CHAPTER ONE

UNITS AND MEASUREMENT

1.1 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**.

1.2 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, developed by the Bureau International des Poids et measures (The International Bureau of Weights and Measures, BIPM) in 1971 were recently revised by the General Conference on Weights and Measures in November 2018. The scheme is now 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 1.1. 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. 1.1(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.

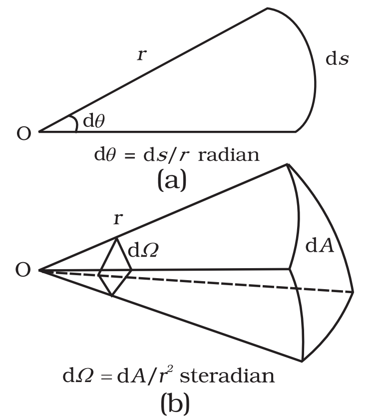


Fig. 1.1 Description of (a) plane angle $d\theta$ and (b) solid angle $d\Omega$.

Table 1.1 SI Base Quantities and Units*

Base quantity	SI Units		
	Name	Symbol	Definition
Length	metre	m	The metre, symbol m, is the SI unit of length. It is defined by taking the fixed numerical value of the speed of light in vacuum c to be 299792458 when expressed in the unit m s^{-1}, where the second is defined in terms of the caesium frequency $\Delta\nu_{cs}$.
Mass	kilogram	kg	The kilogram, symbol kg, is the SI unit of mass. It is defined by taking the fixed numerical value of the Planck constant h to be 6.62607015 10^{-34} when expressed in the unit J s, which is equal to kg $m^2 s^{-1}$, where the metre and the second are defined in terms of c and $\Delta\nu_{cs}$.
Time	second	s	The second, symbol s, is the SI unit of time. It is defined by taking the fixed numerical value of the caesium frequency $\Delta\nu_{cs}$, the unperturbed ground-state hyperfine transition frequency of the caesium-133 atom, to be 9192631770 when expressed in the unit Hz, which is equal to s^{-1}.
Electric	ampere	A	The ampere, symbol A, is the SI unit of electric current. It is defined by taking the fixed numerical value of the elementary charge e to be 1.602176634 10^{-19} when expressed in the unit C, which is equal to A s, where the second is defined in terms of $\Delta\nu_{cs}$.
Thermo dynamic Temperature	kelvin	K	The kelvin, symbol K, is the SI unit of thermodynamic temperature. It is defined by taking the fixed numerical value of the Boltzmann constant k to be 1.380649 10^{-23} when expressed in the unit J K^{-1}, which is equal to kg $m^2 s^{-2} k^{-1}$, where the kilogram, metre and second are defined in terms of h, c and $\Delta\nu_{cs}$.
Amount of substance	mole	mol	The mole, symbol mol, is the SI unit of amount of substance. One mole contains exactly 6.02214076 10^{23} elementary entities. This number is the fixed numerical value of the Avogadro constant, N_A, when expressed in the unit mol^{-1} and is called the Avogadro number. The amount of substance, symbol n, of a system is a measure of the number of specified elementary entities. An elementary entity may be an atom, a molecule, an ion, an electron, any other particle or specified group of particles.
Luminous intensity	candela	cd	The candela, symbol cd, is the SI unit of luminous intensity in given direction. It is defined by taking the fixed numerical value of the luminous efficacy of monochromatic radiation of frequency 540 10^{12} Hz, K_{cd}, to be 683 when expressed in the unit lm W^{-1}, which is equal to cd sr W^{-1}, or cd sr $kg^{-1} m^{-2} s^3$, where the kilogram, metre and second are defined in terms of h, c and $\Delta\nu_{cs}$.

* *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 1.2 **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 1.2.

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.

1.3 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 $a \le 5$) and to 10 (for $5 < a \le 10$). 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.

1.3.1 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 figure) 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 :

$$0.307\ \text{m} - 0.304\ \text{m} = 0.003\ \text{m} = 3 \times 10^{-3}\ \text{m}.$$

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.

1.3.2 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 2.746 rounded off to three significant figures is 1.75, while the number 1.743 would be 1.74. 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 2.745 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 2.745 rounded off to three significant figures becomes 1.74. On the other hand, the number 2.735 rounded off to three significant figures becomes 1.74 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 1.99792458×10^8 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 1.1 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 1.2 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.

$$\text{Density} = \frac{5.74}{1.2}\ \text{g cm}^{-3}$$
$$= 4.8\ \text{g cm}^{-3}.$$

1.3.3 Rules for Determining the Uncertainty in the Results of Arithmetic 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.

1.4 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°], zero dimension in time [T°] and three dimensions in length.

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

Force = mass × acceleration
= mass × (length)/(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 $[L\ T^{-1}]$.

1.5 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 $[\upsilon]$, force $[F]$ and mass density $[\rho]$ 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.

1.6 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.

1.6.1 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

$$x = x_0 + v_0\, t + (1/2)\, a\, t^2$$

for the distance x travelled by a particle or body in time t which starts from the position x_0 with an initial velocity v_0 at time $t = 0$ and has uniform acceleration a along the direction of motion.

The dimensions of each term may be written as

$$[x] = [L]$$
$$[x_0] = [L]$$
$$[v_0\, t] = [L\, T^{-1}]\ [T] = [L]$$
$$[(1/2)\, a\, t^2] = [L\, T^{-2}]\ [T^2] = [L]$$

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 1.3 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 1.4 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); $[M\,L\,T^{-2}]$ 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 5). The correct formula for kinetic energy is given by (b). ◀

1.6.2 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 1.5 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

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

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, $T = k\, l^{½}\, g^{-½}$

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. 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.
11. 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.

1.1 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}.

1.2 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}.

1.3 A calorie is a unit of heat (energy in transit) 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 α^{-1} β^{-2} γ^2 in terms of the new units.

1.4 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.

1.5 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 ?

1.6 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 ?

1.7 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 ?

1.8 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 ?

1.9 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.

1.10 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

1.11 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.

1.12 The mass of a box measured by a grocer's balance is 2.30 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 ?

1.13 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 :

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

Guess where to put the missing c.

1.14 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 ?

1.15 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 ?

1.16 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).

1.17 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.

CHAPTER TWO

11086CH03

MOTION IN A STRAIGHT LINE

2.1 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 4.

2.2 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,

$$v = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} \qquad (2.1a)$$

$$= \frac{dx}{dt} \qquad (2.1b)$$

where the symbol $\lim\limits_{\Delta t \to 0}$ 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. (2.1a) is the differential coefficient of x with respect to t and is denoted by $\frac{dx}{dt}$ (see Appendix 2.1). It is the rate of change of position with respect to time, at that instant.

We can use Eq. (2.1a) 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.2.1 calculation. Let us take $\Delta t = 2$ s centred at $t = 4$ s. Then, by the definition of the average velocity, the slope of line P_1P_2 (Fig. 2.1) gives the value of average velocity over the interval 3 s to 5 s.

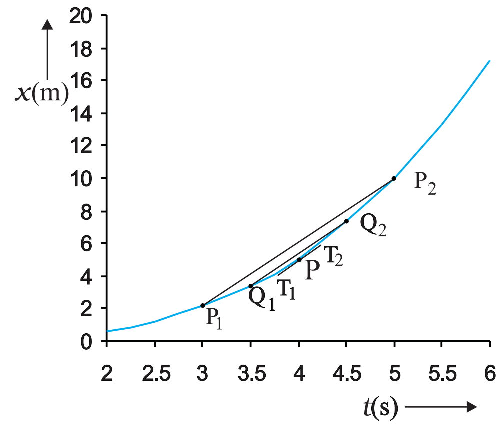


Fig. 2.1 Determining velocity from position-time graph. Velocity at t = 4 s is the slope of the tangent to the graph at that instant.

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. 2.1, $x = 0.08\, t^3$. Table 2.1 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

Table 2.1 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

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 2.1 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.

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 2.1 or use differential calculus for the given expression and calculate $\frac{dx}{dt}$ at different instants as done in the following example.

Example 2.1 *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} .

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

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

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

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 ?

2.3 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 $\overline{a}$ over a time interval is defined as the change of velocity divided by the time interval :

$$\overline{a} = \frac{v_2 - v_1}{t_2 - t_1} = \frac{\Delta v}{\Delta t} \tag{2.2}$$

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).

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

$$a = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} = \frac{dv}{dt} \qquad (2.3)$$

The acceleration at an instant is the slope of the tangent to the *v–t* curve at that instant.

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. 2.4 (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.

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

$$\bar{a} = \frac{v - v_0}{t - 0} \quad \text{or,} \quad v = v_0 + a\,t \qquad (2.4)$$

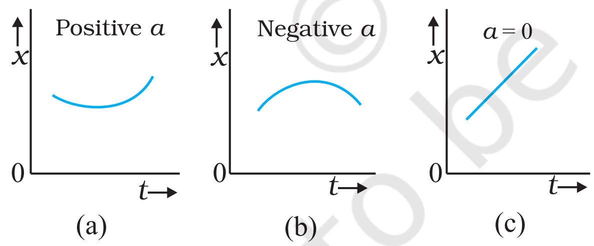


Fig. 2.2 *Position-time graph for motion with (a) positive acceleration; (b) negative acceleration, and (c) zero acceleration.*

Let us see how velocity-time graph looks like for some simple cases. Fig. 2.3 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.

(b) An object is moving in positive direction with a negative acceleration.

(c) An object is moving in negative direction with a negative acceleration.

(d) An object is moving in positive direction till time t_1, and then turns back 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 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. 2.4.

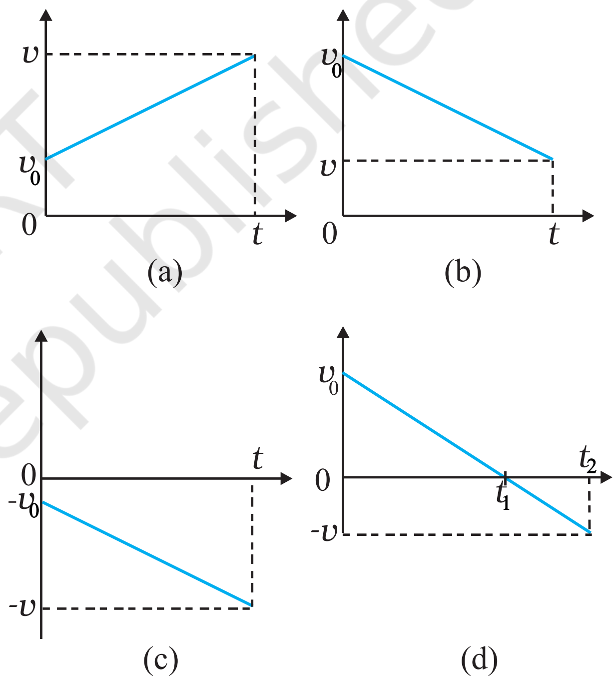


Fig. 2.3 *Velocity–time graph for motions with constant acceleration. (a) Motion in positive direction with positive acceleration, (b) Motion in positive direction with negative acceleration, (c) Motion in negative direction with negative acceleration, (d) Motion of an object with negative acceleration that changes direction at time t_1. Between times 0 to t_1, it moves in positive x - direction and between t_1 and t_2 it moves in the opposite direction.*

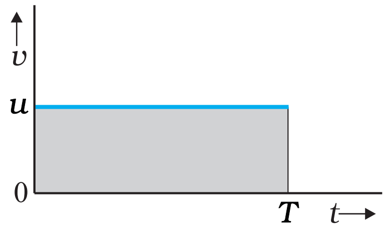


Fig. 2.4 *Area under v–t curve equals displacement of the object over a given time interval.*

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.

2.4 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 (2.4) already obtained gives a relation between final and initial velocities v and v_0 of an object moving with uniform acceleration a :

$$v = v_0 + at \qquad (2.4)$$

This relation is graphically represented in Fig. 2.5. The area under this curve is :
Area between instants 0 and t = Area of triangle ABC + Area of rectangle OACD

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

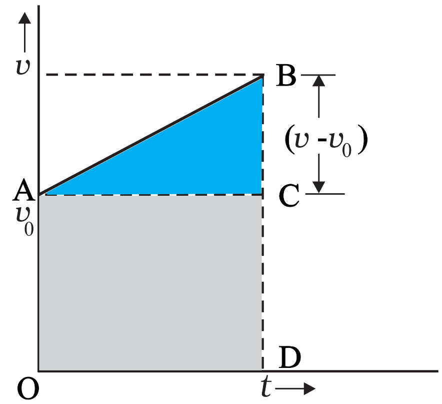


Fig. 2.5 *Area under v-t curve for an object with uniform acceleration.*

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

$$x = \frac{1}{2}(v-v_0)t + v_0 t \qquad (2.5)$$

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

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

or, $$x = v_0 t + \frac{1}{2}at^2 \qquad (2.6)$$

Equation (2.5) can also be written as

$$x = \frac{v + v_0}{2}t = \bar{v}\,t \qquad (2.7a)$$

where,

$$\bar{v} = \frac{v + v_0}{2} \quad \textbf{(constant acceleration only)} \qquad (2.7b)$$

Equations (2.7a) and (2.7b) 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. (2.4), $t = (v - v_0)/a$. Substituting this in Eq. (2.7a), we get

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

$$v^2 = v_0^2 + 2ax \qquad (2.8)$$

This equation can also be obtained by substituting the value of t from Eq. (2.4) into Eq. (2.6). Thus, we have obtained three important equations :

$$v = v_0 + at$$

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

$$v^2 = v_0^2 + 2ax \qquad (2.9a)$$

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

The set of Eq. (2.9a) 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. (2.9a) are modified (replacing x by $x - x_0$) to :

$$v = v_0 + at$$

$$x = x_0 + v_0 t + \frac{1}{2} a t^2 \qquad (2.9b)$$

$$v^2 = v_0^2 + 2a(x - x_0) \qquad (2.9c)$$

Example 2.2 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 2.3 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. 2.6.

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**.

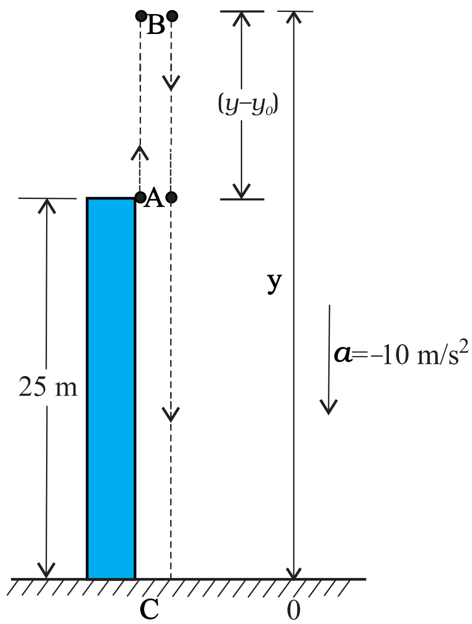


Fig. 2.6

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, $$t_1 = 2 \text{ s}$$

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 y = 0 m
v_o = 20 m s^{-1}, a = –10m s^{-2}, t = ?

$$0 = 25 + 20\,t + (½)\,(-10)\,t^2$$
Or, $$5t^2 - 20t - 25 = 0$$

Solving this quadratic equation for t, we get

$$t = 5\text{s}$$

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 2.4*** **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 $\boldsymbol{g}$. 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.

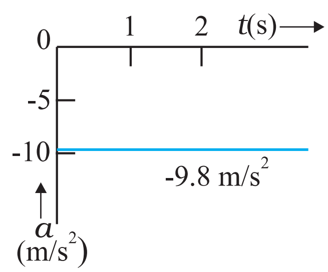
We assume that the motion is in y-direction, more correctly in –y-direction because we choose upward direction as positive. Since the acceleration due to gravity is always downward, it is in the negative direction and we have

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

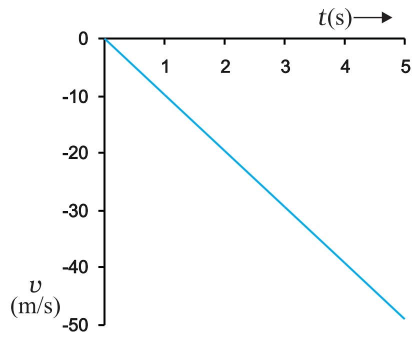
The object is released from rest at y = 0. Therefore, v_0 = 0 and the equations of motion become:

$$v = 0 - g\,t \quad = -9.8\,t \quad \text{m s}^{-1}$$
$$y = 0 - ½\,g\,t^2 \quad = -4.9\,t^2 \quad \text{m}$$
$$v^2 = 0 - 2\,g\,y \quad = -19.6\,y \quad \text{m}^2\,\text{s}^{-2}$$

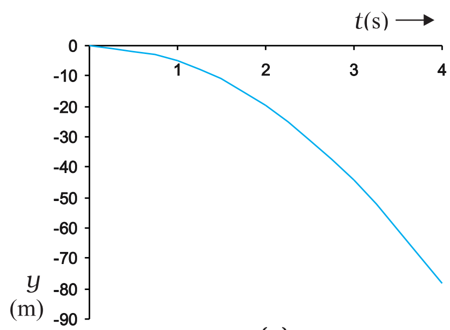
These equations give the velocity and the distance travelled as a function of time and also the variation of velocity with distance. The variation of acceleration, velocity, and distance, with time have been plotted in Fig. 2.7(a), (b) and (c).



(a)



(b)



(c)

Fig. 2.7 *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 2.5 **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 2.2. 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 2.6 **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

$$d_s = \frac{-v_0^2}{2a}$$

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

Table 2.2

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	$-(1/2)\, g\, \tau^2$	y_o	y_o	1
$2\,\tau$	$-4(1/2)\, g\, \tau^2$	$4\, y_o$	$3\, y_o$	3
$3\,\tau$	$-9(1/2)\, g\, \tau^2$	$9\, y_o$	$5\, y_o$	5
$4\,\tau$	$-16(1/2)\, g\, \tau^2$	$16\, y_o$	$7\, y_o$	7
$5\,\tau$	$-25(1/2)\, g\, \tau^2$	$25\, y_o$	$9\, y_o$	9
$6\,\tau$	$-36(1/2)\, g\, \tau^2$	$36\, y_o$	$11\, y_o$	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 2.7*** **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. 2.8). 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.

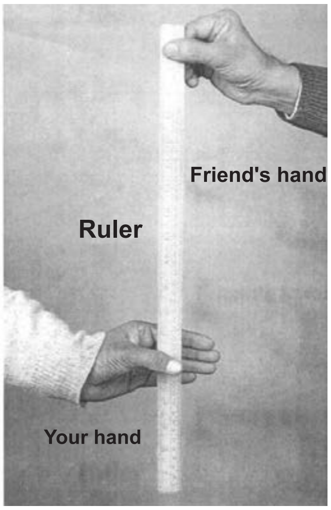


Fig. 2.8 *Measuring the reaction time.*

Answer The ruler drops under free fall. Therefore, $v_o = 0$, and $a = -g = -9.8$ 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

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

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.

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

2. *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.

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

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

4. *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.

5. 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.

6. 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 (a) Average (b) Instantaneous		$[LT^{-1}]$	$m\ s^{-1}$	$= \frac{\textit{Path length}}{\textit{Time interval}}$ $= \frac{dx}{dt}$
Acceleration (a) Average (b) Instantaneous	 $\overline{a}$ a	$[LT^{-2}]$	$m\ s^{-2}$	$= \frac{\Delta v}{\Delta t}$ $= \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} = \frac{dv}{dt}$ In one dimension, its sign indicates the direction.

POINTS TO PONDER

1. The 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.
2. 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.
3. 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.
4. 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.
5. In the kinematic equations of motion [Eq. (2.9)], 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.
6. The definitions of instantaneous velocity and acceleration (Eqs. (2.1) and (2.3)) are exact and are always correct while the kinematic equations (Eq. (2.9)) are true only for motion in which the magnitude and the direction of acceleration are constant during the course of motion.

EXERCISES

2.1 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.

2.2 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. 2.9. 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).

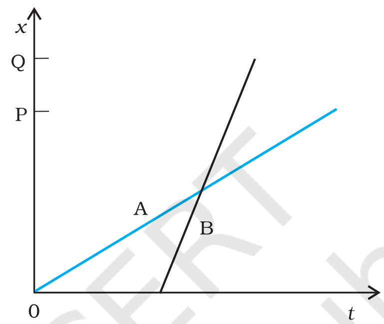


Fig. 2.9

2.3 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.

2.4 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.

2.5 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 ?

2.6 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).

2.7 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.

2.8 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.

2.9 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].

2.10 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 !]

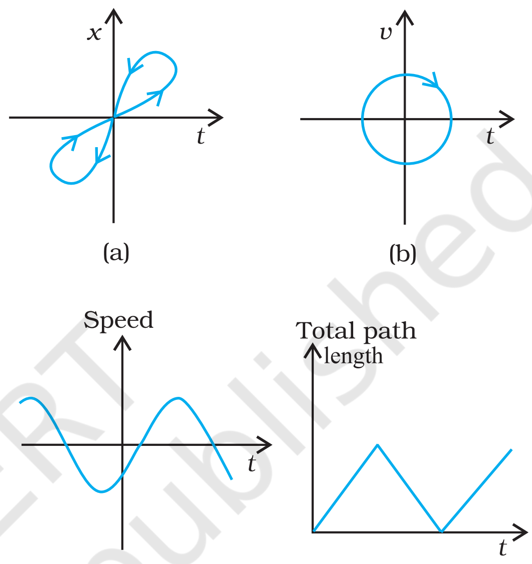


Fig. 2.10

2.11 In Exercises 2.9 and 2.10, 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?

2.12 Look at the graphs (a) to (d) (Fig. 2.10) carefully and state, with reasons, which of these *cannot* possibly represent one-dimensional motion of a particle.

2.13 Figure 2.11shows 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.

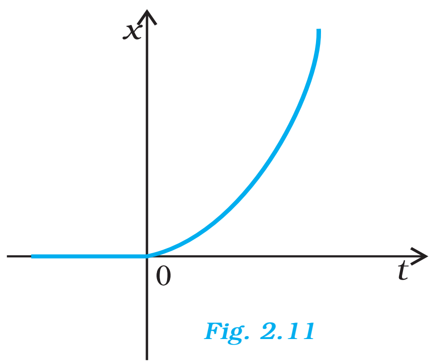


Fig. 2.11

2.14 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).

2.15 Suggest a suitable physical situation for each of the following graphs (Fig 2.12):

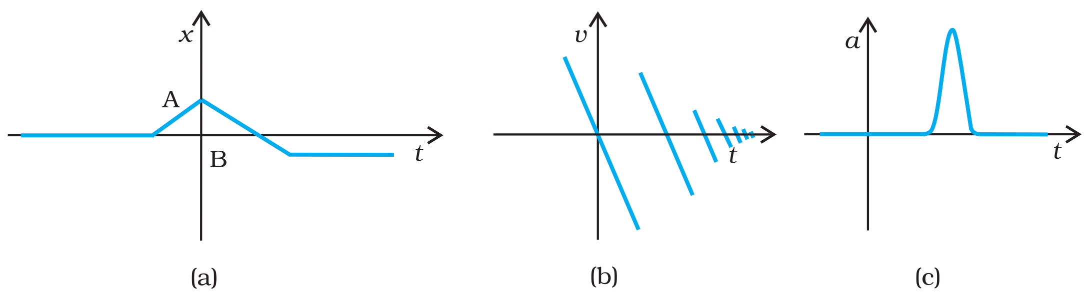


Fig. 2.12

2.16 Figure 2.13 gives the *x-t* plot of a particle executing one-dimensional simple harmonic motion. (You will learn about this motion in more detail in Chapter13). Give the signs of position, velocity and acceleration variables of the particle at t = 0.3 s, 1.2 s, – 1.2 s.

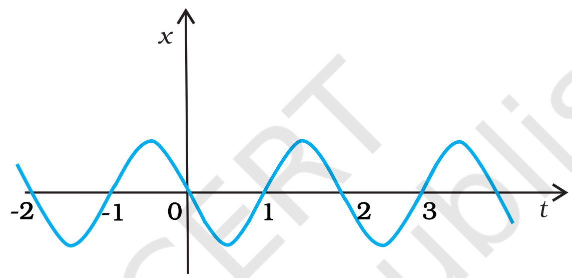


Fig. 2.13

2.17 Figure 2.14 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.

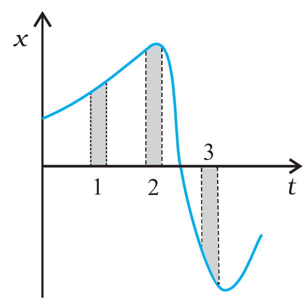


Fig. 2.14

2.18 Figure 2.15 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 ?

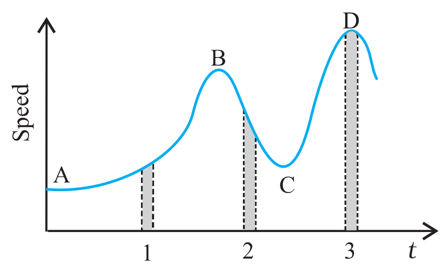


Fig. 2.15

11086CH04

Chapter Three

Motion in a Plane

3.1 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.

3.2 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 |**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*.

3.2.1 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. 3.1(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*′).

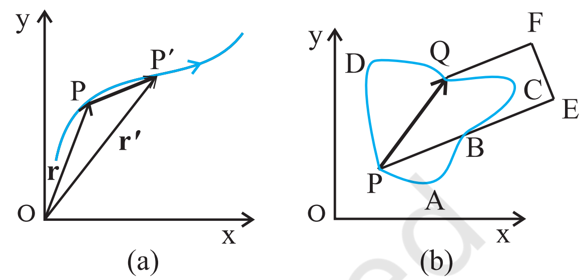


Fig. 3.1 (a) *Position and displacement vectors.* (b) *Displacement vector* **PQ** *and different courses of motion.*

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. 3.1(b), 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.

3.2.2 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 3.2(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.

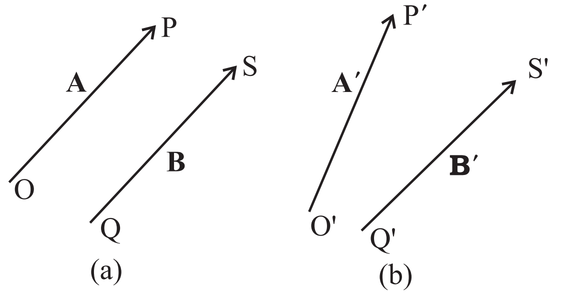


Fig. 3.2 (a) *Two equal vectors* **A** *and* **B**. (b) *Two vectors* **A′** *and* **B′** *are unequal though they are of the same length.*

as **A** = **B**. Note that in Fig. 3.2(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′**.

3.3 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** :

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

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. 3.3(a).

Multiplying a vector **A** by a negative number –λ gives another vector 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 3.3(b).

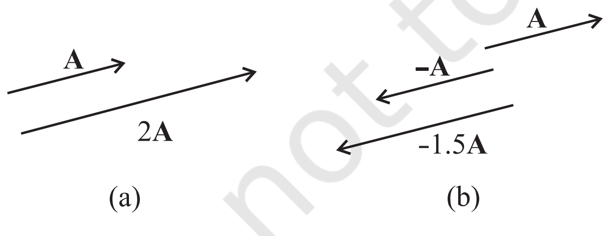


Fig. 3.3 (a) *Vector* **A** *and the resultant vector after multiplying* **A** *by a positive number 2.* (b) *Vector* **A** *and resultant vectors after multiplying it by a negative number –1 and –1.5.*

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.

