Delta L_c/L_c) = Y_s \times (\Delta L_s/L_s)$$

where the subscripts c and s refer to copper and stainless steel respectively. Or,

$$\Delta L_c/\Delta L_s = (Y_s/Y_c) \times (L_c/L_s)$$

Given $L_c = 2.2$ m, $L_s = 1.6$ m,

From Table 9.1 $Y_c = 1.1 \times 10^{11}$ N.m^{-2}, and

$Y_s = 2.0 \times 10^{11}$ N.m^{-2}.

$\Delta L_c/\Delta L_s = (2.0 \times 10^{11}/1.1 \times 10^{11}) \times (2.2/1.6) = 2.5.$

The total elongation is given to be

$$\Delta L_c + \Delta L_s = 7.0 \times 10^{-4} \text{ m}$$

Solving the above equations,

$\Delta L_c = 5.0 \times 10^{-4}$ m, and $\Delta L_s = 2.0 \times 10^{-4}$ m.

Therefore

$$W = (A \times Y_c \times \Delta L_c)/L_c$$

$$= \pi(1.5 \times 10^{-3})^2 \times [(5.0 \times 10^{-4} \times 1.1 \times 10^{11})/2.2]$$

$$= 1.8 \times 10^2 \text{ N}$$

Example 9.3 In a human pyramid in a circus, the entire weight of the balanced group is supported by the legs of a performer who is lying on his back (as shown in Fig. 9.5). The combined mass of all the persons performing the act, and the tables, plaques etc. involved is 280 kg. The mass of the performer lying on his back at the bottom of the pyramid is 60 kg. Each thighbone (femur) of this performer has a length of 50 cm and an effective radius of 2.0 cm. Determine the amount by which each thighbone gets compressed under the extra load.

Fig. 9.5 *Human pyramid in a circus.*

Answer Total mass of all the performers, tables, plaques etc. = 280 kg

Mass of the performer = 60 kg

Mass supported by the legs of the performer at the bottom of the pyramid

= 280 – 60 = 220 kg

Weight of this supported mass

= 220 kg wt. = 220 × 9.8 N = 2156 N.

Weight supported by each thighbone of the performer = ½ (2156) N = 1078 N.

From Table 9.1, the Young's modulus for bone is given by

$Y = 9.4 \times 10^9$ N m^{-2}.

Length of each thighbone L = 0.5 m

the radius of thighbone = 2.0 cm

Thus the cross-sectional area of the thighbone $A = \pi \times (2 \times 10^{-2})^2$ m^2 = 1.26×10^{-3} m^2.

Using Eq. (9.8), the compression in each thighbone (ΔL) can be computed as

$$\Delta L = [(F \times L)/(Y \times A)]$$
$$= [(1078 \times 0.5)/(9.4 \times 10^9 \times 1.26 \times 10^{-3})]$$
$$= 4.55 \times 10^{-5} \text{ m or } 4.55 \times 10^{-3} \text{ cm.}$$

This is a very small change! The fractional decrease in the thighbone is $\Delta L/L$ = 0.000091 or 0.0091%. t

9.6.2 Determination of Young's Modulus of the Material of a Wire

A typical experimental arrangement to determine the Young's modulus of a material of wire under tension is shown in Fig. 9.6. It consists of two long straight wires of same length and equal radius suspended side by side from a fixed rigid support. The wire A (called the *reference wire*) carries a millimetre main scale M and a pan to place a weight. The wire B (called the *experimental wire*) of uniform area of cross-section also carries a pan in which known weights can be placed. A vernier scale V is attached to a pointer at the bottom of the experimental wire B, and the main scale M is fixed to the reference wire A. The weights placed in the pan exert a downward force and stretch the experimental wire under a tensile stress. The elongation of the wire (increase in length) is measured by the vernier arrangement. The reference wire is used to compensate for any change in length that may occur due to change in room temperature, since any change in length of the reference wire due to temperature change will be accompanied by an equal change in experimental wire. (We shall study these temperature effects in detail in Chapter 11.)

Fig. 9.6 An arrangement for the determination of Young's modulus of the material of a wire.

Both the reference and experimental wires are given an initial small load to keep the wires straight and the vernier reading is noted. Now the experimental wire is gradually loaded with more weights to bring it under a tensile stress and the vernier reading is noted again. The difference between two vernier readings gives the elongation produced in the wire. Let r and L be the initial radius and length of the experimental wire, respectively. Then the area of cross-section of the wire would be πr^2. Let M be the mass that produced an elongation ΔL in the wire. Thus the applied force is equal to Mg, where g is the acceleration due to gravity. From Eq. (9.8), the Young's modulus of the material of the experimental wire is given by

$$Y = \frac{\sigma}{\varepsilon} = \frac{Mg}{\pi r^2} \cdot \frac{L}{\Delta L}$$
$$= Mg \times L/(\pi r^2 \times \Delta L) \qquad (9.9)$$

9.6.3 Shear Modulus

The ratio of shearing stress to the corresponding shearing strain is called the *shear modulus* of the material and is represented by G. It is also called the *modulus of rigidity*.

G = shearing stress (σ_s)/shearing strain

$$G = (F/A)/(\Delta x/L)$$
$$= (F \times L)/(A \times \Delta x) \quad (9.10)$$

Similarly, from Eq. (9.4)

$$G = (F/A)/\theta$$
$$= F/(A \times \theta) \quad (9.11)$$

The shearing stress σ_s can also be expressed as

$$\sigma_s = G \times \theta \quad (9.12)$$

SI unit of shear modulus is N m^{-2} or Pa. The shear moduli of a few common materials are given in Table 9.2. It can be seen that shear modulus (or modulus of rigidity) is generally less than Young's modulus (from Table 9.1). For most materials $G \approx Y/3$.

Table 9.2 Shear moduli *(G)* of some common materials

Material	G (10^9 Nm^{-2} or GPa)
Aluminium	25
Brass	36
Copper	42
Glass	23
Iron	70
Lead	5.6
Nickel	77
Steel	84
Tungsten	150
Wood	10

Example 9.4 A square lead slab of side 50 cm and thickness 10 cm is subject to a shearing force (on its narrow face) of 9.0×10^4 N. The lower edge is riveted to the floor. How much will the upper edge be displaced?

Answer The lead slab is fixed and the force is applied parallel to the narrow face as shown in Fig. 9.7. The area of the face parallel to which this force is applied is

$$A = 50 \text{ cm} \times 10 \text{ cm}$$
$$= 0.5 \text{ m} \times 0.1 \text{ m}$$
$$= 0.05 \text{ m}^2$$

Therefore, the stress applied is

$$= (9.4 \times 10^4 \text{ N}/0.05 \text{ m}^2)$$
$$= 1.80 \times 10^6 \text{ N.m}^{-2}$$

Fig. 9.7

We know that shearing strain = ($\Delta x/L$)= Stress /G. Therefore the displacement Δx = (Stress $\times$ L)/G

$$= (1.8 \times 10^6 \text{ N m}^{-2} \times 0.5\text{m})/(5.6 \times 10^9 \text{ N m}^{-2})$$
$$= 1.6 \times 10^{-4} \text{ m} = 0.16 \text{ mm}$$

9.6.4 Bulk Modulus

In Section (9.3), we have seen that when a body is submerged in a fluid, it undergoes a hydraulic stress (equal in magnitude to the hydraulic pressure). This leads to the decrease in the volume of the body thus producing a strain called volume strain [Eq. (9.5)]. The ratio of hydraulic stress to the corresponding hydraulic strain is called *bulk modulus*. It is denoted by symbol B.

$$B = -p/(\Delta V/V) \quad (9.13)$$

The negative sign indicates the fact that with an increase in pressure, a decrease in volume occurs. That is, if p is positive, ΔV is negative. Thus for a system in equilibrium, the value of bulk modulus B is always positive. SI unit of bulk modulus is the same as that of pressure *i.e.*, N m^{-2} or Pa. The bulk moduli of a few common materials are given in Table 9.3.

The reciprocal of the bulk modulus is called *compressibility* and is denoted by k. It is defined as the fractional change in volume per unit increase in pressure.

$$k = (1/B) = -(1/\Delta p) \times (\Delta V/V) \quad (9.14)$$

It can be seen from the data given in Table 9.3 that the bulk moduli for solids are much larger than for liquids, which are again much larger than the bulk modulus for gases (air).

Table 9.3 Bulk moduli *(B)* of some common Materials

Material Solids	B (10^9 N m^{-2} or GPa)
Aluminium	72
Brass	61
Copper	140
Glass	37
Iron	100
Nickel	260
Steel	160
Liquids	
Water	2.2
Ethanol	0.9
Carbon disulphide	1.56
Glycerine	4.76
Mercury	25
Gases	
Air (at STP)	1.0×10^{-4}

Thus solids are least compressible whereas gases are most compressible. Gases are about a million times more compressible than solids! Gases have large compressibilities, which vary with pressure and temperature. The incompressibility of the solids is primarily due to the tight coupling between the neighbouring atoms. The molecules in liquids are also bound with their neighbours but not as strong as in solids. Molecules in gases are very poorly coupled to their neighbours.

Table 9.4 shows the various types of stress, strain, elastic moduli, and the applicable state of matter at a glance.

Example 9.5 The average depth of Indian Ocean is about 3000 m. Calculate the fractional compression, $\Delta V/V$, of water at the bottom of the ocean, given that the bulk modulus of water is 2.2×10^9 N m^{-2}. (Take g = 10 m s^{-2})

Answer The pressure exerted by a 3000 m column of water on the bottom layer

$$p = h\rho g = 3000 \text{ m} \times 1000 \text{ kg m}^{-3} \times 10 \text{ m s}^{-2}$$
$$= 3 \times 10^7 \text{ kg m}^{-1} \text{ s}^{-2}$$
$$= 3 \times 10^7 \text{ N m}^{-2}$$

Fractional compression $\Delta V/V$, is

$$\Delta V/V = \text{stress}/B = (3 \times 10^7 \text{ N m}^{-2})/(2.2 \times 10^9 \text{ N m}^{-2})$$
$$= 1.36 \times 10^{-2} \text{ or } 1.36\ \%$$

Table 9.4 Stress, strain and various elastic moduli

Type of stress	Stress	Strain	Change in shape	Change in volume	Elastic modulus	Name of modulus	State of Mater
Tensile or compressive	Two equal and opposite forces perpendicular to opposite faces ($\sigma = F/A$)	Elongation or compression parallel to force direction ($\Delta L/L$) (longitudinal strain)	Yes	No	$Y = (F \times L)/(A \times \Delta L)$	Young's modulus	Solid
Shearing	Two equal and opposite forces parallel to oppoiste surfaces [forces in each case such that total force and total torque on the body vanishes ($\sigma_s = F/A$)	Pure shear, θ	Yes	No	$G = (F \times \theta)/A$	Shear modulus	Solid
Hydraulic	Forces perpendicular everywhere to the surface, force per unit area (pressure) same everywhere.	Volume change (compression or elongation ($\Delta V/V$)	No	Yes	$B = -p/(\Delta V/V)$	Bulk modulus	Solid, liquid and gas

9.7 APPLICATIONS OF ELASTIC BEHAVIOUR OF MATERIALS

The elastic behaviour of materials plays an important role in everyday life. All engineering designs require precise knowledge of the elastic behaviour of materials. For example while designing a building, the structural design of the columns, beams and supports require knowledge of strength of materials used. Have you ever thought why the beams used in construction of bridges, as supports etc. have a cross-section of the type **I**? Why does a heap of sand or a hill have a pyramidal shape? Answers to these questions can be obtained from the study of structural engineering which is based on concepts developed here.

Cranes used for lifting and moving heavy loads from one place to another have a thick metal rope to which the load is attached. The rope is pulled up using pulleys and motors. Suppose we want to make a crane, which has a lifting capacity of 10 tonnes or metric tons (1 metric ton = 1000 kg). How thick should the steel rope be? We obviously want that the load does not deform the rope permanently. Therefore, the extension should not exceed the elastic limit. From Table 9.1, we find that mild steel has a yield strength (S_y) of about 300×10^6 N m^{-2}. Thus, the area of cross-section (A) of the rope should at least be

$$A \geq W/S_y = Mg/S_y \qquad (9.15)$$
$$= (10^4 \text{ kg} \times 10 \text{ m s}^{-2})/(300 \times 10^6 \text{ N m}^{-2})$$
$$= 3.3 \times 10^{-4} \text{ m}^2$$

corresponding to a radius of about 1 cm for a rope of circular cross-section. Generally a large margin of safety (of about a factor of ten in the load) is provided. Thus a thicker rope of radius about 3 cm is recommended. A single wire of this radius would practically be a rigid rod. So the ropes are always made of a number of thin wires braided together, like in pigtails, for ease in manufacture, flexibility and strength.

A bridge has to be designed such that it can withstand the load of the flowing traffic, the force of winds and its own weight. Similarly, in the design of buildings use of beams and columns is very common. In both the cases, the overcoming of the problem of bending of beam under a load is of prime importance. The beam should not bend too much or break. Let us consider the case of a beam loaded at the centre and supported near its ends as shown in Fig. 9.8. A bar of length l, breadth b, and depth d when loaded at the centre by a load W sags by an amount given by

$$\delta = W l^3/(4bd^3Y) \qquad (9.16)$$

Fig. 9.8 *A beam supported at the ends and loaded at the centre.*

This relation can be derived using what you have already learnt and a little calculus. From Eq. (9.16), we see that to reduce the bending for a given load, one should use a material with a large Young's modulus Y. For a given material, increasing the depth d rather than the breadth b is more effective in reducing the bending, since δ is proportional to d^{-3} and only to b^{-1}(of course the length l of the span should be as small as possible). But on increasing the depth, unless the load is exactly at the right place (difficult to arrange in a bridge with moving traffic), the deep bar may bend as shown in Fig. 9.9(b). This is called buckling. To avoid this, a common compromise is the cross-sectional shape shown in Fig. 9.9(c). This section provides a large load-bearing surface and enough depth to prevent bending. This shape reduces the weight of the beam without sacrificing the strength and hence reduces the cost.

Fig. 9.9 *Different cross-sectional shapes of a beam. (a) Rectangular section of a bar; (b) A thin bar and how it can buckle; (c) Commonly used section for a load bearing bar.*

Use of pillars or columns is also very common in buildings and bridges. A pillar with rounded ends as shown in Fig. 9.10(a) supports less load than that with a distributed shape at the ends [Fig. 9.10(b)]. The precise design of a bridge or a building has to take into account the conditions under which it will function, the cost and long period, reliability of usable materials etc.

Fig. 9.10 *Pillars or columns: (a) a pillar with rounded ends, (b) Pillar with distributed ends.*

The answer to the question why the maximum height of a mountain on earth is ~10 km can also be provided by considering the elastic properties of rocks. A mountain base is not under uniform compression and this provides some shearing stress to the rocks under which they can flow. The stress due to all the material on the top should be less than the critical shearing stress at which the rocks flow.

At the bottom of a mountain of height h, the force per unit area due to the weight of the mountain is $h\rho g$ where ρ is the density of the material of the mountain and g is the acceleration due to gravity. The material at the bottom experiences this force in the vertical direction, and the sides of the mountain are free. Therefore this is not a case of pressure or bulk compression. There is a shear component, approximately $h\rho g$ itself. Now the elastic limit for a typical rock is 30×10^7 N m^{-2}. Equating this to $h\rho g$, with $\rho = 3 \times 10^3$ kg m^{-3} gives

$$h\rho g = 30 \times 10^7 \text{ N m}^{-2}. \quad \text{Or}$$

$$h = 30 \times 10^7 \text{ N m}^{-2}/(3 \times 10^3 \text{ kg m}^{-3} \times 10 \text{ m s}^{-2})$$
$$= 10 \text{ km}$$

which is more than the height of Mt. Everest!

SUMMARY

1. Stress is the restoring force per unit area and strain is the fractional change in dimension. In general there are three types of stresses (a) tensile stress — longitudinal stress (associated with stretching) or compressive stress (associated with compression), (b) shearing stress, and (c) hydraulic stress.
2. For small deformations, stress is directly proportional to the strain for many materials. This is known as Hooke's law. The constant of proportionality is called modulus of elasticity. Three elastic moduli *viz.*, Young's modulus, shear modulus and bulk modulus are used to describe the elastic behaviour of objects as they respond to deforming forces that act on them.
   A class of solids called elastomers does not obey Hooke's law.
3. When an object is under tension or compression, the Hooke's law takes the form
   $$F/A = Y\Delta L/L$$
   where $\Delta L/L$ is the tensile or compressive strain of the object, F is the magnitude of the applied force causing the strain, A is the cross-sectional area over which F is applied (perpendicular to A) and Y is the Young's modulus for the object. The stress is F/A.
4. A pair of forces when applied parallel to the upper and lower faces, the solid deforms so that the upper face moves sideways with respect to the lower. The horizontal displacement ΔL of the upper face is perpendicular to the vertical height L. This type of deformation is called shear and the corresponding stress is the shearing stress. This type of stress is possible only in solids.
   In this kind of deformation the Hooke's law takes the form
   $$F/A = G \times \Delta L/L$$
   where ΔL is the displacement of one end of object in the direction of the applied force F, and G is the shear modulus.

5. When an object undergoes hydraulic compression due to a stress exerted by a surrounding fluid, the Hooke's law takes the form

$$p = B\,(\Delta V/V),$$

where p is the pressure (hydraulic stress) on the object due to the fluid, $\Delta V/V$ (the volume strain) is the absolute fractional change in the object's volume due to that pressure and B is the bulk modulus of the object.

POINTS TO PONDER

1. In the case of a wire, suspended from celing and stretched under the action of a weight (F) suspended from its other end, the force exerted by the ceiling on it is equal and opposite to the weight. However, the tension at any cross-section A of the wire is just F and not $2F$. Hence, tensile stress which is equal to the tension per unit area is equal to F/A.
2. Hooke's law is valid only in the linear part of stress-strain curve.
3. The Young's modulus and shear modulus are relevant only for solids since only solids have lengths and shapes.
4. Bulk modulus is relevant for solids, liquid and gases. It refers to the change in volume when every part of the body is under the uniform stress so that the shape of the body remains unchanged.
5. Metals have larger values of Young's modulus than alloys and elastomers. A material with large value of Young's modulus requires a large force to produce small changes in its length.
6. In daily life, we feel that a material which stretches more is more elastic, but it a is misnomer. In fact material which stretches to a lesser extent for a given load is considered to be more elastic.
7. In general, a deforming force in one direction can produce strains in other directions also. The proportionality between stress and strain in such situations cannot be described by just one elastic constant. For example, for a wire under longitudinal strain, the lateral dimensions (radius of cross section) will undergo a small change, which is described by another elastic constant of the material (called *Poisson ratio*).
8. Stress is not a vector quantity since, unlike a force, the stress cannot be assigned a specific direction. Force acting on the portion of a body on a specified side of a section has a definite direction.

EXERCISES

9.1 A steel wire of length 4.7 m and cross-sectional area 3.0×10^{-5} m^2 stretches by the same amount as a copper wire of length 3.5 m and cross-sectional area of 4.0×10^{-5} m^2 under a given load. What is the ratio of the Young's modulus of steel to that of copper?

9.2 Figure 9.11 shows the strain-stress curve for a given material. What are (a) Young's modulus and (b) approximate yield strength for this material?

Fig. 9.11

9.3 The stress-strain graphs for materials *A* and *B* are shown in Fig. 9.12.

Fig. 9.12

The graphs are drawn to the same scale.

(a) Which of the materials has the greater Young's modulus?

(b) Which of the two is the stronger material?

9.4 Read the following two statements below carefully and state, with reasons, if it is true or false.

(a) The Young's modulus of rubber is greater than that of steel;

(b) The stretching of a coil is determined by its shear modulus.

9.5 Two wires of diameter 0.25 cm, one made of steel and the other made of brass are loaded as shown in Fig. 9.13. The unloaded length of steel wire is 1.5 m and that of brass wire is 1.0 m. Compute the elongations of the steel and the brass wires.

Fig. 9.13

9.6 The edge of an aluminium cube is 10 cm long. One face of the cube is firmly fixed to a vertical wall. A mass of 100 kg is then attached to the opposite face of the cube. The shear modulus of aluminium is 25 GPa. What is the vertical deflection of this face?

9.7 Four identical hollow cylindrical columns of mild steel support a big structure of mass 50,000 kg. The inner and outer radii of each column are 30 and 60 cm respectively. Assuming the load distribution to be uniform, calculate the compressional strain of each column.

9.8 A piece of copper having a rectangular cross-section of 15.2 mm × 19.1 mm is pulled in tension with 44,500 N force, producing only elastic deformation. Calculate the resulting strain?

9.9 A steel cable with a radius of 1.5 cm supports a chairlift at a ski area. If the maximum stress is not to exceed 10^8 N m^{-2}, what is the maximum load the cable can support ?

9.10 A rigid bar of mass 15 kg is supported symmetrically by three wires each 2.0 m long. Those at each end are of copper and the middle one is of iron. Determine the ratios of their diameters if each is to have the same tension.

9.11 A 14.5 kg mass, fastened to the end of a steel wire of unstretched length 1.0 m, is whirled in a vertical circle with an angular velocity of 2 rev/s at the bottom of the circle. The cross-sectional area of the wire is 0.065 cm^2. Calculate the elongation of the wire when the mass is at the lowest point of its path.

9.12 Compute the bulk modulus of water from the following data: Initial volume = 100.0 litre, Pressure increase = 100.0 atm (1 atm = 1.013 × 10^5 Pa), Final volume = 100.5 litre. Compare the bulk modulus of water with that of air (at constant temperature). Explain in simple terms why the ratio is so large.

9.13 What is the density of water at a depth where pressure is 80.0 atm, given that its density at the surface is 1.03 × 103 kg m^{-3}?

9.14 Compute the fractional change in volume of a glass slab, when subjected to a hydraulic pressure of 10 atm.

9.15 Determine the volume contraction of a solid copper cube, 10 cm on an edge, when subjected to a hydraulic pressure of 7.0 × 10^6 Pa.

9.16 How much should the pressure on a litre of water be changed to compress it by 0.10%?

Additional Exercises

9.17 Anvils made of single crystals of diamond, with the shape as shown in Fig. 9.14, are used to investigate behaviour of materials under very high pressures. Flat faces at the narrow end of the anvil have a diameter of 0.50 mm, and the wide ends are subjected to a compressional force of 50,000 N. What is the pressure at the tip of the anvil?

Fig. 9.14

9.18 A rod of length 1.05 m having negligible mass is supported at its ends by two wires of steel (wire A) and aluminium (wire B) of equal lengths as shown in Fig. 9.15. The cross-sectional areas of wires A and B are 1.0 mm^2 and 2.0 mm^2, respectively. At what point along the rod should a mass m be suspended in order to produce (a) equal stresses and (b) equal strains in both steel and aluminium wires.

Fig. 9.15

9.19 A mild steel wire of length 1.0 m and cross-sectional area 0.50×10^{-2} cm^2 is stretched, well within its elastic limit, horizontally between two pillars. A mass of 100 g is suspended from the mid-point of the wire. Calculate the depression at the mid-point.

9.20 Two strips of metal are riveted together at their ends by four rivets, each of diameter 6.0 mm. What is the maximum tension that can be exerted by the riveted strip if the shearing stress on the rivet is not to exceed 6.9×10^7 Pa? Assume that each rivet is to carry one quarter of the load.

9.21 The Marina trench is located in the Pacific Ocean, and at one place it is nearly eleven km beneath the surface of water. The water pressure at the bottom of the trench is about 1.1×10^8 Pa. A steel ball of initial volume 0.32 m^3 is dropped into the ocean and falls to the bottom of the trench. What is the change in the volume of the ball when it reaches to the bottom?

ANSWERS

Chapter 9

9.1 1.8

9.2 (a) From the given graph for a stress of 150×10^6 N m^{-2} the strain is 0.002

(b) Approximate yield strength of the material is 3×10^8 N m^{-2}

9.3 (a) Material A

(b) Strength of a material is determined by the amount of stress required to cause fracture: material A is stronger than material B.

9.4 (a) False (b) True

9.5 1.5×10^{-4} m (steel); 1.3×10^{-4} m (brass)

9.6 Deflection = 4×10^{-6} m

9.7 2.8×10^{-6}

9.8 0.127

9.9 7.07×10^4 N

9.10 $D_{copper}/D_{iron} = 1.25$

9.11 1.539×10^{-4} m

9.12 2.026×10^9 Pa

9.13 1.034×10^3 kg/m^3

9.14 0.0027

9.15 0.058 cm^3

9.16 2.2×10^6 N/m^2

9.17 Pressure at the tip of anvil is 2.5×10^{11} Pa

9.18 (a) 0.7 m (b) 0.43 m from steel wire

9.19 Approximately 0.01 m

9.20 260 kN

9.21 2.51×10^{-4} m^3

Chapter 10

10.3 (a) decreases (b) η of gases increases, η of liquid decreases with temperature (c) shear strain, rate of shear strain (d) conservation of mass, Bernoulli's equation (e) greater.

10.5 6.2×10^6 Pa

10.6 10.5 m

10.7 Pressure at that depth in the sea is about 3×10^7 Pa. The structure is suitable since it can withstand far greater pressure or stress.

10.8 6.92×10^5 Pa

10.9 0.800

10.10 Mercury will rise in the arm containing spirit; the difference in levels of mercury will be 0.221 cm.

10.11 No, Bernoulli's principle applies to streamline flow only.

10.12 No, unless the atmospheric pressures at the two points where Bernoulli's equation is applied are significantly different.

10.13 9.8×10^2 Pa (The Reynolds number is about 0.3 so the flow is laminar).

10.14 1.5×10^3 N

10.15 Fig (a) is incorrect [Reason: at a constriction (i.e. where the area of cross-section of the tube is smaller), flow speed is larger due to mass conservation. Consequently pressure there is smaller according to Bernoulli's equation. We assume the fluid to be incompressible].

10.16 0.64 m s^{-1}

10.17 2.5×10^{-2} N m^{-1}

10.18 4.5×10^{-2} N for (b) and (c), the same as in (a).

10.19 Excess pressure = 310 Pa, total pressure = 1.0131×10^5 Pa. However, since data are correct to three significant figures, we should write total pressure inside the drop as 1.01×10^5 Pa.

10.20 Excess pressure inside the soap bubble = 20.0 Pa; excess pressure inside the air bubble in soap solution = 10.0 Pa. Outside pressure for air bubble = $1.01 \times 10^5 + 0.4 \times 10^3 \times 9.8$

$\times$ 1.2 = 1.06 $\times$ 10^5 Pa. The excess pressure is so small that up to three significant figures, total pressure inside the air bubble is 1.06 $\times$ 10^5 Pa.

10.21 55 N (Note, the base area does not affect the answer)

10.22 (a) absolute pressure = 96 cm of Hg; gauge pressure = 20 cm of Hg for (a), absolute pressure = 58 cm of Hg, gauge pressure = -18 cm of Hg for (b); (b) mercury would rise in the left limb such that the difference in its levels in the two limbs becomes19 cm.

10.23 Pressure (and therefore force) on the two equal base areas are identical. But force is exerted by water on the sides of the vessels also, which has a nonzero vertical component when the sides of the vessel are not perfectly normal to the base. This net vertical component of force by water on sides of the vessel is greater for the first vessel than the second. Hence the vessels weigh different even when the force on the base is the same in the two cases.

10.24 0.2 m

10.25 (a) The pressure drop is greater (b) More important with increasing flow velocity.

10.26 (a) 0.98 m s^{-1}; (b) 1.24 $\times$ 10^{-5} m^3 s^{-1}

10.27 4393 kg

10.28 5.8 cm s^{-1}, 3.9 $\times$ 10^{-10} N

10.29 5.34 mm

10.30 For the first bore, pressure difference (between the concave and convex side) = 2 $\times$ 7.3 $\times$ 10^{-2} / 3 $\times$ 10^{-3} = 48.7 Pa. Similarly for the second bore, pressure difference = 97.3 Pa. Consequently, the level difference in the two bores is [48.7 / (10^3 x 9.8)] m = 5.0 mm.

The level in the narrower bore is higher. (Note, for zero angle of contact, the radius of the meniscus equals radius of the bore. The concave side of the surface in each bore is at 1 atm).

10.31 (b) 8 km. If we consider the variation of *g* with altitude the height is somewhat more, about 8.2 km.

Chapter 11

11.1 Neon: –248.58 °C = – 415.44 °F;

CO_2: – 56.60 °C = – 69.88 °F

(use $t_F = \frac{9}{5} t_c + 32$)

11.2 T_A = (4/7) T_B

11.3 384.8 K

11.4 (a) Triple-point has a *unique* temperature; fusion point and boiling point temperatures depend on pressure; (b) The other fixed point is the absolute zero itself; (c) Triple-point is 0.01°C, not 0 °C; (d) 491.69.

11.5 (a) T_A = 392.69 K, T_B = 391.98 K; (b) The discrepancy arises because the gases are not perfectly ideal. To reduce the discrepancy, readings should be taken for lower and

lower pressures and the plot between temperature measured versus absolute pressure of the gas at triple point should be extrapolated to obtain temperature in the limit pressure tends to zero, when the gases approach ideal gas behaviour.

11.6 Actual length of the rod at 45.0 °C = (63.0 + 0.0136) cm = 63.0136 cm. (However, we should say that change in length up to three significant figures is 0.0136 cm, but the total length is 63.0 cm, up to three significant places. Length of the same rod at 27.0 °C = 63.0 cm.

11.7 When the shaft is cooled to temperature – 69^0C the wheel can slip on the shaft.

11.8 The diameter increases by an amount = 1.44×10^{-2} cm.

11.9 3.8×10^2 N

11.10 Since the ends of the combined rod are not clamped, each rod expands freely.

Δl_{brass} = 0.21 cm, Δl_{steel} = 0.126 cm = 0.13 cm

Total change in length = 0.34 cm. No 'thermal stress' is developed at the junction since the rods freely expand.

11.11 0.0147 = 1.5×10^{-2}

11.12 103 °C

11.13 1.5 kg

11.14 0.43 J g^{-1} K^{-1} ; smaller

11.15 The gases are diatomic, and have other degrees of freedom (i.e. have other modes of motion) possible besides the translational degrees of freedom. To raise the temperature of the gas by a certain amount, heat is to be supplied to increase the average energy of all the modes. Consequently, molar specific heat of diatomic gases is more than that of monatomic gases. It can be shown that if only rotational modes of motion are considered, the molar specific heat of diatomic gases is nearly (5/2) R which agrees with the observations for all the gases listed in the table, except chlorine. The higher value of molar specific heat of chlorine indicates that besides rotational modes, vibrational modes are also present in chlorine at room temperature.

11.16 (a) At the triple point temperature = – 56.6 °C and pressure = 5.11 atm.

(b) Both the boiling point and freezing point of CO_2 decrease if pressure decreases.

(c) The critical temperature and pressure of CO_2 are 31.1 °C and 73.0 atm respectively. Above this temperature, CO_2 will not liquefy even if compressed to high pressures.

(d) (a) vapour (b) solid (c) liquid

11.17 (a) No, vapour condenses to solid directly.

(b) It condenses to solid directly without passing through the liquid phase.

(c) It turns to liquid phase and then to vapour phase. The fusion and boiling points are where the horizontal line on *P-T* diagram at the constant pressure of 10 atm intersects the fusion and vaporization curves.

(d) It will not exhibit any clear transition to the liquid phase, but will depart more and more from ideal gas behaviour as its pressure increases.

11.18 4.3 g/min

11.19 3.7 kg

11.20 238 °C

11.22 9 min

Chapter 12

12.1 16 g per min

12.2 934 J

12.4 2.64

12.5 16.9 J

12.6 (a) 0.5 atm (b) zero (c) zero (assuming the gas to be ideal) (d) No, since the process (called free expansion) is rapid and cannot be controlled. The intermediate states are non-equilibrium states and do not satisfy the gas equation. In due course, the gas does return to an equilibrium state.

12.7 15%, 3.1×10^9 J

12.8 25 W

12.9 450 J

12.10 10.4

Chapter 13

13.1 4×10^{-4}

13.3 (a) The dotted plot corresponds to 'ideal' gas behaviour; (b) $T_1 > T_2$; (c) 0.26 J K^{-1}; (d) No, 6.3×10^{-5} kg of H_2 would yield the same value

13.4 0.14 kg

13.5 5.3×10^{-6} m^3

13.6 6.10×10^{26}

13.7 (a) 6.2×10^{-21} J (b) 1.24×10^{-19} J (c) 2.1×10^{-16} J

13.8 Yes, according to Avogadro's law. No, v_{rms} is largest for the lightest of the three gases; neon.

13.9 2.52×10^3 K

13.10 Use the formula for mean free path :

$$\bar{l} \quad \frac{1}{\sqrt{2}\ nd^2}$$

where d is the diameter of a molecule. For the given pressure and temperature $N/V = 5.10 \times 10^{25}$ m^{-3} and $= 1.0 \times 10^{-7}$ m. $v_{rms} = 5.1 \times 10^2$ m s^{-1}.

collisional frequency = $\frac{v_{rms}}{\bar{l}}$ $5.1 \quad 10^9 \text{ s}^{-1}$. Time taken for the collision = $d / v_{rms} = 4 \times 10^{-13}$ s. Time taken between successive collisions = $l / v_{rms} = 2 \times 10^{-10}$ s. Thus the time taken between successive collisions is 500 times the time taken for a collision. Thus a molecule in a gas moves essentially free for most of the time.

13.11 Nearly 24 cm of mercury flows out, and the remaining 52 cm of mercury thread plus the 48 cm of air above it remain in equilibrium with the outside atmospheric pressure (We assume there is no change in temperature throughout).

13.12 Oxygen

13.14 Carbon[1.29 Å]; Gold [1.59 Å]; Liquid Nitrogen [1.77 Å]; Lithium [1.73 Å]; Liquid fluorine[1.88 Å]

Chapter 14

14.1 (b), (c)

14.2 (b) and (c): SHM; (a) and (d) represent periodic but not SHM [A polyatomic molecule has a number of natural frequencies; so in general, its vibration is a superposition of SHM's of a number of different frequencies. This superposition is periodic but not SHM].

14.3 (b) and (d) are periodic, each with a period of 2 s; (a) and (c) are not periodic. [Note in (c), repetition of merely one position is not enough for motion to be periodic; the entire motion during one period must be repeated successively].

14.4 (a) Simple harmonic, $T = (2\pi/\omega)$; (b) periodic, $T = (2\pi/\omega)$ but not simple harmonic; (c) simple harmonic, $T = (\pi/\omega)$; (d) periodic, $T = (2\pi/\omega)$ but not simple harmonic; (e) non-periodic; (f) non-periodic (physically not acceptable as the function $\to\infty$ as $t\to\infty$.

14.5 (a) 0, +, + ; (b) 0, –, – ; (c) –, 0,0 ; (d) –, –, – ; (e) +, +, + ; (f) –, –, –.

14.6 (c) represents a simple harmonic motion.

14.7 A = $\sqrt{2}$ cm, $\phi = 7\pi/4$; B = $\sqrt{2}$ cm, $\alpha = \pi/4$.

14.8 219 N

14.9 Frequency 3.2 s^{-1}; maximum acceleration of the mass 8.0 m s^{-2}; maximum speed of the mass 0.4 m s^{-1}.

14.10 (a) $x = 2 \sin 20t$
(b) $x = 2 \cos 20t$
(c) $x = -2 \cos 20t$

where x is in cm. These functions differ neither in amplitude nor frequency. They differ in initial phase.

14.11 (a) $x = -3 \sin \pi t$ where x is in cm.

(b) $x = -2 \cos \frac{\pi}{2} t$ where x is in cm.

14.13 (a) F/k for both (a) and (b).

(b) $T = 2\pi$ for (a) and $2\pi \sqrt{\frac{m}{2k}}$ for (b)

14.14 100 m/min

14.15 8.4 s

14.16 (a) For a simple pendulum, k itself is proportional to m, so m cancels out.

(b) $\sin \theta < \theta$; if the restoring force, $mg \sin \theta$ is replaced by $mg\theta$, this amounts to effective reduction in g for large angles and hence an increase in time period T over that given by the formula $T = 2\pi \sqrt{\frac{l}{g}}$ where one assumes $\sin\theta = \theta$.

$\sqrt{\frac{m}{k}}$

(c) Yes, the motion in the wristwatch depends on spring action and has nothing to do with acceleration due to gravity.

(d) Gravity disappears for a man under free fall, so frequency is zero.

14.17 $T = 2\pi \sqrt{\frac{l}{\sqrt{g^2 + v^4/R^2}}}$. Hint: Effective acceleration due to gravity will get reduced due to radial acceleration v^2/R acting in the horizontal plane.

14.18 In equilibrium, weight of the cork equals the up thrust. When the cork is depressed by an amount x, the net upward force is $Ax\rho_l g$. Thus the force constant $k = A\rho_l g$.

Using $m = Ah\rho$, and $T = 2\pi \sqrt{\frac{m}{k}}$ one gets the given expression.

14.19 When both the ends are open to the atmosphere, and the difference in levels of the liquid in the two arms is h, the net force on the liquid column is $Ah\rho g$ where A is the area of cross-section of the tube and ρ is the density of the liquid. Since restoring force is proportional to h, motion is simple harmonic.

14.20 $T = 2\pi\sqrt{\frac{Vm}{Ba^2}}$ where B is the bulk modulus of air. For isothermal changes $B = P$.

14.21 (a) 5×10^4 N m^{-1}; (b) 1344.6 kg s^{-1}

14.22 Hint: Average K.E. = $\frac{1}{T}\int_0^T \frac{1}{2}mv^2 dt$; Average P.E.= $\frac{1}{T}\int_0^T \frac{1}{2}kx^2 dt$

14.23 Hint: The time period of a torsional pendulum is given by $T = 2\pi\sqrt{\frac{I}{\alpha}}$, where I is the moment of inertia about the axis of rotation. In our case $I = \frac{1}{2}MR^2$, where M is the mass of the disk and R its radius. Substituting the given values, $\alpha = 2.0$ N m rad^{-1}.

14.24 (a) $-5\pi^2$ m s^{-2} ; 0; (b) $-3\pi^2$ m s^{-2}; 0.4π m s^{-1}; (c) 0 ; $0.5\,\pi$ m s^{-1}

14.25 $\sqrt{\left(x_0^2 + \frac{v_0^2}{\omega^2}\right)}$

Chapter 15

15.1 0.5 s

15.2 8.7 s

15.3 2.06×10^4 N

15.4 Assume ideal gas law: $P = \frac{\rho RT}{M}$, where ρ is the density, M is the molecular mass, and T is the temperature of the gas. This gives $v = \sqrt{\frac{\gamma RT}{M}}$. This shows that v is:

(a) Independent of pressure.

(b) Increases as $\sqrt{T}$.

(c) The molecular mass of water (18) is less than that of N_2 (28) and O_2 (32).

Therefore as humidity increases, the effective molecular mass of air decreases and hence v increases.

15.5 The converse is not true. An obvious requirement for an acceptable function for a travelling wave is that it should be finite everywhere and at all times. Only function (c) satisfies this condition, the remaining functions cannot possibly represent a travelling wave.

15.6 (a) 3.4×10^{-4} m (b) 1.49×10^{-3} m

15.7 4.1×10^{-4} m

15.8 (a) A travelling wave. It travels from right to left with a speed of 20 ms^{-1}.

(b) 3.0 cm, 5.7 Hz

(c) $\pi/4$

(d) 3.5 m

15.9 All the graphs are sinusoidal. They have same amplitude and frequency, but different initial phases.

15.10 (a) 6.4 π rad

(b) 0.8 π rad

(c) π rad

(d) $(\pi/2)$ rad

15.11 (a) Stationary wave

(b) $l = 3$ m, n = 60 Hz, and $v = 180$ m s^{-1} for each wave

(c) 648 N

15.12 (a) All the points except the nodes on the string have the same frequency and phase, but not the same amplitude.

(b) 0.042 m

15.13 (a) Stationary wave.

(b) Unacceptable function for any wave.

(c) Travelling harmonic wave.

(d) Superposition of two stationary waves.

15.14 (a) 79 m s^{-1}

(b) 248 N

15.15 347 m s^{-1}

Hint : $v_n = \dfrac{(2n-1)v}{4l}$; $n = 1,2,3,....$for a pipe with one end closed

15.16 5.06 km s^{-1}

15.17 First harmonic (fundamental); No.

15.18 318 Hz

15.20 (i) (a) 412 Hz, (b) 389 Hz, (ii) 340 m s^{-1} in each case.

15.21 400 Hz, 0.875 m, 350 m s^{-1}. No, because in this case, with respect to the medium, both the observer and the source are in motion.

15.22 (a) 1.666 cm, 87.75 cm s^{-1}; No, the velocity of wave propagation is –24 m s^{-1}

(b) All points at distances of $n\lambda$ ($n = \pm 1, \pm 2, \pm 3, \ldots$) where λ = 12.6 m from the point x = 1 cm.

15.23 (a) The pulse does not have a definite wavelength or frequency, but has a definite speed of propagation (in a non-dispersive medium).

(b) No

15.24 $y = 0.05 \sin(\omega t - kx)$; here $\omega = 1.61 \times 10^3$ s^{-1}, $k = 4.84$ m^{-1}; x and y are in m.

15.25 45.9 kHz

15.26 1920 km

15.27 42.47 kHz

BIBLIOGRAPHY

TEXTBOOKS

For additional reading on the topics covered in this book, you may like to consult one or more of the following books. Some of these books however are more advanced and contain many more topics than this book.

1. **Ordinary Level Physics**, A.F. Abbott, Arnold-Heinemann (1984).
2. **Advanced Level Physics**, M. Nelkon and P. Parker, 6th Edition Arnold-Heinemann (1987).
3. **Advanced Physics**, Tom Duncan, John Murray (2000).
4. **Fundamentals of Physics**, David Halliday, Robert Resnick and Jearl Walker, 7th Edition John Wily (2004).
5. **University Physics**, H.D. Young, M.W. Zemansky and F.W. Sears, Narosa Pub. House (1982).
6. **Problems in Elementary Physics**, B. Bukhovtsa, V. Krivchenkov, G. Myakishev and V. Shalnov, MIR Publishers, (1971).
7. **Lectures on Physics** (3 volumes), R.P. Feynman, Addision – Wesley (1965).
8. **Berkeley Physics Course** (5 volumes) McGraw Hill (1965).
   a. Vol. 1 – Mechanics: (Kittel, Knight and Ruderman)
   b. Vol. 2 – Electricity and Magnetism (E.M. Purcell)
   c. Vol. 3 – Waves and Oscillations (Frank S. Craw-ford)
   d. Vol. 4 – Quantum Physics (Wichmann)
   e. Vol. 5 – Statistical Physics (F. Reif)
9. **Fundamental University Physics**, M. Alonso and E. J. Finn, Addison – Wesley (1967).
10. **College Physics**, R.L. Weber, K.V. Manning, M.W. White and G.A. Weygand, Tata McGraw Hill (1977).
11. **Physics**: **Foundations and Frontiers**, G. Gamow and J.M. Cleveland, Tata McGraw Hill (1978).
12. **Physics for the Inquiring Mind**, E.M. Rogers, Princeton University Press (1960)
13. **PSSC Physics Course**, DC Heath and Co. (1965) Indian Edition, NCERT (1967)
14. **Physics Advanced Level**, Jim Breithampt, Stanley Thornes Publishers (2000).
15. **Physics,** Patrick Fullick, Heinemann (2000).

16 **Conceptual Physics,** Paul G. Hewitt, Addision-Wesley (1998).
17 **College Physics,** Raymond A. Serway and Jerry S. Faughn, Harcourt Brace and Co. (1999).
18 **University Physics,** Harris Benson, John Wiley (1996).
19 **University Physics,** William P. Crummet and Arthur B. Western, Wm.C. Brown (1994).
20 **General Physics,** Morton M. Sternheim and Joseph W. Kane, John Wiley (1988).
21 **Physics,** Hans C. Ohanian, W.W. Norton (1989).
22 **Advanced Physics,** Keith Gibbs, Cambridge University Press(1996).
23 **Understanding Basic Mechanics,** F. Reif, John Wiley (1995).
24 **College Physics,** Jerry D. Wilson and Anthony J. Buffa, Prentice-Hall (1997).
25 **Senior Physics, Part – I,** I.K. Kikoin and A.K. Kikoin, Mir Publishers (1987).
26 **Senior Physics, Part – II,** B. Bekhovtsev, Mir Publishers (1988).
27 **Understanding Physics,** K. Cummings, Patrick J. Cooney, Priscilla W. Laws and Edward F. Redish, John Wiley (2005)
28 **Essentials of Physics,** John D. Cutnell and Kenneth W. Johnson, John Wiley (2005)

GENERAL BOOKS

For instructive and entertaining general reading on science, you may like to read some of the following books. Remember however, that many of these books are written at a level far beyond the level of the present book.

1 **Mr. Tompkins** in paperback, G. Gamow, Cambridge University Press (1967).
2 **The Universe and Dr. Einstein**, C. Barnett, Time Inc. New York (1962).
3 **Thirty years that Shook Physics**, G. Gamow, Double Day, New York (1966).
4 **Surely You're Joking, Mr. Feynman**, R.P. Feynman, Bantam books (1986).
5 **One, Two, Three... Infinity**, G. Gamow, Viking Inc. (1961).
6 **The Meaning of Relativity**, A. Einstein, (Indian Edition) Oxford and IBH Pub. Co (1965).
7 **Atomic Theory and the Description of Nature**, Niels Bohr, Cambridge (1934).
8 **The Physical Principles of Quantum Theory**, W. Heisenberg, University of Chicago Press (1930).
9 **The Physics- Astronomy Frontier**, F. Hoyle and J.V. Narlikar, W.H. Freeman (1980).
10 **The Flying Circus of Physics with Answer**, J. Walker, John Wiley and Sons (1977).
11 **Physics for Everyone** (series), L.D. Landau and A.I. Kitaigorodski, MIR Publisher (1978).
Book 1: Physical Bodies
Book 2: Molecules
Book 3: Electrons
Book 4: Photons and Nuclei.
12 **Physics can be Fun**, Y. Perelman, MIR Publishers (1986).
13 **Power of Ten**, Philip Morrison and Eames, W.H. Freeman (1985).
14 **Physics in your Kitchen Lab**., I.K. Kikoin, MIR Publishers (1985).
15 **How Things Work : The Physics of Everyday Life**, Louis A. Bloomfield, John Wiley (2005)
16 **Physics Matters : An Introduction to Conceptual Physics,** James Trefil and Robert M. Hazen, John Wiley (2004).

INDEX

A

Absolute scale temperature 276
Absolute zero 276
Acceleration (linear) 45
Acceleration due to gravity 49,189
Accuracy 22
Action-reaction 97
Addition of vectors 67
Adiabatic process 306, 307
Aerofoil 258
Air resistance 79
Amplitude 340, 368
Angle of contact 263, 264
Angstrom 21
Angular Acceleration 154
Angular displacement 338
Angular frequency 340, 369
Angular momentum 155
Angular velocity 152
Angular wave number 368
Antinodes 376
Archimedes Principle 251
Area expansion 277
Atmospheric pressure 249
Average acceleration 45, 74
Average speed 42
Average velocity 42
Avogardo's law 320

B

Banked road 104
Barometer 250
Beat frequency 338, 380, 381
Beats 379, 380
Bending of beam 240
Bernoulli's Principle 254
Blood pressure 272, 273
Boiling point 283
Boyle's law 321
Buckling 240
Bulk modulus 238
Buoyant force 251

C

Calorimeter 281
Capillary rise 264
Capillary waves 366
Carnot engine 311
Central forces 186
Centre of Gravity 161
Centre of mass 144
Centripetal acceleration 81
Centripetal force 104
Change of state 282
Charle's law 321
Chemical Energy 126
Circular motion 104
Clausius statement 310
Coefficient of area expansion 279
Coefficient of linear expansion 277
Coefficient of performance 309
Coefficient of static friction 101
Coefficient of viscosity 258
Coefficient of volume expansion 277
Cold reservoir 308
Collision 129
Collision in two dimensions 131
Compressibility 238
Compressions 364, 365, 371
Compressive stress 232, 239
Conduction 286
Conservation laws 12
Conservation of angular momentum 157, 173
Conservation of Mechanical Energy 121
Conservation of momentum 98
Conservative force 121
Constant acceleration 46,75
Contact force 100
Convection 289
Couple 159
Crest 367
Cyclic process 307

D

Dalton's law of partial pressure 321
Damped oscillations 351
Damped simple Harmonic motion 351
Damping constant 351
Damping force 351
Derived units 16
Detergent action 265
Diastolic pressure 273
Differential calculus 61
Dimensional analysis 32
Dimensions 31
Displacement vector 66
Displacement 40
Doppler effect 381, 382
Doppler shift 383
Driving frequency 354
Dynamics of rotational motion 169

E

Efficiency of heat engine 308
Elastic Collision 129
Elastic deformation 232, 234
Elastic limit 234
Elastic moduli 235
Elasticity 231
Elastomers 235
Electromagnetic force 8
Energy 117
Equality of vectors 66
Equation of continuity 253
Equilibrium of a particle 99
Equilibrium of Rigid body 158
Equilibrium position 337, 338, 349
Errors in measurement 22
Escape speed 193

F

First law of Thermodynamics 302
Fluid pressure 247
Force 94
Forced frequency 353
Forced oscillations 353
Fracture point 234
Free Fall 49
Free-body diagram 100
Frequency of periodic motion 338, 369
Friction 101
Fundamental Forces 6
Fundamental mode 377
Fusion 282

G

Gauge pressure 249
Geocentric model 183
Geostationary satellite 196
Gravitational constant 189
Gravitational Force 8, 192
Gravitational potential energy 191
Gravity waves 366

H

Harmonic frequency 377, 378
Harmonics 377, 378
Heat capacity 280
Heat engines 308
Heat pumps 308
Heat 274
Heliocentric model 183
Hertz 338
Hooke's law 234
Horizontal range 78
Hot reservoir 308
Hydraulic brakes 251, 252
Hydraulic lift 251, 252
Hydraulic machines 251
Hydraulic pressure 234
Hydraulic stress 234, 239
Hydrostatic paradox 249

I

Ideal gas equation 276
Ideal gas 321
Impulse 96
Inelastic collision 129
Initial phase angle 368
Instantaneous acceleration 74
Instantaneous speed 45
Instantaneous velocity 43
Interference 374
Internal energy 301, 325
Irreversible engine 311, 313
Irreversible processes 310
Isobaric process 306, 307
Isochoric process 306, 307
Isotherm 305
Isothermal process 306

K

Kelvin-Planck statement 310
Kepler's laws of planetary motion 184
Kinematics of Rotational Motion 167
Kinematics 39
Kinetic energy of rolling motion 174
Kinetic Energy 117
Kinetic interpretation of temperature 324
Kinetic theory of gases 323

L

Laminar flow 254, 260
Laplace correction 373

Latent heat of fusion 285
Latent heat of vaporisation 285
Latent heat 284
Law of cosine 72
Law of equipartition of energy 327
Law of Inertia 90
Law of sine 72
Linear expansion 277
Linear harmonic oscillator 345, 347
Linear momentum 155
Longitudinal strain 232
Longitudinal strain 232, 235
Longitudinal stress 232
Longitudinal Wave 366, 372

M

Magnus effect 257
Manometer 250
Mass Energy Equivalence 126
Maximum height of projectile 78
Maxwell Distribution 326
Mean free path 319, 330
Measurement of length 18
Measurement of mass 21
Measurement of temperature 275
Measurement of time 22
Melting point 282
Modes 376
Modulus of elasticity 234
Modulus of rigidity 238
Molar specific heat capacity at constant pressure 280, 304
Molar specific heat capacity at constant volume 280, 304
Molar specific heat capacity 280
Molecular nature of matter 318
Moment of Inertia 163
Momentum 93
Motion in a plane 72
Multiplication of vectors 67
Musical instruments 380

N

Natural frequency 354
Newton's first law of motion 91
Newton's Law of cooling 290
Newton's law of gravitation 185
Newton's second law of motion 93
Newton's third law of motion 96
Newtons' formula for speed of sound 372
Nodes 376
Normal Modes 376, 377, 379
Note 380, 381
Nuclear Energy 126
Null vector 68

O

Odd harmonics 378
Orbital velocity/speed 194
Order of magnitude 28
Oscillations 337
Oscillatory motion 337

P

Parallax method 18
Parallelogram law of addition of vectors 66
Pascal's law 248
Path length 40
Path of projectile 78
Periodic force 353
Periodic motion 337
Periodic time 337
Permanent set 234
Phase angle 340
Phase constant 340
Phase diagram 283
Pipe open at both ends 380
Pipe open at one end 379
Pitch 382
Plastic deformation 234
Plasticity 231
Polar satellite 196
Position vector and displacement 73
Potential energy of a spring 123
Potential energy 120
Power 128
Precession 143
Pressure gauge 249
Pressure of an ideal gas 323
Pressure pulse 366
Pressure 246
Principle of Conservation of Energy 128
Principle of moments 160
Progressive wave 366
Projectile motion 77
Projectile 77
Propagation constant 368
Pulse 365

Q

Quasi-static process 305, 306

R

Radial acceleration 344
Radiation 290
Radius of Gyration 164
Raman effect 11
Rarefactions 364, 365, 371
Ratio of specific heat capacities 329
Reaction time 51
Real gases 321
Rectilinear motion 39

Reductionism 2
Reflected wave 375
Reflection of waves 374
Refracted wave 375
Refrigerator 308
Regelation 282
Relative velocity in two dimensions 76
Relative velocity 51
Resolution of vectors 69
Resonance 354
Restoring force 232, 345, 364
Reversible engine 311, 312
Reversible processes 310
Reynolds number 260
Rigid body 141
Rolling motion 173
Root mean square speed 325
Rotation 142

S

S.H.M. (Simple Harmonic Motion) 339
Scalar-product 114
Scalars 65
Scientific Method 1
Second law of Thermodynamics 309
Shear modulus 238
Shearing strain 233
Shearing stress 232, 239
SI units 16
Significant figures 27
Simple pendulum 338, 349
Soap bubbles 264
Sonography 383
Sound 371
Specific heat capacity of Solids 303, 330
Specific heat capacity of Gases 328, 329
Specific heat capacity of Water 330
Specific heat capacity 280, 303
Speed of efflux 255
Speed of Sound 371, 372
Speed of Transverse wave on a stretched string 370, 371
Sphygmomanometer 273
Spring constant 348, 351
Standing waves 376
Stationary waves 378
Steady flow 253
Stethoscope 273
Stokes' law 259
Stopping distance 50
Strain 232
Streamline flow 253, 254
Streamline 253, 254
Stress 232
Stress-strain curve 234
Stretched string 370
Sublimation 284
Subtraction of vectors 67
Superposition principle 374
Surface energy 261
Surface tension 261
Symmetry 146
System of units 16
Systolic pressure 273

T

Temperature 275
Tensile strength 234
Tensile stress 232
Terminal velocity 260
Theorem of parallel axes 167
Theorem of perpendicular axes 165
Thermal conductivity 287
Thermal equilibrium 299
Thermal expansion 276
Thermal stress 279
Thermodynamic processes 305
Thermodynamic state variables 304
Thermodynamics 3, 298
Time of flight 78
Torque 154
Torricelli's Law 255, 256
Trade wind 289
Transmitted wave 375
Travelling wave 376
Triangle law of addition of vectors 66
Triple point 283
Trough 367
Tune 380
Turbulent flow 253, 254

U

Ultimate strength 234
Ultrasonic waves 383
Unification of Forces 10
Unified Atomic Mass Unit 21
Uniform circular motion 79
Uniform Motion 41
Uniformly accelerated motion 47
Unit vectors 70

V

Vane 351
Vaporisation 283
Vector-product 151
Vectors 66
Velocity amplitude 344
Venturi meter 256
Vibration 337
Viscosity 258
Volume expansion 277
Volume Strain 234

W

Wave equation 370
Wave length 368
Wave speed 370
Waves 363
Waxing and waning of sound 381
Weak nuclear force 9
Weightlessness 197
Work done by variable force 118
Work 116
Work-Energy Theorem 116
Working substance 308

Y

Yield Point 234
Yield strength 234
Young's modulus 235

Z

Zeroth law of Thermodynamics 300

PHYSICS

Textbook for Class XII

PART I

PHYSICS Part I Class XII NCERT

CONTENTS

Foreword *v*
Preface *xi*

CHAPTER ONE
Electric Charges and Fields

1.1 Introduction 1
1.2 Electric Charges 1
1.3 Conductors and Insulators 5
1.4 Charging by Induction 6
1.5 Basic Properties of Electric Charge 8
1.6 Coulomb's Law 10
1.7 Forces between Multiple Charges 15
1.8 Electric Field 18
1.9 Electric Field Lines 23
1.10 Electric Flux 25
1.11 Electric Dipole 27
1.12 Dipole in a Uniform External Field 31
1.13 Continuous Charge Distribution 32
1.14 Gauss's Law 33
1.15 Application of Gauss's Law 37

CHAPTER TWO
Electrostatic Potential and Capacitance

2.1 Introduction 51
2.2 Electrostatic Potential 53
2.3 Potential due to a Point Charge 54
2.4 Potential due to an Electric Dipole 55
2.5 Potential due to a System of Charges 57
2.6 Equipotential Surfaces 60
2.7 Potential Energy of a System of Charges 61
2.8 Potential Energy in an External Field 64
2.9 Electrostatics of Conductors 67
2.10 Dielectrics and Polarisation 71
2.11 Capacitors and Capacitance 73
2.12 The Parallel Plate Capacitor 74
2.13 Effect of Dielectric on Capacitance 75

2.14 Combination of Capacitors 78
2.15 Energy Stored in a Capacitor 80
2.16 Van de Graaff Generator 83

CHAPTER THREE
CURRENT ELECTRICITY

3.1 Introduction 93
3.2 Electric Current 93
3.3 Electric Currents in Conductors 94
3.4 Ohm's law 95
3.5 Drift of Electrons and the Origin of Resistivity 97
3.6 Limitations of Ohm's Law 101
3.7 Resistivity of various Materials 101
3.8 Temperature Dependence of Resistivity 103
3.9 Electrical Energy, Power 105
3.10 Combination of Resistors — Series and Parallel 107
3.11 Cells, emf, Internal Resistance 110
3.12 Cells in Series and in Parallel 113
3.13 Kirchhoff's Laws 115
3.14 Wheatstone Bridge 118
3.15 Meter Bridge 120
3.16 Potentiometer 122

CHAPTER FOUR
MOVING CHARGES AND MAGNETISM

4.1 Introduction 132
4.2 Magnetic Force 133
4.3 Motion in a Magnetic Field 137
4.4 Motion in Combined Electric and Magnetic Fields 140
4.5 Magnetic Field due to a Current Element, Biot-Savart Law 143
4.6 Magnetic Field on the Axis of a Circular Current Loop 145
4.7 Ampere's Circuital Law 147
4.8 The Solenoid and the Toroid 150
4.9 Force between Two Parallel Currents, the Ampere 154
4.10 Torque on Current Loop, Magnetic Dipole 157
4.11 The Moving Coil Galvanometer 163

CHAPTER FIVE
MAGNETISM AND MATTER

5.1 Introduction 173
5.2 The Bar Magnet 174

5.3 Magnetism and Gauss's Law 181
5.4 The Earth's Magnetism 185
5.5 Magnetisation and Magnetic Intensity 189
5.6 Magnetic Properties of Materials 191
5.7 Permanent Magnets and Electromagnets 195

CHAPTER SIX
ELECTROMAGNETIC INDUCTION

6.1 Introduction 204
6.2 The Experiments of Faraday and Henry 205
6.3 Magnetic Flux 206
6.4 Faraday's Law of Induction 207
6.5 Lenz's Law and Conservation of Energy 210
6.6 Motional Electromotive Force 212
6.7 Energy Consideration: A Quantitative Study 215
6.8 Eddy Currents 218
6.9 Inductance 219
6.10 AC Generator 224

CHAPTER SEVEN
ALTERNATING CURRENT

7.1 Introduction 233
7.2 AC Voltage Applied to a Resistor 234
7.3 Representation of AC Current and Voltage by Rotating Vectors — Phasors 237
7.4 AC Voltage Applied to an Inductor 237
7.5 AC Voltage Applied to a Capacitor 241
7.6 AC Voltage Applied to a Series LCR Circuit 244
7.7 Power in AC Circuit: The Power Factor 252
7.8 LC Oscillations 255
7.9 Transformers 259

CHAPTER EIGHT
ELECTROMAGNETIC WAVES

8.1 Introduction 269
8.2 Displacement Current 270
8.3 Electromagnetic Waves 274
8.4 Electromagnetic Spectrum 280

ANSWERS 288

Chapter One

ELECTRIC CHARGES AND FIELDS

1.1 Introduction

All of us have the experience of seeing a spark or hearing a crackle when we take off our synthetic clothes or sweater, particularly in dry weather. This is almost inevitable with ladies garments like a polyester saree. Have you ever tried to find any explanation for this phenomenon? Another common example of electric discharge is the lightning that we see in the sky during thunderstorms. We also experience a sensation of an electric shock either while opening the door of a car or holding the iron bar of a bus after sliding from our seat. The reason for these experiences is discharge of electric charges through our body, which were accumulated due to rubbing of insulating surfaces. You might have also heard that this is due to generation of static electricity. This is precisely the topic we are going to discuss in this and the next chapter. Static means anything that does not move or change with time. *Electrostatics deals with the study of forces, fields and potentials arising from static charges.*

1.2 Electric Charge

Historically the credit of discovery of the fact that amber rubbed with wool or silk cloth attracts light objects goes to Thales of Miletus, Greece, around 600 BC. The name electricity is coined from the Greek word *elektron* meaning *amber*. Many such pairs of materials were known which

FIGURE 1.1 Rods and pith balls: like charges repel and unlike charges attract each other.

on rubbing could attract light objects like straw, pith balls and bits of papers. You can perform the following activity at home to experience such an effect. Cut out long thin strips of white paper and lightly iron them. Take them near a TV screen or computer monitor. You will see that the strips get attracted to the screen. In fact they remain stuck to the screen for a while.

It was observed that if two glass rods rubbed with wool or silk cloth are brought close to each other, they repel each other [Fig. 1.1(a)]. The two strands of wool or two pieces of silk cloth, with which the rods were rubbed, also repel each other. However, the glass rod and wool attracted each other. Similarly, two plastic rods rubbed with cat's fur repelled each other [Fig. 1.1(b)] but attracted the fur. On the other hand, the plastic rod attracts the glass rod [Fig. 1.1(c)] and repel the silk or wool with which the glass rod is rubbed. The glass rod repels the fur.

If a plastic rod rubbed with fur is made to touch two small pith balls (now-a-days we can use polystyrene balls) suspended by silk or nylon thread, then the balls repel each other [Fig. 1.1(d)] and are also repelled by the rod. A similar effect is found if the pith balls are touched with a glass rod rubbed with silk [Fig. 1.1(e)]. A dramatic observation is that a pith ball touched with glass rod attracts another pith ball touched with plastic rod [Fig. 1.1(f)].

These seemingly simple facts were established from years of efforts and careful experiments and their analyses. It was concluded, after many careful studies by different scientists, that there were only two kinds of an entity which is called the *electric charge*. We say that the bodies like glass or plastic rods, silk, fur and pith balls are electrified. They acquire an electric charge on rubbing. The experiments on pith balls suggested that there are two kinds of electrification and we find that (i) *like charges repel* and (ii) *unlike charges attract* each other. The experiments also demonstrated that the charges are transferred from the rods to the pith balls on contact. It is said that the pith balls are electrified or are charged by contact. The property which differentiates the two kinds of charges is called the *polarity* of charge.

When a glass rod is rubbed with silk, the rod acquires one kind of charge and the silk acquires the second kind of charge. This is true for any pair of objects that are rubbed to be electrified. Now if the electrified glass rod is brought in contact with silk, with which it was rubbed, they no longer attract each other. They also do not attract or repel other light objects as they did on being electrified.

Thus, the charges acquired after rubbing are lost when the charged bodies are brought in contact. What can you conclude from these observations? It just tells us that unlike charges acquired by the objects

Interactive animation on simple electrostatic experiments:
http://ephysics.physics.ucla.edu/travoltage/HTML/

PHYSICS

neutralise or nullify each other's effect. Therefore the charges were named as *positive* and *negative* by the American scientist Benjamin Franklin. We know that when we add a positive number to a negative number of the same magnitude, the sum is zero. This might have been the philosophy in naming the charges as positive and negative. By convention, the charge on glass rod or cat's fur is called positive and that on plastic rod or silk is termed negative. If an object possesses an electric charge, it is said to be electrified or charged. When it has no charge it is said to be neutral.

Unification of electricity and magnetism

In olden days, electricity and magnetism were treated as separate subjects. Electricity dealt with charges on glass rods, cat's fur, batteries, lightning, etc., while magnetism described interactions of magnets, iron filings, compass needles, etc. In 1820 Danish scientist Oersted found that a compass needle is deflected by passing an electric current through a wire placed near the needle. Ampere and Faraday supported this observation by saying that electric charges in motion produce magnetic fields and moving magnets generate electricity. The unification was achieved when the Scottish physicist Maxwell and the Dutch physicist Lorentz put forward a theory where they showed the interdependence of these two subjects. This field is called *electromagnetism*. Most of the phenomena occurring around us can be described under electromagnetism. Virtually every force that we can think of like friction, chemical force between atoms holding the matter together, and even the forces describing processes occurring in cells of living organisms, have its origin in electromagnetic force. Electromagnetic force is one of the fundamental forces of nature.

Maxwell put forth four equations that play the same role in classical electromagnetism as Newton's equations of motion and gravitation law play in mechanics. He also argued that light is electromagnetic in nature and its speed can be found by making purely electric and magnetic measurements. He claimed that the science of optics is intimately related to that of electricity and magnetism.

The science of electricity and magnetism is the foundation for the modern technological civilisation. Electric power, telecommunication, radio and television, and a wide variety of the practical appliances used in daily life are based on the principles of this science. Although charged particles in motion exert both electric and magnetic forces, in the frame of reference where all the charges are at rest, the forces are purely electrical. You know that gravitational force is a long-range force. Its effect is felt even when the distance between the interacting particles is very large because the force decreases inversely as the square of the distance between the interacting bodies. We will learn in this chapter that electric force is also as pervasive and is in fact stronger than the gravitational force by several orders of magnitude (refer to Chapter 1 of Class XI Physics Textbook).

A simple apparatus to detect charge on a body is the *gold-leaf electroscope* [Fig. 1.2(a)]. It consists of a vertical metal rod housed in a box, with two thin gold leaves attached to its bottom end. When a charged object touches the metal knob at the top of the rod, charge flows on to the leaves and they diverge. The degree of divergance is an indicator of the amount of charge.

FIGURE 1.2 Electroscopes: (a) The gold leaf electroscope, (b) Schematics of a simple electroscope.

Students can make a simple electroscope as follows [Fig. 1.2(b)]: Take a thin aluminium curtain rod with ball ends fitted for hanging the curtain. Cut out a piece of length about 20 cm with the ball at one end and flatten the cut end. Take a large bottle that can hold this rod and a cork which will fit in the opening of the bottle. Make a hole in the cork sufficient to hold the curtain rod snugly. Slide the rod through the hole in the cork with the cut end on the lower side and ball end projecting above the cork. Fold a small, thin aluminium foil (about 6 cm in length) in the middle and attach it to the flattened end of the rod by cellulose tape. This forms the leaves of your electroscope. Fit the cork in the bottle with about 5 cm of the ball end projecting above the cork. A paper scale may be put inside the bottle in advance to measure the separation of leaves. The separation is a rough measure of the amount of charge on the electroscope.

To understand how the electroscope works, use the white paper strips we used for seeing the attraction of charged bodies. Fold the strips into half so that you make a mark of fold. Open the strip and iron it lightly with the mountain fold up, as shown in Fig. 1.3. Hold the strip by pinching it at the fold. You would notice that the two halves move apart. This shows that the strip has acquired charge on ironing. When you fold it into half, both the halves have the same charge. Hence they repel each other. The same effect is seen in the leaf electroscope. On charging the curtain rod by touching the ball end with an electrified body, charge is transferred to the curtain rod and the attached aluminium foil. Both the halves of the foil get similar charge and therefore repel each other. The divergence in the leaves depends on the amount of charge on them. Let us first try to understand why material bodies acquire charge.

FIGURE 1.3 Paper strip experiment.

You know that all matter is made up of atoms and/or molecules. Although normally the materials are electrically neutral, they do contain charges; but their charges are exactly balanced. Forces that hold the molecules together, forces that hold atoms together in a solid, the adhesive force of glue, forces associated with surface tension, all are basically electrical in nature, arising from the forces between charged particles. Thus the electric force is all pervasive and it encompasses almost each and every field associated with our life. It is therefore essential that we learn more about such a force.

To electrify a neutral body, we need to add or remove one kind of charge. When we say that a body is charged, we always refer to this excess charge or deficit of charge. In solids, some of the electrons, being less tightly bound in the atom, are the charges which are transferred from one body to the other. A body can thus be charged positively by losing some of its electrons. Similarly, a body can be charged negatively

by gaining electrons. When we rub a glass rod with silk, some of the electrons from the rod are transferred to the silk cloth. Thus the rod gets positively charged and the silk gets negatively charged. No new charge is created in the process of rubbing. Also the number of electrons, that are transferred, is a very small fraction of the total number of electrons in the material body. Also only the less tightly bound electrons in a material body can be transferred from it to another by rubbing. Therefore, when a body is rubbed with another, the bodies get charged and that is why we have to stick to certain pairs of materials to notice charging on rubbing the bodies.

1.3 Conductors and Insulators

A metal rod held in hand and rubbed with wool will not show any sign of being charged. However, if a metal rod with a wooden or plastic handle is rubbed without touching its metal part, it shows signs of charging. Suppose we connect one end of a copper wire to a neutral pith ball and the other end to a negatively charged plastic rod. We will find that the pith ball acquires a negative charge. If a similar experiment is repeated with a nylon thread or a rubber band, no transfer of charge will take place from the plastic rod to the pith ball. Why does the transfer of charge not take place from the rod to the ball?

Some substances readily allow passage of electricity through them, others do not. Those which allow electricity to pass through them easily are called *conductors*. They have electric charges (electrons) that are comparatively free to move inside the material. Metals, human and animal bodies and earth are conductors. Most of the non-metals like glass, porcelain, plastic, nylon, wood offer high resistance to the passage of electricity through them. They are called *insulators*. Most substances fall into one of the two classes stated above*.

When some charge is transferred to a conductor, it readily gets distributed over the entire surface of the conductor. In contrast, if some charge is put on an insulator, it stays at the same place. You will learn why this happens in the next chapter.

This property of the materials tells you why a nylon or plastic comb gets electrified on combing dry hair or on rubbing, but a metal article like spoon does not. The charges on metal leak through our body to the ground as both are conductors of electricity.

When we bring a charged body in contact with the earth, all the excess charge on the body disappears by causing a momentary current to pass to the ground through the connecting conductor (such as our body). This process of sharing the charges with the earth is called *grounding or earthing*. Earthing provides a safety measure for electrical circuits and appliances. A thick metal plate is buried deep into the earth and thick wires are drawn from this plate; these are used in buildings for the purpose of earthing near the mains supply. The electric wiring in our houses has three wires: live, neutral and earth. The first two carry

* There is a third category called *semiconductors*, which offer resistance to the movement of charges which is intermediate between the conductors and insulators.

electric current from the power station and the third is earthed by connecting it to the buried metal plate. Metallic bodies of the electric appliances such as electric iron, refrigerator, TV are connected to the earth wire. When any fault occurs or live wire touches the metallic body, the charge flows to the earth without damaging the appliance and without causing any injury to the humans; this would have otherwise been unavoidable since the human body is a conductor of electricity.

1.4 Charging by Induction

When we touch a pith ball with an electrified plastic rod, some of the negative charges on the rod are transferred to the pith ball and it also gets charged. Thus the pith ball is *charged by contact*. It is then repelled by the plastic rod but is attracted by a glass rod which is oppositely charged. However, why a electrified rod attracts light objects, is a question we have still left unanswered. Let us try to understand what could be happening by performing the following experiment.

(i) Bring two metal spheres, A and B, supported on insulating stands, in contact as shown in Fig. 1.4(a).

(ii) Bring a positively charged rod near one of the spheres, say A, taking care that it does not touch the sphere. The free electrons in the spheres are attracted towards the rod. This leaves an excess of positive charge on the rear surface of sphere B. Both kinds of charges are bound in the metal spheres and cannot escape. They, therefore, reside on the surfaces, as shown in Fig. 1.4(b). The left surface of sphere A, has an excess of negative charge and the right surface of sphere B, has an excess of positive charge. However, not all of the electrons in the spheres have accumulated on the left surface of A. As the negative charge starts building up at the left surface of A, other electrons are repelled by these. In a short time, equilibrium is reached under the action of force of attraction of the rod and the force of repulsion due to the accumulated charges. Fig. 1.4(b) shows the equilibrium situation. The process is called *induction of charge* and happens almost instantly. The accumulated charges remain on the surface, as shown, till the glass rod is held near the sphere. If the rod is removed, the charges are not acted by any outside force and they redistribute to their original neutral state.

(iii) Separate the spheres by a small distance while the glass rod is still held near sphere A, as shown in Fig. 1.4(c). The two spheres are found to be oppositely charged and attract each other.

(iv) Remove the rod. The charges on spheres rearrange themselves as shown in Fig. 1.4(d). Now, separate the spheres quite apart. The charges on them get uniformly distributed over them, as shown in Fig. 1.4(e).

In this process, the metal spheres will each be equal and oppositely charged. This is *charging by induction*. The positively charged glass rod does not lose any of its charge, contrary to the process of charging by contact.

When electrified rods are brought near light objects, a similar effect takes place. The rods induce opposite charges on the near surfaces of the objects and similar charges move to the farther side of the object.

FIGURE 1.4 Charging by induction.

[This happens even when the light object is not a conductor. The mechanism for how this happens is explained later in Sections 1.10 and 2.10.] The centres of the two types of charges are slightly separated. We know that opposite charges attract while similar charges repel. However, the magnitude of force depends on the distance between the charges and in this case the force of attraction overweighs the force of repulsion. As a result the particles like bits of paper or pith balls, being light, are pulled towards the rods.

Example 1.1 How can you charge a metal sphere positively without touching it?

Solution Figure 1.5(a) shows an uncharged metallic sphere on an insulating metal stand. Bring a negatively charged rod close to the metallic sphere, as shown in Fig. 1.5(b). As the rod is brought close to the sphere, the free electrons in the sphere move away due to repulsion and start piling up at the farther end. The near end becomes positively charged due to deficit of electrons. This process of charge distribution stops when the net force on the free electrons inside the metal is zero. Connect the sphere to the ground by a conducting wire. The electrons will flow to the ground while the positive charges at the near end will remain held there due to the attractive force of the negative charges on the rod, as shown in Fig. 1.5(c). Disconnect the sphere from the ground. The positive charge continues to be held at the near end [Fig. 1.5(d)]. Remove the electrified rod. The positive charge will spread uniformly over the sphere as shown in Fig. 1.5(e).

FIGURE 1.5

In this experiment, the metal sphere gets charged by the process of induction and the rod does not lose any of its charge.

Similar steps are involved in charging a metal sphere negatively by induction, by bringing a positively charged rod near it. In this case the electrons will flow from the ground to the sphere when the sphere is connected to the ground with a wire. Can you explain why?

EXAMPLE 1.1

PHYSICS

Interactive animation on charging a two-sphere system by induction:
http://www.physicsclassroom.com/mmedia/estatics/estaticTOC.html

1.5 Basic Properties of Electric Charge

We have seen that there are two types of charges, namely positive and negative and their effects tend to cancel each other. Here, we shall now describe some other properties of the electric charge.

If the sizes of charged bodies are very small as compared to the distances between them, we treat them as *point charges*. All the charge content of the body is assumed to be concentrated at one point in space.

1.5.1 Additivity of charges

We have not as yet given a quantitative definition of a charge; we shall follow it up in the next section. We shall tentatively assume that this can be done and proceed. If a system contains two point charges q_1 and q_2, the total charge of the system is obtained simply by adding algebraically q_1 and q_2, i.e., charges add up like real numbers or they are scalars like the mass of a body. If a system contains n charges $q_1, q_2, q_3, \ldots, q_n$, then the total charge of the system is $q_1 + q_2 + q_3 + \ldots + q_n$. Charge has magnitude but no direction, similar to the mass. However, there is one difference between mass and charge. Mass of a body is always positive whereas a charge can be either positive or negative. Proper signs have to be used while adding the charges in a system. For example, the total charge of a system containing five charges +1, +2, –3, +4 and –5, in some arbitrary unit, is (+1) + (+2) + (–3) + (+4) + (–5) = –1 in the same unit.

1.5.2 Charge is conserved

We have already hinted to the fact that when bodies are charged by rubbing, there is transfer of electrons from one body to the other; no new charges are either created or destroyed. A picture of particles of electric charge enables us to understand the idea of conservation of charge. When we rub two bodies, what one body gains in charge the other body loses. Within an isolated system consisting of many charged bodies, due to interactions among the bodies, charges may get redistributed but it is found that *the total charge of the isolated system is always conserved.* Conservation of charge has been established experimentally.

It is not possible to create or destroy net charge carried by any isolated system although the charge carrying particles may be created or destroyed in a process. Sometimes nature creates charged particles: a neutron turns into a proton and an electron. The proton and electron thus created have equal and opposite charges and the total charge is zero before and after the creation.

1.5.3 Quantisation of charge

Experimentally it is established that all free charges are integral multiples of a basic unit of charge denoted by e. Thus charge q on a body is always given by

$$q = ne$$

where n is any integer, positive or negative. This basic unit of charge is the charge that an electron or proton carries. By convention, the charge on an electron is taken to be negative; therefore charge on an electron is written as $-e$ and that on a proton as $+e$.

The fact that electric charge is always an integral multiple of e is termed as *quantisation of charge*. There are a large number of situations in physics where certain physical quantities are quantised. The quantisation of charge was first suggested by the experimental laws of electrolysis discovered by English experimentalist Faraday. It was experimentally demonstrated by Millikan in 1912.

In the International System (SI) of Units, a unit of charge is called a *coulomb* and is denoted by the symbol C. A coulomb is defined in terms the unit of the electric current which you are going to learn in a subsequent chapter. In terms of this definition, one coulomb is the charge flowing through a wire in 1 s if the current is 1 A (ampere), (see Chapter 2 of Class XI, Physics Textbook , Part I). In this system, the value of the basic unit of charge is

$$e = 1.602192 \times 10^{-19}\ \text{C}$$

Thus, there are about 6×10^{18} electrons in a charge of -1C. In electrostatics, charges of this large magnitude are seldom encountered and hence we use smaller units 1 μC (micro coulomb) = 10^{-6} C or 1 mC (milli coulomb) = 10^{-3} C.

If the protons and electrons are the only basic charges in the universe, all the observable charges have to be integral multiples of e. Thus, if a body contains n_1 electrons and n_2 protons, the total amount of charge on the body is $n_2 \times e + n_1 \times (-e) = (n_2 - n_1)\, e$. Since n_1 and n_2 are integers, their difference is also an integer. Thus the charge on any body is always an integral multiple of e and can be increased or decreased also in steps of e.

The step size e is, however, very small because at the macroscopic level, we deal with charges of a few μC. At this scale the fact that charge of a body can increase or decrease in units of e is not visible. The grainy nature of the charge is lost and it appears to be continuous.

This situation can be compared with the geometrical concepts of points and lines. A dotted line viewed from a distance appears continuous to us but is not continuous in reality. As many points very close to each other normally give an impression of a continuous line, many small charges taken together appear as a continuous charge distribution.

At the macroscopic level, one deals with charges that are enormous compared to the magnitude of charge e. Since $e = 1.6 \times 10^{-19}$ C, a charge of magnitude, say 1 μC, contains something like 10^{13} times the electronic charge. At this scale, the fact that charge can increase or decrease only in units of e is not very different from saying that charge can take continuous values. Thus, at the macroscopic level, the quantisation of charge has no practical consequence and can be ignored. At the microscopic level, where the charges involved are of the order of a few tens or hundreds of e, i.e.,

they can be counted, they appear in discrete lumps and quantisation of charge cannot be ignored. It is the scale involved that is very important.

EXAMPLE 1.2

Example 1.2 If 10^9 electrons move out of a body to another body every second, how much time is required to get a total charge of 1 C on the other body?

Solution In one second 10^9 electrons move out of the body. Therefore the charge given out in one second is $1.6 \times 10^{-19} \times 10^9$ C = 1.6×10^{-10} C. The time required to accumulate a charge of 1 C can then be estimated to be 1 C ÷ (1.6×10^{-10} C/s) = 6.25×10^9 s = 6.25×10^9 ÷ (365 × 24 × 3600) years = 198 years. Thus to collect a charge of one coulomb, from a body from which 10^9 electrons move out every second, we will need approximately 200 years. One coulomb is, therefore, a very large unit for many practical purposes.

It is, however, also important to know what is roughly the number of electrons contained in a piece of one cubic centimetre of a material. A cubic piece of copper of side 1 cm contains about 2.5×10^{24} electrons.

EXAMPLE 1.3

Example 1.3 How much positive and negative charge is there in a cup of water?

Solution Let us assume that the mass of one cup of water is 250 g. The molecular mass of water is 18g. Thus, one mole (= 6.02×10^{23} molecules) of water is 18 g. Therefore the number of molecules in one cup of water is (250/18) × 6.02×10^{23}.

Each molecule of water contains two hydrogen atoms and one oxygen atom, i.e., 10 electrons and 10 protons. Hence the total positive and total negative charge has the same magnitude. It is equal to (250/18) × $6.02 \times 10^{23} \times 10 \times 1.6 \times 10^{-19}$ C = 1.34×10^7 C.

1.6 COULOMB'S LAW

Coulomb's law is a quantitative statement about the force between two point charges. When the linear size of charged bodies are much smaller than the distance separating them, the size may be ignored and the charged bodies are treated as *point charges*. Coulomb measured the force between two point charges and found that *it varied inversely as the square of the distance between the charges and was directly proportional to the product of the magnitude of the two charges and acted along the line joining the two charges*. Thus, if two point charges q_1, q_2 are separated by a distance r in vacuum, the magnitude of the force (**F**) between them is given by

$$F = k \frac{|q_1 \; q_2|}{r^2} \tag{1.1}$$

How did Coulomb arrive at this law from his experiments? Coulomb used a torsion balance* for measuring the force between two charged metallic

* A torsion balance is a sensitive device to measure force. It was also used later by Cavendish to measure the very feeble gravitational force between two objects, to verify Newton's Law of Gravitation.

spheres. When the separation between two spheres is much larger than the radius of each sphere, the charged spheres may be regarded as point charges. However, the charges on the spheres were unknown, to begin with. How then could he discover a relation like Eq. (1.1)? Coulomb thought of the following simple way: Suppose the charge on a metallic sphere is q. If the sphere is put in contact with an identical uncharged sphere, the charge will spread over the two spheres. By symmetry, the charge on each sphere will be $q/2$*. Repeating this process, we can get charges $q/2$, $q/4$, etc. Coulomb varied the distance for a fixed pair of charges and measured the force for different separations. He then varied the charges in pairs, keeping the distance fixed for each pair. Comparing forces for different pairs of charges at different distances, Coulomb arrived at the relation, Eq. (1.1).

Coulomb's law, a simple mathematical statement, was initially experimentally arrived at in the manner described above. While the original experiments established it at a macroscopic scale, it has also been established down to subatomic level ($r \sim 10^{-10}$ m).

Coulomb discovered his law without knowing the *explicit* magnitude of the charge. In fact, it is the other way round: Coulomb's law can *now* be employed to furnish a definition for a unit of charge. In the relation, Eq. (1.1), k is so far arbitrary. We can choose any positive value of k. The choice of k determines the size of the unit of charge. In SI units, the value of k is about 9×10^9. The unit of charge that results from this choice is called a coulomb which we defined earlier in Section 1.4. Putting this value of k in Eq. (1.1), we see that for $q_1 = q_2 = 1$ C, $r = 1$ m

$$F = 9 \times 10^9 \text{ N}$$

That is, 1 C is the charge that when placed at a distance of 1 m from another charge of the same magnitude *in vacuum* experiences an electrical force of repulsion of magnitude 9×10^9 N. One coulomb is evidently too big a unit to be used. In practice, in electrostatics, one uses smaller units like 1 mC or 1 μC.

The constant k in Eq. (1.1) is usually put as $k = 1/4\pi\varepsilon_0$ for later convenience, so that Coulomb's law is written as

$$F = \frac{1}{4\pi\varepsilon_0} \frac{|q_1 q_2|}{r^2} \qquad (1.2)$$

ε_0 is called the *permittivity of free space* . The value of ε_0 in SI units is

$$\varepsilon_0 = 8.854 \times 10^{-12} \text{ C}^2 \text{ N}^{-1}\text{m}^{-2}$$

Charles Augustin de Coulomb (1736 – 1806) Coulomb, a French physicist, began his career as a military engineer in the West Indies. In 1776, he returned to Paris and retired to a small estate to do his scientific research. He invented a torsion balance to measure the quantity of a force and used it for determination of forces of electric attraction or repulsion between small charged spheres. He thus arrived in 1785 at the inverse square law relation, now known as Coulomb's law. The law had been anticipated by Priestley and also by Cavendish earlier, though Cavendish never published his results. Coulomb also found the inverse square law of force between unlike and like magnetic poles.

CHARLES AUGUSTIN DE COULOMB (1736 –1806)

* Implicit in this is the assumption of additivity of charges and conservation: two charges ($q/2$ each) add up to make a total charge q.

FIGURE 1.6 (a) Geometry and (b) Forces between charges.

Since force is a vector, it is better to write Coulomb's law in the vector notation. Let the position vectors of charges q_1 and q_2 be $\mathbf{r}_1$ and $\mathbf{r}_2$ respectively [see Fig.1.6(a)]. We denote force on q_1 due to q_2 by $\mathbf{F}_{12}$ and force on q_2 due to q_1 by $\mathbf{F}_{21}$. The two point charges q_1 and q_2 have been numbered 1 and 2 for convenience and the vector leading from 1 to 2 is denoted by $\mathbf{r}_{21}$:

$$\mathbf{r}_{21} = \mathbf{r}_2 - \mathbf{r}_1$$

In the same way, the vector leading from 2 to 1 is denoted by $\mathbf{r}_{12}$:

$$\mathbf{r}_{12} = \mathbf{r}_1 - \mathbf{r}_2 = -\mathbf{r}_{21}$$

The magnitude of the vectors $\mathbf{r}_{21}$ and $\mathbf{r}_{12}$ is denoted by r_{21} and r_{12}, respectively ($r_{12} = r_{21}$). The direction of a vector is specified by a unit vector along the vector. To denote the direction from 1 to 2 (or from 2 to 1), we define the unit vectors:

$$\hat{\mathbf{r}}_{21} = \frac{\mathbf{r}_{21}}{r_{21}}, \quad \hat{\mathbf{r}}_{12} = \frac{\mathbf{r}_{12}}{r_{12}}, \quad \hat{\mathbf{r}}_{21} = \hat{\mathbf{r}}_{12}$$

Coulomb's force law between two point charges q_1 and q_2 located at $\mathbf{r}_1$ and $\mathbf{r}_2$ is then expressed as

$$\mathbf{F}_{21} = \frac{1}{4\pi\varepsilon_o} \frac{q_1 \; q_2}{r_{21}^2} \hat{\mathbf{r}}_{21} \tag{1.3}$$

Some remarks on Eq. (1.3) are relevant:

- Equation (1.3) is valid for any sign of q_1 and q_2 whether positive or negative. If q_1 and q_2 are of the same sign (either both positive or both negative), $\mathbf{F}_{21}$ is along $\hat{\mathbf{r}}_{21}$, which denotes repulsion, as it should be for like charges. If q_1 and q_2 are of opposite signs, $\mathbf{F}_{21}$ is along $-\hat{\mathbf{r}}_{21} (= \hat{\mathbf{r}}_{12})$, which denotes attraction, as expected for unlike charges. Thus, we do not have to write separate equations for the cases of like and unlike charges. Equation (1.3) takes care of both cases correctly [Fig. 1.6(b)].
- The force $\mathbf{F}_{12}$ on charge q_1 due to charge q_2, is obtained from Eq. (1.3), by simply interchanging 1 and 2, i.e.,

$$\mathbf{F}_{12} = \frac{1}{4\pi\varepsilon_0} \frac{q_1 \; q_2}{r_{12}^2} \hat{\mathbf{r}}_{12} = -\mathbf{F}_{21}$$

  Thus, Coulomb's law agrees with the Newton's third law.
- Coulomb's law [Eq. (1.3)] gives the force between two charges q_1 and q_2 in vacuum. If the charges are placed in matter or the intervening space has matter, the situation gets complicated due to the presence of charged constituents of matter. We shall consider electrostatics in matter in the next chapter.

Example 1.4 Coulomb's law for electrostatic force between two point charges and Newton's law for gravitational force between two stationary point masses, both have inverse-square dependence on the distance between the charges/masses. (a) Compare the strength of these forces by determining the ratio of their magnitudes (i) for an electron and a proton and (ii) for two protons. (b) Estimate the accelerations of electron and proton due to the electrical force of their mutual attraction when they are 1 Å (= 10^{-10} m) apart? (m_p = 1.67 × 10^{-27} kg, m_e = 9.11 × 10^{-31} kg)

Solution

(a) (i) The electric force between an electron and a proton at a distance r apart is:

$$F_e = -\frac{1}{4\pi\varepsilon_0}\frac{e^2}{r^2}$$

where the negative sign indicates that the force is attractive. The corresponding gravitational force (always attractive) is:

$$F_G = -G\frac{m_p\ m_e}{r^2}$$

where m_p and m_e are the masses of a proton and an electron respectively.

$$\left|\frac{F_e}{F_G}\right| = \frac{e^2}{4\pi\varepsilon_0 G m_p m_e} = 2.4\times10^{39}$$

(ii) On similar lines, the ratio of the magnitudes of electric force to the gravitational force between two protons at a distance r apart is :

$$\left|\frac{F_e}{F_G}\right| = \frac{e^2}{4\pi\varepsilon_0 G m_p m_p} = 1.3 \times 10^{36}$$

However, it may be mentioned here that the signs of the two forces are different. For two protons, the gravitational force is attractive in nature and the Coulomb force is repulsive . The actual values of these forces between two protons inside a nucleus (distance between two protons is ~ 10^{-15} m inside a nucleus) are F_e ~ 230 N whereas F_G ~ 1.9 × 10^{-34} N.

The (dimensionless) ratio of the two forces shows that electrical forces are enormously stronger than the gravitational forces.

(b) The electric force **F** exerted by a proton on an electron is same in magnitude to the force exerted by an electron on a proton; however the masses of an electron and a proton are different. Thus, the magnitude of force is

$$|\mathbf{F}| = \frac{1}{4\pi\varepsilon_0}\frac{e^2}{r^2} = 8.987 \times 10^9\ \text{Nm}^2/\text{C}^2 \times (1.6 \times 10^{-19}\text{C})^2\ /\ (10^{-10}\text{m})^2$$

$$= 2.3 \times 10^{-8}\ \text{N}$$

Using Newton's second law of motion, $F = ma$, the acceleration that an electron will undergo is

a = 2.3×10^{-8} N / 9.11 ×10^{-31} kg = 2.5 × 10^{22} m/s^2

Comparing this with the value of acceleration due to gravity, we can conclude that the effect of gravitational field is negligible on the motion of electron and it undergoes very large accelerations under the action of Coulomb force due to a proton.

The value for acceleration of the proton is

2.3 × 10^{-8} N / 1.67 × 10^{-27} kg = 1.4 × 10^{19} m/s^2

EXAMPLE 1.4

PHYSICS

Interactive animation on Coulomb's law:
http://webphysics.davidson.edu/physlet_resources/bu_semester2/c01_coulomb.html

Example 1.5 A charged metallic sphere A is suspended by a nylon thread. Another charged metallic sphere B held by an insulating handle is brought close to A such that the distance between their centres is 10 cm, as shown in Fig. 1.7(a). The resulting repulsion of A is noted (for example, by shining a beam of light and measuring the deflection of its shadow on a screen). Spheres A and B are touched by uncharged spheres C and D respectively, as shown in Fig. 1.7(b). C and D are then removed and B is brought closer to A to a distance of 5.0 cm between their centres, as shown in Fig. 1.7(c). What is the expected repulsion of A on the basis of Coulomb's law? Spheres A and C and spheres B and D have identical sizes. Ignore the sizes of A and B in comparison to the separation between their centres.

FIGURE 1.7

EXAMPLE 1.5

Solution Let the original charge on sphere A be q and that on B be q'. At a distance r between their centres, the magnitude of the electrostatic force on each is given by

$$F = \frac{1}{4\pi\varepsilon_0}\frac{qq'}{r^2}$$

neglecting the sizes of spheres A and B in comparison to r. When an identical but uncharged sphere C touches A, the charges redistribute on A and C and, by symmetry, each sphere carries a charge $q/2$. Similarly, after D touches B, the redistributed charge on each is $q'/2$. Now, if the separation between A and B is halved, the magnitude of the electrostatic force on each is

$$F' = \frac{1}{4\pi\varepsilon_0}\frac{(q/2)(q'/2)}{(r/2)^2} = \frac{1}{4\pi\varepsilon_0}\frac{(qq')}{r^2} = F$$

Thus the electrostatic force on A, due to B, remains unaltered.

EXAMPLE 1.5

1.7 Forces between Multiple Charges

The mutual electric force between two charges is given by Coulomb's law. How to calculate the force on a charge where there are not one but several charges around? Consider a system of n stationary charges $q_1, q_2, q_3, \ldots, q_n$ in vacuum. What is the force on q_1 due to $q_2, q_3, \ldots, q_n$? Coulomb's law is not enough to answer this question. Recall that forces of mechanical origin add according to the parallelogram law of addition. Is the same true for forces of electrostatic origin?

Experimentally it is verified that *force on any charge due to a number of other charges is the vector sum of all the forces on that charge due to the other charges, taken one at a time. The individual forces are unaffected due to the presence of other charges.* This is termed as the *principle of superposition.*

To better understand the concept, consider a system of three charges q_1, q_2 and q_3, as shown in Fig. 1.8(a). The force on one charge, say q_1, due to two other charges q_2, q_3 can therefore be obtained by performing a vector addition of the forces due to each one of these charges. Thus, if the force on q_1 due to q_2 is denoted by $\mathbf{F}_{12}$, $\mathbf{F}_{12}$ is given by Eq. (1.3) even though other charges are present.

$$\text{Thus,}\quad \mathbf{F}_{12} = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r_{12}^2}\hat{\mathbf{r}}_{12}$$

In the same way, the force on q_1 due to q_3, denoted by $\mathbf{F}_{13}$, is given by

$$\mathbf{F}_{13} = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_3}{r_{13}^2}\hat{\mathbf{r}}_{13}$$

FIGURE 1.8 A system of (a) three charges (b) multiple charges.

which again is the Coulomb force on q_1 due to q_3, even though other charge q_2 is present.

Thus the total force $\mathbf{F}_1$ on q_1 due to the two charges q_2 and q_3 is given as

$$\mathbf{F}_1 = \mathbf{F}_{12} + \mathbf{F}_{13} = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r_{12}^2}\hat{\mathbf{r}}_{12} + \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_3}{r_{13}^2}\hat{\mathbf{r}}_{13} \tag{1.4}$$

The above calculation of force can be generalised to a system of charges more than three, as shown in Fig. 1.8(b).

The principle of superposition says that in a system of charges q_1, q_2, ..., q_n, the force on q_1 due to q_2 is the same as given by Coulomb's law, i.e., it is unaffected by the presence of the other charges q_3, q_4, ..., q_n. The total force $\mathbf{F}_1$ on the charge q_1, due to all other charges, is then given by the vector sum of the forces $\mathbf{F}_{12}$, $\mathbf{F}_{13}$, ..., $\mathbf{F}_{1n}$:

i.e.,

$$\mathbf{F}_1 = \mathbf{F}_{12} + \mathbf{F}_{13} + \ldots + \mathbf{F}_{1n} = \frac{1}{4\pi\varepsilon_0}\left[\frac{q_1 q_2}{r_{12}^2}\hat{\mathbf{r}}_{12} + \frac{q_1 q_3}{r_{13}^2}\hat{\mathbf{r}}_{13} + \ldots + \frac{q_1 q_n}{r_{1n}^2}\hat{\mathbf{r}}_{1n}\right]$$

$$= \frac{q_1}{4\pi\varepsilon_0}\sum_{i=2}^{n}\frac{q_i}{r_{1i}^2}\hat{\mathbf{r}}_{1i} \tag{1.5}$$

The vector sum is obtained as usual by the parallelogram law of addition of vectors. All of electrostatics is basically a consequence of Coulomb's law and the superposition principle.

Example 1.6 Consider three charges q_1, q_2, q_3 each equal to q at the vertices of an equilateral triangle of side l. What is the force on a charge Q (with the same sign as q) placed at the centroid of the triangle, as shown in Fig. 1.9?

FIGURE 1.9

Solution In the given equilateral triangle ABC of sides of length l, if we draw a perpendicular AD to the side BC,

AD = AC cos 30° = $(\sqrt{3}/2)\,l$ and the distance AO of the centroid O from A is (2/3) AD = $(1/\sqrt{3})\,l$. By symmatry AO = BO = CO.

EXAMPLE 1.6

Thus,

Force $\mathbf{F_1}$ on Q due to charge q at A $= \dfrac{3}{4\pi\varepsilon_0}\dfrac{Qq}{l^2}$ along AO

Force $\mathbf{F_2}$ on Q due to charge q at B $= \dfrac{3}{4\pi\varepsilon_0}\dfrac{Qq}{l^2}$ along BO

Force $\mathbf{F_3}$ on Q due to charge q at C $= \dfrac{3}{4\pi\varepsilon_0}\dfrac{Qq}{l^2}$ along CO

The resultant of forces $\mathbf{F_2}$ and $\mathbf{F_3}$ is $\dfrac{3}{4\pi\varepsilon_0}\dfrac{Qq}{l^2}$ along OA, by the parallelogram law. Therefore, the total force on $Q = \dfrac{3}{4\pi\varepsilon_0}\dfrac{Qq}{l^2}(\hat{\mathbf{r}} - \hat{\mathbf{r}})$

$= 0$, where $\hat{\mathbf{r}}$ is the unit vector along OA.
It is clear also by symmetry that the three forces will sum to zero. Suppose that the resultant force was non-zero but in some direction. Consider what would happen if the system was rotated through 60° about O.

EXAMPLE 1.6

Example 1.7 Consider the charges q, q, and $-q$ placed at the vertices of an equilateral triangle, as shown in Fig. 1.10. What is the force on each charge?

FIGURE 1.10

Solution The forces acting on charge q at A due to charges q at B and $-q$ at C are $\mathbf{F}_{12}$ along BA and $\mathbf{F}_{13}$ along AC respectively, as shown in Fig. 1.10. By the parallelogram law, the total force $\mathbf{F}_1$ on the charge q at A is given by

$\mathbf{F}_1 = F\,\hat{\mathbf{r}}_1$ where $\hat{\mathbf{r}}_1$ is a unit vector along BC.

The force of attraction or repulsion for each pair of charges has the same magnitude $F = \dfrac{q^2}{4\pi\varepsilon_0 l^2}$

The total force $\mathbf{F}_2$ on charge q at B is thus $\mathbf{F_2} = F\,\hat{\mathbf{r}}_2$, where $\hat{\mathbf{r}}_2$ is a unit vector along AC.

EXAMPLE 1.7

EXAMPLE 1.7

Similarly the total force on charge $-q$ at C is $\mathbf{F}_3 = \sqrt{3}\ F\ \hat{\mathbf{n}}$, where $\hat{\mathbf{n}}$ is the unit vector along the direction bisecting the $\angle$BCA.
It is interesting to see that the sum of the forces on the three charges is zero, i.e.,

$\mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3 = 0$

The result is not at all surprising. It follows straight from the fact that Coulomb's law is consistent with Newton's third law. The proof is left to you as an exercise.

1.8 ELECTRIC FIELD

Let us consider a point charge Q placed in vacuum, at the origin O. If we place another point charge q at a point P, where **OP** = **r**, then the charge Q will exert a force on q as per Coulomb's law. We may ask the question: If charge q is removed, then what is left in the surrounding? Is there nothing? If there is nothing at the point P, then how does a force act when we place the charge q at P. In order to answer such questions, the early scientists introduced the concept of *field*. According to this, we say that the charge Q produces an electric field everywhere in the surrounding. When another charge q is brought at some point P, the field there acts on it and produces a force. The electric field produced by the charge Q at a point **r** is given as

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0}\frac{Q}{r^2}\hat{\mathbf{r}} = \frac{1}{4\pi\varepsilon_0}\frac{Q}{r^2}\hat{\mathbf{r}} \tag{1.6}$$

where $\hat{\mathbf{r}} = \mathbf{r}/r$, is a unit vector from the origin to the point **r**. Thus, Eq.(1.6) specifies the value of the electric field for each value of the position vector **r**. The word "field" signifies how some distributed quantity (which could be a scalar or a vector) varies with position. The effect of the charge has been incorporated in the existence of the electric field. We obtain the force **F** exerted by a charge Q on a charge q, as

$$\mathbf{F} = \frac{1}{4\pi\varepsilon_0}\frac{Qq}{r^2}\hat{\mathbf{r}} \tag{1.7}$$

Note that the charge q also exerts an equal and opposite force on the charge Q. The electrostatic force between the charges Q and q can be looked upon as an interaction between charge q and the electric field of Q and *vice versa*. If we denote the position of charge q by the vector **r**, it experiences a force **F** equal to the charge q multiplied by the electric field **E** at the location of q. Thus,

$$\mathbf{F}(\mathbf{r}) = q\,\mathbf{E}(\mathbf{r}) \tag{1.8}$$

Equation (1.8) defines the SI unit of electric field as N/C*.

Some important remarks may be made here:

(i) From Eq. (1.8), we can infer that if q is unity, the electric field due to a charge Q is numerically equal to the force exerted by it. Thus, the *electric field due to a charge Q at a point in space may be defined as the force that a unit positive charge would experience if placed*

FIGURE 1.11 Electric field (a) due to a charge Q, (b) due to a charge $-Q$.

* An alternate unit V/m will be introduced in the next chapter.

at that point. The charge Q, which is producing the electric field, is called a *source charge* and the charge q, which tests the effect of a source charge, is called a *test charge*. Note that the source charge Q must remain at its original location. However, if a charge q is brought at any point around Q, Q itself is bound to experience an electrical force due to q and will tend to move. A way out of this difficulty is to make q negligibly small. The force $\mathbf{F}$ is then negligibly small but the ratio $\mathbf{F}/q$ is finite and defines the electric field:

$$\mathbf{E} = \lim_{q \to 0} \left(\frac{\mathbf{F}}{q} \right) \tag{1.9}$$

A practical way to get around the problem (of keeping Q undisturbed in the presence of q) is to hold Q to its location by unspecified forces! This may look strange but actually this is what happens in practice. When we are considering the electric force on a test charge q due to a charged planar sheet (Section 1.15), the charges on the sheet are held to their locations by the forces due to the unspecified charged constituents inside the sheet.

(ii) Note that the electric field $\mathbf{E}$ due to Q, though defined operationally in terms of some test charge q, is independent of q. This is because $\mathbf{F}$ is proportional to q, so the ratio $\mathbf{F}/q$ does not depend on q. The force $\mathbf{F}$ on the charge q due to the charge Q depends on the particular location of charge q which may take any value in the space around the charge Q. Thus, the electric field $\mathbf{E}$ due to Q is also dependent on the space coordinate $\mathbf{r}$. For different positions of the charge q all over the space, we get different values of electric field $\mathbf{E}$. The field exists at every point in three-dimensional space.

(iii) For a positive charge, the electric field will be directed radially outwards from the charge. On the other hand, if the source charge is negative, the electric field vector, at each point, points radially inwards.

(iv) Since the magnitude of the force $\mathbf{F}$ on charge q due to charge Q depends only on the distance r of the charge q from charge Q, the magnitude of the electric field $\mathbf{E}$ will also depend only on the distance r. Thus at equal distances from the charge Q, the magnitude of its electric field $\mathbf{E}$ is same. The magnitude of electric field $\mathbf{E}$ due to a point charge is thus same on a sphere with the point charge at its centre; in other words, it has a spherical symmetry.

1.8.1 Electric field due to a system of charges

Consider a system of charges $q_1, q_2, \ldots, q_n$ with position vectors $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n$ relative to some origin O. Like the electric field at a point in space due to a single charge, electric field at a point in space due to the system of charges is defined to be the force experienced by a unit test charge placed at that point, without disturbing the original positions of charges $q_1, q_2, \ldots, q_n$. We can use Coulomb's law and the superposition principle to determine this field at a point P denoted by position vector $\mathbf{r}$.

FIGURE 1.12 Electric field at a point due to a system of charges is the vector sum of the electric fields at the point due to individual charges.

Electric field $\mathbf{E}_1$ at $\mathbf{r}$ due to q_1 at $\mathbf{r}_1$ is given by

$$\mathbf{E}_1 = \frac{1}{4\pi\varepsilon_0}\frac{q_1}{r_{1P}^2}\hat{\mathbf{r}}_{1P}$$

where $\hat{\mathbf{r}}_{1P}$ is a unit vector in the direction from q_1 to P, and r_{1P} is the distance between q_1 and P.

In the same manner, electric field $\mathbf{E}_2$ at $\mathbf{r}$ due to q_2 at $\mathbf{r}_2$ is

$$\mathbf{E}_2 = \frac{1}{4\pi\varepsilon_0}\frac{q_2}{r_{2P}^2}\hat{\mathbf{r}}_{2P}$$

where $\hat{\mathbf{r}}_{2P}$ is a unit vector in the direction from q_2 to P and r_{2P} is the distance between q_2 and P. Similar expressions hold good for fields $\mathbf{E}_3$, $\mathbf{E}_4$, ..., $\mathbf{E}_n$ due to charges q_3, q_4, ..., q_n.

By the superposition principle, the electric field $\mathbf{E}$ at $\mathbf{r}$ due to the system of charges is (as shown in Fig. 1.12)

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}_1(\mathbf{r}) + \mathbf{E}_2(\mathbf{r}) + \ldots + \mathbf{E}_n(\mathbf{r})$$

$$= \frac{1}{4\pi\varepsilon_0}\frac{q_1}{r_{1P}^2}\hat{\mathbf{r}}_{1P} + \frac{1}{4\pi\varepsilon_0}\frac{q_2}{r_{2P}^2}\hat{\mathbf{r}}_{2P} + \ldots + \frac{1}{4\pi\varepsilon_0}\frac{q_n}{r_{nP}^2}\hat{\mathbf{r}}_{nP}$$

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0}\sum_{i=1}^{n}\frac{q_i}{r_{iP}^2}\hat{\mathbf{r}}_{iP} \qquad (1.10)$$

$\mathbf{E}$ is a vector quantity that varies from one point to another point in space and is determined from the positions of the source charges.

1.8.2 Physical significance of electric field

You may wonder why the notion of electric field has been introduced here at all. After all, for any system of charges, the measurable quantity is the force on a charge which can be directly determined using Coulomb's law and the superposition principle [Eq. (1.5)]. Why then introduce this intermediate quantity called the electric field?

For electrostatics, the concept of electric field is convenient, but not really necessary. Electric field is an elegant way of characterising the electrical environment of a system of charges. Electric field at a point in the space around a system of charges tells you the force a unit positive test charge would experience if placed at that point (without disturbing the system). Electric field is a characteristic of the system of charges and is independent of the test charge that you place at a point to determine the field. The term *field* in physics generally refers to a quantity that is defined at every point in space and may vary from point to point. Electric field is a vector field, since force is a vector quantity.

The true physical significance of the concept of electric field, however, emerges only when we go beyond electrostatics and deal with time-dependent electromagnetic phenomena. Suppose we consider the force between two distant charges q_1, q_2 in accelerated motion. Now the greatest speed with which a signal or information can go from one point to another is c, the speed of light. Thus, the effect of any motion of q_1 on q_2 cannot

arise instantaneously. There will be some time delay between the effect (force on q_2) and the cause (motion of q_1). It is precisely here that the notion of electric field (strictly, electromagnetic field) is natural and very useful. *The field picture is this: the accelerated motion of charge q_1 produces electromagnetic waves, which then propagate with the speed c, reach q_2 and cause a force on q_2.* The notion of field elegantly accounts for the time delay. Thus, even though electric and magnetic fields can be detected only by their effects (forces) on charges, they are regarded as physical entities, not merely mathematical constructs. They have an *independent dynamics* of their own, i.e., they evolve according to laws of their *own*. They can also transport energy. Thus, a source of time-dependent electromagnetic fields, turned on briefly and switched off, leaves behind propagating electromagnetic fields transporting energy. The concept of field was first introduced by Faraday and is now among the central concepts in physics.

Example 1.8 An electron falls through a distance of 1.5 cm in a uniform electric field of magnitude 2.0×10^4 N C^{-1} [Fig. 1.13(a)]. The direction of the field is reversed keeping its magnitude unchanged and a proton falls through the same distance [Fig. 1.13(b)]. Compute the time of fall in each case. Contrast the situation with that of 'free fall under gravity'.

FIGURE 1.13

Solution In Fig. 1.13(a) the field is upward, so the negatively charged electron experiences a downward force of magnitude eE where E is the magnitude of the electric field. The acceleration of the electron is

$$a_e = eE/m_e$$

where m_e is the mass of the electron.

Starting from rest, the time required by the electron to fall through a distance h is given by $t_e = \sqrt{\frac{2h}{a_e}} = \sqrt{\frac{2hm_e}{eE}}$

For $e = 1.6 \times 10^{-19}$C, $m_e = 9.11 \times 10^{-31}$ kg,

$E = 2.0 \times 10^4$ N C^{-1}, $h = 1.5 \times 10^{-2}$ m,

$t_e = 2.9 \times 10^{-9}$s

In Fig. 1.13 (b), the field is downward, and the positively charged proton experiences a downward force of magnitude eE. The acceleration of the proton is

$$a_p = eE/m_p$$

where m_p is the mass of the proton; $m_p = 1.67 \times 10^{-27}$ kg. The time of fall for the proton is

EXAMPLE 1.8

EXAMPLE 1.8

$$t_p = \sqrt{\frac{2h}{a_p}} = \sqrt{\frac{2hm_p}{eE}} = 1.3\times10^{-7}\,\text{s}$$

Thus, the heavier particle (proton) takes a greater time to fall through the same distance. This is in basic contrast to the situation of 'free fall under gravity' where the time of fall is independent of the mass of the body. Note that in this example we have ignored the acceleration due to gravity in calculating the time of fall. To see if this is justified, let us calculate the acceleration of the proton in the given electric field:

$$a_p = \frac{eE}{m_p}$$

$$= \frac{(1.6 \times 10^{-19}\ \text{C}) \times (2.0 \times 10^{4}\ \text{N C}^{-1})}{1.67 \times 10^{-27}\ \text{kg}}$$

$$= 1.9 \times 10^{12}\ \text{m s}^{-2}$$

which is enormous compared to the value of g (9.8 m s^{-2}), the acceleration due to gravity. The acceleration of the electron is even greater. Thus, the effect of acceleration due to gravity can be ignored in this example.

EXAMPLE 1.9

Example 1.9 Two point charges q_1 and q_2, of magnitude $+10^{-8}$ C and -10^{-8} C, respectively, are placed 0.1 m apart. Calculate the electric fields at points A, B and C shown in Fig. 1.14.

FIGURE 1.14

Solution The electric field vector $\mathbf{E}_{1A}$ at A due to the positive charge q_1 points towards the right and has a magnitude

$$E_{1A} = \frac{(9\times10^{9}\,\text{Nm}^2\text{C}^{-2})\times(10^{-8}\,\text{C})}{(0.05\,\text{m})^2} = 3.6 \times 10^{4}\ \text{N C}^{-1}$$

The electric field vector $\mathbf{E}_{2A}$ at A due to the negative charge q_2 points towards the right and has the same magnitude. Hence the magnitude of the total electric field E_A at A is

$E_A = E_{1A} + E_{2A} = 7.2 \times 10^{4}$ N C^{-1}

$\mathbf{E}_A$ is directed toward the right.

The electric field vector $\mathbf{E}_{1B}$ at B due to the positive charge q_1 points towards the left and has a magnitude

$$E_{1B} = \frac{(9 \times 10^9 \text{Nm}^2\text{C}^{-2}) \times (10^{-8}\text{C})}{(0.05\,\text{m})^2} = 3.6 \times 10^4 \text{ N C}^{-1}$$

The electric field vector $\mathbf{E}_{2B}$ at B due to the negative charge q_2 points towards the right and has a magnitude

$$E_{2B} = \frac{(9 \times 10^9 \text{Nm}^2\text{C}^{-2}) \times (10^{-8}\text{C})}{(0.15\,\text{m})^2} = 4 \times 10^3 \text{ N C}^{-1}$$

The magnitude of the total electric field at B is
$E_B = E_{1B} - E_{2B} = 3.2 \times 10^4 \text{ N C}^{-1}$
$\mathbf{E}_B$ is directed towards the left.
The magnitude of each electric field vector at point C, due to charge q_1 and q_2 is

$$E_{1C} = E_{2C} = \frac{(9 \times 10^9 \text{Nm}^2\text{C}^{-2}) \times (10^{-8}\text{C})}{(0.10\,\text{m})^2} = 9 \times 10^3 \text{ N C}^{-1}$$

The directions in which these two vectors point are indicated in Fig. 1.14. The resultant of these two vectors is

$$E_C = E_1 \cos\frac{\pi}{3} + E_2 \cos\frac{\pi}{3} = 9 \times 10^3 \text{ N C}^{-1}$$

$\mathbf{E}_C$ points towards the right.

EXAMPLE 1.9

1.9 Electric Field Lines

We have studied electric field in the last section. It is a vector quantity and can be represented as we represent vectors. Let us try to represent **E** due to a point charge pictorially. Let the point charge be placed at the origin. Draw vectors pointing along the direction of the electric field with their lengths proportional to the strength of the field at each point. Since the magnitude of electric field at a point decreases inversely as the square of the distance of that point from the charge, the vector gets shorter as one goes away from the origin, always pointing radially outward. Figure 1.15 shows such a picture. In this figure, each arrow indicates the electric field, i.e., the force acting on a unit positive charge, placed at the tail of that arrow. Connect the arrows pointing in one direction and the resulting figure represents a field line. We thus get many field lines, all pointing outwards from the point charge. Have we lost the information about the strength or magnitude of the field now, because it was contained in the length of the arrow? No. Now the magnitude of the field is represented by the density of field lines. **E** is strong near the charge, so the density of field lines is more near the charge and the lines are closer. Away from the charge, the field gets weaker and the density of field lines is less, resulting in well-separated lines.

FIGURE 1.15 Field of a point charge.

Another person may draw more lines. But the number of lines is not important. In fact, an infinite number of lines can be drawn in any region.

FIGURE 1.16 Dependence of electric field strength on the distance and its relation to the number of field lines.

It is the relative density of lines in different regions which is important.

We draw the figure on the plane of paper, *i.e.*, in two-dimensions but we live in three-dimensions. So if one wishes to estimate the density of field lines, one has to consider the number of lines per unit cross-sectional area, perpendicular to the lines. Since the electric field decreases as the square of the distance from a point charge and the area enclosing the charge increases as the square of the distance, the number of field lines crossing the enclosing area remains constant, whatever may be the distance of the area from the charge.

We started by saying that the field lines carry information about the direction of electric field at different points in space. Having drawn a certain set of field lines, the relative density (i.e., closeness) of the field lines at different points indicates the relative strength of electric field at those points. The field lines crowd where the field is strong and are spaced apart where it is weak. Figure 1.16 shows a set of field lines. We can imagine two equal and small elements of area placed at points R and S normal to the field lines there. The number of field lines in our picture cutting the area elements is proportional to the magnitude of field at these points. The picture shows that the field at R is stronger than at S.

To understand the dependence of the field lines on the area, or rather the *solid angle* subtended by an area element, let us try to relate the area with the solid angle, a generalization of angle to three dimensions. Recall how a (plane) angle is defined in two-dimensions. Let a small transverse line element Δl be placed at a distance r from a point O. Then the angle subtended by Δl at O can be approximated as $\Delta\theta = \Delta l/r$. Likewise, in three-dimensions the solid angle* subtended by a small perpendicular plane area ΔS, at a distance r, can be written as $\Delta\Omega = \Delta S/r^2$. We know that in a given solid angle the number of radial field lines is the same. In Fig. 1.16, for two points P_1 and P_2 at distances r_1 and r_2 from the charge, the element of area subtending the solid angle $\Delta\Omega$ is $r_1^2\,\Delta\Omega$ at P_1 and an element of area $r_2^2\,\Delta\Omega$ at P_2, respectively. The number of lines (say n) cutting these area elements are the same. The number of field lines, cutting unit area element is therefore $n/(r_1^2\,\Delta\Omega)$ at P_1 and $n/(r_2^2\,\Delta\Omega)$ at P_2, respectively. Since n and $\Delta\Omega$ are common, the strength of the field clearly has a $1/r^2$ dependence.

The picture of field lines was invented by Faraday to develop an intuitive non- mathematical way of visualizing electric fields around charged configurations. Faraday called them *lines of force. This term is somewhat misleading, especially in case of magnetic fields. The more appropriate term is field lines (electric or magnetic) that we have adopted in this book.*

Electric field lines are thus a way of pictorially mapping the electric field around a configuration of charges. An electric field line is, in general,

* Solid angle is a measure of a cone. Consider the intersection of the given cone with a sphere of radius R. The solid angle $\Delta\Omega$ of the cone is defined to be equal to $\Delta S/R^2$, where ΔS is the area on the sphere cut out by the cone.

a curve drawn in such a way that the tangent to it at each point is in the direction of the net field at that point. An arrow on the curve is obviously necessary to specify the direction of electric field from the two possible directions indicated by a tangent to the curve. A field line is a space curve, i.e., a curve in three dimensions.

Figure 1.17 shows the field lines around some simple charge configurations. As mentioned earlier, the field lines are in 3-dimensional space, though the figure shows them only in a plane. The field lines of a single positive charge are radially outward while those of a single negative charge are radially inward. The field lines around a system of two positive charges (q, q) give a vivid pictorial description of their mutual repulsion, while those around the configuration of two equal and opposite charges (q, $-q$), a dipole, show clearly the mutual attraction between the charges. The field lines follow some important general properties:

(i) Field lines start from positive charges and end at negative charges. If there is a single charge, they may start or end at infinity.

(ii) In a charge-free region, electric field lines can be taken to be continuous curves without any breaks.

(iii) Two field lines can never cross each other. (If they did, the field at the point of intersection will not have a unique direction, which is absurd.)

(iv) Electro*static* field lines do not form any closed loops. This follows from the conservative nature of electric field (Chapter 2).

1.10 Electric Flux

Consider flow of a liquid with velocity **v**, through a small flat surface dS, in a direction normal to the surface. The rate of flow of liquid is given by the volume crossing the area per unit time v dS and represents the flux of liquid flowing across the plane. If the normal to the surface is not parallel to the direction of flow of liquid, *i.e.*, to **v**, but makes an angle θ with it, the projected area in a plane perpendicular to **v** is v dS cos θ. Therefore the flux going out of the surface dS is $\mathbf{v}\cdot\hat{\mathbf{n}}$ dS.

For the case of the electric field, we define an analogous quantity and call it *electric flux*.

We should however note that there is no *flow* of a physically observable quantity unlike the case of liquid flow.

In the picture of electric field lines described above, we saw that the number of field lines crossing a unit area, placed normal to the field at a point is a measure of the strength of electric field at that point. This means that if

FIGURE 1.17 Field lines due to some simple charge configurations.

FIGURE 1.18 Dependence of flux on the inclination θ between **E** and $\hat{\mathbf{n}}$.

we place a small planar element of area ΔS normal to **E** at a point, the number of field lines crossing it is proportional* to $E\,\Delta S$. Now suppose we tilt the area element by angle θ. Clearly, the number of field lines crossing the area element will be smaller. The projection of the area element normal to E is $\Delta S \cos\theta$. Thus, the number of field lines crossing ΔS is proportional to $E\,\Delta S \cos\theta$. When $\theta = 90°$, field lines will be parallel to ΔS and will not cross it at all (Fig. 1.18).

The orientation of area element and not merely its magnitude is important in many contexts. For example, in a stream, the amount of water flowing through a ring will naturally depend on how you hold the ring. If you hold it normal to the flow, maximum water will flow through it than if you hold it with some other orientation. This shows that an area element should be treated as a vector. It has a magnitude and also a direction. How to specify the direction of a planar area? Clearly, the normal to the plane specifies the orientation of the plane. Thus the direction of a planar area vector is along its normal.

How to associate a vector to the area of a curved surface? We imagine dividing the surface into a large number of very small area elements. Each small area element may be treated as planar and a vector associated with it, as explained before.

Notice one ambiguity here. The direction of an area element is along its normal. But a normal can point in two directions. Which direction do we choose as the direction of the vector associated with the area element? This problem is resolved by some convention appropriate to the given context. For the case of a closed surface, this convention is very simple. The vector associated with every area element of a closed surface is taken to be in the direction of the *outward* normal. This is the convention used in Fig. 1.19. Thus, the area element vector $\Delta\mathbf{S}$ at a point on a closed surface equals $\Delta S\,\hat{\mathbf{n}}$ where ΔS is the magnitude of the area element and $\hat{\mathbf{n}}$ is a unit vector in the direction of outward normal at that point.

FIGURE 1.19 Convention for defining normal $\hat{\mathbf{n}}$ and ΔS.

We now come to the definition of electric flux. Electric flux $\Delta\phi$ through an area element $\Delta\mathbf{S}$ is defined by

$$\Delta\phi = \mathbf{E}\cdot\Delta\mathbf{S} = E\,\Delta S\,\cos\theta \qquad (1.11)$$

which, as seen before, is proportional to the number of field lines cutting the area element. The angle θ here is the angle between **E** and $\Delta\mathbf{S}$. For a closed surface, with the convention stated already, θ is the angle between **E** and the outward normal to the area element. Notice we could look at the expression $E\,\Delta S\cos\theta$ in two ways: E $(\Delta S \cos\theta)$ i.e., E times the

* It will not be proper to say that the number of field lines is equal to $E\Delta S$. The number of field lines is after all, a matter of how many field lines we choose to draw. What is physically significant is the relative number of field lines crossing a given area at different points.

projection of area normal to **E**, or $E_{\perp}\,\Delta S$, i.e., component of **E** along the normal to the area element times the magnitude of the area element. The unit of electric flux is N C^{-1} m^2.

The basic definition of electric flux given by Eq. (1.11) can be used, in principle, to calculate the total flux through any given surface. All we have to do is to divide the surface into small area elements, calculate the flux at each element and add them up. Thus, the total flux ϕ through a surface S is

$$\phi \simeq \Sigma\, \mathbf{E}\cdot\Delta\mathbf{S} \qquad (1.12)$$

The approximation sign is put because the electric field **E** is taken to be constant over the small area element. This is mathematically exact only when you take the limit $\Delta S \rightarrow 0$ and the sum in Eq. (1.12) is written as an integral.

1.11 Electric Dipole

An electric dipole is a pair of equal and opposite point charges q and $-q$, separated by a distance $2a$. The line connecting the two charges defines a direction in space. By convention, the direction from $-q$ to q is said to be the direction of the dipole. The mid-point of locations of $-q$ and q is called the centre of the dipole.

The total charge of the electric dipole is obviously zero. This does not mean that the field of the electric dipole is zero. Since the charge q and $-q$ are separated by some distance, the electric fields due to them, when added, do not exactly cancel out. However, at distances much larger than the separation of the two charges forming a dipole ($r >> 2a$), the fields due to q and $-q$ nearly cancel out. The electric field due to a dipole therefore falls off, at large distance, faster than like $1/r^2$ (the dependence on r of the field due to a single charge q). These qualitative ideas are borne out by the explicit calculation as follows:

1.11.1 The field of an electric dipole

The electric field of the pair of charges ($-q$ and q) at any point in space can be found out from Coulomb's law and the superposition principle. The results are simple for the following two cases: (i) when the point is on the dipole axis, and (ii) when it is in the *equatorial plane* of the dipole, i.e., on a plane perpendicular to the dipole axis through its centre. The electric field at any general point P is obtained by adding the electric fields $\mathbf{E}_{-q}$ due to the charge $-q$ and $\mathbf{E}_{+q}$ due to the charge q, by the parallelogram law of vectors.

(i) For points on the axis

Let the point P be at distance r from the centre of the dipole on the side of the charge q, as shown in Fig. 1.20(a). Then

$$\mathbf{E}_{-q} = -\frac{q}{4\pi\varepsilon_0 (r+a)^2}\hat{\mathbf{p}} \qquad [1.13(a)]$$

where $\hat{\mathbf{p}}$ is the unit vector along the dipole axis (from $-q$ to q). Also

$$\mathbf{E}_{+q} = \frac{q}{4\pi\varepsilon_0 (r-a)^2}\hat{\mathbf{p}} \qquad [1.13(b)]$$

FIGURE 1.20 Electric field of a dipole at (a) a point on the axis, (b) a point on the equatorial plane of the dipole. **p** is the dipole moment vector of magnitude $p = q \times 2a$ and directed from $-q$ to q.

The total field at P is

$$\mathbf{E} = \mathbf{E}_{+q} + \mathbf{E}_{-q} = \frac{q}{4\pi\varepsilon_0}\left[\frac{1}{(r-a)^2} - \frac{1}{(r+a)^2}\right]\hat{\mathbf{p}}$$

$$= \frac{q}{4\pi\varepsilon_o}\frac{4ar}{(r^2-a^2)^2}\hat{\mathbf{p}} \qquad (1.14)$$

For $r >> a$

$$\mathbf{E} = \frac{4qa}{4\pi\varepsilon_0 r^3}\hat{\mathbf{p}} \qquad (r >> a) \qquad (1.15)$$

(ii) For points on the equatorial plane

The magnitudes of the electric fields due to the two charges $+q$ and $-q$ are given by

$$E_{+q} = \frac{q}{4\pi\varepsilon_0}\frac{1}{r^2+a^2} \qquad [1.16(a)]$$

$$E_{-q} = \frac{q}{4\pi\varepsilon_0}\frac{1}{r^2+a^2} \qquad [1.16(b)]$$

and are equal.

The directions of $\mathbf{E}_{+q}$ and $\mathbf{E}_{-q}$ are as shown in Fig. 1.20(b). Clearly, the components normal to the dipole axis cancel away. The components along the dipole axis add up. The total electric field is opposite to $\hat{\mathbf{p}}$. We have

$$\mathbf{E} = -(E_{+q} + E_{-q})\cos\theta\ \hat{\mathbf{p}}$$

$$= -\frac{2qa}{4\pi\varepsilon_o(r^2+a^2)^{3/2}}\hat{\mathbf{p}} \qquad (1.17)$$

At large distances ($r >> a$), this reduces to

$$\mathbf{E} = -\frac{2qa}{4\pi\varepsilon_o r^3}\hat{\mathbf{p}} \qquad (r >> a) \qquad (1.18)$$

From Eqs. (1.15) and (1.18), it is clear that the dipole field at large distances does not involve q and a separately; it depends on the product qa. This suggests the definition of dipole moment. The *dipole moment vector* **p** of an electric dipole is defined by

$$\mathbf{p} = q \times 2a\ \hat{\mathbf{p}} \qquad (1.19)$$

that is, it is a vector whose magnitude is charge q times the separation $2a$ (between the pair of charges q, $-q$) and the direction is along the line from $-q$ to q. In terms of **p**, the electric field of a dipole at large distances takes simple forms:

At a point on the dipole axis

$$\mathbf{E} = \frac{2\mathbf{p}}{4\pi\varepsilon_o r^3} \qquad (r >> a) \qquad (1.20)$$

At a point on the equatorial plane

$$\mathbf{E} = -\frac{\mathbf{p}}{4\pi\varepsilon_o r^3} \qquad (r >> a) \qquad (1.21)$$

Notice the important point that the dipole field at large distances falls off not as $1/r^2$ but as $1/r^3$. Further, the magnitude and the direction of the dipole field depends not only on the distance r but also on the *angle* between the position vector **r** and the dipole moment **p**.

We can think of the limit when the dipole size $2a$ approaches zero, the charge q approaches infinity in such a way that the product $p = q \times 2a$ is finite. Such a dipole is referred to as a *point dipole*. For a point dipole, Eqs. (1.20) and (1.21) are exact, true for any r.

1.11.2 Physical significance of dipoles

In most molecules, the centres of positive charges and of negative charges* lie at the same place. Therefore, their dipole moment is zero. CO_2 and CH_4 are of this type of molecules. However, they develop a dipole moment when an electric field is applied. But in some molecules, the centres of negative charges and of positive charges do not coincide. Therefore they have a permanent electric dipole moment, even in the absence of an electric field. Such molecules are called polar molecules. Water molecules, H_2O, is an example of this type. Various materials give rise to interesting properties and important applications in the presence or absence of electric field.

Example 1.10 Two charges ±10 μC are placed 5.0 mm apart. Determine the electric field at (a) a point P on the axis of the dipole 15 cm away from its centre O on the side of the positive charge, as shown in Fig. 1.21(a), and (b) a point Q, 15 cm away from O on a line passing through O and normal to the axis of the dipole, as shown in Fig. 1.21(b).

FIGURE 1.21

EXAMPLE 1.10

* Centre of a collection of positive point charges is defined much the same way as the centre of mass: $\mathbf{r}_{cm} = \frac{\sum_i q_i \mathbf{r}_i}{\sum_i q_i}$.

Solution (a) Field at P due to charge +10 μC

$$= \frac{10^{-5}\,\text{C}}{4\pi(8.854\times10^{-12}\,\text{C}^2\,\text{N}^{-1}\,\text{m}^{-2})} \times \frac{1}{(15-0.25)^2\times10^{-4}\,\text{m}^2}$$

$= 4.13 \times 10^6$ N C^{-1} along BP

Field at P due to charge –10 μC

$$= \frac{10^{-5}\,\text{C}}{4\pi(8.854\times10^{-12}\,\text{C}^2\,\text{N}^{-1}\,\text{m}^{-2})} \times \frac{1}{(15+0.25)^2\times10^{-4}\,\text{m}^2}$$

$= 3.86 \times 10^6$ N C^{-1} along PA

The resultant electric field at P due to the two charges at A and B is $= 2.7 \times 10^5$ N C^{-1} along BP.

In this example, the ratio OP/OB is quite large (= 60). Thus, we can expect to get approximately the same result as above by directly using the formula for electric field at a far-away point on the axis of a dipole. For a dipole consisting of charges ± q, $2a$ distance apart, the electric field at a distance r from the centre on the axis of the dipole has a magnitude

$$E = \frac{2p}{4\pi\varepsilon_0 r^3} \qquad (r/a >> 1)$$

where $p = 2a\,q$ is the magnitude of the dipole moment.

The direction of electric field on the dipole axis is always along the direction of the dipole moment vector (i.e., from $-q$ to q). Here, $p = 10^{-5}$ C $\times 5 \times 10^{-3}$ m $= 5 \times 10^{-8}$ C m

Therefore,

$$E = \frac{2\times5\times10^{-8}\,\text{C m}}{4\pi(8.854\times10^{-12}\,\text{C}^2\,\text{N}^{-1}\,\text{m}^{-2})} \times \frac{1}{(15)^3\times10^{-6}\,\text{m}^3} = 2.6 \times 10^5 \text{ N C}^{-1}$$

along the dipole moment direction AB, which is close to the result obtained earlier.

(b) Field at Q due to charge + 10 μC at B

$$= \frac{10^{-5}\,\text{C}}{4\,\pi\,(8.854\times10^{-12}\,\text{C}^2\,\text{N}^{-1}\,\text{m}^{-2})} \times \frac{1}{[15^2+(0.25)^2]\times10^{-4}\,\text{m}^2}$$

$= 3.99 \times 10^6$ N C^{-1} along BQ

Field at Q due to charge –10 μC at A

$$= \frac{10^{-5}\,\text{C}}{4\,\pi(8.854\times10^{-12}\,\text{C}^2\,\text{N}^{-1}\,\text{m}^{-2})} \times \frac{1}{[15^2+(0.25)^2]\times10^{-4}\,\text{m}^2}$$

$= 3.99 \times 10^6$ N C^{-1} along QA.

Clearly, the components of these two forces with equal magnitudes cancel along the direction OQ but add up along the direction parallel to BA. Therefore, the resultant electric field at Q due to the two charges at A and B is

$$= 2 \times \frac{0.25}{\sqrt{15^2 + (0.25)^2}} \times 3.99 \times 10^6\,\text{N C}^{-1} \text{ along BA}$$

$= 1.33 \times 10^5$ N C^{-1} along BA.

As in (a), we can expect to get approximately the same result by directly using the formula for dipole field at a point on the normal to the axis of the dipole:

EXAMPLE 1.10

$$E = \frac{p}{4\pi\varepsilon_0 r^3} \qquad (r/a >> 1)$$

$$= \frac{5\times10^{-8}\ \text{C m}}{4\pi(8.854\times10^{-12}\ \text{C}^2\ \text{N}^{-1}\ \text{m}^{-2})} \times \frac{1}{(15)^3\times10^{-6}\ \text{m}^3}$$

$= 1.33 \times 10^5\ \text{N C}^{-1}$.

The direction of electric field in this case is opposite to the direction of the dipole moment vector. Again the result agrees with that obtained before.

EXAMPLE 1.10

1.12 Dipole in a Uniform External Field

Consider a permanent dipole of dipole moment **p** in a uniform external field **E**, as shown in Fig. 1.22. (By permanent dipole, we mean that **p** exists irrespective of **E**; it has not been induced by **E**.)

There is a force q**E** on q and a force $-q$**E** on $-q$. The net force on the dipole is zero, since **E** is uniform. However, the charges are separated, so the forces act at different points, resulting in a torque on the dipole. When the net force is zero, the torque (couple) is independent of the origin. Its magnitude equals the magnitude of each force multiplied by the arm of the couple (perpendicular distance between the two antiparallel forces).

FIGURE 1.22 Dipole in a uniform electric field.

Magnitude of torque $= q\,E \times 2\,a \sin\theta$

$= 2\,q\,a\,E \sin\theta$

Its direction is normal to the plane of the paper, coming out of it.

The magnitude of $\mathbf{p} \times \mathbf{E}$ is also $p\,E \sin\theta$ and its direction is normal to the paper, coming out of it. Thus,

$$\tau = \mathbf{p} \times \mathbf{E} \qquad (1.22)$$

This torque will tend to align the dipole with the field **E**. When **p** is aligned with **E**, the torque is zero.

What happens if the field is not uniform? In that case, the net force will evidently be non-zero. In addition there will, in general, be a torque on the system as before. The general case is involved, so let us consider the simpler situations when **p** is parallel to **E** or antiparallel to **E**. In either case, the net torque is zero, but there is a net force on the dipole if **E** is not uniform.

Figure 1.23 is self-explanatory. It is easily seen that when **p** is parallel to **E**, the dipole has a net force in the direction of increasing field. When **p** is antiparallel to **E**, the net force on the dipole is in the direction of decreasing field. In general, the force depends on the orientation of **p** with respect to **E**.

FIGURE 1.23 Electric force on a dipole: (a) **E** parallel to **p**, (b) **E** antiparallel to **p**.

This brings us to a common observation in frictional electricity. A comb run through dry hair attracts pieces of paper. The comb, as we know, acquires charge through friction. But the paper is not charged. What then explains the attractive force? Taking the clue from the preceding

discussion, the charged comb 'polarizes' the piece of paper, i.e., induces a net dipole moment in the direction of field. Further, the electric field due to the comb is not uniform. In this situation, it is easily seen that the paper should move in the direction of the comb!

1.13 Continuous Charge Distribution

We have so far dealt with charge configurations involving discrete charges $q_1, q_2, \ldots, q_n$. One reason why we restricted to discrete charges is that the mathematical treatment is simpler and does not involve calculus. For many purposes, however, it is impractical to work in terms of discrete charges and we need to work with continuous charge distributions. For example, on the surface of a charged conductor, it is impractical to specify the charge distribution in terms of the locations of the microscopic charged constituents. It is more feasible to consider an area element ΔS (Fig. 1.24) on the surface of the conductor (which is very small on the macroscopic scale but big enough to include a very large number of electrons) and specify the charge ΔQ on that element. We then define a *surface charge density* σ at the area element by

$$\sigma = \frac{\Delta Q}{\Delta S} \tag{1.23}$$

We can do this at different points on the conductor and thus arrive at a continuous function σ, called the surface charge density. The surface charge density σ so defined ignores the quantisation of charge and the discontinuity in charge distribution at the microscopic level*. σ represents macroscopic surface charge density, which in a sense, is a smoothed out average of the microscopic charge density over an area element ΔS which, as said before, is large microscopically but small macroscopically. The units for σ are C/m^2.

Similar considerations apply for a line charge distribution and a volume charge distribution. The *linear charge density* λ of a wire is defined by

$$\lambda = \frac{\Delta Q}{\Delta l} \tag{1.24}$$

where Δl is a small line element of wire on the macroscopic scale that, however, includes a large number of microscopic charged constituents, and ΔQ is the charge contained in that line element. The units for λ are C/m. The *volume charge density* (sometimes simply called charge density) is defined in a similar manner:

$$\rho = \frac{\Delta Q}{\Delta V} \tag{1.25}$$

where ΔQ is the charge included in the macroscopically small volume element ΔV that includes a large number of microscopic charged constituents. The units for ρ are C/m^3.

The notion of continuous charge distribution is similar to that we adopt for continuous mass distribution in mechanics. When we refer to

FIGURE 1.24 Definition of linear, surface and volume charge densities. In each case, the element (Δl, ΔS, ΔV) chosen is small on the macroscopic scale but contains a very large number of microscopic constituents.

* At the microscopic level, charge distribution is discontinuous, because they are discrete charges separated by intervening space where there is no charge.

the density of a liquid, we are referring to its macroscopic density. We regard it as a continuous fluid and ignore its discrete molecular constitution.

The field due to a continuous charge distribution can be obtained in much the same way as for a system of discrete charges, Eq. (1.10). Suppose a continuous charge distribution in space has a charge density ρ. Choose any convenient origin O and let the position vector of any point in the charge distribution be **r**. The charge density ρ may vary from point to point, i.e., it is a function of **r**. Divide the charge distribution into small volume elements of size ΔV. The charge in a volume element ΔV is $\rho\Delta V$.

Now, consider any general point P (inside or outside the distribution) with position vector **R** (Fig. 1.24). Electric field due to the charge $\rho\Delta V$ is given by Coulomb's law:

$$\Delta\mathbf{E} = \frac{1}{4\pi\varepsilon_0}\frac{\rho\,\Delta V}{r'^2}\hat{\mathbf{r}}' \tag{1.26}$$

where r' is the distance between the charge element and P, and $\hat{\mathbf{r}}'$ is a unit vector in the direction from the charge element to P. By the superposition principle, the total electric field due to the charge distribution is obtained by summing over electric fields due to different volume elements:

$$\mathbf{E} \cong \frac{1}{4\pi\varepsilon_0}\sum_{all\ \Delta V}\frac{\rho\,\Delta V}{r'^2}\hat{\mathbf{r}}' \tag{1.27}$$

Note that ρ, r', $\hat{\mathbf{r}}'$ all can vary from point to point. In a strict mathematical method, we should let $\Delta V \to 0$ and the sum then becomes an integral; but we omit that discussion here, for simplicity. In short, using Coulomb's law and the superposition principle, electric field can be determined for any charge distribution, discrete or continuous or part discrete and part continuous.

1.14 Gauss's Law

As a simple application of the notion of electric flux, let us consider the total flux through a sphere of radius r, which encloses a point charge q at its centre. Divide the sphere into small area elements, as shown in Fig. 1.25.

The flux through an area element $\Delta\mathbf{S}$ is

$$\Delta\phi = \mathbf{E}\cdot\Delta\mathbf{S} = \frac{q}{4\pi\varepsilon_0 r^2}\hat{\mathbf{r}}\cdot\Delta\mathbf{S} \tag{1.28}$$

where we have used Coulomb's law for the electric field due to a single charge q. The unit vector $\hat{\mathbf{r}}$ is along the radius vector from the centre to the area element. Now, since the normal to a sphere at every point is along the radius vector at that point, the area element $\Delta\mathbf{S}$ and $\hat{\mathbf{r}}$ have the same direction. Therefore,

$$\Delta\phi = \frac{q}{4\pi\varepsilon_0 r^2}\Delta S \tag{1.29}$$

since the magnitude of a unit vector is 1.

The total flux through the sphere is obtained by adding up flux through all the different area elements:

FIGURE 1.25 Flux through a sphere enclosing a point charge q at its centre.

FIGURE 1.26 Calculation of the flux of uniform electric field through the surface of a cylinder.

$$\phi = \sum_{all\ \Delta S} \frac{q}{4\pi\varepsilon_0 r^2} \Delta S$$

Since each area element of the sphere is at the same distance r from the charge,

$$\phi = \frac{q}{4\pi\varepsilon_o r^2} \sum_{all\ \Delta S} \Delta S = \frac{q}{4\pi\varepsilon_0 r^2} S$$

Now S, the total area of the sphere, equals $4\pi r^2$. Thus,

$$\phi = \frac{q}{4\pi\varepsilon_0 r^2} \times 4\pi r^2 = \frac{q}{\varepsilon_0} \qquad (1.30)$$

Equation (1.30) is a simple illustration of a general result of electrostatics called Gauss's law.

We state *Gauss's law* without proof:

Electric flux through a closed surface S

$$= q/\varepsilon_0 \qquad (1.31)$$

q = total charge enclosed by S.

The law implies that the total electric flux through a closed surface is zero if no charge is enclosed by the surface. We can see that explicitly in the simple situation of Fig. 1.26.

Here the electric field is uniform and we are considering a closed cylindrical surface, with its axis parallel to the uniform field **E**. The total flux ϕ through the surface is $\phi = \phi_1 + \phi_2 + \phi_3$, where ϕ_1 and ϕ_2 represent the flux through the surfaces 1 and 2 (of circular cross-section) of the cylinder and ϕ_3 is the flux through the curved cylindrical part of the closed surface. Now the normal to the surface 3 at every point is perpendicular to **E**, so by definition of flux, $\phi_3 = 0$. Further, the outward normal to 2 is along **E** while the outward normal to 1 is opposite to **E**. Therefore,

$$\phi_1 = -E\,S_1, \qquad \phi_2 = +E\,S_2$$
$$S_1 = S_2 = S$$

where S is the area of circular cross-section. Thus, the total flux is zero, as expected by Gauss's law. Thus, whenever you find that the net electric flux through a closed surface is zero, we conclude that the total charge contained in the closed surface is zero.

The great significance of Gauss's law Eq. (1.31), is that it is true in general, and not only for the simple cases we have considered above. Let us note some important points regarding this law:

(i) Gauss's law is true for any closed surface, no matter what its shape or size.

(ii) The term q on the right side of Gauss's law, Eq. (1.31), includes the sum of all charges enclosed by the surface. The charges may be located anywhere inside the surface.

(iii) In the situation when the surface is so chosen that there are some charges inside and some outside, the electric field [whose flux appears on the left side of Eq. (1.31)] is due to all the charges, both inside and outside S. The term q on the right side of Gauss's law, however, represents only the total charge inside S.

(iv) The surface that we choose for the application of Gauss's law is called the Gaussian surface. You may choose any Gaussian surface and apply Gauss's law. However, take care not to let the Gaussian surface pass through any discrete charge. This is because electric field due to a system of discrete charges is not well defined at the location of any charge. (As you go close to the charge, the field grows without any bound.) However, the Gaussian surface can pass through a continuous charge distribution.

(v) Gauss's law is often useful towards a much easier calculation of the electrostatic field *when the system has some symmetry.* This is facilitated by the choice of a suitable Gaussian surface.

(vi) Finally, Gauss's law is based on the inverse square dependence on distance contained in the Coulomb's law. Any violation of Gauss's law will indicate departure from the inverse square law.

Example 1.11 The electric field components in Fig. 1.27 are $E_x = \alpha x^{1/2}$, $E_y = E_z = 0$, in which $\alpha = 800$ N/C m$^{1/2}$. Calculate (a) the flux through the cube, and (b) the charge within the cube. Assume that $a = 0.1$ m.

FIGURE 1.27

Solution

(a) Since the electric field has only an x component, for faces perpendicular to x direction, the angle between **E** and Δ**S** is $\pm \pi/2$. Therefore, the flux $\phi = \mathbf{E}.\Delta\mathbf{S}$ is separately zero for each face of the cube except the two shaded ones. Now the magnitude of the electric field at the left face is

$E_L = \alpha x^{1/2} = \alpha a^{1/2}$

($x = a$ at the left face).

The magnitude of electric field at the right face is

$E_R = \alpha\, x^{1/2} = \alpha\, (2a)^{1/2}$

($x = 2a$ at the right face).

The corresponding fluxes are

$\phi_L = \mathbf{E}_L \cdot \Delta\mathbf{S} = \Delta S\, \mathbf{E}_L \cdot \hat{\mathbf{n}}_L = E_L\, \Delta S \cos\theta = -E_L\, \Delta S$, since $\theta = 180°$

$= -E_L a^2$

$\phi_R = \mathbf{E}_R \cdot \Delta\mathbf{S} = E_R\, \Delta S \cos\theta = E_R\, \Delta S$, since $\theta = 0°$

$= E_R a^2$

Net flux through the cube

EXAMPLE 1.11

EXAMPLE 1.11

$= \phi_R + \phi_L = E_R a^2 - E_L a^2 = a^2 (E_R - E_L) = \alpha a^2 [(2a)^{1/2} - a^{1/2}]$

$= \alpha a^{5/2} \left(\sqrt{2} - 1\right)$

$= 800\ (0.1)^{5/2} \left(\sqrt{2} - 1\right)$

$= 1.05\ \text{N m}^2\ \text{C}^{-1}$

(b) We can use Gauss's law to find the total charge q inside the cube. We have $\phi = q/\varepsilon_0$ or $q = \phi\varepsilon_0$. Therefore,

$q = 1.05 \times 8.854 \times 10^{-12}\ \text{C} = 9.27 \times 10^{-12}\ \text{C}.$

EXAMPLE 1.12

Example 1.12 An electric field is uniform, and in the positive x direction for positive x, and uniform with the same magnitude but in the negative x direction for negative x. It is given that $\mathbf{E} = 200\ \hat{\mathbf{i}}$ N/C for $x > 0$ and $\mathbf{E} = -200\ \hat{\mathbf{i}}$ N/C for $x < 0$. A right circular cylinder of length 20 cm and radius 5 cm has its centre at the origin and its axis along the x-axis so that one face is at $x = +10$ cm and the other is at $x = -10$ cm (Fig. 1.28). (a) What is the net outward flux through each flat face? (b) What is the flux through the side of the cylinder? (c) What is the net outward flux through the cylinder? (d) What is the net charge inside the cylinder?

Solution

(a) We can see from the figure that on the left face $\mathbf{E}$ and $\Delta\mathbf{S}$ are parallel. Therefore, the outward flux is

$\phi_L = \mathbf{E}\cdot\Delta\mathbf{S} = -200\ \hat{\mathbf{i}}\cdot\Delta\mathbf{S}$

$= +200\ \Delta S$, since $\hat{\mathbf{i}}\cdot\Delta\mathbf{S} = -\Delta S$

$= +200 \times \pi\ (0.05)^2 = +1.57\ \text{N m}^2\ \text{C}^{-1}$

On the right face, $\mathbf{E}$ and $\Delta\mathbf{S}$ are parallel and therefore

$\phi_R = \mathbf{E}\cdot\Delta\mathbf{S} = +1.57\ \text{N m}^2\ \text{C}^{-1}$.

(b) For any point on the side of the cylinder $\mathbf{E}$ is perpendicular to $\Delta\mathbf{S}$ and hence $\mathbf{E}\cdot\Delta\mathbf{S} = 0$. Therefore, the flux out of the side of the cylinder is zero.

(c) Net outward flux through the cylinder

$\phi = 1.57 + 1.57 + 0 = 3.14\ \text{N m}^2\ \text{C}^{-1}$

FIGURE 1.28

(d) The net charge within the cylinder can be found by using Gauss's law which gives

$q = \varepsilon_0\phi$

$= 3.14 \times 8.854 \times 10^{-12}\ \text{C}$

$= 2.78 \times 10^{-11}\ \text{C}$

1.15 Applications of Gauss's Law

The electric field due to a general charge distribution is, as seen above, given by Eq. (1.27). In practice, except for some special cases, the summation (or integration) involved in this equation cannot be carried out to give electric field at every point in space. For some symmetric charge configurations, however, it is possible to obtain the electric field in a simple way using the Gauss's law. This is best understood by some examples.

1.15.1 Field due to an infinitely long straight uniformly charged wire

Consider an infinitely long thin straight wire with uniform linear charge density λ. The wire is obviously an axis of symmetry. Suppose we take the radial vector from O to P and rotate it around the wire. The points P, P′, P″ so obtained are completely equivalent with respect to the charged wire. This implies that the electric field must have the same magnitude at these points. The direction of electric field at every point must be radial (outward if $\lambda > 0$, inward if $\lambda < 0$). This is clear from Fig. 1.29.

Consider a pair of line elements P_1 and P_2 of the wire, as shown. The electric fields produced by the two elements of the pair when summed give a resultant electric field which is radial (the components normal to the radial vector cancel). This is true for any such pair and hence the total field at any point P is radial. Finally, since the wire is infinite, electric field does not depend on the position of P along the length of the wire. In short, the electric field is everywhere radial in the plane cutting the wire normally, and its magnitude depends only on the radial distance r.

To calculate the field, imagine a cylindrical Gaussian surface, as shown in the Fig. 1.29(b). Since the field is everywhere radial, flux through the two ends of the cylindrical Gaussian surface is zero. At the cylindrical part of the surface, **E** is normal to the surface at every point, and its magnitude is constant, since it depends only on r. The surface area of the curved part is $2\pi rl$, where l is the length of the cylinder.

(a)

(b)

FIGURE 1.29 (a) Electric field due to an infinitely long thin straight wire is radial, (b) The Gaussian surface for a long thin wire of uniform linear charge density.

Flux through the Gaussian surface

= flux through the curved cylindrical part of the surface

= $E \times 2\pi rl$

The surface includes charge equal to $\lambda\, l$. Gauss's law then gives $E \times 2\pi rl = \lambda l/\varepsilon_0$

i.e., $$E = \frac{\lambda}{2\pi\varepsilon_0 r}$$

Vectorially, **E** at any point is given by

$$\mathbf{E} = \frac{\lambda}{2\pi\varepsilon_0 r}\ \hat{\mathbf{n}} \tag{1.32}$$

where $\hat{\mathbf{n}}$ is the radial unit vector in the plane normal to the wire passing through the point. **E** is directed outward if λ is positive and inward if λ is negative.

Note that when we write a vector **A** as a scalar multiplied by a unit vector, i.e., as $\mathbf{A} = A\ \hat{\mathbf{a}}$, the scalar A is an algebraic number. It can be negative or positive. The direction of **A** will be the same as that of the unit vector $\hat{\mathbf{a}}$ if $A > 0$ and opposite to $\hat{\mathbf{a}}$ if $A < 0$. When we want to restrict to non-negative values, we use the symbol $|\mathbf{A}|$ and call it the modulus of **A**. Thus, $|\mathbf{A}| \geq 0$.

Also note that though only the charge enclosed by the surface (λl) was included above, the electric field **E** is due to the charge on the entire wire. Further, the assumption that the wire is infinitely long is crucial. Without this assumption, we cannot take **E** to be normal to the curved part of the cylindrical Gaussian surface. However, Eq. (1.32) is approximately true for electric field around the central portions of a long wire, where the end effects may be ignored.

1.15.2 Field due to a uniformly charged infinite plane sheet

Let σ be the uniform surface charge density of an infinite plane sheet (Fig. 1.30). We take the x-axis normal to the given plane. By symmetry, the electric field will not depend on y and z coordinates and its direction at every point must be parallel to the x-direction.

FIGURE 1.30 Gaussian surface for a uniformly charged infinite plane sheet.

We can take the Gaussian surface to be a rectangular parallelepiped of cross sectional area A, as shown. (A cylindrical surface will also do.) As seen from the figure, only the two faces 1 and 2 will contribute to the flux; electric field lines are parallel to the other faces and they, therefore, do not contribute to the total flux.

The unit vector normal to surface 1 is in $-x$ direction while the unit vector normal to surface 2 is in the $+x$ direction. Therefore, flux $\mathbf{E}.\Delta\mathbf{S}$ through both the surfaces are equal and add up. Therefore the net flux through the Gaussian surface is $2\,EA$. The charge enclosed by the closed surface is σA. Therefore by Gauss's law,

$2\,EA = \sigma A/\varepsilon_0$
or, $E = \sigma/2\varepsilon_0$
Vectorically,

$$\mathbf{E} = \frac{\sigma}{2\varepsilon_0}\hat{\mathbf{n}} \tag{1.33}$$

where $\hat{\mathbf{n}}$ is a unit vector normal to the plane and going away from it.

E is directed away from the plate if σ is positive and toward the plate if σ is negative. Note that the above application of the Gauss' law has brought out an additional fact: E is independent of x also.

For a finite large planar sheet, Eq. (1.33) is approximately true in the middle regions of the planar sheet, away from the ends.

1.15.3 Field due to a uniformly charged thin spherical shell

Let σ be the uniform surface charge density of a thin spherical shell of radius R (Fig. 1.31). The situation has obvious spherical symmetry. The field at any point P, outside or inside, can depend only on r (the radial distance from the centre of the shell to the point) and must be radial (i.e., along the radius vector).

(i) Field outside the shell: Consider a point P outside the shell with radius vector **r**. To calculate **E** at P, we take the Gaussian surface to be a sphere of radius r and with centre O, passing through P. All points on this sphere are equivalent relative to the given charged configuration. (That is what we mean by spherical symmetry.) The electric field at each point of the Gaussian surface, therefore, has the same magnitude E and is along the radius vector at each point. Thus, **E** and $\Delta\mathbf{S}$ at every point are parallel and the flux through each element is $E\,\Delta S$. Summing over all ΔS, the flux through the Gaussian surface is $E \times 4\,\pi\,r^2$. The charge enclosed is $\sigma \times 4\,\pi\,R^2$. By Gauss's law

$$E \times 4\,\pi\,r^2 = \frac{\sigma}{\varepsilon_0}\,4\,\pi\,R^2$$

$$\text{Or,}\quad E = \frac{\sigma R^2}{\varepsilon_0\,r^2} = \frac{q}{4\pi\varepsilon_0 r^2}$$

where $q = 4\,\pi R^2\,\sigma$ is the total charge on the spherical shell. Vectorially,

$$\mathbf{E} = \frac{q}{4\pi\varepsilon_0\,r^2}\hat{\mathbf{r}} \tag{1.34}$$

FIGURE 1.31 Gaussian surfaces for a point with (a) $r > R$, (b) $r < R$.

The electric field is directed outward if $q > 0$ and inward if $q < 0$. This, however, is exactly the field produced by a charge q placed at the centre O. Thus for points outside the shell, the field due to a uniformly charged shell is as if the entire charge of the shell is concentrated at its centre.

(ii) *Field inside the shell:* In Fig. 1.31(b), the point P is inside the shell. The Gaussian surface is again a sphere through P centred at O.

The flux through the Gaussian surface, calculated as before, is $E \times 4\pi r^2$. However, in this case, the Gaussian surface encloses no charge. Gauss's law then gives

$$E \times 4\pi r^2 = 0$$

i.e., $E = 0 \qquad (r < R) \qquad (1.35)$

that is, the field due to a uniformly charged thin shell is zero at all points inside the shell*. This important result is a direct consequence of Gauss's law which follows from Coulomb's law. The experimental verification of this result confirms the $1/r^2$ dependence in Coulomb's law.

EXAMPLE 1.13

Example 1.13 An early model for an atom considered it to have a positively charged point nucleus of charge Ze, surrounded by a uniform density of negative charge up to a radius R. The atom as a whole is neutral. For this model, what is the electric field at a distance r from the nucleus?

FIGURE 1.32

Solution The charge distribution for this model of the atom is as shown in Fig. 1.32. The total negative charge in the uniform spherical charge distribution of radius R must be $-Ze$, since the atom (nucleus of charge Ze + negative charge) is neutral. This immediately gives us the negative charge density ρ, since we must have

$$\frac{4\pi R^3}{3}\rho = 0 - Ze$$

or $$\rho = -\frac{3Ze}{4\pi R^3}$$

To find the electric field **E(r)** at a point P which is a distance r away from the nucleus, we use Gauss's law. Because of the spherical symmetry of the charge distribution, the magnitude of the electric field **E(r)** depends only on the radial distance, no matter what the direction of **r**. Its direction is along (or opposite to) the radius vector **r** from the origin to the point P. The obvious Gaussian surface is a spherical surface centred at the nucleus. We consider two situations, namely, $r < R$ and $r > R$.

(i) $r < R$: The electric flux ϕ enclosed by the spherical surface is

$$\phi = E(r) \times 4\pi r^2$$

where $E(r)$ is the magnitude of the electric field at r. This is because

* Compare this with a uniform mass shell discussed in Section 8.5 of Class XI Textbook of Physics.

the field at any point on the spherical Gaussian surface has the same direction as the normal to the surface there, and has the same magnitude at all points on the surface.

The charge q enclosed by the Gaussian surface is the positive nuclear charge and the negative charge within the sphere of radius r,

i.e., $q = Ze + \frac{4\pi r^3}{3}\rho$

Substituting for the charge density ρ obtained earlier, we have

$$q = Ze - Ze\frac{r^3}{R^3}$$

Gauss's law then gives,

$$E(r) = \frac{Ze}{4\pi\varepsilon_0}\left(\frac{1}{r^2} - \frac{r}{R^3}\right); \quad r < R$$

The electric field is directed radially outward.

(ii) $r > R$: In this case, the total charge enclosed by the Gaussian spherical surface is zero since the atom is neutral. Thus, from Gauss's law,

$E(r) \times 4\pi r^2 = 0$ or $E(r) = 0$; $r > R$

At $r = R$, both cases give the same result: $E = 0$.