3.4 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. 3.4(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. 3.4(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

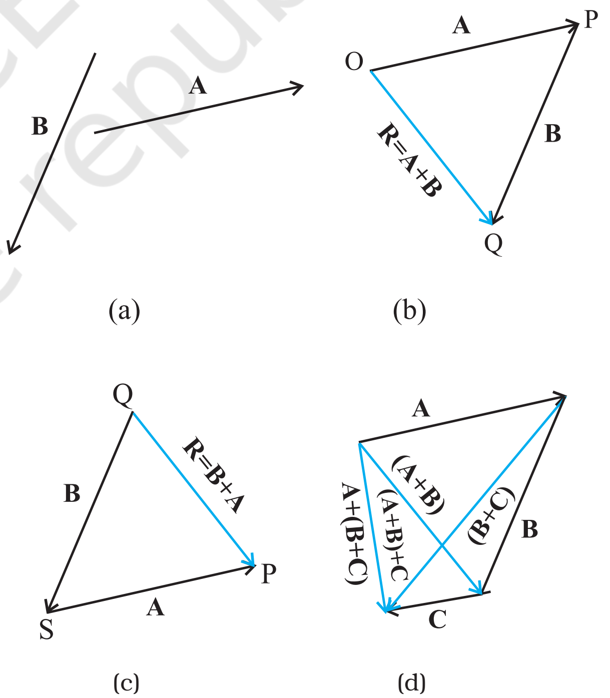


Fig. 3.4 (a) *Vectors* **A** *and* **B**. (b) *Vectors* **A** *and* **B** *added graphically.* (c) *Vectors* **B** *and* **A** *added graphically.* (d) *Illustrating the associative law of vector addition.*

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. 3.4(c), the same vector **R** is obtained. Thus, vector addition is **commutative**:

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A} \tag{3.1}$$

The addition of vectors also obeys the associative law as illustrated in Fig. 3.4(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** :

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C}) \tag{3.2}$$

What is the result of adding two equal and opposite vectors ? Consider two vectors **A** and –**A** shown in Fig. 3.3(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** :

$$\mathbf{A} - \mathbf{A} = \mathbf{0} \qquad |\mathbf{0}| = 0 \tag{3.3}$$

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 :

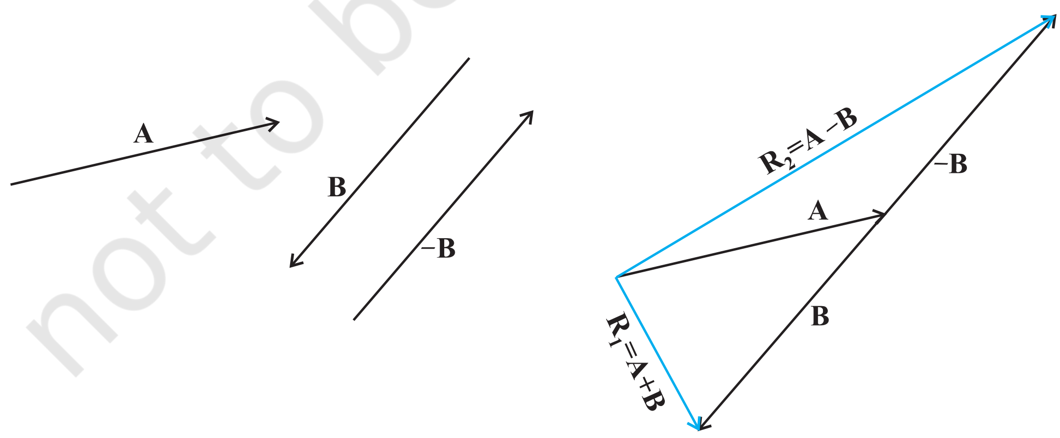
$$\mathbf{A} + \mathbf{0} = \mathbf{A}$$
$$\lambda\,\mathbf{0} = \mathbf{0}$$
$$0\,\mathbf{A} = \mathbf{0} \tag{3.4}$$

What is the physical meaning of a zero vector? Consider the position and displacement vectors in a plane as shown in Fig. 3.1(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** :

$$\mathbf{A} - \mathbf{B} = \mathbf{A} + (-\mathbf{B}) \tag{3.5}$$

It is shown in Fig 3.5. 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. 3.6(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. 3.6(b)]. In Fig.3.6(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.



(a) (b)

Fig. 3.5 (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.*

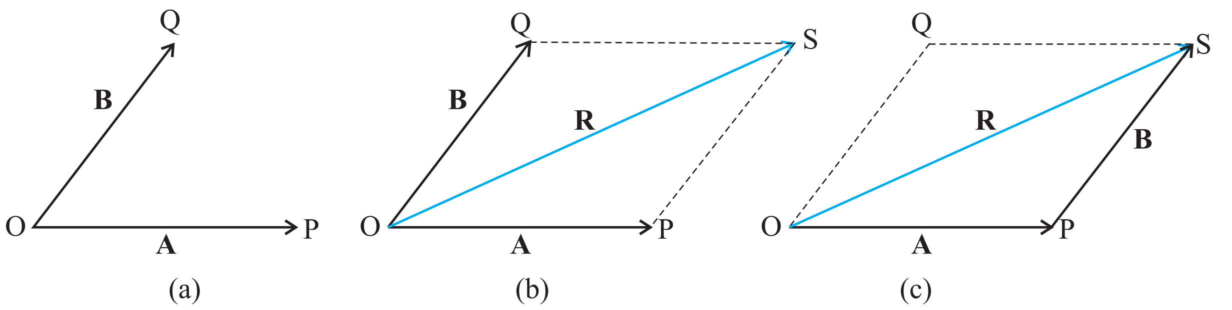


Fig. 3.6 (a) *Two vectors* **A** *and* **B** *with their tails brought to a common origin.* (b) *The sum* **A** + **B** *obtained using the parallelogram method.* (c) *The parallelogram method of vector addition is equivalent to the triangle method.*

Example 3.1 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 ?

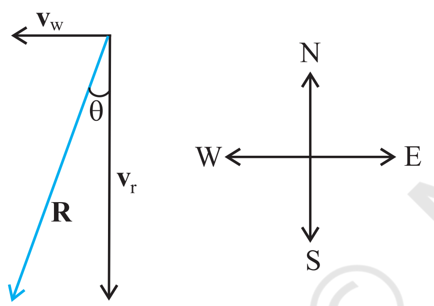


Fig. 3.7

Answer The velocity of the rain and the wind are represented by the vectors $\mathbf{v}_r$ and $\mathbf{v}_w$ in Fig. 3.7 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.

3.5 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. 3.8). **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

$$\mathbf{A} = \mathbf{OP} = \mathbf{OQ} + \mathbf{QP} \qquad (3.6)$$

But since **OQ** is parallel to **a**, and **QP** is parallel to **b**, we can write :

$$\mathbf{OQ} = \lambda\,\mathbf{a}, \text{ and } \mathbf{QP} = \mu\,\mathbf{b} \qquad (3.7)$$

where λ and μ are real numbers.

Therefore, $\mathbf{A} = \lambda\,\mathbf{a} + \mu\,\mathbf{b}$ (3.8)

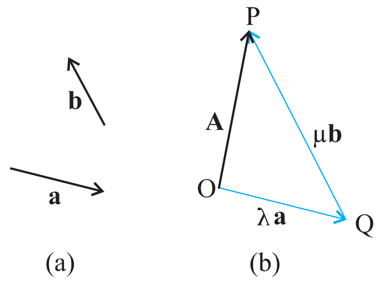


Fig. 3.8 (a) *Two non-colinear vectors* **a** *and* **b**. (b) *Resolving a vector* **A** *in terms of vectors* **a** *and* **b**.

We say that **A** has been resolved into two component vectors $\lambda\,\mathbf{a}$ and $\mu\,\mathbf{b}$ 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. 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. 3.9(a).

Since these are unit vectors, we have

$$|\hat{\mathbf{i}}| = |\hat{\mathbf{j}}| = |\hat{\mathbf{k}}| = 1 \qquad (3.9)$$

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 $\hat{\mathbf{n}}$ by a scalar, the result is a vector $\boldsymbol{\lambda} = \lambda\hat{\mathbf{n}}$. In general, a vector **A** can be written as

$$\mathbf{A} = |\mathbf{A}|\,\hat{\mathbf{n}} \qquad (3.10)$$

where $\hat{\mathbf{n}}$ 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. 3.9(b). We draw lines from the head of **A** perpendicular to the coordinate axes as in Fig. 3.9(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 :

$$\mathbf{A_1} = A_x\,\hat{\mathbf{i}}, \quad \mathbf{A_2} = A_y\,\hat{\mathbf{j}} \qquad (3.11)$$

where A_x and A_y are real numbers.

Thus, $$\mathbf{A} = A_x\,\hat{\mathbf{i}} + A_y\,\hat{\mathbf{j}} \qquad (3.12)$$

This is represented in Fig. 3.9(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 :

$$A_x = A\cos\theta$$
$$A_y = A\sin\theta \qquad (3.13)$$

As is clear from Eq. (3.13), 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. (3.13). 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, $$A = \sqrt{A_x^2 + A_y^2} \qquad (3.14)$$

And $$\tan\theta = \frac{A_y}{A_x}, \quad \theta = \tan^{-1}\frac{A_y}{A_x} \qquad (3.15)$$

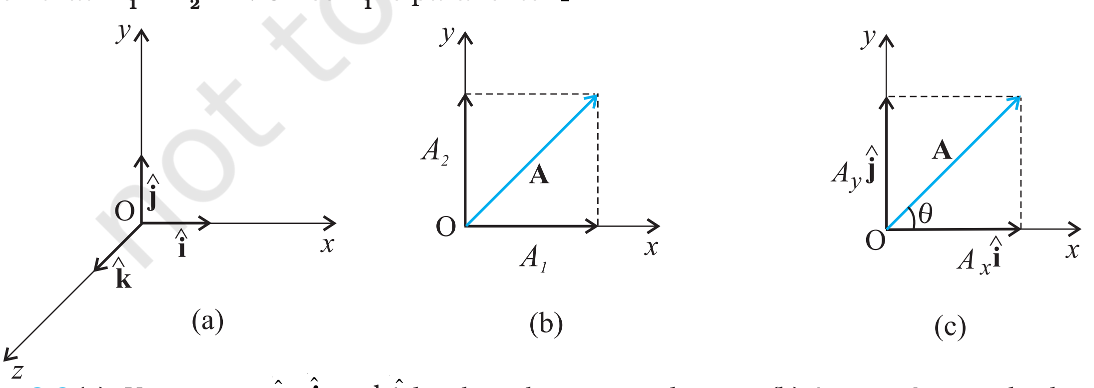
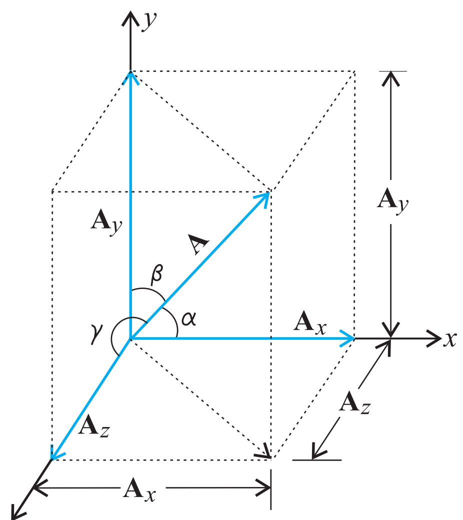


Fig. 3.9 (a) *Unit vectors* $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$ *and* $\hat{\mathbf{k}}$ *lie along the x-, y-, and z-axes.* (b) *A vector* **A** *is resolved into its components* A_x *and* A_y *along x-, and y- axes.* (c) $\mathbf{A_1}$ *and* $\mathbf{A_2}$ *expressed in terms of* $\hat{\mathbf{i}}$ *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. 3.9(d)], we have



(d)

Fig. 3.9 (d) *A vector* **A** *resolved into components along x-, y-, and z-axes*

$$A_x = A\cos\alpha,\ A_y = A\cos\beta,\ A_z = A\cos\gamma \qquad (3.16a)$$

In general, we have

$$\mathbf{A} = A_x\hat{\mathbf{i}} + A_y\hat{\mathbf{j}} + A_z\hat{\mathbf{k}} \qquad (3.16b)$$

The magnitude of vector **A** is

$$A = \sqrt{A_x^2 + A_y^2 + A_z^2} \qquad (3.16c)$$

A position vector **r** can be expressed as

$$\mathbf{r} = x\,\hat{\mathbf{i}} + y\,\hat{\mathbf{j}} + z\,\hat{\mathbf{k}} \qquad (3.17)$$

where x, y, and z are the components of **r** along x-, y-, z-axes, respectively.

3.6 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 :

$$\mathbf{A} = A_x\hat{\mathbf{i}} + A_y\hat{\mathbf{j}} \qquad (3.18)$$

$$\mathbf{B} = B_x\hat{\mathbf{i}} + B_y\hat{\mathbf{j}}$$

Let **R** be their sum. We have

$$\mathbf{R} = \mathbf{A} + \mathbf{B}$$

$$= \left(A_x\hat{\mathbf{i}} + A_y\hat{\mathbf{j}}\right) + \left(B_x\hat{\mathbf{i}} + B_y\hat{\mathbf{j}}\right) \qquad (3.19a)$$

Since vectors obey the commutative and associative laws, we can arrange and regroup the vectors in Eq. (3.19a) as convenient to us :

$$\mathbf{R} = \left(A_x + B_x\right)\hat{\mathbf{i}} + \left(A_y + B_y\right)\hat{\mathbf{j}} \qquad (3.19b)$$

$$\text{Since}\ \mathbf{R} = R_x\hat{\mathbf{i}} + R_y\hat{\mathbf{j}} \qquad (3.20)$$

$$\text{we have,}\ R_x = A_x + B_x,\ R_y = A_y + B_y \qquad (3.21)$$

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

$$R_x = A_x + B_x$$
$$R_y = A_y + B_y$$
$$R_z = A_z + B_z \qquad (3.22)$$

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

$$\mathbf{a} = a_x\hat{\mathbf{i}} + a_y\hat{\mathbf{j}} + a_z\hat{\mathbf{k}}$$

$$\mathbf{b} = b_x\hat{\mathbf{i}} + b_y\hat{\mathbf{j}} + b_z\hat{\mathbf{k}}$$

$$\mathbf{c} = c_x\hat{\mathbf{i}} + c_y\hat{\mathbf{j}} + c_z\hat{\mathbf{k}} \qquad (3.23a)$$

then, a vector **T** = **a** + **b** – **c** has components :

$$T_x = a_x + b_x - c_x$$
$$T_y = a_y + b_y - c_y \qquad (3.23b)$$
$$T_z = a_z + b_z - c_z.$$

▶ ***Example 3.2*** 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.*

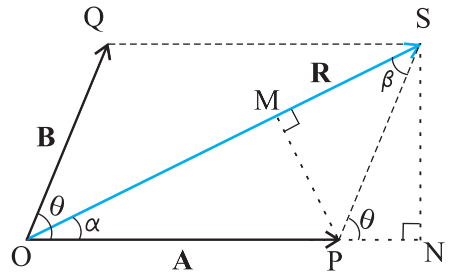


Fig. 3.10

Answer Let **OP** and **OQ** represent the two vectors **A** and **B** making an angle θ (Fig. 3.10). 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 $\quad ON = OP + PN = A + B\cos\theta$

$$SN = B\sin\theta$$

$$OS^2 = (A + B\cos\theta)^2 + (B\sin\theta)^2$$

or, $\quad R^2 = A^2 + B^2 + 2AB\cos\theta$

$$R = \sqrt{A^2 + B^2 + 2AB\cos\theta} \qquad (3.24a)$$

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, $$\frac{R}{\sin\theta} = \frac{B}{\sin\alpha} \qquad (3.24b)$$

Similarly,

$$PM = A\sin\alpha = B\sin\beta$$

or, $$\frac{A}{\sin\beta} = \frac{B}{\sin\alpha} \qquad (3.24c)$$

Combining Eqs. (3.24b) and (3.24c), we get

$$\frac{R}{\sin\theta} = \frac{A}{\sin\beta} = \frac{B}{\sin\alpha} \qquad (3.24d)$$

Using Eq. (3.24d), we get:

$$\sin\alpha = \frac{B}{R}\sin\theta \qquad (3.24e)$$

where *R* is given by Eq. (3.24a).

or, $$\tan\alpha = \frac{SN}{OP + PN} = \frac{B\sin\theta}{A + B\cos\theta} \qquad (3.24f)$$

Equation (3.24a) gives the magnitude of the resultant and Eqs. (3.24e) and (3.24f) its direction. Equation (3.24a) is known as the **Law of cosines** and Eq. (3.24d) as the **Law of sines**. ◀

▶ ***Example 3.3*** 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. 3.11 in directions specified by the problem. Using the parallelogram method of addition, the resultant **R** is obtained in the direction shown in the figure.

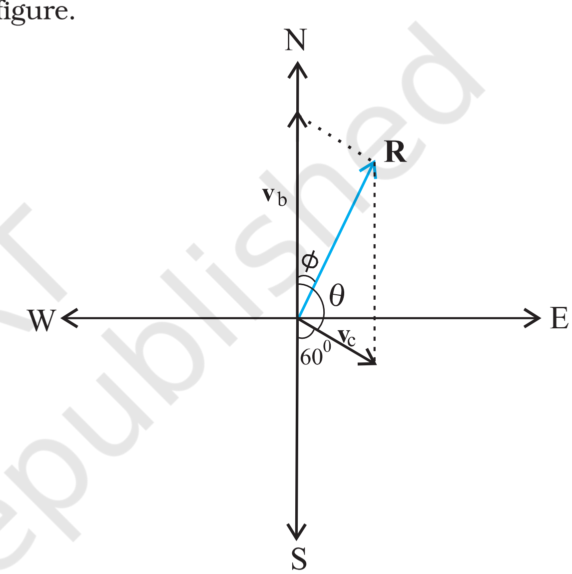


Fig. 3.11

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

$$\frac{R}{\sin\theta} = \frac{v_c}{\sin\phi} \quad \text{or,}\ \sin\phi = \frac{v_c}{R}\sin\theta$$

$$= \frac{10\times\sin 120^\circ}{21.8} = \frac{10\sqrt{3}}{2\times 21.8} \cong 0.397$$

$$\phi \cong 23.4^\circ$$ ◀

3.7 MOTION IN A PLANE

In this section we shall see how to describe motion in two dimensions using vectors.

3.7.1 Position Vector and Displacement

The position vector $\mathbf{r}$ of a particle P located in a plane with reference to the origin of an *x-y* reference frame (Fig. 3.12) is given by

$$\mathbf{r} = x\,\hat{\mathbf{i}} + y\,\hat{\mathbf{j}}$$

where x and y are components of $\mathbf{r}$ along x-, and y- axes or simply they are the coordinates of the object.

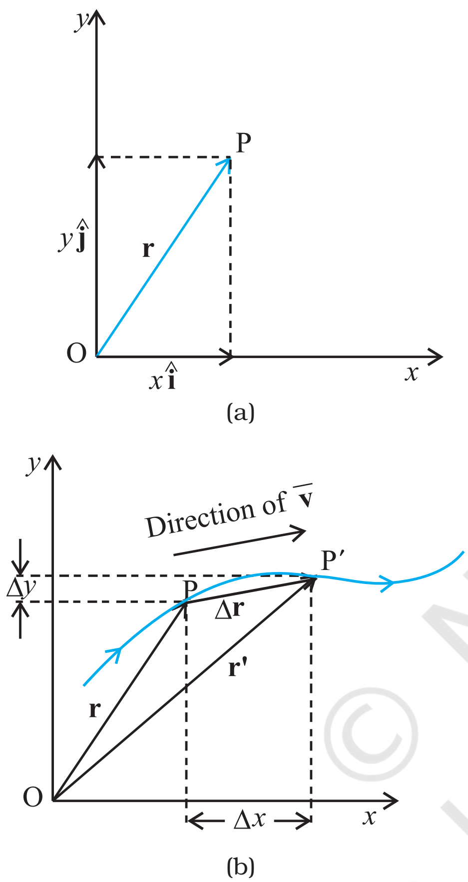


Fig. 3.12 (a) *Position vector* $\mathbf{r}$. (b) *Displacement* $\Delta\mathbf{r}$ *and average velocity* $\bar{\mathbf{v}}$ *of a particle.*

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. 3.12(b)]. Then, the displacement is :

$$\Delta\mathbf{r} = \mathbf{r}' - \mathbf{r} \qquad (3.25)$$

and is directed from P to P′.

We can write Eq. (3.25) 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 $\Delta x = x' - x,\ \Delta y = y' - y \qquad (3.26)$

Velocity

The average velocity $(\bar{\mathbf{v}})$ of an object is the ratio of the displacement and the corresponding time interval :

$$\bar{\mathbf{v}} = \frac{\Delta\mathbf{r}}{\Delta t} = \frac{\Delta x\,\hat{\mathbf{i}} + \Delta y\,\hat{\mathbf{j}}}{\Delta t} = \hat{\mathbf{i}}\frac{\Delta x}{\Delta t} + \hat{\mathbf{j}}\frac{\Delta y}{\Delta t} \qquad (3.27)$$

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 $\Delta\mathbf{r}$ (Fig. 3.12). The **velocity** (instantaneous velocity) is given by the limiting value of the average velocity as the time interval approaches zero :

$$\mathbf{v} = \lim_{\Delta t \to 0}\frac{\Delta\mathbf{r}}{\Delta t} = \frac{d\mathbf{r}}{dt} \qquad (3.28)$$

The meaning of the limiting process can be easily understood with the help of Fig 3.13(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 $\Delta t_1, \Delta t_2$, 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

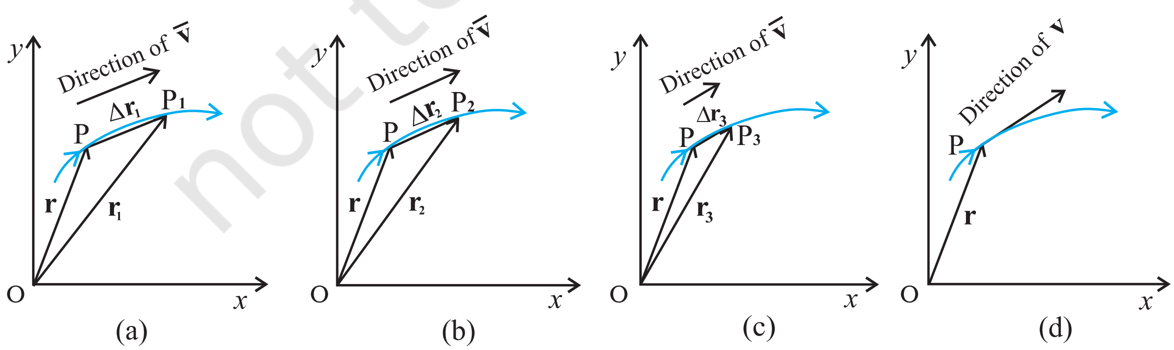


Fig. 3.13 *As the time interval* Δt *approaches zero, the average velocity approaches the velocity* $\mathbf{v}$. *The direction of* $\bar{\mathbf{v}}$ *is parallel to the line tangent to the path.*

Δt_3, respectively. The direction of the average velocity $\overline{\mathbf{v}}$ 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. 3.13(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 :

$$\mathbf{v} = \frac{d\mathbf{r}}{dt}$$

$$= \lim_{\Delta t \to 0}\left(\frac{\Delta x}{\Delta t}\hat{\mathbf{i}} + \frac{\Delta y}{\Delta t}\hat{\mathbf{j}}\right) \qquad (3.29)$$

$$= \hat{\mathbf{i}} \lim_{\Delta t \to 0}\frac{\Delta x}{\Delta t} + \hat{\mathbf{j}} \lim_{\Delta t \to 0}\frac{\Delta y}{\Delta t}$$

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 $$v_x = \frac{dx}{dt}, v_y = \frac{dy}{dt} \qquad (3.30a)$$

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

$$v = \sqrt{v_x^2 + v_y^2} \qquad (3.30b)$$

and the direction of **v** is given by the angle θ :

$$\tan\theta = \frac{v_y}{v_x}, \quad \theta = \tan^{-1}\left(\frac{v_y}{v_x}\right) \qquad (3.30c)$$

v_x, v_y and angle θ are shown in Fig. 3.14 for a velocity vector **v** at point **p**.

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 :

$$\overline{\mathbf{a}} = \frac{\Delta \mathbf{v}}{\Delta t} = \frac{\Delta\left(v_x\hat{\mathbf{i}} + v_y\hat{\mathbf{j}}\right)}{\Delta t} = \frac{\Delta v_x}{\Delta t}\hat{\mathbf{i}} + \frac{\Delta v_y}{\Delta t}\hat{\mathbf{j}} \qquad (3.31a)$$

Or, $$\overline{\mathbf{a}} = a_x\hat{\mathbf{i}} + a_y\hat{\mathbf{j}}. \qquad (3.31b)$$

* *In terms of x and y, a_x and a_y can be expressed as*

$$a_x = \frac{d}{dt}\left(\frac{dx}{dt}\right) = \frac{d^2x}{dt^2}, \quad a_y = \frac{d}{dt}\left(\frac{dy}{dt}\right) = \frac{d^2y}{dt^2}$$

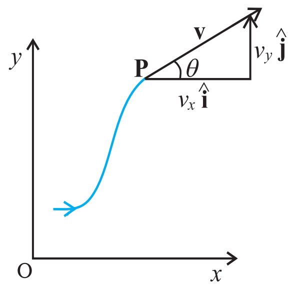


Fig. 3.14 The components v_x and v_y of velocity **v** *and the angle θ it makes with x-axis. Note that $v_x = v\cos\theta$, $v_y = v\sin\theta$.*

The **acceleration** (instantaneous acceleration) is the limiting value of the average acceleration as the time interval approaches zero :

$$\mathbf{a} = \lim_{\Delta t \to 0}\frac{\Delta \mathbf{v}}{\Delta t} \qquad (3.32a)$$

Since $\Delta \boldsymbol{v} = \Delta v_x\hat{\mathbf{i}} + \Delta v_y\hat{\mathbf{j}}$, 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, $$\mathbf{a} = a_x\hat{\mathbf{i}} + a_y\hat{\mathbf{j}} \qquad (3.32b)$$

where, $$a_x = \frac{dv_x}{dt}, \; a_y = \frac{dv_y}{dt} \qquad (3.32c)^*$$

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. 3.15(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. 3.15 (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 $\Delta t \to 0$ [Fig. 3.15(d)], the average acceleration becomes the instantaneous acceleration and has the direction as shown.

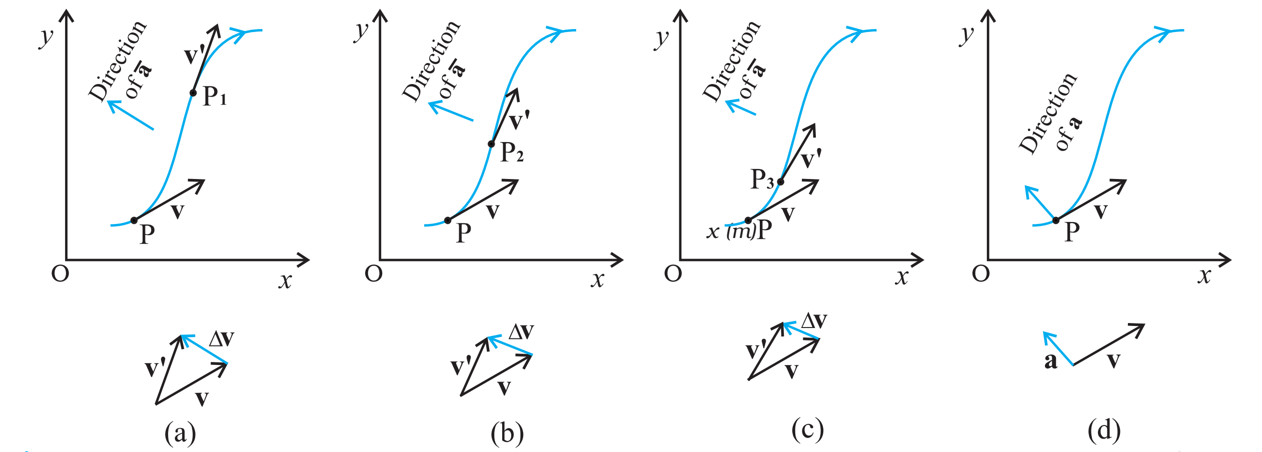


Fig. 3.15 *The average acceleration for three time intervals* (a) Δt_1, (b) Δt_2, *and* (c) Δt_3, *(* $\Delta t_1 > \Delta t_2 > \Delta t_3$ *).* (d) *In the limit* $\Delta t \to 0$, *the average acceleration becomes the acceleration.*

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 3.4 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

$$\theta = \tan^{-1}\left(\frac{v_y}{v_x}\right) = \tan^{-1}\left(\frac{4}{3}\right) \cong 53^\circ \text{ with } x\text{-axis.}$$

3.8 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, $\mathbf{v} = \mathbf{v_0} + \mathbf{a}t$ (3.33a)

In terms of components :

$$v_x = v_{ox} + a_x t$$

$$v_y = v_{oy} + a_y t \qquad (3.33b)$$

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, $\mathbf{r}=\mathbf{r_0}+\mathbf{v_0}t+\frac{1}{2}\mathbf{a}t^2$ (3.34a)

It can be easily verified that the derivative of Eq. (3.34a), i.e. $\frac{d\mathbf{r}}{dt}$ gives Eq.(3.33a) and it also satisfies the condition that at t=0, $\mathbf{r} = \mathbf{r}_o$. Equation (3.34a) can be written in component form as

$$x=x_0+v_{ox}t+\frac{1}{2}a_x t^2$$

$$y=y_0+v_{oy}t+\frac{1}{2}a_y t^2 \qquad (3.34b)$$

One immediate interpretation of Eq.(3.34b) 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 3.9 for projectile motion.

Example 3.5 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 From Eq. (3.34a) for $\mathbf{r}_0 = 0$, 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, $t = ?$

$5.0\,t + 1.5\,t^2 = 84 \Rightarrow t = 6$ 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 $=|\mathbf{v}|=\sqrt{23^2+12^2}\cong 26\ \text{m s}^{-1}$. ◀

3.9 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. 3.16.

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, $a_x = 0,\ a_y = -g$ (3.35)

The components of initial velocity $\mathbf{v}_o$ are :

$v_{ox} = v_o\cos\theta_o$
$v_{oy} = v_o\sin\theta_o$ (3.36)

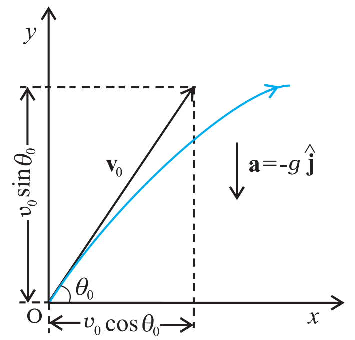


Fig 3.16 Motion of an object projected with velocity $\mathbf{v}_o$ at angle θ_o.

If we take the initial position to be the origin of the reference frame as shown in Fig. 3.16, we have :

$$x_o = 0,\ y_o = 0$$

Then, Eq.(3.34b) becomes :

$$x = v_{ox}\, t = (v_o \cos\theta_o)\, t$$

and $\quad y = (v_o \sin\theta_o)\, t - (\tfrac{1}{2})g\,t^2 \qquad (3.37)$

The components of velocity at time t can be obtained using Eq.(3.33b) :

$$v_x = v_{ox} = v_o \cos\theta_o$$

$$v_y = v_o \sin\theta_o - g\,t \qquad (3.38)$$

Equation (3.37) 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. 3.17. 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. (3.37). We obtain:

$$y = (\tan\theta_o)\,x - \frac{g}{2\,(v_o\cos\theta_o)^2}\,x^2 \qquad (3.39)$$

Now, since g, θ_o and v_o are constants, Eq. (3.39) 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. 3.17).

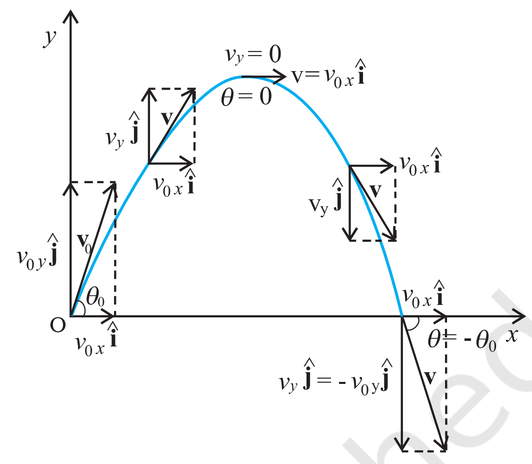


Fig. 3.17 The path of a projectile is a parabola.

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. (3.38):

$$v_y = v_o \sin\theta_o - g\,t_m = 0$$

Or, $\quad t_m = v_o \sin\theta_o / g \qquad (3.40a)$

The total time T_f during which the projectile is in flight can be obtained by putting $y = 0$ in Eq. (3.37). We get :

$$T_f = 2\,(v_o \sin\theta_o)/g \qquad (3.40b)$$

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. (3.37) :

$$y = h_m = \left(v_0 \sin\theta_0\right)\left(\frac{v_0\sin\theta_0}{g}\right) - \frac{g}{2}\left(\frac{v_0\sin\theta_0}{g}\right)^2$$

Or, $\quad h_m = \dfrac{\left(v_0\sin\theta_0\right)^2}{2g} \qquad (3.41)$

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, $$R = \frac{v_0^2 \sin 2\theta_0}{g} \qquad (3.42a)$$

Equation (3.42a) 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,

$$R_m = \frac{v_0^2}{g} \qquad (3.42b)$$

> ***Example 3.6*** 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 3.7*** 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 :

$$v_{ox} = 15 \text{ m s}^{-1}$$

$$v_{oy} = 0 - 9.8 \quad 10 = -\,98 \text{ m s}^{-1}$$

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 3.8*** 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

$$T_f = (2\ v_o \sin \theta_o)/g = (2 \quad 28 \quad \sin 30\ )/9.8$$
$$= 28/9.8 \text{ s} = 2.9 \text{ s}$$

(c) The distance from the thrower to the point where the ball returns to the same level is

$$R = \frac{(v_o^2 \sin 2\theta_o)}{g} = \frac{28 \times 28 \times \sin 60^o}{9.8} = 69 \text{ m}$$ ◀

3.10 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. 3.18. 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.

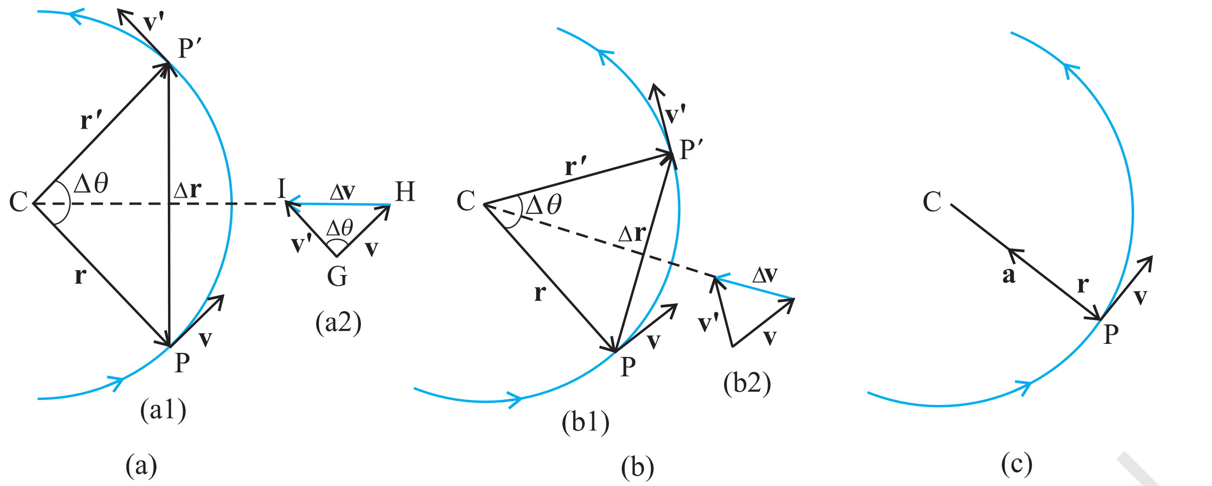


Fig. 3.18 *Velocity and acceleration of an object in uniform circular motion. The time interval Δt decreases from (a) to (c) where it is zero. The acceleration is directed, at each point of the path, towards the centre of the circle.*

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. 3.18(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. 3.18(a1). Δ**v** is obtained in Fig. 3.18 (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 3.18(b) shows the same quantities for smaller time interval. Δ**v** and hence $\bar{\mathbf{a}}$ is again directed towards the centre. In Fig. 3.18(c), Δ*t*→ 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 Δ*θ*. Since the velocity vectors **v** and **v**′ are always perpendicular to the position vectors, the angle between them is also Δ*θ* . 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. 3.18a). 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.*

$$a_c = \left(\frac{v}{R}\right)v = v^2/R \qquad (3.43)$$

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. 3.18) 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 :

$$\omega = \frac{\Delta\theta}{\Delta t} \qquad (3.44)$$

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 :

$$v = R\frac{\Delta\theta}{\Delta t} = R\,\omega$$

$$v = R\,\omega \qquad (3.45)$$

We can express centripetal acceleration a_c in terms of angular speed :

$$a_c = \frac{v^2}{R} = \frac{\omega^2 R^2}{R} = \omega^2 R$$

$$a_c = \omega^2 R \qquad (3.46)$$

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, $v = 2\pi R/T = 2\pi R\nu$ (3.47)

In terms of frequency ν, we have

$$\omega = 2\pi\nu$$

$$v = 2\pi R\nu$$

$$a_c = 4\pi^2\nu^2 R \qquad (3.48)$$

▶ ***Example 3.9*** 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

$$\omega = 2\pi/T = 2\pi \;\; 7/100 = 0.44 \text{ rad/s}$$

The linear speed v is :

$$v = \omega R = 0.44 \text{ s}^{-1} \;\; 12 \text{ cm} = 5.3 \text{ cm s}^{-1}$$

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 1 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 $\overline{\mathbf{v}} = \dfrac{\Delta\mathbf{r}}{\Delta t}$. 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 :

$$h_m = \frac{(v_o \sin\theta_o)^2}{2g}$$

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 (3.33a) and (3.34a) 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

3.1 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.

3.2 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.

3.3 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.

3.4 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.

3.5 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.

3.6 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) $|\mathbf{a}-\mathbf{b}| \le |\mathbf{a}| + |\mathbf{b}|$

(d) $|\mathbf{a}-\mathbf{b}| \ge ||\mathbf{a}| - |\mathbf{b}||$

When does the equality sign above apply?

3.7 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 ?

3.8 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. 3.19. What is the magnitude of the displacement vector for each ? For which girl is this equal to the actual length of path skate ?

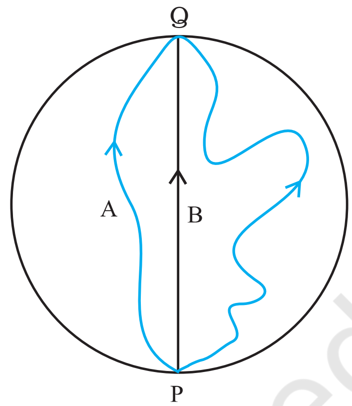


Fig. 3.19

3.9 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. 3.20. If the round trip takes 10 min, what is the (a) net displacement, (b) average velocity, and (c) average speed of the cyclist ?

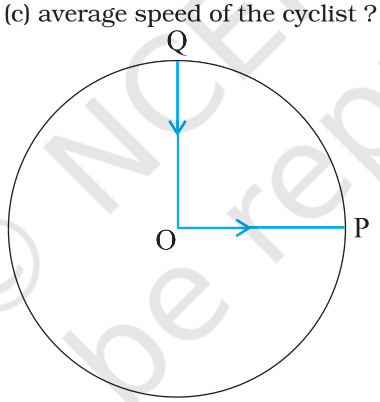


Fig. 3.20

3.10 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.

3.11 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 ?

3.12 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 ?

3.13 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 ?

3.14 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 ?

3.15 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.

3.16 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

3.17 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 ?

3.18 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 ?

3.19 $\hat{\mathbf{i}}$ 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]

3.20 For any arbitrary motion in space, which of the following relations are true :

(a) $\mathbf{v}_{\text{average}} = (1/2)\,(\mathbf{v}\,(t_1) + \mathbf{v}\,(t_2))$

(b) $\mathbf{v}_{\text{average}} = [\mathbf{r}(t_2) - \mathbf{r}(t_1)\,] / (t_2 - t_1)$

(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) $\mathbf{a}_{\text{average}} = [\,\mathbf{v}\,(t_2) - \mathbf{v}\,(t_1)\,] / (\,t_2 - t_1)$

(The 'average' stands for average of the quantity over the time interval t_1 to t_2)

3.21 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.

3.22 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 a part is 30°, wat is the speed of the aircraft ?

11086CH05

Chapter Four

Laws of Motion

4.1 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?**

4.2 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.

4.3 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. 4.1(a)).

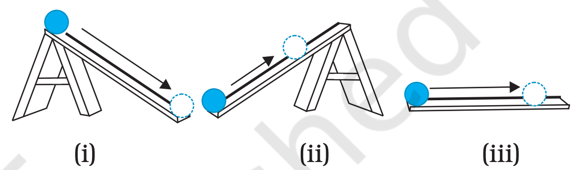


Fig. 4.1(a)

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. 4.1(b)).

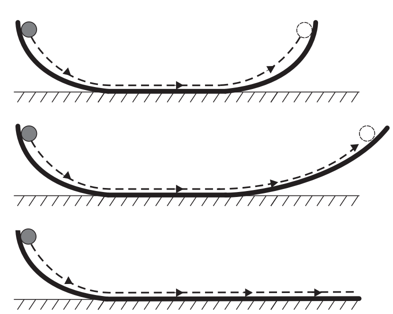


Fig. 4.1(b) *The law of inertia was inferred by Galileo from observations of motion of a ball on a double inclined plane.*

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.

4.4 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.

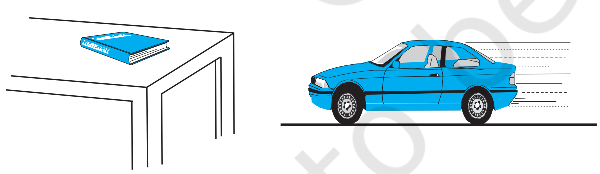
The state of rest or uniform linear motion both imply zero acceleration. The first law of motion can, therefore, be simply expressed as:

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.

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. (4.2(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* ".



(a) (b)

Fig. 4.2 (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. (4.2(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 4.9). 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 4.1 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. ◀

4.5 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 of a body is defined to be the product of its mass m and velocity $\mathbf{v}$, and is denoted by $\mathbf{p}$:

$$\mathbf{p} = m\,\mathbf{v} \tag{4.1}$$

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. 4.3). 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.

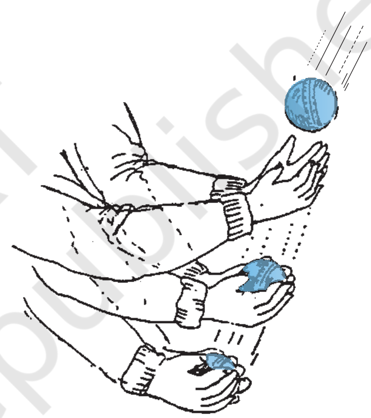


Fig. 4.3 *Force not only depends on the change in momentum but also on how fast the change is brought about. A seasoned cricketer draws in his hands during a catch, allowing greater time for the ball to stop and hence requires a smaller force.*

- 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. 4.4). 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.

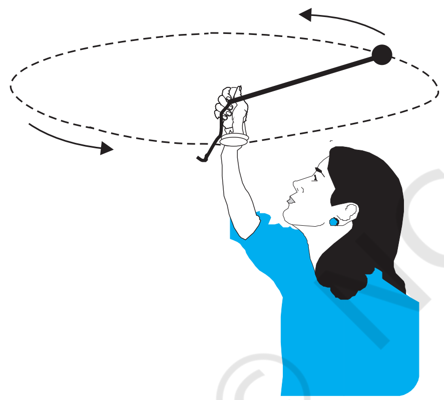


Fig. 4.4 *Force is necessary for changing the direction of momentum, even if its magnitude is constant. We can feel this while rotating a stone in a horizontal circle with uniform speed by means of a string.*

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 $\mathbf{F}$ 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 $\mathbf{p} = m\,\mathbf{v}$ changes by $\Delta\mathbf{p} = m\Delta\mathbf{v}$. 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

$$\mathbf{F} = k\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} \tag{4.2}$$

For a body of fixed mass m,

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}(m\,\mathbf{v}) = m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = m\,\mathbf{a} \tag{4.3}$$

i.e the Second Law can also be written as

$$\mathbf{F} = k\,m\,\mathbf{a} \tag{4.4}$$

which shows that force is proportional to the product of mass m and acceleration $\mathbf{a}$.

The unit of force has not been defined so far. In fact, we use Eq. (4.4) 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

$$\mathbf{F} = \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = m\,\mathbf{a} \tag{4.5}$$

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, $\mathbf{F} = 0$ implies $\mathbf{a} = 0$. 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 :

$$F_x = \frac{\mathrm{d}p_x}{\mathrm{d}t} = ma_x$$

$$F_y = \frac{\mathrm{d}p_y}{\mathrm{d}t} = ma_y$$

$$F_z = \frac{\mathrm{d}p_z}{\mathrm{d}t} = m\,a_z \tag{4.6}$$

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. 4.5).

3. The second law of motion given by Eq. (4.5) 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 about which we shall study in detail in Chapter 6. **Any internal forces in the system are not to be included in F**.

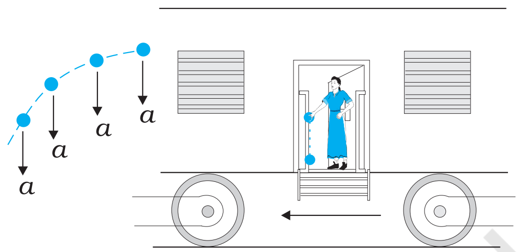


Fig. 4.5 *Acceleration at an instant is determined by the force at that instant. The moment after a stone is dropped out of an accelerated train, it has no horizontal acceleration or force, if air resistance is neglected. The stone carries no memory of its acceleration with the train a moment ago.*

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. 4.5)**.

Example 4.2 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

$$= 0.04 \text{ kg} \quad 6750 \text{ m s}^{-2} = 270 \text{ N}$$

The actual resistive force, and therefore, retardation of the bullet may not be uniform. The answer therefore, only indicates the average resistive force.

Example 4.3 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,

$$v = \frac{dy}{dt} = u + gt$$

$$\text{acceleration, } a = \frac{dv}{dt} = g$$

Then the force is given by Eq. (4.5)

$$F = ma = mg$$

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:

Impulse = Force   time duration
= Change in momentum (4.7)

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 4.4 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

$= 0.15 \times 12 - (-0.15 \times 12)$

$= 3.6$ 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.

4.6 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,

$$\mathbf{F}_{AB} = -\mathbf{F}_{BA} \qquad (4.8)$$

(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 6).

Example 4.5 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. 4.6. 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 ?

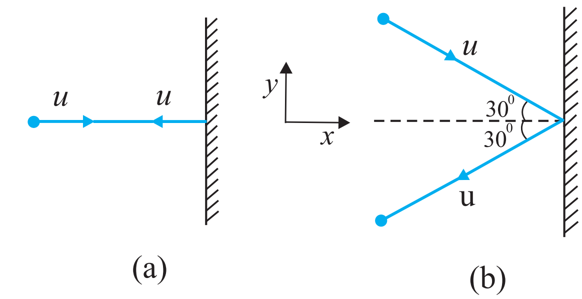


Fig. 4.6

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)

$$\left(p_x\right)_{\text{initial}} = mu \qquad \left(p_y\right)_{\text{initial}} = 0$$

$$\left(p_x\right)_{\text{final}} = -mu \qquad \left(p_y\right)_{\text{final}} = 0$$

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)

$$\left(p_x\right)_{initial} = m\,u\cos 30^\circ,\ \left(p_y\right)_{initial} = -\ m\,u\sin 30^\circ$$

$$\left(p_x\right)_{final} = -\,m\,u\cos 30^\circ,\ \left(p_y\right)_{final} = -\ m\,u\sin 30^\circ$$

Note, while p_x changes sign after collision, p_y does not. Therefore,

x-component of impulse = $-2\ m\,u\ \cos\ 30$
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

$$2\,m\,u/\left(2\,m\,u\cos 30^\circ\right) = \frac{2}{\sqrt{3}} \approx 1.2$$

4.7 CONSERVATION OF MOMENTUM

The second and 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 $-$**F**, according to the third law. The two forces act for a common interval of time Δt. According to the second law, **F** Δt is the change in momentum of the bullet and $-$**F** Δt 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. $$\mathbf{p}'_A + \mathbf{p}'_B = \mathbf{p}_A + \mathbf{p}_B \qquad (4.9)$$

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 5).

4.8 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

$$\mathbf{F}_1 = -\mathbf{F}_2 \qquad (4.10)$$

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.

$$\mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3 = 0 \qquad (4.11)$$

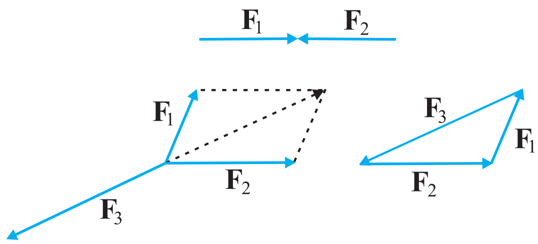


Fig. 4.7 Equilibrium under concurrent forces.

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. 4.7, 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 (4.11) implies that

$$F_{1x} + F_{2x} + F_{3x} = 0$$
$$F_{1y} + F_{2y} + F_{3y} = 0$$
$$F_{1z} + F_{2z} + F_{3z} = 0 \qquad (4.12)$$

where F_{1x}, F_{1y} and F_{1z} are the components of F_1 along x, y and z directions respectively.

▶ ***Example 4.6*** See Fig. 4.8. 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.

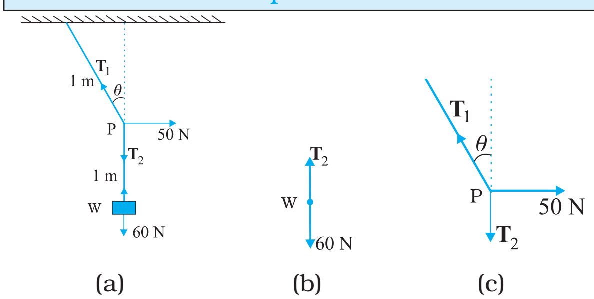


Fig. 4.8

* *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 6.*

Answer Figures 4.8(b) and 4.8(c) are known as free-body diagrams. Figure 4.8(b) is the free-body diagram of W and Fig. 4.8(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. ◀

4.9 COMMON FORCES IN MECHANICS

In mechanics, we encounter several kinds of forces. The gravitational force is, of course, 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. 4.9).

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.

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

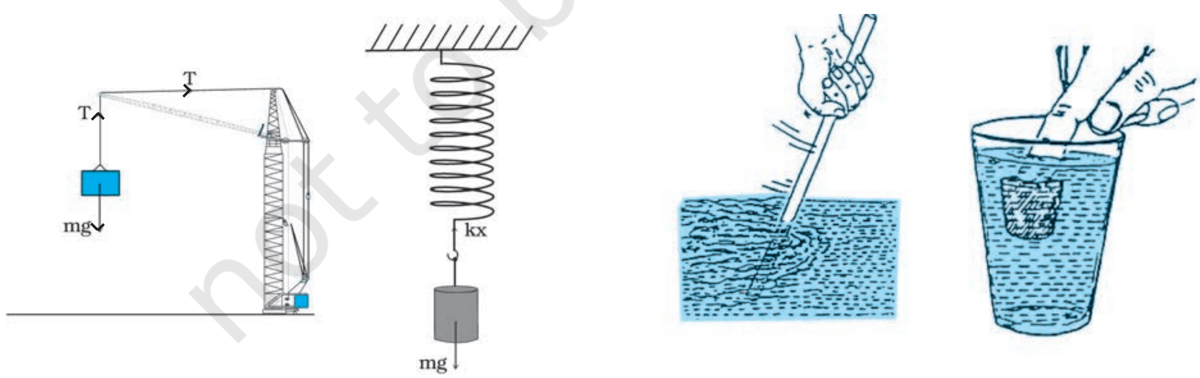


Fig. 4.9 *Some examples of contact forces in mechanics.*

* *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.

4.9.1 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. 4.10(a)). The subscript stands for static friction to distinguish it from kinetic friction f_k that we consider later (Fig. 4.10(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.

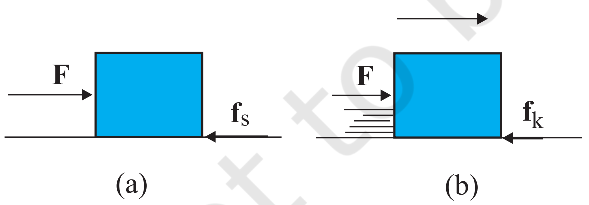


Fig. 4.10 *Static and sliding friction: (a) Impending motion of the body is opposed by static friction. When external force exceeds the maximum limit of static friction, the body begins to move. (b) Once the body is in motion, it is subject to sliding or kinetic friction which opposes relative motion 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 :

$$(f_s)_{\max} = \mu_s N \tag{4.13}$$

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

$$f_s \le \mu_s N \tag{4.14}$$

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 $\mathbf{f}_k$. 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:

$$\mathbf{f}_k = \mu_k \mathbf{N} \tag{4.15}$$

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 4.7 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 4.8 See Fig. 4.11. 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 ?

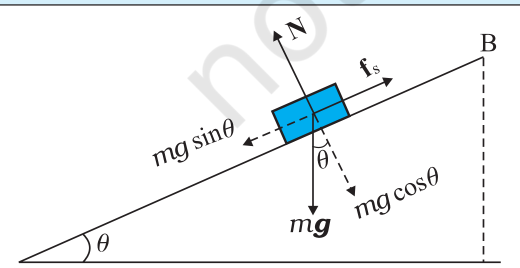


Fig. 4.11

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

$$m g \sin\theta = f_s \quad , \quad m g \cos\theta = N$$

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,

$$\tan\theta_{max} = \mu_s \quad or \quad \theta_{max} = \tan^{-1} \mu_s$$

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 4.9 What is the acceleration of the block and trolley system shown in a Fig. 4.12(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.

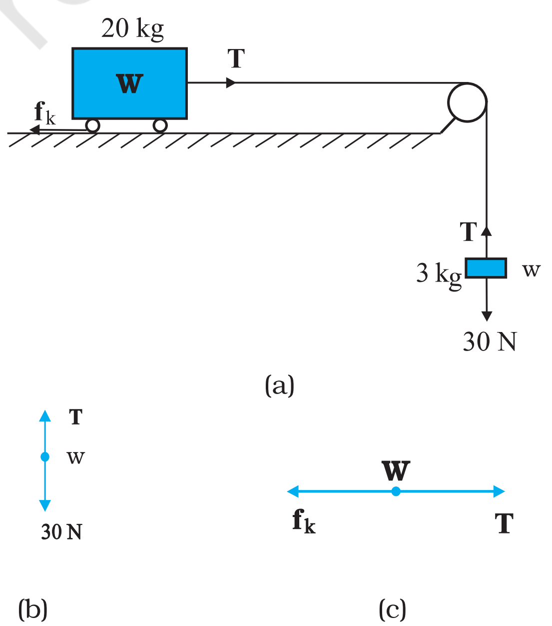


Fig. 4.12

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. 4.12(b)),

$$30 - T = 3a$$

Apply the second law to motion of the trolley (Fig. 4.12(c)),

$$T - f_k = 20\ a.$$

Now $\quad f_k = \mu_k N,$

Here $\quad \mu_k = 0.04,$

$$N = 20 \times 10 = 200\ \text{N}.$$

Thus the equation for the motion of the trolley is

$T - 0.04 \times 200 = 20\ a \qquad$ Or $\quad T - 8 = 20a.$

These equations give $a = \frac{22}{23}\ \text{m s}^{-2} = 0.96\ \text{m s}^{-2}$ 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. 4.13(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. 4.13(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.