EXAMPLE 1.13

ON SYMMETRY OPERATIONS

In Physics, we often encounter systems with various symmetries. Consideration of these symmetries helps one arrive at results much faster than otherwise by a straightforward calculation. Consider, for example an infinite uniform sheet of charge (surface charge density σ) along the y-z plane. This system is unchanged if (a) translated parallel to the y-z plane in any direction, (b) rotated about the x-axis through any angle. As the system is unchanged under such symmetry operation, so must its properties be. In particular, in this example, the electric field **E** must be unchanged.

Translation symmetry along the y-axis shows that the electric field must be the same at a point $(0, y_1, 0)$ as at $(0, y_2, 0)$. Similarly translational symmetry along the z-axis shows that the electric field at two point $(0, 0, z_1)$ and $(0, 0, z_2)$ must be the same. By using rotation symmetry around the x-axis, we can conclude that **E** must be perpendicular to the y-z plane, that is, it must be parallel to the x-direction.

Try to think of a symmetry now which will tell you that the magnitude of the electric field is a constant, independent of the x-coordinate. It thus turns out that the magnitude of the electric field due to a uniformly charged infinite conducting sheet is the same at all points in space. The direction, however, is opposite of each other on either side of the sheet.

Compare this with the effort needed to arrive at this result by a direct calculation using Coulomb's law.

SUMMARY

1. Electric and magnetic forces determine the properties of atoms, molecules and bulk matter.
2. From simple experiments on frictional electricity, one can infer that there are two types of charges in nature; and that like charges repel and unlike charges attract. By convention, the charge on a glass rod rubbed with silk is positive; that on a plastic rod rubbed with fur is then negative.
3. Conductors allow movement of electric charge through them, insulators do not. In metals, the mobile charges are electrons; in electrolytes both positive and negative ions are mobile.
4. Electric charge has three basic properties: quantisation, additivity and conservation.

   Quantisation of electric charge means that total charge (q) of a body is always an integral multiple of a basic quantum of charge (e) i.e., $q = n\,e$, where $n = 0, \pm1, \pm2, \pm3, \ldots$. Proton and electron have charges $+e$, $-e$, respectively. For macroscopic charges for which n is a very large number, quantisation of charge can be ignored.

   Additivity of electric charges means that the total charge of a system is the algebraic sum (i.e., the sum taking into account proper signs) of all individual charges in the system.

   Conservation of electric charges means that the total charge of an isolated system remains unchanged with time. This means that when bodies are charged through friction, there is a transfer of electric charge from one body to another, but no creation or destruction of charge.
5. *Coulomb's Law*: The mutual electrostatic force between two point charges q_1 and q_2 is proportional to the product $q_1 q_2$ and inversely proportional to the square of the distance r_{21} separating them. Mathematically,

   $$\mathbf{F}_{21} = \text{force on } q_2 \text{ due to } q_1 = \frac{k\,(q_1 q_2)}{r_{21}^2}\,\hat{\mathbf{r}}_{21}$$

   where $\hat{\mathbf{r}}_{21}$ is a unit vector in the direction from q_1 to q_2 and $k = \dfrac{1}{4\pi\varepsilon_0}$ is the constant of proportionality.

   In SI units, the unit of charge is coulomb. The experimental value of the constant ε_0 is

   $$\varepsilon_0 = 8.854 \times 10^{-12}\ \mathrm{C^2\,N^{-1}\,m^{-2}}$$

   The approximate value of k is

   $$k = 9 \times 10^9\ \mathrm{N\,m^2\,C^{-2}}$$
6. The ratio of electric force and gravitational force between a proton and an electron is

   $$\frac{k\,e^2}{G\,m_e m_p} \cong 2.4 \times 10^{39}$$
7. *Superposition Principle*: The principle is based on the property that the forces with which two charges attract or repel each other are not affected by the presence of a third (or more) additional charge(s). For an assembly of charges $q_1, q_2, q_3, \ldots$, the force on any charge, say q_1, is

the vector sum of the force on q_1 due to q_2, the force on q_1 due to q_3, and so on. For each pair, the force is given by the Coulomb's law for two charges stated earlier.

8. The electric field **E** at a point due to a charge configuration is the force on a small positive test charge q placed at the point divided by the magnitude of the charge. Electric field due to a point charge q has a magnitude $|q|/4\pi\varepsilon_0 r^2$; it is radially outwards from q, if q is positive, and radially inwards if q is negative. Like Coulomb force, electric field also satisfies superposition principle.
9. An electric field line is a curve drawn in such a way that the tangent at each point on the curve gives the direction of electric field at that point. The relative closeness of field lines indicates the relative strength of electric field at different points; they crowd near each other in regions of strong electric field and are far apart where the electric field is weak. In regions of constant electric field, the field lines are uniformly spaced parallel straight lines.
10. Some of the important properties of field lines are: (i) Field lines are continuous curves without any breaks. (ii) Two field lines cannot cross each other. (iii) Electrostatic field lines start at positive charges and end at negative charges —they cannot form closed loops.
11. An electric dipole is a pair of equal and opposite charges q and $-q$ separated by some distance $2a$. Its dipole moment vector **p** has magnitude $2qa$ and is in the direction of the dipole axis from $-q$ to q.
12. Field of an electric dipole in its equatorial plane (i.e., the plane perpendicular to its axis and passing through its centre) at a distance r from the centre:

$$\mathbf{E} = \frac{-\mathbf{p}}{4\pi\varepsilon_o}\frac{1}{(a^2+r^2)^{3/2}}$$

$$\cong \frac{-\mathbf{p}}{4\pi\varepsilon_o r^3}, \qquad \text{for } r >> a$$

Dipole electric field on the axis at a distance r from the centre:

$$\mathbf{E} = \frac{2\mathbf{p}r}{4\pi\varepsilon_0(r^2-a^2)^2}$$

$$\cong \frac{2\mathbf{p}}{4\pi\varepsilon_0 r^3} \qquad \text{for } r >> a$$

The $1/r^3$ dependence of dipole electric fields should be noted in contrast to the $1/r^2$ dependence of electric field due to a point charge.

13. In a uniform electric field **E**, a dipole experiences a torque $\boldsymbol{\tau}$ given by

$$\boldsymbol{\tau} = \mathbf{p} \times \mathbf{E}$$

but experiences no net force.

14. The flux $\Delta\phi$ of electric field **E** through a small area element $\Delta\mathbf{S}$ is given by

$$\Delta\phi = \mathbf{E}\cdot\Delta\mathbf{S}$$

The vector area element $\Delta\mathbf{S}$ is

$$\Delta\mathbf{S} = \Delta S\,\hat{\mathbf{n}}$$

where ΔS is the magnitude of the area element and $\hat{\mathbf{n}}$ is normal to the area element, which can be considered planar for sufficiently small ΔS.

For an area element of a closed surface, $\hat{\mathbf{n}}$ is taken to be the direction of *outward* normal, by convention.

15. *Gauss's law*: The flux of electric field through any closed surface S is $1/\varepsilon_0$ times the total charge enclosed by S. The law is especially useful in determining electric field **E**, when the source distribution has simple symmetry:

(i) *Thin infinitely long straight wire of uniform linear charge density* λ

$$\mathbf{E} = \frac{\lambda}{2\,\pi\varepsilon_0\, r}\,\hat{\mathbf{n}}$$

where r is the perpendicular distance of the point from the wire and $\hat{\mathbf{n}}$ is the radial unit vector in the plane normal to the wire passing through the point.

(ii) Infinite thin plane sheet of uniform surface charge density σ

$$\mathbf{E} = \frac{\sigma}{2\,\varepsilon_0}\,\hat{\mathbf{n}}$$

where $\hat{\mathbf{n}}$ is a unit vector normal to the plane, outward on either side.

(iii) *Thin spherical shell of uniform surface charge density* σ

$$\mathbf{E} = \frac{q}{4\,\pi\varepsilon_0\, r^2}\,\hat{\mathbf{r}} \qquad (r \geq R)$$

$$\mathbf{E} = 0 \qquad (r < R)$$

where r is the distance of the point from the centre of the shell and R the radius of the shell. q is the total charge of the shell: $q = 4\pi R^2\sigma$.

The electric field outside the shell is as though the total charge is concentrated at the centre. The same result is true for a solid sphere of uniform volume charge density. The field is zero at all points inside the shell

Physical quantity	Symbol	Dimensions	Unit	Remarks
Vector area element	$\Delta\mathbf{S}$	$[L^2]$	m^2	$\Delta\mathbf{S} = \Delta S\,\hat{\mathbf{n}}$
Electric field	$\mathbf{E}$	$[MLT^{-3}A^{-1}]$	$V\,m^{-1}$	
Electric flux	ϕ	$[ML^3\,T^{-3}A^{-1}]$	V m	$\Delta\phi = \mathbf{E}\cdot\Delta\mathbf{S}$
Dipole moment	$\mathbf{p}$	[LTA]	C m	Vector directed from negative to positive charge
Charge density				
linear	λ	$[L^{-1}\,TA]$	$C\,m^{-1}$	Charge/length
surface	σ	$[L^{-2}\,TA]$	$C\,m^{-2}$	Charge/area
volume	ρ	$[L^{-3}\,TA]$	$C\,m^{-3}$	Charge/volume

POINTS TO PONDER

1. You might wonder why the protons, all carrying positive charges, are compactly residing inside the nucleus. Why do they not fly away? You will learn that there is a third kind of a fundamental force, called the strong force which holds them together. The range of distance where this force is effective is, however, very small ~10^{-14} m. This is precisely the size of the nucleus. Also the electrons are not allowed to sit on top of the protons, i.e. inside the nucleus, due to the laws of quantum mechanics. This gives the atoms their structure as they exist in nature.
2. Coulomb force and gravitational force follow the same inverse-square law. But gravitational force has only one sign (always attractive), while Coulomb force can be of both signs (attractive and repulsive), allowing possibility of cancellation of electric forces. This is how gravity, despite being a much weaker force, can be a dominating and more pervasive force in nature.
3. The constant of proportionality k in Coulomb's law is a matter of choice if the unit of charge is to be defined using Coulomb's law. In SI units, however, what is defined is the unit of current (A) via its magnetic effect (Ampere's law) and the unit of charge (coulomb) is simply defined by (1C = 1 A s). In this case, the value of k is no longer arbitrary; it is approximately 9×10^9 N m^2 C^{-2}.
4. The rather large value of k, i.e., the large size of the unit of charge (1C) from the point of view of electric effects arises because (as mentioned in point 3 already) the unit of charge is defined in terms of magnetic forces (forces on current–carrying wires) which are generally much weaker than the electric forces. Thus while 1 ampere is a unit of reasonable size for magnetic effects, 1 C = 1 A s, is too big a unit for electric effects.
5. The additive property of charge is not an 'obvious' property. It is related to the fact that electric charge has no direction associated with it; charge is a scalar.
6. Charge is not only a scalar (or invariant) under rotation; it is also invariant for frames of reference in relative motion. This is not always true for every scalar. For example, kinetic energy is a scalar under rotation, but is not invariant for frames of reference in relative motion.
7. Conservation of total charge of an isolated system is a property independent of the scalar nature of charge noted in point 6. Conservation refers to invariance in time in a given frame of reference. A quantity may be scalar but not conserved (like kinetic energy in an inelastic collision). On the other hand, one can have conserved vector quantity (e.g., angular momentum of an isolated system).
8. Quantisation of electric charge is a basic (unexplained) law of nature; interestingly, there is no analogous law on quantisation of mass.
9. Superposition principle should not be regarded as 'obvious', or equated with the law of addition of vectors. It says two things: force on one charge due to another charge is unaffected by the presence of other charges, and there are no additional three-body, four-body, etc., forces which arise only when there are more than two charges.
10. The electric field due to a discrete charge configuration is not defined at the locations of the discrete charges. For continuous volume charge distribution, it is defined at any point in the distribution. For a surface charge distribution, electric field is discontinuous across the surface.

11. The electric field due to a charge configuration with total charge zero is not zero; but for distances large compared to the size of the configuration, its field falls off faster than $1/r^2$, typical of field due to a single charge. An electric dipole is the simplest example of this fact.

EXERCISES

1.1 What is the force between two small charged spheres having charges of 2×10^{-7}C and 3×10^{-7}C placed 30 cm apart in air?

1.2 The electrostatic force on a small sphere of charge 0.4 μC due to another small sphere of charge –0.8 μC in air is 0.2 N. (a) What is the distance between the two spheres? (b) What is the force on the second sphere due to the first?

1.3 Check that the ratio $ke^2/G\, m_e m_p$ is dimensionless. Look up a Table of Physical Constants and determine the value of this ratio. What does the ratio signify?

1.4 (a) Explain the meaning of the statement 'electric charge of a body is quantised'.
(b) Why can one ignore quantisation of electric charge when dealing with macroscopic i.e., large scale charges?

1.5 When a glass rod is rubbed with a silk cloth, charges appear on both. A similar phenomenon is observed with many other pairs of bodies. Explain how this observation is consistent with the law of conservation of charge.

1.6 Four point charges $q_A = 2$ μC, $q_B = -5$ μC, $q_C = 2$ μC, and $q_D = -5$ μC are located at the corners of a square ABCD of side 10 cm. What is the force on a charge of 1 μC placed at the centre of the square?

1.7 (a) An electrostatic field line is a continuous curve. That is, a field line cannot have sudden breaks. Why not?
(b) Explain why two field lines never cross each other at any point?

1.8 Two point charges $q_A = 3$ μC and $q_B = -3$ μC are located 20 cm apart in vacuum.
(a) What is the electric field at the midpoint O of the line AB joining the two charges?
(b) If a negative test charge of magnitude 1.5×10^{-9} C is placed at this point, what is the force experienced by the test charge?

1.9 A system has two charges $q_A = 2.5 \times 10^{-7}$ C and $q_B = -2.5 \times 10^{-7}$ C located at points A: (0, 0, –15 cm) and B: (0,0, +15 cm), respectively. What are the total charge and electric dipole moment of the system?

1.10 An electric dipole with dipole moment 4×10^{-9} C m is aligned at 30° with the direction of a uniform electric field of magnitude 5×10^4 NC^{-1}. Calculate the magnitude of the torque acting on the dipole.

1.11 A polythene piece rubbed with wool is found to have a negative charge of 3×10^{-7} C.
(a) Estimate the number of electrons transferred (from which to which?)
(b) Is there a transfer of mass from wool to polythene?

1.12 (a) Two insulated charged copper spheres A and B have their centres separated by a distance of 50 cm. What is the mutual force of

electrostatic repulsion if the charge on each is 6.5×10^{-7} C? The radii of A and B are negligible compared to the distance of separation.

(b) What is the force of repulsion if each sphere is charged double the above amount, and the distance between them is halved?

1.13 Suppose the spheres A and B in Exercise 1.12 have identical sizes. A third sphere of the same size but uncharged is brought in contact with the first, then brought in contact with the second, and finally removed from both. What is the new force of repulsion between A and B?

1.14 Figure 1.33 shows tracks of three charged particles in a uniform electrostatic field. Give the signs of the three charges. Which particle has the highest charge to mass ratio?

FIGURE 1.33

1.15 Consider a uniform electric field $\mathbf{E} = 3 \times 10^3\ \hat{\mathbf{i}}$ N/C. (a) What is the flux of this field through a square of 10 cm on a side whose plane is parallel to the *yz* plane? (b) What is the flux through the same square if the normal to its plane makes a 60° angle with the *x*-axis?

1.16 What is the net flux of the uniform electric field of Exercise 1.15 through a cube of side 20 cm oriented so that its faces are parallel to the coordinate planes?

1.17 Careful measurement of the electric field at the surface of a black box indicates that the net outward flux through the surface of the box is 8.0×10^3 Nm^2/C. (a) What is the net charge inside the box? (b) If the net outward flux through the surface of the box were zero, could you conclude that there were no charges inside the box? Why or Why not?

1.18 A point charge +10 μC is a distance 5 cm directly above the centre of a square of side 10 cm, as shown in Fig. 1.34. What is the magnitude of the electric flux through the square? (*Hint:* Think of the square as one face of a cube with edge 10 cm.)

FIGURE 1.34

1.19 A point charge of 2.0 μC is at the centre of a cubic Gaussian surface 9.0 cm on edge. What is the net electric flux through the surface?

1.20 A point charge causes an electric flux of -1.0×10^3 Nm^2/C to pass through a spherical Gaussian surface of 10.0 cm radius centred on the charge. (a) If the radius of the Gaussian surface were doubled, how much flux would pass through the surface? (b) What is the value of the point charge?

1.21 A conducting sphere of radius 10 cm has an unknown charge. If the electric field 20 cm from the centre of the sphere is 1.5×10^3 N/C and points radially inward, what is the net charge on the sphere?

1.22 A uniformly charged conducting sphere of 2.4 m diameter has a surface charge density of 80.0 μC/m^2. (a) Find the charge on the sphere. (b) What is the total electric flux leaving the surface of the sphere?

1.23 An infinite line charge produces a field of 9×10^4 N/C at a distance of 2 cm. Calculate the linear charge density.

1.24 Two large, thin metal plates are parallel and close to each other. On their inner faces, the plates have surface charge densities of opposite signs and of magnitude 17.0×10^{-22} C/m^2. What is **E**: (a) in the outer region of the first plate, (b) in the outer region of the second plate, and (c) between the plates?

ADDITIONAL EXERCISES

1.25 An oil drop of 12 excess electrons is held stationary under a constant electric field of 2.55×10^4 NC^{-1} in Millikan's oil drop experiment. The density of the oil is 1.26 g cm^{-3}. Estimate the radius of the drop. (g = 9.81 m s^{-2}; $e = 1.60 \times 10^{-19}$ C).

1.26 Which among the curves shown in Fig. 1.35 cannot possibly represent electrostatic field lines?

(a) (b) (c)

(d) (e)

FIGURE 1.35

1.27 In a certain region of space, electric field is along the z-direction throughout. The magnitude of electric field is, however, not constant but increases uniformly along the positive z-direction, at the rate of 10^5 NC^{-1} per metre. What are the force and torque experienced by a system having a total dipole moment equal to 10^{-7} Cm in the negative z-direction ?

1.28 (a) A conductor A with a cavity as shown in Fig. 1.36(a) is given a charge Q. Show that the entire charge must appear on the outer surface of the conductor. (b) Another conductor B with charge q is inserted into the cavity keeping B insulated from A. Show that the total charge on the outside surface of A is $Q + q$ [Fig. 1.36(b)]. (c) A sensitive instrument is to be shielded from the strong electrostatic fields in its environment. Suggest a possible way.

(a) (b)

FIGURE 1.36

1.29 A hollow charged conductor has a tiny hole cut into its surface. Show that the electric field in the hole is $(\sigma/2\varepsilon_0)\,\hat{\mathbf{n}}$, where $\hat{\mathbf{n}}$ is the unit vector in the outward normal direction, and σ is the surface charge density near the hole.

1.30 Obtain the formula for the electric field due to a long thin wire of uniform linear charge density λ without using Gauss's law. [*Hint:* Use Coulomb's law directly and evaluate the necessary integral.]

1.31 It is now believed that protons and neutrons (which constitute nuclei of ordinary matter) are themselves built out of more elementary units called quarks. A proton and a neutron consist of three quarks each. Two types of quarks, the so called 'up' quark (denoted by u) of charge + (2/3) e, and the 'down' quark (denoted by d) of charge (–1/3) e, together with electrons build up ordinary matter. (Quarks of other types have also been found which give rise to different unusual varieties of matter.) Suggest a possible quark composition of a proton and neutron.

1.32 (a) Consider an arbitrary electrostatic field configuration. A small test charge is placed at a null point (i.e., where $\mathbf{E} = 0$) of the configuration. Show that the equilibrium of the test charge is necessarily unstable.

(b) Verify this result for the simple configuration of two charges of the same magnitude and sign placed a certain distance apart.

1.33 A particle of mass m and charge $(-q)$ enters the region between the two charged plates initially moving along x-axis with speed v_x (like particle 1 in Fig. 1.33). The length of plate is L and an uniform electric field E is maintained between the plates. Show that the vertical deflection of the particle at the far edge of the plate is $qEL^2/(2m\, v_x^2)$.

Compare this motion with motion of a projectile in gravitational field discussed in Section 4.10 of Class XI Textbook of Physics.

1.34 Suppose that the particle in Exercise in 1.33 is an electron projected with velocity $v_x = 2.0 \times 10^6$ m s^{-1}. If E between the plates separated by 0.5 cm is 9.1×10^2 N/C, where will the electron strike the upper plate? ($|e| = 1.6 \times 10^{-19}$ C, $m_e = 9.1 \times 10^{-31}$ kg.)

Chapter Two

ELECTROSTATIC POTENTIAL AND CAPACITANCE

2.1 Introduction

In Chapters 6 and 8 (Class XI), the notion of potential energy was introduced. When an external force does work in taking a body from a point to another against a force like spring force or gravitational force, that work gets stored as potential energy of the body. When the external force is removed, the body moves, gaining kinetic energy and losing an equal amount of potential energy. The sum of kinetic and potential energies is thus conserved. Forces of this kind are called conservative forces. Spring force and gravitational force are examples of conservative forces.

Coulomb force between two (stationary) charges, like the gravitational force, is also a conservative force. This is not surprising, since both have inverse-square dependence on distance and differ mainly in the proportionality constants – the masses in the gravitational law are replaced by charges in Coulomb's law. Thus, like the potential energy of a mass in a gravitational field, we can define electrostatic potential energy of a charge in an electrostatic field.

Consider an electrostatic field $\mathbf{E}$ due to some charge configuration. First, for simplicity, consider the field $\mathbf{E}$ due to a charge Q placed at the origin. Now, imagine that we bring a test charge q from a point R to a point P against the repulsive force on it due to the charge Q. With reference

FIGURE 2.1 A test charge q (> 0) is moved from the point R to the point P against the repulsive force on it by the charge Q (> 0) placed at the origin.

to Fig. 2.1, this will happen if Q and q are both positive or both negative. For definiteness, let us take Q, $q > 0$.

Two remarks may be made here. First, we assume that the test charge q is so small that it does not disturb the original configuration, namely the charge Q at the origin (or else, we keep Q fixed at the origin by some unspecified force). Second, in bringing the charge q from R to P, we apply an external force $\mathbf{F}_{ext}$ just enough to counter the repulsive electric force $\mathbf{F}_E$ (i.e, $\mathbf{F}_{ext} = -\mathbf{F}_E$). This means there is no net force on or acceleration of the charge q when it is brought from R to P, i.e., it is brought with infinitesimally slow constant speed. In this situation, work done by the external force is the negative of the work done by the electric force, and gets fully stored in the form of potential energy of the charge q. If the external force is removed on reaching P, the electric force will take the charge away from Q – the stored energy (potential energy) at P is used to provide kinetic energy to the charge q in such a way that the sum of the kinetic and potential energies is conserved.

Thus, work done by external forces in moving a charge q from R to P is

$$W_{RP} = \int_R^P \mathbf{F}_{ext} \cdot d\mathbf{r}$$

$$= -\int_R^P \mathbf{F}_E \cdot d\mathbf{r} \qquad (2.1)$$

This work done is against electrostatic repulsive force and gets stored as potential energy.

At every point in electric field, a particle with charge q possesses a certain electrostatic potential energy, this work done increases its potential energy by an amount equal to potential energy difference between points R and P.

Thus, potential energy difference

$$\Delta U = U_P - U_R = W_{RP} \qquad (2.2)$$

(Note here that this displacement is in an opposite sense to the electric force and hence work done by electric field is negative, i.e., $-W_{RP}$.)

Therefore, we can define electric potential energy difference between two points as the work required to be done by an external force in moving (without accelerating) charge q from one point to another for electric field of any arbitrary charge configuration.

Two important comments may be made at this stage:

(i) The right side of Eq. (2.2) depends only on the initial and final positions of the charge. It means that the work done by an electrostatic field in moving a charge from one point to another depends only on the initial and the final points and is independent of the path taken to go from one point to the other. This is the fundamental characteristic of a conservative force. The concept of the potential energy would not be meaningful if the work depended on the path. The path-independence of work done by an electrostatic field can be proved using the Coulomb's law. We omit this proof here.

(ii) Equation (2.2) defines *potential energy difference* in terms of the physically meaningful quantity *work*. Clearly, potential energy so defined is undetermined to within an additive constant.What this means is that the actual value of potential energy is not physically significant; it is only the difference of potential energy that is significant. We can always add an arbitrary constant α to potential energy at every point, since this will not change the potential energy difference:

$$(U_P + \alpha) - (U_R + \alpha) = U_P - U_R$$

Put it differently, there is a freedom in choosing the point where potential energy is zero. A convenient choice is to have electrostatic potential energy zero at infinity. With this choice, if we take the point R at infinity, we get from Eq. (2.2)

$$W_{\infty P} = U_P - U_\infty = U_P \qquad (2.3)$$

Since the point P is arbitrary, Eq. (2.3) provides us with a definition of potential energy of a charge q at any point. *Potential energy of charge q at a point* (in the presence of field due to any charge configuration) *is the work done by the external force* (equal and opposite to the electric force) *in bringing the charge q from infinity to that point.*

Count Alessandro Volta (1745 – 1827) Italian physicist, professor at Pavia. Volta established that the *animal electricity* observed by Luigi Galvani, 1737–1798, in experiments with frog muscle tissue placed in contact with dissimilar metals, was not due to any exceptional property of animal tissues but was also generated whenever any wet body was sandwiched between dissimilar metals. This led him to develop the first *voltaic pile*, or battery, consisting of a large stack of moist disks of cardboard (electrolyte) sandwiched between disks of metal (electrodes).

COUNT ALESSANDRO VOLTA (1745 –1827)

2.2 Electrostatic Potential

Consider any general static charge configuration. We define potential energy of a test charge q in terms of the work done on the charge q. This work is obviously proportional to q, since the force at any point is $q\mathbf{E}$, where $\mathbf{E}$ is the electric field at that point due to the given charge configuration. It is, therefore, convenient to divide the work by the amount of charge q, so that the resulting quantity is independent of q. In other words, work done per unit test charge is characteristic of the electric field associated with the charge configuration. This leads to the idea of electrostatic potential V due to a given charge configuration. From Eq. (2.1), we get:

Work done by external force in bringing a unit positive charge from point R to P

$$= V_P - V_R \quad \left(= \frac{U_P - U_R}{q}\right) \qquad (2.4)$$

where V_P and V_R are the electrostatic potentials at P and R, respectively. Note, as before, that it is not the actual value of potential but the potential difference that is physically significant. If, as before, we choose the potential to be zero at infinity, Eq. (2.4) implies:

Work done by an external force in bringing a unit positive charge from infinity to a point = electrostatic potential (V) at that point.

FIGURE 2.2 Work done on a test charge q by the electrostatic field due to any given charge configuration is independent of the path, and depends only on its initial and final positions.

In other words, the electrostatic potential (V) at any point in a region with electrostatic field is the work done in bringing a unit positive charge (without acceleration) from infinity to that point.

The qualifying remarks made earlier regarding potential energy also apply to the definition of potential. To obtain the work done per unit test charge, we should take an infinitesimal test charge δq, obtain the work done δW in bringing it from infinity to the point and determine the ratio $\delta W/\delta q$. Also, the external force at every point of the path is to be equal and opposite to the electrostatic force on the test charge at that point.

2.3 Potential due to a Point Charge

Consider a point charge Q at the origin (Fig. 2.3). For definiteness, take Q to be positive. We wish to determine the potential at any point P with position vector **r** from the origin. For that we must calculate the work done in bringing a unit positive test charge from infinity to the point P. For $Q > 0$, the work done against the repulsive force on the test charge is positive. Since work done is independent of the path, we choose a convenient path – along the radial direction from infinity to the point P.

FIGURE 2.3 Work done in bringing a unit positive test charge from infinity to the point P, against the repulsive force of charge Q ($Q > 0$), is the potential at P due to the charge Q.

At some intermediate point P′ on the path, the electrostatic force on a unit positive charge is

$$\frac{Q \times 1}{4\pi\varepsilon_0 r'^2}\hat{\mathbf{r}}' \tag{2.5}$$

where $\hat{\mathbf{r}}'$ is the unit vector along OP′. Work done against this force from $\mathbf{r}'$ to $\mathbf{r}' + \Delta\mathbf{r}'$ is

$$\Delta W = -\frac{Q}{4\pi\varepsilon_0 r'^2}\Delta r' \tag{2.6}$$

The negative sign appears because for $\Delta r' < 0$, ΔW is positive . Total work done (W) by the external force is obtained by integrating Eq. (2.6) from $r' = \infty$ to $r' = r$,

$$W = -\int_{\infty}^{r} \frac{Q}{4\pi\varepsilon_0 r'^2} dr' = \left.\frac{Q}{4\pi\varepsilon_0 r'}\right|_{\infty}^{r} = \frac{Q}{4\pi\varepsilon_0 r} \tag{2.7}$$

This, by definition is the potential at P due to the charge Q

$$V(r) = \frac{Q}{4\pi\varepsilon_0 r} \tag{2.8}$$

Equation (2.8) is true for any sign of the charge Q, though we considered $Q > 0$ in its derivation. For $Q < 0$, $V < 0$, i.e., work done (by the external force) per unit positive test charge in bringing it from infinity to the point is negative. This is equivalent to saying that work done by the electrostatic force in bringing the unit positive charge form infinity to the point P is positive. [This is as it should be, since for $Q < 0$, the force on a unit positive test charge is attractive, so that the electrostatic force and the displacement (from infinity to P) are in the same direction.] Finally, we note that Eq. (2.8) is consistent with the choice that potential at infinity be zero.

FIGURE 2.4 Variation of potential V with r [in units of $(Q/4\pi\varepsilon_0)$ m^{-1}] (blue curve) and field with r [in units of $(Q/4\pi\varepsilon_0)$ m^{-2}] (black curve) for a point charge Q.

Figure (2.4) shows how the electrostatic potential ($\propto 1/r$) and the electrostatic field ($\propto 1/r^2$) varies with r.

Example 2.1

(a) Calculate the potential at a point P due to a charge of 4×10^{-7}C located 9 cm away.

(b) Hence obtain the work done in bringing a charge of 2×10^{-9} C from infinity to the point P. Does the answer depend on the path along which the charge is brought?

Solution

(a) $$V = \frac{1}{4\pi\varepsilon_0}\frac{Q}{r} = 9\times10^9\ \text{Nm}^2\ \text{C}^{-2} \times \frac{4\times10^{-7}\text{C}}{0.09\,\text{m}}$$

$$= 4 \times 10^4\ \text{V}$$

(b) $$W = qV = 2\times10^{-9}\text{C}\times4\times10^4\,\text{V}$$

$$= 8 \times 10^{-5}\,\text{J}$$

No, work done will be path independent. Any arbitrary infinitesimal path can be resolved into two perpendicular displacements: One along **r** and another perpendicular to **r**. The work done corresponding to the later will be zero.

EXAMPLE 2.1

2.4 Potential due to an Electric Dipole

As we learnt in the last chapter, an electric dipole consists of two charges q and $-q$ separated by a (small) distance $2a$. Its total charge is zero. It is characterised by a dipole moment vector **p** whose magnitude is $q \times 2a$ and which points in the direction from $-q$ to q (Fig. 2.5). We also saw that the electric field of a dipole at a point with position vector **r** depends not just on the magnitude r, but also on the angle between **r** and **p**. Further,

FIGURE 2.5 Quantities involved in the calculation of potential due to a dipole.

the field falls off, at large distance, not as $1/r^2$ (typical of field due to a single charge) but as $1/r^3$. We, now, determine the electric potential due to a dipole and contrast it with the potential due to a single charge.

As before, we take the origin at the centre of the dipole. Now we know that the electric field obeys the superposition principle. Since potential is related to the work done by the field, electrostatic potential also follows the superposition principle. Thus, the potential due to the dipole is the sum of potentials due to the charges q and $-q$

$$V = \frac{1}{4\pi\varepsilon_0}\left(\frac{q}{r_1} - \frac{q}{r_2}\right) \tag{2.9}$$

where r_1 and r_2 are the distances of the point P from q and $-q$, respectively.

Now, by geometry,

$$r_1^2 = r^2 + a^2 - 2ar\cos\theta$$

$$r_2^2 = r^2 + a^2 + 2ar\cos\theta \tag{2.10}$$

We take r much greater than a ($r >> a$) and retain terms only upto the first order in a/r

$$r_1^2 = r^2\left(1 - \frac{2a\cos\theta}{r} + \frac{a^2}{r^2}\right)$$

$$\cong r^2\left(1 - \frac{2a\cos\theta}{r}\right) \tag{2.11}$$

Similarly,

$$r_2^2 \cong r^2\left(1 + \frac{2a\cos\theta}{r}\right) \tag{2.12}$$

Using the Binomial theorem and retaining terms upto the first order in a/r; we obtain,

$$\frac{1}{r_1} \cong \frac{1}{r}\left(1 - \frac{2a\cos\theta}{r}\right)^{-1/2} \cong \frac{1}{r}\left(1 + \frac{a}{r}\cos\theta\right) \tag{2.13(a)}$$

$$\frac{1}{r_2} \cong \frac{1}{r}\left(1 + \frac{2a\cos\theta}{r}\right)^{-1/2} \cong \frac{1}{r}\left(1 - \frac{a}{r}\cos\theta\right) \tag{2.13(b)}$$

Using Eqs. (2.9) and (2.13) and $p = 2qa$, we get

$$V = \frac{q}{4\pi\varepsilon_0}\frac{2a\cos\theta}{r^2} = \frac{p\cos\theta}{4\pi\varepsilon_0 r^2} \tag{2.14}$$

Now, $p\cos\theta = \mathbf{p}\cdot\hat{\mathbf{r}}$

where $\hat{\mathbf{r}}$ is the unit vector along the position vector **OP**.

The electric potential of a dipole is then given by

$$V = \frac{1}{4\pi\varepsilon_0}\frac{\mathbf{p}\cdot\hat{\mathbf{r}}}{r^2}; \qquad (r >> a) \tag{2.15}$$

Equation (2.15) is, as indicated, approximately true only for distances large compared to the size of the dipole, so that higher order terms in a/r are negligible. For a point dipole **p** at the origin, Eq. (2.15) is, however, exact.

From Eq. (2.15), potential on the dipole axis ($\theta = 0, \pi$) is given by

$$V = \pm\frac{1}{4\pi\varepsilon_0}\frac{p}{r^2} \tag{2.16}$$

(Positive sign for $\theta = 0$, negative sign for $\theta = \pi$.) The potential in the equatorial plane ($\theta = \pi/2$) is zero.

The important contrasting features of electric potential of a dipole from that due to a single charge are clear from Eqs. (2.8) and (2.15):

(i) The potential due to a dipole depends not just on r but also on the angle between the position vector **r** and the dipole moment vector **p**. (It is, however, axially symmetric about **p**. That is, if you rotate the position vector **r** about **p**, keeping θ fixed, the points corresponding to P on the cone so generated will have the same potential as at P.)

(ii) The electric dipole potential falls off, at large distance, as $1/r^2$, not as $1/r$, characteristic of the potential due to a single charge. (You can refer to the Fig. 2.5 for graphs of $1/r^2$ versus r and $1/r$ versus r, drawn there in another context.)

2.5 Potential due to a System of Charges

Consider a system of charges $q_1, q_2, .., q_n$ with position vectors $\mathbf{r}_1, \mathbf{r}_2, .., \mathbf{r}_n$ relative to some origin (Fig. 2.6). The potential V_1 at P due to the charge q_1 is

$$V_1 = \frac{1}{4\pi\varepsilon_0}\frac{q_1}{r_{1P}}$$

where r_{1P} is the distance between q_1 and P.

Similarly, the potential V_2 at P due to q_2 and V_3 due to q_3 are given by

$$V_2 = \frac{1}{4\pi\varepsilon_0}\frac{q_2}{r_{2P}}, \quad V_3 = \frac{1}{4\pi\varepsilon_0}\frac{q_3}{r_{3P}}$$

where r_{2P} and r_{3P} are the distances of P from charges q_2 and q_3, respectively; and so on for the potential due to other charges. By the superposition principle, the potential V at P due to the total charge configuration is the algebraic sum of the potentials due to the individual charges

$$V = V_1 + V_2 + ... + V_n \tag{2.17}$$

FIGURE 2.6 Potential at a point due to a system of charges is the sum of potentials due to individual charges.

$$= \frac{1}{4\pi\varepsilon_0}\left(\frac{q_1}{r_{1P}} + \frac{q_2}{r_{2P}} + \ldots\ldots + \frac{q_n}{r_{nP}}\right) \tag{2.18}$$

If we have a continuous charge distribution characterised by a charge density $\rho(\mathbf{r})$, we divide it, as before, into small volume elements each of size Δv and carrying a charge $\rho\Delta v$. We then determine the potential due to each volume element and sum (strictly speaking , integrate) over all such contributions, and thus determine the potential due to the entire distribution.

We have seen in Chapter 1 that for a uniformly charged spherical shell, the electric field outside the shell is as if the entire charge is concentrated at the centre. Thus, the potential outside the shell is given by

$$V = \frac{1}{4\pi\varepsilon_0}\frac{q}{r} \quad (r \geq R) \qquad [2.19(a)]$$

where q is the total charge on the shell and R its radius. The electric field inside the shell is zero. This implies (Section 2.6) that potential is constant inside the shell (as no work is done in moving a charge inside the shell), and, therefore, equals its value at the surface, which is

$$V = \frac{1}{4\pi\varepsilon_0}\frac{q}{R} \qquad [2.19(b)]$$

EXAMPLE 2.2

Example 2.2 Two charges 3×10^{-8} C and -2×10^{-8} C are located 15 cm apart. At what point on the line joining the two charges is the electric potential zero? Take the potential at infinity to be zero.

Solution Let us take the origin O at the location of the positive charge. The line joining the two charges is taken to be the x-axis; the negative charge is taken to be on the right side of the origin (Fig. 2.7).

FIGURE 2.7

Let P be the required point on the x-axis where the potential is zero. If x is the x-coordinate of P, obviously x must be positive. (There is no possibility of potentials due to the two charges adding up to zero for $x < 0$.) If x lies between O and A, we have

$$\frac{1}{4\pi\varepsilon_0}\left[\frac{3\times10^{-8}}{x\times10^{-2}} - \frac{2\times10^{-8}}{(15-x)\times10^{-2}}\right] = 0$$

where x is in cm. That is,

$$\frac{3}{x} - \frac{2}{15-x} = 0$$

which gives x = 9 cm.

If x lies on the extended line OA, the required condition is

$$\frac{3}{x} - \frac{2}{x-15} = 0$$

which gives

$x = 45$ cm

Thus, electric potential is zero at 9 cm and 45 cm away from the positive charge on the side of the negative charge. Note that the formula for potential used in the calculation required choosing potential to be zero at infinity.

EXAMPLE 2.2

Example 2.3 Figures 2.8 (a) and (b) show the field lines of a positive and negative point charge respectively.

FIGURE 2.8

(a) Give the signs of the potential difference $V_P - V_Q$; $V_B - V_A$.

(b) Give the sign of the potential energy difference of a small negative charge between the points Q and P; A and B.

(c) Give the sign of the work done by the field in moving a small positive charge from Q to P.

(d) Give the sign of the work done by the external agency in moving a small negative charge from B to A.

(e) Does the kinetic energy of a small negative charge increase or decrease in going from B to A?

Solution

(a) As $V \propto \frac{1}{r}$, $V_P > V_Q$. Thus, $(V_P - V_Q)$ is positive. Also V_B is less negative than V_A. Thus, $V_B > V_A$ or $(V_B - V_A)$ is positive.

(b) A small negative charge will be attracted towards positive charge. The negative charge moves from higher potential energy to lower potential energy. Therefore the sign of potential energy difference of a small negative charge between Q and P is positive.

Similarly, $(\text{P.E.})_A > (\text{P.E.})_B$ and hence sign of potential energy differences is positive.

(c) In moving a small positive charge from Q to P, work has to be done by an external agency against the electric field. Therefore, work done by the field is negative.

(d) In moving a small negative charge from B to A work has to be done by the external agency. It is positive.

(e) Due to force of repulsion on the negative charge, velocity decreases and hence the kinetic energy decreases in going from B to A.

EXAMPLE 2.3

PHYSICS

Electric potential, equipotential surfaces:
http://sol.sci.uop.edu/~jfalward/electricpotential/electricpotential.html

2.6 Equipotential Surfaces

FIGURE 2.9 For a single charge q (a) equipotential surfaces are spherical surfaces centred at the charge, and (b) electric field lines are radial, starting from the charge if $q > 0$.

An equipotential surface is a surface with a constant value of potential at all points on the surface. For a single charge q, the potential is given by Eq. (2.8):

$$V = \frac{1}{4\pi\varepsilon_o}\frac{q}{r}$$

This shows that V is a constant if r is constant . Thus, equipotential surfaces of a single point charge are concentric spherical surfaces centred at the charge.

Now the electric field lines for a single charge q are radial lines starting from or ending at the charge, depending on whether q is positive or negative. Clearly, the electric field at every point is normal to the equipotential surface passing through that point. This is true in general: *for any charge configuration, equipotential surface through a point is normal to the electric field at that point*. The proof of this statement is simple.

If the field were not normal to the equipotential surface, it would have non-zero component along the surface. To move a unit test charge against the direction of the component of the field, work would have to be done. But this is in contradiction to the definition of an equipotential surface: there is no potential difference between any two points on the surface and no work is required to move a test charge on the surface. The electric field must, therefore, be normal to the equipotential surface at every point. Equipotential surfaces offer an alternative visual picture in addition to the picture of electric field lines around a charge configuration.

FIGURE 2.10 Equipotential surfaces for a uniform electric field.

For a uniform electric field **E**, say, along the x-axis, the equipotential surfaces are planes normal to the x-axis, i.e., planes parallel to the y-z plane (Fig. 2.10). Equipotential surfaces for (a) a dipole and (b) two identical positive charges are shown in Fig. 2.11.

FIGURE 2.11 Some equipotential surfaces for (a) a dipole, (b) two identical positive charges.

2.6.1 Relation between field and potential

Consider two closely spaced equipotential surfaces A and B (Fig. 2.12) with potential values V and $V + \delta V$, where δV is the change in V in the direction of the electric field **E**. Let P be a point on the surface B. δl is the perpendicular distance of the surface A from P. Imagine that a unit positive charge is moved along this perpendicular from the surface B to surface A against the electric field. The work done in this process is $|\mathbf{E}|\,\delta l$.

This work equals the potential difference $V_A - V_B$.

Thus,

$$|\mathbf{E}|\,\delta l = V - (V + \delta V) = -\delta V$$

$$\text{i.e., } |\mathbf{E}| = -\frac{\delta V}{\delta l} \qquad (2.20)$$

FIGURE 2.12 From the potential to the field.

Since δV is negative, $\delta V = -|\delta V|$. we can rewrite Eq (2.20) as

$$|\mathbf{E}| = -\frac{\delta V}{\delta l} = +\frac{|\delta V|}{\delta l} \qquad (2.21)$$

We thus arrive at two important conclusions concerning the relation between electric field and potential:

(i) *Electric field is in the direction in which the potential decreases steepest.*

(ii) *Its magnitude is given by the change in the magnitude of potential per unit displacement normal to the equipotential surface at the point.*

2.7 Potential Energy of a System of Charges

Consider first the simple case of two charges q_1 and q_2 with position vector $\mathbf{r}_1$ and $\mathbf{r}_2$ relative to some origin. Let us calculate the work done (externally) in building up this configuration. This means that we consider the charges q_1 and q_2 initially at infinity and determine the work done by an external agency to bring the charges to the given locations. Suppose, first the charge q_1 is brought from infinity to the point $\mathbf{r}_1$. There is no external field against which work needs to be done, so work done in bringing q_1 from infinity to $\mathbf{r}_1$ is zero. This charge produces a potential in space given by

$$V_1 = \frac{1}{4\pi\varepsilon_0}\frac{q_1}{r_{1P}}$$

where r_{1P} is the distance of a point P in space from the location of q_1. From the definition of potential, work done in bringing charge q_2 from infinity to the point $\mathbf{r}_2$ is q_2 times the potential at $\mathbf{r}_2$ due to q_1:

$$\text{work done on } q_2 = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r_{12}}$$

FIGURE 2.13 Potential energy of a system of charges q_1 and q_2 is directly proportional to the product of charges and inversely to the distance between them.

where r_{12} is the distance between points 1 and 2.

Since electrostatic force is conservative, this work gets stored in the form of potential energy of the system. Thus, the potential energy of a system of two charges q_1 and q_2 is

$$U = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r_{12}} \tag{2.22}$$

Obviously, if q_2 was brought first to its present location and q_1 brought later, the potential energy U would be the same. More generally, the potential energy expression, Eq. (2.22), is unaltered whatever way the charges are brought to the specified locations, because of path-independence of work for electrostatic force.

Equation (2.22) is true for any sign of q_1 and q_2. If $q_1 q_2 > 0$, potential energy is positive. This is as expected, since for like charges ($q_1 q_2 > 0$), electrostatic force is repulsive and a positive amount of work is needed to be done against this force to bring the charges from infinity to a finite distance apart. For unlike charges ($q_1 q_2 < 0$), the electrostatic force is attractive. In that case, a positive amount of work is needed against this force to take the charges from the given location to infinity. In other words, a negative amount of work is needed for the reverse path (from infinity to the present locations), so the potential energy is negative.

Equation (2.22) is easily generalised for a system of any number of point charges. Let us calculate the potential energy of a system of three charges q_1, q_2 and q_3 located at $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$, respectively. To bring q_1 first from infinity to $\mathbf{r}_1$, no work is required. Next we bring q_2 from infinity to $\mathbf{r}_2$. As before, work done in this step is

$$q_2 V_1(\mathbf{r}_2) = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r_{12}} \tag{2.23}$$

The charges q_1 and q_2 produce a potential, which at any point P is given by

$$V_{1,2} = \frac{1}{4\pi\varepsilon_0}\left(\frac{q_1}{r_{1P}} + \frac{q_2}{r_{2P}}\right) \tag{2.24}$$

Work done next in bringing q_3 from infinity to the point $\mathbf{r}_3$ is q_3 times $V_{1,2}$ at $\mathbf{r}_3$

FIGURE 2.14 Potential energy of a system of three charges is given by Eq. (2.26), with the notation given in the figure.

$$q_3 V_{1,2}(\mathbf{r}_3) = \frac{1}{4\pi\varepsilon_0}\left(\frac{q_1 q_3}{r_{13}} + \frac{q_2 q_3}{r_{23}}\right) \tag{2.25}$$

The total work done in assembling the charges at the given locations is obtained by adding the work done in different steps [Eq. (2.23) and Eq. (2.25)],

$$U = \frac{1}{4\pi\varepsilon_0}\left(\frac{q_1 q_2}{r_{12}} + \frac{q_1 q_3}{r_{13}} + \frac{q_2 q_3}{r_{23}}\right) \tag{2.26}$$

Again, because of the conservative nature of the electrostatic force (or equivalently, the path independence of work done), the final expression for U, Eq. (2.26), is independent of the manner in which the configuration is assembled. *The potential energy*

is characteristic of the present state of configuration, and not the way the state is achieved.

Example 2.4 Four charges are arranged at the corners of a square ABCD of side d, as shown in Fig. 2.15.(a) Find the work required to put together this arrangement. (b) A charge q_0 is brought to the centre E of the square, the four charges being held fixed at its corners. How much extra work is needed to do this?

FIGURE 2.15

Solution

(a) Since the work done depends on the final arrangement of the charges, and not on how they are put together, we calculate work needed for one way of putting the charges at A, B, C and D. Suppose, first the charge $+q$ is brought to A, and then the charges $-q$, $+q$, and $-q$ are brought to B, C and D, respectively. The total work needed can be calculated in steps:

(i) Work needed to bring charge $+q$ to A when no charge is present elsewhere: this is zero.

(ii) Work needed to bring $-q$ to B when $+q$ is at A. This is given by (charge at B) × (electrostatic potential at B due to charge $+q$ at A)

$$= -q \times \left(\frac{q}{4\pi\varepsilon_0 d}\right) = -\frac{q^2}{4\pi\varepsilon_0 d}$$

(iii) Work needed to bring charge $+q$ to C when $+q$ is at A and $-q$ is at B. This is given by (charge at C) × (potential at C due to charges at A and B)

$$= +q\left(\frac{+q}{4\pi\varepsilon_0 d\sqrt{2}} + \frac{-q}{4\pi\varepsilon_0 d}\right)$$

$$= \frac{-q^2}{4\pi\varepsilon_0 d}\left(1 - \frac{1}{\sqrt{2}}\right)$$

(iv) Work needed to bring $-q$ to D when $+q$ at A, $-q$ at B, and $+q$ at C. This is given by (charge at D) × (potential at D due to charges at A, B and C)

$$= -q\left(\frac{+q}{4\pi\varepsilon_0 d} + \frac{-q}{4\pi\varepsilon_0 d\sqrt{2}} + \frac{q}{4\pi\varepsilon_0 d}\right)$$

$$= \frac{-q^2}{4\pi\varepsilon_0 d}\left(2 - \frac{1}{\sqrt{2}}\right)$$

EXAMPLE 2.4

EXAMPLE 2.4

Add the work done in steps (i), (ii), (iii) and (iv). The total work required is

$$= \frac{-q^2}{4\pi\varepsilon_0 d}\left\{(0)+(1)+\left(1-\frac{1}{\sqrt{2}}\right)+\left(2-\frac{1}{\sqrt{2}}\right)\right\}$$

$$= \frac{-q^2}{4\pi\varepsilon_0 d}\left(4-\sqrt{2}\right)$$

The work done depends only on the arrangement of the charges, and not how they are assembled. By definition, this is the total electrostatic energy of the charges.

(Students may try calculating same work/energy by taking charges in any other order they desire and convince themselves that the energy will remain the same.)

(b) The extra work necessary to bring a charge q_0 to the point E when the four charges are at A, B, C and D is q_0 × (electrostatic potential at E due to the charges at A, B, C and D). The electrostatic potential at E is clearly zero since potential due to A and C is cancelled by that due to B and D. Hence no work is required to bring any charge to point E.

2.8 Potential Energy in an External Field

2.8.1 Potential energy of a single charge

In Section 2.7, the source of the electric field was specified – the charges and their locations - and the potential energy of the system of those charges was determined. In this section, we ask a related but a distinct question. What is the potential energy of a charge q in a given field? This question was, in fact, the starting point that led us to the notion of the electrostatic potential (Sections 2.1 and 2.2). But here we address this question again to clarify in what way it is different from the discussion in Section 2.7.

The main difference is that we are now concerned with the potential energy of a charge (or charges) in an *external* field. The external field **E** is *not* produced by the given charge(s) whose potential energy we wish to calculate. **E** is produced by sources external to the given charge(s).The external sources may be known, but often they are unknown or unspecified; what is specified is the electric field **E** or the electrostatic potential V due to the external sources. We assume that the charge q does not significantly affect the sources producing the external field. This is true if q is very small, or the external sources are held fixed by other unspecified forces. Even if q is finite, its influence on the external sources may still be ignored in the situation when very strong sources far away at infinity produce a finite field **E** in the region of interest. Note again that we are interested in determining the potential energy of a given charge q (and later, a system of charges) in the external field; we are not interested in the potential energy of the sources producing the external electric field.

The external electric field **E** and the corresponding external potential V may vary from point to point. By definition, V at a point P is the work done in bringing a unit positive charge from infinity to the point P.

(We continue to take potential at infinity to be zero.) Thus, work done in bringing a charge q from infinity to the point P in the external field is qV. This work is stored in the form of potential energy of q. If the point P has position vector **r** relative to some origin, we can write:

Potential energy of q at **r** in an external field

$$= qV(\mathbf{r}) \tag{2.27}$$

where $V(\mathbf{r})$ is the external potential at the point **r**.

Thus, if an electron with charge $q = e = 1.6\times10^{-19}$ C is accelerated by a potential difference of ΔV = 1 volt, it would gain energy of $q\Delta V = 1.6 \times 10^{-19}$J. This unit of energy is defined as 1 *electron volt* or 1eV, i.e., 1 eV=1.6×10^{-19}J. The units based on eV are most commonly used in atomic, nuclear and particle physics, (1 keV = 10^3eV = 1.6×10^{-16}J, 1 MeV = 10^6eV = 1.6×10^{-13}J, 1 GeV = 10^9eV = 1.6×10^{-10}J and 1 TeV = 10^{12}eV = 1.6×10^{-7}J). [This has already been defined on Page 117, XI Physics Part I, Table 6.1.]

2.8.2 Potential energy of a system of two charges in an external field

Next, we ask: what is the potential energy of a system of two charges q_1 and q_2 located at $\mathbf{r}_1$and $\mathbf{r}_2$, respectively, in an external field? First, we calculate the work done in bringing the charge q_1 from infinity to $\mathbf{r}_1$. Work done in this step is $q_1\, V(\mathbf{r}_1)$, using Eq. (2.27). Next, we consider the work done in bringing q_2 to $\mathbf{r}_2$. In this step, work is done not only against the external field **E** but also against the field due to q_1.

Work done on q_2 against the external field

$= q_2\, V(\mathbf{r}_2)$

Work done on q_2 against the field due to q_1

$$= \frac{q_1 q_2}{4\pi\varepsilon_o r_{12}}$$

where r_{12} is the distance between q_1 and q_2. We have made use of Eqs. (2.27) and (2.22). By the superposition principle for fields, we add up the work done on q_2 against the two fields (**E** and that due to q_1):

Work done in bringing q_2 to $\mathbf{r}_2$

$$= q_2 V(\mathbf{r}_2) + \frac{q_1 q_2}{4\pi\varepsilon_o r_{12}} \tag{2.28}$$

Thus,

Potential energy of the system

= the total work done in assembling the configuration

$$= q_1 V(\mathbf{r}_1) + q_2 V(\mathbf{r}_2) + \frac{q_1 q_2}{4\pi\varepsilon_0 r_{12}} \tag{2.29}$$

Example 2.5

(a) Determine the electrostatic potential energy of a system consisting of two charges 7 μC and –2 μC (and with no external field) placed at (–9 cm, 0, 0) and (9 cm, 0, 0) respectively.

(b) How much work is required to separate the two charges infinitely away from each other?

EXAMPLE 2.5

EXAMPLE 2.5

(c) Suppose that the same system of charges is now placed in an external electric field $E = A\,(1/r^2)$; $A = 9 \times 10^5$ C m^{-2}. What would the electrostatic energy of the configuration be?

Solution

(a) $U = \dfrac{1}{4\pi\varepsilon_0}\dfrac{q_1 q_2}{r} = 9\times10^9 \times \dfrac{7\times(-2)\times10^{-12}}{0.18} = -0.7$ J.

(b) $W = U_2 - U_1 = 0 - U = 0 - (-0.7) = 0.7$ J.

(c) The mutual interaction energy of the two charges remains unchanged. In addition, there is the energy of interaction of the two charges with the external electric field. We find,

$$q_1 V(\mathbf{r}_1) + q_2 V(\mathbf{r}_2) = A\frac{7\mu\text{C}}{0.09\text{m}} + A\frac{-2\mu\text{C}}{0.09\text{m}}$$

and the net electrostatic energy is

$$q_1 V(\mathbf{r}_1) + q_2 V(\mathbf{r}_2) + \frac{q_1 q_2}{4\pi\varepsilon_0 r_{12}} = A\frac{7\,\mu\text{C}}{0.09\,\text{m}} + A\frac{-2\,\mu\text{C}}{0.09\,\text{m}} - 0.7\,\text{J}$$

$$= 70 - 20 - 0.7 = 49.3\,\text{J}$$

2.8.3 Potential energy of a dipole in an external field

Consider a dipole with charges $q_1 = +q$ and $q_2 = -q$ placed in a uniform electric field **E**, as shown in Fig. 2.16.

FIGURE 2.16 Potential energy of a dipole in a uniform external field.

As seen in the last chapter, in a uniform electric field, the dipole experiences no net force; but experiences a torque $\boldsymbol{\tau}$ given by

$$\boldsymbol{\tau} = \mathbf{p}\times\mathbf{E} \tag{2.30}$$

which will tend to rotate it (unless **p** is parallel or antiparallel to **E**). Suppose an external torque $\boldsymbol{\tau}_{\text{ext}}$ is applied in such a manner that it just neutralises this torque and rotates it in the plane of paper from angle θ_0 to angle θ_1 at an infinitesimal angular speed and *without angular acceleration*. The amount of work done by the external torque will be given by

$$W = \int_{\theta_0}^{\theta_1} \tau_{\text{ext}}(\theta)d\theta = \int_{\theta_0}^{\theta_1} pE\sin\theta\, d\theta$$

$$= pE(\cos\theta_0 - \cos\theta_1) \tag{2.31}$$

This work is stored as the potential energy of the system. We can then associate potential energy $U(\theta)$ with an inclination θ of the dipole. Similar to other potential energies, there is a freedom in choosing the angle where the potential energy U is taken to be zero. A natural choice is to take $\theta_0 = \pi/2$. (An explanation for it is provided towards the end of discussion.) We can then write,

$$U(\theta) = pE\left(\cos\frac{\pi}{2} - \cos\theta\right) = -pE\cos\theta = -\mathbf{p}\cdot\mathbf{E} \tag{2.32}$$

This expression can alternately be understood also from Eq. (2.29). We apply Eq. (2.29) to the present system of two charges $+q$ and $-q$. The potential energy expression then reads

$$U'(\theta) = q[V(\mathbf{r}_1) - V(\mathbf{r}_2)] - \frac{q^2}{4\pi\varepsilon_0 \times 2a} \tag{2.33}$$

Here, $\mathbf{r}_1$ and $\mathbf{r}_2$ denote the position vectors of $+q$ and $-q$. Now, the potential difference between positions $\mathbf{r}_1$ and $\mathbf{r}_2$ equals the work done in bringing a unit positive charge against field from $\mathbf{r}_2$ to $\mathbf{r}_1$. The displacement parallel to the force is $2a\cos\theta$. Thus, $[V(\mathbf{r}_1)-V(\mathbf{r}_2)] = -E \times 2a\cos\theta$. We thus obtain,

$$U'(\theta) = -pE\cos\theta - \frac{q^2}{4\pi\varepsilon_0 \times 2a} = -\mathbf{p}\cdot\mathbf{E} - \frac{q^2}{4\pi\varepsilon_0 \times 2a} \tag{2.34}$$

We note that $U'(\theta)$ differs from $U(\theta)$ by a quantity which is just a constant for a given dipole. Since a constant is insignificant for potential energy, we can drop the second term in Eq. (2.34) and it then reduces to Eq. (2.32).

We can now understand why we took $\theta_0=\pi/2$. In this case, the work done against the *external* field $\mathbf{E}$ in bringing $+q$ and $-q$ are equal and opposite and cancel out, i.e., $q[V(\mathbf{r}_1) - V(\mathbf{r}_2)]=0$.

Example 2.6 A molecule of a substance has a permanent electric dipole moment of magnitude 10^{-29} C m. A mole of this substance is polarised (at low temperature) by applying a strong electrostatic field of magnitude 10^6 V m^{-1}. The direction of the field is suddenly changed by an angle of 60°. Estimate the heat released by the substance in aligning its dipoles along the new direction of the field. For simplicity, assume 100% polarisation of the sample.

Solution Here, dipole moment of each molecules = 10^{-29} C m

As 1 mole of the substance contains 6×10^{23} molecules,

total dipole moment of all the molecules, $p = 6 \times 10^{23} \times 10^{-29}$ C m

$= 6 \times 10^{-6}$ C m

Initial potential energy, $U_i = -pE\cos\theta = -6\times10^{-6}\times10^{6}\cos 0° = -6$ J

Final potential energy (when $\theta = 60°$), $U_f = -6 \times 10^{-6} \times 10^{6}\cos 60° = -3$ J

Change in potential energy = –3 J – (–6J) = 3 J

So, there is loss in potential energy. This must be the energy released by the substance in the form of heat in aligning its dipoles.

EXAMPLE 2.6

2.9 ELECTROSTATICS OF CONDUCTORS

Conductors and insulators were described briefly in Chapter 1. Conductors contain mobile charge carriers. In metallic conductors, these charge carriers are electrons. In a metal, the outer (valence) electrons part away from their atoms and are free to move. These electrons are free within the metal but not free to leave the metal. The free electrons form a kind of 'gas'; they collide with each other and with the ions, and move randomly in different directions. In an external electric field, they drift against the direction of the field. The positive ions made up of the nuclei and the bound electrons remain held in their fixed positions. In electrolytic conductors, the charge carriers are both positive and negative ions; but

the situation in this case is more involved – the movement of the charge carriers is affected both by the external electric field as also by the so-called chemical forces (see Chapter 3). We shall restrict our discussion to metallic solid conductors. Let us note important results regarding electrostatics of conductors.

1. Inside a conductor, electrostatic field is zero

Consider a conductor, neutral or charged. There may also be an external electrostatic field. In the static situation, when there is no current inside or on the surface of the conductor, the electric field is zero everywhere inside the conductor. This fact can be taken as the defining property of a conductor. A conductor has free electrons. As long as electric field is not zero, the free charge carriers would experience force and drift. In the static situation, the free charges have so distributed themselves that the electric field is zero everywhere inside. *Electrostatic field is zero inside a conductor*.

2. At the surface of a charged conductor, electrostatic field must be normal to the surface at every point

If **E** were not normal to the surface, it would have some non-zero component along the surface. Free charges on the surface of the conductor would then experience force and move. In the static situation, therefore, **E** should have no tangential component. Thus *electrostatic field at the surface of a charged conductor must be normal to the surface at every point*. (For a conductor without any surface charge density, field is zero even at the surface.) See result 5.

3. The interior of a conductor can have no excess charge in the static situation

A neutral conductor has equal amounts of positive and negative charges in every small volume or surface element. When the conductor is charged, the excess charge can reside only on the surface in the static situation. This follows from the Gauss's law. Consider any arbitrary volume element v inside a conductor. On the closed surface S bounding the volume element v, electrostatic field is zero. Thus the total electric flux through S is zero. Hence, by Gauss's law, there is no net charge enclosed by S. But the surface S can be made as small as you like, i.e., the volume v can be made vanishingly small. This means *there is no net charge at any point inside the conductor*, and *any excess charge must reside at the surface*.

4. Electrostatic potential is constant throughout the volume of the conductor and has the same value (as inside) on its surface

This follows from results 1 and 2 above. Since **E** = 0 inside the conductor and has no tangential component on the surface, no work is done in moving a small test charge within the conductor and on its surface. That is, there is no potential difference between any two points inside or on the surface of the conductor. Hence, the result. If the conductor is charged,

electric field normal to the surface exists; this means potential will be different for the surface and a point just outside the surface.

In a system of conductors of arbitrary size, shape and charge configuration, each conductor is characterised by a constant value of potential, but this constant may differ from one conductor to the other.

5. Electric field at the surface of a charged conductor

$$\mathbf{E} = \frac{\sigma}{\varepsilon_0}\hat{\mathbf{n}} \tag{2.35}$$

where σ is the surface charge density and $\hat{\mathbf{n}}$ is a unit vector normal to the surface in the outward direction.

To derive the result, choose a pill box (a short cylinder) as the Gaussian surface about any point P on the surface, as shown in Fig. 2.17. The pill box is partly inside and partly outside the surface of the conductor. It has a small area of cross section δS and negligible height.

Just inside the surface, the electrostatic field is zero; just outside, the field is normal to the surface with magnitude E. Thus, the contribution to the total flux through the pill box comes only from the outside (circular) cross-section of the pill box. This equals $\pm E\delta S$ (positive for $\sigma > 0$, negative for $\sigma < 0$), since over the small area δS, **E** may be considered constant and **E** and δS are parallel or antiparallel. The charge enclosed by the pill box is $\sigma\delta S$.

By Gauss's law

$$E\delta S = \frac{|\sigma|\,\delta S}{\varepsilon_0}$$

$$E = \frac{|\sigma|}{\varepsilon_0} \tag{2.36}$$

FIGURE 2.17 The Gaussian surface (a pill box) chosen to derive Eq. (2.35) for electric field at the surface of a charged conductor.

Including the fact that electric field is normal to the surface, we get the vector relation, Eq. (2.35), which is true for both signs of σ. For $\sigma > 0$, electric field is normal to the surface outward; for $\sigma < 0$, electric field is normal to the surface inward.

6. Electrostatic shielding

Consider a conductor with a cavity, with no charges inside the cavity. A remarkable result is that the electric field inside the cavity is zero, whatever be the size and shape of the cavity and whatever be the charge on the conductor and the external fields in which it might be placed. We have proved a simple case of this result already: the electric field inside a charged spherical shell is zero. The proof of the result for the shell makes use of the spherical symmetry of the shell (see Chapter 1). But the vanishing of electric field in the (charge-free) cavity of a conductor is, as mentioned above, a very general result. A related result is that even if the conductor

FIGURE 2.18 The electric field inside a cavity of any conductor is zero. All charges reside only on the outer surface of a conductor with cavity. (There are no charges placed in the cavity.)

is charged or charges are induced on a neutral conductor by an external field, all charges reside only on the outer surface of a conductor with cavity.

The proofs of the results noted in Fig. 2.18 are omitted here, but we note their important implication. Whatever be the charge and field configuration outside, any cavity in a conductor remains shielded from outside electric influence: *the field inside the cavity is always zero*. This is known as *electrostatic shielding*. The effect can be made use of in protecting sensitive instruments from outside electrical influence. Figure 2.19 gives a summary of the important electrostatic properties of a conductor.

FIGURE 2.19 Some important electrostatic properties of a conductor.

EXAMPLE 2.7

Example 2.7

(a) A comb run through one's dry hair attracts small bits of paper. Why?
What happens if the hair is wet or if it is a rainy day? (Remember, a paper does not conduct electricity.)

(b) Ordinary rubber is an insulator. But special rubber tyres of aircraft are made slightly conducting. Why is this necessary?

(c) Vehicles carrying inflammable materials usually have metallic ropes touching the ground during motion. Why?

(d) A bird perches on a bare high power line, and nothing happens to the bird. A man standing on the ground touches the same line and gets a fatal shock. Why?

Solution

(a) This is because the comb gets charged by friction. The molecules in the paper gets polarised by the charged comb, resulting in a net force of attraction. If the hair is wet, or if it is rainy day, friction between hair and the comb reduces. The comb does not get charged and thus it will not attract small bits of paper.

EXAMPLE 2.7

(b) To enable them to conduct charge (produced by friction) to the ground; as too much of static electricity accumulated may result in spark and result in fire.

(c) Reason similar to (b).

(d) Current passes only when there is difference in potential.

2.10 DIELECTRICS AND POLARISATION

Dielectrics are non-conducting substances. In contrast to conductors, they have no (or negligible number of) charge carriers. Recall from Section 2.9 what happens when a conductor is placed in an external electric field. The free charge carriers move and charge distribution in the conductor adjusts itself in such a way that the electric field due to induced charges opposes the external field within the conductor. This happens until, in the static situation, the two fields cancel each other and the net electrostatic field in the conductor is zero. In a dielectric, this free movement of charges is not possible. It turns out that the external field induces dipole moment by stretching or re-orienting molecules of the dielectric. The collective effect of all the molecular dipole moments is net charges on the surface of the dielectric which produce a field that opposes the external field. Unlike in a conductor, however, the opposing field so induced does not exactly cancel the external field. It only reduces it. The extent of the effect depends on the nature of the dielectric. To understand the effect, we need to look at the charge distribution of a dielectric at the molecular level.

FIGURE 2.20 Difference in behaviour of a conductor and a dielectric in an external electric field.

The molecules of a substance may be polar or non-polar. In a non-polar molecule, the centres of positive and negative charges coincide. The molecule then has no permanent (or intrinsic) dipole moment. Examples of non-polar molecules are oxygen (O_2) and hydrogen (H_2) molecules which, because of their symmetry, have no dipole moment. On the other hand, a polar molecule is one in which the centres of positive and negative charges are separated (even when there is no external field). Such molecules have a permanent dipole moment. An ionic molecule such as HCl or a molecule of water (H_2O) are examples of polar molecules.

FIGURE 2.21 Some examples of polar and non-polar molecules.

FIGURE 2.22 A dielectric develops a net dipole moment in an external electric field. (a) Non-polar molecules, (b) Polar molecules.

In an external electric field, the positive and negative charges of a non-polar molecule are displaced in opposite directions. The displacement stops when the external force on the constituent charges of the molecule is balanced by the restoring force (due to internal fields in the molecule). The non-polar molecule thus develops an induced dipole moment. The dielectric is said to be polarised by the external field. We consider only the simple situation when the induced dipole moment is in the direction of the field and is proportional to the field strength. (Substances for which this assumption is true are called *linear isotropic dielectrics*.) The induced dipole moments of different molecules add up giving a net dipole moment of the dielectric in the presence of the external field.

A dielectric with polar molecules also develops a net dipole moment in an external field, but for a different reason. In the absence of any external field, the different permanent dipoles are oriented randomly due to thermal agitation; so the total dipole moment is zero. When an external field is applied, the individual dipole moments tend to align with the field. When summed over all the molecules, there is then a net dipole moment in the direction of the external field, i.e., the dielectric is polarised. The extent of polarisation depends on the relative strength of two mutually opposite factors: the dipole potential energy in the external field tending to align the dipoles with the field and thermal energy tending to disrupt the alignment. There may be, in addition, the 'induced dipole moment' effect as for non-polar molecules, but generally the alignment effect is more important for polar molecules.

Thus in either case, whether polar or non-polar, a dielectric develops a net dipole moment in the presence of an external field. The dipole moment per unit volume is called *polarisation* and is denoted by **P**. For linear isotropic dielectrics,

$$\mathbf{P} = \chi_e \mathbf{E} \tag{2.37}$$

where χ_e is a constant characteristic of the dielectric and is known as the *electric susceptibility* of the dielectric medium.

It is possible to relate χ_e to the molecular properties of the substance, but we shall not pursue that here.

The question is: how does the polarised dielectric modify the original external field inside it? Let us consider, for simplicity, a rectangular dielectric slab placed in a uniform external field $\mathbf{E}_0$ parallel to two of its faces. The field causes a uniform polarisation **P** of the dielectric. Thus

every volume element Δv of the slab has a dipole moment $\mathbf{P}\,\Delta v$ in the direction of the field. The volume element Δv is macroscopically small but contains a very large number of molecular dipoles. Anywhere inside the dielectric, the volume element Δv has no net charge (though it has net dipole moment). This is, because, the positive charge of one dipole sits close to the negative charge of the adjacent dipole. However, at the surfaces of the dielectric normal to the electric field, there is evidently a net charge density. As seen in Fig 2.23, the positive ends of the dipoles remain unneutralised at the right surface and the negative ends at the left surface. The unbalanced charges are the induced charges due to the external field.

Thus the polarised dielectric is equivalent to two charged surfaces with induced surface charge densities, say σ_p and $-\sigma_p$. Clearly, the field produced by these surface charges opposes the external field. The total field in the dielectric is, thereby, reduced from the case when no dielectric is present. We should note that the surface charge density $\pm\sigma_p$ arises from bound (not free charges) in the dielectric.

FIGURE 2.23 A uniformly polarised dielectric amounts to induced surface charge density, but no volume charge density.

2.11 Capacitors and Capacitance

A capacitor is a system of two conductors separated by an insulator (Fig. 2.24). The conductors have charges, say Q_1 and Q_2, and potentials V_1 and V_2. Usually, in practice, the two conductors have charges Q and $-Q$, with potential difference $V = V_1 - V_2$ between them. We shall consider only this kind of charge configuration of the capacitor. (Even a single conductor can be used as a capacitor by assuming the other at infinity.) The conductors may be so charged by connecting them to the two terminals of a battery. Q is called the charge of the capacitor, though this, in fact, is the charge on one of the conductors – the total charge of the capacitor is zero.

The electric field in the region between the conductors is proportional to the charge Q. That is, if the charge on the capacitor is, say doubled, the electric field will also be doubled at every point. (This follows from the direct proportionality between field and charge implied by Coulomb's law and the superposition principle.) Now, potential difference V is the work done per unit positive charge in taking a small test charge from the conductor 2 to 1 against the field. Consequently, V is also proportional to Q, and the ratio Q/V is a constant:

FIGURE 2.24 A system of two conductors separated by an insulator forms a capacitor.

$$C = \frac{Q}{V} \qquad (2.38)$$

The constant C is called the *capacitance* of the capacitor. C is independent of Q or V, as stated above. The capacitance C depends only on the

geometrical configuration (shape, size, separation) of the system of two conductors. [As we shall see later, it also depends on the nature of the insulator (dielectric) separating the two conductors.] The SI unit of capacitance is 1 farad (=1 coulomb volt^{-1}) or 1 F = 1 C V^{-1}. A capacitor with fixed capacitance is symbolically shown as ⊣ ⊢, while the one with variable capacitance is shown as ⫽.

Equation (2.38) shows that for large C, V is small for a given Q. This means a capacitor with large capacitance can hold large amount of charge Q at a relatively small V. This is of practical importance. High potential difference implies strong electric field around the conductors. A strong electric field can ionise the surrounding air and accelerate the charges so produced to the oppositely charged plates, thereby neutralising the charge on the capacitor plates, at least partly. In other words, the charge of the capacitor leaks away due to the reduction in insulating power of the intervening medium.

The maximum electric field that a dielectric medium can withstand without break-down (of its insulating property) is called its *dielectric strength*; for air it is about 3×10^6 Vm^{-1}. For a separation between conductors of the order of 1 cm or so, this field corresponds to a potential difference of 3×10^4 V between the conductors. Thus, for a capacitor to store a large amount of charge without leaking, its capacitance should be high enough so that the potential difference and hence the electric field do not exceed the break-down limits. Put differently, there is a limit to the amount of charge that can be stored on a given capacitor without significant leaking. In practice, a farad is a very big unit; the most common units are its sub-multiples 1 μF = 10^{-6} F, 1 nF = 10^{-9} F, 1 pF = 10^{-12} F, etc. Besides its use in storing charge, a capacitor is a key element of most ac circuits with important functions, as described in Chapter 7.

2.12 The Parallel Plate Capacitor

A parallel plate capacitor consists of two large plane parallel conducting plates separated by a small distance (Fig. 2.25). We first take the intervening medium between the plates to be vacuum. The effect of a dielectric medium between the plates is discussed in the next section. Let A be the area of each plate and d the separation between them. The two plates have charges Q and $-Q$. Since d is much smaller than the linear dimension of the plates ($d^2 << A$), we can use the result on electric field by an infinite plane sheet of uniform surface charge density (Section 1.15). Plate 1 has surface charge density $\sigma = Q/A$ and plate 2 has a surface charge density $-\sigma$. Using Eq. (1.33), the electric field in different regions is:

FIGURE 2.25 The parallel plate capacitor.

Outer region I (region above the plate 1),

$$E = \frac{\sigma}{2\varepsilon_0} - \frac{\sigma}{2\varepsilon_0} = 0 \qquad (2.39)$$

Outer region II (region below the plate 2),

$$E = \frac{\sigma}{2\varepsilon_0} - \frac{\sigma}{2\varepsilon_0} = 0 \tag{2.40}$$

In the inner region between the plates 1 and 2, the electric fields due to the two charged plates add up, giving

$$E = \frac{\sigma}{2\varepsilon_0} + \frac{\sigma}{2\varepsilon_0} = \frac{\sigma}{\varepsilon_0} = \frac{Q}{\varepsilon_0 A} \tag{2.41}$$

The direction of electric field is from the positive to the negative plate.

Thus, the electric field is localised between the two plates and is uniform throughout. For plates with finite area, this will not be true near the outer boundaries of the plates. The field lines bend outward at the edges – an effect called 'fringing of the field'. By the same token, σ will not be strictly uniform on the entire plate. [E and σ are related by Eq. (2.35).] However, for $d^2 << A$, these effects can be ignored in the regions sufficiently far from the edges, and the field there is given by Eq. (2.41). Now for uniform electric field, potential difference is simply the electric field times the distance between the plates, that is,

$$V = E\,d = \frac{1}{\varepsilon_0}\frac{Qd}{A} \tag{2.42}$$

The capacitance C of the parallel plate capacitor is then

$$C = \frac{Q}{V} = = \frac{\varepsilon_0 A}{d} \tag{2.43}$$

which, as expected, depends only on the geometry of the system. For typical values like $A = 1\ \text{m}^2$, $d = 1$ mm, we get

$$C = \frac{8.85\times10^{-12}\,\text{C}^2\text{N}^{-1}\text{m}^{-2}\times1\,\text{m}^2}{10^{-3}\,\text{m}} = 8.85\times10^{-9}\,\text{F} \tag{2.44}$$

(You can check that if 1F= 1C V^{-1} = 1C $(NC^{-1}m)^{-1}$ = 1 $C^2\,N^{-1}m^{-1}$.) This shows that 1F is too big a unit in practice, as remarked earlier. Another way of seeing the 'bigness' of 1F is to calculate the area of the plates needed to have C = 1F for a separation of, say 1 cm:

$$A = \frac{Cd}{\varepsilon_0} = \frac{1\text{F}\times10^{-2}\,\text{m}}{8.85\times10^{-12}\,\text{C}^2\text{N}^{-1}\text{m}^{-2}} = 10^9\,\text{m}^2 \tag{2.45}$$

which is a plate about 30 km in length and breadth!

2.13 Effect of Dielectric on Capacitance

With the understanding of the behavior of dielectrics in an external field developed in Section 2.10, let us see how the capacitance of a parallel plate capacitor is modified when a dielectric is present. As before, we have two large plates, each of area A, separated by a distance d. The charge on the plates is $\pm Q$, corresponding to the charge density $\pm\sigma$ (with $\sigma = Q/A$). When there is vacuum between the plates,

$$E_0 = \frac{\sigma}{\varepsilon_0}$$

PHYSICS

Factors affecting capacitance, capacitors in action
Interactive Java tutorial
http://micro.magnet.fsa.edu/electromag/java/capacitance/

and the potential difference V_0 is

$V_0 = E_0 d$

The capacitance C_0 in this case is

$$C_0 = \frac{Q}{V_0} = \varepsilon_0 \frac{A}{d} \tag{2.46}$$

Consider next a dielectric inserted between the plates fully occupying the intervening region. The dielectric is polarised by the field and, as explained in Section 2.10, the effect is equivalent to two charged sheets (at the surfaces of the dielectric normal to the field) with surface charge densities σ_p and $-\sigma_p$. The electric field in the dielectric then corresponds to the case when the net surface charge density on the plates is $\pm(\sigma - \sigma_p)$. That is,

$$E = \frac{\sigma - \sigma_P}{\varepsilon_0} \tag{2.47}$$

so that the potential difference across the plates is

$$V = E\,d = \frac{\sigma - \sigma_P}{\varepsilon_0} d \tag{2.48}$$

For linear dielectrics, we expect σ_p to be proportional to E_0, i.e., to σ. Thus, $(\sigma - \sigma_p)$ is proportional to σ and we can write

$$\sigma - \sigma_P = \frac{\sigma}{K} \tag{2.49}$$

where K is a constant characteristic of the dielectric. Clearly, $K > 1$. We then have

$$V = \frac{\sigma d}{\varepsilon_0 K} = \frac{Qd}{A\varepsilon_0 K} \tag{2.50}$$

The capacitance C, with dielectric between the plates, is then

$$C = \frac{Q}{V} = \frac{\varepsilon_0 KA}{d} \tag{2.51}$$

The product $\varepsilon_0 K$ is called the *permittivity* of the medium and is denoted by ε

$$\varepsilon = \varepsilon_0 K \tag{2.52}$$

For vacuum $K = 1$ and $\varepsilon = \varepsilon_0$; ε_0 is called the *permittivity of the vacuum.* The dimensionless ratio

$$K = \frac{\varepsilon}{\varepsilon_0} \tag{2.53}$$

is called the *dielectric constant* of the substance. As remarked before, from Eq. (2.49), it is clear that K is greater than 1. From Eqs. (2.46) and (2. 51)

$$K = \frac{C}{C_0} \tag{2.54}$$

Thus, the dielectric constant of a substance is the factor (>1) by which the capacitance increases from its vacuum value, when the dielectric is inserted fully between the plates of a capacitor. Though we arrived at

Eq. (2.54) for the case of a parallel plate capacitor, it holds good for any type of capacitor and can, in fact, be viewed in general as a definition of the dielectric constant of a substance.

ELECTRIC DISPLACEMENT

We have introduced the notion of dielectric constant and arrived at Eq. (2.54), without giving the explicit relation between the induced charge density σ_p and the polarisation **P**.

We take without proof the result that

$$\sigma_P = \mathbf{P}\cdot\hat{\mathbf{n}}$$

where $\hat{\mathbf{n}}$ is a unit vector along the outward normal to the surface. Above equation is general, true for any shape of the dielectric. For the slab in Fig. 2.23, **P** is along $\hat{\mathbf{n}}$ at the right surface and opposite to $\hat{\mathbf{n}}$ at the left surface. Thus at the right surface, induced charge density is positive and at the left surface, it is negative, as guessed already in our qualitative discussion before. Putting the equation for electric field in vector form

$$\mathbf{E}\cdot\hat{\mathbf{n}} = \frac{\sigma - \mathbf{P}\cdot\hat{\mathbf{n}}}{\varepsilon_0}$$

or $(\varepsilon_0 \mathbf{E} + \mathbf{P})\cdot\hat{\mathbf{n}} = \sigma$

The quantity $\varepsilon_0 \mathbf{E} + \mathbf{P}$ is called the *electric displacement* and is denoted by **D**. It is a vector quantity. Thus,

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P},\ \mathbf{D}\cdot\hat{\mathbf{n}} = \sigma,$$

The significance of **D** is this : in vacuum, **E** is related to the free charge density σ. When a dielectric medium is present, the corresponding role is taken up by **D**. For a dielectric medium, it is **D** not **E** that is directly related to free charge density σ, as seen in above equation. Since **P** is in the same direction as **E**, all the three vectors **P**, **E** and **D** are parallel.

The ratio of the magnitudes of **D** and **E** is

$$\frac{D}{E} = \frac{\sigma\varepsilon_0}{\sigma - \sigma_P} = \varepsilon_0 K$$

Thus,

$$\mathbf{D} = \varepsilon_0 K \mathbf{E}$$

and $\mathbf{P} = \mathbf{D} - \varepsilon_0\mathbf{E} = \varepsilon_0 (K-1)\mathbf{E}$

This gives for the electric susceptibility χ_e defined in Eq. (2.37)

$$\chi_e = \varepsilon_0 (K-1)$$

Example 2.8 A slab of material of dielectric constant K has the same area as the plates of a parallel-plate capacitor but has a thickness $(3/4)d$, where d is the separation of the plates. How is the capacitance changed when the slab is inserted between the plates?

Solution Let $E_0 = V_0/d$ be the electric field between the plates when there is no dielectric and the potential difference is V_0. If the dielectric is now inserted, the electric field in the dielectric will be $E = E_0/K$. The potential difference will then be

EXAMPLE 2.8

EXAMPLE 2.8

$$V = E_0(\frac{1}{4}d) + \frac{E_0}{K}(\frac{3}{4}d)$$

$$= E_0 d(\frac{1}{4} + \frac{3}{4K}) = V_0 \frac{K+3}{4K}$$

The potential difference decreases by the factor $(K + 3)/4K$ while the free charge Q_0 on the plates remains unchanged. The capacitance thus increases

$$C = \frac{Q_0}{V} = \frac{4K}{K+3}\frac{Q_0}{V_0} = \frac{4K}{K+3}C_0$$

2.14 COMBINATION OF CAPACITORS

We can combine several capacitors of capacitance C_1, C_2,.., C_n to obtain a system with some effective capacitance C. The effective capacitance depends on the way the individual capacitors are combined. Two simple possibilities are discussed below.

2.14.1 Capacitors in series

Figure 2.26 shows capacitors C_1 and C_2 combined in series.

The left plate of C_1 and the right plate of C_2 are connected to two terminals of a battery and have charges Q and $-Q$, respectively. It then follows that the right plate of C_1 has charge $-Q$ and the left plate of C_2 has charge Q. If this was not so, the net charge on each capacitor would not be zero. This would result in an electric field in the conductor connecting C_1 and C_2. Charge would flow until the net charge on both C_1 and C_2 is zero and there is no electric field in the conductor connecting C_1 and C_2. Thus, in the series combination, charges on the two plates ($\pm Q$) are the same on each capacitor. The total potential drop V across the combination is the sum of the potential drops V_1 and V_2 across C_1 and C_2, respectively.

FIGURE 2.26 Combination of two capacitors in series.

$$V = V_1 + V_2 = \frac{Q}{C_1} + \frac{Q}{C_2} \tag{2.55}$$

$$\text{i.e.,}\ \frac{V}{Q} = \frac{1}{C_1} + \frac{1}{C_2}, \tag{2.56}$$

Now we can regard the combination as an effective capacitor with charge Q and potential difference V. The *effective capacitance* of the combination is

$$C = \frac{Q}{V} \tag{2.57}$$

We compare Eq. (2.57) with Eq. (2.56), and obtain

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} \tag{2.58}$$

FIGURE 2.27 Combination of n capacitors in series.

The proof clearly goes through for any number of capacitors arranged in a similar way. Equation (2.55), for n capacitors arranged in series, generalises to

$$V = V_1 + V_2 + \ldots + V_n = \frac{Q}{C_1} + \frac{Q}{C_2} + \ldots + \frac{Q}{C_n} \qquad (2.59)$$

Following the same steps as for the case of two capacitors, we get the general formula for effective capacitance of a series combination of n capacitors:

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \ldots + \frac{1}{C_n} \qquad (2.60)$$

2.14.2 Capacitors in parallel

Figure 2.28 (a) shows two capacitors arranged in parallel. In this case, the same potential difference is applied across both the capacitors. But the plate charges ($\pm Q_1$) on capacitor 1 and the plate charges ($\pm Q_2$) on the capacitor 2 are not necessarily the same:

$$Q_1 = C_1 V,\ Q_2 = C_2 V \qquad (2.61)$$

The equivalent capacitor is one with charge

$$Q = Q_1 + Q_2 \qquad (2.62)$$

and potential difference V.

$$Q = CV = C_1 V + C_2 V \qquad (2.63)$$

The effective capacitance C is, from Eq. (2.63),

$$C = C_1 + C_2 \qquad (2.64)$$

The general formula for effective capacitance C for parallel combination of n capacitors [Fig. 2.28 (b)] follows similarly,

$$Q = Q_1 + Q_2 + \ldots + Q_n \qquad (2.65)$$

$$\text{i.e., } CV = C_1 V + C_2 V + \ldots C_n V \qquad (2.66)$$

which gives

$$C = C_1 + C_2 + \ldots C_n \qquad (2.67)$$

FIGURE 2.28 Parallel combination of (a) two capacitors, (b) n capacitors.

Example 2.9 A network of four 10 μF capacitors is connected to a 500 V supply, as shown in Fig. 2.29. Determine (a) the equivalent capacitance of the network and (b) the charge on each capacitor. (Note, the *charge on a capacitor* is the charge on the plate with higher potential, equal and opposite to the charge on the plate with lower potential.)

FIGURE 2.29

EXAMPLE 2.9

EXAMPLE 2.9

Solution

(a) In the given network, C_1, C_2 and C_3 are connected in series. The effective capacitance C' of these three capacitors is given by

$$\frac{1}{C'} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}$$

For $C_1 = C_2 = C_3 = 10\ \mu F$, $C' = (10/3)\ \mu F$. The network has C' and C_4 connected in parallel. Thus, the equivalent capacitance C of the network is

$$C = C' + C_4 = \left(\frac{10}{3} + 10\right)\ \mu F = 13.3 \mu F$$

(b) Clearly, from the figure, the charge on each of the capacitors, C_1, C_2 and C_3 is the same, say Q. Let the charge on C_4 be Q'. Now, since the potential difference across AB is Q/C_1, across BC is Q/C_2, across CD is Q/C_3, we have

$$\frac{Q}{C_1} + \frac{Q}{C_2} + \frac{Q}{C_3} = 500\ V$$

Also, $Q'/C_4 = 500$ V.

This gives for the given value of the capacitances,

$$Q = 500\,V \times \frac{10}{3}\mu F = 1.7 \times 10^{-3} C \text{ and}$$

$$Q' = 500\,V \times 10\,\mu F = 5.0 \times 10^{-3}\ C$$

2.15 ENERGY STORED IN A CAPACITOR

A capacitor, as we have seen above, is a system of two conductors with charge Q and $-Q$. To determine the energy stored in this configuration, consider initially two uncharged conductors 1 and 2. Imagine next a process of transferring charge from conductor 2 to conductor 1 bit by bit, so that at the end, conductor 1 gets charge Q. By charge conservation, conductor 2 has charge $-Q$ at the end (Fig 2.30).

FIGURE 2.30 (a) Work done in a small step of building charge on conductor 1 from Q' to $Q' + \delta Q'$. (b) Total work done in charging the capacitor may be viewed as stored in the energy of electric field between the plates.

In transferring positive charge from conductor 2 to conductor 1, work will be done externally, since at any stage conductor 1 is at a higher potential than conductor 2. To calculate the total work done, we first calculate the work done in a small step involving transfer of an infinitesimal (i.e., vanishingly small) amount of charge. Consider the intermediate situation when the conductors 1 and 2 have charges Q' and $-Q'$ respectively. At this stage, the potential difference V' between conductors 1 to 2 is Q'/C, where C is the capacitance of the system. Next imagine that a small charge $\delta Q'$ is transferred from conductor 2 to 1. Work done in this step ($\delta W'$), resulting in charge Q' on conductor 1 increasing to $Q' + \delta Q'$, is given by

$$\delta W = V' \delta Q' = \frac{Q'}{C} \delta Q' \qquad (2.68)$$

Since $\delta Q'$ can be made as small as we like, Eq. (2.68) can be written as

$$\delta W = \frac{1}{2C}[(Q' + \delta Q')^2 - Q'^2] \tag{2.69}$$

Equations (2.68) and (2.69) are identical because the term of second order in $\delta Q'$, i.e., $\delta Q'^2/2C$, is negligible, since $\delta Q'$ is arbitrarily small. The total work done (W) is the sum of the small work (δW) over the very large number of steps involved in building the charge Q' from zero to Q.

$$W = \sum_{\text{sum over all steps}} \delta W$$

$$= \sum_{\text{sum over all steps}} \frac{1}{2C}[(Q' + \delta Q')^2 - Q'^2] \tag{2.70}$$

$$= \frac{1}{2C}[\{\delta Q'^2 - 0\} + \{(2\delta Q')^2 - \delta Q'^2\} + \{(3\delta Q')^2 - (2\delta Q')^2\} + \ldots$$
$$+ \{Q^2 - (Q - \delta Q)^2\}] \tag{2.71}$$

$$= \frac{1}{2C}[Q^2 - 0] = \frac{Q^2}{2C} \tag{2.72}$$

The same result can be obtained directly from Eq. (2.68) by integration

$$W = \int_0^Q \frac{Q'}{C}\delta Q' = \frac{1}{C}\frac{Q'^2}{2}\Big|_0^Q = \frac{Q^2}{2C}$$

This is not surprising since integration is nothing but summation of a large number of small terms.

We can write the final result, Eq. (2.72) in different ways

$$W = \frac{Q^2}{2C} = \frac{1}{2}CV^2 = \frac{1}{2}QV \tag{2.73}$$

Since electrostatic force is conservative, this work is stored in the form of potential energy of the system. For the same reason, the final result for potential energy [Eq. (2.73)] is independent of the manner in which the charge configuration of the capacitor is built up. When the capacitor discharges, this stored-up energy is released. It is possible to view the potential energy of the capacitor as 'stored' in the electric field between the plates. To see this, consider for simplicity, a parallel plate capacitor [of area A(of each plate) and separation d between the plates].

Energy stored in the capacitor

$$= \frac{1}{2}\frac{Q^2}{C} = \frac{(A\sigma)^2}{2} \times \frac{d}{\varepsilon_0 A} \tag{2.74}$$

The surface charge density σ is related to the electric field E between the plates,

$$E = \frac{\sigma}{\varepsilon_0} \tag{2.75}$$

From Eqs. (2.74) and (2.75) , we get

Energy stored in the capacitor

$$U = (1/2)\varepsilon_0 E^2 \times A\,d \tag{2.76}$$

Note that *Ad* is the volume of the region between the plates (where electric field alone exists). If we define *energy density as energy stored per unit volume of space*, Eq (2.76) shows that

Energy density of electric field,

$$u = (1/2)\varepsilon_0 E^2 \qquad (2.77)$$

Though we derived Eq. (2.77) for the case of a parallel plate capacitor, the result on energy density of an electric field is, in fact, very general and holds true for electric field due to any configuration of charges.

Example 2.10 (a) A 900 pF capacitor is charged by 100 V battery [Fig. 2.31(a)]. How much electrostatic energy is stored by the capacitor? (b) The capacitor is disconnected from the battery and connected to another 900 pF capacitor [Fig. 2.31(b)]. What is the electrostatic energy stored by the system?

FIGURE 2.31

Solution

(a) The charge on the capacitor is

$Q = CV = 900 \times 10^{-12}$ F $\times$ 100 V $= 9 \times 10^{-8}$ C

The energy stored by the capacitor is

$= (1/2)\, CV^2 = (1/2)\, QV$

$= (1/2) \times 9 \times 10^{-8}$C $\times$ 100 V $= 4.5 \times 10^{-6}$ J

(b) In the steady situation, the two capacitors have their positive plates at the same potential, and their negative plates at the same potential. Let the common potential difference be V′. The charge on each capacitor is then $Q' = CV'$. By charge conservation, $Q' = Q/2$. This implies $V' = V/2$. The total energy of the system is

$$= 2 \times \frac{1}{2} Q'V' = \frac{1}{4} QV = 2.25 \times 10^{-6}\,\text{J}$$

Thus in going from (a) to (b), though no charge is lost; the final energy is only half the initial energy. *Where has the remaining energy gone?*

There is a transient period before the system settles to the situation (b). During this period, a transient current flows from the first capacitor to the second. Energy is lost during this time in the form of heat and electromagnetic radiation.

EXAMPLE 2.10

2.16 Van De Graaff Generator

This is a machine that can build up high voltages of the order of a few million volts. The resulting large electric fields are used to accelerate charged particles (electrons, protons, ions) to high energies needed for experiments to probe the small scale structure of matter. The principle underlying the machine is as follows.

Suppose we have a large spherical conducting shell of radius R, on which we place a charge Q. This charge spreads itself uniformly all over the sphere. As we have seen in Section 1.14, the field outside the sphere is just that of a point charge Q at the centre; while the field inside the sphere vanishes. So the potential outside is that of a point charge; and inside it is constant, namely the value at the radius R. We thus have:

Potential inside conducting spherical shell of radius R carrying charge Q = constant

$$= \frac{1}{4\pi\varepsilon_0}\frac{Q}{R} \tag{2.78}$$

Now, as shown in Fig. 2.32, let us suppose that in some way we introduce a small sphere of radius r, carrying some charge q, into the large one, and place it at the centre. The potential due to this new charge clearly has the following values at the radii indicated:

Potential due to small sphere of radius r carrying charge q

$$= \frac{1}{4\pi\varepsilon_0}\frac{q}{r} \text{ at surface of small sphere}$$

$$= \frac{1}{4\pi\varepsilon_0}\frac{q}{R} \text{ at large shell of radius } R. \tag{2.79}$$

Taking both charges q and Q into account we have for the total potential V and the potential difference the values

$$V(R) = \frac{1}{4\pi\varepsilon_0}\left(\frac{Q}{R} + \frac{q}{R}\right)$$

$$V(r) = \frac{1}{4\pi\varepsilon_0}\left(\frac{Q}{R} + \frac{q}{r}\right)$$

$$V(r) - V(R) = \frac{q}{4\pi\varepsilon_0}\left(\frac{1}{r} - \frac{1}{R}\right) \tag{2.80}$$

Assume now that q is positive. We see that, independent of the amount of charge Q that may have accumulated on the larger sphere and even if it is positive, the inner sphere is always at a higher potential: the difference $V(r)-V(R)$ is positive. The potential due to Q is constant upto radius R and so cancels out in the difference!

This means that if we now connect the smaller and larger sphere by a wire, the charge q on the former

FIGURE 2.32 Illustrating the principle of the electrostatic generator.

PHYSICS

Van de Graaff generator, principle and demonstration:
http://amasce.com/emotor/vdg.html
http://www.coe.ufrj.br/~acmg/myvdg.html

FIGURE 2.33 Principle of construction of Van de Graaff generator.

will immediately flow onto the matter, even though the charge Q may be quite large. The natural tendency is for positive charge to move from higher to lower potential. Thus, provided we are somehow able to introduce the small charged sphere into the larger one, we can in this way keep piling up larger and larger amount of charge on the latter. The potential (Eq. 2.78) at the outer sphere would also keep rising, at least until we reach the breakdown field of air.

This is the principle of the van de Graaff generator. It is a machine capable of building up potential difference of a few million volts, and fields close to the breakdown field of air which is about 3×10^6 V/m. A schematic diagram of the van de Graaff generator is given in Fig. 2.33. A large spherical conducting shell (of few metres radius) is supported at a height several meters above the ground on an insulating column. A long narrow endless belt insulating material, like rubber or silk, is wound around two pulleys – one at ground level, one at the centre of the shell. This belt is kept continuously moving by a motor driving the lower pulley. It continuously carries positive charge, sprayed on to it by a brush at ground level, to the top. There it transfers its positive charge to another conducting brush connected to the large shell. Thus positive charge is transferred to the shell, where it spreads out uniformly on the outer surface. In this way, voltage differences of as much as 6 or 8 million volts (with respect to ground) can be built up.

SUMMARY

1. Electrostatic force is a conservative force. Work done by an external force (equal and opposite to the electrostatic force) in bringing a charge q from a point R to a point P is $V_P - V_R$, which is the difference in potential energy of charge q between the final and initial points.
2. Potential at a point is the work done per unit charge (by an external agency) in bringing a charge from infinity to that point. Potential at a point is arbitrary to within an additive constant, since it is the potential difference between two points which is physically significant. If potential at infinity is chosen to be zero; potential at a point with position vector **r** due to a point charge Q placed at the origin is given is given by

$$V(\mathbf{r}) = \frac{1}{4\pi\varepsilon_o}\frac{Q}{r}$$

3. The electrostatic potential at a point with position vector **r** due to a point dipole of dipole moment **p** placed at the origin is

$$V(\mathbf{r}) = \frac{1}{4\pi\varepsilon_o}\frac{\mathbf{p}\cdot\hat{\mathbf{r}}}{r^2}$$

The result is true also for a dipole (with charges $-q$ and q separated by $2a$) for $r >> a$.

4. For a charge configuration q_1, q_2, ..., q_n with position vectors $\mathbf{r}_1$, $\mathbf{r}_2$, ... $\mathbf{r}_n$, the potential at a point P is given by the superposition principle

$$V = \frac{1}{4\pi\varepsilon_0}\left(\frac{q_1}{r_{1P}} + \frac{q_2}{r_{2P}} + \ldots + \frac{q_n}{r_{nP}}\right)$$

where r_{1P} is the distance between q_1 and P, as and so on.

5. An equipotential surface is a surface over which potential has a constant value. For a point charge, concentric spheres centered at a location of the charge are equipotential surfaces. The electric field **E** at a point is perpendicular to the equipotential surface through the point. **E** is in the direction of the steepest decrease of potential.

6. Potential energy stored in a system of charges is the work done (by an external agency) in assembling the charges at their locations. Potential energy of two charges q_1, q_2 at $\mathbf{r}_1$, $\mathbf{r}_2$ is given by

$$U = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r_{12}}$$

where r_{12} is distance between q_1 and q_2.

7. The potential energy of a charge q in an external potential $V(\mathbf{r})$ is $qV(\mathbf{r})$. The potential energy of a dipole moment **p** in a uniform electric field **E** is $-\mathbf{p}\cdot\mathbf{E}$.

8. Electrostatics field **E** is zero in the interior of a conductor; just outside the surface of a charged conductor, **E** is normal to the surface given by

$\mathbf{E} = \frac{\sigma}{\varepsilon_0}\hat{\mathbf{n}}$ where $\hat{\mathbf{n}}$ is the unit vector along the outward normal to the surface and σ is the surface charge density. Charges in a conductor can reside only at its surface. Potential is constant within and on the surface of a conductor. In a cavity within a conductor (with no charges), the electric field is zero.

9. A capacitor is a system of two conductors separated by an insulator. Its capacitance is defined by $C = Q/V$, where Q and $-Q$ are the charges on the two conductors and V is the potential difference between them. C is determined purely geometrically, by the shapes, sizes and relative positions of the two conductors. The unit of capacitance is farad:, 1 F = 1 C V^{-1}. For a parallel plate capacitor (with vacuum between the plates),

$$C = \varepsilon_0 \frac{A}{d}$$

where A is the area of each plate and d the separation between them.

10. If the medium between the plates of a capacitor is filled with an insulating substance (dielectric), the electric field due to the charged plates induces a net dipole moment in the dielectric. This effect, called polarisation, gives rise to a field in the opposite direction. The net electric field inside the dielectric and hence the potential difference between the plates is thus reduced. Consequently, the capacitance C increases from its value C_0 when there is no medium (vacuum),

$C = KC_0$

where K is the dielectric constant of the insulating substance.

11. For capacitors in the series combination, the total capacitance C is given by

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \ldots$$

In the parallel combination, the total capacitance C is:

$C = C_1 + C_2 + C_3 + \ldots$

where C_1, C_2, C_3... are individual capacitances.

12. The energy U stored in a capacitor of capacitance C, with charge Q and voltage V is

$$U = \frac{1}{2}QV = \frac{1}{2}CV^2 = \frac{1}{2}\frac{Q^2}{C}$$

The electric energy density (energy per unit volume) in a region with electric field is $(1/2)\varepsilon_0 E^2$.

13. A Van de Graaff generator consists of a large spherical conducting shell (a few metre in diameter). By means of a moving belt and suitable brushes, charge is continuously transferred to the shell and potential difference of the order of several million volts is built up, which can be used for accelerating charged particles.

Physical quantity	Symbol	Dimensions	Unit	Remark
Potential	ϕ or V	$[M^1 L^2 T^{-3} A^{-1}]$	V	Potential difference is physically significant
Capacitance	C	$[M^{-1} L^{-2} T^{-4} A^2]$	F	
Polarisation	$\mathbf{P}$	$[L^{-2} AT]$	$C\ m^{-2}$	Dipole moment per unit volume
Dielectric constant	K	[Dimensionless]		

POINTS TO PONDER

1. Electrostatics deals with forces between charges at rest. But if there is a force on a charge, how can it be at rest? Thus, when we are talking of electrostatic force between charges, it should be understood that each charge is being kept at rest by some unspecified force that opposes the net Coulomb force on the charge.
2. A capacitor is so configured that it confines the electric field lines within a small region of space. Thus, even though field may have considerable strength, the potential difference between the two conductors of a capacitor is small.
3. Electric field is discontinuous across the surface of a spherical charged shell. It is zero inside and $\frac{\sigma}{\varepsilon_0}\hat{\mathbf{n}}$ outside. Electric potential is, however continuous across the surface, equal to $q/4\pi\varepsilon_0 R$ at the surface.
4. The torque $\mathbf{p} \times \mathbf{E}$ on a dipole causes it to oscillate about $\mathbf{E}$. Only if there is a dissipative mechanism, the oscillations are damped and the dipole eventually aligns with $\mathbf{E}$.

5. Potential due to a charge q at its own location is not defined – it is infinite.
6. In the expression $qV(\mathbf{r})$ for potential energy of a charge q, $V(\mathbf{r})$ is the potential due to external charges and not the potential due to q. As seen in point 5, this expression will be ill-defined if $V(\mathbf{r})$ includes potential due to a charge q itself.
7. A cavity inside a conductor is shielded from outside electrical influences. It is worth noting that electrostatic shielding does not work the other way round; that is, if you put charges inside the cavity, the exterior of the conductor is not shielded from the fields by the inside charges.

EXERCISES

2.1 Two charges 5×10^{-8} C and -3×10^{-8} C are located 16 cm apart. At what point(s) on the line joining the two charges is the electric potential zero? Take the potential at infinity to be zero.

2.2 A regular hexagon of side 10 cm has a charge 5 μC at each of its vertices. Calculate the potential at the centre of the hexagon.

2.3 Two charges 2 μC and –2 μC are placed at points A and B 6 cm apart.

(a) Identify an equipotential surface of the system.

(b) What is the direction of the electric field at every point on this surface?

2.4 A spherical conductor of radius 12 cm has a charge of 1.6×10^{-7}C distributed uniformly on its surface. What is the electric field

(a) inside the sphere

(b) just outside the sphere

(c) at a point 18 cm from the centre of the sphere?

2.5 A parallel plate capacitor with air between the plates has a capacitance of 8 pF (1pF = 10^{-12} F). What will be the capacitance if the distance between the plates is reduced by half, and the space between them is filled with a substance of dielectric constant 6?

2.6 Three capacitors each of capacitance 9 pF are connected in series.

(a) What is the total capacitance of the combination?

(b) What is the potential difference across each capacitor if the combination is connected to a 120 V supply?

2.7 Three capacitors of capacitances 2 pF, 3 pF and 4 pF are connected in parallel.

(a) What is the total capacitance of the combination?

(b) Determine the charge on each capacitor if the combination is connected to a 100 V supply.

2.8 In a parallel plate capacitor with air between the plates, each plate has an area of 6×10^{-3} m^2 and the distance between the plates is 3 mm. Calculate the capacitance of the capacitor. If this capacitor is connected to a 100 V supply, what is the charge on each plate of the capacitor?

2.9 Explain what would happen if in the capacitor given in Exercise 2.8, a 3 mm thick mica sheet (of dielectric constant = 6) were inserted between the plates,

(a) while the voltage supply remained connected.

(b) after the supply was disconnected.

2.10 A 12pF capacitor is connected to a 50V battery. How much electrostatic energy is stored in the capacitor?

2.11 A 600pF capacitor is charged by a 200V supply. It is then disconnected from the supply and is connected to another uncharged 600 pF capacitor. How much electrostatic energy is lost in the process?

ADDITIONAL EXERCISES

2.12 A charge of 8 mC is located at the origin. Calculate the work done in taking a small charge of -2×10^{-9} C from a point P (0, 0, 3 cm) to a point Q (0, 4 cm, 0), via a point R (0, 6 cm, 9 cm).

2.13 A cube of side b has a charge q at each of its vertices. Determine the potential and electric field due to this charge array at the centre of the cube.

2.14 Two tiny spheres carrying charges 1.5 μC and 2.5 μC are located 30 cm apart. Find the potential and electric field:

(a) at the mid-point of the line joining the two charges, and

(b) at a point 10 cm from this midpoint in a plane normal to the line and passing through the mid-point.

2.15 A spherical conducting shell of inner radius r_1 and outer radius r_2 has a charge Q.

(a) A charge q is placed at the centre of the shell. What is the surface charge density on the inner and outer surfaces of the shell?

(b) Is the electric field inside a cavity (with no charge) zero, even if the shell is not spherical, but has any irregular shape? Explain.

2.16 (a) Show that the normal component of electrostatic field has a discontinuity from one side of a charged surface to another given by

$$(\mathbf{E}_2 - \mathbf{E}_1)\cdot\hat{\mathbf{n}} = \frac{\sigma}{\varepsilon_0}$$

where $\hat{\mathbf{n}}$ is a unit vector normal to the surface at a point and σ is the surface charge density at that point. (The direction of $\hat{\mathbf{n}}$ is from side 1 to side 2.) Hence show that just outside a conductor, the electric field is $\sigma\,\hat{\mathbf{n}}/\varepsilon_0$.

(b) Show that the tangential component of electrostatic field is continuous from one side of a charged surface to another. [Hint: For (a), use Gauss's law. For, (b) use the fact that work done by electrostatic field on a closed loop is zero.]

2.17 A long charged cylinder of linear charged density λ is surrounded by a hollow co-axial conducting cylinder. What is the electric field in the space between the two cylinders?

2.18 In a hydrogen atom, the electron and proton are bound at a distance of about 0.53 Å:

(a) Estimate the potential energy of the system in eV, taking the zero of the potential energy at infinite separation of the electron from proton.

(b) What is the minimum work required to free the electron, given that its kinetic energy in the orbit is half the magnitude of potential energy obtained in (a)?

(c) What are the answers to (a) and (b) above if the zero of potential energy is taken at 1.06 Å separation?

2.19 If one of the two electrons of a H_2 molecule is removed, we get a hydrogen molecular ion H_2^+. In the ground state of an H_2^+, the two protons are separated by roughly 1.5 Å, and the electron is roughly 1 Å from each proton. Determine the potential energy of the system. Specify your choice of the zero of potential energy.

2.20 Two charged conducting spheres of radii a and b are connected to each other by a wire. What is the ratio of electric fields at the surfaces of the two spheres? Use the result obtained to explain why charge density on the sharp and pointed ends of a conductor is higher than on its flatter portions.

2.21 Two charges $-q$ and $+q$ are located at points (0, 0, $-a$) and (0, 0, a), respectively.

(a) What is the electrostatic potential at the points (0, 0, z) and (x, y, 0) ?

(b) Obtain the dependence of potential on the distance r of a point from the origin when $r/a >> 1$.

(c) How much work is done in moving a small test charge from the point (5,0,0) to (–7,0,0) along the x-axis? Does the answer change if the path of the test charge between the same points is not along the x-axis?

2.22 Figure 2.34 shows a charge array known as an *electric quadrupole*. For a point on the axis of the quadrupole, obtain the dependence of potential on r for $r/a >> 1$, and contrast your results with that due to an electric dipole, and an electric monopole (i.e., a single charge).

FIGURE 2.34

2.23 An electrical technician requires a capacitance of 2 μF in a circuit across a potential difference of 1 kV. A large number of 1 μF capacitors are available to him each of which can withstand a potential difference of not more than 400 V. Suggest a possible arrangement that requires the minimum number of capacitors.

2.24 What is the area of the plates of a 2 F parallel plate capacitor, given that the separation between the plates is 0.5 cm? [You will realise from your answer why ordinary capacitors are in the range of μF or less. However, electrolytic capacitors do have a much larger capacitance (0.1 F) because of very minute separation between the conductors.]

2.25 Obtain the equivalent capacitance of the network in Fig. 2.35. For a 300 V supply, determine the charge and voltage across each capacitor.

FIGURE 2.35

2.26 The plates of a parallel plate capacitor have an area of 90 cm^2 each and are separated by 2.5 mm. The capacitor is charged by connecting it to a 400 V supply.

(a) How much electrostatic energy is stored by the capacitor?

(b) View this energy as stored in the electrostatic field between the plates, and obtain the energy per unit volume u. Hence arrive at a relation between u and the magnitude of electric field E between the plates.

2.27 A 4 μF capacitor is charged by a 200 V supply. It is then disconnected from the supply, and is connected to another uncharged 2 μF capacitor. How much electrostatic energy of the first capacitor is lost in the form of heat and electromagnetic radiation?

2.28 Show that the force on each plate of a parallel plate capacitor has a magnitude equal to (½) QE, where Q is the charge on the capacitor, and E is the magnitude of electric field between the plates. Explain the origin of the factor ½.

2.29 A spherical capacitor consists of two concentric spherical conductors, held in position by suitable insulating supports (Fig. 2.36). Show

FIGURE 2.36

that the capacitance of a spherical capacitor is given by

$$C = \frac{4\pi\varepsilon_0 r_1 r_2}{r_1 - r_2}$$

where r_1 and r_2 are the radii of outer and inner spheres, respectively.

2.30 A spherical capacitor has an inner sphere of radius 12 cm and an outer sphere of radius 13 cm. The outer sphere is earthed and the inner sphere is given a charge of 2.5 μC. The space between the concentric spheres is filled with a liquid of dielectric constant 32.

(a) Determine the capacitance of the capacitor.

(b) What is the potential of the inner sphere?

(c) Compare the capacitance of this capacitor with that of an isolated sphere of radius 12 cm. Explain why the latter is much smaller.

2.31 Answer carefully:

(a) Two large conducting spheres carrying charges Q_1 and Q_2 are brought close to each other. Is the magnitude of electrostatic force between them exactly given by $Q_1 Q_2/4\pi\varepsilon_0 r^2$, where r is the distance between their centres?

(b) If Coulomb's law involved $1/r^3$ dependence (instead of $1/r^2$), would Gauss's law be still true ?

(c) A small test charge is released at rest at a point in an electrostatic field configuration. Will it travel along the field line passing through that point?

(d) What is the work done by the field of a nucleus in a complete circular orbit of the electron? What if the orbit is elliptical?

(e) We know that electric field is discontinuous across the surface of a charged conductor. Is electric potential also discontinuous there?

(f) What meaning would you give to the capacitance of a single conductor?

(g) Guess a possible reason why water has a much greater dielectric constant (= 80) than say, mica (= 6).

2.32 A cylindrical capacitor has two co-axial cylinders of length 15 cm and radii 1.5 cm and 1.4 cm. The outer cylinder is earthed and the inner cylinder is given a charge of 3.5 μC. Determine the capacitance of the system and the potential of the inner cylinder. Neglect end effects (i.e., bending of field lines at the ends).

2.33 A parallel plate capacitor is to be designed with a voltage rating 1 kV, using a material of dielectric constant 3 and dielectric strength about 10^7 Vm^{-1}. (Dielectric strength is the maximum electric field a material can tolerate without breakdown, i.e., without starting to conduct electricity through partial ionisation.) For safety, we should like the field never to exceed, say 10% of the dielectric strength. What minimum area of the plates is required to have a capacitance of 50 pF?

2.34 Describe schematically the equipotential surfaces corresponding to

(a) a constant electric field in the z-direction,

(b) a field that uniformly increases in magnitude but remains in a constant (say, z) direction,

(c) a single positive charge at the origin, and

(d) a uniform grid consisting of long equally spaced parallel charged wires in a plane.

2.35 In a Van de Graaff type generator a spherical metal shell is to be a 15×10^6 V electrode. The dielectric strength of the gas surrounding the electrode is 5×10^7 Vm^{-1}. What is the minimum radius of the spherical shell required? (You will learn from this exercise why one cannot build an electrostatic generator using a very small shell which requires a small charge to acquire a high potential.)

2.36 A small sphere of radius r_1 and charge q_1 is enclosed by a spherical shell of radius r_2 and charge q_2. Show that if q_1 is positive, charge will necessarily flow from the sphere to the shell (when the two are connected by a wire) no matter what the charge q_2 on the shell is.

2.37 Answer the following:

(a) The top of the atmosphere is at about 400 kV with respect to the surface of the earth, corresponding to an electric field that decreases with altitude. Near the surface of the earth, the field is about 100 Vm^{-1}. Why then do we not get an electric shock as we step out of our house into the open? (Assume the house to be a steel cage so there is no field inside!)

(b) A man fixes outside his house one evening a two metre high insulating slab carrying on its top a large aluminium sheet of area $1m^2$. Will he get an electric shock if he touches the metal sheet next morning?

(c) The discharging current in the atmosphere due to the small conductivity of air is known to be 1800 A on an average over the globe. Why then does the atmosphere not discharge itself completely in due course and become electrically neutral? In other words, what keeps the atmosphere charged?

(d) What are the forms of energy into which the electrical energy of the atmosphere is dissipated during a lightning?
(Hint: The earth has an electric field of about 100 Vm^{-1} at its surface in the downward direction, corresponding to a surface charge density = -10^{-9} C m^{-2}. Due to the slight conductivity of the atmosphere up to about 50 km (beyond which it is good conductor), about + 1800 C is pumped every second into the earth as a whole. The earth, however, does not get discharged since thunderstorms and lightning occurring continually all over the globe pump an equal amount of negative charge on the earth.)

Chapter Three

CURRENT ELECTRICITY

3.1 Introduction

In Chapter 1, all charges whether free or bound, were considered to be at rest. Charges in motion constitute an electric current. Such currents occur naturally in many situations. Lightning is one such phenomenon in which charges flow from the clouds to the earth through the atmosphere, sometimes with disastrous results. The flow of charges in lightning is not steady, but in our everyday life we see many devices where charges flow in a steady manner, like water flowing smoothly in a river. A torch and a cell-driven clock are examples of such devices. In the present chapter, we shall study some of the basic laws concerning steady electric currents.

3.2 Electric Current

Imagine a small area held normal to the direction of flow of charges. Both the positive and the negative charges may flow forward and backward across the area. In a given time interval t, let q_+ be the net amount (*i.e.*, forward *minus* backward) of positive charge that flows in the forward direction across the area. Similarly, let q_- be the net amount of negative charge flowing across the area in the forward direction. The net amount of charge flowing across the area in the forward direction in the time interval t, then, is $q = q_+ - q_-$. This is proportional to t for steady current

and the quotient

$$I = \frac{q}{t} \tag{3.1}$$

is defined to be the *current* across the area in the forward direction. (If it turn out to be a negative number, it implies a current in the backward direction.)

Currents are not always steady and hence more generally, we define the current as follows. Let ΔQ be the net charge flowing across a cross-section of a conductor during the time interval Δt [i.e., between times t and $(t + \Delta t)$]. Then, the current at time t across the cross-section of the conductor is defined as the value of the ratio of ΔQ to Δt in the limit of Δt tending to zero,

$$I(t) \equiv \lim_{\Delta t \to 0} \frac{\Delta Q}{\Delta t} \tag{3.2}$$

In SI units, the unit of current is ampere. An ampere is defined through magnetic effects of currents that we will study in the following chapter. An ampere is typically the order of magnitude of currents in domestic appliances. An average lightning carries currents of the order of tens of thousands of amperes and at the other extreme, currents in our nerves are in microamperes.

3.3 Electric Currents in Conductors

An electric charge will experience a force if an electric field is applied. If it is free to move, it will thus move contributing to a current. In nature, free charged particles do exist like in upper strata of atmosphere called the *ionosphere*. However, in atoms and molecules, the negatively charged electrons and the positively charged nuclei are bound to each other and are thus not free to move. Bulk matter is made up of many molecules, a gram of water, for example, contains approximately 10^{22} molecules. These molecules are so closely packed that the electrons are no longer attached to individual nuclei. In some materials, the electrons will still be bound, i.e., they will not accelerate even if an electric field is applied. In other materials, notably metals, some of the electrons are practically free to move within the bulk material. These materials, generally called conductors, develop electric currents in them when an electric field is applied.

If we consider solid conductors, then of course the atoms are tightly bound to each other so that the current is carried by the negatively charged electrons. There are, however, other types of conductors like electrolytic solutions where positive and negative charges both can move. In our discussions, we will focus only on solid conductors so that the current is carried by the negatively charged electrons in the background of fixed positive ions.

Consider first the case when no electric field is present. The electrons will be moving due to thermal motion during which they collide with the fixed ions. An electron colliding with an ion emerges with the same speed as before the collision. However, the direction of its velocity after the collision is completely random. At a given time, there is no preferential direction for the velocities of the electrons. Thus on the average, the

number of electrons travelling in any direction will be equal to the number of electrons travelling in the opposite direction. So, there will be no net electric current.

Let us now see what happens to such a piece of conductor if an electric field is applied. To focus our thoughts, imagine the conductor in the shape of a cylinder of radius R (Fig. 3.1). Suppose we now take two thin circular discs of a dielectric of the same radius and put positive charge $+Q$ distributed over one disc and similarly $-Q$ at the other disc. We attach the two discs on the two flat surfaces of the cylinder. An electric field will be created and is directed from the positive towards the negative charge. The electrons will be accelerated due to this field towards $+Q$. They will thus move to neutralise the charges. The electrons, as long as they are moving, will constitute an electric current. Hence in the situation considered, there will be a current for a very short while and no current thereafter.

FIGURE 3.1 Charges $+Q$ and $-Q$ put at the ends of a metallic cylinder. The electrons will drift because of the electric field created to neutralise the charges. The current thus will stop after a while unless the charges $+Q$ and $-Q$ are continuously replenished.

We can also imagine a mechanism where the ends of the cylinder are supplied with fresh charges to make up for any charges neutralised by electrons moving inside the conductor. In that case, there will be a steady electric field in the body of the conductor. This will result in a continuous current rather than a current for a short period of time. Mechanisms, which maintain a steady electric field are cells or batteries that we shall study later in this chapter. In the next sections, we shall study the steady current that results from a steady electric field in conductors.

3.4 Ohm's Law

A basic law regarding flow of currents was discovered by G.S. Ohm in 1828, long before the physical mechanism responsible for flow of currents was discovered. Imagine a conductor through which a current I is flowing and let V be the potential difference between the ends of the conductor. Then Ohm's law states that

$$V \propto I$$

$$\text{or, } V = R\,I \tag{3.3}$$

where the constant of proportionality R is called the *resistance* of the conductor. The SI units of resistance is *ohm*, and is denoted by the symbol Ω. The resistance R not only depends on the material of the conductor but also on the dimensions of the conductor. The dependence of R on the dimensions of the conductor can easily be determined as follows.

FIGURE 3.2 Illustrating the relation $R = \rho l/A$ for a rectangular slab of length l and area of cross-section A.

Consider a conductor satisfying Eq. (3.3) to be in the form of a slab of length l and cross sectional area A [Fig. 3.2(a)]. Imagine placing two such identical slabs side by side [Fig. 3.2(b)], so that the length of the combination is $2l$. The current flowing through the combination is the same as that flowing through either of the slabs. If V is the potential difference across the ends of the first slab, then V is also the potential difference across the ends of the second slab since the second slab is

GEORG SIMON OHM (1787–1854)

Georg Simon Ohm (1787–1854) German physicist, professor at Munich. Ohm was led to his law by an analogy between the conduction of heat: the electric field is analogous to the temperature gradient, and the electric current is analogous to the heat flow.

identical to the first and the same current I flows through both. The potential difference across the ends of the combination is clearly sum of the potential difference across the two individual slabs and hence equals $2V$. The current through the combination is I and the resistance of the combination R_C is [from Eq. (3.3)],

$$R_C = \frac{2V}{I} = 2R \tag{3.4}$$

since $V/I = R$, the resistance of either of the slabs. Thus, doubling the length of a conductor doubles the resistance. In general, then resistance is proportional to length,

$$R \propto l \tag{3.5}$$

Next, imagine dividing the slab into two by cutting it lengthwise so that the slab can be considered as a combination of two identical slabs of length l, but each having a cross sectional area of $A/2$ [Fig. 3.2(c)].

For a given voltage V across the slab, if I is the current through the entire slab, then clearly the current flowing through each of the two half-slabs is $I/2$. Since the potential difference across the ends of the half-slabs is V, i.e., the same as across the full slab, the resistance of each of the half-slabs R_1 is

$$R_1 = \frac{V}{(I/2)} = 2\frac{V}{I} = 2R. \tag{3.6}$$

Thus, halving the area of the cross-section of a conductor doubles the resistance. In general, then the resistance R is inversely proportional to the cross-sectional area,

$$R \propto \frac{1}{A} \tag{3.7}$$

Combining Eqs. (3.5) and (3.7), we have

$$R \propto \frac{l}{A} \tag{3.8}$$

and hence for a given conductor

$$R = \rho \frac{l}{A} \tag{3.9}$$

where the constant of proportionality ρ depends on the material of the conductor but not on its dimensions. ρ is called *resistivity*.

Using the last equation, Ohm's law reads

$$V = I \cdot R = \frac{I\rho l}{A} \tag{3.10}$$

Current per unit area (taken normal to the current), I/A, is called *current density* and is denoted by j. The SI units of the current density are A/m^2. Further, if E is the magnitude of uniform electric field in the conductor whose length is l, then the potential difference V across its ends is El. Using these, the last equation reads

$$E\,l = j\,\rho\,l$$

or, $$E = j\,\rho \tag{3.11}$$

The above relation for *magnitudes* E and j can indeed be cast in a *vector* form. The current density, (which we have defined as the current through unit area *normal* to the current) is also directed along **E**, and is also a vector **j** ($\equiv j\,\mathbf{E}/E$). Thus, the last equation can be written as,

$$\mathbf{E} = \mathbf{j}\rho \tag{3.12}$$

or, $$\mathbf{j} = \sigma\,\mathbf{E} \tag{3.13}$$

where $\sigma \equiv 1/\rho$ is called the *conductivity*. Ohm's law is often stated in an equivalent form, Eq. (3.13) in addition to Eq.(3.3). In the next section, we will try to understand the origin of the Ohm's law as arising from the characteristics of the drift of electrons.

3.5 Drift of Electrons and the Origin of Resistivity

As remarked before, an electron will suffer collisions with the heavy fixed ions, but after collision, it will emerge with the same speed but in random directions. If we consider all the electrons, their average velocity will be zero since their directions are random. Thus, if there are N electrons and the velocity of the ith electron (i = 1, 2, 3, ... N) at a given time is $\mathbf{v}_i$, then

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{v}_i = 0 \tag{3.14}$$

Consider now the situation when an electric field is present. Electrons will be accelerated due to this field by

$$\mathbf{a} = \frac{-e\,\mathbf{E}}{m} \tag{3.15}$$

where $-e$ is the charge and m is the mass of an electron. Consider again the ith electron at a given time t. This electron would have had its last collision some time before t, and let t_i be the time elapsed after its last collision. If $\mathbf{v}_i$ was its velocity immediately after the last collision, then its velocity $\mathbf{V}_i$ at time t is

$$\mathbf{V}_i = \mathbf{v}_i - \frac{e\,\mathbf{E}}{m}\,t_i \tag{3.16}$$

since starting with its last collision it was accelerated (Fig. 3.3) with an acceleration given by Eq. (3.15) for a time interval t_i. The average velocity of the electrons at time t is the average of all the $\mathbf{V}_i$'s. The average of $\mathbf{v}_i$'s is zero [Eq. (3.14)] since immediately after any collision, the direction of the velocity of an electron is completely random. The collisions of the electrons do not occur at regular intervals but at random times. Let us denote by τ, the average time between successive collisions. Then at a given time, some of the electrons would have spent

FIGURE 3.3 A schematic picture of an electron moving from a point A to another point B through repeated collisions, and straight line travel between collisions (full lines). If an electric field is applied as shown, the electron ends up at point B′ (dotted lines). A slight drift in a direction opposite the electric field is visible.

time more than τ and some less than τ. In other words, the time t_i in Eq. (3.16) will be less than τ for some and more than τ for others as we go through the values of i = 1, 2 ..... N. The average value of t_i then is τ (known as *relaxation time*). Thus, averaging Eq. (3.16) over the N-electrons at any given time t gives us for the average velocity $\mathbf{v}_d$

$$\mathbf{v}_d \equiv (\mathbf{V}_i)_{average} = (\mathbf{v}_i)_{average} - \frac{e\mathbf{E}}{m}(t_i)_{average}$$

$$= 0 - \frac{e\mathbf{E}}{m}\tau = -\frac{e\mathbf{E}}{m}\tau \qquad (3.17)$$

This last result is surprising. It tells us that the electrons move with an average velocity which is independent of time, although electrons are accelerated. This is the phenomenon of drift and the velocity $\mathbf{v}_d$ in Eq. (3.17) is called the *drift velocity*.

FIGURE 3.4 Current in a metallic conductor. The magnitude of current density in a metal is the magnitude of charge contained in a cylinder of unit area and length v_d.

Because of the drift, there will be net transport of charges across any area perpendicular to $\mathbf{E}$. Consider a planar area A, located inside the conductor such that the normal to the area is parallel to $\mathbf{E}$ (Fig. 3.4). Then because of the drift, in an infinitesimal amount of time Δt, all electrons to the left of the area at distances upto $|\mathbf{v}_d|\Delta t$ would have crossed the area. If n is the number of free electrons per unit volume in the metal, then there are $n\,\Delta t\,|\mathbf{v}_d|A$ such electrons. Since each electron carries a charge $-e$, the total charge transported across this area A to the right in time Δt is $-ne\,A|\mathbf{v}_d|\Delta t$. $\mathbf{E}$ is directed towards the left and hence the total charge transported along $\mathbf{E}$ across the area is negative of this. The amount of charge crossing the area A in time Δt is by definition [Eq. (3.2)] $I\,\Delta t$, where I is the magnitude of the current. Hence,

$$I\,\Delta t = +n\,e\,A\,|\mathbf{v}_d|\,\Delta t \qquad (3.18)$$

Substituting the value of $|\mathbf{v}_d|$ from Eq. (3.17)

$$I\,\Delta t = \frac{e^2 A}{m}\tau\, n\,\Delta t\,|\mathbf{E}| \qquad (3.19)$$

By definition I is related to the magnitude $|\mathbf{j}|$ of the current density by

$$I = |\mathbf{j}|A \qquad (3.20)$$

Hence, from Eqs.(3.19) and (3.20),

$$|\mathbf{j}| = \frac{ne^2}{m}\tau|\mathbf{E}| \qquad (3.21)$$

The vector $\mathbf{j}$ is parallel to $\mathbf{E}$ and hence we can write Eq. (3.21) in the vector form

$$\mathbf{j} = \frac{ne^2}{m}\tau\mathbf{E} \qquad (3.22)$$

Comparison with Eq. (3.13) shows that Eq. (3.22) is exactly the Ohm's law, if we identify the conductivity σ as

$$\sigma = \frac{ne^2}{m}\tau \qquad (3.23)$$

We thus see that a very simple picture of electrical conduction reproduces Ohm's law. We have, of course, made assumptions that τ and n are constants, independent of E. We shall, in the next section, discuss the limitations of Ohm's law.

Example 3.1 (a) Estimate the average drift speed of conduction electrons in a copper wire of cross-sectional area 1.0×10^{-7} m^2 carrying a current of 1.5 A. Assume that each copper atom contributes roughly one conduction electron. The density of copper is 9.0×10^3 kg/m^3, and its atomic mass is 63.5 u. (b) Compare the drift speed obtained above with, (i) thermal speeds of copper atoms at ordinary temperatures, (ii) speed of propagation of electric field along the conductor which causes the drift motion.

Solution

(a) The direction of drift velocity of conduction electrons is opposite to the electric field direction, i.e., electrons drift in the direction of increasing potential. The drift speed v_d is given by Eq. (3.18)

$v_d = (I/neA)$

Now, $e = 1.6 \times 10^{-19}$ C, $A = 1.0 \times 10^{-7}$m^2, $I = 1.5$ A. The density of conduction electrons, n is equal to the number of atoms per cubic metre (assuming one conduction electron per Cu atom as is reasonable from its valence electron count of one). A cubic metre of copper has a mass of 9.0×10^3 kg. Since 6.0×10^{23} copper atoms have a mass of 63.5 g,

$$n = \frac{6.0 \cdot 10^{23}}{63.5} \cdot 9.0 \cdot 10^6$$

$$= 8.5 \times 10^{28}\ \text{m}^{-3}$$

which gives,

$$v_d = \frac{1.5}{8.5 \cdot 10^{28} \cdot 1.6 \quad 10^{-19} \quad 1.0 \quad 10^{-7}}$$

$$= 1.1 \times 10^{-3}\ \text{m s}^{-1} = 1.1\ \text{mm s}^{-1}$$

(b) (i) At a temperature T, the thermal speed* of a copper atom of mass M is obtained from $[\langle(1/2)\ Mv^2\rangle = (3/2)\ k_B T]$ and is thus typically of the order of $\sqrt{k_B T/M}$, where k_B is the Boltzmann constant. For copper at 300 K, this is about 2×10^2 m/s. This figure indicates the random vibrational speeds of copper atoms in a conductor. Note that the drift speed of electrons is much smaller, about 10^{-5} times the typical thermal speed at ordinary temperatures.

(ii) An electric field travelling along the conductor has a speed of an electromagnetic wave, namely equal to 3.0×10^8 m s^{-1} (You will learn about this in Chapter 8). The drift speed is, in comparison, extremely small; smaller by a factor of 10^{-11}.

EXAMPLE 3.1

* See Eq. (13.23) of Chapter 13 from Class XI book.

Example 3.2

(a) In Example 3.1, the electron drift speed is estimated to be only a few mm s^{-1} for currents in the range of a few amperes? How then is current established almost the instant a circuit is closed?

(b) The electron drift arises due to the force experienced by electrons in the electric field inside the conductor. But force should cause acceleration. Why then do the electrons acquire a steady average drift speed?

(c) If the electron drift speed is so small, and the electron's charge is small, how can we still obtain large amounts of current in a conductor?

(d) When electrons drift in a metal from lower to higher potential, does it mean that all the 'free' electrons of the metal are moving in the same direction?

(e) Are the paths of electrons straight lines between successive collisions (with the positive ions of the metal) in the (i) absence of electric field, (ii) presence of electric field?

Solution

(a) Electric field is established throughout the circuit, almost instantly (with the speed of light) causing at every point a *local electron drift.* Establishment of a current does not have to wait for electrons from one end of the conductor travelling to the other end. However, it does take a little while for the current to reach its steady value.

(b) Each 'free' electron does accelerate, increasing its drift speed until it collides with a positive ion of the metal. It loses its drift speed after collision but starts to accelerate and increases its drift speed again only to suffer a collision again and so on. On the average, therefore, electrons acquire only a drift speed.

(c) Simple, because the electron number density is enormous, $\sim 10^{29}$ m^{-3}.

(d) By no means. The drift velocity is superposed over the large random velocities of electrons.

(e) In the absence of electric field, the paths are straight lines; in the presence of electric field, the paths are, in general, curved.

Example 3.2

3.5.1 Mobility

As we have seen, conductivity arises from mobile charge carriers. In metals, these mobile charge carriers are electrons; in an ionised gas, they are electrons and positive charged ions; in an electrolyte, these can be both positive and negative ions.

An important quantity is the *mobility* $\propto$ defined as the magnitude of the drift velocity per unit electric field:

$$\propto = \frac{|\mathbf{v}_d|}{E} \tag{3.24}$$

The SI unit of mobility is m^2/Vs and is 10^4 of the mobility in practical units (cm^2/Vs). Mobility is positive. From Eq. (3.17), we have

$$v_d = \frac{e\,\tau\,E}{m}$$

Hence,

$$\propto = \frac{v_d}{E} = \frac{e\tau}{m} \qquad (3.25)$$

where τ is the average collision time for electrons.

FIGURE 3.5 The dashed line represents the linear Ohm's law. The solid line is the voltage V versus current I for a good conductor.

3.6 Limitations of Ohm's Law

Although Ohm's law has been found valid over a large class of materials, there do exist materials and devices used in electric circuits where the proportionality of V and I does not hold. The deviations broadly are one or more of the following types:

(a) V ceases to be proportional to I (Fig. 3.5).

(b) The relation between V and I depends on the sign of V. In other words, if I is the current for a certain V, then reversing the direction of V keeping its magnitude fixed, does not produce a current of the same magnitude as I in the opposite direction (Fig. 3.6). This happens, for example, in a diode which we will study in Chapter 14.

FIGURE 3.6 Characteristic curve of a diode. Note the different scales for negative and positive values of the voltage and current.

FIGURE 3.7 Variation of current versus voltage for GaAs.

(c) The relation between V and I is not unique, i.e., there is more than one value of V for the same current I (Fig. 3.7). A material exhibiting such behaviour is GaAs.

Materials and devices not obeying Ohm's law in the form of Eq. (3.3) are actually widely used in electronic circuits. In this and a few subsequent chapters, however, we will study the electrical currents in materials that obey Ohm's law.

3.7 Resistivity of Various Materials

The resistivities of various common materials are listed in Table 3.1. The materials are classified as conductors, semiconductors and insulators

depending on their resistivities, in an increasing order of their values. Metals have low resistivities in the range of 10^{-8} Ωm to 10^{-6} Ωm. At the other end are insulators like ceramic, rubber and plastics having resistivities 10^{18} times greater than metals or more. In between the two are the semiconductors. These, however, have resistivities characteristically decreasing with a rise in temperature. The resistivities of semiconductors are also affected by presence of small amount of impurities. This last feature is exploited in use of semiconductors for electronic devices.

TABLE 3.1 RESISTIVITIES OF SOME MATERIALS

Material	Resistivity, ρ (Ω m) at 0°C	Temperature coefficient of resistivity, α (°C)$^{-1}$ $\frac{1}{\rho}\left(\frac{d\rho}{dT}\right)$ at 0°C
Conductors		
Silver	1.6×10^{-8}	0.0041
Copper	1.7×10^{-8}	0.0068
Aluminium	2.7×10^{-8}	0.0043
Tungsten	5.6×10^{-8}	0.0045
Iron	10×10^{-8}	0.0065
Platinum	11×10^{-8}	0.0039
Mercury	98×10^{-8}	0.0009
Nichrome (alloy of Ni, Fe, Cr)	$\sim 100 \times 10^{-8}$	0.0004
Manganin (alloy)	48×10^{-8}	0.002×10^{-3}
Semiconductors		
Carbon (graphite)	3.5×10^{-5}	– 0.0005
Germanium	0.46	– 0.05
Silicon	2300	– 0.07
Insulators		
Pure Water	2.5×10^{5}	
Glass	$10^{10} - 10^{14}$	
Hard Rubber	$10^{13} - 10^{16}$	
NaCl	$\sim 10^{14}$	
Fused Quartz	$\sim 10^{16}$	

Commercially produced resistors for domestic use or in laboratories are of two major types: *wire bound resistors* and *carbon resistors*. Wire bound resistors are made by winding the wires of an alloy, viz., manganin, constantan, nichrome or similar ones. The choice of these materials is dictated mostly by the fact that their resistivities are relatively insensitive to temperature. These resistances are typically in the range of a fraction of an ohm to a few hundred ohms.

Resistors in the higher range are made mostly from carbon. Carbon resistors are compact, inexpensive and thus find extensive use in electronic circuits. Carbon resistors are small in size and hence their values are given using a colour code.

TABLE 3.2 RESISTOR COLOUR CODES

Colour	Number	Multiplier	Tolerance (%)
Black	0	1	
Brown	1	10^1	
Red	2	10^2	
Orange	3	10^3	
Yellow	4	10^4	
Green	5	10^5	
Blue	6	10^6	
Violet	7	10^7	
Gray	8	10^8	
White	9	10^9	
Gold		10^{-1}	5
Silver		10^{-2}	10
No colour			20

The resistors have a set of co-axial coloured rings on them whose significance are listed in Table 3.2. The first two *bands* from the end indicate the first two significant figures of the resistance in ohms. The third band indicates the decimal multiplier (as listed in Table 3.2). The last band stands for tolerance or possible variation in percentage about the indicated values. Sometimes, this last band is absent and that indicates a tolerance of 20% (Fig. 3.8). For example, if the four colours are orange, blue, yellow and gold, the resistance value is $36 \times 10^4\ \Omega$, with a tolerence value of 5%.

FIGURE 3.8 Colour coded resistors (a) $(22 \times 10^2\ \Omega) \pm 10\%$, (b) $(47 \times 10\ \Omega) \pm 5\%$.

3.8 TEMPERATURE DEPENDENCE OF RESISTIVITY

The resistivity of a material is found to be dependent on the temperature. Different materials do not exhibit the same dependence on temperatures. Over a limited range of temperatures, that is not too large, the resistivity of a metallic conductor is approximately given by,

$$\rho_T = \rho_0 [1 + \alpha (T - T_0)] \qquad (3.26)$$

where ρ_T is the resistivity at a temperature T and ρ_0 is the same at a reference temperature T_0. α is called the *temperature co-efficient of resistivity*, and from Eq. (3.26), the dimension of α is (Temperature)$^{-1}$.

For metals, α is positive and values of α for some metals at $T_0 = 0°C$ are listed in Table 3.1.

The relation of Eq. (3.26) implies that a graph of ρ_T plotted against T would be a straight line. At temperatures much lower than 0°C, the graph, however, deviates considerably from a straight line (Fig. 3.9).

Equation (3.26) thus, can be used approximately over a limited range of T around any reference temperature T_0, where the graph can be approximated as a straight line.

FIGURE 3.9 Resistivity ρ_T of copper as a function of temperature T.

FIGURE 3.10 Resistivity ρ_T of nichrome as a function of absolute temperature T.

FIGURE 3.11 Temperature dependence of resistivity for a typical semiconductor.

Some materials like Nichrome (which is an alloy of nickel, iron and chromium) exhibit a very weak dependence of resistivity with temperature (Fig. 3.10). Manganin and constantan have similar properties. These materials are thus widely used in wire bound standard resistors since their resistance values would change very little with temperatures.

Unlike metals, the resistivities of semiconductors decrease with increasing temperatures. A typical dependence is shown in Fig. 3.11.

We can qualitatively understand the temperature dependence of resistivity, in the light of our derivation of Eq. (3.23). From this equation, resistivity of a material is given by

$$\rho = \frac{1}{\sigma} = \frac{m}{n e^2 \tau} \tag{3.27}$$

ρ thus depends inversely both on the number n of free electrons per unit volume and on the average time τ between collisions. As we increase temperature, average speed of the electrons, which act as the carriers of current, increases resulting in more frequent collisions. The average time of collisions τ, thus decreases with temperature.

In a metal, n is not dependent on temperature to any appreciable extent and thus the decrease in the value of τ with rise in temperature causes ρ to increase as we have observed.

For insulators and semiconductors, however, n increases with temperature. This increase more than compensates any decrease in τ in Eq.(3.23) so that for such materials, ρ decreases with temperature.

Example 3.3 An electric toaster uses nichrome for its heating element. When a negligibly small current passes through it, its resistance at room temperature (27.0 °C) is found to be 75.3 Ω. When the toaster is connected to a 230 V supply, the current settles, after a few seconds, to a steady value of 2.68 A. What is the steady temperature of the nichrome element? The temperature coefficient of resistance of nichrome averaged over the temperature range involved, is 1.70×10^{-4} °C^{-1}.

Solution When the current through the element is very small, heating effects can be ignored and the temperature T_1 of the element is the same as room temperature. When the toaster is connected to the supply, its initial current will be slightly higher than its steady value of 2.68 A. But due to heating effect of the current, the temperature will rise. This will cause an increase in resistance and a slight decrease in current. In a few seconds, a steady state will be reached when temperature will rise no further, and both the resistance of the element and the current drawn will achieve steady values. The resistance R_2 at the steady temperature T_2 is

$$R_2 = \frac{230\,\text{V}}{2.68\,\text{A}} = 85.8\ \Omega$$

Using the relation

$$R_2 = R_1\,[1 + \alpha\,(T_2 - T_1)]$$

with $\alpha = 1.70 \times 10^{-4}$ °C^{-1}, we get

$$T_2 - T_1 = \frac{(85.8 - 75.3)}{(75.3)\cdot\ 1.70\ \ 10^{-4}} = 820\ °\text{C}$$

that is, $T_2 = (820 + 27.0)$ °C = 847 °C

Thus, the steady temperature of the heating element (when heating effect due to the current equals heat loss to the surroundings) is 847 °C.

EXAMPLE 3.3

Example 3.4 The resistance of the platinum wire of a platinum resistance thermometer at the ice point is 5 Ω and at steam point is 5.23 Ω. When the thermometer is inserted in a hot bath, the resistance of the platinum wire is 5.795 Ω. Calculate the temperature of the bath.

Solution $R_0 = 5\ \Omega$, $R_{100} = 5.23\ \Omega$ and $R_t = 5.795\ \Omega$

Now, $$t = \frac{R_t - R_0}{R_{100} - R_0}\cdot 100, \qquad R_t = R_0\,(1 + \alpha t)$$

$$= \frac{5.795 - 5}{5.23 - 5}\cdot 100$$

$$= \frac{0.795}{0.23}\cdot 100 = 345.65\ °\text{C}$$

EXAMPLE 3.4

3.9 Electrical Energy, Power

Consider a conductor with end points A and B, in which a current I is flowing from A to B. The electric potential at A and B are denoted by V(A)

and V(B) respectively. Since current is flowing from A to B, V(A) > V(B) and the potential difference across AB is $V = V(A) - V(B) > 0$.

In a time interval Δt, an amount of charge $\Delta Q = I\,\Delta t$ travels from A to B. The potential energy of the charge at A, by definition, was $Q\,V$(A) and similarly at B, it is $Q\,V$(B). Thus, change in its potential energy ΔU_{pot} is

$$\Delta U_{pot} = \text{Final potential energy} - \text{Initial potential energy}$$
$$= \Delta Q[(V(B) - V(A)] = -\Delta Q\,V$$
$$= -I\,V\Delta t < 0 \qquad (3.28)$$

If charges moved without collisions through the conductor, their kinetic energy would also change so that the total energy is unchanged. Conservation of total energy would then imply that,

$$\Delta K = -\Delta U_{pot} \qquad (3.29)$$

that is,

$$\Delta K = I\,V\Delta t > 0 \qquad (3.30)$$

Thus, in case charges were moving freely through the conductor under the action of electric field, their kinetic energy would increase as they move. We have, however, seen earlier that on the average, charge carriers do not move with acceleration but with a steady drift velocity. This is because of the collisions with ions and atoms during transit. During collisions, the energy gained by the charges thus is shared with the atoms. The atoms vibrate more vigorously, i.e., the conductor heats up. Thus, in an actual conductor, an amount of energy dissipated as heat in the conductor during the time interval Δt is,

$$\Delta W = I\,V\Delta t \qquad (3.31)$$

The energy dissipated per unit time is the power dissipated $P = \Delta W/\Delta t$ and we have,

$$P = I\,V \qquad (3.32)$$

Using Ohm's law $V = IR$, we get

$$P = I^2 R = V^2/R \qquad (3.33)$$

as the power loss ("ohmic loss") in a conductor of resistance R carrying a current I. It is this power which heats up, for example, the coil of an electric bulb to incandescence, radiating out heat and light.

FIGURE 3.12 Heat is produced in the resistor R which is connected across the terminals of a cell. The energy dissipated in the resistor R comes from the chemical energy of the electrolyte.

Where does the power come from? As we have reasoned before, we need an external source to keep a steady current through the conductor. It is clearly this source which must supply this power. In the simple circuit shown with a cell (Fig.3.12), it is the chemical energy of the cell which supplies this power for as long as it can.

The expressions for power, Eqs. (3.32) and (3.33), show the dependence of the power dissipated in a resistor R on the current through it and the voltage across it.

Equation (3.33) has an important application to power transmission. Electrical power is transmitted from power stations to homes and factories, which

may be hundreds of miles away, via transmission cables. One obviously wants to minimise the power loss in the transmission cables connecting the power stations to homes and factories. We shall see now how this can be achieved. Consider a device R, to which a power P is to be delivered via transmission cables having a resistance R_c to be dissipated by it finally. If V is the voltage across R and I the current through it, then

$$P = V I \tag{3.34}$$

The connecting wires from the power station to the device has a finite resistance R_c. The power dissipated in the connecting wires, which is wasted is P_c with

$$P_c = I^2 R_c$$

$$= \frac{P^2 R_c}{V^2} \tag{3.35}$$

from Eq. (3.32). Thus, to drive a device of power P, the power wasted in the connecting wires is inversely proportional to V^2. The transmission cables from power stations are hundreds of miles long and their resistance R_c is considerable. To reduce P_c, these wires carry current at enormous values of V and this is the reason for the high voltage danger signs on transmission lines — a common sight as we move away from populated areas. Using electricity at such voltages is not safe and hence at the other end, a device called a transformer lowers the voltage to a value suitable for use.

3.10 Combination of Resistors – Series and Parallel

The current through a single resistor R across which there is a potential difference V is given by Ohm's law $I = V/R$. Resistors are sometimes joined together and there are simple rules for calculation of equivalent resistance of such combination.

FIGURE 3.13 A series combination of two resistor R_1 and R_2.

Two resistors are said to be in *series* if only one of their end points is joined (Fig. 3.13). If a third resistor is joined with the series combination of the two (Fig. 3.14), then all three are said to be in series. Clearly, we can extend this definition to series combination of any number of resistors.

FIGURE 3.14 A series combination of three resistors R_1, R_2, R_3.

Two or more resistors are said to be in *parallel* if one end of all the resistors is joined together and similarly the other ends joined together (Fig. 3.15).

FIGURE 3.15 Two resistors R_1 and R_2 connected in parallel.

Consider two resistors R_1 and R_2 in series. The charge which leaves R_1 must be entering R_2. Since current measures the rate of flow of charge, this means that the same current I flows through R_1 and R_2. By Ohm's law:

Potential difference across $R_1 = V_1 = I R_1$, and

Potential difference across $R_2 = V_2 = I R_2$.

The potential difference V across the combination is V_1+V_2. Hence,

$$V = V_1 + V_2 = I(R_1 + R_2) \tag{3.36}$$

This is as if the combination had an equivalent resistance R_{eq}, which by Ohm's law is

$$R_{eq} \equiv \frac{V}{I} = (R_1 + R_2) \tag{3.37}$$

If we had three resistors connected in series, then similarly

$$V = I R_1 + I R_2 + I R_3 = I(R_1 + R_2 + R_3). \tag{3.38}$$

This obviously can be extended to a series combination of any number n of resistors R_1, R_2 ....., R_n. The equivalent resistance R_{eq} is

$$R_{eq} = R_1 + R_2 + \ldots + R_n \tag{3.39}$$

Consider now the parallel combination of two resistors (Fig. 3.15). The charge that flows in at A from the left flows out partly through R_1 and partly through R_2. The currents I, I_1, I_2 shown in the figure are the rates of flow of charge at the points indicated. Hence,

$$I = I_1 + I_2 \tag{3.40}$$

The potential difference between A and B is given by the Ohm's law applied to R_1

$$V = I_1 R_1 \tag{3.41}$$

Also, Ohm's law applied to R_2 gives

$$V = I_2 R_2 \tag{3.42}$$

$$\therefore I = I_1 + I_2 = \frac{V}{R_1} + \frac{V}{R_2} = V\left(\frac{1}{R_1} + \frac{1}{R_2}\right) \tag{3.43}$$

If the combination was replaced by an equivalent resistance R_{eq}, we would have, by Ohm's law

$$I = \frac{V}{R_{eq}} \tag{3.44}$$

Hence,

$$\frac{1}{R_{eq}} = \frac{1}{R_1} + \frac{1}{R_2} \tag{3.45}$$

We can easily see how this extends to three resistors in parallel (Fig. 3.16).

FIGURE 3.16 Parallel combination of three resistors R_1, R_2 and R_3.

Exactly as before

$$I = I_1 + I_2 + I_3 \tag{3.46}$$

and applying Ohm's law to R_1, R_2 and R_3 we get,

$$V = I_1 R_1,\ V = I_2 R_2,\ V = I_3 R_3 \tag{3.47}$$

So that

$$I = I_1 + I_2 + I_3 = V\left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}\right) \tag{3.48}$$

An equivalent resistance R_{eq} that replaces the combination, would be such that

$$I = \frac{V}{R_{eq}} \tag{3.49}$$

and hence

$$\frac{1}{R_{eq}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \tag{3.50}$$

We can reason similarly for any number of resistors in parallel. The equivalent resistance of n resistors $R_1, R_2 \ldots, R_n$ is

$$\frac{1}{R_{eq}} = \frac{1}{R_1} + \frac{1}{R_2} + \ldots + \frac{1}{R_n} \tag{3.51}$$

These formulae for equivalent resistances can be used to find out currents and voltages in more complicated circuits. Consider for example, the circuit in Fig. (3.17), where there are three resistors R_1, R_2 and R_3. R_2 and R_3 are in parallel and hence we can replace them by an equivalent R_{eq}^{23} between point B and C with

$$\frac{1}{R_{eq}^{23}} = \frac{1}{R_2} + \frac{1}{R_3}$$

or, $$R_{eq}^{23} = \frac{R_2 R_3}{R_2 + R_3} \tag{3.52}$$

The circuit now has R_1 and R_{eq}^{23} in series and hence their combination can be replaced by an equivalent resistance R_{eq}^{123} with

$$R_{eq}^{123} = R_{eq}^{23} + R_1 \tag{3.53}$$

If the voltage between A and C is V, the current I is given by

$$I = \frac{V}{R_{eq}^{123}} = \frac{V}{R_1 + \left[R_2 R_3 / (R_2 + R_3)\right]}$$

$$= \frac{V(R_2 + R_3)}{R_1 R_2 + R_1 R_3 + R_2 R_3} \tag{3.54}$$

FIGURE 3.17 A combination of three resistors R_1, R_2 and R_3. R_2, R_3 are in parallel with an equivalent resistance R_{eq}^{23}. R_1 and R_{eq}^{23} are in series with an equivalent resistance R_{eq}^{123}.

3.11 Cells, emf, Internal Resistance

We have already mentioned that a simple device to maintain a steady current in an electric circuit is the electrolytic cell. Basically a cell has two electrodes, called the positive (P) and the negative (N), as shown in Fig. 3.18. They are immersed in an electrolytic solution. Dipped in the solution, the electrodes exchange charges with the electrolyte. The positive electrode has a potential difference V_+ ($V_+ > 0$) between itself and the electrolyte solution immediately adjacent to it marked A in the figure. Similarly, the negative electrode develops a negative potential $-(V_-)$ ($V_- \geq 0$) relative to the electrolyte adjacent to it, marked as *B* in the figure. When there is no current, the electrolyte has the same potential throughout, so that the potential difference between P and N is $V_+ - (-V_-) = V_+ + V_-$. This difference is called the *electromotive force* (emf) of the cell and is denoted by ε. Thus

$$\varepsilon = V_+ + V_- > 0 \qquad (3.55)$$

FIGURE 3.18 (a) Sketch of an electrolyte cell with positive terminal P and negative terminal N. The gap between the electrodes is exaggerated for clarity. A and B are points in the electrolyte typically close to P and N. (b) the symbol for a cell, + referring to P and – referring to the N electrode. Electrical connections to the cell are made at P and N.

Note that ε is, actually, a potential difference and *not a force*. The name emf, however, is used because of historical reasons, and was given at a time when the phenomenon was not understood properly.

To understand the significance of ε, consider a resistor R connected across the cell (Fig. 3.18). A current I flows across R from C to D. As explained before, a steady current is maintained because current flows from N to P through the electrolyte. Clearly, across the electrolyte the same current flows through the electrolyte but from N to P, whereas through R, it flows from P to N.

The electrolyte through which a current flows has a finite resistance r, called the *internal resistance*. Consider first the situation when R is infinite so that $I = V/R = 0$, where V is the potential difference between P and N. Now,

V = Potential difference between P and A
+ Potential difference between A and B
+ Potential difference between B and N

$$= \varepsilon \qquad (3.56)$$

Thus, emf ε is the potential difference between the positive and negative electrodes in an open circuit, i.e., when no current is flowing through the cell.

If however R is finite, I is not zero. In that case the potential difference between P and N is

$$V = V_+ + V_- - I\,r$$

$$= \varepsilon - I\,r \qquad (3.57)$$

Note the negative sign in the expression ($I\,r$) for the potential difference between A and B. This is because the current I flows from B to A in the electrolyte.

In practical calculations, internal resistances of cells in the circuit may be neglected when the current I is such that $\varepsilon >> I\,r$. The actual values of the internal resistances of cells vary from cell to cell. The internal resistance of dry cells, however, is much higher than the common electrolytic cells.

We also observe that since V is the potential difference across R, we have from Ohm's law

$$V = I\ R \tag{3.58}$$

Combining Eqs. (3.57) and (3.58), we get

$$I\ R\ = \varepsilon - I\ r$$

$$\text{Or, } I = \frac{\varepsilon}{R+r} \tag{3.59}$$

The maximum current that can be drawn from a cell is for $R = 0$ and it is $I_{max} = \varepsilon/r$. However, in most cells the maximum allowed current is much lower than this to prevent permanent damage to the cell.

Charges in clouds

In olden days lightning was considered as an atmospheric flash of supernatural origin. It was believed to be the great weapon of Gods. But today the phenomenon of lightning can be explained scientifically by elementary principles of physics.

Atmospheric electricity arises due to the separation of electric charges. In the ionosphere and magnetosphere strong electric current is generated from the solar-terrestrial interaction. In the lower atmosphere the current is weaker and is maintained by thunderstorm.

There are ice particles in the clouds, which grow, collide, fracture and break apart. The smaller particles acquire positive charge and the larger ones negative charge. These charged particles get separated by updrafts in the clouds and gravity. The upper portion of the cloud becomes positively charged and the middle negatively charged, leading to dipole structure. Sometimes a very weak positive charge is found near the base of the cloud. The ground is positively charged at the time of thunderstorm development. Also cosmic and radioactive radiations ionise air into positive and negative ions and air becomes (weakly) electrically conductive. The separation of charges produce tremendous amount of electrical potential within the cloud as well as between the cloud and ground. This can amount to millions of volts and eventually the electrical resistance in the air breaks down and lightning flash begins and thousands of amperes of current flows. The electric field is of the order of 10^5 V/m. A lightning flash is composed of a series of strokes with an average of about four and the duration of each flash is about 30 seconds. The average peak power per stroke is about 10^{12} watts.

During fair weather also there is charge in the atmosphere. The fair weather electric field arises due to the existence of a surface charge density at ground and an atmospheric conductivity as well as due to the flow of current from the ionosphere to the earth's surface, which is of the order of picoampere / square metre. The surface charge density at ground is negative; the electric field is directed downward. Over land the average electric field is about 120 V/m, which corresponds to a surface charge density of -1.2×10^{-9} C/m^2. Over the entire earth's surface, the total negative charge amount to about 600 kC. An equal positive charge exists in the atmosphere. This electric field is not noticeable in daily life. The reason why it is not noticed is that virtually everything, including our bodies, is conductor compared to air.

Example 3.5 A network of resistors is connected to a 16 V battery with internal resistance of 1Ω, as shown in Fig. 3.19: (a) Compute the equivalent resistance of the network. (b) Obtain the current in each resistor. (c) Obtain the voltage drops V_{AB}, V_{BC} and V_{CD}.

FIGURE 3.19

Solution

(a) The network is a simple series and parallel combination of resistors. First the two 4Ω resistors in parallel are equivalent to a resistor = [(4 × 4)/(4 + 4)] Ω = 2 Ω.

In the same way, the 12 Ω and 6 Ω resistors in parallel are equivalent to a resistor of

[(12 × 6)/(12 + 6)] Ω = 4 Ω.

The equivalent resistance R of the network is obtained by combining these resistors (2 Ω and 4 Ω) with 1 Ω in series, that is,

R = 2 Ω + 4 Ω + 1 Ω = 7 Ω.