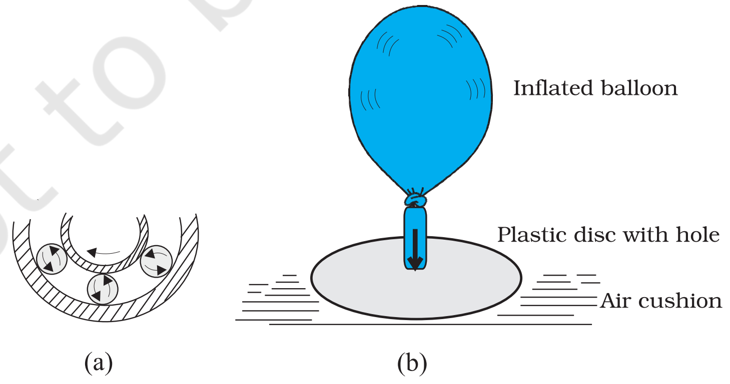


Fig. 4.13 *Some ways of reducing friction. (a) Ball bearings placed between moving parts of a machine. (b) Compressed cushion of air between surfaces in relative motion.*

4.10 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_c providing this acceleration is :

$$f_c = \frac{mv^2}{R} \qquad (4.16)$$

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. 4.14(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

$N - mg = 0$

$N = mg \qquad (4.17)$

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 (4.14) & (4.16) we get the result

$$f = \frac{mv^2}{R} \le \mu_s N$$

$$v^2 \le \frac{\mu_s RN}{m} = \mu_s Rg \qquad [\because N = mg]$$

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

$$v_{max} = \sqrt{\mu_s Rg} \qquad (4.18)$$

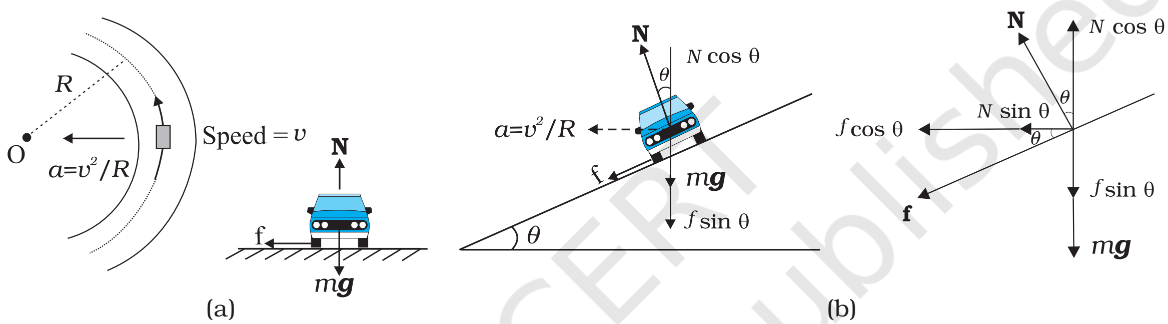


Fig. 4.14 Circular motion of a car on (a) a level road, (b) a banked road.

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. 4.14(b)). Since there is no acceleration along the vertical direction, the net force along this direction must be zero. Hence,

$$N \cos\theta = mg + f \sin\theta \qquad (4.19a)$$

The centripetal force is provided by the horizontal components of N and f.

$$N \sin\theta + f \cos\theta = \frac{mv^2}{R} \qquad (4.19b)$$

But $f \le \mu_s N$

Thus to obtain v_{max} we put

$f = \mu_s N$.

Then Eqs. (4.19a) and (4.19b) become

$$N \cos\theta = mg + \mu_s N \sin\theta \qquad (4.20a)$$

$N\sin\theta + \mu_s N\cos\theta = mv^2/R$ (4.20b)

From Eq. (4.20a), we obtain

$$N = \frac{mg}{\cos\theta - \mu_s \sin\theta}$$

Substituting value of N in Eq. (4.20b), we get

$$\frac{mg(\sin\theta + \mu_s\ \cos\theta)}{\cos\theta - \mu_s\ \sin\theta} = \frac{mv_{max}^2}{R}$$

$$\text{or } v_{max} = \left(Rg\frac{\mu_s + \tan\theta}{1 - \mu_s \tan\theta} \right)^{\frac{1}{2}} \qquad (4.21)$$

Comparing this with Eq. (4.18) 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. (4.21),

$$v_o = (R g\ \tan\theta)^{1/2} \qquad (4.22)$$

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 4.10*** 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. (4.18) :

$$v^2 \le \mu_s R g$$

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 4.11*** 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. (4.22):

$$v_o = (R\, g \tan\theta)^{1/2}$$

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. (4.21):

$$v_{max} = \left(R\, g\frac{\mu_s + \tan\theta}{1 - \mu_s \tan\theta} \right)^{1/2} = 38.1 \text{ m s}^{-1}$$ ◀

4.11 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 4.12*** See Fig. 4.15. 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 $2 \times 10 = 20$ 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,

$$270 - R' = 27 \quad 0.1\text{N}$$
$$\text{ie. } R' = 267.3 \text{ N}$$

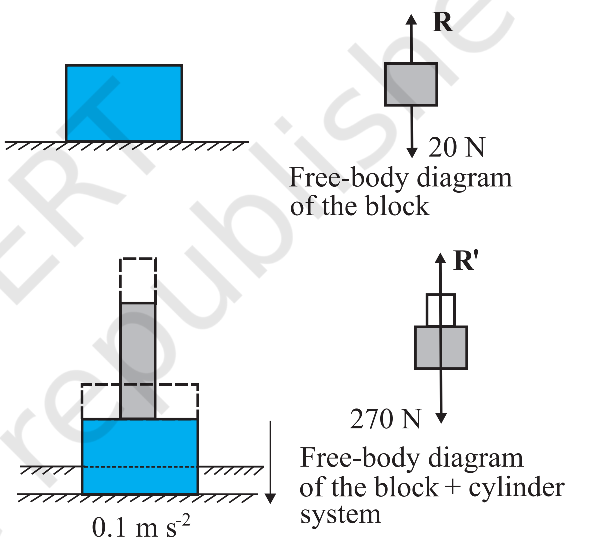


Fig. 4.15

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 (**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

$$\mathbf{F} = k\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = k\,m\,\mathbf{a}$$

where **F** is the net external force on the body and **a** its acceleration. We set the constant of proportionality k = 1 in SI units. Then

$$\mathbf{F} = \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = m\mathbf{a}$$

The SI unit of force is newton : 1 N = 1 kg m s^{-2}.

(a) The second law is consistent with the First Law (**F** = 0 implies **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 **F** is the total external force on the system and **a** is the acceleration of the system as a whole.

(d) **F** at a point at a certain instant determines **a** at the same point at that instant. That is the Second Law is a local law; **a** at an instant does not depend on the history of motion.

4. 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** = m **a** Second Law
Impulse		kg m s^{-1} or N s	[M LT^{-1}]	Impulse = force × time = change in momentum
Static friction	$\mathbf{f}_s$	N	[MLT^{-2}]	$\mathbf{f}_s \le \mu_s \mathbf{N}$
Kinetic friction	$\mathbf{f}_k$	N	[MLT^{-2}]	$\mathbf{f}_k = \mu_k \mathbf{N}$

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 $\mu_s N$ ($f_s \le \mu_s N$). 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 $g = 10 \text{ m s}^{-2}$)

4.1 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.

4.2 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.

4.3 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.

4.4 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].

4.5 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 ?

4.6 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 ?

4.7 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.

4.8 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.

4.9 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.

4.10 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.

4.11 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.)

4.12 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.

4.13 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 ?

4.14 Figure 4.16 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).

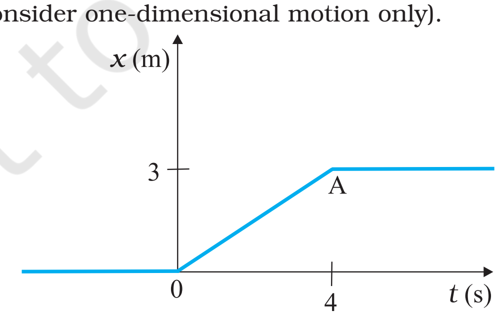


Fig. 4.16

4.15 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?

4.16 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.

4.17 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.

4.18 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 ?

4.19 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 ?

4.20 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.)

4.21 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 ?

4.22 If, in Exercise 4.21, 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 ?

4.23 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.

11086CH06

Chapter Five

Work, Energy and Power

5.1 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.

5.1.1 The Scalar Product

We have learnt about vectors and their use in Chapter 3. 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 6. 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

$$\mathbf{A}\cdot\mathbf{B} = A\,B\cos\theta \qquad (5.1a)$$

where θ is the angle between the two vectors as shown in Fig. 5.1(a). Since A, B and cos θ 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. (5.1a), 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.5.1 (b) and $A\cos\theta$ is the projection of **A** onto **B** in Fig. 5.1 (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 (5.1a) 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

$$\mathbf{A}\cdot\mathbf{B} = \left(A_x\hat{\mathbf{i}} + A_y\hat{\mathbf{j}} + A_z\hat{\mathbf{k}}\right)\cdot\left(B_x\hat{\mathbf{i}} + B_y\hat{\mathbf{j}} + B_z\hat{\mathbf{k}}\right)$$
$$= A_xB_x + A_yB_y + A_zB_z \qquad (5.1b)$$

From the definition of scalar product and (Eq. 5.1b) we have :

(i) $\mathbf{A}\cdot\mathbf{A} = A_xA_x + A_yA_y + A_zA_z$

Or, $A^2 = A_x^2 + A_y^2 + A_z^2 \qquad (5.1c)$

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 5.1*** Find the angle between force $\mathbf{F} = (3\hat{\mathbf{i}} + 4\hat{\mathbf{j}} - 5\hat{\mathbf{k}})$ unit and displacement $\mathbf{d} = (5\hat{\mathbf{i}} + 4\hat{\mathbf{j}} + 3\hat{\mathbf{k}})$ unit. Also find the projection of **F** on **d**.

Answer $\mathbf{F}\cdot\mathbf{d} = F_xd_x + F_yd_y + F_zd_z$
$= 3\,(5) + 4\,(4) + (-5)\,(3)$
$= 16$ 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

$$\therefore \cos\theta = \frac{16}{\sqrt{50}\sqrt{50}} = \frac{16}{50} = 0.32,$$

$$\theta = \cos^{-1} 0.32$$

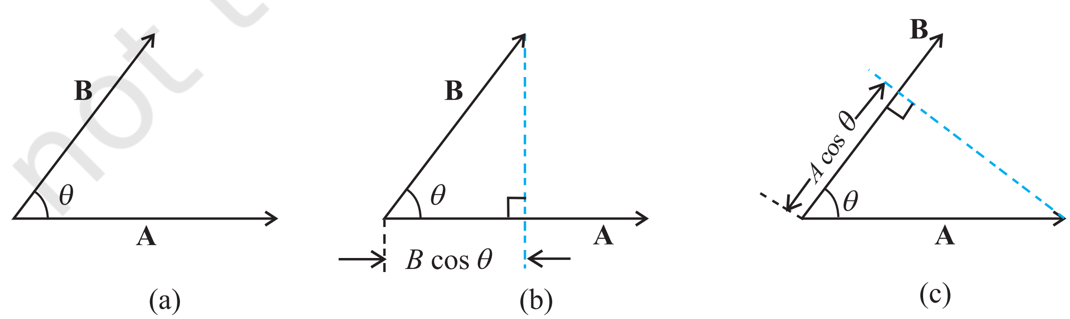


Fig. 5.1 (a) *The scalar product of two vectors* **A** *and* **B** *is a scalar :* $\mathbf{A}\cdot\mathbf{B} = A\,B\cos\theta$. (b) $B\cos\theta$ *is the projection of* **B** *onto* **A**. (c) $A\cos\theta$ *is the projection of* **A** *onto* **B**.

5.2 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,

$$v^2 - u^2 = 2\,as \tag{5.2}$$

where u and v are the initial and final speeds and s the distance traversed. Multiplying both sides by $m/2$, we have

$$\frac{1}{2}mv^2 - \frac{1}{2}mu^2 = mas = Fs \tag{5.2a}$$

where the last step follows from Newton's Second Law. We can generalise Eq. (5.2) to three dimensions by employing vectors

$$v^2 - u^2 = 2\,\mathbf{a.d}$$

Here **a** and **d** are acceleration and displacement vectors of the object respectively.
Once again multiplying both sides by $m/2$, we obtain

$$\frac{1}{2}mv^2 - \frac{1}{2}mu^2 = m\,\mathbf{a.d} = \mathbf{F.d} \tag{5.2b}$$

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. (5.2b) is then

$$K_f - K_i = W \tag{5.3}$$

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 (5.2) 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 5.2 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.

5.3 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. 5.2.

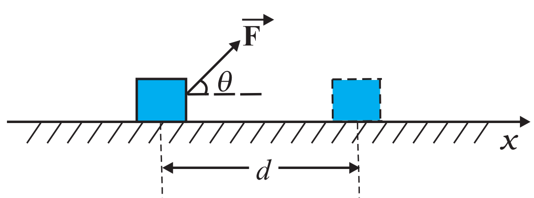


Fig. 5.2 *An object undergoes a displacement* **d** *under the influence of the force* **F**.

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

$$W = (F\cos\theta)d = \mathbf{F.d} \tag{5.4}$$

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. (5.4) 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. (5.4) 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 5.1.

Table 5.1 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 5.3*** 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 Example 5.3 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.

5.4 KINETIC ENERGY

As noted earlier, if an object of mass m has velocity $\mathbf{v}$, its kinetic energy K is

$$K = \frac{1}{2}m\,\mathbf{v.v} = \frac{1}{2}mv^2 \tag{5.5}$$

Kinetic energy is a scalar quantity. The kinetic energy of an object is a measure of the work an

Table 5.2 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	$\simeq10^{-26}$	500	$\simeq10^{-21}$

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 5.2 lists the kinetic energies for various objects.

▶ ***Example 5.4*** In a ballistics demonstration a police officer fires a bullet of mass 50.0 g with speed 200 m s^{-1} (see Table 5.2) 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%). ◀

5.5 WORK DONE BY A VARIABLE FORCE

A constant force is rare. It is the variable force, which is more commonly encountered. Fig. 5.3 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. 5.3(a). Adding successive rectangular areas in Fig. 5.3(a) we get the total work done as

$$W \cong \sum_{x_i}^{x_f} F(x)\Delta x \tag{5.6}$$

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. 5.3(b). Then the work done is

$$W = \lim_{\Delta x \to 0} \sum_{x_i}^{x_f} F(x)\Delta x$$

$$= \int_{x_i}^{x_f} F(x)\,\mathrm{d}x \tag{5.7}$$

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).

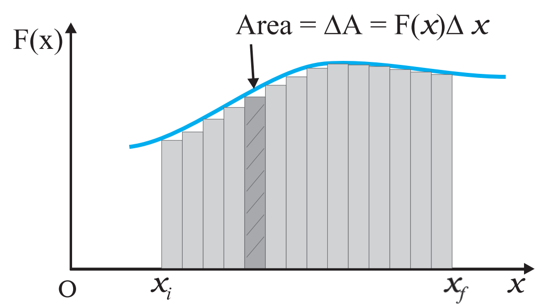
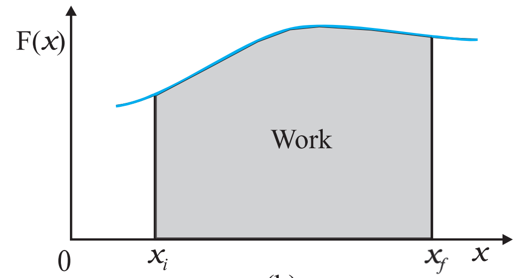


Fig. 5.3(a)



(b)

Fig. 5.3 *(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 5.5 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 versus displacement. Calculate the work done by the two forces over 20 m.

Answer

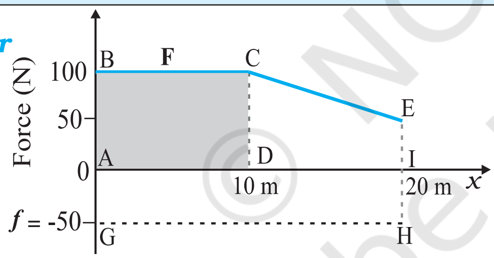


Fig. 5.4 *Plot of the force F applied by the woman and the opposing frictional force f versus* displacement.

The plot of the applied force is shown in Fig. 5.4. At x = 20 m, F = 50 N ($\neq 0$). 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

$W_F \rightarrow$ area of the rectangle ABCD + area of the trapezium CEID

$$W_F = 100 \times 10 + \frac{1}{2}(100 + 50) \times 10$$
$$= 1000 + 750$$
$$= 1750 \text{ J}$$

The work done by the frictional force is

$W_f \rightarrow$ 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. ◀

5.6 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

$$\frac{dK}{dt} = \frac{d}{dt}\left(\frac{1}{2} m\, v^2\right)$$

$$= m \frac{dv}{dt} v$$

$$= F\, v \quad \text{(from Newton's Second Law)}$$

$$= F \frac{dx}{dt}$$

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 $$K_f - K_i = \int_{x_i}^{x_f} F dx \qquad (5.8a)$$

From Eq. (5.7), it follows that

$$K_f - K_i = W \qquad (5.8b)$$

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 5.6*** A block of mass $m = 1$ kg, moving on a horizontal surface with speed $v_i = 2 \text{ m s}^{-1}$ 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,

$$F_r = \frac{-k}{x} \text{ for } 0.1 < x < 2.01 \text{ m}$$

$= 0$ for $x < 0.1$m and $x > 2.01$ 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. (5.8a)

$$K_f = K_i + \int_{0.1}^{2.01} \frac{(-k)}{x} \mathrm{d}x$$

$$= \frac{1}{2} m v_i^2 - k \ln(x) \Big|_{0.1}^{2.01}$$

$$= \frac{1}{2} m v_i^2 - k \ln(2.01/0.1)$$

$$= 2 - 0.5 \ln(20.1)$$

$$= 2 - 1.5 = 0.5 \text{ J}$$

$$v_f = \sqrt{2K_f / m} = 1 \text{ m s}^{-1}$$

Here, note that ln is a symbol for the natural logarithm to the base e and not the logarithm to the base 10 [$\ln X = \log_e X = 2.303 \log_{10} X$]. ◀

5.7 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 ($h \ll R_E$) 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 7 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

$$\Delta V = -F(x)\,\Delta x \tag{5.9}$$

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.

5.8 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

$$\Delta K + \Delta V = 0$$

$$\Delta(K + V) = 0 \tag{5.10}$$

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

$$K_i + V(x_i) = K_f + V(x_f) \tag{5.11}$$

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. (5.9). 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. (5.11) 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. 5.5 depicts a ball of mass m being dropped from a cliff of height H.

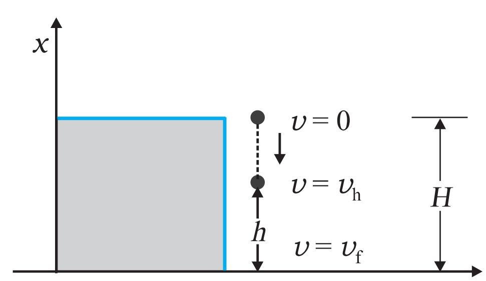


Fig. 5.5 *The conversion of potential energy to kinetic energy for a ball of mass m dropped from a height H.*

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

$$E_H = mgH \tag{5.11 a}$$

$$E_h = mgh + \frac{1}{2}mv_h^2 \tag{5.11 b}$$

$$E_0 = (1/2)\, mv_f^2 \tag{5.11 c}$$

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 5.7 for a freely falling body.

Further,

$$E_H = E_h$$

which implies,

$$v_h^2 = 2g(H-h) \tag{5.11 d}$$

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 5.7*** 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. 5.6. 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.

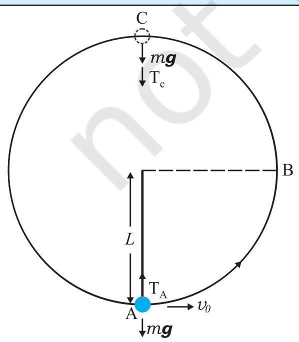


Fig. 5.6

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 :

$$E = \frac{1}{2}mv_0^2 \tag{5.12}$$

$$T_A - mg = \frac{mv_0^2}{L} \text{ [Newton's Second Law]}$$

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

$$E = \frac{1}{2}mv_c^2 + 2mgL \tag{5.13}$$

$$mg = \frac{mv_c^2}{L} \quad \text{[Newton's Second Law]} \tag{5.14}$$

where v_C is the speed at C. From Eqs. (5.13) and (5.14)

$$E = \frac{5}{2}mgL$$

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. (5.14)

$$v_C = \sqrt{gL}$$

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. ◀

5.9 THE POTENTIAL ENERGY OF A SPRING

The spring force is an example of a variable force which is conservative. Fig. 5.7 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. 5.7(b)] or negative [Fig. 5.7(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. 5.7(b). If the extension is x_m, the work done by the spring force is

$$W_s = \int_0^{x_m} F_s \, dx = -\int_0^{x_m} kx \, dx$$

$$= -\frac{k x_m^2}{2} \qquad (5.15)$$

This expression may also be obtained by considering the area of the triangle as in Fig. 5.7(d). Note that the work done by the external pulling force F is positive since it overcomes the spring force.

$$W = +\frac{k x_m^2}{2} \qquad (5.16)$$

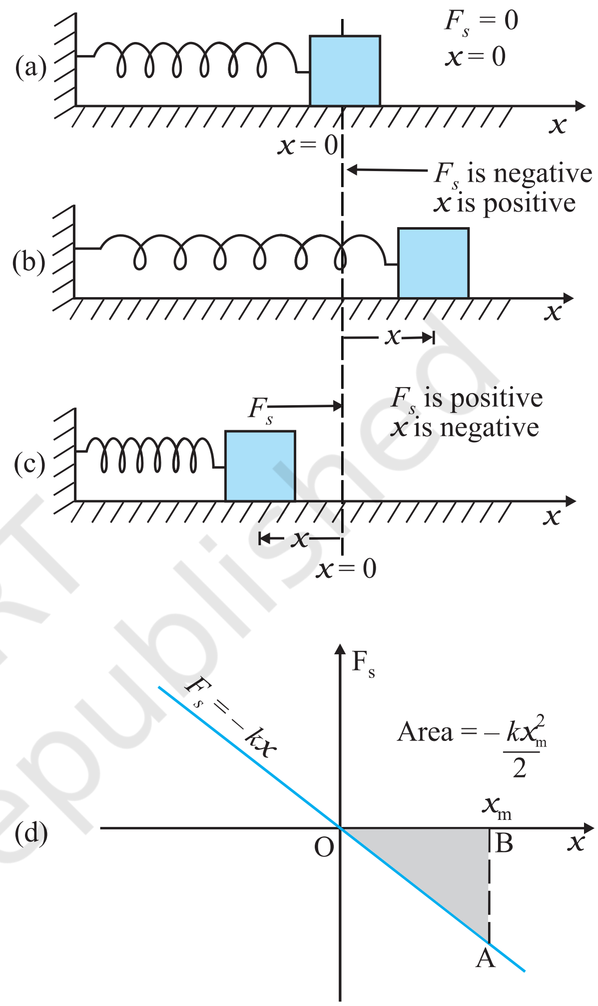


Fig. 5.7 *Illustration of the spring force with a block attached to the free end of the spring. (a) The spring force F_s is zero when the displacement x from the equilibrium position is zero. (b) For the stretched spring $x > 0$ and $F_s < 0$ (c) For the compressed spring $x < 0$ and $F_s > 0$.(d) The 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

$$W_s = -\int_{x_i}^{x_f} k\,x\,\mathrm{d}x \quad = \frac{k\,x_i^2}{2} - \frac{k\,x_f^2}{2} \qquad (5.17)$$

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;

$$W_s = -\int_{x_i}^{x_i} k\,x\,\mathrm{d}x \quad = \frac{k\,x_i^2}{2} - \frac{k\,x_i^2}{2}$$

$$= 0 \qquad (5.18)$$

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. (5.17). 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

$$V(x) = \frac{kx^2}{2} \qquad (5.19)$$

You may easily verify that $-\,\mathrm{d}V/\mathrm{d}x = -\,k\,x$, the spring force. If the block of mass m in Fig. 5.7 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. 5.8.

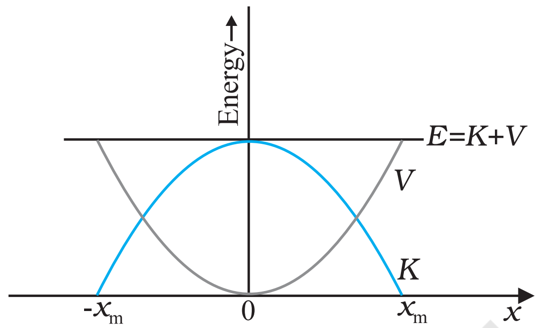


Fig. 5.8 *Parabolic plots of the potential energy V and kinetic energy K of a block attached to a spring obeying Hooke's law. The two plots are complementary, one decreasing as the other increases. The total mechanical energy E = K + V remains constant.*

▶ ***Example 5.8*** 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 5.25×10^3 N m^{-1}. What is the maximum compression of the spring ?

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 5.9.

(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 5.9*** Consider Example 5.8 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. 5.9.

We invoke the work-energy theorem, rather than the conservation of mechanical energy.

The change in kinetic energy is

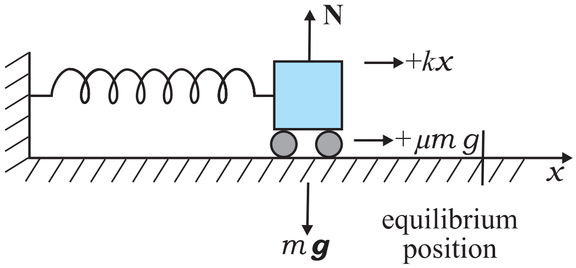


Fig. 5.9 *The forces acting on the car.*

$$\Delta K = K_f - K_i = 0 - \frac{1}{2} m v^2$$

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 $g = 10.0 \text{ m s}^{-2}$). 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 5.8.

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. ◀

5.10 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,

$$P = \frac{\mathrm{d}W}{\mathrm{d}t} \tag{5.20}$$

The work dW done by a force F for a displacement $d\mathbf{r}$ is $\mathrm{d}W = \mathbf{F}.d\mathbf{r}$. The instantaneous power can also be expressed as

$$P = \mathbf{F}.\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}$$

$$= \mathbf{F.v} \tag{5.21}$$

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 5.10*** 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,

$P = \mathbf{F}.\ \mathbf{v} = 22000 \times 2 = 44000$ W = 59 hp ◀

5.11 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. 5.10.

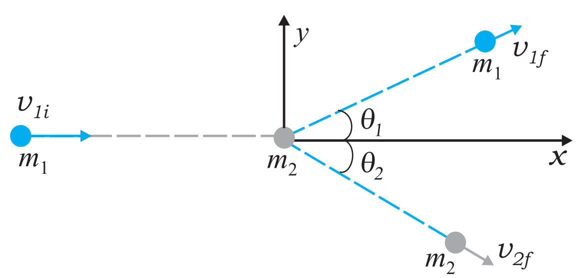


Fig. 5.10 *Collision of mass m_1, with a stationary mass m_2.*

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.

5.11.1 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**.

5.11.2 Collisions in One Dimension

Consider first a **completely inelastic collision** in one dimension. Then, in Fig. 5.10,

$$\theta_1 = \theta_2 = 0$$

$$m_1 v_{1i} = (m_1 + m_2) v_f \quad \text{(momentum conservation)}$$

$$v_f = \frac{m_1}{m_1 + m_2} v_{1i} \tag{5.22}$$

The loss in kinetic energy on collision is

$$\Delta K = \frac{1}{2} m_1 v_{1i}^2 - \frac{1}{2}(m_1 + m_2) v_f^2$$

$$= \frac{1}{2} m_1 v_{1i}^2 - \frac{1}{2} \frac{m_1^2}{m_1 + m_2} v_{1i}^2 \quad \text{[using Eq. (5.22)]}$$

$$= \frac{1}{2} m_1 v_{1i}^2 \left[1 - \frac{m_1}{m_1 + m_2}\right]$$

$$= \frac{1}{2} \frac{m_1 m_2}{m_1 + m_2} v_{1i}^2$$

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

$$m_1 v_{1i} = m_1 v_{1f} + m_2 v_{2f} \tag{5.23}$$

$$m_1 v_{1i}^2 = m_1 v_{1f}^2 + m_2 v_{2f}^2 \tag{5.24}$$

From Eqs. (5.23) and (5.24) 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, $$\therefore v_{2f} = v_{1i} + v_{1f} \tag{5.25}$$

Substituting this in Eq. (5.23), we obtain

$$v_{1f} = \frac{(m_1 - m_2)}{m_1 + m_2} v_{1i} \tag{5.26}$$

and $$v_{2f} = \frac{2 m_1 v_{1i}}{m_1 + m_2} \tag{5.27}$$

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 5.11*** **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. (5.26)

$$K_{1f} = \frac{1}{2} m_1 v_{1f}^2 \quad = \frac{1}{2} m_1 \left(\frac{m_1 - m_2}{m_1 + m_2} \right)^2 v_{1i}^2$$

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. (5.27).

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.

5.11.3 Collisions in Two Dimensions

Fig. 5.10 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

$$m_1 v_{1i} = m_1 v_{1f} \cos \theta_1 + m_2 v_{2f} \cos \theta_2 \qquad (5.28)$$

$$0 = m_1 v_{1f} \sin \theta_1 - m_2 v_{2f} \sin \theta_2 \qquad (5.29)$$

One knows $\{m_1, m_2, v_{1i}\}$ in most situations. There are thus four unknowns $\{v_{1f}, v_{2f}, \theta_1 \textit{ and } \theta_2\}$, and only two equations. If $\theta_1 = \theta_2 = 0$, we regain Eq. (5.23) for one dimensional collision.

If, further the collision is elastic,

$$\frac{1}{2} m_1 v_{1i}^{\ 2} = \frac{1}{2} m_1 v_{1f}^{\ 2} + \frac{1}{2} m_2 v_{2f}^{\ 2} \qquad (5.30)$$

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. (5.28)-(5.30).

▶ ***Example 5.12*** Consider the collision depicted in Fig. 5.10 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

$$v_{1i}^{\ 2} = \left(\mathbf{v}_{1f} + \mathbf{v}_{2f}\right) \cdot \left(\mathbf{v}_{1f} + \mathbf{v}_{2f}\right)$$

$$= v_{1f}^{\ 2} + v_{2f}^{\ 2} + 2\mathbf{v}_{1f}.\mathbf{v}_{2f}$$

$$= \left\{ v_{1f}^{\ 2} + v_{2f}^{\ 2} + 2v_{1f}v_{2f} \cos\ (\theta_1 + 37°) \right\} \quad (5.31)$$

Since the collision is elastic and $m_1 = m_2$ it follows from conservation of kinetic energy that

$$v_{1i}^{\ 2} = v_{1f}^{\ 2} + v_{2f}^{\ 2} \quad (5.32)$$

Comparing Eqs. (5.31) and (5.32), 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

   $$F(x) = -\frac{dV(x)}{dx}$$

   or $$V_i - V_f = \int_{x_i}^{x_f} F(x)\,dx$$

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.
5. 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 : $\mathbf{A \cdot B} = AB \cos\theta$, 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 5.2 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.

5\. 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

5.1 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.

5.2 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.

5.3 Given in Fig. 5.11 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.

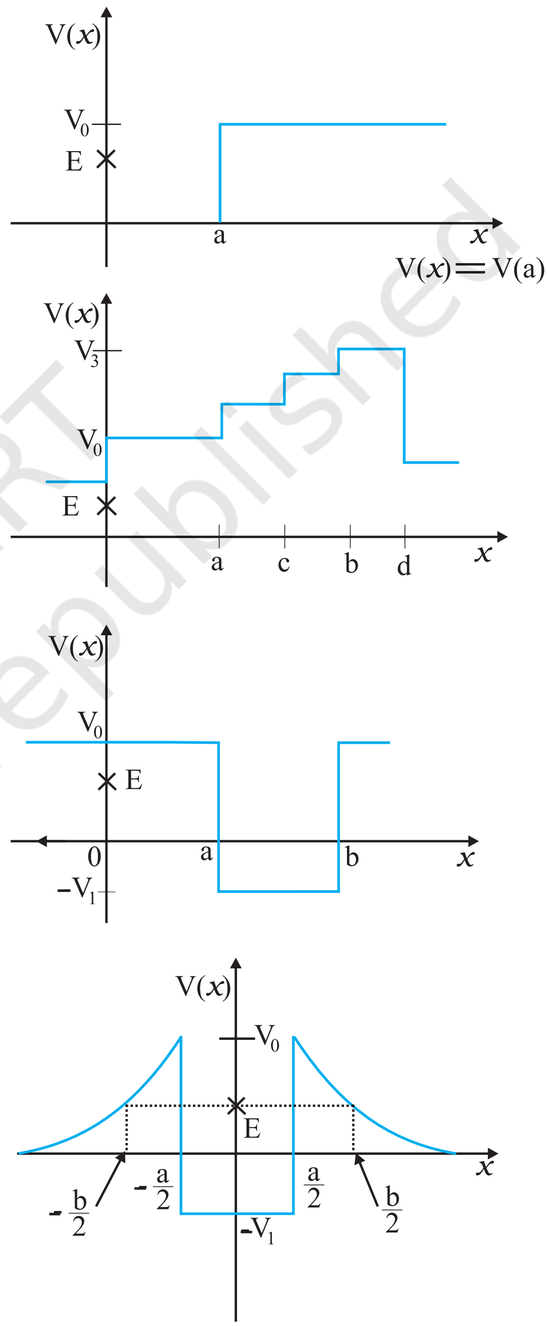


Fig. 5.11

5.4 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. 5.12. Show that a particle of total energy 1 J moving under this potential must 'turn back' when it reaches $x = \pm 2$ m.

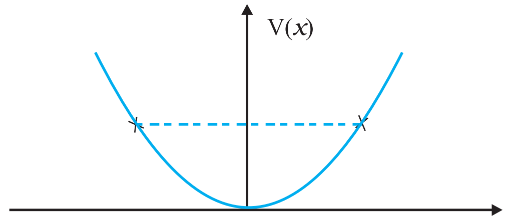
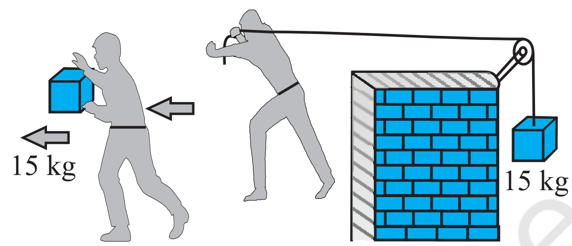


Fig. 5.12

5.5 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 ?



(i) ***Fig. 5.13*** (ii)

(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. 5.13(i) the man walks 2 m carrying a mass of 15 kg on his hands. In Fig. 5.13(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 ?

5.6 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.

5.7 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.

5.8 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).

5.9 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

5.10 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

5.11 A body constrained to move along the z-axis of a coordinate system is subject to a constant force **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 ?

5.12 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).

5.13 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} ?

5.14 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 ?

5.15 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 ?

5.16 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. 5.14) is a possible result after collision ?

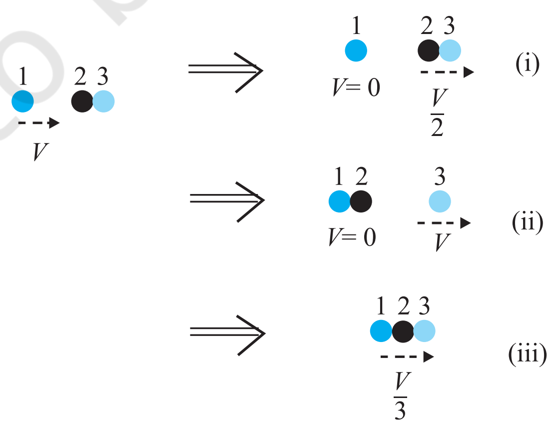


Fig. 5.14

5.17 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. 5.15. How high does the bob A rise after the collision ? Neglect the size of the bobs and assume the collision to be elastic.

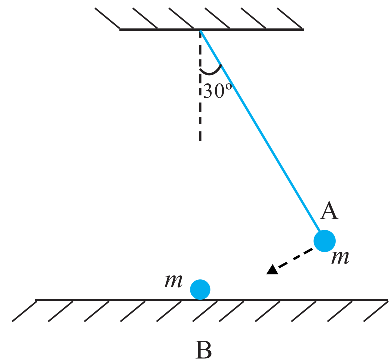


Fig. 5.15

5.18 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 ?

5.19 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 ?

5.20 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 ?

5.21 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 ?

5.22 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?

5.23 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.

11086CH07

CHAPTER SIX

SYSTEMS OF PARTICLES AND ROTATIONAL MOTION

6.1 INTRODUCTION

In the earlier chapters we primarily considered the motion of a single particle. (A particle is ideally represented as a point mass having 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 all pairs of particles 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. 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 (twist out of shape), bend or vibrate and treat them as rigid.

6.1.1 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

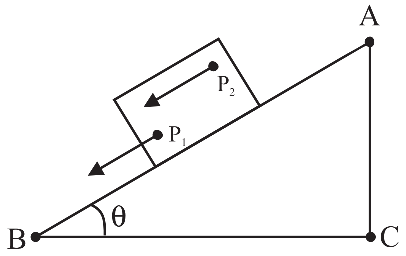


Fig 6.1 *Translational (sliding) motion of a block down an inclined plane.*
(Any point like P_1 or P_2 of the block moves with the same velocity at any instant of time.)

block sliding down an inclined plane without any sidewise movement. The block is taken as 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. 6.1).

In pure translational motion at any instant of time, all particles of the body have the same velocity.

Consider now the rolling motion of a solid metallic or wooden cylinder down the same inclined plane (Fig. 6.2). The rigid body in this problem, namely the cylinder, shifts from the top to the bottom of the inclined plane, and thus, seems to have translational motion. But as Fig. 6.2 shows, all its particles are not moving with the same velocity at any instant. The body, therefore, is not in pure translational motion. Its motion is translational plus 'something else.'

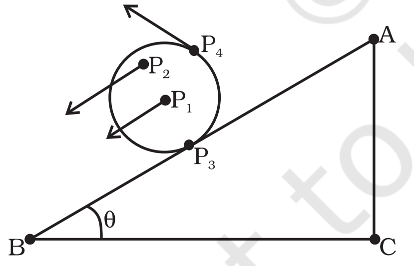


Fig. 6.2 *Rolling motion of a cylinder. It is not pure translational motion. Points P_1, P_2, P_3 and P_4 have different velocities (shown by arrows) at any instant of time. In fact, the velocity of the point of contact P_3 is zero at any instant, if the cylinder rolls without slipping.*

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 or fixed axis about which the body is rotating is 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 6.3(a) and (b)).

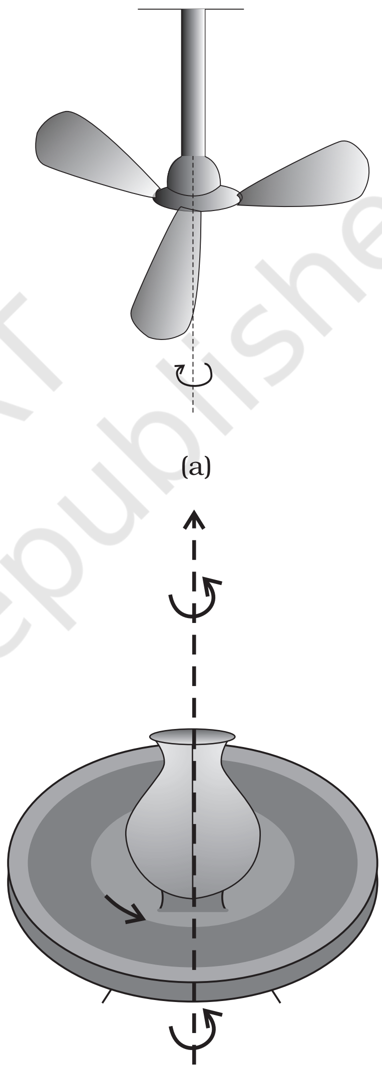


Fig. 6.3 *Rotation about a fixed axis*
(a) A ceiling fan
(b) A potter's wheel.

Let us try to understand what rotation is, what characterises rotation. You may notice that **in rotation of a rigid body about a fixed axis,**

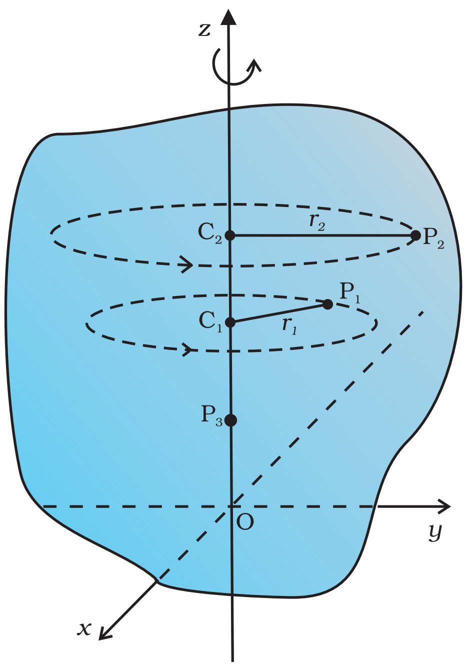


Fig. 6.4 *A rigid body rotation about the z-axis (Each point of the body such as P_1 or P_2 describes a circle with its centre (C_1 or C_2) on the axis of rotation. The radius of the circle (r_1 or r_2) is the perpendicular distance of the point (P_1 or P_2) from the axis. A point on the axis like P_3 remains stationary).*

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. 6.4 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 of rotation is fixed.

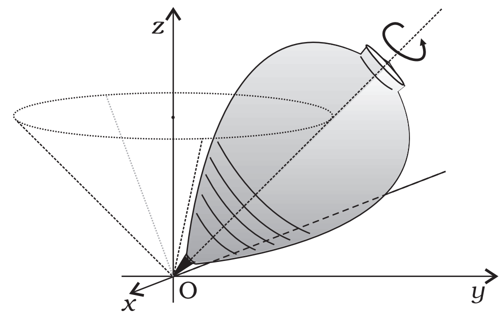


Fig. 6.5 (a) *A spinning top (The point of contact of the top with the ground, its tip O, is fixed.)*

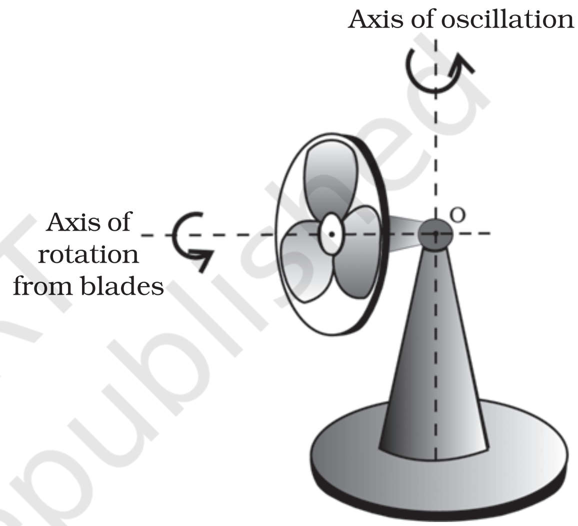


Fig. 6.5 (b) *An oscillating table fan with rotating blades. The pivot of the fan, point O, is fixed. The blades of the fan are under rotational motion, whereas, the axis of rotation of the fan blades is oscillating.*

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. 6.5(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. 6.5(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 [Fig.6.5(b)]. 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. 6.5(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.

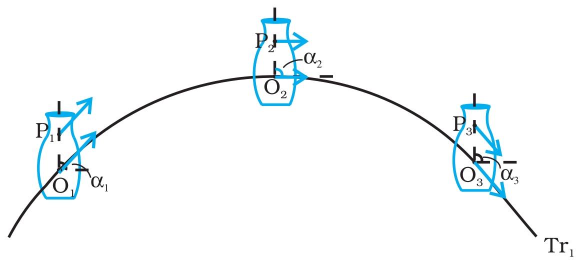


Fig. 6.6(a) *Motion of a rigid body which is pure translation.*

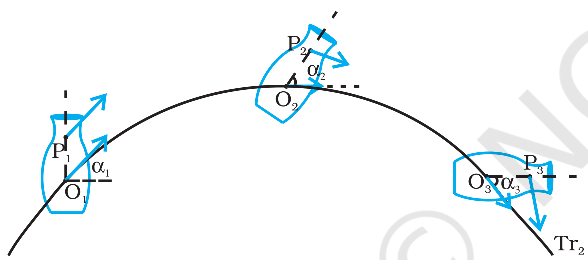


Fig. 6.6(b) *Motion of a rigid body which is a combination of translation and rotation.*

Fig 6.6 (a) and 6.6 (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 both Figs. 6.6 (a) and (b) . As seen from Fig. 6.6(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. 6.6 (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.

Thus, for us rotation will be about a fixed axis only unless stated otherwise.

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. 6.6(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. 6.6(a), the motion is a pure translation; in the other case [Fig. 6.6(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 [Fig.6.5(b)]). We shall, in the present chapter, consider rotational motion about a fixed axis only.

6.2 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.

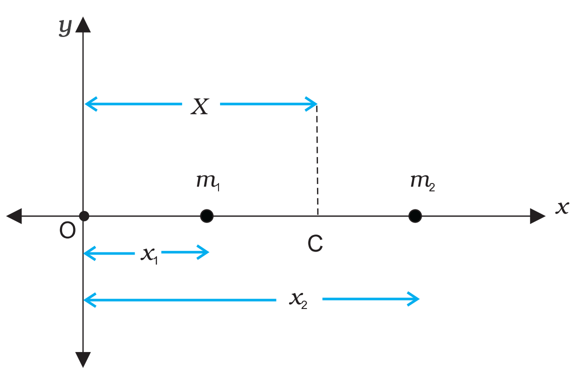


Fig. 6.7

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

$$X = \frac{m_1 x_1 + m_2 x_2}{m_1 + m_2} \tag{6.1}$$

In Eq. (6.1), 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.

$$X = \frac{m_1 x_1 + m_2 x_2 + \ldots + m_n x_n}{m_1 + m_2 + \ldots + m_n} = \frac{\sum_{i=1}^{n} m_i x_i}{\sum_{i=1}^{n} m_i} = \frac{\sum m_i x_i}{\sum m_i} \tag{6.2}$$

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

$$X = \frac{m_1 x_1 + m_2 x_2 + m_3 x_3}{m_1 + m_2 + m_3} \tag{6.3a}$$

$$Y = \frac{m_1 y_1 + m_2 y_2 + m_3 y_3}{m_1 + m_2 + m_3} \tag{6.3b}$$

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. (6.3a) and (6.3b) 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

$$X = \frac{\sum m_i x_i}{M} \tag{6.4a}$$

$$Y = \frac{\sum m_i y_i}{M} \tag{6.4b}$$

and $$Z = \frac{\sum m_i z_i}{M} \tag{6.4c}$$

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 i^{th} particle and the position of the i^{th} particle is given by (x_i, y_i, z_i).

Eqs. (6.4a), (6.4b) and (6.4c) can be combined into one equation using the notation of position vectors. Let $\mathbf{r}_i$ be the position vector of the i^{th} 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 $$\mathbf{R} = \frac{\sum m_i \mathbf{r}_i}{M} \tag{6.4d}$$

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. (6.4a), (6.4b), (6.4c) and (6.4d) 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 i^{th} 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,

$$\sum \Delta m_i \to \int \mathrm{d}m = M,$$

$$\sum (\Delta m_i) x_i \to \int x\, \mathrm{d}m,$$

$$\sum (\Delta m_i) y_i \to \int y\, \mathrm{d}m,$$

and $$\sum (\Delta m_i) z_i \to \int z\, \mathrm{d}m$$

Here M is the total mass of the body. The coordinates of the centre of mass now are

$$X = \frac{1}{M}\int x\, \mathrm{d}m, Y = \frac{1}{M}\int y\, \mathrm{d}m \text{ and } Z = \frac{1}{M}\int z\, \mathrm{d}m \qquad (6.5a)$$

The vector expression equivalent to these three scalar expressions is

$$\mathbf{R} = \frac{1}{M}\int \mathbf{r}\, \mathrm{d}m \qquad (6.5b)$$

If we choose, the centre of mass as the origin of our coordinate system,

$$\mathbf{R} = \mathbf{0}$$

i.e., $\int \mathbf{r}\, \mathrm{d}m = \mathbf{0}$

or $$\int x\, \mathrm{d}m = \int y\, \mathrm{d}m = \int z\, \mathrm{d}m = 0 \qquad (6.6)$$

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.

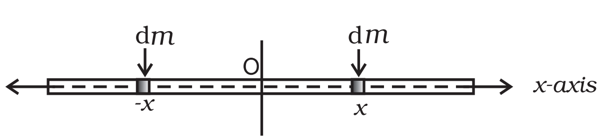


Fig. 6.8 *Determining the CM of a thin rod.*

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. 6.8).

The net contribution of every such pair to the integral and hence the integral $\int x\, \mathrm{d}m$ itself is zero. From Eq. (6.6), 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. (6.5 a) all are zero. This means that for all the above bodies, their centre of mass coincides with their geometric centre.

▶ ***Example 6.1*** 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

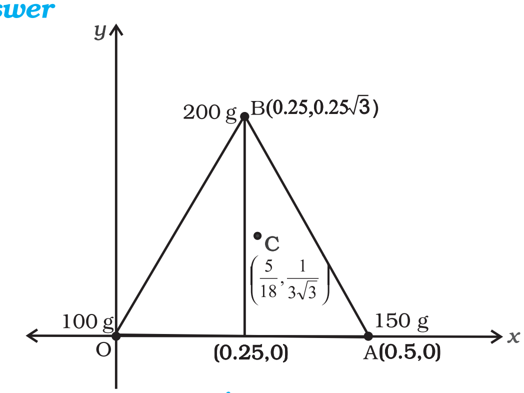


Fig. 6.9

With the x–and y–axes chosen as shown in Fig. 6.9, 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,

$$X = \frac{m_1 x_1 + m_2 x_2 + m_3 x_3}{m_1 + m_2 + m_3}$$

$$= \frac{100(0) + 150(0.5) + 200(0.25) \text{ g m}}{(100 + 150 + 200) \text{ g}}$$

$$= \frac{75 + 50}{450}\text{m} = \frac{125}{450}\text{m} = \frac{5}{18}\text{m}$$

$$Y = \frac{100(0) + 150(0) + 200(0.25\sqrt{3}) \text{ g m}}{450 \text{ g}}$$

$$= \frac{50\sqrt{3}}{450}\text{m} = \frac{\sqrt{3}}{9}\text{m} = \frac{1}{3\sqrt{3}}\text{m}$$

The centre of mass C is shown in the figure. Note that it is not the geometric centre of the triangle OAB. Why? ◀

▶ ***Example 6.2*** 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. 6.10

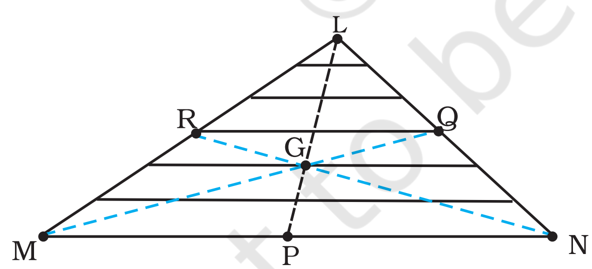


Fig. 6.10

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 6.3*** 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. 6.11 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.

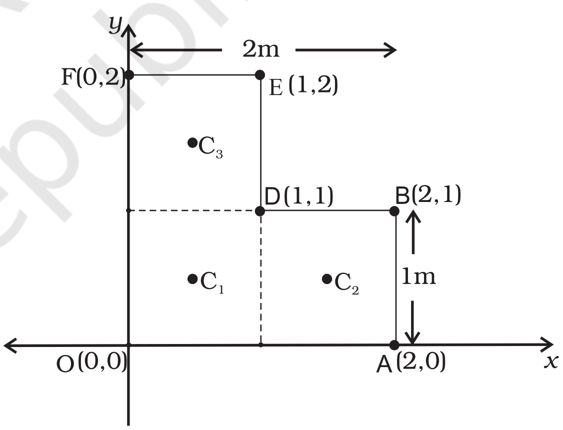


Fig. 6.11

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. 6.11 had different masses. How will you then determine the centre of mass of the lamina?
◀

6.3 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 n particles. We may rewrite Eq.(6.4d) as

$$M\mathbf{R} = \sum m_i \mathbf{r}_i = m_1\mathbf{r}_1 + m_2\mathbf{r}_2 + \ldots + m_n\mathbf{r}_n \qquad (6.7)$$

Differentiating the two sides of the equation with respect to time we get

$$M\frac{\mathrm{d}\mathbf{R}}{\mathrm{d}t} = m_1\frac{\mathrm{d}\mathbf{r}_1}{\mathrm{d}t} + m_2\frac{\mathrm{d}\mathbf{r}_2}{\mathrm{d}t} + \ldots + m_n\frac{\mathrm{d}\mathbf{r}_n}{\mathrm{d}t}$$

or

$$M\mathbf{V} = m_1\mathbf{v}_1 + m_2\mathbf{v}_2 + \ldots + m_n\mathbf{v}_n \qquad (6.8)$$

where $\mathbf{v}_1 \left(= \mathrm{d}\mathbf{r}_1 / \mathrm{d}t\right)$ is the velocity of the first particle $\mathbf{v}_2 \left(= d\mathbf{r}_2/dt\right)$ 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.(6.8) 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

$$M\mathbf{A} = m_1\mathbf{a}_1 + m_2\mathbf{a}_2 + \ldots + m_n\mathbf{a}_n \qquad (6.9)$$

where $\mathbf{a}_1 \left(= \mathrm{d}\mathbf{v}_1 / \mathrm{d}t\right)$ is the acceleration of the first particle, $\mathbf{a}_2 \left(= \mathrm{d}\mathbf{v}_2 / \mathrm{d}t\right)$ is the acceleration of the second particle etc. and $\mathbf{A} \left(= \mathrm{d}\mathbf{V} / \mathrm{d}t\right)$ 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. (6.9) may be written as

$$M\mathbf{A} = \mathbf{F}_1 + \mathbf{F}_2 + \ldots + \mathbf{F}_n \qquad (6.10)$$

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. (6.10), their contribution is zero. Only the external forces contribute to the equation. We can then rewrite Eq. (6.10) as

$$M\mathbf{A} = \mathbf{F}_{ext} \qquad (6.11)$$

where $\mathbf{F}_{ext}$ represents the sum of all external forces acting on the particles of the system.

Eq. (6.11) 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. (6.11) 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. (6.11).

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 of the 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.

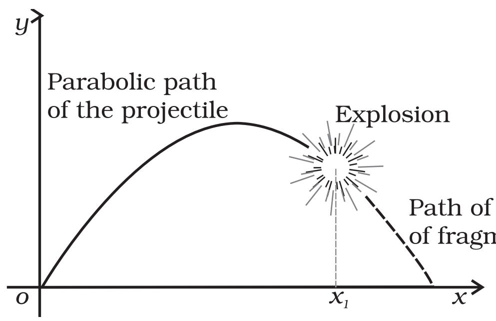


Fig. 6.12 *The centre of mass of the fragments of the projectile continues along the same parabolic path which it would have followed if there were no explosion.*

Figure 6.12 is a good illustration of Eq. (6.11). 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.

6.4 LINEAR MOMENTUM OF A SYSTEM OF PARTICLES

Let us recall that the linear momentum of a particle is defined as

$$\mathbf{p} = m\,\mathbf{v} \tag{6.12}$$

Let us also recall that Newton's second law written in symbolic form for a single particle is

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} \tag{6.13}$$

where **F** is the force on the particle. Let us consider a system of n particles with masses m_1, m_2,...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,

$$\mathbf{P} = \mathbf{p}_1 + \mathbf{p}_2 + \ldots + \mathbf{p}_n$$

$$= m_1\mathbf{v}_1 + m_2\mathbf{v}_2 + \ldots + m_n\mathbf{v}_n \tag{6.14}$$

Comparing this with Eq. (6.8)

$$\mathbf{P} = M\,\mathbf{V} \tag{6.15}$$

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. (6.15) with respect to time,

$$\frac{d\mathbf{P}}{dt} = M\frac{d\mathbf{V}}{dt} = M\mathbf{A} \tag{6.16}$$

Comparing Eq.(6.16) and Eq. (6.11),

$$\frac{d\mathbf{P}}{dt} = \mathbf{F}_{ext} \tag{6.17}$$

This is the statement of **Newton's second law of motion 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.(6.17)

$$\frac{d\mathbf{P}}{dt} = 0 \quad \text{or} \quad \mathbf{P} = \text{Constant} \tag{6.18a}$$

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. (6.15), 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. (6.18a) is equivalent to three scalar equations,

$$P_x = c_1,\ P_y = c_2 \text{ and } P_z = c_3 \qquad (6.18\text{ b})$$

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.

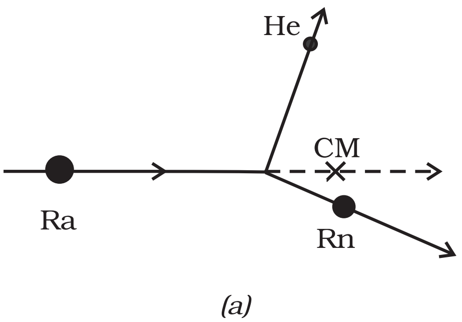


Fig. 6.13 (a) A heavy nucleus radium (Ra) splits into a lighter nucleus radon (Rn) and an alpha particle (nucleus of helium atom). The CM of the system is in uniform motion.
(b) The same spliting of the heavy nucleus radium (Ra) with the centre of mass at rest. The two product particles fly back to back.