(b) The total current I in the circuit is

$$I = \frac{\varepsilon}{R+r} = \frac{16\,V}{(7+1)\,\Omega} = 2\text{ A}$$

Consider the resistors between A and B. If I_1 is the current in one of the 4 Ω resistors and I_2 the current in the other,

$I_1 \times 4 = I_2 \times 4$

that is, $I_1 = I_2$, which is otherwise obvious from the symmetry of the two arms. But $I_1 + I_2 = I = 2$ A. Thus,

$I_1 = I_2 = 1$ A

that is, current in each 4 Ω resistor is 1 A. Current in 1 Ω resistor between B and C would be 2 A.

Now, consider the resistances between C and D. If I_3 is the current in the 12 Ω resistor, and I_4 in the 6 Ω resistor,

$I_3 \times 12 = I_4 \times 6$, i.e., $I_4 = 2I_3$

But, $I_3 + I_4 = I = 2$ A

Thus, $I_3 = \left(\frac{2}{3}\right)$ A, $I_4 = \left(\frac{4}{3}\right)$ A

that is, the current in the 12 Ω resistor is (2/3) A, while the current in the 6 Ω resistor is (4/3) A.

(c) The voltage drop across AB is

$V_{AB} = I_1 \times 4 = 1\text{ A} \times 4\ \Omega = 4$ V,

This can also be obtained by multiplying the total current between A and B by the equivalent resistance between A and B, that is,

EXAMPLE 3.5

V_{AB} = 2 A × 2 Ω = 4 V

The voltage drop across BC is

V_{BC} = 2 A × 1 Ω = 2 V

Finally, the voltage drop across CD is

$$V_{CD} = 12\ \Omega \times I_3 = 12\ \Omega \times \left(\frac{2}{3}\right) \text{ A} = 8 \text{ V}.$$

This can alternately be obtained by multiplying total current between C and D by the equivalent resistance between C and D, that is,

V_{CD} = 2 A × 4 Ω = 8 V

Note that the total voltage drop across AD is 4 V + 2 V + 8 V = 14 V. Thus, the terminal voltage of the battery is 14 V, while its emf is 16 V. The loss of the voltage (= 2 V) is accounted for by the internal resistance 1 Ω of the battery [2 A × 1 Ω = 2 V].

EXAMPLE 3.5

3.12 Cells in Series and in Parallel

Like resistors, cells can be combined together in an electric circuit. And like resistors, one can, for calculating currents and voltages in a circuit, replace a combination of cells by an equivalent cell.

FIGURE 3.20 Two cells of emf's ε_1 and ε_2 in the series. r_1, r_2 are their internal resistances. For connections across A and C, the combination can be considered as one cell of emf ε_{eq} and an internal resistance r_{eq}.

Consider first two cells in series (Fig. 3.20), where one terminal of the two cells is joined together leaving the other terminal in either cell free. ε_1, ε_2 are the emf's of the two cells and r_1, r_2 their internal resistances, respectively.

Let V(A), V(B), V(C) be the potentials at points A, B and C shown in Fig. 3.20. Then V(A) – V(B) is the potential difference between the positive and negative terminals of the first cell. We have already calculated it in Eq. (3.57) and hence,

$$V_{AB} \equiv V(A) - V(B) = \varepsilon_1 - I r_1 \tag{3.60}$$

Similarly,

$$V_{BC} \equiv V(B) - V(C) = \varepsilon_2 - I r_2 \tag{3.61}$$

Hence, the potential difference between the terminals A and C of the combination is

$$V_{AC} \equiv V(A) - V(C) = [V(A) - V(B)] + [V(B) - V(C)]$$

$$= (\varepsilon_1 + \varepsilon_2) - I(r_1 + r_2) \tag{3.62}$$

If we wish to replace the combination by a single cell between A and C of emf ε_{eq} and internal resistance r_{eq}, we would have

$$V_{AC} = \varepsilon_{eq} - I\, r_{eq} \tag{3.63}$$

Comparing the last two equations, we get

$$\varepsilon_{eq} = \varepsilon_1 + \varepsilon_2 \tag{3.64}$$

and $$r_{eq} = r_1 + r_2 \tag{3.65}$$

In Fig.3.20, we had connected the negative electrode of the first to the positive electrode of the second. If instead we connect the two negatives, Eq. (3.61) would change to $V_{BC} = -\varepsilon_2 - Ir_2$ and we will get

$$\varepsilon_{eq} = \varepsilon_1 - \varepsilon_2 \qquad (\varepsilon_1 > \varepsilon_2) \tag{3.66}$$

The rule for series combination clearly can be extended to any number of cells:

(i) The equivalent emf of a series combination of n cells is just the sum of their individual emf's, and

(ii) The equivalent internal resistance of a series combination of n cells is just the sum of their internal resistances.

This is so, when the current leaves each cell from the positive electrode. If in the combination, the current leaves any cell from the *negative* electrode, the emf of the cell enters the expression for ε_{eq} with a *negative* sign, as in Eq. (3.66).

FIGURE 3.21 Two cells in parallel. For connections across A and C, the combination can be replaced by one cell of emf ε_{eq} and internal resistances r_{eq} whose values are given in Eqs. (3.64) and (3.65).

Next, consider a parallel combination of the cells (Fig. 3.21). I_1 and I_2 are the currents leaving the positive electrodes of the cells. At the point B_1, I_1 and I_2 flow in whereas the current I flows out. Since as much charge flows in as out, we have

$$I = I_1 + I_2 \tag{3.67}$$

Let $V(B_1)$ and $V(B_2)$ be the potentials at B_1 and B_2, respectively. Then, considering the first cell, the potential difference across its terminals is $V(B_1) - V(B_2)$. Hence, from Eq. (3.57)

$$V \equiv V(B_1) - V(B_2) = \varepsilon_1 - I_1 r_1 \tag{3.68}$$

Points B_1 and B_2 are connected exactly similarly to the second cell. Hence considering the second cell, we also have

$$V \equiv V(B_1) - V(B_2) = \varepsilon_2 - I_2 r_2 \tag{3.69}$$

Combining the last three equations

$$I = I_1 + I_2$$

$$= \frac{\varepsilon_1 - V}{r_1} + \frac{\varepsilon_2 - V}{r_2} = \left(\frac{\varepsilon_1}{r_1} + \frac{\varepsilon_2}{r_2}\right) - V\left(\frac{1}{r_1} + \frac{1}{r_2}\right) \tag{3.70}$$

Hence, V is given by,

$$V = \frac{\varepsilon_1 r_2 + \varepsilon_2 r_1}{r_1 + r_2} - I\frac{r_1 r_2}{r_1 + r_2} \tag{3.71}$$

If we want to replace the combination by a single cell, between B_1 and B_2, of emf ε_{eq} and internal resistance r_{eq}, we would have

$$V = \varepsilon_{eq} - I\, r_{eq} \tag{3.72}$$

The last two equations should be the same and hence

$$\varepsilon_{eq} = \frac{\varepsilon_1 r_2 + \varepsilon_2 r_1}{r_1 + r_2} \tag{3.73}$$

$$r_{eq} = \frac{r_1 r_2}{r_1 + r_2} \tag{3.74}$$

We can put these equations in a simpler way,

$$\frac{1}{r_{eq}} = \frac{1}{r_1} + \frac{1}{r_2} \tag{3.75}$$

$$\frac{\varepsilon_{eq}}{r_{eq}} = \frac{\varepsilon_1}{r_1} + \frac{\varepsilon_2}{r_2} \tag{3.76}$$

In Fig. (3.21), we had joined the positive terminals together and similarly the two negative ones, so that the currents I_1, I_2 flow out of positive terminals. If the negative terminal of the second is connected to positive terminal of the first, Eqs. (3.75) and (3.76) would still be valid with $\varepsilon_2 \rightarrow -\varepsilon_2$

Equations (3.75) and (3.76) can be extended easily. If there an n cells of emf $\varepsilon_1, \ldots \varepsilon_n$ and of internal resistances $r_1, \ldots r_n$ respectively, connected in parallel, the combination is equivalent to a single cell of emf ε_{eq} and internal resistance r_{eq}, such that

$$\frac{1}{r_{eq}} = \frac{1}{r_1} + \cdots + \frac{1}{r_n} \tag{3.77}$$

$$\frac{\varepsilon_{eq}}{r_{eq}} = \frac{\varepsilon_1}{r_1} + \cdots + \frac{\varepsilon_n}{r_n} \tag{3.78}$$

Gustav Robert Kirchhoff (1824 – 1887) German physicist, professor at Heidelberg and at Berlin. Mainly known for his development of spectroscopy, he also made many important contributions to mathematical physics, among them, his first and second rules for circuits.

GUSTAV ROBERT KIRCHHOFF (1824 – 1887)

3.13 Kirchhoff's Rules

Electric circuits generally consist of a number of resistors and cells interconnected sometimes in a complicated way. The formulae we have derived earlier for series and parallel combinations of resistors are not always sufficient to determine all the currents and potential differences in the circuit. Two rules, called *Kirchhoff's rules*, are very useful for analysis of electric circuits.

Given a circuit, we start by labelling currents in each resistor by a symbol, say I, and a directed arrow to indicate that a current I flows along the resistor in the direction indicated. If ultimately I is determined to be positive, the actual current in the resistor is in the direction of the arrow. If I turns out to be negative, the current actually flows in a direction opposite to the arrow. Similarly, for each source (i.e., cell or some other source of electrical power) the positive and negative electrodes are labelled as well as a directed arrow with a symbol for the current flowing through the cell. This will tell us the potential difference, $V = V(\text{P}) - V(\text{N}) = \varepsilon - I\,r$

[Eq. (3.57) between the positive terminal P and the negative terminal N; I here is the current flowing from N to P through the cell]. If, while labelling the current I through the cell one goes from P to N, then of course

$$V = \varepsilon + I\,r \qquad (3.79)$$

FIGURE 3.22 At junction a the current leaving is $I_1 + I_2$ and current entering is I_3. The junction rule says $I_3 = I_1 + I_2$. At point h current entering is I_1. There is only one current leaving h and by junction rule that will also be I_1. For the loops 'ahdcba' and 'ahdefga', the loop rules give $-30I_1 - 41\,I_3 + 45 = 0$ and $-30I_1 + 21\,I_2 - 80 = 0$.

Having clarified labelling, we now state the rules and the proof:

(a) *Junction rule: At any junction, the sum of the currents entering the junction is equal to the sum of currents leaving the junction* (Fig. 3.22). This applies equally well if instead of a junction of several lines, we consider a point in a line.

The proof of this rule follows from the fact that when currents are steady, there is no accumulation of charges at any junction or at any point in a line. Thus, the total current flowing in, (which is the rate at which charge flows into the junction), must equal the total current flowing out.

(b) *Loop rule: The algebraic sum of changes in potential around any closed loop involving resistors and cells in the loop is zero* (Fig. 3.22). This rule is also obvious, since electric potential is dependent on the location of the point. Thus starting with any point if we come back to the same point, the total change must be zero. In a closed loop, we do come back to the starting point and hence the rule.

EXAMPLE 3.6

Example 3.6 A battery of 10 V and negligible internal resistance is connected across the diagonally opposite corners of a cubical network consisting of 12 resistors each of resistance 1 Ω (Fig. 3.23). Determine the equivalent resistance of the network and the current along each edge of the cube.

FIGURE 3.23

Solution The network is not reducible to a simple series and parallel combinations of resistors. There is, however, a clear symmetry in the problem which we can exploit to obtain the equivalent resistance of the network.

The paths AA′, AD and AB are obviously symmetrically placed in the network. Thus, the current in each must be the same, say, I. Further, at the corners A′, B and D, the incoming current I must split equally into the two outgoing branches. In this manner, the current in all the 12 edges of the cube are easily written down in terms of I, using Kirchhoff's first rule and the symmetry in the problem.

Next take a closed loop, say, ABCC′EA, and apply Kirchhoff's second rule:

$$-IR - (1/2)IR - IR + \varepsilon = 0$$

where R is the resistance of each edge and ε the emf of battery. Thus,

$$\varepsilon = \frac{5}{2} I R$$

The equivalent resistance R_{eq} of the network is

$$R_{eq} = \frac{\varepsilon}{3I} = \frac{5}{6} R$$

For $R = 1\ \Omega$, $R_{eq} = (5/6)\ \Omega$ and for $\varepsilon = 10$ V, the total current ($= 3I$) in the network is

$$3I = 10\ \text{V}/(5/6)\ \Omega = 12\ \text{A, i.e., } I = 4\ \text{A}$$

The current flowing in each edge can now be read off from the Fig. 3.23.

EXAMPLE 3.6

It should be noted that because of the symmetry of the network, the great power of Kirchhoff's rules has not been very apparent in Example 3.6. In a general network, there will be no such simplification due to symmetry, and only by application of Kirchhoff's rules to junctions and closed loops (as many as necessary to solve the unknowns in the network) can we handle the problem. This will be illustrated in Example 3.7.

Example 3.7 Determine the current in each branch of the network shown in Fig. 3.24.

FIGURE 3.24

EXAMPLE 3.7

PHYSICS

Simulation for application of Kirchhoff's rules:
http://www.phys.hawaii.edu/~teb/optics/java/kirch3/

EXAMPLE 3.7

Solution Each branch of the network is assigned an unknown current to be determined by the application of Kirchhoff's rules. To reduce the number of unknowns at the outset, the first rule of Kirchhoff is used at every junction to assign the unknown current in each branch. We then have three unknowns I_1, I_2 and I_3 which can be found by applying the second rule of Kirchhoff to three different closed loops. Kirchhoff's second rule for the closed loop ADCA gives,

$$10 - 4(I_1 - I_2) + 2(I_2 + I_3 - I_1) - I_1 = 0 \qquad [3.80(a)]$$

that is, $7I_1 - 6I_2 - 2I_3 = 10$

For the closed loop ABCA, we get

$$10 - 4I_2 - 2\,(I_2 + I_3) - I_1 = 0$$

that is, $I_1 + 6I_2 + 2I_3 = 10$ [3.80(b)]

For the closed loop BCDEB, we get

$$5 - 2\,(I_2 + I_3) - 2\,(I_2 + I_3 - I_1) = 0$$

that is, $2I_1 - 4I_2 - 4I_3 = -5$ [3.80(c)]

Equations (3.80 a, b, c) are three simultaneous equations in three unknowns. These can be solved by the usual method to give

$$I_1 = 2.5\text{A}, \quad I_2 = \frac{5}{8}\text{ A}, \quad I_3 = 1\frac{7}{8}\text{ A}$$

The currents in the various branches of the network are

$$\text{AB}: \frac{5}{8}\text{ A}, \quad \text{CA}: 2\frac{1}{2}\text{ A}, \quad \text{DEB}: 1\frac{7}{8}\text{ A}$$

$$\text{AD}: 1\frac{7}{8}\text{ A}, \quad \text{CD}: 0\text{ A}, \quad \text{BC}: 2\frac{1}{2}\text{ A}$$

It is easily verified that Kirchhoff's second rule applied to the remaining closed loops does not provide any additional independent equation, that is, the above values of currents satisfy the second rule for every closed loop of the network. For example, the total voltage drop over the closed loop BADEB

$$5\text{ V} + \left(\frac{5}{8}\cdot 4\right)\text{V} - \left(\frac{15}{8}\quad 4\right)\text{V}$$

equal to zero, as required by Kirchhoff's second rule.

3.14 WHEATSTONE BRIDGE

As an application of Kirchhoff's rules consider the circuit shown in Fig. 3.25, which is called the *Wheatstone bridge*. The bridge has four resistors R_1, R_2, R_3 and R_4. Across one pair of diagonally opposite points (A and C in the figure) a source is connected. This (*i.e.*, AC) is called the battery arm. Between the other two vertices, B and D, a galvanometer G (which is a device to detect currents) is connected. This line, shown as BD in the figure, is called the galvanometer arm.

For simplicity, we assume that the cell has no internal resistance. In general there will be currents flowing across all the resistors as well as a current I_g through G. Of special interest, is the case of a *balanced* bridge where the resistors are such that $I_g = 0$. We can easily get the balance condition, such that there is no current through G. In this case, the Kirchhoff's junction rule applied to junctions D and B (see the figure)

immediately gives us the relations $I_1 = I_3$ and $I_2 = I_4$. Next, we apply Kirchhoff's loop rule to closed loops ADBA and CBDC. The first loop gives

$$-I_1 R_1 + 0 + I_2 R_2 = 0 \qquad (I_g = 0) \qquad (3.81)$$

and the second loop gives, upon using $I_3 = I_1$, $I_4 = I_2$

$$I_2 R_4 + 0 - I_1 R_3 = 0 \qquad (3.82)$$

From Eq. (3.81), we obtain,

$$\frac{I_1}{I_2} = \frac{R_2}{R_1}$$

whereas from Eq. (3.82), we obtain,

$$\frac{I_1}{I_2} = \frac{R_4}{R_3}$$

Hence, we obtain the condition

$$\frac{R_2}{R_1} = \frac{R_4}{R_3} \qquad [3.83(a)]$$

FIGURE 3.25

This last equation relating the four resistors is called the *balance condition* for the galvanometer to give zero or null deflection.

The Wheatstone bridge and its balance condition provide a practical method for determination of an unknown resistance. Let us suppose we have an unknown resistance, which we insert in the fourth arm; R_4 is thus not known. Keeping known resistances R_1 and R_2 in the first and second arm of the bridge, we go on varying R_3 till the galvanometer shows a null deflection. The bridge then is balanced, and from the balance condition the value of the unknown resistance R_4 is given by,

$$R_4 = R_3 \frac{R_2}{R_1} \qquad [3.83(b)]$$

A practical device using this principle is called the *meter bridge*. It will be discussed in the next section.

Example 3.8 The four arms of a Wheatstone bridge (Fig. 3.26) have the following resistances:

AB = 100Ω, BC = 10Ω, CD = 5Ω, and DA = 60Ω.

FIGURE 3.26

EXAMPLE 3.8

EXAMPLE 3.8

A galvanometer of 15Ω resistance is connected across BD. Calculate the current through the galvanometer when a potential difference of 10 V is maintained across AC.

Solution Considering the mesh BADB, we have

$$100I_1 + 15I_g - 60I_2 = 0$$

or $$20I_1 + 3I_g - 12I_2 = 0 \qquad [3.84(a)]$$

Considering the mesh BCDB, we have

$$10\,(I_1 - I_g) - 15I_g - 5\,(I_2 + I_g) = 0$$

$$10I_1 - 30I_g - 5I_2 = 0$$

$$2I_1 - 6I_g - I_2 = 0 \qquad [3.84(b)]$$

Considering the mesh ADCEA,

$$60I_2 + 5\,(I_2 + I_g) = 10$$

$$65I_2 + 5I_g = 10$$

$$13I_2 + I_g = 2 \qquad [3.84(c)]$$

Multiplying Eq. (3.84b) by 10

$$20I_1 - 60I_g - 10I_2 = 0 \qquad [3.84(d)]$$

From Eqs. (3.84d) and (3.84a) we have

$$63I_g - 2I_2 = 0$$

$$I_2 = 31.5I_g \qquad [3.84(e)]$$

Substituting the value of I_2 into Eq. [3.84(c)], we get

$$13\,(31.5I_g) + I_g = 2$$

$$410.5\,I_g = 2$$

$$I_g = 4.87 \text{ mA}.$$

FIGURE 3.27 A meter bridge. Wire AC is 1 m long. *R* is a resistance to be measured and *S* is a standard resistance.

3.15 METER BRIDGE

The meter bridge is shown in Fig. 3.27. It consists of a wire of length 1 m and of uniform cross sectional area stretched taut and clamped between two thick metallic strips bent at right angles, as shown. The metallic strip has two gaps across which resistors can be connected. The end points where the wire is clamped are connected to a cell through a key. One end of a galvanometer is connected to the metallic strip midway between the two gaps. The other end of the galvanometer is connected to a 'jockey'. The jockey is essentially a metallic rod whose one end has a knife-edge which can slide over the wire to make electrical connection.

R is an unknown resistance whose value we want to determine. It is connected across one of the gaps. Across the other gap, we connect a

standard known resistance S. The jockey is connected to some point D on the wire, a distance l cm from the end A. The jockey can be moved along the wire. The portion AD of the wire has a resistance $R_{cm}l$, where R_{cm} is the resistance of the wire per unit centimetre. The portion DC of the wire similarly has a resistance R_{cm} (100-l).

The four arms AB, BC, DA and CD [with resistances R, S, $R_{cm}\,l$ and R_{cm}(100-l)] obviously form a Wheatstone bridge with AC as the battery arm and BD the galvanometer arm. If the jockey is moved along the wire, then there will be one position where the galvanometer will show no current. Let the distance of the jockey from the end A at the balance point be $l = l_1$. The four resistances of the bridge at the balance point then are R, S, $R_{cm}\,l_1$ and $R_{cm}(100-l_1)$. The balance condition, Eq. [3.83(a)] gives

$$\frac{R}{S} = \frac{R_{cm}\,l_1}{R_{cm}\left(100 - l_1\right)} = \frac{l_1}{100 - l_1} \tag{3.85}$$

Thus, once we have found out l_1, the unknown resistance R is known in terms of the standard known resistance S by

$$R = S\,\frac{l_1}{100 - l_1} \tag{3.86}$$

By choosing various values of S, we would get various values of l_1, and calculate R each time. An error in measurement of l_1 would naturally result in an error in R. It can be shown that the percentage error in R can be minimised by adjusting the balance point near the middle of the bridge, i.e., when l_1 is close to 50 cm. (This requires a suitable choice of S.)

Example 3.9 In a metre bridge (Fig. 3.27), the null point is found at a distance of 33.7 cm from A. If now a resistance of 12Ω is connected in parallel with S, the null point occurs at 51.9 cm. Determine the values of R and S.

Solution From the first balance point, we get

$$\frac{R}{S} = \frac{33.7}{66.3} \tag{3.87}$$

After S is connected in parallel with a resistance of 12Ω , the resistance across the gap changes from S to S_{eq}, where

$$S_{eq} = \frac{12S}{S+12}$$

and hence the new balance condition now gives

$$\frac{51.9}{48.1} = \frac{R}{S_{eq}} = \frac{R(S+12)}{12\,S} \tag{3.88}$$

Substituting the value of R/S from Eq. (3.87), we get

$$\frac{51.9}{48.1} = \frac{S+12}{12} \cdot \frac{33.7}{66.3}$$

which gives S = 13.5Ω. Using the value of R/S above, we get R = 6.86 Ω.

EXAMPLE 3.9

3.16 POTENTIOMETER

This is a versatile instrument. It is basically a long piece of uniform wire, sometimes a few meters in length across which a standard cell is connected. In actual design, the wire is sometimes cut in several pieces placed side by side and connected at the ends by thick metal strip. (Fig. 3.28). In the figure, the wires run from A to C. The small vertical portions are the thick metal strips connecting the various sections of the wire.

A current I flows through the wire which can be varied by a variable resistance (rheostat, R) in the circuit. Since the wire is uniform, the potential difference between A and any point at a distance l from A is

$$\varepsilon(l) = \phi l \tag{3.89}$$

where ϕ is the potential drop per unit length.

Figure 3.28 (a) shows an application of the potentiometer to compare the emf of two cells of emf ε_1 and ε_2 . The points marked 1, 2, 3 form a two way key. Consider first a position of the key where 1 and 3 are connected so that the galvanometer is connected to ε_1. The jockey is moved along the wire till at a point N_1, at a distance l_1 from A, there is no deflection in the galvanometer. We can apply Kirchhoff's loop rule to the closed loop AN_1G31A and get,

$$\phi\, l_1 + 0 - \varepsilon_1 = 0 \tag{3.90}$$

Similarly, if another emf ε_2 is balanced against l_2 (AN_2)

$$\phi\, l_2 + 0 - \varepsilon_2 = 0 \tag{3.91}$$

From the last two equations

$$\frac{\varepsilon_1}{\varepsilon_2} = \frac{l_1}{l_2} \tag{3.92}$$

This simple mechanism thus allows one to compare the emf's of any two sources. In practice one of the cells is chosen as a standard cell whose emf is known to a high degree of accuracy. The emf of the other cell is then easily calculated from Eq. (3.92).

We can also use a potentiometer to measure internal resistance of a cell [Fig. 3.28 (b)]. For this the cell (emf ε) whose internal resistance (r) is to be determined is connected across a resistance box through a key K_2, as shown in the figure. With key K_2 open, balance is obtained at length l_1 (AN_1). Then,

$$\varepsilon = \phi\, l_1 \tag{3.93(a)}$$

When key K_2 is closed, the cell sends a current (I) through the resistance box (R). If V is the terminal potential difference of the cell and balance is obtained at length l_2 (AN_2),

$$V = \phi\, l_2 \tag{3.93(b)}$$

FIGURE 3.28 A potentiometer. G is a galvanometer and R a variable resistance (rheostat). 1, 2, 3 are terminals of a two way key (a) circuit for comparing emfs of two cells; (b) circuit for determining internal resistance of a cell.

So, we have $\varepsilon/V = l_1/l_2$ [3.94(a)]

But, $\varepsilon = I(r + R)$ and $V = IR$. This gives

$\varepsilon/V = (r+R)/R$ [3.94(b)]

From Eq. [3.94(a)] and [3.94(b)] we have

$(R+r)/R = l_1/l_2$

$$r = R\left(\frac{l_1}{l_2} - 1\right) \quad (3.95)$$

Using Eq. (3.95) we can find the internal resistance of a given cell.

The potentiometer has the advantage that it draws *no current* from the voltage source being measured. As such it is unaffected by the internal resistance of the source.

Example 3.10 A resistance of R Ω draws current from a potentiometer. The potentiometer has a total resistance R_0 Ω (Fig. 3.29). A voltage V is supplied to the potentiometer. Derive an expression for the voltage across R when the sliding contact is in the middle of the potentiometer.

FIGURE 3.29

Solution While the slide is in the middle of the potentiometer only half of its resistance ($R_0/2$) will be between the points A and B. Hence, the total resistance between A and B, say, R_1, will be given by the following expression:

$$\frac{1}{R_1} = \frac{1}{R} + \frac{1}{(R_0/2)}$$

$$R_1 = \frac{R_0 R}{R_0 + 2R}$$

The total resistance between A and C will be sum of resistance between A and B and B and C, i.e., $R_1 + R_0/2$

∴ The current flowing through the potentiometer will be

$$I = \frac{V}{R_1 + R_0/2} = \frac{2V}{2R_1 + R_0}$$

The voltage V_1 taken from the potentiometer will be the product of current I and resistance R_1,

$$V_1 = I R_1 = \left(\frac{2V}{2R_1 + R_0}\right) \cdot R_1$$

EXAMPLE 3.10

EXAMPLE 3.10

Substituting for R_1, we have a

$$V_1 = \frac{2V}{2\left(\frac{R_0 \cdot R}{R_0 + 2R}\right) + R_0} \cdot \frac{R_0 \cdot R}{R_0 + 2R}$$

$$V_1 = \frac{2VR}{2R + R_0 + 2R}$$

or $$V_1 = \frac{2VR}{R_0 + 4R}.$$

SUMMARY

1. *Current* through a given area of a conductor is the net charge passing per unit time through the area.
2. To maintain a steady current, we must have a closed circuit in which an external agency moves electric charge from lower to higher potential energy. The work done per unit charge by the source in taking the charge from lower to higher potential energy (i.e., from one terminal of the source to the other) is called the electromotive force, or *emf*, of the source. Note that the emf is not a force; it is the voltage difference between the two terminals of a source in open circuit.
3. *Ohm's law*: The electric current I flowing through a substance is proportional to the voltage V across its ends, i.e., $V \propto I$ or $V = RI$, where R is called the *resistance* of the substance. The unit of resistance is ohm: $1\Omega = 1\ \text{V A}^{-1}$.
4. The *resistance* R of a conductor depends on its length l and constant cross-sectional area A through the relation,

   $$R = \frac{\rho l}{A}$$

   where ρ, called *resistivity* is a property of the material and depends on temperature and pressure.
5. *Electrical resistivity* of substances varies over a very wide range. Metals have low resistivity, in the range of $10^{-8}\ \Omega$ m to $10^{-6}\ \Omega$ m. Insulators like glass and rubber have 10^{22} to 10^{24} times greater resistivity. Semiconductors like Si and Ge lie roughly in the middle range of resistivity on a logarithmic scale.
6. In most substances, the carriers of current are electrons; in some cases, for example, ionic crystals and electrolytic liquids, positive and negative ions carry the electric current.
7. *Current density* $\mathbf{j}$ gives the amount of charge flowing per second per unit area normal to the flow,

   $$\mathbf{j} = nq\,\mathbf{v}_d$$

   where n is the number density (number per unit volume) of charge carriers each of charge q, and $\mathbf{v}_d$ is the *drift velocity* of the charge carriers. For electrons $q = -e$. If $\mathbf{j}$ is normal to a cross-sectional area $\mathbf{A}$ and is constant over the area, the magnitude of the current I through the area is $nev_d A$.

8. Using $E = V/l$, $I = nev_d A$, and Ohm's law, one obtains

$$\frac{eE}{m} = \rho \frac{ne^2}{m} v_d$$

The proportionality between the *force eE* on the electrons in a metal due to the external field E and the drift velocity v_d (not acceleration) can be understood, if we assume that the electrons suffer collisions with ions in the metal, which deflect them randomly. If such collisions occur on an average at a time interval τ,

$v_d = a\tau = eE\tau/m$

where a is the acceleration of the electron. This gives

$$\rho = \frac{m}{ne^2\tau}$$

9. In the temperature range in which resistivity increases linearly with temperature, the *temperature coefficient of resistivity* α is defined as the fractional increase in resistivity per unit increase in temperature.

10. Ohm's law is obeyed by many substances, but it is not a fundamental law of nature. It fails if
    (a) V depends on I non-linearly.
    (b) the relation between V and I depends on the sign of V for the same absolute value of V.
    (c) The relation between V and I is non-unique.

    An example of (a) is when ρ increases with I (even if temperature is kept fixed). A rectifier combines features (a) and (b). GaAs shows the feature (c).

11. When a source of emf ε is connected to an external resistance R, the voltage V_{ext} across R is given by

$$V_{ext} = IR = \frac{\varepsilon}{R+r} R$$

where r is the *internal resistance* of the source.

12. (a) Total resistance R of n resistors connected in *series* is given by

    $R = R_1 + R_2 + \ldots + R_n$

    (b) Total resistance R of n resistors connected in *parallel* is given by

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \ldots\ldots + \frac{1}{R_n}$$

13. *Kirchhoff's Rules* –
    (a) *Junction Rule:* At any junction of circuit elements, the sum of currents entering the junction must equal the sum of currents leaving it.
    (b) *Loop Rule:* The algebraic sum of changes in potential around any closed loop must be zero.

14. The *Wheatstone bridge* is an arrangement of four resistances – R_1, R_2, R_3, R_4 as shown in the text. The null-point condition is given by

$$\frac{R_1}{R_2} = \frac{R_3}{R_4}$$

using which the value of one resistance can be determined, knowing the other three resistances.

15. The *potentiometer* is a device to compare potential differences. Since the method involves a condition of *no* current flow, the device can be used to measure potential difference; internal resistance of a cell and compare emf's of two sources.

Physical Quantity	Symbol	Dimensions	Unit	Remark
Electric current	I	[A]	A	SI base unit
Charge	Q, q	[T A]	C	
Voltage, Electric potential difference	V	$[M\ L^2\ T^{-3}\ A^{-1}]$	V	Work/charge
Electromotive force	ε	$[M\ L^2\ T^{-3}\ A^{-1}]$	V	Work/charge
Resistance	R	$[M\ L^2 T^{-3}\ A^{-2}]$	Ω	$R = V/I$
Resistivity	ρ	$[M\ L^3 T^{-3}\ A^{-2}]$	Ω m	$R = \rho l/A$
Electrical conductivity	σ	$[M^{-1}\ L^{-3}\ T^3\ A^2]$	S	$\sigma = 1/\rho$
Electric field	**E**	$[M\ L\ T^{-3}\ A^{-1}]$	$V\ m^{-1}$	$\frac{\text{Electric force}}{\text{charge}}$
Drift speed	v_d	$[L\ T^{-1}]$	$m\ s^{-1}$	$v_d = \frac{e\,E\,\tau}{m}$
Relaxation time	τ	[T]	s	
Current density	**j**	$[L^{-2}\ A]$	$A\ m^{-2}$	current/area
Mobility	$\propto$	$[M\ L^3 T^{-4}\ A^{-1}]$	$m^2\ V^{-1} s^{-1}$	v_d / E

POINTS TO PONDER

1. Current is a scalar although we represent current with an arrow. Currents do not obey the law of vector addition. That current is a scalar also follows from it's definition. The current I through an area of cross-section is given by the scalar product of two vectors:

   $I = \mathbf{j} \cdot \Delta\mathbf{S}$

   where **j** and Δ**S** are vectors.
2. Refer to V-I curves of a resistor and a diode as drawn in the text. A resistor obeys Ohm's law while a diode does not. The assertion that $V = IR$ is a statement of Ohm's law is not true. This equation defines resistance and it may be applied to all conducting devices whether they obey Ohm's law or not. The Ohm's law asserts that the plot of I versus V is linear i.e., R is independent of V.

   Equation $\mathbf{E} = \rho\,\mathbf{j}$ leads to another statement of *Ohm's* law, i.e., a conducting material obeys Ohm's law when the resistivity of the material does not depend on the magnitude and direction of applied electric field.
3. Homogeneous conductors like silver or semiconductors like pure germanium or germanium containing impurities obey Ohm's law within some range of electric field values. If the field becomes too strong, there are departures from Ohm's law in all cases.
4. Motion of conduction electrons in electric field **E** is the sum of (i) motion due to random collisions and (ii) that due to **E**. The motion

due to random collisions averages to zero and does not contribute to v_d (Chapter 11, Textbook of Class XI). v_d , thus is only due to applied electric field on the electron.

5. The relation $\mathbf{j} = \rho\,\mathbf{v}$ should be applied to each type of charge carriers separately. In a conducting wire, the total current and charge density arises from both positive and negative charges:

   $\mathbf{j} = \rho_+ \mathbf{v}_+ + \rho_- \mathbf{v}_-$

   $\rho = \rho_+ + \rho_-$

   Now in a neutral wire carrying electric current,

   $\rho_+ = -\rho_-$

   Further, $v_+ \sim 0$ which gives

   $\rho = 0$

   $\mathbf{j} = \rho_- \mathbf{v}$

   Thus, the relation $\mathbf{j} = \rho\,\mathbf{v}$ does not apply to the total current charge density.

6. Kirchhoff's junction rule is based on conservation of charge and the outgoing currents add up and are equal to incoming current at a junction. Bending or reorienting the wire does not change the validity of Kirchhoff's junction rule.

EXERCISES

3.1 The storage battery of a car has an emf of 12 V. If the internal resistance of the battery is 0.4 Ω, what is the maximum current that can be drawn from the battery?

3.2 A battery of emf 10 V and internal resistance 3 Ω is connected to a resistor. If the current in the circuit is 0.5 A, what is the resistance of the resistor? What is the terminal voltage of the battery when the circuit is closed?

3.3 (a) Three resistors 1 Ω, 2 Ω, and 3 Ω are combined in series. What is the total resistance of the combination?
(b) If the combination is connected to a battery of emf 12 V and negligible internal resistance, obtain the potential drop across each resistor.

3.4 (a) Three resistors 2 Ω, 4 Ω and 5 Ω are combined in parallel. What is the total resistance of the combination?
(b) If the combination is connected to a battery of emf 20 V and negligible internal resistance, determine the current through each resistor, and the total current drawn from the battery.

3.5 At room temperature (27.0 °C) the resistance of a heating element is 100 Ω. What is the temperature of the element if the resistance is found to be 117 Ω, given that the temperature coefficient of the material of the resistor is 1.70×10^{-4} °C^{-1}.

3.6 A negligibly small current is passed through a wire of length 15 m and uniform cross-section 6.0×10^{-7} m^2, and its resistance is measured to be 5.0 Ω. What is the resistivity of the material at the temperature of the experiment?

3.7 A silver wire has a resistance of 2.1 Ω at 27.5 °C, and a resistance of 2.7 Ω at 100 °C. Determine the temperature coefficient of resistivity of silver.

3.8 A heating element using nichrome connected to a 230 V supply draws an initial current of 3.2 A which settles after a few seconds to

a steady value of 2.8 A. What is the steady temperature of the heating element if the room temperature is 27.0 °C? Temperature coefficient of resistance of nichrome averaged over the temperature range involved is 1.70×10^{-4} °C^{-1}.

3.9 Determine the current in each branch of the network shown in Fig. 3.30:

FIGURE 3.30

3.10 (a) In a metre bridge [Fig. 3.27], the balance point is found to be at 39.5 cm from the end *A*, when the resistor *Y* is of 12.5 Ω. Determine the resistance of *X*. Why are the connections between resistors in a Wheatstone or meter bridge made of thick copper strips?

(b) Determine the balance point of the bridge above if *X* and *Y* are interchanged.

(c) What happens if the galvanometer and cell are interchanged at the balance point of the bridge? Would the galvanometer show any current?

3.11 A storage battery of emf 8.0 V and internal resistance 0.5 Ω is being charged by a 120 V dc supply using a series resistor of 15.5 Ω. What is the terminal voltage of the battery during charging? What is the purpose of having a series resistor in the charging circuit?

3.12 In a potentiometer arrangement, a cell of emf 1.25 V gives a balance point at 35.0 cm length of the wire. If the cell is replaced by another cell and the balance point shifts to 63.0 cm, what is the emf of the second cell?

3. 13 The number density of free electrons in a copper conductor estimated in Example 3.1 is 8.5×10^{28} m^{-3}. How long does an electron take to drift from one end of a wire 3.0 m long to its other end? The area of cross-section of the wire is 2.0×10^{-6} m^2 and it is carrying a current of 3.0 A.

ADDITIONAL EXERCISES

3. 14 The earth's surface has a negative surface charge density of 10^{-9} C m^{-2}. The potential difference of 400 kV between the top of the atmosphere and the surface results (due to the low conductivity of the lower atmosphere) in a current of only 1800 A over the entire globe. If there were no mechanism of sustaining atmospheric electric

field, how much time (roughly) would be required to neutralise the earth's surface? (This never happens in practice because there is a mechanism to replenish electric charges, namely the continual thunderstorms and lightning in different parts of the globe). (Radius of earth = 6.37×10^6 m.)

3.15 (a) Six lead-acid type of secondary cells each of emf 2.0 V and internal resistance 0.015 Ω are joined in series to provide a supply to a resistance of 8.5 Ω. What are the current drawn from the supply and its terminal voltage?

(b) A secondary cell after long use has an emf of 1.9 V and a large internal resistance of 380 Ω. What maximum current can be drawn from the cell? Could the cell drive the starting motor of a car?

3.16 Two wires of equal length, one of aluminium and the other of copper have the same resistance. Which of the two wires is lighter? Hence explain why aluminium wires are preferred for overhead power cables. ($\rho_{Al} = 2.63 \times 10^{-8}$ Ω m, $\rho_{Cu} = 1.72 \times 10^{-8}$ Ω m, Relative density of Al = 2.7, of Cu = 8.9.)

3.17 What conclusion can you draw from the following observations on a resistor made of alloy manganin?

Current A	Voltage V	Current A	Voltage V
0.2	3.94	3.0	59.2
0.4	7.87	4.0	78.8
0.6	11.8	5.0	98.6
0.8	15.7	6.0	118.5
1.0	19.7	7.0	138.2
2.0	39.4	8.0	158.0

3.18 Answer the following questions:

(a) A steady current flows in a metallic conductor of non-uniform cross-section. Which of these quantities is constant along the conductor: current, current density, electric field, drift speed?

(b) Is Ohm's law universally applicable for all conducting elements? If not, give examples of elements which do not obey Ohm's law.

(c) A low voltage supply from which one needs high currents must have very low internal resistance. Why?

(d) A high tension (HT) supply of, say, 6 kV must have a very large internal resistance. Why?

3.19 Choose the correct alternative:

(a) Alloys of metals usually have (greater/less) resistivity than that of their constituent metals.

(b) Alloys usually have much (lower/higher) temperature coefficients of resistance than pure metals.

(c) The resistivity of the alloy manganin is nearly independent of/increases rapidly with increase of temperature.

(d) The resistivity of a typical insulator (e.g., amber) is greater than that of a metal by a factor of the order of ($10^{22}/10^3$).

3.20 (a) Given n resistors each of resistance R, how will you combine them to get the (i) maximum (ii) minimum effective resistance? What is the ratio of the maximum to minimum resistance?

(b) Given the resistances of 1 Ω, 2 Ω, 3 Ω, how will be combine them to get an equivalent resistance of (i) (11/3) Ω (ii) (11/5) Ω, (iii) 6 Ω, (iv) (6/11) Ω?

(c) Determine the equivalent resistance of networks shown in Fig. 3.31.

FIGURE 3.31

3.21 Determine the current drawn from a 12V supply with internal resistance 0.5Ω by the infinite network shown in Fig. 3.32. Each resistor has 1Ω resistance.

FIGURE 3.32

3.22 Figure 3.33 shows a potentiometer with a cell of 2.0 V and internal resistance 0.40 Ω maintaining a potential drop across the resistor wire AB. A standard cell which maintains a constant emf of 1.02 V (for very moderate currents upto a few mA) gives a balance point at 67.3 cm length of the wire. To ensure very low currents drawn from the standard cell, a very high resistance of 600 kΩ is put in series with it, which is shorted close to the balance point. The standard cell is then replaced by a cell of unknown emf ε and the balance point found similarly, turns out to be at 82.3 cm length of the wire.

FIGURE 3.33

(a) What is the value ε?
(b) What purpose does the high resistance of 600 kΩ have?

(c) Is the balance point affected by this high resistance?
(d) Is the balance point affected by the internal resistance of the driver cell?
(e) Would the method work in the above situation if the driver cell of the potentiometer had an emf of 1.0V instead of 2.0V?
(f) Would the circuit work well for determining an extremely small emf, say of the order of a few mV (such as the typical emf of a thermo-couple)? If not, how will you modify the circuit?

3.23 Figure 3.34 shows a potentiometer circuit for comparison of two resistances. The balance point with a standard resistor R = 10.0 Ω is found to be 58.3 cm, while that with the unknown resistance *X* is 68.5 cm. Determine the value of *X*. What might you do if you failed to find a balance point with the given cell of emf ε?

FIGURE 3.34

3.24 Figure 3.35 shows a 2.0 V potentiometer used for the determination of internal resistance of a 1.5 V cell. The balance point of the cell in open circuit is 76.3 cm. When a resistor of 9.5 Ω is used in the external circuit of the cell, the balance point shifts to 64.8 cm length of the potentiometer wire. Determine the internal resistance of the cell.

FIGURE 3.35

Chapter Four

MOVING CHARGES AND MAGNETISM

4.1 Introduction

Both Electricity and Magnetism have been known for more than 2000 years. However, it was only about 200 years ago, in 1820, that it was realised that they were intimately related*. During a lecture demonstration in the summer of 1820, the Danish physicist Hans Christian Oersted noticed that a current in a straight wire caused a noticeable deflection in a nearby magnetic compass needle. He investigated this phenomenon. He found that the alignment of the needle is tangential to an imaginary circle which has the straight wire as its centre and has its plane perpendicular to the wire. This situation is depicted in Fig.4.1(a). It is noticeable when the current is large and the needle sufficiently close to the wire so that the earth's magnetic field may be ignored. Reversing the direction of the current reverses the orientation of the needle [Fig. 4.1(b)]. The deflection increases on increasing the current or bringing the needle closer to the wire. Iron filings sprinkled around the wire arrange themselves in concentric circles with the wire as the centre [Fig. 4.1(c)]. Oersted concluded that *moving charges or currents produced a magnetic field in the surrounding space.*

Following this there was intense experimentation. In 1864, the laws obeyed by electricity and magnetism were unified and formulated by

* See the box in Chapter 1, Page 3.

James Maxwell who then realised that light was electromagnetic waves. Radio waves were discovered by Hertz, and produced by J.C.Bose and G. Marconi by the end of the 19th century. A remarkable scientific and technological progress has taken place in the 20th century. This is due to our increased understanding of electromagnetism and the invention of devices for production, amplification, transmission and detection of electromagnetic waves.

FIGURE 4.1 The magnetic field due to a straight long current-carrying wire. The wire is perpendicular to the plane of the paper. A ring of compass needles surrounds the wire. The orientation of the needles is shown when (a) the current emerges out of the plane of the paper, (b) the current moves into the plane of the paper. (c) The arrangement of iron filings around the wire. The darkened ends of the needle represent north poles. The effect of the earth's magnetic field is neglected.

In this chapter, we will see how magnetic field exerts forces on moving charged particles, like electrons, protons, and current-carrying wires. We shall also learn how currents produce magnetic fields. We shall see how particles can be accelerated to very high energies in a cyclotron. We shall study how currents and voltages are detected by a galvanometer.

In this and subsequent Chapter on magnetism, we adopt the following convention: A current or a field (electric or magnetic) emerging out of the plane of the paper is depicted by a dot (⊙). A current or a field going into the plane of the paper is depicted by a cross (⊗)*. Figures. 4.1(a) and 4.1(b) correspond to these two situations, respectively.

Hans Christian Oersted (1777–1851) Danish physicist and chemist, professor at Copenhagen. He observed that a compass needle suffers a deflection when placed near a wire carrying an electric current. This discovery gave the first empirical evidence of a connection between electric and magnetic phenomena.

HANS CHRISTIAN OERSTED (1777–1851)

4.2 Magnetic Force

4.2.1 Sources and fields

Before we introduce the concept of a magnetic field **B**, we shall recapitulate what we have learnt in Chapter 1 about the electric field **E.** We have seen that the interaction between two charges can be considered in two stages. The charge Q, the source of the field, produces an electric field **E**, where

* A dot appears like the tip of an arrow pointed at you, a cross is like the feathered tail of an arrow moving away from you.

HENDRIK ANTOON LORENTZ (1853 – 1928)

Hendrik Antoon Lorentz (1853 – 1928) Dutch theoretical physicist, professor at Leiden. He investigated the relationship between electricity, magnetism, and mechanics. In order to explain the observed effect of magnetic fields on emitters of light (Zeeman effect), he postulated the existence of electric charges in the atom, for which he was awarded the Nobel Prize in 1902. He derived a set of transformation equations (known after him, as Lorentz transformation equations) by some tangled mathematical arguments, but he was not aware that these equations hinge on a new concept of space and time.

$$\mathbf{E} = Q\,\hat{\mathbf{r}} / (4\pi\varepsilon_0) r^2 \qquad (4.1)$$

where $\hat{\mathbf{r}}$ is unit vector along $\mathbf{r}$, and the field $\mathbf{E}$ is a vector field. A charge q interacts with this field and experiences a force $\mathbf{F}$ given by

$$\mathbf{F} = q\,\mathbf{E} = q\,Q\,\hat{\mathbf{r}} / (4\pi\varepsilon_0)\, r^2 \qquad (4.2)$$

As pointed out in the Chapter 1, the field $\mathbf{E}$ is not just an artefact but has a physical role. It can convey energy and momentum and is not established instantaneously but takes finite time to propagate. The concept of a field was specially stressed by Faraday and was incorporated by Maxwell in his unification of electricity and magnetism. In addition to depending on each point in space, it can also vary with time, i.e., be a function of time. In our discussions in this chapter, we will assume that the fields do not change with time.

The field at a particular point can be due to one or more charges. If there are more charges the fields add vectorially. You have already learnt in Chapter 1 that this is called the principle of superposition. Once the field is known, the force on a test charge is given by Eq. (4.2).

Just as static charges produce an electric field, the currents or moving charges produce (in addition) a magnetic field, denoted by $\mathbf{B}(\mathbf{r})$, again a vector field. It has several basic properties identical to the electric field. It is defined at each point in space (and can in addition depend on time). Experimentally, it is found to obey the principle of superposition: *the magnetic field of several sources is the vector addition of magnetic field of each individual source.*

4.2.2 Magnetic Field, Lorentz Force

Let us suppose that there is a point charge q (moving with a velocity $\mathbf{v}$ and, located at $\mathbf{r}$ at a given time t) in presence of both the electric field $\mathbf{E}(\mathbf{r})$ and the magnetic field $\mathbf{B}(\mathbf{r})$. The force on an electric charge q due to both of them can be written as

$$\mathbf{F} = q\,[\,\mathbf{E}(\mathbf{r}) + \mathbf{v} \times \mathbf{B}(\mathbf{r})] \equiv \mathbf{F}_{\text{electric}} + \mathbf{F}_{\text{magnetic}} \qquad (4.3)$$

This force was given first by H.A. Lorentz based on the extensive experiments of Ampere and others. It is called the *Lorentz force*. You have already studied in detail the force due to the electric field. If we look at the interaction with the magnetic field, we find the following features.

(i) It depends on q, $\mathbf{v}$ and $\mathbf{B}$ (charge of the particle, the velocity and the magnetic field). *Force on a negative charge is opposite to that on a positive charge.*

(ii) The magnetic force $q\,[\,\mathbf{v} \times \mathbf{B}\,]$ includes a vector product of velocity and magnetic field. The vector product makes the force due to magnetic

field vanish (become zero) if velocity and magnetic field are parallel or anti-parallel. The force acts in a (sideways) direction perpendicular to both the velocity and the magnetic field. Its direction is given by the screw rule or right hand rule for vector (or cross) product as illustrated in Fig. 4.2.

(iii) The magnetic force is zero if charge is not moving (as then $|\mathbf{v}| = 0$). Only a moving charge feels the magnetic force.

The expression for the magnetic force helps us to define the unit of the magnetic field, if one takes q, **F** and **v**, all to be unity in the force equation $\mathbf{F} = q\,[\mathbf{v} \times \mathbf{B}] = q\,v\,B \sin\theta\,\hat{\mathbf{n}}$, where θ is the angle between **v** and **B** [see Fig. 4.2 (a)]. The magnitude of magnetic field B is 1 SI unit, when the force acting on a unit charge (1 C), moving perpendicular to **B** with a speed 1m/s, is one newton.

Dimensionally, we have $[B] = [F/qv]$ and the unit of **B** are Newton second / (coulomb metre). This unit is called *tesla* (T) named after Nikola Tesla (1856 – 1943). Tesla is a rather large unit. A smaller unit (non-SI) called *gauss* ($=10^{-4}$ tesla) is also often used. The earth's magnetic field is about 3.6×10^{-5} T. Table 4.1 lists magnetic fields over a wide range in the universe.

FIGURE 4.2 The direction of the magnetic force acting on a charged particle. (a) The force on a positively charged particle with velocity **v** and making an angle θ with the magnetic field **B** is given by the right-hand rule. (b) A moving charged particle q is deflected in an opposite sense to $-q$ in the presence of magnetic field.

TABLE 4.1 ORDER OF MAGNITUDES OF MAGNETIC FIELDS IN A VARIETY OF PHYSICAL SITUATIONS

Physical situation	Magnitude of B (in tesla)
Surface of a neutron star	10^{8}
Typical large field in a laboratory	1
Near a small bar magnet	10^{-2}
On the earth's surface	10^{-5}
Human nerve fibre	10^{-10}
Interstellar space	10^{-12}

4.2.3 Magnetic force on a current-carrying conductor

We can extend the analysis for force due to magnetic field on a single moving charge to a straight rod carrying current. Consider a rod of a uniform cross-sectional area A and length l. We shall assume one kind of mobile carriers as in a conductor (here electrons). Let the number density of these mobile charge carriers in it be n. Then the total number of mobile charge carriers in it is nAl. For a steady current I in this conducting rod, we may assume that each mobile carrier has an average

drift velocity $\mathbf{v}_d$ (see Chapter 3). In the presence of an external magnetic field **B**, the force on these carriers is:

$$\mathbf{F} = (nAl)q\,\mathbf{v}_d \times \mathbf{B}$$

where q is the value of the charge on a carrier. Now $nq\mathbf{v}_d$ is the current density $\mathbf{j}$ and $|(nq\,\mathbf{v}_d)|A$ is the current I (see Chapter 3 for the discussion of current and current density). Thus,

$$\mathbf{F} = [(nqe\mathbf{v}_d)Al] \times \mathbf{B} = [\,\mathbf{j}Al\,] \times \mathbf{B}$$

$$= I\mathbf{l} \times \mathbf{B} \qquad (4.4)$$

where $\mathbf{l}$ is a vector of magnitude l, the length of the rod, and with a direction identical to the current I. Note that the current I is not a vector. In the last step leading to Eq. (4.4), we have transferred the vector sign from $\mathbf{j}$ to $\mathbf{l}$.

Equation (4.4) holds for a straight rod. In this equation, **B** is the external magnetic field. It is not the field produced by the current-carrying rod. If the wire has an arbitrary shape we can calculate the Lorentz force on it by considering it as a collection of linear strips $d\mathbf{l}_j$ and summing

$$\mathbf{F} = \sum_j I d\mathbf{l}_j \times \mathbf{B}$$

This summation can be converted to an integral in most cases.

On permittivity and permeability

In the universal law of gravitation, we say that any two point masses exert a force on each other which is proportional to the product of the masses m_1, m_2 and inversely proportional to the square of the distance r between them. We write it as $F = Gm_1m_2/r^2$ where G is the universal constant of gravitation. Similarly in Coulomb's law of electrostatics we write the force between two point charges q_1, q_2, separated by a distance r as $F = kq_1q_2/r^2$ where k is a constant of proportionality. In SI units, k is taken as $1/4\pi\varepsilon$ where ε is the permittivity of the medium. Also in magnetism, we get another constant, which in SI units, is taken as $\propto/4\pi$ where $\propto$ is the permeability of the medium.

Although G, ε and $\propto$ arise as proportionality constants, there is a difference between gravitational force and electromagnetic force. While the gravitational force does not depend on the intervening medium, the electromagnetic force depends on the medium between the two charges or magnets. Hence while G is a universal constant, ε and $\propto$ depend on the medium. They have different values for different media. The product $\varepsilon\propto$ turns out to be related to the speed v of electromagnetic radiation in the medium through $\varepsilon\propto = 1/v^2$.

Electric permittivity ε is a physical quantity that describes how an electric field affects and is affected by a medium. It is determined by the ability of a material to polarise in response to an applied field, and thereby to cancel, partially, the field inside the material. Similarly, magnetic permeability $\propto$ is the ability of a substance to acquire magnetisation in magnetic fields. It is a measure of the extent to which magnetic field can penetrate matter.

Example 4.1

Example 4.1 A straight wire of mass 200 g and length 1.5 m carries a current of 2 A. It is suspended in mid-air by a uniform horizontal magnetic field **B** (Fig. 4.3). What is the magnitude of the magnetic field?

FIGURE 4.3

Solution From Eq. (4.4), we find that there is an upward force **F**, of magnitude IlB,. For mid-air suspension, this must be balanced by the force due to gravity:

$$m\,g = I\,lB$$

$$B = \frac{m\,g}{I\,l}$$

$$= \frac{0.2 \cdot 9.8}{2 \cdot 1.5} = 0.65 \text{ T}$$

Note that it would have been sufficient to specify m/l, the mass per unit length of the wire. The earth's magnetic field is approximately 4×10^{-5} T and we have ignored it.

EXAMPLE 4.1

Example 4.2 If the magnetic field is parallel to the positive y-axis and the charged particle is moving along the positive x-axis (Fig. 4.4), which way would the Lorentz force be for (a) an electron (negative charge), (b) a proton (positive charge).

FIGURE 4.4

Solution The velocity **v** of particle is along the x-axis, while **B**, the magnetic field is along the y-axis, so **v** × **B** is along the z-axis (screw rule or right-hand thumb rule). So, (a) for electron it will be along $-z$ axis. (b) for a positive charge (proton) the force is along $+z$ axis.

EXAMPLE 4.2

PHYSICS

Charged particles moving in a magnetic field. Interactive demonstration:
http://www.phys.hawaii.edu/~teb/optics/java/partmagn/index.html

4.3 Motion in a Magnetic Field

We will now consider, in greater detail, the motion of a charge moving in a magnetic field. We have learnt in Mechanics (see Class XI book, Chapter 6) that a force on a particle does work if the force has a component along (or opposed to) the direction of motion of the particle. In the case of motion

FIGURE 4.5 Circular motion

of a charge in a magnetic field, the magnetic force is perpendicular to the velocity of the particle. So no work is done and no change in the magnitude of the velocity is produced (though the direction of momentum may be changed). [Notice that this is unlike the force due to an electric field, $q\mathbf{E}$, which *can* have a component parallel (or antiparallel) to motion and thus can transfer energy in addition to momentum.]

We shall consider motion of a charged particle in a *uniform* magnetic field. First consider the case of **v** perpendicular to **B**. The perpendicular force, $q\,\mathbf{v} \times \mathbf{B}$, acts as a centripetal force and produces a circular motion perpendicular to the magnetic field. *The particle will describe a circle if* **v** *and* **B** *are perpendicular to each other* (Fig. 4.5).

If velocity has a component along **B**, this component remains unchanged as the motion along the magnetic field will not be affected by the magnetic field. The motion in a plane perpendicular to **B** is as before a circular one, thereby producing a *helical motion* (Fig. 4.6).

FIGURE 4.6 Helical motion

You have already learnt in earlier classes (See Class XI, Chapter 4) that if r is the radius of the circular path of a particle, then a force of $m v^2 / r$, acts perpendicular to the path towards the centre of the circle, and is called the centripetal force. If the velocity **v** is perpendicular to the magnetic field **B,** the magnetic force is perpendicular to both **v** and **B** and acts like a centripetal force. It has a magnitude $q\,v\,B$. Equating the two expressions for centripetal force,

$m v^2/r = q\,v\,B$, which gives

$$r = m v / qB \qquad (4.5)$$

for the radius of the circle described by the charged particle. The larger the momentum, the larger is the radius and bigger the circle described. If ω is the angular frequency, then $v = \omega\, r$. So,

$$\omega = 2\pi\, \nu = q B / m \qquad [4.6(a)]$$

which is independent of the velocity or energy . Here ν is the frequency of rotation. The independence of ν from energy has important application in the design of a cyclotron (see Section 4.4.2).

The time taken for one revolution is $T = 2\pi/\omega \equiv 1/\nu$. If there is a component of the velocity parallel to the magnetic field (denoted by $v_{||}$), it will make the particle move along the field and the path of the particle would be a helical one (Fig. 4.6). The distance moved along the magnetic field in one rotation is called pitch p. Using Eq. [4.6 (a)], we have

$$p = v_{||} T = 2\pi m\, v_{||} / q B \qquad [4.6(b)]$$

The radius of the circular component of motion is called the *radius* of the *helix*.

Example 4.3 What is the radius of the path of an electron (mass 9×10^{-31} kg and charge 1.6×10^{-19} C) moving at a speed of 3×10^{7} m/s in a magnetic field of 6×10^{-4} T perpendicular to it? What is its frequency? Calculate its energy in keV. (1 eV = 1.6×10^{-19} J).

Solution Using Eq. (4.5) we find

$r = mv / (qB) = 9 \times 10^{-31}\ \text{kg} \times 3 \times 10^{7}\ \text{m s}^{-1} / (1.6 \times 10^{-19}\ \text{C} \times 6 \times 10^{-4}\ \text{T})$

$= 26 \times 10^{-2}$ m = 26 cm

$\nu = v / (2\pi r) = 2 \times 10^{6}\ \text{s}^{-1} = 2 \times 10^{6}$ Hz =2 MHz.

$E = (½)mv^2 = (½)\ 9 \times 10^{-31}\ \text{kg} \times 9 \times 10^{14}\ \text{m}^2/\text{s}^2 = 40.5 \times 10^{-17}$ J

$\approx 4 \times 10^{-16}$ J = 2.5 keV.

EXAMPLE 4.3

HELICAL MOTION OF CHARGED PARTICLES AND AURORA BORIOLIS

In polar regions like Alaska and Northern Canada, a splendid display of colours is seen in the sky. The appearance of dancing green pink lights is fascinating, and equally puzzling. An explanation of this natural phenomenon is now found in physics, in terms of what we have studied here.

Consider a charged particle of mass m and charge q, entering a region of magnetic field **B** with an initial velocity **v**. Let this velocity have a component $\mathbf{v}_p$ parallel to the magnetic field and a component $\mathbf{v}_n$ normal to it. There is no force on a charged particle in the direction of the field. Hence the particle continues to travel with the velocity $\mathbf{v}_p$ parallel to the field. The normal component $\mathbf{v}_n$ of the particle results in a Lorentz force $(\mathbf{v}_n \times \mathbf{B})$ which is perpendicular to both $\mathbf{v}_n$ and **B**. As seen in Section 4.3.1 the particle thus has a tendency to perform a circular motion in a plane perpendicular to the magnetic field. When this is coupled with the velocity parallel to the field, the resulting trajectory will be a helix along the magnetic field line, as shown in Figure (a) here. Even if the field line bends, the helically moving particle is trapped and guided to move around the field line. Since the Lorentz force is normal to the velocity of each point, the field does no work on the particle and the magnitude of velocity remains the same.

(a) (b)

During a solar flare, a large number of electrons and protons are ejected from the sun. Some of them get trapped in the earth's magnetic field and move in helical paths along the field lines. The field lines come closer to each other near the magnetic poles; see figure (b). Hence the density of charges increases near the poles. These particles collide with atoms and molecules of the atmosphere. Excited oxygen atoms emit green light and excited nitrogen atoms emits pink light. This phenomenon is called *Aurora Boriolis* in physics.

4.4 Motion in Combined Electric and Magnetic Fields

4.4.1 Velocity selector

You know that a charge q moving with velocity $\mathbf{v}$ in presence of both electric and magnetic fields experiences a force given by Eq. (4.3), that is,

$$\mathbf{F} = q\,(\mathbf{E} + \boldsymbol{v} \times \mathbf{B}) = \mathbf{F}_E + \mathbf{F}_B$$

FIGURE 4.7

We shall consider the simple case in which electric and magnetic fields are perpendicular to each other and also perpendicular to the velocity of the particle, as shown in Fig. 4.7. We have,

$$\mathbf{E} = E\,\hat{\mathbf{j}},\ \mathbf{B} = B\,\hat{\mathbf{k}},\ \mathbf{v} = v\,\hat{\mathbf{i}}$$

$$\mathbf{F}_E = q\mathbf{E} = qE\,\hat{\mathbf{j}},\ \mathbf{F}_B = q\mathbf{v} \times \mathbf{B},\ = q\left(v\,\hat{\mathbf{i}} \times B\hat{\mathbf{k}}\right) = -qB\,\hat{\mathbf{j}}$$

Therefore, $\mathbf{F} = q\left(E - vB\right)\hat{\mathbf{j}}$.

Thus, electric and magnetic forces are in opposite directions as shown in the figure. Suppose, we adjust the value of $\mathbf{E}$ and $\mathbf{B}$ such that magnitude of the two forces are equal. Then, total force on the charge is zero and the charge will move in the fields undeflected. This happens when,

$$qE = qvB \quad \text{or} \quad v = \frac{E}{B} \tag{4.7}$$

This condition can be used to select charged particles of a particular velocity out of a beam containing charges moving with different speeds (irrespective of their charge and mass). The crossed E and B fields, therefore, serve as a *velocity selector*. Only particles with speed E/B pass undeflected through the region of crossed fields. This method was employed by J. J. Thomson in 1897 to measure the charge to mass ratio (e/m) of an electron. The principle is also employed in Mass Spectrometer – a device that separates charged particles, usually ions, according to their charge to mass ratio.

4.4.2 Cyclotron

Cyclotron
Interactive demonstration:
http://www.phy.ntnu.edu.tw/ntnujava/viewtopic.php?t=50

The cyclotron is a machine to accelerate charged particles or ions to high energies. It was invented by E.O. Lawrence and M.S. Livingston in 1934 to investigate nuclear structure. The cyclotron uses both electric and magnetic fields in combination to increase the energy of charged particles. As the fields are perpendicular to each other they are called *crossed fields*. Cyclotron uses the fact that the frequency of revolution of the charged particle in a magnetic field is independent of its energy. The particles move most of the time inside two semicircular disc-like metal containers, D_1 and D_2, which are called *dees* as they look like the letter D. Figure 4.8 shows a schematic view of the cyclotron. Inside the metal boxes the particle is shielded and is not acted on by the electric field. The magnetic field, however, acts on the particle and makes it go round in a circular path inside a dee. Every time the particle moves from one dee to another it is acted upon by the electric field. The sign of the electric field is changed alternately in tune with the circular motion of the particle. This ensures that the particle is always accelerated by the electric field. Each time the acceleration increases the energy of the particle. As energy

increases, the radius of the circular path increases. So the path is a spiral one.

The whole assembly is evacuated to minimise collisions between the ions and the air molecules. A high frequency alternating voltage is applied to the dees. In the sketch shown in Fig. 4.8, positive ions or positively charged particles (e.g., protons) are released at the centre P. They move in a semi-circular path in one of the dees and arrive in the gap between the dees in a time interval $T/2$; where T, the period of revolution, is given by Eq. (4.6),

$$T = \frac{1}{\nu_c} = \frac{2\pi m}{qB}$$

$$\text{or } \nu_c = \frac{qB}{2\pi m} \tag{4.8}$$

This frequency is called the *cyclotron frequency* for obvious reasons and is denoted by ν_c.

The frequency ν_a of the applied voltage is adjusted so that the polarity of the dees is reversed in the same time that it takes the ions to complete one half of the revolution. The requirement $\nu_a = \nu_c$ is called the *resonance condition*. The phase of the supply is adjusted so that when the positive ions arrive at the edge of D_1, D_2 is at a lower potential and the ions are accelerated across the gap. Inside the dees the particles travel in a region free of the electric field. The increase in their kinetic energy is qV each time they cross from one dee to another (V refers to the voltage across the dees at that time). From Eq. (4.5), it is clear that the radius of their path goes on increasing each time their kinetic energy increases. The ions are repeatedly accelerated across the dees until they have the required energy to have a radius approximately that of the dees. They are then deflected by a magnetic field and leave the system via an exit slit. From Eq. (4.5) we have,

$$v = \frac{qBR}{m} \tag{4.9}$$

where R is the radius of the trajectory at exit, and equals the radius of a dee.

Hence, the kinetic energy of the ions is,

$$\frac{1}{2}mv^2 = \frac{q^2B^2R^2}{2m} \tag{4.10}$$

The operation of the cyclotron is based on the fact that the time for one revolution of an ion is independent of its speed or radius of its orbit. The cyclotron is used to bombard nuclei with energetic particles, so accelerated by it, and study

FIGURE 4.8 A schematic sketch of the cyclotron. There is a source of charged particles or ions at P which move in a circular fashion in the dees, D_1 and D_2, on account of a uniform perpendicular magnetic field B. An alternating voltage source accelerates these ions to high speeds. The ions are eventually 'extracted' at the exit port.

the resulting nuclear reactions. It is also used to implant ions into solids and modify their properties or even synthesise new materials. It is used in hospitals to produce radioactive substances which can be used in diagnosis and treatment.

EXAMPLE 4.4

Example 4.4 A cyclotron's oscillator frequency is 10 MHz. What should be the operating magnetic field for accelerating protons? If the radius of its 'dees' is 60 cm, what is the kinetic energy (in MeV) of the proton beam produced by the accelerator.

($e = 1.60 \times 10^{-19}$ C, $m_p = 1.67 \times 10^{-27}$ kg, 1 MeV = 1.6×10^{-13} J).

Solution The oscillator frequency should be same as proton's cyclotron frequency.

Using Eqs. (4.5) and [4.6(a)] we have

$B = 2\pi\, m\, v/q = 6.3 \times 1.67 \times 10^{-27} \times 10^{7} / (1.6 \times 10^{-19}) = 0.66$ T

Final velocity of protons is

$v = r \times 2\pi\, v = 0.6 \text{ m} \times 6.3 \times 10^{7} = 3.78 \times 10^{7}$ m/s.

$E = \frac{1}{2} mv^2 = 1.67 \times 10^{-27} \times 14.3 \times 10^{14} / (2 \times 1.6 \times 10^{-13}) = 7$ MeV.

ACCELERATORS IN INDIA

India has been an early entrant in the area of accelerator- based research. The vision of Dr. Meghnath Saha created a 37" Cyclotron in the Saha Institute of Nuclear Physics in Kolkata in 1953. This was soon followed by a series of Cockroft-Walton type of accelerators established in Tata Institute of Fundamental Research (TIFR), Mumbai, Aligarh Muslim University (AMU), Aligarh, Bose Institute, Kolkata and Andhra University, Waltair.

The sixties saw the commissioning of a number of Van de Graaff accelerators: a 5.5 MV terminal machine in Bhabha Atomic Research Centre (BARC), Mumbai (1963); a 2 MV terminal machine in Indian Institute of Technology (IIT), Kanpur; a 400 kV terminal machine in Banaras Hindu University (BHU), Varanasi; and Punjabi University, Patiala. One 66 cm Cyclotron donated by the Rochester University of USA was commissioned in Panjab University, Chandigarh. A small electron accelerator was also established in University of Pune, Pune.

In a major initiative taken in the seventies and eighties, a Variable Energy Cyclotron was built indigenously in Variable Energy Cyclotron Centre (VECC), Kolkata; 2 MV Tandem Van de Graaff accelerator was developed and built in BARC and a 14 MV Tandem Pelletron accelerator was installed in TIFR.

This was soon followed by a 15 MV Tandem Pelletron established by University Grants Commission (UGC), as an inter-university facility in Inter-University Accelerator Centre (IUAC), New Delhi; a 3 MV Tandem Pelletron in Institute of Physics, Bhubaneshwar; and two 1.7 MV Tandetrons in Atomic Minerals Directorate for Exploration and Research, Hyderabad and Indira Gandhi Centre for Atomic Research, Kalpakkam. Both TIFR and IUAC are augmenting their facilities with the addition of superconducting LINAC modules to accelerate the ions to higher energies.

Besides these ion accelerators, the Department of Atomic Energy (DAE) has developed many electron accelerators. A 2 GeV Synchrotron Radiation Source is being built in Raja Ramanna Centre for Advanced Technologies, Indore.

The Department of Atomic Energy is considering Accelerator Driven Systems (ADS) for power production and fissile material breeding as future options.

4.5 Magnetic Field due to a Current Element, Biot-Savart Law

All magnetic fields that we know are due to currents (or moving charges) and due to intrinsic magnetic moments of particles. Here, we shall study the relation between current and the magnetic field it produces. It is given by the Biot-Savart's law. Figure 4.9 shows a finite conductor XY carrying current I. Consider an infinitesimal element dl of the conductor. The magnetic field d**B** due to this element is to be determined at a point P which is at a distance r from it. Let θ be the angle between d**l** and the displacement vector **r**. According to Biot-Savart's law, the magnitude of the magnetic field d**B** is proportional to the current I, the element length |d**l**|, and inversely proportional to the square of the distance r. Its direction* is perpendicular to the plane containing d**l** and **r**. Thus, in vector notation,

FIGURE 4.9 Illustration of the Biot-Savart law. The current element I d**l** produces a field d**B** at a distance r. The ⊗ sign indicates that the field is perpendicular to the plane of this page and directed into it.

$$d\mathbf{B} \propto \frac{I\,d\mathbf{l}\cdot\mathbf{r}}{r^3}$$

$$= \frac{\mu_0}{4\pi}\frac{I\,d\mathbf{l}\cdot\mathbf{r}}{r^3} \qquad [4.11(a)]$$

where $\mu_0/4\pi$ is a constant of proportionality. The above expression holds when the medium is vacuum.

The magnitude of this field is,

$$|d\mathbf{B}| = \frac{\mu_0}{4\pi}\frac{I\,dl\sin\theta}{r^2} \qquad [4.11(b)]$$

where we have used the property of cross-product. Equation [4.11 (a)] constitutes our basic equation for the magnetic field. The proportionality constant in SI units has the exact value,

$$\frac{\mu_0}{4\pi} = 10^{-7}\,\text{Tm/A} \qquad [4.11(c)]$$

We call μ_0 the *permeability* of free space (or vacuum).

The Biot-Savart law for the magnetic field has certain similarities as well as differences with the Coulomb's law for the electrostatic field. Some of these are:

(i) Both are long range, since both depend inversely on the square of distance from the source to the point of interest. The principle of superposition applies to both fields. [In this connection, note that the magnetic field is *linear* in the *source* I d**l** just as the electrostatic field is linear in its source: the electric charge.]

(ii) The electrostatic field is produced by a scalar source, namely, the electric charge. The magnetic field is produced by a vector source I d**l**.

* The sense of **dl×r** is also given by the *Right Hand Screw rule* : Look at the plane containing vectors **dl** and **r**. Imagine moving from the first vector towards second vector. If the movement is anticlockwise, the resultant is towards you. If it is clockwise, the resultant is away from you.

(iii) The electrostatic field is along the displacement vector joining the source and the field point. The magnetic field is perpendicular to the plane containing the displacement vector **r** and the current element $I\,d\mathbf{l}$.

(iv) There is an angle dependence in the Biot-Savart law which is not present in the electrostatic case. In Fig. 4.9, the magnetic field at any point in the direction of $d\mathbf{l}$ (the dashed line) is zero. Along this line, $\theta = 0$, $\sin\theta = 0$ and from Eq. [4.11(a)], $|d\mathbf{B}| = 0$.

There is an interesting relation between ε_0, the permittivity of free space; $\propto_0$, the permeability of free space; and c, the speed of light in vacuum:

$$\varepsilon_0 \propto_0 = (4\pi\varepsilon_0)\left(\frac{\propto_0}{4\pi}\right) = \left(\frac{1}{9\cdot 10^9}\right)(10^{-7}) = \frac{1}{(3\cdot 10^8)^2} = \frac{1}{c^2}$$

We will discuss this connection further in Chapter 8 on the electromagnetic waves. Since the speed of light in vacuum is constant, the product $\propto_0\varepsilon_0$ is fixed in magnitude. Choosing the value of either ε_0 or $\propto_0$, fixes the value of the other. In SI units, $\propto_0$ is fixed to be equal to $4\pi \times 10^{-7}$ in magnitude.

EXAMPLE 4.5

Example 4.5 An element $\Delta\mathbf{l} = \Delta x\,\hat{\mathbf{i}}$ is placed at the origin and carries a large current I = 10 A (Fig. 4.10). What is the magnetic field on the y-axis at a distance of 0.5 m. Δx = 1 cm.

FIGURE 4.10

Solution

$$|d\mathbf{B}| = \frac{\propto_0}{4\pi}\frac{I\,dl\sin\theta}{r^2} \quad \text{[using Eq. (4.11)]}$$

$dl = \Delta x = 10^{-2}$ m, I = 10 A, r = 0.5 m = y, $\propto_0/4\pi = 10^{-7}\,\dfrac{\text{T m}}{\text{A}}$

$\theta = 90°$; $\sin\theta = 1$

$$|d\mathbf{B}| = \frac{10^{-7}\cdot 10\cdot 10^{-2}}{25\cdot 10^{-2}} = 4\times 10^{-8}\ \text{T}$$

The direction of the field is in the +z-direction. This is so since,

$$d\mathbf{l}\times\mathbf{r} = \Delta x\,\hat{\mathbf{i}}\times y\,\hat{\mathbf{j}} = y\,\Delta x\left(\hat{\mathbf{i}}\times\hat{\mathbf{j}}\right) = y\,\Delta x\,\hat{\mathbf{k}}$$

We remind you of the following cyclic property of cross-products,

$\hat{\mathbf{i}}\cdot\hat{\mathbf{j}} = \hat{\mathbf{k}}$; $\hat{\mathbf{j}}\cdot\hat{\mathbf{k}} = \hat{\mathbf{i}}$; $\hat{\mathbf{k}}\cdot\hat{\mathbf{i}} = \hat{\mathbf{j}}$

Note that the field is small in magnitude.

In the next section, we shall use the Biot-Savart law to calculate the magnetic field due to a circular loop.

4.6 Magnetic Field on the Axis of a Circular Current Loop

In this section, we shall evaluate the magnetic field due to a circular coil along its axis. The evaluation entails summing up the effect of infinitesimal current elements ($I\,d\mathbf{l}$) mentioned in the previous section. We assume that the current I is steady and that the evaluation is carried out in free space (i.e., vacuum).

Figure 4.11 depicts a circular loop carrying a steady current I. The loop is placed in the y-z plane with its centre at the origin O and has a radius R. The x-axis is the axis of the loop. We wish to calculate the magnetic field at the point P on this axis. Let x be the distance of P from the centre O of the loop.

FIGURE 4.11 Magnetic field on the axis of a current carrying circular loop of radius R. Shown are the magnetic field $d\mathbf{B}$ (due to a line element $d\mathbf{l}$) and its components along and perpendicular to the axis.

Consider a conducting element $d\mathbf{l}$ of the loop. This is shown in Fig. 4.11. The magnitude dB of the magnetic field due to $d\mathbf{l}$ is given by the Biot-Savart law [Eq. 4.11(a)],

$$dB = \frac{\mu_0}{4\pi}\frac{I\,|d\mathbf{l}\times\mathbf{r}|}{r^3} \tag{4.12}$$

Now $r^2 = x^2 + R^2$. Further, any element of the loop will be perpendicular to the displacement vector from the element to the axial point. For example, the element $d\mathbf{l}$ in Fig. 4.11 is in the y-z plane whereas the displacement vector $\mathbf{r}$ from $d\mathbf{l}$ to the axial point P is in the x-y plane. Hence $|d\mathbf{l}\times\mathbf{r}| = r\,dl$. Thus,

$$dB = \frac{\mu_0}{4\pi}\frac{I\,dl}{\left(x^2+R^2\right)} \tag{4.13}$$

The direction of $d\mathbf{B}$ is shown in Fig. 4.11. It is perpendicular to the plane formed by $d\mathbf{l}$ and $\mathbf{r}$. It has an x-component $d\mathbf{B}_x$ and a component perpendicular to x-axis, $d\mathbf{B}_\perp$. When the components perpendicular to the x-axis are summed over, they cancel out and we obtain a null result. For example, the $d\mathbf{B}_\perp$ component due to $d\mathbf{l}$ is cancelled by the contribution due to the diametrically opposite $d\mathbf{l}$ element, shown in Fig. 4.11. Thus, only the x-component survives. The net contribution along x-direction can be obtained by integrating $dB_x = dB\cos\theta$ over the loop. For Fig. 4.11,

$$\cos\theta = \frac{R}{(x^2+R^2)^{1/2}} \tag{4.14}$$

From Eqs. (4.13) and (4.14),

$$dB_x = \frac{\mu_0 I\,dl}{4\pi}\frac{R}{\left(x^2+R^2\right)^{3/2}}$$

The summation of elements dl over the loop yields $2\pi R$, the circumference of the loop. Thus, the magnetic field at P due to entire circular loop is

$$\mathbf{B} = B_x \hat{\mathbf{i}} = \frac{\mu_0 I R^2}{2\left(x^2 + R^2\right)^{3/2}} \hat{\mathbf{i}} \tag{4.15}$$

As a special case of the above result, we may obtain the field at the centre of the loop. Here $x = 0$, and we obtain,

$$\mathbf{B}_0 = \frac{\mu_0 I}{2R} \hat{\mathbf{i}} \tag{4.16}$$

The magnetic field lines due to a circular wire form closed loops and are shown in Fig. 4.12. The direction of the magnetic field is given by (another) *right-hand thumb rule* stated below:

Curl the palm of your right hand around the circular wire with the fingers pointing in the direction of the current. The right-hand thumb gives the direction of the magnetic field.

FIGURE 4.12 The magnetic field lines for a current loop. The direction of the field is given by the right-hand thumb rule described in the text. The upper side of the loop may be thought of as the north pole and the lower side as the south pole of a magnet.

EXAMPLE 4.6

Example 4.6 A straight wire carrying a current of 12 A is bent into a semi-circular arc of radius 2.0 cm as shown in Fig. 4.13(a). Consider the magnetic field **B** at the centre of the arc. (a) What is the magnetic field due to the straight segments? (b) In what way the contribution to **B** from the semicircle differs from that of a circular loop and in what way does it resemble? (c) Would your answer be different if the wire were bent into a semi-circular arc of the same radius but in the opposite way as shown in Fig. 4.13(b)?

FIGURE 4.13

Solution

(a) d**l** and **r** for each element of the straight segments are parallel. Therefore, d**l** × **r** = 0. Straight segments do not contribute to |**B**|.

(b) For all segments of the semicircular arc, d**l** × **r** are all parallel to each other (into the plane of the paper). All such contributions add up in magnitude. Hence direction of **B** for a semicircular arc is given by the right-hand rule and magnitude is half that of a circular loop. Thus **B** is 1.9×10^{-4} T normal to the plane of the paper going into it.

(c) Same magnitude of **B** but opposite in direction to that in (b).

EXAMPLE 4.6

Example 4.7 Consider a tightly wound 100 turn coil of radius 10 cm, carrying a current of 1 A. What is the magnitude of the magnetic field at the centre of the coil?

Solution Since the coil is tightly wound, we may take each circular element to have the same radius R = 10 cm = 0.1 m. The number of turns N = 100. The magnitude of the magnetic field is,

$$B = \frac{\propto_0 NI}{2R} = \frac{4\pi \cdot 10^{-7} \cdot 10^{2} \cdot 1}{2 \cdot 10^{-1}} = 2\pi \cdot 10^{-4} = 6.28 \cdot 10^{-4} \text{ T}$$

EXAMPLE 4.7

4.7 Ampere's Circuital Law

There is an alternative and appealing way in which the Biot-Savart law may be expressed. Ampere's circuital law considers an open surface with a boundary (Fig. 4.14). The surface has current passing through it. We consider the boundary to be made up of a number of small line elements. Consider one such element of length dl. We take the value of the tangential component of the magnetic field, B_t, at this element and multiply it by the length of that element dl [Note: $B_t dl = \mathbf{B} \cdot d\mathbf{l}$]. All such products are added together. We consider the limit as the lengths of elements get smaller and their number gets larger. The sum then tends to an integral. Ampere's law states that this integral is equal to $\propto_0$ times the total current passing through the surface, i.e.,

FIGURE 4.14

$$\oint \mathbf{B} \cdot d\mathbf{l} = \propto_0 I \qquad [4.17(a)]$$

where I is the total current through the surface. The integral is taken over the closed loop coinciding with the boundary C of the surface. The relation above involves a sign-convention, given by the right-hand rule. Let the fingers of the right-hand be curled in the sense the boundary is traversed in the loop integral $\oint \mathbf{B} \cdot d\mathbf{l}$. Then the direction of the thumb gives the sense in which the current I is regarded as positive.

For several applications, a much simplified version of Eq. [4.17(a)] proves sufficient. We shall assume that, in such cases, it is possible to choose the loop (called an *amperian loop*) such that at each point of the loop, *either*

Andre Ampere (1775 – 1836) Andre Marie Ampere was a French physicist, mathematician and chemist who founded the science of electrodynamics. Ampere was a child prodigy who mastered advanced mathematics by the age of 12. Ampere grasped the significance of Oersted's discovery. He carried out a large series of experiments to explore the relationship between current electricity and magnetism. These investigations culminated in 1827 with the publication of the 'Mathematical Theory of Electrodynamic Phenomena Deduced Solely from Experiments'. He hypothesised that *all* magnetic phenomena are due to circulating electric currents. Ampere was humble and absent-minded. He once forgot an invitation to dine with the Emperor Napoleon. He died of pneumonia at the age of 61. His gravestone bears the epitaph: *Tandem Felix* (Happy at last).

ANDRE AMPERE (1775 –1836)

(i) **B** is tangential to the loop and is a non-zero *constant* B, *or*

(ii) **B** is normal to the loop, *or*

(iii) **B** vanishes.

Now, let L be the length (part) of the loop for which **B** is tangential. Let I_e be the current enclosed by the loop. Then, Eq. (4.17) reduces to,

$$BL = \propto_0 I_e \qquad [4.17(b)]$$

When there is a system with a symmetry such as for a *straight infinite current-carrying wire* in Fig. 4.15, the Ampere's law enables an easy evaluation of the magnetic field, much the same way Gauss' law helps in determination of the electric field. This is exhibited in the Example 4.8 below. The boundary of the loop chosen is a circle and magnetic field is tangential to the circumference of the circle. The law gives, for the left hand side of Eq. [4.17 (b)], B. $2\pi r$. We find that the magnetic field at a distance r outside the wire is *tangential* and given by

$$B \times 2\pi r = \propto_0 I,$$

$$B = \propto_0 I/(2\pi r) \qquad (4.18)$$

The above result for the infinite wire is interesting from several points of view.

(i) It implies that the field at every point on a circle of radius r, (with the wire along the axis), is same in magnitude. In other words, the magnetic field possesses what is called a *cylindrical symmetry*. The field that normally can depend on three coordinates depends only on one: r. Whenever there is symmetry, the solutions simplify.

(ii) The field direction at any point on this circle is tangential to it. Thus, the lines of constant magnitude of magnetic field form concentric circles. Notice now, in Fig. 4.1(c), the iron filings form concentric circles. These lines called *magnetic field lines* form closed loops. This is unlike the electrostatic field lines which originate from positive charges and end at negative charges. The expression for the magnetic field of a straight wire provides a theoretical justification to Oersted's experiments.

(iii) Another interesting point to note is that even though the wire is infinite, the field due to it at a nonzero distance is *not* infinite. It tends to blow up only when we come very close to the wire. The field is directly proportional to the current and inversely proportional to the distance from the (infinitely long) current source.

(iv) There exists a simple rule to determine the direction of the magnetic field due to a long wire. This rule, called the *right-hand rule**, is:

Grasp the wire in your right hand with your extended thumb pointing in the direction of the current. Your fingers will curl around in the direction of the magnetic field.

Ampere's circuital law is not new in content from Biot-Savart law. Both relate the magnetic field and the current, and both express the same physical consequences of a steady electrical current. Ampere's law is to Biot-Savart law, what Gauss's law is to Coulomb's law. Both, Ampere's and Gauss's law relate a physical quantity on the periphery or boundary (magnetic or electric field) to another physical quantity, namely, the source, in the interior (current or charge). We also note that Ampere's circuital law holds for steady currents which do not fluctuate with time. The following example will help us understand what is meant by the term *enclosed* current.

Example 4.8 Figure 4.15 shows a long straight wire of a circular cross-section (radius a) carrying steady current I. The current I is uniformly distributed across this cross-section. Calculate the magnetic field in the region $r < a$ and $r > a$.

FIGURE 4.15

Solution (a) Consider the case $r > a$. The Amperian loop, labelled 2, is a circle concentric with the cross-section. For this loop,

$L = 2\pi r$

I_e = Current enclosed by the loop = I

The result is the familiar expression for a long straight wire

$B(2\pi r) = \propto_0 I$

$$B = \frac{\propto_0 I}{2\pi r} \qquad [4.19(a)]$$

$$B \propto \frac{1}{r} \qquad (r > a)$$

(b) Consider the case $r < a$. The Amperian loop is a circle labelled 1. For this loop, taking the radius of the circle to be r,

$L = 2\pi r$

EXAMPLE 4.8

* Note that there are *two distinct* right-hand rules: One which gives the direction of **B** on the axis of current-loop and the other which gives direction of **B** for a straight conducting wire. Fingers and thumb play different roles in the two.

EXAMPLE 4.8

Now the current enclosed I_e is not I, but is less than this value. Since the current distribution is uniform, the current enclosed is,

$$I_e = I\left(\frac{\pi r^2}{\pi a^2}\right) = \frac{Ir^2}{a^2}$$

Using Ampere's law, $B(2\pi r) = \mu_0 \frac{I r^2}{a^2}$

$$B = \left(\frac{\mu_0 I}{2\pi a^2}\right) r \qquad [4.19(b)]$$

$B \propto r \qquad (r < a)$

FIGURE 4.16

Figure (4.16) shows a plot of the magnitude of **B** with distance r from the centre of the wire. The direction of the field is tangential to the respective circular loop (1 or 2) and given by the right-hand rule described earlier in this section.

This example possesses the required symmetry so that Ampere's law can be applied readily.

It should be noted that while Ampere's circuital law holds for any loop, it may not always facilitate an evaluation of the magnetic field in every case. For example, for the case of the circular loop discussed in Section 4.6, it cannot be applied to extract the simple expression $B = \mu_0 I/2R$ [Eq. (4.16)] for the field at the centre of the loop. However, there exists a large number of situations of high symmetry where the law can be conveniently applied. We shall use it in the next section to calculate the magnetic field produced by two commonly used and very useful magnetic systems: the *solenoid* and the *toroid*.

4.8 THE SOLENOID AND THE TOROID

The solenoid and the toroid are two pieces of equipment which generate magnetic fields. The television uses the solenoid to generate magnetic fields needed. The synchrotron uses a combination of both to generate the high magnetic fields required. In both, solenoid and toroid, we come across a situation of high symmetry where Ampere's law can be conveniently applied.

4.8.1 The solenoid

We shall discuss a long solenoid. By long solenoid we mean that the solenoid's length is large compared to its radius. It consists of a long wire wound in the form of a helix where the neighbouring turns are closely spaced. So each turn can be regarded as a circular loop. The net magnetic field is the vector sum of the fields due to all the turns. Enamelled wires are used for winding so that turns are insulated from each other.

FIGURE 4.17 (a) The magnetic field due to a section of the solenoid which has been stretched out for clarity. Only the exterior semi-circular part is shown. Notice how the circular loops between neighbouring turns tend to cancel. (b) The magnetic field of a finite solenoid.

Figure 4.17 displays the magnetic field lines for a finite solenoid. We show a section of this solenoid in an enlarged manner in Fig. 4.17(a). Figure 4.17(b) shows the entire finite solenoid with its magnetic field. In Fig. 4.17(a), it is clear from the circular loops that the field between two neighbouring turns vanishes. In Fig. 4.17(b), we see that the field at the interior mid-point P is uniform, strong and along the axis of the solenoid. The field at the exterior mid-point Q is weak and moreover is along the axis of the solenoid with no perpendicular or normal component. As the solenoid is made longer it appears like a long cylindrical metal sheet. Figure 4.18 represents this idealised picture. The field outside the solenoid approaches zero. We shall assume that the field outside is zero. The field inside becomes everywhere parallel to the axis.

FIGURE 4.18 The magnetic field of a very long solenoid. We consider a rectangular Amperian loop abcd to determine the field.

Consider a rectangular Amperian loop abcd. Along cd the field is zero as argued above. Along transverse sections bc and ad, the field component is zero. Thus, these two sections make no contribution. Let the field along ab be B. Thus, the relevant length of the Amperian loop is, $L = h$.

Let n be the number of turns per unit length, then the total number of turns is nh. The enclosed current is, $I_e = I\,(n\,h)$, where I is the current in the solenoid. From Ampere's circuital law [Eq. 4.17 (b)]

$$BL = \propto_0 I_e, \quad B\,h = \propto_0 I\,(n\,h)$$

$$B = \propto_0 n\,I \tag{4.20}$$

The direction of the field is given by the right-hand rule. The solenoid is commonly used to obtain a uniform magnetic field. We shall see in the next chapter that a large field is possible by inserting a soft iron core inside the solenoid.

FIGURE 4.19 (a) A toroid carrying a current I. (b) A sectional view of the toroid. The magnetic field can be obtained at an arbitrary distance r from the centre O of the toroid by Ampere's circuital law. The dashed lines labelled 1, 2 and 3 are three circular Amperian loops.

4.8.2 The toroid

The toroid is a hollow circular ring on which a large number of turns of a wire are closely wound. It can be viewed as a solenoid which has been bent into a circular shape to close on itself. It is shown in Fig. 4.19(a) carrying a current I. We shall see that the magnetic field in the open space inside (point P) and exterior to the toroid (point Q) is zero. The field **B** inside the toroid is constant in magnitude for the *ideal* toroid of closely wound turns.

Figure 4.19(b) shows a sectional view of the toroid. The direction of the magnetic field inside is clockwise as per the right-hand thumb rule for circular loops. Three circular Amperian loops 1, 2 and 3 are shown by dashed lines. By symmetry, the magnetic field should be tangential to each of them and constant in magnitude for a given loop. The circular areas bounded by loops 2 and 3 both cut the toroid: so that each turn of current carrying wire is cut once by the loop 2 and twice by the loop 3.

Let the magnetic field along loop 1 be B_1 in magnitude. Then in Ampere's circuital law [Eq. 4.17(a)], $L = 2\pi\, r_1$. However, the loop encloses no current, so $I_e = 0$. Thus,

$$B_1\,(2\,\pi\,r_1) = \propto_0 (0), \quad B_1 = 0$$

Thus, the magnetic field at any point P in the open space inside the toroid is zero.

We shall now show that magnetic field at Q is likewise zero. Let the magnetic field along loop 3 be B_3. Once again from Ampere's law $L = 2\,\pi\,r_3$. However, from the sectional cut, we see that the current coming out of the plane of the paper is cancelled exactly by the current going into it. Thus, $I_e = 0$, and $B_3 = 0$. Let the magnetic field inside the solenoid be B. We shall now consider the magnetic field at S. Once again we employ Ampere's law in the form of Eq. [4.17 (a)]. We find, $L = 2\pi\, r$.

The current enclosed I_e is (for N turns of toroidal coil) $N\,I$.

$$B\,(2\pi r) = \propto_0 NI$$

$$B = \frac{\propto_0 NI}{2\pi r} \tag{4.21}$$

We shall now compare the two results: for a toroid and solenoid. We re-express Eq. (4.21) to make the comparison easier with the solenoid result given in Eq. (4.20). Let r be the average radius of the toroid and n be the number of turns per unit length. Then

$N = 2\pi r\, n$ = *(average) perimeter of the toroid* × *number of turns per unit length*

and thus,

$$B = \propto_0 n\, I, \tag{4.22}$$

i.e., the result for the solenoid!

In an ideal toroid the coils are circular. In reality the turns of the toroidal coil form a helix and there is always a small magnetic field external to the toroid.

MAGNETIC CONFINEMENT

We have seen in Section 4.3 (see also the box on helical motion of charged particles earlier in this chapter) that orbits of charged particles are helical. If the magnetic field is non-uniform, but does not change much during one circular orbit, then the radius of the helix will decrease as it enters stronger magnetic field and the radius will increase when it enters weaker magnetic fields. We consider two solenoids at a distance from each other, enclosed in an evacuated container (see figure below where we have not shown the container). Charged particles moving in the region between the two solenoids will start with a small radius. The radius will increase as field decreases and the radius will decrease again as field due to the second solenoid takes over. The solenoids act as a mirror or reflector. [See the direction of **F** as the particle approaches coil 2 in the figure. It has a horizontal component against the forward motion.] This makes the particles turn back when they approach the solenoid. Such an arrangement will act like *magnetic bottle* or magnetic container. The particles will never touch the sides of the container. Such magnetic bottles are of great use in confining the high energy plasma in fusion experiments. The plasma will destroy any other form of material container because of it's high temperature. Another useful container is a toroid. Toroids are expected to play a key role in the *tokamak*, an equipment for plasma confinement in fusion power reactors. There is an international collaboration called the *International Thermonuclear Experimental Reactor* (ITER), being set up in France, for achieving controlled fusion, of which India is a collaborating nation. For details of ITER collaboration and the project, you may visit http://www.iter.org.

EXAMPLE 4.9

Example 4.9 A solenoid of length 0.5 m has a radius of 1 cm and is made up of 500 turns. It carries a current of 5 A. What is the magnitude of the magnetic field inside the solenoid?

Solution The number of turns per unit length is,

$$n = \frac{500}{0.5} = 1000 \text{ turns/m}$$

The length l = 0.5 m and radius r = 0.01 m. Thus, l/a = 50 i.e., $l >> a$. Hence, we can use the *long* solenoid formula, namely, Eq. (4.20)

$$B = \propto_0 n I$$
$$= 4\pi \times 10^{-7} \times 10^3 \times 5$$
$$= 6.28 \times 10^{-3} \text{ T}$$

4.9 Force between Two Parallel Currents, the Ampere

We have learnt that there exists a magnetic field due to a conductor carrying a current which obeys the Biot-Savart law. Further, we have learnt that an external magnetic field will exert a force on a current-carrying conductor. This follows from the Lorentz force formula. Thus, it is logical to expect that two current-carrying conductors placed near each other will exert (magnetic) forces on each other. In the period 1820-25, Ampere studied the nature of this magnetic force and its dependence on the magnitude of the current, on the shape and size of the conductors as well as the distances between the conductors. In this section, we shall take the simple example of two parallel current-carrying conductors, which will perhaps help us to appreciate Ampere's painstaking work.

FIGURE 4.20 Two long straight parallel conductors carrying steady currents I_a and I_b and separated by a distance d. $\mathbf{B}_a$ is the magnetic field set up by conductor 'a' at conductor 'b'.

Figure 4.20 shows two long parallel conductors a and b separated by a distance d and carrying (parallel) currents I_a and I_b, respectively. The conductor 'a' produces, the same magnetic field $\mathbf{B}_a$ at all points along the conductor 'b'. The right-hand rule tells us that the direction of this field is downwards (when the conductors are placed horizontally). Its magnitude is given by Eq. [4.19(a)] or from Ampere's circuital law,

$$B_a = \frac{\propto_0 I_a}{2\pi d}$$

The conductor 'b' carrying a current I_b will experience a sideways force due to the field $\mathbf{B}_a$. The direction of this force is towards the conductor 'a' (Verify this). We label this force as $\mathbf{F}_{ba}$, the force on a segment L of 'b' due to 'a'. The magnitude of this force is given by Eq. (4.4),

$$F_{ba} = I_b L B_a$$

$$= \frac{\mu_0 I_a I_b}{2\pi d} L \qquad (4.23)$$

It is of course possible to compute the force on 'a' due to 'b'. From considerations similar to above we can find the force $\mathbf{F}_{ab}$, on a segment of length L of 'a' due to the current in 'b'. It is equal in magnitude to F_{ba}, and directed towards 'b'. Thus,

$$\mathbf{F}_{ba} = -\mathbf{F}_{ab} \qquad (4.24)$$

Note that this is consistent with Newton's third Law. Thus, at least for parallel conductors and steady currents, we have shown that the Biot-Savart law and the Lorentz force yield results in accordance with Newton's third Law*.

We have seen from above that currents flowing in the same direction attract each other. One can show that oppositely directed currents repel each other. Thus,

Parallel currents attract, and antiparallel currents repel.

This rule is the opposite of what we find in electrostatics. Like (same sign) charges repel each other, but like (parallel) currents attract each other.

Let f_{ba} represent the magnitude of the force $\mathbf{F}_{ba}$ per unit length. Then, from Eq. (4.23),

$$f_{ba} = \frac{\mu_0 I_a I_b}{2\pi d} \qquad (4.25)$$

The above expression is used to define the ampere (A), which is one of the seven SI base units.

The *ampere* is the value of that steady current which, when maintained in each of the two very long, straight, parallel conductors of negligible cross-section, and placed one metre apart in vacuum, would produce on each of these conductors a force equal to 2×10^{-7} newtons per metre of length.

This definition of the ampere was adopted in 1946. It is a theoretical definition. In practice one must eliminate the effect of the earth's magnetic field and substitute very long wires by multiturn coils of appropriate geometries. An instrument called the current balance is used to measure this mechanical force.

The SI unit of charge, namely, the coulomb, can now be defined in terms of the ampere.

When a steady current of 1A is set up in a conductor, the quantity of charge that flows through its cross-section in 1s is one coulomb (1C).

---

* It turns out that when we have time-dependent currents and/or charges in motion, Newton's third law may not hold for forces between charges and/or conductors. An essential consequence of the Newton's third law in mechanics is conservation of momentum of an isolated system. This, however, holds even for the case of time-dependent situations with electromagnetic fields, provided the momentum carried by fields is also taken into account.

ROGET'S SPIRAL FOR ATTRACTION BETWEEN PARALLEL CURRENTS

Magnetic effects are generally smaller than electric effects. As a consequence, the force between currents is rather small, because of the smallness of the factor $\propto$. Hence it is difficult to demonstrate attraction or repulsion between currents. Thus for 5 A current in each wire at a separation of 1cm, the force per metre would be 5×10^{-4} N, which is about 50 mg weight. It would be like pulling a wire by a string going over a pulley to which a 50 mg weight is attached. The displacement of the wire would be quite unnoticeable.

With the use of a soft spring, we can increase the effective length of the parallel current and by using mercury, we can make the displacement of even a few mm observable very dramatically. You will also need a constant-current supply giving a constant current of about 5 A.

Take a soft spring whose natural period of oscillations is about 0.5 – 1s. Hang it vertically and attach a pointed tip to its lower end, as shown in the figure here. Take some mercury in a dish and adjust the spring such that the tip is just above the mercury surface. Take the DC current source, connect one of its terminals to the upper end of the spring, and dip the other terminal in mercury. If the tip of the spring touches mercury, the circuit is completed through mercury.

Let the DC source be put off to begin with. Let the tip be adjusted so that it just touches the mercury surface. Switch on the constant current supply, and watch the fascinating outcome. The spring shrinks with a jerk, the tip comes out of mercury (just by a mm or so), the circuit is broken, the current stops, the spring relaxes and tries to come back to its original position, the tip again touches mercury establishing a current in the circuit, and the cycle continues with tick, tick, tick, ... In the beginning, you may require some small adjustments to get a good effect.

Keep your face away from mercury vapours as they are poisonous. Do not inhale mercury vapours for long.

EXAMPLE 4.10

Example 4.10 The horizontal component of the earth's magnetic field at a certain place is 3.0×10^{-5} T and the direction of the field is from the geographic south to the geographic north. A very long straight conductor is carrying a steady current of 1A. What is the force per unit length on it when it is placed on a horizontal table and the direction of the current is (a) east to west; (b) south to north?

Solution $\mathbf{F} = I\mathbf{l} \times \mathbf{B}$

$F = IlB \sin\theta$

The force per unit length is

$f = F/l = I\,B \sin\theta$

(a) When the current is flowing from east to west,

$\theta = 90°$

Hence,

$f = I\,B$

$= 1 \times 3 \times 10^{-5} = 3 \times 10^{-5}$ N m^{-1}

This is larger than the value 2×10^{-7} Nm^{-1} quoted in the definition of the ampere. Hence it is important to eliminate the effect of the earth's magnetic field and other stray fields while standardising the ampere.

The direction of the force is downwards. This direction may be obtained by the directional property of cross product of vectors.

(b) When the current is flowing from south to north,

$\theta = 0^{\circ}$

$f = 0$

Hence there is no force on the conductor.

EXAMPLE 4.10

4.10 Torque on Current Loop, Magnetic Dipole

4.10.1 Torque on a rectangular current loop in a uniform magnetic field

We now show that a rectangular loop carrying a steady current I and placed in a uniform magnetic field experiences a torque. It does not experience a net force. This behaviour is analogous to that of electric dipole in a uniform electric field (Section 1.10).

We first consider the simple case when the rectangular loop is placed such that the uniform magnetic field **B** is in the plane of the loop. This is illustrated in Fig. 4.21(a).

The field exerts no force on the two arms AD and BC of the loop. It is perpendicular to the arm AB of the loop and exerts a force $\mathbf{F}_1$ on it which is directed into the plane of the loop. Its magnitude is,

$$F_1 = I\,b\,B$$

Similarly it exerts a force $\mathbf{F}_2$ on the arm CD and $\mathbf{F}_2$ is directed out of the plane of the paper.

$$F_2 = I\,b\,B = F_1$$

Thus, the *net force* on the loop *is zero*. There is a torque on the loop due to the pair of forces $\mathbf{F}_1$ and $\mathbf{F}_2$. Figure 4.21(b) shows a view of the loop from the AD end. It shows that the torque on the loop tends to rotate it anti-clockwise. This torque is (in magnitude),

$$\tau = F_1\frac{a}{2} + F_2\frac{a}{2}$$

$$= IbB\frac{a}{2} + IbB\frac{a}{2} = I(ab)B$$

$$= I\,A\,B \qquad (4.26)$$

where $A = ab$ is the area of the rectangle.

We next consider the case when the plane of the loop, is not along the magnetic field, but makes an angle with it. We take the angle between the field and the normal to

FIGURE 4.21 (a) A rectangular current-carrying coil in uniform magnetic field. The magnetic moment **m** points downwards. The torque τ is along the axis and tends to rotate the coil anticlockwise. (b) The couple acting on the coil.

FIGURE 4.22 (a) The area vector of the loop ABCD makes an arbitrary angle θ with the magnetic field. (b) Top view of the loop. The forces $\mathbf{F_1}$ and $\mathbf{F_2}$ acting on the arms AB and CD are indicated.

the coil to be angle θ (The previous case corresponds to $\theta = \pi/2$). Figure 4.22 illustrates this general case.

The forces on the arms BC and DA are equal, opposite, and act along the axis of the coil, which connects the centres of mass of BC and DA. Being collinear along the axis they cancel each other, resulting in no net force or torque. The forces on arms AB and CD are $\mathbf{F_1}$ and $\mathbf{F_2}$. They too are equal and opposite, with magnitude,

$$F_1 = F_2 = I\,b\,B$$

But they are not collinear! This results in a couple as before. The torque is, however, less than the earlier case when plane of loop was along the magnetic field. This is because the perpendicular distance between the forces of the couple has decreased. Figure 4.22(b) is a view of the arrangement from the AD end and it illustrates these two forces constituting a couple. The magnitude of the torque on the loop is,

$$\tau = F_1 \frac{a}{2}\sin\theta + F_2 \frac{a}{2}\sin\theta$$

$$= I\,ab\,B\sin\theta$$

$$= I\,A\,B\sin\theta \qquad (4.27)$$

As $\theta \rightarrow 0$, the perpendicular distance between the forces of the couple also approaches zero. This makes the forces collinear and the net force and torque zero. The torques in Eqs. (4.26) and (4.27) can be expressed as vector product of the magnetic moment of the coil and the magnetic field. We define the *magnetic moment* of the current loop as,

$$\mathbf{m} = I\,\mathbf{A} \qquad (4.28)$$

where the direction of the area vector $\mathbf{A}$ is given by the right-hand thumb rule and is directed into the plane of the paper in Fig. 4.21. Then as the angle between $\mathbf{m}$ and $\mathbf{B}$ is θ, Eqs. (4.26) and (4.27) can be expressed by one expression

$$\boldsymbol{\tau} = \mathbf{m} \cdot \mathbf{B} \qquad (4.29)$$

This is analogous to the electrostatic case (Electric dipole of dipole moment $\mathbf{p}_e$ in an electric field $\mathbf{E}$).

$$\boldsymbol{\tau} = \mathbf{p}_e \cdot \mathbf{E}$$

As is clear from Eq. (4.28), the dimensions of the magnetic moment are $[A][L^2]$ and its unit is Am^2.

From Eq. (4.29), we see that the torque $\boldsymbol{\tau}$ vanishes when $\mathbf{m}$ is either parallel or antiparallel to the magnetic field $\mathbf{B}$. This indicates a state of equilibrium as there is no torque on the coil (this also applies to any object with a magnetic moment $\mathbf{m}$). When $\mathbf{m}$ and $\mathbf{B}$ are parallel the

equilibrium is a stable one. Any small rotation of the coil produces a torque which brings it back to its original position. When they are antiparallel, the equilibrium is unstable as any rotation produces a torque which increases with the amount of rotation. The presence of this torque is also the reason why a small magnet or any magnetic dipole aligns itself with the external magnetic field.

If the loop has N closely wound turns, the expression for torque, Eq. (4.29), still holds, with

$$\mathbf{m} = N I \mathbf{A} \tag{4.30}$$

Example 4.11 A 100 turn closely wound circular coil of radius 10 cm carries a current of 3.2 A. (a) What is the field at the centre of the coil? (b) What is the magnetic moment of this coil?

The coil is placed in a vertical plane and is free to rotate about a horizontal axis which coincides with its diameter. A uniform magnetic field of 2T in the horizontal direction exists such that initially the axis of the coil is in the direction of the field. The coil rotates through an angle of 90° under the influence of the magnetic field. (c) What are the magnitudes of the torques on the coil in the initial and final position? (d) What is the angular speed acquired by the coil when it has rotated by 90°? The moment of inertia of the coil is 0.1 kg m^2.

Solution

(a) From Eq. (4.16)

$$B = \frac{\mu_0 N I}{2R}$$

Here, $N = 100$; $I = 3.2$ A, and $R = 0.1$ m. Hence,

$$B = \frac{4\pi \cdot 10^{-7} \cdot 10^2 \cdot 3.2}{2 \cdot 10^{-1}} = \frac{4 \cdot 10^{-5} \cdot 10}{2 \cdot 10^{-1}} \quad \text{(using } \pi \cdot 3.2 = 10\text{)}$$

$$= 2 \times 10^{-3}\,\text{T}$$

The direction is given by the right-hand thumb rule.

(b) The magnetic moment is given by Eq. (4.30),

$$m = N I A = N I \pi r^2 = 100 \times 3.2 \times 3.14 \times 10^{-2} = 10 \text{ A m}^2$$

The direction is once again given by the right hand thumb rule.

(c) $\tau = |\mathbf{m} \cdot \mathbf{B}|$ [from Eq. (4.29)]

$$= m B \sin\theta$$

Initially, $\theta = 0$. Thus, initial torque $\tau_i = 0$. Finally, $\theta = \pi/2$ (or 90°). Thus, final torque $\tau_f = m B = 10 \cdot 2 = 20$ N m.

(d) From Newton's second law,

$$\mathcal{I}\frac{d\omega}{dt} = m B \sin\theta$$

where $\mathcal{I}$ is the moment of inertia of the coil. From chain rule,

$$\frac{d\omega}{dt} = \frac{d\omega}{d\theta}\frac{d\theta}{dt} = \frac{d\omega}{d\theta}\omega$$

Using this,

$$\mathcal{I}\,\omega\, d\omega = m B \sin\theta\, d\theta$$

EXAMPLE 4.11

EXAMPLE 4.11

Integrating from $\theta = 0$ to $\theta = \pi/2$,

$$\mathcal{I}\int_0^{\omega_f} \omega\, d\omega = m\,B\int_0^{\pi/2} \sin\theta\, d\theta$$

$$\mathcal{I}\frac{\omega_f^2}{2} = -m\,B\cos\theta\,\Big|_0^{\pi/2} = m\,B$$

$$\omega_f = \left(\frac{2\,m\,B}{\mathcal{I}}\right)^{1/2} = \left(\frac{2\cdot 20}{10^{-1}}\right)^{1/2} = 20\ \text{s}^{-1}.$$

EXAMPLE 4.12

Example 4.12

(a) A current-carrying circular loop lies on a smooth horizontal plane. Can a uniform magnetic field be set up in such a manner that the loop turns around itself (i.e., turns about the vertical axis).

(b) A current-carrying circular loop is located in a uniform external magnetic field. If the loop is free to turn, what is its orientation of stable equilibrium? Show that in this orientation, the flux of the total field (external field + field produced by the loop) is maximum.

(c) A loop of irregular shape carrying current is located in an external magnetic field. If the wire is flexible, why does it change to a circular shape?

Solution

(a) No, because that would require $\boldsymbol{\tau}$ to be in the vertical direction. But $\boldsymbol{\tau} = I\,\mathbf{A}\times\mathbf{B}$, and since $\mathbf{A}$ of the horizontal loop is in the vertical direction, τ would be in the plane of the loop for any $\mathbf{B}$.

(b) Orientation of stable equilibrium is one where the area vector $\mathbf{A}$ of the loop is in the direction of external magnetic field. In this orientation, the magnetic field produced by the loop is in the same direction as external field, both normal to the plane of the loop, thus giving rise to maximum flux of the total field.

(c) It assumes circular shape with its plane normal to the field to maximize flux, since for a given perimeter, a circle encloses greater area than any other shape.

4.10.2 Circular current loop as a magnetic dipole

In this section, we shall consider the elementary magnetic element: the current loop. We shall show that the magnetic field (at large distances) due to current in a circular current loop is very similar in behavior to the electric field of an electric dipole. In Section 4.6, we have evaluated the magnetic field on the axis of a circular loop, of a radius R, carrying a steady current I. The magnitude of this field is [(Eq. (4.15)],

$$B = \frac{\propto_0 I\,R^2}{2\left(x^2+R^2\right)^{3/2}}$$

and its direction is along the axis and given by the right-hand thumb rule (Fig. 4.12). Here, x is the distance along the axis from the centre of the loop. For $x >> R$, we may drop the R^2 term in the denominator. Thus,

$$B = \frac{\propto_0 R^2}{2x^3}$$

Note that the area of the loop $A = \pi R^2$. Thus,

$$B = \frac{\propto_0 IA}{2\pi x^3}$$

As earlier, we define the magnetic moment **m** to have a magnitude IA, $\boldsymbol{m} = I\mathbf{A}$. Hence,

$$\mathbf{B} \simeq \frac{\propto_0 \mathbf{m}}{2\pi x^3}$$

$$= \frac{\propto_0}{4\pi}\frac{2\mathbf{m}}{x^3} \qquad [4.31(a)]$$

The expression of Eq. [4.31(a)] is very similar to an expression obtained earlier for the electric field of a dipole. The similarity may be seen if we substitute,

$\propto_0 \rightarrow 1/\varepsilon_0$

$\mathbf{m} \rightarrow \mathbf{p}_e$ (electrostatic dipole)

$\mathbf{B} \rightarrow \mathbf{E}$ (electrostatic field)

We then obtain,

$$\mathbf{E} = \frac{2\mathbf{p}_e}{4\pi\varepsilon_0 x^3}$$

which is precisely the field for an electric dipole at a point on its axis. considered in Chapter 1, Section 1.10 [Eq. (1.20)].

It can be shown that the above analogy can be carried further. We had found in Chapter 1 that the electric field on the perpendicular bisector of the dipole is given by [See Eq.(1.21)],

$$\mathbf{E} \simeq \frac{\mathbf{p}_e}{4\pi\varepsilon_0 x^3}$$

where x is the distance from the dipole. If we replace $\mathbf{p} \rightarrow \mathbf{m}$ and $\propto_0 \rightarrow 1/\varepsilon_0$ in the above expression, we obtain the result for **B** for a point *in the plane of the loop* at a distance x from the centre. For $x >> R$,

$$\mathbf{B} \simeq \frac{\propto_0}{4\pi}\frac{\mathbf{m}}{x^3}; \qquad x >> R \qquad [4.31(b)]$$

The results given by Eqs. [4.31(a)] and [4.31(b)] become exact for a *point* magnetic dipole.

The results obtained above can be shown to apply to any planar loop: a planar current loop is equivalent to a magnetic dipole of dipole moment $\mathbf{m} = I\mathbf{A}$, which is the analogue of electric dipole moment **p**. Note, however, a fundamental difference: an electric dipole is built up of two elementary units — the charges (or electric monopoles). In magnetism, a magnetic dipole (or a current loop) is the most elementary element. The equivalent of electric charges, i.e., magnetic monopoles, are not known to exist.

We have shown that a current loop (i) produces a magnetic field (see Fig. 4.12) and behaves like a magnetic dipole at large distances, and

(ii) is subject to torque like a magnetic needle. This led Ampere to suggest that all magnetism is due to circulating currents. This seems to be partly true and no magnetic monopoles have been seen so far. However, elementary particles such as an electron or a proton also carry an *intrinsic* magnetic moment, not accounted by circulating currents.

4.10.3 The magnetic dipole moment of a revolving electron

In Chapter 12 we shall read about the Bohr model of the hydrogen atom. You may perhaps have heard of this model which was proposed by the Danish physicist Niels Bohr in 1911 and was a stepping stone to a new kind of mechanics, namely, quantum mechanics. In the Bohr model, the electron (a negatively charged particle) revolves around a positively charged nucleus much as a planet revolves around the sun. The force in the former case is electrostatic (Coulomb force) while it is gravitational for the planet-Sun case. We show this Bohr picture of the electron in Fig. 4.23.

FIGURE 4.23 In the Bohr model of hydrogen-like atoms, the negatively charged electron is revolving with uniform speed around a centrally placed positively charged ($+Z\,e$) nucleus. The uniform circular motion of the electron constitutes a current. The direction of the magnetic moment is into the plane of the paper and is indicated separately by ⊗.

The electron of charge ($-e$) ($e = +\,1.6 \cdot 10^{-19}$ C) performs uniform circular motion around a stationary heavy nucleus of charge $+Ze$. This constitutes a current I, where,

$$I = \frac{e}{T} \tag{4.32}$$

and T is the time period of revolution. Let r be the orbital radius of the electron, and v the orbital speed. Then,

$$T = \frac{2\pi r}{v} \tag{4.33}$$

Substituting in Eq. (4.32), we have $I = ev/2\pi r$.

There will be a magnetic moment, usually denoted by $\propto_l$, associated with this circulating current. From Eq. (4.28) its magnitude is, $\propto_l = I\pi r^2 = evr/2$.

The direction of this magnetic moment is into the plane of the paper in Fig. 4.23. [This follows from the right-hand rule discussed earlier and the fact that the negatively charged electron is moving anti-clockwise, leading to a clockwise current.] Multiplying and dividing the right-hand side of the above expression by the electron mass m_e, we have,

$$\propto_l = \frac{e}{2m_e}(m_e v r)$$

$$= \frac{e}{2m_e} l \tag{4.34(a)}$$

Here, l is the magnitude of the angular momentum of the electron about the central nucleus ("orbital" angular momentum). Vectorially,

$$\boldsymbol{\propto}_l = -\frac{e}{2m_e}\mathbf{l} \tag{4.34(b)}$$

The negative sign indicates that the angular momentum of the electron is opposite in direction to the magnetic moment. Instead of electron with

charge $(-e)$, if we had taken a particle with charge $(+q)$, the angular momentum and magnetic moment would be in the same direction. The ratio

$$\frac{\propto_l}{l} = \frac{e}{2m_e} \tag{4.35}$$

is called the *gyromagnetic ratio* and is a constant. Its value is $8.8 \cdot 10^{10}$ C /kg for an electron, which has been verified by experiments.

The fact that even at an atomic level there is a magnetic moment, confirms Ampere's bold hypothesis of atomic magnetic moments. This according to Ampere, would help one to explain the magnetic properties of materials. Can one assign a value to this atomic dipole moment? The answer is Yes. One can do so within the Bohr model. Bohr hypothesised that the angular momentum assumes a discrete set of values, namely,

$$l = \frac{nh}{2\pi} \tag{4.36}$$

where n is a natural number, $n = 1, 2, 3, \ldots$ and h is a constant named after Max Planck (Planck's constant) with a value $h = 6.626 \cdot 10^{-34}$ J s. This condition of discreteness is called the *Bohr quantisation condition*. We shall discuss it in detail in Chapter 12. Our aim here is merely to use it to calculate the elementary dipole moment. Take the value $n = 1$, we have from Eq. (4.34) that,

$$(\propto_l)_{\min} = \frac{e}{4\pi m_e} h$$

$$= \frac{1.60 \cdot 10^{-19} \cdot 6.63 \quad 10^{-34}}{4 \cdot 3.14 \cdot 9.11 \quad 10^{-31}}$$

$$= 9.27 \times 10^{-24} \text{ Am}^2 \tag{4.37}$$

where the subscript 'min' stands for minimum. This value is called the *Bohr magneton*.

Any charge in uniform circular motion would have an associated magnetic moment given by an expression similar to Eq. (4.34). This dipole moment is labelled as the *orbital magnetic moment*. Hence the subscript 'l' in $\propto_l$. Besides the orbital moment, the electron has an *intrinsic* magnetic moment, which has the same numerical value as given in Eq. (4.37). It is called the *spin magnetic moment*. But we hasten to add that it is not as though the electron is spinning. The electron is an elementary particle and it does not have an axis to spin around like a top or our earth. Nevertheless it does possess this *intrinsic* magnetic moment. The microscopic roots of magnetism in iron and other materials can be traced back to this intrinsic spin magnetic moment.

4.11 The Moving Coil Galvanometer

Currents and voltages in circuits have been discussed extensively in Chapters 3. But how do we measure them? How do we claim that current in a circuit is 1.5 A or the voltage drop across a resistor is 1.2 V? Figure 4.24 exhibits a very useful instrument for this purpose: the *moving*

PHYSICS

Conversion of galvanometer into ameter and voltmeter:
http://patsy.hunter.cuny.edu/CORE/CORE4/LectureNotes/Electricity/electric6.htm

coil galvanometer (MCG). It is a device whose principle can be understood on the basis of our discussion in Section 4.10.

The galvanometer consists of a coil, with many turns, free to rotate about a fixed axis (Fig. 4.24), in a uniform radial magnetic field. There is a cylindrical soft iron core which not only makes the field radial but also increases the strength of the magnetic field. When a current flows through the coil, a torque acts on it. This torque is given by Eq. (4.26) to be

$$\tau = NI\,AB$$

where the symbols have their usual meaning. Since the field is radial by design, we have taken sin $\theta = 1$ in the above expression for the torque. The magnetic torque *NIAB* tends to rotate the coil. A spring S_p provides a counter torque $k\phi$ that balances the magnetic torque *NIAB;* resulting in a steady angular deflection ϕ. In equilibrium

$$k\phi = NI\,AB$$

where k is the torsional constant of the spring; i.e. the restoring torque per unit twist. The deflection ϕ is indicated on the scale by a pointer attached to the spring. We have

$$\phi = \left(\frac{NAB}{k}\right) I \qquad (4.38)$$

FIGURE 4.24 The moving coil galvanometer. Its elements are described in the text. Depending on the requirement, this device can be used as a current detector or for measuring the value of the current (ammeter) or voltage (voltmeter).

The quantity in brackets is a constant for a given galvanometer.

The galvanometer can be used in a number of ways. It can be used as a detector to check if a current is flowing in the circuit. We have come across this usage in the Wheatstone's bridge arrangement. In this usage the neutral position of the pointer (when no current is flowing through the galvanometer) is in the middle of the scale and not at the left end as shown in Fig.4.24. Depending on the direction of the current, the pointer deflection is either to the right or the left.

The galvanometer cannot as such be used as an ammeter to measure the value of the current in a given circuit. This is for two reasons: (i) Galvanometer is a very sensitive device, it gives a full-scale deflection for a current of the order of ∝A. (ii) For measuring currents, the galvanometer has to be connected in series, and as it has a large resistance, this will change the value of the current in the circuit. To overcome these difficulties, one attaches a small resistance r_s, called *shunt resistance*, in parallel with the galvanometer coil; so that most of the current passes through the shunt. The resistance of this arrangement is,

$$R_G\, r_s / (R_G + r_s) \simeq r_s \qquad \text{if} \quad R_G >> r_s$$

If r_s has small value, in relation to the resistance of the rest of the circuit R_c, the effect of introducing the measuring instrument is also small and negligible. This

arrangement is schematically shown in Fig. 4.25. The scale of this ammeter is calibrated and then graduated to read off the current value with ease. We define the *current sensitivity of the galvanometer as the deflection per unit current*. From Eq. (4.38) this current sensitivity is,

$$\frac{\phi}{I} = \frac{NAB}{k} \tag{4.39}$$

A convenient way for the manufacturer to increase the sensitivity is to increase the number of turns N. We choose galvanometers having sensitivities of value, required by our experiment.

FIGURE 4.25 Conversion of a galvanometer (G) to an ammeter by the introduction of a shunt resistance r_s of very small value in parallel.

The galvanometer can also be used as a voltmeter to measure the voltage across a given section of the circuit. For this it must be connected *in parallel* with that section of the circuit. Further, it must draw a very small current, otherwise the voltage measurement will disturb the original set up by an amount which is very large. Usually we like to keep the disturbance due to the measuring device below one per cent. To ensure this, a large resistance R is connected *in series* with the galvanometer. This arrangement is schematically depicted in Fig.4.26. Note that the resistance of the voltmeter is now,

$$R_G + R \simeq R: \text{ large}$$

The scale of the voltmeter is calibrated to read off the voltage value with ease. We define the *voltage sensitivity as the deflection per unit voltage*. From Eq. (4.38),

$$\frac{\phi}{V} = \left(\frac{NAB}{k}\right)\frac{I}{V} = \left(\frac{NAB}{k}\right)\frac{1}{R} \tag{4.40}$$

FIGURE 4.26 Conversion of a galvanometer (G) to a voltmeter by the introduction of a resistance R of large value in series.

An interesting point to note is that increasing the current sensitivity may not necessarily increase the voltage sensitivity. Let us take Eq. (4.39) which provides a measure of current sensitivity. If $N \rightarrow 2N$, i.e., we double the number of turns, then

$$\frac{\phi}{I} \rightarrow 2\frac{\phi}{I}$$

Thus, the current sensitivity doubles. However, the resistance of the galvanometer is also likely to double, since it is proportional to the length of the wire. In Eq. (4.40), $N \rightarrow 2N$, and $R \rightarrow 2R$, thus the voltage sensitivity,

$$\frac{\phi}{V} \rightarrow \frac{\phi}{V}$$

remains unchanged. So in general, the modification needed for conversion of a galvanometer to an ammeter will be different from what is needed for converting it into a voltmeter.

Example 4.13 In the circuit (Fig. 4.27) the current is to be measured. What is the value of the current if the ammeter shown (a) is a galvanometer with a resistance R_G = 60.00 Ω; (b) is a galvanometer described in (a) but converted to an ammeter by a shunt resistance r_s = 0.02 Ω; (c) is an ideal ammeter with zero resistance?

EXAMPLE 4.13

FIGURE 4.27

Solution

(a) Total resistance in the circuit is,

$R_G + 3 = 63\,\Omega$. Hence, I = 3/63 = 0.048 A.

(b) Resistance of the galvanometer converted to an ammeter is,

$$\frac{R_G\, r_s}{R_G + r_s} = \frac{60\ \ \Omega \cdot 0.02\Omega}{(60 + 0.02)\Omega} \simeq 0.02\,\Omega$$

Total resistance in the circuit is,

$0.02\,\Omega + 3\,\Omega = 3.02\,\Omega$. Hence, I = 3/3.02 = 0.99 A.

(c) For the ideal ammeter with zero resistance,

I = 3/3 = 1.00 A

EXAMPLE 4.13

SUMMARY

1. The total force on a charge q moving with velocity **v** in the presence of magnetic and electric fields **B** and **E,** respectively is called the *Lorentz force*. It is given by the expression:
   F = q (**v** × **B** + **E**)
   The magnetic force q (**v** × **B**) is normal to **v** and work done by it is zero.
2. A straight conductor of length l and carrying a steady current I experiences a force **F** in a uniform external magnetic field **B**,
   F = I **l** × **B**
   where |**l**| = l and the direction of **l** is given by the direction of the current.
3. In a uniform magnetic field **B**, a charge q executes a circular orbit in a plane normal to **B**. Its frequency of uniform circular motion is called the *cyclotron frequency* and is given by:
   $$\nu_c = \frac{qB}{2\pi m}$$
   This frequency is independent of the particle's speed and radius. This fact is exploited in a machine, the cyclotron, which is used to accelerate charged particles.
4. The *Biot-Savart* law asserts that the magnetic field d**B** due to an element d**l** carrying a steady current I at a point P at a distance r from the current element is:
   $$\mathrm{d}\mathbf{B} = \frac{\mu_0}{4\pi} I \frac{\mathrm{d}\mathbf{l} \cdot \mathbf{r}}{r^3}$$

To obtain the total field at P, we must integrate this vector expression over the entire length of the conductor.

5. The magnitude of the magnetic field due to a circular coil of radius R carrying a current I at an axial distance x from the centre is

$$B = \frac{\propto_0 IR^2}{2(x^2 + R^2)^{3/2}}$$

At the center this reduces to

$$B = \frac{\propto_0 I}{2R}$$

6. *Ampere's Circuital Law:* Let an open surface S be bounded by a loop C. Then the Ampere's law states that $\oint_C \mathbf{B} \cdot d\mathbf{l} = \propto_0 I$ where I refers to the current passing through S. The sign of I is determined from the right-hand rule. We have discussed a simplified form of this law. If **B** is directed along the tangent to every point on the perimeter L of a closed curve and is constant in magnitude along perimeter then,

$BL = \propto_0 I_e$

where I_e is the net current enclosed by the closed circuit.

7. The magnitude of the magnetic field at a distance R from a long, straight wire carrying a current I is given by:

$$B = \frac{\propto_0 I}{2\pi R}$$

The field lines are circles concentric with the wire.

8. The magnitude of the field B inside a *long solenoid* carrying a current I is

$B = \propto_0 nI$

where n is the number of turns per unit length. For a *toroid* one obtains,

$$B = \frac{\propto_0 NI}{2\pi r}$$

where N is the total number of turns and r is the average radius.

9. Parallel currents attract and anti-parallel currents repel.

10. A planar loop carrying a current I, having N closely wound turns, and an area A possesses a magnetic moment **m** where,

m = N I **A**

and the direction of **m** is given by the right-hand thumb rule : curl the palm of your right hand along the loop with the fingers pointing in the direction of the current. The thumb sticking out gives the direction of **m** (and **A**)

When this loop is placed in a uniform magnetic field **B**, the force **F** on it is: $F = 0$

And the torque on it is,

$\boldsymbol{\tau} = \mathbf{m} \times \mathbf{B}$

In a moving coil galvanometer, this torque is balanced by a counter-torque due to a spring, yielding

$k\phi = NI\,AB$

where ϕ is the equilibrium deflection and k the torsion constant of the spring.

11. An electron moving around the central nucleus has a magnetic moment $\propto_l$ given by:

$$\propto_l = \frac{e}{2m} l$$

where l is the magnitude of the angular momentum of the circulating electron about the central nucleus. The smallest value of $\propto_l$ is called the Bohr magneton $\propto_B$ and it is $\propto_B = 9.27 \cdot 10^{-24}$ J/T

12. A moving coil galvanometer can be converted into a ammeter by introducing a shunt resistance r_s, of small value in parallel. It can be converted into a voltmeter by introducing a resistance of a large value in series.

Physical Quantity	Symbol	Nature	Dimensions	Units	Remarks
Permeability of free space	$\propto_0$	Scalar	$[MLT^{-2}A^{-2}]$	$T\ m\ A^{-1}$	$4\pi \cdot 10^{-7}\ T\ m\ A^{-1}$
Magnetic Field	**B**	Vector	$[M\ T^{-2}A^{-1}]$	T (telsa)	
Magnetic Moment	**m**	Vector	$[L^2A]$	$A\ m^2$ or J/T	
Torsion Constant	k	Scalar	$[M\ L^2T^{-2}]$	$N\ m\ rad^{-1}$	Appears in MCG

POINTS TO PONDER

1. Electrostatic field lines originate at a positive charge and terminate at a negative charge or fade at infinity. Magnetic field lines always form closed loops.
2. The discussion in this Chapter holds only for steady currents which do not vary with time.
   When currents vary with time Newton's third law is valid only if momentum carried by the electromagnetic field is taken into account.
3. Recall the expression for the Lorentz force,

   $$\mathbf{F} = q\,(\mathbf{v} \times \mathbf{B} + \mathbf{E})$$

   This velocity dependent force has occupied the attention of some of the greatest scientific thinkers. If one switches to a frame with instantaneous velocity **v**, the magnetic part of the force vanishes. The motion of the charged particle is then explained by arguing that there exists an appropriate electric field in the new frame. We shall not discuss the details of this mechanism. However, we stress that the resolution of this paradox implies that electricity and magnetism are linked phenomena (*electromagnetism*) and that the Lorentz force expression *does not* imply a universal preferred frame of reference in nature.
4. Ampere's Circuital law is not independent of the Biot-Savart law. It can be derived from the Biot-Savart law. Its relationship to the Biot-Savart law is similar to the relationship between Gauss's law and Coulomb's law.

EXERCISES

4.1 A circular coil of wire consisting of 100 turns, each of radius 8.0 cm carries a current of 0.40 A. What is the magnitude of the magnetic field **B** at the centre of the coil?

4.2 A long straight wire carries a current of 35 A. What is the magnitude of the field **B** at a point 20 cm from the wire?

4.3 A long straight wire in the horizontal plane carries a current of 50 A in north to south direction. Give the magnitude and direction of **B** at a point 2.5 m east of the wire.

4.4 A horizontal overhead power line carries a current of 90 A in east to west direction. What is the magnitude and direction of the magnetic field due to the current 1.5 m below the line?

4.5 What is the magnitude of magnetic force per unit length on a wire carrying a current of 8 A and making an angle of 30° with the direction of a uniform magnetic field of 0.15 T?

4.6 A 3.0 cm wire carrying a current of 10 A is placed inside a solenoid perpendicular to its axis. The magnetic field inside the solenoid is given to be 0.27 T. What is the magnetic force on the wire?

4.7 Two long and parallel straight wires A and B carrying currents of 8.0 A and 5.0 A in the same direction are separated by a distance of 4.0 cm. Estimate the force on a 10 cm section of wire A.

4.8 A closely wound solenoid 80 cm long has 5 layers of windings of 400 turns each. The diameter of the solenoid is 1.8 cm. If the current carried is 8.0 A, estimate the magnitude of **B** inside the solenoid near its centre.

4.9 A square coil of side 10 cm consists of 20 turns and carries a current of 12 A. The coil is suspended vertically and the normal to the plane of the coil makes an angle of 30° with the direction of a uniform horizontal magnetic field of magnitude 0.80 T. What is the magnitude of torque experienced by the coil?

4.10 Two moving coil meters, M_1 and M_2 have the following particulars:

$R_1 = 10\ \Omega$, $N_1 = 30$,

$A_1 = 3.6 \times 10^{-3}\ m^2$, $B_1 = 0.25$ T

$R_2 = 14\ \Omega$, $N_2 = 42$,

$A_2 = 1.8 \times 10^{-3}\ m^2$, $B_2 = 0.50$ T

(The spring constants are identical for the two meters).

Determine the ratio of (a) current sensitivity and (b) voltage sensitivity of M_2 and M_1.

4.11 In a chamber, a uniform magnetic field of 6.5 G (1 G = 10^{-4} T) is maintained. An electron is shot into the field with a speed of $4.8 \times 10^6\ m\ s^{-1}$ normal to the field. Explain why the path of the electron is a circle. Determine the radius of the circular orbit. ($e = 1.6 \times 10^{-19}$ C, $m_e = 9.1\times10^{-31}$ kg)

4.12 In Exercise 4.11 obtain the frequency of revolution of the electron in its circular orbit. Does the answer depend on the speed of the electron? Explain.

4.13 (a) A circular coil of 30 turns and radius 8.0 cm carrying a current of 6.0 A is suspended vertically in a uniform horizontal magnetic field of magnitude 1.0 T. The field lines make an angle of 60°

with the normal of the coil. Calculate the magnitude of the counter torque that must be applied to prevent the coil from turning.

(b) Would your answer change, if the circular coil in (a) were replaced by a planar coil of some irregular shape that encloses the same area? (All other particulars are also unaltered.)

ADDITIONAL EXERCISES

4.14 Two concentric circular coils X and Y of radii 16 cm and 10 cm, respectively, lie in the same vertical plane containing the north to south direction. Coil X has 20 turns and carries a current of 16 A; coil Y has 25 turns and carries a current of 18 A. The sense of the current in X is anticlockwise, and clockwise in Y, for an observer looking at the coils facing west. Give the magnitude and direction of the net magnetic field due to the coils at their centre.

4.15 A magnetic field of 100 G (1 G = 10^{-4} T) is required which is uniform in a region of linear dimension about 10 cm and area of cross-section about 10^{-3} m^2. The maximum current-carrying capacity of a given coil of wire is 15 A and the number of turns per unit length that can be wound round a core is at most 1000 turns m^{-1}. Suggest some appropriate design particulars of a solenoid for the required purpose. Assume the core is not ferromagnetic.

4.16 For a circular coil of radius R and N turns carrying current I, the magnitude of the magnetic field at a point on its axis at a distance x from its centre is given by,

$$B = \frac{\propto_0 I R^2 N}{2\left(x^2 + R^2\right)^{3/2}}$$

(a) Show that this reduces to the familiar result for field at the centre of the coil.

(b) Consider two parallel co-axial circular coils of equal radius R, and number of turns N, carrying equal currents in the same direction, and separated by a distance R. Show that the field on the axis around the mid-point between the coils is uniform over a distance that is small as compared to R, and is given by,

$$B = 0.72 \frac{\propto_0 N I}{R}, \quad \text{approximately.}$$

[Such an arrangement to produce a nearly uniform magnetic field over a small region is known as *Helmholtz coils*.]

4.17 A toroid has a core (non-ferromagnetic) of inner radius 25 cm and outer radius 26 cm, around which 3500 turns of a wire are wound. If the current in the wire is 11 A, what is the magnetic field (a) outside the toroid, (b) inside the core of the toroid, and (c) in the empty space surrounded by the toroid.

4.18 Answer the following questions:

(a) A magnetic field that varies in magnitude from point to point but has a constant direction (east to west) is set up in a chamber. A charged particle enters the chamber and travels undeflected

along a straight path with constant speed. What can you say about the initial velocity of the particle?

(b) A charged particle enters an environment of a strong and non-uniform magnetic field varying from point to point both in magnitude and direction, and comes out of it following a complicated trajectory. Would its final speed equal the initial speed if it suffered no collisions with the environment?

(c) An electron travelling west to east enters a chamber having a uniform electrostatic field in north to south direction. Specify the direction in which a uniform magnetic field should be set up to prevent the electron from deflecting from its straight line path.

4.19 An electron emitted by a heated cathode and accelerated through a potential difference of 2.0 kV, enters a region with uniform magnetic field of 0.15 T. Determine the trajectory of the electron if the field (a) is transverse to its initial velocity, (b) makes an angle of 30° with the initial velocity.

4.20 A magnetic field set up using Helmholtz coils (described in Exercise 4.16) is uniform in a small region and has a magnitude of 0.75 T. In the same region, a uniform electrostatic field is maintained in a direction normal to the common axis of the coils. A narrow beam of (single species) charged particles all accelerated through 15 kV enters this region in a direction perpendicular to both the axis of the coils and the electrostatic field. If the beam remains undeflected when the electrostatic field is 9.0×10^{-5} V m^{-1}, make a simple guess as to what the beam contains. Why is the answer not unique?

4.21 A straight horizontal conducting rod of length 0.45 m and mass 60 g is suspended by two vertical wires at its ends. A current of 5.0 A is set up in the rod through the wires.

(a) What magnetic field should be set up normal to the conductor in order that the tension in the wires is zero?

(b) What will be the total tension in the wires if the direction of current is reversed keeping the magnetic field same as before? (Ignore the mass of the wires.) $g = 9.8$ m s^{-2}.

4.22 The wires which connect the battery of an automobile to its starting motor carry a current of 300 A (for a short time). What is the force per unit length between the wires if they are 70 cm long and 1.5 cm apart? Is the force attractive or repulsive?

4.23 A uniform magnetic field of 1.5 T exists in a cylindrical region of radius10.0 cm, its direction parallel to the axis along east to west. A wire carrying current of 7.0 A in the north to south direction passes through this region. What is the magnitude and direction of the force on the wire if,

(a) the wire intersects the axis,

(b) the wire is turned from N-S to northeast-northwest direction,

(c) the wire in the N-S direction is lowered from the axis by a distance of 6.0 cm?

4.24 A uniform magnetic field of 3000 G is established along the positive *z*-direction. A rectangular loop of sides 10 cm and 5 cm carries a current of 12 A. What is the torque on the loop in the different cases shown in Fig. 4.28? What is the force on each case? Which case corresponds to stable equilibrium?

FIGURE 4.28

4.25 A circular coil of 20 turns and radius 10 cm is placed in a uniform magnetic field of 0.10 T normal to the plane of the coil. If the current in the coil is 5.0 A, what is the

(a) total torque on the coil,

(b) total force on the coil,

(c) average force on each electron in the coil due to the magnetic field?

(The coil is made of copper wire of cross-sectional area 10^{-5} m^2, and the free electron density in copper is given to be about 10^{29} m^{-3}.)

4.26 A solenoid 60 cm long and of radius 4.0 cm has 3 layers of windings of 300 turns each. A 2.0 cm long wire of mass 2.5 g lies inside the solenoid (near its centre) normal to its axis; both the wire and the axis of the solenoid are in the horizontal plane. The wire is connected through two leads parallel to the axis of the solenoid to an external battery which supplies a current of 6.0 A in the wire. What value of current (with appropriate sense of circulation) in the windings of the solenoid can support the weight of the wire? g = 9.8 m s^{-2}.

4.27 A galvanometer coil has a resistance of 12 Ω and the metre shows full scale deflection for a current of 3 mA. How will you convert the metre into a voltmeter of range 0 to 18 V?

4.28 A galvanometer coil has a resistance of 15 Ω and the metre shows full scale deflection for a current of 4 mA. How will you convert the metre into an ammeter of range 0 to 6 A?

Chapter Five

MAGNETISM AND MATTER

5.1 Introduction

Magnetic phenomena are universal in nature. Vast, distant galaxies, the tiny invisible atoms, men and beasts all are permeated through and through with a host of magnetic fields from a variety of sources. The earth's magnetism predates human evolution. The word magnet is derived from the name of an island in Greece called *magnesia* where magnetic ore deposits were found, as early as 600 BC. Shepherds on this island complained that their wooden shoes (which had nails) at times stayed struck to the ground. Their iron-tipped rods were similarly affected. This attractive property of magnets made it difficult for them to move around.

The directional property of magnets was also known since ancient times. A thin long piece of a magnet, when suspended freely, pointed in the north-south direction. A similar effect was observed when it was placed on a piece of cork which was then allowed to float in still water. The name *lodestone* (or *loadstone*) given to a naturally occurring ore of iron-magnetite means leading stone. The technological exploitation of this property is generally credited to the Chinese. Chinese texts dating 400 BC mention the use of magnetic needles for navigation on ships. Caravans crossing the Gobi desert also employed magnetic needles.

A Chinese legend narrates the tale of the victory of the emperor Huang-ti about four thousand years ago, which he owed to his craftsmen (whom

FIGURE 5.1 The arm of the statuette mounted on the chariot always points south. This is an artist's sketch of one of the earliest known compasses, thousands of years old.

nowadays you would call engineers). These 'engineers' built a chariot on which they placed a magnetic figure with arms outstretched. Figure 5.1 is an artist's description of this chariot. The figure swiveled around so that the finger of the statuette on it always pointed south. With this chariot, Huang-ti's troops were able to attack the enemy from the rear in thick fog, and to defeat them.

In the previous chapter we have learned that moving charges or electric currents produce magnetic fields. This discovery, which was made in the early part of the nineteenth century is credited to Oersted, Ampere, Biot and Savart, among others.

In the present chapter, we take a look at magnetism as a subject in its own right.

Some of the commonly known ideas regarding magnetism are:

(i) The earth behaves as a magnet with the magnetic field pointing approximately from the geographic south to the north.

(ii) When a bar magnet is freely suspended, it points in the north-south direction. The tip which points to the geographic north is called the *north pole* and the tip which points to the geographic south is called the *south pole* of the magnet.

(iii) There is a repulsive force when north poles (or south poles) of two magnets are brought close together. Conversely, there is an attractive force between the north pole of one magnet and the south pole of the other.

(iv) We cannot isolate the north, or south pole of a magnet. If a bar magnet is broken into two halves, we get two similar bar magnets with somewhat weaker properties. Unlike electric charges, isolated magnetic north and south poles known as *magnetic monopoles* do not exist.

(v) It is possible to make magnets out of iron and its alloys.

We begin with a description of a bar magnet and its behaviour in an external magnetic field. We describe Gauss's law of magnetism. We then follow it up with an account of the earth's magnetic field. We next describe how materials can be classified on the basis of their magnetic properties. We describe para-, dia-, and ferromagnetism. We conclude with a section on electromagnets and permanent magnets.

5.2 The Bar Magnet

One of the earliest childhood memories of the famous physicist Albert Einstein was that of a magnet gifted to him by a relative. Einstein was fascinated, and played endlessly with it. He wondered how the magnet could affect objects such as nails or pins placed away from it and not in any way *connected* to it by a spring or string.

We begin our study by examining iron filings sprinkled on a sheet of glass placed over a short bar magnet. The arrangement of iron filings is shown in Fig. 5.2.

The pattern of iron filings suggests that the magnet has two poles similar to the positive and negative charge of an electric dipole. As mentioned in the introductory section, one pole is designated the *North pole* and the other, the *South pole*. When suspended freely, these poles point approximately towards the geographic north and south poles, respectively. A similar pattern of iron filings is observed around a current carrying solenoid.

FIGURE 5.2 The arrangement of iron filings surrounding a bar magnet. The pattern mimics magnetic field lines. The pattern suggests that the bar magnet is a magnetic dipole.

5.2.1 The magnetic field lines

The pattern of iron filings permits us to plot the magnetic field lines*. This is shown both for the bar-magnet and the current-carrying solenoid in Fig. 5.3. For comparison refer to the Chapter 1, Figure 1.17(d). Electric field lines of an electric dipole are also displayed in Fig. 5.3(c). The magnetic field lines are a visual and intuitive realisation of the magnetic field. Their properties are:

(i) The magnetic field lines of a magnet (or a solenoid) form continuous closed loops. This is unlike the electric dipole where these field lines begin from a positive charge and end on the negative charge or escape to infinity.

(ii) The tangent to the field line at a given point represents the direction of the net magnetic field **B** at that point.

FIGURE 5.3 The field lines of (a) a bar magnet, (b) a current-carrying finite solenoid and (c) electric dipole. At large distances, the field lines are very similar. The curves labelled (i) and (ii) are closed Gaussian surfaces.

* In some textbooks the magnetic field lines are called *magnetic lines of force*. This nomenclature is avoided since it can be confusing. Unlike electrostatics the field lines in magnetism do not indicate the direction of the force on a (moving) charge.

(iii) The larger the number of field lines crossing per unit area, the stronger is the magnitude of the magnetic field **B**. In Fig. 5.3(a), **B** is larger around region (ii) than in region (i).

(iv) The magnetic field lines do not intersect, for if they did, the direction of the magnetic field would not be unique at the point of intersection.

One can plot the magnetic field lines in a variety of ways. One way is to place a small magnetic compass needle at various positions and note its orientation. This gives us an idea of the magnetic field direction at various points in space.

5.2.2 Bar magnet as an equivalent solenoid

In the previous chapter, we have explained how a current loop acts as a magnetic dipole (Section 4.10). We mentioned Ampere's hypothesis that all magnetic phenomena can be explained in terms of circulating currents. Recall that the magnetic dipole moment **m** associated with a current loop was defined to be **m** = NI**A** where N is the number of turns in the loop, I the current and **A** the area vector (Eq. 4.30).

FIGURE 5.4 (a) Calculation of the axial field of a finite solenoid in order to demonstrate its similarity to that of a bar magnet. (b) A magnetic needle in a uniform magnetic field **B**. The arrangement may be used to determine either **B** or the magnetic moment **m** of the needle.

The resemblance of magnetic field lines for a bar magnet and a solenoid suggest that a bar magnet may be thought of as a large number of circulating currents in analogy with a solenoid. Cutting a bar magnet in half is like cutting a solenoid. We get two smaller solenoids with weaker magnetic properties. The field lines remain continuous, emerging from one face of the solenoid and entering into the other face. One can test this analogy by moving a small compass needle in the neighbourhood of a bar magnet and a current-carrying finite solenoid and noting that the deflections of the needle are similar in both cases.

To make this analogy more firm we calculate the axial field of a finite solenoid depicted in Fig. 5.4 (a). We shall demonstrate that at large distances this axial field resembles that of a bar magnet.

Let the solenoid of Fig. 5.4(a) consists of n turns per unit length. Let its length be $2l$ and radius a. We can evaluate the axial field at a point P, at a distance r from the centre O of the solenoid. To do this, consider a circular element of thickness dx of the solenoid at a distance x from its centre. It consists of $n\,d\,x$ turns. Let I be the current in the solenoid. In Section 4.6 of the previous chapter we have calculated the magnetic field on the axis of a circular current loop. From Eq. (4.13), the magnitude of the field at point P due to the circular element is

$$dB = \frac{\propto_0 n\, dx\, I a^2}{2[(r-x)^2 + a^2]^{3/2}}$$

The magnitude of the total field is obtained by summing over all the elements — in other words by integrating from $x = -l$ to $x = +l$. Thus,

$$B = \frac{\propto_0 nIa^2}{2} \int_{-l}^{l} \frac{dx}{[(r-x)^2 + a^2]^{3/2}}$$

This integration can be done by trigonometric substitutions. This exercise, however, is not necessary for our purpose. Note that the range of x is from $-l$ to $+l$. Consider the far axial field of the solenoid, i.e., $r >> a$ and $r >> l$. Then the denominator is approximated by

$$[(r-x)^2 + a^2]^{3/2} \approx r^3$$

and $$B = \frac{\propto_0 n I a^2}{2r^3} \int_{-l}^{l} dx$$

$$= \frac{\propto_0 n I}{2} \frac{2la^2}{r^3} \tag{5.1}$$

Note that the magnitude of the magnetic moment of the solenoid is, $m = n\,(2l)\,I\,(\pi a^2)$ — (total number of turns × current × cross-sectional area). Thus,

$$B = \frac{\propto_0}{4\pi} \frac{2m}{r^3} \tag{5.2}$$

This is also the far axial magnetic field of a bar magnet which one may obtain experimentally. Thus, a bar magnet and a solenoid produce similar magnetic fields. The magnetic moment of a bar magnet is thus equal to the magnetic moment of an equivalent solenoid that produces the same magnetic field.

Some textbooks assign a *magnetic charge* (also called *pole strength*) $+q_m$ to the north pole and $-q_m$ to the south pole of a bar magnet of length $2l$, and magnetic moment $q_m(2l)$. The field strength due to q_m at a distance r from it is given by $\propto_0 q_m/4\pi r^2$. The magnetic field due to the bar magnet is then obtained, both for the axial and the equatorial case, in a manner analogous to that of an electric dipole (Chapter 1). The method is simple and appealing. However, *magnetic monopoles do not exist*, and *we have avoided this approach for that reason.*

5.2.3 The dipole in a uniform magnetic field

The pattern of iron filings, i.e., the magnetic field lines gives us an approximate idea of the magnetic field **B**. We may at times be required to determine the magnitude of **B** accurately. This is done by placing a small compass needle of known magnetic moment **m** and moment of inertia $\mathcal{I}$ and allowing it to oscillate in the magnetic field. This arrangement is shown in Fig. 5.4(b).

The torque on the needle is [see Eq. (4.29)],

$$\tau = \mathbf{m} \times \mathbf{B} \tag{5.3}$$

In magnitude $\tau = mB\sin\theta$
Here τ is restoring torque and θ is the angle between **m** and **B**.

Therefore, in equilibrium $\mathcal{I}\dfrac{d^2\theta}{dt^2} = -mB\sin\theta$

Negative sign with $mB\sin\theta$ implies that restoring torque is in opposition to deflecting torque. For small values of θ in radians, we approximate $\sin\theta \approx \theta$ and get

$$\mathcal{I}\frac{d^2\theta}{dt^2} \approx -mB\theta$$

or, $$\frac{d^2\theta}{dt^2} = -\frac{mB}{\mathcal{I}}\theta$$

This represents a simple harmonic motion. The square of the angular frequency is $\omega^2 = mB/\mathcal{I}$ and the time period is,

$$T = 2\pi\sqrt{\frac{\mathcal{I}}{mB}} \tag{5.4}$$

or $$B = \frac{4\pi^2\mathcal{I}}{mT^2} \tag{5.5}$$

An expression for magnetic potential energy can also be obtained on lines similar to electrostatic potential energy.
The magnetic potential energy U_m is given by

$$U_m = \int \tau(\theta)d\theta$$

$$= \int mB\sin\theta = -mB\cos\theta$$

$$= -\mathbf{m}\cdot\mathbf{B} \tag{5.6}$$

We have emphasised in Chapter 2 that the zero of potential energy can be fixed at one's convenience. Taking the constant of integration to be zero means fixing the zero of potential energy at $\theta = 90°$, i.e., when the needle is perpendicular to the field. Equation (5.6) shows that potential energy is minimum (= $-mB$) at $\theta = 0°$ (most stable position) and maximum (= $+mB$) at $\theta = 180°$ (most unstable position).

Example 5.1 In Fig. 5.4(b), the magnetic needle has magnetic moment 6.7×10^{-2} Am^2 and moment of inertia $\mathcal{I} = 7.5 \times 10^{-6}$ kg m^2. It performs 10 complete oscillations in 6.70 s. What is the magnitude of the magnetic field?

Solution The time period of oscillation is,

$$T = \frac{6.70}{10} = 0.67s$$

From Eq. (5.5)

$$B = \frac{4\pi^2\mathcal{I}}{mT^2}$$

$$= \frac{4\cdot(3.14)^2\cdot\ 7.5\ \ 10^{-6}}{6.7\cdot 10^{-2}\cdot\ (0.67)^2}$$

$$= 0.01\ \text{T}$$

EXAMPLE 5.1

Example 5.2 A short bar magnet placed with its axis at 30° with an external field of 800 G experiences a torque of 0.016 Nm. (a) What is the magnetic moment of the magnet? (b) What is the work done in moving it from its most stable to most unstable position? (c) The bar magnet is replaced by a solenoid of cross-sectional area 2×10^{-4} m^2 and 1000 turns, but of the same magnetic moment. Determine the current flowing through the solenoid.

Solution

(a) From Eq. (5.3), $\tau = m B \sin\theta$, $\theta = 30°$, hence $\sin\theta = 1/2$.

Thus, $0.016 = m \times (800 \times 10^{-4}\ \text{T}) \times (1/2)$

$m = 160 \times 2/800 = 0.40\ \text{A m}^2$

(b) From Eq. (5.6), the most stable position is $\theta = 0°$ and the most unstable position is $\theta = 180°$. Work done is given by

$$W = U_m(\theta = 180°) - U_m(\theta = 0°)$$

$$= 2\, m B = 2 \times 0.40 \times 800 \times 10^{-4} = 0.064\ \text{J}$$

(c) From Eq. (4.30), $m_s = NIA$. From part (a), $m_s = 0.40$ A m^2

$$0.40 = 1000 \times I \times 2 \times 10^{-4}$$

$I = 0.40 \times 10^4/(1000 \times 2) = 2\text{A}$

EXAMPLE 5.2

Example 5.3

(a) What happens if a bar magnet is cut into two pieces: (i) transverse to its length, (ii) along its length?

(b) A magnetised needle in a uniform magnetic field experiences a torque but no net force. An iron nail near a bar magnet, however, experiences a force of attraction in addition to a torque. Why?

(c) Must every magnetic configuration have a north pole and a south pole? What about the field due to a toroid?

(d) Two identical looking iron bars A and B are given, one of which is definitely known to be magnetised. (We do not know which one.) How would one ascertain whether or not both are magnetised? If only one is magnetised, how does one ascertain which one? [Use nothing else but the bars A and B.]

Solution

(a) In either case, one gets two magnets, each with a north and south pole.

(b) No force if the field is uniform. The iron nail experiences a non-uniform field due to the bar magnet. There is induced magnetic moment in the nail, therefore, it experiences both force and torque. The net force is attractive because the induced south pole (say) in the nail is closer to the north pole of magnet than induced north pole.

(c) Not necessarily. True only if the source of the field has a net non-zero magnetic moment. This is not so for a toroid or even for a straight infinite conductor.

(d) Try to bring different ends of the bars closer. A repulsive force in some situation establishes that both are magnetised. If it is always attractive, then one of them is not magnetised. In a bar magnet the intensity of the magnetic field is the strongest at the two ends (poles) and weakest at the central region. This fact may be used to determine whether A or B is the magnet. In this case, to see which

EXAMPLE 5.3

EXAMPLE 5.3

one of the two bars is a magnet, pick up one, (say, A) and lower one of its ends; first on one of the ends of the other (say, B), and then on the middle of B. If you notice that in the middle of B, A experiences no force, then B is magnetised. If you do not notice any change from the end to the middle of B, then A is magnetised.

5.2.4 The electrostatic analog

Comparison of Eqs. (5.2), (5.3) and (5.6) with the corresponding equations for electric dipole (Chapter 1), suggests that magnetic field at large distances due to a bar magnet of magnetic moment **m** can be obtained from the equation for electric field due to an electric dipole of dipole moment **p**, by making the following replacements:

$$\mathbf{E} \rightarrow \mathbf{B}\,,\ \mathbf{p} \rightarrow \mathbf{m}\,,\ \frac{1}{4\pi\varepsilon_0} \rightarrow \frac{\propto_0}{4\pi}$$

In particular, we can write down the equatorial field ($\mathbf{B}_E$) of a bar magnet at a distance r, for $r >> l$, where l is the size of the magnet:

$$\mathbf{B}_E = -\frac{\propto_0 \mathbf{m}}{4\pi r^3} \tag{5.7}$$

Likewise, the axial field ($\mathbf{B}_A$) of a bar magnet for $r >> l$ is:

$$\mathbf{B}_A = \frac{\propto_0}{4\pi}\frac{2\mathbf{m}}{r^3} \tag{5.8}$$

Equation (5.8) is just Eq. (5.2) in the vector form. Table 5.1 summarises the analogy between electric and magnetic dipoles.

TABLE 5.1 THE DIPOLE ANALOGY

	Electrostatics	Magnetism
	$1/\varepsilon_0$	$\propto_0$
Dipole moment	$\mathbf{p}$	$\mathbf{m}$
Equatorial Field for a short dipole	$-\mathbf{p}/4\pi\varepsilon_0 r^3$	$-\propto_0 \mathbf{m} / 4\pi r^3$
Axial Field for a short dipole	$2\mathbf{p}/4\pi\varepsilon_0 r^3$	$\propto_0 2\mathbf{m} / 4\pi r^3$
External Field: torque	$\mathbf{p} \times \mathbf{E}$	$\mathbf{m} \times \mathbf{B}$
External Field: Energy	$-\mathbf{p}\cdot\mathbf{E}$	$-\mathbf{m}\cdot\mathbf{B}$

EXAMPLE 5.4

Example 5.4 What is the magnitude of the equatorial and axial fields due to a bar magnet of length 5.0 cm at a distance of 50 cm from its mid-point? The magnetic moment of the bar magnet is 0.40 A m^2, the same as in Example 5.2.

Solution From Eq. (5.7)

$$B_E = \frac{\propto_0 m}{4\pi r^3} = \frac{10^{-7}\cdot 0.4}{(0.5)^3} = \frac{10^{-7}\cdot 0.4}{0.125} = 3.2\cdot 10^{-7}\ \text{T}$$

From Eq. (5.8), $B_A = \dfrac{\propto_0 2m}{4\pi r^3} = 6.4\cdot 10^{-7}\ \text{T}$

Example 5.5 Figure 5.5 shows a small magnetised needle P placed at a point O. The arrow shows the direction of its magnetic moment. The other arrows show different positions (and orientations of the magnetic moment) of another identical magnetised needle Q.

(a) In which configuration the system is not in equilibrium?

(b) In which configuration is the system in (i) stable, and (ii) unstable equilibrium?