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. 6.13(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 move back to back with their centre of mass remaining at rest as shown in Fig.6.13 (b).

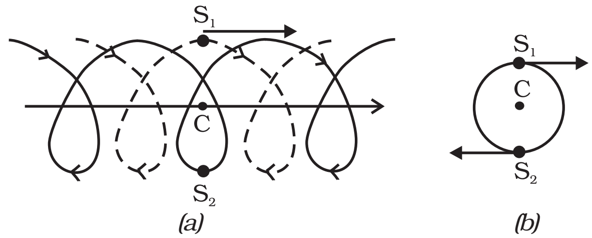


Fig. 6.14 (a) Trajectories of two stars, S_1 (dotted line) and S_2 (solid line) forming a binary system with their centre of mass C in uniform motion.
(b) The same binary system, with the centre of mass C at rest.

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.6.14 (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. 6.14(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.

6.5 VECTOR PRODUCT OF TWO VECTORS

We are already familiar with vectors and their use in physics. In chapter 5 (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. 6.15a.

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. 6.15b.

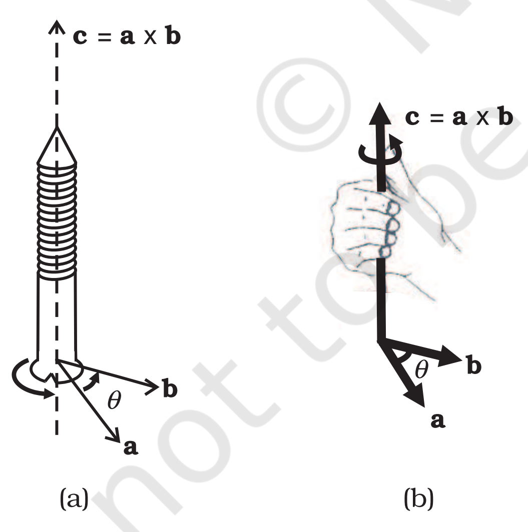


Fig. 6.15 (a) Rule of the right handed screw for defining the direction of the vector product of two vectors.
(b) Rule of the right hand for defining the direction of the vector product.

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. 6.15 (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. **a** × **b** ≠ **b** × **a**

The magnitude of both **a** × **b** and **b** × **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 **a** × **b** is from **a** to **b**, whereas in case of **b** × **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 plane 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 **a** × **b** under reflection?

$$\mathbf{a}\times\mathbf{b} \to (-\mathbf{a})\times(-\mathbf{b}) = \mathbf{a}\times\mathbf{b}$$

Thus, **a** × **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) **a** × **a** = **0** (**0** is a null vector, i.e. a vector with zero magnitude)

This follows since magnitude of **a** × **a** is $a^2\sin 0° = 0$.

From this follow the results

(i) $\hat{\mathbf{i}} \times \hat{\mathbf{i}} = \mathbf{0}$, $\hat{\mathbf{j}} \times \hat{\mathbf{j}} = \mathbf{0}$, $\hat{\mathbf{k}} \times \hat{\mathbf{k}} = \mathbf{0}$

(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,

$$\mathbf{a} \times \mathbf{b} = (a_x\hat{\mathbf{i}} + a_y\hat{\mathbf{j}} + a_z\hat{\mathbf{k}}) \times (b_x\hat{\mathbf{i}} + b_y\hat{\mathbf{j}} + b_z\hat{\mathbf{k}})$$

$$= a_x b_y \hat{\mathbf{k}} - a_x b_z \hat{\mathbf{j}} - a_y b_x \hat{\mathbf{k}} + a_y b_z \hat{\mathbf{i}} + a_z b_x \hat{\mathbf{j}} - a_z b_y \hat{\mathbf{i}}$$

$$= (a_y b_z - a_z b_y)\hat{\mathbf{i}} + (a_z b_x - a_x b_z)\hat{\mathbf{j}} + (a_x b_y - a_y b_x)\hat{\mathbf{k}}$$

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 6.4 *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 $\mathbf{b} \times \mathbf{a} = -7\hat{\mathbf{i}} + \hat{\mathbf{j}} + 5\hat{\mathbf{k}}$

6.6 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. 6.4. As said above, in rotational motion of a rigid body about a fixed axis, every particle of the body moves in a circle,

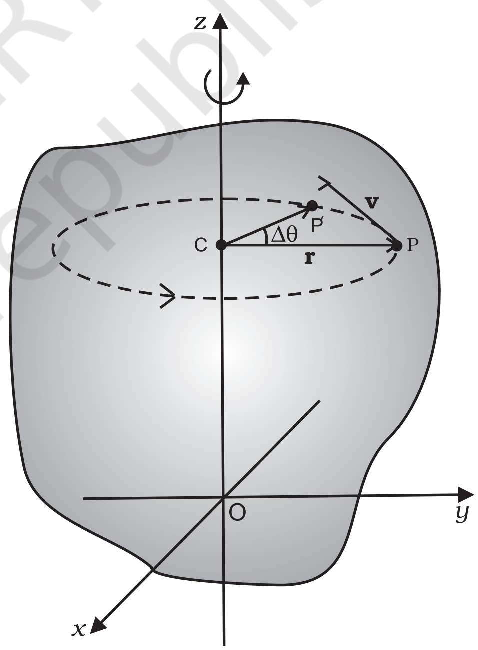


Fig. 6.16 *Rotation about a fixed axis. (A particle (P) of the rigid body rotating about the fixed (z-) axis moves in a circle with centre (C) on the axis.)*

which lies in a plane perpendicular to the axis and has its centre on the axis. In Fig. 6.16 we redraw Fig. 6.4, 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 $\mathbf{v}$ 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. 6.16). 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 $\upsilon = \omega r$, where r is the radius of the circle.

We observe that at any given instant the relation $\upsilon = \omega r$ 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

$$\upsilon_i = \omega r_i \tag{6.19}$$

The index i runs from 1 to n, where n is the total number of particles of the body.

For particles on the axis, $r = 0$, 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 $\boldsymbol{\omega}$ 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. 6.1 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. 6.17a).

The magnitude of this vector is $\omega = d\theta/dt$ referred as above.

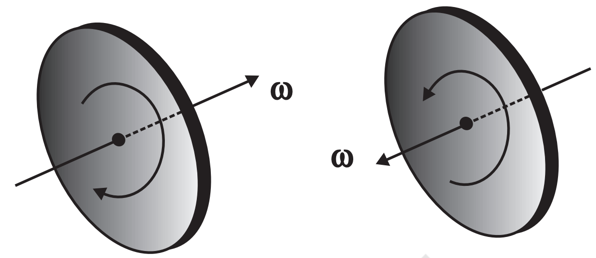


Fig. 6.17 (a) *If the head of a right handed screw rotates with the body, the screw advances in the direction of the angular velocity* $\boldsymbol{\omega}$. *If the sense (clockwise or anticlockwise) of rotation of the body changes, so does the direction of* $\boldsymbol{\omega}$.

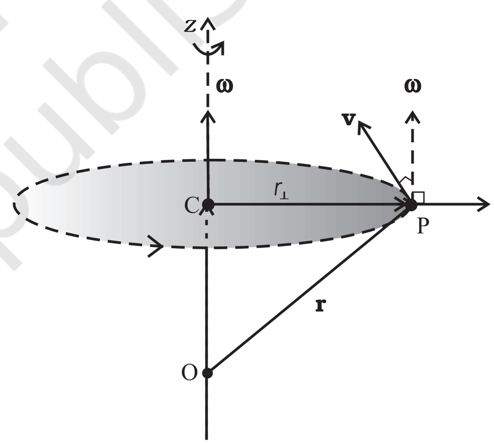


Fig. 6.17 (b) *The angular velocity vector* $\boldsymbol{\omega}$ *is directed along the fixed axis as shown. The linear velocity of the particle at P is* $\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}$. *It is perpendicular to both* $\boldsymbol{\omega}$ *and* $\mathbf{r}$ *and is directed along 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. 6.17(b) which is a part of Fig. 6.16 reproduced to show the path of the particle P. The figure shows the vector $\boldsymbol{\omega}$ 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 $\boldsymbol{\omega} \times \mathbf{OC} = 0$ 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 $\boldsymbol{\omega}$, i.e. to the z-axis and also to **CP**, 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 $\boldsymbol{\omega} \times \mathbf{CP}$ is ω (CP) since $\boldsymbol{\omega}$ and **CP** are perpendicular to each other. We shall denote **CP** by $\mathbf{r}_\perp$ and not by **r**, as we did earlier.

Thus, $\boldsymbol{\omega} \times \mathbf{r}$ 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,

$$\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r} \tag{6.20}$$

In fact, the relation, Eq. (6.20), holds good even for rotation of a rigid body with one point fixed, such as the rotation of the top [Fig. 6.6(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.**

6.6.1 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 (**s**) and velocity (**v**) in translational motion, we have angular displacement ($\boldsymbol{\theta}$) 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 $\boldsymbol{\alpha}$ as the time rate of change of angular velocity. Thus,

$$\boldsymbol{\alpha} = \frac{d\boldsymbol{\omega}}{dt} \tag{6.21}$$

If the axis of rotation is fixed, the direction of $\boldsymbol{\omega}$ and hence, that of $\boldsymbol{\alpha}$ is fixed. In this case the vector equation reduces to a scalar equation

$$\alpha = \frac{d\omega}{dt} \tag{6.22}$$

6.7 TORQUE AND ANGULAR MOMENTUM

In this section, we shall acquaint ourselves with two physical quantities (torque and angular momentum) 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.

6.7.1 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 in linear motion is **moment of force**. It is also referred to as **torque** or **couple**. (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.

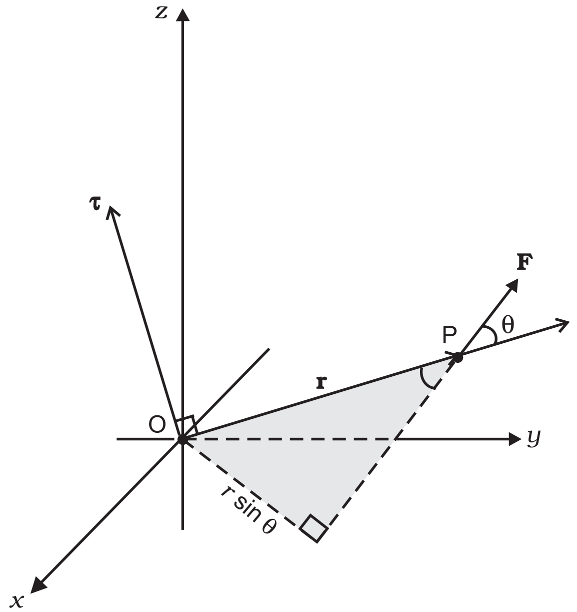


Fig. 6.18 $\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F}$, *$\tau$ is perpendicular to the plane containing* $\mathbf{r}$ *and* $\mathbf{F}$, *and its direction is given by the right handed screw rule.*

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 $\mathbf{r}$ (Fig. 6.18), the moment of the force acting on the particle with respect to the origin O is defined as the vector product

$$\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F} \qquad (6.23)$$

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

$$\tau = r F \sin\theta \qquad (6.24a)$$

where r is the magnitude of the position vector $\mathbf{r}$, i.e. the length OP, F is the magnitude of force $\mathbf{F}$ and θ is the angle between $\mathbf{r}$ and $\mathbf{F}$ 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 (N m). The magnitude of the moment of force may be written

$$\tau = (r \sin\theta) F = r_\perp F \qquad (6.24b)$$

$$\text{or } \tau = r F \sin\theta = r F_\perp \qquad (6.24c)$$

where $r_\perp = r \sin\theta$ is the perpendicular distance of the line of action of $\mathbf{F}$ from the origin and $F_\perp (= F \sin\theta)$ is the component of $\mathbf{F}$ in the direction perpendicular to $\mathbf{r}$. 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 $\mathbf{F}$ is reversed, the direction of the moment of force is reversed. If directions of both $\mathbf{r}$ and $\mathbf{F}$ are reversed, the direction of the moment of force remains the same.

6.7.2 Angular momentum of a particle

Just as the moment of a force is the rotational analogue of force in linear motion, 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 $\mathbf{p}$ at a position $\mathbf{r}$ relative to the origin O. The angular momentum $\mathbf{l}$ of the particle with respect to the origin O is defined to be

$$\mathbf{l} = \mathbf{r} \times \mathbf{p} \qquad (6.25a)$$

The magnitude of the angular momentum vector is

$$l = r p \sin\theta \qquad (6.26a)$$

where p is the magnitude of $\mathbf{p}$ and θ is the angle between $\mathbf{r}$ and $\mathbf{p}$. We may write

$$l = r p_\perp \text{ or } r_\perp p \qquad (6.26b)$$

where $r_\perp$ (= $r \sin\theta$) is the perpendicular distance of the directional line of $\mathbf{p}$ from the origin and $p_\perp (= p \sin\theta)$ is the component of $\mathbf{p}$ in a direction perpendicular to $\mathbf{r}$. 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 $\mathbf{p}$ 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 $\boldsymbol{l} = \mathbf{r} \times \mathbf{p}$ with respect to time,

$$\frac{\mathrm{d}\boldsymbol{l}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{r} \times \mathbf{p})$$

Applying the product rule for differentiation to the right hand side,

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{r} \times \mathbf{p}) = \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} \times \mathbf{p} + \mathbf{r} \times \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t}$$

Now, the velocity of the particle is $\mathbf{v} = d\mathbf{r}/dt$ and $\mathbf{p} = m\,\mathbf{v}$

Because of this $\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} \times \mathbf{p} = \mathbf{v} \times m\,\mathbf{v} = 0$,

as the vector product of two parallel vectors vanishes. Further, since $\mathrm{d}\mathbf{p} / \mathrm{d}t = \mathbf{F}$,

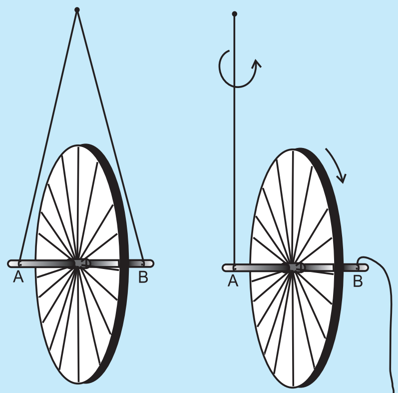
$$\mathbf{r} \times \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = \mathbf{r} \times \mathbf{F} = \mathbf{t}$$

Hence $\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{r} \times \mathbf{p}) = \tau$

or $\frac{\mathrm{d}\boldsymbol{l}}{\mathrm{d}t} = \tau$ (6.27)

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 $\mathbf{F} = \mathrm{d}\mathbf{p}/\mathrm{d}t$, which expresses Newton's second law for the translational motion of a single particle.

An experiment with the bicycle rim



Initially After

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,

$$\mathbf{L} = \boldsymbol{l}_1 + \boldsymbol{l}_2 + \ldots + \boldsymbol{l}_n = \sum_{i=1}^{n} \boldsymbol{l}_i$$

The angular momentum of the i^{th} particle is given by

$\boldsymbol{l}_i = \mathbf{r}_i \ \mathbf{p}_i$

where $\mathbf{r}_i$ is the position vector of the i^{th} 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

$$\mathbf{L} = \sum \boldsymbol{l}_i = \sum_i \mathbf{r}_i \times \mathbf{p}_i \qquad (6.25b)$$

This is a generalisation of the definition of angular momentum (Eq. 6.25a) for a single particle to a system of particles.

Using Eqs. (6.23) and (6.25b), we get

$$\frac{d\mathbf{L}}{dt} = \frac{d}{dt}\left(\Sigma \boldsymbol{l}_i\right) = \sum_i \frac{d\boldsymbol{l}}{dt} = \sum_i \tau_i \qquad (6.28a)$$

where $\boldsymbol{\tau}_i$ is the torque acting on the i^{th} particle;

$\tau_i = \mathbf{r}_i \times \mathbf{F}_i$

The force $\mathbf{F}_i$ on the i^{th} particle is the vector sum of external forces $\mathbf{F}_i^{ext}$ acting on the particle and the internal forces $\mathbf{F}_i^{\text{int}}$ 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

$\tau = \sum_i \tau_i = \sum_i \mathbf{r}_i \times \mathbf{F}_i$ as

$$\tau = \tau_{ext} + \tau_{\text{int}},$$

where $\tau_{ext} = \sum_i \mathbf{r}_i \times \mathbf{F}_i^{ext}$

and $\tau_{\text{int}} = \sum_i \mathbf{r}_i \times \mathbf{F}_i^{\text{int}}$

We shall assume not only Newton's third law of motion, 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, $\boldsymbol{\tau}_{\text{int}} = \mathbf{0}$ and therefore $\boldsymbol{\tau} = \boldsymbol{\tau}_{ext}$.

Since $\tau = \sum \tau_i$, it follows from Eq. (6.28a) that

$$\frac{d\mathbf{L}}{dt} = \tau_{ext} \qquad (6.28\text{ b})$$

Thus, **the time rate of the total angular momentum of a system of particles about a point** (taken as the origin of our frame of reference) **is equal to the sum of the external torques** (i.e. the torques due to external forces) **acting on the system taken about the same point**. Eq. (6.28 b) is the generalisation of the single particle case of Eq. (6.23) to a system of particles. Note that when we have only one particle, there are no internal forces or torques. Eq.(6.28 b) is the rotational analogue of

$$\frac{d\mathbf{P}}{dt} = \mathbf{F}_{ext} \qquad (6.17)$$

Note that like Eq.(6.17), Eq.(6.28b) 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. (6.28b) reduces to

$$\frac{d\mathbf{L}}{dt} = 0$$

or $\mathbf{L}$ = constant. (6.29a)

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. (6.29a) is equivalent to three scalar equations,

$L_x = K_1$, $L_y = K_2$ and $L_z = K_3$ (6.29 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. (6.29a) is the rotational analogue of Eq. (6.18a), i.e. the conservation law of the total linear momentum for a system of particles. Like Eq. (6.18a), it has applications in many practical situations. We shall look at a few of the interesting applications later on in this chapter.

Example 6.5 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 $\mathbf{F} = 7\hat{\mathbf{i}} + 3\hat{\mathbf{j}} - 5\hat{\mathbf{k}}$.

We shall use the determinant rule to find the torque $\tau = \mathbf{r} \times \mathbf{F}$

$$\tau = \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ 1 & -1 & 1 \\ 7 & 3 & -5 \end{vmatrix} = (5-3)\hat{\mathbf{i}} - (-5-7)\hat{\mathbf{j}} + (3-(-7))\hat{\mathbf{k}}$$

or $\tau = 2\hat{\mathbf{i}} + 12\hat{\mathbf{j}} + 10\hat{\mathbf{k}}$ ◀

Example 6.6 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.

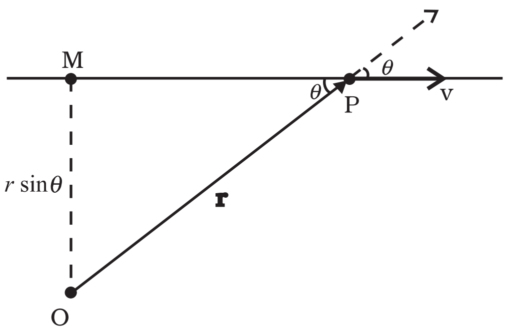


Fig 6.19

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. 6.19. 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? ◀

6.8 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. (6.17). 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. (6.28 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;

$$\mathbf{F}_1 + \mathbf{F}_2 + \ldots + \mathbf{F}_n = \sum_{i=1}^{n} \mathbf{F}_i = \mathbf{0} \qquad (6.30a)$$

If the total force on the body is zero, then the total linear momentum of the body does not change with time. Eq. (6.30a) 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,

$$\boldsymbol{\tau}_1 + \boldsymbol{\tau}_2 + \cdots + \boldsymbol{\tau}_n = \sum_{i=1}^{n} \boldsymbol{\tau}_i = \mathbf{0} \qquad (6.30b)$$

If the total torque on the rigid body is zero, the total angular momentum of the body does not change with time. Eq. (6.30 b) gives the condition for the rotational equilibrium of the body.

One may raise a question, whether the rotational equilibrium condition [Eq. 6.30(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. 6.30(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 6.7 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. (6.30a) and Eq. (6.30b), both, are vector equations. They are equivalent to three scalar equations each. Eq. (6.30a) corresponds to

$$\sum_{i=1}^{n} F_{ix} = 0, \quad \sum_{i=1}^{n} F_{iy} = 0 \ \text{and} \ \sum_{i=1}^{n} F_{iz} = 0 \qquad (6.31a)$$

where F_{ix}, F_{iy} and F_{iz} are respectively the *x*, *y* and *z* components of the forces $\mathbf{F}_i$. Similarly, Eq. (6.30b) is equivalent to three scalar equations

$$\sum_{i=1}^{n} \tau_{ix} = 0, \quad \sum_{i=1}^{n} \tau_{iy} = 0 \ \text{and} \ \sum_{i=1}^{n} \tau_{iz} = 0 \qquad (6.31b)$$

where τ_{ix}, τ_{iy} and τ_{iz} are respectively the *x*, *y* and *z* components of the torque $\boldsymbol{\tau}_i$.

Eq. (6.31a) and (6.31b) 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. 6.30 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) as shown in Fig. 6.20(a). At the two ends (A and B) of which two parallel forces, both equal in magnitude and acting along same direction are applied perpendicular to the rod.

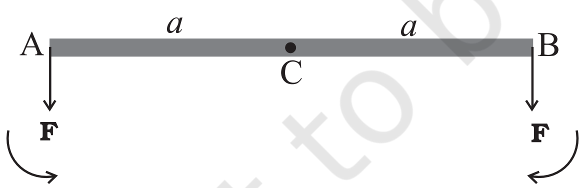


Fig. 6.20 (a)

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; $\sum \mathbf{F} \neq \mathbf{0}$

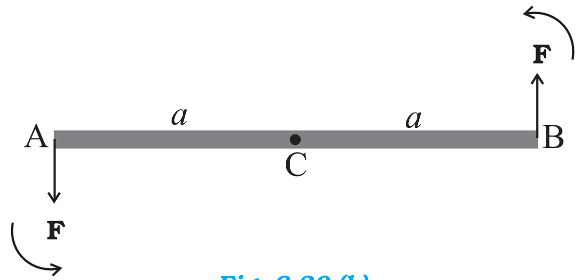


Fig. 6.20 (b)

The force at B in Fig. 6.20(a) is reversed in Fig. 6.20(b). Thus, we have the same rod with two forces of equal magnitude but acting in opposite diretions 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 forces of equal magnitude but acting in opposite directions with different lines of action is known as a **couple** or **torque**. 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. 6.21(a)]. Another known example is a compass needle in the earth's magnetic field as shown in the Fig. 6.21(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. 6.21(a) *Our fingers apply a couple to turn the lid.*

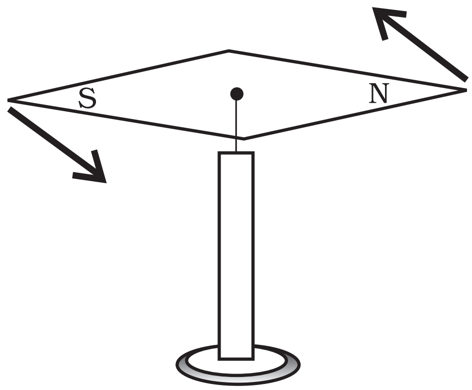


Fig. 6.21(b) *The Earth's magnetic field exerts equal and opposite forces on the poles of a compass needle. These two forces form a couple.*

Example 6.7 Show that moment of a couple does not depend on the point about which you take the moments.

Answer

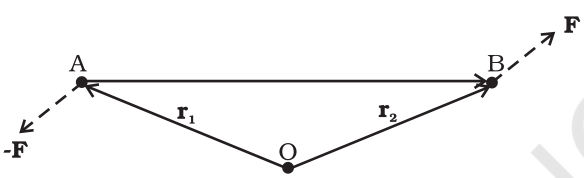


Fig. 6.22

Consider a couple as shown in Fig. 6.22 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.

6.8.1 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. 6.23.

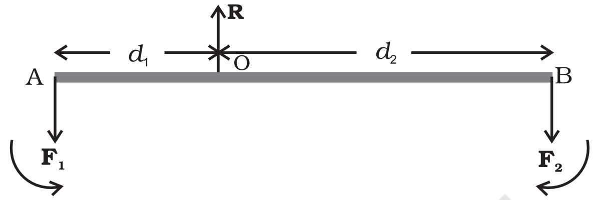


Fig. 6.23

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,

$$d_1 F_1 - d_2 F_2 = 0 \qquad \text{(ii)}$$

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

$$d_1 F_1 = d_2 F_2 \qquad \text{(6.32a)}$$

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.);

$$\text{M.A.} = \frac{F_1}{F_2} = \frac{d_2}{d_1} \qquad \text{(6.32b)}$$

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.

6.8.2 Centre of gravity

Many of you may have the experience of balancing your notebook on the tip of a finger. Figure 6.24 illustrates a similar experiment that you can easily perform. Take an irregular-shaped cardboard having mass M 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. 6.24, the reaction of the tip is equal and opposite to $M\mathbf{g}$ 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.

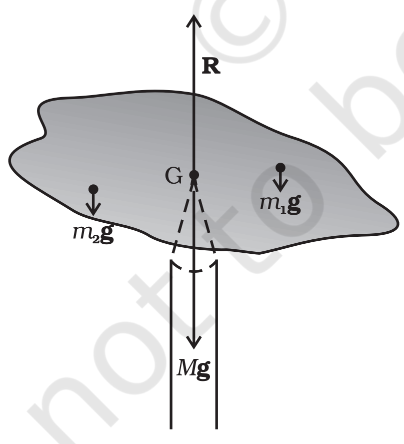


Fig. 6.24 *Balancing a cardboard on the tip of a pencil. The point of support, G, is the centre of gravity.*

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.

$$\boldsymbol{\tau}_g = \sum \boldsymbol{\tau}_i = \sum \mathbf{r}_i \times m_i\mathbf{g} = \mathbf{0} \qquad (6.33)$$

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. (6.33), $\mathbf{g}$ is the same for all particles, and hence it comes out of the summation. This gives, since $\mathbf{g}$ is non-zero, $\sum m_i\mathbf{r}_i = \mathbf{0}$. Remember that the position vectors ($\mathbf{r}_i$) are taken with respect to the CG. Now, in accordance with the reasoning given below Eq. (6.4a) in Sec. 6.2, 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 in uniform gravity or gravity-

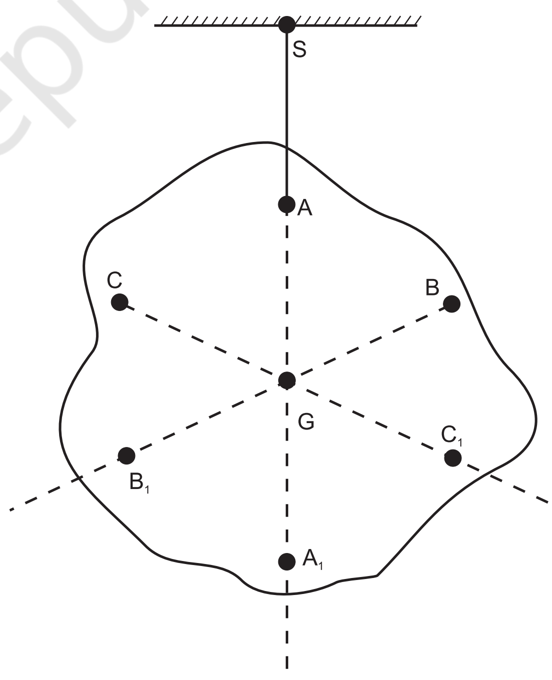


Fig. 6.25 *Determining the centre of gravity of a body of irregular shape. The centre of gravity G lies on the vertical AA_1 through the point of suspension of the body A.*

free space. We note that this is true because the body being small, **g** does not 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. 6.2 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 6.25 illustrates another way of determining the CG of an irregular 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 6.8 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 load 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

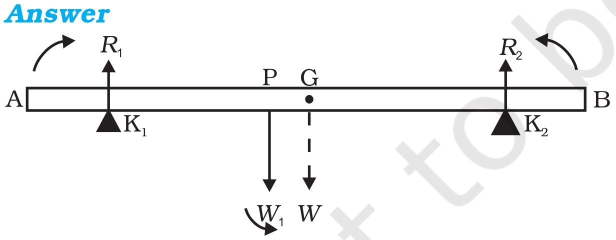


Fig. 6.26

Figure 6.26 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 load 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 load = 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 $R_1 - R_2 = 1.2g$ 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 6.9 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.6.27. Find the reaction forces of the wall and the floor.

Answer

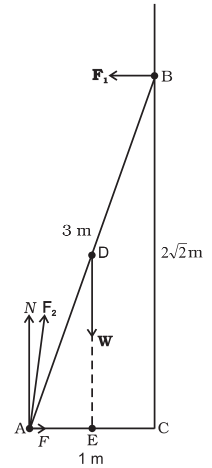


Fig. 6.27

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,

$$N - W = 0 \qquad \text{(i)}$$

Taking the forces in the horizontal direction,

$$F - F_1 = 0 \qquad \text{(ii)}$$

For rotational equilibrium, taking the moments of the forces about A,

$$2\sqrt{2}\,F_1 - (1/2)\,W = 0 \qquad \text{(iii)}$$

Now $W = 20\text{ g} = 20 \quad 9.8\text{ N} = 196.0\text{ N}$

From (i) $N = 196.0$ N

From (iii) $F_1 = W/4\sqrt{2} = 196.0/4\sqrt{2} = 34.6\text{ N}$

From (ii) $F = F_1 = 34.6\text{ N}$

$$F_2 = \sqrt{F^2 + N^2} = 199.0\text{ N}$$

The force F_2 makes an angle α with the horizontal,

$$\tan\alpha = N/F = 4\sqrt{2}\,, \quad \alpha = \tan^{-1}(4\sqrt{2}) \approx 80^\circ$$ ◀

6.9 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. (6.19). (Refer to Fig. 6.16). For a particle at a distance from the axis, the linear velocity is $\upsilon_i = r_i\omega$. The kinetic energy of motion of this particle is

$$k_i = \frac{1}{2}m_i\upsilon_i^2 = \frac{1}{2}m_i r_i^2\omega^2$$

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

$$I = \sum_{i=1}^{n} m_i r_i^2 \qquad (6.34)$$

With this definition,

$$K = \frac{1}{2}I\omega^2 \qquad (6.35)$$

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. (6.35) 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 linear motion. 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. (6.34), 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,

$$K = \frac{1}{2}M\upsilon^2 = \frac{1}{2}MR^2\omega^2$$

Comparing with Eq. (6.35) we get $I = MR^2$ for the ring.

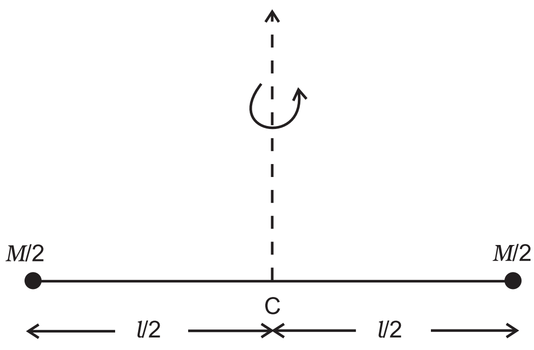


Fig. 6.28 *A light rod of length l with a pair of masses rotating about an axis through the centre of mass of the system and perpendicular to the rod. The total mass of the system is M.*

(b) Next, take a rigid rod of negligible mass of length of length l with a pair of small masses, rotating about an axis through the centre of mass perpendicular to the rod (Fig. 6.28). 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 6.1 simply gives the moment of inertia of various familiar regular shaped bodies about specific axes. (The derivations of these expressions are beyond the scope of this textbook and you will study them in higher classes.)

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 distribution of mass about the axis of rotation, and 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.

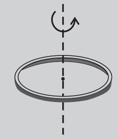
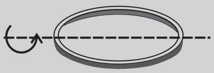
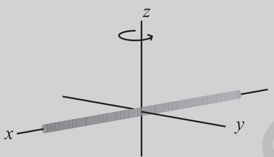
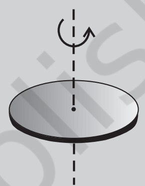
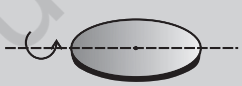
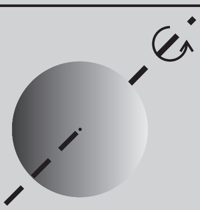
Notice from the Table 6.1 that in all cases, we can write I = Mk^2, where k has the dimension of length. For a rod, about the perpendicular axis at its midpoint, $k^2 = L^2/12$, i.e. $k = L/\sqrt{12}$. Similarly, $k = R/2$ for the circular disc about its diameter. The length k is a geometric property 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. (6.34), we can infer that the dimensions of moments of inertia are 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.

Table 6.1 **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		MR^2
(2)	Thin circular ring, radius R	Diameter		$MR^2/2$
(3)	Thin rod, length L	Perpendicular to rod, at mid point		$ML^2/12$
(4)	Circular disc, radius R	Perpendicular to disc at centre		$MR^2/2$
(5)	Circular disc, radius R	Diameter		$MR^2/4$
(6)	Hollow cylinder, radius R	Axis of cylinder		MR^2
(7)	Solid cylinder, radius R	Axis of cylinder		$MR^2/2$
(8)	Solid sphere, radius R	Diameter		$2MR^2/5$

6.10 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 $\boldsymbol{\omega}$ 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.6.29) 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. 6.29 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 are analogous 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:

$$v = v_0 + at \qquad \text{(a)}$$

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

$$v^2 = v_0^2 + 2ax \qquad \text{(c)}$$

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:

$$\omega = \omega_0 + \alpha t \qquad (6.36)$$

$$\theta = \theta_0 + \omega_0 t + \frac{1}{2}\alpha t^2 \qquad (6.37)$$

$$\text{and } \omega^2 = \omega_0^2 + 2\alpha(\theta - \theta_0) \qquad (6.38)$$

where θ_0= initial angular displacement of the rotating body, and ω_0 = initial angular velocity of the body.

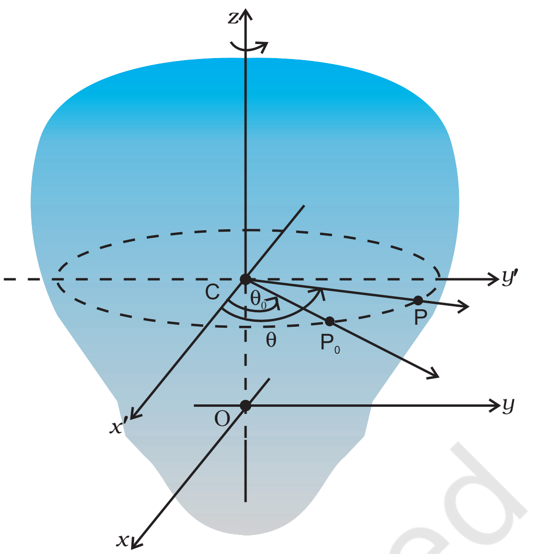


Fig.6.29 *Specifying the angular position of a rigid body.*

Example 6.10 Obtain Eq. (6.36) from first principles.

Answer The angular acceleration is uniform, hence

$$\frac{d\omega}{dt} = \alpha = constant \qquad \text{(i)}$$

Integrating this equation,

$$\omega = \int \alpha \, dt + c$$

$$= \alpha t + c \qquad (\text{as } \alpha \text{ is constant})$$

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. (6.36) to get Eq. (6.37). This derivation and the derivation of Eq. (6.38) is left as an exercise.

Example 6.11 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

$$\alpha = \frac{\omega - \omega_0}{t} \qquad = 4\ \pi \text{ rad/s}^2$$

The angular acceleration of the engine $= 4\pi$ rad/s^2

(ii) The angular displacement in time t is given by

$$\theta = \omega_0 t + \frac{1}{2}\alpha t^2$$

$$= (40\pi \times 16 + \frac{1}{2} \times 4\pi \times 16^2) \text{ rad}$$

$$= (640\pi + 512\pi) \text{ rad}$$

$= 1152\pi$ rad

Number of revolutions $= \frac{1152\pi}{2\pi} = 576$ ◀

6.11 DYNAMICS OF ROTATIONAL MOTION ABOUT A FIXED AXIS

Table 6.2 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 $dx \rightarrow d\theta$ and $F \rightarrow \tau$. 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

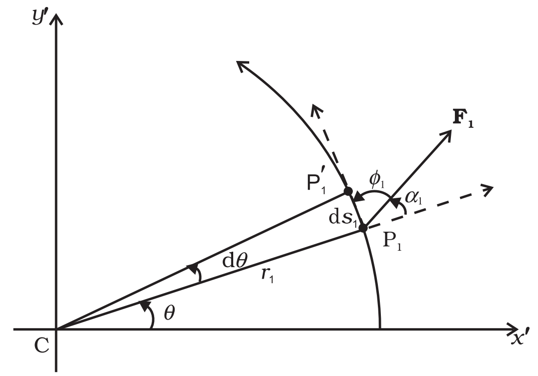


Fig. 6.30 Work done by a force $\mathbf{F}_1$ *acting on a particle of a body rotating about a fixed axis; the particle describes a circular path with centre C on the axis; arc* $P_1P'_1(ds_1)$ *gives the displacement of the particle.*

Table 6.2 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 $L = I\omega$

Figure 6.30 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. 6.29). 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 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. 6.17(b).] Since $\mathbf{OC}$ 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 $\tau_1, \tau_2, \ldots$ etc,

$$dW = (\tau_1 + \tau_2 + \ldots)d\theta$$

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., $\tau = \tau_1 + \tau_2 + \ldots..$ We, therefore, have

$$dW = \tau d\theta \tag{6.39}$$

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. (6.39) by dt gives

$$P = \frac{dW}{dt} = \tau\frac{d\theta}{dt} = \tau\omega$$

or $$P = \tau\omega \tag{6.40}$$

This is the instantaneous power. Compare this expression for power in the case of rotational motion about a fixed axis with that of 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. (6.40). 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,

$$\tau\omega = I\,\omega\,\alpha$$

$$\tau = I\alpha \qquad (6.41)$$

Eq. (6.41) 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. In this respect, Eq.(6.41) can be called Newton's second law for rotational motion about a fixed axis.

Example 6.12 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. 6.31. 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

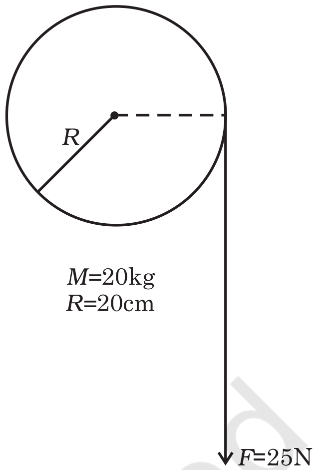


Fig. 6.31

(a) We use $I\,\alpha = \tau$

the torque $\tau = F\,R$

$= 25 \times 0.20$ Nm (as $R = 0.20$m)

$= 5.0$ Nm

I = Moment of inertia 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 = 12.5 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

$$\omega^2 = 2 \times 12.5 \times 10.0 = 250\,(\text{rad/s})^2$$

$$\therefore \text{K.E. gained} = \frac{1}{2} \times 0.4 \times 250 = 50\text{ J}$$

(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.

6.12 ANGULAR MOMENTUM IN CASE OF ROTATION ABOUT A FIXED AXIS

We have studied in section 6.7, 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 of n particles is

$$\mathbf{L} = \sum_{i=1}^{N} \mathbf{r}_i \times \mathbf{p}_i \qquad (6.25b)$$

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 $\boldsymbol{l} = \mathbf{r} \times \mathbf{p}$. As seen in the last section $\mathbf{r} = \mathbf{OP} = \mathbf{OC} + \mathbf{CP}$ [Fig. 6.17(b)]. With $\mathbf{p} = m\,\mathbf{v}$,

$$\boldsymbol{l} = (\mathbf{OC} \times m\,\mathbf{v}) + (\mathbf{CP} \times m\,\mathbf{v})$$

The magnitude of the linear velocity **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, **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 $\hat{\mathbf{k}}$. Hence

$$\mathbf{CP} \times m\,\mathbf{v} = r_\perp (mv)\,\hat{\mathbf{k}}$$

$$= m r_\perp^2 \omega\,\hat{\mathbf{k}} \quad (\text{since } \upsilon = \omega r_\perp)$$

Similarly, we can check that $\mathbf{OC} \times \mathbf{v}$ is perpendicular to the fixed axis. Let us denote the part of $\boldsymbol{l}$ along the fixed axis (i.e. the z-axis) by $\boldsymbol{l}_z$, then

$$\boldsymbol{l}_z = \mathbf{CP} \times m\,\mathbf{v} = m r_\perp^2 \omega \hat{\mathbf{k}}$$

and $\boldsymbol{l} = \boldsymbol{l}_z + \mathbf{OC} \times m\,\mathbf{v}$

We note that $\boldsymbol{l}_z$ is parallel to the fixed axis, but $\boldsymbol{l}$ is not. In general, for a particle, the angular momentum $\boldsymbol{l}$ is not along the axis of rotation, i.e. for a particle, $\boldsymbol{l}$ and $\boldsymbol{\omega}$ are not necessarily parallel. Compare this with the corresponding fact in translation. For a particle, **p** and **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 $\quad \mathbf{L} = \sum \boldsymbol{l}_i = \sum \boldsymbol{l}_{iz} + \sum \mathbf{OC}_i \times m_i \mathbf{v}_i$

We denote by $\mathbf{L}_\perp$ and $\mathbf{L}_z$ the components of **L** respectively perpendicular to the z-axis and along the z-axis;

$$\mathbf{L}_\perp = \sum \mathbf{OC}_i \times m_i \mathbf{v}_i \qquad (6.42a)$$

where m_i and $\mathbf{v}_i$ are respectively the mass and the velocity of the i^{th} particle and C_i is the centre of the circle described by the particle;

and $\quad \mathbf{L}_z = \sum \boldsymbol{l}_{iz} = \left(\sum_i m_i r_i^2\right) w\,\mathbf{k}$

or $\quad \mathbf{L}_z = I\omega\hat{\mathbf{k}} \qquad (6.42b)$

The last step follows since the perpendicular distance of the i^{th} 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 $\quad \mathbf{L} = \mathbf{L}_z + \mathbf{L}_\perp \qquad (6.42c)$

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

$$\mathbf{L} = \mathbf{L}_z = I\omega\hat{\mathbf{k}} \qquad (6.42d)$$

For bodies, which are not symmetric about the axis of rotation, **L** is not equal to $\mathbf{L}_z$ and hence **L** does not lie along the axis of rotation.

Referring to Table 6.1, can you tell in which cases $\mathbf{L} = \mathbf{L}_z$ will not apply?

Let us differentiate Eq. (6.42b). Since $\hat{\mathbf{k}}$ is a fixed (constant) vector, we get

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{L}_z) = \left(\frac{\mathrm{d}}{\mathrm{d}t}(I\omega)\right)\hat{\mathbf{k}}$$

Now, Eq. (6.28b) states

$$\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t} = \boldsymbol{\tau}$$

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 $\boldsymbol{\tau} = \tau\hat{\mathbf{k}}$. Since $\mathbf{L} = \mathbf{L}_z + \mathbf{L}_\perp$ and the direction of $\mathbf{L}_z$ (vector $\hat{\mathbf{k}}$) is fixed, it follows that for rotation about a fixed axis,

$$\frac{\mathrm{d}\mathbf{L}_z}{\mathrm{d}t} = \tau\hat{\mathbf{k}} \qquad (6.43a)$$

and $$\frac{\mathrm{d}\mathbf{L}_\perp}{\mathrm{d}t} = 0 \qquad (6.43b)$$

Thus, for rotation about a fixed axis, the component of angular momentum perpendicular to the fixed axis is constant. As $\mathbf{L}_z = I\omega\hat{\mathbf{k}}$, we get from Eq. (6.43a),

$$\frac{\mathrm{d}}{\mathrm{d}t}(I\omega) = \tau \qquad (6.43c)$$

If the moment of inertia I does not change with time,

$$\frac{\mathrm{d}}{\mathrm{d}t}(I\omega) = I\frac{\mathrm{d}\omega}{\mathrm{d}t} = I\alpha$$

and we get from Eq. (6.43c),

$$\tau = I\alpha \qquad (6.41)$$

We have already derived this equation using the work - kinetic energy route.

6.12.1 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. (6.43c), if the external torque is zero,

$$L_z = I\omega = \text{constant} \qquad (6.44)$$

For symmetric bodies, from Eq. (6.42d), 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. (6.29a), which expresses the general law of conservation of angular momentum of a system of particles. Eq. (6.44) applies to many situations that we come across in daily life. You may do this experiment with your friend. Sit on a swivel chair (a chair with a seat, free to rotate about a pivot) 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

Fig 6.32 (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.*

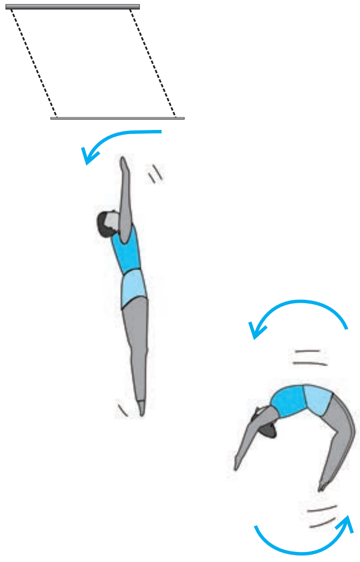


Fig 6.32 (b) *An acrobat employing the principle of conservation of angular momentum in her performance.*

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 (a spinning about a tip–top) on the toes of one foot display 'mastery' over this principle. Can you explain?

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 translational motion or a combination of translational and rotational motions.
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 $\boldsymbol{\omega}$ has a fixed direction.
6. The vector or cross product of two vector $\mathbf{a}$ and $\mathbf{b}$ 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 $\mathbf{r}$ 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 $\mathbf{r}$ 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 n 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 $\mathbf{P}$ 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

$$\tau = \sum_{1} \mathbf{r}_i \times \mathbf{F}_i$$

The force $\mathbf{F}_i$ acting on the i^{th} particle includes the external as well as internal forces. Assuming Newton's third law of motion 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 *i*th point of the body from the axis. The kinetic energy of rotation is $K = \frac{1}{2} I\omega^2$.

Quantity	Symbols	Dimensions	Units	Remarks
Angular velocity	ω	$[T^{-1}]$	rad s^{-}	$v = \omega \times r$
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 the motion of the centre of mass, (i.e., the translational 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 particles 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 **ω** 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 **L** = I**ω** holds good, where I is the moment of the inertia of the body about the rotation axis.

EXERCISES

6.1 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 ?

6.2 In the HCl 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.

6.3 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 ?

6.4 Show that the area of the triangle contained between the vectors **a** and **b** is one half of the magnitude of **a** × **b**.

6.5 Show that **a**·(**b** × **c**) is equal in magnitude to the volume of the parallelepiped formed on the three vectors , **a**, **b** and **c**.

6.6 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.

6.7 Two particles, each of mass m and speed v, travel in opposite directions along parallel lines separated by a distance d. Show that the angular momentum vector of the two particle system is the same whatever be the point about which the angular momentum is taken.

6.8 A non-uniform bar of weight W is suspended at rest by two strings of negligible weight as shown in Fig.6.33. 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.

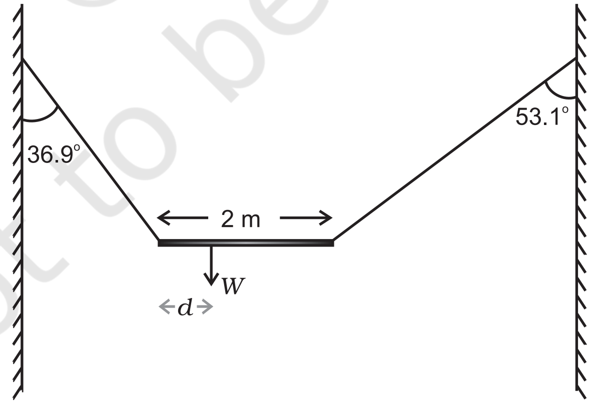


Fig. 6.33

6.9 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.

6.10 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.

6.11 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?

6.12 (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?

6.13 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.

6.14 To maintain a rotor at a uniform angular speed of 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.

6.15 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.

6.16 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?

6.17 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.

CHAPTER SEVEN

11086CH08

GRAVITATION

7.1 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 centre 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 centre around which the planets revolved – the 'heliocentric' model – was already mentioned by Aryabhatta (5^{th} 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.

7.2 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

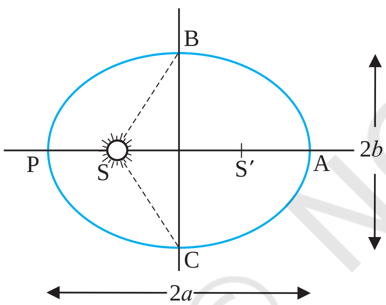


Fig. 7.1(a) *An ellipse traced out by a planet around the sun. The closest point is P and the farthest point is A, P is called the perihelion and A the aphelion. The semimajor axis is half the distance AP.*

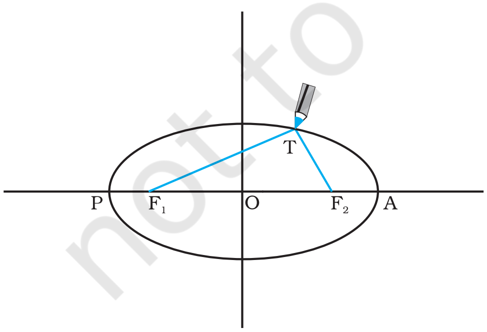


Fig. 7.1(b) *Drawing an ellipse. A string has its ends fixed at F_1 and F_2. The tip of a pencil holds the string taut and is moved around.*

of the ellipse (Fig. 7.1a). 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.

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. 7.1(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. 7.1(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. 7.2). This law comes from the observations that planets appear to move slower when they are farther from the sun than when they are nearer.

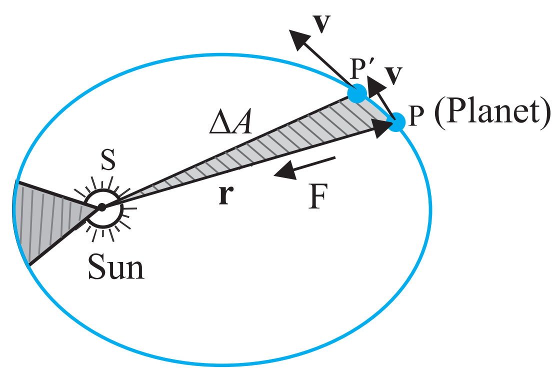


Fig. 7.2 *The planet P moves around the sun in an elliptical orbit. The shaded area is the area ΔA swept out in a small interval of time Δt.*

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.

Table 7.1 gives the approximate time periods of revolution of eight* planets around the sun along with values of their semi-major axes.

Table 7.1 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

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. 7.2) ΔA given by

$$\Delta \mathbf{A} = ½ \; (\mathbf{r} \times \mathbf{v}\Delta t) \qquad (7.1)$$

Hence

$$\Delta \mathbf{A} / \Delta t = ½ \; (\mathbf{r} \times \mathbf{p})/m, \; (\text{since } \mathbf{v} = \mathbf{p}/m)$$
$$= \mathbf{L} / (2\, m) \qquad (7.2)$$

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.

▶ ***Example 7.1*** Let the speed of the planet at the perihelion *P* in Fig. 7.1(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. 7.1. 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*. ◀

7.3 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

$$a_m = \frac{V^2}{R_m} = \frac{4\pi^2 R_m}{T^2} \qquad (7.3)$$

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 Eq. (7.3), 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.

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 centre of the earth, we will have $a_m \propto R_m^{-2}$; $g \propto R_E^{-2}$ and we get

$$\frac{g}{a_m} = \frac{R_m^2}{R_E^2} \simeq 3600 \tag{7.4}$$

in agreement with a value of $g \simeq 9.8$ m s^{-2} and the value of a_m from Eq. (7.3). 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

$$|\mathbf{F}| = G\,\frac{m_1 \; m_2}{r^2} \tag{7.5}$$

Equation (7.5) 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. 7.3.

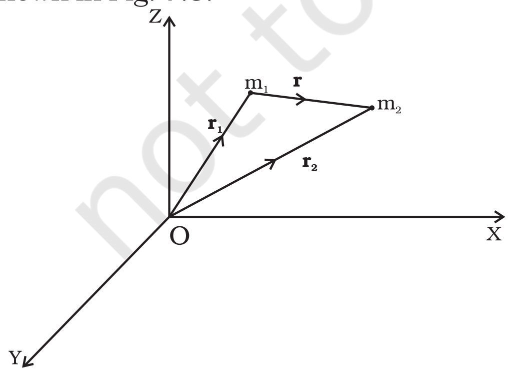


Fig. 7.3 *Gravitational force on m_1 due to m_2 is along* **r** *where the vector* **r** *is ($\mathbf{r}_2 - \mathbf{r}_1$).*

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. (7.5) 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 7.4.

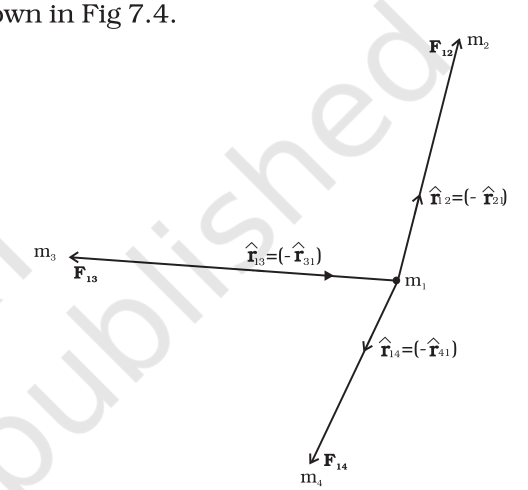


Fig. 7.4 *Gravitational force on point mass m_1 is the vector sum of the gravitational forces exerted by m_2, m_3 and m_4.*

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 7.2*** 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 $2m$ 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 = 1 m (see Fig. 7.5)

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

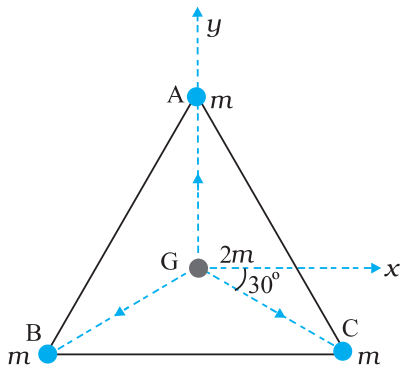


Fig. 7.5 Three equal masses are placed at the three vertices of the Δ ABC. A mass 2m is placed at the centroid G.

$$\mathbf{F}_{GA} = \frac{Gm(2m)}{1}\hat{\mathbf{j}}$$

$$\mathbf{F}_{GB} = \frac{Gm(2m)}{1}\left(-\hat{\mathbf{i}}\cos 30^o - \hat{\mathbf{j}}\sin 30^o\right)$$

$$\mathbf{F}_{GC} = \frac{Gm(2m)}{1}\left(+\hat{\mathbf{i}}\cos 30^o - \hat{\mathbf{j}}\sin 30^o\right)$$

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) Now if the mass at vertex A is doubled then

$$F'_{GA} = \frac{G2m.2m}{1}\hat{j} = 4Gm^2\hat{j}$$

$$F'_{GB} = F_{GB} \text{ and } F'_{GC} = F_{GC}$$

$$F'_R = F'_{GA} + F'_{GB} + F'_{GC}$$

$$F'_R = 2\,Gm^2\,\hat{j}$$

For the gravitational force between an extended object (like the earth) and a point mass, Eq. (7.5) 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.

(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.