(c) Which configuration corresponds to the lowest potential energy among all the configurations shown?

FIGURE 5.5

Solution

Potential energy of the configuration arises due to the potential energy of one dipole (say, Q) in the magnetic field due to other (P). Use the result that the field due to P is given by the expression [Eqs. (5.7) and (5.8)]:

$$\mathbf{B}_P = -\frac{\mu_0}{4\pi}\frac{\mathbf{m}_P}{r^3} \quad \text{(on the normal bisector)}$$

$$\mathbf{B}_P = \frac{\mu_0}{4\pi}\frac{2\,\mathbf{m}_P}{r^3} \quad \text{(on the axis)}$$

where $\mathbf{m}_P$ is the magnetic moment of the dipole P.

Equilibrium is stable when $\mathbf{m}_Q$ is parallel to $\mathbf{B}_P$, and unstable when it is anti-parallel to $\mathbf{B}_P$.

For instance for the configuration Q_3 for which Q is along the perpendicular bisector of the dipole P, the magnetic moment of Q is parallel to the magnetic field at the position 3. Hence Q_3 is stable. Thus,

(a) PQ_1 and PQ_2

(b) (i) PQ_3, PQ_6 (stable); (ii) PQ_5, PQ_4 (unstable)

(c) PQ_6

EXAMPLE 5.5

5.3 Magnetism and Gauss's Law

In Chapter 1, we studied Gauss's law for electrostatics. In Fig 5.3(c), we see that for a closed surface represented by ⓘ, the number of lines leaving the surface is equal to the number of lines entering it. This is consistent with the fact that no net charge is enclosed by the surface. However, in the same figure, for the closed surface ⓘⓘ, there is a net outward flux, since it does include a net (positive) charge.

KARL FRIEDRICH GAUSS (1777 – 1855)

Karl Friedrich Gauss (1777 – 1855) He was a child prodigy and was gifted in mathematics, physics, engineering, astronomy and even land surveying. The properties of numbers fascinated him, and in his work he anticipated major mathematical development of later times. Along with Wilhelm Welser, he built the first electric telegraph in 1833. His mathematical theory of curved surface laid the foundation for the later work of Riemann.

The situation is radically different for magnetic fields which are continuous and form closed loops. Examine the Gaussian surfaces represented by ⓘ or ⓘⓘ in Fig 5.3(a) or Fig. 5.3(b). Both cases visually demonstrate that the number of magnetic field lines leaving the surface is balanced by the number of lines entering it. The *net magnetic flux is zero for both the surfaces*. This is true for any closed surface.

FIGURE 5.6

Consider a small vector area element $\Delta\mathbf{S}$ of a closed surface S as in Fig. 5.6. The magnetic flux through **ÄS** is defined as $\Delta\phi_B = \mathbf{B}\cdot\Delta\mathbf{S}$, where **B** is the field at $\Delta\mathbf{S}$. We divide S into many small area elements and calculate the individual flux through each. Then, the net flux ϕ_B is,

$$\phi_B = \sum_{'all'} \Delta\phi_B = \sum_{'all'} \mathbf{B}\cdot\Delta\mathbf{S} = 0 \qquad (5.9)$$

where 'all' stands for 'all area elements $\Delta\mathbf{S}$'. Compare this with the Gauss's law of electrostatics. The flux through a closed surface in that case is given by

$$\sum \mathbf{E}\cdot\Delta\mathbf{S} = \frac{q}{\varepsilon_0}$$

where q is the electric charge enclosed by the surface.

The difference between the Gauss's law of magnetism and that for electrostatics is a reflection of the fact that isolated magnetic poles (also called monopoles) are not known to exist. There are no sources or sinks of **B**; the simplest magnetic element is a dipole or a current loop. All magnetic phenomena can be explained in terms of an arrangement of dipoles and/or current loops.

Thus, Gauss's law for magnetism is:

The net magnetic flux through any closed surface is zero.

EXAMPLE 5.6

Example 5.6 Many of the diagrams given in Fig. 5.7 show magnetic field lines (thick lines in the figure) *wrongly*. Point out what is wrong with them. Some of them may describe electrostatic field lines correctly. Point out which ones.

FIGURE 5.7

Solution

(a) *Wrong*. Magnetic field lines can never emanate from a point, as shown in figure. Over any closed surface, the net flux of **B** must always be zero, i.e., pictorially as many field lines should seem to enter the surface as the number of lines leaving it. The field lines shown, in fact, represent electric field of a long positively charged wire. The correct magnetic field lines are circling the straight conductor, as described in Chapter 4.

EXAMPLE 5.6

EXAMPLE 5.6

(b) *Wrong*. Magnetic field lines (like electric field lines) can never cross each other, because otherwise the direction of field at the point of intersection is ambiguous. There is further error in the figure. Magnetostatic field lines can never form closed loops around empty space. A closed loop of static magnetic field line must enclose a region across which a current is passing. By contrast, electrostatic field lines can never form closed loops, neither in empty space, nor when the loop encloses charges.

(c) *Right*. Magnetic lines are completely confined within a toroid. Nothing wrong here in field lines forming closed loops, since each loop encloses a region across which a current passes. Note, for clarity of figure, only a few field lines within the toroid have been shown. Actually, the entire region enclosed by the windings contains magnetic field.

(d) *Wrong*. Field lines due to a solenoid at its ends and outside cannot be so completely straight and confined; such a thing violates Ampere's law. The lines should curve out at both ends, and meet eventually to form closed loops.

(e) *Right*. These are field lines outside and inside a bar magnet. Note carefully the direction of field lines inside. Not all field lines emanate out of a north pole (or converge into a south pole). Around both the N-pole, and the S-pole, the net flux of the field is zero.

(f) *Wrong*. These field lines cannot possibly represent a magnetic field. Look at the upper region. All the field lines seem to emanate out of the shaded plate. The net flux through a surface surrounding the shaded plate is not zero. This is impossible for a magnetic field. The given field lines, in fact, show the electrostatic field lines around a positively charged upper plate and a negatively charged lower plate. The difference between Fig. [5.7(e) and (f)] should be carefully grasped.

(g) *Wrong*. Magnetic field lines between two pole pieces cannot be precisely straight at the ends. Some fringing of lines is inevitable. Otherwise, Ampere's law is violated. This is also true for electric field lines.

EXAMPLE 5.7

Example 5.7

(a) Magnetic field lines show the direction (at every point) along which a small magnetised needle aligns (at the point). Do the magnetic field lines also represent the *lines of force* on a moving charged particle at every point?

(b) Magnetic field lines can be entirely confined within the core of a toroid, but not within a straight solenoid. Why?

(c) If magnetic monopoles existed, how would the Gauss's law of magnetism be modified?

(d) Does a bar magnet exert a torque on itself due to its own field? Does one element of a current-carrying wire exert a force on another element of the *same wire*?

(e) Magnetic field arises due to charges in motion. Can a system have magnetic moments even though its net charge is zero?

Solution

(a) *No*. The magnetic force is always normal to **B** (remember magnetic force = $q\mathbf{v} \times \mathbf{B}$). It is misleading to call *magnetic field lines* as *lines of force*.

(b) If field lines were entirely confined between two ends of a straight solenoid, the flux through the cross-section at each end would be non-zero. But the flux of field **B** through any closed surface must always be zero. For a toroid, this difficulty is absent because it has no 'ends'.

(c) Gauss's law of magnetism states that the flux of **B** through any closed surface is always zero $\oint_S \mathbf{B} \cdot d\mathbf{s} = 0$.

If monopoles existed, the right hand side would be equal to the monopole (magnetic charge) q_m enclosed by S. [Analogous to Gauss's law of electrostatics, $\int_S \mathbf{B} \cdot d\mathbf{s} = \mu_0 q_m$ where q_m is the (monopole) magnetic charge enclosed by S.]

(d) *No*. There is no force or torque on an element due to the field produced by that element itself. But there is a force (or torque) on an element of the same wire. (For the special case of a straight wire, this force is zero.)

(e) *Yes*. The average of the charge in the system may be zero. Yet, the mean of the magnetic moments due to various current loops may not be zero. We will come across such examples in connection with paramagnetic material where atoms have net dipole moment through their net charge is zero.

EXAMPLE 5.7

5.4 THE EARTH'S MAGNETISM

Earlier we have referred to the magnetic field of the earth. The strength of the earth's magnetic field varies from place to place on the earth's surface; its value being of the order of 10^{-5} T.

What causes the earth to have a magnetic field is not clear. Originally the magnetic field was thought of as arising from a giant bar magnet placed approximately along the axis of rotation of the earth and deep in the interior. However, this simplistic picture is certainly not correct. The magnetic field is now thought to arise due to electrical currents produced by convective motion of metallic fluids (consisting mostly of molten iron and nickel) in the outer core of the earth. This is known as the *dynamo effect*.

The magnetic field lines of the earth resemble that of a (hypothetical) magnetic dipole located at the centre of the earth. The axis of the dipole does not coincide with the axis of rotation of the earth but is presently titled by approximately 11.3° with respect to the later. In this way of looking at it, the magnetic poles are located where the magnetic field lines due to the dipole enter or leave the earth. The location of the north magnetic pole is at a latitude of 79.74° N and a longitude of 71.8° W, a place somewhere in north Canada. The magnetic south pole is at 79.74° S, 108.22° E in the Antarctica.

The pole near the geographic north pole of the earth is called the *north magnetic pole*. Likewise, the pole near the geographic south pole is called

PHYSICS

Geomagnetic field frequently asked questions
http://www.ngdc.noaa.gov/seg/geomag/

FIGURE 5.8 The earth as a giant magnetic dipole.

the *south magnetic pole*. There is some confusion in the nomenclature of the poles. If one looks at the magnetic field lines of the earth (Fig. 5.8), one sees that unlike in the case of a bar magnet, the field lines go into the earth at the north magnetic pole (N_m) and come out from the south magnetic pole (S_m). The convention arose because the magnetic north was the direction to which the north pole of a magnetic needle pointed; the north pole of a magnet was so named as it was the *north seeking pole*. Thus, in reality, the north magnetic pole behaves like the south pole of a bar magnet inside the earth and vice versa.

EXAMPLE 5.8

Example 5.8 The earth's magnetic field at the equator is approximately 0.4 G. Estimate the earth's dipole moment.

Solution From Eq. (5.7), the equatorial magnetic field is,

$$B_E = \frac{\propto_0 m}{4\pi r^3}$$

We are given that $B_E \sim 0.4$ G $= 4 \times 10^{-5}$ T. For r, we take the radius of the earth 6.4×10^6 m. Hence,

$$m = \frac{4 \cdot 10^{-5} \cdot (6.4 \quad 10^6)^3}{\propto_0 / 4\pi} = 4 \times 10^2 \times (6.4 \times 10^6)^3 \qquad (\propto_0/4\pi = 10^{-7})$$

$$= 1.05 \times 10^{23} \text{ A m}^2$$

This is close to the value 8×10^{22} A m^2 quoted in geomagnetic texts.

5.4.1 Magnetic declination and dip

Consider a point on the earth's surface. At such a point, the direction of the longitude circle determines the geographic north-south direction, the line of longitude towards the north pole being the direction of true north. The vertical plane containing the longitude circle and the axis of rotation of the earth is called the *geographic meridian*. In a similar way, one can define *magnetic meridian* of a place as the vertical plane which passes through the imaginary line joining the magnetic north and the south poles. This plane would intersect the surface of the earth in a longitude like circle. A magnetic needle, which is free to swing horizontally, would then lie in the magnetic meridian and the north pole of the needle would point towards the magnetic north pole. Since the line joining the magnetic poles is titled with respect to the geographic axis of the earth, the magnetic meridian at a point makes angle with the geographic meridian. This, then, is the angle between the true geographic north and the north shown by a compass needle. This angle is called the *magnetic declination* or simply *declination* (Fig. 5.9).

FIGURE 5.9 A magnetic needle free to move in horizontal plane, points toward the magnetic north-south direction.

The declination is greater at higher latitudes and smaller near the equator. The declination in India is small, it being

0°41′ E at Delhi and 0°58′ W at Mumbai. Thus, at both these places a magnetic needle shows the true north quite accurately.

There is one more quantity of interest. If a magnetic needle is perfectly balanced about a horizontal axis so that it can swing in a plane of the magnetic meridian, the needle would make an angle with the horizontal (Fig. 5.10). This is known as the *angle of dip* (also known as *inclination*). Thus, dip is the angle that the total magnetic field $\mathbf{B}_E$ of the earth makes with the surface of the earth. Figure 5.11 shows the magnetic meridian plane at a point P on the surface of the earth. The plane is a section through the earth. The total magnetic field at P can be resolved into a horizontal component $\mathbf{H}_E$ and a vertical component $\mathbf{Z}_E$. The angle that $\mathbf{B}_E$ makes with $\mathbf{H}_E$ is the angle of dip, I.

FIGURE 5.10 The circle is a section through the earth containing the magnetic meridian. The angle between $\mathbf{B}_E$ and the horizontal component $\mathbf{H}_E$ is the angle of dip.

FIGURE 5.11 The earth's magnetic field, $\mathbf{B}_E$, its horizontal and vertical components, $\mathbf{H}_E$ and $\mathbf{Z}_E$. Also shown are the declination, D and the inclination or angle of dip, I.

In most of the northern hemisphere, the north pole of the dip needle tilts downwards. Likewise in most of the southern hemisphere, the south pole of the dip needle tilts downwards.

To describe the magnetic field of the earth at a point on its surface, we need to specify three quantities, viz., the declination D, the angle of dip or the inclination I and the horizontal component of the earth's field H_E. These are known as the *element of the earth's magnetic field*.

Representing the verticle component by Z_E, we have

$$Z_E = B_E \sin I \qquad [5.10(a)]$$

$$H_E = B_E \cos I \qquad [5.10(b)]$$

which gives,

$$\tan I = \frac{Z_E}{H_E} \qquad [5.10(c)]$$

WHAT HAPPENS TO MY COMPASS NEEDLES AT THE POLES?

A compass needle consists of a magnetic needle which floats on a pivotal point. When the compass is held level, it points along the direction of the horizontal component of the earth's magnetic field at the location. Thus, the compass needle would stay along the magnetic meridian of the place. In some places on the earth there are deposits of magnetic minerals which cause the compass needle to deviate from the magnetic meridian. Knowing the magnetic declination at a place allows us to correct the compass to determine the direction of true north.

So what happens if we take our compass to the magnetic pole? At the poles, the magnetic field lines are converging or diverging vertically so that the horizontal component is negligible. If the needle is only capable of moving in a horizontal plane, it can point along any direction, rendering it useless as a direction finder. What one needs in such a case is a *dip needle* which is a compass pivoted to move in a vertical plane containing the magnetic field of the earth. The needle of the compass then shows the angle which the magnetic field makes with the vertical. At the magnetic poles such a needle will point straight down.

EXAMPLE 5.9

Example 5.9 In the magnetic meridian of a certain place, the horizontal component of the earth's magnetic field is 0.26G and the dip angle is 60°. What is the magnetic field of the earth at this location?

Solution

It is given that H_E = 0.26 G. From Fig. 5.11, we have

$$\cos 60^0 = \frac{H_E}{B_E}$$

$$B_E = \frac{H_E}{\cos 60^0}$$

$$= \frac{0.26}{(1/2)} = 0.52\,\text{G}$$

EARTH'S MAGNETIC FIELD

It must not be assumed that there is a giant bar magnet deep inside the earth which is causing the earth's magnetic field. Although there are large deposits of iron inside the earth, it is highly unlikely that a large solid block of iron stretches from the magnetic north pole to the magnetic south pole. The earth's core is very hot and molten, and the ions of iron and nickel are responsible for earth's magnetism. This hypothesis seems very probable. Moon, which has no molten core, has no magnetic field, Venus has a slower rate of rotation, and a weaker magnetic field, while Jupiter, which has the fastest rotation rate among planets, has a fairly strong magnetic field. However, the precise mode of these circulating currents and the energy needed to sustain them are not very well understood. These are several open questions which form an important area of continuing research.

The variation of the earth's magnetic field with position is also an interesting area of study. Charged particles emitted by the sun flow towards the earth and beyond, in a stream called the solar wind. Their motion is affected by the earth's magnetic field, and in turn, they affect the pattern of the earth's magnetic field. The pattern of magnetic field near the poles is quite different from that in other regions of the earth.

The variation of earth's magnetic field with time is no less fascinating. There are short term variations taking place over centuries and long term variations taking place over a period of a million years. In a span of 240 years from 1580 to 1820 AD, over which records are available, the magnetic declination at London has been found to change by 3.5°, suggesting that the magnetic poles inside the earth change position with time. On the scale of a million years, the earth's magnetic fields has been found to reverse its direction. Basalt contains iron, and basalt is emitted during volcanic activity. The little iron magnets inside it align themselves parallel to the magnetic field at that place as the basalt cools and solidifies. Geological studies of basalt containing such pieces of magnetised region have provided evidence for the change of direction of earth's magnetic field, several times in the past.

5.5 MAGNETISATION AND MAGNETIC INTENSITY

The earth abounds with a bewildering variety of elements and compounds. In addition, we have been synthesising new alloys, compounds and even elements. One would like to classify the magnetic properties of these substances. In the present section, we define and explain certain terms which will help us to carry out this exercise.

We have seen that a circulating electron in an atom has a magnetic moment. In a bulk material, these moments add up vectorially and they can give a net magnetic moment which is non-zero. We define *magnetisation* **M** of a sample to be equal to its net magnetic moment per unit volume:

$$\mathbf{M} = \frac{\mathbf{m}_{net}}{V} \tag{5.11}$$

M is a vector with dimensions L^{-1} A and is measured in a units of A m^{-1}.

Consider a long solenoid of n turns per unit length and carrying a current I. The magnetic field in the interior of the solenoid was shown to be given by

$$\mathbf{B}_0 = \propto_0 nI \tag{5.12}$$

If the interior of the solenoid is filled with a material with non-zero magnetisation, the field inside the solenoid will be greater than $\mathbf{B}_0$. The net **B** field in the interior of the solenoid may be expressed as

$$\mathbf{B} = \mathbf{B}_0 + \mathbf{B}_m \tag{5.13}$$

where $\mathbf{B}_m$ is the field contributed by the material core. It turns out that this additional field $\mathbf{B}_m$ is proportional to the magnetisation **M** of the material and is expressed as

$$\mathbf{B}_m = \propto_0 \mathbf{M} \tag{5.14}$$

where $\propto_0$ is the same constant (permeability of vacuum) that appears in Biot-Savart's law.

It is convenient to introduce another vector field **H**, called the *magnetic intensity*, which is defined by

$$\mathbf{H} = \frac{\mathbf{B}}{\propto_0} - \mathbf{M} \tag{5.15}$$

where **H** has the same dimensions as **M** and is measured in units of A m^{-1}. Thus, the total magnetic field **B** is written as

$$\mathbf{B} = \propto_0 (\mathbf{H} + \mathbf{M}) \tag{5.16}$$

We repeat our defining procedure. We have partitioned the contribution to the total magnetic field inside the sample into two parts: *one*, due to external factors such as the current in the solenoid. This is represented by **H**. The *other* is due to the specific nature of the magnetic material, namely **M**. The latter quantity can be influenced by external factors. This influence is mathematically expressed as

$$\mathbf{M} = \chi \mathbf{H} \tag{5.17}$$

where χ, a dimensionless quantity, is appropriately called the *magnetic susceptibility*. It is a measure of how a magnetic material responds to an external field. Table 5.2 lists χ for some elements. It is small and positive for materials, which are called *paramagnetic*. It is small and negative for materials, which are termed *diamagnetic*. In the latter case **M** and **H** are opposite in direction. From Eqs. (5.16) and (5.17) we obtain,

$$\mathbf{B} = \propto_0 (1 + \chi)\mathbf{H} \tag{5.18}$$

$$= \propto_0 \propto_r \mathbf{H}$$

$$= \propto \mathbf{H} \tag{5.19}$$

where $\propto_r = 1 + \chi$, is a dimensionless quantity called the *relative magnetic permeability* of the substance. It is the analog of the dielectric constant in electrostatics. The *magnetic permeability* of the substance is $\propto$ and it has the same dimensions and units as $\propto_0$;

$$\propto = \propto_0 \propto_r = \propto_0 (1+\chi).$$

The three quantities χ, $\propto_r$ and $\propto$ are interrelated and only one of them is independent. Given one, the other two may be easily determined.

TABLE 5.2 MAGNETIC SUSCEPTIBILITY OF SOME ELEMENTS AT 300 K

Diamagnetic substance	χ	Paramagnetic substance	χ
Bismuth	-1.66×10^{-5}	Aluminium	2.3×10^{-5}
Copper	-9.8×10^{-6}	Calcium	1.9×10^{-5}
Diamond	-2.2×10^{-5}	Chromium	2.7×10^{-4}
Gold	-3.6×10^{-5}	Lithium	2.1×10^{-5}
Lead	-1.7×10^{-5}	Magnesium	1.2×10^{-5}
Mercury	-2.9×10^{-5}	Niobium	2.6×10^{-5}
Nitrogen (STP)	-5.0×10^{-9}	Oxygen (STP)	2.1×10^{-6}
Silver	-2.6×10^{-5}	Platinum	2.9×10^{-4}
Silicon	-4.2×10^{-6}	Tungsten	6.8×10^{-5}

Example 5.10 A solenoid has a core of a material with relative permeability 400. The windings of the solenoid are insulated from the core and carry a current of 2A. If the number of turns is 1000 per metre, calculate (a) H, (b) M, (c) B and (d) the magnetising current I_m.

Solution

(a) The field H is dependent of the material of the core, and is
$H = nI = 1000 \times 2.0 = 2 \times 10^3$ A/m.

(b) The magnetic field B is given by
$B = \propto_r \propto_0 H$
$= 400 \times 4\pi \times 10^{-7}$ (N/A^2) $\times 2 \times 10^3$ (A/m)
$= 1.0$ T

(c) Magnetisation is given by
$M = (B - \propto_0 H)/\propto_0$
$= (\propto_r \propto_0 H - \propto_0 H)/\propto_0 = (\propto_r - 1)H = 399 \times H$
$\cong 8 \times 10^5$ A/m

(d) The magnetising current I_M is the additional current that needs to be passed through the windings of the solenoid in the absence of the core which would give a B value as in the presence of the core. Thus $B = \propto_r n_0 (I + I_M)$. Using $I = 2$A, $B = 1$ T, we get $I_M = 794$ A.

EXAMPLE 5.10

5.6 MAGNETIC PROPERTIES OF MATERIALS

The discussion in the previous section helps us to classify materials as diamagnetic, paramagnetic or ferromagnetic. In terms of the susceptibility χ, a material is diamagnetic if χ is negative, para- if χ is positive and small, and ferro- if χ is large and positive.

A glance at Table 5.3 gives one a better feeling for these materials. Here ε is a small positive number introduced to quantify paramagnetic materials. Next, we describe these materials in some detail.

TABLE 5.3

Diamagnetic	Paramagnetic	Ferromagnetic
$-1 \le \chi < 0$	$0 < \chi < \varepsilon$	$\chi >> 1$
$0 \le \propto_r < 1$	$1 < \propto_r < 1+\varepsilon$	$\propto_r >> 1$
$\propto < \propto_0$	$\propto > \propto_0$	$\propto >> \propto_0$

5.6.1 Diamagnetism

Diamagnetic substances are those which have tendency to move from stronger to the weaker part of the external magnetic field. In other words, unlike the way a magnet attracts metals like iron, it would repel a diamagnetic substance.

Figure 5.12(a) shows a bar of diamagnetic material placed in an external magnetic field. The field lines are repelled or expelled and the field inside the material is reduced. In most cases, as is evident from Table 5.2, this reduction is slight, being one part in 10^5. When placed in a non-uniform magnetic field, the bar will tend to move from high to low field.

FIGURE 5.12 Behaviour of magnetic field lines near a (a) diamagnetic, (b) paramagnetic substance.

The simplest explanation for diamagnetism is as follows. Electrons in an atom orbiting around nucleus possess orbital angular momentum. These orbiting electrons are equivalent to current-carrying loop and thus possess orbital magnetic moment. Diamagnetic substances are the ones in which resultant magnetic moment in an atom is zero. When magnetic field is applied, those electrons having orbital magnetic moment in the same direction slow down and those in the opposite direction speed up. This happens due to induced current in accordance with Lenz's law which you will study in Chapter 6. Thus, the substance develops a net magnetic moment in direction opposite to that of the applied field and hence repulsion.

Some diamagnetic materials are bismuth, copper, lead, silicon, nitrogen (at STP), water and sodium chloride. Diamagnetism is present in all the substances. However, the effect is so weak in most cases that it gets shifted by other effects like paramagnetism, ferromagnetism, etc.

The most exotic diamagnetic materials are *superconductors*. These are metals, cooled to very low temperatures which exhibits both *perfect conductivity* and *perfect diamagnetism*. Here the field lines are completely expelled! $\chi = -1$ and $\propto_r = 0$. A superconductor repels a magnet and (by Newton's third law) is repelled by the magnet. The phenomenon of perfect diamagnetism in superconductors is called the *Meissner effect*, after the name of its discoverer. Superconducting magnets can be gainfully exploited in variety of situations, for example, for running magnetically levitated superfast trains.

5.6.2 Paramagnetism

Paramagnetic substances are those which get weakly magnetised when placed in an external magnetic field. They have tendency to move from a region of weak magnetic field to strong magnetic field, i.e., they get weakly attracted to a magnet.

The individual atoms (or ions or molecules) of a paramagnetic material possess a permanent magnetic dipole moment of their own. On account of the ceaseless random thermal motion of the atoms, no net magnetisation is seen. In the presence of an external field $\mathbf{B}_0$, which is strong enough, and at low temperatures, the individual atomic dipole moment can be made to align and point in the same direction as $\mathbf{B}_0$. Figure 5.12(b) shows a bar of paramagnetic material placed in an external field. The field lines gets concentrated inside the material, and the field inside is enhanced. In most cases, as is evident from Table 5.2, this enhancement is slight, being one part in 10^5. When placed in a non-uniform magnetic field, the bar will tend to move from weak field to strong.

Some paramagnetic materials are aluminium, sodium, calcium, oxygen (at STP) and copper chloride. Experimentally, one finds that the magnetisation of a paramagnetic material is inversely proportional to the absolute temperature T,

$$M = C\frac{B_0}{T} \qquad [5.20(a)]$$

or equivalently, using Eqs. (5.12) and (5.17)

$$\chi = C\frac{\propto_0}{T} \qquad [5.20(b)]$$

This is known as *Curie's law*, after its discoverer Pieree Curie (1859-1906). The constant C is called *Curie's constant*. Thus, for a paramagnetic material both χ and $\propto_r$ depend not only on the material, but also (in a simple fashion) on the sample temperature. As the field is increased or the temperature is lowered, the magnetisation increases until it reaches the saturation value M_s, at which point all the dipoles are perfectly aligned with the field. Beyond this, Curie's law [Eq. (5.20)] is no longer valid.

PHYSICS

Magnetic materials, domain, etc.:
http://www.aacg.bham.ac.uk/magnetic_materials/

5.6.3 Ferromagnetism

Ferromagnetic substances are those which gets strongly magnetised when placed in an external magnetic field. They have strong tendency to move from a region of weak magnetic field to strong magnetic field, i.e., they get strongly attracted to a magnet.

The individual atoms (or ions or molecules) in a ferromagnetic material possess a dipole moment as in a paramagnetic material. However, they interact with one another in such a way that they spontaneously align themselves in a common direction over a macroscopic volume called *domain*. The explanation of this cooperative effect requires quantum mechanics and is beyond the scope of this textbook. Each domain has a net magnetisation. Typical domain size is 1mm and the domain contains about 10^{11} atoms. In the first instant, the magnetisation varies randomly from domain to domain and there is no bulk magnetisation. This is shown in Fig. 5.13(a). When we apply an external magnetic field $\mathbf{B}_0$, the domains orient themselves in the direction of $\mathbf{B}_0$ and simultaneously the domain oriented in the direction of $\mathbf{B}_0$ grow in size. This existence of domains and their motion in $\mathbf{B}_0$ are not speculations. One may observe this under a microscope after sprinkling a liquid suspension of powdered

(a)

(b)

FIGURE 5.13
(a) Randomly oriented domains, (b) Aligned domains.

ferromagnetic substance of samples. This motion of suspension can be observed. Figure 5.12(b) shows the situation when the domains have aligned and amalgamated to form a single 'giant' domain.

Thus, in a ferromagnetic material the field lines are highly concentrated. In non-uniform magnetic field, the sample tends to move towards the region of high field. We may wonder as to what happens when the external field is removed. In some ferromagnetic materials the magnetisation persists. Such materials are called *hard* magnetic materials or *hard ferromagnets*. Alnico, an alloy of iron, aluminium, nickel, cobalt and copper, is one such material. The naturally occurring lodestone is another. Such materials form permanent magnets to be used among other things as a compass needle. On the other hand, there is a class of ferromagnetic materials in which the magnetisation disappears on removal of the external field. Soft iron is one such material. Appropriately enough, such materials are called *soft ferromagnetic materials*. There are a number of elements, which are ferromagnetic: iron, cobalt, nickel, gadolinium, etc. The relative magnetic permeability is >1000!

The ferromagnetic property depends on temperature. At high enough temperature, a ferromagnet becomes a paramagnet. The domain structure disintegrates with temperature. This disappearance of magnetisation with temperature is gradual. It is a phase transition reminding us of the melting of a solid crystal. The temperature of transition from ferromagnetic to paramagnetism is called the *Curie temperature* T_c. Table 5.4 lists the Curie temperature of certain ferromagnets. The susceptibility above the Curie temperature, i.e., in the paramagnetic phase is described by,

$$\chi = \frac{C}{T - T_c} \quad (T > T_c) \tag{5.21}$$

Hysterisis in magnetic materials:
http://hyperphysics.phy-astr.gsu.edu/hbase/solids/hyst.html

PHYSICS

TABLE 5.4 CURIE TEMPERATURE T_C OF SOME FERROMAGNETIC MATERIALS

Material	T_c (K)
Cobalt	1394
Iron	1043
Fe_2O_3	893
Nickel	631
Gadolinium	317

EXAMPLE 5.11

Example 5.11 A domain in ferromagnetic iron is in the form of a cube of side length 1∝m. Estimate the number of iron atoms in the domain and the maximum possible dipole moment and magnetisation of the domain. The molecular mass of iron is 55 g/mole and its density is 7.9 g/cm^3. Assume that each iron atom has a dipole moment of 9.27×10^{-24} A m^2.

Solution The volume of the cubic domain is
$V = (10^{-6}\ \text{m})^3 = 10^{-18}\ \text{m}^3 = 10^{-12}\ \text{cm}^3$
Its mass is volume × density $= 7.9\ \text{g cm}^{-3} \times 10^{-12}\ \text{cm}^3 = 7.9 \times 10^{-12}$ g
It is given that Avagadro number (6.023×10^{23}) of iron atoms have a mass of 55 g. Hence, the number of atoms in the domain is

$$N = \frac{7.9 \cdot 10^{-12} \cdot 6.023 \quad 10^{23}}{55}$$

$$= 8.65 \times 10^{10} \text{ atoms}$$

The maximum possible dipole moment m_{max} is achieved for the (unrealistic) case when all the atomic moments are perfectly aligned. Thus,

$$m_{max} = (8.65 \times 10^{10}) \times (9.27 \times 10^{-24})$$
$$= 8.0 \times 10^{-13}\ \text{A m}^2$$

The consequent magnetisation is

$$M_{max} = m_{max}/\textit{Domain volume}$$
$$= 8.0 \times 10^{-13}\ \text{Am}^2/10^{-18}\ \text{m}^3$$
$$= 8.0 \times 10^5\ \text{Am}^{-1}$$

EXAMPLE 5.11

The relation between **B** and **H** in ferromagnetic materials is complex. It is often not linear and it depends on the magnetic history of the sample. Figure 5.14 depicts the behaviour of the material as we take it through one cycle of magnetisation. Let the material be unmagnetised initially. We place it in a solenoid and increase the current through the solenoid. The magnetic field B in the material rises and saturates as depicted in the curve Oa. This behaviour represents the alignment and merger of domains until no further enhancement is possible. It is pointless to increase the current (and hence the magnetic intensity H) beyond this. Next, we decrease H and reduce it to zero. At $H = 0$, B $\neq 0$. This is represented by the curve ab. The value of B at $H = 0$ is called *retentivity* or *remanence*. In Fig. 5.14, $B_R \sim 1.2$ T, where the subscript R denotes retentivity. The domains are not completely randomised even though the external driving field has been removed. Next, the current in the solenoid is reversed and slowly increased. Certain domains are flipped until the net field inside stands nullified. This is represented by the curve bc. The value of H at c is called *coercivity*. In Fig. 5.14 $H_c \sim -90\ \text{A m}^{-1}$. As the reversed current is increased in magnitude, we once again obtain saturation. The curve cd depicts this. The saturated magnetic field $B_s \sim 1.5$ T. Next, the current is reduced (curve de) and reversed (curve ea). The cycle repeats itself. Note that the curve Oa does not retrace itself as H is reduced. For a given value of H, B is not unique but depends on previous history of the sample. This phenomenon is called *hysterisis*. The word *hysterisis* means *lagging behind* (and not 'history').

FIGURE 5.14 The magnetic hysteresis loop is the B-H curve for ferromagnetic materials.

5.7 Permanent Magnets and Electromagnets

Substances which at room temperature retain their ferromagnetic property for a long period of time are called *permanent magnets*. Permanent

FIGURE 5.15 A blacksmith forging a permanent magnet by striking a red-hot rod of iron kept in the north-south direction with a hammer. The sketch is recreated from an illustration in De Magnete, a work published in 1600 and authored by William Gilbert, the court physician to Queen Elizabeth of England.

magnets can be made in a variety of ways. One can hold an iron rod in the north-south direction and hammer it repeatedly. The method is illustrated in Fig. 5.15. The illustration is from a 400 year old book to emphasise that the making of permanent magnets is an old *art*. One can also hold a steel rod and stroke it with one end of a bar magnet a large number of times, always in the same sense to make a permanent magnet.

An efficient way to make a permanent magnet is to place a ferromagnetic rod in a solenoid and pass a current. The magnetic field of the solenoid magnetises the rod.

The hysteresis curve (Fig. 5.14) allows us to select suitable materials for permanent magnets. The material should have high retentivity so that the magnet is strong and high coercivity so that the magnetisation is not erased by stray magnetic fields, temperature fluctuations or minor mechanical damage. Further, the material should have a high permeability. Steel is one-favoured choice. It has a slightly smaller retentivity than soft iron but this is outweighed by the much smaller coercivity of soft iron. Other suitable materials for permanent magnets are alnico, cobalt steel and ticonal.

Core of electromagnets are made of ferromagnetic materials which have high permeability and low retentivity. Soft iron is a suitable material for electromagnets. On placing a soft iron rod in a solenoid and passing a current, we increase the magnetism of the solenoid by a thousand fold. When we switch off the solenoid current, the magnetism is effectively switched off since the soft iron core has a low retentivity. The arrangement is shown in Fig. 5.16.

Indiais Magnetic Field:
http://iigs.iigm.res.in
PHYSICS

FIGURE 5.16 A soft iron core in solenoid acts as an electromagnet.

In certain applications, the material goes through an ac cycle of magnetisation for a long period. This is the case in transformer cores and telephone diaphragms. The hysteresis curve of such materials must be narrow. The energy dissipated and the heating will consequently be small. The material must have a high resistivity to lower eddy current losses. You will study about eddy currents in Chapter 6.

Electromagnets are used in electric bells, loudspeakers and telephone diaphragms. Giant electromagnets are used in cranes to lift machinery, and bulk quantities of iron and steel.

Mapping India's magnetic field

Because of its practical application in prospecting, communication, and navigation, the magnetic field of the earth is mapped by most nations with an accuracy comparable to geographical mapping. In India over a dozen observatories exist, extending from Trivandrum (now Thrivuvananthapuram) in the south to Gulmarg in the north. These observatories work under the aegis of the Indian Institute of Geomagnetism (IIG), in Colaba, Mumbai. The IIG grew out of the Colaba and Alibag observatories and was formally established in 1971. The IIG monitors (via its nation-wide observatories), the geomagnetic fields and fluctuations on land, and under the ocean and in space. Its services are used by the Oil and Natural Gas Corporation Ltd. (ONGC), the National Institute of Oceanography (NIO) and the Indian Space Research Organisation (ISRO). It is a part of the world-wide network which ceaselessly updates the geomagnetic data. Now India has a permanent station called Gangotri.

SUMMARY

1. The science of magnetism is old. It has been known since ancient times that magnetic materials tend to point in the north-south direction; like magnetic poles repel and unlike ones attract; and cutting a bar magnet in two leads to two smaller magnets. Magnetic poles cannot be isolated.
2. When a bar magnet of dipole moment **m** is placed in a uniform magnetic field **B**,
   (a) the force on it is zero,
   (b) the torque on it is **m** × **B**,
   (c) its potential energy is **–m·B**, where we choose the zero of energy at the orientation when **m** is perpendicular to **B**.
3. Consider a bar magnet of size l and magnetic moment **m**, at a distance r from its mid-point, where $r >> l$, the magnetic field **B** due to this bar is,

$$\mathbf{B} = \frac{\mu_0 \mathbf{m}}{2\pi r^3} \quad \text{(along axis)}$$

$$= -\frac{\mu_0 \mathbf{m}}{4\pi r^3} \quad \text{(along equator)}$$

4. Gauss's law for magnetism states that the net magnetic flux through any closed surface is zero

$$\phi_B = \sum_{\substack{\text{all area} \\ \text{elements } \Delta\mathbf{S}}} \mathbf{B}\cdot\Delta\mathbf{S} = 0$$

5. The earth's magnetic field resembles that of a (hypothetical) magnetic dipole located at the centre of the earth. The pole near the geographic north pole of the earth is called the north magnetic pole. Similarly, the pole near the geographic south pole is called the south magnetic pole. This dipole is aligned making a small angle with the rotation axis of the earth. The magnitude of the field on the earth's surface $\approx 4 \times 10^{-5}$ T.

6. Three quantities are needed to specify the magnetic field of the earth on its surface – the horizontal component, the magnetic declination, and the magnetic dip. These are known as the elements of the earth's magnetic field.
7. Consider a material placed in an external magnetic field $\mathbf{B}_0$. The magnetic intensity is defined as,

   $$\mathbf{H} = \frac{\mathbf{B}_0}{\propto_0}$$

   The magnetisation **M** of the material is its dipole moment per unit volume. The magnetic field **B** in the material is,

   $\mathbf{B} = \propto_0 (\mathbf{H} + \mathbf{M})$
8. For a linear material $\mathbf{M} = \chi \mathbf{H}$. So that $\mathbf{B} = \propto \mathbf{H}$ and χ is called the magnetic susceptibility of the material. The three quantities, χ, the relative magnetic permeability $\propto_r$, and the magnetic permeability $\propto$ are related as follows:

   $\propto = \propto_0 \propto_r$

   $\propto_r = 1 + \chi$
9. Magnetic materials are broadly classified as: diamagnetic, paramagnetic, and ferromagnetic. For diamagnetic materials χ is negative and small and for paramagnetic materials it is positive and small. Ferromagnetic materials have large χ and are characterised by non-linear relation between **B** and **H**. They show the property of hysteresis.
10. Substances, which at room temperature, retain their ferromagnetic property for a long period of time are called permanent magnets.

Physical quantity	Symbol	Nature	Dimensions	Units	Remarks
Permeability of free space	$\propto_0$	Scalar	$[MLT^{-2} A^{-2}]$	$T\ m\ A^{-1}$	$\propto_0/4\pi = 10^{-7}$
Magnetic field, Magnetic induction, Magnetic flux density	**B**	Vector	$[MT^{-2} A^{-1}]$	T (tesla)	10^4 G (gauss) = 1 T
Magnetic moment	**m**	Vector	$[L^{-2} A]$	$A\ m^2$	
Magnetic flux	ϕ_B	Scalar	$[ML^2T^{-2} A^{-1}]$	W (weber)	$W = T\ m^2$
Magnetisation	**M**	Vector	$[L^{-1} A]$	$A\ m^{-1}$	$\frac{\text{Magnetic moment}}{\text{Volume}}$
Magnetic intensity Magnetic field strength	**H**	Vector	$[L^{-1} A]$	$A\ m^{-1}$	$\mathbf{B} = \propto_0 (\mathbf{H} + \mathbf{M})$
Magnetic susceptibility	χ	Scalar	-	-	$\mathbf{M} = \chi\mathbf{H}$
Relative magnetic permeability	$\propto_r$	Scalar	-	-	$\mathbf{B} = \propto_0 \propto_r \mathbf{H}$
Magnetic permeability	$\propto$	Scalar	$[MLT^{-2} A^{-2}]$	$T\ m\ A^{-1}$ $N\ A^{-2}$	$\propto = \propto_0 \propto_r$ $\mathbf{B} = \propto\mathbf{H}$

POINTS TO PONDER

1. A satisfactory understanding of magnetic phenomenon in terms of moving charges/currents was arrived at after 1800 AD. But technological exploitation of the directional properties of magnets predates this scientific understanding by two thousand years. Thus, scientific understanding is not a necessary condition for engineering applications. Ideally, science and engineering go hand-in-hand, one leading and assisting the other in tandem.
2. Magnetic monopoles do not exist. If you slice a magnet in half, you get two smaller magnets. On the other hand, isolated positive and negative charges exist. There exists a smallest unit of charge, for example, the electronic charge with value $|e| = 1.6 \times 10^{-19}$ C. All other charges are integral multiples of this smallest unit charge. In other words, charge is quantised. We do not know why magnetic monopoles do not exist or why electric charge is quantised.
3. A consequence of the fact that magnetic monopoles do not exist is that the magnetic field lines are continuous and form closed loops. In contrast, the electrostatic lines of force begin on a positive charge and terminate on the negative charge (or fade out at infinity).
4. The earth's magnetic field is not due to a huge bar magnet inside it. The earth's core is hot and molten. Perhaps convective currents in this core are responsible for the earth's magnetic field. As to what 'dynamo' effect sustains this current, and why the earth's field reverses polarity every million years or so, we do not know.
5. A miniscule difference in the value of χ, the magnetic susceptibility, yields radically different behaviour: diamagnetic versus paramagnetic. For diamagnetic materials $\chi = -10^{-5}$ whereas $\chi = +10^{-5}$ for paramagnetic materials.
6. There exists *a perfect diamagnet*, namely, a superconductor. This is a metal at very low temperatures. In this case $\chi = -1$, $\propto_r = 0$, $\propto = 0$. The external magnetic field is totally expelled. Interestingly, this material is also a perfect conductor. However, there exists no classical theory which ties these two properties together. A quantum-mechanical theory by Bardeen, Cooper, and Schrieffer (BCS theory) explains these effects. The BCS theory was proposed in1957 and was eventually recognised by a Nobel Prize in physics in 1970.
7. The phenomenon of magnetic hysteresis is reminiscent of similar behaviour concerning the elastic properties of materials. Strain may not be proportional to stress; here H and B (or M) are not linearly related. The stress-strain curve exhibits hysteresis and area enclosed by it represents the energy dissipated per unit volume. A similar interpretation can be given to the B-H magnetic hysteresis curve.
8. Diamagnetism is universal. It is present in all materials. But it is weak and hard to detect if the substance is para- or ferromagnetic.
9. We have classified materials as diamagnetic, paramagnetic, and ferromagnetic. However, there exist additional types of magnetic material such as ferrimagnetic, anti-ferromagnetic, spin glass, etc. with properties which are exotic and mysterious.

EXERCISES

5.1 Answer the following questions regarding earth's magnetism:

(a) A vector needs three quantities for its specification. Name the three independent quantities conventionally used to specify the earth's magnetic field.

(b) The angle of dip at a location in southern India is about 18°. Would you expect a greater or smaller dip angle in Britain?

(c) If you made a map of magnetic field lines at Melbourne in Australia, would the lines seem to go into the ground or come out of the ground?

(d) In which direction would a compass free to move in the vertical plane point to, if located right on the geomagnetic north or south pole?

(e) The earth's field, it is claimed, roughly approximates the field due to a dipole of magnetic moment 8×10^{22} J T^{-1} located at its centre. Check the order of magnitude of this number in some way.

(f) Geologists claim that besides the main magnetic N-S poles, there are several local poles on the earth's surface oriented in different directions. How is such a thing possible at all?

5.2 Answer the following questions:

(a) The earth's magnetic field varies from point to point in space. Does it also change with time? If so, on what time scale does it change appreciably?

(b) The earth's core is known to contain iron. Yet geologists do not regard this as a source of the earth's magnetism. Why?

(c) The charged currents in the outer conducting regions of the earth's core are thought to be responsible for earth's magnetism. What might be the 'battery' (i.e., the source of energy) to sustain these currents?

(d) The earth may have even reversed the direction of its field several times during its history of 4 to 5 billion years. How can geologists know about the earth's field in such distant past?

(e) The earth's field departs from its dipole shape substantially at large distances (greater than about 30,000 km). What agencies may be responsible for this distortion?

(f) Interstellar space has an extremely weak magnetic field of the order of 10^{-12} T. Can such a weak field be of any significant consequence? Explain.

[**Note:** Exercise 5.2 is meant mainly to arouse your curiosity. Answers to some questions above are tentative or unknown. Brief answers wherever possible are given at the end. For details, you should consult a good text on geomagnetism.]

5.3 A short bar magnet placed with its axis at 30° with a uniform external magnetic field of 0.25 T experiences a torque of magnitude equal to 4.5×10^{-2} J. What is the magnitude of magnetic moment of the magnet?

5.4 A short bar magnet of magnetic moment m = 0.32 JT^{-1} is placed in a uniform magnetic field of 0.15 T. If the bar is free to rotate in the plane of the field, which orientation would correspond to its (a) stable, and (b) unstable equilibrium? What is the potential energy of the magnet in each case?

5.5 A closely wound solenoid of 800 turns and area of cross section 2.5×10^{-4} m^2 carries a current of 3.0 A. Explain the sense in which the solenoid acts like a bar magnet. What is its associated magnetic moment?

5.6 If the solenoid in Exercise 5.5 is free to turn about the vertical direction and a uniform horizontal magnetic field of 0.25 T is applied, what is the magnitude of torque on the solenoid when its axis makes an angle of 30° with the direction of applied field?

5.7 A bar magnet of magnetic moment 1.5 J T^{-1} lies aligned with the direction of a uniform magnetic field of 0.22 T.

(a) What is the amount of work required by an external torque to turn the magnet so as to align its magnetic moment: (i) normal to the field direction, (ii) opposite to the field direction?

(b) What is the torque on the magnet in cases (i) and (ii)?

5.8 A closely wound solenoid of 2000 turns and area of cross-section 1.6×10^{-4} m^2, carrying a current of 4.0 A, is suspended through its centre allowing it to turn in a horizontal plane.

(a) What is the magnetic moment associated with the solenoid?

(b) What is the force and torque on the solenoid if a uniform horizontal magnetic field of 7.5×10^{-2} T is set up at an angle of 30° with the axis of the solenoid?

5.9 A circular coil of 16 turns and radius 10 cm carrying a current of 0.75 A rests with its plane normal to an external field of magnitude 5.0×10^{-2} T. The coil is free to turn about an axis in its plane perpendicular to the field direction. When the coil is turned slightly and released, it oscillates about its stable equilibrium with a frequency of 2.0 s^{-1}. What is the moment of inertia of the coil about its axis of rotation?

5.10 A magnetic needle free to rotate in a vertical plane parallel to the magnetic meridian has its north tip pointing down at 22° with the horizontal. The horizontal component of the earth's magnetic field at the place is known to be 0.35 G. Determine the magnitude of the earth's magnetic field at the place.

5.11 At a certain location in Africa, a compass points 12° west of the geographic north. The north tip of the magnetic needle of a dip circle placed in the plane of magnetic meridian points 60° above the horizontal. The horizontal component of the earth's field is measured to be 0.16 G. Specify the direction and magnitude of the earth's field at the location.

5.12 A short bar magnet has a magnetic moment of 0.48 J T^{-1}. Give the direction and magnitude of the magnetic field produced by the magnet at a distance of 10 cm from the centre of the magnet on (a) the axis, (b) the equatorial lines (normal bisector) of the magnet.

5.13 A short bar magnet placed in a horizontal plane has its axis aligned along the magnetic north-south direction. Null points are found on the axis of the magnet at 14 cm from the centre of the magnet. The earth's magnetic field at the place is 0.36 G and the angle of dip is zero. What is the total magnetic field on the normal bisector of the magnet at the same distance as the null–point (i.e., 14 cm) from the centre of the magnet? (At *null points*, field due to a magnet is equal and opposite to the horizontal component of earth's magnetic field.)

5.14 If the bar magnet in exercise 5.13 is turned around by 180°, where will the new null points be located?

5.15 A short bar magnet of magnetic moment 5.25×10^{-2} J T^{-1} is placed with its axis perpendicular to the earth's field direction. At what distance from the centre of the magnet, the resultant field is inclined at 45° with earth's field on (a) its normal bisector and (b) its axis. Magnitude of the earth's field at the place is given to be 0.42 G. Ignore the length of the magnet in comparison to the distances involved.

ADDITIONAL EXERCISES

5.16 Answer the following questions:

(a) Why does a paramagnetic sample display greater magnetisation (for the same magnetising field) when cooled?

(b) Why is diamagnetism, in contrast, almost independent of temperature?

(c) If a toroid uses bismuth for its core, will the field in the core be (slightly) greater or (slightly) less than when the core is empty?

(d) Is the permeability of a ferromagnetic material independent of the magnetic field? If not, is it more for lower or higher fields?

(e) Magnetic field lines are always nearly normal to the surface of a ferromagnet at every point. (This fact is analogous to the static electric field lines being normal to the surface of a conductor at every point.) Why?

(f) Would the maximum possible magnetisation of a paramagnetic sample be of the same order of magnitude as the magnetisation of a ferromagnet?

5.17 Answer the following questions:

(a) Explain qualitatively on the basis of domain picture the irreversibility in the magnetisation curve of a ferromagnet.

(b) The hysteresis loop of a soft iron piece has a much smaller area than that of a carbon steel piece. If the material is to go through repeated cycles of magnetisation, which piece will dissipate greater heat energy?

(c) 'A system displaying a hysteresis loop such as a ferromagnet, is a device for storing memory?' Explain the meaning of this statement.

(d) What kind of ferromagnetic material is used for coating magnetic tapes in a cassette player, or for building 'memory stores' in a modern computer?

(e) A certain region of space is to be shielded from magnetic fields. Suggest a method.

5.18 A long straight horizontal cable carries a current of 2.5 A in the direction 10° south of west to 10° north of east. The magnetic meridian of the place happens to be 10° west of the geographic meridian. The earth's magnetic field at the location is 0.33 G, and the angle of dip is zero. Locate the line of neutral points (ignore the thickness of the cable). (At *neutral points*, magnetic field due to a current-carrying cable is equal and opposite to the horizontal component of earth's magnetic field.)

5.19 A telephone cable at a place has four long straight horizontal wires carrying a current of 1.0 A in the same direction east to west. The

earth's magnetic field at the place is 0.39 G, and the angle of dip is 35°. The magnetic declination is nearly zero. What are the resultant magnetic fields at points 4.0 cm below the cable?

5.20 A compass needle free to turn in a horizontal plane is placed at the centre of circular coil of 30 turns and radius 12 cm. The coil is in a vertical plane making an angle of 45° with the magnetic meridian. When the current in the coil is 0.35 A, the needle points west to east.

(a) Determine the horizontal component of the earth's magnetic field at the location.

(b) The current in the coil is reversed, and the coil is rotated about its vertical axis by an angle of 90° in the anticlockwise sense looking from above. Predict the direction of the needle. Take the magnetic declination at the places to be zero.

5.21 A magnetic dipole is under the influence of two magnetic fields. The angle between the field directions is 60°, and one of the fields has a magnitude of 1.2×10^{-2} T. If the dipole comes to stable equilibrium at an angle of 15° with this field, what is the magnitude of the other field?

5.22 A monoenergetic (18 keV) electron beam initially in the horizontal direction is subjected to a horizontal magnetic field of 0.04 G normal to the initial direction. Estimate the up or down deflection of the beam over a distance of 30 cm ($m_e = 9.11 \times 10^{-19}$ C). [**Note:** *Data* in this exercise are so chosen that the answer will give you an idea of the effect of earth's magnetic field on the motion of the electron beam from the electron gun to the screen in a TV set.]

5.23 A sample of paramagnetic salt contains 2.0×10^{24} atomic dipoles each of dipole moment 1.5×10^{-23} J T^{-1}. The sample is placed under a homogeneous magnetic field of 0.64 T, and cooled to a temperature of 4.2 K. The degree of magnetic saturation achieved is equal to 15%. What is the total dipole moment of the sample for a magnetic field of 0.98 T and a temperature of 2.8 K? (Assume Curie's law)

5.24 A Rowland ring of mean radius 15 cm has 3500 turns of wire wound on a ferromagnetic core of relative permeability 800. What is the magnetic field **B** in the core for a magnetising current of 1.2 A?

5.25 The magnetic moment vectors $\propto_s$ and $\propto_l$ associated with the intrinsic spin angular momentum **S** and orbital angular momentum **l**, respectively, of an electron are predicted by quantum theory (and verified experimentally to a high accuracy) to be given by:

$\propto_s = -(e/m)\ \mathbf{S}$,

$\propto_l = -(e/2m)\mathbf{l}$

Which of these relations is in accordance with the result expected *classically*? Outline the derivation of the classical result.

Chapter Six

ELECTROMAGNETIC INDUCTION

6.1 Introduction

Electricity and magnetism were considered separate and unrelated phenomena for a long time. In the early decades of the nineteenth century, experiments on electric current by Oersted, Ampere and a few others established the fact that electricity and magnetism are inter-related. They found that moving electric charges produce magnetic fields. For example, an electric current deflects a magnetic compass needle placed in its vicinity. This naturally raises the questions like: Is the converse effect possible? Can moving magnets produce electric currents? Does the nature permit such a relation between electricity and magnetism? The answer is resounding yes! The experiments of Michael Faraday in England and Joseph Henry in USA, conducted around 1830, demonstrated conclusively that electric currents were induced in closed coils when subjected to changing magnetic fields. In this chapter, we will study the phenomena associated with changing magnetic fields and understand the underlying principles. The phenomenon in which electric current is generated by varying magnetic fields is appropriately called *electromagnetic induction*.

When Faraday first made public his discovery that relative motion between a bar magnet and a wire loop produced a small current in the latter, he was asked, "What is the use of it?" His reply was: "What is the use of a new born baby?" The phenomenon of electromagnetic induction

is not merely of theoretical or academic interest but also of practical utility. Imagine a world where there is no electricity – no electric lights, no trains, no telephones and no personal computers. The pioneering experiments of Faraday and Henry have led directly to the development of modern day generators and transformers. Today's civilisation owes its progress to a great extent to the discovery of electromagnetic induction.

Josheph Henry [1797 – 1878] American experimental physicist professor at Princeton University and first director of the Smithsonian Institution. He made important improvements in electro-magnets by winding coils of insulated wire around iron pole pieces and invented an electromagnetic motor and a new, efficient telegraph. He discoverd self-induction and investigated how currents in one circuit induce currents in another.

JOSEPH HENRY (1797 – 1878)

6.2 The Experiments of Faraday and Henry

The discovery and understanding of electromagnetic induction are based on a long series of experiments carried out by Faraday and Henry. We shall now describe some of these experiments.

Experiment 6.1

Figure 6.1 shows a coil C_1* connected to a galvanometer G. When the North-pole of a bar magnet is pushed towards the coil, the pointer in the galvanometer deflects, indicating the presence of electric current in the coil. The deflection lasts as long as the bar magnet is in motion. The galvanometer does not show any deflection when the magnet is held stationary. When the magnet is pulled away from the coil, the galvanometer shows deflection in the opposite direction, which indicates reversal of the current's direction. Moreover, when the South-pole of the bar magnet is moved towards or away from the coil, the deflections in the galvanometer are opposite to that observed with the North-pole for similar movements. Further, the deflection (and hence current) is found to be larger when the magnet is pushed towards or pulled away from the coil faster. Instead, when the bar magnet is held fixed and the coil C_1 is moved towards or away from the magnet, the same effects are observed. It shows that *it is the relative motion between the magnet and the coil that is responsible for generation (induction) of electric current in the coil.*

FIGURE 6.1 When the bar magnet is pushed towards the coil, the pointer in the galvanometer G deflects.

Experiment 6.2

In Fig. 6.2 the bar magnet is replaced by a second coil C_2 connected to a battery. The steady current in the coil C_2 produces a steady magnetic field. As coil C_2 is

* Wherever the term 'coil or 'loop' is used, it is assumed that they are made up of conducting material and are prepared using wires which are coated with insulating material.

FIGURE 6.2 Current is induced in coil C_1 due to motion of the current carrying coil C_2.

moved towards the coil C_1, the galvanometer shows a deflection. This indicates that electric current is induced in coil C_1. When C_2 is moved away, the galvanometer shows a deflection again, but this time in the opposite direction. The deflection lasts as long as coil C_2 is in motion. When the coil C_2 is held fixed and C_1 is moved, the same effects are observed. Again, *it is the relative motion between the coils that induces the electric current.*

Experiment 6.3

The above two experiments involved relative motion between a magnet and a coil and between two coils, respectively. Through another experiment, Faraday showed that this relative motion is not an absolute requirement. Figure 6.3 shows two coils C_1 and C_2 held stationary. Coil C_1 is connected to galvanometer G while the second coil C_2 is connected to a battery through a tapping key K.

FIGURE 6.3 Experimental set-up for Experiment 6.3.

Interactive animation on Faraday's experiments and Lenz's law: http://micro.magnet.fsu.edu/electromagnet/java/faraday/index.html

It is observed that the galvanometer shows a momentary deflection when the tapping key K is pressed. The pointer in the galvanometer returns to zero immediately. If the key is held pressed continuously, there is no deflection in the galvanometer. When the key is released, a momentory deflection is observed again, but in the opposite direction. It is also observed that the deflection increases dramatically when an iron rod is inserted into the coils along their axis.

6.3 Magnetic Flux

Faraday's great insight lay in discovering a simple mathematical relation to explain the series of experiments he carried out on electromagnetic induction. However, before we state and appreciate his laws, we must get familiar with the notion of magnetic flux, Φ_B. Magnetic flux is defined in the same way as electric flux is defined in Chapter 1. Magnetic flux through

a plane of area A placed in a uniform magnetic field **B** (Fig. 6.4) can be written as

$$\Phi_B = \mathbf{B} \cdot \mathbf{A} = BA \cos \theta \qquad (6.1)$$

where θ is angle between **B** and **A**. The notion of the area as a vector has been discussed earlier in Chapter 1. Equation (6.1) can be extended to curved surfaces and nonuniform fields.

If the magnetic field has different magnitudes and directions at various parts of a surface as shown in Fig. 6.5, then the magnetic flux through the surface is given by

$$\Phi_B = \mathbf{B}_1 \cdot d\mathbf{A}_1 + \mathbf{B}_2 \cdot d\mathbf{A}_2 + \cdots = \sum_{\text{all}} \mathbf{B}_i \cdot d\mathbf{A}_i \qquad (6.2)$$

where 'all' stands for summation over all the area elements $d\mathbf{A}_i$ comprising the surface and $\mathbf{B}_i$ is the magnetic field at the area element $d\mathbf{A}_i$. The SI unit of magnetic flux is weber (Wb) or tesla meter squared (T m^2). Magnetic flux is a scalar quantity.

FIGURE 6.4 A plane of surface area **A** placed in a uniform magnetic field **B**.

6.4 Faraday's Law of Induction

From the experimental observations, Faraday arrived at a conclusion that an emf is induced in a coil when magnetic flux through the coil changes with time. Experimental observations discussed in Section 6.2 can be explained using this concept.

The motion of a magnet towards or away from coil C_1 in Experiment 6.1 and moving a current-carrying coil C_2 towards or away from coil C_1 in Experiment 6.2, change the magnetic flux associated with coil C_1. The change in magnetic flux induces emf in coil C_1. It was this induced emf which caused electric current to flow in coil C_1 and through the galvanometer. A plausible explanation for the observations of Experiment 6.3 is as follows: When the tapping key K is pressed, the current in coil C_2 (and the resulting magnetic field) rises from zero to a maximum value in a short time. Consequently, the magnetic flux through the neighbouring coil C_1 also increases. It is the change in magnetic flux through coil C_1 that produces an induced emf in coil C_1. When the key is held pressed, current in coil C_2 is constant. Therefore, there is no change in the magnetic flux through coil C_1 and the current in coil C_1 drops to zero. When the key is released, the current in C_2 and the resulting magnetic field decreases from the maximum value to zero in a short time. This results in a decrease in magnetic flux through coil C_1 and hence again induces an electric current in coil C_1*. The common point in all these observations is that the time rate of change of magnetic flux through a circuit induces emf in it. Faraday stated experimental observations in the form of a law called *Faraday's law of electromagnetic induction*. The law is stated below.

FIGURE 6.5 Magnetic field $\mathbf{B}_i$ at the ith area element. $d\mathbf{A}_i$ represents area vector of the ith area element.

* Note that sensitive electrical instruments in the vicinity of an electromagnet can be damaged due to the induced emfs (and the resulting currents) when the electromagnet is turned on or off.

MICHAEL FARADAY (1791–1867)

Michael Faraday [1791–1867] Faraday made numerous contributions to science, viz., the discovery of electromagnetic induction, the laws of electrolysis, benzene, and the fact that the plane of polarisation is rotated in an electric field. He is also credited with the invention of the electric motor, the electric generator and the transformer. He is widely regarded as the greatest experimental scientist of the nineteenth century.

The magnitude of the induced emf *in a circuit is equal to the time rate of change of magnetic flux through the circuit.*

Mathematically, the induced emf is given by

$$\varepsilon = -\frac{d\Phi_B}{dt} \tag{6.3}$$

The negative sign indicates the direction of ε and hence the direction of current in a closed loop. This will be discussed in detail in the next section.

In the case of a closely wound coil of N turns, change of flux associated with each turn, is the same. Therefore, the expression for the total induced emf is given by

$$\varepsilon = -N\frac{d\Phi_B}{dt} \tag{6.4}$$

The induced emf can be increased by increasing the number of turns N of a closed coil.

From Eqs. (6.1) and (6.2), we see that the flux can be varied by changing any one or more of the terms **B**, **A** and θ. In Experiments 6.1 and 6.2 in Section 6.2, the flux is changed by varying **B**. The flux can also be altered by changing the shape of a coil (that is, by shrinking it or stretching it) in a magnetic field, or rotating a coil in a magnetic field such that the angle θ between **B** and **A** changes. In these cases too, an emf is induced in the respective coils.

EXAMPLE 6.1

Example 6.1 Consider Experiment 6.2. (a) What would you do to obtain a large deflection of the galvanometer? (b) How would you demonstrate the presence of an induced current in the absence of a galvanometer?

Solution

(a) To obtain a large deflection, one or more of the following steps can be taken: (i) Use a rod made of soft iron inside the coil C_2, (ii) Connect the coil to a powerful battery, and (iii) Move the arrangement rapidly towards the test coil C_1.

(b) Replace the galvanometer by a small bulb, the kind one finds in a small torch light. The relative motion between the two coils will cause the bulb to glow and thus demonstrate the presence of an induced current.

In experimental physics one must learn to innovate. Michael Faraday who is ranked as one of the best experimentalists ever, was legendary for his innovative skills.

EXAMPLE 6.2

Example 6.2 A square loop of side 10 cm and resistance 0.5 Ω is placed vertically in the east-west plane. A uniform magnetic field of 0.10 T is set up across the plane in the north-east direction. The magnetic field is decreased to zero in 0.70 s at a steady rate. Determine the magnitudes of induced emf and current during this time-interval.

Solution The angle θ made by the area vector of the coil with the magnetic field is 45°. From Eq. (6.1), the initial magnetic flux is

$$\Phi = BA\cos\theta$$

$$= \frac{0.1 \times 10^{-2}}{\sqrt{2}}\,\text{Wb}$$

Final flux, $\Phi_{min} = 0$

The change in flux is brought about in 0.70 s. From Eq. (6.3), the magnitude of the induced emf is given by

$$\varepsilon = \frac{|\Delta\Phi_B|}{\Delta t} = \frac{|(\Phi - 0)|}{\Delta t} = \frac{10^{-3}}{\sqrt{2} \times 0.7} = 1.0\,\text{mV}$$

And the magnitude of the current is

$$I = \frac{\varepsilon}{R} = \frac{10^{-3}\,\text{V}}{0.5\Omega} = 2\,\text{mA}$$

Note that the earth's magnetic field also produces a flux through the loop. But it is a steady field (which does not change within the time span of the experiment) and hence does not induce any emf.

EXAMPLE 6.2

Example 6.3

A circular coil of radius 10 cm, 500 turns and resistance 2 Ω is placed with its plane perpendicular to the horizontal component of the earth's magnetic field. It is rotated about its vertical diameter through 180° in 0.25 s. Estimate the magnitudes of the emf and current induced in the coil. Horizontal component of the earth's magnetic field at the place is 3.0×10^{-5} T.

Solution

Initial flux through the coil,

$$\Phi_{B\,(initial)} = BA\cos\theta$$

$$= 3.0 \times 10^{-5} \times (\pi \times 10^{-2}) \times \cos 0°$$

$$= 3\pi \times 10^{-7}\ \text{Wb}$$

Final flux after the rotation,

$$\Phi_{B\,(final)} = 3.0 \times 10^{-5} \times (\pi \times 10^{-2}) \times \cos 180°$$

$$= -3\pi \times 10^{-7}\ \text{Wb}$$

Therefore, estimated value of the induced emf is,

$$\varepsilon = N\frac{\Delta\Phi}{\Delta t}$$

$$= 500 \times (6\pi \times 10^{-7})/0.25$$

$$= 3.8 \times 10^{-3}\ \text{V}$$

$$I = \varepsilon/R = 1.9 \times 10^{-3}\,\text{A}$$

Note that the magnitudes of ε and I are the estimated values. Their instantaneous values are different and depend upon the speed of rotation at the particular instant.

EXAMPLE 6.3

6.5 Lenz's Law and Conservation of Energy

In 1834, German physicist Heinrich Friedrich Lenz (1804-1865) deduced a rule, known as *Lenz's law* which gives the polarity of the induced emf in a clear and concise fashion. The statement of the law is:

The polarity of induced emf is such that it tends to produce a current which opposes the change in magnetic flux that produced it.

The negative sign shown in Eq. (6.3) represents this effect. We can understand Lenz's law by examining Experiment 6.1 in Section 6.2.1. In Fig. 6.1, we see that the North-pole of a bar magnet is being pushed towards the closed coil. As the North-pole of the bar magnet moves towards the coil, the magnetic flux through the coil increases. Hence current is induced in the coil in such a direction that it opposes the increase in flux. This is possible only if the current in the coil is in a counter-clockwise direction with respect to an observer situated on the side of the magnet. Note that magnetic moment associated with this current has North polarity towards the North-pole of the approaching magnet. Similarly, if the North-pole of the magnet is being withdrawn from the coil, the magnetic flux through the coil will decrease. To counter this decrease in magnetic flux, the induced current in the coil flows in clockwise direction and its South-pole faces the receding North-pole of the bar magnet. This would result in an attractive force which opposes the motion of the magnet and the corresponding decrease in flux.

What will happen if an open circuit is used in place of the closed loop in the above example? In this case too, an emf is induced across the open ends of the circuit. The direction of the induced emf can be found using Lenz's law. Consider Figs. 6.6 (a) and (b). They provide an easier way to understand the direction of induced currents. Note that the direction shown by ⇄ and ⇄ indicate the directions of the induced currents.

FIGURE 6.6 Illustration of Lenz's law.

A little reflection on this matter should convince us on the correctness of Lenz's law. Suppose that the induced current was in the direction opposite to the one depicted in Fig. 6.6(a). In that case, the South-pole due to the induced current will face the approaching North-pole of the magnet. The bar magnet will then be attracted towards the coil at an ever increasing acceleration. A gentle push on the magnet will initiate the process and its velocity and kinetic energy will continuously increase without expending any energy. If this can happen, one could construct a perpetual-motion machine by a suitable arrangement. This violates the law of conservation of energy and hence can not happen.

Now consider the correct case shown in Fig. 6.6(a). In this situation, the bar magnet experiences a repulsive force due to the induced current. Therefore, a person has to do work in moving the magnet. Where does the energy spent by the person go? This energy is dissipated by Joule heating produced by the induced current.

Example 6.4

Figure 6.7 shows planar loops of different shapes moving out of or into a region of a magnetic field which is directed normal to the plane of the loop away from the reader. Determine the direction of induced current in each loop using Lenz's law.

FIGURE 6.7

Solution

(i) The magnetic flux through the rectangular loop abcd increases, due to the motion of the loop into the region of magnetic field, The induced current must flow along the path bcdab so that it opposes the increasing flux.

(ii) Due to the outward motion, magnetic flux through the triangular loop abc decreases due to which the induced current flows along bacb, so as to oppose the change in flux.

(iii) As the magnetic flux decreases due to motion of the irregular shaped loop abcd out of the region of magnetic field, the induced current flows along cdabc, so as to oppose change in flux.

Note that there are no induced current as long as the loops are completely inside or outside the region of the magnetic field.

EXAMPLE 6.4

Example 6.5

(a) A closed loop is held stationary in the magnetic field between the north and south poles of two permanent magnets held fixed. Can we hope to generate current in the loop by using very strong magnets?

(b) A closed loop moves normal to the constant electric field between the plates of a large capacitor. Is a current induced in the loop (i) when it is wholly inside the region between the capacitor plates (ii) when it is partially outside the plates of the capacitor? The electric field is normal to the plane of the loop.

(c) A rectangular loop and a circular loop are moving out of a uniform magnetic field region (Fig. 6.8) to a field-free region with a *constant velocity* **v**. In which loop do you expect the induced emf to be constant *during* the passage out of the field region? The field is normal to the loops.

EXAMPLE 6.5

FIGURE 6.8

(d) Predict the polarity of the capacitor in the situation described by Fig. 6.9.

FIGURE 6.9

Solution

(a) No. However strong the magnet may be, current can be induced only by changing the magnetic flux through the loop.

(b) No current is induced in *either* case. Current can not be induced by changing the electric flux.

(c) The induced emf is expected to be constant only in the case of the rectangular loop. In the case of circular loop, the rate of change of area of the loop during its passage out of the field region is not constant, hence induced *emf* will vary accordingly.

(d) The polarity of plate 'A' will be positive with respect to plate 'B' in the capacitor.

EXAMPLE 6.5

6.6 MOTIONAL ELECTROMOTIVE FORCE

Let us consider a straight conductor moving in a uniform and time-independent magnetic field. Figure 6.10 shows a rectangular conductor PQRS in which the conductor PQ is free to move. The rod PQ is moved towards the left with a constant velocity **v** as shown in the figure. Assume that there is no loss of energy due to friction. PQRS forms a closed circuit enclosing an area that changes as PQ moves. It is placed in a uniform magnetic field **B** which is perpendicular to the plane of this system. If the length RQ = x and RS = l, the magnetic flux Φ_B enclosed by the loop PQRS will be

FIGURE 6.10 The arm PQ is moved to the left side, thus decreasing the area of the rectangular loop. This movement induces a current I as shown.

$$\Phi_B = Blx$$

Since x is changing with time, the rate of change of flux Φ_B will induce an emf given by:

$$\varepsilon = \frac{-d\Phi_B}{dt} = -\frac{d}{dt}(Blx)$$

$$= -Bl\frac{dx}{dt} = Blv \qquad (6.5)$$

where we have used $dx/dt = -v$ which is the speed of the conductor PQ. The induced emf Blv is called *motional emf*. Thus, we are able to produce induced emf by moving a conductor instead of varying the magnetic field, that is, by changing the magnetic flux enclosed by the circuit.

It is also possible to explain the motional emf expression in Eq. (6.5) by invoking the Lorentz force acting on the free charge carriers of conductor PQ. Consider any arbitrary charge q in the conductor PQ. When the rod moves with speed v, the charge will also be moving with speed v in the magnetic field **B**. The Lorentz force on this charge is qvB in magnitude, and its direction is towards Q. All charges experience the same force, in magnitude and direction, irrespective of their position in the rod PQ. The work done in moving the charge from P to Q is,

$$W = qvBl$$

Since emf is the work done per unit charge,

$$\varepsilon = \frac{W}{q}$$

$$= Blv$$

This equation gives emf induced across the rod PQ and is identical to Eq. (6.5). We stress that our presentation is not wholly rigorous. But it does help us to understand the basis of Faraday's law when the conductor is moving in a uniform and time-independent magnetic field.

On the other hand, it is not obvious how an emf is induced when a conductor is stationary and the magnetic field is changing – a fact which Faraday verified by numerous experiments. In the case of a stationary conductor, the force on its charges is given by

$$\mathbf{F} = q\,(\mathbf{E} + \mathbf{v} \times \mathbf{B}) = q\mathbf{E} \qquad (6.6)$$

since $\mathbf{v} = 0$. Thus, any force on the charge must arise from the electric field term **E** alone. Therefore, to explain the existence of induced emf or induced current, we must assume that a time-varying magnetic field generates an electric field. However, we hasten to add that electric fields produced by static electric charges have properties different from those produced by time-varying magnetic fields. In Chapter 4, we learnt that charges in motion (current) can exert force/torque on a stationary magnet. Conversely, a bar magnet in motion (or more generally, a changing magnetic field) can exert a force on the stationary charge. This is the fundamental significance of the Faraday's discovery. Electricity and magnetism are related.

PHYSICS

Interactive animation on motional emf:
http://www.ngsir.netfirms.com/englishhtm/Induction.htm

Example 6.6 A metallic rod of 1 m length is rotated with a frequency of 50 rev/s, with one end hinged at the centre and the other end at the circumference of a circular metallic ring of radius 1 m, about an axis passing through the centre and perpendicular to the plane of the ring (Fig. 6.11). A constant and uniform magnetic field of 1 T parallel to the axis is present everywhere. What is the emf between the centre and the metallic ring?

EXAMPLE 6.6

FIGURE 6.11

Solution

Method I

As the rod is rotated, free electrons in the rod move towards the outer end due to Lorentz force and get distributed over the ring. Thus, the resulting separation of charges produces an emf across the ends of the rod. At a certain value of emf, there is no more flow of electrons and a steady state is reached. Using Eq. (6.5), the magnitude of the emf generated across a length dr of the rod as it moves at right angles to the magnetic field is given by

$d\varepsilon = Bv\,dr$. Hence,

$$\varepsilon = \int d\varepsilon = \int_0^R Bv\,dr = \int_0^R B\omega r\,dr = \frac{B\omega R^2}{2}$$

Note that we have used $v = \omega r$. This gives

$$\varepsilon = \frac{1}{2} \times 1.0 \times 2\pi \times 50 \times (1^2)$$

$= 157$ V

Method II

To calculate the emf, we can imagine a closed loop OPQ in which point O and P are connected with a resistor R and OQ is the rotating rod. The potential difference across the resistor is then equal to the induced emf and equals $B \times$ (rate of change of area of loop). If θ is the angle between the rod and the radius of the circle at P at time t, the area of the sector OPQ is given by

$$\pi R^2 \times \frac{\theta}{2\pi} = \frac{1}{2} R^2 \theta$$

where R is the radius of the circle. Hence, the induced emf is

$$\varepsilon = B \times \frac{d}{dt}\left[\frac{1}{2} R^2 \theta\right] = \frac{1}{2} BR^2 \frac{d\theta}{dt} = \frac{B\omega R^2}{2}$$

[Note: $\frac{d\theta}{dt} = \omega = 2\pi\nu$]

This expression is identical to the expression obtained by Method I and we get the same value of ε.

EXAMPLE 6.6

Example 6.7
A wheel with 10 metallic spokes each 0.5 m long is rotated with a speed of 120 rev/min in a plane normal to the horizontal component of earth's magnetic field H_E at a place. If H_E = 0.4 G at the place, what is the induced emf between the axle and the rim of the wheel? Note that 1 G = 10^{-4} T.

Solution

Induced emf = $(1/2)\ \omega B R^2$

$= (1/2) \times 4\pi \times 0.4 \times 10^{-4} \times (0.5)^2$

$= 6.28 \times 10^{-5}$ V

The number of spokes is immaterial because the emf's across the spokes are *in parallel*.

EXAMPLE 6.7

6.7 ENERGY CONSIDERATION: A QUANTITATIVE STUDY

In Section 6.5, we discussed qualitatively that Lenz's law is consistent with the law of conservation of energy. Now we shall explore this aspect further with a concrete example.

Let r be the resistance of movable arm PQ of the rectangular conductor shown in Fig. 6.10. We assume that the remaining arms QR, RS and SP have negligible resistances compared to r. Thus, the overall resistance of the rectangular loop is r and this does not change as PQ is moved. The current I in the loop is,

$$I = \frac{\varepsilon}{r}$$

$$= \frac{Blv}{r} \qquad (6.7)$$

On account of the presence of the magnetic field, there will be a force on the arm PQ. This force $I(\mathbf{l} \times \mathbf{B})$, is directed outwards in the direction opposite to the velocity of the rod. The magnitude of this force is,

$$F = IlB = \frac{B^2 l^2 v}{r}$$

where we have used Eq. (6.7). Note that this force arises due to drift velocity of charges (responsible for current) along the rod and the consequent Lorentz force acting on them.

Alternatively, the arm PQ is being pushed with a constant speed v, the power required to do this is,

$$P = Fv$$

$$= \frac{B^2 l^2 v^2}{r} \qquad (6.8)$$

The agent that does this work is mechanical. Where does this mechanical energy go? The answer is: it is dissipated as Joule heat, and is given by

$$P_J = I^2 r = \left(\frac{Blv}{r}\right)^2 r = \frac{B^2 l^2 v^2}{r}$$

which is identical to Eq. (6.8).

Thus, mechanical energy which was needed to move the arm PQ is converted into electrical energy (the induced emf) and then to thermal energy.

There is an interesting relationship between the charge flow through the circuit and the change in the magnetic flux. From Faraday's law, we have learnt that the magnitude of the induced emf is,

$$|\varepsilon| = \frac{\Delta \Phi_B}{\Delta t}$$

However,

$$|\varepsilon| = Ir = \frac{\Delta Q}{\Delta t} r$$

Thus,

$$\Delta Q = \frac{\Delta \Phi_B}{r}$$

EXAMPLE 6.8

Example 6.8 Refer to Fig. 6.12(a). The arm PQ of the rectangular conductor is moved from $x = 0$, outwards. The uniform magnetic field is perpendicular to the plane and extends from $x = 0$ to $x = b$ and is zero for $x > b$. Only the arm PQ possesses substantial resistance r. Consider the situation when the arm PQ is pulled outwards from $x = 0$ to $x = 2b$, and is then moved back to $x = 0$ with constant speed v. Obtain expressions for the flux, the induced emf, the force necessary to pull the arm and the power dissipated as Joule heat. Sketch the variation of these quantities with distance.

(a)

FIGURE 6.12

Solution Let us first consider the forward motion from $x = 0$ to $x = 2b$ The flux Φ_B linked with the circuit SPQR is

$$\Phi_B = Blx \qquad 0 \le x < b$$

$$= Blb \qquad b \le x < 2b$$

The induced emf is,

$$\varepsilon = -\frac{d\Phi_B}{dt}$$

$$= -Blv \qquad 0 \le x < b$$

$$= 0 \qquad b \le x < 2b$$

When the induced emf is non-zero, the current I is (in magnitude)

$$I = \frac{Blv}{r}$$

(b)

FIGURE 6.12

The force required to keep the arm PQ in constant motion is $I\,lB$. Its direction is to the left. In magnitude

$$F = \frac{B^2l^2v}{r} \qquad 0 \le x < b$$

$$= 0 \qquad b \le x < 2b$$

The Joule heating loss is

$$P_J = I^2 r$$

$$= \frac{B^2l^2v^2}{r} \qquad 0 \le x < b$$

$$= 0 \qquad b \le x < 2b$$

One obtains similar expressions for the inward motion from $x = 2b$ to $x = 0$. One can appreciate the whole process by examining the sketch of various quantities displayed in Fig. 6.12(b).

EXAMPLE 6.8

6.8 Eddy Currents

So far we have studied the electric currents induced in well defined paths in conductors like circular loops. Even when bulk pieces of conductors are subjected to changing magnetic flux, induced currents are produced in them. However, their flow patterns resemble swirling eddies in water. This effect was discovered by physicist Foucault (1819-1868) and these currents are called *eddy currents*.

FIGURE 6.13 Eddy currents are generated in the copper plate, while entering and leaving the region of magnetic field.

Consider the apparatus shown in Fig. 6.13. A copper plate is allowed to swing like a simple pendulum between the pole pieces of a strong magnet. It is found that the motion is damped and in a little while the plate comes to a halt in the magnetic field. We can explain this phenomenon on the basis of electromagnetic induction. Magnetic flux associated with the plate keeps on changing as the plate moves in and out of the region between magnetic poles. The flux change induces eddy currents in the plate. Directions of eddy currents are opposite when the plate swings into the region between the poles and when it swings out of the region.

If rectangular slots are made in the copper plate as shown in Fig. 6.14, area available to the flow of eddy currents is less. Thus, the pendulum plate with holes or slots reduces electromagnetic damping and the plate swings more freely. Note that magnetic moments of the induced currents (which oppose the motion) depend upon the area enclosed by the currents (recall equation $\mathbf{m} = I\mathbf{A}$ in Chapter 4).

FIGURE 6.14 Cutting slots in the copper plate reduces the effect of eddy currents.

This fact is helpful in reducing eddy currents in the metallic cores of transformers, electric motors and other such devices in which a coil is to be wound over metallic core. Eddy currents are undesirable since they heat up the core and dissipate electrical energy in the form of heat. Eddy currents are minimised by using laminations of metal to make a metal core. The laminations are separated by an insulating material like lacquer. The plane of the laminations must be arranged parallel to the magnetic field, so that they cut across the eddy current paths. This arrangement reduces the strength of the eddy currents. Since the dissipation of electrical energy into heat depends on the square of the strength of electric current, heat loss is substantially reduced.

Eddy currents are used to advantage in certain applications like:

(i) *Magnetic braking in trains*: Strong electromagnets are situated above the rails in some electrically powered trains. When the electromagnets are activated, the eddy currents induced in the rails oppose the motion of the train. As there are no mechanical linkages, the braking effect is smooth.

(ii) *Electromagnetic damping*: Certain galvanometers have a fixed core made of nonmagnetic metallic material. When the coil oscillates, the eddy currents generated in the core oppose the motion and bring the coil to rest quickly.

(iii) *Induction furnace*: Induction furnace can be used to produce high temperatures and can be utilised to prepare alloys, by melting the constituent metals. A high frequency alternating current is passed through a coil which surrounds the metals to be melted. The eddy currents generated in the metals produce high temperatures sufficient to melt it.

(iv) *Electric power meters*: The shiny metal disc in the electric power meter (analogue type) rotates due to the eddy currents. Electric currents are induced in the disc by magnetic fields produced by sinusoidally varying currents in a coil.

You can observe the rotating shiny disc in the power meter of your house.

ELECTROMAGNETIC DAMPING

Take two hollow thin cylindrical pipes of equal internal diameters made of aluminium and PVC, respectively. Fix them vertically with clamps on retort stands. Take a small cylinderical magnet having diameter slightly smaller than the inner diameter of the pipes and drop it through each pipe in such a way that the magnet does not touch the sides of the pipes during its fall. You will observe that the magnet dropped through the PVC pipe takes the same time to come out of the pipe as it would take when dropped through the same height without the pipe. Note the time it takes to come out of the pipe in each case. You will see that the magnet takes much longer time in the case of aluminium pipe. Why is it so? It is due to the eddy currents that are generated in the aluminium pipe which oppose the change in magnetic flux, i.e., the motion of the magnet. The retarding force due to the eddy currents inhibits the motion of the magnet. Such phenomena are referred to as *electromagnetic damping*. Note that eddy currents are not generated in PVC pipe as its material is an insulator whereas aluminium is a conductor.

6.9 INDUCTANCE

An electric current can be induced in a coil by flux change produced by another coil in its vicinity or flux change produced by the same coil. These two situations are described separately in the next two sub-sections. However, in both the cases, the flux through a coil is proportional to the current. That is, $\Phi_B \propto I$.

Further, if the geometry of the coil does not vary with time then,

$$\frac{d\Phi_B}{dt} \propto \frac{dI}{dt}$$

For a closely wound coil of N turns, the same magnetic flux is linked with all the turns. When the flux Φ_B through the coil changes, each turn contributes to the induced emf. Therefore, a term called *flux linkage* is used which is equal to $N\Phi_B$ for a closely wound coil and in such a case

$$N\Phi_B \propto I$$

The constant of proportionality, in this relation, is called *inductance*. We shall see that inductance depends only on the geometry of the coil

and intrinsic material properties. This aspect is akin to capacitance which for a parallel plate capacitor depends on the plate area and plate separation (geometry) and the dielectric constant K of the intervening medium (intrinsic material property).

Inductance is a scalar quantity. It has the dimensions of $[M L^2 T^{-2} A^{-2}]$ given by the dimensions of flux divided by the dimensions of current. The SI unit of inductance is *henry* and is denoted by H. It is named in honour of Joseph Henry who discovered electromagnetic induction in USA, independently of Faraday in England.

6.9.1 Mutual inductance

Consider Fig. 6.15 which shows two long co-axial solenoids each of length l. We denote the radius of the inner solenoid S_1 by r_1 and the number of turns per unit length by n_1. The corresponding quantities for the outer solenoid S_2 are r_2 and n_2, respectively. Let N_1 and N_2 be the total number of turns of coils S_1 and S_2, respectively.

When a current I_2 is set up through S_2, it in turn sets up a magnetic flux through S_1. Let us denote it by Φ_1. The corresponding flux linkage with solenoid S_1 is

$$N_1 \Phi_1 = M_{12} I_2 \tag{6.9}$$

M_{12} is called the *mutual inductance* of solenoid S_1 with respect to solenoid S_2. It is also referred to as the *coefficient of mutual induction*.

For these simple co-axial solenoids it is possible to calculate M_{12}. The magnetic field due to the current I_2 in S_2 is $\mu_0 n_2 I_2$. The resulting flux linkage with coil S_1 is,

$$N_1 \Phi_1 = (n_1 l)\left(\pi r_1^2\right)\left(\mu_0 n_2 I_2\right)$$

$$= \mu_0 n_1 n_2 \pi r_1^2 l\, I_2 \tag{6.10}$$

where $n_1 l$ is the total number of turns in solenoid S_1. Thus, from Eq. (6.9) and Eq. (6.10),

$$M_{12} = \mu_0 n_1 n_2 \pi r_1^2 l \tag{6.11}$$

FIGURE 6.15 Two long co-axial solenoids of same length l.

Note that we neglected the edge effects and considered the magnetic field $\mu_0 n_2 I_2$ to be uniform throughout the length and width of the solenoid S_2. This is a good approximation keeping in mind that the solenoid is long, implying $l >> r_2$.

We now consider the reverse case. A current I_1 is passed through the solenoid S_1 and the flux linkage with coil S_2 is,

$$N_2 \Phi_2 = M_{21}\, I_1 \tag{6.12}$$

M_{21} is called the *mutual inductance* of solenoid S_2 with respect to solenoid S_1.

The flux due to the current I_1 in S_1 can be assumed to be confined solely inside S_1 since the solenoids are very long. Thus, flux linkage with solenoid S_2 is

$$N_2 \Phi_2 = (n_2 l)\left(\pi r_1^2\right)\left(\mu_0 n_1 I_1\right)$$

where $n_2 l$ is the total number of turns of S_2. From Eq. (6.12),

$$M_{21} = \mu_0 n_1 n_2 \pi r_1^2 l \qquad (6.13)$$

Using Eq. (6.11) and Eq. (6.12), we get

$$M_{12} = M_{21} = M \text{ (say)} \qquad (6.14)$$

We have demonstrated this equality for long co-axial solenoids. However, the relation is far more general. Note that if the inner solenoid was much shorter than (and placed well inside) the outer solenoid, then we could still have calculated the flux linkage $N_1 \Phi_1$ because the inner solenoid is effectively immersed in a uniform magnetic field due to the outer solenoid. In this case, the calculation of M_{12} would be easy. However, it would be extremely difficult to calculate the flux linkage with the outer solenoid as the magnetic field due to the inner solenoid would vary across the length as well as cross section of the outer solenoid. Therefore, the calculation of M_{21} would also be extremely difficult in this case. The equality $M_{12}=M_{21}$ is very useful in such situations.

We explained the above example with air as the medium within the solenoids. Instead, if a medium of relative permeability μ_r had been present, the mutual inductance would be

$$M = \mu_r \mu_0 n_1 n_2 \pi r_1^2 l$$

It is also important to know that the mutual inductance of a pair of coils, solenoids, etc., depends on their separation as well as their relative orientation.

Example 6.9 Two concentric circular coils, one of small radius r_1 and the other of large radius r_2, such that $r_1 << r_2$, are placed co-axially with centres coinciding. Obtain the mutual inductance of the arrangement.

Solution Let a current I_2 flow through the outer circular coil. The field at the centre of the coil is $B_2 = \mu_0 I_2 / 2r_2$. Since the other co-axially placed coil has a very small radius, B_2 may be considered constant over its cross-sectional area. Hence,

$$\Phi_1 = \pi r_1^2 B_2$$

$$= \frac{\mu_0 \pi r_1^2}{2r_2} I_2$$

$$= M_{12} I_2$$

Thus,

$$M_{12} = \frac{\mu_0 \pi r_1^2}{2r_2}$$

From Eq. (6.14)

$$M_{12} = M_{21} = \frac{\mu_0 \pi r_1^2}{2r_2}$$

Note that we calculated M_{12} from an approximate value of Φ_1, assuming the magnetic field B_2 to be uniform over the area πr_1^2. However, we can accept this value because $r_1 << r_2$.

EXAMPLE 6.9

Now, let us recollect Experiment 6.3 in Section 6.2. In that experiment, emf is induced in coil C_1 wherever there was any change in current through coil C_2. Let Φ_1 be the flux through coil C_1 (say of N_1 turns) when current in coil C_2 is I_2.

Then, from Eq. (6.9), we have

$N_1\Phi_1 = MI_2$

For currents varrying with time,

$$\frac{d(N_1\Phi_1)}{dt} = \frac{d(MI_2)}{dt}$$

Since induced emf in coil C_1 is given by

$$\varepsilon_1 = -\frac{d(N_1\Phi_1)}{dt}$$

We get,

$$\varepsilon_1 = -M\frac{dI_2}{dt}$$

It shows that varying current in a coil can induce emf in a neighbouring coil. The magnitude of the induced emf depends upon the rate of change of current and mutual inductance of the two coils.

6.9.2 Self-inductance

In the previous sub-section, we considered the flux in one solenoid due to the current in the other. It is also possible that emf is induced in a single isolated coil due to change of flux through the coil by means of varying the current through the same coil. This phenomenon is called *self-induction*. In this case, flux linkage through a coil of N turns is proportional to the current through the coil and is expressed as

$$N\Phi_B \propto I$$

$$N\Phi_B = L\,I \qquad (6.15)$$

where constant of proportionality L is called *self-inductance* of the coil. It is also called the *coefficient of self-induction* of the coil. When the current is varied, the flux linked with the coil also changes and an emf is induced in the coil. Using Eq. (6.15), the induced emf is given by

$$\varepsilon = -\frac{d(N\Phi_B)}{dt}$$

$$\varepsilon = -L\frac{dI}{dt} \qquad (6.16)$$

Thus, the self-induced emf always opposes any change (increase or decrease) of current in the coil.

It is possible to calculate the self-inductance for circuits with simple geometries. Let us calculate the self-inductance of a long solenoid of cross-sectional area A and length l, having n turns per unit length. The magnetic field due to a current I flowing in the solenoid is $B = \mu_0\, n\, I$ (neglecting edge effects, as before). The total flux linked with the solenoid is

$$N\Phi_B = (nl)(\mu_0 n\, I)(A)$$

$$= \mu_0 n^2 A l I$$

where nl is the total number of turns. Thus, the self-inductance is,

$$L = \frac{N\Phi_B}{I}$$

$$= \mu_0 n^2 A l \qquad (6.17)$$

If we fill the inside of the solenoid with a material of relative permeability μ_r (for example soft iron, which has a high value of relative permiability), then,

$$L = \mu_r \, \mu_0 \, n^2 A l \qquad (6.18)$$

The self-inductance of the coil depends on its geometry and on the permeability of the medium.

The self-induced emf is also called the *back emf* as it opposes any change in the current in a circuit. Physically, the *self-inductance plays the role of inertia*. It is the electromagnetic analogue of mass in mechanics. So, work needs to be done against the back emf (ε) in establishing the current. This work done is stored as magnetic potential energy. For the current I at an instant in a circuit, the rate of work done is

$$\frac{dW}{dt} = |\varepsilon| I$$

If we ignore the resistive losses and consider only inductive effect, then using Eq. (6.16),

$$\frac{dW}{dt} = L \, I \frac{dI}{dt}$$

Total amount of work done in establishing the current I is

$$W = \int dW = \int_0^I L \, I \, dI$$

Thus, the energy required to build up the current I is,

$$W = \frac{1}{2} L I^2 \qquad (6.19)$$

This expression reminds us of $mv^2/2$ for the (mechanical) kinetic energy of a particle of mass m, and shows that L is analogus to m (i.e., L is electrical inertia and opposes growth and decay of current in the circuit).

Consider the general case of currents flowing simultaneously in two nearby coils. The flux linked with one coil will be the sum of two fluxes which exist independently. Equation (6.9) would be modified into

$$N_1 \Phi_1 = M_{11} I_1 + M_{12} I_2$$

where M_{11} represents inductance due to the same coil.

Therefore, using Faraday's law,

$$\varepsilon_1 = -M_{11} \frac{dI_1}{dt} - M_{12} \frac{dI_2}{dt}$$

M_{11} is the *self-inductance* and is written as L_1. Therefore,

$$\varepsilon_1 = -L_1 \frac{dI_1}{dt} - M_{12} \frac{dI_2}{dt}$$

Example 6.10 (a) Obtain the expression for the magnetic energy stored in a solenoid in terms of magnetic field B, area A and length l of the solenoid. (b) How does this magnetic energy compare with the electrostatic energy stored in a capacitor?

Solution

(a) From Eq. (6.19), the magnetic energy is

$$U_B = \frac{1}{2} LI^2$$

$$= \frac{1}{2} L \left(\frac{B}{\mu_0 n}\right)^2 \qquad (\text{since } B = \mu_0 nI, \text{for a solenoid})$$

$$= \frac{1}{2} (\mu_0 n^2 Al) \left(\frac{B}{\mu_0 n}\right)^2 \qquad \text{[from Eq. (6.17)]}$$

$$= \frac{1}{2\mu_0} B^2 Al$$

(b) The magnetic energy per unit volume is,

$$u_B = \frac{U_B}{V} \qquad \text{(where } V \text{ is volume that contains flux)}$$

$$= \frac{U_B}{Al}$$

$$= \frac{B^2}{2\mu_0} \qquad (6.20)$$

We have already obtained the relation for the electrostatic energy stored per unit volume in a parallel plate capacitor (refer to Chapter 2, Eq. 2.77),

$$u_E = \frac{1}{2} \varepsilon_0 E^2 \qquad (2.77)$$

In both the cases energy is proportional to the square of the field strength. Equations (6.20) and (2.77) have been derived for special cases: a solenoid and a parallel plate capacitor, respectively. But they are general and valid for any region of space in which a magnetic field or/and an electric field exist.

EXAMPLE 6.10

Interactive animation on ac generator: http://micro.magnet.fsu.edu/electromagnet~java/generator/ac.html

PHYSICS

6.10 AC GENERATOR

The phenomenon of electromagnetic induction has been technologically exploited in many ways. An exceptionally important application is the generation of alternating currents (ac). The modern ac generator with a typical output capacity of 100 MW is a highly evolved machine. In this section, we shall describe the basic principles behind this machine. The Yugoslav inventor Nicola Tesla is credited with the development of the machine. As was pointed out in Section 6.3, one method to induce an emf

or current in a loop is through a change in the loop's orientation or a change in its effective area. As the coil rotates in a magnetic field **B**, the effective area of the loop (the face perpendicular to the field) is $A \cos \theta$, where θ is the angle between **A** and **B**. This method of producing a flux change is the principle of operation of a simple ac generator. An ac generator converts mechanical energy into electrical energy.

The basic elements of an ac generator are shown in Fig. 6.16. It consists of a coil mounted on a rotor shaft. The axis of rotation of the coil is perpendicular to the direction of the magnetic field. The coil (called armature) is mechanically rotated in the uniform magnetic field by some external means. The rotation of the coil causes the magnetic flux through it to change, so an emf is induced in the coil. The ends of the coil are connected to an external circuit by means of slip rings and brushes.

FIGURE 6.16 AC Generator

When the coil is rotated with a constant angular speed ω, the angle θ between the magnetic field vector **B** and the area vector **A** of the coil at any instant t is $\theta = \omega t$ (assuming $\theta = 0°$ at $t = 0$). As a result, the effective area of the coil exposed to the magnetic field lines changes with time, and from Eq. (6.1), the flux at any time t is

$$\Phi_B = BA \cos \theta = BA \cos \omega t$$

From Faraday's law, the induced emf for the rotating coil of N turns is then,

$$\varepsilon = -N \frac{d\Phi_B}{dt} = -NBA \frac{d}{dt}(\cos \omega t)$$

Thus, the instantaneous value of the emf is

$$\varepsilon = NBA\,\omega \sin \omega t \qquad (6.21)$$

where $NBA\omega$ is the maximum value of the emf, which occurs when $\sin \omega t = \pm 1$. If we denote $NBA\omega$ as ε_0, then

$$\varepsilon = \varepsilon_0 \sin \omega t \qquad (6.22)$$

Since the value of the sine fuction varies between +1 and –1, the sign, or polarity of the emf changes with time. Note from Fig. 6.17 that the emf has its extremum value when $\theta = 90°$ or $\theta = 270°$, as the change of flux is greatest at these points.

The direction of the current changes periodically and therefore the current is called *alternating current* (ac). Since $\omega = 2\pi\nu$, Eq (6.22) can be written as

$$\varepsilon = \varepsilon_0 \sin 2\pi\, \nu\, t \qquad (6.23)$$

where ν is the frequency of revolution of the generator's coil.

Note that Eq. (6.22) and (6.23) give the instantaneous value of the emf and ε varies between $+\varepsilon_0$ and $-\varepsilon_0$ periodically. We shall learn how to determine the time-averaged value for the alternating voltage and current in the next chapter.

FIGURE 6.17 An alternating emf is generated by a loop of wire rotating in a magnetic field.

In commercial generators, the mechanical energy required for rotation of the armature is provided by water falling from a height, for example, from dams. These are called *hydro-electric generators*. Alternatively, water is heated to produce steam using coal or other sources. The steam at high pressure produces the rotation of the armature. These are called *thermal generators*. Instead of coal, if a nuclear fuel is used, we get *nuclear power generators*. Modern day generators produce electric power as high as 500 MW, i.e., one can light up 5 million 100 W bulbs! In most generators, the coils are held stationary and it is the electromagnets which are rotated. The frequency of rotation is 50 Hz in India. In certain countries such as USA, it is 60 Hz.

EXAMPLE 6.11

Example 6.11 Kamla peddles a stationary bicycle the pedals of the bicycle are attached to a 100 turn coil of area 0.10 m^2. The coil rotates at half a revolution per second and it is placed in a uniform magnetic field of 0.01 T perpendicular to the axis of rotation of the coil. What is the maximum voltage generated in the coil?

Solution Here $f = 0.5$ Hz; $N = 100$, $A = 0.1$ m^2 and $B = 0.01$ T. Employing Eq. (6.21)

$\varepsilon_0 = NBA\,(2\,\pi\,\nu)$

$= 100 \times 0.01 \times 0.1 \times 2 \times 3.14 \times 0.5$

$= 0.314$ V

The maximum voltage is 0.314 V.

We urge you to explore such alternative possibilities for power generation.

MIGRATION OF BIRDS

The migratory pattern of birds is one of the mysteries in the field of biology, and indeed all of science. For example, every winter birds from Siberia fly unerringly to water spots in the Indian subcontinent. There has been a suggestion that electromagnetic induction may provide a clue to these migratory patterns. The earth's magnetic field has existed throughout evolutionary history. It would be of great benefit to migratory birds to use this field to determine the direction. As far as we know birds contain no ferromagnetic material. So electromagnetic induction seems to be the only reasonable mechanism to determine direction. Consider the optimal case where the magnetic field **B**, the velocity of the bird **v**, and two relevant points of its anatomy separated by a distance l, all three are mutually perpendicular. From the formula for motional emf, Eq. (6.5),

$$\varepsilon = Blv$$

Taking $B = 4 \times 10^{-5}$ T, $l = 2$ cm wide, and $v = 10$ m/s, we obtain

$$\varepsilon = 4 \times 10^{-5} \times 2 \times 10^{-2} \times 10 \text{ V} = 8 \times 10^{-6} \text{ V}$$

$$= 8 \ \mu\text{V}$$

This extremely small potential difference suggests that our hypothesis is of doubtful validity. Certain kinds of fish are able to detect small potential differences. However, in these fish, special cells have been identified which detect small voltage differences. In birds no such cells have been identified. Thus, the migration patterns of birds continues to remain a mystery.

SUMMARY

1. The magnetic flux through a surface of area **A** placed in a uniform magnetic field **B** is defined as,

   $\Phi_B = \mathbf{B} \cdot \mathbf{A} = BA \cos \theta$

   where θ is the angle between **B** and **A**.
2. Faraday's laws of induction imply that the emf induced in a coil of N turns is directly related to the rate of change of flux through it,

   $$\varepsilon = -N \frac{d\Phi_B}{dt}$$

   Here Φ_B is the flux linked with one turn of the coil. If the circuit is closed, a current $I = \varepsilon/R$ is set up in it, where R is the resistance of the circuit.
3. Lenz's law states that the polarity of the induced emf is such that it tends to produce a current which opposes the change in magnetic flux that produces it. The negative sign in the expression for Faraday's law indicates this fact.
4. When a metal rod of length l is placed normal to a uniform magnetic field B and moved with a velocity v perpendicular to the field, the induced emf (called motional emf) across its ends is

   $\varepsilon = Blv$
5. Changing magnetic fields can set up current loops in nearby metal (any conductor) bodies. They dissipate electrical energy as heat. Such currents are called eddy currents.
6. Inductance is the ratio of the flux-linkage to current. It is equal to $N\Phi/I$.

7. A changing current in a coil (coil 2) can induce an emf in a nearby coil (coil 1). This relation is given by,

$$\varepsilon_1 = -M_{12}\frac{dI_2}{dt}$$

The quantity M_{12} is called mutual inductance of coil 1 with respect to coil 2. One can similarly define M_{21}. There exists a general equality,

$M_{12} = M_{21}$

8. When a current in a coil changes, it induces a back emf in the same coil. The self-induced emf is given by,

$$\varepsilon = -L\frac{dI}{dt}$$

L is the self-inductance of the coil. It is a measure of the inertia of the coil against the change of current through it.

9. The self-inductance of a long solenoid, the core of which consists of a magnetic material of permeability μ_r, is given by

$L = \mu_r \mu_0 n^2 A l$

where A is the area of cross-section of the solenoid, l its length and n the number of turns per unit length.

10. In an ac generator, mechanical energy is converted to electrical energy by virtue of electromagnetic induction. If coil of N turn and area A is rotated at ν revolutions per second in a uniform magnetic field B, then the motional emf produced is

$\varepsilon = NBA\,(2\pi\nu)\sin(2\pi\nu t)$

where we have assumed that at time $t = 0$ s, the coil is perpendicular to the field.

Quantity	Symbol	Units	Dimensions	Equations
Magnetic Flux	Φ_B	Wb (weber)	$[M L^2 T^{-2} A^{-1}]$	$\Phi_B = \mathbf{B}\cdot\mathbf{A}$
EMF	ε	V (volt)	$[M L^2 T^{-3} A^{-1}]$	$\varepsilon = -d(N\Phi_B)/dt$
Mutual Inductance	M	H (henry)	$[M L^2 T^{-2} A^{-2}]$	$\varepsilon_1 = -M_{12}(dI_2/dt)$
Self Inductance	L	H (henry)	$[M L^2 T^{-2} A^{-2}]$	$\varepsilon = -L(dI/dt)$

POINTS TO PONDER

1. Electricity and magnetism are intimately related. In the early part of the nineteenth century, the experiments of Oersted, Ampere and others established that moving charges (currents) produce a magnetic field. Somewhat later, around 1830, the experiments of Faraday and Henry demonstrated that a moving magnet can induce electric current.
2. In a closed circuit, electric currents are induced so as to oppose the changing magnetic flux. It is as per the law of conservation of energy. However, in case of an open circuit, an emf is induced across its ends. How is it related to the flux change?
3. The motional emf discussed in Section 6.5 can be argued independently from Faraday's law using the Lorentz force on moving charges. However,

even if the charges are stationary [and the q ($\mathbf{v} \times \mathbf{B}$) term of the Lorentz force is not operative], an emf is nevertheless induced in the presence of a time-varying magnetic field. Thus, moving charges in static field and static charges in a time-varying field seem to be symmetric situation for Faraday's law. This gives a tantalising hint on the relevance of the principle of relativity for Faraday's law.

4. The motion of a copper plate is damped when it is allowed to oscillate between the magnetic pole-pieces. How is the damping force, produced by the eddy currents?

EXERCISES

6.1 Predict the direction of induced current in the situations described by the following Figs. 6.18(a) to (f).

FIGURE 6.18

6.2 Use Lenz's law to determine the direction of induced current in the situations described by Fig. 6.19:
(a) A wire of irregular shape turning into a circular shape;
(b) A circular loop being deformed into a narrow straight wire.

FIGURE 6.19

6.3 A long solenoid with 15 turns per cm has a small loop of area 2.0 cm^2 placed inside the solenoid normal to its axis. If the current carried by the solenoid changes steadily from 2.0 A to 4.0 A in 0.1 s, what is the induced emf in the loop while the current is changing?

6.4 A rectangular wire loop of sides 8 cm and 2 cm with a small cut is moving out of a region of uniform magnetic field of magnitude 0.3 T directed normal to the loop. What is the emf developed across the cut if the velocity of the loop is 1 cm s^{-1} in a direction normal to the (a) longer side, (b) shorter side of the loop? For how long does the induced voltage last in each case?

6.5 A 1.0 m long metallic rod is rotated with an angular frequency of 400 rad s^{-1} about an axis normal to the rod passing through its one end. The other end of the rod is in contact with a circular metallic ring. A constant and uniform magnetic field of 0.5 T parallel to the axis exists everywhere. Calculate the emf developed between the centre and the ring.

6.6 A circular coil of radius 8.0 cm and 20 turns is rotated about its vertical diameter with an angular speed of 50 rad s^{-1} in a uniform horizontal magnetic field of magnitude 3.0×10^{-2} T. Obtain the maximum and average emf induced in the coil. If the coil forms a closed loop of resistance 10 Ω, calculate the maximum value of current in the coil. Calculate the average power loss due to Joule heating. Where does this power come from?

6.7 A horizontal straight wire 10 m long extending from east to west is falling with a speed of 5.0 m s^{-1}, at right angles to the horizontal component of the earth's magnetic field, 0.30×10^{-4} Wb m^{-2}.
(a) What is the instantaneous value of the emf induced in the wire?
(b) What is the direction of the emf?
(c) Which end of the wire is at the higher electrical potential?

6.8 Current in a circuit falls from 5.0 A to 0.0 A in 0.1 s. If an average emf of 200 V induced, give an estimate of the self-inductance of the circuit.

6.9 A pair of adjacent coils has a mutual inductance of 1.5 H. If the current in one coil changes from 0 to 20 A in 0.5 s, what is the change of flux linkage with the other coil?

6.10 A jet plane is travelling towards west at a speed of 1800 km/h. What is the voltage difference developed between the ends of the wing

having a span of 25 m, if the Earth's magnetic field at the location has a magnitude of 5×10^{-4} T and the dip angle is 30°.

ADDITIONAL EXERCISES

6.11 Suppose the loop in Exercise 6.4 is stationary but the current feeding the electromagnet that produces the magnetic field is gradually reduced so that the field decreases from its initial value of 0.3 T at the rate of 0.02 T s^{-1}. If the cut is joined and the loop has a resistance of 1.6 Ω, how much power is dissipated by the loop as heat? What is the source of this power?

6.12 A square loop of side 12 cm with its sides parallel to X and Y axes is moved with a velocity of 8 cm s^{-1} in the positive *x*-direction in an environment containing a magnetic field in the positive *z*-direction. The field is neither uniform in space nor constant in time. It has a gradient of 10^{-3} T cm^{-1} along the negative *x*-direction (that is it increases by 10^{-3} T cm^{-1} as one moves in the negative *x*-direction), and it is decreasing in time at the rate of 10^{-3} T s^{-1}. Determine the direction and magnitude of the induced current in the loop if its resistance is 4.50 mΩ.

6.13 It is desired to measure the magnitude of field between the poles of a powerful loud speaker magnet. A small flat search coil of area 2 cm^2 with 25 closely wound turns, is positioned normal to the field direction, and then quickly snatched out of the field region. Equivalently, one can give it a quick 90° turn to bring its plane parallel to the field direction). The total charge flown in the coil (measured by a ballistic galvanometer connected to coil) is 7.5 mC. The combined resistance of the coil and the galvanometer is 0.50 Ω. Estimate the field strength of magnet.

6.14 Figure 6.20 shows a metal rod PQ resting on the smooth rails AB and positioned between the poles of a permanent magnet. The rails, the rod, and the magnetic field are in three mutual perpendicular directions. A galvanometer G connects the rails through a switch K. Length of the rod = 15 cm, B = 0.50 T, resistance of the closed loop containing the rod = 9.0 mΩ. Assume the field to be uniform.

(a) Suppose K is open and the rod is moved with a speed of 12 cm s^{-1} in the direction shown. Give the polarity and magnitude of the induced emf.

FIGURE 6.20

(b) Is there an excess charge built up at the ends of the rods when K is open? What if K is closed?

(c) With K open and the rod moving uniformly, there is *no net force* on the electrons in the rod PQ even though they do

experience magnetic force due to the motion of the rod. Explain.

(d) What is the retarding force on the rod when K is closed?

(e) How much power is required (by an external agent) to keep the rod moving at the same speed (=12 cm s^{-1}) when K is closed? How much power is required when K is open?

(f) How much power is dissipated as heat in the closed circuit? What is the source of this power?

(g) What is the induced emf in the moving rod if the magnetic field is parallel to the rails instead of being perpendicular?

6.15 An air-cored solenoid with length 30 cm, area of cross-section 25 cm^2 and number of turns 500, carries a current of 2.5 A. The current is suddenly switched off in a brief time of 10^{-3} s. How much is the average back emf induced across the ends of the open switch in the circuit? Ignore the variation in magnetic field near the ends of the solenoid.

6.16 (a) Obtain an expression for the mutual inductance between a long straight wire and a square loop of side a as shown in Fig. 6.21.

(b) Now assume that the straight wire carries a current of 50 A and the loop is moved to the right with a constant velocity, v = 10 m/s. Calculate the induced emf in the loop at the instant when x = 0.2 m. Take a = 0.1 m and assume that the loop has a large resistance.

FIGURE 6.21

6.17 A line charge λ per unit length is lodged uniformly onto the rim of a wheel of mass M and radius R. The wheel has light non-conducting spokes and is free to rotate without friction about its axis (Fig. 6.22). A uniform magnetic field extends over a circular region within the rim. It is given by,

$$\mathbf{B} = -B_0\mathbf{k} \qquad (r \le a;\ a < R)$$
$$= 0 \qquad \text{(otherwise)}$$

What is the angular velocity of the wheel after the field is suddenly switched off?

FIGURE 6.22

Chapter Seven

ALTERNATING CURRENT

7.1 Introduction

We have so far considered direct current (dc) sources and circuits with dc sources. These currents do not change direction with time. But voltages and currents that vary with time are very common. The electric mains supply in our homes and offices is a voltage that varies like a sine function with time. Such a voltage is called *alternating voltage* (ac voltage) and the current driven by it in a circuit is called the *alternating current* (ac current)*. Today, most of the electrical devices we use require ac voltage. This is mainly because most of the electrical energy sold by power companies is transmitted and distributed as alternating current. The main reason for preferring use of ac voltage over dc voltage is that ac voltages can be easily and efficiently converted from one voltage to the other by means of transformers. Further, electrical energy can also be transmitted economically over long distances. AC circuits exhibit characteristics which are exploited in many devices of daily use. For example, whenever we tune our radio to a favourite station, we are taking advantage of a special property of ac circuits – one of many that you will study in this chapter.

* The phrases *ac voltage* and *ac current* are contradictory and redundant, respectively, since they mean, literally, *alternating current voltage* and *alternating current current*. Still, the abbreviation *ac* to designate an electrical quantity displaying simple harmonic time dependance has become so universally accepted that we follow others in its use. Further, *voltage* – another phrase commonly used means potential difference between two points.

NICOLA TESLA (1836 – 1943)

Nicola Tesla (1836 – 1943) Yugoslov scientist, inventor and genius. He conceived the idea of the rotating magnetic field, which is the basis of practically all alternating current machinery, and which helped usher in the age of electric power. He also invented among other things the induction motor, the polyphase system of ac power, and the high frequency induction coil (the Tesla coil) used in radio and television sets and other electronic equipment. The SI unit of magnetic field is named in his honour.

7.2 AC Voltage Applied to a Resistor

Figure 7.1 shows a resistor connected to a source ε of ac voltage. The symbol for an ac source in a circuit diagram is ⊖. We consider a source which produces sinusoidally varying potential difference across its terminals. Let this potential difference, also called ac voltage, be given by

$$v = v_m \sin \omega t \tag{7.1}$$

where v_m is the amplitude of the oscillating potential difference and ω is its angular frequency.

FIGURE 7.1 AC voltage applied to a resistor.

To find the value of current through the resistor, we apply Kirchhoff's loop rule $\sum \varepsilon(t) = 0$, to the circuit shown in Fig. 7.1 to get

$$v_m \sin \omega t = i R$$

or $$i = \frac{v_m}{R} \sin \omega t$$

Since R is a constant, we can write this equation as

$$i = i_m \sin \omega t \tag{7.2}$$

where the current amplitude i_m is given by

$$i_m = \frac{v_m}{R} \tag{7.3}$$

FIGURE 7.2 In a pure resistor, the voltage and current are in phase. The minima, zero and maxima occur at the same respective times.

Equation (7.3) is just Ohm's law which for resistors works equally well for both ac and dc voltages. The voltage across a pure resistor and the current through it, given by Eqs. (7.1) and (7.2) are plotted as a function of time in Fig. 7.2. Note, in particular that both v and i reach zero, minimum and maximum values at the same time. Clearly, *the voltage* and *current are in phase with each other*.

We see that, like the applied voltage, the current varies sinusoidally and has corresponding positive and negative values during each cycle. Thus, the sum of the instantaneous current values over one complete cycle is zero, and the average current is zero. The fact that the average current is zero, however, does

not mean that the average power consumed is zero and that there is no dissipation of electrical energy. As you know, Joule heating is given by i^2R and depends on i^2 (which is always positive whether i is positive or negative) and not on i. Thus, there is Joule heating and dissipation of electrical energy when an ac current passes through a resistor.

The instantaneous power dissipated in the resistor is

$$p = i^2R = i_m^2 R \sin^2 \omega t \qquad (7.4)$$

The average value of p over a cycle is*

$$\bar{p} = \langle i^2R \rangle = \langle i_m^2 R \sin^2 \omega t \rangle \qquad [7.5(a)]$$

where the bar over a letter(here, p) denotes its average value and <.....> denotes taking average of the quantity inside the bracket. Since, i_m^2 and R are constants,

$$\bar{p} = i_m^2 R \langle \sin^2 \omega t \rangle \qquad [7.5(b)]$$

Using the trigonometric identity, $\sin^2 \omega t = 1/2\,(1 - \cos 2\omega t)$, we have $\langle \sin^2 \omega t \rangle = (1/2)\,(1 - \langle \cos 2\omega t \rangle)$ and since $\langle \cos 2\omega t \rangle = 0$**, we have,

$$\langle \sin^2 \omega t \rangle = \frac{1}{2}$$

Thus,

$$\bar{p} = \frac{1}{2} i_m^2 R \qquad [7.5(c)]$$

To express ac power in the same form as dc power ($P = I^2R$), a special value of current is defined and used. It is called, *root mean square* (rms) or *effective current* (Fig. 7.3) and is denoted by I_{rms} *or* I.

FIGURE 7.3 The rms current I is related to the peak current i_m by $I = i_m / \sqrt{2} = 0.707\ i_m$.

George Westinghouse (1846 – 1914) A leading proponent of the use of alternating current over direct current. Thus, he came into conflict with Thomas Alva Edison, an advocate of direct current. Westinghouse was convinced that the technology of alternating current was the key to the electrical future. He founded the famous Company named after him and enlisted the services of Nicola Tesla and other inventors in the development of alternating current motors and apparatus for the transmission of high tension current, pioneering in large scale lighting.

GEORGE WESTINGHOUSE (1846 – 1914)

* The average value of a function $F(t)$ over a period T is given by $\langle F(t) \rangle = \frac{1}{T}\int_0^T F(t)\,dt$

** $\langle \cos 2\omega t \rangle = \frac{1}{T}\int_0^T \cos 2\omega t\, dt = \frac{1}{T}\left[\frac{\sin 2\omega t}{2\omega}\right]_0^T = \frac{1}{2\omega T}[\sin 2\omega T - 0] = 0$

It is defined by

$$I = \sqrt{\overline{i^2}} = \sqrt{\frac{1}{2} i_m^2} = \frac{i_m}{\sqrt{2}}$$

$$= 0.707\ i_m \tag{7.6}$$

In terms of I, the average power, denoted by P is

$$P = \overline{p} = \frac{1}{2} i_m^2 R = I^2 R \tag{7.7}$$

Similarly, we define the *rms voltage* or *effective voltage* by

$$V = \frac{v_m}{\sqrt{2}} = 0.707\ v_m \tag{7.8}$$

From Eq. (7.3), we have

$$v_m = i_m R$$

or, $$\frac{v_m}{\sqrt{2}} = \frac{i_m}{\sqrt{2}} R$$

or, $$V = IR \tag{7.9}$$

Equation (7.9) gives the relation between ac current and ac voltage and is similar to that in the dc case. This shows the advantage of introducing the concept of rms values. In terms of rms values, the equation for power [Eq. (7.7)] and relation between current and voltage in ac circuits are essentially the same as those for the dc case.

It is customary to measure and specify rms values for ac quantities. For example, the household line voltage of 220 V is an rms value with a peak voltage of

$$v_m = \sqrt{2}\ V = (1.414)(220\ \text{V}) = 311\ \text{V}$$

In fact, the I or rms current is the equivalent dc current that would produce the same average power loss as the alternating current. Equation (7.7) can also be written as

$$P = V^2 / R = I V \quad (\text{since } V = I R)$$

EXAMPLE 7.1

Example 7.1 A light bulb is rated at 100W for a 220 V supply. Find (a) the resistance of the bulb; (b) the peak voltage of the source; and (c) the rms current through the bulb.

Solution

(a) We are given P = 100 W and V = 220 V. The resistance of the bulb is

$$R = \frac{V^2}{P} = \frac{(220\ \text{V})^2}{100\ \text{W}} = 484\Omega$$

(b) The peak voltage of the source is

$$v_m = \sqrt{2} V = 311\ \text{V}$$

(c) Since, $P = I V$

$$I = \frac{P}{V} = \frac{100\ \text{W}}{220\ \text{V}} = 0.450\text{A}$$

7.3 Representation of AC Current and Voltage by Rotating Vectors — Phasors

In the previous section, we learnt that the current through a resistor is in phase with the ac voltage. But this is not so in the case of an inductor, a capacitor or a combination of these circuit elements. In order to show phase relationship between voltage and current in an ac circuit, we use the notion of *phasors*. The analysis of an ac circuit is facilitated by the use of a phasor diagram. A phasor* is a vector which rotates about the origin with angular speed ω, as shown in Fig. 7.4. The vertical components of phasors **V** and **I** represent the sinusoidally varying quantities v and i. The magnitudes of phasors **V** and **I** represent the amplitudes or the peak values v_m and i_m of these oscillating quantities. Figure 7.4(a) shows the voltage and current phasors and their relationship at time t_1 for the case of an ac source connected to a resistor i.e., corresponding to the circuit shown in Fig. 7.1. The projection of voltage and current phasors on vertical axis, i.e., $v_m \sin\omega t$ and $i_m \sin\omega t$, respectively represent the value of voltage and current at that instant. As they rotate with frequency ω, curves in Fig. 7.4(b) are generated.

FIGURE 7.4 (a) A phasor diagram for the circuit in Fig 7.1. (b) Graph of v and i versus ωt.

From Fig. 7.4(a) we see that phasors **V** and **I** for the case of a resistor are in the same direction. This is so for all times. This means that the phase angle between the voltage and the current is zero.

7.4 AC Voltage Applied to an Inductor

Figure 7.5 shows an ac source connected to an inductor. Usually, inductors have appreciable resistance in their windings, but we shall assume that this inductor has negligible resistance. Thus, the circuit is a purely inductive ac circuit. Let the voltage across the source be $v = v_m \sin\omega t$. Using the Kirchhoff's loop rule, $\sum \varepsilon(t) = 0$, and since there is no resistor in the circuit,

$$v - L\frac{di}{dt} = 0 \qquad (7.10)$$

where the second term is the self-induced Faraday emf in the inductor; and L is the self-inductance of

FIGURE 7.5 An ac source connected to an inductor.

* Though voltage and current in ac circuit are represented by phasors – rotating vectors, they are not vectors themselves. They are scalar quantities. It so happens that the amplitudes and phases of harmonically varying scalars combine mathematically in the same way as do the projections of rotating vectors of corresponding magnitudes and directions. The *rotating vectors* that *represent* harmonically varying scalar quantities are introduced only to provide us with a simple way of adding these quantities using a rule that we already know.

the inductor. The negative sign follows from Lenz's law (Chapter 6). Combining Eqs. (7.1) and (7.10), we have

$$\frac{di}{dt} = \frac{v}{L} = \frac{v_m}{L}\sin\omega t \tag{7.11}$$

Equation (7.11) implies that the equation for $i(t)$, the current as a function of time, must be such that its slope di/dt is a sinusoidally varying quantity, with the same phase as the source voltage and an amplitude given by v_m/L. To obtain the current, we integrate di/dt with respect to time:

$$\int \frac{di}{dt}dt = \frac{v_m}{L}\int \sin(\omega t)dt$$

and get,

$$i = -\frac{v_m}{\omega L}\cos(\omega t) + \text{constant}$$

The integration constant has the dimension of current and is time-independent. Since the source has an emf which oscillates symmetrically about zero, the current it sustains also oscillates symmetrically about zero, so that no constant or time-independent component of the current exists. Therefore, the integration constant is zero.

Using

$$-\cos(\omega t) = \sin\left(\omega t - \frac{\pi}{2}\right), \text{ we have}$$

$$i = i_m \sin\left(\omega t - \frac{\pi}{2}\right) \tag{7.12}$$

where $i_m = \frac{v_m}{\omega L}$ is the amplitude of the current. The quantity ωL is analogous to the resistance and is called *inductive reactance*, denoted by X_L:

$$X_L = \omega L \tag{7.13}$$

The amplitude of the current is, then

$$i_m = \frac{v_m}{X_L} \tag{7.14}$$

The dimension of inductive reactance is the same as that of resistance and its SI unit is ohm (Ω). The inductive reactance limits the current in a purely inductive circuit in the same way as the resistance limits the current in a purely resistive circuit. The inductive reactance is directly proportional to the inductance and to the frequency of the current.

A comparison of Eqs. (7.1) and (7.12) for the source voltage and the current in an inductor shows that the current lags the voltage by $\pi/2$ or one-quarter (1/4) cycle. Figure 7.6 (a) shows the voltage and the current phasors in the present case at instant t_1. The current phasor **I** is $\pi/2$ behind the voltage phasor **V**. When rotated with frequency ω counter-clockwise, they generate the voltage and current given by Eqs. (7.1) and (7.12), respectively and as shown in Fig. 7.6(b).

Interactive animation on Phasor diagrams of ac circuits containing, R, L, C and RLC series circuits:
http://www.phys.unsw.edu.au/~jw/AC.html

(a) (b)

FIGURE 7.6 (a) A Phasor diagram for the circuit in Fig. 7.5. (b) Graph of v and i versus ωt.

We see that the current reaches its maximum value later than the voltage by one-fourth of a period $\left[\frac{T}{4} = \frac{\pi/2}{\omega}\right]$. You have seen that an inductor has reactance that limits current similar to resistance in a dc circuit. Does it also consume power like a resistance? Let us try to find out.

The instantaneous power supplied to the inductor is

$$p_L = iv = i_m \sin\left(\omega t - \frac{\pi}{2}\right) \times v_m \sin(\omega t)$$

$$= -i_m v_m \cos(\omega t)\sin(\omega t)$$

$$= -\frac{i_m v_m}{2}\sin(2\omega t)$$

So, the average power over a complete cycle is

$$P_L = \left\langle -\frac{i_m v_m}{2}\sin(2\omega t)\right\rangle$$

$$= -\frac{i_m v_m}{2}\langle \sin(2\omega t)\rangle = 0,$$

since the average of sin ($2\omega t$) over a complete cycle is zero.

Thus, the *average power supplied to an inductor over one complete cycle is zero.*

Figure 7.7 explains it in detail.

Example 7.2 A pure inductor of 25.0 mH is connected to a source of 220 V. Find the inductive reactance and rms current in the circuit if the frequency of the source is 50 Hz.

Solution The inductive reactance,

$$X_L = 2\pi\nu L = 2 \times 3.14 \times 50 \times 25 \times 10^{-3}\ \text{W}$$
$$= 7.85\Omega$$

The rms current in the circuit is

$$I = \frac{V}{X_L} = \frac{220\,\text{V}}{7.85\,\Omega} = 28\text{A}$$

EXAMPLE 7.2

0-1 Current i through the coil entering at A increase from zero to a maximum value. Flux lines are set up i.e., the core gets magnetised. With the polarity shown voltage and current are both positive. So their product p is positive. ENERGY IS ABSORBED FROM THE SOURCE.

1-2 Current in the coil is still positive but is decreasing. The core gets demagnetised and the net flux becomes zero at the end of a half cycle. The voltage v is negative (since di/dt is negative). The product of voltage and current is negative, and ENERGY IS BEING RETURNED TO SOURCE.

One complete cycle of voltage/current. Note that the current lags the voltage.

2-3 Current i becomes negative i.e., it enters at B and comes out of A. Since the direction of current has changed, the polarity of the magnet changes. The current and voltage are both negative. So their product p is positive. ENERGY IS ABSORBED.

3-4 Current i decreases and reaches its zero value at 4 when core is demagnetised and flux is zero. The voltage is positive but the current is negative. The power is, therefore, negative. ENERGY ABSORBED DURING THE 1/4 CYCLE 2-3 IS RETURNED TO THE SOURCE.

FIGURE 7.7 Magnetisation and demagnetisation of an inductor.

7.5 AC Voltage Applied to a Capacitor

Figure 7.8 shows an ac source ε generating ac voltage $v = v_m \sin \omega t$ connected to a capacitor only, a purely capacitive ac circuit.

When a capacitor is connected to a voltage source in a dc circuit, current will flow for the short time required to charge the capacitor. As charge accumulates on the capacitor plates, the voltage across them increases, opposing the current. That is, a capacitor in a dc circuit will limit or oppose the current as it charges. When the capacitor is fully charged, the current in the circuit falls to zero.

FIGURE 7.8 An ac source connected to a capacitor.

When the capacitor is connected to an ac source, as in Fig. 7.8, it limits or regulates the current, but does not completely prevent the flow of charge. The capacitor is alternately charged and discharged as the current reverses each half cycle. Let q be the charge on the capacitor at any time t. The instantaneous voltage v across the capacitor is

$$v = \frac{q}{C} \tag{7.15}$$

From the Kirchhoff's loop rule, the voltage across the source and the capacitor are equal,

$$v_m \sin \omega t = \frac{q}{C}$$

To find the current, we use the relation $i = \frac{dq}{dt}$

$$i = \frac{d}{dt}\left(v_m C \sin \omega t\right) = \omega C v_m \cos(\omega t)$$

Using the relation, $\cos(\omega t) = \sin\left(\omega t + \frac{\pi}{2}\right)$, we have

$$i = i_m \sin\left(\omega t + \frac{\pi}{2}\right) \tag{7.16}$$

where the amplitude of the oscillating current is $i_m = \omega C v_m$. We can rewrite it as

$$i_m = \frac{v_m}{(1/\omega C)}$$

Comparing it to $i_m = v_m/R$ for a purely resistive circuit, we find that $(1/\omega C)$ plays the role of resistance. It is called *capacitive reactance* and is denoted by X_c,

$$X_c = 1/\omega C \tag{7.17}$$

so that the amplitude of the current is

$$i_m = \frac{v_m}{X_C} \tag{7.18}$$

FIGURE 7.9 (a) A Phasor diagram for the circuit in Fig. 7.8. (b) Graph of v and i versus ωt.

The dimension of capacitive reactance is the same as that of resistance and its SI unit is ohm (Ω). The capacitive reactance limits the amplitude of the current in a purely capacitive circuit in the same way as the resistance limits the current in a purely resistive circuit. But it is inversely proportional to the frequency and the capacitance.

A comparison of Eq. (7.16) with the equation of source voltage, Eq. (7.1) shows that the current is $\pi/2$ ahead of voltage. Figure 7.9(a) shows the phasor diagram at an instant t_1. Here the current phasor **I** is $\pi/2$ ahead of the voltage phasor **V** as they rotate counterclockwise. Figure 7.9(b) shows the variation of voltage and current with time. We see that the current reaches its maximum value earlier than the voltage by one-fourth of a period.

The instantaneous power supplied to the capacitor is

$$p_c = i\,v = i_m \cos(\omega t) v_m \sin(\omega t)$$
$$= i_m v_m \cos(\omega t) \sin(\omega t)$$
$$= \frac{i_m v_m}{2} \sin(2\omega t) \qquad (7.19)$$

So, as in the case of an inductor, the average power

$$P_C = \left\langle \frac{i_m v_m}{2} \sin(2\omega t) \right\rangle = \frac{i_m v_m}{2} \langle \sin(2\omega t) \rangle = 0$$

since <sin $(2\omega t)$> = 0 over a complete cycle. Figure 7.10 explains it in detail. Thus, we see that in the case of an inductor, the current lags the voltage by $\pi/2$ and in the case of a capacitor, the current leads the voltage by $\pi/2$.

EXAMPLE 7.3

Example 7.3 A lamp is connected in series with a capacitor. Predict your observations for dc and ac connections. What happens in each case if the capacitance of the capacitor is reduced?

Solution When a dc source is connected to a capacitor, the capacitor gets charged and after charging no current flows in the circuit and the lamp will not glow. There will be no change even if C is reduced. With ac source, the capacitor offers capacitative reactance $(1/\omega C)$ and the current flows in the circuit. Consequently, the lamp will shine. Reducing C will increase reactance and the lamp will shine less brightly than before.

EXAMPLE 7.4

Example 7.4 A 15.0 μF capacitor is connected to a 220 V, 50 Hz source. Find the capacitive reactance and the current (rms and peak) in the circuit. If the frequency is doubled, what happens to the capacitive reactance and the current?

Solution The capacitive reactance is

$$X_C = \frac{1}{2\pi \nu C} = \frac{1}{2\pi(50\text{Hz})(15.0 \times 10^{-6}\text{F})} = 212\,\Omega$$

The rms current is

0-1 The current i flows as shown and from the maximum at 0, reaches a zero value at 1. The plate A is charged to positive polarity while negative charge q builds up in B reaching a maximum at 1 until the current becomes zero. The voltage $v_c = q/C$ is in phase with q and reaches maximum value at *1*. Current and voltage are both positive. So $p = v_c i$ is positive. ENERGY IS ABSORBED FROM THE SOURCE DURING THIS QUARTER CYCLE AS THE CAPACITOR IS CHARGED.

1-2 The current i reverses its direction. The accumulated charge is depleted i.e., the capacitor is discharged during this quarter cycle.The voltage gets reduced but is still positive. The current is negative. Their product, the power is negative.
THE ENERGY ABSORBED DURING THE 1/4 CYCLE **0-1** IS RETURNED DURING THIS QUARTER.

One complete cycle of voltage/current. Note that the current leads the voltage.

2-3 As i continues to flow from A to B, the capacitor is charged to reversed polarity i.e., the plate B acquires positive and A acquires negative charge. Both the current and the voltage are negative. Their product p is positive. The capacitor ABSORBS ENERGY during this 1/4 cycle.

3-4 The current i reverses its direction at **3** and flows from B to A. The accumulated charge is depleted and the magnitude of the voltage v_c is reduced. v_c becomes zero at **4** when the capacitor is fully discharged. The power is negative.ENERGY ABSORBED DURING **2-3** IS RETURNED TO THE SOURCE. NET ENERGY ABSORBED IS ZERO.

FIGURE 7.10 Charging and discharging of a capacitor.

EXAMPLE 7.4

$$I = \frac{V}{X_C} = \frac{220\,\text{V}}{212\,\Omega} = 1.04\,\text{A}$$

The peak current is

$$i_m = \sqrt{2}I = (1.41)(1.04\,A) = 1.47A$$

This current oscillates between +1.47A and –1.47 A, and is ahead of the voltage by $\pi/2$.

If the frequency is doubled, the capacitive reactance is halved and consequently, the current is doubled.

EXAMPLE 7.5

Example 7.5 A light bulb and an open coil inductor are connected to an ac source through a key as shown in Fig. 7.11.

FIGURE 7.11

The switch is closed and after sometime, an iron rod is inserted into the interior of the inductor. The glow of the light bulb (a) increases; (b) decreases; (c) is unchanged, as the iron rod is inserted. Give your answer with reasons.

Solution As the iron rod is inserted, the magnetic field inside the coil magnetizes the iron increasing the magnetic field inside it. Hence, the inductance of the coil increases. Consequently, the inductive reactance of the coil increases. As a result, a larger fraction of the applied ac voltage appears across the inductor, leaving less voltage across the bulb. Therefore, the glow of the light bulb decreases.

7.6 AC VOLTAGE APPLIED TO A SERIES LCR CIRCUIT

Figure 7.12 shows a series LCR circuit connected to an ac source ε. As usual, we take the voltage of the source to be $v = v_m \sin \omega t$.

FIGURE 7.12 A series *LCR* circuit connected to an ac source.

If q is the charge on the capacitor and i the current, at time t, we have, from Kirchhoff's loop rule:

$$L\frac{di}{dt} + iR + \frac{q}{C} = v \qquad (7.20)$$

We want to determine the instantaneous current i and its phase relationship to the applied alternating voltage v. We shall solve this problem by two methods. First, we use the technique of phasors and in the second method, we solve Eq. (7.20) analytically to obtain the time–dependence of i.

7.6.1 Phasor-diagram solution

From the circuit shown in Fig. 7.12, we see that the resistor, inductor and capacitor are in series. Therefore, the ac current in each element is the same at any time, having the same amplitude and phase. Let it be

$$i = i_m \sin(\omega t + \phi) \tag{7.21}$$

where ϕ is the phase difference between the voltage across the source and the current in the circuit. On the basis of what we have learnt in the previous sections, we shall construct a phasor diagram for the present case.

Let **I** be the phasor representing the current in the circuit as given by Eq. (7.21). Further, let $\mathbf{V_L}$, $\mathbf{V_R}$, $\mathbf{V_C}$, and **V** represent the voltage across the inductor, resistor, capacitor and the source, respectively. From previous section, we know that $\mathbf{V_R}$ is parallel to **I**, $\mathbf{V_C}$ is $\pi/2$ behind **I** and $\mathbf{V_L}$ is $\pi/2$ ahead of **I**. $\mathbf{V_L}$, $\mathbf{V_R}$, $\mathbf{V_C}$ and **I** are shown in Fig. 7.13(a) with apppropriate phase-relations.

The length of these phasors or the amplitude of $\mathbf{V_R}$, $\mathbf{V_C}$ and $\mathbf{V_L}$ are:

$$v_{Rm} = i_m R,\ v_{Cm} = i_m X_C,\ v_{Lm} = i_m X_L \tag{7.22}$$

The voltage Equation (7.20) for the circuit can be written as

$$v_L + v_R + v_C = v \tag{7.23}$$

The phasor relation whose vertical component gives the above equation is

$$\mathbf{V_L} + \mathbf{V_R} + \mathbf{V_C} = \mathbf{V} \tag{7.24}$$

FIGURE 7.13 (a) Relation between the phasors $\mathbf{V}_L$, $\mathbf{V}_R$, $\mathbf{V}_C$, and *I*, (b) Relation between the phasors $\mathbf{V}_L$, $\mathbf{V}_R$, and $(\mathbf{V}_L + \mathbf{V}_C)$ for the circuit in Fig. 7.11.

This relation is represented in Fig. 7.13(b). Since $\mathbf{V_C}$ and $\mathbf{V_L}$ are always along the same line and in opposite directions, they can be combined into a single phasor $(\mathbf{V_C} + \mathbf{V_L})$ which has a magnitude $|v_{Cm} - v_{Lm}|$. Since **V** is represented as the hypotenuse of a right-traingle whose sides are $\mathbf{V_R}$ and $(\mathbf{V_C} + \mathbf{V_L})$, the pythagorean theorem gives:

$$v_m^2 = v_{Rm}^2 + \left(v_{Cm} - v_{Lm}\right)^2$$

Substituting the values of v_{Rm}, v_{Cm}, and v_{Lm} from Eq. (7.22) into the above equation, we have

$$v_m^2 = (i_m R)^2 + (i_m X_C - i_m X_L)^2$$

$$= i_m^2 \left[R^2 + (X_C - X_L)^2\right]$$

or, $$i_m = \frac{v_m}{\sqrt{R^2 + (X_C - X_L)^2}} \tag{7.25(a)}$$

By analogy to the resistance in a circuit, we introduce the *impedance Z* in an ac circuit:

$$i_m = \frac{v_m}{Z} \tag{7.25(b)}$$

where $$Z = \sqrt{R^2 + (X_C - X_L)^2} \tag{7.26}$$

FIGURE 7.14 Impedance diagram.

Since phasor **I** is always parallel to phasor $\mathbf{V_R}$, the phase angle ϕ is the angle between $\mathbf{V_R}$ and **V** and can be determined from Fig. 7.14:

$$\tan\phi = \frac{v_{Cm} - v_{Lm}}{v_{Rm}}$$

Using Eq. (7.22), we have

$$\tan\phi = \frac{X_C - X_L}{R} \tag{7.27}$$

Equations (7.26) and (7.27) are graphically shown in Fig. (7.14). This is called *Impedance diagram* which is a right-triangle with Z as its hypotenuse.

Equation 7.25(a) gives the amplitude of the current and Eq. (7.27) gives the phase angle. With these, Eq. (7.21) is completely specified.

If $X_C > X_L$, ϕ is positive and the circuit is predominantly capacitive. Consequently, the current in the circuit leads the source voltage. If $X_C < X_L$, ϕ is negative and the circuit is predominantly inductive. Consequently, the current in the circuit lags the source voltage.

Figure 7.15 shows the phasor diagram and variation of v and i with ωt for the case $X_C > X_L$.

FIGURE 7.15 (a) Phasor diagram of **V** and **I**. (b) Graphs of v and i versus ωt for a series *LCR* circuit where $X_C > X_L$.

Thus, we have obtained the amplitude and phase of current for an *LCR* series circuit using the technique of phasors. But this method of analysing ac circuits suffers from certain disadvantages. First, the phasor diagram say nothing about the initial condition. One can take any arbitrary value of t (say, t_1, as done throughout this chapter) and draw different phasors which show the relative angle between different phasors. The solution so obtained is called the *steady-state solution*. This is not a general solution. Additionally, we do have a *transient solution* which exists even for $v = 0$. The general solution is the sum of the transient solution and the steady-state solution. After a sufficiently long time, the effects of the transient solution die out and the behaviour of the circuit is described by the steady-state solution.

7.6.2 Analytical solution

The voltage equation for the circuit is

$$L\frac{di}{dt} + Ri + \frac{q}{C} = v$$

$$= v_m \sin \omega t$$

We know that $i = dq/dt$. Therefore, $di/dt = d^2q/dt^2$. Thus, in terms of q, the voltage equation becomes

$$L\frac{d^2q}{dt^2} + R\frac{dq}{dt} + \frac{q}{C} = v_m \sin\omega t \qquad (7.28)$$

This is like the equation for a forced, damped oscillator, [see Eq. {14.37(b)} in Class XI Physics Textbook]. Let us assume a solution

$$q = q_m \sin(\omega t + \theta) \qquad [7.29(a)]$$

$$\text{so that } \frac{dq}{dt} = q_m\omega\cos(\omega t + \theta) \qquad [7.29(b)]$$

$$\text{and } \frac{d^2q}{dt^2} = -q_m\omega^2 \sin(\omega t + \theta) \qquad [7.29(c)]$$

Substituting these values in Eq. (7.28), we get

$$q_m\omega\left[R\cos(\omega t + \theta) + (X_C - X_L)\sin(\omega t + \theta)\right] = v_m \sin\omega t \qquad (7.30)$$

where we have used the relation $X_c = 1/\omega C$, $X_L = \omega L$. Multiplying and dividing Eq. (7.30) by $Z = \sqrt{R^2 + (X_c - X_L)^2}$, we have

$$q_m\omega Z\left[\frac{R}{Z}\cos(\omega t + \theta) + \frac{(X_C - X_L)}{Z}\sin(\omega t + \theta)\right] = v_m \sin\omega t \qquad (7.31)$$

Now, let $\frac{R}{Z} = \cos\phi$

and $\frac{(X_C - X_L)}{Z} = \sin\phi$

$$\text{so that } \phi = \tan^{-1}\frac{X_C - X_L}{R} \qquad (7.32)$$

Substituting this in Eq. (7.31) and simplifying, we get:

$$q_m\omega Z\cos(\omega t + \theta - \phi) = v_m \sin\omega t \qquad (7.33)$$

Comparing the two sides of this equation, we see that

$$v_m = q_m\,\omega Z = i_m Z$$

where

$$i_m = q_m\,\omega \qquad [7.33(a)]$$

$$\text{and } \theta - \phi = -\frac{\pi}{2} \text{ or } \theta = -\frac{\pi}{2} + \phi \qquad [7.33(b)]$$

Therefore, the current in the circuit is

$$i = \frac{dq}{dt} = q_m\,\omega\cos(\omega t + \theta)$$

$$= i_m \cos(\omega t + \theta)$$

$$\text{or } i = i_m \sin(\omega t + \phi) \qquad (7.34)$$

$$\text{where } i_m = \frac{v_m}{Z} = \frac{v_m}{\sqrt{R^2 + (X_C - X_L)^2}} \qquad [7.34(a)]$$

$$\text{and } \phi = \tan^{-1}\frac{X_C - X_L}{R}$$

Thus, the analytical solution for the amplitude and phase of the current in the circuit agrees with that obtained by the technique of phasors.

7.6.3 Resonance

An interesting characteristic of the series *RLC* circuit is the phenomenon of resonance. The phenomenon of resonance is common among systems that have a tendency to oscillate at a particular frequency. This frequency is called the system's *natural frequency*. If such a system is driven by an energy source at a frequency that is near the natural frequency, the amplitude of oscillation is found to be large. A familiar example of this phenomenon is a child on a swing. The the swing has a natural frequency for swinging back and forth like a pendulum. If the child pulls on the rope at regular intervals and the frequency of the pulls is almost the same as the frequency of swinging, the amplitude of the swinging will be large (Chapter 14, Class XI).

For an *RLC* circuit driven with voltage of amplitude v_m and frequency ω, we found that the current amplitude is given by

$$i_m = \frac{v_m}{Z} = \frac{v_m}{\sqrt{R^2 + (X_C - X_L)^2}}$$

with $X_c = 1/\omega C$ and $X_L = \omega L$. So if ω is varied, then at a particular frequency ω_0, $X_c = X_L$, and the impedance is minimum $\left(Z = \sqrt{R^2 + 0^2} = R\right)$. This frequency is called the *resonant frequency*:

$$X_c = X_L \text{ or } \frac{1}{\omega_0 C} = \omega_0 L$$

$$\text{or} \quad \omega_0 = \frac{1}{\sqrt{LC}} \tag{7.35}$$

At resonant frequency, the current amplitude is maximum; $i_m = v_m/R$.

FIGURE 7.16 Variation of i_m with ω for two cases: (i) R = 100 Ω, (ii) R = 200 Ω, L = 1.00 mH.

Figure 7.16 shows the variation of i_m with ω in a *RLC* series circuit with L = 1.00 mH, C = 1.00 nF for two values of R: (i) R = 100 Ω and (ii) R = 200 Ω. For the source applied v_m = 100 V. ω_0 for this case is $\left(\frac{1}{\sqrt{LC}}\right)$ = 1.00×10^6 rad/s.

We see that the current amplitude is maximum at the resonant frequency. Since $i_m = v_m / R$ at resonance, the current amplitude for case (i) is twice to that for case (ii).

Resonant circuits have a variety of applications, for example, in the tuning mechanism of a radio or a TV set. The antenna of a radio accepts signals from many broadcasting stations. The signals picked up in the antenna acts as a source in the tuning circuit of the radio, so the circuit can be driven at many frequencies.

But to hear one particular radio station, we tune the radio. In tuning, we vary the capacitance of a capacitor in the tuning circuit such that the resonant frequency of the circuit becomes nearly equal to the frequency of the radio signal received. When this happens, the amplitude of the current with the frequency of the signal of the particular radio station in the circuit is maximum.

It is important to note that resonance phenomenon is exhibited by a circuit only if both L and C are present in the circuit. Only then do the voltages across L and C cancel each other (both being out of phase) and the current amplitude is v_m/R, the total source voltage appearing across R. This means that we cannot have resonance in a RL or RC circuit.

Sharpness of resonance

The amplitude of the current in the series *LCR* circuit is given by

$$i_m = \frac{v_m}{\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}}$$

and is maximum when $\omega = \omega_0 = 1/\sqrt{LC}$. The maximum value is $i_m^{\max} = v_m / R$.

For values of ω other than ω_0, the amplitude of the current is less than the maximum value. Suppose we choose a value of ω for which the current amplitude is $1/\sqrt{2}$ times its maximum value. At this value, the power dissipated by the circuit becomes half. From the curve in Fig. (7.16), we see that there are two such values of ω, say, ω_1 and ω_2, one greater and the other smaller than ω_0 and symmetrical about ω_0. We may write

$$\omega_1 = \omega_0 + \Delta\omega$$

$$\omega_2 = \omega_0 - \Delta\omega$$

The difference $\omega_1 - \omega_2 = 2\Delta\omega$ is often called the *bandwidth* of the circuit. The quantity $(\omega_0 / 2\Delta\omega)$ is regarded as a measure of the sharpness of resonance. The smaller the $\Delta\omega$, the sharper or narrower is the resonance. To get an expression for $\Delta\omega$, we note that the current amplitude i_m is $\left(1/\sqrt{2}\right) i_m^{\max}$ for $\omega_1 = \omega_0 + \Delta\omega$. Therefore,

$$\text{at } \omega_1, \quad i_m = \frac{v_m}{\sqrt{R^2 + \left(\omega_1 L - \frac{1}{\omega_1 C}\right)^2}}$$

$$= \frac{i_m^{\max}}{\sqrt{2}} = \frac{v_m}{R\sqrt{2}}$$

or $$\sqrt{R^2 + \left(\omega_1 L - \frac{1}{\omega_1 C}\right)^2} = R\sqrt{2}$$

or $$R^2 + \left(\omega_1 L - \frac{1}{\omega_1 C}\right)^2 = 2R^2$$

$$\omega_1 L - \frac{1}{\omega_1 C} = R$$

which may be written as,

$$(\omega_0 + \Delta\omega) L - \frac{1}{(\omega_0 + \Delta\omega)C} = R$$

$$\omega_0 L \left(1 + \frac{\Delta\omega}{\omega_0}\right) - \frac{1}{\omega_0 C \left(1 + \frac{\Delta\omega}{\omega_0}\right)} = R$$

Using $\omega_0^2 = \frac{1}{LC}$ in the second term on the left hand side, we get

$$\omega_0 L \left(1 + \frac{\Delta\omega}{\omega_0}\right) - \frac{\omega_0 L}{\left(1 + \frac{\Delta\omega}{\omega_0}\right)} = R$$

We can approximate $\left(1 + \frac{\Delta\omega}{\omega_0}\right)^{-1}$ as $\left(1 - \frac{\Delta\omega}{\omega_0}\right)$ since $\frac{\Delta\omega}{\omega_0} << 1$. Therefore,

$$\omega_0 L \left(1 + \frac{\Delta\omega}{\omega_0}\right) - \omega_0 L \left(1 - \frac{\Delta\omega}{\omega_0}\right) = R$$

or $$\omega_0 L \frac{2\Delta\omega}{\omega_0} = R$$

$$\Delta\omega = \frac{R}{2L} \qquad [7.36(a)]$$

The sharpness of resonance is given by,

$$\frac{\omega_0}{2\Delta\omega} = \frac{\omega_0 L}{R} \qquad [7.36(b)]$$

The ratio $\frac{\omega_0 L}{R}$ is also called the *quality factor*, Q of the circuit.

$$Q = \frac{\omega_0 L}{R} \qquad [7.36(c)]$$

From Eqs. [7.36 (b)] and [7.36 (c)], we see that $2\Delta\omega = \frac{\omega_0}{Q}$. So, larger the

value of Q, the smaller is the value of $2\Delta\omega$ or the bandwidth and sharper is the resonance. Using $\omega_0^2 = 1/LC$, Eq. [7.36(c)] can be equivalently expressed as $Q = 1/\omega_0 CR$.

We see from Fig. 7.15, that if the resonance is less sharp, not only is the maximum current less, the circuit is close to resonance for a larger range $\Delta\omega$ of frequencies and the tuning of the circuit will not be good. So, less sharp the resonance, less is the selectivity of the circuit or vice versa. From Eq. (7.36), we see that if quality factor is large, i.e., R is low or L is large, the circuit is more selective.

Example 7.6 A resistor of 200 Ω and a capacitor of 15.0 μF are connected in series to a 220 V, 50 Hz ac source. (a) Calculate the current in the circuit; (b) Calculate the voltage (rms) across the resistor and the capacitor. Is the algebraic sum of these voltages more than the source voltage? If yes, resolve the paradox.

Solution

Given

$R = 200\,\Omega,\ C = 15.0\,\mu\text{F} = 15.0 \times 10^{-6}\,\text{F}$

$V = 220\,\text{V},\ \nu = 50\,\text{Hz}$

(a) In order to calculate the current, we need the impedance of the circuit. It is

$$Z = \sqrt{R^2 + X_C^2} = \sqrt{R^2 + (2\pi\nu C)^{-2}}$$

$$= \sqrt{(200\,\Omega)^2 + (2 \times 3.14 \times 50 \times 10^{-6}\,\text{F})^{-2}}$$

$$= \sqrt{(200\,\Omega)^2 + (212\,\Omega)^2}$$

$$= 291.5\,\Omega$$

Therefore, the current in the circuit is

$$I = \frac{V}{Z} = \frac{220\,\text{V}}{291.5\,\Omega} = 0.755\,\text{A}$$

(b) Since the current is the same throughout the circuit, we have

$$V_R = I\,R = (0.755\,\text{A})(200\,\Omega) = 151\,\text{V}$$

$$V_C = I\,X_C = (0.755\,\text{A})(212.3\,\Omega) = 160.3\,\text{V}$$

The algebraic sum of the two voltages, V_R and V_C is 311.3 V which is more than the source voltage of 220 V. How to resolve this paradox? As you have learnt in the text, the two voltages are not in the same phase. Therefore, *they cannot be added like ordinary numbers*. The two voltages are out of phase by ninety degrees. Therefore, the total of these voltages must be obtained using the Pythagorean theorem:

$$V_{R+C} = \sqrt{V_R^2 + V_C^2}$$

$$= 220\ \text{V}$$

Thus, if the phase difference between two voltages is properly taken into account, the total voltage across the resistor and the capacitor is equal to the voltage of the source.

EXAMPLE 7.6

7.7 Power in AC Circuit: The Power Factor

We have seen that a voltage $v = v_m \sin\omega t$ applied to a series *RLC* circuit drives a current in the circuit given by $i = i_m \sin(\omega t + \phi)$ where

$$i_m = \frac{v_m}{Z} \quad \text{and} \quad \phi = \tan^{-1}\left(\frac{X_C - X_L}{R}\right)$$

Therefore, the instantaneous power p supplied by the source is

$$p = v\,i = \left(v_m \sin\omega t\right) \times \left[i_m \sin(\omega t + \phi)\right]$$

$$= \frac{v_m i_m}{2}\left[\cos\phi - \cos(2\omega t + \phi)\right] \qquad (7.37)$$

The average power over a cycle is given by the average of the two terms in R.H.S. of Eq. (7.37). It is only the second term which is time-dependent. Its average is zero (the positive half of the cosine cancels the negative half). Therefore,

$$P = \frac{v_m i_m}{2}\cos\phi = \frac{v_m}{\sqrt{2}}\frac{i_m}{\sqrt{2}}\cos\phi$$

$$= V\,I\cos\phi \qquad [7.38(a)]$$

This can also be written as,

$$P = I^2\,Z\cos\phi \qquad [7.38(b)]$$

So, the average power dissipated depends not only on the voltage and current but also on the cosine of the phase angle ϕ between them. The quantity $\cos\phi$ is called the *power factor*. Let us discuss the following cases:

Case (i) *Resistive circuit*: If the circuit contains only pure *R*, it is called *resistive*. In that case $\phi = 0$, $\cos\phi = 1$. There is maximum power dissipation.

Case (ii) *Purely inductive or capacitive circuit*: If the circuit contains only an inductor or capacitor, we know that the phase difference between voltage and current is $\pi/2$. Therefore, $\cos\phi = 0$, and no power is dissipated even though a current is flowing in the circuit. This current is sometimes referred to as *wattless current*.

Case (iii) *LCR series circuit*: In an *LCR* series circuit, power dissipated is given by Eq. (7.38) where $\phi = \tan^{-1}(X_c - X_L)/R$. So, ϕ may be non-zero in a *RL* or *RC* or *RCL* circuit. Even in such cases, power is dissipated only in the resistor.

Case (iv) *Power dissipated at resonance in LCR circuit*: At resonance $X_c - X_L = 0$, and $\phi = 0$. Therefore, $\cos\phi = 1$ and $P = I^2Z = I^2 R$. That is, maximum power is dissipated in a circuit (through *R*) at resonance.

Example 7.7

Example7.7 (a) For circuits used for transporting electric power, a low power factor implies large power loss in transmission. Explain.

(b) Power factor can often be improved by the use of a capacitor of appropriate capacitance in the circuit. Explain.

Solution (a) We know that $P = I\,V \cos\phi$ where $\cos\phi$ is the power factor. To supply a given power at a given voltage, if $\cos\phi$ is small, we have to increase current accordingly. But this will lead to large power loss (I^2R) in transmission.

(b)Suppose in a circuit, current I lags the voltage by an angle ϕ. Then power factor $\cos\phi = R/Z$.

We can improve the power factor (tending to 1) by making Z tend to R. Let us understand, with the help of a phasor diagram (Fig. 7.17) how this can be achieved. Let us resolve **I** into two components. $\mathbf{I}_p$ along

FIGURE 7.17

the applied voltage **V** and $\mathbf{I}_q$ perpendicular to the applied voltage. $\mathbf{I}_q$ as you have learnt in Section 7.7, is called the wattless component since corresponding to this component of current, there is no power loss. $\mathbf{I}_P$ is known as the power component because it is in phase with the voltage and corresponds to power loss in the circuit.

It's clear from this analysis that if we want to improve power factor, we must completely neutralize the lagging wattless current $\mathbf{I}_q$ by an equal leading wattless current $\mathbf{I}'_q$. This can be done by connecting a capacitor of appropriate value in parallel so that $\mathbf{I}_q$ and $\mathbf{I}'_q$ cancel each other and P is effectively $I_p\,V$.

EXAMPLE 7.7

Example 7.8 A sinusoidal voltage of peak value 283 V and frequency 50 Hz is applied to a series *LCR* circuit in which R = 3 Ω, L = 25.48 mH, and C = 796 μF. Find (a) the impedance of the circuit; (b) the phase difference between the voltage across the source and the current; (c) the power dissipated in the circuit; and (d) the power factor.

Solution

(a) To find the impedance of the circuit, we first calculate X_L and X_C.

$$X_L = 2\,\pi\nu L$$
$$= 2 \times 3.14 \times 50 \times 25.48 \times 10^{-3}\ \Omega = 8\ \Omega$$
$$X_C = \frac{1}{2\,\pi\,\nu C}$$

EXAMPLE 7.8

EXAMPLE 7.8

$$= \frac{1}{2 \times 3.14 \times 50 \times 796 \times 10^{-6}} = 4\Omega$$

Therefore,

$$Z = \sqrt{R^2 + (X_L - X_C)^2} = \sqrt{3^2 + (8-4)^2}$$

$$= 5\ \Omega$$

(b) Phase difference, $\phi = \tan^{-1} \frac{X_C - X_L}{R}$

$$= \tan^{-1}\left(\frac{4-8}{3}\right) = -53.1°$$

Since ϕ is negative, the current in the circuit lags the voltage across the source.

(c) The power dissipated in the circuit is

$$P = I^2 R$$

Now, $I = \frac{i_m}{\sqrt{2}} = \frac{1}{\sqrt{2}}\left(\frac{283}{5}\right) = 40\text{A}$

Therefore, $P = (40\text{A})^2 \times 3\Omega = 4800\,\text{W}$

(d) Power factor $= \cos\phi = \cos 53.1° = 0.6$

EXAMPLE 7.9

Example 7.9 Suppose the frequency of the source in the previous example can be varied. (a) What is the frequency of the source at which resonance occurs? (b) Calculate the impedance, the current, and the power dissipated at the resonant condition.

Solution

(a) The frequency at which the resonance occurs is

$$\omega_0 = \frac{1}{\sqrt{LC}} = \frac{1}{\sqrt{25.48 \times 10^{-3} \times 796 \times 10^{-6}}}$$

$$= 222.1\,\text{rad/s}$$

$$\nu_r = \frac{\omega_0}{2\pi} = \frac{221.1}{2 \times 3.14}\text{Hz} = 35.4\text{Hz}$$

(b) The impedance Z at resonant condition is equal to the resistance:

$$Z = R = 3\Omega$$

The rms current at resonance is

$$= \frac{V}{Z} = \frac{V}{R} = \left(\frac{283}{\sqrt{2}}\right)\frac{1}{3} = 66.7\text{A}$$

The power dissipated at resonance is

$$P = I^2 \times R = (66.7)^2 \times 3 = 13.35\ \text{kW}$$

You can see that in the present case, power dissipated at resonance is more than the power dissipated in Example 7.8.

Example 7.10 At an airport, a person is made to walk through the doorway of a metal detector, for security reasons. If she/he is carrying anything made of metal, the metal detector emits a sound. On what principle does this detector work?

Solution The metal detector works on the principle of resonance in ac circuits. When you walk through a metal detector, you are, in fact, walking through a coil of many turns. The coil is connected to a capacitor tuned so that the circuit is in resonance. When you walk through with metal in your pocket, the impedance of the circuit changes – resulting in significant change in current in the circuit. This change in current is detected and the electronic circuitry causes a sound to be emitted as an alarm.

EXAMPLE 7.10

7.8 LC OSCILLATIONS

We know that a capacitor and an inductor can store electrical and magnetic energy, respectively. When a capacitor (initially charged) is connected to an inductor, the charge on the capacitor and the current in the circuit exhibit the phenomenon of electrical oscillations similar to oscillations in mechanical systems (Chapter 14, Class XI).

Let a capacitor be charged q_m (at $t = 0$) and connected to an inductor as shown in Fig. 7.18.

The moment the circuit is completed, the charge on the capacitor starts decreasing, giving rise to current in the circuit. Let q and i be the charge and current in the circuit at time t. Since $\mathrm{d}i/\mathrm{d}t$ is positive, the induced emf in L will have polarity as shown, i.e., $v_b < v_a$. According to Kirchhoff's loop rule,

FIGURE 7.18 At the instant shown, the current is increasing so the polarity of induced emf in the inductor is as shown.

$$\frac{q}{C} - L\frac{\mathrm{d}i}{\mathrm{d}t} = 0 \qquad (7.39)$$

$i = -(\mathrm{d}q/\mathrm{d}t)$ in the present case (as q decreases, i increases). Therefore, Eq. (7.39) becomes:

$$\frac{\mathrm{d}^2 q}{\mathrm{d}t^2} + \frac{1}{LC}q = 0 \qquad (7.40)$$

This equation has the form $\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} + \omega_0^2 x = 0$ for a simple harmonic oscillator. The charge, therefore, oscillates with a natural frequency

$$\omega_0 = \frac{1}{\sqrt{LC}} \qquad (7.41)$$

and varies sinusoidally with time as

$$q = q_m \cos(\omega_0 t + \phi) \qquad (7.42)$$

where q_m is the maximum value of q and ϕ is a phase constant. Since $q = q_m$ at $t = 0$, we have $\cos\phi = 1$ or $\phi = 0$. Therefore, in the present case,

$$q = q_m \cos(\omega_0 t) \tag{7.43}$$

The current $i\left(= -\frac{dq}{dt}\right)$ is given by

$$i = i_m \sin(\omega_0 t) \tag{7.44}$$

where $i_m = \omega_0 q_m$

Let us now try to visualise how this oscillation takes place in the circuit.

Figure 7.19(a) shows a capacitor with initial charge q_m connected to an ideal inductor. The electrical energy stored in the charged capacitor is $U_E = \frac{1}{2}\frac{q_m^2}{C}$. Since, there is no current in the circuit, energy in the inductor is zero. Thus, the total energy of *LC* circuit is,

$$U = U_E = \frac{1}{2}\frac{q_m^2}{C}$$

FIGURE 7.19 The oscillations in an *LC* circuit are analogous to the oscillation of a block at the end of a spring. The figure depicts one-half of a cycle.

At $t = 0$, the switch is closed and the capacitor starts to discharge [Fig. 7.19(b)]. As the current increases, it sets up a magnetic field in the inductor and thereby, some energy gets stored in the inductor in the form of magnetic energy: $U_B = (1/2)\ Li^2$. As the current reaches its maximum value i_m, (at $t = T/4$) as in Fig. 7.19(c), all the energy is stored in the magnetic field: $U_B = (1/2)\ Li_m^2$. You can easily check that the maximum electrical energy equals the maximum magnetic energy. The capacitor now has no charge and hence no energy. The current now starts charging the capacitor, as in Fig. 7.19(d). This process continues till the capacitor is fully charged (at $t = T/2$) [Fig. 7.19(e)]. But it is charged with a polarity opposite to its initial state in Fig. 7.19(a). The whole process just described will now repeat itself till the system reverts to its original state. Thus, the energy in the system oscillates between the capacitor and the inductor.

The LC oscillation is similar to the mechanical oscillation of a block attached to a spring. The lower part of each figure in Fig. 7.19 depicts the corresponding stage of a mechanical system (a block attached to a spring). As noted earlier, for a block of a mass m oscillating with frequency ω_0, the equation is

$$\frac{d^2x}{dt^2} + \omega_0^2 x = 0$$

Here, $\omega_0 = \sqrt{k/m}$, and k is the spring constant. So, x corresponds to q. In case of a mechanical system $F = ma = m\,(dv/dt) = m\,(d^2x/dt^2)$. For an electrical system, $\varepsilon = -L\,(di/dt) = -L\,(d^2q/dt^2)$. Comparing these two equations, we see that L is analogous to mass m: L is a measure of resistance to change in current. In case of LC circuit, $\omega_0 = 1/\sqrt{LC}$ and for mass on a spring, $\omega_0 = \sqrt{k/m}$. So, $1/C$ is analogous to k. The constant k $(=F/x)$ tells us the (external) force required to produce a unit displacement whereas $1/C$ $(=V/q)$ tells us the potential difference required to store a unit charge. Table 7.1 gives the analogy between mechanical and electrical quantities.

TABLE 7.1 ANALOGIES BETWEEN MECHANICAL AND ELECTRICAL QUANTITIES

Mechanical system	Slectrical system
Mass m	Inductance L
Force constant k	Reciprocal capacitance $1/C$
Displacement x	Charge q
Velocity $v = dx/dt$	Current $i = dq/dt$
Mechanical energy	Electromagnetic energy
$E = \frac{1}{2}kx^2 + \frac{1}{2}mv^2$	$U = \frac{1}{2}\frac{q^2}{C} + \frac{1}{2}Li^2$

Note that the above discussion of LC oscillations is not realistic for two reasons:

(i) Every inductor has some resistance. The effect of this resistance is to introduce a damping effect on the charge and current in the circuit and the oscillations finally die away.

(ii) Even if the resistance were zero, the total energy of the system would not remain constant. It is radiated away from the system in the form of electromagnetic waves (discussed in the next chapter). In fact, radio and TV transmitters depend on this radiation.

Two different phenomena, same mathematical treatment

You may like to compare the treatment of a forced damped oscillator discussed in Section 14.10 of Class XI physics textbook, with that of an *LCR* circuit when an ac voltage is applied in it. We have already remarked that Eq. [14.37(b)] of Class XI Textbook is exactly similar to Eq. (7.28) here, although they use different symbols and parameters. Let us therefore list the equivalence between different quantities in the two situations:

Forced oscillations	*Driven LCR circuit*
$m\frac{d^2x}{dt^2} + b\frac{dx}{dt} + kx = F\cos\omega_d\, t$	$L\frac{d^2q}{dt^2} + R\frac{dq}{dt} + \frac{q}{C} = v_m \sin\omega t$
Displacement, x	Charge on capacitor, q
Time, t	Time, t
Mass, m	Self inductance, L
Damping constant, b	Resistance, R
Spring constant, k	Inverse capacitance, $1/C$
Driving frequency, ω_d	Driving frequency, ω
Natural frequency of oscillations, ω	Natural frequency of LCR circuit, ω_0
Amplitude of forced oscillations, A	Maximum charge stored, q_m
Amplitude of driving force, F_0	Amplitude of applied voltage, v_m

You must note that since x corresponds to q, the amplitude A (maximum displacement) will correspond to the maximum charge stored, q_m. Equation [14.39 (a)] of Class XI gives the amplitude of oscillations in terms of other parameters, which we reproduce here for convenience:

$$A = \frac{F_0}{\left\{m^2(\omega^2 - \omega_d^2)^2 + \omega_d^2 b^2\right\}^{1/2}}$$

Replace each parameter in the above equation by the corresponding electrical quantity, and see what happens. Eliminate L, C, ω, and ω_0, using $X_L = \omega L$, $X_C = 1/\omega C$, and $\omega_0^2 = 1/LC$. When you use Eqs. (7.33) and (7.34), you will see that there is a perfect match.

You will come across numerous such situations in physics where diverse physical phenomena are represented by the same mathematical equation. If you have dealt with one of them, and you come across another situation, you may simply replace the corresponding quantities and interpret the result in the new context. We suggest that you may try to find more such parallel situations from different areas of physics. One must, of course, be aware of the differences too.

Example 7.11 Show that in the free oscillations of an *LC* circuit, the sum of energies stored in the capacitor and the inductor is constant in time.

Solution Let q_0 be the initial charge on a capacitor. Let the charged capacitor be connected to an inductor of inductance *L*. As you have studied in Section 7.8, this LC circuit will sustain an oscillation with frquency

$$\omega \left(= 2\pi\nu = \frac{1}{\sqrt{LC}} \right)$$

At an instant *t*, charge *q* on the capacitor and the current *i* are given by:

$q(t) = q_0 \cos \omega t$

$i(t) = -q_0 \omega \sin \omega t$

Energy stored in the capacitor at time *t* is

$$U_E = \frac{1}{2} C V^2 = \frac{1}{2}\frac{q^2}{C} = \frac{q_0^2}{2C} \cos^2(\omega t)$$

Energy stored in the inductor at time *t* is

$$U_M = \frac{1}{2} L i^2$$

$$= \frac{1}{2} L q_0^2 \omega^2 \sin^2(\omega t)$$

$$= \frac{q_0^2}{2C} \sin^2(\omega t) \quad \left(\because \omega^2 = 1/\sqrt{LC}\right)$$

Sum of energies

$$U_E + U_M = \frac{q_0^2}{2C}\left[\cos^2 \omega t + \sin^2 \omega t\right]$$

$$= \frac{q_0^2}{2C}$$

This sum is constant in time as q_o and *C*, both are time-independent. Note that it is equal to the initial energy of the capacitor. Why it is so? Think!

EXAMPLE 7.11

7.9 TRANSFORMERS

For many purposes, it is necessary to change (or transform) an alternating voltage from one to another of greater or smaller value. This is done with a device called *transformer* using the principle of mutual induction.

A transformer consists of two sets of coils, insulated from each other. They are wound on a soft-iron core, either one on top of the other as in Fig. 7.20(a) or on separate limbs of the core as in Fig. 7.20(b). One of the coils called the *primary coil* has N_p turns. The other coil is called the *secondary coil*; it has N_s turns. Often the primary coil is the input coil and the secondary coil is the output coil of the transformer.

FIGURE 7.20 Two arrangements for winding of primary and secondary coil in a transformer: (a) two coils on top of each other, (b) two coils on separate limbs of the core.

When an alternating voltage is applied to the primary, the resulting current produces an alternating magnetic flux which links the secondary and induces an emf in it. The value of this emf depends on the number of turns in the secondary. We consider an ideal transformer in which the primary has negligible resistance and all the flux in the core links both primary and secondary windings. Let ϕ be the flux in each turn in the core at time t due to current in the primary when a voltage v_p is applied to it.

Then the induced emf or voltage ε_s, in the secondary with N_s turns is

$$\varepsilon_s = -N_s \frac{d\phi}{dt} \tag{7.45}$$

The alternating flux ϕ also induces an emf, called back emf in the primary. This is

$$\varepsilon_p = -N_p \frac{d\phi}{dt} \tag{7.46}$$

But $\varepsilon_p = v_p$. If this were not so, the primary current would be infinite since the primary has zero resistance(as assumed). If the secondary is an open circuit or the current taken from it is small, then to a good approximation

$$\varepsilon_s = v_s$$

where v_s is the voltage across the secondary. Therefore, Eqs. (7.45) and (7.46) can be written as

$$v_s = -N_s \frac{d\phi}{dt} \qquad [7.45(a)]$$

$$v_p = -N_p \frac{d\phi}{dt} \qquad [7.46(a)]$$

From Eqs. [7.45 (a)] and [7.45 (a)], we have

$$\frac{v_s}{v_p} = \frac{N_s}{N_p} \tag{7.47}$$

Note that the above relation has been obtained using three assumptions: (i) the primary resistance and current are small; (ii) the same flux links both the primary and the secondary as very little flux escapes from the core, and (iii) the secondary current is small.

If the transformer is assumed to be 100% efficient (no energy losses), the power input is equal to the power output, and since $p = i\,v$,

$$i_p v_p = i_s v_s \tag{7.48}$$

Although some energy is always lost, this is a good approximation, since a well designed transformer may have an efficiency of more than 95%. Combining Eqs. (7.47) and (7.48), we have

$$\frac{i_p}{i_s} = \frac{v_s}{v_p} = \frac{N_s}{N_p} \tag{7.49}$$

Since i and v both oscillate with the same frequency as the ac source, Eq. (7.49) also gives the ratio of the amplitudes or rms values of corresponding quantities.

Now, we can see how a transformer affects the voltage and current. We have:

$$V_s = \left(\frac{N_s}{N_p}\right) V_p \quad \text{and} \quad I_s = \left(\frac{N_p}{N_s}\right) I_p \tag{7.50}$$

That is, if the secondary coil has a greater number of turns than the primary ($N_s > N_p$), the voltage is stepped up($V_s > V_p$). This type of arrangement is called a *step-up transformer*. However, in this arrangement, there is less current in the secondary than in the primary ($N_p/N_s < 1$ and $I_s < I_p$). For example, if the primary coil of a transformer has 100 turns and the secondary has 200 turns, $N_s/N_p = 2$ and $N_p/N_s = 1/2$. Thus, a 220V input at 10A will step-up to 440 V output at 5.0 A.

If the secondary coil has less turns than the primary($N_s < N_p$), we have a *step-down transformer*. In this case, $V_s < V_p$ and $I_s > I_p$. That is, the voltage is stepped down, or reduced, and the current is increased.

The equations obtained above apply to ideal transformers (without any energy losses). But in actual transformers, small energy losses do occur due to the following reasons:

(i) *Flux Leakage*: There is always some flux leakage; that is, not all of the flux due to primary passes through the secondary due to poor design of the core or the air gaps in the core. It can be reduced by winding the primary and secondary coils one over the other.

(ii) *Resistance of the windings*: The wire used for the windings has some resistance and so, energy is lost due to heat produced in the wire (I^2R). In high current, low voltage windings, these are minimised by using thick wire.

(iii) *Eddy currents*: The alternating magnetic flux induces eddy currents in the iron core and causes heating. The effect is reduced by having a laminated core.

(iv) *Hysteresis*: The magnetisation of the core is repeatedly reversed by the alternating magnetic field. The resulting expenditure of energy in the core appears as heat and is kept to a minimum by using a magnetic material which has a low hysteresis loss.

The large scale transmission and distribution of electrical energy over long distances is done with the use of transformers. The voltage output of the generator is stepped-up (so that current is reduced and consequently, the I^2R loss is cut down). It is then transmitted over long distances to an area sub-station near the consumers. There the voltage is stepped down. It is further stepped down at distributing sub-stations and utility poles before a power supply of 240 V reaches our homes.

SUMMARY

1. An alternating voltage $v = v_m \sin \omega t$ applied to a resistor R drives a current $i = i_m \sin\omega t$ in the resistor, $i_m = \frac{v_m}{R}$. The current is in phase with the applied voltage.
2. For an alternating current $i = i_m \sin\omega t$ passing through a resistor R, the average power loss P (averaged over a cycle) due to joule heating is $(1/2)i_m^2R$. To express it in the same form as the dc power ($P = I^2R$), a special value of current is used. It is called *root mean square* (rms) *current* and is donoted by I:

   $$I = \frac{i_m}{\sqrt{2}} = 0.707\, i_m$$

   Similarly, the *rms voltage* is defined by

   $$V = \frac{v_m}{\sqrt{2}} = 0.707\, v_m$$

   We have $P = IV = I^2R$
3. An ac voltage $v = v_m \sin\omega t$ applied to a pure inductor L, drives a current in the inductor $i = i_m \sin(\omega t - \pi/2)$, where $i_m = v_m/X_L$. $X_L = \omega L$ is called *inductive reactance*. The current in the inductor lags the voltage by $\pi/2$. The average power supplied to an inductor over one complete cycle is zero.
4. An ac voltage $v = v_m \sin\omega t$ applied to a capacitor drives a current in the capacitor: $i = i_m \sin(\omega t + \pi/2)$. Here,

   $i_m = \frac{v_m}{X_C}$, $X_C = \frac{1}{\omega C}$ is called *capacitive reactance*.

   The current through the capacitor is $\pi/2$ ahead of the applied voltage. As in the case of inductor, the average power supplied to a capacitor over one complete cycle is zero.
5. For a series RLC circuit driven by voltage $v = v_m \sin\omega t$, the current is given by $i = i_m \sin(\omega t + \phi)$

   where $i_m = \frac{v_m}{\sqrt{R^2 + (X_C - X_L)^2}}$

   and $\phi = \tan^{-1}\frac{X_C - X_L}{R}$

   $Z = \sqrt{R^2 + (X_C - X_L)^2}$ is called the *impedance* of the circuit.

The average power loss over a complete cycle is given by

$$P = V I \cos\phi$$

The term $\cos\phi$ is called the *power factor*.

6. In a purely inductive or capacitive circuit, $\cos\phi = 0$ and no power is dissipated even though a current is flowing in the circuit. In such cases, current is referred to as a *wattless current*.
7. The phase relationship between current and voltage in an ac circuit can be shown conveniently by representing voltage and current by rotating vectors called *phasors*. A phasor is a vector which rotates about the origin with angular speed ω. The magnitude of a phasor represents the amplitude or peak value of the quantity (voltage or current) represented by the phasor.

   The analysis of an ac circuit is facilitated by the use of a phasor diagram.
8. An interesting characteristic of a series RLC circuit is the phenomenon of *resonance*. The circuit exhibits resonance, i.e., the amplitude of the current is maximum at the resonant frequency, $\omega_0 = \frac{1}{\sqrt{LC}}$. The *quality factor* Q defined by

   $Q = \frac{\omega_0 L}{R} = \frac{1}{\omega_0 CR}$ is an indicator of the sharpness of the resonance, the higher value of Q indicating sharper peak in the current.
9. A circuit containing an inductor L and a capacitor C (initially charged) with no ac source and no resistors exhibits *free oscillations*. The charge q of the capacitor satisfies the equation of simple harmonic motion:

$$\frac{d^2q}{dt^2} + \frac{1}{LC}q = 0$$

   and therefore, the frequency ω of free oscillation is $\omega_0 = \frac{1}{\sqrt{LC}}$. The energy in the system oscillates between the capacitor and the inductor but their sum or the total energy is constant in time.
10. A transformer consists of an iron core on which are bound a primary coil of N_p turns and a secondary coil of N_s turns. If the primary coil is connected to an ac source, the primary and secondary voltages are related by

$$V_s = \left(\frac{N_s}{N_p}\right) V_p$$

   and the currents are related by

$$I_s = \left(\frac{N_p}{N_s}\right) I_p$$

   If the secondary coil has a greater number of turns than the primary, the voltage is stepped-up ($V_s > V_p$). This type of arrangement is called a *step-up transformer*. If the secondary coil has turns less than the primary, we have a *step-down transformer*.

Physical quantity	Symbol	Dimensions	Unit	Remarks
rms voltage	V	$[M L^2 T^{-3} A^{-1}]$	V	$V = \frac{v_m}{\sqrt{2}}$, v_m is the amplitude of the ac voltage.
rms current	I	[A]	A	$I = \frac{i_m}{\sqrt{2}}$, i_m is the amplitude of the ac current.
Reactance:				
Inductive	X_L	$[M L^2 T^{-3} A^{-2}]$	Ω	$X_L = \omega L$
Capacitive	X_C	$[M L^2 T^{-3} A^{-2}]$	Ω	$X_C = 1/\omega C$
Impedance	Z	$[M L^2 T^{-3} A^{-2}]$	Ω	Depends on elements present in the circuit.
Resonant frequency	ω_r or ω_0	$[T^{-1}]$	Hz	$\omega_0 = \frac{1}{\sqrt{LC}}$ for a series *RLC* circuit
Quality factor	Q	Dimensionless		$Q = \frac{\omega_0 L}{R} = \frac{1}{\omega_0 C R}$ for a series *RLC* circuit.
Power factor		Dimensionless		$= \cos\phi$, ϕ is the phase difference between voltage applied and current in the circuit.

POINTS TO PONDER

1. When a value is given for ac voltage or current, it is ordinarily the rms value. The voltage across the terminals of an outlet in your room is normally 240 V. This refers to the *rms* value of the voltage. The amplitude of this voltage is

$$v_m = \sqrt{2}V = \sqrt{2}(240) = 340 \text{ V}$$

2. The power rating of an element used in ac circuits refers to its average power rating.
3. The power consumed in an circuit is never negative.
4. Both alternating current and direct current are measured in amperes. But how is the ampere defined for an alternating current? It cannot be derived from the mutual attraction of two parallel wires carrying ac currents, as the dc ampere is derived. An ac current changes direction

with the source frequency and the attractive force would average to zero. Thus, the ac ampere must be defined in terms of some property that is independent of the direction of the current. Joule heating is such a property, and there is one ampere of *rms* value of alternating current in a circuit if the current produces the same average heating effect as one ampere of dc current would produce under the same conditions.

5. In an ac circuit, while adding voltages across different elements, one should take care of their phases properly. For example, if V_R and V_C are voltages across R and C, respectively in an RC circuit, then the total voltage across RC combination is $V_{RC} = \sqrt{V_R^2 + V_C^2}$ and not $V_R + V_C$ since V_C is $\pi/2$ out of phase of V_R.
6. Though in a phasor diagram, voltage and current are represented by vectors, these quantities are not really vectors themselves. They are scalar quantities. It so happens that the amplitudes and phases of harmonically varying scalars combine mathematically in the same way as do the projections of rotating vectors of corresponding magnitudes and directions. The 'rotating vectors' that represent harmonically varying scalar quantities are introduced only to provide us with a simple way of adding these quantities using a rule that we already know as the law of vector addition.
7. There are no power losses associated with pure capacitances and pure inductances in an ac circuit. The only element that dissipates energy in an ac circuit is the resistive element.
8. In a RLC circuit, resonance phenomenon occur when $X_L = X_C$ or $\omega_0 = \frac{1}{\sqrt{LC}}$. For resonance to occur, the presence of both L and C elements in the circuit is a must. With only one of these (L or C) elements, there is no possibility of voltage cancellation and hence, no resonance is possible.
9. The power factor in a RLC circuit is a measure of how close the circuit is to expending the maximum power.
10. In generators and motors, the roles of input and output are reversed. In a motor, electric energy is the input and mechanical energy is the output. In a generator, mechanical energy is the input and electric energy is the output. Both devices simply transform energy from one form to another.
11. A transformer (step-up) changes a low-voltage into a high-voltage. This does not violate the law of conservation of energy. The current is reduced by the same proportion.
12. The choice of whether the description of an oscillatory motion is by means of sines or cosines or by their linear combinations is unimportant, since changing the zero-time position transforms the one to the other.

EXERCISES

7.1 A 100 Ω resistor is connected to a 220 V, 50 Hz ac supply.

(a) What is the rms value of current in the circuit?

(b) What is the net power consumed over a full cycle?

7.2 (a) The peak voltage of an ac supply is 300 V. What is the rms voltage?

(b) The rms value of current in an ac circuit is 10 A. What is the peak current?

7.3 A 44 mH inductor is connected to 220 V, 50 Hz ac supply. Determine the rms value of the current in the circuit.

7.4 A 60 μF capacitor is connected to a 110 V, 60 Hz ac supply. Determine the rms value of the current in the circuit.

7.5 In Exercises 7.3 and 7.4, what is the net power absorbed by each circuit over a complete cycle. Explain your answer.

7.6 Obtain the resonant frequency ω_r of a series *LCR* circuit with $L = 2.0$H, $C = 32$ μF and $R = 10$ Ω. What is the *Q*-value of this circuit?

7.7 A charged 30 μF capacitor is connected to a 27 mH inductor. What is the angular frequency of free oscillations of the circuit?

7.8 Suppose the initial charge on the capacitor in Exercise 7.7 is 6 mC. What is the total energy stored in the circuit initially? What is the total energy at later time?

7.9 A series *LCR* circuit with $R = 20$ Ω, $L = 1.5$ H and $C = 35$ μF is connected to a variable-frequency 200 V ac supply. When the frequency of the supply equals the natural frequency of the circuit, what is the average power transferred to the circuit in one complete cycle?

7.10 A radio can tune over the frequency range of a portion of MW broadcast band: (800 kHz to 1200 kHz). If its *LC* circuit has an effective inductance of 200 μH, what must be the range of its variable capacitor?

[*Hint:* For tuning, the natural frequency i.e., the frequency of free oscillations of the *LC* circuit should be equal to the frequency of the radiowave.]

7.11 Figure 7.21 shows a series *LCR* circuit connected to a variable frequency 230 V source. $L = 5.0$ H, $C = 80$μF, $R = 40$ Ω.

FIGURE 7.21

(a) Determine the source frequency which drives the circuit in resonance.

(b) Obtain the impedance of the circuit and the amplitude of current at the resonating frequency.

(c) Determine the rms potential drops across the three elements of the circuit. Show that the potential drop across the *LC* combination is zero at the resonating frequency.

ADDITIONAL EXERCISES

7.12 An *LC* circuit contains a 20 mH inductor and a 50 μF capacitor with an initial charge of 10 mC. The resistance of the circuit is negligible. Let the instant the circuit is closed be $t = 0$.

(a) What is the total energy stored initially? Is it conserved during *LC* oscillations?
(b) What is the natural frequency of the circuit?
(c) At what time is the energy stored
(i) completely electrical (i.e., stored in the capacitor)? (ii) completely magnetic (i.e., stored in the inductor)?
(d) At what times is the total energy shared equally between the inductor and the capacitor?
(e) If a resistor is inserted in the circuit, how much energy is eventually dissipated as heat?

7.13 A coil of inductance 0.50 H and resistance 100 Ω is connected to a 240 V, 50 Hz ac supply.

(a) What is the maximum current in the coil?
(b) What is the time lag between the voltage maximum and the current maximum?

7.14 Obtain the answers (a) to (b) in Exercise 7.13 if the circuit is connected to a high frequency supply (240 V, 10 kHz). Hence, explain the statement that at very high frequency, an inductor in a circuit nearly amounts to an open circuit. How does an inductor behave in a dc circuit after the steady state?

7.15 A 100 μF capacitor in series with a 40 Ω resistance is connected to a 110 V, 60 Hz supply.

(a) What is the maximum current in the circuit?
(b) What is the time lag between the current maximum and the voltage maximum?

7.16 Obtain the answers to (a) and (b) in Exercise 7.15 if the circuit is connected to a 110 V, 12 kHz supply? Hence, explain the statement that a capacitor is a conductor at very high frequencies. Compare this behaviour with that of a capacitor in a dc circuit after the steady state.

7.17 Keeping the source frequency equal to the resonating frequency of the series *LCR* circuit, if the three elements, *L*, *C* and *R* are arranged in parallel, show that the total current in the parallel *LCR* circuit is minimum at this frequency. Obtain the current rms value in each branch of the circuit for the elements and source specified in Exercise 7.11 for this frequency.

7.18 A circuit containing a 80 mH inductor and a 60 μF capacitor in series is connected to a 230 V, 50 Hz supply. The resistance of the circuit is negligible.

(a) Obtain the current amplitude and rms values.
(b) Obtain the rms values of potential drops across each element.
(c) What is the average power transferred to the inductor?
(d) What is the average power transferred to the capacitor?
(e) What is the total average power absorbed by the circuit? ['Average' implies 'averaged over one cycle'.]

7.19 Suppose the circuit in Exercise 7.18 has a resistance of 15 Ω. Obtain the average power transferred to each element of the circuit, and the total power absorbed.

7.20 A series *LCR* circuit with $L = 0.12$ H, $C = 480$ nF, $R = 23\ \Omega$ is connected to a 230 V variable frequency supply.

(a) What is the source frequency for which current amplitude is maximum. Obtain this maximum value.

(b) What is the source frequency for which average power absorbed by the circuit is maximum. Obtain the value of this maximum power.

(c) For which frequencies of the source is the power transferred to the circuit half the power at resonant frequency? What is the current amplitude at these frequencies?

(d) What is the *Q*-factor of the given circuit?

7.21 Obtain the resonant frequency and *Q*-factor of a series *LCR* circuit with $L = 3.0$ H, $C = 27\ \mu$F, and $R = 7.4\ \Omega$. It is desired to improve the sharpness of the resonance of the circuit by reducing its 'full width at half maximum' by a factor of 2. Suggest a suitable way.

7.22 Answer the following questions:

(a) In any ac circuit, is the applied instantaneous voltage equal to the algebraic sum of the instantaneous voltages across the series elements of the circuit? Is the same true for rms voltage?

(b) A capacitor is used in the primary circuit of an induction coil.

(c) An applied voltage signal consists of a superposition of a dc voltage and an ac voltage of high frequency. The circuit consists of an inductor and a capacitor in series. Show that the dc signal will appear across *C* and the ac signal across *L*.

(d) A choke coil in series with a lamp is connected to a dc line. The lamp is seen to shine brightly. Insertion of an iron core in the choke causes no change in the lamp's brightness. Predict the corresponding observations if the connection is to an ac line.

(e) Why is choke coil needed in the use of fluorescent tubes with ac mains? Why can we not use an ordinary resistor instead of the choke coil?

7.23 A power transmission line feeds input power at 2300 V to a step-down transformer with its primary windings having 4000 turns. What should be the number of turns in the secondary in order to get output power at 230 V?

7.24 At a hydroelectric power plant, the water pressure head is at a height of 300 m and the water flow available is 100 $m^3 s^{-1}$. If the turbine generator efficiency is 60%, estimate the electric power available from the plant ($g = 9.8\ ms^{-2}$).

7.25 A small town with a demand of 800 kW of electric power at 220 V is situated 15 km away from an electric plant generating power at 440 V. The resistance of the two wire line carrying power is 0.5 Ω per km. The town gets power from the line through a 4000-220 V step-down transformer at a sub-station in the town.

(a) Estimate the line power loss in the form of heat.

(b) How much power must the plant supply, assuming there is negligible power loss due to leakage?

(c) Characterise the step up transformer at the plant.

7.26 Do the same exercise as above with the replacement of the earlier transformer by a 40,000-220 V step-down transformer (Neglect, as before, leakage losses though this may not be a good assumption any longer because of the very high voltage transmission involved). Hence, explain why high voltage transmission is preferred?

Chapter Eight

ELECTROMAGNETIC WAVES

8.1 Introduction

In Chapter 4, we learnt that an electric current produces magnetic field and that two current-carrying wires exert a magnetic force on each other. Further, in Chapter 6, we have seen that a magnetic field changing with time gives rise to an electric field. Is the converse also true? Does an electric field changing with time give rise to a magnetic field? James Clerk Maxwell (1831-1879), argued that this was indeed the case – not only an electric current but also a time-varying electric field generates magnetic field. While applying the Ampere's circuital law to find magnetic field at a point outside a capacitor connected to a time-varying current, Maxwell noticed an inconsistency in the Ampere's circuital law. He suggested the existence of an additional current, called by him, the displacement current to remove this inconsistency.

Maxwell formulated a set of equations involving electric and magnetic fields, and their sources, the charge and current densities. These equations are known as Maxwell's equations. Together with the Lorentz force formula (Chapter 4), they mathematically express all the basic laws of electromagnetism.

The most important prediction to emerge from Maxwell's equations is the existence of electromagnetic waves, which are (coupled) time-varying electric and magnetic fields that propagate in space. The speed of the waves, according to these equations, turned out to be very close to

James Clerk Maxwell (1831 – 1879) Born in Edinburgh, Scotland, was among the greatest physicists of the nineteenth century. He derived the thermal velocity distribution of molecules in a gas and was among the first to obtain reliable estimates of molecular parameters from measurable quantities like viscosity, etc. Maxwell's greatest acheivement was the unification of the laws of electricity and magnetism (discovered by Coulomb, Oersted, Ampere and Faraday) into a consistent set of equations now called Maxwell's equations. From these he arrived at the most important conclusion that light is an electromagnetic wave. Interestingly, Maxwell did not agree with the idea (strongly suggested by the Faraday's laws of electrolysis) that electricity was particulate in nature.

JAMES CLERK MAXWELL (1831–1879)

the speed of light(3×10^8 m/s), obtained from optical measurements. This led to the remarkable conclusion that light is an electromagnetic wave. Maxwell's work thus unified the domain of electricity, magnetism and light. Hertz, in 1885, experimentally demonstrated the existence of electromagnetic waves. Its technological use by Marconi and others led in due course to the revolution in communication that we are witnessing today.

In this chapter, we first discuss the need for displacement current and its consequences. Then we present a descriptive account of electromagnetic waves. The broad spectrum of electromagnetic waves, stretching from γ rays (wavelength $\sim 10^{-12}$ m) to long radio waves (wavelength $\sim 10^6$ m) is described. How the electromagnetic waves are sent and received for communication is discussed in Chapter 15.

8.2 Displacement Current

We have seen in Chapter 4 that an electrical current produces a magnetic field around it. Maxwell showed that for logical consistency, a changing electric field *must also* produce a magnetic field. This effect is of great importance because it explains the existence of radio waves, gamma rays and visible light, as well as all other forms of electromagnetic waves.

To see how a changing electric field gives rise to a magnetic field, let us consider the process of charging of a capacitor and apply Ampere's circuital law given by (Chapter 4)

$$\oint \mathbf{B} \cdot d\mathbf{l} = \mu_0 \, i(t) \qquad (8.1)$$

to find magnetic field at a point outside the capacitor. Figure 8.1(a) shows a parallel plate capacitor C which is a part of circuit through which a time-dependent current $i(t)$ flows . Let us find the magnetic field at a point such as P, in a region outside the parallel plate capacitor. For this, we consider a plane circular loop of radius r whose plane is perpendicular to the direction of the current-carrying wire, and which is centred symmetrically with respect to the wire [Fig. 8.1(a)]. From symmetry, the magnetic field is directed along the circumference of the circular loop and is the same in magnitude at all points on the loop so that if B is the magnitude of the field, the left side of Eq. (8.1) is $B(2\pi r)$. So we have

$$B(2\pi r) = \mu_0 i(t) \qquad (8.2)$$

Now, consider a different surface, which has the same boundary. This is a pot like surface [Fig. 8.1(b)] which nowhere touches the current, but has its bottom between the capacitor plates; its mouth is the circular loop mentioned above. Another such surface is shaped like a tiffin box (without the lid) [Fig. 8.1(c)]. On applying Ampere's circuital law to such surfaces with the *same* perimeter, we find that the left hand side of Eq. (8.1) has not changed but the right hand side is *zero* and *not* $\mu_0 i$, since *no* current passes through the surface of Fig. 8.1(b) and (c). So we have a *contradiction*; calculated one way, there is a magnetic field at a point P; calculated another way, the magnetic field at P is zero.
Since the contradiction arises from our use of Ampere's circuital law, this law must be missing something. The missing term must be such that one gets the same magnetic field at point P, no matter what surface is used.

We can actually guess the missing term by looking carefully at Fig. 8.1(c). Is there anything passing through the surface S *between* the plates of the capacitor? Yes, of course, the electric field! If the plates of the capacitor have an area A, and a total charge Q, the magnitude of the electric field **E** between the plates is $(Q/A)/\varepsilon_0$ (see Eq. 2.41). The field is perpendicular to the surface S of Fig. 8.1(c). It has the same magnitude over the area A of the capacitor plates, and vanishes outside it. So what is the *electric* flux Φ_E through the surface S? Using Gauss's law, it is

$$\Phi_E = |\mathbf{E}|\, A = \frac{1}{\varepsilon_0}\frac{Q}{A} A = \frac{Q}{\varepsilon_0} \tag{8.3}$$

Now if the charge Q on the capacitor plates changes with time, there is a current $i = (\mathrm{d}Q/\mathrm{d}t)$, so that using Eq. (8.3), we have

$$\frac{\mathrm{d}\Phi_E}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{Q}{\varepsilon_0}\right) = \frac{1}{\varepsilon_0}\frac{\mathrm{d}Q}{\mathrm{d}t}$$

This implies that for consistency,

$$\varepsilon_0 \left(\frac{\mathrm{d}\Phi_E}{\mathrm{d}t}\right) = i \tag{8.4}$$

This is the missing term in Ampere's circuital law. If we generalise this law by adding to the total current carried by conductors through the surface, another term which is ε_0 times the rate of change of electric flux through the same surface, the *total* has the same value of current i for all surfaces. If this is done, there is no contradiction in the value of B obtained anywhere using the generalised Ampere's law. B at the point P is non-zero no matter which surface is used for calculating it. B at a point P outside the plates [Fig. 8.1(a)] is the same as at a point M just inside, as it should be. The current carried by conductors due to flow of charges is called *conduction current*. The current, given by Eq. (8.4), is a new term, and is due to changing electric field (or electric *displacement*, an old term still used sometimes). It is, therefore, called *displacement current* or Maxwell's displacement current. Figure 8.2 shows the electric and magnetic fields inside the parallel plate capacitor discussed above.

The generalisation made by Maxwell then is the following. The source of a magnetic field is not *just* the conduction electric current due to flowing

FIGURE 8.1 A parallel plate capacitor C, as part of a circuit through which a time dependent current $i(t)$ flows, (a) a loop of radius r, to determine magnetic field at a point P on the loop; (b) a pot-shaped surface passing through the interior between the capacitor plates with the loop shown in (a) as its rim; (c) a tiffin-shaped surface with the circular loop as its rim and a flat circular bottom S between the capacitor plates. The arrows show uniform electric field between the capacitor plates.

C

(a)

(b)

FIGURE 8.2 (a) The electric and magnetic fields **E** and **B** between the capacitor plates, at the point M. (b) A cross sectional view of Fig. (a).

charges, but also the time rate of change of electric field. More precisely, the total current i is the sum of the conduction current denoted by i_c, and the displacement current denoted by i_d (= ε_0 (dΦ_E/dt)). So we have

$$i = i_e + i_d = i_c + \varepsilon_0 \frac{d\Phi_E}{dt} \tag{8.5}$$

In explicit terms, this means that outside the capacitor plates, we have only conduction current $i_c = i$, and no displacement current, i.e., $i_d = 0$. On the other hand, inside the capacitor, there is no conduction current, i.e., $i_c = 0$, and there is only displacement current, so that $i_d = i$.

The generalised (and correct) Ampere's circuital law has the same form as Eq. (8.1), with one difference: "the *total current* passing through any surface of which the closed loop is the perimeter" is the sum of the conduction current and the displacement current. The generalised law is

$$\oint \mathbf{B} \cdot d\mathbf{l} = \mu_0 i_c + \mu_0 \varepsilon_0 \frac{d\Phi_E}{dt} \tag{8.6}$$

and is known as Ampere-Maxwell law.

In all respects, the displacement current has the same physical effects as the conduction current. In some cases, for example, steady electric fields in a conducting wire, the displacement current may be zero since the electric field **E** does not change with time. In other cases, for example, the charging capacitor above, both conduction and displacement currents may be present in different regions of space. In most of the cases, they both may be present in the same region of space, as there exist no perfectly conducting or perfectly insulating medium. Most interestingly, there may be large regions of space where there is *no* conduction current, but there is only a displacement current due to time-varying electric fields. In such a region, we expect a magnetic field, though there is no (conduction) current source nearby! The prediction of such a displacement current can be verified experimentally. For example, a *magnetic* field (say at point M) between the plates of the capacitor in Fig. 8.2(a) can be measured and is seen to be the same as that just outside (at P).

The displacement current has (literally) far reaching consequences. One thing we immediately notice is that the laws of electricity and magnetism are now more symmetrical*. Faraday's law of induction states that there is an induced emf *equal to the rate of* change of magnetic flux. Now, since the emf between two points 1 and 2 is the work done per unit charge in taking it from 1 to 2, the existence of an emf implies the existence of an electric field. So, we can rephrase Faraday's law of electromagnetic induction by saying that a *magnetic field*, changing with time, gives rise to an *electric field*. Then, the fact that an *electric field* changing with time gives rise to a *magnetic field*, is the symmetrical counterpart, and is

* They are still not perfectly symmetrical; there are no known sources of magnetic field (magnetic monopoles) analogous to electric charges which are sources of electric field.

a consequence of the displacement current being a source of a magnetic field. Thus, time- dependent electric and magnetic fields give rise to each other! Faraday's law of electromagnetic induction and Ampere-Maxwell law give a quantitative expression of this statement, with the current being the total current, as in Eq. (8.5). One very important consequence of this symmetry is the existence of electromagnetic waves, which we discuss qualitatively in the next section.

MAXWELL'S EQUATIONS

1. $\oint \mathbf{E}\cdot d\mathbf{A} = Q/\varepsilon_0$ (Gauss's Law for electricity)
2. $\oint \mathbf{B}\cdot d\mathbf{A} = 0$ (Gauss's Law for magnetism)
3. $\oint \mathbf{E}\cdot d\mathbf{l} = \dfrac{-d\Phi_B}{dt}$ (Faraday's Law)
4. $\oint \mathbf{B}\cdot d\mathbf{l} = \mu_0 i_c + \mu_0 \varepsilon_0 \dfrac{d\Phi_E}{dt}$ (Ampere – Maxwell Law)

Example 8.1 A parallel plate capacitor with circular plates of radius 1 m has a capacitance of 1 nF. At $t = 0$, it is connected for charging in series with a resistor $R = 1\ \text{M}\Omega$ across a 2V battery (Fig. 8.3). Calculate the magnetic field at a point P, halfway between the centre and the periphery of the plates, after $t = 10^{-3}$ s. (The charge on the capacitor at time t is $q(t) = CV\,[1 - \exp(-t/\tau)]$, where the time constant τ is equal to CR.)

FIGURE 8.3

Solution The time constant of the CR circuit is $\tau = CR = 10^{-3}$ s. Then, we have

$q(t) = CV\,[1 - \exp(-t/\tau)]$

$= 2 \times 10^{-9}\,[1 - \exp(-t/10^{-3})]$

The electric field in between the plates at time t is

$$E = \frac{q(t)}{\varepsilon_0 A} = \frac{q}{\pi\varepsilon_0}\ ;\ A = \pi\,(1)^2\ \text{m}^2 = \text{area of the plates.}$$

Consider now a circular loop of radius (1/2) m parallel to the plates passing through P. The magnetic field **B** at all points on the loop is

EXAMPLE 8.1

EXAMPLE 8.1

along the loop and of the same value.
The flux Φ_E through this loop is
$\Phi_E = E \times$ area of the loop

$$= E \times \pi \times \left(\frac{1}{2}\right)^2 = \frac{\pi E}{4} = \frac{q}{4\varepsilon_0}$$

The displacement current

$$i_d = \varepsilon_0 \frac{d\Phi_E}{dt} = \frac{1}{4}\frac{dq}{dt} = 0.5 \times 10^{-6} \exp(-1)$$

at $t = 10^{-3}$s. Now, applying Ampere-Maxwell law to the loop, we get

$$B \times 2\pi \times \left(\frac{1}{2}\right) = \mu_0 (i_c + i_d) = \mu_0 (0 + i_d) = 0.5\times10^{-6}\ \mu_0 \exp(-1)$$

or, $B = 0.74 \times 10^{-13}$ T

8.3 ELECTROMAGNETIC WAVES

8.3.1 Sources of electromagnetic waves

How are electromagnetic waves produced? Neither stationary charges nor charges in uniform motion (steady currents) can be sources of electromagnetic waves. The former produces only electrostatic fields, while the latter produces magnetic fields that, however, do not vary with time. It is an important result of Maxwell's theory that accelerated charges radiate electromagnetic waves. The proof of this basic result is beyond the scope of this book, but we can accept it on the basis of rough, qualitative reasoning. Consider a charge oscillating with some frequency. (An oscillating charge is an example of accelerating charge.) This produces an oscillating electric field in space, which produces an oscillating magnetic field, which in turn, is a source of oscillating electric field, and so on. The oscillating electric and magnetic fields thus regenerate each other, so to speak, as the wave propagates through the space. The frequency of the electromagnetic wave naturally equals the frequency of oscillation of the charge. The energy associated with the propagating wave comes at the expense of the energy of the source – the accelerated charge.

From the preceding discussion, it might appear easy to test the prediction that light is an electromagnetic wave. We might think that all we needed to do was to set up an ac circuit in which the current oscillate at the frequency of visible light, say, yellow light. But, alas, that is not possible. The frequency of yellow light is about 6×10^{14} Hz, while the frequency that we get even with modern electronic circuits is hardly about 10^{11} Hz. This is why the experimental demonstration of electromagnetic wave had to come in the low frequency region (the radio wave region), as in the Hertz's experiment (1887).

Hertz's successful experimental test of Maxwell's theory created a sensation and sparked off other important works in this field. Two important achievements in this connection deserve mention. Seven years after Hertz, Jagdish Chandra Bose, working at Calcutta (now Kolkata),

succeeded in producing and observing electromagnetic waves of much shorter wavelength (25 mm to 5 mm). His experiment, like that of Hertz's, was confined to the laboratory.

At around the same time, Guglielmo Marconi in Italy followed Hertz's work and succeeded in transmitting electromagnetic waves over distances of many kilometres. Marconi's experiment marks the beginning of the field of communication using electromagnetic waves.

8.3.2 Nature of electromagnetic waves

It can be shown from Maxwell's equations that electric and magnetic fields in an electromagnetic wave are perpendicular to each other, *and* to the direction of propagation. It appears reasonable, say from our discussion of the displacement current. Consider Fig. 8.2. The electric field inside the plates of the capacitor is directed perpendicular to the plates. The magnetic field this gives rise to via the displacement current is along the perimeter of a circle parallel to the capacitor plates. So **B** and **E** are perpendicular in this case. This is a general feature.

In Fig. 8.4, we show a typical example of a plane electromagnetic wave propagating along the z direction (the fields are shown as a function of the z coordinate, at a given time t). The electric field E_x is along the x-axis, and varies sinusoidally with z, at a given time. The magnetic field B_y is along the y-axis, and again varies sinusoidally with z. The electric and magnetic fields E_x and B_y are perpendicular to each other, and to the direction z of propagation. We can write E_x and B_y as follows:

$$E_x = E_0 \sin(kz - \omega t) \qquad [8.7(a)]$$

$$B_y = B_0 \sin(kz - \omega t) \qquad [8.7(b)]$$

Here k is related to the wave length λ of the wave by the usual equation

$$k = \frac{2\pi}{\lambda} \qquad (8.8)$$

and ω is the angular frequency. k is the magnitude of the wave vector (or propagation vector) **k** and its direction describes the direction of propagation of the wave. The speed of propagation of the wave is (ω/k). Using Eqs. [8.7(a) and (b)] for E_x and B_y and Maxwell's equations, one finds that

Heinrich Rudolf Hertz (1857 – 1894) German physicist who was the first to broadcast and receive radio waves. He produced electro-magnetic waves, sent them through space, and measured their wave-length and speed. He showed that the nature of their vibration, reflection and refraction was the same as that of light and heat waves, establishing their identity for the first time. He also pioneered research on discharge of electricity through gases, and discovered the photoelectric effect.

HEINRICH RUDOLF HERTZ (1857–1894)

FIGURE 8.4 A linearly polarised electromagnetic wave, propagating in the z-direction with the oscillating electric field **E** along the x-direction and the oscillating magnetic field **B** along the y-direction.

$\omega = ck$, where, $c = 1/\sqrt{\mu_0 \varepsilon_0}$ [8.9(a)]

The relation $\omega = ck$ is the standard one for waves (see for example, Section 15.4 of class XI Physics textbook). This relation is often written in terms of frequency, ν $(=\omega/2\pi)$ and wavelength, λ $(=2\pi/k)$ as

$$2\pi\nu = c\left(\frac{2\pi}{\lambda}\right) \quad \text{or}$$

$$\nu\lambda = c \qquad [8.9(b)]$$

It is also seen from Maxwell's equations that the magnitude of the electric and the magnetic fields in an electromagnetic wave are related as

$$B_0 = (E_0/c) \qquad (8.10)$$

We here make remarks on some features of electromagnetic waves. They are self-sustaining oscillations of electric and magnetic fields in free space, or vacuum. They differ from all the other waves we have studied so far, in respect that *no material medium* is involved in the vibrations of the electric and magnetic fields. Sound waves in air are longitudinal waves of compression and rarefaction. Transverse waves on the surface of water consist of water moving up and down as the wave spreads horizontally and radially onwards. Transverse elastic (sound) waves can also propagate in a solid, which is rigid and that resists shear. Scientists in the nineteenth century were so much used to this mechanical picture that they thought that there must be some medium pervading all space and all matter, which responds to electric and magnetic fields just as any elastic medium does. They called this medium *ether*. They were so convinced of the reality of this medium, that there is even a novel called *The Poison Belt* by *Sir Arthur Conan Doyle* (the creator of the famous detective *Sherlock Holmes*) where the solar system is supposed to pass through a poisonous region of ether! We now accept that no such physical medium is needed. The famous experiment of Michelson and Morley in 1887 demolished conclusively the hypothesis of ether. Electric and magnetic fields, oscillating in space and time, can sustain each other in vacuum.

But what if a material medium is actually there? We know that light, an electromagnetic wave, does propagate through glass, for example. We have seen earlier that the total electric and magnetic fields inside a medium are described in terms of a permittivity ε and a magnetic permeability μ (these describe the factors by which the total fields differ from the external fields). These replace ε_0 and μ_0 in the description to electric and magnetic fields in Maxwell's equations with the result that in a material medium of permittivity ε and magnetic permeability μ, the velocity of light becomes,

$$v = \frac{1}{\sqrt{\mu\varepsilon}} \qquad (8.11)$$

Thus, the velocity of light depends on electric and magnetic properties of the medium. We shall see in the next chapter that the *refractive index* of one medium with respect to the other is equal to the ratio of velocities of light in the two media.

The velocity of electromagnetic waves in free space or vacuum is an important fundamental constant. It has been shown by experiments on electromagnetic waves of different wavelengths that this velocity is the

Simulate propagation of electromagnetic waves
(i) http://www.amanogawa.com/waves.html
(ii) http://www.phys.hawaii.edu/~teb/java/ntnujava/emWave/emWave.html

PHYSICS

same (independent of wavelength) to within a few metres per second, out of a value of 3×10^8 m/s. The constancy of the velocity of em waves in vacuum is so strongly supported by experiments and the actual value is so well known now that this is used to define a standard of *length*. Namely, the metre is now *defined* as the distance travelled by light in vacuum in a time $(1/c)$ seconds = $(2.99792458 \times 10^8)^{-1}$ seconds. This has come about for the following reason. The basic unit of time can be defined very accurately in terms of some atomic frequency, i.e., frequency of light emitted by an atom in a particular process. The basic unit of length is harder to define as accurately in a direct way. Earlier measurement of c using earlier units of length (metre rods, etc.) converged to a value of about 2.9979246×10^8 m/s. Since c is such a strongly fixed number, unit of length can be defined in terms of c and the unit of time!

Hertz not only showed the existence of electromagnetic waves, but also demonstrated that the waves, which had wavelength ten million times that of the light waves, could be diffracted, refracted and polarised. Thus, he conclusively established the wave nature of the radiation. Further, he produced stationary electromagnetic waves and determined their wavelength by measuring the distance between two successive nodes. Since the frequency of the wave was known (being equal to the frequency of the oscillator), he obtained the speed of the wave using the formula $\upsilon = \nu\lambda$ and found that the waves travelled with the same speed as the speed of light.

The fact that electromagnetic waves are polarised can be easily seen in the response of a portable AM radio to a broadcasting station. If an AM radio has a telescopic antenna, it responds to the electric part of the signal. When the antenna is turned horizontal, the signal will be greatly diminished. Some portable radios have horizontal antenna (usually inside the case of radio), which are sensitive to the magnetic component of the electromagnetic wave. Such a radio must remain horizontal in order to receive the signal. In such cases, response also depends on the orientation of the radio with respect to the station.

Do electromagnetic waves carry energy and momentum like other waves? Yes, they do. We have seen in chapter 2 that in a region of free space with electric field E, there is an energy density $(\varepsilon_0 E^2/2)$. Similarly, as seen in Chapter 6, associated with a magnetic field B is a magnetic energy density $(B^2/2\mu_0)$. As electromagnetic wave contains both electric and magnetic fields, there is a non-zero energy density associated with it. Now consider a plane perpendicular to the direction of propagation of the electromagnetic wave (Fig. 8.4). If there are, on this plane, electric charges, they will be set and sustained in motion by the electric and magnetic fields of the electromagnetic wave. The charges thus acquire energy and momentum from the waves. This just illustrates the fact that an electromagnetic wave (like other waves) carries energy and momentum. Since it carries momentum, an electromagnetic wave also exerts pressure, called *radiation pressure*.

If the total energy transferred to a surface in time t is U, it can be shown that the magnitude of the total momentum delivered to this surface (*for complete absorption*) is,

$$p = \frac{U}{c} \tag{8.12}$$

When the sun shines on your hand, you feel the energy being absorbed from the electromagnetic waves (your hands get warm). Electromagnetic waves also transfer momentum to your hand but because c is very large, the amount of momentum transferred is extremely small and you do not feel the pressure. In 1903, the American scientists Nicols and Hull succeeded in measuring radiation pressure of visible light and verified Eq. (8.12). It was found to be of the order of 7×10^{-6} N/m^2. Thus, on a surface of area 10 cm^2, the force due to radiation is only about 7×10^{-9} N.

The great technological importance of electromagnetic waves stems from their capability to carry energy from one place to another. The radio and TV signals from broadcasting stations carry energy. Light carries energy from the sun to the earth, thus making life possible on the earth.

EXAMPLE 8.2

Example 8.2 A plane electromagnetic wave of frequency 25 MHz travels in free space along the x-direction. At a particular point in space and time, $\mathbf{E} = 6.3\ \hat{\mathbf{j}}$ V/m. What is $\mathbf{B}$ at this point?

Solution Using Eq. (8.10), the magnitude of $\mathbf{B}$ is

$$B = \frac{E}{c}$$

$$= \frac{6.3\ \text{V/m}}{3 \times 10^8\ \text{m/s}} = 2.1 \times 10^{-8}\ \text{T}$$

To find the direction, we note that $\mathbf{E}$ is along y-direction and the wave propagates along x-axis. Therefore, $\mathbf{B}$ should be in a direction perpendicular to both x- and y-axes. Using vector algebra, $\mathbf{E} \times \mathbf{B}$ should be along x-direction. Since, $(+\hat{\mathbf{j}}) \times (+\hat{\mathbf{k}}) = \hat{\mathbf{i}}$, $\mathbf{B}$ is along the z-direction. Thus, $\mathbf{B} = 2.1 \times 10^{-8}\ \hat{\mathbf{k}}$ T

EXAMPLE 8.3

Example 8.3 The magnetic field in a plane electromagnetic wave is given by $B_y = 2 \times 10^{-7} \sin(0.5\times10^3 x + 1.5\times10^{11} t)$ T.

(a) What is the wavelength and frequency of the wave?

(b) Write an expression for the electric field.

Solution

(a) Comparing the given equation with

$$B_y = B_0 \sin\left[2\pi\left(\frac{x}{\lambda} + \frac{t}{T}\right)\right]$$

We get, $\lambda = \dfrac{2\pi}{0.5 \times 10^3}$ m = 1.26 cm,

and $\dfrac{1}{T} = \nu = \left(1.5 \times 10^{11}\right)/2\pi = 23.9$ GHz

(b) $E_0 = B_0 c = 2\times10^{-7}\ \text{T} \times 3 \times 10^8\ \text{m/s} = 6 \times 10^1$ V/m

The electric field component is perpendicular to the direction of propagation and the direction of magnetic field. Therefore, the electric field component along the z-axis is obtained as

$E_z = 60 \sin(0.5 \times 10^3 x + 1.5 \times 10^{11}\ t)$ V/m

Example 8.4 Light with an energy flux of 18 W/cm^2 falls on a non-reflecting surface at normal incidence. If the surface has an area of 20 cm^2, find the average force exerted on the surface during a 30 minute time span.

Solution

The total energy falling on the surface is

$$U = (18\ \text{W/cm}^2) \times (20\ \text{cm}^2) \times (30 \times 60)$$
$$= 6.48 \times 10^5\ \text{J}$$

Therefore, the total momentum delivered (for complete absorption) is

$$p = \frac{U}{c} = \frac{6.48 \times 10^5\ \text{J}}{3 \times 10^8\ \text{m/s}} = 2.16 \times 10^{-3}\ \text{kg m/s}$$

The average force exerted on the surface is

$$F = \frac{p}{t} = \frac{2.16 \times 10^{-3}}{0.18 \times 10^4} = 1.2 \times 10^{-6}\ \text{N}$$

How will your result be modified if the surface is a perfect reflector?

EXAMPLE 8.4

Example 8.5 Calculate the electric and magnetic fields produced by the radiation coming from a 100 W bulb at a distance of 3 m. Assume that the efficiency of the bulb is 2.5% and it is a point source.

Solution The bulb, as a point source, radiates light in all directions uniformly. At a distance of 3 m, the surface area of the surrounding sphere is

$$A = 4\pi r^2 = 4\pi(3)^2 = 113\ \text{m}^2$$

The intensity at this distance is

$$I = \frac{\text{Power}}{\text{Area}} = \frac{100\ \text{W} \times 2.5\%}{113\ \text{m}^2}$$
$$= 0.022\ \text{W/m}^2$$

Half of this intensity is provided by the electric field and half by the magnetic field.

$$\frac{1}{2} I = \frac{1}{2}\left(\varepsilon_0 E_{rms}^2 c\right)$$
$$= \frac{1}{2}\left(0.022\ \text{W/m}^2\right)$$

$$E_{rms} = \sqrt{\frac{0.022}{\left(8.85 \times 10^{-12}\right)\left(3 \times 10^8\right)}}\ \text{V/m}$$
$$= 2.9\ \text{V/m}$$

The value of E found above is the root mean square value of the electric field. Since the electric field in a light beam is sinusoidal, the peak electric field, E_0 is

$$E_0 = \sqrt{2} E_{\text{rms}} = \sqrt{2} \times 2.9\ \text{V/m}$$
$$= 4.07\ \text{V/m}$$

Thus, you see that the electric field strength of the light that you use for reading is fairly large. Compare it with electric field strength of TV or FM waves, which is of the order of a few microvolts per metre.

EXAMPLE 8.5

EXAMPLE 8.5

Now, let us calculate the strength of the magnetic field. It is

$$B_{rms} = \frac{E_{rms}}{c} = \frac{2.9 \ \text{V m}^{-1}}{3 \times 10^8 \text{ m s}^{-1}} = 9.6 \times 10^{-9} \text{ T}$$

Again, since the field in the light beam is sinusoidal, the peak magnetic field is $B_0 = \sqrt{2}\ B_{rms} = 1.4 \times 10^{-8}$ T. Note that although the energy in the magnetic field is equal to the energy in the electric field, the magnetic field strength is evidently very weak.

8.4 ELECTROMAGNETIC SPECTRUM

At the time Maxwell predicted the existence of electromagnetic waves, the only familiar electromagnetic waves were the visible light waves. The existence of ultraviolet and infrared waves was barely established. By the end of the nineteenth century, X-rays and gamma rays had also been discovered. We now know that, electromagnetic waves include visible light waves, X-rays, gamma rays, radio waves, microwaves, ultraviolet and infrared waves. The classification of em waves according to frequency is the electromagnetic spectrum (Fig. 8.5). *There is no sharp division between one kind of wave and the next.* The classification is based roughly on how the waves are produced and/or detected.

Electromagnetic spectrum
http://www.fnal.gov/pub/inquiring/more/light
http://imagine.gsfc.nasa.gov/docs/science/

FIGURE 8.5 The electromagnetic spectrum, with common names for various part of it. The various regions do not have sharply defined boundaries.

We briefly describe these different types of electromagnetic waves, in order of decreasing wavelengths.

8.4.1 Radio waves

Radio waves are produced by the accelerated motion of charges in conducting wires. They are used in radio and television communication systems. They are generally in the frequency range from 500 kHz to about 1000 MHz. The AM (amplitude modulated) band is from 530 kHz to 1710 kHz. Higher frequencies upto 54 MHz are used for *short wave* bands. TV waves range from 54 MHz to 890 MHz. The FM (frequency modulated) radio band extends from 88 MHz to 108 MHz. Cellular phones use radio waves to transmit voice communication in the ultrahigh frequency (UHF) band. How these waves are transmitted and received is described in Chapter 15.

8.4.2 Microwaves

Microwaves (short-wavelength radio waves), with frequencies in the gigahertz (GHz) range, are produced by special vacuum tubes (called klystrons, magnetrons and Gunn diodes). Due to their short wavelengths, they are suitable for the radar systems used in aircraft navigation. Radar also provides the basis for the speed guns used to time fast balls, tennis-serves, and automobiles. Microwave ovens are an interesting domestic application of these waves. In such ovens, the frequency of the microwaves is selected to match the resonant frequency of water molecules so that energy from the waves is transferred efficiently to the kinetic energy of the molecules. This raises the temperature of any food containing water.

MICROWAVE OVEN

The spectrum of *electromagnetic radiation* contains a part known as *microwaves*. These waves have frequency and energy smaller than visible light and wavelength larger than it. What is the principle of a microwave oven and how does it work?

Our objective is to cook food or warm it up. All food items such as fruit, vegetables, meat, cereals, etc., contain water as a constituent. Now, what does it mean when we say that a certain object has become warmer? When the temperature of a body rises, the energy of the random motion of atoms and molecules increases and the molecules travel or vibrate or rotate with higher energies. The frequency of rotation of water molecules is about 300 crore hertz, which is 3 gigahertz (GHz). If water receives microwaves of this frequency, its molecules absorb this radiation, which is equivalent to heating up water. These molecules share this energy with neighbouring food molecules, heating up the food.

One should use porcelain vessels and not metal containers in a microwave oven because of the danger of getting a shock from accumulated electric charges. Metals may also melt from heating. The porcelain container remains unaffected and cool, because its large molecules vibrate and rotate with much smaller frequencies, and thus cannot absorb microwaves. Hence, they do not get heated up.

Thus, the basic principle of a microwave oven is to generate microwave radiation of appropriate frequency in the working space of the oven where we keep food. This way energy is not wasted in heating up the vessel. In the conventional heating method, the vessel on the burner gets heated first, and then the food inside gets heated because of transfer of energy from the vessel. In the microwave oven, on the other hand, energy is directly delivered to water molecules which is shared by the entire food.

8.4.3 Infrared waves

Infrared waves are produced by hot bodies and molecules. This band lies adjacent to the low-frequency or long-wave length end of the visible spectrum. Infrared waves are sometimes referred to as *heat waves*. This is because water molecules present in most materials readily absorb infrared waves (many other molecules, for example, CO_2, NH_3, also absorb infrared waves). After absorption, their thermal motion increases, that is, they heat up and heat their surroundings. Infrared lamps are used in physical therapy. Infrared radiation also plays an important role in maintaining the earth's warmth or average temperature through the greenhouse effect. Incoming visible light (which passes relatively easily through the atmosphere) is absorbed by the earth's surface and re-radiated as infrared (longer wavelength) radiations. This radiation is trapped by greenhouse gases such as carbon dioxide and water vapour. Infrared detectors are used in Earth satellites, both for military purposes and to observe growth of crops. Electronic devices (for example semiconductor light emitting diodes) also emit infrared and are widely used in the remote switches of household electronic systems such as TV sets, video recorders and hi-fi systems.

8.4.4 Visible rays

It is the most familiar form of electromagnetic waves. It is the part of the spectrum that is detected by the human eye. It runs from about 4×10^{14} Hz to about 7×10^{14} Hz or a wavelength range of about 700 – 400 nm. Visible light emitted or reflected from objects around us provides us information about the world. Our eyes are sensitive to this range of wavelengths. Different animals are sensitive to different range of wavelengths. For example, snakes can detect infrared waves, and the 'visible' range of many insects extends well into the utraviolet.

8.4.5 Ultraviolet rays

It covers wavelengths ranging from about 4×10^{-7} m (400 nm) down to 6×10^{-10}m (0.6 nm). Ultraviolet (UV) radiation is produced by special lamps and very hot bodies. The sun is an important source of ultraviolet light. But fortunately, most of it is absorbed in the ozone layer in the atmosphere at an altitude of about 40 – 50 km. UV light in large quantities has harmful effects on humans. Exposure to UV radiation induces the production of more melanin, causing tanning of the skin. UV radiation is absorbed by ordinary glass. Hence, one cannot get tanned or sunburn through glass windows.

Welders wear special glass goggles or face masks with glass windows to protect their eyes from large amount of UV produced by welding arcs. Due to its shorter wavelengths, UV radiations can be focussed into very narrow beams for high precision applications such as LASIK (*Laser-assisted in situ keratomileusis*) eye surgery. UV lamps are used to kill germs in water purifiers.

Ozone layer in the atmosphere plays a protective role, and hence its depletion by chlorofluorocarbons (CFCs) gas (such as freon) is a matter of international concern.

8.4.6 X-rays

Beyond the UV region of the electromagnetic spectrum lies the X-ray region. We are familiar with X-rays because of its medical applications. It covers wavelengths from about 10^{-8} m (10 nm) down to 10^{-13} m (10^{-4} nm). One common way to generate X-rays is to bombard a metal target by high energy electrons. X-rays are used as a diagnostic tool in medicine and as a treatment for certain forms of cancer. Because X-rays damage or destroy living tissues and organisms, care must be taken to avoid unnecessary or over exposure.

8.4.7 Gamma rays

They lie in the upper frequency range of the electromagnetic spectrum and have wavelengths of from about 10^{-10}m to less than 10^{-14}m. This high frequency radiation is produced in nuclear reactions and also emitted by radioactive nuclei. They are used in medicine to destroy cancer cells.

Table 8.1 summarises different types of electromagnetic waves, their production and detections. As mentioned earlier, the demarcation between different region is not sharp and there are over laps.

TABLE 8.1 DIFFERENT TYPES OF ELECTROMAGNETIC WAVES

Type	Wavelength range	Production	Detection
Radio	> 0.1 m	Rapid acceleration and decelerations of electrons in aerials	Receiver's aerials
Microwave	0.1m to 1 mm	Klystron valve or magnetron valve	Point contact diodes
Infra-red	1mm to 700 nm	Vibration of atoms and molecules	Thermopiles Bolometer, Infrared photographic film
Light	700 nm to 400 nm	Electrons in atoms emit light when they move from one energy level to a lower energy level	The eye Photocells Photographic film
Ultraviolet	400 nm to 1nm	Inner shell electrons in atoms moving from one energy level to a lower level	Photocells Photographic film
X-rays	1nm to 10^{-3} nm	X-ray tubes or inner shell electrons	Photographic film Geiger tubes Ionisation chamber
Gamma rays	$<10^{-3}$ nm	Radioactive decay of the nucleus	-do-

SUMMARY

1. Maxwell found an inconsistency in the Ampere's law and suggested the existence of an additional current, called displacement current, to remove this inconsistency. This displacement current is due to time-varying electric field and is given by

$$i_d = \varepsilon_0 \frac{d\Phi_E}{dt}$$

and acts as a source of magnetic field in exactly the same way as conduction current.

2. An accelerating charge produces electromagnetic waves. An electric charge oscillating harmonically with frequency ν, produces electromagnetic waves of the same frequency ν. An electric dipole is a basic source of electromagnetic waves.
3. Electromagnetic waves with wavelength of the order of a few metres were first produced and detected in the laboratory by Hertz in 1887. He thus verified a basic prediction of Maxwell's equations.
4. Electric and magnetic fields oscillate sinusoidally in space and time in an electromagnetic wave. The oscillating electric and magnetic fields, **E** and **B** are perpendicular to each other, and to the direction of propagation of the electromagnetic wave. For a wave of frequency ν, wavelength λ, propagating along z-direction, we have

$$E = E_x(t) = E_0 \sin(kz - \omega t)$$

$$= E_0 \sin\left[2\pi\left(\frac{z}{\lambda} - \nu t\right)\right] = E_0 \sin\left[2\pi\left(\frac{z}{\lambda} - \frac{t}{T}\right)\right]$$

$$B = B_y(t) = B_0 \sin(kz - \omega t)$$

$$= B_0 \sin\left[2\pi\left(\frac{z}{\lambda} - \nu t\right)\right] = B_0 \sin\left[2\pi\left(\frac{z}{\lambda} - \frac{t}{T}\right)\right]$$

They are related by $E_0/B_0 = c$.

5. The speed c of electromagnetic wave in vacuum is related to μ_0 and ε_0 (the free space permeability and permittivity constants) as follows: $c = 1/\sqrt{\mu_0 \varepsilon_0}$. The value of c equals the speed of light obtained from optical measurements.

   Light is an electromagnetic wave; c is, therefore, also the speed of light. Electromagnetic waves other than light also have the same velocity c in free space.

   The speed of light, or of electromagnetic waves in a material medium is given by $v = 1/\sqrt{\mu \varepsilon}$

   where μ is the permeability of the medium and ε its permittivity.
6. Electromagnetic waves carry energy as they travel through space and this energy is shared equally by the electric and magnetic fields.

   Electromagnetic waves transport momentum as well. When these waves strike a surface, a pressure is exerted on the surface. If total energy transferred to a surface in time t is U, total momentum delivered to this surface is $p = U/c$.
7. The spectrum of electromagnetic waves stretches, in principle, over an infinite range of wavelengths. Different regions are known by different

names; γ-rays, X-rays, ultraviolet rays, visible rays, infrared rays, microwaves and radio waves in order of increasing wavelength from 10^{-2} Å or 10^{-12} m to 10^{6} m.

They interact with matter via their electric and magnetic fields which set in oscillation charges present in all matter. The detailed interaction and so the mechanism of absorption, scattering, etc., depend on the wavelength of the electromagnetic wave, and the nature of the atoms and molecules in the medium.

POINTS TO PONDER

1. The basic difference between various types of electromagnetic waves lies in their wavelengths or frequencies since all of them travel through vacuum with the same speed. Consequently, the waves differ considerably in their mode of interaction with matter.
2. Accelerated charged particles radiate electromagnetic waves. The wavelength of the electromagnetic wave is often correlated with the characteristic size of the system that radiates. Thus, gamma radiation, having wavelength of 10^{-14} m to 10^{-15} m, typically originate from an atomic nucleus. X-rays are emitted from heavy atoms. Radio waves are produced by accelerating electrons in a circuit. A transmitting antenna can most efficiently radiate waves having a wavelength of about the same size as the antenna. Visible radiation emitted by atoms is, however, much longer in wavelength than atomic size.
3. The oscillating fields of an electromagnetic wave can accelerate charges and can produce oscillating currents. Therefore, an apparatus designed to detect electromagnetic waves is based on this fact. Hertz original 'receiver' worked in exactly this way. The same basic principle is utilised in practically all modern receiving devices. High frequency electromagnetic waves are detected by other means based on the physical effects they produce on interacting with matter.
4. Infrared waves, with frequencies lower than those of visible light, vibrate not only the electrons, but entire atoms or molecules of a substance. This vibration increases the internal energy and consequently, the temperature of the substance. This is why infrared waves are often called *heat waves*.
5. The centre of sensitivity of our eyes coincides with the centre of the wavelength distribution of the sun. It is because humans have evolved with visions most sensitive to the strongest wavelengths from the sun.

EXERCISES

8.1 Figure 8.6 shows a capacitor made of two circular plates each of radius 12 cm, and separated by 5.0 cm. The capacitor is being charged by an external source (not shown in the figure). The charging current is constant and equal to 0.15A.

(a) Calculate the capacitance and the rate of charge of potential difference between the plates.

(b) Obtain the displacement current across the plates.

(c) Is Kirchhoff's first rule (junction rule) valid at each plate of the capacitor? Explain.

FIGURE 8.6

8.2 A parallel plate capacitor (Fig. 8.7) made of circular plates each of radius R = 6.0 cm has a capacitance C = 100 pF. The capacitor is connected to a 230 V ac supply with a (angular) frequency of 300 rad s^{-1}.

(a) What is the rms value of the conduction current?

(b) Is the conduction current equal to the displacement current?

(c) Determine the amplitude of **B** at a point 3.0 cm from the axis between the plates.

FIGURE 8.7

8.3 What physical quantity is the same for X-rays of wavelength 10^{-10} m, red light of wavelength 6800 Å and radiowaves of wavelength 500m?

8.4 A plane electromagnetic wave travels in vacuum along z-direction. What can you say about the directions of its electric and magnetic field vectors? If the frequency of the wave is 30 MHz, what is its wavelength?

8.5 A radio can tune in to any station in the 7.5 MHz to 12 MHz band. What is the corresponding wavelength band?

8.6 A charged particle oscillates about its mean equilibrium position with a frequency of 10^9 Hz. What is the frequency of the electromagnetic waves produced by the oscillator?

8.7 The amplitude of the magnetic field part of a harmonic electromagnetic wave in vacuum is B_0 = 510 nT. What is the amplitude of the electric field part of the wave?

8.8 Suppose that the electric field amplitude of an electromagnetic wave is E_0 = 120 N/C and that its frequency is ν = 50.0 MHz. (a) Determine, B_0, ω, k, and λ. (b) Find expressions for **E** and **B**.

8.9 The terminology of different parts of the electromagnetic spectrum is given in the text. Use the formula $E = h\nu$ (for energy of a quantum of radiation: photon) and obtain the photon energy in units of eV for different parts of the electromagnetic spectrum. In what way are the different scales of photon energies that you obtain related to the sources of electromagnetic radiation?

8.10 In a plane electromagnetic wave, the electric field oscillates sinusoidally at a frequency of 2.0×10^{10} Hz and amplitude 48 V m^{-1}.

(a) What is the wavelength of the wave?
(b) What is the amplitude of the oscillating magnetic field?
(c) Show that the average energy density of the **E** field equals the average energy density of the **B** field. [$c = 3 \times 10^8$ m s^{-1}.]

ADDITIONAL EXERCISES

8.11 Suppose that the electric field part of an electromagnetic wave in vacuum is $\mathbf{E} = \{(3.1\ \text{N/C}) \cos [(1.8\ \text{rad/m})\ y + (5.4 \times 10^6\ \text{rad/s})t]\}\ \hat{\mathbf{i}}$.

(a) What is the direction of propagation?
(b) What is the wavelength λ?
(c) What is the frequency ν?
(d) What is the amplitude of the magnetic field part of the wave?
(e) Write an expression for the magnetic field part of the wave.

8.12 About 5% of the power of a 100 W light bulb is converted to visible radiation. What is the average intensity of visible radiation

(a) at a distance of 1m from the bulb?
(b) at a distance of 10 m?

Assume that the radiation is emitted isotropically and neglect reflection.

8.13 Use the formula $\lambda_m T = 0.29$ cmK to obtain the characteristic temperature ranges for different parts of the electromagnetic spectrum. What do the numbers that you obtain tell you?

8.14 Given below are some famous numbers associated with electromagnetic radiations in different contexts in physics. State the part of the electromagnetic spectrum to which each belongs.

(a) 21 cm (wavelength emitted by atomic hydrogen in interstellar space).
(b) 1057 MHz (frequency of radiation arising from two close energy levels in hydrogen; known as Lamb shift).
(c) 2.7 K [temperature associated with the isotropic radiation filling all space-thought to be a relic of the 'big-bang' origin of the universe].
(d) 5890 Å - 5896 Å [double lines of sodium]
(e) 14.4 keV [energy of a particular transition in ^{57}Fe nucleus associated with a famous high resolution spectroscopic method (Mössbauer spectroscopy)].

8.15 Answer the following questions:

(a) Long distance radio broadcasts use short-wave bands. Why?
(b) It is necessary to use satellites for long distance TV transmission. Why?
(c) Optical and radiotelescopes are built on the ground but X-ray astronomy is possible only from satellites orbiting the earth. Why?
(d) The small ozone layer on top of the stratosphere is crucial for human survival. Why?
(e) If the earth did not have an atmosphere, would its average surface temperature be higher or lower than what it is now?
(f) Some scientists have predicted that a global nuclear war on the earth would be followed by a severe 'nuclear winter' with a devastating effect on life on earth. What might be the basis of this prediction?

PHYSICS

Textbook for Class XII

PART II

PHYSICS PART II CLASS XII NCERT

CONTENTS

FOREWORD v
PREFACE vii

CHAPTER NINE
RAY OPTICS AND OPTICAL INSTRUMENTS

9.1 Introduction 309
9.2 Reflection of Light by Spherical Mirrors 310
9.3 Refraction 316
9.4 Total Internal Reflection 319
9.5 Refraction at Spherical Surfaces and by Lenses 323
9.6 Refraction through a Prism 330
9.7 Dispersion by a Prism 332
9.8 Some Natural Phenomena due to Sunlight 333
9.9 Optical Instruments 335

CHAPTER TEN
WAVE OPTICS

10.1 Introduction 351
10.2 Huygens Principle 353
10.3 Refraction and reflection of plane waves using Huygens Principle 355
10.4 Coherent and Incoherent Addition of Waves 360
10.5 Interference of Light Waves and Young's Experiment 362
10.6 Diffraction 367
10.7 Polarisation 376

CHAPTER ELEVEN
DUAL NATURE OF RADIATION AND MATTER

11.1 Introduction 386
11.2 Electron Emission 387
11.3 Photoelectric Effect 388
11.4 Experimental Study of Photoelectric Effect 389
11.5 Photoelectric Effect and Wave Theory of Light 393
11.6 Einstein's Photoelectric Equation: Energy Quantum of Radiation 393
11.7 Particle Nature of Light: The Photon 395
11.8 Wave Nature of Matter 398
11.9 Davisson and Germer Experiment 403

CHAPTER TWELVE
ATOMS

12.1 Introduction 414
12.2 Alpha-particle Scattering and Rutherford's Nuclear Model of Atom 415
12.3 Atomic Spectra 420
12.4 Bohr Model of the Hydrogen Atom 422
12.5 The Line Spectra of the Hydrogen Atom 428
12.6 DE Broglie's Explanation of Bohr's Second Postulate of Quantisation 430

CHAPTER THIRTEEN
NUCLEI

13.1 Introduction 438
13.2 Atomic Masses and Composition of Nucleus 438
13.3 Size of the Nucleus 441
13.5 Nuclear Force 445
13.6 Radioactivity 446
13.7 Nuclear Energy 451

CHAPTER FOURTEEN
SEMICONDUCTOR ELECTRONICS: MATERIALS, DEVICES AND SIMPLE CIRCUITS

14.1 Introduction 467
14.2 Classification of Metals, Conductors and Semiconductors 468
14.3 Intrinsic Semiconductor 472
14.4 Extrinsic Semiconductor 474
14.5 p-n Junction 478
14.6 Semiconductor diode 479
14.7 Application of Junction Diode as a Rectifier 483
14.8 Special Purpose p-n Junction Diodes 485
14.9 Junction Transistor 490
14.10 Digital Electronics and Logic Gates 501
14.11 Integrated Circuits 505

CHAPTER FIFTEEN
COMMUNICATION SYSTEMS

15.1 Introduction 513
15.2 Elements of a Communication System 513
15.3 Basic Terminology Used in Electronic Communication Systems 515
15.4 Bandwidth of Signals 517
15.5 Bandwidth of Transmission Medium 518
15.6 Propagation of Electromagnetic Waves 519

15.7 Modulation and its Necessity 522
15.8 Amplitude Modulation 524
15.9 Production of Amplitude Modulated Wave 525
15.10 Detection of Amplitude Modulated Wave 526

ADDITIONAL INFORMATION 528

APPENDICES 532

ANSWERS 534

BIBLIOGRAPHY 552

INDEX 554

Chapter Nine

RAY OPTICS AND OPTICAL INSTRUMENTS

9.1 Introduction

Nature has endowed the human eye (retina) with the sensitivity to detect electromagnetic waves within a small range of the electromagnetic spectrum. Electromagnetic radiation belonging to this region of the spectrum (wavelength of about 400 nm to 750 nm) is called light. It is mainly through light and the sense of vision that we know and interpret the world around us.

There are two things that we can intuitively mention about light from common experience. First, that it travels with enormous speed and second, that it travels in a straight line. It took some time for people to realise that the speed of light is finite and measurable. Its presently accepted value in vacuum is $c = 2.99792458 \times 10^8$ m s^{-1}. For many purposes, it suffices to take $c = 3 \times 10^8$ m s^{-1}. The speed of light in vacuum is the highest speed attainable in nature.

The intuitive notion that light travels in a straight line seems to contradict what we have learnt in Chapter 8, that light is an electromagnetic wave of wavelength belonging to the visible part of the spectrum. How to reconcile the two facts? The answer is that the wavelength of light is very small compared to the size of ordinary objects that we encounter commonly (generally of the order of a few cm or larger). In this situation, as you will learn in Chapter 10, a light wave can be considered to travel from one point to another, along a straight line joining

them. The path is called a *ray* of light, and a bundle of such rays constitutes a *beam* of light.

In this chapter, we consider the phenomena of reflection, refraction and dispersion of light, using the ray picture of light. Using the basic laws of reflection and refraction, we shall study the image formation by plane and spherical reflecting and refracting surfaces. We then go on to describe the construction and working of some important optical instruments, including the human eye.

PARTICLE MODEL OF LIGHT

Newton's fundamental contributions to mathematics, mechanics, and gravitation often blind us to his deep experimental and theoretical study of light. He made pioneering contributions in the field of optics. He further developed the corpuscular model of light proposed by Descartes. It presumes that light energy is concentrated in tiny particles called *corpuscles*. He further assumed that corpuscles of light were massless elastic particles. With his understanding of mechanics, he could come up with a simple model of reflection and refraction. It is a common observation that a ball bouncing from a smooth plane surface obeys the laws of reflection. When this is an elastic collision, the magnitude of the velocity remains the same. As the surface is smooth, there is no force acting parallel to the surface, so the component of momentum in this direction also remains the same. Only the component perpendicular to the surface, i.e., the normal component of the momentum, gets reversed in reflection. Newton argued that smooth surfaces like mirrors reflect the corpuscles in a similar manner.

In order to explain the phenomena of refraction, Newton postulated that the speed of the corpuscles was greater in water or glass than in air. However, later on it was discovered that the speed of light is less in water or glass than in air.

In the field of optics, Newton – the experimenter, was greater than Newton – the theorist. He himself observed many phenomena, which were difficult to understand in terms of particle nature of light. For example, the colours observed due to a thin film of oil on water. Property of partial reflection of light is yet another such example. Everyone who has looked into the water in a pond sees image of the face in it, but also sees the bottom of the pond. Newton argued that some of the corpuscles, which fall on the water, get reflected and some get transmitted. But what property could distinguish these two kinds of corpuscles? Newton had to postulate some kind of unpredictable, chance phenomenon, which decided whether an individual corpuscle would be reflected or not. In explaining other phenomena, however, the corpuscles were presumed to behave as if they are identical. Such a dilemma does not occur in the wave picture of light. An incoming wave can be divided into two weaker waves at the boundary between air and water.

9.2 REFLECTION OF LIGHT BY SPHERICAL MIRRORS

We are familiar with the laws of reflection. The angle of reflection (i.e., the angle between reflected ray and the normal to the reflecting surface or the mirror) equals the angle of incidence (angle between incident ray and the normal). Also that the incident ray, reflected ray and the normal to the reflecting surface at the point of incidence lie in the same plane (Fig. 9.1). These laws are valid at each point on any reflecting surface whether plane or curved. However, we shall restrict our discussion to the special case of curved surfaces, that is, spherical surfaces. The normal in

this case is to be taken as normal to the tangent to surface at the point of incidence. That is, the normal is along the radius, the line joining the centre of curvature of the mirror to the point of incidence.

We have already studied that the geometric centre of a spherical mirror is called its pole while that of a spherical lens is called its optical centre. The line joining the pole and the centre of curvature of the spherical mirror is known as the *principal axis*. In the case of spherical lenses, the principal axis is the line joining the optical centre with its principal focus as you will see later.

FIGURE 9.1 The incident ray, reflected ray and the normal to the reflecting surface lie in the same plane.

9.2.1 Sign convention

To derive the relevant formulae for reflection by spherical mirrors and refraction by spherical lenses, we must first adopt a sign convention for measuring distances. In this book, we shall follow the *Cartesian sign convention*. According to this convention, all distances are measured from the pole of the mirror or the optical centre of the lens. The distances measured in the same direction as the incident light are taken as positive and those measured in the direction opposite to the direction of incident light are taken as negative (Fig. 9.2). The heights measured upwards with respect to *x*-axis and normal to the principal axis (*x*-axis) of the mirror/lens are taken as positive (Fig. 9.2). The heights measured downwards are taken as negative.

FIGURE 9.2 The Cartesian Sign Convention.

With a common accepted convention, it turns out that a single formula for spherical mirrors and a single formula for spherical lenses can handle all different cases.

9.2.2 Focal length of spherical mirrors

Figure 9.3 shows what happens when a parallel beam of light is incident on (a) a concave mirror, and (b) a convex mirror. We assume that the rays are *paraxial*, i.e., they are incident at points close to the pole P of the mirror and make small angles with the principal axis. The reflected rays converge at a point F on the principal axis of a concave mirror [Fig. 9.3(a)]. For a convex mirror, the reflected rays appear to diverge from a point F on its principal axis [Fig. 9.3(b)]. The point F is called the *principal focus* of the mirror. If the parallel paraxial beam of light were incident, making some angle with the principal axis, the reflected rays would converge (or appear to diverge) from a point in a plane through F normal to the principal axis. This is called the *focal plane* of the mirror [Fig. 9.3(c)].

FIGURE 9.3 Focus of a concave and convex mirror.

The distance between the focus F and the pole P of the mirror is called the *focal length* of the mirror, denoted by f. We now show that $f = R/2$, where R is the radius of curvature of the mirror. The geometry of reflection of an incident ray is shown in Fig. 9.4.

FIGURE 9.4 Geometry of reflection of an incident ray on (a) concave spherical mirror, and (b) convex spherical mirror.

Let C be the centre of curvature of the mirror. Consider a ray parallel to the principal axis striking the mirror at M. Then CM will be perpendicular to the mirror at M. Let θ be the angle of incidence, and MD be the perpendicular from M on the principal axis. Then,

$\angle MCP = \theta$ and $\angle MFP = 2\theta$

Now,

$$\tan\theta = \frac{MD}{CD} \text{ and } \tan 2\theta = \frac{MD}{FD} \qquad (9.1)$$

For small θ, which is true for paraxial rays, $\tan\theta \approx \theta$, $\tan 2\theta \approx 2\theta$. Therefore, Eq. (9.1) gives

$$\frac{MD}{FD} = 2\,\frac{MD}{CD}$$

$$\text{or, } FD = \frac{CD}{2} \qquad (9.2)$$

Now, for small θ, the point D is very close to the point P. Therefore, FD = f and CD = R. Equation (9.2) then gives

$$f = R/2 \qquad (9.3)$$

9.2.3 The mirror equation

If rays emanating from a point actually meet at another point after reflection and/or refraction, that point is called the *image* of the first point. The image is *real* if the rays actually converge to the point; it is

virtual if the rays do not actually meet but appear to diverge from the point when produced backwards. An image is thus a point-to-point correspondence with the object established through reflection and/or refraction.

In principle, we can take any two rays emanating from a point on an object, trace their paths, find their point of intersection and thus, obtain the image of the point due to reflection at a spherical mirror. In practice, however, it is convenient to choose any two of the following rays:

(i) The ray from the point which is parallel to the principal axis. The reflected ray goes through the focus of the mirror.

(ii) The ray passing through the centre of curvature of a concave mirror or appearing to pass through it for a convex mirror. The reflected ray simply retraces the path.

(iii) The ray passing through (or directed towards) the focus of the concave mirror or appearing to pass through (or directed towards) the focus of a convex mirror. The reflected ray is parallel to the principal axis.

(iv) The ray incident at any angle at the pole. The reflected ray follows laws of reflection.

FIGURE 9.5 Ray diagram for image formation by a concave mirror.

Figure 9.5 shows the ray diagram considering three rays. It shows the image A′B′ (in this case, real) of an object AB formed by a concave mirror. It does not mean that only three rays emanate from the point A. An infinite number of rays emanate from any source, in all directions. Thus, point A′ is image point of A if every ray originating at point A and falling on the concave mirror after reflection passes through the point A′.

We now derive the mirror equation or the relation between the object distance (u), image distance (v) and the focal length (f).

From Fig. 9.5, the two right-angled triangles A′B′F and MPF are similar. (For paraxial rays, MP can be considered to be a straight line perpendicular to CP.) Therefore,

$$\frac{B'A'}{PM} = \frac{B'F}{FP}$$

$$\text{or } \frac{B'A'}{BA} = \frac{B'F}{FP} \quad (\because PM = AB) \tag{9.4}$$

Since ∠ APB = ∠ A′PB′, the right angled triangles A′B′P and ABP are also similar. Therefore,

$$\frac{B'A'}{BA} = \frac{B'P}{BP} \tag{9.5}$$

Comparing Eqs. (9.4) and (9.5), we get

$$\frac{B'F}{FP} = \frac{B'P - FP}{FP} = \frac{B'P}{BP} \tag{9.6}$$

Equation (9.6) is a relation involving magnitude of distances. We now apply the sign convention. We note that light travels from the object to the mirror MPN. Hence this is taken as the positive direction. To reach

the object AB, image A′B′ as well as the focus F from the pole P, we have to travel opposite to the direction of incident light. Hence, all the three will have negative signs. Thus,

B′ P = –v, FP = –f, BP = –u

Using these in Eq. (9.6), we get

$$\frac{-v+f}{-f} = \frac{-v}{-u}$$

or $$\frac{v-f}{f} = \frac{v}{u}$$

$$\frac{1}{v}+\frac{1}{u} = \frac{1}{f} \tag{9.7}$$

This relation is known as the *mirror equation*.

The size of the image relative to the size of the object is another important quantity to consider. We define linear *magnification* (m) as the ratio of the height of the image (h') to the height of the object (h):

$$m = \frac{h'}{h} \tag{9.8}$$

h and h' will be taken positive or negative in accordance with the accepted sign convention. In triangles A′B′P and ABP, we have,

$$\frac{B'A'}{BA} = \frac{B'P}{BP}$$

With the sign convention, this becomes

$$\frac{-h'}{h} = \frac{-v}{-u}$$

so that

$$m = \frac{h'}{h} = -\frac{v}{u} \tag{9.9}$$

We have derived here the mirror equation, Eq. (9.7), and the magnification formula, Eq. (9.9), for the case of real, inverted image formed by a concave mirror. With the proper use of sign convention, these are, in fact, valid for all the cases of reflection by a spherical mirror (concave or convex) whether the image formed is real or virtual. Figure 9.6 shows the ray diagrams for virtual image formed by a concave and convex mirror. You should verify that Eqs. (9.7) and (9.9) are valid for these cases as well.

FIGURE 9.6 Image formation by (a) a concave mirror with object between P and F, and (b) a convex mirror.

Example 9.1 Suppose that the lower half of the concave mirror's reflecting surface in Fig. 9.5 is covered with an opaque (non-reflective) material. What effect will this have on the image of an object placed in front of the mirror?

Solution You may think that the image will now show only half of the object, but taking the laws of reflection to be true for all points of the remaining part of the mirror, the image will be that of the whole object. However, as the area of the reflecting surface has been reduced, the intensity of the image will be low (in this case, half).

EXAMPLE 9.1

Example 9.2 A mobile phone lies along the principal axis of a concave mirror, as shown in Fig. 9.7. Show by suitable diagram, the formation of its image. Explain why the magnification is not uniform. Will the distortion of image depend on the location of the phone with respect to the mirror?

FIGURE 9.7

Solution
The ray diagram for the formation of the image of the phone is shown in Fig. 9.7. The image of the part which is on the plane perpendicular to principal axis will be on the same plane. It will be of the same size, i.e., B′C = BC. You can yourself realise why the image is distorted.

EXAMPLE 9.2

Example 9.3 An object is placed at (i) 10 cm, (ii) 5 cm in front of a concave mirror of radius of curvature 15 cm. Find the position, nature, and magnification of the image in each case.

Solution

The focal length $f = -15/2$ cm = –7.5 cm

(i) The object distance $u = -10$ cm. Then Eq. (9.7) gives

$$\frac{1}{v}+\frac{1}{10}=\frac{1}{7.5}$$

or $$v=\frac{10\times 7.5}{-2.5} = -30 \text{ cm}$$

The image is 30 cm from the mirror on the same side as the object.

Also, magnification $m = -\frac{v}{u}=-\frac{(-30)}{(-10)}=-3$

The image is magnified, real and inverted.

EXAMPLE 9.3

EXAMPLE 9.3

(ii) The object distance u = –5 cm. Then from Eq. (9.7),

$$\frac{1}{v}+\frac{1}{-5}=\frac{1}{-7.5}$$

or $$v=\frac{5\times 7.5}{(7.5\quad 5)}=15\text{ cm}$$

This image is formed at 15 cm behind the mirror. It is a virtual image.

Magnification $m = -\frac{v}{u}=-\frac{15}{(-5)}=3$

The image is magnified, virtual and erect.

EXAMPLE 9.4

Example 9.4 Suppose while sitting in a parked car, you notice a jogger approaching towards you in the side view mirror of R = 2 m. If the jogger is running at a speed of 5 m s^{-1}, how fast the image of the jogger appear to move when the jogger is (a) 39 m, (b) 29 m, (c) 19 m, and (d) 9 m away.

Solution

From the mirror equation, Eq. (9.7), we get

$$v=\frac{fu}{u-f}$$

For convex mirror, since R = 2 m, f = 1 m. Then

for u = –39 m, $v=\frac{(-39)\times 1}{-39-1}=\frac{39}{40}\text{m}$

Since the jogger moves at a constant speed of 5 m s^{-1}, after 1 s the position of the image v (for u = –39 + 5 = –34) is (34/35)m.

The shift in the position of image in 1 s is

$$\frac{39}{40}-\frac{34}{35}=\frac{1365-1360}{1400}=\frac{5}{1400}=\frac{1}{280}\text{m}$$

Therefore, the average speed of the image when the jogger is between 39 m and 34 m from the mirror, is (1/280) m s^{-1}

Similarly, it can be seen that for u = –29 m, –19 m and –9 m, the speed with which the image appears to move is

$\frac{1}{150}\text{m s}^{-1}$, $\frac{1}{60}\text{m s}^{-1}$ and $\frac{1}{10}\text{m s}^{-1}$, respectively.

Although the jogger has been moving with a constant speed, the speed of his/her image appears to increase substantially as he/she moves closer to the mirror. This phenomenon can be noticed by any person sitting in a stationary car or a bus. In case of moving vehicles, a similar phenomenon could be observed if the vehicle in the rear is moving closer with a constant speed.

9.3 REFRACTION

When a beam of light encounters another transparent medium, a part of light gets reflected back into the first medium while the rest enters the other. A ray of light represents a beam. The direction of propagation of an obliquely incident ray of light that enters the other medium, changes

at the interface of the two media. This phenomenon is called *refraction of light*. Snell experimentally obtained the following laws of refraction:

(i) The incident ray, the refracted ray and the normal to the interface at the point of incidence, all lie in the same plane.

(ii) The ratio of the sine of the angle of incidence to the sine of angle of refraction is constant. Remember that the angles of incidence (i) and refraction (r) are the angles that the incident and its refracted ray make with the normal, respectively. We have

$$\frac{\sin i}{\sin r} = n_{21} \qquad (9.10)$$

FIGURE 9.8 Refraction and reflection of light.

where n_{21} is a constant, called the *refractive index* of the second medium with respect to the first medium. Equation (9.10) is the well-known Snell's law of refraction. We note that n_{21} is a characteristic of the pair of media (and also depends on the wavelength of light), but is independent of the angle of incidence.

From Eq. (9.10), if $n_{21} > 1$, $r < i$, i.e., the refracted ray bends towards the normal. In such a case medium 2 is said to be *optically denser* (or *denser*, in short) than medium 1. On the other hand, if $n_{21} < 1$, $r > i$, the refracted ray bends away from the normal. This is the case when incident ray in a denser medium refracts into a rarer medium.

Note: *Optical density should not be confused with mass density, which is mass per unit volume. It is possible that mass density of an optically denser medium may be less than that of an optically rarer medium (optical density is the ratio of the speed of light in two media). For example, turpentine and water. Mass density of turpentine is less than that of water but its optical density is higher.*

If n_{21} is the refractive index of medium 2 with respect to medium 1 and n_{12} the refractive index of medium 1 with respect to medium 2, then it should be clear that

$$n_{12} = \frac{1}{n_{21}} \qquad (9.11)$$

It also follows that if n_{32} is the refractive index of medium 3 with respect to medium 2 then $n_{32} = n_{31} \times n_{12}$, where n_{31} is the refractive index of medium 3 with respect to medium 1.

FIGURE 9.9 Lateral shift of a ray refracted through a parallel-sided slab.

Some elementary results based on the laws of refraction follow immediately. For a rectangular slab, refraction takes place at two interfaces (air-glass and glass-air). It is easily seen from Fig. 9.9 that $r_2 = i_1$, i.e., the emergent ray is parallel to the incident ray—there is no

FIGURE 9.10 Apparent depth for (a) normal, and (b) oblique viewing.

deviation, but it does suffer lateral displacement/ shift with respect to the incident ray. Another familiar observation is that the bottom of a tank filled with water appears to be raised (Fig. 9.10). For viewing near the normal direction, it can be shown that the apparent depth, (h_1) is real depth (h_2) divided by the refractive index of the medium (water).

The refraction of light through the atmosphere is responsible for many interesting phenomena. For example, the sun is visible a little before the actual sunrise and until a little after the actual sunset due to refraction of light through the atmosphere (Fig. 9.11). By actual sunrise we mean the actual crossing of the horizon by the sun. Figure 9.11 shows the actual and apparent positions of the sun with respect to the horizon. The figure is highly exaggerated to show the effect. The refractive index of air with respect to vacuum is 1.00029. Due to this, the apparent shift in the direction of the sun is by about half a degree and the corresponding time difference between actual sunset and apparent sunset is about 2 minutes (see Example 9.5). The apparent flattening (oval shape) of the sun at sunset and sunrise is also due to the same phenomenon.

FIGURE 9.11 Advance sunrise and delayed sunset due to atmospheric refraction.

EXAMPLE 9.5

Example 9.5 The earth takes 24 h to rotate once about its axis. How much time does the sun take to shift by 1° when viewed from the earth?

Solution

Time taken for 360° shift = 24 h

Time taken for 1° shift = 24/360 h = 4 min.

THE DROWNING CHILD, LIFEGUARD AND SNELL'S LAW

Consider a rectangular swimming pool PQSR; see figure here. A lifeguard sitting at G outside the pool notices a child drowning at a point C. The guard wants to reach the child in the shortest possible time. Let SR be the side of the pool between G and C. Should he/she take a straight line path GAC between G and C or GBC in which the path BC in water would be the shortest, or some other path GXC? The guard knows that his/her running speed v_1 on ground is higher than his/her swimming speed v_2.

Suppose the guard enters water at X. Let GX $= l_1$ and XC $= l_2$. Then the time taken to reach from G to C would be

$$t = \frac{l_1}{v_1} + \frac{l_2}{v_2}$$

To make this time minimum, one has to differentiate it (with respect to the coordinate of X) and find the point X when t is a minimum. On doing all this algebra (which we skip here), we find that the guard should enter water at a point where Snell's law is satisfied. To understand this, draw a perpendicular LM to side SR at X. Let $\angle$GXM = i and $\angle$CXL = r. Then it can be seen that t is minimum when

$$\frac{\sin i}{\sin r} = \frac{v_1}{v_2}$$

In the case of light v_1/v_2, the ratio of the velocity of light in vacuum to that in the medium, is the refractive index n of the medium.

In short, whether it is a wave or a particle or a human being, whenever two mediums and two velocities are involved, one must follow Snell's law if one wants to take the shortest time.

9.4 TOTAL INTERNAL REFLECTION

When light travels from an optically denser medium to a rarer medium at the interface, it is partly reflected back into the same medium and partly refracted to the second medium. This reflection is called the *internal reflection.*

When a ray of light enters from a denser medium to a rarer medium, it bends away from the normal, for example, the ray AO_1B in Fig. 9.12. The incident ray AO_1 is partially reflected (O_1C) and partially transmitted (O_1B) or refracted, the angle of refraction (r) being larger than the angle of incidence (i). As the angle of incidence increases, so does the angle of refraction, till for the ray AO_3, the angle of refraction is $\pi/2$. The refracted ray is bent so much away from the normal that it grazes the surface at the interface between the two media. This is shown by the ray AO_3D in Fig. 9.12. If the angle of incidence is increased still further (e.g., the ray AO_4), refraction is not possible, and the incident ray is totally reflected.

FIGURE 9.12 Refraction and internal reflection of rays from a point A in the denser medium (water) incident at different angles at the interface with a rarer medium (air).

This is called *total internal reflection*. When light gets reflected by a surface, normally some fraction of it gets transmitted. The reflected ray, therefore, is always less intense than the incident ray, howsoever smooth the reflecting surface may be. In total internal reflection, on the other hand, no transmission of light takes place.

The angle of incidence corresponding to an angle of refraction 90°, say $\angle AO_3N$, is called the *critical angle* (i_c) for the given pair of media. We see from Snell's law [Eq. (9.10)] that if the relative refractive index is less than one then, since the maximum value of sin r is unity, there is an upper limit to the value of sin i for which the law can be satisfied, that is, $i = i_c$ such that

$$\sin i_c = n_{21} \tag{9.12}$$

For values of i larger than i_c, Snell's law of refraction cannot be satisfied, and hence no refraction is possible.

The refractive index of denser medium 2 with respect to rarer medium 1 will be $n_{12} = 1/\sin i_c$. Some typical critical angles are listed in Table 9.1.

TABLE 9.1 CRITICAL ANGLE OF SOME TRANSPARENT MEDIA

Substance medium	Refractive index	Critical angle
Water	1.33	48.75°
Crown glass	1.52	41.14°
Dense flint glass	1.62	37.31°
Diamond	2.42	24.41°

A demonstration for total internal reflection

All optical phenomena can be demonstrated very easily with the use of a laser torch or pointer, which is easily available nowadays. Take a glass beaker with clear water in it. Stir the water a few times with a piece of soap, so that it becomes a little turbid. Take a laser pointer and shine its beam through the turbid water. You will find that the path of the beam inside the water shines brightly.

Shine the beam from below the beaker such that it strikes at the upper water surface at the other end. Do you find that it undergoes partial reflection (which is seen as a spot on the table below) and partial refraction [which comes out in the air and is seen as a spot on the roof; Fig. 9.13(a)]? Now direct the laser beam from one side of the beaker such that it strikes the upper surface of water more obliquely [Fig. 9.13(b)]. Adjust the direction of laser beam until you find the angle for which the refraction

above the water surface is totally absent and the beam is totally reflected back to water. This is total internal reflection at its simplest.

Pour this water in a long test tube and shine the laser light from top, as shown in Fig. 9.13(c). Adjust the direction of the laser beam such that it is totally internally reflected every time it strikes the walls of the tube. This is similar to what happens in optical fibres.

Take care not to look into the laser beam directly and not to point it at anybody's face.

FIGURE 9.13 Observing total internal reflection in water with a laser beam (refraction due to glass of beaker neglected being very thin).

9.4.1 Total internal reflection in nature and its technological applications

(i) *Mirage:* On hot summer days, the air near the ground becomes hotter than the air at higher levels. The refractive index of air increases with its density. Hotter air is less dense, and has smaller refractive index than the cooler air. If the air currents are small, that is, the air is still, the optical density at different layers of air increases with height. As a result, light from a tall object such as a tree, passes through a medium whose refractive index decreases towards the ground. Thus, a ray of light from such an object successively bends away from the normal and undergoes total internal reflection, if the angle of incidence for the air near the ground exceeds the critical angle. This is shown in Fig. 9.14(b). To a distant observer, the light appears to be coming from somewhere below the ground. The observer naturally assumes that light is being reflected from the ground, say, by a pool of water near the tall object. Such inverted images of distant tall objects cause an optical illusion to the observer. This phenomenon is called *mirage*. This type of mirage is especially common in hot deserts. Some of you might have noticed that while moving in a bus or a car during a hot summer day, a distant patch of road, especially on a highway, appears to be wet. But, you do not find any evidence of wetness when you reach that spot. This is also due to mirage.

FIGURE 9.14 (a) A tree is seen by an observer at its place when the air above the ground is at uniform temperature, (b) When the layers of air close to the ground have varying temperature with hottest layers near the ground, light from a distant tree may undergo total internal reflection, and the apparent image of the tree may create an illusion to the observer that the tree is near a pool of water.

FIGURE 9.15 Prisms designed to bend rays by 90° and 180° or to invert image without changing its size make use of total internal reflection.

(ii) *Diamond*: Diamonds are known for their spectacular brilliance. Their brilliance is mainly due to the total internal reflection of light inside them. The critical angle for diamond-air interface (≅ 24.4°) is very small, therefore once light enters a diamond, it is very likely to undergo total internal reflection inside it. Diamonds found in nature rarely exhibit the brilliance for which they are known. It is the technical skill of a diamond cutter which makes diamonds to sparkle so brilliantly. By cutting the diamond suitably, multiple total internal reflections can be made to occur.

(iii) *Prism*: Prisms designed to bend light by 90° or by 180° make use of total internal reflection [Fig. 9.15(a) and (b)]. Such a prism is also used to invert images without changing their size [Fig. 9.15(c)]. In the first two cases, the critical angle i_c for the material of the prism must be less than 45°. We see from Table 9.1 that this is true for both crown glass and dense flint glass.

(iv) *Optical fibres*: Now-a-days optical fibres are extensively used for transmitting audio and video signals through long distances. Optical fibres too make use of the phenomenon of total internal reflection. Optical fibres are fabricated with high quality composite glass/quartz fibres. Each fibre consists of a core and cladding. The refractive index of the material of the core is higher than that of the cladding.

FIGURE 9.16 Light undergoes successive total internal reflections as it moves through an optical fibre.

When a signal in the form of light is directed at one end of the fibre at a suitable angle, it undergoes repeated total internal reflections along the length of the fibre and finally comes out at the other end (Fig. 9.16). Since light undergoes total internal reflection at each stage, there is no appreciable loss in the intensity of the light signal. Optical fibres are fabricated such that light reflected at one side of inner surface strikes the other at an angle larger than the critical angle. Even if the fibre is bent, light can easily travel along its length. Thus, an optical fibre can be used to act as an optical pipe.

A bundle of optical fibres can be put to several uses. Optical fibres are extensively used for transmitting and receiving electrical signals which are converted to light by suitable transducers. Obviously, optical fibres can also be used for transmission of optical signals. For example, these are used as a 'light pipe' to facilitate visual examination of internal organs like esophagus, stomach and intestines. You might have seen a commonly

available decorative lamp with fine plastic fibres with their free ends forming a fountain like structure. The other end of the fibres is fixed over an electric lamp. When the lamp is switched on, the light travels from the bottom of each fibre and appears at the tip of its free end as a dot of light. The fibres in such decorative lamps are optical fibres.

The main requirement in fabricating optical fibres is that there should be very little absorption of light as it travels for long distances inside them. This has been achieved by purification and special preparation of materials such as quartz. In silica glass fibres, it is possible to transmit more than 95% of the light over a fibre length of 1 km. (Compare with what you expect for a block of ordinary window glass 1 km thick.)

9.5 Refraction at Spherical Surfaces and by Lenses

We have so far considered refraction at a plane interface. We shall now consider refraction at a spherical interface between two transparent media. An infinitesimal part of a spherical surface can be regarded as planar and the same laws of refraction can be applied at every point on the surface. Just as for reflection by a spherical mirror, the normal at the point of incidence is perpendicular to the tangent plane to the spherical surface at that point and, therefore, passes through its centre of curvature. We first consider refraction by a single spherical surface and follow it by thin lenses. A thin lens is a transparent optical medium bounded by two surfaces; at least one of which should be spherical. Applying the formula for image formation by a single spherical surface successively at the two surfaces of a lens, we shall obtain the lens maker's formula and then the lens formula.

9.5.1 Refraction at a spherical surface

Figure 9.17 shows the geometry of formation of image I of an object O on the principal axis of a spherical surface with centre of curvature C, and radius of curvature R. The rays are incident from a medium of refractive index n_1, to another of refractive index n_2. As before, we take the aperture (or the lateral size) of the surface to be small compared to other distances involved, so that small angle approximation can be made. In particular, NM will be taken to be nearly equal to the length of the perpendicular from the point N on the principal axis. We have, for small angles,

$$\tan \angle NOM = \frac{MN}{OM}$$

$$\tan \angle NCM = \frac{MN}{MC}$$

$$\tan \angle NIM = \frac{MN}{MI}$$

FIGURE 9.17 Refraction at a spherical surface separating two media.

Light sources and photometry

It is known that a body above absolute zero temperature emits electromagnetic radiation. The wavelength region in which the body emits the radiation depends on its absolute temperature. Radiation emitted by a hot body, for example, a tungsten filament lamp having temperature 2850 K are partly invisible and mostly in infrared (or heat) region. As the temperature of the body increases radiation emitted by it is in visible region. The sun with temperature of about 5500 K emits radiation whose energy versus wavelength graph peaks approximately at 550 nm corresponding to green light and is almost in the middle of the visible region. The energy versus wavelength distribution graph for a given body peaks at some wavelength, which is inversely proportional to the absolute temperature of that body.

The measurement of light as perceived by human eye is called *photometry*. Photometry is measurement of a physiological phenomenon, being the stimulus of light as received by the human eye, transmitted by the optic nerves and analysed by the brain. The main physical quantities in photometry are (i) the *luminous intensity* of the source, (ii) the *luminous flux* or flow of light from the source, and (iii) *illuminance* of the surface. The SI unit of *luminous intensity* (I) is candela (cd). The candela is the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency 540×10^{12} Hz and that has a radiant intensity in that direction of 1/683 watt per steradian. If a light source emits one candela of luminous intensity into a solid angle of one steradian, the total luminous flux emitted into that solid angle is one *lumen* (lm). A standard 100 watt incadescent light bulb emits approximately 1700 lumens.

In photometry, the only parameter, which can be measured directly is *illuminance*. It is defined as luminous flux incident per unit area on a surface (lm/m^2 or *lux*). Most light meters measure this quantity. The illuminance E, produced by a source of luminous intensity I, is given by $E = I/r^2$, where r is the normal distance of the surface from the source. A quantity named *luminance* (L), is used to characterise the brightness of emitting or reflecting flat surfaces. Its unit is cd/m^2 (sometimes called 'nit' in industry) . A good LCD computer monitor has a brightness of about 250 nits.

Now, for ΔNOC, i is the exterior angle. Therefore, $i = \angle NOM + \angle NCM$

$$i = \frac{MN}{OM} + \frac{MN}{MC} \tag{9.13}$$

Similarly,

$$r = \angle NCM - \angle NIM$$

$$\text{i.e., } r = \frac{MN}{MC} - \frac{MN}{MI} \tag{9.14}$$

Now, by Snell's law

$$n_1 \sin i = n_2 \sin r$$

or for small angles

$$n_1 i = n_2 r$$

Substituting i and r from Eqs. (9.13) and (9.14), we get

$$\frac{n_1}{\text{OM}} + \frac{n_2}{\text{MI}} = \frac{n_2 - n_1}{\text{MC}} \tag{9.15}$$

Here, OM, MI and MC represent magnitudes of distances. Applying the Cartesian sign convention,

OM = $-u$, MI = $+v$, MC = $+R$

Substituting these in Eq. (9.15), we get

$$\frac{n_2}{v} - \frac{n_1}{u} = \frac{n_2 - n_1}{R} \tag{9.16}$$

Equation (9.16) gives us a relation between object and image distance in terms of refractive index of the medium and the radius of curvature of the curved spherical surface. It holds for any curved spherical surface.

Example 9.6 Light from a point source in air falls on a spherical glass surface (n = 1.5 and radius of curvature = 20 cm). The distance of the light source from the glass surface is 100 cm. At what position the image is formed?

Solution

We use the relation given by Eq. (9.16). Here
u = – 100 cm, v = ?, R = + 20 cm, n_1 = 1, and n_2 = 1.5.
We then have

$$\frac{1.5}{v} + \frac{1}{100} = \frac{0.5}{20}$$

or v = +100 cm

The image is formed at a distance of 100 cm from the glass surface, in the direction of incident light.

EXAMPLE 9.6

9.5.2 Refraction by a lens

Figure 9.18(a) shows the geometry of image formation by a double convex lens. The image formation can be seen in terms of two steps: (i) The first refracting surface forms the image I_1 of the object O [Fig. 9.18(b)]. The image I_1 acts as a virtual object for the second surface that forms the image at I [Fig. 9.18(c)]. Applying Eq. (9.15) to the first interface ABC, we get

$$\frac{n_1}{\text{OB}} + \frac{n_2}{\text{BI}_1} = \frac{n_2 - n_1}{\text{BC}_1} \tag{9.17}$$

A similar procedure applied to the second interface* ADC gives,

$$-\frac{n_2}{\text{DI}_1} + \frac{n_1}{\text{DI}} = \frac{n_2 - n_1}{\text{DC}_2} \tag{9.18}$$

* Note that now the refractive index of the medium on the right side of ADC is n_1 while on its left it is n_2. Further DI_1 is negative as the distance is measured against the direction of incident light.

FIGURE 9.18 (a) The position of object, and the image formed by a double convex lens, (b) Refraction at the first spherical surface and (c) Refraction at the second spherical surface.

For a thin lens, $BI_1 = DI_1$. Adding Eqs. (9.17) and (9.18), we get

$$\frac{n_1}{OB}+\frac{n_1}{DI}=(n_2-n_1)\left(\frac{1}{BC_1}+\frac{1}{DC_2}\right) \qquad (9.19)$$

Suppose the object is at infinity, i.e., $OB \to \infty$ and $DI = f$, Eq. (9.19) gives

$$\frac{n_1}{f}=(n_2-n_1)\left(\frac{1}{BC_1}+\frac{1}{DC_2}\right) \qquad (9.20)$$

The point where image of an object placed at infinity is formed is called the *focus* F, of the lens and the distance f gives its *focal length*. A lens has two foci, F and F′, on either side of it (Fig. 9.19). By the sign convention,

$BC_1 = +R_1$,

$DC_2 = -R_2$

So Eq. (9.20) can be written as

$$\frac{1}{f}=(n_{21}-1)\left(\frac{1}{R_1}-\frac{1}{R_2}\right) \quad \left(\because n_{21}=\frac{n_2}{n_1}\right) \qquad (9.21)$$

Equation (9.21) is known as the *lens maker's formula*. It is useful to design lenses of desired focal length using surfaces of suitable radii of curvature. Note that the formula is true for a concave lens also. In that case R_1 is negative, R_2 positive and therefore, f is negative.

From Eqs. (9.19) and (9.20), we get

$$\frac{n_1}{OB}+\frac{n_1}{DI} = \frac{n_1}{f} \qquad (9.22)$$

Again, in the thin lens approximation, B and D are both close to the optical centre of the lens. Applying the sign convention,

$BO = -u$, $DI = +v$, we get

$$\frac{1}{v}-\frac{1}{u}=\frac{1}{f} \qquad (9.23)$$

Equation (9.23) is the familiar *thin lens formula*. Though we derived it for a real image formed by a convex lens, the formula is valid for both convex as well as concave lenses and for both real and virtual images.

It is worth mentioning that the two foci, F and F′, of a double convex or concave lens are equidistant from the optical centre. The focus on the side of the (original) source of light is called the *first focal point*, whereas the other is called the *second focal point*.

To find the image of an object by a lens, we can, in principle, take any two rays emanating from a point on an object; trace their paths using

the laws of refraction and find the point where the refracted rays meet (or appear to meet). In practice, however, it is convenient to choose any two of the following rays:

(i) A ray emanating from the object parallel to the principal axis of the lens after refraction passes through the second principal focus F′ (in a convex lens) or appears to diverge (in a concave lens) from the first principal focus F.

(ii) A ray of light, passing through the optical centre of the lens, emerges without any deviation after refraction.

(iii) A ray of light passing through the first principal focus (for a convex lens) or appearing to meet at it (for a concave lens) emerges parallel to the principal axis after refraction.

Figures 9.19(a) and (b) illustrate these rules for a convex and a concave lens, respectively. You should practice drawing similar ray diagrams for different positions of the object with respect to the lens and also verify that the lens formula, Eq. (9.23), holds good for all cases.

FIGURE 9.19 Tracing rays through (a) convex lens (b) concave lens.

Here again it must be remembered that each point on an object gives out infinite number of rays. All these rays will pass through the same image point after refraction at the lens.

Magnification (m) produced by a lens is defined, like that for a mirror, as the ratio of the size of the image to that of the object. Proceeding in the same way as for spherical mirrors, it is easily seen that for a lens

$$m = \frac{h'}{h} = \frac{v}{u} \tag{9.24}$$

When we apply the sign convention, we see that, for erect (and virtual) image formed by a convex or concave lens, m is positive, while for an inverted (and real) image, m is negative.

Example 9.7 A magician during a show makes a glass lens with n = 1.47 disappear in a trough of liquid. What is the refractive index of the liquid? Could the liquid be water?

Solution

The refractive index of the liquid must be equal to 1.47 in order to make the lens disappear. This means $n_1 = n_2$. This gives $1/f = 0$ or $f \to \infty$. The lens in the liquid will act like a plane sheet of glass. No, the liquid is not water. It could be glycerine.

EXAMPLE 9.7

9.5.3 Power of a lens

Power of a lens is a measure of the convergence or divergence, which a lens introduces in the light falling on it. Clearly, a lens of shorter focal

FIGURE 9.20 Power of a lens.

length bends the incident light more, while converging it in case of a convex lens and diverging it in case of a concave lens. The *power P* of a lens is defined as the tangent of the angle by which it converges or diverges a beam of light falling at unit distant from the optical centre (Fig. 9.20).

$\tan\delta = \frac{h}{f}$; if $h = 1$ $\tan\delta = \frac{1}{f}$ or $\delta = \frac{1}{f}$ for small value of δ. Thus,

$$P = \frac{1}{f} \tag{9.25}$$

The SI unit for power of a lens is dioptre (D): 1D = 1m^{-1}. The power of a lens of focal length of 1 metre is one dioptre. Power of a lens is positive for a converging lens and negative for a diverging lens. Thus, when an optician prescribes a corrective lens of power + 2.5 D, the required lens is a convex lens of focal length + 40 cm. A lens of power of – 4.0 D means a concave lens of focal length – 25 cm.

EXAMPLE 9.8

Example 9.8 (i) If f = 0.5 m for a glass lens, what is the power of the lens? (ii) The radii of curvature of the faces of a double convex lens are 10 cm and 15 cm. Its focal length is 12 cm. What is the refractive index of glass? (iii) A convex lens has 20 cm focal length in air. What is focal length in water? (Refractive index of air-water = 1.33, refractive index for air-glass = 1.5.)

Solution

(i) Power = +2 dioptre.

(ii) Here, we have f = +12 cm, R_1 = +10 cm, R_2 = –15 cm.
Refractive index of air is taken as unity.
We use the lens formula of Eq. (9.22). The sign convention has to be applied for f, R_1 and R_2.
Substituting the values, we have

$$\frac{1}{12} = (n-1)\left(\frac{1}{10} - \frac{1}{-15}\right)$$

This gives n = 1.5.

(iii) For a glass lens in air, n_2 = 1.5, n_1 = 1, f = +20 cm. Hence, the lens formula gives

$$\frac{1}{20} = 0.5\left[\frac{1}{R_1} - \frac{1}{R_2}\right]$$

For the same glass lens in water, n_2 = 1.5, n_1 = 1.33. Therefore,

$$\frac{1.33}{f} = (1.5 - 1.33)\left[\frac{1}{R_1} - \frac{1}{R_2}\right] \tag{9.26}$$

Combining these two equations, we find f = + 78.2 cm.

9.5.4 Combination of thin lenses in contact

Consider two lenses A and B of focal length f_1 and f_2 placed in contact with each other. Let the object be placed at a point O beyond the focus of

the first lens A (Fig. 9.21). The first lens produces an image at I_1. Since image I_1 is real, it serves as a virtual object for the second lens B, producing the final image at I. It must, however, be borne in mind that formation of image by the first lens is presumed only to facilitate determination of the position of the final image. In fact, the direction of rays emerging from the first lens gets modified in accordance with the angle at which they strike the second lens. Since the lenses are thin, we assume the optical centres of the lenses to be coincident. Let this central point be denoted by P.

FIGURE 9.21 Image formation by a combination of two thin lenses in contact.

For the image formed by the first lens A, we get

$$\frac{1}{v_1} - \frac{1}{u} = \frac{1}{f_1} \tag{9.27}$$

For the image formed by the second lens B, we get

$$\frac{1}{v} - \frac{1}{v_1} = \frac{1}{f_2} \tag{9.28}$$

Adding Eqs. (9.27) and (9.28), we get

$$\frac{1}{v} - \frac{1}{u} = \frac{1}{f_1} + \frac{1}{f_2} \tag{9.29}$$

If the two lens-system is regarded as equivalent to a single lens of focal length f, we have

$$\frac{1}{v} - \frac{1}{u} = \frac{1}{f}$$

so that we get

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} \tag{9.30}$$

The derivation is valid for any number of thin lenses in contact. If several thin lenses of focal length $f_1, f_2, f_3, \ldots$ are in contact, the effective focal length of their combination is given by

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} + \frac{1}{f_3} + \ldots \tag{9.31}$$

In terms of power, Eq. (9.31) can be written as

$$P = P_1 + P_2 + P_3 + \ldots \tag{9.32}$$

where P is the net power of the lens combination. Note that the sum in Eq. (9.32) is an algebraic sum of individual powers, so some of the terms on the right side may be positive (for convex lenses) and some negative (for concave lenses). Combination of lenses helps to obtain diverging or converging lenses of desired magnification. It also enhances sharpness of the image. Since the image formed by the first lens becomes the object for the second, Eq. (9.25) implies that the total magnification m of the combination is a product of magnification (m_1, m_2, m_3,...) of individual lenses

$$m = m_1 m_2 m_3 \ldots \tag{9.33}$$

Such a system of combination of lenses is commonly used in designing lenses for cameras, microscopes, telescopes and other optical instruments.

Example 9.9 Find the position of the image formed by the lens combination given in the Fig. 9.22.

FIGURE 9.22

Solution Image formed by the first lens

$$\frac{1}{v_1} - \frac{1}{u_1} = \frac{1}{f_1}$$

$$\frac{1}{v_1} - \frac{1}{-30} = \frac{1}{10}$$

or $v_1 = 15$ cm

The image formed by the first lens serves as the object for the second. This is at a distance of (15 – 5) cm = 10 cm to the right of the second lens. Though the image is real, it serves as a virtual object for the second lens, which means that the rays appear to come from it for the second lens.

$$\frac{1}{v_2} - \frac{1}{10} = \frac{1}{-10}$$

or $v_2 = \infty$

The virtual image is formed at an infinite distance to the left of the second lens. This acts as an object for the third lens.

$$\frac{1}{v_3} - \frac{1}{u_3} = \frac{1}{f_3}$$

or $$\frac{1}{v_3} = \frac{1}{\infty} + \frac{1}{30}$$

or $v_3 = 30$ cm

The final image is formed 30 cm to the right of the third lens.

EXAMPLE 9.9

9.6 Refraction through a Prism

Figure 9.23 shows the passage of light through a triangular prism ABC. The angles of incidence and refraction at the first face AB are i and r_1, while the angle of incidence (from glass to air) at the second face AC is r_2 and the angle of refraction or emergence e. The angle between the emergent ray RS and the direction of the incident ray PQ is called the *angle of deviation*, δ.

In the quadrilateral AQNR, two of the angles (at the vertices Q and R) are right angles. Therefore, the sum of the other angles of the quadrilateral is 180°.

$\angle A + \angle QNR = 180°$

From the triangle QNR,

$r_1 + r_2 + \angle QNR = 180°$

Comparing these two equations, we get

$$r_1 + r_2 = A \tag{9.34}$$

FIGURE 9.23 A ray of light passing through a triangular glass prism.

The total deviation δ is the sum of deviations at the two faces,

$\delta = (i - r_1) + (e - r_2)$

that is,

$$\delta = i + e - A \tag{9.35}$$

Thus, the angle of deviation depends on the angle of incidence. A plot between the angle of deviation and angle of incidence is shown in Fig. 9.24. You can see that, in general, any given value of δ, except for $i = e$, corresponds to two values i and hence of e. This, in fact, is expected from the symmetry of i and e in Eq. (9.35), i.e., δ remains the same if i and e are interchanged. Physically, this is related to the fact that the path of ray in Fig. 9.23 can be traced back, resulting in the same angle of deviation. At the minimum deviation D_m, the refracted ray inside the prism becomes parallel to its base. We have

$\delta = D_m$, $i = e$ which implies $r_1 = r_2$.

Equation (9.34) gives

$$2r = A \text{ or } r = \frac{A}{2} \tag{9.36}$$

In the same way, Eq. (9.35) gives

$$D_m = 2i - A, \text{ or } i = (A + D_m)/2 \tag{9.37}$$

The refractive index of the prism is

$$n_{21} = \frac{n_2}{n_1} = \frac{\sin[(A + D_m)/2]}{\sin[A/2]} \tag{9.38}$$

FIGURE 9.24 Plot of angle of deviation (δ) versus angle of incidence (i) for a triangular prism.

The angles A and D_m can be measured experimentally. Equation (9.38) thus provides a method of determining refractive index of the material of the prism.

For a small angle prism, i.e., a thin prism, D_m is also very small, and we get

$$n_{21} = \frac{\sin[(A + D_m)/2]}{\sin[A/2]} \simeq \frac{(A + D_m)/2}{A/2}$$

$D_m = (n_{21} - 1)A$

It implies that, thin prisms do not deviate light much.

9.7 Dispersion by a Prism

It has been known for a long time that when a narrow beam of sunlight, usually called white light, is incident on a glass prism, the emergent light is seen to be consisting of several colours. There is actually a continuous variation of colour, but broadly, the different component colours that appear in sequence are: **v**iolet, **i**ndigo, **b**lue, **g**reen, **y**ellow, **o**range and **r**ed (given by the acronym VIBGYOR). The red light bends the least, while the violet light bends the most (Fig. 9.25).

FIGURE 9.25 Dispersion of sunlight or white light on passing through a glass prism. The relative deviation of different colours shown is highly exaggerated.

The phenomenon of splitting of light into its component colours is known as *dispersion*. The pattern of colour components of light is called the spectrum of light. The word *spectrum* is now used in a much more general sense: we discussed in Chapter 8 the electro-magnetic spectrum over the large range of wavelengths, from γ-rays to radio waves, of which the spectrum of light (visible spectrum) is only a small part.

Though the reason for appearance of spectrum is now common knowledge, it was a matter of much debate in the history of physics. Does the prism itself create colour in some way or does it only separate the colours already present in white light?

In a classic experiment known for its simplicity but great significance, Isaac Newton settled the issue once for all. He put another similar prism, but in an inverted position, and let the emergent beam from the first prism fall on the second prism (Fig. 9.26). The resulting emergent beam was found to be white light. The explanation was clear— the first prism splits the white light into its component colours, while the inverted prism recombines them to give white light. Thus, white light itself consists of light of different colours, which are separated by the prism.

FIGURE 9.26 Schematic diagram of Newton's classic experiment on dispersion of white light.

It must be understood here that a ray of light, as defined mathematically, does not exist. An actual ray is really a beam of many rays of light. Each ray splits into component colours when it enters the glass prism. When those coloured rays come out on the other side, they again produce a white beam.

We now know that colour is associated with wavelength of light. In the visible spectrum, red light is at the long wavelength end (~700 nm) while the violet light is at the short wavelength end (~ 400 nm). Dispersion takes place because the refractive index of medium for different wavelengths (colours) is different. For example, the bending

of red component of white light is least while it is most for the violet. Equivalently, red light travels faster than violet light in a glass prism. Table 9.2 gives the refractive indices for different wavelength for crown glass and flint glass. Thick lenses could be assumed as made of many prisms, therefore, thick lenses show *chromatic aberration* due to dispersion of light.

TABLE 9.2 REFRACTIVE INDICES FOR DIFFERENT WAVELENGTHS

Colour	Wavelength (nm)	Crown glass	Flint glass
Violet	396.9	1.533	1.663
Blue	486.1	1.523	1.639
Yellow	589.3	1.517	1.627
Red	656.3	1.515	1.622

The variation of refractive index with wavelength may be more pronounced in some media than the other. In vacuum, of course, the speed of light is independent of wavelength. Thus, vacuum (or air approximately) is a non-dispersive medium in which all colours travel with the same speed. This also follows from the fact that sunlight reaches us in the form of white light and not as its components. On the other hand, glass is a dispersive medium.

9.8 SOME NATURAL PHENOMENA DUE TO SUNLIGHT

The interplay of light with things around us gives rise to several beautiful phenomena. The spectacle of colour that we see around us all the time is possible only due to sunlight. The blue of the sky, white clouds, the red-hue at sunrise and sunset, the rainbow, the brilliant colours of some pearls, shells, and wings of birds, are just a few of the natural wonders we are used to. We describe some of them here from the point of view of physics.

9.8.1 The rainbow

The rainbow is an example of the dispersion of sunlight by the water drops in the atmosphere. This is a phenomenon due to combined effect of dispersion, refraction and reflection of sunlight by spherical water droplets of rain. The conditions for observing a rainbow are that the sun should be shining in one part of the sky (say near western horizon) while it is raining in the opposite part of the sky (say eastern horizon). An observer can therefore see a rainbow only when his back is towards the sun.

In order to understand the formation of rainbows, consider Fig. (9.27(a). Sunlight is first refracted as it enters a raindrop, which causes the different wavelengths (colours) of white light to separate. Longer wangelength of light (red) are bent the least while the shorter wavelength (violet) are bent the most. Next, these component rays strike

PHYSICS

Formation of rainbows
http://www.eo.ucar.edu/rainbows
http://www.atoptics.co.uk/bows.htm

FIGURE 9.27 Rainbow: (a) The sun rays incident on a water drop get refracted twice and reflected internally by a drop; (b) Enlarge view of internal reflection and refraction of a ray of light inside a drop form primary rainbow; and (c) secondary rainbow is formed by rays undergoing internal reflection twice inside the drop.

the inner surface of the water drop and get internally reflected if the angle between the refracted ray and normal to the drop surface is greater then the critical angle (48°, in this case). The reflected light is refracted again as it comes out of the drop as shown in the figure. It is found that the violet light emerges at an angle of 40° related to the incoming sunlight and red light emerges at an angle of 42°. For other colours, angles lie in between these two values.

Figure 9.27(b) explains the formation of primary rainbow. We see that red light from drop 1 and violet light from drop 2 reach the observers eye. The violet from drop 1 and red light from drop 2 are directed at level above or below the observer. Thus the observer sees a rainbow with red colour on the top and violet on the bottom. Thus, the primary rainbow is a result of three-step process, that is, refraction, reflection and refraction.

When light rays undergoes *two* internal reflections inside a raindrop, instead of *one* as in the primary rainbow, a secondary rainbow is formed as shown in Fig. 9.27(c). It is due to four-step process. The intensity of light is reduced at the second reflection and hence the secondary rainbow is fainter than the primary rainbow. Further, the order of the colours is reversed in it as is clear from Fig. 9.27(c).

9.8.2 Scattering of light

As sunlight travels through the earth's atmosphere, it gets *scattered* (changes its direction) by the atmospheric particles. Light of shorter wavelengths is scattered much more than light of longer wavelengths. (The amount of scattering is inversely proportional to the fourth power of the wavelength. This is known as Rayleigh scattering). Hence, the bluish colour predominates in a clear sky, since blue has a shorter wavelength than red and is scattered much more strongly. In fact, violet gets scattered even more than blue, having a shorter wavelength. But since our eyes are more sensitive to blue than violet, we see the sky blue.

Large particles like dust and water droplets present in the atmosphere behave differently. The relevant quantity here is the relative size of the wavelength of light λ, and the scatterer (of typical size, say, a). For $a << \lambda$, one has Rayleigh scattering which is proportional to $(1/\lambda)^4$. For $a >> \lambda$, i.e., large scattering objects (for example, raindrops, large dust or ice particles) this is not true; all wavelengths are scattered nearly equally. Thus, clouds which have droplets of water with $a >> \lambda$ are generally white.

FIGURE 9.28 Sunlight travels through a longer distance in the atmosphere at sunset and sunrise.

At sunset or sunrise, the sun's rays have to pass through a larger distance in the atmosphere (Fig. 9.28). Most of the blue and other shorter wavelengths are removed by scattering. The least scattered light reaching our eyes, therefore, the sun looks reddish. This explains the reddish appearance of the sun and full moon near the horizon.

9.9 Optical Instruments

A number of optical devices and instruments have been designed utilising reflecting and refracting properties of mirrors, lenses and prisms. Periscope, kaleidoscope, binoculars, telescopes, microscopes are some

examples of optical devices and instruments that are in common use. Our eye is, of course, one of the most important optical device the nature has endowed us with. Starting with the eye, we then go on to describe the principles of working of the microscope and the telescope.

9.9.1 The eye

Figure 9.29 (a) shows the eye. Light enters the eye through a curved front surface, the cornea. It passes through the pupil which is the central hole in the iris. The size of the pupil can change under control of muscles. The light is further focussed by the eye lens on the retina. The retina is a film of nerve fibres covering the curved back surface of the eye. The retina contains rods and cones which sense light intensity and colour, respectively, and transmit electrical signals via the optic nerve to the brain which finally processes this information. The shape (curvature) and therefore the focal length of the lens can be modified somewhat by the ciliary muscles. For example, when the muscle is relaxed, the focal length is about 2.5 cm and objects at infinity are in sharp focus on the retina. When the object is brought closer to the eye, in order to maintain the same image-lens distance ($\cong$ 2.5 cm), the focal length of the eye lens becomes shorter by the action of the ciliary muscles. This property of the eye is called *accommodation*. If the object is too close to the eye, the lens cannot curve enough to focus the image on to the retina, and the image is blurred. The closest distance for which the lens can focus light on the retina is called the *least distance of distinct vision*, or the *near point*. The standard value for normal vision is taken as 25 cm. (Often the near point is given the symbol D.) This distance increases with age, because of the decreasing effectiveness of the ciliary muscle and the loss of flexibility of the lens. The near point may be as close as about 7 to 8 cm in a child ten years of age, and may increase to as much as 200 cm at 60 years of age. Thus, if an elderly person tries to read a book at about 25 cm from the eye, the image appears blurred. This condition (defect of the eye) is called *presbyopia*. It is corrected by using a converging lens for reading.

Thus, our eyes are marvellous organs that have the capability to interpret incoming electromagnetic waves as images through a complex process. These are our greatest assets and we must take proper care to protect them. Imagine the world without a pair of functional eyes. Yet many amongst us bravely face this challenge by effectively overcoming their limitations to lead a normal life. They deserve our appreciation for their courage and conviction.

In spite of all precautions and proactive action, our eyes may develop some defects due to various reasons. We shall restrict our discussion to some common optical defects of the eye. For example, the light from a distant object arriving at the eye-lens may get converged at a point in front of the retina. This type of defect is called *nearsightedness* or *myopia*. This means that the eye is producing too much convergence in the incident beam. To compensate this, we interpose a concave lens between the eye and the object, with the diverging effect desired to get the image focussed on the retina [Fig. 9.29(b)].

FIGURE 9.29 (a) The structure of the eye; (b) shortsighted or myopic eye and its correction; (c) farsighted or hypermetropic eye and its correction; and (d) astigmatic eye and its correction.

Similarly, if the eye-lens focusses the incoming light at a point behind the retina, a convergent lens is needed to compensate for the defect in vision. This defect is called *farsightedness* or *hypermetropia* [Fig. 9.29(c)].

Another common defect of vision is called *astigmatism.* This occurs when the cornea is not spherical in shape. For example, the cornea could have a larger curvature in the vertical plane than in the horizontal plane or vice-versa. If a person with such a defect in eye-lens looks at a wire mesh or a grid of lines, focussing in either the vertical or the horizontal plane may not be as sharp as in the other plane. Astigmatism results in lines in one direction being well focussed while those in a perpendicular direction may appear distorted [Fig. 9.29(d)]. Astigmatism can be corrected by using a cylindrical lens of desired radius of curvature with an appropriately directed axis. This defect can occur along with myopia or hypermetropia.

Example 9.10 What focal length should the reading spectacles have for a person for whom the least distance of distinct vision is 50 cm?

Solution The distance of normal vision is 25 cm. So if a book is at $u = -25$ cm, its image should be formed at $v = -50$ cm. Therefore, the desired focal length is given by

$$\frac{1}{f} = \frac{1}{v} - \frac{1}{u}$$

$$\text{or } \frac{1}{f} = \frac{1}{-50} - \frac{1}{-25} = \frac{1}{50}$$

or $f = +50$ cm (convex lens).

EXAMPLE 9.10

EXAMPLE 9.11

Example 9.11

(a) The far point of a myopic person is 80 cm in front of the eye. What is the power of the lens required to enable him to see very distant objects clearly?

(b) In what way does the corrective lens help the above person? Does the lens magnify very distant objects? Explain carefully.

(c) The above person prefers to remove his spectacles while reading a book. Explain why?

Solution

(a) Solving as in the previous example, we find that the person should use a concave lens of focal length = – 80 cm, i.e., of power = – 1.25 dioptres.

(b) No. The concave lens, in fact, reduces the size of the object, but the angle subtended by the distant object at the eye is the same as the angle subtended by the image (at the far point) at the eye. The eye is able to see distant objects not because the corrective lens magnifies the object, but because it brings the object (i.e., it produces virtual image of the object) at the far point of the eye which then can be focussed by the eye-lens on the retina.

(c) The myopic person may have a normal near point, i.e., about 25 cm (or even less). In order to read a book with the spectacles, such a person must keep the book at a distance greater than 25 cm so that the image of the book by the concave lens is produced not closer than 25 cm. The angular size of the book (or its image) at the greater distance is evidently less than the angular size when the book is placed at 25 cm and no spectacles are needed. Hence, the person prefers to remove the spectacles while reading.

EXAMPLE 9.12

Example 9.12 (a) The near point of a hypermetropic person is 75 cm from the eye. What is the power of the lens required to enable the person to read clearly a book held at 25 cm from the eye? (b) In what way does the corrective lens help the above person? Does the lens magnify objects held near the eye? (c) The above person prefers to remove the spectacles while looking at the sky. Explain why?

Solution

(a) $u = -25$ cm, $v = -75$ cm

$1/f = 1/25 - 1/75$, i.e., $f = 37.5$ cm.

The corrective lens needs to have a converging power of +2.67 dioptres.

(b) The corrective lens produces a virtual image (at 75 cm) of an object at 25 cm. The angular size of this image is the same as that of the object. In this sense the lens does not magnify the object but merely brings the object to the near point of the hypermetric eye, which then gets focussed on the retina. However, the angular size is greater than that of the same object at the near point (75 cm) viewed without the spectacles.

(c) A hypermetropic eye may have normal far point i.e., it may have enough converging power to focus parallel rays from infinity on the retina of the shortened eyeball. Wearing spectacles of converging lenses (used for near vision) will amount to more converging power than needed for parallel rays. Hence the person prefers not to use the spectacles for far objects.

9.9.2 The microscope

A simple magnifier or microscope is a converging lens of small focal length (Fig. 9.30). In order to use such a lens as a microscope, the lens is held near the object, one focal length away or less, and the eye is positioned close to the lens on the other side. The idea is to get an erect, magnified and virtual image of the object at a distance so that it can be viewed comfortably, i.e., at 25 cm or more. If the object is at a distance f, the image is at infinity. However, if the object is at a distance slightly less than the focal length of the lens, the image is virtual and closer than infinity. Although the closest comfortable distance for viewing the image is when it is at the near point (distance $D \cong 25$ cm), it causes some strain on the eye. Therefore, the image formed at infinity is often considered most suitable for viewing by the relaxed eye. We show both cases, the first in Fig. 9.30(a), and the second in Fig. 9.30(b) and (c).

FIGURE 9.30 A simple microscope; (a) the magnifying lens is located such that the image is at the near point, (b) the angle subtanded by the object, is the same as that at the near point, and (c) the object near the focal point of the lens; the image is far off but closer than infinity.

The linear magnification m, for the image formed at the near point D, by a simple microscope can be obtained by using the relation

$$m = \frac{v}{u} = v\left(\frac{1}{v} - \frac{1}{f}\right) = \left(1 - \frac{v}{f}\right)$$

Now according to our sign convention, v is negative, and is equal in magnitude to D. Thus, the magnification is

$$m = \left(1 + \frac{D}{f}\right) \qquad (9.39)$$

Since D is about 25 cm, to have a magnification of six, one needs a convex lens of focal length, $f = 5$ cm.

Note that $m = h'/h$ where h is the size of the object and h' the size of the image. This is also the ratio of the angle subtended by the image to that subtended by the object, if placed at D for comfortable viewing. (Note that this is not the angle actually subtended by the object at the eye, which is h/u.) What a single-lens simple magnifier achieves is that it allows the object to be brought closer to the eye than D.

We will now find the magnification when the image is at infinity. In this case we will have to obtained the *angular* magnification. Suppose the object has a height h. The maximum angle it can subtend, and be clearly visible (without a lens), is when it is at the near point, i.e., a distance D. The angle subtended is then given by

$$\tan\ \theta_o = \left(\frac{h}{D}\right) \approx \theta_o \qquad (9.40)$$

We now find the angle subtended at the eye by the image when the object is at u. From the relations

$$\frac{h'}{h} = m = \frac{v}{u}$$

we have the angle subtended by the image

$\tan\theta_i = \frac{h'}{-v} = \frac{h}{-v}\cdot\frac{v}{u} = \frac{h}{-u} \approx \theta$. The angle subtended by the object, when it is at $u = -f$.

$$\theta_i = \left(\frac{h}{f}\right) \tag{9.41}$$

as is clear from Fig. 9.29(c). The angular magnification is, therefore

$$m = \left(\frac{\theta_i}{\theta_o}\right) = \frac{D}{f} \tag{9.42}$$

This is one less than the magnification when the image is at the near point, Eq. (9.39), but the viewing is more comfortable and the difference in magnification is usually small. In subsequent discussions of optical instruments (microscope and telescope) we shall assume the image to be at infinity.

A simple microscope has a limited maximum magnification (≤ 9) for realistic focal lengths. For much larger magnifications, one uses two lenses, one compounding the effect of the other. This is known as a *compound microscope*. A schematic diagram of a compound microscope is shown in Fig. 9.31. The lens nearest the object, called the *objective*, forms a real, inverted, magnified image of the object. This serves as the object for the second lens, the *eyepiece*, which functions essentially like a simple microscope or magnifier, produces the final image, which is enlarged and virtual. The first inverted image is thus near (at or within) the focal plane of the eyepiece, at a distance appropriate for final image formation at infinity, or a little closer for image formation at the near point. Clearly, the final image is inverted with respect to the original object.

FIGURE 9.31 Ray diagram for the formation of image by a compound microscope.

We now obtain the magnification due to a compound microscope. The ray diagram of Fig. 9.31 shows that the (linear) magnification due to the objective, namely h'/h, equals

$$m_O = \frac{h'}{h} = \frac{L}{f_o} \tag{9.43}$$

where we have used the result

$$\tan\beta = \left(\frac{h}{f_o}\right) = \left(\frac{h'}{L}\right)$$

Here h' is the size of the first image, the object size being h and f_o being the focal length of the objective. The first image is formed near the focal point of the eyepiece. The distance L, i.e., the distance between the second focal point of the objective and the first focal point of the eyepiece (focal length f_e) is called the tube length of the compound microscope.

As the first inverted image is near the focal point of the eyepiece, we use the result from the discussion above for the simple microscope to obtain the (angular) magnification m_e due to it [Eq. (9.39)], when the final image is formed at the near point, is

$$m_e = \left(1 + \frac{D}{f_e}\right) \qquad [9.44(a)]$$

When the final image is formed at infinity, the angular magnification due to the eyepiece [Eq. (9.42)] is

$$m_e = (D/f_e) \qquad [9.44(b)]$$

Thus, the total magnification [(according to Eq. (9.33)], when the image is formed at infinity, is

$$m = m_o m_e = \left(\frac{L}{f_o}\right) \left(\frac{D}{f_e}\right) \qquad (9.45)$$

Clearly, to achieve a large magnification of a *small* object (hence the name microscope), the objective and eyepiece should have small focal lengths. In practice, it is difficult to make the focal length much smaller than 1 cm. Also large lenses are required to make L large.

For example, with an objective with f_o = 1.0 cm, and an eyepiece with focal length f_e = 2.0 cm, and a tube length of 20 cm, the magnification is

$$m = m_o m_e = \left(\frac{L}{f_o}\right) \left(\frac{D}{f_e}\right)$$

$$= \frac{20}{1} \times \frac{25}{2} = 250$$

Various other factors such as illumination of the object, contribute to the quality and visibility of the image. In modern microscopes, multi-component lenses are used for both the objective and the eyepiece to improve image quality by minimising various optical aberrations (defects) in lenses.

9.9.3 Telescope

The telescope is used to provide angular magnification of distant objects (Fig. 9.32). It also has an objective and an eyepiece. But here, the objective has a large focal length and a much larger aperture than the eyepiece. Light from a distant object enters the objective and a real image is formed in the tube at its second focal point. The eyepiece magnifies this image producing a final inverted image. The magnifying power m is the ratio of the angle β subtended at the eye by the final image to the angle α which the object subtends at the lens or the eye. Hence

$$m \approx \frac{\beta}{\alpha} \approx \frac{h}{f_e} \cdot \frac{f_o}{h} = \frac{f_o}{f_e} \qquad (9.46)$$

In this case, the length of the telescope tube is $f_o + f_e$.

PHYSICS

The world's largest optical telescopes
http://astro.nineplanets.org/bigeyes.html

FIGURE 9.32 A refracting telescope.

Terrestrial telescopes have, in addition, a pair of inverting lenses to make the final image erect. Refracting telescopes can be used both for terrestrial and astronomical observations. For example, consider a telescope whose objective has a focal length of 100 cm and the eyepiece a focal length of 1 cm. The magnifying power of this telescope is m = 100/1 = 100.

Let us consider a pair of stars of actual separation 1′ (one minute of arc). The stars appear as though they are separated by an angle of 100 × 1′ = 100′ =1.67°.

The main considerations with an astronomical telescope are its light gathering power and its resolution or resolving power. The former clearly depends on the area of the objective. With larger diameters, fainter objects can be observed. The resolving power, or the ability to observe two objects distinctly, which are in very nearly the same direction, also depends on the diameter of the objective. So, the desirable aim in optical telescopes is to make them with objective of large diameter. The largest lens objective in use has a diameter of 40 inch (~1.02 m). It is at the Yerkes Observatory in Wisconsin, USA. Such big lenses tend to be very heavy and therefore, difficult to make and support by their edges. Further, it is rather difficult and expensive to make such large sized lenses which form images that are free from any kind of chromatic aberration and distortions.

For these reasons, modern telescopes use a concave mirror rather than a lens for the objective. Telescopes with mirror objectives are called *reflecting* telescopes. They have several advantages. First, there is no chromatic aberration in a mirror. Second, if a parabolic reflecting surface is chosen, spherical aberration is also removed. Mechanical support is much less of a problem since a mirror weighs much less than a lens of equivalent optical quality, and can be supported over its entire back surface, not just over its rim. One obvious problem with a reflecting telescope is that the objective mirror focusses light inside the telescope tube. One must have an eyepiece and the observer right there, obstructing some light (depending on the size of the observer cage). This is what is done in the very large 200 inch (~5.08 m) diameters, Mt. Palomar telescope, California. The viewer sits near the focal point of the mirror, in a small cage. Another solution to the problem is to deflect the light being focussed by another mirror. One such arrangement using a convex secondary mirror to focus the incident light, which now passes through a hole in the objective primary mirror, is shown

FIGURE 9.33 Schematic diagram of a reflecting telescope (Cassegrain).

in Fig. 9.33. This is known as a *Cassegrain* telescope, after its inventor. It has the advantages of a large focal length in a short telescope. The largest telescope in India is in Kavalur, Tamil Nadu. It is a 2.34 m diameter reflecting telescope (Cassegrain). It was ground, polished, set up, and is being used by the Indian Institute of Astrophysics, Bangalore. The largest reflecting telescopes in the world are the pair of Keck telescopes in Hawaii, USA, with a reflector of 10 metre in diameter.

SUMMARY

1. Reflection is governed by the equation $\angle i = \angle r'$ and refraction by the Snell's law, $\sin i/\sin r = n$, where the incident ray, reflected ray, refracted ray and normal lie in the same plane. Angles of incidence, reflection and refraction are i, r' and r, respectively.
2. The *critical angle of incidence* i_c for a ray incident from a denser to rarer medium, is that angle for which the angle of refraction is 90°. For $i > i_c$, total internal reflection occurs. Multiple internal reflections in diamond ($i_c \cong 24.4°$), totally reflecting prisms and mirage, are some examples of total internal reflection. Optical fibres consist of glass fibres coated with a thin layer of material of *lower* refractive index. Light incident at an angle at one end comes out at the other, after multiple internal reflections, even if the fibre is bent.
3. *Cartesian sign convention*: Distances measured in the same direction as the incident light are positive; those measured in the opposite direction are negative. All distances are measured from the pole/optic centre of the mirror/lens on the principal axis. The heights measured upwards above x-axis and normal to the principal axis of the mirror/lens are taken as positive. The heights measured downwards are taken as negative.
4. *Mirror equation*:

   $$\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$$

   where u and v are object and image distances, respectively and f is the focal length of the mirror. f is (approximately) half the radius of curvature R. f is negative for concave mirror; f is positive for a convex mirror.
5. For a prism of the angle A, of refractive index n_2 placed in a medium of refractive index n_1,

   $$n_{21} = \frac{n_2}{n_1} = \frac{\sin\left[\left(A + D_m\right)/2\right]}{\sin\left(A/2\right)}$$

   where D_m is the angle of minimum deviation.
6. *For refraction through a spherical interface* (from medium 1 to 2 of refractive index n_1 and n_2, respectively)

   $$\frac{n_2}{v} - \frac{n_1}{u} = \frac{n_2 - n_1}{R}$$

   Thin lens formula

   $$\frac{1}{v} - \frac{1}{u} = \frac{1}{f}$$

Lens maker's formula

$$\frac{1}{f} = \frac{(n_2 - n_1)}{n_1}\left(\frac{1}{R_1} - \frac{1}{R_2}\right)$$

R_1 and R_2 are the radii of curvature of the lens surfaces. f is positive for a converging lens; f is negative for a diverging lens. The power of a lens $P = 1/f$.

The SI unit for power of a lens is dioptre (D): 1 D = 1 m^{-1}.

If several thin lenses of focal length f_1, f_2, f_3,.. are in contact, the effective focal length of their combination, is given by

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} + \frac{1}{f_3} + \ldots$$

The total power of a combination of several lenses is

$P = P_1 + P_2 + P_3 + \ldots$

7. *Dispersion* is the splitting of light into its constituent colours.
8. *The Eye*: The eye has a convex lens of focal length about 2.5 cm. This focal length can be varied somewhat so that the image is always formed on the retina. This ability of the eye is called *accommodation*. In a defective eye, if the image is focussed before the retina (myopia), a diverging corrective lens is needed; if the image is focussed beyond the retina (hypermetropia), a converging corrective lens is needed. Astigmatism is corrected by using cylindrical lenses.
9. *Magnifying power m of a simple microscope* is given by $m = 1 + (D/f)$, where D = 25 cm is the least distance of distinct vision and f is the focal length of the convex lens. If the image is at infinity, $m = D/f$. For a compound microscope, the magnifying power is given by $m = m_e \times m_0$ where $m_e = 1 + (D/f_e)$, is the magnification due to the eyepiece and m_o is the magnification produced by the objective. *Approximately,*

$$m = \frac{L}{f_o} \times \frac{D}{f_e}$$

where f_o and f_e are the focal lengths of the objective and eyepiece, respectively, and L is the distance between their focal points.

10. *Magnifying power m of a telescope* is the ratio of the angle β subtended at the eye by the image to the angle α subtended at the eye by the object.

$$m = \frac{\beta}{\alpha} = \frac{f_o}{f_e}$$

where f_0 and f_e are the focal lengths of the objective and eyepiece, respectively.

POINTS TO PONDER

1. The laws of reflection and refraction are true for all surfaces and pairs of media at the point of the incidence.
2. The real image of an object placed between f and $2f$ from a convex lens can be seen on a screen placed at the image location. If the screen is removed, is the image still there? This question puzzles many, because it is difficult to reconcile ourselves with an image suspended in air

without a screen. But the image does exist. Rays from a given point on the object are converging to an image point in space and diverging away. The screen simply diffuses these rays, some of which reach our eye and we see the image. This can be seen by the images formed in air during a laser show.

3. Image formation needs regular reflection/refraction. In principle, all rays from a given point should reach the same image point. This is why you do not see your image by an irregular reflecting object, say the page of a book.
4. Thick lenses give coloured images due to dispersion. The variety in colour of objects we see around us is due to the constituent colours of the light incident on them. A monochromatic light may produce an entirely different perception about the colours on an object as seen in white light.
5. For a simple microscope, the angular size of the object equals the angular size of the image. Yet it offers magnification because we can keep the small object much closer to the eye than 25 cm and hence have it subtend a large angle. The image is at 25 cm which we can see. Without the microscope, you would need to keep the small object at 25 cm which would subtend a very small angle.

EXERCISES

9.1 A small candle, 2.5 cm in size is placed at 27 cm in front of a concave mirror of radius of curvature 36 cm. At what distance from the mirror should a screen be placed in order to obtain a sharp image? Describe the nature and size of the image. If the candle is moved closer to the mirror, how would the screen have to be moved?

9.2 A 4.5 cm needle is placed 12 cm away from a convex mirror of focal length 15 cm. Give the location of the image and the magnification. Describe what happens as the needle is moved farther from the mirror.

9.3 A tank is filled with water to a height of 12.5 cm. The apparent depth of a needle lying at the bottom of the tank is measured by a microscope to be 9.4 cm. What is the refractive index of water? If water is replaced by a liquid of refractive index 1.63 up to the same height, by what distance would the microscope have to be moved to focus on the needle again?

9.4 Figures 9.34(a) and (b) show refraction of a ray in air incident at 60° with the normal to a glass-air and water-air interface, respectively. Predict the angle of refraction in glass when the angle of incidence in water is 45° with the normal to a water-glass interface [Fig. 9.34(c)].

FIGURE 9.34

9.5 A small bulb is placed at the bottom of a tank containing water to a depth of 80cm. What is the area of the surface of water through which light from the bulb can emerge out? Refractive index of water is 1.33. (Consider the bulb to be a point source.)

9.6 A prism is made of glass of unknown refractive index. A parallel beam of light is incident on a face of the prism. The angle of minimum deviation is measured to be 40°. What is the refractive index of the material of the prism? The refracting angle of the prism is 60°. If the prism is placed in water (refractive index 1.33), predict the new angle of minimum deviation of a parallel beam of light.

9.7 Double-convex lenses are to be manufactured from a glass of refractive index 1.55, with both faces of the same radius of curvature. What is the radius of curvature required if the focal length is to be 20cm?

9.8 A beam of light converges at a point P. Now a lens is placed in the path of the convergent beam 12cm from P. At what point does the beam converge if the lens is (a) a convex lens of focal length 20cm, and (b) a concave lens of focal length 16cm?

9.9 An object of size 3.0cm is placed 14cm in front of a concave lens of focal length 21cm. Describe the image produced by the lens. What happens if the object is moved further away from the lens?

9.10 What is the focal length of a convex lens of focal length 30cm in contact with a concave lens of focal length 20cm? Is the system a converging or a diverging lens? Ignore thickness of the lenses.

9.11 A compound microscope consists of an objective lens of focal length 2.0cm and an eyepiece of focal length 6.25cm separated by a distance of 15cm. How far from the objective should an object be placed in order to obtain the final image at (a) the least distance of distinct vision (25cm), and (b) at infinity? What is the magnifying power of the microscope in each case?

9.12 A person with a normal near point (25cm) using a compound microscope with objective of focal length 8.0 mm and an eyepiece of focal length 2.5cm can bring an object placed at 9.0mm from the objective in sharp focus. What is the separation between the two lenses? Calculate the magnifying power of the microscope,

9.13 A small telescope has an objective lens of focal length 144cm and an eyepiece of focal length 6.0cm. What is the magnifying power of the telescope? What is the separation between the objective and the eyepiece?

9.14 (a) A giant refracting telescope at an observatory has an objective lens of focal length 15m. If an eyepiece of focal length 1.0cm is used, what is the angular magnification of the telescope?

(b) If this telescope is used to view the moon, what is the diameter of the image of the moon formed by the objective lens? The diameter of the moon is 3.48×10^6m, and the radius of lunar orbit is 3.8×10^8m.

9.15 Use the mirror equation to deduce that:

(a) an object placed between f and $2f$ of a concave mirror produces a real image beyond $2f$.

(b) a convex mirror always produces a virtual image independent of the location of the object.

(c) the virtual image produced by a convex mirror is always diminished in size and is located between the focus and the pole.

(d) an object placed between the pole and focus of a concave mirror produces a virtual and enlarged image.

[*Note:* This exercise helps you deduce algebraically properties of images that one obtains from explicit ray diagrams.]

9.16 A small pin fixed on a table top is viewed from above from a distance of 50 cm. By what distance would the pin appear to be raised if it is viewed from the same point through a 15 cm thick glass slab held parallel to the table? Refractive index of glass = 1.5. Does the answer depend on the location of the slab?

9.17 (a) Figure 9.35 shows a cross-section of a 'light pipe' made of a glass fibre of refractive index 1.68. The outer covering of the pipe is made of a material of refractive index 1.44. What is the range of the angles of the incident rays with the axis of the pipe for which total reflections inside the pipe take place, as shown in the figure.

(b) What is the answer if there is no outer covering of the pipe?

FIGURE 9.35

9.18 Answer the following questions:

(a) You have learnt that plane and convex mirrors produce virtual images of objects. Can they produce real images under some circumstances? Explain.

(b) A virtual image, we always say, cannot be caught on a screen. Yet when we 'see' a virtual image, we are obviously bringing it on to the 'screen' (i.e., the retina) of our eye. Is there a contradiction?

(c) A diver under water, looks obliquely at a fisherman standing on the bank of a lake. Would the fisherman look taller or shorter to the diver than what he actually is?

(d) Does the apparent depth of a tank of water change if viewed obliquely? If so, does the apparent depth increase or decrease?

(e) The refractive index of diamond is much greater than that of ordinary glass. Is this fact of some use to a diamond cutter?

9.19 The image of a small electric bulb fixed on the wall of a room is to be obtained on the opposite wall 3 m away by means of a large convex lens. What is the maximum possible focal length of the lens required for the purpose?

9.20 A screen is placed 90 cm from an object. The image of the object on the screen is formed by a convex lens at two different locations separated by 20 cm. Determine the focal length of the lens.

9.21 (a) Determine the 'effective focal length' of the combination of the two lenses in Exercise 9.10, if they are placed 8.0 cm apart with their principal axes coincident. Does the answer depend on which side of the combination a beam of parallel light is incident? Is the notion of effective focal length of this system useful at all?

(b) An object 1.5 cm in size is placed on the side of the convex lens in the arrangement (a) above. The distance between the object

and the convex lens is 40 cm. Determine the magnification produced by the two-lens system, and the size of the image.

9.22 At what angle should a ray of light be incident on the face of a prism of refracting angle 60° so that it just suffers total internal reflection at the other face? The refractive index of the material of the prism is 1.524.

9.23 You are given prisms made of crown glass and flint glass with a wide variety of angles. Suggest a combination of prisms which will

(a) deviate a pencil of white light without much dispersion,

(b) disperse (and displace) a pencil of white light without much deviation.

9.24 For a normal eye, the far point is at infinity and the near point of distinct vision is about 25 cm in front of the eye. The cornea of the eye provides a converging power of about 40 dioptres, and the least converging power of the eye-lens behind the cornea is about 20 dioptres. From this rough data estimate the range of accommodation (i.e., the range of converging power of the eye-lens) of a normal eye.

9.25 Does short-sightedness (myopia) or long-sightedness (hyper-metropia) imply necessarily that the eye has partially lost its ability of accommodation? If not, what might cause these defects of vision?

9.26 A myopic person has been using spectacles of power –1.0 dioptre for distant vision. During old age he also needs to use separate reading glass of power + 2.0 dioptres. Explain what may have happened.

9.27 A person looking at a person wearing a shirt with a pattern comprising vertical and horizontal lines is able to see the vertical lines more distinctly than the horizontal ones. What is this defect due to? How is such a defect of vision corrected?

9.28 A man with normal near point (25 cm) reads a book with small print using a magnifying glass: a thin convex lens of focal length 5 cm.

(a) What is the closest and the farthest distance at which he should keep the lens from the page so that he can read the book when viewing through the magnifying glass?

(b) What is the maximum and the minimum angular magnification (magnifying power) possible using the above simple microscope?

9.29 A card sheet divided into squares each of size 1 mm^2 is being viewed at a distance of 9 cm through a magnifying glass (a converging lens of focal length 9 cm) held close to the eye.

(a) What is the magnification produced by the lens? How much is the area of each square in the virtual image?

(b) What is the angular magnification (magnifying power) of the lens?

(c) Is the magnification in (a) equal to the magnifying power in (b)? Explain.

9.30 (a) At what distance should the lens be held from the figure in Exercise 9.29 in order to view the squares distinctly with the maximum possible magnifying power?

(b) What is the magnification in this case?

(c) Is the magnification equal to the magnifying power in this case? Explain.

9.31 What should be the distance between the object in Exercise 9.30 and the magnifying glass if the virtual image of each square in the figure is to have an area of 6.25 mm^2. Would you be able to see the squares distinctly with your eyes very close to the magnifier?

[*Note:* Exercises 9.29 to 9.31 will help you clearly understand the difference between magnification in absolute size and the angular magnification (or magnifying power) of an instrument.]

9.32 Answer the following questions:

(a) The angle subtended at the eye by an object is equal to the angle subtended at the eye by the virtual image produced by a magnifying glass. In what sense then does a magnifying glass provide angular magnification?

(b) In viewing through a magnifying glass, one usually positions one's eyes very close to the lens. Does angular magnification change if the eye is moved back?

(c) Magnifying power of a simple microscope is inversely proportional to the focal length of the lens. What then stops us from using a convex lens of smaller and smaller focal length and achieving greater and greater magnifying power?

(d) Why must both the objective and the eyepiece of a compound microscope have short focal lengths?

(e) When viewing through a compound microscope, our eyes should be positioned not on the eyepiece but a short distance away from it for best viewing. Why? How much should be that short distance between the eye and eyepiece?

9.33 An angular magnification (magnifying power) of 30X is desired using an objective of focal length 1.25 cm and an eyepiece of focal length 5 cm. How will you set up the compound microscope?

9.34 A small telescope has an objective lens of focal length 140 cm and an eyepiece of focal length 5.0 cm. What is the magnifying power of the telescope for viewing distant objects when

(a) the telescope is in normal adjustment (i.e., when the final image is at infinity)?

(b) the final image is formed at the least distance of distinct vision (25 cm)?

9.35 (a) For the telescope described in Exercise 9.34 (a), what is the separation between the objective lens and the eyepiece?

(b) If this telescope is used to view a 100 m tall tower 3 km away, what is the height of the image of the tower formed by the objective lens?

(c) What is the height of the final image of the tower if it is formed at 25 cm?

9.36 A Cassegrain telescope uses two mirrors as shown in Fig. 9.33. Such a telescope is built with the mirrors 20 mm apart. If the radius of curvature of the large mirror is 220 mm and the small mirror is 140 mm, where will the final image of an object at infinity be?

9.37 Light incident normally on a plane mirror attached to a galvanometer coil retraces backwards as shown in Fig. 9.36. A current in the coil produces a deflection of 3.5° of the mirror. What is the displacement of the reflected spot of light on a screen placed 1.5 m away?

FIGURE 9.36

9.38 Figure 9.37 shows an equiconvex lens (of refractive index 1.50) in contact with a liquid layer on top of a plane mirror. A small needle with its tip on the principal axis is moved along the axis until its inverted image is found at the position of the needle. The distance of the needle from the lens is measured to be 45.0 cm. The liquid is removed and the experiment is repeated. The new distance is measured to be 30.0 cm. What is the refractive index of the liquid?

FIGURE 9.37

Chapter Ten

WAVE OPTICS

10.1 Introduction

In 1637 Descartes gave the corpuscular model of light and derived Snell's law. It explained the laws of reflection and refraction of light at an interface. The corpuscular model predicted that if the ray of light (on refraction) bends towards the normal then the speed of light would be greater in the second medium. This corpuscular model of light was further developed by Isaac Newton in his famous book entitled *OPTICKS* and because of the tremendous popularity of this book, the corpuscular model is very often attributed to Newton.

In 1678, the Dutch physicist Christiaan Huygens put forward the wave theory of light – it is this wave model of light that we will discuss in this chapter. As we will see, the wave model could satisfactorily explain the phenomena of reflection and refraction; however, it predicted that on refraction if the wave bends towards the normal then the speed of light would be less in the second medium. This is in contradiction to the prediction made by using the corpuscular model of light. It was much later confirmed by experiments where it was shown that the speed of light in water is less than the speed in air confirming the prediction of the wave model; Foucault carried out this experiment in 1850.

The wave theory was not readily accepted primarily because of Newton's authority and also because light could travel through vacuum

and it was felt that a wave would always require a medium to propagate from one point to the other. However, when Thomas Young performed his famous interference experiment in 1801, it was firmly established that light is indeed a wave phenomenon. The wavelength of visible light was measured and found to be extremely small; for example, the wavelength of yellow light is about 0.5 μm. Because of the smallness of the wavelength of visible light (in comparison to the dimensions of typical mirrors and lenses), light can be assumed to approximately travel in straight lines. This is the field of geometrical optics, which we had discussed in the previous chapter. Indeed, the branch of optics in which one completely neglects the finiteness of the wavelength is called geometrical optics and a ray is defined as the path of energy propagation in the limit of wavelength tending to zero.

After the interference experiment of Young in 1801, for the next 40 years or so, many experiments were carried out involving the interference and diffraction of lightwaves; these experiments could only be satisfactorily explained by assuming a wave model of light. Thus, around the middle of the nineteenth century, the wave theory seemed to be very well established. The only major difficulty was that since it was thought that a wave required a medium for its propagation, how could light waves propagate through vacuum. This was explained when Maxwell put forward his famous electromagnetic theory of light. Maxwell had developed a set of equations describing the laws of electricity and magnetism and using these equations he derived what is known as the wave equation from which he *predicted* the existence of electromagnetic waves*. From the wave equation, Maxwell could calculate the speed of electromagnetic waves in free space and he found that the theoretical value was very close to the measured value of speed of light. From this, he propounded that *light must be an electromagnetic wave*. Thus, according to Maxwell, light waves are associated with changing electric and magnetic fields; changing electric field produces a time and space varying magnetic field and a changing magnetic field produces a time and space varying electric field. The changing electric and magnetic fields result in the propagation of electromagnetic waves (or light waves) even in vacuum.

In this chapter we will first discuss the original formulation of the *Huygens principle* and derive the laws of reflection and refraction. In Sections 10.4 and 10.5, we will discuss the phenomenon of interference which is based on the principle of superposition. In Section 10.6 we will discuss the phenomenon of diffraction which is based on Huygens-Fresnel principle. Finally in Section 10.7 we will discuss the phenomenon of polarisation which is based on the fact that the light waves are *transverse electromagnetic waves*.

* Maxwell had predicted the existence of electromagnetic waves around 1855; it was much later (around 1890) that Heinrich Hertz produced radiowaves in the laboratory. J.C. Bose and G. Marconi made practical applications of the *Hertzian waves*

DOES LIGHT TRAVEL IN A STRAIGHT LINE?

Light travels in a straight line in Class VI; it does not do so in Class XII and beyond! Surprised, aren't you?

In school, you are shown an experiment in which you take three cardboards with pinholes in them, place a candle on one side and look from the other side. If the flame of the candle and the three pinholes are in a straight line, you can see the candle. Even if one of them is displaced a little, you cannot see the candle. *This proves*, so your teacher says, *that light travels in a straight line*.

In the present book, there are two consecutive chapters, one on ray optics and the other on wave optics. Ray optics is based on rectilinear propagation of light, and deals with mirrors, lenses, reflection, refraction, etc. Then you come to the chapter on wave optics, and you are told that light travels as a wave, that it can bend around objects, it can diffract and interfere, etc.

In optical region, light has a wavelength of about half a micrometre. If it encounters an obstacle of about this size, it can bend around it and can be seen on the other side. Thus a micrometre size obstacle will not be able to stop a light ray. If the obstacle is much larger, however, light will not be able to bend to that extent, and will not be seen on the other side.

This is a property of a wave in general, and can be seen in sound waves too. The sound wave of our speech has a wavelength of about 50 cm to 1 m. If it meets an obstacle of the size of a few metres, it bends around it and reaches points behind the obstacle. But when it comes across a larger obstacle of a few hundred metres, such as a hillock, most of it is reflected and is heard as an echo.

Then what about the primary school experiment? What happens there is that when we move any cardboard, the displacement is of the order of a few millimetres, which is much larger than the wavelength of light. Hence the candle cannot be seen. If we are able to move one of the cardboards by a micrometer or less, light will be able to diffract, and the candle will still be seen.

One could add to the first sentence in this box: *It learns how to bend as it grows up*!

10.2 HUYGENS PRINCIPLE

We would first define a wavefront: when we drop a small stone on a calm pool of water, waves spread out from the point of impact. Every point on the surface starts oscillating with time. At any instant, a photograph of the surface would show circular rings on which the disturbance is maximum. Clearly, all points on such a circle are oscillating in phase because they are at the same distance from the source. Such a locus of points, which oscillate in phase is called a *wavefront*; thus *a wavefront is defined as a surface of constant phase*. The speed with which the wavefront moves outwards from the source is called the speed of the wave. The energy of the wave travels in a direction perpendicular to the wavefront.

FIGURE 10.1 (a) A diverging spherical wave emanating from a point source. The wavefronts are spherical.

If we have a point source emitting waves uniformly in all directions, then the locus of points which have the same amplitude and vibrate in the same phase are spheres and we have what is known as a *spherical wave* as shown in Fig. 10.1(a). At a large distance from the source, a

FIGURE 10.1 (b) At a large distance from the source, a small portion of the spherical wave can be approximated by a plane wave.

small portion of the sphere can be considered as a plane and we have what is known as a *plane wave* [Fig. 10.1(b)].

Now, if we know the shape of the wavefront at $t = 0$, then Huygens principle allows us to determine the shape of the wavefront at a later time τ. Thus, Huygens principle is essentially a geometrical construction, which given the shape of the wafefront at any time allows us to determine the shape of the wavefront at a later time. Let us consider a diverging wave and let F_1F_2 represent a portion of the spherical wavefront at $t = 0$ (Fig. 10.2). Now, according to Huygens principle, *each point of the wavefront is the source of a secondary disturbance and the wavelets emanating from these points spread out in all directions with the speed of the wave. These wavelets emanating from the wavefront are usually referred to as secondary wavelets and if we draw a common tangent to all these spheres, we obtain the new position of the wavefront at a later time.*

FIGURE 10.2 F_1F_2 represents the spherical wavefront (with O as centre) at $t = 0$. The envelope of the secondary wavelets emanating from F_1F_2 produces the forward moving wavefront G_1G_2. The backwave D_1D_2 does not exist.

FIGURE 10.3 Huygens geometrical construction for a plane wave propagating to the right. F_1F_2 is the plane wavefront at $t = 0$ and G_1G_2 is the wavefront at a later time τ. The lines A_1A_2, B_1B_2 ... etc, are normal to both F_1F_2 and G_1G_2 and represent rays.

Thus, if we wish to determine the shape of the wavefront at $t = \tau$, we draw spheres of radius $v\tau$ from each point on the spherical wavefront where v represents the speed of the waves in the medium. If we now draw a common tangent to all these spheres, we obtain the new position of the wavefront at $t = \tau$. The new wavefront shown as G_1G_2 in Fig. 10.2 is again spherical with point O as the centre.

The above model has one shortcoming: we also have a backwave which is shown as D_1D_2 in Fig. 10.2. Huygens argued that the amplitude of the secondary wavelets is maximum in the forward direction and zero in the backward direction; by making this adhoc assumption, Huygens could explain the absence of the backwave. However, this adhoc assumption is not satisfactory and the absence of the backwave is really justified from more rigorous wave theory.

In a similar manner, we can use Huygens principle to determine the shape of the wavefront for a plane wave propagating through a medium (Fig. 10.3).

10.3 Refraction and Reflection of Plane Waves using Huygens Principle

10.3.1 Refraction of a plane wave

We will now use Huygens principle to derive the laws of refraction. Let PP′ represent the surface separating medium 1 and medium 2, as shown in Fig. 10.4. Let v_1 and v_2 represent the speed of light in medium 1 and medium 2, respectively. We assume a plane wavefront AB propagating in the direction A′A incident on the interface at an angle i as shown in the figure. Let τ be the time taken by the wavefront to travel the distance BC. Thus,

$$BC = v_1 \tau$$

FIGURE 10.4 *A plane wave* AB is incident at an angle i on the surface PP′ separating medium 1 and medium 2. The plane wave undergoes refraction and CE represents the refracted wavefront. The figure corresponds to $v_2 < v_1$ so that the refracted waves bends towards the normal.

Christiaan Huygens (1629 – 1695) Dutch physicist, astronomer, mathematician and the founder of the wave theory of light. His book, *Treatise on light*, makes fascinating reading even today. He brilliantly explained the double refraction shown by the mineral calcite in this work in addition to reflection and refraction. He was the first to analyse circular and simple harmonic motion and designed and built improved clocks and telescopes. He discovered the true geometry of Saturn's rings.

CHRISTIAAN HUYGENS (1629 – 1695)

In order to determine the shape of the refracted wavefront, we draw a sphere of radius $v_2\tau$ from the point A in the second medium (the speed of the wave in the second medium is v_2). Let CE represent a tangent plane drawn from the point C on to the sphere. Then, AE = $v_2 \tau$ and CE would represent the refracted wavefront. If we now consider the triangles ABC and AEC, we readily obtain

$$\sin i = \frac{BC}{AC} = \frac{v_1 \tau}{AC} \qquad (10.1)$$

and

$$\sin r = \frac{AE}{AC} = \frac{v_2 \tau}{AC} \qquad (10.2)$$

where i and r are the angles of incidence and refraction, respectively.

Thus we obtain

$$\frac{\sin i}{\sin r} = \frac{v_1}{v_2} \tag{10.3}$$

From the above equation, we get the important result that if $r < i$ (i.e., if the ray bends toward the normal), the speed of the light wave in the second medium (v_2) will be less then the speed of the light wave in the first medium (v_1). This prediction is opposite to the prediction from the corpuscular model of light and as later experiments showed, the prediction of the wave theory is correct. Now, if c represents the speed of light in vacuum, then,

$$n_1 = \frac{c}{v_1} \tag{10.4}$$

and

$$n_2 = \frac{c}{v_2} \tag{10.5}$$

are known as the refractive indices of medium 1 and medium 2, respectively. In terms of the refractive indices, Eq. (10.3) can be written as

$$n_1 \sin i = n_2 \sin r \tag{10.6}$$

This is the *Snell's law of refraction*. Further, if λ_1 and λ_2 denote the wavelengths of light in medium 1 and medium 2, respectively and if the distance BC is equal to λ_1 then the distance AE will be equal to λ_2 (because if the crest from B has reached C in time τ, then the crest from A should have also reached E in time τ); thus,

$$\frac{\lambda_1}{\lambda_2} = \frac{\text{BC}}{\text{AE}} = \frac{v_1}{v_2}$$

or

$$\frac{v_1}{\lambda_1} = \frac{v_2}{\lambda_2} \tag{10.7}$$

The above equation implies that when a wave gets refracted into a denser medium ($v_1 > v_2$) the wavelength and the speed of propagation decrease but the *frequency* ν (= v/λ) *remains the same*.

10.3.2 Refraction at a rarer medium

We now consider refraction of a plane wave at a rarer medium, i.e., $v_2 > v_1$. Proceeding in an exactly similar manner we can construct a refracted wavefront as shown in Fig. 10.5. The angle of refraction will now be greater than angle of incidence; however, we will still have $n_1 \sin i = n_2 \sin r$. We define an angle i_c by the following equation

$$\sin i_c = \frac{n_2}{n_1} \tag{10.8}$$

Thus, if $i = i_c$ then $\sin r = 1$ and $r = 90°$. Obviously, for $i > i_c$, there can not be any refracted wave. The angle i_c is known as the *critical angle* and for all angles of incidence greater than the critical angle, we will not have

Demonstration of interference, diffraction, refraction, resonance and Doppler effect
http://www.falstad.com/ripple/
PHYSICS

any refracted wave and the wave will undergo what is known as *total internal reflection*. The phenomenon of total internal reflection and its applications was discussed in Section 9.4.

FIGURE 10.5 Refraction of a plane wave incident on a rarer medium for which $v_2 > v_1$. The plane wave bends away from the normal.

10.3.3 Reflection of a plane wave by a plane surface

We next consider a plane wave AB incident at an angle i on a reflecting surface MN. If v represents the speed of the wave in the medium and if τ represents the time taken by the wavefront to advance from the point B to C then the distance

$$BC = v\tau$$

In order the construct the reflected wavefront we draw a sphere of radius $v\tau$ from the point A as shown in Fig. 10.6. Let CE represent the tangent plane drawn from the point C to this sphere. Obviously

$$AE = BC = v\tau$$

FIGURE 10.6 Reflection of a plane wave AB by the reflecting surface MN. AB and CE represent incident and reflected wavefronts.

If we now consider the triangles EAC and BAC we will find that they are congruent and therefore, the angles i and r (as shown in Fig. 10.6) would be equal. This is the *law of reflection*.

Once we have the laws of reflection and refraction, the behaviour of prisms, lenses, and mirrors can be understood. These phenomena were

discussed in detail in Chapter 9 on the basis of rectilinear propagation of light. Here we just describe the behaviour of the wavefronts as they undergo reflection or refraction. In Fig. 10.7(a) we consider a plane wave passing through a thin prism. Clearly, since the speed of light waves is less in glass, the lower portion of the incoming wavefront (which travels through the greatest thickness of glass) will get delayed resulting in a tilt in the emerging wavefront as shown in the figure. In Fig. 10.7(b) we consider a plane wave incident on a thin convex lens; the central part of the incident plane wave traverses the thickest portion of the lens and is delayed the most. The emerging wavefront has a depression at the centre and therefore the wavefront becomes spherical and converges to the point F which is known as the *focus*. In Fig. 10.7(c) a plane wave is incident on a concave mirror and on reflection we have a spherical wave converging to the focal point F. In a similar manner, we can understand refraction and reflection by concave lenses and convex mirrors.

FIGURE 10.7 Refraction of a plane wave by (a) a thin prism, (b) a convex lens. (c) Reflection of a plane wave by a concave mirror.

From the above discussion it follows that the total time taken from a point on the object to the corresponding point on the image is the same measured along any ray. For example, when a convex lens focusses light to form a real image, although the ray going through the centre traverses a shorter path, but because of the slower speed in glass, the time taken is the same as for rays travelling near the edge of the lens.

10.3.4 The doppler effect

We should mention here that one should be careful in constructing the wavefronts if the source (or the observer) is moving. For example, if there is no medium and the source moves away from the observer, then later wavefronts have to travel a greater distance to reach the observer and hence take a longer time. The time taken between the arrival of two successive wavefronts is hence longer at the observer than it is at the source. Thus, when the source moves away from the observer the frequency as measured by the source will be smaller. This is known as the *Doppler effect*. Astronomers call the increase in wavelength due to doppler effect as *red shift* since a wavelength in the middle of the visible region of the spectrum moves towards the red end of the spectrum. When waves are received from a source moving towards the observer, there is an apparent decrease in wavelength, this is referred to as *blue shift*.

You have already encountered Doppler effect for sound waves in Chapter 15 of Class XI textbook. For velocities small compared to the speed of light, we can use the same formulae which we use for sound waves. The fractional change in frequency $\Delta\nu/\nu$ is given by $-\upsilon_{radial}/c$, where υ_{radial} is the component of the source velocity along the line joining the observer to the source relative to the observer; υ_{radial} is considered positive when the source moves away from the observer. Thus, the Doppler shift can be expressed as:

$$\frac{\Delta\nu}{\nu} = -\frac{\upsilon_{radial}}{c} \tag{10.9}$$

The formula given above is valid only when the speed of the source is small compared to that of light. A more accurate formula for the Doppler effect which is valid even when the speeds are close to that of light, requires the use of Einstein's special theory of relativity. The Doppler effect for light is very important in astronomy. It is the basis for the measurements of the radial velocities of distant galaxies.

Example 10.1 What speed should a galaxy move with respect to us so that the sodium line at 589.0 nm is observed at 589.6 nm?

Solution Since $\nu\lambda = c$, $\frac{\Delta\nu}{\nu} = -\frac{\Delta\lambda}{\lambda}$ (for small changes in ν and λ). For

$\Delta\lambda = 589.6 - 589.0 = +0.6$ nm

we get [using Eq. (10.9)]

$$\frac{\Delta\nu}{\nu} = -\frac{\Delta\lambda}{\lambda} = -\frac{\upsilon_{radial}}{c}$$

or, $\upsilon_{radial} \cong +c\left(\frac{0.6}{589.0}\right) = +3.06\times10^{5}\,\text{m s}^{-1}$

$= 306$ km/s

Therefore, the galaxy is moving away from us.

EXAMPLE 10.1

Example 10.2

(a) When monochromatic light is incident on a surface separating two media, the reflected and refracted light both have the same frequency as the incident frequency. Explain why?

(b) When light travels from a rarer to a denser medium, the speed decreases. Does the reduction in speed imply a reduction in the energy carried by the light wave?

(c) In the wave picture of light, intensity of light is determined by the square of the amplitude of the wave. What determines the intensity of light in the photon picture of light.

Solution

(a) Reflection and refraction arise through interaction of incident light with the atomic constituents of matter. Atoms may be viewed as

EXAMPLE 10.2

EXAMPLE 10.2

oscillators, which take up the frequency of the external agency (light) causing forced oscillations. The frequency of light emitted by a charged oscillator equals its frequency of oscillation. Thus, the frequency of scattered light equals the frequency of incident light.

(b) No. Energy carried by a wave depends on the amplitude of the wave, not on the speed of wave propagation.

(c) For a given frequency, intensity of light in the photon picture is determined by the number of photons crossing an unit area per unit time.

FIGURE 10.8 (a) Two needles oscillating in phase in water represent two coherent sources.
(b) The pattern of displacement of water molecules at an instant on the surface of water showing nodal N (no displacement) and antinodal A (maximum displacement) lines.

10.4 COHERENT AND INCOHERENT ADDITION OF WAVES

In this section we will discuss the interference pattern produced by the superposition of two waves. You may recall that we had discussed the superposition principle in Chapter 15 of your Class XI textbook. Indeed the entire field of interference is based on the *superposition principle* according to which *at a particular point in the medium, the resultant displacement produced by a number of waves is the vector sum of the displacements produced by each of the waves*.

Consider two needles S_1 and S_2 moving periodically up and down in an identical fashion in a trough of water [Fig. 10.8(a)]. They produce two water waves, and at a particular point, the phase difference between the displacements produced by each of the waves does not change with time; when this happens the two sources are said to be *coherent*. Figure 10.8(b) shows the position of crests (solid circles) and troughs (dashed circles) at a given instant of time. Consider a point P for which

$$S_1 P = S_2 P$$

Since the distances S_1 P and S_2 P are equal, waves from S_1 and S_2 will take the same time to travel to the point P and waves that emanate from S_1 and S_2 in phase will also arrive, at the point P, in phase.

Thus, if the displacement produced by the source S_1 at the point P is given by

$$y_1 = a \cos \omega t$$

then, the displacement produced by the source S_2 (at the point P) will also be given by

$$y_2 = a \cos \omega t$$

Thus, the resultant of displacement at P would be given by

$$y = y_1 + y_2 = 2\, a \cos \omega t$$

Since the intensity is the proportional to the square of the amplitude, the resultant intensity will be given by

$$I = 4\, I_0$$

where I_0 represents the intensity produced by each one of the individual sources; I_0 is proportional to a^2. In fact at any point on the perpendicular bisector of S_1S_2, the intensity will be $4I_0$. The two sources are said to

interfere constructively and we have what is referred to as *constructive interference*. We next consider a point Q [Fig. 10.9(a)] for which

$$S_2Q - S_1Q = 2\lambda$$

The waves emanating from S_1 will arrive exactly two cycles earlier than the waves from S_2 and will again be in phase [Fig. 10.9(a)]. Thus, if the displacement produced by S_1 is given by

$$y_1 = a \cos \omega t$$

then the displacement produced by S_2 will be given by

$$y_2 = a \cos (\omega t - 4\pi) = a \cos \omega t$$

where we have used the fact that a path difference of 2λ corresponds to a phase difference of 4π. The two displacements are once again in phase and the intensity will again be $4\,I_0$ giving rise to constructive interference. In the above analysis we have assumed that the distances S_1Q and S_2Q are much greater than d (which represents the distance between S_1 and S_2) so that although S_1Q and S_2Q are not equal, the amplitudes of the displacement produced by each wave are very nearly the same.

We next consider a point R [Fig. 10.9(b)] for which

$$S_2R - S_1R = -2.5\lambda$$

The waves emanating from S_1 will arrive exactly two and a half cycles later than the waves from S_2 [Fig. 10.10(b)]. Thus if the displacement produced by S_1 is given by

$$y_1 = a \cos \omega t$$

then the displacement produced by S_2 will be given by

$$y_2 = a \cos (\omega t + 5\pi) = -\, a \cos \omega t$$

where we have used the fact that a path difference of 2.5λ corresponds to a phase difference of 5π. The two displacements are now out of phase and the two displacements will cancel out to give zero intensity. This is referred to as *destructive interference*.

To summarise: If we have two coherent sources S_1 and S_2 vibrating in phase, then for an arbitrary point P whenever the path difference,

$$S_1P \sim S_2P = n\lambda \quad (n = 0, 1, 2, 3, \ldots) \tag{10.10}$$

we will have constructive interference and the resultant intensity will be $4I_0$; the sign ~ between S_1P and S_2P represents the difference between S_1P and S_2P. On the other hand, if the point P is such that the path difference,

$$S_1P \sim S_2P = (n + \frac{1}{2})\,\lambda \quad (n = 0, 1, 2, 3, \ldots) \tag{10.11}$$

we will have *destructive interference* and the resultant intensity will be zero. Now, for any other arbitrary point G (Fig. 10.10) let the phase difference between the two displacements be ϕ. Thus, if the displacement produced by S_1 is given by

$$y_1 = a \cos \omega t$$

FIGURE 10.9 (a) Constructive interference at a point Q for which the path difference is 2λ. (b) Destructive interference at a point R for which the path difference is $2.5\,\lambda$.

FIGURE 10.10 Locus of points for which $S_1P - S_2P$ is equal to zero, $\pm\lambda$, $\pm 2\lambda$, $\pm 3\lambda$.

then, the displacement produced by S_2 would be

$$y_2 = a \cos(\omega t + \phi)$$

and the resultant displacement will be given by

$$y = y_1 + y_2$$

$$= a[\cos \omega t + \cos(\omega t + \phi)]$$

$$= 2a \cos(\phi/2) \cos(\omega t + \phi/2)$$

The amplitude of the resultant displacement is $2a \cos(\phi/2)$ and therefore the intensity at that point will be

$$I = 4 I_0 \cos^2(\phi/2) \tag{10.12}$$

If $\phi = 0, \pm 2\pi, \pm 4\pi, \ldots$ which corresponds to the condition given by Eq. (10.10) we will have constructive interference leading to maximum intensity. On the other hand, if $\phi = \pm\pi, \pm 3\pi, \pm 5\pi \ldots$ [which corresponds to the condition given by Eq. (10.11)] we will have destructive interference leading to zero intensity.

Now if the two sources are coherent (i.e., if the two needles are going up and down regularly) then the phase difference ϕ at any point will not change with time and we will have a stable interference pattern; i.e., the positions of maxima and minima will not change with time. However, if the two needles do not maintain a constant phase difference, then the interference pattern will also change with time and, if the phase difference changes very rapidly with time, the positions of maxima and minima will also vary rapidly with time and we will see a "time-averaged" intensity distribution. When this happens, we will observe an average intensity that will be given by

$$< I >= 4I_0 < \cos^2(\phi/2) > \tag{10.13}$$

where angular brackets represent time averaging. Indeed it is shown in Section 7.2 that if $\phi(t)$ varies randomly with time, the time-averaged quantity $< \cos^2(\phi/2) >$ will be 1/2. This is also intuitively obvious because the function $\cos^2(\phi/2)$ will randomly vary between 0 and 1 and the average value will be 1/2. The resultant intensity will be given by

$$I = 2 I_0 \tag{10.14}$$

at all points.

When the phase difference between the two vibrating sources changes rapidly with time, we say that the two sources are incoherent and when this happens the intensities just add up. This is indeed what happens when two separate light sources illuminate a wall.

Ripple Tank experiments on wave interference
http://www.colorado.edu/physics/2000/applets/fourier.html

10.5 Interference of Light Waves and Young's Experiment

We will now discuss interference using light waves. If we use two sodium lamps illuminating two pinholes (Fig. 10.11) we will not observe any interference fringes. This is because of the fact that the light wave emitted from an ordinary source (like a sodium lamp) undergoes abrupt phase

changes in times of the order of 10^{-10} seconds. Thus the light waves coming out from two independent sources of light will not have any fixed phase relationship and would be incoherent, when this happens, as discussed in the previous section, the intensities on the screen will add up.

FIGURE 10.11 If two sodium lamps illuminate two pinholes S_1 and S_2, the intensities will add up and no interference fringes will be observed on the screen.

The British physicist Thomas Young used an ingenious technique to "lock" the phases of the waves emanating from S_1 and S_2. He made two pinholes S_1 and S_2 (very close to each other) on an opaque screen [Fig. 10.12(a)]. These were illuminated by another pinholes that was in turn, lit by a bright source. Light waves spread out from S and fall on both S_1 and S_2. S_1 and S_2 then behave like two coherent sources because light waves coming out from S_1 and S_2 are derived from the same original source and any abrupt phase change in S will manifest in exactly similar phase changes in the light coming out from S_1 and S_2. Thus, the two sources S_1 and S_2 will be *locked* in phase; i.e., they will be coherent like the two vibrating needle in our water wave example [Fig. 10.8(a)].

FIGURE 10.12 Young's arrangement to produce interference pattern.

Thus spherical waves emanating from S_1 and S_2 will produce interference fringes on the screen GG′, as shown in Fig. 10.12(b). The positions of maximum and minimum intensities can be calculated by using the analysis given in Section 10.4 where we had shown that for an arbitrary point P on the line GG′ [Fig. 10.12(b)] to correspond to a maximum, we must have

$$S_2P - S_1P = n\lambda; \quad n = 0, 1, 2 \ldots \tag{10.15}$$

Now,

$$(S_2P)^2 - (S_1P)^2 = \left[D^2 + \left(x + \frac{d}{2}\right)^2\right] - \left[D^2 + \left(x - \frac{d}{2}\right)^2\right] = 2xd$$

THOMAS YOUNG (1773 – 1829)

Thomas Young (1773 – 1829) English physicist, physician and Egyptologist. Young worked on a wide variety of scientific problems, ranging from the structure of the eye and the mechanism of vision to the decipherment of the Rosetta stone. He revived the wave theory of light and recognised that interference phenomena provide proof of the wave properties of light.

where $S_1S_2 = d$ and $OP = x$. Thus

$$S_2P - S_1P = \frac{2xd}{S_2P+S_1P} \tag{10.16}$$

If x, $d<<D$ then negligible error will be introduced if $S_2P + S_1P$ (in the denominator) is replaced by $2D$. For example, for $d = 0.1$ cm, $D = 100$ cm, OP = 1 cm (which correspond to typical values for an interference experiment using light waves), we have

$$S_2P + S_1P = [(100)^2 + (1.05)^2]^{½} + [(100)^2 + (0.95)^2]^{½}$$
$$\approx 200.01 \text{ cm}$$

Thus if we replace $S_2P + S_1P$ by $2D$, the error involved is about 0.005%. In this approximation, Eq. (10.16) becomes

$$S_2P - S_1P \approx \tag{10.17}$$

Hence we will have constructive interference resulting in a bright region when

$$x = x_n = \frac{n\lambda D}{d};\ n = 0, \pm 1, \pm 2, \ldots \tag{10.18}$$

On the other hand, we will have a dark region near

$$x = x_n = (n+\quad)\frac{\lambda D}{d};\quad n = 0, \pm 1, \pm 2 \tag{10.19}$$

Thus dark and bright bands appear on the screen, as shown in Fig. 10.13. Such bands are called *fringes*. Equations (10.18) and (10.19) show that dark and bright fringes are equally spaced and the distance between two consecutive bright and dark fringes is given by

$$\beta = x_{n+1} - x_n$$

$$\text{or}\quad \beta = \frac{\lambda D}{d} \tag{10.20}$$

which is the expression for the *fringe width*. Obviously, the central point O (in Fig. 10.12) will be bright because $S_1O = S_2O$ and it will correspond to $n = 0$. If we consider the line perpendicular to the plane of the paper and passing through O [i.e., along the y-axis] then all points on this line will be equidistant from S_1 and S_2 and we will have a bright central fringe which is a straight line as shown in Fig. 10.13. In order to determine the shape of the interference pattern on the screen we note that a particular fringe would correspond to the locus of points with a constant value of $S_2P - S_1P$. Whenever this constant is an integral multiple of λ, the fringe will be bright and whenever it is an odd integral multiple of $\lambda/2$ it will be a dark fringe. Now, the locus of the point P lying in the x-y plane such that $S_2P - S_1P$ ($= \Delta$) is a constant, is a hyperbola. Thus the fringe pattern will strictly be a hyperbola; however, if the distance D is very large compared to the fringe width, the fringes will be very nearly straight lines as shown in Fig. 10.13.

FIGURE 10.13 Computer generated fringe pattern produced by two point source S_1 and S_2 on the screen GG′ (Fig. 10.12); (a) and (b) correspond to d = 0.005 mm and 0.025 mm, respectively (both figures correspond to D = 5 cm and $\lambda = 5 \times 10^{-5}$ cm.) (Adopted from OPTICS by A. Ghatak, Tata McGraw Hill Publishing Co. Ltd., New Delhi, 2000.)

In the double-slit experiment shown in Fig. 10.12, we have taken the source hole S on the perpendicular bisector of the two slits, which is shown as the line SO. What happens if the source S is slightly away from the perpendicular bisector. Consider that the source is moved to some new point S′ and suppose that Q is the mid-point of S_1 and S_2. If the angle S′QS is ϕ, then the central bright fringe occurs at an angle $-\phi$, on the other side. Thus, if the source S is on the perpendicular bisector, then the central fringe occurs at O, also on the perpendicular bisector. If S is shifted by an angle ϕ to point S′, then the central fringe appears at a point O' at an angle $-\phi$, which means that it is shifted by the same angle on the other side of the bisector. This also means that the source S′, the mid-point Q and the point O′ of the central fringe are in a straight line.

We end this section by quoting from the Nobel lecture of Dennis Gabor*

The wave nature of light was demonstrated convincingly for the first time in 1801 by Thomas Young by a wonderfully simple experiment. He let a ray of sunlight into a dark room, placed a dark screen in front of it, pierced with two small pinholes, and beyond this, at some distance, a white screen. He then saw two darkish lines at both sides of a bright line, which gave him sufficient encouragement to repeat the experiment, this time with spirit flame as light source, with a little salt in it to produce the bright yellow sodium light. This time he saw a number of dark lines, regularly spaced; the first clear proof that light added to light can produce darkness. This phenomenon is called

* Dennis Gabor received the 1971 Nobel Prize in Physics for discovering the principles of holography.

interference. Thomas Young had expected it because he believed in the wave theory of light.

We should mention here that the fringes are straight lines although S_1 and S_2 are point sources. If we had slits instead of the point sources (Fig. 10.14), each pair of points would have produced straight line fringes resulting in straight line fringes with increased intensities.

FIGURE 10.14 Photograph and the graph of the intensity distribution in Young's double-slit experiment.

Interactive animation of Young's experiment
http://vsg.quasihome.com/interfer.html

PHYSICS

EXAMPLE 10.3

Example 10.3 Two slits are made one millimetre apart and the screen is placed one metre away. What is the fringe separation when blue-green light of wavelength 500 nm is used?

Solution Fringe spacing $= \dfrac{D\lambda}{d} = \dfrac{1\times 5\times 10^{-7}}{1\times 10^{-3}}\,\text{m}$

$= 5 \times 10^{-4}$ m = 0.5 mm

EXAMPLE 10.4

Example 10.4 What is the effect on the interference fringes in a Young's double-slit experiment due to each of the following operations:

(a) the screen is moved away from the plane of the slits;
(b) the (monochromatic) source is replaced by another (monochromatic) source of shorter wavelength;
(c) the separation between the two slits is increased;
(d) the source slit is moved closer to the double-slit plane;
(e) the width of the source slit is increased;
(f) the monochromatic source is replaced by a source of white light?

(In each operation, take all parameters, other than the one specified, to remain unchanged.)

Solution

(a) Angular separation of the fringes remains constant ($= \lambda/d$). The actual separation of the fringes increases in proportion to the distance of the screen from the plane of the two slits.

(b) The separation of the fringes (and also angular separation) decreases. See, however, the condition mentioned in (d) below.

(c) The separation of the fringes (and also angular separation) decreases. See, however, the condition mentioned in (d) below.

(d) Let s be the size of the source and S its distance from the plane of the two slits. For interference fringes to be seen, the condition $s/S < \lambda/d$ should be satisfied; otherwise, interference patterns produced by different parts of the source overlap and no fringes are seen. Thus, as S decreases (i.e., the source slit is brought closer), the interference pattern gets less and less sharp, and when the source is brought too close for this condition to be valid, the fringes disappear. Till this happens, the fringe separation remains fixed.

(e) Same as in (d). As the source slit width increases, fringe pattern gets less and less sharp. When the source slit is so wide that the condition $s/S \leq \lambda/d$ is not satisfied, the interference pattern disappears.

(f) The interference patterns due to different component colours of white light overlap (incoherently). The central bright fringes for different colours are at the same position. Therefore, the central fringe is white. For a point P for which $S_2P - S_1P = \lambda_b/2$, where λ_b ($\approx$ 4000 Å) represents the wavelength for the blue colour, the blue component will be absent and the fringe will appear red in colour. Slightly farther away where $S_2Q - S_1Q = \lambda_b = \lambda_r/2$ where λ_r ($\approx$ 8000 Å) is the wavelength for the red colour, the fringe will be predominantly blue.

Thus, the fringe closest on either side of the central white fringe is red and the farthest will appear blue. After a few fringes, no clear fringe pattern is seen.

EXAMPLE 10.4

10.6 DIFFRACTION

If we look clearly at the shadow cast by an opaque object, close to the region of geometrical shadow, there are alternate dark and bright regions just like in interference. This happens due to the phenomenon of diffraction. Diffraction is a general characteristic exhibited by all types of waves, be it sound waves, light waves, water waves or matter waves. Since the wavelength of light is much smaller than the dimensions of most obstacles; we do not encounter diffraction effects of light in everyday observations. However, the finite resolution of our eye or of optical

instruments such as telescopes or microscopes is limited due to the phenomenon of diffraction. Indeed the colours that you see when a CD is viewed is due to diffraction effects. We will now discuss the phenomenon of diffraction.

10.6.1 The single slit

In the discussion of Young's experiment, we stated that a single narrow slit acts as a new source from which light spreads out. Even before Young, early experimenters – including Newton – had noticed that light spreads out from narrow holes and slits. It seems to turn around corners and enter regions where we would expect a shadow. These effects, known as *diffraction*, can only be properly understood using wave ideas. After all, you are hardly surprised to hear sound waves from someone talking around a corner!

When the double slit in Young's experiment is replaced by a single narrow slit (illuminated by a monochromatic source), a broad pattern with a central bright region is seen. On both sides, there are alternate dark and bright regions, the intensity becoming weaker away from the centre (Fig. 10.16). To understand this, go to Fig. 10.15, which shows a parallel beam of light falling normally on a single slit LN of width a. The diffracted light goes on to meet a screen. The midpoint of the slit is M.

A straight line through M perpendicular to the slit plane meets the screen at C. We want the intensity at any point P on the screen. As before, straight lines joining P to the different points L,M,N, etc., can be treated as parallel, making an angle θ with the normal MC.

The basic idea is to divide the slit into much smaller parts, and add their contributions at P with the proper phase differences. We are treating different parts of the wavefront at the slit as secondary sources. Because the incoming wavefront is parallel to the plane of the slit, these sources are in phase.

The path difference NP – LP between the two edges of the slit can be calculated exactly as for Young's experiment. From Fig. 10.15,

$$\begin{aligned} NP - LP &= NQ \\ &= a \sin \theta \\ &\approx a\theta \end{aligned} \tag{10.21}$$

Similarly, if two points M_1 and M_2 in the slit plane are separated by y, the path difference $M_2P - M_1P \approx y\theta$. We now have to sum up equal, coherent contributions from a large number of sources, each with a different phase. This calculation was made by Fresnel using integral calculus, so we omit it here. The main features of the diffraction pattern can be understood by simple arguments.

At the central point C on the screen, the angle θ is zero. All path differences are zero and hence all the parts of the slit contribute in phase. This gives maximum intensity at C. Experimental observation shown in

Fig. 10.15 indicates that the intensity has a central maximum at $\theta = 0$ and other secondary maxima at $\theta \approx (n+1/2)\,\lambda/a$, and has minima (zero intensity) at $\theta \approx n\lambda/a$, $n = \pm1, \pm2, \pm3, \ldots$. It is easy to see why it has minima at these values of angle. Consider first the angle θ where the path difference $a\theta$ is λ. Then,

$$\theta \approx \lambda / a\,. \qquad (10.22)$$

FIGURE 10.15 The geometry of path differences for diffraction by a single slit.

Now, divide the slit into two equal halves LM and MN each of size $a/2$. For every point M_1 in LM, there is a point M_2 in MN such that $M_1M_2 = a/2$. The path difference between M_1 and M_2 at P = $M_2P - M_1P$ = $\theta a/2 = \lambda/2$ for the angle chosen. This means that the contributions from M_1 and M_2 are 180° out of phase and cancel in the direction $\theta = \lambda/a$. Contributions from the two halves of the slit LM and MN, therefore, cancel each other. Equation (10.22) gives the angle at which the intensity falls to zero. One can similarly show that the intensity is zero for $\theta = n\lambda/a$, with n being any integer (except zero!). Notice that the angular size of the central maximum increases when the slit width a decreases.

It is also easy to see why there are maxima at $\theta \simeq (n + 1/2)\,\lambda/a$ and why they go on becoming weaker and weaker with increasing n. Consider an angle $\theta = 3\lambda/2a$ which is midway between two of the dark fringes. Divide the slit into three equal parts. If we take the first two thirds of the slit, the path difference between the two ends would be

$$\frac{2}{3}a \times \theta = \frac{2a}{3} \times \frac{3\lambda}{2a} = \lambda \qquad (10.23)$$

The first two-thirds of the slit can therefore be divided into two halves which have a $\lambda/2$ path difference. The contributions of these two halves cancel in the same manner as described earlier. Only the remaining one-third of the slit contributes to the intensity at a point between the two minima. Clearly, this will be much weaker than the central maximum (where the entire slit contributes in phase). One can similarly show that there are maxima at $(n + 1/2)\,\theta/a$ with $n = 2, 3$, etc. These become weaker with increasing n, since only one-fifth, one-seventh, etc., of the slit contributes in these cases. The photograph and intensity pattern corresponding to it is shown in Fig. 10.16.

FIGURE 10.16 Intensity distribution and photograph of fringes due to diffraction at single slit.

There has been prolonged discussion about difference between intereference and diffraction among scientists since the discovery of these phenomena. In this context, it is

interesting to note what Richard Feynman* has said in his famous Feynman Lectures on Physics:

> *No one has ever been able to define the difference between interference and diffraction satisfactorily. It is just a question of usage, and there is no specific, important physical difference between them. The best we can do is, roughly speaking, is to say that when there are only a few sources, say two interfering sources, then the result is usually called interference, but if there is a large number of them, it seems that the word diffraction is more often used.*

In the double-slit experiment, we must note that the pattern on the screen is actually a superposition of single-slit diffraction from each slit or hole, and the double-slit interference pattern. This is shown in Fig. 10.17. It shows a broader diffraction peak in which there appear several fringes of smaller width due to double-slit interference. The number of interference fringes occuring in the broad diffraction peak depends on the ratio d/a, that is the ratio of the distance between the two slits to the width of a slit. In the limit of a becoming very small, the diffraction pattern will become very flat and we will obsrve the two-slit interference pattern [see Fig. 10.13(b)].

Interactive animation on single slit diffraction pattern
http://www.phys.hawaii.edu/~teb/optics/java/slitdiffr/

FIGURE 10.17 The actual double-slit interference pattern. The envelope shows the single slit diffraction.

EXAMPLE 10.5

Example 10.5 In Example 10.3, what should the width of each slit be to obtain 10 maxima of the double slit pattern within the central maximum of the single slit pattern?

Solution We want $a\theta = \lambda, \theta = \dfrac{\lambda}{a}$

$$10\frac{\lambda}{d} = 2\frac{\lambda}{a} \quad a = \frac{d}{5} = 0.2 \text{ mm}$$

Notice that the wavelength of light and distance of the screen do not enter in the calculation of a.

In the double-slit interference experiment of Fig. 10.12, what happens if we close one slit? You will see that it now amounts to a single slit. But you will have to take care of some shift in the pattern. We now have a source at S, and only one hole (or slit) S_1 or S_2. This will produce a single-

* Richand Feynman was one of the recipients of the 1965 Nobel Prize in Physics for his fundamental work in quantum electrodynamics.

slit diffraction pattern on the screen. The centre of the central bright fringe will appear at a point which lies on the straight line SS_1 or SS_2, as the case may be.

We now compare and contrast the interference pattern with that seen for a coherently illuminated single slit (usually called the single slit diffraction pattern).

(i) The interference pattern has a number of equally spaced bright and dark bands. The diffraction pattern has a central bright maximum which is twice as wide as the other maxima. The intensity falls as we go to successive maxima away from the centre, on either side.

(ii) We calculate the interference pattern by superposing two waves originating from the two narrow slits. The diffraction pattern is a superposition of a continuous family of waves originating from each point on a single slit.

(iii) For a single slit of width a, the first null of the interference pattern occurs at an angle of λ/a. At the same angle of λ/a, we get a maximum (not a null) for two narrow slits separated by a distance a.

One must understand that both d and a have to be quite small, to be able to observe good interference and diffraction patterns. For example, the separation d between the two slits must be of the order of a milimetre or so. The width a of each slit must be even smaller, of the order of 0.1 or 0.2 mm.

In our discussion of Young's experiment and the single-slit diffraction, we have assumed that the screen on which the fringes are formed is at a large distance. The two or more paths from the slits to the screen were treated as parallel. This situation also occurs when we place a converging lens after the slits and place the screen at the focus. Parallel paths from the slit are combined at a single point on the screen. *Note that the lens does not introduce any extra path differences in a parallel beam.* This arrangement is often used since it gives more intensity than placing the screen far away. If f is the focal length of the lens, then we can easily work out the size of the central bright maximum. In terms of angles, the separation of the central maximum from the first null of the diffraction pattern is λ/a. Hence, the size on the screen will be $f\lambda/a$.

10.6.2 Seeing the single slit diffraction pattern

It is surprisingly easy to see the single-slit diffraction pattern for oneself. The equipment needed can be found in most homes — two razor blades and one clear glass electric bulb preferably with a straight filament. One has to hold the two blades so that the edges are parallel and have a narrow slit in between. This is easily done with the thumb and forefingers (Fig. 10.18).

FIGURE 10.18 Holding two blades to form a single slit. A bulb filament viewed through this shows clear diffraction bands.

Keep the slit parallel to the filament, right in front of the eye. Use spectacles if you normally do. With slight adjustment of the width of the slit and the parallelism of the edges, the pattern should be seen with its bright and dark bands. Since the position of all the bands (except the central one) depends on wavelength, they will show some colours. Using a filter for red or blue will make the fringes clearer. With both filters available, the wider fringes for red compared to blue can be seen.

In this experiment, the filament plays the role of the first slit S in Fig. 10.16. The lens of the eye focuses the pattern on the screen (the retina of the eye).

With some effort, one can cut a double slit in an aluminium foil with a blade. The bulb filament can be viewed as before to repeat Young's experiment. In daytime, there is another suitable bright source subtending a small angle at the eye. This is the reflection of the Sun in any shiny convex surface (e.g., a cycle bell). Do not try direct sunlight – it can damage the eye and will not give fringes anyway as the Sun subtends an angle of (1/2)°.

In interference and diffraction, light energy is redistributed. If it reduces in one region, producing a dark fringe, it increases in another region, producing a bright fringe. There is no gain or loss of energy, which is consistent with the principle of conservation of energy.

10.6.3 Resolving power of optical instruments

In Chapter 9 we had discussed about telescopes. The angular resolution of the telescope is determined by the objective of the telescope. The stars which are not resolved in the image produced by the objective cannot be resolved by any further magnification produced by the eyepiece. The primary purpose of the eyepiece is to provide magnification of the image produced by the objective.

Consider a parallel beam of light falling on a convex lens. If the lens is well corrected for aberrations, then geometrical optics tells us that the beam will get focused to a point. However, because of diffraction, the beam instead of getting focused to a point gets focused to a spot of finite area. In this case the effects due to diffraction can be taken into account by considering a plane wave incident on a circular aperture followed by a convex lens (Fig. 10.19). The analysis of the corresponding diffraction pattern is quite involved; however, in principle, it is similar to the analysis carried out to obtain the single-slit diffraction pattern. Taking into account the effects due to diffraction, the pattern on the focal plane would consist of a central bright region surrounded by concentric dark and bright rings (Fig. 10.19). A detailed analysis shows that the radius of the central bright region is approximately given by

$$r_0 \approx \frac{1.22\,\lambda f}{2a} = \frac{0.61\lambda f}{a} \tag{10.24}$$

FIGURE 10.19 A parallel beam of light is incident on a convex lens. Because of diffraction effects, the beam gets focused to a spot of radius ≈ 0.61 $\lambda f/a$.

where f is the focal length of the lens and $2a$ is the diameter of the circular aperture or the diameter of the lens, whichever is smaller. Typically if

$$\lambda \approx 0.5\ \mu\text{m},\ f \approx 20\ \text{cm}\ \text{ and }\ a \approx 5\ \text{cm}$$

we have

$$r_0 \approx 1.2\ \mu\text{m}$$

Although the size of the spot is very small, it plays an important role in determining the limit of resolution of optical instruments like a telescope or a microscope. For the two stars to be just resolved

$$f\Delta\theta \approx r_0 \approx \frac{0.61\lambda f}{a}$$

implying

$$\Delta\theta \approx \frac{0.61\lambda}{a} \tag{10.25}$$

Thus $\Delta\theta$ will be small if the diameter of the objective is large. This implies that the telescope will have better resolving power if a is large. It is for this reason that for better resolution, a telescope must have a large diameter objective.

Example 10.6 Assume that light of wavelength 6000Å is coming from a star. What is the limit of resolution of a telescope whose objective has a diameter of 100 inch?

Solution A 100 inch telescope implies that $2a$ = 100 inch = 254 cm. Thus if,

$$\lambda \approx 6000\text{Å} = 6\times10^{-5}\ \text{cm}$$

then

$$\Delta\theta \approx \frac{0.61\times6\times10^{-5}}{127} \approx 2.9\times10^{-7}\ \text{radians}$$

EXAMPLE 10.6

We can apply a similar argument to the objective lens of a microscope. In this case, the object is placed slightly beyond f, so that a real image is formed at a distance v [Fig. 10.20]. The magnification – ratio of image size to object size – is given by $m \approx v/f$. It can be seen from Fig. 10.20 that

$$D/f \approx 2\tan\beta \tag{10.26}$$

where 2β is the angle subtended by the diameter of the objective lens at the focus of the microscope.

FIGURE 10.20 Real image formed by the objective lens of the microscope.

Determine the resolving power of your eye

You can estimate the resolving power of your eye with a simple experiment. Make black stripes of equal width separated by white stripes; see figure here. All the black stripes should be of equal width, while the width of the intermediate white stripes should increase as you go from the left to the right. For example, let all black stripes have a width of 5 mm. Let the width of the first two white stripes be 0.5 mm each, the next two white stripes be 1 mm each, the next two 1.5 mm each, etc. Paste this pattern on a wall in a room or laboratory, at the height of your eye.

Now watch the pattern, preferably with one eye. By moving away or closer to the wall, find the position where you can just see some two black stripes as separate stripes. All the black stripes to the left of this stripe would merge into one another and would not be distinguishable. On the other hand, the black stripes to the right of this would be more and more clearly visible. Note the width d of the white stripe which separates the two regions, and measure the distance D of the wall from your eye. Then d/D is the resolution of your eye.

You have watched specks of dust floating in air in a sunbeam entering through your window. Find the distance (of a speck) which you can clearly see and distinguish from a neighbouring speck. Knowing the resolution of your eye and the distance of the speck, estimate the size of the speck of dust.

When the separation between two points in a microscopic specimen is comparable to the wavelength λ of the light, the diffraction effects become important. The image of a point object will again be a diffraction pattern whose size in the image plane will be

$$v\theta = v\left(\frac{1.22\,\lambda}{D}\right) \tag{10.27}$$

Two objects whose images are closer than this distance will not be resolved, they will be seen as one. The corresponding minimum separation, d_{min}, in the object plane is given by

$$d_{min} = \left[v\left(\frac{1.22\,\lambda}{D}\right)\right]\Big/m$$

$$= \frac{1.22\,\lambda}{D}\cdot\frac{v}{m}$$

$$= \frac{1.22\,f\,\lambda}{D} \tag{10.28}$$

Now, combining Eqs. (10.26) and (10.28), we get

$$d_{min} = \frac{1.22\,\lambda}{2\tan\beta}$$

$$\simeq \frac{1.22\,\lambda}{2\sin\beta} \tag{10.29}$$

If the medium between the object and the objective lens is not air but a medium of refractive index n, Eq. (10.29) gets modified to

$$d_{\min} = \frac{1.22\,\lambda}{2\,n\sin\beta} \tag{10.30}$$

The product $n\sin\beta$ is called the *numerical aperture* and is sometimes marked on the objective.

The resolving power of the microscope is given by the reciprocal of the minimum separation of two points seen as distinct. It can be seen from Eq. (10.30) that the resolving power can be increased by choosing a medium of higher refractive index. Usually an oil having a refractive index close to that of the objective glass is used. Such an arrangement is called an *'oil immersion objective'*. Notice that it is not possible to make $\sin\beta$ larger than unity. Thus, we see that the resolving power of a microscope is basically determined by the wavelength of the light used.

There is a likelihood of confusion between resolution and magnification, and similarly between the role of a telescope and a microscope to deal with these parameters. A telescope produces images of far objects nearer to our eye. Therefore objects which are not resolved at far distance, can be resolved by looking at them through a telescope. A microscope, on the other hand, magnifies objects (which are near to us) and produces their larger image. We may be looking at two stars or two satellites of a far-away planet, or we may be looking at different regions of a living cell. In this context, it is good to remember that a telescope resolves whereas a microscope magnifies.

10.6.4 The validity of ray optics

An aperture (i.e., slit or hole) of size a illuminated by a parallel beam sends diffracted light into an angle of approximately $\approx \lambda/a$. This is the angular size of the bright central maximum. In travelling a distance z, the diffracted beam therefore acquires a width $z\lambda/a$ due to diffraction. It is interesting to ask at what value of z the spreading due to diffraction becomes comparable to the size a of the aperture. We thus approximately equate $z\lambda/a$ with a. This gives the distance beyond which divergence of the beam of width a becomes significant. Therefore,

$$z \simeq \frac{a^2}{\lambda} \tag{10.31}$$

We define a quantity z_F called the *Fresnel distance* by the following equation

$$z_F \simeq a^2/\lambda$$

Equation (10.31) shows that for distances much smaller than z_F, the spreading due to diffraction is smaller compared to the size of the beam. It becomes comparable when the distance is approximately z_F. For distances much greater than z_F, the spreading due to diffraction

dominates over that due to ray optics (i.e., the size a of the aperture). Equation (10.31) also shows that ray optics is valid in the limit of wavelength tending to zero.

EXAMPLE 10.7

Example 10.7 For what distance is ray optics a good approximation when the aperture is 3 mm wide and the wavelength is 500 nm?

Solution $$z_F = \frac{a^2}{\lambda} = \frac{\left(3\times10^{-3}\right)^2}{5\times10^{-7}} = 18 \text{ m}$$

This example shows that even with a small aperture, diffraction spreading can be neglected for rays many metres in length. Thus, ray optics is valid in many common situations.

10.7 POLARISATION

Consider holding a long string that is held horizontally, the other end of which is assumed to be fixed. If we move the end of the string up and down in a periodic manner, we will generate a wave propagating in the $+x$ direction (Fig. 10.22). Such a wave could be described by the following equation

FIGURE 10.21 (a) The curves represent the displacement of a string at $t = 0$ and at $t = \Delta t$, respectively when a sinusoidal wave is propagating in the $+x$-direction. (b) The curve represents the time variation of the displacement at $x = 0$ when a sinusoidal wave is propagating in the $+x$-direction. At $x = \Delta x$, the time variation of the displacement will be slightly displaced to the right.

$$y\,(x,t) = a \sin\,(kx - \omega t) \tag{10.32}$$

where a and $\omega (= 2\pi\nu)$ represent the amplitude and the angular frequency of the wave, respectively; further,

$$\lambda = \frac{2\pi}{k} \tag{10.33}$$

represents the wavelength associated with the wave. We had discussed propagation of such waves in Chapter 15 of Class XI textbook. Since the displacement (which is along the y direction) is at right angles to the direction of propagation of the wave, we have what is known as a *transverse wave*. Also, since the displacement is in the y direction, it is often referred to as a y-polarised wave. Since each point on the string moves on a straight line, the wave is also referred to as a linearly polarised wave. Further, the string always remains confined to the x-y plane and therefore it is also referred to as a *plane polarised wave*.

In a similar manner we can consider the vibration of the string in the x-z plane generating a z-polarised wave whose displacement will be given by

$$z\,(x,t) = a \sin\,(kx - \omega t) \tag{10.34}$$

It should be mentioned that the linearly polarised waves [described by Eqs. (10.33) and (10.34)] are all transverse waves; i.e., the displacement of each point of the string is always at right angles to the direction of propagation of the wave. Finally, if the plane of vibration of the string is changed randomly in very short intervals of time, then we have what is known as an *unpolarised wave*. Thus, for an unpolarised wave the displacement will be randomly changing with time though it will always be perpendicular to the direction of propagation.

Light waves are transverse in nature; i.e., the electric field associated with a propagating light wave is always at right angles to the direction of propagation of the wave. This can be easily demonstrated using a simple polaroid. You must have seen thin plastic like sheets, which are called *polaroids*. A polaroid consists of long chain molecules aligned in a particular direction. The electric vectors (associated with the propagating light wave) along the direction of the aligned molecules get absorbed. Thus, if an unpolarised light wave is incident on such a polaroid then the light wave will get linearly polarised with the electric vector oscillating along a direction perpendicular to the aligned molecules; this direction is known as the *pass-axis* of the polaroid.

Thus, if the light from an ordinary source (like a sodium lamp) passes through a polaroid sheet P_1, it is observed that its intensity is reduced by half. Rotating P_1 has no effect on the transmitted beam and transmitted intensity remains constant. Now, let an identical piece of polaroid P_2 be placed before P_1. As expected, the light from the lamp is reduced in intensity on passing through P_2 alone. But now rotating P_1 has a dramatic effect on the light coming from P_2. In one position, the intensity transmitted

by P_2 followed by P_1 is nearly zero. When turned by 90° from this position, P_1 transmits nearly the full intensity emerging from P_2 (Fig. 10.22).

FIGURE 10.22 (a) Passage of light through two polaroids P_2 and P_1. The transmitted fraction falls from 1 to 0 as the angle between them varies from 0° to 90°. Notice that the light seen through a single polaroid P_1 does not vary with angle. (b) Behaviour of the electric vector when light passes through two polaroids. The transmitted polarisation is the component parallel to the polaroid axis. The double arrows show the oscillations of the electric vector.

The above experiment can be easily understood by assuming that light passing through the polaroid P_2 gets polarised along the pass-axis of P_2. If the pass-axis of P_2 makes an angle θ with the pass-axis of P_1, then when the polarised beam passes through the polaroid P_2, the component $E \cos \theta$ (along the pass-axis of P_2) will pass through P_2. Thus, as we rotate the polaroid P_1 (or P_2), the intensity will vary as:

$$I = I_0 \cos^2\theta \qquad (10.35)$$

where I_0 is the intensity of the polarized light after passing through P_1. This is known as *Malus' law*. The above discussion shows that the intensity coming out of a single polaroid is half of the incident intensity. By putting a second polaroid, the intensity can be further controlled from 50% to zero of the incident intensity by adjusting the angle between the pass-axes of two polaroids.

Polaroids can be used to control the intensity, in sunglasses, windowpanes, etc. Polaroids are also used in photographic cameras and 3D movie cameras.

EXAMPLE 10.8

Example 10.8 Discuss the intensity of transmitted light when a polaroid sheet is rotated between two crossed polaroids?

Solution Let I_0 be the intensity of polarised light after passing through the first polariser P_1. Then the intensity of light after passing through second polariser P_2 will be

$$I = I_0 \cos^2\theta,$$

EXAMPLE 10.8

where θ is the angle between pass axes of P_1 and P_2. Since P_1 and P_3 are crossed the angle between the pass axes of P_2 and P_3 will be $(\pi/2-\theta)$. Hence the intensity of light emerging from P_3 will be

$$I = I_0 \cos^2\theta \cos^2\left(\frac{\pi}{2}-\theta\right)$$

$$= I_0 \cos^2\theta \ \sin^2\theta = (I_0/4) \sin^2 2\theta$$

Therefore, the transmitted intensity will be maximum when $\theta = \pi/4$.

10.7.1 Polarisation by scattering

The light from a clear blue portion of the sky shows a rise and fall of intensity when viewed through a polaroid which is rotated. This is nothing but sunlight, which has changed its direction (having been scattered) on encountering the molecules of the earth's atmosphere. As Fig. 10.24(a) shows, the incident sunlight is unpolarised. The dots stand for polarisation perpendicular to the plane of the figure. The double arrows show polarisation in the plane of the figure. (There is no phase relation between these two in unpolarised light). Under the influence of the electric field of the incident wave the electrons in the molecules acquire components of motion in both these directions. We have drawn an observer looking at 90° to the direction of the sun. Clearly, charges accelerating parallel to the double arrows do not radiate energy towards this observer since their acceleration has no transverse component. The radiation scattered by the molecule is therefore represented by dots. It is polarised perpendicular to the plane of the figure. This explains the polarisation of scattered light from the sky.

FIGURE 10.23 (a) Polarisation of the blue scattered light from the sky. The incident sunlight is unpolarised (dots and arrows). A typical molecule is shown. It scatters light by 90° polarised normal to the plane of the paper (dots only). (b) Polarisation of light reflected from a transparent medium at the Brewster angle (reflected ray perpendicular to refracted ray).

The scattering of light by molecules was intensively investigated by C.V. Raman and his collaborators in Kolkata in the 1920s. Raman was awarded the Nobel Prize for Physics in 1930 for this work.

A SPECIAL CASE OF TOTAL TRANSMISSION

When light is incident on an interface of two media, it is observed that some part of it gets reflected and some part gets transmitted. Consider a related question: *Is it possible that under some conditions a monochromatic beam of light incident on a surface (which is normally reflective) gets completely transmitted with no reflection?* To your surprise, the answer is *yes*.

Let us try a simple experiment and check what happens. Arrange a laser, a good polariser, a prism and screen as shown in the figure here.

Let the light emitted by the laser source pass through the polariser and be incident on the surface of the prism at the Brewster's angle of incidence i_B. Now rotate the polariser carefully and you will observe that for a specific alignment of the polariser, the light incident on the prism is completely transmitted and no light is reflected from the surface of the prism. The reflected spot will completely vanish.

10.7.2 Polarisation by reflection

Figure 10.23(b) shows light reflected from a transparent medium, say, water. As before, the dots and arrows indicate that both polarisations are present in the incident and refracted waves. We have drawn a situation in which the reflected wave travels at right angles to the refracted wave. The oscillating electrons in the water produce the reflected wave. These move in the two directions transverse to the radiation from wave in the medium, i.e., the *refracted wave*. The arrows are parallel to the direction of the *reflected* wave. Motion in this direction does not contribute to the reflected wave. As the figure shows, the reflected light is therefore linearly polarised perpendicular to the plane of the figure (represented by dots). This can be checked by looking at the reflected light through an analyser. The transmitted intensity will be zero when the axis of the analyser is in the plane of the figure, i.e., the plane of incidence.

When unpolarised light is incident on the boundary between two transparent media, the reflected light is polarised with its electric vector perpendicular to the plane of incidence when the refracted and reflected rays make a right angle with each other. Thus we have seen that when reflected wave is perpendicular to the refracted wave, the reflected wave is a totally polarised wave. The angle of incidence in this case is called *Brewster's angle* and is denoted by i_B. We can see that i_B is related to the refractive index of the denser medium. Since we have $i_B + r = \pi/2$, we get from Snell's law

$$\mu = \frac{\sin i_B}{\sin r} = \frac{\sin i_B}{\sin(\pi/2 - i_B)}$$

$$= \frac{\sin i_B}{\cos i_B} = \tan i_B \qquad (10.36)$$

This is known as *Brewster's law*.

Example 10.9 Unpolarised light is incident on a plane glass surface. What should be the angle of incidence so that the reflected and refracted rays are perpendicular to each other?

Solution For $i + r$ to be equal to $\pi/2$, we should have $\tan i_B = \mu = 1.5$. This gives $i_B = 57°$. This is the Brewster's angle for air to glass interface.

EXAMPLE 10.9

For simplicity, we have discussed scattering of light by 90°, and reflection at the Brewster angle. In this special situation, one of the two perpendicular components of the electric field is zero. At other angles, both components are present but one is stronger than the other. There is no stable phase relationship between the two perpendicular components since these are derived from two perpendicular components of an unpolarised beam. When such light is viewed through a rotating analyser, one sees a maximum and a minimum of intensity but not complete darkness. This kind of light is called *partially polarised*.

Let us try to understand the situation. When an unpolarised beam of light is incident at the Brewster's angle on an interface of two media, only part of light with electric field vector perpendicular to the plane of incidence will be reflected. Now by using a good polariser, if we completely remove all the light with its electric vector perpendicular to the plane of incidence and let this light be incident on the surface of the prism at Brewster's angle, you will then observe no reflection and there will be total transmission of light.

We began this chapter by pointing out that there are some phenomena which can be explained only by the wave theory. In order to develop a proper understanding, we first described how some phenomena like reflection and refraction, which were studied on this basis of Ray Optics in Chapter 9, can also be understood on the basis of Wave Optics. Then we described Young's double slit experiment which was a turning point in the study of optics. Finally, we described some associated points such as diffraction, resolution, polarisation, and validity of ray optics. In the next chapter, you will see how new experiments led to new theories at the turn of the century around 1900 A.D.

SUMMARY

1. Huygens' principle tells us that each point on a wavefront is a source of secondary waves, which add up to give the wavefront at a later time.
2. Huygens' construction tells us that the new wavefront is the forward envelope of the secondary waves. When the speed of light is independent of direction, the secondary waves are spherical. The rays are then perpendicular to both the wavefronts and the time of travel

is the same measured along any ray. This principle leads to the well known laws of reflection and refraction.

3. The principle of superposition of waves applies whenever two or more sources of light illuminate the same point. When we consider the intensity of light due to these sources at the given point, there is an interference term in addition to the sum of the individual intensities. But this term is important only if it has a non-zero average, which occurs only if the sources have the same frequency and a stable phase difference.
4. Young's double slit of separation d gives equally spaced fringes of angular separation λ/d. The source, mid-point of the slits, and central bright fringe lie in a straight line. An extended source will destroy the fringes if it subtends angle more than λ/d at the slits.
5. A single slit of width a gives a diffraction pattern with a central maximum. The intensity falls to zero at angles of $\pm\frac{\lambda}{a}, \pm\frac{2\lambda}{a}$, etc., with successively weaker secondary maxima in between. Diffraction limits the angular resolution of a telescope to λ/D where D is the diameter. Two stars closer than this give strongly overlapping images. Similarly, a microscope objective subtending angle 2β at the focus, in a medium of refractive index n, will just separate two objects spaced at a distance $\lambda/(2n \sin \beta)$, which is the resolution limit of a microscope. Diffraction determines the limitations of the concept of light rays. A beam of width a travels a distance a^2/λ, called the Fresnel distance, before it starts to spread out due to diffraction.
6. Natural light, e.g., from the sun is unpolarised. This means the electric vector takes all possible directions in the transverse plane, rapidly and randomly, during a measurement. A polaroid transmits only one component (parallel to a special axis). The resulting light is called linearly polarised or plane polarised. When this kind of light is viewed through a second polaroid whose axis turns through 2π, two maxima and minima of intensity are seen. Polarised light can also be produced by reflection at a special angle (called the Brewster angle) and by scattering through $\pi/2$ in the earth's atmosphere.

POINTS TO PONDER

1. Waves from a point source spread out in all directions, while light was seen to travel along narrow rays. It required the insight and experiment of Huygens, Young and Fresnel to understand how a wave theory could explain all aspects of the behaviour of light.
2. The crucial new feature of waves is interference of amplitudes from different sources which can be both constructive and destructive, as shown in Young's experiment.
3. Even a wave falling on single slit should be regarded as a large number of sources which interefere constructively in the forward direction ($\theta = 0$), and destructively in other directions.
4. Diffraction phenomena define the limits of ray optics. The limit of the ability of microscopes and telescopes to distinguish very close objects is set by the wavelength of light.
5. Most interference and diffraction effects exist even for longitudinal waves like sound in air. But polarisation phenomena are special to transverse waves like light waves.

EXERCISES

10.1 Monochromatic light of wavelength 589 nm is incident from air on a water surface. What are the wavelength, frequency and speed of (a) reflected, and (b) refracted light? Refractive index of water is 1.33.

10.2 What is the shape of the wavefront in each of the following cases:

(a) Light diverging from a point source.

(b) Light emerging out of a convex lens when a point source is placed at its focus.

(c) The portion of the wavefront of light from a distant star intercepted by the Earth.

10.3 (a) The refractive index of glass is 1.5. What is the speed of light in glass? (Speed of light in vacuum is 3.0×10^8 m s^{-1})

(b) Is the speed of light in glass independent of the colour of light? If not, which of the two colours red and violet travels slower in a glass prism?

10.4 In a Young's double-slit experiment, the slits are separated by 0.28 mm and the screen is placed 1.4 m away. The distance between the central bright fringe and the fourth bright fringe is measured to be 1.2 cm. Determine the wavelength of light used in the experiment.

10.5 In Young's double-slit experiment using monochromatic light of wavelength λ, the intensity of light at a point on the screen where path difference is λ, is K units. What is the intensity of light at a point where path difference is $\lambda/3$?

10.6 A beam of light consisting of two wavelengths, 650 nm and 520 nm, is used to obtain interference fringes in a Young's double-slit experiment.

(a) Find the distance of the third bright fringe on the screen from the central maximum for wavelength 650 nm.

(b) What is the least distance from the central maximum where the bright fringes due to both the wavelengths coincide?

10.7 In a double-slit experiment the angular width of a fringe is found to be 0.2° on a screen placed 1 m away. The wavelength of light used is 600 nm. What will be the angular width of the fringe if the entire experimental apparatus is immersed in water? Take refractive index of water to be 4/3.

10.8 What is the Brewster angle for air to glass transition? (Refractive index of glass = 1.5.)

10.9 Light of wavelength 5000 Å falls on a plane reflecting surface. What are the wavelength and frequency of the reflected light? For what angle of incidence is the reflected ray normal to the incident ray?

10.10 Estimate the distance for which ray optics is good approximation for an aperture of 4 mm and wavelength 400 nm.

ADDITIONAL EXERCISES

10.11 The 6563 Å Hα line emitted by hydrogen in a star is found to be red-shifted by 15 Å. Estimate the speed with which the star is receding from the Earth.

10.12 Explain how Corpuscular theory predicts the speed of light in a medium, say, water, to be greater than the speed of light in vacuum. Is the prediction confirmed by experimental determination of the speed of light in water? If not, which alternative picture of light is consistent with experiment?

10.13 You have learnt in the text how Huygens' principle leads to the laws of reflection and refraction. Use the same principle to deduce directly that a point object placed in front of a plane mirror produces a virtual image whose distance from the mirror is equal to the object distance from the mirror.

10.14 Let us list some of the factors, which could possibly influence the speed of wave propagation:

(i) nature of the source.

(ii) direction of propagation.

(iii) motion of the source and/or observer.

(iv) wavelength.

(v) intensity of the wave.

On which of these factors, if any, does

(a) the speed of light in vacuum,

(b) the speed of light in a medium (say, glass or water),

depend?

10.15 For sound waves, the Doppler formula for frequency shift differs slightly between the two situations: (i) source at rest; observer moving, and (ii) source moving; observer at rest. The exact Doppler formulas for the case of light waves in vacuum are, however, strictly identical for these situations. Explain why this should be so. Would you expect the formulas to be strictly identical for the two situations in case of light travelling in a medium?

10.16 In double-slit experiment using light of wavelength 600 nm, the angular width of a fringe formed on a distant screen is 0.1°. What is the spacing between the two slits?

10.17 Answer the following questions:

(a) In a single slit diffraction experiment, the width of the slit is made double the original width. How does this affect the size and intensity of the central diffraction band?

(b) In what way is diffraction from each slit related to the interference pattern in a double-slit experiment?

(c) When a tiny circular obstacle is placed in the path of light from a distant source, a bright spot is seen at the centre of the shadow of the obstacle. Explain why?

(d) Two students are separated by a 7 m partition wall in a room 10 m high. If both light and sound waves can bend around

obstacles, how is it that the students are unable to see each other even though they can converse easily.

(e) Ray optics is based on the assumption that light travels in a straight line. Diffraction effects (observed when light propagates through small apertures/slits or around small obstacles) disprove this assumption. Yet the ray optics assumption is so commonly used in understanding location and several other properties of images in optical instruments. What is the justification?

10.18 Two towers on top of two hills are 40 km apart. The line joining them passes 50 m above a hill halfway between the towers. What is the longest wavelength of radio waves, which can be sent between the towers without appreciable diffraction effects?

10.19 A parallel beam of light of wavelength 500 nm falls on a narrow slit and the resulting diffraction pattern is observed on a screen 1 m away. It is observed that the first minimum is at a distance of 2.5 mm from the centre of the screen. Find the width of the slit.

10.20 Answer the following questions:

(a) When a low flying aircraft passes overhead, we sometimes notice a slight shaking of the picture on our TV screen. Suggest a possible explanation.

(b) As you have learnt in the text, the principle of linear superposition of wave displacement is basic to understanding intensity distributions in diffraction and interference patterns. What is the justification of this principle?

10.21 In deriving the single slit diffraction pattern, it was stated that the intensity is zero at angles of $n\lambda/a$. Justify this by suitably dividing the slit to bring out the cancellation.

Chapter Eleven

DUAL NATURE OF RADIATION AND MATTER

11.1 Introduction

The Maxwell's equations of electromagnetism and Hertz experiments on the generation and detection of electromagnetic waves in 1887 strongly established the wave nature of light. Towards the same period at the end of 19th century, experimental investigations on conduction of electricity (electric discharge) through gases at low pressure in a discharge tube led to many historic discoveries. The discovery of X-rays by Roentgen in 1895, and of electron by J. J. Thomson in 1897, were important milestones in the understanding of atomic structure. It was found that at sufficiently low pressure of about 0.001 mm of mercury column, a discharge took place between the two electrodes on applying the electric field to the gas in the discharge tube. A fluorescent glow appeared on the glass opposite to cathode. The colour of glow of the glass depended on the type of glass, it being yellowish-green for soda glass. The cause of this fluorescence was attributed to the radiation which appeared to be coming from the cathode. These *cathode rays* were discovered, in 1870, by William Crookes who later, in 1879, suggested that these rays consisted of streams of fast moving negatively charged particles. The British physicist J. J. Thomson (1856-1940) confirmed this hypothesis. By applying mutually perpendicular electric and magnetic fields across the discharge tube, J. J. Thomson was the first to determine experimentally the speed and the specific charge [charge to mass ratio (e/m)] of the cathode ray

particles. They were found to travel with speeds ranging from about 0.1 to 0.2 times the speed of light (3×10^8 m/s). The presently accepted value of e/m is 1.76×10^{11} C/kg. Further, the value of e/m was found to be independent of the nature of the material/metal used as the cathode (emitter), or the gas introduced in the discharge tube. This observation suggested the universality of the cathode ray particles.

Around the same time, in 1887, it was found that certain metals, when irradiated by ultraviolet light, emitted negatively charged particles having small speeds. Also, certain metals when heated to a high temperature were found to emit negatively charged particles. The value of e/m of these particles was found to be the same as that for cathode ray particles. These observations thus established that all these particles, although produced under different conditions, were identical in nature. J. J. Thomson, in 1897, named these particles as *electrons*, and suggested that they were fundamental, universal constituents of matter. For his epoch-making discovery of electron, through his theoretical and experimental investigations on conduction of electricity by gasses, he was awarded the Nobel Prize in Physics in 1906. In 1913, the American physicist R. A. Millikan (1868-1953) performed the pioneering oil-drop experiment for the precise measurement of the charge on an electron. He found that the charge on an oil-droplet was always an integral multiple of an elementary charge, 1.602×10^{-19} C. Millikan's experiment established that *electric charge is quantised*. From the values of charge (e) and specific charge (e/m), the mass (m) of the electron could be determined.

11.2 Electron Emission

We know that metals have free electrons (negatively charged particles) that are responsible for their conductivity. However, the free electrons cannot normally escape out of the metal surface. If an electron attempts to come out of the metal, the metal surface acquires a positive charge and pulls the electron back to the metal. The free electron is thus held inside the metal surface by the attractive forces of the ions. Consequently, the electron can come out of the metal surface only if it has got sufficient energy to overcome the attractive pull. A certain minimum amount of energy is required to be given to an electron to pull it out from the surface of the metal. This minimum energy required by an electron to escape from the metal surface is called the *work function* of the metal. It is generally denoted by ϕ_0 and measured in eV (electron volt). One electron volt is the energy gained by an electron when it has been accelerated by a potential difference of 1 volt, so that $1 \text{ eV} = 1.602 \times 10^{-19}$ J.

This unit of energy is commonly used in atomic and nuclear physics. The work function (ϕ_0) depends on the properties of the metal and the nature of its surface. The values of work function of some metals are given in Table 11.1. These values are approximate as they are very sensitive to surface impurities.

Note from Table 11.1 that the work function of platinum is the highest ($\phi_0 = 5.65$ eV) while it is the lowest ($\phi_0 = 2.14$ eV) for caesium.

The minimum energy required for the electron emission from the metal surface can be supplied to the free electrons by any one of the following physical processes:

TABLE 11.1 WORK FUNCTIONS OF SOME METALS

Metal	Work function ϕ_0 (eV)	Metal	Work function ϕ_0 (eV)
Cs	2.14	Al	4.28
K	2.30	Hg	4.49
Na	2.75	Cu	4.65
Ca	3.20	Ag	4.70
Mo	4.17	Ni	5.15
Pb	4.25	Pt	5.65

(i) *Thermionic emission:* By suitably heating, sufficient thermal energy can be imparted to the free electrons to enable them to come out of the metal.

(ii) *Field emission*: By applying a very strong electric field (of the order of 10^8 V m^{-1}) to a metal, electrons can be pulled out of the metal, as in a spark plug.

(iii) *Photo-electric emission*: When light of suitable frequency illuminates a metal surface, electrons are emitted from the metal surface. These photo(light)-generated electrons are called *photoelectrons*.

11.3 PHOTOELECTRIC EFFECT

11.3.1 Hertz's observations

The phenomenon of photoelectric emission was discovered in 1887 by Heinrich Hertz (1857-1894), during his electromagnetic wave experiments. In his experimental investigation on the production of electromagnetic waves by means of a spark discharge, Hertz observed that high voltage sparks across the detector loop were enhanced when the emitter plate was illuminated by ultraviolet light from an arc lamp.

Light shining on the metal surface somehow facilitated the escape of free, charged particles which we now know as electrons. When light falls on a metal surface, some electrons near the surface absorb enough energy from the incident radiation to overcome the attraction of the positive ions in the material of the surface. After gaining sufficient energy from the incident light, the electrons escape from the surface of the metal into the surrounding space.

11.3.2 Hallwachs' and Lenard's observations

Wilhelm Hallwachs and Philipp Lenard investigated the phenomenon of photoelectric emission in detail during 1886-1902.

Lenard (1862-1947) observed that when ultraviolet radiations were allowed to fall on the emitter plate of an evacuated glass tube enclosing two electrodes (metal plates), current flows in the circuit (Fig. 11.1). As soon as the ultraviolet radiations were stopped, the current flow also

stopped. These observations indicate that when ultraviolet radiations fall on the emitter plate C, electrons are ejected from it which are attracted towards the positive, collector plate A by the electric field. The electrons flow through the evacuated glass tube, resulting in the current flow. Thus, light falling on the surface of the emitter causes current in the external circuit. Hallwachs and Lenard studied how this photo current varied with collector plate potential, and with frequency and intensity of incident light.

Hallwachs, in 1888, undertook the study further and connected a negatively charged zinc plate to an electroscope. He observed that the zinc plate lost its charge when it was illuminated by ultraviolet light. Further, the uncharged zinc plate became positively charged when it was irradiated by ultraviolet light. Positive charge on a positively charged zinc plate was found to be further enhanced when it was illuminated by ultraviolet light. From these observations he concluded that negatively charged particles were emitted from the zinc plate under the action of ultraviolet light.

After the discovery of the electron in 1897, it became evident that the incident light causes electrons to be emitted from the emitter plate. Due to negative charge, the emitted electrons are pushed towards the collector plate by the electric field. Hallwachs and Lenard also observed that when ultraviolet light fell on the emitter plate, no electrons were emitted at all when the frequency of the incident light was smaller than a certain minimum value, called the *threshold frequency*. This minimum frequency depends on the nature of the material of the emitter plate.

It was found that certain metals like zinc, cadmium, magnesium, etc., responded only to ultraviolet light, having short wavelength, to cause electron emission from the surface. However, some alkali metals such as lithium, sodium, potassium, caesium and rubidium were sensitive even to visible light. All these *photosensitive substances* emit electrons when they are illuminated by light. After the discovery of electrons, these electrons were termed as *photoelectrons.* The phenomenon is called *photoelectric effect.*

PHYSICS

Simulate experiments on photoelectric effect
http://physics.berea.edu/~king/Teaching/ModPhys/QM/Photoelectric/Photoelectric.html

11.4 Experimental Study of Photoelectric Effect

Figure 11.1 depicts a schematic view of the arrangement used for the experimental study of the photoelectric effect. It consists of an evacuated glass/quartz tube having a photosensitive plate C and another metal plate A. Monochromatic light from the source S of sufficiently short wavelength passes through the window W and falls on the photosensitive plate C (emitter). A transparent quartz window is sealed on to the glass tube, which permits ultraviolet radiation to pass through it and irradiate the photosensitive plate C. The electrons are emitted by the plate C and are collected by the plate A (collector), by the electric field created by the battery. The battery maintains the potential difference between the plates C and A, that can be varied. The polarity of the plates C and A can be reversed by a commutator. Thus, the plate A can be maintained at a desired positive or negative potential with respect to emitter C. When the collector plate A is positive with respect to the emitter plate C, the electrons are

FIGURE 11.1 Experimental arrangement for study of photoelectric effect.

attracted to it. The emission of electrons causes flow of electric current in the circuit. The potential difference between the emitter and collector plates is measured by a voltmeter (V) whereas the resulting photo current flowing in the circuit is measured by a microammeter (μA). The photoelectric current can be increased or decreased by varying the potential of collector plate A with respect to the emitter plate C. The intensity and frequency of the incident light can be varied, as can the potential difference V between the emitter C and the collector A.

We can use the experimental arrangement of Fig. 11.1 to study the variation of photocurrent with (a) intensity of radiation, (b) frequency of incident radiation, (c) the potential difference between the plates A and C, and (d) the nature of the material of plate C. Light of different frequencies can be used by putting appropriate coloured filter or coloured glass in the path of light falling on the emitter C. The intensity of light is varied by changing the distance of the light source from the emitter.

FIGURE 11.2 Variation of Photoelectric current with intensity of light.

11.4.1 Effect of intensity of light on photocurrent

The collector A is maintained at a positive potential with respect to emitter C so that electrons ejected from C are attracted towards collector A. Keeping the frequency of the incident radiation and the accelerating potential fixed, the intensity of light is varied and the resulting photoelectric current is measured each time. It is found that the photocurrent increases linearly with intensity of incident light as shown graphically in Fig. 11.2. The photocurrent is directly proportional to the number of photoelectrons emitted per second. This implies that *the number of photoelectrons emitted per second is directly proportional to the intensity of incident radiation.*

11.4.2 Effect of potential on photoelectric current

We first keep the plate A at some positive accelerating potential with respect to the plate C and illuminate the plate C with light of fixed frequency ν and fixed intensity I_1. We next vary the positive potential of plate A gradually and measure the resulting photocurrent each time. It is found that the photoelectric current increases with increase in accelerating (positive) potential. At some stage, for a certain positive potential of plate A, all the emitted electrons are collected by the plate A and the photoelectric current becomes maximum or saturates. If we increase the accelerating potential of plate A further, the photocurrent does not increase. This maximum value of the photoelectric current is called *saturation current.* Saturation current corresponds to the case when all the photoelectrons emitted by the emitter plate C reach the collector plate A.

We now apply a negative (retarding) potential to the plate A with respect to the plate C and make it increasingly negative gradually. When the

polarity is reversed, the electrons are repelled and only the most energetic electrons are able to reach the collector A. The photocurrent is found to decrease rapidly until it drops to zero at a certain sharply defined, critical value of the negative potential V_0 on the plate A. For a particular frequency of incident radiation, *the minimum negative (retarding) potential V_0 given to the plate A for which the photocurrent stops or becomes zero is called the cut-off or stopping potential.*

The interpretation of the observation in terms of photoelectrons is straightforward. All the photoelectrons emitted from the metal do not have the same energy. Photoelectric current is zero when the stopping potential is sufficient to repel even the most energetic photoelectrons, with the maximum kinetic energy (K_{max}), so that

FIGURE 11.3 Variation of photocurrent with collector plate potential for different intensity of incident radiation.

$$K_{max} = e\,V_0 \qquad (11.1)$$

We can now repeat this experiment with incident radiation of the same frequency but of higher intensity I_2 and I_3 ($I_3 > I_2 > I_1$). We note that the saturation currents are now found to be at higher values. This shows that more electrons are being emitted per second, proportional to the intensity of incident radiation. But the stopping potential remains the same as that for the incident radiation of intensity I_1, as shown graphically in Fig. 11.3. Thus, *for a given frequency of the incident radiation, the stopping potential is independent of its intensity.* In other words, the maximum kinetic energy of photoelectrons depends on the light source and the emitter plate material, but is independent of intensity of incident radiation.

11.4.3 Effect of frequency of incident radiation on stopping potential

We now study the relation between the frequency ν of the incident radiation and the stopping potential V_0. We suitably adjust the same intensity of light radiation at various frequencies and study the variation of photocurrent with collector plate potential. The resulting variation is shown in Fig. 11.4. We obtain different values of stopping potential but the same value of the saturation current for incident radiation of different frequencies. The energy of the emitted electrons depends on the frequency of the incident radiations. The stopping potential is more negative for higher frequencies of incident radiation. Note from

FIGURE 11.4 Variation of photoelectric current with collector plate potential for different frequencies of incident radiation.

FIGURE 11.5 Variation of stopping potential V_0 with frequency ν of incident radiation for a given photosensitive material.

Fig. 11.4 that the stopping potentials are in the order $V_{03} > V_{02} > V_{01}$ if the frequencies are in the order $\nu_3 > \nu_2 > \nu_1$. This implies that greater the frequency of incident light, greater is the maximum kinetic energy of the photoelectrons. Consequently, we need greater retarding potential to stop them completely. If we plot a graph between the frequency of incident radiation and the corresponding stopping potential for different metals we get a straight line, as shown in Fig. 11.5.

The graph shows that

(i) the stopping potential V_0 varies linearly with the frequency of incident radiation for a given photosensitive material.

(ii) there exists a certain minimum cut-off frequency ν_0 for which the stopping potential is zero.

These observations have two implications:

(i) *The maximum kinetic energy of the photoelectrons varies linearly with the frequency of incident radiation, but is independent of its intensity.*

(ii) *For a frequency ν of incident radiation, lower than the cut-off frequency ν_0, no photoelectric emission is possible even if the intensity is large.*

This minimum, cut-off frequency ν_0, is called the *threshold frequency*. It is different for different metals.

Different photosensitive materials respond differently to light. Selenium is more sensitive than zinc or copper. The same photosensitive substance gives different response to light of different wavelengths. For example, ultraviolet light gives rise to photoelectric effect in copper while green or red light does not.

Note that in all the above experiments, it is found that, if frequency of the incident radiation exceeds the threshold frequency, the photoelectric emission starts instantaneously without any apparent time lag, even if the incident radiation is very dim. It is now known that emission starts in a time of the order of 10^{-9} s or less.

We now summarise the experimental features and observations described in this section.

(i) For a given photosensitive material and frequency of incident radiation (above the threshold frequency), the photoelectric current is directly proportional to the intensity of incident light (Fig. 11.2).

(ii) For a given photosensitive material and frequency of incident radiation, saturation current is found to be proportional to the intensity of incident radiation whereas the stopping potential is independent of its intensity (Fig. 11.3).

(iii) For a given photosensitive material, there exists a certain minimum cut-off frequency of the incident radiation, called the *threshold frequency*, below which no emission of photoelectrons takes place, no matter how intense the incident light is. Above the threshold frequency, the stopping potential or equivalently the maximum kinetic

energy of the emitted photoelectrons increases linearly with the frequency of the incident radiation, but is independent of its intensity (Fig. 11.5).

(iv) The photoelectric emission is an instantaneous process without any apparent time lag (~10^{-9}s or less), even when the incident radiation is made exceedingly dim.

11.5 Photoelectric Effect and Wave Theory of Light

The wave nature of light was well established by the end of the nineteenth century. The phenomena of interference, diffraction and polarisation were explained in a natural and satisfactory way by the wave picture of light. According to this picture, light is an electromagnetic wave consisting of electric and magnetic fields with continuous distribution of energy over the region of space over which the wave is extended. Let us now see if this wave picture of light can explain the observations on photoelectric emission given in the previous section.

According to the wave picture of light, the free electrons at the surface of the metal (over which the beam of radiation falls) absorb the radiant energy continuously. The greater the intensity of radiation, the greater are the amplitude of electric and magnetic fields. Consequently, the greater the intensity, the greater should be the energy absorbed by each electron. In this picture, the maximum kinetic energy of the photoelectrons on the surface is then expected to increase with increase in intensity. Also, no matter what the frequency of radiation is, a sufficiently intense beam of radiation (over sufficient time) should be able to impart enough energy to the electrons, so that they exceed the minimum energy needed to escape from the metal surface . A threshold frequency, therefore, should not exist. These expectations of the wave theory directly contradict observations (i), (ii) and (iii) given at the end of sub-section 11.4.3.

Further, we should note that in the wave picture, the absorption of energy by electron takes place continuously over the entire wavefront of the radiation. Since a large number of electrons absorb energy, the energy absorbed per electron per unit time turns out to be small. Explicit calculations estimate that it can take hours or more for a single electron to pick up sufficient energy to overcome the work function and come out of the metal. This conclusion is again in striking contrast to observation (iv) that the photoelectric emission is instantaneous. In short, the wave picture is unable to explain the most basic features of photoelectric emission.

11.6 Einstein's Photoelectric Equation: Energy Quantum of Radiation

In 1905, Albert Einstein (1879-1955) proposed a radically new picture of electromagnetic radiation to explain photoelectric effect. In this picture, photoelectric emission does not take place by continuous absorption of energy from radiation. Radiation energy is built up of discrete units – the so called *quanta of energy of radiation*. Each quantum of radiant energy

Albert Einstein (1879 – 1955) Einstein, one of the greatest physicists of all time, was born in Ulm, Germany. In 1905, he published three path-breaking papers. In the first paper, he introduced the notion of light quanta (now called photons) and used it to explain the features of photoelectric effect. In the second paper, he developed a theory of Brownian motion, confirmed experimentally a few years later and provided a convincing evidence of the atomic picture of matter. The third paper gave birth to the special theory of relativity. In 1916, he published the general theory of relativity. Some of Einstein's most significant later contributions are: the notion of stimulated emission introduced in an alternative derivation of Planck's blackbody radiation law, static model of the universe which started modern cosmology, quantum statistics of a gas of massive bosons, and a critical analysis of the foundations of quantum mechanics. In 1921, he was awarded the Nobel Prize in physics for his contribution to theoretical physics and the photoelectric effect.

ALBERT EINSTEIN (1879 – 1955)

has energy $h\nu$, where h is Planck's constant and ν the frequency of light. In photoelectric effect, an electron absorbs a quantum of energy ($h\nu$) of radiation. If this quantum of energy absorbed exceeds the minimum energy needed for the electron to escape from the metal surface (work function ϕ_0), the electron is emitted with maximum kinetic energy

$$K_{max} = h\nu - \phi_0 \qquad (11.2)$$

More tightly bound electrons will emerge with kinetic energies less than the maximum value. Note that the intensity of light of a given frequency is determined by the number of photons incident per second. Increasing the intensity will increase the number of emitted electrons per second. However, the maximum kinetic energy of the emitted photoelectrons is determined by the energy of each photon.

Equation (11.2) is known as *Einstein's photoelectric equation*. We now see how this equation accounts in a simple and elegant manner all the observations on photoelectric effect given at the end of sub-section 11.4.3.

- According to Eq. (11.2), K_{max} depends linearly on ν, and is independent of intensity of radiation, in agreement with observation. This has happened because in Einstein's picture, photoelectric effect arises from the absorption of a single quantum of radiation by a single electron. The intensity of radiation (that is proportional to the number of energy quanta per unit area per unit time) is irrelevant to this basic process.
- Since K_{max} must be non-negative, Eq. (11.2) implies that photoelectric emission is possible only if
  $h\nu > \phi_0$
  or $\nu > \nu_0$, where
  $$\nu_0 = \frac{\phi_0}{h} \qquad (11.3)$$
  Equation (11.3) shows that the greater the work function ϕ_0, the higher the minimum or threshold frequency ν_0 needed to emit photoelectrons. Thus, there exists a threshold frequency ν_0 (= ϕ_0/h) for the metal surface, below which no photoelectric emission is possible, no matter how intense the incident radiation may be or how long it falls on the surface.
- In this picture, intensity of radiation as noted above, is proportional to the number of energy quanta per unit area per unit time. The greater the number of energy quanta available, the greater is the number of electrons absorbing the energy quanta and greater, therefore, is the number of electrons coming out of the metal (for $\nu > \nu_0$). This explains why, for $\nu > \nu_0$, photoelectric current is proportional to intensity.

- In Einstein's picture, the basic elementary process involved in photoelectric effect is the absorption of a light quantum by an electron. This process is instantaneous. Thus, whatever may be the intensity i.e., the number of quanta of radiation per unit area per unit time, photoelectric emission is instantaneous. Low intensity does not mean delay in emission, since the basic elementary process is the same. Intensity only determines how many electrons are able to participate in the elementary process (absorption of a light quantum by a single electron) and, therefore, the photoelectric current.

Using Eq. (11.1), the photoelectric equation, Eq. (11.2), can be written as

$$e V_0 = h \nu - \phi_0; \text{ for } \nu \geq \nu_0$$

$$\text{or } V_0 = \left(\frac{h}{e}\right) \nu - \frac{\phi_0}{e} \tag{11.4}$$

This is an important result. It predicts that the V_0 versus ν curve is a straight line with slope = (h/e), independent of the nature of the material. During 1906-1916, Millikan performed a series of experiments on photoelectric effect, aimed at disproving Einstein's photoelectric equation. He measured the slope of the straight line obtained for sodium, similar to that shown in Fig. 11.5. Using the known value of e, he determined the value of Planck's constant h. This value was close to the value of Planck's contant (= 6.626×10^{-34} J s) determined in an entirely different context. In this way, in 1916, Millikan proved the validity of Einstein's photoelectric equation, instead of disproving it.

The successful explanation of photoelectric effect using the hypothesis of light quanta and the experimental determination of values of h and ϕ_0, in agreement with values obtained from other experiments, led to the acceptance of Einstein's picture of photoelectric effect. Millikan verified photoelectric equation with great precision, for a number of alkali metals over a wide range of radiation frequencies.

11.7 Particle Nature of Light: The Photon

Photoelectric effect thus gave evidence to the strange fact that light in interaction with matter behaved as if it was made of quanta or packets of energy, each of energy $h\nu$.

Is the light quantum of energy to be associated with a particle? Einstein arrived at the important result, that the light quantum can also be associated with momentum $(h\nu/c)$. A definite value of energy as well as momentum is a strong sign that the light quantum can be associated with a particle. This particle was later named *photon*. The particle-like behaviour of light was further confirmed, in 1924, by the experiment of A.H. Compton (1892-1962) on scattering of X-rays from electrons. In 1921, Einstein was awarded the Nobel Prize in Physics for his contribution to theoretical physics and the photoelectric effect. In 1923, Millikan was awarded the Nobel Prize in physics for his work on the elementary charge of electricity and on the photoelectric effect.

We can summarise the photon picture of electromagnetic radiation as follows:

(i) In interaction of radiation with matter, radiation behaves as if it is made up of particles called photons.

(ii) Each photon has energy E (=$h\nu$) and momentum p (= $h\nu/c$), and speed c, the speed of light.

(iii) All photons of light of a particular frequency ν, or wavelength λ, have the same energy E (=$h\nu = hc/\lambda$) and momentum p (= $h\nu/c = h/\lambda$), whatever the intensity of radiation may be. By increasing the intensity of light of given wavelength, there is only an increase in the number of photons per second crossing a given area, with each photon having the same energy. Thus, photon energy is independent of intensity of radiation.

(iv) Photons are electrically neutral and are not deflected by electric and magnetic fields.

(v) In a photon-particle collision (such as photon-electron collision), the total energy and total momentum are conserved. However, the number of photons may not be conserved in a collision. The photon may be absorbed or a new photon may be created.

EXAMPLE 11.1

Example 11.1 Monochromatic light of frequency 6.0×10^{14} Hz is produced by a laser. The power emitted is 2.0×10^{-3} W. (a) What is the energy of a photon in the light beam? (b) How many photons per second, on an average, are emitted by the source?

Solution

(a) Each photon has an energy

$$E = h\nu = (6.63 \times 10^{-34}\ \text{J s})\ (6.0 \times 10^{14}\ \text{Hz})$$
$$= 3.98 \times 10^{-19}\ \text{J}$$

(b) If N is the number of photons emitted by the source per second, the power P transmitted in the beam equals N times the energy per photon E, so that $P = N E$. Then

$$N = \frac{P}{E} = \frac{2.0 \times 10^{-3}\ \text{W}}{3.98 \times 10^{-19}\ \text{J}}$$
$$= 5.0 \times 10^{15}\ \text{photons per second.}$$

EXAMPLE 11.2

Example 11.2 The work function of caesium is 2.14 eV. Find (a) the threshold frequency for caesium, and (b) the wavelength of the incident light if the photocurrent is brought to zero by a stopping potential of 0.60 V.

Solution

(a) For the cut-off or threshold frequency, the energy $h\nu_0$ of the incident radiation must be equal to work function ϕ_0, so that

$$\nu_0 = \frac{\phi_0}{h} = \frac{2.14\ \text{eV}}{6.63 \times 10^{-34}\ \text{J s}}$$
$$= \frac{2.14 \times 1.6 \times 10^{-19}\ \text{J}}{6.63 \times 10^{-34}\ \text{J s}} = 5.16 \times 10^{14}\ \text{Hz}$$

Thus, for frequencies less than this threshold frequency, no photoelectrons are ejected.

(b) Photocurrent reduces to zero, when maximum kinetic energy of the emitted photoelectrons equals the potential energy eV_0 by the retarding potential V_0. Einstein's Photoelectric equation is

$$eV_0 = h\nu - \phi_0 = \frac{hc}{\lambda} - \phi_0$$

or, $\lambda = hc/(eV_0 + \phi_0)$

$$= \frac{(6.63 \times 10^{-34}\,\text{J s}) \times (3 \times 10^8\,\text{m/s})}{(0.60\,\text{eV} + 2.14\,\text{eV})}$$

$$= \frac{19.89 \times 10^{-26}\,\text{J m}}{(2.74\,\text{eV})}$$

$$\lambda = \frac{19.89 \times 10^{-26}\,\text{J m}}{2.74 \times 1.6 \times 10^{-19}\,\text{J}} = 454\ \text{nm}$$

EXAMPLE 11.2

Example 11.3 The wavelength of light in the visible region is about 390 nm for violet colour, about 550 nm (average wavelength) for yellow-green colour and about 760 nm for red colour.

(a) What are the energies of photons in (eV) at the (i) violet end, (ii) average wavelength, yellow-green colour, and (iii) red end of the visible spectrum? (Take $h = 6.63\times10^{-34}$ J s and 1 eV = 1.6×10^{-19}J.)

(b) From which of the photosensitive materials with work functions listed in Table 11.1 and using the results of (i), (ii) and (iii) of (a), can you build a photoelectric device that operates with visible light?

Solution

(a) Energy of the incident photon, $E = h\nu = hc/\lambda$

$E = (6.63\times10^{-34}\text{J s})\ (3\times10^8\ \text{m/s})/\lambda$

$$= \frac{1.989 \times 10^{-25}\,\text{J m}}{\lambda}$$

(i) For violet light, λ_1 = 390 nm (lower wavelength end)

Incident photon energy, $E_1 = \dfrac{1.989 \times 10^{-25}\,\text{J m}}{390\times10^{-9}\ \text{m}}$

$= 5.10 \times 10^{-19}$J

$= \dfrac{5.10 \times 10^{-19}\,\text{J}}{1.6\times10^{-19}\ \text{J/eV}}$

= 3.19 eV

(ii) For yellow-green light, λ_2 = 550 nm (average wavelength)

Incident photon energy, $E_2 = \dfrac{1.989 \times 10^{-25}\,\text{J m}}{550\times10^{-9}\ \text{m}}$

$= 3.62\times10^{-19}$ J = 2.26 eV

(iii) For red light, λ_3 = 760 nm (higher wavelength end)

Incident photon energy, $E_3 = \dfrac{1.989 \times 10^{-25}\,\text{J m}}{760\times10^{-9}\ \text{m}}$

$= 2.62\times10^{-19}$ J = 1.64 eV

(b) For a photoelectric device to operate, we require incident light energy E to be equal to or greater than the work function ϕ_0 of the material. Thus, the photoelectric device will operate with violet light (with E = 3.19 eV) photosensitive material Na (with ϕ_0 = 2.75 eV), K (with ϕ_0 = 2.30 eV) and Cs (with ϕ_0 = 2.14 eV). It will also operate with yellow-green light (with E = 2.26 eV) for Cs (with ϕ_0 = 2.14 eV) only. However, it will not operate with red light (with E = 1.64 eV) for any of these photosensitive materials.

EXAMPLE 11.3

11.8 Wave Nature of Matter

The dual (wave-particle) nature of light (electromagnetic radiation, in general) comes out clearly from what we have learnt in this and the preceding chapters. The wave nature of light shows up in the phenomena of interference, diffraction and polarisation. On the other hand, in photoelectric effect and Compton effect which involve energy and momentum transfer, radiation behaves as if it is made up of a bunch of particles – the photons. Whether a particle or wave description is best suited for understanding an experiment depends on the nature of the experiment. For example, in the familiar phenomenon of seeing an object by our eye, both descriptions are important. The gathering and focussing mechanism of light by the eye-lens is well described in the wave picture. But its absorption by the rods and cones (of the retina) requires the photon picture of light.

A natural question arises: If radiation has a dual (wave-particle) nature, might not the particles of nature (the electrons, protons, etc.) also exhibit wave-like character? In 1924, the French physicist Louis Victor de Broglie (pronounced as de Broy) (1892-1987) put forward the bold hypothesis that moving particles of matter should display wave-like properties under suitable conditions. He reasoned that nature was symmetrical and that the two basic physical entities – matter and energy, must have symmetrical character. If radiation shows dual aspects, so should matter. De Broglie proposed that the wave length λ associated with a particle of momentum p is given as

$$\lambda = \frac{h}{p} = \frac{h}{mv} \tag{11.5}$$

where m is the mass of the particle and v its speed. Equation (11.5) is known as the *de Broglie relation* and the wavelength λ of the *matter wave* is called *de Broglie wavelength*. The dual aspect of matter is evident in the de Broglie relation. On the left hand side of Eq. (11.5), λ is the attribute of a wave while on the right hand side the momentum p is a typical attribute of a particle. Planck's constant h relates the two attributes.

Equation (11.5) for a material particle is basically a hypothesis whose validity can be tested only by experiment. However, it is interesting to see that it is satisfied also by a photon. For a photon, as we have seen,

$$p = h\nu / c \tag{11.6}$$

Therefore,

$$\frac{h}{p} = \frac{c}{\nu} = \lambda \tag{11.7}$$

That is, the de Broglie wavelength of a photon given by Eq. (11.5) equals the wavelength of electromagnetic radiation of which the photon is a quantum of energy and momentum.

Clearly, from Eq. (11.5), λ is smaller for a heavier particle (large m) or more energetic particle (large v). For example, the de Broglie wavelength of a ball of mass 0.12 kg moving with a speed of 20 m s^{-1} is easily calculated:

PHOTOCELL

A photocell is a technological application of the photoelectric effect. It is a device whose electrical properties are affected by light. It is also sometimes called an electric eye. A photocell consists of a semi-cylindrical photo-sensitive metal plate C (emitter) and a wire loop A (collector) supported in an evacuated glass or quartz bulb. It is connected to the external circuit having a high-tension battery B and microammeter (μA) as shown in the Figure. Sometimes, instead of the plate C, a thin layer of photosensitive material is pasted on the inside of the bulb. A part of the bulb is left clean for the light to enter it.

When light of suitable wavelength falls on the emitter C, photoelectrons are emitted. These photoelectrons are drawn to the collector A. Photocurrent of the order of a few microampere can be normally obtained from a photo cell.

A photocell converts a change in intensity of illumination into a change in photocurrent. This current can be used to operate control systems and in light measuring devices. A photocell of lead sulphide sensitive to infrared radiation is used in electronic ignition circuits.

A photo cell

In scientific work, photo cells are used whenever it is necessary to measure the intensity of light. Light meters in photographic cameras make use of photo cells to measure the intensity of incident light. The photocells, inserted in the door light electric circuit, are used as automatic door opener. A person approaching a doorway may interrupt a light beam which is incident on a photocell. The abrupt change in photocurrent may be used to start a motor which opens the door or rings an alarm. They are used in the control of a counting device which records every interruption of the light beam caused by a person or object passing across the beam. So photocells help count the persons entering an auditorium, provided they enter the hall one by one. They are used for detection of traffic law defaulters: an alarm may be sounded whenever a beam of *(invisible)* radiation is intercepted.

In burglar alarm, (invisible) ultraviolet light is continuously made to fall on a photocell installed at the doorway. A person entering the door interrupts the beam falling on the photocell. The abrupt change in photocurrent is used to start an electric bell ringing. In fire alarm, a number of photocells are installed at suitable places in a building. In the event of breaking out of fire, light radiations fall upon the photocell. This completes the electric circuit through an electric bell or a siren which starts operating as a warning signal.

Photocells are used in the reproduction of sound in motion pictures and in the television camera for scanning and telecasting scenes. They are used in industries for detecting minor flaws or holes in metal sheets.

$$p = m\,v = 0.12 \text{ kg} \times 20 \text{ m s}^{-1} = 2.40 \text{ kg m s}^{-1}$$

$$\lambda = \frac{h}{p} = \frac{6.63 \times 10^{-34} \text{ J s}}{2.40 \text{ kg m s}^{-1}} = 2.76 \times 10^{-34} \text{ m}$$

LOUIS VICTOR DE BROGLIE (1892 – 1987)

Louis Victor de Broglie (1892 – 1987) French physicist who put forth revolutionary idea of wave nature of matter. This idea was developed by Erwin Schródinger into a full-fledged theory of quantum mechanics commonly known as wave mechanics. In 1929, he was awarded the Nobel Prize in Physics for his discovery of the wave nature of electrons.

This wavelength is so small that it is beyond any measurement. This is the reason why macroscopic objects in our daily life do not show wave-like properties. On the other hand, in the sub-atomic domain, the wave character of particles is significant and measurable.

Consider an electron (mass m, charge e) accelerated from rest through a potential V. The kinetic energy K of the electron equals the work done (eV) on it by the electric field:

$$K = e\,V \tag{11.8}$$

Now, $K = \frac{1}{2} m v^2 = \frac{p^2}{2m}$, so that

$$p = \sqrt{2\,m\,K} = \sqrt{2\,m\,eV} \tag{11.9}$$

The de Broglie wavelength λ of the electron is then

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{2\,mK}} = \frac{h}{\sqrt{2\,meV}} \tag{11.10}$$

Substituting the numerical values of h, m, e, we get

$$\lambda = \frac{1.227}{\sqrt{V}}\ \text{nm} \tag{11.11}$$

where V is the magnitude of accelerating potential in volts. For a 120 V accelerating potential, Eq. (11.11) gives λ = 0.112 nm. This wavelength is of the same order as the spacing between the atomic planes in crystals. This suggests that matter waves associated with an electron could be verified by crystal diffraction experiments analogous to X-ray diffraction. We describe the experimental verification of the de Broglie hypothesis in the next section. In 1929, de Broglie was awarded the Nobel Prize in Physics for his discovery of the wave nature of electrons.

The matter–wave picture elegantly incorporated the Heisenberg's *uncertainty principle*. According to the principle, it is not possible to measure *both* the position and momentum of an electron (or any other particle) *at the same time* exactly. There is always some uncertainty (Δx) in the specification of position and some uncertainty (Δp) in the specification of momentum. The product of Δx and Δp is of the order of $\hbar$* (with $\hbar = h/2\pi$), i.e.,

$$\Delta x\,\Delta p \approx \hbar \tag{11.12}$$

Equation (11.12) allows the possibility that Δx is zero; but then Δp must be infinite in order that the product is non-zero. Similarly, if Δp is zero, Δx must be infinite. Ordinarily, both Δx and Δp are non-zero such that their product is of the order of $\hbar$.

Now, if an electron has a definite momentum p, (i.e.$\Delta p = 0$), by the de Broglie relation, it has a definite wavelength λ. A wave of definite (single)

* A more rigorous treatment gives $\Delta x\,\Delta p \geq \hbar/2$.

wavelength extends all over space. By Born's probability interpretation this means that the electron is not localised in any finite region of space. That is, its position uncertainty is infinite ($\Delta x \to \infty$), which is consistent with the uncertainty principle.

In general, the matter wave associated with the electron is not extended all over space. It is a wave packet extending over some finite region of space. In that case Δx is not infinite but has some finite value depending on the extension of the wave packet. Also, you must appreciate that a wave packet of finite extension does not have a single wavelength. It is built up of wavelengths spread around some central wavelength.

By de Broglie's relation, then, the momentum of the electron will also have a spread – an uncertainty Δp. This is as expected from the uncertainty principle. It can be shown that the wave packet description together with de Broglie relation and Born's probability interpretation reproduce the Heisenberg's uncertainty principle exactly.

In Chapter 12, the de Broglie relation will be seen to justify Bohr's postulate on quantisation of angular momentum of electron in an atom.

Figure 11.6 shows a schematic diagram of (a) a localised wave packet, and (b) an extended wave with fixed wavelength.

FIGURE 11.6 (a) The wave packet description of an electron. The wave packet corresponds to a spread of wavelength around some central wavelength (and hence by de Broglie relation, a spread in momentum). Consequently, it is associated with an uncertainty in position (Δx) and an uncertainty in momentum (Δp). (b) The matter wave corresponding to a definite momentum of an electron extends all over space. In this case, $\Delta p = 0$ and $\Delta x \to \infty$.

Example 11.4 What is the de Broglie wavelength associated with (a) an electron moving with a speed of 5.4×10^6 m/s, and (b) a ball of mass 150 g travelling at 30.0 m/s?

Solution

(a) For the electron:

Mass $m = 9.11\times10^{-31}$ kg, speed $v = 5.4\times10^6$ m/s. Then, momentum

$p = m\,v = 9.11\times10^{-31}$ (kg) $\times$ 5.4×10^6 (m/s)

$p = 4.92 \times 10^{-24}$ kg m/s

de Broglie wavelength, $\lambda = h/p$

$$= \frac{6.63\times10^{-34}\ \text{J s}}{4.92\times10^{-24}\ \text{kg m/s}}$$

$\lambda = 0.135$ nm

(b) For the ball:

Mass $m' = 0.150$ kg, speed $v' = 30.0$ m/s.

Then momentum $p' = m'\,v' = 0.150$ (kg) $\times$ 30.0 (m/s)

$p' = 4.50$ kg m/s

de Broglie wavelength $\lambda' = h/p'$.

EXAMPLE 11.4

EXAMPLE 11.4

$$= \frac{6.63 \times 10^{\pm 34}\ \text{J s}}{4.50 \times\ \text{kg m/s}}$$

$\lambda' = 1.47 \times 10^{-34}$ m

The de Broglie wavelength of electron is comparable with X-ray wavelengths. However, for the ball it is about 10^{-19} times the size of the proton, quite beyond experimental measurement.

EXAMPLE 11.5

Example 11.5 An electron, an α-particle, and a proton have the same kinetic energy. Which of these particles has the shortest de Broglie wavelength?

Solution

For a particle, de Broglie wavelength, $\lambda = h/p$

Kinetic energy, $K = p^2/2m$

Then, $\lambda = h/\sqrt{2mK}$

For the same kinetic energy K, the de Broglie wavelength associated with the particle is inversely proportional to the square root of their masses. A proton $\left({}^1_1\text{H}\right)$ is 1836 times massive than an electron and an α-particle $\left({}^4_2\text{He}\right)$ four times that of a proton.

Hence, α – particle has the shortest de Broglie wavelength.

PROBABILITY INTERPRETATION TO MATTER WAVES

It is worth pausing here to reflect on just what a matter wave associated with a particle, say, an electron, means. Actually, a truly satisfactory physical understanding of the dual nature of matter and radiation has not emerged so far. The great founders of quantum mechanics (Niels Bohr, Albert Einstein, and many others) struggled with this and related concepts for long. Still the deep physical interpretation of quantum mechanics continues to be an area of active research. Despite this, the concept of matter wave has been mathematically introduced in modern quantum mechanics with great success. An important milestone in this connection was when Max Born (1882-1970) suggested a probability interpretation to the matter wave amplitude. According to this, the intensity (square of the amplitude) of the matter wave at a point determines the probability density of the particle at that point. Probability density means probability per unit volume. Thus, if A is the amplitude of the wave at a point, $|A|^2\ \Delta V$ is the probability of the particle being found in a small volume ΔV around that point. Thus, if the intensity of matter wave is large in a certain region, there is a greater probability of the particle being found there than where the intensity is small.

EXAMPLE 11.6

Example 11.6 A particle is moving three times as fast as an electron. The ratio of the de Broglie wavelength of the particle to that of the electron is 1.813×10^{-4}. Calculate the particle's mass and identify the particle.

Solution

de Broglie wavelength of a moving particle, having mass m and velocity v:

$$\lambda = \frac{h}{p} = \frac{h}{mv}$$

Mass, $m = h/\lambda v$

For an electron, mass $m_e = h/\lambda_e v_e$

Now, we have $v/v_e = 3$ and

$\lambda/\lambda_e = 1.813 \times 10^{-4}$

Then, mass of the particle, $m = m_e \left(\frac{\lambda_e}{\lambda}\right)\left(\frac{v_e}{v}\right)$

$m = (9.11\times10^{-31}\ \text{kg}) \times (1/3) \times (1/1.813 \times 10^{-4})$

$m = 1.675 \times 10^{-27}$ kg.

Thus, the particle, with this mass could be a proton or a neutron.

EXAMPLE 11.6

Example 11.7 What is the de Broglie wavelength associated with an electron, accelerated through a potential differnece of 100 volts?

Solution Accelerating potential V = 100 V. The de Broglie wavelength λ is

$$\lambda = h/p = \frac{1.227}{\sqrt{V}}\ \text{nm}$$

$$\lambda = \frac{1.227}{\sqrt{100}}\ \text{nm} = 0.123\ \text{nm}$$

The de Broglie wavelength associated with an electron in this case is of the order of X-ray wavelengths.

EXAMPLE 11.7

11.9 Davisson and Germer Experiment

The wave nature of electrons was first experimentally verified by C.J. Davisson and L.H. Germer in 1927 and independently by G.P. Thomson, in 1928, who observed diffraction effects with beams of electrons scattered by crystals. Davisson and Thomson shared the Nobel Prize in 1937 for their experimental discovery of diffraction of electrons by crystals.

The experimental arrangement used by Davisson and Germer is schematically shown in Fig. 11.7. It consists of an electron gun which comprises of a tungsten filament F, coated with barium oxide and heated by a low voltage power supply (L.T. or battery). Electrons emitted by the filament are accelerated to a desired velocity

FIGURE 11.7 Davisson-Germer electron diffraction arrangement.

by applying suitable potential/voltage from a high voltage power supply (H.T. or battery). They are made to pass through a cylinder with fine holes along its axis, producing a fine collimated beam. The beam is made to fall on the surface of a nickel crystal. The electrons are scattered in all directions by the atoms of the crystal. The intensity of the electron beam, scattered in a given direction, is measured by the electron detector (collector). The detector can be moved on a circular scale and is connected to a sensitive galvanometer, which records the current. The deflection of the galvanometer is proportional to the intensity of the electron beam entering the collector. The apparatus is enclosed in an evacuated chamber. By moving the detector on the circular scale at different positions, the intensity of the scattered electron beam is measured for different values of angle of scattering θ which is the angle between the incident and the scattered electron beams. The variation of the intensity (I) of the scattered electrons with the angle of scattering θ is obtained for different accelerating voltages.

The experiment was performed by varying the accelarating voltage from 44 V to 68 V. It was noticed that a strong peak appeared in the intensity (I) of the scattered electron for an accelarating voltage of 54V at a scattering angle $\theta = 50°$

The appearance of the peak in a particular direction is due to the constructive interference of electrons scattered from different layers of the regularly spaced atoms of the crystals. From the electron diffraction measurements, the wavelength of matter waves was found to be 0.165 nm.

The de Broglie wavelength λ associated with electrons, using Eq. (11.11), for $V = 54$ V is given by

$$\lambda = h/p = \frac{1.227}{\sqrt{V}} \text{ nm}$$

$$\lambda = \frac{1.227}{\sqrt{54}} \text{ nm} = 0.167 \text{ nm}$$

Thus, there is an excellent agreement between the theoretical value and the experimentally obtained value of de Broglie wavelength. Davisson-Germer experiment thus strikingly confirms the wave nature of electrons and the de Broglie relation. More recently, in 1989, the wave nature of a beam of electrons was experimentally demonstrated in a double-slit experiment, similar to that used for the wave nature of light. Also, in an experiment in 1994, interference fringes were obtained with the beams of iodine molecules, which are about a million times more massive than electrons.

The de Broglie hypothesis has been basic to the development of modern quantum mechanics. It has also led to the field of electron optics. The wave properties of electrons have been utilised in the design of electron microscope which is a great improvement, with higher resolution, over the optical microscope.

Development of electron microscope
http://nobelprize.org
PHYSICS

SUMMARY

1. The minimum energy needed by an electron to come out from a metal surface is called the work function of the metal. Energy (greater than the work function (ϕ_0) required for electron emission from the metal surface can be supplied by suitably heating or applying strong electric field or irradiating it by light of suitable frequency.
2. Photoelectric effect is the phenomenon of emission of electrons by metals when illuminated by light of suitable frequency. Certain metals respond to ultraviolet light while others are sensitive even to the visible light. Photoelectric effect involves conversion of light energy into electrical energy. It follows the law of conservation of energy. The photoelectric emission is an instantaneous process and possesses certain special features.
3. Photoelectric current depends on (i) the intensity of incident light, (ii) the potential difference applied between the two electrodes, and (iii) the nature of the emitter material.
4. The stopping potential (V_0) depends on (i) the frequency of incident light, and (ii) the nature of the emitter material. For a given frequency of incident light, it is independent of its intensity. The stopping potential is directly related to the maximum kinetic energy of electrons emitted: $e\,V_0 = (1/2)\,m\,v^2_{max} = K_{max}$.
5. Below a certain frequency (threshold frequency) ν_0, characteristic of the metal, no photoelectric emission takes place, no matter how large the intensity may be.
6. The classical wave theory could not explain the main features of photoelectric effect. Its picture of continuous absorption of energy from radiation could not explain the independence of K_{max} on intensity, the existence of ν_0 and the instantaneous nature of the process. Einstein explained these features on the basis of photon picture of light. According to this, light is composed of discrete packets of energy called quanta or photons. Each photon carries an energy E (= $h\,\nu$) and momentum p (= h/λ), which depend on the frequency (ν) of incident light and not on its intensity. Photoelectric emission from the metal surface occurs due to absorption of a photon by an electron.
7. Einstein's photoelectric equation is in accordance with the energy conservation law as applied to the photon absorption by an electron in the metal. The maximum kinetic energy $(1/2)m\,v^2_{max}$ is equal to the photon energy ($h\nu$) minus the work function ϕ_0 (= $h\nu_0$) of the target metal:

$$\frac{1}{2}\,m\,v^2_{max} = V_0\,e = h\nu - \phi_0 = h\,(\nu - \nu_0)$$

   This photoelectric equation explains all the features of the photoelectric effect. Millikan's first precise measurements confirmed the Einstein's photoelectric equation and obtained an accurate value of Planck's constant h. This led to the acceptance of particle or photon description (nature) of electromagnetic radiation, introduced by Einstein.
8. Radiation has dual nature: wave and particle. The nature of experiment determines whether a wave or particle description is best suited for understanding the experimental result. Reasoning that radiation and matter should be symmetrical in nature, Louis Victor de Broglie

attributed a wave-like character to matter (material particles). The waves associated with the moving material particles are called matter waves or de Broglie waves.

9. The de Broglie wavelength (λ) associated with a moving particle is related to its momentum p as: $\lambda = h/p$. The dualism of matter is inherent in the de Broglie relation which contains a wave concept (λ) and a particle concept (p). The de Broglie wavelength is independent of the charge and nature of the material particle. It is significantly measurable (of the order of the atomic-planes spacing in crystals) only in case of sub-atomic particles like electrons, protons, etc. (due to smallness of their masses and hence, momenta). However, it is indeed very small, quite beyond measurement, in case of macroscopic objects, commonly encountered in everyday life.
10. Electron diffraction experiments by Davisson and Germer, and by G. P. Thomson, as well as many later experiments, have verified and confirmed the wave-nature of electrons. The de Broglie hypothesis of matter waves supports the Bohr's concept of stationary orbits.

Physical Quantity	Symbol	Dimensions	Unit	Remarks
Planck's constant	h	$[ML^2T^{-1}]$	J s	$E = h\nu$
Stopping potential	V_0	$[ML^2T^{-3}A^{-1}]$	V	$eV_0 = K_{max}$
Work function	ϕ_0	$[ML^2T^{-2}]$	J; eV	$K_{max} = E - \phi_0$
Threshold frequency	ν_0	$[T^{-1}]$	Hz	$\nu_0 = \phi_0/h$
de Broglie wavelength	λ	[L]	m	$\lambda = h/p$

POINTS TO PONDER

1. Free electrons in a metal are free in the sense that they move inside the metal in a constant potential (This is only an approximation). They are not free to move out of the metal. They need additional energy to get out of the metal.
2. Free electrons in a metal do not all have the same energy. Like molecules in a gas jar, the electrons have a certain energy distribution at a given temperature. This distribution is different from the usual Maxwell's distribution that you have learnt in the study of kinetic theory of gases. You will learn about it in later courses, but the difference has to do with the fact that electrons obey Pauli's exclusion principle.
3. Because of the energy distribution of free electrons in a metal, the energy required by an electron to come out of the metal is different for different electrons. Electrons with higher energy require less additional energy to come out of the metal than those with lower energies. Work function is the least energy required by an electron to come out of the metal.

4. Observations on photoelectric effect imply that in the event of matter-light interaction, *absorption of energy takes place in discrete units of hν*. This is not quite the same as saying that light consists of particles, each of energy *hν*.
5. Observations on the stopping potential (its independence of intensity and dependence on frequency) are the crucial discriminator between the wave-picture and photon-picture of photoelectric effect.
6. The wavelength of a matter wave given by $\lambda = \frac{h}{p}$ has physical significance; its phase velocity v_p has no physical significance. However, the group velocity of the matter wave is physically meaningful and equals the velocity of the particle.

EXERCISES

11.1 Find the
(a) maximum frequency, and
(b) minimum wavelength of X-rays produced by 30 kV electrons.

11.2 The work function of caesium metal is 2.14 eV. When light of frequency 6×10^{14} Hz is incident on the metal surface, photoemission of electrons occurs. What is the
(a) maximum kinetic energy of the emitted electrons,
(b) Stopping potential, and
(c) maximum speed of the emitted photoelectrons?

11.3 The photoelectric cut-off voltage in a certain experiment is 1.5 V. What is the maximum kinetic energy of photoelectrons emitted?

11.4 Monochromatic light of wavelength 632.8 nm is produced by a helium-neon laser. The power emitted is 9.42 mW.
(a) Find the energy and momentum of each photon in the light beam,
(b) How many photons per second, on the average, arrive at a target irradiated by this beam? (Assume the beam to have uniform cross-section which is less than the target area), and
(c) How fast does a hydrogen atom have to travel in order to have the same momentum as that of the photon?

11.5 The energy flux of sunlight reaching the surface of the earth is 1.388×10^3 W/m^2. How many photons (nearly) per square metre are incident on the Earth per second? Assume that the photons in the sunlight have an average wavelength of 550 nm.

11.6 In an experiment on photoelectric effect, the slope of the cut-off voltage versus frequency of incident light is found to be 4.12×10^{-15} V s. Calculate the value of Planck's constant.

11.7 A 100W sodium lamp radiates energy uniformly in all directions. The lamp is located at the centre of a large sphere that absorbs all the sodium light which is incident on it. The wavelength of the sodium light is 589 nm. (a) What is the energy per photon associated

with the sodium light? (b) At what rate are the photons delivered to the sphere?

11.8 The threshold frequency for a certain metal is 3.3×10^{14} Hz. If light of frequency 8.2×10^{14} Hz is incident on the metal, predict the cut-off voltage for the photoelectric emission.

11.9 The work function for a certain metal is 4.2 eV. Will this metal give photoelectric emission for incident radiation of wavelength 330 nm?

11.10 Light of frequency 7.21×10^{14} Hz is incident on a metal surface. Electrons with a maximum speed of 6.0×10^{5} m/s are ejected from the surface. What is the threshold frequency for photoemission of electrons?

11.11 Light of wavelength 488 nm is produced by an argon laser which is used in the photoelectric effect. When light from this spectral line is incident on the emitter, the stopping (cut-off) potential of photoelectrons is 0.38 V. Find the work function of the material from which the emitter is made.

11.12 Calculate the
(a) momentum, and
(b) de Broglie wavelength of the electrons accelerated through a potential difference of 56 V.

11.13 What is the
(a) momentum,
(b) speed, and
(c) de Broglie wavelength of an electron with kinetic energy of 120 eV.

11.14 The wavelength of light from the spectral emission line of sodium is 589 nm. Find the kinetic energy at which
(a) an electron, and
(b) a neutron, would have the same de Broglie wavelength.

11.15 What is the de Broglie wavelength of
(a) a bullet of mass 0.040 kg travelling at the speed of 1.0 km/s,
(b) a ball of mass 0.060 kg moving at a speed of 1.0 m/s, and
(c) a dust particle of mass 1.0×10^{-9} kg drifting with a speed of 2.2 m/s?

11.16 An electron and a photon each have a wavelength of 1.00 nm. Find
(a) their momenta,
(b) the energy of the photon, and
(c) the kinetic energy of electron.

11.17 (a) For what kinetic energy of a neutron will the associated de Broglie wavelength be 1.40×10^{-10} m?
(b) Also find the de Broglie wavelength of a neutron, in thermal equilibrium with matter, having an average kinetic energy of $(3/2)\ kT$ at 300 K.

11.18 Show that the wavelength of electromagnetic radiation is equal to the de Broglie wavelength of its quantum (photon).

11.19 What is the de Broglie wavelength of a nitrogen molecule in air at 300 K? Assume that the molecule is moving with the root-mean-square speed of molecules at this temperature. (Atomic mass of nitrogen = 14.0076 u)

ADDITIONAL EXERCISES

11.20 (a) Estimate the speed with which electrons emitted from a heated emitter of an evacuated tube impinge on the collector maintained at a potential difference of 500 V with respect to the emitter. Ignore the small initial speeds of the electrons. The *specific charge* of the electron, i.e., its e/m is given to be 1.76×10^{11} C kg^{-1}.

(b) Use the same formula you employ in (a) to obtain electron speed for an collector potential of 10 MV. Do you see what is wrong ? In what way is the formula to be modified?

11.21 (a) A monoenergetic electron beam with electron speed of 5.20×10^{6} m s^{-1} is subject to a magnetic field of 1.30×10^{-4} T normal to the beam velocity. What is the radius of the circle traced by the beam, given e/m for electron equals 1.76×10^{11}C kg^{-1}.

(b) Is the formula you employ in (a) valid for calculating radius of the path of a 20 MeV electron beam? If not, in what way is it modified?

[**Note:** Exercises 11.20(b) and 11.21(b) take you to relativistic mechanics which is beyond the scope of this book. They have been inserted here simply to emphasise the point that the formulas you use in part (a) of the exercises are not valid at very high speeds or energies. See answers at the end to know what 'very high speed or energy' means.]

11.22 An electron gun with its collector at a potential of 100 V fires out electrons in a spherical bulb containing hydrogen gas at low pressure (~10^{-2} mm of Hg). A magnetic field of 2.83×10^{-4} T curves the path of the electrons in a circular orbit of radius 12.0 cm. (The path can be viewed because the gas ions in the path focus the beam by attracting electrons, and emitting light by electron capture; this method is known as the 'fine beam tube' method.) Determine e/m from the data.

11.23 (a) An X-ray tube produces a continuous spectrum of radiation with its short wavelength end at 0.45 Å. What is the maximum energy of a photon in the radiation?

(b) From your answer to (a), guess what order of accelerating voltage (for electrons) is required in such a tube?

11.24 In an accelerator experiment on high-energy collisions of electrons with positrons, a certain event is interpreted as annihilation of an electron-positron pair of total energy 10.2 BeV into two γ-rays of equal energy. What is the wavelength associated with each γ-ray? (1BeV = 10^{9} eV)

11.25 Estimating the following two numbers should be interesting. The first number will tell you why radio engineers do not need to worry much about photons! The second number tells you why our eye can never 'count photons', even in barely detectable light.

(a) The number of photons emitted per second by a Medium wave transmitter of 10 kW power, emitting radiowaves of wavelength 500 m.

(b) The number of photons entering the pupil of our eye per second corresponding to the minimum intensity of white light that we

humans can perceive (~10^{-10} W m^{-2}). Take the area of the pupil to be about 0.4 cm^2, and the average frequency of white light to be about 6×10^{14} Hz.

11.26 Ultraviolet light of wavelength 2271 Å from a 100 W mercury source irradiates a photo-cell made of molybdenum metal. If the stopping potential is –1.3 V, estimate the work function of the metal. How would the photo-cell respond to a high intensity (~10^5 W m^{-2}) red light of wavelength 6328 Å produced by a He-Ne laser?

11.27 Monochromatic radiation of wavelength 640.2 nm (1nm = 10^{-9} m) from a neon lamp irradiates photosensitive material made of caesium on tungsten. The stopping voltage is measured to be 0.54 V. The source is replaced by an iron source and its 427.2 nm line irradiates the same photo-cell. Predict the new stopping voltage.

11.28 A mercury lamp is a convenient source for studying frequency dependence of photoelectric emission, since it gives a number of spectral lines ranging from the UV to the red end of the visible spectrum. In our experiment with rubidium photo-cell, the following lines from a mercury source were used:

λ_1 = 3650 Å, λ_2= 4047 Å, λ_3= 4358 Å, λ_4= 5461 Å, λ_5= 6907 Å,

The stopping voltages, respectively, were measured to be:

V_{01} = 1.28 V, V_{02} = 0.95 V, V_{03} = 0.74 V, V_{04} = 0.16 V, V_{05} = 0 V

Determine the value of Planck's constant h, the threshold frequency and work function for the material.

[*Note:* You will notice that to get h from the data, you will need to know e (which you can take to be 1.6×10^{-19} C). Experiments of this kind on Na, Li, K, etc. were performed by Millikan, who, using his own value of e (from the oil-drop experiment) confirmed Einstein's photoelectric equation and at the same time gave an independent estimate of the value of h.]

11.29 The work function for the following metals is given:

Na: 2.75 eV; K: 2.30 eV; Mo: 4.17 eV; Ni: 5.15 eV. Which of these metals will not give photoelectric emission for a radiation of wavelength 3300 Å from a He-Cd laser placed 1 m away from the photocell? What happens if the laser is brought nearer and placed 50 cm away?

11.30 Light of intensity 10^{-5} W m^{-2} falls on a sodium photo-cell of surface area 2 cm^2. Assuming that the top 5 layers of sodium absorb the incident energy, estimate time required for photoelectric emission in the wave-picture of radiation. The work function for the metal is given to be about 2 eV. What is the implication of your answer?

11.31 Crystal diffraction experiments can be performed using X-rays, or electrons accelerated through appropriate voltage. Which probe has greater energy? (For quantitative comparison, take the wavelength of the probe equal to 1 Å, which is of the order of inter-atomic spacing in the lattice) (m_e=9.11 × 10^{-31} kg).

11.32 (a) Obtain the de Broglie wavelength of a neutron of kinetic energy 150 eV. As you have seen in Exercise 11.31, an electron beam of this energy is suitable for crystal diffraction experiments. Would a neutron beam of the same energy be equally suitable? Explain. (m_n = 1.675 × 10^{-27} kg)

(b) Obtain the de Broglie wavelength associated with thermal neutrons at room temperature (27 °C). Hence explain why a fast neutron beam needs to be thermalised with the environment before it can be used for neutron diffraction experiments.

11.33 An electron microscope uses electrons accelerated by a voltage of 50 kV. Determine the de Broglie wavelength associated with the electrons. If other factors (such as numerical aperture, etc.) are taken to be roughly the same, how does the resolving power of an electron microscope compare with that of an optical microscope which uses yellow light?

11.34 The wavelength of a probe is roughly a measure of the size of a structure that it can probe in some detail. The quark structure of protons and neutrons appears at the minute length-scale of 10^{-15} m or less. This structure was first probed in early 1970's using high energy electron beams produced by a linear accelerator at Stanford, USA. Guess what might have been the order of energy of these electron beams. (Rest mass energy of electron = 0.511 MeV.)

11.35 Find the typical de Broglie wavelength associated with a He atom in helium gas at room temperature (27 °C) and 1 atm pressure; and compare it with the mean separation between two atoms under these conditions.

11.36 Compute the typical de Broglie wavelength of an electron in a metal at 27 °C and compare it with the mean separation between two electrons in a metal which is given to be about 2×10^{-10} m.

[*Note:* Exercises 11.35 and 11.36 reveal that while the wave-packets associated with gaseous molecules under ordinary conditions are non-overlapping, the electron wave-packets in a metal strongly overlap with one another. This suggests that whereas molecules in an ordinary gas can be distinguished apart, electrons in a metal cannot be distintguished apart from one another. This indistinguishibility has many fundamental implications which you will explore in more advanced Physics courses.]

11.37 Answer the following questions:

(a) Quarks inside protons and neutrons are thought to carry fractional charges [$(+2/3)e$; $(-1/3)e$]. Why do they not show up in Millikan's oil-drop experiment?

(b) What is so special about the combination e/m? Why do we not simply talk of e and m separately?

(c) Why should gases be insulators at ordinary pressures and start conducting at very low pressures?

(d) Every metal has a definite work function. Why do all photoelectrons not come out with the same energy if incident radiation is monochromatic? Why is there an energy distribution of photoelectrons?

(e) The energy and momentum of an electron are related to the frequency and wavelength of the associated matter wave by the relations:

$$E = h\nu,\ p = \frac{h}{\lambda}$$

But while the value of λ is physically significant, the value of ν (and therefore, the value of the phase speed $\nu\lambda$) has no physical significance. Why?

APPENDIX

11.1 The history of wave-particle flip-flop

What is light? This question has haunted mankind for a long time. But systematic experiments were done by scientists since the dawn of the scientific and industrial era, about four centuries ago. Around the same time, theoretical models about what light is made of were developed. While building a model in any branch of science, it is essential to see that it is able to explain all the experimental observations existing at that time. It is therefore appropriate to summarize some observations about light that were known in the seventeenth century.

The properties of light known at that time included (a) rectilinear propagation of light, (b) reflection from plane and curved surfaces, (c) refraction at the boundary of two media, (d) dispersion into various colours, (e) high speed. Appropriate laws were formulated for the first four phenomena. For example, Snell formulated his laws of refraction in 1621. Several scientists right from the days of Galileo had tried to measure the speed of light. But they had not been able to do so. They had only concluded that it was higher than the limit of their measurement.

Two models of light were also proposed in the seventeenth century. Descartes, in early decades of seventeenth century, proposed that light consists of particles, while Huygens, around 1650-60, proposed that light consists of waves. Descartes' proposal was merely a philosophical model, devoid of any experiments or scientific arguments. Newton soon after, around 1660-70, extended Descartes' particle model, known as *corpuscular theory*, built it up as a scientific theory, and explained various known properties with it. These models, light as waves and as particles, in a sense, are quite opposite of each other. But both models could explain all the known properties of light. There was nothing to choose between them.

The history of the development of these models over the next few centuries is interesting. Bartholinus, in 1669, discovered double refraction of light in some crystals, and Huygens, in 1678, was quick to explain it on the basis of his wave theory of light. In spite of this, for over one hundred years, Newton's particle model was firmly believed and preferred over the wave model. This was partly because of its simplicity and partly because of Newton's influence on contemporary physics.

Then in 1801, Young performed his double-slit experiment and observed interference fringes. This phenomenon could be explained only by wave theory. It was realized that diffraction was also another phenomenon which could be explained only by wave theory. In fact, it was a natural consequence of Huygens idea of secondary wavelets emanating from every point in the path of light. These experiments could not be explained by assuming that light consists of particles. Another phenomenon of polarisation was discovered around 1810, and this too could be naturally explained by the wave theory. Thus wave theory of Huygens came to the forefront and Newton's particle theory went into the background. This situation again continued for almost a century.

Better experiments were performed in the nineteenth century to determine the speed of light. With more accurate experiments, a value of 3×10^8 m/s for speed of light in vacuum was arrived at. Around 1860, Maxwell proposed his equations of electromagnetism and it was realized that *all* electromagnetic phenomena known at that time could be explained by Maxwell's four equations. Soon Maxwell showed that electric and magnetic fields could propagate through empty space (vacuum) in the form of electromagnetic waves. He calculated the speed of these waves and arrived at a theoretical value of 2.998×10^8 m/s. The close agreement of this value with the experimental value suggested that light consists of electromagnetic waves. In 1887 Hertz demonstrated the generation and detection of such waves. This established the wave theory of light on a firm footing. We might say that while eighteenth century belonged to the particle model, the nineteenth century belonged to the wave model of light.

Vast amounts of experiments were done during the period 1850-1900 on heat and related phenomena, an altogether different area of physics. Theories and models like kinetic theory and thermodynamics were developed which quite successfully explained the various phenomena, except one.

Dual Nature of Radiation and Matter

Every body at any temperature emits radiation of all wavelengths. It also absorbs radiation falling on it. A body which absorbs all the radiation falling on it is called a *black body*. It is an ideal concept in physics, like concepts of a point mass or uniform motion. A graph of the intensity of radiation emitted by a body versus wavelength is called the *black body spectrum*. No theory in those days could explain the complete black body spectrum!

In 1900, Planck hit upon a novel idea. If we assume, he said, that radiation is emitted in packets of energy instead of continuously as in a wave, then we can explain the black body spectrum. Planck himself regarded these quanta, or packets, as a property of emission and absorption, rather than that of light. He derived a formula which agreed with the entire spectrum. This was a confusing mixture of wave and particle pictures – radiation is emitted as a particle, it travels as a wave, and is again absorbed as a particle! Moreover, this put physicists in a dilemma. Should we again accept the particle picture of light just to explain one phenomenon? Then what happens to the phenomena of interference and diffraction which cannot be explained by the particle model?

But soon in 1905, Einstein explained the photoelectric effect by assuming the particle picture of light. In 1907, Debye explained the low temperature specific heats of solids by using the particle picture for lattice vibrations in a crystalline solid. Both these phenomena belonging to widely diverse areas of physics could be explained only by the particle model and not by the wave model. In 1923, Compton's x-ray scattering experiments from atoms also went in favour of the particle picture. This increased the dilemma further.

Thus by 1923, physicists faced with the following situation. (a) There were some phenomena like rectilinear propagation, reflection, refraction, which could be explained by either particle model or by wave model. (b) There were some phenomena such as diffraction and interference which could be explained only by the wave model but *not* by the particle model. (c) There were some phenomena such as black body radiation, photoelectric effect, and Compton scattering which could be explained only by the particle model but *not* by the wave model. Somebody in those days aptly remarked that light behaves as a particle on Mondays, Wednesdays and Fridays, and as a wave on Tuesdays, Thursdays and Saturdays, and we don't talk of light on Sundays!

In 1924, de Broglie proposed his theory of wave-particle duality in which he said that not only photons of light but also 'particles' of matter such as electrons and atoms possess a dual character, sometimes behaving like a particle and sometimes as a wave. He gave a formula connecting their mass, velocity, momentum (particle characteristics), with their wavelength and frequency (wave characteristics)! In 1927 Thomson, and Davisson and Germer, in separate experiments, showed that electrons did behave like waves with a wavelength which agreed with that given by de Broglie's formula. Their experiment was on diffraction of electrons through crystalline solids, in which the regular arrangement of atoms acted like a grating. Very soon, diffraction experiments with other 'particles' such as neutrons and protons were performed and these too confirmed with de Broglie's formula. This confirmed wave-particle duality as an established principle of physics. Here was a principle, physicists thought, which explained all the phenomena mentioned above not only for light but also for the so-called particles.

But there was no basic theoretical foundation for wave-particle duality. De Broglie's proposal was merely a qualitative argument based on symmetry of nature. Wave-particle duality was at best a principle, not an outcome of a sound fundamental theory. It is true that all experiments whatever agreed with de Broglie formula. But physics does not work that way. On the one hand, it needs experimental confirmation, while on the other hand, it also needs sound theoretical basis for the models proposed. This was developed over the next two decades. Dirac developed his theory of radiation in about 1928, and Heisenberg and Pauli gave it a firm footing by 1930. Tomonaga, Schwinger, and Feynman, in late 1940s, produced further refinements and cleared the theory of inconsistencies which were noticed. All these theories mainly put wave-particle duality on a theoretical footing.

Although the story continues, it grows more and more complex and beyond the scope of this note. But we have here the essential structure of what happened, and let us be satisfied with it at the moment. Now it is regarded as a natural consequence of present theories of physics that electromagnetic radiation as well as particles of matter exhibit both wave and particle properties in different experiments, and sometimes even in the different parts of the same experiment.

Chapter Twelve

ATOMS

12.1 Introduction

By the nineteenth century, enough evidence had accumulated in favour of atomic hypothesis of matter. In 1897, the experiments on electric discharge through gases carried out by the English physicist J. J. Thomson (1856 – 1940) revealed that atoms of different elements contain negatively charged constituents (electrons) that are identical for all atoms. However, atoms on a whole are electrically neutral. Therefore, an atom must also contain some positive charge to neutralise the negative charge of the electrons. But what is the arrangement of the positive charge and the electrons inside the atom? In other words, what is the structure of an atom?

The first model of atom was proposed by J. J. Thomson in 1898. According to this model, the positive charge of the atom is uniformly distributed throughout the volume of the atom and the negatively charged electrons are embedded in it like seeds in a watermelon. This model was picturesquely called *plum pudding model* of the atom. However subsequent studies on atoms, as described in this chapter, showed that the distribution of the electrons and positive charges are very different from that proposed in this model.

We know that condensed matter (solids and liquids) and dense gases at all temperatures emit electromagnetic radiation in which a continuous distribution of several wavelengths is present, though with different intensities. This radiation is considered to be due to oscillations of atoms

and molecules, governed by the interaction of each atom or molecule with its neighbours. *In contrast*, light emitted from rarefied gases heated in a flame, or excited electrically in a glow tube such as the familiar neon sign or mercury vapour light has only certain discrete wavelengths. The spectrum appears as a series of bright lines. In such gases, the average spacing between atoms is large. Hence, the radiation emitted can be considered due to individual atoms rather than because of interactions between atoms or molecules.

In the early nineteenth century it was also established that each element is associated with a characteristic spectrum of radiation, for example, hydrogen always gives a set of lines with fixed relative position between the lines. This fact suggested an intimate relationship between the internal structure of an atom and the spectrum of radiation emitted by it. In 1885, Johann Jakob Balmer (1825 – 1898) obtained a simple empirical formula which gave the wavelengths of a group of lines emitted by atomic hydrogen. Since hydrogen is simplest of the elements known, we shall consider its spectrum in detail in this chapter.

Ernst Rutherford (1871–1937), a former research student of J. J. Thomson, was engaged in experiments on α-particles emitted by some radioactive elements. In 1906, he proposed a classic experiment of scattering of these α-particles by atoms to investigate the atomic structure. This experiment was later performed around 1911 by Hans Geiger (1882–1945) and Ernst Marsden (1889–1970, who was 20 year-old student and had not yet earned his bachelor's degree). The details are discussed in Section 12.2. The explanation of the results led to the birth of Rutherford's planetary model of atom (also called the *nuclear model of the atom*). According to this the entire positive charge and most of the mass of the atom is concentrated in a small volume called the nucleus with electrons revolving around the nucleus just as planets revolve around the sun.

Ernst Rutherford (1871 – 1937) British physicist who did pioneering work on radioactive radiation. He discovered alpha-rays and beta-rays. Along with Federick Soddy, he created the modern theory of radioactivity. He studied the 'emanation' of thorium and discovered a new noble gas, an isotope of radon, now known as thoron. By scattering alpha-rays from the metal foils, he discovered the atomic nucleus and proposed the plenatery model of the atom. He also estimated the approximate size of the nucleus.

ERNST RUTHERFORD (1871 – 1937)

Rutherford's nuclear model was a major step towards how we see the atom today. However, it could not explain why atoms emit light of only discrete wavelengths. How could an atom as simple as hydrogen, consisting of a single electron and a single proton, emit a complex spectrum of specific wavelengths? In the classical picture of an atom, the electron revolves round the nucleus much like the way a planet revolves round the sun. However, we shall see that there are some serious difficulties in accepting such a model.

12.2 Alpha-particle Scattering and Rutherford's Nuclear Model of Atom

At the suggestion of Ernst Rutherford, in 1911, H. Geiger and E. Marsden performed some experiments. In one of their experiments, as shown in

FIGURE 12.1 Geiger-Marsden scattering experiment. The entire apparatus is placed in a vacuum chamber (not shown in this figure).

Fig. 12.1, they directed a beam of 5.5 MeV α-particles emitted from a $^{214}_{83}$Bi radioactive source at a thin metal foil made of gold. Figure 12.2 shows a schematic diagram of this experiment. Alpha-particles emitted by a $^{214}_{83}$Bi radioactive source were collimated into a narrow beam by their passage through lead bricks. The beam was allowed to fall on a thin foil of gold of thickness 2.1×10^{-7} m. The scattered alpha-particles were observed through a rotatable detector consisting of zinc sulphide screen and a microscope. The scattered alpha-particles on striking the screen produced brief light flashes or scintillations. These flashes may be viewed through a microscope and the distribution of the number of scattered particles may be studied as a function of angle of scattering.

FIGURE 12.2 Schematic arrangement of the Geiger-Marsden experiment.

A typical graph of the total number of α-particles scattered at different angles, in a given interval of time, is shown in Fig. 12.3. The dots in this figure represent the data points and the solid curve is the theoretical prediction based on the assumption that the target atom has a small, dense, positively charged nucleus. Many of the α-particles pass through the foil. It means that they do not suffer any collisions. Only about 0.14% of the incident α-particles scatter by more than 1°; and about 1 in 8000 deflect by more than 90°. Rutherford argued that, to deflect the α-particle backwards, it must experience a large repulsive force. This force could

be provided if the greater part of the mass of the atom and its positive charge were concentrated tightly at its centre. Then the incoming α-particle could get very close to the positive charge without penetrating it, and such a close encounter would result in a large deflection. This agreement supported the hypothesis of the nuclear atom. This is why Rutherford is credited with the *discovery* of the nucleus.

FIGURE 12.3 The dots are α-particle scattering data for a thin gold foil obtained by Geiger and Marsden using the experiment shown in Figs. 12.1 and 12.2. The solid curve is the theoretical prediction based on the assumption that atom has a small, dense, positively charged nucleus.

In Rutherford's nuclear model of the atom, the entire positive charge and most of the mass of the atom are concentrated in the nucleus with the electrons some distance away. The electrons would be moving in orbits about the nucleus just as the planets do around the sun. Rutherford's experiments suggested the size of the nucleus to be about 10^{-15} m to 10^{-14} m. From kinetic theory, the size of an atom was known to be 10^{-10} m, about 10,000 to 100,000 times larger than the size of the nucleus (see Chapter 11, Section 11.6 in Class XI Physics textbook). Thus, the electrons would seem to be at a distance from the nucleus of about 10,000 to 100,000 times the size of the nucleus itself. Thus, most of an atom is empty space. With the atom being largely empty space, it is easy to see why most α-particles go right through a thin metal foil. However, when α-particle happens to come near a nucleus, the intense electric field there scatters it through a large angle. The atomic electrons, being so light, do not appreciably affect the α-particles.

The scattering data shown in Fig. 12.3 can be analysed by employing Rutherford's nuclear model of the atom. As the gold foil is very thin, it can be assumed that α-particles will suffer not more than one scattering during their passage through it. Therefore, computation of the trajectory of an alpha-particle scattered by a single nucleus is enough. Alpha-particles are nuclei of helium atoms and, therefore, carry two units, $2e$, of positive charge and have the mass of the helium atom. The charge of the gold nucleus is Ze, where Z is the atomic number of the atom; for gold $Z = 79$. Since the nucleus of gold is about 50 times heavier than an α-particle, it is reasonable to assume that it remains stationary throughout the scattering process. Under these assumptions, the trajectory of an alpha-particle can be computed employing Newton's second law of motion and the Coulomb's law for electrostatic force of repulsion between the alpha-particle and the positively charged nucleus.

The magnitude of this force is

$$F = \frac{1}{4\pi\varepsilon_0}\frac{(2e)(Ze)}{r^2} \tag{12.1}$$

where r is the distance between the α-particle and the nucleus. The force is directed along the line joining the α-particle and the nucleus. The magnitude and direction of the force on an α-particle continuously changes as it approaches the nucleus and recedes away from it.

12.2.1 Alpha-particle trajectory

The trajectory traced by an α-particle depends on the impact parameter, b of collision. The *impact parameter* is the perpendicular distance of the initial velocity vector of the α-particle from the centre of the nucleus (Fig. 12.4). A given beam of α-particles has a distribution of impact parameters b, so that the beam is scattered in various directions with different probabilities (Fig. 12.4). (In a beam, all particles have nearly same kinetic energy.) It is seen that an α-particle close to the nucleus (small impact parameter) suffers large scattering. In case of head-on collision, the impact parameter is minimum and the α-particle rebounds back ($\theta \cong \pi$). For a large impact parameter, the α-particle goes nearly undeviated and has a small deflection ($\theta \cong 0$).

FIGURE 12.4 Trajectory of α-particles in the coulomb field of a target nucleus. The impact parameter, b and scattering angle θ are also depicted.

The fact that only a small fraction of the number of incident particles rebound back indicates that the number of α-particles undergoing head on collision is small. This, in turn, implies that the mass of the atom is concentrated in a small volume. Rutherford scattering therefore, is a powerful way to determine an upper limit to the size of the nucleus.

EXAMPLE 12.1

Example 12.1 In the Rutherford's nuclear model of the atom, the nucleus (radius about 10^{-15} m) is analogous to the sun about which the electron move in orbit (radius ≈ 10^{-10} m) like the earth orbits around the sun. If the dimensions of the solar system had the same proportions as those of the atom, would the earth be closer to or farther away from the sun than actually it is? The radius of earth's orbit is about 1.5×10^{11} m. The radius of sun is taken as 7×10^8 m.

Solution The ratio of the radius of electron's orbit to the radius of nucleus is $(10^{-10}\text{ m})/(10^{-15}\text{ m}) = 10^5$, that is, the radius of the electron's orbit is 10^5 times larger than the radius of nucleus. If the radius of the earth's orbit around the sun were 10^5 times larger than the radius of the sun, the radius of the earth's orbit would be $10^5 \times 7 \times 10^8$ m = 7×10^{13} m. This is more than 100 times greater than the actual orbital radius of earth. Thus, the earth would be much farther away from the sun.

It implies that an atom contains a much greater fraction of empty space than our solar system does.

Example 12.2 In a Geiger-Marsden experiment, what is the distance of closest approach to the nucleus of a 7.7 MeV α-particle before it comes momentarily to rest and reverses its direction?

Solution The key idea here is that throughout the scattering process, the total mechanical energy of the system consisting of an α-particle and a gold nucleus is conserved. The system's initial mechanical energy is E_i, before the particle and nucleus interact, and it is equal to its mechanical energy E_f when the α-particle momentarily stops. The initial energy E_i is just the kinetic energy K of the incoming α- particle. The final energy E_f is just the electric potential energy U of the system. The potential energy U can be calculated from Eq. (12.1).

Let d be the centre-to-centre distance between the α-particle and the gold nucleus when the α-particle is at its stopping point. Then we can write the conservation of energy $E_i = E_f$ as

$$K = \frac{1}{4\pi\varepsilon_0}\frac{(2e)(Ze)}{d} = \frac{2Ze^2}{4\pi\varepsilon_0 d}$$

Thus the distance of closest approach d is given by

$$d = \frac{2Ze^2}{4\pi\varepsilon_0 K}$$

The maximum kinetic energy found in α-particles of natural origin is 7.7 MeV or 1.2×10^{-12} J. Since $1/4\pi\varepsilon_0 = 9.0 \times 10^9$ N m^2/C^2. Therefore with $e = 1.6 \times 10^{-19}$ C, we have,

$$d = \frac{(2)(9.0\times10^9\,\text{Nm}^2/C^2)(1.6\times10^{-19}C)^2 Z}{1.2\times10^{-12}\,\text{J}}$$

$$= 3.84 \times 10^{-16}\, Z\,\text{m}$$

The atomic number of foil material gold is $Z = 79$, so that

d (Au) = 3.0×10^{-14} m = 30 fm. (1 fm (i.e. fermi) = 10^{-15} m.)

The radius of gold nucleus is, therefore, less than 3.0×10^{-14} m. This is not in very good agreement with the observed result as the actual radius of gold nucleus is 6 fm. The cause of discrepancy is that the distance of closest approach is considerably larger than the sum of the radii of the gold nucleus and the α-particle. Thus, the α-particle reverses its motion without ever actually *touching* the gold nucleus.

EXAMPLE 12.2

PHYSICS

Simulate Rutherford scattering experiment
http://www outreach.phy.cam.ac.uk/camphy/nucleus/nucleus_exp.htm

12.2.2 Electron orbits

The Rutherford nuclear model of the atom which involves classical concepts, pictures the atom as an electrically neutral sphere consisting of a very small, massive and positively charged nucleus at the centre surrounded by the revolving electrons in their respective dynamically stable orbits. The electrostatic force of attraction, F_e between the revolving electrons and the nucleus provides the requisite centripetal force (F_c) to keep them in their orbits. Thus, for a dynamically stable orbit in a hydrogen atom

$$F_e = F_c$$

$$\frac{mv^2}{r} = \frac{1}{4\pi\varepsilon_0}\frac{e^2}{r^2} \qquad (12.2)$$

Thus the relation between the orbit radius and the electron velocity is

$$r = \frac{e^2}{4\pi\varepsilon_0 m v^2} \tag{12.3}$$

The kinetic energy (K) and electrostatic potential energy (U) of the electron in hydrogen atom are

$$K = \frac{1}{2}mv^2 = \frac{e^2}{8\pi\varepsilon_0 r} \text{ and } U = -\frac{e^2}{4\pi\varepsilon_0 r}$$

(The negative sign in U signifies that the electrostatic force is in the $-r$ direction.) Thus the total energy E of the electron in a hydrogen atom is

$$E = K + U = \frac{e^2}{8\pi\varepsilon_0 r} - \frac{e^2}{4\pi\varepsilon_0 r}$$

$$= -\frac{e^2}{8\pi\varepsilon_0 r} \tag{12.4}$$

The total energy of the electron is negative. This implies the fact that the electron is bound to the nucleus. If E were positive, an electron will not follow a closed orbit around the nucleus.

EXAMPLE 12.3

Example 12.3 It is found experimentally that 13.6 eV energy is required to separate a hydrogen atom into a proton and an electron. Compute the orbital radius and the velocity of the electron in a hydrogen atom.

Solution Total energy of the electron in hydrogen atom is –13.6 eV = $-13.6 \times 1.6 \times 10^{-19}$ J = -2.2×10^{-18} J. Thus from Eq. (12.4), we have

$$-\frac{e^2}{8\pi\varepsilon_0 r} = -2.2 \times 10^{-18} \text{ J}$$

This gives the orbital radius

$$r = -\frac{e^2}{8\pi\varepsilon_0 E} = -\frac{(9 \times 10^9 \text{ N m}^2/\text{C}^2)(1.6 \times 10^{-19}\text{C})^2}{(2)(-2.2 \times 10^{-18} \text{ J})}$$

$= 5.3 \times 10^{-11}$ m.

The velocity of the revolving electron can be computed from Eq. (12.3) with $m = 9.1 \times 10^{-31}$ kg,

$$v = \frac{e}{\sqrt{4\pi\varepsilon_0 m r}} = 2.2 \times 10^6 \text{ m/s.}$$

12.3 Atomic Spectra

As mentioned in Section 12.1, each element has a characteristic spectrum of radiation, which it emits. When an atomic gas or vapour is excited at low pressure, usually by passing an electric current through it, the emitted radiation has a spectrum which contains certain specific wavelengths only. A spectrum of this kind is termed as emission line spectrum and it

consists of bright lines on a dark background. The spectrum emitted by atomic hydrogen is shown in Fig. 12.5. Study of emission line spectra of a material can therefore serve as a type of "fingerprint" for identification of the gas. When white light passes through a gas and we analyse the transmitted light using a spectrometer we find some dark lines in the spectrum. These dark lines correspond precisely to those wavelengths which were found in the emission line spectrum of the gas. This is called the *absorption spectrum* of the material of the gas.

FIGURE 12.5 Emission lines in the spectrum of hydrogen.

12.3.1 Spectral series

We might expect that the frequencies of the light emitted by a particular element would exhibit some regular pattern. Hydrogen is the simplest atom and therefore, has the simplest spectrum. In the observed spectrum, however, at first sight, there does not seem to be any resemblance of order or regularity in spectral lines. But the spacing between lines within certain sets of the hydrogen spectrum decreases in a regular way (Fig. 12.5). Each of these sets is called a *spectral series*. In 1885, the first such series was observed by a Swedish school teacher Johann Jakob Balmer (1825–1898) in the visible region of the hydrogen spectrum. This series is called *Balmer series* (Fig. 12.6). The line with the longest wavelength, 656.3 nm in the red is called H_α; the next line with wavelength 486.1 nm in the blue-green is called H_β, the third line 434.1 nm in the violet is called H_γ; and so on. As the wavelength decreases, the lines appear closer together and are weaker in intensity. Balmer found a simple empirical formula for the observed wavelengths

FIGURE 12.6 Balmer series in the emission spectrum of hydrogen.

$$\frac{1}{\lambda} = R\left(\frac{1}{2^2} - \frac{1}{n^2}\right) \qquad (12.5)$$

where λ is the wavelength, R is a constant called the *Rydberg constant*, and n may have integral values 3, 4, 5, etc. The value of R is $1.097 \times 10^7\ \text{m}^{-1}$. This equation is also called Balmer formula.

Taking $n = 3$ in Eq. (12.5), one obtains the wavelength of the H_α line:

$$\frac{1}{\lambda} = 1.097 \times 10^7 \left(\frac{1}{2^2} - \frac{1}{3^2}\right)\ \text{m}^{-1}$$

$$= 1.522 \times 10^6\ \text{m}^{-1}$$

i.e., $\lambda = 656.3$ nm

For $n = 4$, one obtains the wavelength of H_β line, etc. For $n = \infty$, one obtains the limit of the series, at $\lambda = 364.6$ nm. This is the shortest wavelength in the Balmer series. Beyond this limit, no further distinct lines appear, instead only a faint continuous spectrum is seen.

Other series of spectra for hydrogen were subsequently discovered. These are known, after their discoverers, as Lyman, Paschen, Brackett, and Pfund series. These are represented by the formulae:

Lyman series:

$$\frac{1}{\lambda} = R\left(\frac{1}{1^2} - \frac{1}{n^2}\right) \qquad n = 2,3,4... \tag{12.6}$$

Paschen series:

$$\frac{1}{\lambda} = R\left(\frac{1}{3^2} - \frac{1}{n^2}\right) \qquad n = 4,5,6... \tag{12.7}$$

Brackett series:

$$\frac{1}{\lambda} = R\left(\frac{1}{4^2} - \frac{1}{n^2}\right) \qquad n = 5,6,7... \tag{12.8}$$

Pfund series:

$$\frac{1}{\lambda} = R\left(\frac{1}{5^2} - \frac{1}{n^2}\right) \qquad n = 6,7,8... \tag{12.9}$$

The Lyman series is in the ultraviolet, and the Paschen and Brackett series are in the infrared region.

The Balmer formula Eq. (12.5) may be written in terms of frequency of the light, recalling that

$$c = \nu\lambda$$

or $\frac{1}{\lambda} = \frac{\nu}{c}$

Thus, Eq. (12.5) becomes

$$\nu = Rc\left(\frac{1}{2^2} - \frac{1}{n^2}\right) \tag{12.10}$$

There are only a few elements (hydrogen, singly ionised helium, and doubly ionised lithium) whose spectra can be represented by simple formula like Eqs. (12.5) – (12.9).

Equations (12.5) – (12.9) are useful as they give the wavelengths that hydrogen atoms radiate or absorb. However, these results are empirical and do not give any reasoning why only certain frequencies are observed in the hydrogen spectrum.

12.4 Bohr Model of the Hydrogen Atom

The model of the atom proposed by Rutherford assumes that the atom, consisting of a central nucleus and revolving electron is stable much like sun-planet system which the model imitates. However, there are some fundamental differences between the two situations. While the planetary system is held by gravitational force, the nucleus-electron system being charged objects, interact by Coulomb's Law of force. We know that an

object which moves in a circle is being constantly accelerated – the acceleration being centripetal in nature. According to classical electromagnetic theory, an accelerating charged particle emits radiation in the form of electromagnetic waves. The energy of an accelerating electron should therefore, continuously decrease. The electron would spiral inward and eventually fall into the nucleus (Fig. 12.7). Thus, such an atom can not be stable. Further, according to the classical electromagnetic theory, the frequency of the electromagnetic waves emitted by the revolving electrons is equal to the frequency of revolution. As the electrons spiral inwards, their angular velocities and hence their frequencies would change continuously, and so will the frequency of the light emitted. Thus, they would emit a continuous spectrum, in contradiction to the line spectrum actually observed. Clearly Rutherford model tells only a part of the story implying that the classical ideas are not sufficient to explain the atomic structure.

Niels Henrik David Bohr (1885 – 1962) Danish physicist who explained the spectrum of hydrogen atom based on quantum ideas. He gave a theory of nuclear fission based on the liquid-drop model of nucleus. Bohr contributed to the clarification of conceptual problems in quantum mechanics, in particular by proposing the complementary principle.

NIELS HENRIK DAVID BOHR (1885 – 1962)

FIGURE 12.7 An accelerated atomic electron must spiral into the nucleus as it loses energy.

Example 12.4 According to the classical electromagnetic theory, calculate the initial frequency of the light emitted by the electron revolving around a proton in hydrogen atom.

Solution From Example 12.3 we know that velocity of electron moving around a proton in hydrogen atom in an orbit of radius 5.3×10^{-11} m is 2.2×10^{-6} m/s. Thus, the frequency of the electron moving around the proton is

$$\nu = \frac{v}{2\pi r} = \frac{2.2 \times 10^{6}\ \text{m s}^{-1}}{2\pi\left(5.3 \times 10^{-11}\ \text{m}\right)}$$

$$\approx 6.6 \times 10^{15}\ \text{Hz}.$$

According to the classical electromagnetic theory we know that the frequency of the electromagnetic waves emitted by the revolving electrons is equal to the frequency of its revolution around the nucleus. Thus the initial frequency of the light emitted is 6.6×10^{15} Hz.

EXAMPLE 12.4

It was Niels Bohr (1885 – 1962) who made certain modifications in this model by adding the ideas of the newly developing quantum hypothesis. Niels Bohr studied in Rutherford's laboratory for several months in 1912 and he was convinced about the validity of Rutherford nuclear model. Faced with the dilemma as discussed above, Bohr, in 1913, concluded that in spite of the success of electromagnetic theory in explaining large-scale phenomena, it could not be applied to the processes at the atomic scale. It became clear that a fairly radical departure from the established principles of classical mechanics and electromagnetism would be needed to understand the structure of atoms and the relation of atomic structure to atomic spectra. Bohr combined classical and early quantum concepts and gave his theory in the form of three postulates. These are :

(i) Bohr's first postulate was that *an electron in an atom could revolve in certain stable orbits without the emission of radiant energy*, contrary to the predictions of electromagnetic theory. According to this postulate, each atom has certain definite stable states in which it can exist, and each possible state has definite total energy. These are called the stationary states of the atom.

(ii) Bohr's second postulate defines these stable orbits. This postulate states that the *electron* revolves around the nucleus *only in those orbits for which the angular momentum is some integral multiple of* $h/2\pi$ where h is the Planck's constant (= 6.6×10^{-34} J s). Thus the angular momentum (L) of the orbiting electron is quantised. That is

$$L = nh/2\pi \tag{12.11}$$

(iii) Bohr's third postulate incorporated into atomic theory the early quantum concepts that had been developed by Planck and Einstein. It states that *an electron might make a transition from one of its specified non-radiating orbits to another of lower energy. When it does so, a photon is emitted having energy equal to the energy difference between the initial and final states. The frequency of the emitted photon is then given by*

$$h\nu = E_i - E_f \tag{12.12}$$

where E_i and E_f are the energies of the initial and final states and $E_i > E_f$.

For a hydrogen atom, Eq. (12.4) gives the expression to determine the energies of different energy states. But then this equation requires the radius r of the electron orbit. To calculate r, Bohr's second postulate about the angular momentum of the electron–the quantisation condition – is used. The angular momentum L is given by

$$L = mvr$$

Bohr's second postulate of quantisation [Eq. (12.11)] says that the allowed values of angular momentum are integral multiples of $h/2\pi$.

$$L_n = mv_n r_n = \frac{nh}{2\pi} \tag{12.13}$$

where n is an integer, r_n is the radius of nth possible orbit and v_n is the speed of moving electron in the nth orbit. The allowed orbits are numbered

1, 2, 3 ..., according to the values of n, which is called the *principal quantum number* of the orbit.

From Eq. (12.3), the relation between v_n and r_n is

$$v_n = \frac{e}{\sqrt{4\pi\varepsilon_0 m r_n}}$$

Combining it with Eq. (12.13), we get the following expressions for v_n and r_n,

$$v_n = \frac{1}{n}\frac{e^2}{4\pi\varepsilon_0}\frac{1}{(h/2\pi)} \tag{12.14}$$

and

$$r_n = \left(\frac{n^2}{m}\right)\left(\frac{h}{2\pi}\right)^2 \frac{4\pi\varepsilon_0}{e^2} \tag{12.15}$$

Eq. (12.14) depicts that the orbital speed in the nth orbit falls by a factor of n. Using Eq. (12.15), the size of the innermost orbit (n = 1) can be obtained as

$$r_1 = \frac{h^2\varepsilon_0}{\pi m e^2}$$

This is called the *Bohr radius*, represented by the symbol a_0. Thus,

$$a_0 = \frac{h^2\varepsilon_0}{\pi m e^2} \tag{12.16}$$

Substitution of values of h, m, ε_0 and e gives $a_0 = 5.29 \times 10^{-11}$ m. From Eq. (12.15), it can also be seen that the radii of the orbits increase as n^2.

The total energy of the electron in the stationary states of the hydrogen atom can be obtained by substituting the value of orbital radius in Eq. (12.4) as

$$E_n = -\left(\frac{e^2}{8\pi\varepsilon_0}\right)\left(\frac{m}{n^2}\right)\left(\frac{2\pi}{h}\right)^2\left(\frac{e^2}{4\pi\varepsilon_0}\right)$$

or $$E_n = -\frac{me^4}{8n^2\varepsilon_0^2 h^2} \tag{12.17}$$

Substituting values, Eq. (12.17) yields

$$E_n = -\frac{2.18 \times 10^{-18}}{n^2}\text{J} \tag{12.18}$$

Atomic energies are often expressed in electron volts (eV) rather than joules. Since 1 eV = 1.6×10^{-19} J, Eq. (12.18) can be rewritten as

$$E_n = -\frac{13.6}{n^2}\ \text{eV} \tag{12.19}$$

The negative sign of the total energy of an electron moving in an orbit means that the electron is bound with the nucleus. Energy will thus be required to remove the electron from the hydrogen atom to a distance infinitely far away from its nucleus (or proton in hydrogen atom).

The derivation of Eqs. (12.17) – (12.19) involves the assumption that the electronic orbits are circular, though orbits under inverse square force are, in general elliptical. (Planets move in elliptical orbits under the inverse square gravitational force of the sun.) However, it was shown by the German physicist Arnold Sommerfeld (1868 – 1951) that, when the restriction of circular orbit is relaxed, these equations continue to hold even for elliptic orbits.

Orbit vs state (orbital picture) of electron in atom

We are introduced to the Bohr Model of atom one time or the other in the course of physics. This model has its place in the history of quantum mechanics and particularly in explaining the structure of an atom. It has become a milestone since Bohr introduced the revolutionary idea of definite energy orbits for the electrons, contrary to the classical picture requiring an accelerating particle to radiate. Bohr also introduced the idea of quantisation of angular momentum of electrons moving in definite orbits. Thus it was a semi-classical picture of the structure of atom.

Now with the development of quantum mechanics, we have a better understanding of the structure of atom. Solutions of the Schrödinger wave equation assign a wave-like description to the electrons bound in an atom due to attractive forces of the protons.

An orbit of the electron in the Bohr model is the circular path of motion of an electron around the nucleus. But according to quantum mechanics, we cannot associate a definite path with the motion of the electrons in an atom. We can only talk about the probability of finding an electron in a certain region of space around the nucleus. This probability can be inferred from the one-electron wave function called the *orbital*. This function depends only on the coordinates of the electron.

It is therefore essential that we understand the subtle differences that exist in the two models:

- Bohr model is valid for only one-electron atoms/ions; an energy value, assigned to each orbit, depends on the principal quantum number n in this model. We know that energy associated with a stationary state of an electron depends on n only, for one-electron atoms/ions. For a multi-electron atom/ion, this is not true.
- The solution of the Schrödinger wave equation, obtained for hydrogen-like atoms/ions, called the wave function, gives information about the probability of finding an electron in various regions around the nucleus. This *orbital* has no resemblance whatsoever with the *orbit* defined for an electron in the Bohr model.

Example 12.5

Example 12.5 A 10 kg satellite circles earth once every 2 h in an orbit having a radius of 8000 km. Assuming that Bohr's angular momentum postulate applies to satellites just as it does to an electron in the hydrogen atom, find the quantum number of the orbit of the satellite.

Solution

From Eq. (12.13), we have

$m\, v_n\, r_n = nh/2\pi$

EXAMPLE 12.5

Here m = 10 kg and $r_n = 8 \times 10^6$ m. We have the time period T of the circling satellite as 2 h. That is T = 7200 s.
Thus the velocity $v_n = 2\pi\, r_n/T$.
The quantum number of the orbit of satellite
$n = (2\pi\, r_n)^2 \times m/(T \times h)$.
Substituting the values,
$n = (2\pi \times 8 \times 10^6 \text{ m})^2 \times 10/(7200 \text{ s} \times 6.64 \times 10^{-34} \text{ J s})$
$= 5.3 \times 10^{45}$
Note that the quantum number for the satellite motion is extremely large! In fact for such large quantum numbers the results of quantisation conditions tend to those of classical physics.

12.4.1 Energy levels

The energy of an atom is the *least* (largest negative value) when its electron is revolving in an orbit closest to the nucleus i.e., the one for which n = 1. For n = 2, 3, ... the absolute value of the energy E is smaller, hence the energy is progressively larger in the outer orbits. The *lowest* state of the atom, called the *ground state*, is that of the lowest energy, with the electron revolving in the orbit of smallest radius, the Bohr radius, a_0. The energy of this state (n = 1), E_1 is –13.6 eV. Therefore, the minimum energy required to free the electron from the ground state of the hydrogen atom is 13.6 eV. It is called the *ionisation energy* of the hydrogen atom. This prediction of the Bohr's model is in excellent agreement with the experimental value of ionisation energy.

At room temperature, most of the hydrogen atoms are in *ground state*. When a hydrogen atom receives energy by processes such as electron collisions, the atom may acquire sufficient energy to raise the electron to higher energy states. The atom is then said to be in an *excited state*. From Eq. (12.19), for n = 2; the energy E_2 is –3.40 eV. It means that the energy required to excite an electron in hydrogen atom to its first excited state, is an energy equal to $E_2 - E_1$ = –3.40 eV – (–13.6) eV = 10.2 eV. Similarly, E_3 = –1.51 eV and $E_3 - E_1$ = 12.09 eV, or to excite the hydrogen atom from its ground state (n = 1) to second excited state (n = 3), 12.09 eV energy is required, and so on. From these excited states the electron can then fall back to a state of lower energy, emitting a photon in the process. Thus, as the excitation of hydrogen atom increases (that is as n increases) the value of minimum energy required to free the electron from the excited atom decreases.

The energy level diagram* for the stationary states of a hydrogen atom, computed from Eq. (12.19), is given in

FIGURE 12.8 The energy level diagram for the hydrogen atom. The electron in a hydrogen atom at room temperature spends most of its time in the ground state. To ionise a hydrogen atom an electron from the ground state, 13.6 eV of energy must be supplied. (The horizontal lines specify the presence of allowed energy states.)

* An electron can have any total energy above E = 0 eV. In such situations the electron is free. Thus there is a continuum of energy states above E = 0 eV, as shown in Fig. 12.8.

Fig. 12.8. The principal quantum number n labels the stationary states in the ascending order of energy. In this diagram, the highest energy state corresponds to $n = \infty$ in Eq, (12.19) and has an energy of 0 eV. This is the energy of the atom when the electron is completely removed ($r = \infty$) from the nucleus and is at rest. Observe how the energies of the excited states come closer and closer together as n increases.

Franck – Hertz experiment

The existence of discrete energy levels in an atom was directly verified in 1914 by James Franck and Gustav Hertz. They studied the spectrum of mercury vapour when electrons having different kinetic energies passed through the vapour. The electron energy was varied by subjecting the electrons to electric fields of varying strength. The electrons collide with the mercury atoms and can transfer energy to the mercury atoms. This can only happen when the energy of the electron is higher than the energy difference between an energy level of Hg occupied by an electron and a higher unoccupied level (see Figure). For instance, the difference between an occupied energy level of Hg and a higher unoccupied level is 4.9 eV. If an electron of having an energy of 4.9 eV or more passes through mercury, an electron in mercury atom can absorb energy from the bombarding electron and get excited to the higher level [Fig (a)]. The colliding electron's kinetic energy would reduce by this amount.

(a) (b)

The excited electron would subsequently fall back to the ground state by emission of radiation [Fig. (b)]. The wavelength of emitted radiation is:

$$\lambda = \frac{hc}{E} = \frac{6.625 \times 10^{-34} \times 3 \times 10^{8}}{4.9 \times 1.6 \times 10^{-19}} = 253 \text{ nm}$$

By direct measurement, Franck and Hertz found that the emission spectrum of mercury has a line corresponding to this wavelength. For this experimental verification of Bohr's basic ideas of discrete energy levels in atoms and the process of photon emission, Frank and Hertz were awarded the Nobel prize in 1925.

12.5 The Line Spectra of the Hydrogen Atom

According to the third postulate of Bohr's model, when an atom makes a transition from the higher energy state with quantum number n_i to the lower energy state with quantum number n_f ($n_f < n_i$), the difference of energy is carried away by a photon of frequency ν_{if} such that

$$h\nu_{if} = E_{n_i} - E_{n_f} \tag{12.20}$$

Using Eq. (12.16), for E_{n_f} and E_{n_i}, we get

$$h\nu_{if} = \frac{me^4}{8\varepsilon_0^2 h^2}\left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right) \tag{12.21}$$

$$\text{or } \nu_{if} = \frac{me^4}{8\varepsilon_0^2 h^3}\left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right) \tag{12.22}$$

Equation (12.21) is the Rydberg formula, for the spectrum of the hydrogen atom. In this relation, if we take $n_f = 2$ and $n_i = 3, 4, 5...$, it reduces to a form similar to Eq. (12.10) for the Balmer series. The Rydberg constant R is readily identified to be

$$R = \frac{me^4}{8\varepsilon_0^2 h^3 c} \tag{12.23}$$

If we insert the values of various constants in Eq. (12.23), we get

$R = 1.03 \times 10^7 \text{ m}^{-1}$

This is a value very close to the value ($1.097 \times 10^7 \text{ m}^{-1}$) obtained from the empirical Balmer formula. This agreement between the theoretical and experimental values of the Rydberg constant provided a direct and striking confirmation of the Bohr's model.

Since both n_f and n_i are integers, this immediately shows that in transitions between different atomic levels, light is radiated in various discrete frequencies. For hydrogen spectrum, the Balmer formula corresponds to $n_f = 2$ and $n_i = 3, 4, 5$, etc. The results of the Bohr's model suggested the presence of other series spectra for hydrogen atom–those corresponding to transitions resulting from $n_f = 1$ and $n_i = 2, 3$, etc.; $n_f = 3$ and $n_i = 4, 5$, etc., and so on. Such series were identified in the course of spectroscopic investigations and are known as the Lyman, Balmer, Paschen, Brackett, and Pfund series. The electronic transitions corresponding to these series are shown in Fig. 12.9.

FIGURE 12.9 Line spectra originate in transitions between energy levels.

The various lines in the atomic spectra are produced when electrons jump from higher energy state to a lower energy state and photons are emitted. These spectral lines are called emission lines. But when an atom absorbs a photon that has precisely

the same energy needed by the electron in a lower energy state to make transitions to a higher energy state, the process is called absorption. Thus if photons with a continuous range of frequencies pass through a rarefied gas and then are analysed with a spectrometer, a series of dark spectral absorption lines appear in the continuous spectrum. The dark lines indicate the frequencies that have been absorbed by the atoms of the gas.

The explanation of the hydrogen atom spectrum provided by Bohr's model was a brilliant achievement, which greatly stimulated progress towards the modern quantum theory. In 1922, Bohr was awarded Nobel Prize in Physics.

Example 12.6 Using the Rydberg formula, calculate the wavelengths of the first four spectral lines in the Lyman series of the hydrogen spectrum.

Solution The Rydberg formula is

$$hc/\lambda_{if} = \frac{me^4}{8\varepsilon_0^2 h^2}\left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right)$$

The wavelengths of the first four lines in the Lyman series correspond to transitions from n_i = 2,3,4,5 to n_f = 1. We know that

$$\frac{me^4}{8\varepsilon_0^2 h^2} = 13.6 \text{ eV} = 21.76 \times 10^{-19} \text{ J}$$

Therefore,

$$\lambda_{i1} = \frac{hc}{21.76 \times 10^{-19}\left(\frac{1}{1} - \frac{1}{n_i^2}\right)} \text{ m}$$

$$= \frac{6.625 \times 10^{-34} \times 3 \times 10^8 \times n_i^2}{21.76 \times 10^{-19} \times (n_i^2 - 1)} \text{ m} = \frac{0.9134\ n_i^2}{(n_i^2 - 1)} \times 10^{-7} \text{ m}$$

$$= 913.4\ n_i^2/(n_i^2 - 1) \text{ Å}$$

Substituting n_i = 2,3,4,5, we get λ_{21} = 1218 Å, λ_{31} = 1028 Å, λ_{41} = 974.3 Å, and λ_{51} = 951.4 Å.

EXAMPLE 12.6

12.6 De Broglie's Explanation of Bohr's Second Postulate of Quantisation

Of all the postulates, Bohr made in his model of the atom, perhaps the most puzzling is his second postulate. It states that the angular momentum of the electron orbiting around the nucleus is quantised (that is, $L_n = nh/2\pi$; n = 1, 2, 3 ...). Why should the angular momentum have only those values that are integral multiples of $h/2\pi$? The French physicist Louis de Broglie explained this puzzle in 1923, ten years after Bohr proposed his model.

We studied, in Chapter 11, about the de Broglie's hypothesis that material particles, such as electrons, also have a wave nature. C. J. Davisson and L. H. Germer later experimentally verified the wave nature of electrons

in 1927. Louis de Broglie argued that the electron in its circular orbit, as proposed by Bohr, must be seen as a particle wave. In analogy to waves travelling on a string, particle waves too can lead to standing waves under resonant conditions. From Chapter 15 of Class XI Physics textbook, we know that when a string is plucked, a vast number of wavelengths are excited. However only those wavelengths survive which have nodes at the ends and form the standing wave in the string. It means that in a string, standing waves are formed when the total distance travelled by a wave down the string and back is one wavelength, two wavelengths, or any integral number of wavelengths. Waves with other wavelengths interfere with themselves upon reflection and their amplitudes quickly drop to zero. For an electron moving in n^{th} circular orbit of radius r_n, the total distance is the circumference of the orbit, $2\pi r_n$. Thus

$$2\pi\, r_n = n\lambda, \qquad n = 1, 2, 3... \tag{12.24}$$

FIGURE 12.10 A standing wave is shown on a circular orbit where four de Broglie wavelengths fit into the circumference of the orbit.

Figure 12.10 illustrates a standing particle wave on a circular orbit for $n = 4$, i.e., $2\pi r_n = 4\lambda$, where λ is the de Broglie wavelength of the electron moving in n^{th} orbit. From Chapter 11, we have $\lambda = h/p$, where p is the magnitude of the electron's momentum. If the speed of the electron is much less than the speed of light, the momentum is mv_n. Thus, $\lambda = h/mv_n$. From Eq. (12.24), we have

$$2\pi\, r_n = n\, h/mv_n \qquad \text{or} \qquad m\, v_n\, r_n = nh/2\pi$$

This is the quantum condition proposed by Bohr for the angular momentum of the electron [Eq. (12.13)]. In Section 12.5, we saw that this equation is the basis of explaining the discrete orbits and energy levels in hydrogen atom. Thus de Broglie hypothesis provided an explanation for Bohr's second postulate for the quantisation of angular momentum of the orbiting electron. The quantised electron orbits and energy states are due to the wave nature of the electron and only resonant standing waves can persist.

Bohr's model, involving classical trajectory picture (planet-like electron orbiting the nucleus), correctly predicts the gross features of the hydrogenic atoms*, in particular, the frequencies of the radiation emitted or selectively absorbed. This model however has many limitations. Some are:

(i) The Bohr model is applicable to hydrogenic atoms. It cannot be extended even to mere two electron atoms such as helium. The analysis of atoms with more than one electron was attempted on the lines of Bohr's model for hydrogenic atoms but did not meet with any success. Difficulty lies in the fact that each electron interacts not only with the positively charged nucleus but also with all other electrons.

* Hydrogenic atoms are the atoms consisting of a nucleus with positive charge +Ze and a single electron, where Z is the proton number. Examples are hydrogen atom, singly ionised helium, doubly ionised lithium, and so forth. In these atoms more complex electron-electron interactions are nonexistent.

The formulation of Bohr model involves electrical force between positively charged nucleus and electron. It does not include the electrical forces between electrons which necessarily appear in multi-electron atoms.

(ii) While the Bohr's model correctly predicts the frequencies of the light emitted by hydrogenic atoms, the model is unable to explain the relative intensities of the frequencies in the spectrum. In emission spectrum of hydrogen, some of the visible frequencies have weak intensity, others strong. Why? Experimental observations depict that some transitions are more favoured than others. Bohr's model is unable to account for the intensity variations.

Bohr's model presents an elegant picture of an atom and cannot be generalised to complex atoms. For complex atoms we have to use a new and radical theory based on Quantum Mechanics, which provides a more complete picture of the atomic structure.

LASER LIGHT

Imagine a crowded market place or a railway platform with people entering a gate and going towards all directions. Their footsteps are random and there is no phase correlation between them. On the other hand, think of a large number of soldiers in a regulated march. Their footsteps are very well correlated. See figure here.

This is similar to the difference between light emitted by an ordinary source like a candle or a bulb and that emitted by a laser. The acronym LASER stands for Light Amplification by Stimulated Emission of Radiation. Since its development in 1960, it has entered into all areas of science and technology. It has found applications in physics, chemistry, biology, medicine, surgery, engineering, etc. There are low power lasers, with a power of 0.5 mW, called pencil lasers, which serve as pointers. There are also lasers of different power, suitable for delicate surgery of eye or glands in the stomach. Finally, there are lasers which can cut or weld steel.

(a) Light from a bulb

(b) Laser light

Light is emitted from a source in the form of packets of waves. Light coming out from an ordinary source contains a mixture of many wavelengths. There is also no phase relation between the various waves. Therefore, such light, even if it is passed through an aperture, spreads very fast and the beam size increases rapidly with distance. In the case of laser light, the wavelength of each packet is almost the same. Also the average length of the packet of waves is much larger. This means that there is better phase correlation over a longer duration of time. This results in reducing the divergence of a laser beam substantially.

If there are N atoms in a source, each emitting light with intensity I, then the total intensity produced by an ordinary source is proportional to NI, whereas in a laser source, it is proportional to N^2I. Considering that N is very large, we see that the light from a laser can be much stronger than that from an ordinary source.

When astronauts of the Apollo missions visited the moon, they placed a mirror on its surface, facing the earth. Then scientists on the earth sent a strong laser beam, which was reflected by the mirror on the moon and received back on the earth. The size of the reflected laser beam and the time taken for the round trip were measured. This allowed a very accurate determination of (a) the extremely small divergence of a laser beam and (b) the distance of the moon from the earth.

SUMMARY

1. Atom, as a whole, is electrically neutral and therefore contains equal amount of positive and negative charges.
2. In *Thomson's model*, an atom is a spherical cloud of positive charges with electrons embedded in it.
3. In *Rutherford's model*, most of the mass of the atom and all its positive charge are concentrated in a tiny nucleus (typically one by ten thousand the size of an atom), and the electrons revolve around it.
4. Rutherford nuclear model has two main difficulties in explaining the structure of atom: (a) It predicts that atoms are unstable because the accelerated electrons revolving around the nucleus must spiral into the nucleus. This contradicts the stability of matter. (b) It cannot explain the characteristic line spectra of atoms of different elements.
5. Atoms of each element are stable and emit characteristic spectrum. The spectrum consists of a set of isolated parallel lines termed as line spectrum. It provides useful information about the atomic structure.
6. The atomic hydrogen emits a line spectrum consisting of various series. The frequency of any line in a series can be expressed as a difference of two terms;

   Lyman series: $\nu = Rc\left(\frac{1}{1^2} - \frac{1}{n^2}\right)$; $n = 2, 3, 4, \ldots$

   Balmer series: $\nu = Rc\left(\frac{1}{2^2} - \frac{1}{n^2}\right)$; $n = 3, 4, 5, \ldots$

   Paschen series: $\nu = Rc\left(\frac{1}{3^2} - \frac{1}{n^2}\right)$; $n = 4, 5, 6, \ldots$

   Brackett series: $\nu = Rc\left(\frac{1}{4^2} - \frac{1}{n^2}\right)$; $n = 5, 6, 7, \ldots$

   Pfund series: $\nu = Rc\left(\frac{1}{5^2} - \frac{1}{n^2}\right)$; $n = 6, 7, 8, \ldots$
7. To explain the line spectra emitted by atoms, as well as the stability of atoms, Niel's Bohr proposed a model for hydrogenic (single elctron) atoms. He introduced three postulates and laid the foundations of quantum mechanics:
   (a) In a hydrogen atom, an electron revolves in certain stable orbits (called stationary orbits) without the emission of radiant energy.
   (b) The stationary orbits are those for which the angular momentum is some integral multiple of $h/2\pi$. (Bohr's quantisation condition.) That is $L = nh/2\pi$, where n is an integer called a quantum number.
   (c) The third postulate states that an electron might make a transition from one of its specified non-radiating orbits to another of lower energy. When it does so, a photon is emitted having energy equal to the energy difference between the initial and final states. The frequency (ν) of the emitted photon is then given by

   $h\nu = E_i - E_f$

   An atom absorbs radiation of the same frequency the atom emits, in which case the electron is transferred to an orbit with a higher value of n.

   $E_i + h\nu = E_f$

8. As a result of the quantisation condition of angular momentum, the electron orbits the nucleus at only specific radii. For a hydrogen atom it is given by

$$r_n = \left(\frac{n^2}{m}\right)\left(\frac{h}{2\pi}\right)^2 \frac{4\pi\varepsilon_0}{e^2}$$

The total energy is also quantised:

$$E_n = -\frac{me^4}{8n^2\varepsilon_0^2 h^2}$$

$$= -13.6 \text{ eV}/n^2$$

The n = 1 state is called ground state. In hydrogen atom the ground state energy is –13.6 eV. Higher values of n correspond to excited states (n > 1). Atoms are excited to these higher states by collisions with other atoms or electrons or by absorption of a photon of right frequency.

9. de Broglie's hypothesis that electrons have a wavelength $\lambda = h/mv$ gave an explanation for Bohr's quantised orbits by bringing in the wave-particle duality. The orbits correspond to circular standing waves in which the circumference of the orbit equals a whole number of wavelengths.

10. Bohr's model is applicable only to hydrogenic (single electron) atoms. It cannot be extended to even two electron atoms such as helium. This model is also unable to explain for the relative intensities of the frequencies emitted even by hydrogenic atoms.

POINTS TO PONDER

1. Both the Thomson's as well as the Rutherford's models constitute an unstable system. Thomson's model is unstable electrostatically, while Rutherford's model is unstable because of electromagnetic radiation of orbiting electrons.
2. What made Bohr quantise angular momentum (second postulate) and not some other quantity? Note, h has dimensions of angular momentum, and for circular orbits, angular momentum is a very relevant quantity. The second postulate is then so natural!
3. The orbital picture in Bohr's model of the hydrogen atom was inconsistent with the uncertainty principle. It was replaced by modern quantum mechanics in which Bohr's orbits are regions where the electron may be found with large probability.
4. Unlike the situation in the solar system, where planet-planet gravitational forces are very small as compared to the gravitational force of the sun on each planet (because the mass of the sun is so much greater than the mass of any of the planets), the electron-electron electric force interaction is comparable in magnitude to the electron-nucleus electrical force, because the charges and distances are of the same order of magnitude. This is the reason why the Bohr's model with its planet-like electron is not applicable to many electron atoms.
5. Bohr laid the foundation of the quantum theory by postulating specific orbits in which electrons do not radiate. Bohr's model include only

one quantum number n. The new theory called quantum mechanics supportes Bohr's postulate. However in quantum mechanics (more generally accepted), a given energy level may not correspond to just one quantum state. For example, a state is characterised by four quantum numbers (n, l, m, and s), but for a pure Coulomb potential (as in hydrogen atom) the energy depends only on n.

6. In Bohr model, contrary to ordinary classical expectation, the frequency of revolution of an electron in its orbit is not connected to the frequency of spectral line. The later is the difference between two orbital energies divided by h. For transitions between large quantum numbers (n to $n-1$, n very large), however, the two coincide as expected.
7. Bohr's semiclassical model based on some aspects of classical physics and some aspects of modern physics also does not provide a true picture of the simplest hydrogenic atoms. The true picture is quantum mechanical affair which differs from Bohr model in a number of fundamental ways. But then if the Bohr model is not strictly correct, why do we bother about it? The reasons which make Bohr's model still useful are:
   (i) The model is based on just three postulates but accounts for almost all the general features of the hydrogen spectrum.
   (ii) The model incorporates many of the concepts we have learnt in classical physics.
   (iii) The model demonstrates how a theoretical physicist occasionally must quite literally ignore certain problems of approach in hopes of being able to make some predictions. If the predictions of the theory or model agree with experiment, a theoretician then must somehow hope to explain away or rationalise the problems that were ignored along the way.

EXERCISES

12.1 Choose the correct alternative from the clues given at the end of the each statement:

(a) The size of the atom in Thomson's model is ..... the atomic size in Rutherford's model. (much greater than/no different from/much less than.)

(b) In the ground state of ..... electrons are in stable equilibrium, while in ..... electrons always experience a net force. (Thomson's model/ Rutherford's model.)

(c) A *classical* atom based on ..... is doomed to collapse. (Thomson's model/ Rutherford's model.)

(d) An atom has a nearly continuous mass distribution in a ..... but has a highly non-uniform mass distribution in ..... (Thomson's model/ Rutherford's model.)

(e) The positively charged part of the atom possesses most of the mass in ..... (Rutherford's model/both the models.)

12.2 Suppose you are given a chance to repeat the alpha-particle scattering experiment using a thin sheet of solid hydrogen in place of the gold foil. (Hydrogen is a solid at temperatures below 14 K.) What results do you expect?

12.3 What is the shortest wavelength present in the Paschen series of spectral lines?

12.4 A difference of 2.3 eV separates two energy levels in an atom. What is the frequency of radiation emitted when the atom make a transition from the upper level to the lower level?

12.5 The ground state energy of hydrogen atom is –13.6 eV. What are the kinetic and potential energies of the electron in this state?

12.6 A hydrogen atom initially in the ground level absorbs a photon, which excites it to the n = 4 level. Determine the wavelength and frequency of photon.

12.7 (a) Using the Bohr's model calculate the speed of the electron in a hydrogen atom in the n = 1, 2, and 3 levels. (b) Calculate the orbital period in each of these levels.

12.8 The radius of the innermost electron orbit of a hydrogen atom is 5.3×10^{-11} m. What are the radii of the n = 2 and n =3 orbits?

12.9 A 12.5 eV electron beam is used to bombard gaseous hydrogen at room temperature. What series of wavelengths will be emitted?

12.10 In accordance with the Bohr's model, find the quantum number that characterises the earth's revolution around the sun in an orbit of radius 1.5×10^{11} m with orbital speed 3×10^{4} m/s. (Mass of earth = 6.0×10^{24} kg.)

ADDITIONAL EXERCISES

12.11 Answer the following questions, which help you understand the difference between Thomson's model and Rutherford's model better.

(a) Is the average angle of deflection of α-particles by a thin gold foil predicted by Thomson's model much less, about the same, or much greater than that predicted by Rutherford's model?

(b) Is the probability of backward scattering (i.e., scattering of α-particles at angles greater than 90°) predicted by Thomson's model much less, about the same, or much greater than that predicted by Rutherford's model?

(c) Keeping other factors fixed, it is found experimentally that for small thickness t, the number of α-particles scattered at moderate angles is proportional to t. What clue does this linear dependence on t provide?

(d) In which model is it completely wrong to ignore multiple scattering for the calculation of average angle of scattering of α-particles by a thin foil?

12.12 The gravitational attraction between electron and proton in a hydrogen atom is weaker than the coulomb attraction by a factor of about 10^{-40}. An alternative way of looking at this fact is to estimate the radius of the first Bohr orbit of a hydrogen atom if the electron and proton were bound by gravitational attraction. You will find the answer interesting.

12.13 Obtain an expression for the frequency of radiation emitted when a hydrogen atom de-excites from level n to level (n–1). For large n, show that this frequency equals the classical frequency of revolution of the electron in the orbit.

12.14 Classically, an electron can be in any orbit around the nucleus of an atom. Then what determines the typical atomic size? Why is an atom not, say, thousand times bigger than its typical size? The question had greatly puzzled Bohr before he arrived at his famous model of the atom that you have learnt in the text. To simulate what he might well have done before his discovery, let us play as follows with the basic constants of nature and see if we can get a quantity with the dimensions of length that is roughly equal to the known size of an atom (~ 10^{-10}m).

(a) Construct a quantity with the dimensions of length from the fundamental constants e, m_e, and c. Determine its numerical value.

(b) You will find that the length obtained in (a) is many orders of magnitude smaller than the atomic dimensions. Further, it involves c. But energies of atoms are mostly in non-relativistic domain where c is not expected to play any role. This is what may have suggested Bohr to discard c and look for 'something else' to get the right atomic size. Now, the Planck's constant h had already made its appearance elsewhere. Bohr's great insight lay in recognising that h, m_e, and e will yield the right atomic size. Construct a quantity with the dimension of length from h, me, and e and confirm that its numerical value has indeed the correct order of magnitude.

12.15 The total energy of an electron in the first excited state of the hydrogen atom is about –3.4 eV.

(a) What is the kinetic energy of the electron in this state?

(b) What is the potential energy of the electron in this state?

(c) Which of the answers above would change if the choice of the zero of potential energy is changed?

12.16 If Bohr's quantisation postulate (angular momentum = $nh/2\pi$) is a basic law of nature, it should be equally valid for the case of planetary motion also. Why then do we never speak of quantisation of orbits of planets around the sun?

12.17 Obtain the first Bohr's radius and the ground state energy of a *muonic hydrogen atom* [i.e., an atom in which a negatively charged muon (μ^-) of mass about $207m_e$ orbits around a proton].

Chapter Thirteen

NUCLEI

13.1 Introduction

In the previous chapter, we have learnt that in every atom, the positive charge and mass are densely concentrated at the centre of the atom forming its nucleus. The overall dimensions of a nucleus are much smaller than those of an atom. Experiments on scattering of α-particles demonstrated that the radius of a nucleus was smaller than the radius of an atom by a factor of about 10^4. This means the volume of a nucleus is about 10^{-12} times the volume of the atom. In other words, an atom is almost empty. If an atom is enlarged to the size of a classroom, the nucleus would be of the size of pinhead. Nevertheless, the nucleus contains most (more than 99.9%) of the mass of an atom.

Does the nucleus have a structure, just as the atom does? If so, what are the constituents of the nucleus? How are these held together? In this chapter, we shall look for answers to such questions. We shall discuss various properties of nuclei such as their size, mass and stability, and also associated nuclear phenomena such as radioactivity, fission and fusion.

13.2 Atomic Masses and Composition of Nucleus

The mass of an atom is very small, compared to a kilogram; for example, the mass of a carbon atom, ^{12}C, is 1.992647×10^{-26} kg. Kilogram is not a very convenient unit to measure such small quantities. Therefore, a

different mass unit is used for expressing atomic masses. This unit is the atomic mass unit (u), defined as $1/12^{th}$ of the mass of the carbon (^{12}C) atom. According to this definition

$$1\text{u} = \frac{\text{mass of one } ^{12}\text{C atom}}{12}$$

$$= \frac{1.992647 \times 10^{-26} \text{ kg}}{12}$$

$$= 1.660539 \times 10^{-27} \text{ kg} \qquad (13.1)$$

The atomic masses of various elements expressed in atomic mass unit (u) are close to being integral multiples of the mass of a hydrogen atom. There are, however, many striking exceptions to this rule. For example, the atomic mass of chlorine atom is 35.46 u.

Accurate measurement of atomic masses is carried out with a mass spectrometer, The measurement of atomic masses reveals the existence of different types of atoms of the same element, which exhibit the same chemical properties, but differ in mass. Such atomic species of the same element differing in mass are called *isotopes*. (In Greek, isotope means the same place, i.e. they occur in the same place in the periodic table of elements.) It was found that practically every element consists of a mixture of several isotopes. The relative abundance of different isotopes differs from element to element. Chlorine, for example, has two isotopes having masses 34.98 u and 36.98 u, which are nearly integral multiples of the mass of a hydrogen atom. The relative abundances of these isotopes are 75.4 and 24.6 per cent, respectively. Thus, the average mass of a chlorine atom is obtained by the weighted average of the masses of the two isotopes, which works out to be

$$= \frac{75.4 \times 34.98 + 24.6 \times 36.98}{100}$$

$$= 35.47 \text{ u}$$

which agrees with the atomic mass of chlorine.

Even the lightest element, hydrogen has three isotopes having masses 1.0078 u, 2.0141 u, and 3.0160 u. The nucleus of the lightest atom of hydrogen, which has a relative abundance of 99.985%, is called the proton. The mass of a proton is

$$m_p = 1.00727 \text{ u} = 1.67262 \times 10^{-27} \text{ kg} \qquad (13.2)$$

This is equal to the mass of the hydrogen atom (= 1.00783u), minus the mass of a single electron (m_e = 0.00055 u). The other two isotopes of hydrogen are called deuterium and tritium. Tritium nuclei, being unstable, do not occur naturally and are produced artificially in laboratories.

The positive charge in the nucleus is that of the protons. A proton carries one unit of fundamental charge and is stable. It was earlier thought that the nucleus may contain electrons, but this was ruled out later using arguments based on quantum theory. All the electrons of an atom are outside the nucleus. We know that the number of these electrons outside the nucleus of the atom is Z, the atomic number. The total charge of the

atomic electrons is thus (–*Ze*), and since the atom is neutral, the charge of the nucleus is (+*Ze*). The number of protons in the nucleus of the atom is, therefore, exactly *Z*, the atomic number.

Discovery of Neutron

Since the nuclei of deuterium and tritium are isotopes of hydrogen, they must contain only one proton each. But the masses of the nuclei of hydrogen, deuterium and tritium are in the ratio of 1:2:3. Therefore, the nuclei of deuterium and tritium must contain, in addition to a proton, some neutral matter. The amount of neutral matter present in the nuclei of these isotopes, expressed in units of mass of a proton, is approximately equal to one and two, respectively. This fact indicates that the nuclei of atoms contain, in addition to protons, neutral matter in multiples of a basic unit. This hypothesis was verified in 1932 by James Chadwick who observed emission of neutral radiation when beryllium nuclei were bombarded with alpha-particles. (α-particles are helium nuclei, to be discussed in a later section). It was found that this neutral radiation could knock out protons from light nuclei such as those of helium, carbon and nitrogen. The only neutral radiation known at that time was photons (electromagnetic radiation). Application of the principles of conservation of energy and momentum showed that if the neutral radiation consisted of photons, the energy of photons would have to be much higher than is available from the bombardment of beryllium nuclei with α-particles. The clue to this puzzle, which Chadwick satisfactorily solved, was to assume that the neutral radiation consists of a new type of neutral particles called *neutrons*. From conservation of energy and momentum, he was able to determine the mass of new particle 'as very nearly the same as mass of proton'.

The mass of a neutron is now known to a high degree of accuracy. It is

$$m_n = 1.00866 \text{ u} = 1.6749\times10^{-27} \text{ kg} \qquad (13.3)$$

Chadwick was awarded the 1935 Nobel Prize in Physics for his discovery of the neutron.

A free neutron, unlike a free proton, is unstable. It decays into a proton, an electron and a antineutrino (another elementary particle), and has a mean life of about 1000s. It is, however, stable inside the nucleus.

The composition of a nucleus can now be described using the following terms and symbols:

Z - atomic number = number of protons [13.4(a)]

N - neutron number = number of neutrons [13.4(b)]

A - mass number = *Z* + *N*

= total number of protons and neutrons [13.4(c)]

One also uses the term nucleon for a proton or a neutron. Thus the number of nucleons in an atom is its mass number A.

Nuclear species or nuclides are shown by the notation $^{A}_{Z}X$ where X is the chemical symbol of the species. For example, the nucleus of gold is denoted by $^{197}_{79}Au$. It contains 197 nucleons, of which 79 are protons and the rest118 are neutrons.

The composition of isotopes of an element can now be readily explained. The nuclei of isotopes of a given element contain the same number of protons, but differ from each other in their number of neutrons. Deuterium, ${}^{2}_{1}\text{H}$, which is an isotope of hydrogen, contains one proton and one neutron. Its other isotope tritium, ${}^{3}_{1}\text{H}$, contains one proton and two neutrons. The element gold has 32 isotopes, ranging from A =173 to A = 204. We have already mentioned that chemical properties of elements depend on their electronic structure. As the atoms of isotopes have identical electronic structure they have identical chemical behaviour and are placed in the same location in the periodic table.

All nuclides with same mass number A are called *isobars*. For example, the nuclides ${}^{3}_{1}\text{H}$ and ${}^{3}_{2}\text{He}$ are isobars. Nuclides with same neutron number N but different atomic number Z, for example ${}^{198}_{80}\text{Hg}$ and ${}^{197}_{79}\text{Au}$, are called *isotones*.

13.3 Size of the Nucleus

As we have seen in Chapter 12, Rutherford was the pioneer who postulated and established the existence of the atomic nucleus. At Rutherford's suggestion, Geiger and Marsden performed their classic experiment: on the scattering of α-particles from thin gold foils. Their experiments revealed that the distance of closest approach to a gold nucleus of an α-particle of kinetic energy 5.5 MeV is about 4.0×10^{-14} m. The scattering of α-particle by the gold sheet could be understood by Rutherford by assuming that the coulomb repulsive force was solely responsible for scattering. Since the positive charge is confined to the nucleus, the actual size of the nucleus has to be less than 4.0×10^{-14} m.

If we use α-particles of higher energies than 5.5 MeV, the distance of closest approach to the gold nucleus will be smaller and at some point the scattering will begin to be affected by the short range nuclear forces, and differ from Rutherford's calculations. Rutherford's calculations are based on pure coulomb repulsion between the positive charges of the α-particle and the gold nucleus. From the distance at which deviations set in, nuclear sizes can be inferred.

By performing scattering experiments in which fast electrons, instead of α-particles, are projectiles that bombard targets made up of various elements, the sizes of nuclei of various elements have been accurately measured.

It has been found that a nucleus of mass number A has a radius

$$R = R_0 A^{1/3} \tag{13.5}$$

where $R_0 = 1.2 \times 10^{-15}$ m. This means the volume of the nucleus, which is proportional to R^3 is proportional to A. Thus the density of nucleus is a constant, independent of A, for all nuclei. Different nuclei are likes drop of liquid of constant density. The density of nuclear matter is approximately 2.3×10^{17} kg m^{-3}. This density is very large compared to ordinary matter, say water, which is 10^3 kg m^{-3}. This is understandable, as we have already seen that most of the atom is empty. Ordinary matter consisting of atoms has a large amount of empty space.

EXAMPLE 13.1

Example 13.1 Given the mass of iron nucleus as 55.85u and A=56, find the nuclear density?

Solution

m_{Fe} = 55.85

u = 9.27×10^{-26} kg

$$\text{Nuclear density} = \frac{\text{mass}}{\text{volume}} = \frac{9.27\times10^{-26}}{(4\pi/3)(1.2\times10^{-15})^3}\times\frac{1}{56}$$

$$= 2.29 \times 10^{17}\ \text{kg m}^{-3}$$

The density of matter in neutron stars (an astrophysical object) is comparable to this density. This shows that matter in these objects has been compressed to such an extent that they resemble a *big nucleus*.

13.4.1 Mass – Energy

Einstein showed from his theory of special relativity that it is necessary to treat mass as another form of energy. Before the advent of this theory of special relativity it was presumed that mass and energy were conserved separately in a reaction. However, Einstein showed that mass is another form of energy and one can convert mass-energy into other forms of energy, say kinetic energy and vice-versa.

Einstein gave the famous mass-energy equivalence relation

$$E = mc^2 \tag{13.6}$$

Here the energy equivalent of mass m is related by the above equation and c is the velocity of light in vacuum and is approximately equal to 3×10^8 m s^{-1}.

EXAMPLE 13.2

Example 13.2 Calculate the energy equivalent of 1 g of substance.

Solution

Energy, $E = 10^{-3} \times (3 \times 10^8)^2$ J

$E = 10^{-3} \times 9 \times 10^{16} = 9 \times 10^{13}$ J

Thus, if one gram of matter is converted to energy, there is a release of enormous amount of energy.

Experimental verification of the Einstein's mass-energy relation has been achieved in the study of nuclear reactions amongst nucleons, nuclei, electrons and other more recently discovered particles. In a reaction the conservation law of energy states that the initial energy and the final energy are equal provided the energy associated with mass is also included. This concept is important in understanding nuclear masses and the interaction of nuclei with one another. They form the subject matter of the next few sections.

13.4.2 Nuclear binding energy

In Section 13.2 we have seen that the nucleus is made up of neutrons and protons. Therefore it may be expected that the mass of the nucleus is equal to the total mass of its individual protons and neutrons. However,

the nuclear mass M is found to be always less than this. For example, let us consider $^{16}_{8}O$; a nucleus which has 8 neutrons and 8 protons. We have

Mass of 8 neutrons = 8 × 1.00866 u

Mass of 8 protons = 8 × 1.00727 u

Mass of 8 electrons = 8 × 0.00055 u

Therefore the expected mass of $^{16}_{8}O$ nucleus
= 8 × 2.01593 u = 16.12744 u.

The atomic mass of $^{16}_{8}O$ found from mass spectroscopy experiments is seen to be 15.99493 u. Substracting the mass of 8 electrons (8 × 0.00055 u) from this, we get the experimental mass of $^{16}_{8}O$ nucleus to be 15.99053 u.

Thus, we find that the mass of the $^{16}_{8}O$ nucleus is less than the total mass of its constituents by 0.13691u. The difference in mass of a nucleus and its constituents, ΔM, is called the *mass defect*, and is given by

$$\Delta M = [Zm_p + (A - Z)m_n] - M \tag{13.7}$$

What is the meaning of the mass defect? It is here that Einstein's equivalence of mass and energy plays a role. Since the mass of the oxygen nucleus is less that the sum of the masses of its constituents (8 protons and 8 neutrons, in the unbound state), the equivalent energy of the oxygen nucleus is less than that of the sum of the equivalent energies of its constituents. If one wants to break the oxygen nucleus into 8 protons and 8 neutrons, this extra energy $\Delta M\, c^2$, has to supplied. This energy required E_b is related to the mass defect by

$$E_b = \Delta M\, c^2 \tag{13.8}$$

Example 13.3 Find the energy equivalent of one atomic mass unit, first in Joules and then in MeV. Using this, express the mass defect of $^{16}_{8}O$ in MeV/c^2.

Solution

1u = 1.6605×10^{-27} kg

To convert it into energy units, we multiply it by c^2 and find that energy equivalent = $1.6605 \times 10^{-27} \times (2.9979 \times 10^8)^2$ kg m^2/s^2

$= 1.4924 \times 10^{-10}$ J

$= \dfrac{1.4924 \times 10^{-10}}{1.602 \times 10^{-19}}$ eV

$= 0.9315 \times 10^9$ eV

= 931.5 MeV

or, 1u = 931.5 MeV/c^2

For $^{16}_{8}O$, ΔM = 0.13691 u = 0.13691×931.5 MeV/c^2
= 127.5 MeV/c^2

The energy needed to separate $^{16}_{8}O$ into its constituents is thus 127.5 MeV/c^2.

EXAMPLE 13.3

If a certain number of neutrons and protons are brought together to form a nucleus of a certain charge and mass, an energy E_b will be released

in the process. The energy E_b is called the *binding energy* of the nucleus. If we separate a nucleus into its nucleons, we would have to supply a total energy equal to E_b, to those particles. Although we cannot tear apart a nucleus in this way, the nuclear binding energy is still a convenient measure of how well a nucleus is held together. A more useful measure of the binding between the constituents of the nucleus is the *binding energy per nucleon*, E_{bn}, which is the ratio of the binding energy E_b of a nucleus to the number of the nucleons, A, in that nucleus:

$$E_{bn} = E_b / A \tag{13.9}$$

We can think of binding energy per nucleon as the average energy per nucleon needed to separate a nucleus into its individual nucleons.

FIGURE 13.1 The binding energy per nucleon as a function of mass number.

Figure 13.1 is a plot of the binding energy per nucleon E_{bn} versus the mass number A for a large number of nuclei. We notice the following main features of the plot:

(i) the binding energy per nucleon, E_{bn}, is practically constant, i.e. practically independent of the atomic number for nuclei of middle mass number (30 < A < 170). The curve has a maximum of about 8.75 MeV for A = 56 and has a value of 7.6 MeV for A = 238.

(ii) E_{bn} is lower for both light nuclei (A<30) and heavy nuclei (A>170).

We can draw some conclusions from these two observations:

(i) The force is attractive and sufficiently strong to produce a binding energy of a few MeV per nucleon.

(ii) The constancy of the binding energy in the range 30 < A < 170 is a consequence of the fact that the nuclear force is short-ranged. Consider a particular nucleon inside a sufficiently large nucleus. It will be under the influence of only some of its neighbours, which come within the range of the nuclear force. If any other nucleon is at a distance more than the range of the nuclear force from the particular nucleon it will have no influence on the binding energy of the nucleon under consideration. If a nucleon can have a maximum of p neighbours within the range of nuclear force, its binding energy would be proportional to p. Let the binding energy of the nucleus be pk, where k is a constant having the dimensions of energy. If we increase A by adding nucleons they will not change the binding energy of a nucleon inside. Since most of the nucleons in a large nucleus reside inside it and not on the surface, the change in binding energy per nucleon would be small. The binding energy per nucleon is a constant and is approximately equal to pk. The property that a given nucleon

influences only nucleons close to it is also referred to as saturation property of the nuclear force.

(iii) A very heavy nucleus, say A = 240, has lower binding energy per nucleon compared to that of a nucleus with A = 120. Thus if a nucleus A = 240 breaks into two A = 120 nuclei, nucleons get more tightly bound. This implies energy would be released in the process. It has very important implications for energy production through *fission*, to be discussed later in Section 13.7.1.

(iv) Consider two very light nuclei ($A \leq 10$) joining to form a heavier nucleus. The binding energy per nucleon of the fused heavier nuclei is more than the binding energy per nucleon of the lighter nuclei. This means that the final system is more tightly bound than the initial system. Again energy would be released in such a process of *fusion*. This is the energy source of sun, to be discussed later in Section 13.7.3.

13.5 Nuclear Force

The force that determines the motion of atomic electrons is the familiar Coulomb force. In Section 13.4, we have seen that for average mass nuclei the binding energy per nucleon is approximately 8 MeV, which is much larger than the binding energy in atoms. Therefore, to bind a nucleus together there must be a strong attractive force of a totally different kind. It must be strong enough to overcome the repulsion between the (positively charged) protons and to bind both protons and neutrons into the tiny nuclear volume. We have already seen that the constancy of binding energy per nucleon can be understood in terms of its short-range. Many features of the nuclear binding force are summarised below. These are obtained from a variety of experiments carried out during 1930 to 1950.

(i) The nuclear force is much stronger than the Coulomb force acting between charges or the gravitational forces between masses. The nuclear binding force has to dominate over the Coulomb repulsive force between protons inside the nucleus. This happens only because the nuclear force is much stronger than the coulomb force. The gravitational force is much weaker than even Coulomb force.

(ii) The nuclear force between two nucleons falls rapidly to zero as their distance is more than a few femtometres. This leads to *saturation of forces* in a medium or a large-sized nucleus, which is the reason for the constancy of the binding energy per nucleon.

A rough plot of the potential energy between two nucleons as a function of distance is shown in the Fig. 13.2. The potential energy is a minimum at a distance r_0 of about 0.8 fm. This means that the force is attractive for distances larger than 0.8 fm and repulsive if they are separated by distances less than 0.8 fm.

FIGURE 13.2 Potential energy of a pair of nucleons as a function of their separation. For a separation greater than r_0, the force is attractive and for separations less than r_0, the force is strongly repulsive.

(iii) The nuclear force between neutron-neutron, proton-neutron and proton-proton is approximately the same. The nuclear force does not depend on the electric charge.

Unlike Coulomb's law or the Newton's law of gravitation there is no simple mathematical form of the nuclear force.

13.6 RADIOACTIVITY

A. H. Becquerel discovered radioactivity in 1896 purely by accident. While studying the fluorescence and phosphorescence of compounds irradiated with visible light, Becquerel observed an interesting phenomenon. After illuminating some pieces of uranium-potassium sulphate with visible light, he wrapped them in black paper and separated the package from a photographic plate by a piece of silver. When, after several hours of exposure, the photographic plate was developed, it showed blackening due to something that must have been emitted by the compound and was able to penetrate both black paper and the silver.

Experiments performed subsequently showed that radioactivity was a nuclear phenomenon in which an unstable nucleus undergoes a decay. This is referred to as *radioactive decay*. Three types of radioactive decay occur in nature :

(i) α-decay in which a helium nucleus ${}^{4}_{2}\text{He}$ is emitted;

(ii) β-decay in which electrons or positrons (particles with the same mass as electrons, but with a charge exactly opposite to that of electron) are emitted;

(iii) γ-decay in which high energy (hundreds of keV or more) photons are emitted.

Each of these decay will be considered in subsequent sub-sections.

13.6.1 Law of radioactive decay

In any radioactive sample, which undergoes α, β or γ-decay, it is found that the number of nuclei undergoing the decay per unit time is proportional to the total number of nuclei in the sample. If N is the number of nuclei in the sample and ΔN undergo decay in time Δt then

$$\frac{\Delta N}{\Delta t} \propto N$$

or, $\Delta N/\Delta t = \lambda N$, (13.10)

where λ is called the radioactive *decay constant* or *disintegration constant*.

The change in the number of nuclei in the sample* is $dN = -\Delta N$ in time Δt. Thus the rate of change of N is (in the limit $\Delta t \to 0$)

$$\frac{dN}{dt} = -\lambda N$$

* ΔN is the number of nuclei that decay, and hence is always positive. dN is the change in N, which may have either sign. Here it is negative, because out of original N nuclei, ΔN have decayed, leaving $(N-\Delta N)$ nuclei.

or, $\frac{dN}{N} = -\lambda dt$

Now, integrating both sides of the above equation,we get,

$$\int_{N_0}^{N} \frac{dN}{N} = -\lambda \int_{t_0}^{t} dt \qquad (13.11)$$

or, $\ln N \quad \ln N_0 = \quad \lambda (t - t_0)$ (13.12)

Here N_0 is the number of radioactive nuclei in the sample at some arbitrary time t_0 and N is the number of radioactive nuclei at any subsequent time t. Setting $t_0 = 0$ and rearranging Eq. (13.12) gives us

$$\ln \frac{N}{N_0} = -\lambda t \qquad (13.13)$$

which gives

$$N(t) = N_0 \, e \quad \lambda \, t \qquad (13.14)$$

Note, for example, the light bulbs follow no such exponential decay law. If we test 1000 bulbs for their life (time span before they burn out or fuse), we expect that they will 'decay' (that is, burn out) at more or less the same time. The decay of radionuclides follows quite a different law, the *law of radioactive decay* represented by Eq. (13.14).

We are quite often interested in the decay rate R (= –dN/dt) than in N itself. It gives us the number of nuclei decaying per unit time. For example, suppose we have some amount of a radioactive substance. We need not know the number of nuclei present in it. But we may measure the number of emissions of α, β or γ particles in a given time interval, say 10 seconds or 20 seconds. Suppose we make measurements for a time interval dt and get a decay count in our instruments equal to ΔN (=–dN). Then the decay rate R is defined as

$$R = -\frac{dN}{dt}$$

Differentiating Eq. (13.14), we find

$$R = -\frac{dN}{dt} = \lambda N_0 \, e^{\lambda t}$$

or, $R = R_0 \, e^{-\lambda t}$ (13.15)

which is an alternative form of the law of radioactive decay [Eq. (13.14)]. Here R_0 is the radioactive decay rate at time $t = 0$, and R is the rate at any subsequent time t. We can now write Eq. (13.10) in terms of the decay rate R of the sample as

$$R = \lambda N \qquad (13.16)$$

where R and the number of radioactive nuclei that have not yet undergone decay must be evaluated at the same instant.

FIGURE 13.3 Exponential decay of a radioactive species. After a lapse of $T_{1/2}$, population of the given species drops by a factor of 2.

The total decay rate R of a sample of one or more radionuclides is called the *activity* of that sample. The SI unit for activity is becquerel, named after the discoverer of radioactivity, Henry Becquerel. It is defined as

1 becquerel = 1Bq = 1 decay per second

An older unit, the curie, is still in common use:

1 curie = 1 Ci = 3.7×10^{10} Bq (decays per second)

There are two common time measures of how long any given type of radionuclide lasts. One measure is the *half-life* $T_{1/2}$ of a radionuclide, which is the time at which both N and R have been reduced to one-half their initial values. The other measure is the *mean life* τ, which is the time at which both N and R have been reduced to e^{-1} of their initial values.

To relate $T_{1/2}$ to the disintegration constant λ, we put $R = (1/2)\, R_0$ and $t = T_{1/2}$ in Eq. (13.15) and solve for $T_{1/2}$, we find

$$T_{1/2} = \frac{\ln 2}{\lambda} = \frac{0.693}{\lambda} \qquad (13.17)$$

The average life or mean life, τ can also be obtained from Eq. (13.14). The number of nuclei which decay in the time interval t to $t + \Delta t$ is $R(t)\Delta t = (\lambda N_0 e^{-\lambda t}\Delta t)$. Each of them has lived for time t. Thus the total life of all these nuclei would be $t\,\lambda N_0 e^{-\lambda t}\,\Delta t$. It is clear that some nuclei may live for a short time while others may live longer. Therefore to obtain the meanlife, we have to sum (or integrate) this expression over all times from 0 to ∞, and divide by the total number N_0 of nuclei at $t = 0$. Thus,

$$\tau = \frac{\lambda N_0 \int_0^\infty t e^{-\lambda t} dt}{N_0} = \lambda \int_0^\infty t e^{-\lambda t} dt$$

One can show by performing this integral that

$\tau = 1/\lambda$

We summarise these results with the following:

$$T_{1/2} = \frac{\ln 2}{\lambda} = \tau \ln 2 \qquad (13.18)$$

Those radioactive elements whose half-life is short compared to the age of the universe (13.7 billion years) are not found in observable quantities in nature today. They have, however, been seen in the laboratory in nuclear reactions. Tritium and plutonium belong to this category.

MARIE SKLODOWSKA CURIE (1867-1934)

Marie Sklodowska Curie (1867-1934) Born in Poland. She is recognised both as a physicist and as a chemist. The discovery of radioactivity by Henri Becquerel in 1896 inspired Marie and her husband Pierre Curie in their researches and analyses which led to the isolation of radium and polonium elements. She was the first person to be awarded two Nobel Prizes- for Physics in 1903 and for Chemistry in 1911.

EXAMPLE 13.4

Example 13.4 The half-life of $^{238}_{92}U$ undergoing α-decay is 4.5×10^9 years. What is the activity of 1g sample of $^{238}_{92}U$?

Solution

$T_{1/2} = 4.5 \times 10^9$ y

$= 4.5 \times 10^9$ y x 3.16×10^7 s/y

$= 1.42 \times 10^{17}$s

One k mol of any isotope contains Avogadro's number of atoms, and so 1g of $^{238}_{92}U$ contains

$$\frac{1}{238\times10^{-3}}\text{ kmol} \times 6.025 \times 10^{26}\text{ atoms/kmol}$$

$= 25.3 \times 10^{20}$ atoms.

The decay rate R is

$R = \lambda N$

$$= \frac{0.693}{T_{1/2}}N = \frac{0.693\times25.3\times10^{20}}{1.42\times10^{17}}\,s^{-1}$$

$= 1.23 \times 10^4\ s^{-1}$

$= 1.23 \times 10^4$ Bq

EXAMPLE 13.4

Example 13.5 Tritium has a half-life of 12.5 y undergoing beta decay. What fraction of a sample of pure tritium will remain undecayed after 25 y.

Solution

By definition of half-life, half of the initial sample will remain undecayed after 12.5 y. In the next 12.5 y, one-half of these nuclei would have decayed. Hence, one fourth of the sample of the initial pure tritium will remain undecayed.

EXAMPLE 13.5

13.6.2 Alpha decay

When a nucleus undergoes *alpha-decay*, it transforms to a different nucleus by emitting an alpha-particle (a helium nucleus, $^{4}_{2}He$). For example, when $^{238}_{92}U$ undergoes alpha-decay, it transforms to $^{234}_{90}Th$

$$^{238}_{92}U \rightarrow {}^{234}_{90}Th + {}^{4}_{2}He \qquad (13.19)$$

In this process, it is observed that since $^{4}_{2}He$ contains two protons and two neutrons, the mass number and the atomic number of the daughter nucleus decreases by four and two, respectively. Thus, the transformation of a nucleus $^{A}_{Z}X$ into a nucleus $^{A-4}_{Z-2}Y$ can be expressed as

$$^{A}_{Z}X \rightarrow {}^{A-4}_{Z-2}Y + {}^{4}_{2}He \qquad (13.20)$$

where $^{A}_{Z}X$ is the parent nucleus and $^{A-4}_{Z-2}Y$ is the daughter nucleus.

The alpha-decay of $^{238}_{92}U$ can occur spontaneously (without an external source of energy) because the total mass of the decay products $^{234}_{90}Th$ and $^{4}_{2}He$ is less than the mass of the original $^{238}_{92}U$. Thus, the total mass energy of the decay products is less than the mass energy of the original nuclide. The difference between the initial mass energy and the final mass energy of the decay products is called the *Q value* of the process or the *disintegration energy*. Thus, the Q value of an alpha decay can be expressed as

$$Q = (m_X - m_Y - m_{He})\,c^2 \qquad (13.21)$$

This energy is shared by the daughter nucleus $^{A-4}_{Z-2}Y$ and the alpha-particle, $^{4}_{2}He$ in the form of kinetic energy. Alpha-decay obeys the radioactive law as given in Eqs. (13.14) and (13.15).

Example 13.6 We are given the following atomic masses:

$^{238}_{92}U$ = 238.05079 u $^{4}_{2}He$ = 4.00260 u

$^{234}_{90}Th$ = 234.04363 u $^{1}_{1}H$ = 1.00783 u

$^{237}_{91}Pa$ = 237.05121 u

Here the symbol Pa is for the element protactinium (Z = 91).

(a) Calculate the energy released during the alpha decay of $^{238}_{92}U$.

(b) Show that $^{238}_{92}U$ can not spontaneously emit a proton.

Solution

(a) The alpha decay of $^{238}_{92}U$ is given by Eq. (13.20). The energy released in this process is given by

$Q = (M_U - M_{Th} - M_{He})\, c^2$

Substituting the atomic masses as given in the data, we find

$Q = (238.05079 - 234.04363 - 4.00260)\text{u} \times c^2$

$= (0.00456 \text{ u})\, c^2$

$= (0.00456 \text{ u})\, (931.5 \text{ MeV/u})$

$= 4.25$ MeV.

(b) If $^{238}_{92}U$ spontaneously emits a proton, the decay process would be

$^{238}_{92}U \rightarrow {}^{237}_{91}Pa + {}^{1}_{1}H$

The Q for this process to happen is

$= (M_U - M_{Pa} - M_H)\, c^2$

$= (238.05079 - 237.05121 - 1.00783) \text{ u} \times c^2$

$= (-\, 0.00825 \text{ u})\, c^2$

$= -\, (0.00825 \text{ u})(931.5 \text{ MeV/u})$

$= -\, 7.68$ MeV

Thus, the Q of the process is negative and therefore it cannot proceed spontaneously. We will have to supply an energy of 7.68 MeV to a $^{238}_{92}U$ nucleus to make it emit a proton.

EXAMPLE 13.6

13.6.3 Beta decay

A nucleus that decays spontaneously by emitting an electron or a positron is said to undergo *beta decay*. Like alpha decay, this is a spontaneous process, with a definite disintegration energy and half-life. Again like alpha decay, beta decay is a statistical process governed by Eqs. (13.14) and (13.15). In *beta minus* (β^-) decay, an electron is emitted by the nucleus, as in the decay of $^{32}_{15}P$

$$^{32}_{15}P \rightarrow {}^{32}_{16}S + e^- + \bar{\nu} \qquad (T_{1/2} = 14.3 \text{ d}) \tag{13.22}$$

In *beta plus* (β^+) decay, a positron is emitted by the nucleus, as in the decay of $^{22}_{11}Na$

$$^{22}_{11}Na \rightarrow {}^{22}_{10}Ne + e^+ + \nu \qquad (T_{1/2} = 2.6 \text{ y}) \tag{13.23}$$

The symbols $\bar{\nu}$ and ν represent antineutrino and neutrino, respectively; both are neutral particles, with very little or no mass. These particles are emitted from the nucleus along with the electron or positron during the decay process. Neutrinos interact only very weakly with matter; they can even penetrate the earth without being absorbed. It is for this reason that their detection is extremely difficult and their presence went unnoticed for long.

In beta-minus decay, a neutron transforms into a proton within the nucleus according to

$$n \rightarrow p + e^{-} + \bar{\nu} \tag{13.24}$$

whereas in beta-plus decay, a proton transforms into neutron (inside the nucleus) via

$$p \rightarrow n + e^{+} + \nu \tag{13.25}$$

These processes show why the mass number A of a nuclide undergoing beta decay does not change; one of its constituent nucleons simply changes its character according to Eq. (13.24) or (13.25).

13.6.4 Gamma decay

There are energy levels in a nucleus, just like there are energy levels in atoms. When a nucleus is in an excited state, it can make a transition to a lower energy state by the emission of electromagnetic radiation. As the energy differences between levels in a nucleus are of the order of MeV, the photons emitted by the nuclei have MeV energies and are called gamma rays.

Most radionuclides after an alpha decay or a beta decay leave the daughter nucleus in an excited state. The daughter nucleus reaches the ground state by a single transition or sometimes by successive transitions by emitting one or more gamma rays. A well-known example of such a process is that of $^{60}_{27}\text{Co}$. By beta emission, the $^{60}_{27}\text{Co}$ nucleus transforms into $^{60}_{28}\text{Ni}$ nucleus in its excited state. The excited $^{60}_{28}\text{Ni}$ nucleus so formed then de-excites to its ground state by successive emission of 1.17 MeV and 1.33 MeV gamma rays. This process is depicted in Fig.13.4 through an energy level diagram.

FIGURE 13.4 Energy level diagram showing the emission of γ rays by a $^{60}_{27}\text{Co}$ nucleus subsequent to beta decay.

13.7 Nuclear Energy

The curve of binding energy per nucleon E_{bn}, given in Fig. 13.1, has a long flat middle region between A = 30 and A = 170. In this region the binding energy per nucleon is nearly constant (8.0 MeV). For the lighter nuclei region, A < 30, and for the heavier nuclei region, A > 170, the binding energy per nucleon is less than 8.0 MeV, as we have noted earlier. This feature of the binding energy curve means that nuclei in the middle region $30 \leq A \leq 170$ are more tightly bound than nuclei with A < 30 and A > 170 . Energy then can be released if less tightly bound nuclei are transmuted into more tightly bound nuclei. Two such processes, which we have already referred to, are fission and fusion.

In conventional energy sources like coal or petroleum, energy is released through chemical reactions. The energies involved are of the order of electron volts per atom. As we have seen, energies involved in

nuclear processes are million times larger (in MeVs per nucleon). This means that for the same quantity of matter, nuclear sources will give a million times larger energy than conventional sources. One kilogram of coal on burning gives 10^7 J of energy, whereas 1 kg of uranium, which undergoes fission, will generate on fission 10^{14} J of energy.

13.7.1 Fission

Soon after the discovery of neutron by Chadwick, Enrico Fermi found that when neutrons bombard various elements, new radioactive elements are produced. However, when a neutron was bombared on a uranium target, the uranium nucleus broke into two nearly equal fragments releasing great amount of energy. An example of such a reaction is

$$^{1}_{0}\text{n} + ^{235}_{92}\text{U} \rightarrow ^{236}_{92}\text{U} \rightarrow ^{144}_{56}\text{Ba} + ^{89}_{36}\text{Kr} + 3\,^{1}_{0}\text{n} \qquad (13.26)$$

Fission does not always produce barium and krypton. A different pair can be produced, for example

$$^{1}_{0}\text{n} + ^{235}_{92}\text{U} \rightarrow ^{236}_{92}\text{U} \rightarrow ^{133}_{51}\text{Sb} + ^{99}_{41}\text{Nb} + 4\,^{1}_{0}\text{n} \qquad (13.27)$$

Still another example is

$$^{1}_{0}\text{n} + ^{235}_{92}\text{U} \rightarrow ^{140}_{54}\text{Xe} + ^{94}_{38}\text{Sr} + 2\,^{1}_{0}\text{n} \qquad (13.28)$$

The fragment nuclei produced in fission are highly neutron-rich and unstable. They are radioactive and emit beta particles in succession until each reaches a stable end product.

The energy released (the Q value) in the fission reaction of nuclei like uranium is of the order of 200 MeV per fissioning nucleus. This is estimated as follows:

Let us take a nucleus with A = 240 breaking into two fragments each of A = 120. Then

E_{bn} for A = 240 nucleus is about 7.6 MeV,

E_{bn} for the two A = 120 fragment nuclei is about 8.5 MeV.

$\therefore$ Gain in binding energy for nucleon is about 0.9 MeV.

Hence the total gain in binding energy is 240×0.9 or 216 MeV

The disintegration energy in fission events first appears as the kinetic energy of the fragments and neutrons. Eventually it is transferred to the surrounding matter appearing as heat. The source of energy in nuclear reactors, which produce electricity, is nuclear fission. The enormous energy released in an atom bomb comes from uncontrolled nuclear fission. We discuss some details in the next section how a nuclear reactor functions.

13.7.2 Nuclear reactor

When $^{235}_{92}\text{U}$ undergoes a fission after bombarded by a neutron, it also releases an extra neutron. This extra neutron is then available for initiating fission of another $^{235}_{92}\text{U}$ nucleus. In fact, on an average, 2½ neutrons per fission of uranium nucleus are released. The fact that more neutrons are produced in fission than are consumed raises the possibility of a chain reaction with each neutron that is produced triggering another fission.

India's atomic energy programme

The atomic energy programme in India was launched around the time of independence under the leadership of Homi J. Bhabha (1909-1966). An early historic achievement was the design and construction of the first nuclear reactor in India (named Apsara) which went critical on August 4, 1956. It used enriched uranium as fuel and water as moderator. Following this was another notable landmark: the construction of CIRUS (Canada India Research U.S.) reactor in 1960. This 40 MW reactor used natural uranium as fuel and heavy water as moderator. Apsara and CIRUS spurred research in a wide range of areas of basic and applied nuclear science. An important milestone in the first two decades of the programme was the indigenous design and construction of the plutonium plant at Trombay, which ushered in the technology of fuel reprocessing (separating useful fissile and fertile nuclear materials from the spent fuel of a reactor) in India. Research reactors that have been subsequently commissioned include ZERLINA, PURNIMA (I, II and III), DHRUVA and KAMINI. KAMINI is the country's first large research reactor that uses U-233 as fuel. As the name suggests, the primary objective of a research reactor is not generation of power but to provide a facility for research on different aspects of nuclear science and technology. Research reactors are also an excellent source for production of a variety of radioactive isotopes that find application in diverse fields: industry, medicine and agriculture.

The main objectives of the Indian Atomic Energy programme are to provide safe and reliable electric power for the country's social and economic progress and to be self-reliant in all aspects of nuclear technology. Exploration of atomic minerals in India undertaken since the early fifties has indicated that India has limited reserves of uranium, but fairly abundant reserves of thorium. Accordingly, our country has adopted a three-stage strategy of nuclear power generation. The first stage involves the use of natural uranium as a fuel, with heavy water as moderator. The Plutonium-239 obtained from reprocessing of the discharged fuel from the reactors then serves as a fuel for the second stage — the fast breeder reactors. They are so called because they use fast neutrons for sustaining the chain reaction (hence no moderator is needed) and, besides generating power, also breed more fissile species (plutonium) than they consume. The third stage, most significant in the long term, involves using fast breeder reactors to produce fissile Uranium-233 from Thorium-232 and to build power reactors based on them.

India is currently well into the second stage of the programme and considerable work has also been done on the third — the thorium utilisation — stage. The country has mastered the complex technologies of mineral exploration and mining, fuel fabrication, heavy water production, reactor design, construction and operation, fuel reprocessing, etc. Pressurised Heavy Water Reactors (PHWRs) built at different sites in the country mark the accomplishment of the first stage of the programme. India is now more than self-sufficient in heavy water production. Elaborate safety measures both in the design and operation of reactors, as also adhering to stringent standards of radiological protection are the hallmark of the Indian Atomic Energy Programme.

Enrico Fermi first suggested such a possibility in 1939. The chain reaction can be either uncontrolled and rapid (as in a nuclear bomb) or controlled and steady (as in a nuclear reactor). The first leads to destruction while the latter can be harnessed to generate electric power.

But soon it was discovered that neutrons liberated in fission of a uranium nucleus were so energetic that they would escape instead of triggering another fission reaction. Also, it turns out that slow neutrons have a much higher intrinsic probability of inducing fission in $^{235}_{92}\text{U}$ than fast neutrons.

The average energy of a neutron produced in fission of $^{235}_{92}\text{U}$ is 2 MeV. These neutrons unless slowed down will escape from the reactor without interacting with the uranium nuclei, unless a very large amount of fissionable material is used for sustaining the chain reaction. What one needs to do is to slow down the fast neutrons by elastic scattering with light nuclei. In fact, Chadwick's experiments showed that in an elastic collision with hydrogen the neutron almost comes to rest and proton carries away the energy. This is the same situation as when a marble hits head-on an identical marble at rest. Therefore, in reactors, light nuclei called *moderators* are provided along with the fissionable nuclei for slowing down fast neutrons. The moderators commonly used are water, heavy water (D_2O) and graphite. The Apsara reactor at the Bhabha Atomic Research Centre (BARC), Mumbai, uses water as moderator. The other Indian reactors, which are used for power production, use heavy water as moderator.

Because of the use of moderator, it is possible that the ratio, K, of number of fission produced by a given generation of neutrons to the number of fission of the preceeding generation may be greater than one. This ratio is called the *multiplication factor;* it is the measure of the growth rate of the neutrons in the reactor. For $K = 1$, the operation of the reactor is said to be *critical*, which is what we wish it to be for steady power operation. If K becomes greater than one, the reaction rate and the reactor power increases exponentially. Unless the factor K is brought down very close to unity, the reactor will become supercritical and can even explode. The explosion of the Chernobyl reactor in Ukraine in 1986 is a sad reminder that accidents in a nuclear reactor can be catastrophic.

The reaction rate is controlled through control-rods made out of neutron-absorbing material such as cadmium. In addition to control rods, reactors are provided with *safety rods* which, when required, can be inserted into the reactor and K can be reduced rapidly to less than unity.

The abundant $^{238}_{92}\text{U}$ isotope, which does not fission, on capturing a neutron leads to the formation of plutonium. The series of reactions involved is

$$^{238}_{92}\text{U} + \text{n} \rightarrow {}^{239}_{92}\text{U} \rightarrow {}^{239}_{93}\text{Np} + e^- + \bar{\nu}$$

$$^{239}_{93}\text{Np} \rightarrow {}^{239}_{94}\text{Pu} + e^- + \bar{\nu} \qquad (13.29)$$

Plutonium is highly radioactive and can also undergo fission under bombardment by slow neutrons.

The broad outlines of a typical nuclear power plant based on *pressurised-water reactor* are shown in Fig. 13.5. In such a reactor, water is used both as the moderator and as the heat transfer medium. In the *primary-loop*, water is circulated through the reactor vessel and transfers energy at high temperature and pressure (at about 600 K and 150 atm) to the steam generator, which is part of the *secondary-loop*. In the steam generator, evaporation provides high-pressure steam to operate the

Nuclear power plants in India
http://www.npcil.nic.in/plantsInoperation.asp
PHYSICS

FIGURE 13.5 Simplified outlines of a nuclear power plant.

turbine that drives the electric generator. The low-pressure steam from the turbine is cooled and condensed to water and forced back into the steam generator.

The energy released in nuclear reactions is a million times larger than in chemical reactions. Therefore, the nuclear reactors need fuel a million times less in mass than used in the chemical reactors of the same power capacity. A kilogram of $^{235}_{92}\text{U}$ on complete fission generates about 3×10^4 MW. However, in nuclear reactions highly radioactive elements are continuously produced. Therefore, an unavoidable feature of reactor operation is the accumulation of radioactive waste, including both fission products and heavy *transuranic* elements such as plutonium and americium.

Historically energy has been produced by using chemical reactions, such as burning coal, wood, gas and petroleum products. The environmental pollution produced by these is creating serious problems due to green house effect leading to global warming. The problem encountered with the nuclear power station is that the spent fuel is highly radioactive and extremely hazardous to all forms of life on earth. Elaborate safety measures, both for reactor operation as well as handling and reprocessing the spent fuel, are required. These safety measures are a distinguishing feature of the Indian Atomic Energy programme. An appropriate plan is being evolved to study the possibility of converting radioactive waste into less active and short-lived material.

13.7.3 Nuclear fusion – energy generation in stars

The binding energy curve shown in Fig.13.1 shows that energy can be released if two light nuclei combine to form a single larger nucleus, a process called *nuclear fusion*. Some examples of such energy liberating reactions are

$$^{1}_{1}\text{H} + ^{1}_{1}\text{H} \rightarrow ^{2}_{1}\text{H} + e^{+} + \nu + 0.42 \text{ MeV} \qquad [13.29(a)]$$

PHYSICS

A simplified online simulation of a nuclear reactor
http://www.kernenergie.net/datf/en/interactive/reactor.php

$$^{2}_{1}\text{H} + ^{2}_{1}\text{H} \rightarrow ^{3}_{2}\text{He} + n + 3.27 \text{ MeV} \qquad [13.29(b)]$$

$$^{2}_{1}\text{H} + ^{2}_{1}\text{H} \rightarrow ^{3}_{1}\text{H} + ^{1}_{1}\text{H} + 4.03 \text{ MeV} \qquad [13.29(c)]$$

In reaction (13.29a), two protons combine to form a deuteron and a positron with a release of 0.42 MeV energy. In reaction [13.29(b)], two deuterons combine to form the light isotope of helium. In reaction (13.29c), two deuterons combine to form a triton and a proton. In all these reactions, we find that two positively charged particles combine to form a larger nucleus. It must be realised that such a process is hindered by the Coulomb repulsion that acts to prevent the two positively charged particles from getting close enough to be within the range of their attractive nuclear forces and thus 'fusing'. The height of this *Coulomb barrier* depends on the charges and the radii of the two interacting nuclei. For example, it can be easily shown that for two protons, the barrier height is ~ 400 keV. The barrier height for more highly charged nuclei is higher. The temperature at which protons in a proton gas would have enough energy to overcome the coulomb's barrier is given by $(3/2)k\,T = K \simeq 400$ keV and is about 3×10^9 K.

To generate useful amount of energy, nuclear fusion must occur in bulk matter. What is needed is to raise the temperature of the material until the particles have enough energy – due to their thermal motions alone – to penetrate the coulomb barrier. This process is called *thermonuclear fusion*.

The temperature of the core of the sun is only about 1.5×10^7 K. Therefore, even in the sun if the fusion is to take place, it must involve protons whose energies are *far* above the average energy.

Thus, for thermonuclear fusion to take place, extreme conditions of temperature and pressure are required, which are available only in the interiors of stars including sun. The energy generation in stars takes place via thermonuclear fusion.

The fusion reaction in the sun is a multi-step process in which hydrogen is burned into helium, hydrogen being the 'fuel' and helium the 'ashes'. The *proton-proton* (p, p) *cycle* by which this occurs is represented by the following sets of reactions:

$$^{1}_{1}\text{H} + ^{1}_{1}\text{H} \rightarrow ^{2}_{1}\text{H} + e^{+} + \nu + 0.42 \text{ MeV} \qquad \text{(i)}$$

$$e^{+} + e^{-} \rightarrow \gamma + \gamma + 1.02 \text{ MeV} \qquad \text{(ii)}$$

$$^{2}_{1}\text{H} + ^{1}_{1}\text{H} \rightarrow ^{3}_{2}\text{He} + \gamma + 5.49 \text{ MeV} \qquad \text{(iii)}$$

$$^{3}_{2}\text{H} + ^{3}_{2}\text{H} \rightarrow ^{4}_{2}\text{He} + ^{1}_{1}\text{H} + ^{1}_{1}\text{H} + 12.86 \text{ MeV} \qquad \text{(iv)} \qquad (13.30)$$

For the fourth reaction to occur, the first three reactions must occur twice, in which case two light helium nuclei unite to form ordinary helium or nucleus. If we consider the combination 2(i) + 2(ii) + 2(iii) +(iv), the net effect is

$$4\,^{1}_{1}\text{H} + 2e^{-} \rightarrow ^{4}_{2}\text{He} + 2\nu + 6\gamma + 26.7 \text{ MeV}$$

or $$(4\,^{1}_{1}\text{H} + 4e^{-}) \rightarrow (^{4}_{2}\text{He} + 2e^{-}) + 2\nu + 6\gamma + 26.7 \text{ MeV} \qquad (13.31)$$

Thus, four hydrogen atoms combine to form an $^{4}_{2}\text{He}$ atom with a release of 26.7 MeV of energy.

The burning of hydrogen in the sun's core is alchemy on a grand scale in the sense that one element is turned into another. It has been

going on for about 5×10^9 y, and calculations show that there is enough hydrogen to keep the sun going for about the same time into the future. In about 5 billion years, however, the sun's core, which by that time will be largely helium, will begin to cool and the sun will start to collapse under its own gravity. This will raise the core temperature and cause the outer envelope to expand, turning the sun into what is called a *red giant*.

If the core temperature increases to 10^8 K again, energy can be produced through fusion once more – this time by burning helium to make carbon. As a star evolves further and becomes still hotter, other elements can be formed by other fusion reactions. However, elements more massive than those near the peak of the binding energy curve of Fig. 13.1 cannot be produced by further fusion.

The energy generation in stars takes place via thermonuclear fusion.

NUCLEAR HOLOCAUST

In a single uranium fission about 0.9×235 MeV (≈200 MeV) of energy is liberated. If each nucleus of about 50 kg of ^{235}U undergoes fission the amount of energy involved is about 4×10^{15}J. This energy is equivalent to about 20,000 tons of TNT, enough for a superexplosion. Uncontrolled release of large nuclear energy is called an atomic explosion. On August 6, 1945 an atomic device was used in warfare for the first time. The US dropped an atom bomb on Hiroshima, Japan. The explosion was equivalent to 20,000 tons of TNT. Instantly the radioactive products devastated 10 sq km of the city which had 3,43,000 inhabitants. Of this number 66,000 were killed and 69,000 were injured; more than 67% of the city's structures were destroyed.

High temperature conditions for fusion reactions can be created by exploding a fission bomb. Super-explosions equivalent to 10 megatons of explosive power of TNT were tested in 1954. Such bombs which involve fusion of isotopes of hydrogen, deuterium and tritium are called hydrogen bombs. It is estimated that a nuclear arsenal sufficient to destroy every form of life on this planet several times over is in position to be triggered by the press of a button. Such a nuclear holocaust will not only destroy the life that exists now but its radioactive fallout will make this planet unfit for life for all times. Scenarios based on theoretical calculations predict a long *nuclear winter*, as the radioactive waste will hang like a cloud in the earth's atmosphere and will absorb the sun's radiation.

13.7.4 Controlled thermonuclear fusion

The first thermonuclear reaction on earth occurred at Eniwetok Atoll on November 1, 1952, when USA exploded a fusion device, generating energy equivalent to 10 million tons of TNT (one ton of TNT on explosion releases 2.6×10^{22} MeV of energy).

A sustained and controllable source of fusion power is considerably more difficult to achieve. It is being pursued vigorously in many countries around the world (including India), because fusion reactor is regarded as the future power source.

Example 13.7 Answer the following questions:

(a) Are the equations of nuclear reactions (such as those given in Section 13.7) 'balanced' in the sense a chemical equation (e.g., $2H_2 + O_2 \rightarrow 2\ H_2O$) is? If not, in what sense are they balanced on both sides?

(b) If both the number of protons and the number of neutrons are conserved in each nuclear reaction, in what way is mass converted into energy (or vice-versa) in a nuclear reaction?

(c) A general impression exists that mass-energy interconversion takes place only in nuclear reaction and never in chemical reaction. This is strictly speaking, incorrect. Explain.

Solution

(a) A chemical equation is balanced in the sense that the number of atoms of each element is the same on both sides of the equation. A chemical reaction merely alters the original combinations of atoms. In a nuclear reaction, elements may be transmuted. Thus, the number of atoms of each element is not necessarily conserved in a nuclear reaction. However, the number of protons and the number of neutrons are both separately conserved in a nuclear reaction. [Actually, even this is not strictly true in the realm of very high energies – what is strictly conserved is the total charge and total 'baryon number'. We need not pursue this matter here.]

In nuclear reactions (e.g., Eq. 13.26), the number of protons and the number of neutrons are the same on the two sides of the equation.

(b) We know that the binding energy of a nucleus gives a negative contribution to the mass of the nucleus (mass defect). Now, since proton number and neutron number are conserved in a nuclear reaction, the total rest mass of neutrons and protons is the same on either side of a reaction. But the total binding energy of nuclei on the left side need not be the same as that on the right hand side. The difference in these binding energies appears as energy released or absorbed in a nuclear reaction. Since binding energy contributes to mass, we say that the difference in the total mass of nuclei on the two sides get converted into energy or vice-versa. It is in these sense that a nuclear reaction is an example of mass-energy interconversion.

(c) From the point of view of mass-energy interconversion, a chemical reaction is similar to a nuclear reaction *in principle.* The energy released or absorbed in a chemical reaction can be traced to the difference in chemical (not nuclear) binding energies of atoms and molecules on the two sides of a reaction. Since, strictly speaking, chemical binding energy also gives a negative contribution (mass defect) to the total mass of an atom or molecule, we can equally well say that the difference in the total mass of atoms or molecules, on the two sides of the chemical reaction gets converted into energy or vice-versa. However, the mass defects involved in a chemical reaction are almost a million times smaller than those in a nuclear reaction.This is the reason for the general impression, (which is *incorrect*) that mass-energy interconversion does not take place in a chemical reaction.

SUMMARY

1. An atom has a nucleus. The nucleus is positively charged. The radius of the nucleus is smaller than the radius of an atom by a factor of 10^4. More than 99.9% mass of the atom is concentrated in the nucleus.
2. On the atomic scale, mass is measured in atomic mass units (u). By definition, 1 atomic mass unit (1u) is $1/12^{th}$ mass of one atom of ^{12}C; $1u = 1.660563 \times 10^{-27}$ kg.
3. A nucleus contains a neutral particle called neutron. Its mass is almost the same as that of proton
4. The atomic number Z is the number of protons in the atomic nucleus of an element. The mass number A is the total number of protons and neutrons in the atomic nucleus; $A = Z+N$; Here N denotes the number of neutrons in the nucleus.

   A nuclear species or a nuclide is represented as $^{A}_{Z}X$, where X is the chemical symbol of the species.

   Nuclides with the same atomic number Z, but different neutron number N are called *isotopes*. Nuclides with the same A are *isobars* and those with the same N are *isotones*.

   Most elements are mixtures of two or more isotopes. The atomic mass of an element is a weighted average of the masses of its isotopes. The masses are the relative abundances of the isotopes.
5. A nucleus can be considered to be spherical in shape and assigned a radius. Electron scattering experiments allow determination of the nuclear radius; it is found that radii of nuclei fit the formula

   $R = R_0 A^{1/3}$,

   where R_0 = a constant = 1.2 fm. This implies that the nuclear density is independent of A. It is of the order of 10^{17} kg/m^3.
6. Neutrons and protons are bound in a nucleus by the short-range strong nuclear force. The nuclear force does not distinguish between neutron and proton.
7. The nuclear mass M is always less than the total mass, Σm, of its constituents. The difference in mass of a nucleus and its constituents is called the *mass defect*,

   $\Delta M = (Z\, m_p + (A - Z)m_n) - M$

   Using Einstein's mass energy relation, we express this mass difference in terms of energy as

   $\Delta E_b = \Delta M\, c^2$

   The energy ΔE_b represents the *binding energy* of the nucleus. In the mass number range A = 30 to 170, the binding energy per nucleon is nearly constant, about 8 MeV/nucleon.
8. Energies associated with nuclear processes are about a million times larger than chemical process.
9. The Q-value of a nuclear process is

   Q = final kinetic energy – initial kinetic energy.

   Due to conservation of mass-energy, this is also,

   Q = (sum of initial masses – sum of final masses)c^2
10. Radioactivity is the phenomenon in which nuclei of a given species transform by giving out α or β or γ rays; α-rays are helium nuclei;

β-rays are electrons. γ-rays are electromagnetic radiation of wavelengths shorter than X-rays;

11. Law of radioactive decay : $N(t) = N(0)\, e^{-\lambda t}$

    where λ is the decay constant or disintegration constant.

    The half-life $T_{1/2}$ of a radionuclide is the time in which N has been reduced to one-half of its initial value. The mean life τ is the time at which N has been reduced to e^{-1} of its initial value

    $$T_{1/2} = \frac{\ln 2}{\lambda} = \tau \ln 2$$

12. Energy is released when less tightly bound nuclei are transmuted into more tightly bound nuclei. In fission, a heavy nucleus like ${}^{235}_{92}\mathrm{U}$ breaks into two smaller fragments, e.g., ${}^{235}_{92}\mathrm{U} + {}^{1}_{0}\mathrm{n} \rightarrow {}^{133}_{51}\mathrm{Sb} + {}^{99}_{41}\mathrm{Nb} + 4\,{}^{1}_{0}\mathrm{n}$

13. The fact that more neutrons are produced in fission than are consumed gives the possibility of a chain reaction with each neutron that is produced triggering another fission. The chain reaction is uncontrolled and rapid in a nuclear bomb explosion. It is controlled and steady in a nuclear reactor. In a reactor, the value of the neutron multiplication factor k is maintained at 1.

14. In fusion, lighter nuclei combine to form a larger nucleus. Fusion of hydrogen nuclei into helium nuclei is the source of energy of all stars including our sun.

Physical Quantity	Symbol	Dimensions	Units	Remarks
Atomic mass unit		[M]	u	Unit of mass for expressing atomic or nuclear masses. One atomic mass unit equals $1/12^{th}$ of the mass of ${}^{12}C$ atom.
Disintegration or decay constant	λ	$[T^{-1}]$	s^{-1}	
Half-life	$T_{1/2}$	[T]	s	Time taken for the decay of one-half of the initial number of nuclei present in a radioactive sample.
Mean life	τ	[T]	s	Time at which number of nuclei has been reduced to e^{-1} of its initial value
Activity of a radio-active sample	R	$[T^{-1}]$	Bq	Measure of the activity of a radioactive source.

POINTS TO PONDER

1. The density of nuclear matter is independent of the size of the nucleus. The mass density of the atom does not follow this rule.
2. The radius of a nucleus determined by electron scattering is found to be slightly different from that determined by alpha-particle scattering.

This is because electron scattering senses the charge distribution of the nucleus, whereas alpha and similar particles sense the nuclear matter.

3. After Einstein showed the equivalence of mass and energy, $E = mc^2$, we cannot any longer speak of separate laws of conservation of mass and conservation of energy, but we have to speak of a unified law of conservation of mass and energy. The most convincing evidence that this principle operates in nature comes from nuclear physics. It is central to our understanding of nuclear energy and harnessing it as a source of power. Using the principle, Q of a nuclear process (decay or reaction) can be expressed also in terms of initial and final masses.
4. The nature of the binding energy (per nucleon) curve shows that exothermic nuclear reactions are possible, when two light nuclei fuse or when a heavy nucleus undergoes fission into nuclei with intermediate mass.
5. For fusion, the light nuclei must have sufficient initial energy to overcome the coulomb potential barrier. That is why fusion requires very high temperatures.
6. Although the binding energy (per nucleon) curve is smooth and slowly varying, it shows peaks at nuclides like ^{4}He, ^{16}O etc. This is considered as evidence of atom-like shell structure in nuclei.
7. Electrons and positron are a particle-antiparticle pair. They are identical in mass; their charges are equal in magnitude and opposite. (It is found that when an electron and a positron come together, they annihilate each other giving energy in the form of gamma-ray photons.)
8. In â$^-$-decay (electron emission), the particle emitted along with electron is anti-neutrino ($\bar{\nu}$). On the other hand, the particle emitted in β^+-decay (positron emission) is neutrino (ν). Neutrino and anti-neutrino are a particle-antiparticle pair. There are anti particles associated with every particle. What should be antiproton which is the anti particle of the proton?
9. A free neutron is unstable ($n \rightarrow p + e^- + \bar{\nu}$). But a similar free proton decay is not possible, since a proton is (slightly) lighter than a neutron.
10. Gamma emission usually follows alpha or beta emission. A nucleus in an excited (higher) state goes to a lower state by emitting a gamma photon. A nucleus may be left in an excited state after alpha or beta emission. Successive emission of gamma rays from the same nucleus (as in case of ^{60}Ni, Fig. 13.4) is a clear proof that nuclei also have discrete energy levels as do the atoms.
11. Radioactivity is an indication of the instability of nuclei. Stability requires the ratio of neutron to proton to be around 1:1 for light nuclei. This ratio increases to about 3:2 for heavy nuclei. (More neutrons are required to overcome the effect of repulsion among the protons.) Nuclei which are away from the stability ratio, i.e., nuclei which have an excess of neutrons or protons are unstable. In fact, only about 10% of knon isotopes (of all elements), are stable. Others have been either artificially produced in the laboratory by bombarding α, p, d, n or other particles on targets of stable nuclear species or identified in astronomical observations of matter in the universe.

EXERCISES

You may find the following data useful in solving the exercises:

$e = 1.6\times10^{-19}$C $\quad N = 6.023\times10^{23}$ per mole

$1/(4\pi\varepsilon_0) = 9 \times 10^9$ N m^2/C^2 $\quad k = 1.381\times10^{-23}$J ^{0}K^{-1}

1 MeV $= 1.6\times10^{-13}$J $\quad$ 1 u = 931.5 MeV/c^2

1 year $= 3.154\times10^7$ s

$m_H = 1.007825$ u $\quad m_n = 1.008665$ u

$m(^4_2\text{He}) = 4.002603$ u $\quad m_e = 0.000548$ u

13.1 (a) Two stable isotopes of lithium $^6_3$Li and $^7_3$Li have respective abundances of 7.5% and 92.5%. These isotopes have masses 6.01512 u and 7.01600 u, respectively. Find the atomic mass of lithium.

(b) Boron has two stable isotopes, $^{10}_5$B and $^{11}_5$B. Their respective masses are 10.01294 u and 11.00931 u, and the atomic mass of boron is 10.811 u. Find the abundances of $^{10}_5$B and $^{11}_5$B.

13.2 The three stable isotopes of neon: $^{20}_{10}$Ne, $^{21}_{10}$Ne and $^{22}_{10}$Ne have respective abundances of 90.51%, 0.27% and 9.22%. The atomic masses of the three isotopes are 19.99 u, 20.99 u and 21.99 u, respectively. Obtain the average atomic mass of neon.

13.3 Obtain the binding energy (in MeV) of a nitrogen nucleus $\left(^{14}_7\text{N}\right)$, given $m\left(^{14}_7\text{N}\right)$=14.00307 u

13.4 Obtain the binding energy of the nuclei $^{56}_{26}$Fe and $^{209}_{83}$Bi in units of MeV from the following data:

$m(^{56}_{26}\text{Fe}) = 55.934939$ u $\quad m(^{209}_{83}\text{Bi}) = 208.980388$ u

13.5 A given coin has a mass of 3.0 g. Calculate the nuclear energy that would be required to separate all the neutrons and protons from each other. For simplicity assume that the coin is entirely made of $^{63}_{29}$Cu atoms (of mass 62.92960 u).

13.6 Write nuclear reaction equations for

(i) α-decay of $^{226}_{88}$Ra $\quad$ (ii) α-decay of $^{242}_{94}$Pu

(iii) β^--decay of $^{32}_{15}$P $\quad$ (iv) β^--decay of $^{210}_{83}$Bi

(v) β^+-decay of $^{11}_6$C $\quad$ (vi) β^+-decay of $^{97}_{43}$Tc

(vii) Electron capture of $^{120}_{54}$Xe

13.7 A radioactive isotope has a half-life of T years. How long will it take the activity to reduce to a) 3.125%, b) 1% of its original value?

13.8 The normal activity of living carbon-containing matter is found to be about 15 decays per minute for every gram of carbon. This activity arises from the small proportion of radioactive $^{14}_6$C present with the stable carbon isotope $^{12}_6$C. When the organism is dead, its interaction with the atmosphere (which maintains the above equilibrium activity) ceases and its activity begins to drop. From the known half-life (5730 years) of $^{14}_6$C, and the measured activity, the age of the specimen can be approximately estimated. This is the principle of $^{14}_6$C dating

used in archaeology. Suppose a specimen from Mohenjodaro gives an activity of 9 decays per minute per gram of carbon. Estimate the approximate age of the Indus-Valley civilisation.

13.9 Obtain the amount of $^{60}_{27}\text{Co}$ necessary to provide a radioactive source of 8.0 mCi strength. The half-life of $^{60}_{27}\text{Co}$ is 5.3 years.

13.10 The half-life of $^{90}_{38}\text{Sr}$ is 28 years. What is the disintegration rate of 15 mg of this isotope?

13.11 Obtain approximately the ratio of the nuclear radii of the gold isotope $^{197}_{79}\text{Au}$ and the silver isotope $^{107}_{47}\text{Ag}$.

13.12 Find the Q-value and the kinetic energy of the emitted α-particle in the α-decay of (a) $^{226}_{88}\text{Ra}$ and (b) $^{220}_{86}\text{Rn}$.

Given $m\,(^{226}_{88}\text{Ra}) = 226.02540$ u, $m\,(^{222}_{86}\text{Rn}) = 222.01750$ u,

$m\,(^{222}_{86}\text{Rn}) = 220.01137$ u, $m\,(^{216}_{84}\text{Po}) = 216.00189$ u.

13.13 The radionuclide ^{11}C decays according to

$$^{11}_{6}\text{C} \rightarrow ^{11}_{5}\text{B} + e^{+} + \nu: \quad T_{1/2} = 20.3 \text{ min}$$

The maximum energy of the emitted positron is 0.960 MeV.

Given the mass values:

$m\,(^{11}_{6}\text{C}) = 11.011434$ u and $m\,(^{11}_{6}\text{B}) = 11.009305$ u,

calculate Q and compare it with the maximum energy of the positron emitted.

13.14 The nucleus $^{23}_{10}\text{Ne}$ decays by β^- emission. Write down the β-decay equation and determine the maximum kinetic energy of the electrons emitted. Given that:

$m\,(^{23}_{10}\text{Ne}) = 22.994466$ u

$m\,(^{23}_{11}\text{Na}) = 22.089770$ u.

13.15 The Q value of a nuclear reaction $A + b \rightarrow C + d$ is defined by

$Q = [\,m_A + m_b - m_C - m_d]c^2$

where the masses refer to the respective nuclei. Determine from the given data the Q-value of the following reactions and state whether the reactions are exothermic or endothermic.

(i) $^{1}_{1}\text{H} + ^{3}_{1}\text{H} \rightarrow ^{2}_{1}\text{H} + ^{2}_{1}\text{H}$

(ii) $^{12}_{6}\text{C} + ^{12}_{6}\text{C} \rightarrow ^{20}_{10}\text{Ne} + ^{4}_{2}\text{He}$

Atomic masses are given to be

$m\,(^{2}_{1}\text{H}) = 2.014102$ u

$m\,(^{3}_{1}\text{H}) = 3.016049$ u

$m\,(^{12}_{6}\text{C}) = 12.000000$ u

$m\,(^{20}_{10}\text{Ne}) = 19.992439$ u

13.16 Suppose, we think of fission of a $^{56}_{26}\text{Fe}$ nucleus into two equal fragments, $^{28}_{13}\text{Al}$. Is the fission energetically possible? Argue by working out Q of the process. Given $m\,(^{56}_{26}\text{Fe}) = 55.93494$ u and $m\,(^{28}_{13}\text{Al}) = 27.98191$ u.

13.17 The fission properties of $^{239}_{94}\text{Pu}$ are very similar to those of $^{235}_{92}\text{U}$. The average energy released per fission is 180 MeV. How much energy,

in MeV, is released if all the atoms in 1 kg of pure $^{239}_{94}Pu$ undergo fission?

13.18 A 1000 MW fission reactor consumes half of its fuel in 5.00 y. How much $^{235}_{92}U$ did it contain initially? Assume that the reactor operates 80% of the time, that all the energy generated arises from the fission of $^{235}_{92}U$ and that this nuclide is consumed only by the fission process.

13.19 How long can an electric lamp of 100W be kept glowing by fusion of 2.0 kg of deuterium? Take the fusion reaction as

$$^{2}_{1}H + ^{2}_{1}H \rightarrow ^{3}_{2}He + n + 3.27 \text{ MeV}$$

13.20 Calculate the height of the potential barrier for a head on collision of two deuterons. (Hint: The height of the potential barrier is given by the Coulomb repulsion between the two deuterons when they just touch each other. Assume that they can be taken as hard spheres of radius 2.0 fm.)

13.21 From the relation $R = R_0 A^{1/3}$, where R_0 is a constant and A is the mass number of a nucleus, show that the nuclear matter density is nearly constant (i.e. independent of A).

13.22 For the β^+ (positron) emission from a nucleus, there is another competing process known as electron capture (electron from an inner orbit, say, the K–shell, is captured by the nucleus and a neutrino is emitted).

$$e^+ + ^{A}_{Z}X \rightarrow ^{A}_{Z-1}Y + \nu$$

Show that if β^+ emission is energetically allowed, electron capture is necessarily allowed but not vice–versa.

ADDITIONAL EXERCISES

13.23 In a periodic table the average atomic mass of magnesium is given as 24.312 u. The average value is based on their relative natural abundance on earth. The three isotopes and their masses are $^{24}_{12}Mg$ (23.98504u), $^{25}_{12}Mg$ (24.98584u) and $^{26}_{12}Mg$ (25.98259u). The natural abundance of $^{24}_{12}Mg$ is 78.99% by mass. Calculate the abundances of other two isotopes.

13.24 The neutron separation energy is defined as the energy required to remove a neutron from the nucleus. Obtain the neutron separation energies of the nuclei $^{41}_{20}Ca$ and $^{27}_{13}Al$ from the following data:

$m(^{40}_{20}Ca) = 39.962591$ u

$m(^{41}_{20}Ca) = 40.962278$ u

$m(^{26}_{13}Al) = 25.986895$ u

$m(^{27}_{13}Al) = 26.981541$ u

13.25 A source contains two phosphorous radio nuclides $^{32}_{15}P$ ($T_{1/2}$ = 14.3d) and $^{33}_{15}P$ ($T_{1/2}$ = 25.3d). Initially, 10% of the decays come from $^{33}_{15}P$. How long one must wait until 90% do so?

13.26 Under certain circumstances, a nucleus can decay by emitting a particle more massive than an α-particle. Consider the following decay processes:

$$^{223}_{88}Ra \rightarrow ^{209}_{82}Pb + ^{14}_{6}C$$

$^{223}_{88}\text{Ra} \rightarrow {}^{219}_{86}\text{Rn} + {}^{4}_{2}\text{He}$

Calculate the Q-values for these decays and determine that both are energetically allowed.

13.27 Consider the fission of $^{238}_{92}\text{U}$ by fast neutrons. In one fission event, no neutrons are emitted and the final end products, after the beta decay of the primary fragments, are $^{140}_{58}\text{Ce}$ and $^{99}_{44}\text{Ru}$. Calculate Q for this fission process. The relevant atomic and particle masses are

$m(^{238}_{92}\text{U}) = 238.05079$ u

$m(^{140}_{58}\text{Ce}) = 139.90543$ u

$m(^{99}_{44}\text{Ru}) = 98.90594$ u

13.28 Consider the D–T reaction (deuterium–tritium fusion)

$^{2}_{1}\text{H} + {}^{3}_{1}\text{H} \rightarrow {}^{4}_{2}\text{He} + \text{n}$

(a) Calculate the energy released in MeV in this reaction from the data:

$m(^{2}_{1}\text{H}) = 2.014102$ u

$m(^{3}_{1}\text{H}) = 3.016049$ u

(b) Consider the radius of both deuterium and tritium to be approximately 2.0 fm. What is the kinetic energy needed to overcome the coulomb repulsion between the two nuclei? To what temperature must the gas be heated to initiate the reaction?

(Hint: Kinetic energy required for one fusion event =average thermal kinetic energy available with the interacting particles = $2(3kT/2)$; k = Boltzman's constant, T = absolute temperature.)

13.29 Obtain the maximum kinetic energy of β-particles, and the radiation frequencies of γ decays in the decay scheme shown in Fig. 13.6. You are given that

$m(^{198}\text{Au}) = 197.968233$ u

$m(^{198}\text{Hg}) = 197.966760$ u

FIGURE13.6

13.30 Calculate and compare the energy released by a) fusion of 1.0 kg of hydrogen deep within Sun and b) the fission of 1.0 kg of ^{235}U in a fission reactor.

13.31 Suppose India had a target of producing by 2020 AD, 200,000 MW of electric power, ten percent of which was to be obtained from nuclear power plants. Suppose we are given that, on an average, the efficiency of utilization (i.e. conversion to electric energy) of thermal energy produced in a reactor was 25%. How much amount of fissionable uranium would our country need per year by 2020? Take the heat energy per fission of ^{235}U to be about 200MeV.

Chapter Fourteen

SEMICONDUCTOR ELECTRONICS: MATERIALS, DEVICES AND SIMPLE CIRCUITS

14.1 Introduction

Devices in which a controlled flow of electrons can be obtained are the basic *building blocks* of all the electronic circuits. Before the discovery of transistor in 1948, such devices were mostly vacuum tubes (also called valves) like the vacuum diode which has two electrodes, viz., anode (often called plate) and cathode; triode which has three electrodes – cathode, plate and grid; tetrode and pentode (respectively with 4 and 5 electrodes). In a vacuum tube, the electrons are supplied by a heated cathode and the controlled flow of these electrons *in vacuum* is obtained by varying the voltage between its different electrodes. Vacuum is required in the inter-electrode space; otherwise the moving electrons may lose their energy on collision with the air molecules in their path. In these devices the electrons can flow only from the cathode to the anode (i.e., only in one direction). Therefore, such devices are generally referred to as *valves*. These vacuum tube devices are bulky, consume high power, operate generally at high voltages (~100 V) and have limited life and low reliability. The seed of the development of modern *solid-state semiconductor electronics* goes back to 1930's when it was realised that some solid-state semiconductors and their junctions offer the possibility of controlling the number and the direction of flow of charge carriers through them. Simple excitations like light, heat or small applied voltage can change the number of mobile charges in a semiconductor. Note that the supply

and flow of charge carriers in the semiconductor devices are *within the solid itself*, while in the earlier vacuum tubes/valves, the mobile electrons were obtained from a heated cathode and they were made to flow in an *evacuated* space or vacuum. No external heating or large evacuated space is required by the semiconductor devices. They are small in size, consume low power, operate at low voltages and have long life and high reliability. Even the Cathode Ray Tubes (CRT) used in television and computer monitors which work on the principle of vacuum tubes are being replaced by Liquid Crystal Display (LCD) monitors with supporting solid state electronics. Much before the full implications of the semiconductor devices was formally understood, a naturally occurring crystal of *galena* (Lead sulphide, PbS) with a metal point contact attached to it was used as *detector* of radio waves.

In the following sections, we will introduce the basic concepts of semiconductor physics and discuss some semiconductor devices like junction diodes (a 2-electrode device) and bipolar junction transistor (a 3-electrode device). A few circuits illustrating their applications will also be described.

14.2 Classification of Metals, Conductors and Semiconductors

On the basis of conductivity

On the basis of the relative values of electrical conductivity (σ) or resistivity ($\rho = 1/\sigma$), the solids are broadly classified as:

(i) ***Metals:*** They possess very low resistivity (or high conductivity).
$\rho \sim 10^{-2} - 10^{-8}\ \Omega$ m
$\sigma \sim 10^{2} - 10^{8}$ S m^{-1}

(ii) ***Semiconductors:*** They have resistivity or conductivity intermediate to metals and insulators.
$\rho \sim 10^{-5} - 10^{6}\ \Omega$ m
$\sigma \sim 10^{5} - 10^{-6}$ S m^{-1}

(iii) ***Insulators:*** They have high resistivity (or low conductivity).
$\rho \sim 10^{11} - 10^{19}\ \Omega$ m
$\sigma \sim 10^{-11} - 10^{-19}$ S m^{-1}

The values of ρ and σ given above are indicative of magnitude and could well go outside the ranges as well. Relative values of the resistivity are not the only criteria for distinguishing metals, insulators and semiconductors from each other. There are some other differences, which will become clear as we go along in this chapter.

Our interest in this chapter is in the study of semiconductors which could be:

(i) *Elemental semiconductors*: Si and Ge

(ii) *Compound semiconductors*: Examples are:
- Inorganic: CdS, GaAs, CdSe, InP, etc.
- Organic: anthracene, doped pthalocyanines, etc.
- Organic polymers: polypyrrole, polyaniline, polythiophene, etc.

Most of the currently available semiconductor devices are based on elemental semiconductors Si or Ge and compound *inorganic*

semiconductors. However, after 1990, a few semiconductor devices using organic semiconductors and semiconducting polymers have been developed signalling the birth of a futuristic technology of polymer-electronics and molecular-electronics. In this chapter, we will restrict ourselves to the study of inorganic semiconductors, particularly elemental semiconductors Si and Ge. The general concepts introduced here for discussing the elemental semiconductors, by-and-large, apply to most of the compound semiconductors as well.

On the basis of energy bands

According to the Bohr atomic model, in an *isolated atom* the energy of any of its electrons is decided by the orbit in which it revolves. But when the atoms come together to form a solid they are close to each other. So the outer orbits of electrons from neighbouring atoms would come very close or could even overlap. This would make the nature of electron motion in a solid very different from that in an isolated atom.

Inside the crystal each electron has a unique position and no two electrons see exactly the same pattern of surrounding charges. Because of this, each electron will have a different *energy level*. These different energy levels with continuous energy variation form what are called *energy bands*. The energy band which includes the energy levels of the valence electrons is called the *valence band*. The energy band above the valence band is called the *conduction band*. With no external energy, all the valence electrons will reside in the valence band. If the lowest level in the conduction band happens to be lower than the highest level of the valence band, the electrons from the valence band can easily move into the conduction band. Normally the conduction band is empty. But when it overlaps on the valence band electrons can move freely into it. This is the case with metallic conductors.

If there is some gap between the conduction band and the valence band, electrons in the valence band all remain bound and no free electrons are available in the conduction band. This makes the material an insulator. But some of the electrons from the valence band may gain external energy to cross the gap between the conduction band and the valence band. Then these electrons will move into the conduction band. At the same time they will create vacant energy levels in the valence band where other valence electrons can move. Thus the process creates the possibility of conduction due to electrons in conduction band as well as due to vacancies in the valence band.

Let us consider what happens in the case of Si or Ge crystal containing N atoms. For Si, the outermost orbit is the third orbit ($n = 3$), while for Ge it is the fourth orbit ($n = 4$). The number of electrons in the outermost orbit is 4 ($2s$ and $2p$ electrons). Hence, the total number of outer electrons in the crystal is $4N$. The maximum possible number of electrons in the outer orbit is 8 ($2s + 6p$ electrons). So, for the $4N$ valence electrons there are $8N$ available energy states. These $8N$ discrete energy levels can either form a continuous band or they may be grouped in different bands depending upon the distance between the atoms in the crystal (see box on Band Theory of Solids).

At the distance between the atoms in the crystal lattices of Si and Ge, the energy band of these $8N$ states is split apart into two which are separated by an *energy gap* E_g (Fig. 14.1). The lower band which is

completely occupied by the $4N$ valence electrons at temperature of absolute zero is the *valence band.* The other band consisting of $4N$ energy states, called the *conduction band*, is completely empty at absolute zero.

Band theory of solids

Consider that the Si or Ge crystal contains N atoms. Electrons of each atom will have discrete energies in different orbits. The electron energy will be same if all the atoms are *isolated*, i.e., separated from each other by a large distance. However, in a crystal, the atoms are close to each other (2 to 3 Å) and therefore the electrons interact with each other and also with the neighbouring atomic cores. The overlap (or interaction) will be more felt by the electrons in the outermost orbit while the inner orbit or core electron energies may remain unaffected. Therefore, for understanding electron energies in Si or Ge crystal, we need to consider the changes in the energies of the electrons in the outermost orbit only. For Si, the outermost orbit is the third orbit (n = 3), while for Ge it is the fourth orbit (n = 4). The number of electrons in the outermost orbit is 4 ($2s$ and $2p$ electrons). Hence, the total number of outer electrons in the crystal is $4N$. The maximum possible number of outer electrons in the orbit is 8 ($2s + 6p$ electrons). So, out of the $4N$ electrons, $2N$ electrons are in the $2N$ *s-states* (orbital quantum number $l = 0$) and 2N electrons are in the available $6N$ *p-states*. Obviously, some p-electron states are empty as shown in the extreme right of Figure. This is the case of well separated or isolated atoms [region A of Figure].

Suppose these atoms start coming nearer to each other to form a solid. The energies of these electrons in the outermost orbit may change (both increase and decrease) due to the interaction between the electrons of different atoms. The $6N$ states for $l = 1$, which originally had identical energies in the isolated atoms, spread out and form an *energy band* [region B in Figure]. Similarly, the $2N$ states for $l = 0$, having identical energies in the isolated atoms, split into a second band (carefully see the region B of Figure) separated from the first one by an *energy gap*.

At still smaller spacing, however, there comes a region in which the bands merge with each other. The lowest energy state that is a split from the upper atomic level appears to drop below the upper state that has come from the lower atomic level. In this region (region C in Figure), *no energy gap exists where the upper and lower energy states get mixed.*

Finally, if the distance between the atoms further decreases, the energy bands again split apart and are separated by an *energy gap* E_g (region D in Figure). The total number of available energy states $8N$ has been *re-apportioned* between the two bands ($4N$ states each in the lower and upper energy bands). Here the significant point is that there are exactly as many states in the lower band ($4N$) as there are available valence electrons from the atoms ($4N$).

Therefore, this band (called the *valence band*) is completely filled while the upper band is completely empty. The upper band is called the *conduction band.*

The lowest energy level in the conduction band is shown as E_C and highest energy level in the valence band is shown as E_V. Above E_C and below E_V there are a large number of closely spaced energy levels, as shown in Fig. 14.1.

The gap between the top of the valence band and bottom of the conduction band is called the *energy band gap* (Energy gap E_g). It may be large, small, or zero, depending upon the material. These different situations, are depicted in Fig. 14.2 and discussed below:

FIGURE 14.1 The energy band positions in a semiconductor at 0 K. The upper band, called the conduction band, consists of infinitely large number of closely spaced energy states. The lower band, called the valence band, consists of closely spaced completely filled energy states.

Case I: This refers to a situation, as shown in Fig. 14.2(a). One can have a metal either when the conduction band is partially filled and the balanced band is partially empty or when the conduction and valance bands overlap. When there is overlap electrons from valence band can easily move into the conduction band. This situation makes a large number of electrons available for electrical conduction. When the valence band is partially empty, electrons from its lower level can move to higher level making conduction possible. Therefore, the resistance of such materials is low or the conductivity is high.

FIGURE 14.2 Difference between energy bands of (a) metals, (b) insulators and (c) semiconductors.

Case II: In this case, as shown in Fig. 14.2(b), a large band gap E_g exists ($E_g > 3$ eV). There are no electrons in the conduction band, and therefore no electrical conduction is possible. Note that the energy gap is so large that electrons cannot be excited from the valence band to the conduction band by thermal excitation. This is the case of *insulators*.

Case III: This situation is shown in Fig. 14.2(c). Here a finite but small band gap ($E_g < 3$ eV) exists. Because of the small band gap, at room temperature some electrons from valence band can acquire enough energy to cross the energy gap and enter the *conduction band*. These electrons (though small in numbers) can move in the conduction band. Hence, the resistance of *semiconductors* is not as high as that of the insulators.

In this section we have made a broad classification of metals, conductors and semiconductors. In the section which follows you will learn the conduction process in semiconductors.

14.3 Intrinsic Semiconductor

We shall take the most common case of Ge and Si whose lattice structure is shown in Fig. 14.3. These structures are called the diamond-like structures. Each atom is surrounded by four nearest neighbours. We know that Si and Ge have four valence electrons. In its crystalline structure, every Si or Ge atom tends *to share* one of its four valence electrons with each of its four nearest neighbour atoms, and also *to take share* of one electron from each such neighbour. These shared electron pairs are referred to as forming a *covalent bond* or simply a *valence bond*. The two shared electrons can be assumed to shuttle back-and-forth between the associated atoms holding them together strongly. Figure 14.4 schematically shows the 2-dimensional representation of Si or Ge structure shown in Fig. 14.3 which overemphasises the covalent bond. It shows an idealised picture in which no bonds are broken (all bonds are intact). Such a situation arises at low temperatures. As the temperature increases, more thermal energy becomes available to these electrons and some of these electrons may break–away (becoming *free* electrons contributing to conduction). The thermal energy effectively ionises only a few atoms in the crystalline lattice and creates a *vacancy* in the bond as shown in Fig. 14.5(a). The neighbourhood, from which the free electron (with charge $-q$) has come out leaves a vacancy with an effective charge ($+q$). This *vacancy* with the effective positive electronic charge is called a *hole*. The hole behaves as an *apparent free particle* with effective positive charge.

FIGURE 14.3 Three-dimensional diamond-like crystal structure for Carbon, Silicon or Germanium with respective lattice spacing a equal to 3.56, 5.43 and 5.66 Å.

In intrinsic semiconductors, the number of free electrons, n_e is equal to the number of holes, n_h. That is

$$n_e = n_h = n_i \qquad (14.1)$$

where n_i is called intrinsic carrier concentration.

Semiconductors posses the unique property in which, apart from electrons, the holes also move.

Suppose there is a hole at site 1 as shown in Fig. 14.5(a). The movement of holes can be visualised as shown in Fig. 14.5(b). An electron from the covalent bond at site 2 may jump to the vacant site 1 (hole). Thus, after such a jump, the hole is at site 2 and the site 1 has now an electron. Therefore, apparently, the hole has moved from site 1 to site 2. Note that the electron originally set free [Fig. 14.5(a)] is not involved in this process of hole motion. The free electron moves completely independently as conduction electron and gives rise to an electron current, I_e under an applied electric field. Remember that the motion of hole is only a convenient way of describing the actual motion of *bound* electrons, whenever there is an empty bond anywhere in the crystal. Under the action of an electric field, these holes move towards negative potential giving the hole current, I_h. The total current, I is thus the sum of the electron current I_e and the hole current I_h:

FIGURE 14.4 Schematic two-dimensional representation of Si or Ge structure showing covalent bonds at low temperature (all bonds intact). +4 symbol indicates inner cores of Si or Ge.

$$I = I_e + I_h \tag{14.2}$$

It may be noted that apart from the *process of generation* of conduction electrons and holes, a simultaneous *process of recombination* occurs in which the electrons *recombine* with the holes. At equilibrium, the rate of generation is equal to the rate of recombination of charge carriers. The recombination occurs due to an electron colliding with a hole.

FIGURE 14.5 (a) Schematic model of generation of hole at site 1 and conduction electron due to thermal energy at moderate temperatures. (b) Simplified representation of possible thermal motion of a hole. The electron from the lower left hand covalent bond (site 2) goes to the earlier hole site1, leaving a hole at its site indicating an apparent movement of the hole from site 1 to site 2.

FIGURE 14.6 (a) An intrinsic semiconductor at $T = 0$ K behaves like insulator. (b) At $T > 0$ K, four thermally generated electron-hole pairs. The filled circles (•) represent electrons and empty fields (○) represent holes.

An intrinsic semiconductor will behave like an insulator at $T = 0$ K as shown in Fig. 14.6(a). It is the thermal energy at higher temperatures ($T > 0$K), which excites some electrons from the valence band to the conduction band. These thermally excited electrons at $T > 0$ K, partially occupy the conduction band. Therefore, the energy-band diagram of an intrinsic semiconductor will be as shown in Fig. 14.6(b). Here, some electrons are shown in the conduction band. These have come from the valence band leaving equal number of holes there.

EXAMPLE 14.1

Example 14.1 C, Si and Ge have same lattice structure. Why is C insulator while Si and Ge intrinsic semiconductors?

Solution The 4 bonding electrons of C, Si or Ge lie, respectively, in the second, third and fourth orbit. Hence, energy required to take out an electron from these atoms (i.e., ionisation energy E_g) will be least for Ge, followed by Si and highest for C. Hence, number of free electrons for conduction in Ge and Si are significant but negligibly small for C.

14.4 EXTRINSIC SEMICONDUCTOR

The conductivity of an intrinsic semiconductor depends on its temperature, but at room temperature its conductivity is very low. As such, no important electronic devices can be developed using these semiconductors. Hence there is a necessity of improving their conductivity. This can be done by making use of impurities.

When a small amount, say, a few parts per million (ppm), of a suitable impurity is added to the pure semiconductor, the conductivity of the semiconductor is increased manifold. Such materials are known as *extrinsic semiconductors* or *impurity semiconductors*. The deliberate addition of a desirable impurity is called *doping* and the impurity atoms are called *dopants*. Such a material is also called a *doped semiconductor*. The dopant has to be such that it does not distort the original pure semiconductor lattice. It occupies only a very few of the original semiconductor atom sites in the crystal. A necessary condition to attain this is that the sizes of the dopant and the semiconductor atoms should be nearly the same.

There are two types of dopants used in doping the tetravalent Si or Ge:

(i) Pentavalent (valency 5); like Arsenic (As), Antimony (Sb), Phosphorous (P), etc.

(ii) Trivalent (valency 3); like Indium (In), Boron (B), Aluminium (Al), etc.

We shall now discuss how the doping changes the number of charge carriers (and hence the conductivity) of semiconductors. Si or Ge belongs to the fourth group in the Periodic table and, therefore, we choose the dopant element from nearby fifth or third group, expecting and taking care that the size of the dopant atom is nearly the same as that of Si or Ge. Interestingly, the pentavalent and trivalent dopants in Si or Ge give two entirely different types of semiconductors as discussed below.

FIGURE 14.7 (a) Pentavalent donor atom (As, Sb, P, etc.) doped for tetravalent Si or Ge giving n-type semiconductor, and (b) Commonly used schematic representation of n-type material which shows only the fixed cores of the substituent donors with one additional effective positive charge and its associated extra electron.

(i) n-type semiconductor

Suppose we dope Si or Ge with a pentavalent element as shown in Fig. 14.7. When an atom of +5 valency element occupies the position of an atom in the crystal lattice of Si, four of its electrons bond with the four silicon neighbours while the fifth remains very weakly bound to its parent atom. This is because the four electrons participating in bonding are seen as part of the effective core of the atom by the fifth electron. As a result the ionisation energy required to set this electron free is very small and even at room temperature it will be free to move in the lattice of the semiconductor. For example, the energy required is ~ 0.01 eV for germanium, and 0.05 eV for silicon, to separate this electron from its atom. This is in contrast to the energy required to jump the forbidden band (about 0.72 eV for germanium and about 1.1 eV for silicon) at room temperature in the intrinsic semiconductor. Thus, the pentavalent dopant is donating one extra electron for conduction and hence is known as *donor* impurity. The number of electrons made available for conduction by dopant atoms depends strongly upon the doping level and is independent of any increase in ambient temperature. On the other hand, the number of free electrons (with an equal number of holes) generated by Si atoms, increases weakly with temperature.

In a doped semiconductor the total number of conduction electrons n_e is due to the electrons contributed by donors and those generated intrinsically, while the total number of holes n_h is only due to the holes from the intrinsic source. But the rate of recombination of holes would increase due to the increase in the number of electrons. As a result, the number of holes would get reduced further.

Thus, with proper level of doping the number of conduction electrons can be made much larger than the number of holes. Hence in an extrinsic

FIGURE 14.8 (a) Trivalent acceptor atom (In, Al, B etc.) doped in tetravalent Si or Ge lattice giving p-type semiconductor. (b) Commonly used schematic representation of p-type material which shows only the fixed core of the substituent acceptor with one effective additional negative charge and its associated hole.

semiconductor doped with pentavalent impurity, electrons become the *majority carriers* and holes the *minority carriers*. These semiconductors are, therefore, known as n-*type semiconductors*. For n-type semiconductors, we have,

$$n_e >> n_h \tag{14.3}$$

(ii) p-type semiconductor

This is obtained when Si or Ge is doped with a trivalent impurity like Al, B, In, etc. The dopant has one valence electron less than Si or Ge and, therefore, this atom can form covalent bonds with neighbouring three Si atoms but does not have any electron to offer to the fourth Si atom. So the bond between the fourth neighbour and the trivalent atom has a vacancy or hole as shown in Fig. 14.8. Since the neighbouring Si atom in the lattice wants an electron in place of a hole, an electron in the outer orbit of an atom in the neighbourhood may jump to fill this vacancy, leaving a vacancy or hole at its own site. Thus the *hole* is available for conduction. Note that the trivalent foreign atom becomes effectively negatively charged when it shares fourth electron with neighbouring Si atom. Therefore, the dopant atom of p-type material can be treated as *core of one negative charge* along with its associated hole as shown in Fig. 14.8(b). It is obvious that one *acceptor* atom gives one *hole*. These holes are in addition to the intrinsically generated holes while the source of conduction electrons is only intrinsic generation. Thus, for such a material, the holes are the majority carriers and electrons are minority carriers. Therefore, extrinsic semiconductors doped with trivalent impurity are called p-*type semiconductors*. For p-type semiconductors, the recombination process will reduce the number (n_i)of intrinsically generated electrons to n_e. We have, for p-type semiconductors

$$n_h >> n_e \tag{14.4}$$

Note that *the crystal maintains an overall charge neutrality as the charge of additional charge carriers is just equal and opposite to that of the ionised cores in the lattice.*

In extrinsic semiconductors, because of the abundance of majority current carriers, the minority carriers produced thermally have more chance of meeting majority carriers and thus getting destroyed. Hence, the dopant, by adding a large number of current carriers of one type, which become the majority carriers, indirectly helps to reduce the intrinsic concentration of minority carriers.

The semiconductor's energy band structure is affected by doping. In the case of extrinsic semiconductors, additional energy states due to donor impurities (E_D) and acceptor impurities (E_A) also exist. In the energy band diagram of n-type Si semiconductor, the donor energy level E_D is slightly below the bottom E_C of the conduction band and electrons from this level move into the conduction band with very small supply of energy. At room temperature, most of the donor atoms get ionised but very few (~10^{-12}) atoms of Si get ionised. So the conduction band will have most electrons coming from the donor impurities, as shown in Fig. 14.9(a). Similarly,

for p-type semiconductor, the acceptor energy level E_A is slightly above the top E_V of the valence band as shown in Fig. 14.9(b). With very small supply of energy an electron from the valence band can jump to the level E_A and ionise the acceptor negatively. (Alternately, we can also say that with very small supply of energy the hole from level E_A sinks down into the valence band. Electrons rise up and holes fall down when they gain external energy.) At room temperature, most of the acceptor atoms get ionised leaving holes in the valence band. Thus at room temperature the density of holes in the valence band is predominantly due to impurity in the extrinsic semiconductor. The electron and hole concentration in a semiconductor *in thermal equilibrium* is given by

$$n_e n_h = n_i^2 \tag{14.5}$$

Though the above description is grossly approximate and hypothetical, it helps in understanding the difference between metals, insulators and semiconductors (extrinsic and intrinsic) in a simple manner. The difference in the resistivity of C, Si and Ge depends upon the energy gap between their conduction and valence bands. For C (diamond), Si and Ge, the energy gaps are 5.4 eV, 1.1 eV and 0.7 eV, respectively. Sn also is a group IV element but it is a metal because the energy gap in its case is 0 eV.

FIGURE 14.9 Energy bands of (a) n-type semiconductor at $T > 0$K, (b) p-type semiconductor at $T > 0$K.

Example 14.2 Suppose a pure Si crystal has 5×10^{28} atoms m^{-3}. It is doped by 1 ppm concentration of pentavalent As. Calculate the number of electrons and holes. Given that $n_i = 1.5 \times 10^{16}$ m^{-3}.

Solution Note that thermally generated electrons ($n_i \sim 10^{16}$ m^{-3}) are negligibly small as compared to those produced by doping.
Therefore, $n_e \approx N_D$.
Since $n_e n_h = n_i^2$, The number of holes
$n_h = (2.25 \times 10^{32})/(5 \times 10^{22})$

$\sim 4.5 \times 10^9$ m^{-3}

EXAMPLE 14.2

14.5 p-n Junction

A p-n junction is the basic building block of many semiconductor devices like diodes, transistor, etc. A clear understanding of the junction behaviour is important to analyse the working of other semiconductor devices. We will now try to understand how a junction is formed and how the junction behaves under the influence of external applied voltage (also called *bias*).

14.5.1 p-n junction formation

Formation and working of p-n junction diode
http://www.mtmi.vu.lt/pfk/funkc_dariniai/diod/p n_devices.htm

Consider a thin p-type silicon (p-Si) semiconductor wafer. By adding precisely a small quantity of pentavelent impurity, part of the p-Si wafer can be converted into n-Si. There are several processes by which a semiconductor can be formed. The wafer now contains p-region and n-region and a metallurgical junction between p-, and n- region.

Two important processes occur during the formation of a p-n junction: *diffusion* and *drift*. We know that in an n-type semiconductor, the concentration of electrons (number of electrons per unit volume) is more compared to the concentration of holes. Similarly, in a p-type semiconductor, the concentration of holes is more than the concentration of electrons. During the formation of p-n junction, and due to the concentration gradient across p-, and n- sides, holes diffuse from p-side to n-side ($p \rightarrow n$) and electrons diffuse from n-side to p-side ($n \rightarrow p$). This motion of charge carries gives rise to diffusion current across the junction.

When an electron diffuses from $n \rightarrow p$, it leaves behind an ionised donor on n-side. This ionised donor (positive charge) is immobile as it is bonded to the surrounding atoms. As the electrons continue to diffuse from $n \rightarrow p$, a layer of positive charge (or positive space-charge region) on n-side of the junction is developed.

Similarly, when a hole diffuses from $p \rightarrow n$ due to the concentration gradient, it leaves behind an ionised acceptor (negative charge) which is immobile. As the holes continue to diffuse, a layer of negative charge (or negative space-charge region) on the p-side of the junction is developed. This space-charge region on either side of the junction together is known as *depletion region* as the electrons and holes taking part in the initial movement across the junction *depleted* the region of its free charges (Fig. 14.10). The thickness of depletion region is of the order of one-tenth of a micrometre. Due to the positive space-charge region on n-side of the junction and negative space charge region on p-side of the junction, an electric field directed from positive charge towards negative charge develops. Due to this field, an electron on p-side of the junction moves to n-side and a hole on n-side of the junction moves to p-side. The motion of charge carriers due to the electric field is called drift. Thus a drift current, which is opposite in direction to the diffusion current (Fig. 14.10) starts.

FIGURE 14.10 p-n junction formation process.

Initially, diffusion current is large and drift current is small. As the diffusion process continues, the space-charge regions on either side of the junction extend, thus increasing the electric field strength and hence drift current. This process continues until the diffusion current equals the drift current. Thus a p-n junction is formed. In a p-n junction under equilibrium there is *no net* current.

The loss of electrons from the n-region and the gain of electron by the p-region causes a difference of potential across the junction of the two regions. The polarity of this potential is such as to oppose further flow of carriers so that a condition of equilibrium exists. Figure 14.11 shows the p-n junction at equilibrium and the potential across the junction. The n-material has lost electrons, and p material has acquired electrons. The n material is thus positive relative to the p material. Since this potential tends to prevent the movement of electron from the n region into the p region, it is often called a *barrier potential*.

FIGURE 14.11 (a) Diode under equilibrium (V = 0), (b) Barrier potential under no bias.

EXAMPLE 14.3

Example 14.3 Can we take one slab of p-type semiconductor and physically join it to another n-type semiconductor to get p-n junction?

Solution No! Any slab, howsoever flat, will have roughness much larger than the inter-atomic crystal spacing (~2 to 3 Å) and hence *continuous contact* at the atomic level will not be possible. The junction will behave as a *discontinuity* for the flowing charge carriers.

14.6 Semiconductor Diode

A semiconductor diode [Fig. 14.12(a)] is basically a p-n junction with metallic contacts provided at the ends for the application of an external voltage. It is a two terminal device. A p-n junction diode is symbolically represented as shown in Fig. 14.12(b).

The direction of arrow indicates the conventional direction of current (when the diode is under forward bias). The equilibrium barrier potential can be altered by applying an external voltage *V* across the diode. The situation of p-n junction diode under equilibrium (without bias) is shown in Fig. 14.11(a) and (b).

FIGURE 14.12 (a) Semiconductor diode, (b) Symbol for p-n junction diode.

14.6.1 p-n junction diode under forward bias

When an external voltage *V* is applied across a semiconductor diode such that p-side is connected to the positive terminal of the battery and n-side to the negative terminal [Fig. 14.13(a)], it is said to be *forward biased*.

The applied voltage mostly drops across the depletion region and the voltage drop across the p-side and n-side of the junction is negligible. (This is because the resistance of the depletion region – a region where there are no charges – is very high compared to the resistance of n-side and p-side.) The direction of the applied voltage (*V*) is opposite to the

FIGURE 14.13 (a) p-n junction diode under forward bias, (b) Barrier potential (1) without battery, (2) Low battery voltage, and (3) High voltage battery.

FIGURE 14.14 Forward bias minority carrier injection.

built-in potential V_0. As a result, the depletion layer width decreases and the barrier height is reduced [Fig. 14.13(b)]. The effective barrier height under forward bias is $(V_0 - V)$.

If the applied voltage is small, the barrier potential will be reduced only slightly below the equilibrium value, and only a small number of carriers in the material—those that happen to be in the uppermost energy levels—will possess enough energy to cross the junction. So the current will be small. If we increase the applied voltage significantly, the barrier height will be reduced and more number of carriers will have the required energy. Thus the current increases.

Due to the applied voltage, electrons from n-side cross the depletion region and reach p-side (where they are minority carries). Similarly, holes from p-side cross the junction and reach the n-side (where they are minority carries). This process under forward bias is known as minority carrier injection. At the junction boundary, on each side, the minority carrier concentration increases significantly compared to the locations far from the junction.

Due to this concentration gradient, the injected electrons on p-side diffuse from the junction edge of p-side to the other end of p-side. Likewise, the injected holes on n-side diffuse from the junction edge of n-side to the other end of n-side (Fig. 14.14). This motion of charged carriers on either side gives rise to current. The total diode forward current is sum of hole diffusion current and conventional current due to electron diffusion. The magnitude of this current is usually in mA.

14.6.2 p-n junction diode under reverse bias

When an external voltage (V) is applied across the diode such that n-side is positive and p-side is negative, it is said to be *reverse biased* [Fig.14.15(a)]. The applied voltage mostly drops across the depletion region. The direction of applied voltage is same as the direction of barrier potential. As a result, the barrier height increases and the depletion region widens due to the change in the electric field. The effective barrier height under reverse bias is $(V_0 + V)$, [Fig. 14.15(b)]. This suppresses the flow of electrons from n $\rightarrow$ p and holes from p $\rightarrow$ n. Thus, diffusion current, decreases enormously compared to the diode under forward bias.

The electric field direction of the junction is such that if electrons on p-side or holes on n-side in their random motion come close to the junction, they will be swept to its majority zone. This drift of carriers gives rise to current. The drift current is of the order of a few μA. This is quite low because it is due to the motion of carriers from their minority side to their majority side across the junction. The drift current is also there under forward bias but it is negligible (μA) when compared with current due to injected carriers which is usually in mA.

The diode reverse current is not very much dependent on the applied voltage. Even a small voltage is sufficient to sweep the minority carriers from one side of the junction to the other side of the junction. The current

is not limited by the magnitude of the applied voltage but is limited due to the concentration of the minority carrier on either side of the junction.

The current under reverse bias is essentially voltage independent upto a critical reverse bias voltage, known as breakdown voltage (V_{br}). When $V = V_{br}$, the diode reverse current increases sharply. Even a slight increase in the bias voltage causes large change in the current. If the reverse current is not limited by an external circuit below the rated value (specified by the manufacturer) the p-n junction will get destroyed. Once it exceeds the rated value, the diode gets destroyed due to overheating. This can happen even for the diode under forward bias, if the forward current exceeds the rated value.

FIGURE 14.15 (a) Diode under reverse bias, (b) Barrier potential under reverse bias.

The circuit arrangement for studying the *V-I* characteristics of a diode, (i.e., the variation of current as a function of applied voltage) are shown in Fig. 14.16(a) and (b). The battery is connected to the diode through a potentiometer (or reheostat) so that the applied voltage to the diode can be changed. For different values of voltages, the value of the current is noted. A graph between *V* and *I* is obtained as in Fig. 14.16(c). Note that in forward bias measurement, we use a milliammeter since the expected current is large (as explained in the earlier section) while a micrometer is used in reverse bias to measure the current. You can see in Fig. 14.16(c) that in forward

FIGURE 14.16 Experimental circuit arrangement for studying *V-I* characteristics of a p-n junction diode (a) in forward bias , (b) in reverse bias. (c) Typical *V-I* characteristics of a silicon diode.

bias, the current first increases very slowly, almost negligibly, till the voltage across the diode crosses a certain value. After the characteristic voltage, the diode current increases significantly (exponentially), even for a very small increase in the diode bias voltage. This voltage is called the *threshold voltage* or cut-in voltage (~0.2V for germanium diode and ~0.7 V for silicon diode).

For the diode in reverse bias, the current is very small (~μA) and almost remains constant with change in bias. It is called *reverse saturation current*. However, for special cases, at very high reverse bias (break down voltage), the current suddenly increases. This special action of the diode is discussed later in Section 14.8. The general purpose diode are not used beyond the reverse saturation current region.

The above discussion shows that the p-n junction diode primerly allows the flow of current only in one direction (forward bias). The forward bias resistance is low as compared to the reverse bias resistance. This property is used for rectification of ac voltages as discussed in the next section. For diodes, we define a quantity called *dynamic resistance* as the ratio of small change in voltage ΔV to a small change in current ΔI:

$$r_d = \frac{\Delta V}{\Delta I} \tag{14.6}$$

Example 14.4 The *V-I* characteristic of a silicon diode is shown in the Fig. 14.17. Calculate the resistance of the diode at (a) I_D = 15 mA and (b) V_D = –10 V.

FIGURE 14.17

Solution Considering the diode characteristics as a straight line between I = 10 mA to I = 20 mA passing through the origin, we can calculate the resistance using Ohm's law.

(a) From the curve, at I = 20 mA, V = 0.8 V, I = 10 mA, V = 0.7 V

$r_{fb} = \Delta V/\Delta I$ = 0.1V/10 mA = 10 Ω

(b) From the curve at V = –10 V, I = –1 μA,

Therefore,

r_{rb} = 10 V/1μA= 1.0×10^7 Ω

EXAMPLE 14.4

14.7 Application of Junction Diode as a Rectifier

From the *V-I* characteristic of a junction diode we see that it allows current to pass only when it is forward biased. So if an alternating voltage is applied across a diode the current flows only in that part of the cycle when the diode is forward biased. This property is used *to rectify* alternating voltages and the circuit used for this purpose is called a *rectifier*.

If an alternating voltage is applied across a diode in series with a load, a pulsating voltage will appear across the load only during the half cycles of the ac input during which the diode is forward biased. Such rectifier circuit, as shown in Fig. 14.18, is called a *half-wave rectifier*. The secondary of a transformer supplies the desired ac voltage across terminals A and B. When the voltage at A is positive, the diode is forward biased and it conducts. When A is negative, the diode is reverse-biased and it does not conduct. The reverse saturation current of a diode is negligible and can be considered equal to zero for practical purposes. (The reverse breakdown voltage of the diode must be sufficiently higher than the peak ac voltage at the secondary of the transformer to protect the diode from reverse breakdown.)

FIGURE 14.18 (a) Half-wave rectifier circuit, (b) Input ac voltage and output voltage waveforms from the rectifier circuit.

Therefore, in the positive *half-cycle* of ac there is a current through the load resistor R_L and we get an output voltage, as shown in Fig. 14.18(b), whereas there is no current in the negative half-cycle. In the next positive half-cycle, again we get the output voltage. Thus, the output voltage, though still varying, is restricted to *only one direction* and is said to be *rectified*. Since the rectified output of this circuit is only for half of the input ac wave it is called as *half-wave rectifier*.

The circuit using two diodes, shown in Fig. 14.19(a), gives output rectified voltage corresponding to both the positive as well as negative half of the ac cycle. Hence, it is known as *full-wave rectifier*. Here the p-side of the two diodes are connected to the ends of the secondary of the transformer. The n-side of the diodes are connected together and the output is taken between this common point of diodes and the midpoint of the secondary of the transformer. So for a full-wave rectifier the secondary of the transformer is provided with a centre tapping and so it is called *centre-tap transformer*. As can be seen from Fig.14.19(c) the voltage rectified by each diode is only half the total secondary voltage. Each diode rectifies only for half the cycle, but the two do so for alternate cycles. Thus, the output between their common terminals and the centre-tap of the transformer becomes a full-wave rectifier output. (Note that there is another circuit of full wave rectifier which does not need a centre-tap transformer but needs four diodes.) Suppose the input voltage to A

FIGURE 14.19 (a) A Full-wave rectifier circuit; (b) Input wave forms given to the diode D_1 at A and to the diode D_2 at B; (c) Output waveform across the load R_L connected in the full-wave rectifier circuit.

with respect to the centre tap at any instant is positive. It is clear that, at that instant, voltage at B being out of phase will be negative as shown in Fig.14.19(b). So, diode D_1 gets forward biased and conducts (while D_2 being reverse biased is not conducting). Hence, during this positive half cycle we get an output current (and a output voltage across the load resistor R_L) as shown in Fig.14.19(c). In the course of the ac cycle when the voltage at A becomes negative with respect to centre tap, the voltage at B would be positive. In this part of the cycle diode D_1 would not conduct but diode D_2 would, giving an output current and output voltage (across R_L) during the negative half cycle of the input ac. Thus, we get output voltage during both the positive as well as the negative half of the cycle. Obviously, this is a more efficient circuit for getting rectified voltage or current than the half-wave rectifier

The rectified voltage is in the form of pulses of the shape of half sinusoids. Though it is unidirectional it does not have a steady value. To get steady dc output from the pulsating voltage normally a capacitor is connected across the output terminals (parallel to the load R_L). One can also use an inductor in series with R_L for the same purpose. Since these additional circuits appear to *filter* out the *ac ripple* and give a *pure dc* voltage, so they are called filters.

Now we shall discuss the role of capacitor in filtering. When the voltage across the capacitor is rising, it gets charged. If there is no external load, it remains charged to the peak voltage of the rectified output. When there is a load, it gets discharged through the load and the voltage across it begins to fall. In the next half-cycle of rectified output it again gets charged to the peak value (Fig. 14.20). The rate of fall of the voltage across the capacitor depends upon the inverse product of capacitor C and the effective resistance R_L used in the circuit and is called the *time constant*. To make the time constant large value of C should be large. So capacitor input filters use large capacitors. The *output voltage* obtained by using capacitor input filter is nearer to the *peak voltage* of the rectified voltage. This type of filter is most widely used in power supplies.

FIGURE 14.20 (a) A full-wave rectifier with capacitor filter, (b) Input and output voltage of rectifier in (a).

14.8 Special Purpose p-n Junction Diodes

In the section, we shall discuss some devices which are basically junction diodes but are developed for different applications.

14.8.1 Zener diode

It is a special purpose semiconductor diode, named after its inventor C. Zener. It is designed to operate under reverse bias in the breakdown region and used as a voltage regulator. The symbol for Zener diode is shown in Fig. 14.21(a).

Zener diode is fabricated by heavily doping both p-, and n- sides of the junction. Due to this, depletion region formed is very thin ($<10^{-6}$ m) and the electric field of the junction is extremely high ($\sim 5\times10^{6}$ V/m) even for a small reverse bias voltage of about 5V. The I-V characteristics of a Zener diode is shown in Fig. 14.21(b). It is seen that when the applied reverse bias voltage(V) reaches the breakdown voltage (V_z) of the Zener diode, there is a large change in the current. Note that after the breakdown voltage V_z, a large change in the current can be produced by almost insignificant change in the reverse bias voltage. In other words, Zener voltage remains constant, even though current through the Zener diode varies over a wide range. This property of the Zener diode is used for regulating supply voltages so that they are constant.

Let us understand how reverse current suddenly increases at the breakdown voltage. We know that reverse current is due to the flow of electrons (minority carriers) from p $\rightarrow$ n and holes from n $\rightarrow$ p. As the reverse bias voltage is increased, the electric field at the junction becomes significant. When the reverse bias voltage V = V_z, then the electric field strength is high enough to pull valence electrons from the host atoms on the p-side which are accelerated to n-side. These electrons account for high current observed at the breakdown. The emission of electrons from the host atoms due to the high electric field is known as internal field emission or field ionisation. The electric field required for field ionisation is of the order of 10^6 V/m.

FIGURE 14.21 Zener diode, (a) symbol, (b) *I-V* characteristics.

Zener diode as a voltage regulator

We know that when the ac input voltage of a rectifier fluctuates, its rectified output also fluctuates. To get a constant dc voltage from the dc unregulated output of a rectifier, we use a Zener diode. The circuit diagram of a voltage regulator using a Zener diode is shown in Fig. 14.22.

FIGURE 14.22 Zener diode as DC voltage regulator (to be corrected).

The unregulated dc voltage (filtered output of a rectifier) is connected to the Zener diode through a series resistance R_s such that the Zener diode is reverse biased. If the input voltage increases, the current through R_s and Zener diode also increases. This increases the voltage drop across R_s without any change in the voltage across the Zener diode. This is because in the breakdown region, Zener voltage remains constant even though the current through the Zener diode changes. Similarly, if the input voltage decreases, the current through R_s and Zener diode also decreases. The voltage drop across R_s decreases without any change in the voltage across the Zener diode. Thus any increase/decrease in the input voltage results in, increase/decrease of the voltage drop across R_s without any change in voltage across the Zener diode. Thus the Zener diode acts as a voltage regulator. We have to select the Zener diode according to the required output voltage and accordingly the series resistance R_s.

EXAMPLE 14.5

Example 14.5 In a Zener regulated power supply a Zener diode with $V_Z = 6.0$ V is used for regulation. The load current is to be 4.0 mA and the unregulated input is 10.0 V. What should be the value of series resistor R_S?

Solution

The value of R_S should be such that the current through the Zener diode is much larger than the load current. This is to have good load regulation. Choose Zener current as five times the load current, i.e., I_Z = 20 mA. The total current through R_S is, therefore, 24 mA. The voltage drop across R_S is 10.0 – 6.0 = 4.0 V. This gives $R_S = 4.0\text{V}/(24 \times 10^{-3})\ \text{A} = 167\ \Omega$. The nearest value of carbon resistor is 150 Ω. So, a series resistor of 150 Ω is appropriate. Note that slight variation in the value of the resistor does not matter, what is important is that the current I_Z should be sufficiently larger than I_L.

14.8.2 Optoelectronic junction devices

We have seen so far, how a semiconductor diode behaves under applied electrical inputs. In this section, we learn about semiconductor diodes in which carriers are generated by photons (photo-excitation). All these devices are called *optoelectronic devices*. We shall study the functioning of the following optoelectronic devices:

(i) *Photodiodes* used for detecting optical signal (photodetectors).
(ii) *Light emitting diodes* (LED) which convert electrical energy into light.
(iii) *Photovoltaic devices* which convert optical radiation into electricity (*solar cells*).

(i) *Photodiode*

A Photodiode is again a special purpose p-n junction diode fabricated with a transparent window to allow light to fall on the diode. It is operated under reverse bias. When the photodiode is illuminated with light (photons) with energy ($h\nu$) greater than the energy gap (E_g) of the semiconductor, then electron-hole pairs are generated due to the absorption of photons. The diode is fabricated such that the generation of *e-h* pairs takes place in or near the depletion region of the diode. Due to electric field of the junction, electrons and holes are separated before they recombine. The direction of the electric field is such that electrons reach n-side and holes reach p-side. Electrons are collected on n-side and holes are collected on p-side giving rise to an emf. When an external load is connected, current flows. The magnitude of the photocurrent depends on the intensity of incident light (photocurrent is proportional to incident light intensity).

It is easier to observe the change in the current with change in the light intensity, if a reverse bias is applied. Thus photodiode can be used as a photodetector to detect optical signals. The circuit diagram used for the measurement of *I-V* characteristics of a photodiode is shown in Fig. 14.23(a) and a typical *I-V* characteristics in Fig. 14.23(b).

FIGURE 14.23 (a) An illuminated photodiode under reverse bias , (b) *I-V* characteristics of a photodiode for different illumination intensity $I_4 > I_3 > I_2 > I_1$.

Example 14.6 The current in the forward bias is known to be more (~mA) than the current in the reverse bias (~μA). What is the reason then to operate the photodiodes in reverse bias?

Solution Consider the case of an n-type semiconductor. Obviously, the majority carrier density (n) is considerably larger than the minority hole density p (i.e., $n >> p$). On illumination, let the excess electrons and holes generated be Δn and Δp, respectively:

$$n' = n + \Delta n$$

$$p' = p + \Delta p$$

Here n' and p' are the electron and hole concentrations* at any particular illumination and n and p are carriers concentration when there is no illumination. Remember $\Delta n = \Delta p$ and $n >> p$. Hence, the

EXAMPLE 14.6

* Note that, to create an e-h pair, we spend some energy (photoexcitation, thermal excitation, etc.). Therefore when an electron and hole recombine the energy is released in the form of light (radiative recombination) or heat (non-radiative recombination). It depends on semiconductor and the method of fabrication of the p-n junction. For the fabrication of LEDs, semiconductors like GaAs, GaAs-GaP are used in which radiative recombination dominates.

EXAMPLE 14.6

fractional change in the majority carriers (i.e., $\Delta n/n$) would be much less than that in the minority carriers (i.e., $\Delta p/p$). In general, we can state that the fractional change due to the photo-effects on the *minority carrier dominated reverse bias current* is more easily measurable than the fractional change in the forward bias current. Hence, photodiodes are preferably used in the reverse bias condition for measuring light intensity.

(ii) *Light emitting diode*

It is a heavily doped p-n junction which under forward bias emits spontaneous radiation. The diode is encapsulated with a transparent cover so that emitted light can come out.

When the diode is forward biased, electrons are sent from n $\rightarrow$ p (where they are minority carriers) and holes are sent from p $\rightarrow$ n (where they are minority carriers). At the junction boundary the concentration of minority carriers increases compared to the equilibrium concentration (i.e., when there is no bias). Thus at the junction boundary on either side of the junction, excess minority carriers are there which recombine with majority carriers near the junction. On recombination, the energy is released in the form of photons. Photons with energy equal to or slightly less than the band gap are emitted. When the forward current of the diode is small, the intensity of light emitted is small. As the forward current increases, intensity of light increases and reaches a maximum. Further increase in the forward current results in decrease of light intensity. LEDs are biased such that the light emitting efficiency is maximum.

The *V-I* characteristics of a LED is similar to that of a Si junction diode. But the threshold voltages are much higher and slightly different for each colour. The reverse breakdown voltages of LEDs are very low, typically around 5V. So care should be taken that high reverse voltages do not appear across them.

LEDs that can emit red, yellow, orange, green and blue light are commercially available. The semiconductor used for fabrication of visible LEDs must at least have a band gap of 1.8 eV (spectral range of visible light is from about 0.4 μm to 0.7 μm, i.e., from about 3 eV to 1.8 eV). The compound semiconductor Gallium Arsenide – Phosphide ($GaAs_{1-x}P_x$) is used for making LEDs of different colours. $GaAs_{0.6}\,P_{0.4}$ (E_g ~ 1.9 eV) is used for red LED. GaAs (E_g ~ 1.4 eV) is used for making infrared LED. These LEDs find extensive use in remote controls, burglar alarm systems, optical communication, etc. Extensive research is being done for developing white LEDs which can replace incandescent lamps.

LEDs have the following advantages over conventional incandescent low power lamps:

(i) Low operational voltage and less power.
(ii) Fast action and no warm-up time required.
(iii) The bandwidth of emitted light is 100 Å to 500 Å or in other words it is nearly (but not exactly) monochromatic.
(iv) Long life and ruggedness.
(v) Fast on-off switching capability.

(iii) *Solar cell*

A solar cell is basically a p-n junction which generates emf when solar radiation falls on the p-n junction. It works on the same principle (photovoltaic effect) as the photodiode, except that no external bias is applied and the junction area is kept much larger for solar radiation to be incident because we are interested in more power.

A simple p-n junction solar cell is shown in Fig. 14.24.

A p-Si wafer of about 300 μm is taken over which a thin layer (~0.3 μm) of n-Si is grown on one-side by diffusion process. The other side of p-Si is coated with a metal (back contact). On the top of n-Si layer, metal finger electrode (or metallic grid) is deposited. This acts as a front contact. The metallic grid occupies only a very small fraction of the cell area (<15%) so that light can be incident on the cell from the top.

FIGURE 14.24 (a) Typical p-n junction solar cell; (b) Cross-sectional view.

The generation of emf by a solar cell, when light falls on, it is due to the following three basic processes: generation, separation and collection—(i) generation of e-h pairs due to light (with $h\nu > E_g$) close to the junction; (ii) separation of electrons and holes due to electric field of the depletion region. Electrons are swept to n-side and holes to p-side; (iii) the electrons reaching the n-side are collected by the front contact and holes reaching p-side are collected by the back contact. Thus p-side becomes positive and n-side becomes negative giving rise to *photovoltage*.

When an external load is connected as shown in the Fig. 14.25(a) a photocurrent I_L flows through the load. A typical *I-V* characteristics of a solar cell is shown in the Fig. 14.25(b).

Note that the $I - V$ characteristics of solar cell is drawn in the fourth quadrant of the coordinate axes. This is because a solar cell does not draw current but supplies the same to the load.

FIGURE 14.25 (a) A typical illuminated p-n junction solar cell; (b) *I-V* characteristics of a solar cell.

Semiconductors with band gap close to 1.5 eV are ideal materials for solar cell fabrication. Solar cells are made with semiconductors like Si (E_g = 1.1 eV), GaAs (E_g = 1.43 eV), CdTe (E_g = 1.45 eV), $CuInSe_2$ (E_g = 1.04 eV), etc. The important criteria for the selection of a material for solar cell fabrication are (i) band gap (~1.0 to 1.8 eV), (ii) high optical absorption (~10^4 cm^{-1}), (iii) electrical conductivity, (iv) availability of the raw material, and (v) cost. Note that sunlight is not always required for a solar cell. *Any* light with photon energies greater than the bandgap will do. Solar cells are used to power electronic devices in satellites and space vehicles and also as power supply to some calculators. Production of low-cost photovoltaic cells for large-scale solar energy is a topic for research.

Example 14.7 Why are Si and GaAs are preferred materials for solar cells?

Solution The solar radiation spectrum received by us is shown in Fig. 14.26.

FIGURE 14.26

The maxima is near 1.5 eV. For photo-excitation, $h\nu > E_g$. Hence, semiconductor with band gap ~1.5 eV or lower is likely to give better solar conversion efficiency. Silicon has E_g ~ 1.1 eV while for GaAs it is ~1.53 eV. In fact, GaAs is better (in spite of its higher band gap) than Si because of its relatively higher absorption coefficient. If we choose materials like CdS or CdSe (E_g ~ 2.4 eV), we can use only the high energy component of the solar energy for photo-conversion and a significant part of energy will be of no use.

The question arises: why we do not use material like PbS (E_g ~ 0.4 eV) which satisfy the condition $h\nu > E_g$ for ν maxima corresponding to the solar radiation spectra? If we do so, most of the solar radiation will be absorbed on the *top-layer* of solar cell and will not reach in or near the depletion region. For effective electron-hole separation, due to the junction field, we want the photo-generation to occur in the junction region only.

EXAMPLE 14.7

14.9 Junction Transistor

The credit of inventing the transistor in the year 1947 goes to J. Bardeen and W.H. Brattain of Bell Telephone Laboratories, U.S.A. That transistor was a point-contact transistor. The first junction transistor consisting of two back-to-back p-n junctions was invented by William Schockley in 1951.

As long as only the junction transistor was known, it was known simply as transistor. But over the years new types of transistors were invented and to differentiate it from the new ones it is now called the Bipolar Junction Transistor (BJT). Even now, often the word transistor

is used to mean BJT when there is no confusion. Since our study is limited to only BJT, we shall use the word transistor for BJT without any ambiguity.

14.9.1 Transistor: structure and action

A transistor has three doped regions forming two p-n junctions between them. Obviously, there are two types of transistors, as shown in Fig. 14.27.

(i) n-p-n transistor: Here two segments of n-type semiconductor (emitter and collector) are separated by a segment of p-type semiconductor (base).

(ii) p-n-p transistor: Here two segments of p-type semiconductor (termed as emitter and collector) are separated by a segment of n-type semiconductor (termed as base).

The schematic representations of an n-p-n and a p-n-p configuration are shown in Fig. 14.27(a). All the three segments of a transistor have different thickness and their doping levels are also different. In the schematic symbols used for representing p-n-p and n-p-n transistors [Fig. 14.27(b)] the arrowhead shows the direction of conventional current in the transistor. A brief description of the three segments of a transistor is given below:

- *Emitter*: This is the segment on one side of the transistor shown in Fig. 14.27(a). It is of *moderate size* and *heavily doped*. It supplies a large number of majority carriers for the current flow through the transistor.
- *Base*: This is the central segment. *It is very thin* and *lightly doped*.
- *Collector*: This segment collects a *major* portion of the majority carriers supplied by the emitter. The collector side is *moderately doped* and *larger* in size as compared to the *emitter*.

We have seen earlier in the case of a p-n junction, that there is a formation of depletion region acorss the junction. In case of a transistor depletion regions are formed at the emitter base-junction and the base-collector junction. For understanding the action of a transistor, we have to consider the nature of depletion regions formed at these junctions. The charge carriers move across different regions of the transistor when proper voltages are applied across its terminals.

The biasing of the transistor is done differently for different uses. The transistor can be used in two distinct ways. Basically, it was invented to function as an amplifier, a device which produces a enlarged copy of a signal. But later its use as a switch acquired equal importance. We shall study both these functions and the ways the transistor is biased to achieve these mutually exclusive functions.

First we shall see what gives the transistor its amplifying capabilities. The transistor works as an amplifier, with its emitter-base junction forward biased and the base-collector junction reverse biased. This situation is shown in Fig. 14.28, where V_{CC} and V_{EE} are used for creating the respective biasing. When the transistor is biased in this way it is said to be in *active* state.We represent the voltage between emitter and base as V_{EB} and that between the collector and the base as V_{CB}. In

FIGURE 14.27
(a) Schematic representations of a n-p-n transistor and p-n-p transistor, and (b) Symbols for n-p-n and p-n-p transistors.

FIGURE 14.28 Bias Voltage applied on: (a) p-n-p transistor and (b) n-p-n transistor.

Fig. 14.28, base is a common terminal for the two power supplies whose other terminals are connected to emitter and collector, respectively. So the two power supplies are represented as V_{EE}, and V_{CC}, respectively. In circuits, where emitter is the common terminal, the power supply between the base and the emitter is represented as V_{BB} and that between collector and emitter as V_{CC}.

Let us see now the paths of current carriers in the transistor with emitter-base junction forward biased and base-collector junction reverse biased. The heavily doped emitter has a high concentration of majority carriers, which will be holes in a p-n-p transistor and electrons in an n-p-n transistor. These majority carriers enter the base region in large numbers. The base is thin and lightly doped. So the majority carriers there would be few. In a p-n-p transistor the majority carriers in the base are electrons since base is of n-type semiconductor. The large number of holes entering the base from the emitter swamps the small number of electrons there. As the base collector-junction is reverse-biased, these holes, which appear as minority carriers at the junction, can easily cross the junction and enter the collector. The holes in the base could move either towards the base terminal to combine with the electrons entering from outside or cross the junction to enter into the collector and reach the collector terminal. The base is made thin so that most of the holes find themselves near the reverse-biased base-collector junction and so cross the junction instead of moving to the base terminal.

It is interesting to note that due to forward bias a large current enters the emitter-base junction, but most of it is diverted to adjacent reverse-biased base-collector junction and the current coming out of the base becomes a very small fraction of the current that entered the junction. If we represent the hole current and the electron current crossing the forward biased junction by I_h and I_e respectively then the total current in a forward biased diode is the sum $I_h + I_e$. We see that the emitter current $I_E = I_h + I_e$ but the base current $I_B << I_h + I_e$, because a major part of I_E goes to collector instead of coming out of the base terminal. The base current is thus a small fraction of the emitter current.

The current entering into the emitter from outside is equal to the emitter current I_E. Similarly the current emerging from the base terminal is I_B and that from collector terminal is I_C. It is obvious from the above description and also from a straight forward application of Kirchhoff's law to Fig. 14.28(a) that the emitter current is the sum of collector current and base current:

$$I_E = I_C + I_B \quad (14.7)$$

We also see that $I_C \approx I_E$.

Our description of the direction of motion of the holes is identical with the direction of the conventional current. But the direction of motion of electrons is just opposite to that of the current. Thus in a p-n-p transistor the current enters from emitter into base whereas in a n-p-n transistor it enters from the base into the emitter. The arrowhead in the emitter shows the direction of the conventional current.

The description about the paths followed by the majority and minority carriers in a n-p-n is exactly the same as that for the p-n-p transistor. But the current paths are exactly opposite, as shown in Fig. 14.28. In Fig. 14.28(b) the electrons are the majority carriers supplied by the n-type emitter region. They cross the thin p-base region and are able to reach the collector to give the collector current, I_C . From the above description we can conclude that in the active state of the transistor the emitter-base junction acts as a low resistance while the base collector acts as a high resistance.

14.9.2 Basic transistor circuit configurations and transistor characteristics

In a transistor, only three terminals are available, viz., *Emitter* (E), *Base* (B) and *Collector* (C). Therefore, in a circuit the input/output connections have to be such that one of these (E, B or C) is common to both the input and the output. Accordingly, the transistor can be *connected* in either of the following three configurations:

Common Emitter (CE), Common Base (CB), Common Collector (CC)

The transistor is most widely used in the CE configuration and we shall restrict our discussion to only this configuration. Since more commonly used transistors are n-p-n Si transistors, we shall confine our discussion to such transistors only. With p-n-p transistors the polarities of the external power supplies are to be inverted.

Common emitter transistor characteristics

When a transistor is used in CE configuration, the input is between the base and the emitter and the output is between the collector and the emitter. The variation of the base current I_B with the base-emitter voltage V_{BE} is called the *input characteristic*. Similarly, the variation of the collector current I_C with the collector-emitter voltage V_{CE} is called the *output characteristic*. You will see that the output characteristics are controlled by the input characteristics. This implies that the collector current changes with the base current.

FIGURE 14.29 Circuit arrangement for studying the input and output characteristics of n-p-n transistor in CE configuration.

The input and the output characteristics of an n-p-n transistors can be studied by using the circuit shown in Fig. 14.29.

To study the input characteristics of the transistor in C_E configuration, a curve is plotted between the base current I_B against the base-emitter voltage V_{BE}. The

FIGURE 14.30 (a) Typical input characteristics, and (b) Typical output characteristics.

collector-emitter voltage V_{CE} is kept fixed while studying the dependence of I_B on V_{BE}. We are interested to obtain the input characteristic when the transistor is in active state. So the collector-emitter voltage V_{CE} is kept large enough to make the base collector junction reverse biased. Since $V_{CE} = V_{CB} + V_{BE}$ and for Si transistor V_{BE} is 0.6 to 0.7 V, V_{CE} must be sufficiently larger than 0.7 V. Since the transistor is operated as an amplifier over large range of V_{CE}, the reverse bias across the base-collector junction is high most of the time. Therefore, the input characteristics may be obtained for V_{CE} somewhere in the range of 3 V to 20 V. Since *the increase in V_{CE} appears as increase in V_{CB}, its effect on I_B is negligible.* As a consequence, input characteristics for various values of V_{CE} will give almost identical curves. Hence, *it is enough to determine only one input characteristics*. The input characteristics of a transistor is as shown in Fig. 14.30(a).

The output characteristic is obtained by observing the variation of I_C as V_{CE} is varied keeping I_B constant. It is obvious that if V_{BE} is increased by a small amount, both hole current from the emitter region and the electron current from the base region will increase. As a consequence both I_B and I_C will increase proportionately. This shows that when I_B increases I_C also increases. The plot of I_C versus V_{CE} for different fixed values of I_B gives one output characteristic. So there will be different output characteristics corresponding to different values of I_B as shown in Fig. 14.30(b).

The linear segments of both the input and output characteristics can be used to calculate some important ac parameters of transistors as shown below.

(i) ***Input resistance*** **(r_i):** This is defined as the ratio of change in base-emitter voltage (ΔV_{BE}) to the resulting change in base current (ΔI_B) at constant collector-emitter voltage (V_{CE}). This is dynamic (ac resistance) and as can be seen from the input characteristic, its value varies with the operating current in the transistor:

$$r_i = \left(\frac{\Delta V_{BE}}{\Delta I_B}\right)_{V_{CE}} \tag{14.8}$$

The value of r_i can be anything from a few hundreds to a few thousand ohms.

(ii) ***Output resistance (r_o):*** This is defined as the ratio of change in collector-emitter voltage (ΔV_{CE}) to the change in collector current (ΔI_C) at a constant base current I_B.

$$r_o = \left(\frac{\Delta V_{CE}}{\Delta I_C}\right)_{I_B} \tag{14.9}$$

The output characteristics show that initially for very small values of V_{CE}, I_C increases almost linearly. This happens because the base-collector junction is not reverse biased and the transistor is not in active state. In fact, the transistor is in the saturation state and the current is controlled by the supply voltage V_{CC} (=V_{CE}) in this part of the characteristic. When V_{CE} is more than that required to reverse bias the base-collector junction, I_C increases very little with V_{CE}. The reciprocal of the slope of the linear part of the output characteristic gives the values of *r*o. The output resistance of the transistor is mainly controlled by the bias of the base-collector junction. The high magnitude of the output resistance (of the order of 100 kΩ) is due to the reverse-biased state of this diode. This also explains why the resistance at the initial part of the characteristic, when the transistor is in saturation state, is very low.

(iii) ***Current amplification factor (β):*** This is defined as the ratio of the change in collector current to the change in base current at a constant collector-emitter voltage (V_{CE}) when the transistor is in active state.

$$\beta_{ac} = \left(\frac{\Delta I_C}{\Delta I_B}\right)_{V_{CE}} \tag{14.10}$$

This is also known as *small signal current gain* and its value is very large.

If we simply find the ratio of I_C and I_B we get what is called dc β of the transistor. Hence,

$$\beta_{dc} = \frac{I_C}{I_B} \tag{14.11}$$

Since I_C increases with I_B almost linearly and $I_C = 0$ when $I_B = 0$, the values of both β_{dc} and β_{ac} are nearly equal. So, for most calculations β_{dc} can be used. Both β_{ac} and β_{dc} vary with V_{CE} and I_B (or I_C) slightly.

Example 14.8 From the output characteristics shown in Fig. 14.30(b), calculate the values of β_{ac} and β_{dc} of the transistor when V_{CE} is 10 V and I_C = 4.0 mA.

Solution

$$\beta_{ac} = \left(\frac{\Delta I_C}{\Delta I_B}\right)_{V_{CE}}, \quad \beta_{dc} = \frac{I_C}{I_B}$$

For determining β_{ac} and β_{dc} at the stated values of V_{CE} and I_C one can proceed as follows. Consider any two characteristics for two values of I_B which lie above and below the given value of I_C. Here I_C = 4.0 mA. (Choose characteristics for I_B= 30 and 20 μA.) At V_{CE} = 10 V we read the two values of I_C from the graph. Then

EXAMPLE 14.8

EXAMPLE 14.8

ΔI_B = (30 – 20) μA = 10 μA, ΔI_C = (4.5 – 3.0) mA = 1.5 mA

Therefore, β_{ac} = 1.5 mA/ 10 μA = 150

For determining $\beta_{dc,}$ either estimate the value of I_B corresponding to I_C = 4.0 mA at V_{CE} = 10 V or calculate the two values of β_{dc} for the two characteristics chosen and find their mean.

Therefore, for I_C = 4.5 mA and I_B = 30 μA,

β_{dc} = 4.5 mA/ 30 μA = 150

and for I_C = 3.0 mA and I_B = 20 μA

β_{dc} =3.0 mA / 20 μA = 150

Hence, β_{dc} =(150 + 150) /2 = 150

14.9.3 Transistor as a device

The transistor can be used as a device application depending on the configuration used (namely CB, CC and CE), the biasing of the E-B and B-C junction and the operation region namely cutoff, active region and saturation. As mentioned earlier we have confined only to the CE configuration and will be concentrating on the biasing and the operation region to understand the working of a device.

FIGURE 14.31 (a) Base-biased transistor in CE configuration, (b) Transfer characteristic.

When the transistor is used in the cutoff or saturation state it acts as a *switch*. On the other hand for using the transistor as an *amplifier*, it has to operate in the active region.

(i) Transistor as a switch

We shall try to understand the operation of the transistor as a switch by analysing the behaviour of the base-biased transistor in CE configuration as shown in Fig. 14.31(a).

Applying Kirchhoff's voltage rule to the input and output sides of this circuit, we get

$$V_{BB} = I_B R_B + V_{BE} \quad (14.12)$$

and

$$V_{CE} = V_{CC} - I_C R_C. \quad (14.13)$$

We shall treat V_{BB} as the dc input voltage V_i and V_{CE} as the dc output voltage V_O. So, we have

$V_i = I_B R_B + V_{BE}$ and

$V_o = V_{CC} - I_C R_C$.

Let us see how V_o changes as V_i increases from zero onwards. In the case of Si transistor, as long as input V_i is less than 0.6 V, the transistor will be in cut off state and current I_C will be zero.

Hence $V_o = V_{CC}$

When V_i becomes greater than 0.6 V the transistor is in active state with some current I_C in the output path and the output V_o decrease as the

term $I_C R_C$ increases. With increase of V_i, I_C increases almost linearly and so V_o decreases linearly till its value becomes less than about 1.0 V.

Beyond this, the change becomes non linear and transistor goes into saturation state. With further increase in V_i the output voltage is found to decrease further towards zero though it may never become zero. If we plot the V_o vs V_i curve, [also called the transfer characteristics of the base-biased transistor (Fig. 14.31(b)], we see that between cut off state and active state and also between active state and saturation state there are regions of non-linearity showing that the transition from cutoff state to active state and from active state to saturation state are not sharply defined.

Let us see now how the transistor is operated as a switch. As long as V_i is *low* and unable to forward-bias the transistor, V_o is *high* (at V_{CC}). If V_i is *high* enough to drive the transistor into saturation, then V_o is *low*, very near to zero. When the transistor is not conducting it is said to be *switched off* and when it is driven into saturation it is said to be *switched on*. This shows that if we define low and high states as below and above certain voltage levels corresponding to cutoff and saturation of the transistor, then we can say that a *low* input switches the transistor off and a *high* input switches it on. Alternatively, we can say that a *low* input to the transistor gives a *high* output and a *high* input gives a *low* output. The switching circuits are designed in such a way that the transistor does not remain in active state.

(ii) Transistor as an amplifier

For using the transistor as an amplifier we will use the active region of the V_o versus V_i curve. The slope of the linear part of the curve represents the rate of change of the output with the input. It is negative because the output is $V_{CC} - I_C R_C$ and not $I_C R_C$. That is why as input voltage of the CE amplifier increases its output voltage decreases and the output is said to be out of phase with the input. If we consider ΔV_o and ΔV_i as small changes in the output and input voltages then $\Delta V_o / \Delta V_i$ is called the small signal voltage gain A_V of the amplifier.

If the V_{BB} voltage has a fixed value corresponding to the mid point of the active region, the circuit will behave as a CE amplifier with voltage gain $\Delta V_o / \Delta V_i$. We can express the voltage gain A_V in terms of the resistors in the circuit and the current gain of the transistor as follows.

We have, $V_o = V_{CC} - I_C R_C$

Therefore, $\Delta V_o = 0 - R_C \Delta I_C$

Similarly, from $V_i = I_B R_B + V_{BE}$

$\Delta V_i = R_B \Delta I_B + \Delta V_{BE}$

But ΔV_{BE} is negligibly small in comparison to $\Delta I_B R_B$ in this circuit.

So, the *voltage gain* of this *CE amplifier* (Fig. 14.32) is given by

$$A_V = -R_C \Delta I_C / R_B \Delta I_B$$

$$= -\beta_{ac}(R_C / R_B) \qquad (14.14)$$

where β_{ac} is equal to $\Delta I_C / \Delta I_B$ from Eq. (14.10). Thus the linear portion of the active region of the transistor can be exploited for the use in amplifiers. Transistor as an amplifier (CE configuration) is discussed in detail in the next section.

14.9.4 Transistor as an Amplifier (CE-Configuration)

To operate the transistor as an amplifier it is necessary to fix its operating point somewhere in the middle of its active region. If we fix the value of V_{BB} corresponding to a point in the middle of the linear part of the transfer curve then the dc base current I_B would be constant and corresponding collector current I_C will also be constant. The dc voltage $V_{CE} = V_{CC} - I_C R_C$ would also remain constant. The operating values of V_{CE} and I_B determine the operating point, of the amplifier.

If a small sinusoidal voltage with amplitude v_s is superposed on the dc base bias by connecting the source of that signal in series with the V_{BB} supply, then the base current will have sinusoidal variations superimposed on the value of I_B. As a consequence the collector current also will have sinusoidal variations superimposed on the value of I_C, producing in turn corresponding change in the value of V_O. We can measure the ac variations across the input and output terminals by blocking the dc voltages by large capacitors.

FIGURE 14.32 A simple circuit of a CE-transistor amplifier.

In the discription of the amplifier given above we have not considered any ac signal. In general, amplifiers are used to amplify alternating signals. Now let us superimpose an ac input signal v_i (to be amplified) on the bias V_{BB} (dc) as shown in Fig. 14.32. The output is taken between the collector and the ground.

The working of an amplifier can be easily understood, if we first assume that $v_i = 0$. Then applying Kirchhoff's law to the output loop, we get

$$V_{cc} = V_{CE} + I_c R_L \tag{14.15}$$

Likewise, the input loop gives

$$V_{BB} = V_{BE} + I_B R_B \tag{14.16}$$

When v_i is not zero, we get

$$V_{BE} + v_i = V_{BE} + I_B R_B + \Delta I_B (R_B + r_i)$$

The change in V_{BE} can be related to the input resistance r_i [see Eq. (14.8)] and the change in I_B. Hence

$$v_i = \Delta I_B (R_B + r_i)$$

$$= r \Delta I_B$$

The change in I_B causes a change in I_c. We define a parameter β_{ac}, which is similar to the β_{dc} defined in Eq. (14.11), as

$$\beta_{ac} = \frac{\Delta I_c}{\Delta I_B} = \frac{i_c}{i_b} \tag{14.17}$$

which is also known as the *ac current gain* A_i. Usually β_{ac} is close to β_{dc} in the linear region of the output characteristics.

The change in I_c due to a change in I_B causes a change in V_{CE} and the voltage drop across the resistor R_L because V_{CC} is fixed.

These changes can be given by Eq. (14.15) as

$\Delta V_{CC} = \Delta V_{CE} + R_L \Delta I_C = 0$

or $\Delta V_{CE} = -R_L \Delta I_C$

The change in V_{CE} is the output voltage v_0. From Eq. (14.10), we get

$v_0 = \Delta V_{CE} = -\beta_{ac} R_L \Delta I_B$

The voltage gain of the amplifier is

$$A_v = \frac{v_0}{v_i} = \frac{\Delta V_{CE}}{r \Delta I_B}$$

$$= -\frac{\beta_{ac} R_L}{r} \qquad (14.18)$$

The negative sign represents that output voltage is opposite with phase with the input voltage.

From the discussion of the transistor characteristics you have seen that there is a current gain β_{ac} in the CE configuration. Here we have also seen the voltage gain A_v. Therefore the power gain A_p can be expressed as the product of the current gain and voltage gain. Mathematically

$$A_p = \beta_{ac} \times A_v \qquad (14.19)$$

Since β_{ac} and A_v are greater than 1, we get ac power gain. However it should be realised that transistor is not a power generating device. The energy for the higher ac power at the output is supplied by the battery.

Example 14.9 In Fig. 14.31(a), the V_{BB} supply can be varied from 0V to 5.0 V. The Si transistor has β_{dc} = 250 and R_B = 100 kΩ, R_C = 1 KΩ, V_{CC} = 5.0V. Assume that when the transistor is saturated, V_{CE} = 0V and V_{BE} = 0.8V. Calculate (a) the minimum base current, for which the transistor will reach saturation. Hence, (b) determine V_1 when the transistor is 'switched on'. (c) find the ranges of V_1 for which the transistor is 'switched off' and 'switched on'.

Solution

Given at saturation V_{CE} = 0V, V_{BE} = 0.8V

$V_{CE} = V_{CC} - I_C R_C$

$I_C = V_{CC}/R_C$ = 5.0V/1.0kΩ = 5.0 mA

Therefore $I_B = I_C/\beta$ = 5.0 mA/250 = 20μA

The input voltage at which the transistor will go into saturation is given by

$V_{IH} = V_{BB} = I_B R_B + V_{BE}$

= 20μA × 100 kΩ + 0.8V = 2.8V

The value of input voltage below which the transistor remains cutoff is given by

V_{IL} = 0.6V, V_{IH} = 2.8V

Between 0.0V and 0.6V, the transistor will be in the 'switched off' state. Between 2.8V and 5.0V, it will be in 'switched on' state.

Note that the transistor is in active state when I_B varies from 0.0mA to 20mA. In this range, $I_C = \beta I_B$ is valid. In the saturation range, $I_C \leq \beta I_B$.

EXAMPLE 14.9

EXAMPLE 14.10

Example 14.10 For a CE transistor amplifier, the audio signal voltage across the collector resistance of 2.0 kΩ is 2.0 V. Suppose the current amplification factor of the transistor is 100, What should be the value of R_B in series with V_{BB} supply of 2.0 V if the dc base current has to be 10 times the signal current. Also calculate the dc drop across the collector resistance. (Refer to Fig. 14.33).

Solution The output ac voltage is 2.0 V. So, the ac collector current i_C = 2.0/2000 = 1.0 mA. The signal current through the base is, therefore given by $i_B = i_C / \beta$ = 1.0 mA/100 = 0.010 mA. The dc base current has to be 10× 0.010 = 0.10 mA.
From Eq.14.16, $R_B = (V_{BB} - V_{BE}) / I_B$. Assuming V_{BE} = 0.6 V,
R_B = (2.0 – 0.6)/0.10 = 14 kΩ.
The dc collector current I_C = 100×0.10 = 10 mA.

14.9.5 Feedback amplifier and transistor oscillator

In an amplifier, we have seen that a sinusoidal input is given which appears as an amplified signal in the output. This means that an *external input is necessary to sustain ac signal in the output for an amplifier.* In an oscillator, we get ac output without any external input signal. In other words, the output in an oscillator is *self-sustained.* To attain this, an amplifier is taken. A portion of the output power is returned back (feedback) to the input *in phase* with the starting power (this process is termed *positive feedback)* as shown in Fig. 14.33(a). The feedback can be achieved by inductive coupling (through mutual inductance) or *LC* or *RC* networks. Different types of oscillators essentially use different methods of coupling the output to the input (feedback network), apart from the resonant circuit for obtaining oscillation at a particular frequency. For understanding the oscillator action, we consider the circuit shown in Fig. 14.33(b) in which the feedback is accomplished by *inductive coupling* from one coil winding (T_1) to another coil winding (T_2). Note that the coils T_2 and T_1 are wound on the same core and hence are inductively coupled through their mutual inductance. As in an amplifier, the base-emitter junction is forward biased while the base-collector junction is reverse biased. Detailed biasing circuits actually used have been omitted for simplicity.

FIGURE 14.33 (a) Principle of a transistor amplifier with positive feedback working as an oscillator and (b) Tuned collector oscillator, (c) Rise and fall (or built up) of current I_c and I_e due to the inductive coupling.

Let us try to understand how oscillations are built. Suppose switch S_1 is put *on* to

apply proper bias for the first time. Obviously, a *surge* of collector current flows in the transistor. This current flows through the coil T_2 where terminals are numbered 3 and 4 [Fig. 14.33(b)]. This current does not reach full amplitude instantaneously but increases from X to Y, as shown in Fig. [14.33(c)(i)]. The inductive coupling between coil T_2 and coil T_1 now causes a current to flow in the emitter circuit (note that this actually is the 'feedback' from input to output). As a result of this positive feedback, this current (in T_1; emitter current) also increases from X′ to Y′ [Fig. 14.33(c)(ii)]. The current in T_2 (collector current) connected in the collector circuit acquires the value Y when the transistor becomes *saturated*. This means that maximum collector current is flowing and can increase no further. Since there is no further change in collector current, the magnetic field around T_2 ceases to grow. As soon as the field becomes static, there will be no further feedback from T_2 to T_1. Without continued feedback, the emitter current begins to fall. Consequently, collector current decreases from Y towards Z [Fig. 14.33(c)(i)]. However, a decrease of collector current causes the magnetic field to decay around the coil T_2. Thus, T_1 is now seeing a decaying field in T_2 (opposite from what it saw when the field was growing at the initial *start* operation). This causes a further decrease in the emitter current till it reaches Z′when the transistor is *cut-off*. This means that both I_E and I_C cease to flow. Therefore, the transistor has reverted back to its original state (when the power was first switched on). The whole process now repeats itself. That is, the transistor is driven to saturation, then to cut-off, and then back to saturation. The time for change from saturation to cut-off and back is determined by the constants of the tank circuit or tuned circuit (inductance L of coil T_2 and C connected in parallel to it). The resonance frequency (ν) of this tuned circuit determines the frequency at which the oscillator will oscillate.

$$\nu = \left(\frac{1}{2\pi\sqrt{LC}}\right) \tag{14.20}$$

In the circuit of Fig. 14.33(b), the tank or tuned circuit is connected in the collector side. Hence, it is known as *tuned collector oscillator*. If the tuned circuit is on the base side, it will be known as *tuned base oscillator*. There are many other types of tank circuits (say *RC*) or feedback circuits giving different types of oscillators like Colpitt's oscillator, Hartley oscillator, *RC*-oscillator.

14.10 Digital Electronics and Logic Gates

In electronics circuits like amplifiers, oscillators, introduced to you in earlier sections, the signal (current or voltage) has been in the form of continuous, time-varying voltage or current. Such signals are called continuous or *analogue signals*. A typical analogue signal is shown in Figure. 14.34(a). Fig. 14.34(b) shows a *pulse waveform* in which only discrete values of voltages are possible. It is convenient to use binary numbers to represent such signals. A binary number has only two digits '0' (say, 0V) and '1' (say, 5V). In digital electronics we use only these two levels of voltage as shown in Fig. 14.34(b). Such signals are called *Digital Signals*. In digital circuits only two values (represented by 0 or 1) of the input and output voltage are permissible.

This section is intended to provide the first step in our understanding of digital electronics. We shall restrict our study to some basic building blocks of digital electronics (called *Logic Gates*) which process the digital signals in a specific manner. Logic gates are used in calculators, digital watches, computers, robots, industrial control systems, and in telecommunications.

A light switch in your house can be used as an example of a digital circuit. The light is either ON or OFF depending on the switch position. When the light is ON, the output value is '1'. When the light is OFF the output value is '0'. The inputs are the position of the light switch. The switch is placed either in the ON or OFF position to activate the light.

FIGURE 14.34 (a) Analogue signal, (b) Digital signal.

Input	Output
A	Y
0	1
1	0

(b)

FIGURE 14.35 (a) Logic symbol, (b) Truth table of NOT gate.

14.10.1 Logic gates

A gate is a digital circuit that follows curtain *logical* relationship between the input and output voltages. Therefore, they are generally known as *logic gates* — gates because they control the flow of information. The five common logic gates used are NOT, AND, OR, NAND, NOR. Each logic gate is indicated by a symbol and its function is defined by a *truth table* that shows all the possible input logic level combinations with their respective output logic levels. Truth tables help understand the behaviour of logic gates. These logic gates can be realised using semiconductor devices.

(i) NOT gate

This is the most basic gate, with one input and one output. It produces a '1' output if the input is '0' and vice-versa. That is, it produces an inverted version of the input at its output. This is why it is also known as an *inverter*. The commonly used symbol together with the truth table for this gate is given in Fig. 14.35.

(ii) OR Gate

An *OR* gate has two or more inputs with one output. The logic symbol and truth table are shown in Fig. 14.36. The output Y is 1 when either input A *or* input B *or* both are 1s, that is, if any of the input is high, the output is high.

(a)

Input		Output
A	B	Y
0	0	0
0	1	1
1	0	1
1	1	1

(b)

FIGURE 14.36 (a) Logic symbol (b) Truth table of OR gate.

Apart from carrying out the above mathematical logic operation, this *gate* can be used for modifying the pulse waveform as explained in the following example.

Example 14.11 Justify the output waveform (Y) of the OR gate for the following inputs A and B given in Fig. 14.37.

Solution Note the following:

- At $t < t_1$; A = 0, B = 0; Hence Y = 0
- For t_1 to t_2; A = 1, B = 0; Hence Y = 1
- For t_2 to t_3; A = 1, B = 1; Hence Y = 1
- For t_3 to t_4; A = 0, B = 1; Hence Y = 1
- For t_4 to t_5; A = 0, B = 0; Hence Y = 0
- For t_5 to t_6; A = 1, B = 0; Hence Y = 1
- For $t > t_6$; A = 0, B = 1; Hence Y = 1

Therefore the waveform Y will be as shown in the Fig. 14.37.

FIGURE 14.37

EXAMPLE 14.11

(iii) AND Gate

An *AND* gate has two or more inputs and one output. The output Y of AND gate is 1 only when input A *and* input B are both 1. The logic symbol and truth table for this gate are given in Fig. 14.38

(a)

Input		Output
A	B	Y
0	0	0
0	1	0
1	0	0
1	1	1

(b)

FIGURE 14.38 (a) Logic symbol, (b) Truth table of AND gate.

EXAMPLE 14.12

Example 14.12 Take A and B input waveforms similar to that in Example 14.11. Sketch the output waveform obtained from AND gate.

Solution

- For $t \le t_1$; A = 0, B = 0; Hence Y = 0
- For t_1 to t_2; A = 1, B = 0; Hence Y = 0
- For t_2 to t_3; A = 1, B = 1; Hence Y = 1
- For t_3 to t_4; A = 0, B = 1; Hence Y = 0
- For t_4 to t_5; A = 0, B = 0; Hence Y = 0
- For t_5 to t_6; A = 1, B = 0; Hence Y = 0
- For $t > t_6$; A = 0, B = 1; Hence Y = 0

Based on the above, the output waveform for AND gate can be drawn as given below.

FIGURE 14.39

(iv) NAND Gate

This is an AND gate followed by a NOT gate. If inputs A *and* B are both '1', the output Y is *not* '1'. The gate gets its name from this NOT AND behaviour. Figure 14.40 shows the symbol and truth table of NAND gate.

NAND gates are also called *Universal Gates* since by using these gates you can realise other basic gates like OR, AND and NOT (Exercises 14.16 and 14.17).

(a)

Input		Output
A	B	Y
0	0	1
0	1	1
1	0	1
1	1	0

(b)

FIGURE 14.40 (a) Logic symbol, (b) Truth table of NAND gate.

EXAMPLE 14.13

Example 14.13 Sketch the output Y from a NAND gate having inputs A and B given below:

Solution

- For $t < t_1$; A = 1, B = 1; Hence Y = 0
- For t_1 to t_2; A = 0, B = 0; Hence Y = 1
- For t_2 to t_3; A = 0, B = 1; Hence Y = 1
- For t_3 to t_4; A = 1, B = 0; Hence Y = 1

- For t_4 to t_5; A = 1, B = 1; Hence Y = 0
- For t_5 to t_6; A = 0, B = 0; Hence Y = 1
- For $t > t_6$; A = 0, B = 1; Hence Y = 1

FIGURE 14.41

EXAMPLE 14.13

(v) NOR Gate

It has two or more inputs and one output. A NOT- operation applied *after* OR gate gives a NOT-OR gate (or simply NOR gate). Its output Y is '1' only when both inputs A and B are '0', i.e., neither one input *nor* the other is '1'. The symbol and truth table for NOR gate is given in Fig. 14.42.

(a)

Input		Output
A	B	Y
0	0	1
0	1	0
1	0	0
1	1	0

(b)

FIGURE 14.42 (a) Logic symbol, (b) Truth table of NOR gate.

NOR gates are considered as *universal* gates because you can obtain all the gates like AND, OR, NOT by using only NOR gates (Exercises 14.18 and 14.19).

14.11 Integrated Circuits

The conventional method of making circuits is to choose components like diodes, transistor, *R*, *L*, *C* etc., and connect them by soldering wires in the desired manner. Inspite of the miniaturisation introduced by the discovery of transistors, such circuits were still bulky. Apart from this, such circuits were less reliable and less shock proof. The concept of fabricating *an entire circuit* (consisting of many passive components like *R* and *C* and active devices like diode and transistor) on a small single block (or chip) of a semiconductor has revolutionised the electronics technology. Such a circuit is known as *Integrated Circuit* (IC). The most widely used technology is the *Monolithic Integrated Circuit.* The word

FIGURE 14.43 The casing and connection of a 'chip'.

monolithic is a combination of two greek words, *monos* means single and *lithos* means stone. This, in effect, means that the entire circuit is formed on a single silicon crystal (or *chip*). The *chip* dimensions are as small as 1mm × 1mm or it could even be smaller. Figure 14.43 shows a chip in its protective plastic case, partly removed to reveal the connections coming out from the 'chip' to the pins that enable it to make external connections.

Depending on nature of input signals, IC's can be grouped in two categories: (a) *linear* or *analogue IC's* and (b) digital IC's. The linear IC's process analogue signals which change smoothly and continuously over a range of values between a maximum and a minimum. The output is more or less directly proportional to the input, i.e., it varies *linearly* with the input. One of the most useful linear IC's is the operational amplifier.

The digital IC's process signals that have only two values. They contain circuits such as logic gates. Depending upon the level of integration (i.e., the number of circuit components or logic gates), the ICs are termed as Small Scale Integration, SSI (logic gates $\leq$ 10); Medium Scale Integration, MSI (logic gates $\leq$ 100); Large Scale Integration, LSI (logic gates $\leq$ 1000); and Very Large Scale Integration, VLSI (logic gates > 1000). The technology of fabrication is very involved but large scale industrial production has made them very inexpensive.

FASTER AND SMALLER: THE FUTURE OF COMPUTER TECHNOLOGY

The *Integrated Chip* (IC) is at the heart of all computer systems. In fact ICs are found in almost all electrical devices like cars, televisions, CD players, cell phones etc. The miniaturisation that made the modern personal computer possible could never have happened without the IC. ICs are electronic devices that contain many transistors, resistors, capacitors, connecting wires – all in one package. You must have heard of the *microprocessor*. The microprocessor is an IC that processes all information in a computer, like keeping track of what keys are pressed, running programmes, games etc. The IC was first invented by Jack Kilky at Texas Instruments in 1958 and he was awarded Nobel Prize for this in 2000. ICs are produced on a piece of semiconductor crystal (or chip) by a process called *photolithography*. Thus, the entire Information Technology (IT) industry hinges on semiconductors. Over the years, the complexity of ICs has increased while the size of its features continued to shrink. In the past five decades, a dramatic miniaturisation in computer technology has made modern day computers *faster and smaller*. In the 1970s, Gordon Moore, co-founder of INTEL, pointed out that the memory capacity of a chip (IC) approximately doubled every one and a half years. This is popularly known as *Moore's law*. The number of transistors per chip has risen exponentially and each year computers are becoming more powerful, yet cheaper than the year before. It is intimated from current trends that the computers available in 2020 will operate at 40 GHz (40,000 MHz) and would be much smaller, more efficient and less expensive than present day computers. The explosive growth in the semiconductor industry and computer technology is best expressed by a famous quote from Gordon Moore: "If the auto industry advanced as rapidly as the semiconductor industry, a Rolls Royce would get half a million miles per gallon, and it would be cheaper to throw it away than to park it".

SUMMARY

1. Semiconductors are the basic materials used in the present solid state electronic devices like diode, transistor, ICs, etc.
2. Lattice structure and the atomic structure of constituent elements decide whether a particular material will be insulator, metal or semiconductor.
3. Metals have low resistivity (10^{-2} to 10^{-8} Ωm), insulators have very high resistivity ($>10^{8}\,\Omega\,m^{-1}$), while semiconductors have intermediate values of resistivity.
4. Semiconductors are elemental (Si, Ge) as well as compound (GaAs, CdS, etc.).
5. Pure semiconductors are called 'intrinsic semiconductors'. The presence of charge carriers (electrons and holes) is an 'intrinsic' property of the material and these are obtained as a result of thermal excitation. The number of electrons (n_e) is equal to the number of holes (n_h) in intrinsic conductors. Holes are essentially electron vacancies with an effective positive charge.
6. The number of charge carriers can be changed by 'doping' of a suitable impurity in pure semiconductors. Such semiconductors are known as extrinsic semiconductors. These are of two types (n-type and p-type).
7. In n-type semiconductors, $n_e >> n_h$ while in p-type semiconductors $n_h >> n_e$.
8. n-type semiconducting Si or Ge is obtained by doping with pentavalent atoms (donors) like As, Sb, P, etc., while p-type Si or Ge can be obtained by doping with trivalent atom (acceptors) like B, Al, In etc.
9. $n_e n_h = n_i^2$ in all cases. Further, the material possesses an *overall charge neutrality*.
10. There are two distinct band of energies (called valence band and conduction band) in which the electrons in a material lie. Valence band energies are low as compared to conduction band energies. All energy levels in the valence band are filled while energy levels in the conduction band may be fully empty or partially filled. The electrons in the conduction band are free to move in a solid and are responsible for the conductivity. The extent of conductivity depends upon the energy gap (E_g) between the top of valence band (E_V) and the bottom of the conduction band E_C. The electrons from valence band can be excited by heat, light or electrical energy to the conduction band and thus, produce a change in the current flowing in a semiconductor.
11. For insulators $E_g > 3$ eV, for semiconductors E_g is 0.2 eV to 3 eV, while for metals $E_g \approx 0$.
12. p-n junction is the 'key' to all semiconductor devices. When such a junction is made, a 'depletion layer' is formed consisting of immobile ion-cores devoid of their electrons or holes. This is responsible for a junction potential barrier.
13. By changing the external applied voltage, junction barriers can be changed. In forward bias (n-side is connected to negative terminal of the battery and p-side is connected to the positive), the barrier is decreased while the barrier increases in reverse bias. Hence, forward bias current is more (mA) while it is very small (μA) in a p-n junction diode.
14. Diodes can be used for rectifying an ac voltage (restricting the ac voltage to one direction). With the help of a capacitor or a suitable filter, a dc voltage can be obtained.
15. There are some special purpose diodes.

16. Zener diode is one such special purpose diode. In reverse bias, after a certain voltage, the current suddenly increases (breakdown voltage) in a Zener diode. This property has been used to obtain *voltage regulation*.
17. p-n junctions have also been used to obtain many photonic or optoelectronic devices where one of the participating entity is 'photon': (a) Photodiodes in which photon excitation results in a change of reverse saturation current which helps us to measure light intensity; (b) Solar cells which convert photon energy into electricity; (c) Light Emitting Diode and Diode Laser in which electron excitation by a bias voltage results in the generation of light.
18. Transistor is an n-p-n or p-n-p junction device. The central block (thin and lightly doped) is called 'Base' while the other electrodes are 'Emitter' and 'Collectors'. The emitter-base junction is forward biased while collector-base junction is reverse biased.
19. The transistors can be connected in such a manner that either C or E or B is common to both the input and output. This gives the three configurations in which a transistor is used: Common Emitter (CE), Common Collector (CC) and Common Base (CB). The plot between I_C and V_{CE} for fixed I_B is called output characteristics while the plot between I_B and V_{BE} with fixed V_{CE} is called input characteristics. The important transistor parameters for CE-configuration are:

    input resistance, $r_i = \left(\frac{\Delta V_{BE}}{\Delta I_B}\right)_{V_{CE}}$

    output resistance, $r_o = \left(\frac{\Delta V_{CE}}{\Delta I_C}\right)_{I_B}$

    current amplification factor, $\beta = \left(\frac{\Delta I_C}{\Delta I_B}\right)_{V_{CE}}$
20. Transistor can be used as an amplifier and oscillator. In fact, an oscillator can also be considered as a self-sustained amplifier in which a part of output is fed-back to the input in the same phase (positive feed back). The voltage gain of a transistor amplifier in common emitter configuration is: $A_v = \left(\frac{v_o}{v_i}\right) = \beta\frac{R_C}{R_B}$, where R_C and R_B are respectively the resistances in collector and base sides of the circuit.
21. When the transistor is used in the cutoff or saturation state, it acts as a switch.
22. There are some special circuits which handle the digital data consisting of 0 and 1 levels. This forms the subject of Digital Electronics.
23. The important digital circuits performing special logic operations are called logic gates. These are: OR, AND, NOT, NAND, and NOR gates.
24. In modern day circuit, many logical gates or circuits are integrated in one single 'Chip'. These are known as Intgrated circuits (IC).

POINTS TO PONDER

1. The energy bands (E_C or E_V) in the semiconductors are space delocalised which means that these are not located in any specific place inside the solid. The energies are the overall averages. When you see a picture in which E_C or E_V are drawn as straight lines, then they should be respectively taken simply as the *bottom* of conduction band energy levels and *top* of valence band energy levels.

2. In elemental semiconductors (Si or Ge), the n-type or p-type semiconductors are obtained by introducing 'dopants' as defects. In compound semiconductors, the change in relative stoichiometric ratio can also change the type of semiconductor. For example, in ideal GaAs the ratio of Ga:As is 1:1 but in Ga-rich or As-rich GaAs it could respectively be $Ga_{1.1}$ $As_{0.9}$ or $Ga_{0.9}$ $As_{1.1}$. In general, the presence of defects control the properties of semiconductors in many ways.
3. In transistors, the base region is both narrow and lightly doped, otherwise the electrons or holes coming from the input side (say, emitter in CE-configuration) will not be able to reach the collector.
4. We have described an oscillator as a positive feedback amplifier. For stable oscillations, the voltage feedback (V_{fb}) from the output voltage (V_o) should be such that after amplification (A) it should again become V_o. If a fraction β' is feedback, then $V_{fb} = V_o . \beta'$ and after amplification its value $A(v_o.\beta')$ should be equal to V_o. This means that the criteria for stable oscillations to be sustained is $A\ \beta' = 1$. This is known as Barkhausen's Criteria.
5. In an oscillator, the feedback is in the same phase (positive feedback). If the feedback voltage is in opposite phase (negative feedback), the gain is less than 1 and it can never work as oscillator. It will be an amplifier with reduced gain. However, the negative feedback also reduces noise and distortion in an amplifier which is an advantageous feature.

EXERCISES

14.1 In an n-type silicon, which of the following statement is true:
(a) Electrons are majority carriers and trivalent atoms are the dopants.
(b) Electrons are minority carriers and pentavalent atoms are the dopants.
(c) Holes are minority carriers and pentavalent atoms are the dopants.
(d) Holes are majority carriers and trivalent atoms are the dopants.

14.2 Which of the statements given in Exercise 14.1 is true for p-type semiconductos.

14.3 Carbon, silicon and germanium have four valence electrons each. These are characterised by valence and conduction bands separated by energy band gap respectively equal to $(E_g)_C$, $(E_g)_{Si}$ and $(E_g)_{Ge}$. Which of the following statements is true?
(a) $(E_g)_{Si} < (E_g)_{Ge} < (E_g)_C$
(b) $(E_g)_C < (E_g)_{Ge} > (E_g)_{Si}$
(c) $(E_g)_C > (E_g)_{Si} > (E_g)_{Ge}$
(d) $(E_g)_C = (E_g)_{Si} = (E_g)_{Ge}$

14.4 In an unbiased p-n junction, holes diffuse from the p-region to n-region because
(a) free electrons in the n-region attract them.
(b) they move across the junction by the potential difference.
(c) hole concentration in p-region is more as compared to n-region.
(d) All the above.

14.5 When a forward bias is applied to a p-n junction, it
(a) raises the potential barrier.
(b) reduces the majority carrier current to zero.
(c) lowers the potential barrier.
(d) None of the above.

14.6 For transistor action, which of the following statements are correct:
(a) Base, emitter and collector regions should have similar size and doping concentrations.
(b) The base region must be very thin and lightly doped.
(c) The emitter junction is forward biased and collector junction is reverse biased.
(d) Both the emitter junction as well as the collector junction are forward biased.

14.7 For a transistor amplifier, the voltage gain
(a) remains constant for all frequencies.
(b) is high at high and low frequencies and constant in the middle frequency range.
(c) is low at high and low frequencies and constant at mid frequencies.
(d) None of the above.

14.8 In half-wave rectification, what is the output frequency if the input frequency is 50 Hz. What is the output frequency of a full-wave rectifier for the same input frequency.

14.9 For a CE-transistor amplifier, the audio signal voltage across the collected resistance of 2 kΩ is 2 V. Suppose the current amplification factor of the transistor is 100, find the input signal voltage and base current, if the base resistance is 1 kΩ.

14.10 Two amplifiers are connected one after the other in series (cascaded). The first amplifier has a voltage gain of 10 and the second has a voltage gain of 20. If the input signal is 0.01 volt, calculate the output ac signal.

14.11 A p-n photodiode is fabricated from a semiconductor with band gap of 2.8 eV. Can it detect a wavelength of 6000 nm?

ADDITIONAL EXERCISES

14.12 The number of silicon atoms per m^3 is 5×10^{28}. This is doped simultaneously with 5×10^{22} atoms per m^3 of Arsenic and 5×10^{20} per m^3 atoms of Indium. Calculate the number of electrons and holes. Given that $n_i = 1.5 \times 10^{16}\ m^{-3}$. Is the material n-type or p-type?

14.13 In an intrinsic semiconductor the energy gap E_g is 1.2eV. Its hole mobility is much smaller than electron mobility and independent of temperature. What is the ratio between conductivity at 600K and that at 300K? Assume that the temperature dependence of intrinsic carrier concentration n_i is given by

$$n_i = n_0 \exp\left(-\frac{E_g}{2k_B T}\right)$$

where n_0 is a constant.

14.14 In a p-n junction diode, the current I can be expressed as

$$I = I_0 \exp\left(\frac{eV}{2k_B T} - 1\right)$$

where I_0 is called the reverse saturation current, V is the voltage across the diode and is positive for forward bias and negative for reverse bias, and I is the current through the diode, k_B is the Boltzmann constant (8.6×10^{-5} eV/K) and T is the absolute temperature. If for a given diode $I_0 = 5 \times 10^{-12}$ A and T = 300 K, then

(a) What will be the forward current at a forward voltage of 0.6 V?

(b) What will be the increase in the current if the voltage across the diode is increased to 0.7 V?

(c) What is the dynamic resistance?

(d) What will be the current if reverse bias voltage changes from 1 V to 2 V?

14.15 You are given the two circuits as shown in Fig. 14.44. Show that circuit (a) acts as OR gate while the circuit (b) acts as AND gate.

FIGURE 14.44

14.16 Write the truth table for a NAND gate connected as given in Fig. 14.45.

FIGURE 14.45

Hence identify the exact logic operation carried out by this circuit.

14.17 You are given two circuits as shown in Fig. 14.46, which consist of NAND gates. Identify the logic operation carried out by the two circuits.

FIGURE 14.46

14.18 Write the truth table for circuit given in Fig. 14.47 below consisting of NOR gates and identify the logic operation (OR, AND, NOT) which this circuit is performing.

FIGURE 14.47

(Hint: A = 0, B = 1 then A and B inputs of second NOR gate will be 0 and hence Y=1. Similarly work out the values of Y for other combinations of A and B. Compare with the truth table of OR, AND, NOT gates and find the correct one.)

14.19 Write the truth table for the circuits given in Fig. 14.48 consisting of NOR gates only. Identify the logic operations (OR, AND, NOT) performed by the two circuits.

FIGURE 14.48

Chapter Fifteen

COMMUNICATION SYSTEMS

15.1 Introduction

Communication is the act of transmission of information. Every living creature in the world experiences the need to impart or receive information almost continuously with others in the surrounding world. For communication to be successful, it is essential that the sender and the receiver understand a common *language*. Man has constantly made endeavors to improve the quality of communication with other human beings. Languages and methods used in communication have kept evolving from prehistoric to modern times, to meet the growing demands in terms of speed and complexity of information. It would be worthwhile to look at the major milestones in events that promoted developments in communications, as presented in Table 15.1.

Modern communication has its roots in the 19th and 20th century in the work of scientists like J.C. Bose, F.B. Morse, G. Marconi and Alexander Graham Bell. The pace of development seems to have increased dramatically after the first half of the 20th century. We can hope to see many more accomplishments in the coming decades. The aim of this chapter is to introduce the concepts of communication, namely the mode of communication, the need for modulation, production and deduction of amplitude modulation.

15.2 Elements of a Communication System

Communication pervades all stages of life of all living creatures. Irrespective of its nature, every communication system has three essential elements-

TABLE 15.1 SOME MAJOR MILESTONES IN THE HISTORY OF COMMUNICATION

Year	Event	Remarks
Around 1565 A.D.	The reporting of the delivery of a child by queen using drum beats from a distant place to King Akbar.	It is believed that minister Birbal experimented with the arrangement to decide the number of drummers posted between the place where the queen stayed and the place where the king stayed.
1835	Invention of telegraph by Samuel F.B. Morse and Sir Charles Wheatstone	It resulted in tremendous growth of messages through post offices and reduced physical travel of messengers considerably.
1876	Telephone invented by Alexander Graham Bell and Antonio Meucci	Perhaps the most widely used means of communication in the history of mankind.
1895	Jagadis Chandra Bose and Guglielmo Marconi demonstrated wireless telegraphy.	It meant a giant leap – from an era of communication using wires to communicating without using wires. (wireless)
1936	Television broadcast(John Logi Baird)	First television broadcast by BBC
1955	First radio FAX transmitted across continent.(Alexander Bain)	The idea of FAX transmission was patented by Alexander Bain in 1843.
1968	ARPANET- the first internet came into existence(J.C.R. Licklider)	ARPANET was a project undertaken by the U.S. defence department. It allowed file transfer from one computer to another connected to the network.
1975	Fiber optics developed at Bell Laboratories	Fiber optical systems are superior and more economical compared to traditional communication systems.
1989-91	Tim Berners-Lee invented the **World Wide Web**.	WWW may be regarded as the mammoth encyclopedia of knowledge accessible to everyone round the clock throughout the year.

transmitter, medium/channel and receiver. The block diagram shown in Fig. 15.1 depicts the general form of a communication system.

FIGURE 15.1 Block diagram of a generalised communication system.

In a communication system, the transmitter is located at one place, the receiver is located at some other place (far or near) separate from the transmitter and the channel is the physical medium that connects them. Depending upon the type of communication system, a channel may be in the form of wires or cables connecting the transmitter and the receiver or it may be wireless. The purpose of the transmitter is to convert the message signal produced by the source of information into a form suitable for transmission through the channel. If the output of the information source is a non-electrical signal like a voice signal, a transducer converts it to electrical form before giving it as an input to the transmitter. When a transmitted signal propagates along the channel it may get distorted due to channel imperfection. Moreover, noise adds to the transmitted signal and the receiver receives a corrupted version of the transmitted signal. The receiver has the task of operating on the received signal. It reconstructs a recognisable form of the original message signal for delivering it to the user of information.

There are two basic modes of communication: *point-to-point* and *broadcast*.

In point-to-point communication mode, communication takes place over a link between a single transmitter and a receiver. Telephony is an example of such a mode of communication. In contrast, in the broadcast mode, there are a large number of receivers corresponding to a single transmitter. Radio and television are examples of broadcast mode of communication.

15.3 Basic Terminology Used in Electronic Communication Systems

By now, we have become familiar with some terms like information source, transmitter, receiver, channel, noise, etc. It would be easy to understand the principles underlying any communication, if we get ourselves acquainted with the following basic terminology.

JAGADIS CHANDRA BOSE (1858 – 1937)

Jagadis Chandra Bose (1858 – 1937) He developed an apparatus for generating ultrashort electro-magnetic waves and studied their quasi-optical properties. He was said to be the first to employ a semiconductor like galena as a self-recovering detector of electromagnetic waves. Bose published three papers in the British magazine, 'The Electrician' of 27 Dec. 1895. His invention was published in the 'Proceedings of The Royal Society' on 27 April 1899 over two years before Marconi's first wireless communication on 13 December 1901. Bose also invented highly sensitive instruments for the detection of minute responses by living organisms to external stimulii and established parallelism between animal and plant tissues.

(i) *Transducer:* Any device that converts one form of energy into another can be termed as a transducer. In electronic communication systems, we usually come across devices that have either their inputs or outputs in the electrical form. An electrical transducer may be defined as a device that converts some physical variable (pressure, displacement, force, temperature, etc) into corresponding variations in the electrical signal at its output.

(ii) *Signal:* Information converted in electrical form and suitable for transmission is called a signal. Signals can be either *analog or digital*. Analog signals are continuous variations of voltage or current. *They are essentially single-valued functions of time*. Sine wave is a fundamental analog signal. All other analog signals can be fully understood in terms of their sine wave components. Sound and picture signals in TV are analog in nature. Digital signals are those which can take only discrete stepwise values. Binary system that is extensively used in digital electronics employs just two levels of a signal. '0' corresponds to a low level and '1' corresponds to a high level of voltage/current. There are several coding schemes useful for digital communication. They employ suitable combinations of number systems such as the binary coded decimal (BCD)*. American Standard Code for Information Interchange (ASCII)** is a universally popular digital code to represent numbers, letters and certain characters.

(iii) *Noise:* Noise refers to the unwanted signals that tend to disturb the transmission and processing of message signals in a communication system. The source generating the noise may be located inside or outside the system.

(iv) *Transmitter:* A transmitter processes the incoming message signal so as to make it suitable for transmission through a channel and subsequent reception.

(v) *Receiver:* A receiver extracts the desired message signals from the received signals at the channel output.

(vi) *Attenuation:* The loss of strength of a signal while propagating through a medium is known as attenuation.

---

* In BCD, a digit is usually represented by four binary (0 or 1) bits. For example the numbers 0, 1, 2, 3, 4 in the decimal system are written as 0000, 0001, 0010, 0011 and 0100. 1000 would represent eight.

** It is a character encoding in terms of numbers based on English alphabet since the computer can only understand numbers.

(vii) *Amplification:* It is the process of *increasing the amplitude* (and consequently the strength) of a signal using an electronic circuit called the amplifier (reference Chapter 14). Amplification is necessary to compensate for the attenuation of the signal in communication systems. The energy needed for additional signal strength is obtained from a DC power source. Amplification is done at a place between the source and the destination wherever signal strength becomes weaker than the required strength.

(viii) *Range:* It is the largest distance between a source and a destination up to which the signal is received with sufficient strength.

(ix) *Bandwidth:* Bandwidth refers to the frequency range over which an equipment operates or the portion of the spectrum occupied by the signal.

(x) *Modulation:* The original low frequency message/information signal cannot be transmitted to long distances because of reasons given in Section 15.7. Therefore, at the transmitter, information contained in the low frequency message signal is superimposed on a high frequency wave, which acts as a carrier of the information. This process is known as modulation. As will be explained later, there are several types of modulation, abbreviated as AM, FM and PM.

(xi) *Demodulation:* The process of retrieval of information from the carrier wave at the receiver is termed demodulation. This is the reverse process of modulation.

(xii) *Repeater:* A repeater is a combination of a receiver and a transmitter. A repeater, picks up the signal from the transmitter, amplifies and retransmits it to the receiver sometimes with a change in carrier frequency. Repeaters are used to extend the range of a communication system as shown in Fig. 15.2. A communication satellite is essentially a repeater station in space.

FIGURE 15.2 Use of repeater station to increase the range of communication.

15.4 Bandwidth of Signals

In a communication system, the message signal can be voice, music, picture or computer data. Each of these signals has different ranges of frequencies. The type of communication system needed for a given signal depends on the band of frequencies which is considered essential for the communication process.

For speech signals, frequency range 300 Hz to 3100 Hz is considered adequate. Therefore speech signal requires a bandwidth of 2800 Hz (3100 Hz – 300 Hz) for commercial telephonic communication. To transmit music,

an approximate bandwidth of 20 kHz is required because of the high frequencies produced by the musical instruments. The audible range of frequencies extends from 20 Hz to 20 kHz.

Video signals for transmission of pictures require about 4.2 MHz of bandwidth. A TV signal contains both voice and picture and is usually allocated 6 MHz of bandwidth for transmission.

In the preceeding paragraph, we have considered only analog signals. Digital signals are in the form of rectangular waves as shown in Fig. 15.3. One can show that this rectangular wave can be decomposed into a superposition of sinusoidal waves of frequencies ν_0, $2\nu_0$, $3\nu_0$, $4\nu_0$ ... $n\nu_0$ where n is an integer extending to infinity and $\nu_0 = 1/T_0$. The fundamental (ν_0), fundamental (ν_0) + second harmonic ($2\nu_0$), and fundamental (ν_0) + second harmonic ($2\nu_0$) + third harmonic ($3\nu_0$), are shown in the same figure to illustrate this fact. It is clear that to reproduce the rectangular wave shape exactly we need to superimpose all the harmonics ν_0, $2\nu_0$, $3\nu_0$, $4\nu_0$..., which implies an infinite bandwidth. However, for practical purposes, the contribution from higher harmonics can be neglected, thus limiting the bandwidth. As a result, received waves are a distorted version of the transmitted one. If the bandwidth is large enough to accommodate a few harmonics, the information is not lost and the rectangular signal is more or less recovered. This is so because the higher the harmonic, less is its contribution to the wave form.

FIGURE 15.3 Approximation of a rectangular wave in terms of a fundamental sine wave and its harmonics.

15.5 Bandwidth of Transmission Medium

Similar to message signals, different types of transmission media offer different bandwidths. The commonly used transmission media are wire, free space and fiber optic cable. Coaxial cable is a widely used wire medium, which offers a bandwidth of approximately 750 MHz. Such cables are normally operated below 18 GHz. Communication through free space using radio waves takes place over a very wide range of frequencies: from a few hundreds of kHz to a few GHz. This range of frequencies is further subdivided and allocated for various services as indicated in Table 15.2. Optical communication using fibers is performed in the frequency range of 1 THz to 1000 THz (microwaves to ultraviolet). An optical fiber can offer a transmission bandwidth in excess of 100 GHz.

Spectrum allocations are arrived at by an international agreement. The International Telecommunication Union (ITU) administers the present system of frequency allocations.

TABLE 15.2 SOME IMPORTANT WIRELESS COMMUNICATION FREQUENCY BANDS

Service	Frequency bands	Comments
Standard AM broadcast	540-1600 kHz	
FM broadcast	88-108 MHz	
Television	54-72 MHz	VHF (very high frequencies)
	76-88 MHz	TV
	174-216 MHz	UHF (ultra high frequencies)
	420-890 MHz	TV
Cellular Mobile Radio	896-901 MHz	Mobile to base station
	840-935 MHz	Base station to mobile
Satellite Communication	5.925-6.425 GHz	Uplink
	3.7-4.2 GHz	Downlink

15.6 PROPAGATION OF ELECTROMAGNETIC WAVES

In communication using radio waves, an antenna at the transmitter radiates the Electromagnetic waves (em waves), which travel through the space and reach the receiving antenna at the other end. As the em wave travels away from the transmitter, the strength of the wave keeps on decreasing. Several factors influence the propagation of em waves and the path they follow. At this point, it is also important to understand the composition of the earth's atmosphere as it plays a vital role in the propagation of em waves. A brief discussion on some useful layers of the atmosphere is given in Table 15.3.

15.6.1 Ground wave

To radiate signals with high efficiency, the antennas should have a size comparable to the wavelength λ of the signal (at least ~ $\lambda/4$). At longer wavelengths (i.e., at lower frequencies), the antennas have large physical size and they are located on or very near to the ground. In standard AM broadcast, ground based vertical towers are generally used as transmitting antennas. For such antennas, ground has a strong influence on the propagation of the signal. The mode of propagation is called surface wave propagation and the wave glides over the surface of the earth. A wave induces current in the ground over which it passes and it is attenuated as a result of absorption of energy by the earth. The attenuation of surface waves increases very rapidly with increase in frequency. The maximum range of coverage depends on the transmitted power and frequency (less than a few MHz).

TABLE 15.3 DIFFERENT LAYERS OF ATMOSPHERE AND THEIR INTERACTION WITH THE PROPAGATING ELECTROMAGNETIC WAVES

Name of the stratum (layer)		Approximate height over earth's surface	Exists during	Frequencies most affected
Troposphere		10 km	Day and night	VHF (up to several GHz)
D (part of stratosphere)	PARTS OF IONOSPHERE	65-75 km	Day only	Reflects LF, absorbs MF and HF to some degree
E (part of Stratosphere)		100 km	Day only	Helps surface waves, reflects HF
F_1 (Part of Mesosphere)		170-190 km	Daytime, merges with F_2 at night	Partially absorbs HF waves yet allowing them to reach F_2
F_2 (Thermosphere)		300 km at night, 250-400 km during daytime	Day and night	Efficiently reflects HF waves, particularly at night

15.6.2 Sky waves

In the frequency range from a few MHz up to 30 to 40 MHz, long distance communication can be achieved by ionospheric reflection of radio waves back towards the earth. This mode of propagation is called *sky wave propagation* and is used by short wave broadcast services. The ionosphere is so called because of the presence of a large number of ions or charged particles. It extends from a height of ~ 65 Km to about 400 Km above the earth's surface. Ionisation occurs due to the absorption of the ultraviolet and other high-energy radiation coming from the sun by air molecules. The ionosphere is further subdivided into several layers, the details of which are given in Table 15.3. The degree of ionisation varies with the height. The density of atmosphere decreases with height. At great heights the solar radiation is intense but there are few molecules to be ionised. Close to the earth, even though the molecular concentration is very high, the radiation intensity is low so that the ionisation is again low. However, at some intermediate heights, there occurs a peak of ionisation density. The ionospheric layer acts as a reflector for a certain range of frequencies (3 to 30 MHz). Electromagnetic waves of frequencies higher than 30 MHz penetrate the ionosphere and escape. These phenomena are shown in the Fig. 15.4. The phenomenon of bending of em waves so that they are diverted towards the earth is similar to total internal reflection in optics*.

* Compare this with the phenomenon of mirage.

FIGURE 15.4 Sky wave propagation. The layer nomenclature is given in Table 15.3.

15.6.3 Space wave

Another mode of radio wave propagation is by *space waves*. A space wave travels in a straight line from transmitting antenna to the receiving antenna. Space waves are used for line-of-sight (LOS) communication as well as satellite communication. At frequencies above 40 MHz, communication is essentially limited to line-of-sight paths. At these frequencies, the antennas are relatively smaller and can be placed at heights of many wavelengths above the ground. Because of line-of-sight nature of propagation, direct waves get blocked at some point by the curvature of the earth as illustrated in Fig. 15.5. If the signal is to be received beyond the horizon then the receiving antenna must be high enough to intercept the line-of-sight waves.

FIGURE 15.5 Line of sight communication by space waves.

If the transmitting antenna is at a height h_T, then you can show that the distance to the horizon d_T is given as $d_T = \sqrt{2Rh_T}$, where R is the radius of the earth (approximately 6400 km). d_T is also called the radio horizon of the transmitting antenna. With reference to Fig. 15.5 the maximum line-of-sight distance d_M between the two antennas having heights h_T and h_R above the earth is given by

$$d_M = \sqrt{2Rh_T} + \sqrt{2Rh_R} \qquad (15.1)$$

where h_R is the height of receiving antenna.

Television broadcast, microwave links and satellite communication are some examples of communication systems that use space wave mode of propagation. Figure 15.6 summarises the various modes of wave propagation discussed so far.

FIGURE 15.6 Various propagation modes for em waves.

EXAMPLE 15.1

Example 15.1 A transmitting antenna at the top of a tower has a height 32 m and the height of the receiving antenna is 50 m. What is the maximum distance between them for satisfactory communication in LOS mode? Given radius of earth 6.4×10^6 m.

Solution

$$d_m = \sqrt{2 \times 64 \times 10^5 \times 32} + \sqrt{2 \times 64 \times 10^5 \times 50} \text{ m}$$

$$= 64 \times 10^2 \times \sqrt{10} + 8 \times 10^3 \times \sqrt{10} \text{ m}$$

$$= 144 \times 10^2 \times \sqrt{10} \text{ m} = 45.5 \text{ km}$$

15.7 Modulation and its Necessity

As already mentioned, the purpose of a communication system is to transmit information or message signals. Message signals are also called *baseband signals*, which essentially designate the band of frequencies representing the original signal, as delivered by the source of information. No signal, in general, is a single frequency sinusoid, but it spreads over a range of frequencies called the signal *bandwidth*. Suppose we wish to transmit an electronic signal in the audio frequency (AF) range (baseband signal frequency less than 20 kHz) over a long distance directly. Let us find what factors prevent us from doing so and how we overcome these factors,

15.7.1 Size of the antenna or aerial

For transmitting a signal, we need an antenna or an aerial. This antenna should have a size comparable to the wavelength of the signal (at least $\lambda/4$ in dimension) so that the antenna properly senses the time variation of the signal. For an electromagnetic wave of frequency 20 kHz, the wavelength λ is 15 km. Obviously, such a long antenna is not possible to construct and operate. Hence direct transmission of such baseband signals is not practical. We can obtain transmission with reasonable antenna lengths if transmission frequency is high (for example, if ν is 1 MHz, then λ is 300 m). Therefore, there is a need of *translating the information contained in our original low frequency baseband signal into high or radio frequencies before transmission.*

15.7.2 Effective power radiated by an antenna

A theoretical study of radiation from a linear antenna (length l) shows that the power radiated is proportional to $(l/\lambda)^2$. This implies that for the same antenna length, the power radiated increases with decreasing λ, i.e., increasing frequency. Hence, the effective power radiated by a long wavelength baseband signal would be small. For a good transmission, we need high powers and hence this also points out to *the need* of using high frequency transmission.

15.7.3 Mixing up of signals from different transmitters

Another important argument against transmitting baseband signals directly is more *practical* in nature. Suppose many people are talking at the same time or many transmitters are transmitting baseband information signals simultaneously. All these signals will get mixed up and there is no simple way to distinguish between them. This points out towards a possible solution by using communication at high frequencies and allotting a *band* of frequencies to each message signal for its transmission.

The above arguments suggest that there is a *need for translating the original low frequency baseband message or information signal into high frequency wave before transmission such that the translated signal continues to possess the information contained in the original signal.* In doing so, we take the help of a high frequency signal, known as the *carrier wave*, and a process known as modulation which attaches information to it. The carrier wave may be continuous (sinusoidal) or in the form of pulses as shown in Fig. 15.7.

FIGURE 15.7 (a) Sinusoidal, and (b) pulse shaped signals.

A sinusoidal carrier wave can be represented as

$$c(t) = A_c \sin(\omega_c t + \phi) \qquad (15.2)$$

where $c(t)$ is the signal strength (voltage or current), A_c is the amplitude, $\omega_c (= 2\pi\nu_c)$ is the angular frequency and ϕ is the initial phase of the carrier wave. During the process of modulation, any of the three parameters, *viz* A_c, ω_c and ϕ, of the carrier wave can be controlled by the message or

information signal. This results in three types of modulation: (i) Amplitude modulation (AM), (ii) Frequency modulation (FM) and (iii) Phase modulation (PM), as shown in Fig. 15.8.

FIGURE 15.8 Modulation of a carrier wave: (a) a sinusoidal carrier wave; (b) a modulating signal; (c) amplitude modulation; (d) frequency modulation; and (e) phase modulation.

Similarly, the significant characteristics of a pulse are: pulse amplitude, pulse duration or pulse Width, and pulse position (denoting the time of *rise* or *fall* of the pulse amplitude) as shown in Fig. 15.7(b). Hence, different types of pulse modulation are: (a) pulse amplitude modulation (PAM), (b) pulse duration modulation (PDM) or pulse width modulation (PWM), and (c) pulse position modulation (PPM). In this chapter, we shall confine to amplitude modulation on ly.

Modulation and Demodulation
http://www.williamson labs.com/480_mod.htm
PHYSICS

15.8 Amplitude Modulation

In amplitude modulation the amplitude of the carrier is varied in accordance with the information signal. Here we explain amplitude modulation process using a sinusoidal signal as the modulating signal.

Let $c(t) = A_c \sin \omega_c t$ represent carrier wave and $m(t) = A_m \sin \omega_m t$ represent the message or the modulating signal where $\omega_m = 2\pi f_m$ is the angular frequency of the message signal. The modulated signal $c_m(t)$ can be written as

$$c_m(t) = (A_c + A_m \sin \omega_m t) \sin \omega_c t$$

$$= A_c \left(1 + \frac{A_m}{A_c} \sin\omega_m t\right) \sin\omega_c t \qquad (15.3)$$

Note that the modulated signal now contains the message signal. This can also be seen from Fig. 15.8(c). From Eq. (15.3), we can write,

$$c_m(t) = A_c \sin\omega_c t + \mu A_c \sin\omega_m t \sin\omega_c t \qquad (15.4)$$

Here $\mu = A_m/A_c$ is the *modulation index;* in practice, μ is kept ≤ 1 to avoid distortion.

Using the trignomatric relation $\sin A \sin B = ½ (\cos(A - B) - \cos (A + B)$, we can write $c_m(t)$ of Eq. (15.4) as

$$c_m(t) = A_c \sin \omega_c t + \frac{\mu A_c}{2} \cos(\omega_c - \omega_m)\, t - \frac{\mu A_c}{2} \cos(\omega_c + \omega_m)t \qquad (15.5)$$

Here $\omega_c - \omega_m$ and $\omega_c + \omega_m$ are respectively called the lower side and upper side frequencies. The modulated signal now consists of the carrier wave of frequency ω_c plus two sinusoidal waves each with a frequency slightly different from, known as side bands. The frequency spectrum of the amplitude modulated signal is shown in Fig. 15.9.

FIGURE 15.9 A plot of amplitude versus ω for an amplitude modulated signal.

As long as the broadcast frequencies (carrier waves) are sufficiently spaced out so that sidebands do not overlap, different stations can operate without interfering with each other.

Example 15.2 A message signal of frequency 10 kHz and peak voltage of 10 volts is used to modulate a carrier of frequency 1 MHz and peak voltage of 20 volts. Determine (a) modulation index, (b) the side bands produced.

Solution

(a) Modulation index =10/20 = 0.5

(b) The side bands are at (1000+10 kHz)=1010 kHz and (1000 –10 kHz) = 990 kHz.

EXAMPLE 15.2

15.9 Production of Amplitude Modulated Wave

Amplitude modulation can be produced by a variety of methods. A conceptually simple method is shown in the block diagram of Fig. 15.10.

FIGURE 15.10 Block diagram of a simple modulator for obtaining an AM signal.

Here the modulating signal $A_m \sin \omega_m t$ is added to the carrier signal $A_c \sin \omega_c t$ to produce the signal $x(t)$. This signal $x(t) = A_m \sin\omega_m t + A_c \sin\omega_c t$ is passed through a square law device which is a non-linear device which produces an output

$$y(t) = B\,x(t) + Cx^2(t) \tag{15.6}$$

where B and C are constants. Thus,

$$y(t) = BA_m \sin \omega_m t + BA_c \sin \omega_c t$$

$$+C\left[A_m^2 \sin^2 \omega_m t + A_c^2 \sin^2 \omega_c t + 2A_m A_c \sin \omega_m t \sin \omega_c t\right] \tag{15.7}$$

$$= BA_m \sin \omega_m t + BA_c \sin \omega_c t$$

$$+\frac{C\,A_m^2}{2} + A_c^2 - \frac{C\,A_m^2}{2}\cos 2\omega_m t - \frac{C\,A_c^2}{2}\cos 2\omega_c t$$

$$+ CA_m A_c \cos(\omega_c - \omega_m)\,t - CA_m A_c \cos(\omega_c + \omega_m)\,t \tag{15.8}$$

where the trigonometric relations $\sin^2 A = (1 - \cos 2A)/2$ and the relation for $\sin A \sin B$ mentioned earlier are used.

In Eq. (15.8), there is a dc term $C/2\left(A_m^2 + A_c^2\right)$ and sinusoids of frequencies ω_m, $2\omega_m$, ω_c, $2\omega_c$, $\omega_c - \omega_m$ and $\omega_c + \omega_m$. As shown in Fig. 15.10 this signal is passed through a band pass filter* which rejects dc and the sinusoids of frequencies ω_m, $2\omega_m$ and $2\,\omega_c$ and retains the frequencies ω_c, $\omega_c - \omega_m$ and $\omega_c + \omega_m$. The output of the band pass filter therefore is of the same form as Eq. (15.5) and is therefore an AM wave.

It is to be mentioned that the modulated signal cannot be transmitted as such. The modulator is to be followed by a power amplifier which provides the necessary power and then the modulated signal is fed to an antenna of appropriate size for radiation as shown in Fig. 15.11.

FIGURE 15.11 Block diagram of a transmitter.

15.10 Detection of Amplitude Modulated Wave

The transmitted message gets attenuated in propagating through the channel. The receiving antenna is therefore to be followed by an amplifier and a detector. In addition, to facilitate further processing, the carrier frequency is usually changed to a lower frequency by what is called an *intermediate frequency (IF) stage* preceding the detection. The detected signal may not be strong enough to be made use of and hence is required

* A band pass filter rejects low and high frequencies and allows a band of frequencies to pass through.

to be amplified. A block diagram of a typical receiver is shown in Fig. 15.12

FIGURE 15.12 Block diagram of a receiver.

Detection is the process of recovering the modulating signal from the modulated carrier wave. We just saw that the modulated carrier wave contains the frequencies ω_c and $\omega_c \pm \omega_m$. In order to obtain the original message signal $m(t)$ of angular frequency ω_m, a simple method is shown in the form of a block diagram in Fig. 15.13.

FIGURE 15.13 Block diagram of a detector for AM signal. The quantity on y-axis can be current or voltage.

The modulated signal of the form given in (a) of fig. 15.13 is passed through a rectifier to produce the output shown in (b). This envelope of signal (b) is the message signal. In order to retrieve $m(t)$, the signal is passed through an envelope detector (which may consist of a simple RC circuit).

In the present chapter we have discussed some basic concepts of communication and communication systems. We have also discussed one specific type of analog modulation namely Amplitude Modulation (AM). Other forms of modulation and digital communication systems play an important role in modern communication. These and other exciting developments are taking place everyday.

So far we have restricted our discussion to some basic communication systems. Before we conclude this chapter, it is worth taking a glance at some of the communication systems (see the box) that in recent times have brought major changes in the way we exchange information even in our day-to-day life:

Additional information

The Internet

It is a system with billions of users worldwide. It permits communication and sharing of all types of information between any two or more computers connected through a large and complex network. It was started in 1960's and opened for public use in 1990's. With the passage of time it has witnessed tremendous growth and it is still expanding its reach. Its applications include

(i) *E mail* – It permits exchange of text/graphic material using email software. We can write a letter and send it to the recipient through ISP's (Internet Service Providers) who work like the dispatching and receiving post offices.

(ii) *File transfer* – A FTP (File Transfer Programmes) allows transfer of files/software from one computer to another connected to the Internet.

(iii) *World Wide Web (WWW)* – Computers that store specific information for sharing with others provide *websites* either directly or through web service providers. Government departments, companies, NGO's (Non-Government Organisations) and individuals can post information about their activities for restricted or free use on their websites. This information becomes accessible to the users. Several search engines like Google, Yahoo! etc. help us in finding information by listing the related websites. *Hypertext* is a powerful feature of the web that automatically links relevant information from one page on the web to another using *HTML (hypertext markup language)*.

(iv) *E-commerce* – Use of the Internet to promote business using electronic means such as using credit cards is called E-commerce. Customers view images and receive all the information about various products or services of companies through their websites. They can do *on-line shopping* from home/office. Goods are dispatched or services are provided by the company through mail/courier.

(v) *Chat* – Real time conversation among people with common interests through typed messages is called chat. Everyone belonging to the *chat group* gets the message instantaneously and can respond rapidly.

Facsimile (FAX)

It scans the contents of a document (as an image, not text) to create electronic signals. These signals are then sent to the destination (another FAX machine) in an orderly manner using telephone lines. At the destination, the signals are reconverted into a replica of the original document. Note that FAX provides image of a static document unlike the image provided by television of objects that might be dynamic.

Mobile telephony

The concept of mobile telephony was developed first in 1970's and it was fully implemented in the following decade. The central concept of this system is to divide the service area into a suitable number of *cells* centred on an office called *MTSO (Mobile Telephone Switching Office)*. Each cell contains a low-power transmitter called a *base station* and caters to a large number of mobile receivers (popularly called cell phones). Each cell could have a service area of a few square kilometers or even less depending upon the number of customers. When a mobile receiver crosses the coverage area of one base station, it is necessary for the mobile user to be transferred to another base station. This procedure is called *handover* or *handoff*. This process is carried out very rapidly, to the extent that the consumer does not even notice it. Mobile telephones operate typically in the UHF range of frequencies (about 800-950 MHz).

SUMMARY

1. Electronic communication refers to the faithful transfer of information or message (available in the form of electrical voltage and current) from one point to another point.
2. Transmitter, transmission channel and receiver are three basic units of a communication system.
3. Two important forms of communication system are: Analog and Digital. The information to be transmitted is generally in continuous waveform for the former while for the latter it has only discrete or quantised levels.
4. Every message signal occupies a range of frequencies. The bandwidth of a message signal refers to the band of frequencies, which are necessary for satisfactory transmission of the information contained in the signal. Similarly, any practical communication system permits transmission of a range of frequencies only, which is referred to as the bandwidth of the system.
5. Low frequencies cannot be transmitted to long distances. Therefore, they are superimposed on a high frequency carrier signal by a process known as modulation.
6. In modulation, some characteristic of the carrier signal like amplitude, frequency or phase varies in accordance with the modulating or message signal. Correspondingly, they are called Amplitude Modulated (AM), Frequency Modulated (FM) or Phase Modulated (PM) waves.
7. Pulse modulation could be classified as: Pulse Amplitude Modulation (PAM), Pulse Duration Modulation (PDM) or Pulse Width Modulation (PWM) and Pulse Position Modulation (PPM).
8. For transmission over long distances, signals are radiated into space using devices called antennas. The radiated signals propagate as electromagnetic waves and the mode of propagation is influenced by the presence of the earth and its atmosphere. Near the surface of the earth, electromagnetic waves propagate as surface waves. Surface wave propagation is useful up to a few MHz frequencies.
9. Long distance communication between two points on the earth is achieved through reflection of electromagnetic waves by ionosphere. Such waves are called sky waves. Sky wave propagation takes place up to frequency of about 30 MHz. Above this frequency, electromagnetic waves essentially propagate as space waves. Space waves are used for line-of-sight communication and satellite communication.
10. If an antenna radiates electromagnetic waves from a height h_T, then the range d_T is given by $\sqrt{2Rh_T}$ where R is the radius of the earth.
11. Amplitude modulated signal contains frequencies $(\omega_c \quad \omega_m)$, ω_c and $(\omega_c + \omega_m)$.
12. Amplitude modulated waves can be produced by application of the message signal and the carrier wave to a non-linear device, followed by a band pass filter.
13. AM detection, which is the process of recovering the modulating signal from an AM waveform, is carried out using a rectifier and an envelope detector.

POINTS TO PONDER

1. In the process of transmission of message/ information signal, noise gets added to the signal anywhere between the information source and the receiving end. Can you think of some sources of noise?
2. In the process of modulation, new frequencies called sidebands are generated on either side (higher and lower than the carrier frequency) of the carrier by an amount equal to the highest modulating frequency. Is it possible to retrieve the message by transmitting (a) only the side bands, (b) only one side band?
3. In amplitude modulation, modulation index $\mu \leq 1$ is used. What will happen if $\mu > 1$?

EXERCISES

15.1 Which of the following frequencies will be suitable for beyond-the-horizon communication using sky waves?
(a) 10 kHz
(b) 10 MHz
(c) 1 GHz
(d) 1000 GHz

15.2 Frequencies in the UHF range normally propagate by means of:
(a) Ground waves.
(b) Sky waves.
(c) Surface waves.
(d) Space waves.

15.3 Digital signals
(i) do not provide a continuous set of values,
(ii) represent values as discrete steps,
(iii) can utilize binary system, and
(iv) can utilize decimal as well as binary systems.
Which of the above statements are true?
(a) (i) and (ii) only
(b) (ii) and (iii) only
(c) (i), (ii) and (iii) but not (iv)
(d) All of (i), (ii), (iii) and (iv).

15.4 Is it necessary for a transmitting antenna to be at the same height as that of the receiving antenna for line-of-sight communication? A TV transmitting antenna is 81m tall. How much service area can it cover if the receiving antenna is at the ground level?

15.5 A carrier wave of peak voltage 12V is used to transmit a message signal. What should be the peak voltage of the modulating signal in order to have a modulation index of 75%?

15.6 A modulating signal is a square wave, as shown in Fig. 15.14.

FIGURE 15.14

The carrier wave is given by $c(t) = 2\sin(8\pi t)$ volts.

(i) Sketch the amplitude modulated waveform

(ii) What is the modulation index?

15.7 For an amplitude modulated wave, the maximum amplitude is found to be 10V while the minimum amplitude is found to be 2V. Determine the modulation index, μ.

What would be the value of μ if the minimum amplitude is zero volt?

15.8 Due to economic reasons, only the upper sideband of an AM wave is transmitted, but at the receiving station, there is a facility for generating the carrier. Show that if a device is available which can multiply two signals, then it is possible to recover the modulating signal at the receiver station.

INDEX

Absorption spectra 421
AC current 233
AC Generator 224
AC voltage 233
    applied to a capacitor 241
    applied to a resistor 234
    applied to an inductor 237
    applied to a series LCR circuit 244
Accelerators in India 142
Accommodation of eye 336
Activity of radioactive substances 447
Additivity of charges 8
Alpha decay 449
Alpha particle scattering 415
Ammeter 165
Ampere 155
Amperes circuital law 147
Amplification 517
Amplitude modulation 524
Analog signal 502
AND gate 503
Andre, Ampere 148
Angle
    of deviation 330
    of incidence 355
    of reflection 355
    of refraction 355
Angular magnification 341
Apparent depth 318
Area element vector 26
Astigmatism 337
Atomic
    mass unit 439
    number 440
    spectra 420
Attenuation 516
Aurora Boriolis 139
Band gap 470
Bandwidth of signal 517
Bandwidth of transmission medium 518
Bar magnet 174
    as solenoid 176
Barrier potential 479
Base 491
Becquerel 448
Beta decay 450
Binding energy per nucleon 444
Biot-Savart law 143
Bohr magneton 163
Bohr radius 425
Bohr's model of atom 422
Bohr's postulates 424
Brewster's angle 380
Brewster's law 381
C A. Volta 53
Capacitance 73
Capacitive reactance 241
Capacitive circuit 252
Capacitor
    parallel plate 74
    in parallel 79
    in series 78
Cartesian sign convention 311
Cassegrain telescope 342
Cells 110
    in parallel 114
    in series 113
Chain reaction 452
Channel 515
Charging by induction 6
Charles August de Coulomb 11
Chromatic aberration 333
Ciliary muscles 337
Coercivity 195
Coherent source 360
Collector 491
Colour code of resistors 103
Combination of lenses 328

Index

Combination of resistors
  series 107
  parallel 108
Composition of nucleus 438
Concave mirror 312
Conduction band 469
Conductivity 97, 468
Conductors 5
Conservation of charge 8
Conservative force 51
Continuous charge distribution 32
Control rods 454
Convex mirror 312
Coulomb 11
Coulomb's law 10
Critical angle 320
Curie temperature 194
Curie 448
Current 94
  amplification factor 495
  density 97
  loop as a magnetic dipole 160
  sensitivity of galvanometer 165
Cut-off voltage/Stopping potential 391
Cyclotron 140
  frequency 141
Davisson & Germer Experiment 403
de Broglie
  relation 398
  wavelength 398
  explanation 430
Decay constant 446
Detection of amplitude modulated wave 526
Diamagnetism 192
Dielectrics 71
Dielectric
  constant 76
  strength 74
Diffraction 367
  single slit 368
Digital
  electronics 501
  signal 502
Dioptre 328
Dipole
  moment 28
  moment vector 28
  in uniform electric field 31
  physical significance 29
Dispersion by a prism 332
Displacement current 270
Doppler effect 358
Drift velocity 98
Earth's magnetism 185
Earthing 6
Eddy currents 218
Einstein's photoelectric equation 394
Electric
  charge 1
  current 93
  dipole 27
  displacement 77
  field 18
  field, physical significance 20
  field due to a system of charges 19
  field lines 23
  flux 25
  susceptibility 72
Electrical energy 105
Electromagnetic
  waves, sources 274
  waves, nature 275
  damping 218
  spectrum 280
Electron emission 387
Electrostatic
  analog 180
  potential 53
  shielding 69
Electrostatics 1
  of conductors 67
Electromotive force (emf) 110
Emission spectra 421
Emitter 491
Energy
  bands 469
  generation in stars 455
  levels 427
  stored in a capacitor 80
Equipotential surfaces 60
Excited state 427
Experiments of Faraday & Henry 205
Extrinsic semiconductor 474
Eye 336
Farad 75
Faraday's law of Induction 207
Fast breeder reactor 453
Ferromagnetism 193
Field
  due to infinite plane sheet 38
  due to uniformly charged thin spherical shell 39
Field emission 388
Flemings left hand rule
Flux leakage 261
Focal length 311
Force between two parallel currents 154
Forward bias 479
Franck-Hertz experiment 428

Fringe width 364
Full-wave rectifier 483
G.S. Ohm 96
Gamma
  rays 283
  decay 451
Gauss's law 33
  its applications 37
  in magnetism 181
Gaussian surface 35
Geographic meridian 186
Gold leaf electroscope 4
Ground
  state 427
  wave 519
H.A. Lorentz 134
Half life 448
Half-wave rectifier 483
Hallwachs' and Lenard's observations 388
Henry 220
Hertz Experiment 274
Holes 472
Horizontal component of earth's magnetic field 187
Huygen's Principle 353
Hypermetropia 337
Impact parameter 418
Impedence diagram 246
Inductance 219
  mutual 220
  self 222
Induction 6
  of charge 6
Inductive
  circuit 252
  reactance 238
Input resistance of a transistor 494
Insulators 5
Integrated circuits (IC) 505
Interference
  constructive 361
  destructive 361
  fringes 363
Internal resistance 110
Intrinsic semiconductor 472
Ionisation energy 427
Iris 337
Isobars 441
Isotones 441
Isotopes 439
J.C. Maxwell 270
Junction transistor 490
K.F. Gauss 182
Kirchhoff's rules 115
Lateral shift 317
Law
  of radioactive decay 447
  of reflection 357
  of refraction 356
LC oscillations 255
Least distance of distinct vision 336
Lenz's law 210
Lens maker's formula 326
Light emitting diode 488
Limitations of Ohm's law 101
Linear
  charge density 32
  magnification/Magnifying power 339
Logic gates 502
Lorentz force 134
Magnetic
  declination 186
  dipole 177
  dipole moment of a revolving electron 162
  field 132
  field lines 175
  field on the axis of a circular current loop 145
  flux 182, 206
  force on a current carrying conductor 135
  force 133
  hysteresis 195
  inclination 187
  intensity 190
  meridian 186
  moment of a current loop 158
  moment 178
  permeability 190
  potential energy 178
  susceptibility 190
  torque 178
Magnetisation 189
Majority carriers 476
Mass
  defect 443
  number 440
  energy relation 442
Maxwell's equations 273
Mean life 448
Meter bridge 120
Michael Faraday 208
Microscope 339
  compound 340
Microwaves 281
Minority carriers 476
Mirage 321
Mirror equation 314
Mobility 100
Moderator 454

Index

Modulation 517, 522
  index 525
Motion in a magnetic field 137
Motional emf 212
Moving coil galvanometer 163
Multiplication factor (fission) 454
Myopia 336
NAND gate 504
Near point 336
Neutrons 440
Noise 516
Non-polar molecules 72
NOR gate 505
North pole 174
NOT gate 502
n-p-n transistor 491
n-type semi conductor 475
Nuclear
  binding energy 442
  density 442
  energy 451
  fission 452
  force 445
  fusion 455
  holocaust 457
  reactor 452
  size 441
  winter 457
Numerical aperture 375
Ohm 95
Ohm's law 95
Optical fibers 322
OR gate 502
Orbital magnetic moment 163
Output resistance of a transistor 495
Paramagnetism 192
Permanent magnets 195
Permeability of free space 143
Permittivity
  of free space 11, 76
  of medium 76
Phasors 237
  diagram 237
Photodiode 487
Photoelectric effect 388
Photocell 399
Photoelectric emission 388
Photoelectrons 389
Photon 395
Pith ball 2
Plane polarised wave 377
p-n Junction 478
p-n-p transistor 491
Point charge 10
Polar molecules 72
Polarisation 71, 376
  by reflection 380
  by scattering 379
Polarity of charge 2
Polaroid 378
Potential 53
  due to an electric dipole 55
  due to a point charge 54
  due to a system of charges 57
  energy difference 53
  energy for a system of charges 61
  energy of a dipole 66
  energy of a single charge 64
  energy of a system of two charges 65
  energy 52
Potentiometer 122
Power (electrical) 106
  factor 252
  in ac circuit 252
  of lens 327
Pressurised heavy water reactors 453
Primary coil
Principal focus 311
Principle of superposition 15
Principle quantum number 425
Prism formula 331
Production of amplitude modulated wave 525
Properties of electric charge 8
p-type semi conductor 476
Q factor/quality factor 250
Quanta of energy 393
Quantisation of charge 8
Radio waves 281
Radioactivity 446
Rainbow 333
Ray optics, validity of 375
Rayleigh scattering 335
Rectifier 483
Red shift 358
Reflection of light 310
Refraction 318
  of a plane wave 355
Refractive index 317, 356
Relation between field and potential 61
Relaxation time 98
Rententivity 195
Repeater 517
Resistance 95
Resistivity 96, 468
  of some materials 102
Resolving power 373
  of eye 374
Resonance 248
  Sharpness 249
Resonant frequency 248
Reverse bias 480

Right hand rule 149
Root mean square (rms) or effective
current 235
voltage 235
Roget's spiral 156
Rutherford's model of atom 415
Saturation current 390
Scattering of light 335
Secondary wavelet 354
Semiconductors 469
diode 479
elemental 468
compound 468
Shunt resistance 164
Signal 516
Sky wave 520
Snell's law 317, 356
Solar cell 489
Solenoid 151
South pole 174
Space wave 521
Spectral series 421
Brackett 422
Fund 422
Lyman 422
Paschen 422
Spectrum of light 332
Spherical mirror 310, 311
Spin magnetic moment 163
Surface charge density 32
Telescope 341
Temperature dependence of
resistivity 103
Tesla 135
Thermionic emission 388
Thermonuclear fusion 456
Thin lens formula 326
Threshold frequency 392
Tokamak 153
Toroid 152
Torque
on a current loop 157
on a dipole 31
Total internal reflection 319
Transducer 516
Transformer 259
Step-down 261
Step-up 261
Transistor
as a switch 496
as an amplifier 497
oscillator 500
common emitter configuration 493
Truth table 502
Uncertainty Principle 400
Unpolarised wave 377
Ultraviolet rays 282
Valence band 469
Van de Graaff Generator 83
Velocity selector 140
Visible rays 282
Voltage Regulator 486
Voltage sensitivity of a galvanometer 165
Voltmeter 165
Volume charge density 32
Wattless current 252
Wavefront 353
plane 354
spherical 354
Wheatstone bridge 118
Work function 394
X rays 283
Young's experiment 362
Zener
diode 485
breakdown 485

ANSWERS

CHAPTER 9

9.1 $v = -54$ cm. The image is real, inverted and magnified. The size of the image is 5.0 cm. As $u \to f$, $v \to \infty$; for $u < f$, image is virtual.

9.2 $v = 6.7$ cm. Magnification = 5/9, i.e., the size of the image is 2.5 cm. As $u \to \infty$; $v \to f$ (but never beyond) while $m \to 0$.

9.3 1.33; 1.7 cm

9.4 $n_{ga} = 1.51$; $n_{wa} = 1.32$; $n_{gw} = 1.144$; which gives $\sin r = 0.6181$ i.e., $r \simeq 38°$.

9.5 $r = 0.8 \times \tan i_c$ and $\sin i_c = 1/1.33 \cong 0.75$, where r is the radius (in m) of the largest circle from which light comes out and i_c is the critical angle for water-air interface, Area = 2.6 m^2

9.6 $n \cong 1.53$ and D_m for prism in water $\cong 10°$

9.7 $R = 22$ cm

9.8 Here the object is virtual and the image is real. $u = +12$ cm (object on right; virtual)

(a) $f = +20$ cm. Image is real and at 7.5 cm from the lens on its right side.

(b) $f = -16$ cm. Image is real and at 48 cm from the lens on its right side.

9.9 $v = 8.4$ cm, image is erect and virtual. It is diminished to a size 1.8 cm. As $u \to \infty$, $v \to f$ (but never beyond f while $m \to 0$).

Note that when the object is placed at the focus of the concave lens (21 cm), the image is located at 10.5 cm (not at infinity as one might wrongly think).

9.10 A diverging lens of focal length 60 cm

9.11 (a) $v_e = -25$ cm and $f_e = 6.25$ cm give $u_e = -5$ cm; $v_O = (15 - 5)$ cm = 10 cm,

$f_O = u_O = -2.5$ cm; Magnifying power = 20

(b) $u_O = -2.59$ cm.

Magnifying power = 13.5.

9.12 Angular magnification of the eye-piece for image at 25 cm

$$= \frac{25}{2.5} + 1 = 11; \quad |u_e| = \frac{25}{11}\text{cm} = 2.27\text{cm}; \quad v_O = 7.2\text{cm}$$

Separation = 9.47 cm; Magnifying power = 88

9.13 24; 150 cm

9.14 (a) Angular magnification = 1500

(b) Diameter of the image = 13.7 cm.

9.15 Apply mirror equation and the condition:

(a) $f < 0$ (concave mirror); $u < 0$ (object on left)

(b) $f > 0$; $u < 0$

(c) $f > 0$ (convex mirror) and $u < 0$

(d) $f < 0$ (concave mirror); $f < u < 0$

to deduce the desired result.

9.16 The pin appears raised by 5.0 cm. It can be seen with an explicit ray diagram that the answer is independent of the location of the slab (for small angles of incidence).

9.17 (a) $\sin i'_c = 1.44/1.68$ which gives $i'_c = 59°$. Total internal reflection takes place when $i > 59°$ or when $r < r_{max} = 31°$. Now, $(\sin i_{max}/\sin r_{max}) = 1.68$, which gives $i_{max} \simeq 60°$. Thus, all incident rays of angles in the range $0 < i < 60°$ will suffer total internal reflections in the pipe. (If the length of the pipe is finite, which it is in practice, there will be a lower limit on i determined by the ratio of the diameter to the length of the pipe.)

(b) If there is no outer coating, $i'_c = \sin^{-1}(1/1.68) = 36.5°$. Now, $i = 90°$ will have $r = 36.5°$ and $i' = 53.5°$ which is greater than i'_c. Thus, *all* incident rays (in the range $53.5° < i < 90°$) will suffer total internal reflections.

9.18 (a) Rays converging to a point 'behind' a plane or convex mirror are reflected to a point in front of the mirror on a screen. In other words, a plane or convex mirror can produce a real image if the object is virtual. Convince yourself by drawing an appropriate ray diagram.

(b) When the reflected or refracted rays are divergent, the image is virtual. The divergent rays can be converged on to a screen by means of an appropriate converging lens. The convex lens of the eye does just that. The virtual image here serves as an object for the lens to produce a real image. Note, the screen here is not located at the position of the virtual image. There is no contradiction.

(c) Taller

(d) The apparent depth for oblique viewing decreases from its value for near-normal viewing. Convince yourself of this fact by drawing ray diagrams for different positions of the observer.

(e) Refractive index of a diamond is about 2.42, much larger than that of ordinary glass (about 1.5). The critical angle of diamond is about 24°, much less than that of glass. A skilled diamond-cutter exploits the larger range of angles of incidence (in the diamond), 24° to 90°, to ensure that light entering the diamond is totally reflected from many faces before getting out–thus producing a sparkling effect.

9.19 For fixed distance s between object and screen, the lens equation does not give a real solution for u or v if f is greater than $s/4$.

Therefore, $f_{max} = 0.75$ m.

9.20 21.4 cm

9.21 (a) (i) Let a parallel beam be the incident from the left on the convex lens first.

f_1 = 30 cm and $u_1 = -\infty$, give v_1 = + 30 cm. This image becomes a virtual object for the second lens.

f_2 = –20 cm, u_2 = + (30 – 8) cm = + 22 cm which gives, v_2 = – 220 cm. The parallel incident beam appears to diverge from a point 216 cm from the centre of the two-lens system.

(ii) Let the parallel beam be incident from the left on the concave lens first: f_1 = – 20 cm, $u_1 = -\infty$, give v_1 = – 20 cm. This image becomes a real object for the second lens: f_2 = + 30 cm, u_2 = – (20 + 8) cm = – 28 cm which gives, v_2 = – 420 cm. The parallel incident beam appears to diverge from a point 416 cm on the left of the centre of the two-lens system.

Clearly, the answer depends on which side of the lens system the parallel beam is incident. *Further we do not have a simple lens equation true for all u (and v) in terms of a definite constant of the system (the constant being determined by f_1 and f_2, and the separation between the lenses).* The notion of effective focal length, therefore, does not seem to be meaningful for this system.

(b) u_1 = – 40 cm, f_1 = 30 cm, gives v_1= 120 cm.

Magnitude of magnification due to the first (convex) lens is 3. u_2 = + (120 – 8) cm = +112 cm (object virtual);

f_2 = – 20 cm which gives $v_2 = -\dfrac{112 \times 20}{92}$ cm

Magnitude of magnification due to the second (concave) lens = 20/92.

Net magnitude of magnification = 0.652

Size of the image = 0.98 cm

9.22 If the refracted ray in the prism is incident on the second face at the critical angle i_c, the angle of refraction r at the first face is (60°–i_c).

Now, $i_c = \sin^{-1}(1/1.524) \simeq 41°$

Therefore, r = 19°

$\sin i$ = 0.4962; $i \simeq 30°$

9.23 Two identical prisms made of the same glass placed with their bases on opposite sides (of the incident white light) and faces touching (or parallel) will neither deviate nor disperse, but will mearly produce a parallel displacement of the beam.

(a) To deviate without dispersion, choose, say, the first prism to be of crown glass, and take for the second prism a flint glass prism of suitably chosen refracting angle (smaller than that of crown glass prism because the flint glass prism disperses more) so that dispersion due to the first is nullified by the second.

(b) To disperse without deviation, increase the angle of flint glass prism (i.e., try flint glass prisms of greater and greater angle) so that deviations due to the two prisms are equal and opposite. (The flint glass prism angle will still be smaller than that of crown glass because flint glass has higher refractive index than that of crown glass). Because of the adjustments involved for so many colours, these are not meant to be precise arrangements for the purpose required.

9.24 To see objects at infinity, the eye uses its least converging power = (40+20) dioptres = 60 dioptres. This gives a rough idea of the distance between the retina and cornea-eye lens: (5/3) cm. To focus an object at the near point (u = –25 cm), on the retina (v = 5/3 cm), the focal length should be

$$\left[\frac{1}{25}+\frac{3}{5}\right]^{-1}=\frac{25}{16}\text{cm}$$

corresponding to a converging power of 64 dioptres. The power of the eye lens then is (64 – 40) dioptres = 24 dioptres. The range of accommodation of the eye-lens is roughly 20 to 24 dioptres.

9.25 No, a person may have normal ability of accommodation of the eye-lens and yet may be myopic or hypermetropic. Myopia arises when the eye-ball from front to back gets too elongated; hypermetropia arises when it gets too shortened. In practice, in addition, the eye lens may also lose some of its ability of accommodation. When the eyeball has the normal length but the eye lens loses partially its ability of accommodation (as happens with increasing age for any normal eye), the 'defect' is called presbyopia and is corrected in the same manner as hypermetropia.

9.26 The far point of the person is 100 cm, while his near point may have been normal (about 25 cm). Objects at infinity produce virtual image at 100 cm (using spectacles). To view closer objects i.e., those which are (or whose images using the spectacles are) between 100 cm and 25 cm, the person uses the ability of accommodation of his eye-lens. This ability usually gets partially lost in old age (presbyopia). The near point of the person recedes to 50 cm. To view objects at 25 cm clearly, the person needs converging lens of power +2 dioptres.

9.27 The defect (called astigmatism) arises because the curvature of the cornea plus eye-lens refracting system is not the same in different planes. [The eye-lens is usually spherical i.e., has the same curvature on different planes but the cornea is not spherical in case of an astigmatic eye.] In the present case, the curvature in the vertical plane is enough, so sharp images of vertical lines can be formed on the retina. But the curvature is insufficient in the horizontal plane, so horizontal lines appear blurred. The defect can be corrected by using a cylindrical lens with its axis along the vertical. Clearly, parallel rays in the vertical plane will suffer no extra refraction, but those in the horizontal plane can get the required extra convergence due to refraction by the curved surface of the cylindrical lens if the curvature of the cylindrical surface is chosen appropriately.

9.28 (a) Closest distance = $4\frac{1}{6}\text{cm} \approx 4.2\,\text{cm}$

Farthest distance = 5 cm

(b) Maximum angular magnification = [25/(25/6)] = 6. Minimum angular magnification = (25/5) = 5

9.29 (a) $\frac{1}{v}+\frac{1}{9}=\frac{1}{10}$

i.e., v = – 90 cm,

Magnitude of magnification = 90/9 = 10.

Each square in the virtual image has an area $10 \times 10 \times 1\ mm^2 = 100\ mm^2 = 1\ cm^2$

(b) Magnifying power = 25/9 = 2.8

(c) No, magnification of an image by a lens and angular magnification (or magnifying power) of an optical instrument are two separate things. The latter is the ratio of the angular size of the object (which is equal to the angular size of the image even if the image is magnified) to the angular size of the object if placed at the near point (25 cm). Thus, magnification magnitude is $|(v/u)|$ and magnifying power is $(25/|u|)$. Only when the image is located at the near point $|v| = 25$ cm, are the two quantities equal.

9.30 (a) Maximum magnifying power is obtained when the image is at the near point (25 cm)

$u = -7.14$ cm.

(b) Magnitude of magnification = $(25/|u|) = 3.5$.

(c) Magnifying power = 3.5

Yes, the magnifying power (when the image is produced at 25 cm) is equal to the magnitude of magnification.

9.31 Magnification $= \sqrt{(6.25/1)} = 2.5$

$v = +2.5u$

$$+\frac{1}{2.5u} - \frac{1}{u} = \frac{1}{10}$$

i.e., $u = -6$ cm

$|v| = 15$ cm

The virtual image is closer than the normal near point (25 cm) and cannot be seen by the eye distinctly.

9.32 (a) Even though the absolute image size is bigger than the object size, the angular size of the image is equal to the angular size of the object. The magnifier helps in the following way: without it object would be placed no closer than 25 cm; with it the object can be placed much closer. The closer object has larger angular size than the same object at 25 cm. It is in this sense that angular magnification is achieved.

(b) Yes, it decreases a little because the angle subtended at the eye is then slightly less than the angle subtended at the lens. The effect is negligible if the image is at a very large distance away. [*Note:* When the eye is separated from the lens, the angles subtended at the eye by the first object and its image are not equal.]

(c) First, grinding lens of very small focal length is not easy. More important, if you decrease focal length, aberrations (both spherical and chromatic) become more pronounced. So, in practice, you cannot get a magnifying power of more than 3 or so with a simple convex lens. However, using an aberration corrected lens system, one can increase this limit by a factor of 10 or so.

(d) Angular magnification of eye-piece is $[(25/f_e) + 1]$ (f_e in cm) which increases if f_e is smaller. Further, magnification of the objective is given by $\dfrac{v_O}{|u_O|} = \dfrac{1}{(|u_O|/f_O) - 1}$

which is large when $|u_O|$ is slightly greater than f_O. The microscope is used for viewing very close object. So $|u_O|$ is small, and so is f_O.

(e) The image of the objective in the eye-piece is known as 'eye-ring'. All the rays from the object refracted by objective go through the eye-ring. Therefore, it is an ideal position for our eyes for viewing. If we place our eyes too close to the eye-piece, we shall not collect much of the light and also reduce our field of view. If we position our eyes on the eye-ring and the area of the pupil of our eye is greater or equal to the area of the eye-ring, our eyes will collect all the light refracted by the objective. The precise location of the eye-ring naturally depends on the separation between the objective and the eye-piece. When you view through a microscope by placing your eyes on one end,the ideal distance between the eyes and eye-piece is usually built-in the design of the instrument.

9.33 Assume microscope in normal use i.e., image at 25 cm. Angular magnification of the eye-piece

$$= \frac{25}{5} + 1 = 6$$

Magnification of the objective

$$= \frac{30}{6} = 5$$

$$\frac{1}{5u_O} - \frac{1}{u_O} = \frac{1}{1.25}$$

which gives $u_O = -1.5$ cm; $v_0 = 7.5$ cm. $|u_e| = (25/6)$ cm = 4.17 cm. The separation between the objective and the eye-piece should be (7.5 + 4.17) cm = 11.67 cm. Further the object should be placed 1.5 cm from the objective to obtain the desired magnification.

9.34 (a) $m = (f_O/f_e) = 28$

(b) $m = \frac{f_O}{f_e}\left[1 + \frac{f_O}{25}\right] = 33.6$

9.35 (a) $f_O + f_e = 145$ cm

(b) Angle subtended by the tower = (100/3000) = (1/30) rad.

Angle subtended by the image produced by the objective

$$= \frac{h}{f_O} = \frac{h}{140}$$

Equating the two, $h = 4.7$ cm.

(c) Magnification (magnitude) of the eye-piece = 6. Height of the final image (magnitude) = 28 cm.

9.36 The image formed by the larger (concave) mirror acts as virtual object for the smaller (convex) mirror. Parallel rays coming from the object at infinity will focus at a distance of 110 mm from the larger mirror. The distance of virtual object for the smaller mirror = (110 −20) = 90 mm. The focal length of smaller mirror is 70 mm. Using the mirror formula, image is formed at 315 mm from the smaller mirror.

9.37 The reflected rays get deflected by twice the angle of rotation of the mirror. Therefore, $d/1.5 = \tan 7°$. Hence $d = 18.4$ cm.

9.38 $n = 1.33$

Chapter 10

10.1 (a) Reflected light: (wavelength, frequency, speed same as incident light)

$\lambda = 589$ nm, $\nu = 5.09 \times 10^{14}$ Hz, $c = 3.00 \times 10^{8}$ m s^{-1}

(b) Refracted light: (frequency same as the incident frequency) $\nu = 5.09 \times 10^{14}$ Hz
$v = (c/n) = 2.26 \times 10^{8}$ m s^{-1}, $\lambda = (v/\nu) = 444$ nm

10.2 (a) Spherical
(b) Plane
(c) Plane (a small area on the surface of a large sphere is nearly planar).

10.3 (a) 2.0×10^{8} m s^{-1}
(b) No. The refractive index, and hence the speed of light in a medium, depends on wavelength. [When no particular wavelength or colour of light is specified, we may take the given refractive index to refer to yellow colour.] Now we know violet colour deviates more than red in a glass prism, i.e. $n_v > n_r$. Therefore, the violet component of white light travels slower than the red component.

10.4 $\lambda = \dfrac{1.2 \times 10^{-2} \times 0.28 \times 10^{-3}}{4 \times 1.4}$ m = 600 nm

10.5 K/4

10.6 (a) 1.17 mm (b) 1.56 mm

10.7 0.15°

10.8 $\tan^{-1}(1.5) \simeq 56.3°$

10.9 5000 Å, 6×10^{14} Hz; 45°

10.10 40 m

10.11 Use the formula $\lambda' - \lambda = \dfrac{v}{c}\lambda$

i.e., $v = \dfrac{c}{\lambda}(\lambda' - \lambda) = \dfrac{3 \times 10^{8} \times 15}{6563} = 6.86 \times 10^{5}$ m s^{-1}

10.12 In corpuscular (particle) picture of refraction, particles of light incident from a rarer to a denser medium experience a force of attraction normal to the surface. This results in an increase in the normal component of the velocity but the component along the surface is unchanged. This means

$c \sin i = v \sin r$ or $\dfrac{v}{c} = \dfrac{\sin i}{\sin r} = n$. Since $n > 1$, $v > c$.

The prediction is *opposite* to the experimental results ($v < c$). The wave picture of light is consistent with the experiment.

10.13 With the point object at the centre, draw a circle touching the mirror. This is a plane section of the spherical wavefront from the object that has just reached the mirror. Next draw the locations of this same wavefront after a time t in the presence of the mirror, and in the absence of the mirror. You will get two arcs symmetrically located on either side of the mirror. Using simple geometry, the centre of the reflected wavefront (the image of the object) is seen to be at the same distance from the mirror as the object.

10.14 (a) The speed of light in vacuum is a universal constant independent of all the factors listed and anything else. In particular, note the surprising fact that it is independent of the relative motion between the source and the observer. This fact is a basic axiom of Einstein's special theory of relativity.

(b) Dependence of the speed of light in a medium:

(i) does not depend on the nature of the source (wave speed is determined by the properties of the medium of propagation. This is also true for other waves, e.g., sound waves, water waves, etc.).

(ii) independent of the direction of propagation for *isotropic* media.

(iii) independent of the motion of the source relative to the medium but depends on the motion of the observer relative to the medium.

(iv) depends on wavelength.

(v) independent of intensity. [For high intensity beams, however, the situation is more complicated and need not concern us here.]

10.15 Sound waves require a medium for propagation. Thus even though the situations (i) and (ii) may correspond to the same relative motion (between the source and the observer), they are not identical physically since the motion of the observer *relative to the medium* is different in the two situations. Therefore, we cannot expect Doppler formulas for sound to be identical for (i) and (ii). For light waves in vacuum, there is clearly nothing to distinguish between (i) and (ii). Here only the relative motion between the source and the observer counts and the relativistic Doppler formula is the same for (i) and (ii). For light propagation in a medium, once again like for sound waves, the two situations are *not* identical and we should expect the Doppler formulas for this case to be different for the two situations (i) and (ii).

10.16 3.4×10^{-4} m.

10.17 (a) The size reduces by half according to the relation: size ~ λ/d. Intensity increases four fold.

(b) The intensity of interference fringes in a double-slit arrangement is modulated by the diffraction pattern of each slit.

(c) Waves diffracted from the edge of the circular obstacle interfere constructively at the centre of the shadow producing a bright spot.

(d) For diffraction or bending of waves by obstacles/apertures by a large angle, the size of the latter should be comparable to wavelength. If the size of the obstacle/aperture is much too large compared to wavelength, diffraction is by a small angle. Here

the size is of the order of a few metres. The wavelength of light is about 5×10^{-7} m, while sound waves of, say, 1 kHz frequency have wavelength of about 0.3 m. Thus, sound waves can bend around the partition while light waves cannot.

(e) Justification based on what is explained in (d). Typical sizes of apertures involved in ordinary optical instruments are much larger than the wavelength of light.

10.18 12.5 cm.

10.19 0.2 nm.

10.20 (a) Interference of the direct signal received by the antenna with the (weak) signal reflected by the passing aircraft.

(b) Superposition principle follows from the linear character of the (differential) equation governing wave motion. If y_1 and y_2 are solutions of the wave equation, so is any linear combination of y_1 and y_2. When the amplitudes are large (e.g., high intensity laser beams) and non-linear effects are important, the situation is far more complicated and need not concern us here.

10.21 Divide the single slit into n smaller slits of width $a' = a/n$. The angle $\theta = n\lambda/a = \lambda/a'$. Each of the smaller slits sends zero intensity in the direction θ. The combination gives zero intensity as well.

Chapter 11

11.1 (a) 7.24×10^{18} Hz (b) 0.041 nm

11.2 (a) 0.34 eV = 0.54×10^{-19} J (b) 0.34 V (c) 344 km/s

11.3 1.5 eV = 2.4×10^{-19} J

11.4 (a) 3.14×10^{-19} J, 1.05×10^{-27} kg m/s (b) 3×10^{16} photons/s (c) 0.63 m/s

11.5 4×10^{21} photons/m^2 s

11.6 6.59×10^{-34} J s

11.7 (a) 3.38×10^{-19} J = 2.11 eV (b) 3.0×10^{20} photons/s

11.8 2.0 V

11.9 No, because $\nu < \nu_o$

11.10 4.73×10^{14} Hz

11.11 2.16 eV = 3.46×10^{-19} J

11.12 (a) 4.04×10^{-24} kg m s^{-1} (b) 0.164 nm

11.13 (a) 5.92×10^{-24} kg m s^{-1} (b) 6.50×10^{6} m s^{-1} (c) 0.112 nm

11.14 (a) 6.95×10^{-25} J = 4.34 μeV (b) 3.78×10^{-28} J = 0.236 neV

11.15 (a) 1.7×10^{-35} m (b) 1.1×10^{-32} m (c) 3.0×10^{-23} m

11.16 (a) 6.63×10^{-25} kg m/s (for both) (b) 1.24 keV (c) 1.51 eV

11.17 (a) 6.686×10^{-21} J = 4.174×10^{-2} eV (b) 0.145 nm

11.18 $\lambda = h/p = h/(h\nu/c) = c/\nu$

11.19 0.028 nm

11.20 (a) Use $eV = (m\,v^2/2)$ i.e., $v = [(2eV/m)]^{1/2}$; $v = 1.33 \times 10^7$ m s^{-1}.

(b) If we use the same formula with $V = 10^7$ V, we get $v = 1.88 \times 10^9$ m s^{-1}. This is clearly wrong, since nothing can move with a speed greater than the speed of light ($c = 3 \times 10^8$ m s^{-1}). Actually, the above formula for kinetic energy ($m\,v^2/2$) is valid only when

$(v/c) << 1$. At very high speeds when (v/c) is comparable to (though always less than) 1, we come to the relativistic domain where the following formulae are valid:

Relativistic momentum $p = m\,v$

Total energy $E = m\,c^2$

Kinetic energy $K = m\,c^2 - m_0\,c^2$,

where the relativistic mass m is given by $m = m_0\left(1-\frac{v^2}{c^2}\right)^{-1/2}$

m_0 is called the rest mass of the particle. These relations also imply:

$E = (p^2c^2 + m_0^2\,c^4)^{1/2}$

Note that in the relativisitc domain when v/c is comparable to 1, K or energy $\geq m_0c^2$ (rest mass energy). The rest mass energy of electron is about 0.51 MeV. Thus a kinetic energy of 10 MeV, being much greater than electron's rest mass energy, implies relativistic domain. Using relativistic formulas, v (for 10 MeV kinetic energy) = 0.999 c.

11.21 (a) 22.7 cm

(b) No. As explained above, a 20 MeV electron moves at relativistic speed. Consequently, the non-relativistic formula $R = (m_0v/e\,B)$ is not valid. The relativistic formula is

$$R = p/eB = mv/eB \quad \text{or} \quad R = m_0v/\left(eB\sqrt{1-v^2/c^2}\right)$$

11.22 We have $e\,V = (m\,v^2/2)$ and $R = (m\,v/e\,B)$ which gives $(e/m) = (2V/R^2\,B^2)$; using the given data $(e/m) = 1.73 \times 10^{11}$ C kg^{-1}.

11.23 (a) 27.6 keV (b) of the order of 30 kV

11.24 Use $\lambda = (hc/E)$ with $E = 5.1 \times 1.602 \times 10^{-10}$J to get $\lambda = 2.43 \times 10^{-16}$ m .

11.25 (a) For $\lambda = 500$ m, $E = (h\,c/\,\lambda) = 3.98 \times 10^{-28}$J. Number of photons emitted per second

$= 10^4$J s^{-1}/3.98×10^{-28}J $\simeq 3 \times 10^{31}$ s^{-1}

We see that the energy of a radiophoton is exceedingly small, and the number of photons emitted per second in a radio beam is enormously large. There is, therefore, negligible error involved in ignoring the existence of a minimum quantum of energy (photon) and treating the total energy of a radio wave as continuous.

(b) For $v = 6 \times 10^{14}$ Hz, $E \simeq 4 \times 10^{-19}$J. Photon flux corresponding to minimum intensity

$= 10^{-10}$ W m^{-2}/4×10^{-19}J $= 2.5 \times 10^8$ m^{-2} s^{-1}

Number of photons entering the pupil per second $= 2.5 \times 10^8 \times 0.4 \times 10^{-4}$ s^{-1} $= 10^4$ s^{-1}. Though this number is not as large as in (a) above, it is large enough for us never to 'sense' or 'count' individual photons by our eye.

11.26 $\phi_0 = h\,v - e\,V_0 = 6.7 \times 10^{-19}$J $= 4.2$ eV; $v_0 = \frac{\phi_0}{h} = 1.0 \times 10^{15}$ Hz; $\lambda = 6328$Å

corresponds to $v = 4.7 \times 10^{14}$ Hz $< v_0$. The photo-cell will not respond howsoever high be the intensity of laser light.

11.27 Use $e\,V_0 = h\,v - \phi_0$ for both sources. From the data on the first source, $\phi_0 = 1.40$ eV. Use this value to obtain for the second source $V_0 = 1.50$ V.

11.28 Obtain V_0 versus ν plot. The slope of the plot is (h/e) and its intercept on the ν-axis is ν_0. The first four points lie nearly on a straight line which intercepts the ν-axis at $\nu_0 = 5.0 \times 10^{14}$ Hz (threshold frequency). The fifth point corresponds to $\nu < \nu_0$; there is no photoelectric emission and therefore no stopping voltage is required to stop the current. Slope of the plot is found to be 4.15×10^{-15} V s. Using $e = 1.6 \times 10^{-19}$ C, $h = 6.64 \times 10^{-34}$ J s (standard value $h = 6.626 \times 10^{-34}$ J s), $\phi_0 = h\,\nu_0 = 2.11$ V.

11.29 It is found that the given incident frequency ν is greater than ν_0 (Na), and ν_0 (K); but less than ν_0 (Mo), and ν_0 (Ni). Therefore, Mo and Ni will not give photoelectric emission. If the laser is brought closer, intensity of radiation increases, but this does not affect the result regarding Mo and Ni. However, photoelectric current from Na and K will increase in proportion to intensity.

11.30 Assume one conduction electron per atom. Effective atomic area $\sim 10^{-20}\,\text{m}^2$

Number of electrons in 5 layers

$$= \frac{5 \times 2 \times 10^{-4}\,\text{m}^2}{10^{-20}\,\text{m}^2} = 10^{17}$$

Incident power

$= 10^{-5}\,\text{W m}^{-2} \times 2 \times 10^{-4}\,\text{m}^2 = 2 \times 10^{-9}\,\text{W}$

In the wave picture, incident power is uniformly absorbed by all the electrons continuously. Consequently, energy absorbed per second per electron

$= 2 \times 10^{-9}/10^{17} = 2 \times 10^{-26}\,\text{W}$

Time required for photoelectric emission

$= 2 \times 1.6 \times 10^{-19}\,\text{J}/2 \times 10^{-26}\,\text{W} = 1.6 \times 10^{7}\,\text{s}$

which is about 0.5 year.

Implication: Experimentally, photoelectric emission is observed nearly instantaneously ($\sim 10^{-9}$ s): Thus, the wave picture is in gross disagreement with experiment. In the photon-picture, energy of the radiation is not continuously shared by all the electrons in the top layers. Rather, energy comes in discontinuous 'quanta'. and absorption of energy does not take place gradually. A photon is either not absorbed, or absorbed by an electron nearly instantly.

11.31 For $\lambda = 1$ Å, electron's energy = 150 eV; photon's energy = 12.4 keV. Thus, for the same wavelength, a photon has much greater energy than an electron.

11.32 (a) $\lambda = \dfrac{h}{p} = \dfrac{h}{\sqrt{2\,mK}}$ Thus, for same K, λ decreases with m as $(1/\sqrt{m})$. Now $(m_n/m_e) = 1838.6$; therefore for the same energy, (150 eV) as in Ex. 11.31, wavelength of neutron $= \left(1/\sqrt{1838.6}\right) \times 10^{-10}$ m $= 2.33 \times 10^{-12}$ m. The interatomic spacing is about a hundred times greater. A neutron beam of 150 eV energy is therefore not suitable for diffraction experiments.

(b) $\lambda = 1.45 \times 10^{-10}$ m [Use $\lambda = (h/\sqrt{3\,m\,k\,T})$] which is comparable to interatomic spacing in a crystal.

Clearly, from (a) and (b) above, thermal neutrons are a suitable probe for diffraction experiments; so a high energy neutron beam should be first thermalised before using it for diffraction.

11.33 $\lambda = 5.5 \times 10^{-12}$ m

λ (yellow light) = 5.9×10^{-7}m

Resolving Power (RP) is inversely proportional to wavelength. Thus, RP of an electron microscope is about 10^5 times that of an optical microscope. In practice, differences in other (geometrical) factors can change this comparison somewhat.

11.34 $p = \dfrac{h}{\lambda} = \dfrac{6.63\times10^{-34}\text{Js}}{10^{-15}\text{m}} = 6.63 \times 10^{-19}\ \text{kg m s}^{-1}$

Use the relativistic formula for energy:

$E^2 = c^2p^2 + m_0^2\, c^4 = 9 \times (6.63)^2 \times 10^{-22} + (0.511 \times 1.6)^2 \times 10^{-26}$

$\simeq 9 \times (6.63)^2 \times 10^{-22}$,

the second term (rest mass energy) being negligible.

Therefore, $E = 1.989 \times 10^{-10}$ J = 1.24 BeV. Thus, electron energies from the accelerator must have been of the order of a few BeV.

11.35 Use $\lambda = \dfrac{h}{\sqrt{3\ m\ k\ T}}$; $m_{He} = \dfrac{4\times10^{\ 3}}{6\times10^{23}}$ kg

This gives $\lambda = 0.73 \times 10^{-10}$ m. Mean separation

$r = (V/N)^{1/3} = (kT/p)^{1/3}$

For $T = 300$ K, $p = 1.01 \times 10^5$ Pa, $r = 3.4 \times 10^{-9}$ m. We find $r >> \lambda$.

11.36 Using the same formula as in Exercise 11.35, $\lambda = 6.2 \times 10^{-9}$ m which is much greater than the given inter-electron separation.

11.37 (a) Quarks are thought to be confined within a proton or neutron by forces which grow stronger if one tries to pull them apart. It, therefore, seems that though fractional charges may exist in nature, observable charges are still integral multiples of e.

(b) Both the basic relations $e\ V = (1/2)\ m\ v^2$ or $e\ E = m\ a$ and $e\ B\ v = m\ v^2/r$, for electric and magnetic fields, respectively, show that the dynamics of electrons is determined not by e, and m separately but by the combination e/m.

(c) At low pressures, ions have a chance to reach their respective electrodes and constitute a current. At ordinary pressures, ions have no chance to do so because of collisions with gas molecules and recombination.

(d) Work function merely indicates the minimum energy required for the electron in the highest level of the conduction band to get out of the metal. Not all electrons in the metal belong to this level. They occupy a continuous band of levels. Consequently, for the same incident radiation, electrons knocked off from different levels come out with different energies.

(e) The absolute value of energy E (but not momentum p) of any particle is arbitrary to within an additive constant. Hence, while λ is physically significant, absolute value of v of a matter wave of an electron has no direct physical meaning. The phase speed $v\lambda$ is likewise not physically significant. The group speed given by

$$\frac{dv}{d(1/\lambda)} = \frac{dE}{dp} = \frac{d}{dp}\left(\frac{p^2}{2m}\right) = \frac{p}{m}$$

is physically meaningful.

Chapter 12

12.1 (a) No different from

(b) Thomson's model; Rutherford's model

(c) Rutherford's model

(d) Thomson's model; Rutherford's model

(e) Both the models

12.2 The nucleus of a hydrogen atom is a proton. The mass of it is 1.67×10^{-27} kg, whereas the mass of an incident α-particle is 6.64×10^{-27} kg. Because the scattering particle is more massive than the target nuclei (proton), the α-particle won't bounce back in even in a head-on collision. It is similar to a football colliding with a tenis ball at rest. Thus, there would be no large-angle scattering.

12.3 820 nm.

12.4 5.6×10^{14} Hz

12.5 13.6 eV; –27.2 eV

12.6 9.7×10^{-8} m; 3.1×10^{15} Hz.

12.7 (a) 2.18×10^{6} m/s; 1.09×10^{6} m/s; 7.27×10^{5} m/s

(b) 1.52×10^{-16} s; 1.22×10^{-15} s; 4.11×10^{-15} s.

12.8 2.12×10^{-10} m; 4.77×10^{-10} m

12.9 Lyman series: 103 nm and 122 nm; Balmer series: 656 nm.

12.10 2.6×10^{74}

12.11 (a) About the same.

(b) Much less.

(c) It suggests that the scattering is predominantly due to a single collision, because the chance of a single collision increases linearly with the number of target atoms, and hence linearly with thickness.

(d) In Thomson's model, a single collision causes very little deflection. The observed average scattering angle can be explained only by considering multiple scattering. So it is wrong to ignore multiple scattering in Thomson's model. In Rutherford's model, most of the scattering comes through a single collision and multiple scattering effects can be ignored as a first approximation.

12.12 The first orbit Bohr's model has a radius a_0 given by $a_0 = \dfrac{4\pi\varepsilon_0 (h/2\pi)^2}{m_e e^2}$. If we consider the atom bound by the gravitational force ($Gm_p m_e/r^2$), we should replace ($e^2/4\pi\varepsilon_0$) by $Gm_p m_e$. That is, the radius of the first Bohr orbit is given by $a_0^G = \dfrac{(h/2\pi)^2}{Gm_p m_e^2} \cong 1.2 \times 10^{29}$ m. This is much greater than the estimated size of the whole universe!

12.13 $$\nu = \frac{me^4}{(4\pi)^3 \varepsilon_0^2 (h/2\pi)^3}\left[\frac{1}{(n-1)^2} - \frac{1}{n^2}\right] = \frac{me^4(2n-1)}{(4\pi)^3 \varepsilon_0^2 (h/2\pi)^3 n^2 (n-1)^2}$$

For large n, $\nu \cong \dfrac{me^4}{32\pi^3 \varepsilon_0^2 (h/2\pi)^3 n^3}$

Orbital frequency $\nu_c = (\upsilon/2\pi r)$. In Bohr model $\upsilon = \dfrac{n(h/2\pi)}{mr}$, and

$r = \dfrac{4\pi\varepsilon_0(h/2\pi)^2}{me^2}n^2$. This gives $\nu_c = \dfrac{n(h/2\pi)}{2\pi mr^2} = \dfrac{me^4}{32\pi^3\varepsilon_0^2(h/2\pi)^3n^3}$

which is same as ν for large n.

12.14 (a) The quantity $\left(\dfrac{e^2}{4\pi\varepsilon_0 mc^2}\right)$ has the dimensions of length. Its value is 2.82×10^{-15}m – much smaller than the typical atomic size.

(b) The quantity $\dfrac{4\pi\varepsilon_0(h/2\pi)^2}{me^2}$ has the dimensions of length. Its value is 0.53×10^{-10}m – of the order of atomic sizes.(Note that the dimensional arguments cannot, of course, tell us that we should use 4π and $h/2\pi$ in place of h to arrive at the right size.)

12.15 In Bohr's model, $mvr = nh$ and $\dfrac{mv^2}{r} = \dfrac{Ze^2}{4\pi\varepsilon_0 r^2}$

which give

$$T = \frac{1}{2}mv^2 = \frac{Ze^2}{8\pi\varepsilon_0 r}\ ;\ r = \frac{4\pi\varepsilon_0\hbar^2}{Ze^2 m}n^2$$

These relations have nothing to do with the choice of the zero of potential energy. Now, choosing the zero of potential energy at infinity we have $V = -(Ze^2/4\pi\varepsilon_0 r)$ which gives $V = -2T$ and $E = T + V = -T$

(a) The quoted value of $E = -3.4$ eV is based on the customary choice of zero of potential energy at infinity. Using $E = -T$, the kinetic energy of the electron in this state is + 3.4 eV.

(b) Using $V = -2T$, potential energy of the electron is = – 6.8 eV

(c) If the zero of potential energy is chosen differently, kinetic energy does not change. Its value is + 3.4 eV independent of the choice of the zero of potential energy. The potential energy, and the total energy of the state, however, would alter if a different zero of the potential energy is chosen.

12.16 Angular momenta associated with planetary motion are incomparably large relative to h. For example, angular momentum of the earth in its orbital motion is of the order of $10^{70}h$. In terms of the Bohr's quantisation postulate, this corresponds to a very large value of n (of the order of 10^{70}). For such large values of n, the differences in the successive energies and angular momenta of the quantised levels of the Bohr model are so small compared to the energies and angular momenta respectively for the levels that one can, for all practical purposes, consider the levels continuous.

12.17 All that is needed is to replace m_e by m_μ in the formulas of the Bohr model. We note that keeping other factors fixed, $r \propto (1/m)$ and $E \propto m$. Therefore,

$$r_m = \frac{r_e m_e}{m_m} = \frac{0.53\times10^{13}}{207} = 2.56 \times 10^{-13}\text{ m}$$

$$E_\mu = \frac{E_e m_m}{m_e} = -(13.6 \times 207)\text{ eV} \cong -2.8\text{ keV}$$

Chapter 13

13.1 (a) 6.941 u (b) 19.9%, 80.1%

13.2 20.18 u

13.3 104.7 MeV

13.4 8.79 MeV, 7.84 MeV

13.5 1.584×10^{25} MeV or 2.535×10^{12} J

13.6 i) ${}^{226}_{88}\text{Ra} \rightarrow {}^{222}_{86}\text{Rn} + {}^{4}_{2}\text{He}$ ii) ${}^{242}_{94}\text{Pu} \rightarrow {}^{238}_{92}\text{U} + {}^{4}_{2}\text{He}$

iii) ${}^{32}_{15}\text{P} \rightarrow {}^{32}_{16}\text{S} + e^- + \bar{\nu}$ iv) ${}^{210}_{83}\text{B} \rightarrow {}^{210}_{84}\text{Po} + e^- + \bar{\nu}$

v) ${}^{11}_{6}\text{C} \rightarrow {}^{11}_{5}\text{B} + e^+ + \nu$ vi) ${}^{97}_{43}\text{Tc} \rightarrow {}^{97}_{42}\text{Mo} + e^+ + \nu$

vii) ${}^{120}_{54}\text{Xe} + e^+ \rightarrow {}^{120}_{53}\text{I} + \nu$

13.7 (a) 5 T years (b) 6.65 T years

13.8 4224 years

13.9 7.126×10^{-6} g

13.10 7.877×10^{10} Bq or 2.13 Ci

13.11 1.23

13.12 (a) Q = 4.93 MeV, E_α = 4.85 MeV (b) Q = 6.41 MeV, E_α = 6.29 MeV

13.13 ${}^{11}_{6}\text{C} \rightarrow {}^{11}_{6}\text{B} + e^+ + \nu + Q$

$$Q = \left[m_N\left({}^{11}_{6}\text{C}\right) - m_N\left({}^{11}_{6}\text{B}\right) - m_e\right]c^2,$$

where the masses used are those of nuclei and not of atoms. If we use atomic masses, we have to add $6m_e$ in case of ^{11}C and $5m_e$ in case of ^{11}B. Hence

$$Q = \left[m\left({}^{11}_{6}\text{C}\right) - m\left({}^{11}_{6}\text{B}\right) - 2m_e\right]c^2 \quad \text{(Note } m_e \text{ has been doubled)}$$

Using given masses, Q = 0.961 MeV.

$Q = E_d + E_e + E_\nu$

The daughter nucleus is too heavy compared to e^+ and ν, so it carries negligible energy ($E_d \approx 0$). If the kinetic energy (E_ν) carried by the neutrino is minimum (i.e., zero), the positron carries maximum energy, and this is practically all energy Q; hence maximum $E_e \approx Q$).

13.14 ${}^{23}_{10}\text{Ne} \rightarrow {}^{23}_{11}\text{Na} + e^- + \bar{\nu} + Q$; $Q = \left[m_N\left({}^{23}_{10}\text{Ne}\right) - m_N\left({}^{23}_{11}\text{Na}\right) - m_e\right]c^2$, where the masses used are masses of nuclei and not of atoms as in Exercise 13.13. Using atomic masses $Q = \left[m\left({}^{23}_{10}\text{Ne}\right) - m\left({}^{23}_{11}\text{Na}\right)\right]c^2$. Note m_e has been cancelled. Using given masses, Q = 4.37 MeV. As in Exercise 13.13, maximum kinetic energy of the electron (max E_e) = Q = 4.37 MeV.

13.15 (i) Q = –4.03 MeV; endothermic

(ii) Q = 4.62 MeV; exothermic

13.16 $Q = m\left({}^{56}_{26}\text{Fe}\right) - 2m\left({}^{28}_{13}\text{Al}\right)$ = 26.90 MeV; not possible.

13.17 4.536×10^{26} MeV

13.18 Energy generated per gram of ${}^{235}_{92}\text{U}$ = $\dfrac{6\times10^{23}\times200\times1.6\times10^{-13}}{235}\,\text{J g}^{-1}$

The amount of $^{235}_{92}U$ consumed in 5y with 80% on-time

$$= \frac{5 \times 0.8 \times 3.154 \times 10^{16} \times 235}{1.2 \times 1.6 \times 10^{13}} \text{ g} \quad = 1544 \text{ kg}$$

The initial amount of $^{235}_{92}U$ = 3088 kg.

13.19 About 4.9×10^4 y

13.20 360 KeV

13.22 Consider the competing processes:

$^{A}_{Z}X \rightarrow {}^{A}_{Z-1}Y + e^{+} + \nu_e + Q_1$ (positron capture)

$e^{-} + {}^{A}_{Z}X \rightarrow {}^{A}_{Z-1}Y + \nu_e + Q_2$ (electron capture)

$$Q_1 = \left[m_N\left({}^{A}_{Z}X\right) - m_N\left({}^{A}_{Z-1}Y\right) - m_e\right]c^2$$

$$= \left[m_N\left({}^{A}_{Z}X\right) - Z\,m_e - m\left({}^{A}_{Z-1}Y\right) - (Z-1)m_e - m_e\right]c^2$$

$$= \left[m\left({}^{A}_{Z}X\right) - m\left({}^{A}_{Z-1}Y\right) - 2m_e\right]c^2$$

$$Q_2 = \left[m_N\left({}^{A}_{Z}X\right) + m_e - m_N\left({}^{A}_{Z-1}Y\right)\right]c^2 = \left[m\left({}^{A}_{Z}X\right) - m\left({}^{A}_{Z-1}Y\right)\right]c^2$$

This means $Q_1 > 0$ implies $Q_2 > 0$ but $Q_2 > 0$ does not necessarily mean $Q_1 > 0$. Hence the result.

13.23 $^{25}_{12}Mg$: 9.3%, $^{26}_{12}Mg$: 11.7%

13.24 Neutron separation energy S_n of a nucleus $^{A}_{Z}X$ is

$$S_n = \left[m_N\left({}^{A-1}_{Z}X\right) + m_n - m_N\left({}^{A}_{Z}X\right)\right]c^2$$

From given data , $S_n(^{41}_{20}Ca) = 8.36\text{MeV}, S_n(^{27}_{13}Al) = 13.06\text{MeV}$

13.25 209 d

13.26 For $^{14}_{6}C$ emission

$$Q = [m_N(^{223}_{88}Ra) - m_N(^{209}_{82}Pb) - m_N(^{14}_{6}C)]c^2$$

$$= [m(^{223}_{88}Ra) - m(^{209}_{82}Pb) - m(^{14}_{6}C)]c^2 = 31.85 \text{ MeV}$$

For $^{4}_{2}He$ emission, $Q = [m(^{223}_{88}Ra) - m(^{219}_{86}Rn) - m(^{4}_{2}He)]c^2 = 5.98\text{MeV}$

13.27 $Q = [m(^{238}_{92}U) + m_n - m(^{140}_{58}Ce) - m(^{99}_{44}Ru)]c^2 = 231.1$ MeV

13.28 (a) $Q = [m(^{2}_{1}H) + m(^{3}_{1}H) - m(^{4}_{2}He) - m_n]c^2 = 17.59 \text{ MeV}$

(b) K.E. required to overcome Coulomb repulsion = 480.0 keV

480.0 KeV = 7.68×10^{-14} J = $3kT$

$$\therefore T = \frac{7.68 \times 10^{-14}}{3 \times 1.381 \times 10^{-23}} \quad (as\ k = 1.381 \times 10^{-23}\ \text{J K}^{-1})$$

= 1.85×10^9 K (required temperature)

13.29 $K_{max}(\beta_1^-) = 0.284 \text{ MeV}$, $K_{max}(\beta_2^-) = 0.960 \text{ MeV}$

$\nu(\gamma_1) = 2.627 \times 10^{20} \text{ Hz}$, $\nu(\gamma_2) = 0.995 \times 10^{20} \text{ Hz}$, $\nu(\gamma_3) = 1.632 \times 10^{20} \text{ Hz}$

13.30 (a) Note that in the interior of Sun, four $^{1}_{1}H$ nuclei combine to form one $^{4}_{2}He$ nucleus releasing about 26 MeV of energy per event.

Energy released in fusion of 1kg of hydrogen = 39×10^{26} MeV

(b) Energy released in fission of 1kg of $^{235}_{92}U = 5.1 \times 10^{26}$ MeV

The energy released in fusion of 1 kg of hydrogen is about 8 times that of the energy released in the fusion of 1 kg of uranium.

13.31 3.076×10^{4} kg

Chapter 14

14.1 (c)

14.2 (d)

14.3 (c)

14.4 (c)

14.5 (c)

14.6 (b), (c)

14.7 (c)

14.8 50 Hz for half-wave, 100 Hz for full-wave

14.9 $v_i = 0.01$ V ; $I_B = 10\ \mu$A

14.10 2V

14.11 No ($h\nu$ has to be greater than E_g).

14.12 $n_e \approx 4.95 \times 10^{22}$; $n_h = 4.75 \times 10^{9}$; n-type since $n_e >> n_h$

For charge neutrality $N_D - N_A = n_e - n_h$; $n_e.n_h = n_i^2$

Solving these equations, $n_e = \frac{1}{2}\left[(N_D - N_A) + \sqrt{(N_D - N_A)^2 + 4n_i^2}\right]$

14.13 About 1×10^{5}

14.14 (a) 0.0629 A, (b) 2.97 A, (c) 0.336 Ω

(d) For both the voltages, the current I will be almost equal to I_0, showing almost infinite dynamic resistance in the reverse bias.

14.16 NOT ;

A	Y
0	1
1	0

14.17 (a) AND (b) OR

14.18 OR gate

14.19 (a) NOT, (b) AND

Chapter 15

15.1 (b) 10 kHz cannot be radiated (antenna size), 1 GHz and 1000 GHz will penetrate.

15.2 (d) Consult Table 15.2

15.3 (c) Decimal system implies continuous set of values

15.4 No. Service area will be $A = \pi\, d_T^2 = \frac{22}{7} \times 162 \times 6.4 \times 10^6 = 3258\,\text{km}^2$.

15.5 $\mu = 0.75 = \frac{A_m}{A_c}$

$A_m = 0.75 \times 12 = 9\text{V}$.

15.6 (a)

(b) $\mu = 0.5$

15.7 Since the AM wave is given by $(A_c + A_m \sin\omega_m t) \cos \omega_c t$, the maximum amplitude is $M_1 = A_c + A_m$ while the minimum amplitude is $M_2 = A_c - A_m$. Hence the modulation index is

$$m = \frac{A_m}{A_c} = \frac{M_1 - M_2}{M_1 + M_2} = \frac{8}{12} = \frac{2}{3}.$$

With $M_2 = 0$, clearly, $m = 1$, irrespective of M_1.

15.8 Let, for simplicity, the received signal be
$A_1 \cos(\omega_c + \omega_m)\, t$
The carrier $A_c \cos \omega_c t$ is available at the receiving station. By multiplying the two signals, we get
$A_1 A_c \cos(\omega_c + \omega_a)\, t \cos \omega_c t$

$$= = \frac{A_1 A_c}{2}\left[\cos\left(2\omega_c + \omega_m\right)t + \cos \omega_m t\right]$$

If this signal is passed through a low-pass filter, we can record

the modulating signal $\frac{A_1 A_c}{2} \cos \omega_m t$.

APPENDICES

APPENDIX A 1

THE GREEK ALPHABET

Alpha	Α	α	Iota	Ι	ι	Rho	Ρ	ρ
Beta	Β	β	Kappa	Κ	κ	Sigma	Σ	σ
Gamma	Γ	γ	Lambda	Λ	λ	Tau	Τ	τ
Delta	Δ	δ	Mu	Μ	μ	Upsilon	Υ	υ
Epsilon	Ε	ε	Nu	Ν	ν	Phi	Φ	ϕ, φ
Zeta	Ζ	ς	Xi	Ξ	ξ	Chi	Χ	χ
Eta	Η	η	Omicron	Ο	ο	Psi	Ψ	ψ
Theta	Θ	θ	Pi	Π	π	Omega	Ω	ω

APPENDIX A 2

COMMON SI PREFIXES AND SYMBOLS FOR MULTIPLES AND SUB-MULTIPLES

Multiple			Sub-Multiple		
Factor	**Prefix**	**Symbol**	**Factor**	**Prefix**	**symbol**
10^{18}	Exa	E	10^{-18}	atto	a
10^{15}	Peta	P	10^{-15}	femto	f
10^{12}	Tera	T	10^{-12}	pico	p
10^{9}	Giga	G	10^{-9}	nano	n
10^{6}	Mega	M	10^{-6}	micro	μ
10^{3}	kilo	k	10^{-3}	milli	m
10^{2}	Hecto	h	10^{-2}	centi	c
10^{1}	Deca	da	10^{-1}	deci	d

APPENDIX A 3

SOME IMPORTANT CONSTANTS

Name	Symbol	Value
Speed of light in vacuum	c	2.9979×10^{8} m s^{-1}
Charge of electron	e	1.602×10^{-19} C
Gravitational constant	G	6.673×10^{-11} N m^{2} kg^{-2}
Planck constant	h	6.626×10^{-34} J s
Boltzmann constant	k	1.381×10^{-23} J K^{-1}
Avogadro number	N_A	6.022×10^{23} mol^{-1}
Universal gas constant	R	8.314 J mol^{-1} K^{-1}
Mass of electron	m_e	9.110×10^{-31} kg
Mass of neutron	m_n	1.675×10^{-27} kg
Mass of proton	m_p	1.673×10^{-27} kg
Electron-charge to mass ratio	e/m_e	1.759×10^{11} C/kg
Faraday constant	F	9.648×10^{4} C/mol
Rydberg constant	R	1.097×10^{7} m^{-1}
Bohr radius	a_0	5.292×10^{-11} m
Stefan-Boltzmann constant	σ	5.670×10^{-8} W m^{-2} K^{-4}
Wien's Constant	b	2.898×10^{-3} m K
Permittivity of free space	ε_0 $1/4\pi\,\varepsilon_0$	8.854×10^{-12} C^{2} N^{-1} m^{-2} 8.987×10^{9} N m^{2} C^{-2}
Permeability of free space	μ_0	$4\pi \times 10^{-7}$ T m A^{-1} $\cong 1.257 \times 10^{-6}$ Wb A^{-1} m^{-1}

OTHER USEFUL CONSTANTS

Name	Symbol	Value
Mechanical equivalent of heat	J	4.186 J cal^{-1}
Standard atmospheric pressure	1 atm	1.013×10^{5} Pa
Absolute zero	0 K	−273.15 °C
Electron volt	1 eV	1.602×10^{-19} J
Unified Atomic mass unit	1 u	1.661×10^{-27} kg
Electron rest energy	mc^2	0.511 MeV
Energy equivalent of 1 u	1 uc^2	931.5 MeV
Volume of ideal gas(0 °C and 1atm)	V	22.4 L mol^{-1}
Acceleration due to gravity (sea level, at equator)	g	9.78049 m s^{-2}

BIBLIOGRAPHY

TEXTBOOKS

For additional reading on the topics covered in this book, you may like to consult one or more of the following books. Some of these books however are more advanced and contain many more topics than this book.

1. **Ordinary Level Physics**, A.F. Abbott, Arnold-Heinemann (1984).
2. **Advanced Level Physics**, M. Nelkon and P. Parker, 6th Edition, Arnold-Heinemann (1987).
3. **Advanced Physics**, Tom Duncan, John Murray (2000).
4. **Fundamentals of Physics**, David Halliday, Robert Resnick and Jearl Walker, 7th Edition John Wily (2004).
5. **University Physics** (Sears and Zemansky's), H.D. Young and R A. Freedman, 11th Edition, Addison—Wesley (2004).
6. **Problems in Elementary Physics**, B. Bukhovtsa, V. Krivchenkov, G. Myakishev and V. Shalnov, MIR Publishers, (1971).
7. **Lectures on Physics** (3 volumes), R.P. Feynman, Addision – Wesley (1965).
8. **Berkeley Physics Course** (5 volumes) McGraw Hill (1965).
   a. Vol. 1 – Mechanics: (Kittel, Knight and Ruderman)
   b. Vol. 2 – Electricity and Magnetism (E.M. Purcell)
   c. Vol. 3 – Waves and Oscillations (Frank S. Crawford)
   d. Vol. 4 – Quantum Physics (Wichmann)
   e. Vol. 5 – Statistical Physics (F. Reif)
9. **Fundamental University Physics**, M. Alonso and E. J. Finn, Addison – Wesley (1967).
10. **College Physics**, R.L. Weber, K.V. Manning, M.W. White and G.A. Weygand, Tata McGraw Hill (1977).
11. **Physics**: **Foundations and Frontiers**, G. Gamow and J.M. Cleveland, Tata McGraw Hill (1978).
12. **Physics for the Inquiring Mind**, E.M. Rogers, Princeton University Press (1960).
13. **PSSC Physics Course**, DC Heath and Co. (1965) Indian Edition, NCERT (1967).
14. **Physics Advanced Level**, Jim Breithampt, Stanley Thornes Publishers (2000).
15. **Physics,** Patrick Fullick, Heinemann (2000).
16. **Conceptual Physics,** Paul G. Hewitt, Addision—Wesley (1998).
17. **College Physics,** Raymond A. Serway and Jerry S. Faughn, Harcourt Brace and Co. (1999).
18. **University Physics,** Harris Benson, John Wiley (1996).
19. **University Physics,** William P. Crummet and Arthur B. Western, Wm.C. Brown (1994).
20. **General Physics,** Morton M. Sternheim and Joseph W. Kane, John Wiley (1988).
21. **Physics,** Hans C. Ohanian, W.W. Norton (1989).

22 **Advanced Physics,** Keith Gibbs, Cambridge University Press (1996).
23 **Understanding Basic Mechanics,** F. Reif, John Wiley (1995).
24 **College Physics,** Jerry D. Wilson and Anthony J. Buffa, Prentice Hall (1997).
25 **Senior Physics, Part – I,** I.K. Kikoin and A.K. Kikoin, MIR Publishers (1987).
26 **Senior Physics, Part – II,** B. Bekhovtsev, MIR Publishers (1988).
27 **Understanding Physics,** K. Cummings, Patrick J. Cooney, Priscilla W. Laws and Edward F. Redish, John Wiley (2005).
28 **Essentials of Physics,** John D. Cutnell and Kenneth W. Johnson, John Wiley (2005).

GENERAL BOOKS

For instructive and entertaining general reading on science, you may like to read some of the following books. Remember however, that many of these books are written at a level far beyond the level of the present book.

1 **Mr. Tompkins** in paperback, G. Gamow, Cambridge University Press (1967).
2 **The Universe and Dr. Einstein**, C. Barnett, Time Inc. New York (1962).
3 **Thirty years that Shook Physics**, G. Gamow, Double Day, New York (1966).
4 **Surely You're Joking, Mr. Feynman**, R.P. Feynman, Bantam books (1986).
5 **One, Two, Three... Infinity**, G. Gamow, Viking Inc. (1961).
6 **The Meaning of Relativity**, A. Einstein, (Indian Edition) Oxford and IBH Pub. Co. (1965).
7 **Atomic Theory and the Description of Nature**, Niels Bohr, Cambridge (1934).
8 **The Physical Principles of Quantum Theory**, W. Heisenberg, University of Chicago Press (1930).
9 **The Physics—Astronomy Frontier**, F. Hoyle and J.V. Narlikar, W.H. Freeman (1980).
10 **The Flying Circus of Physics with Answer**, J. Walker, John Wiley and Sons (1977).
11 **Physics for Everyone** (series), L.D. Landau and A.I. Kitaigorodski, MIR Publisher (1978).
Book 1: Physical Bodies
Book 2: Molecules
Book 3: Electrons
Book 4: Photons and Nuclei.
12 **Physics can be Fun**, Y. Perelman, MIR Publishers (1986).
13 **Power of Ten**, Philip Morrison and Eames, W.H. Freeman (1985).
14 **Physics in your Kitchen Lab**., I.K. Kikoin, MIR Publishers (1985).
15 **How Things Work: The Physics of Everyday Life**, Louis A. Bloomfield, John Wiley (2005).
16 **Physics Matters: An Introduction to Conceptual Physics,** James Trefil and Robert M. Hazen, John Wiley (2004).