7.4 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 Fig.7.6

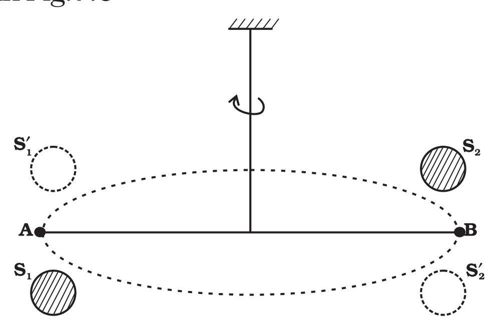


Fig. 7.6 Schematic drawing of Cavendish's experiment. S_1 and S_2 are large spheres which are kept on either side (shown shades) of the masses at A and B. When the big spheres are taken to the other side of the masses (shown by dotted circles), the bar AB rotates a little since the torque reverses direction. 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 centres. Thus if d is the separation between the centres 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.

$$F = G\frac{Mm}{d^2} \tag{7.6}$$

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

$$G\frac{Mm}{d^2}L = \tau\,\theta \tag{7.7}$$

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

$$G = 6.67\times10^{-11}\ \mathrm{N\,m^2/kg^2} \tag{7.8}$$

7.5 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 section 7.3. 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 centre.

For a point inside the earth, the situation is different. This is illustrated in Fig. 7.7.

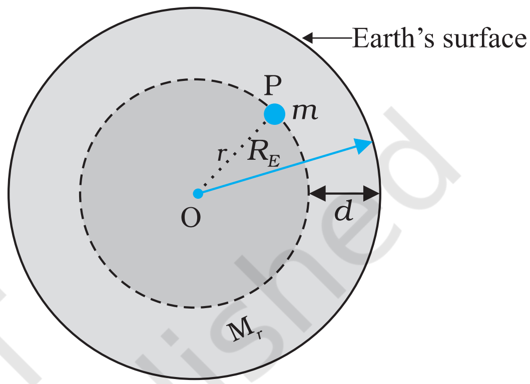


Fig. 7.7 *The mass m is in a mine located at a depth d below the surface of the Earth of mass M_E and radius R_E. We treat the Earth to be spherically symmetric.*

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 centre. Thus the force on the mass m at P has a magnitude

$$F = \frac{Gm\,(M_\mathrm{r})}{r^2} \tag{7.9}$$

We assume that the entire earth is of uniform density and hence its mass is $M_\mathrm{E} = \dfrac{4\pi}{3}R_E^3\,\rho$ 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 $\dfrac{4\pi}{3}\rho r^3$ and hence

$$F = G m\left(\frac{4p}{3} r\right)\frac{r^3}{r^2} = G m\left(\frac{M_E}{R_E^3}\right)\frac{r^3}{r^2}$$

$$= \frac{G m M_E}{R_E^{\ 3}} r \qquad (7.10)$$

If the mass m is situated on the surface of earth, then $r = R_E$ and the gravitational force on it is, from Eq. (7.10)

$$F = G\frac{M_E m}{R_E^2} \qquad (7.11)$$

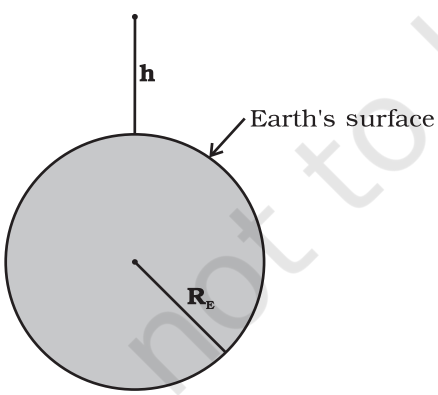
The acceleration experienced by the mass m, which is usually denoted by the symbol g is related to F by Newton's 2nd law by relation $F = mg$. Thus

$$g = \frac{F}{m} = \frac{GM_E}{R_E^2} \qquad (7.12)$$

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. (7.12). This is the reason why there is a popular statement regarding Cavendish : "Cavendish weighed the earth".

7.6 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. 7.8(a). The radius of the earth is denoted by R_E. Since this point is outside the earth,



(a)

Fig. 7.8 (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. (7.5) :

$$F(h) = \frac{GM_E m}{(R_E + h)^2} \qquad (7.13)$$

The acceleration experienced by the point mass is $F(h)/m \equiv g(h)$ and we get

$$g(h) = \frac{F(h)}{m} = \frac{GM_E}{(R_E + h)^2}. \qquad (7.14)$$

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. (7.14) :

$$g(h) = \frac{GM_E}{R_E^2(1 + h/R_E)^2} = g(1 + h/R_E)^{-2}$$

For $\frac{h}{R_E} << 1$, using binomial expression,

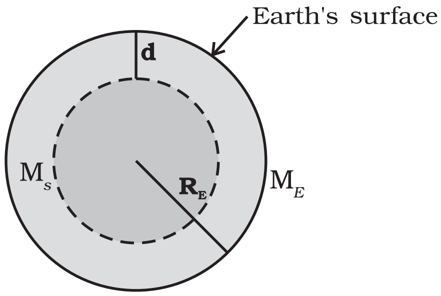
$$g(h) \cong g\left(1 - \frac{2h}{R_E}\right). \qquad (7.15)$$

Equation (7.15) 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. 7.8(b)), so that its distance from the centre 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,

$$M_s/M_E = (R_E - d)^3 \ / \ R_E^{\ 3} \qquad (7.16)$$

Since mass of a sphere is proportional to be cube of its radius.



(b)

Fig. 7.8 (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

$$F(d) = G\ M_s\ m\ /\ (R_E - d)^2 \qquad (7.17)$$

Substituting for M_s from above , we get

$$F(d) = G\ M_E\ m\,(R_E - d)\ /\ R_E^{\ 3} \qquad (7.18)$$

and hence the acceleration due to gravity at a depth d,

$$g(d) = \frac{F(d)}{m} \text{ is}$$

$$g(d) = \frac{F(d)}{m} = \frac{GM_E}{R_E^3}\ (R_E - d)$$

$$= g\frac{R_E - d}{R_E} = g(1 - d/R_E) \qquad (7.19)$$

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.

7.7 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 centre 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}

$$W_{12} = \text{Force} \times \text{displacement}$$
$$= mg\,(h_2 - h_1) \qquad (7.20)$$

If we associate a potential energy $W(h)$ at a point at a height h above the surface such that

$$W(h) = mgh + W_o \qquad (7.21)$$

(where W_o = constant) ;
then it is clear that

$$W_{12} = W(h_2) - W(h_1) \qquad (7.22)$$

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. (7.22). 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 centre of the earth is

$$F = \frac{G M_E m}{r^2} \qquad (7.23)$$

where M_E = mass of earth, m = mass of the particle and r its distance from the centre 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. (7.20)

$$W_{12} = \int_{r_1}^{r_2} \frac{G M m}{r^2} \mathrm{d}r$$

$$= -G M_E m \left(\frac{1}{r_2} - \frac{1}{r_1}\right) \qquad (7.24)$$

In place of Eq. (7.21), we can thus associate a potential energy $W(r)$ at a distance r, such that

$$W(r) = -\frac{G M_E m}{r} + W_1 \, , \qquad (7.25)$$

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. (7.22) and (7.24). 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

$$V = -\frac{G m_1 m_2}{r} \quad \text{(if we choose } V = 0 \text{ as } r \to \infty\text{)}$$

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 7.3 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. 7.9. We have four mass pairs at distance l and two diagonal pairs at distance $\sqrt{2}\, l$

Hence,

$$W(r) = -4\frac{G\,m^2}{l} - 2\frac{G\,m^2}{\sqrt{2}\,l}$$

$$= -\frac{2G\,m^2}{l}\left(2 + \frac{1}{\sqrt{2}}\right) = -5.41\frac{G\,m^2}{l}$$

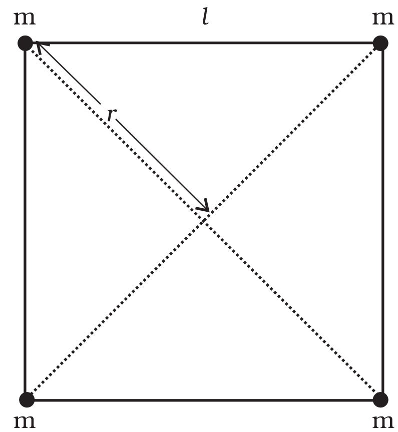


Fig. 7.9

The gravitational potential at the centre of the square $\left(r = \sqrt{2}\, l/2\right)$ is

$$U(r) = -4\sqrt{2}\,\frac{G\,m}{l}.$$

7.8 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

$$E\,(\infty) = W_1 + \frac{mV_f^2}{2} \qquad (7.26)$$

If the object was thrown initially with a speed V_i from a point at a distance $(h+R_E)$ from the centre of the earth (R_E = radius of the earth), its energy initially was

$$E\,(h + R_E) = \frac{1}{2}mV_i^2 - \frac{GmM_E}{(h + R_E)} + W_1 \qquad (7.27)$$

By the principle of energy conservation Eqs. (7.26) and (7.27) must be equal. Hence

$$\frac{mV_i^2}{2} - \frac{GmM_E}{(h+R_E)} = \frac{mV_f^2}{2} \tag{7.28}$$

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

$$\frac{mV_i^2}{2} - \frac{GmM_E}{(h+R_E)} \geq 0 \tag{7.29}$$

The minimum value of V_i corresponds to the case when the L.H.S. of Eq. (7.29) equals zero. Thus, the minimum speed required for an object to reach infinity (i.e. escape from the earth) corresponds to

$$\frac{1}{2} m \left(V_i^2\right)_{\min} = \frac{GmM_E}{h+R_E} \tag{7.30}$$

If the object is thrown from the surface of the earth, $h = 0$, and we get

$$(V_i)_{\min} = \sqrt{\frac{2GM_E}{R_E}} \tag{7.31}$$

Using the relation $g = GM_E / R_E^2$, we get

$$(V_i)_{\min} = \sqrt{2gR_E} \tag{7.32}$$

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 (7.32) 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 7.4 Two uniform solid spheres of equal radii R, but mass M and 4 M have a centre to centre separation 6 R, as shown in Fig. 7.10. 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.

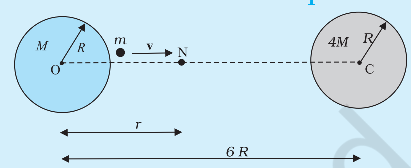


Fig. 7.10

Answer The projectile is acted upon by two mutually opposing gravitational forces of the two spheres. The neutral point N (see Fig. 7.10) is defined as the position where the two forces cancel each other exactly. If ON = r, we have

$$\frac{G\,M\,m}{r^2} = \frac{4\,G\,M\,m}{(6R-r)^2}$$

$(6R - r)^2 = 4r^2$

$6R - r = \pm 2r$

$r = 2R$ 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

$$\frac{1}{2}\upsilon^2 - \frac{GM}{R} - \frac{4GM}{5R} = -\frac{GM}{2R} - \frac{GM}{R}$$

or

$$v^2 = \frac{2GM}{R}\left(\frac{4}{5}-\frac{1}{2}\right)$$

$$v = \left(\frac{3GM}{5R}\right)^{1/2}$$

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.

7.9 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

$$\text{F(centripetal)} = \frac{mV^2}{(R_E + h)} \qquad (7.33)$$

directed towards the centre. This centripetal force is provided by the gravitational force, which is

$$\text{F(gravitation)} = \frac{G\,m\,M_E}{(R_E + h)^2} \qquad (7.34)$$

where M_E is the mass of the earth.

Equating R.H.S of Eqs. (7.33) and (7.34) and cancelling out m, we get

$$V^2 = \frac{G\,M_E}{(R_E + h)} \qquad (7.35)$$

Thus V decreases as h increases. From equation (7.35),the speed V for $h = 0$ is

$$V^2 \ (h=0) = GM/R_E = gR_E \qquad (7.36)$$

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. Its time period T therefore is

$$T = \frac{2\pi(R_E + h)}{V} = \frac{2\pi(R_E + h)^{3/2}}{\sqrt{G\,M_E}} \qquad (7.37)$$

on substitution of value of V from Eq. (7.35). Squaring both sides of Eq. (7.37), we get

$$T^2 = k\ (R_E + h)^3 \ \ (\text{where } k = 4\pi^2 / GM_E) \qquad (7.38)$$

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. (7.38). Hence, for such satellites, T is T_o, where

$$T_0 = 2\pi\sqrt{R_E / g} \qquad (7.39)$$

If we substitute the numerical values $g \simeq 9.8 \text{ m s}^{-2}$ and R_E = 6400 km., we get

$$T_0 = 2\pi\sqrt{\frac{6.4\times10^6}{9.8}} \text{ s}$$

Which is approximately 85 minutes.

Example 7.5 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. (7.38) with the sun's mass replaced by the martian mass M_m

$$T^2 = \frac{4\pi^2}{GM_m}R^3$$

$$M_m = \frac{4\pi^2}{G}\frac{R^3}{T^2}$$

$$= \frac{4\times(3.14)^2\times(9.4)^3\times10^{18}}{6.67\times10^{-11}\times(459\times60)^2}$$

$$M_m = \frac{4\times(3.14)^2\times(9.4)^3\times10^{18}}{6.67\times(4.59\times6)^2\times10^{-5}}$$

$= 6.48 \times 10^{23}$ 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 and Mars 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 7.6 Weighing the Earth : You are given the following data: $g = 9.81\ \text{ms}^{-2}$, $R_E = 6.37\times10^6$ m, the distance to the moon $R = 3.84\times10^8$ 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. (7.12) we have

$$M_E = \frac{g R_E^2}{G}$$

$$= \frac{9.81\times\left(6.37\times10^6\right)^2}{6.67\times10^{-11}}$$

$= 5.97\times 10^{24}$ kg.

The moon is a satellite of the Earth. From the derivation of Kepler's third law [see Eq. (7.38)]

$$T^2 = \frac{4\pi^2 R^3}{G M_E}$$

$$M_E = \frac{4\pi^2 R^3}{G T^2}$$

$$= \frac{4\times3.14\times3.14\times(3.84)^3\times10^{24}}{6.67\times10^{-11}\times(27.3\times24\times60\times60)^2}$$

$$= 6.02\times10^{24}\ \text{kg}$$

Both methods yield almost the same answer, the difference between them being less than 1%.

Example 7.7 Express the constant k of Eq. (7.38) in days and kilometres. Given $k = 10^{-13}\ \text{s}^2\ \text{m}^{-3}$. The moon is at a distance of 3.84×10^5 km from the earth. Obtain its time-period of revolution in days.

Answer Given

$k = 10^{-13}\ \text{s}^2\ \text{m}^{-3}$

$$= 10^{-13}\left[\frac{1}{(24\times60\times60)^2}\text{d}^2\right]\left[\frac{1}{(1/1000)^3\ \text{km}^3}\right]$$

$= 1.33\times10^{-14}\ \text{d}^2\ \text{km}^{-3}$

Using Eq. (7.38) and the given value of k, the time period of the moon is

$T^2 = (1.33\times10^{-14})(3.84\times10^5)^3$

$T = 27.3$ d

Note that Eq. (7.38) 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.

7.10 ENERGY OF AN ORBITING SATELLITE

Using Eq. (7.35), the kinetic energy of the satellite in a circular orbit with speed v is

$$K{\bullet}E = \frac{1}{2}m v^2$$

$$= \frac{Gm M_E}{2(R_E+h)}, \qquad (7.40)$$

Considering gravitational potential energy at infinity to be zero, the potential energy at distance (R_e+h) from the centre of the earth is

$$P.E = -\frac{G m M_E}{(R_E+h)} \qquad (7.41)$$

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

$$E = K.E + P.E = -\frac{G m M_E}{2(R_E+h)} \qquad (7.42)$$

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 7.8 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

$$\Delta V = V_f - V_i = -6.25 \times 10^9\,\text{J}$$

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, ….M_n etc. we use the principle of superposition. Let F_1, F_2, ….F_n be the individual forces due to M_1, M_2, ….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

$$F_R = F_1 + F_2 + \ldots\ldots + F_n = \sum_{i=1}^{n} F_i$$

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.

Physical Quantity	Symbol	Dimensions	Unit	Remarks
Gravitational Constant	G	$[M^{-1}L^3T^{-2}]$	N m^2 kg^{-2}	6.67×10^{-11}
Gravitational Potential Energy	V(r)	$[M\,L^2T^{-2}]$	J	$-\frac{GMm}{r}$ (scalar)
Gravitational Potential	U(r)	$[L^2T^{-2}]$	J kg^{-1}	$-\frac{GM}{r}$ (scalar)
Gravitational Intensity	**E** or **g**	$[LT^{-2}]$	m s^{-2}	$\frac{GM}{r^2}\hat{\mathbf{r}}$ (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. (7.1) 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. (7.38)].
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 $m g h$ 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

7.1 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 ?

7.2 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.

7.3 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 ?

7.4 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.

7.5 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.

7.6 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.

7.7 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?

7.8 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.

7.9 Which of the following symptoms is likely to afflict an astronaut in space (a) swollen feet, (b) swollen face, (c) headache, (d) orientational problem.

7.10 In the following two exercises, choose the correct answer from among the given ones: The gravitational intensity at the centre of a hemispherical shell of uniform mass density has the direction indicated by the arrow (see Fig 7.11) (i) a, (ii) b, (iii) c, (iv) 0.

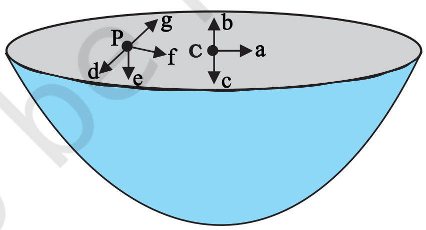


Fig. 7.11

7.11 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.

7.12 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).

7.13 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.

7.14 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 ?

7.15 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 ?

7.16 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 ?

7.17 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 $= 6.0 \times 10^{24}$ kg; mean radius of the earth $= 6.4 \times 10^6$ m; $G = 6.67 \times 10^{-11}$ N m^2 kg^{-2}.

7.18 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.

7.19 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 $= 6.0\times10^{24}$ kg; radius of the earth $= 6.4 \times 10^6$ m; $G = 6.67 \times 10^{-11}$ N m^2 kg^{-2}.

7.20 Two stars each of one solar mass ($= 2\times10^{30}$ 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).

7.21 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 ?

Chapter Eight

Mechanical Properties of Solids

8.1 INTRODUCTION

In Chapter 6, 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.

8.2 STRESS AND STRAIN

When forces are applied on a body in such a manner that the body is still in static equilibrium, 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 normal to the cross–section and A is the area of cross section of the body,

Magnitude of the stress = F/A (8.1)

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. 8.1. In Fig.8.1(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**.

$$\text{Longitudinal strain} = \frac{\Delta L}{L} \qquad (8.2)$$

However, if two equal and opposite deforming forces are applied parallel to the cross-sectional area of the cylinder, as shown in Fig. 8.1(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**.

As a result of applied tangential force, there is a relative displacement Δx between opposite faces of the cylinder as shown in the Fig. 8.1(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.

$$\text{Shearing strain} = \frac{\Delta x}{L} = \tan\theta \qquad (8.3)$$

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. 8.2 (c).

Thus, shearing strain $= \tan\theta \approx \theta$ (8.4)

In Fig. 8.1 (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

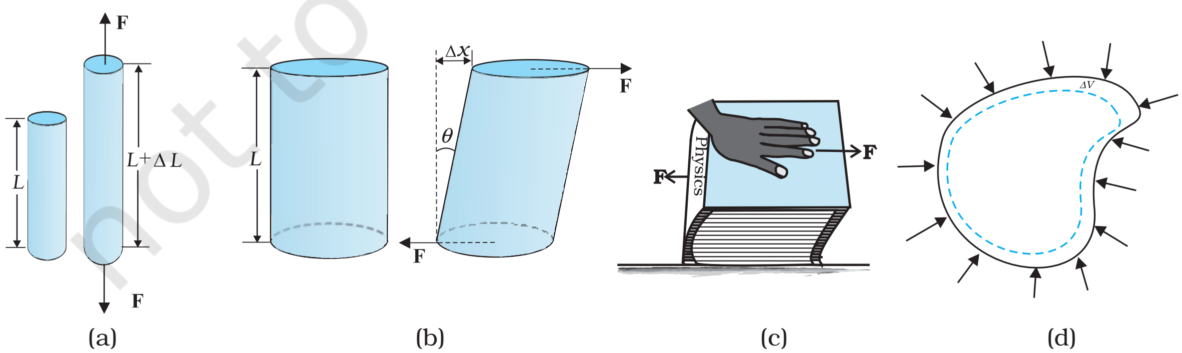


Fig. 8.1 *(a) A cylindrical body under tensile stress elongates by ΔL (b) Shearing stress on a cylinder deforming it by an angle θ (c) A body subjected to shearing stress (d) A solid body under a stress normal to the surface at every point (hydraulic stress). The volumetric strain is $\Delta V/V$, but there is no change in shape.*

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 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).

$$\text{Volume strain} = \frac{\Delta V}{V} \qquad (8.5)$$

Since the strain is a ratio of change in dimension to the original dimension, it has no units or dimensional formula.

8.3 HOOKE'S LAW

Stress and strain take different forms in the situations depicted in the Fig. (8.1). For small deformations the stress and strain are proportional to each other. This is known as Hooke's law.

Thus,

stress $\propto$ strain

$$\text{stress} = k \times \text{strain} \qquad (8.6)$$

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 exhibit this linear relationship.

8.4 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. 8.2. 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.

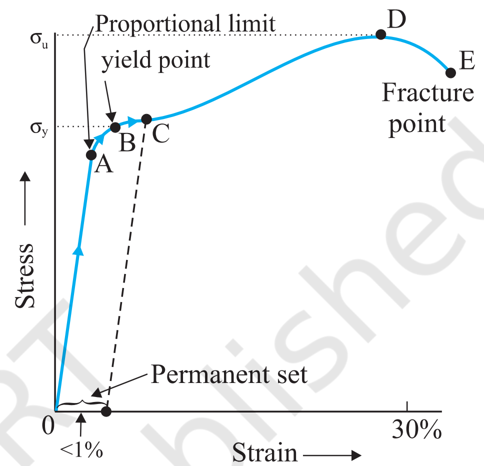


Fig. 8.2 A typical stress-strain curve for a metal.

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** (σ_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** (σ_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**.

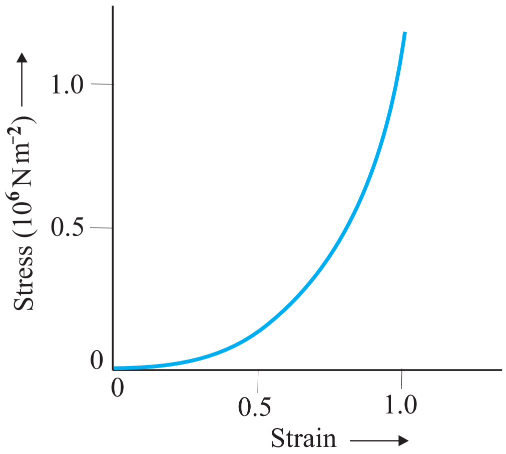


Fig. 8.3 *Stress-strain curve for the elastic tissue of Aorta, the large tube (vessel) carrying blood from the heart.*

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. 8.3 shows stress-strain curve for the elastic tissue of aorta, present in the heart. Note that although elastic region is very large, the material 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**.

8.5 ELASTIC MODULI

The proportional region within the elastic limit of the stress-strain curve (region OA in Fig. 8.2) 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.

8.5.1 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.

$$Y = \frac{\sigma}{\varepsilon} \qquad (8.7)$$

From Eqs. (8.1) and (8.2), we have

$$Y = (F/A)/(\Delta L/L)$$
$$= (F \times L)\ /(A \times \Delta L) \qquad (8.8)$$

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 8.1 gives the values of Young's moduli and yield strengths of some material.

From the data given in Table 8.1, it is noticed that for metals Young's moduli are large.

Table 8.1 **Young's moduli and yield strenghs of some material**

Substance	Density ρ (kg m^{-3})	Young's modulus Y (10^9N m^{-2})	Ultimate strength, σ_u (10^6 N m^{-2})	Yield strength σ_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	9.4	170	—
Polystyrene	1050	3	48	—

 # Substance tested under compression

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 preferred in heavy-duty machines and in structural designs. Wood, bone, concrete and glass have rather small Young's moduli.

Example 8.1 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

$$\text{Stress} = \frac{F}{A} = \frac{F}{\pi r^2}$$

$$= \frac{100 \times 10^3 \text{N}}{3.14 \times \left(10^{-2} \text{m}\right)^2}$$

$$= 3.18 \times 10^8 \text{ N m}^{-2}$$

The elongation,

$$\Delta L = \frac{(F/A)L}{Y}$$

$$= \frac{\left(3.18 \times 10^8 \text{N m}^{-2}\right)(1\text{m})}{2 \times 10^{11} \text{N m}^{-2}}$$

$$= 1.59 \times 10^{-3} \text{ m}$$

$$= 1.59 \text{ mm}$$

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 8.2 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. (8.7) 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 8.3 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. 8.4). 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.

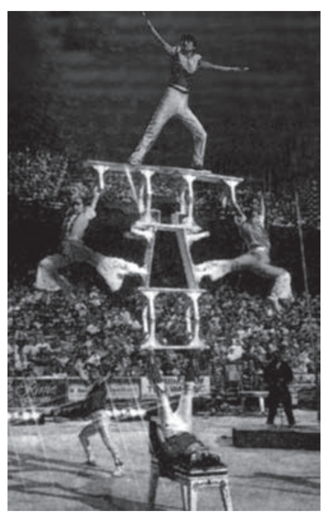


Fig. 8.4 *Human pyramid in a circus.*

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%.

8.5.2 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

$$G = (F/A)/(\Delta x/L)$$
$$= (F \times L)/(A \times \Delta x) \qquad (8.10)$$

Similarly, from Eq. (9.4)

$$G = (F/A)/\theta$$
$$= F/(A \times \theta) \qquad (8.11)$$

The shearing stress σ_s can also be expressed as

$$\sigma_s = G \times \theta \qquad (8.12)$$

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 8.2 **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 8.4 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. 8.6. 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

$$= (9.4 \times 10^4 \text{ N}/0.05 \text{ m}^2)$$
$$= 1.80 \times 10^6 \text{ N.m}^{-2}$$

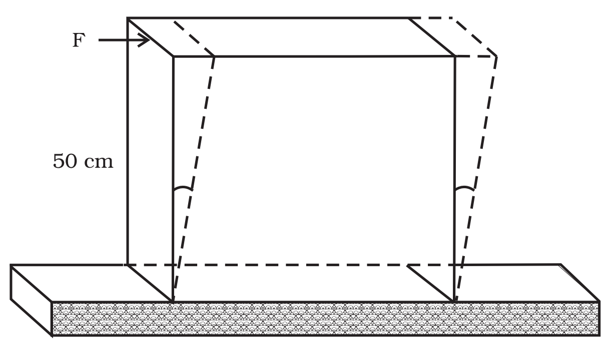


Fig. 8.5

We know that shearing strain = $(\Delta x/L)$= Stress / G.

Therefore the displacement Δx = (Stress × L)/G

$= (1.8 \times 10^6 \text{ N m}^{-2} \times 0.5\text{m})/(5.6 \times 10^9 \text{ N m}^{-2})$

$= 1.6 \times 10^{-4}$ m = 0.16 mm

8.5.3 Bulk Modulus

In Section (8.3), 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. (8.5)]. The ratio of hydraulic stress to the corresponding hydraulic strain is called *bulk modulus*. It is denoted by symbol B.

$$B = -\,p/(\Delta V/V) \qquad (8.12)$$

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 8.3.

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.

$$k = (1/B) = -\,(1/\Delta p) \times (\Delta V/V) \qquad (8.13)$$

It can be seen from the data given in Table 8.3 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 8.3 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}

Table 8.4 Stress, strain and various elastic moduli

Type of stress	Stress	Strain	Change in shape	Change in volume	Elastic Modulus	Name of Modulus	State of Matter
Tensile or compressive ($\sigma = F/A$)	Two equal and opposite forces perpendicular to opposite faces	Elongation or compression parallel to force direction ($\Delta L/L$) (longitudinal strain)	Yes	No	$Y = (F\,L)/(A\,\Delta L)$	Young's modulus	Solid
Shearing ($\sigma_s = F/A$)	Two equal and opposite forces parallel to oppoiste surfaces forces in each case such that total force and total torque on the body vanishes	Pure shear, θ	Yes	No	$G = F/(A\,\theta)$	Shear modulus or modulus of rigidity	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

Thus, solids are the least compressible, whereas, gases are the 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 8.4 shows the various types of stress, strain, elastic moduli, and the applicable state of matter at a glance.

Example 8.5 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

$$\Delta V/V = \text{stress}/B = (3 \times 10^7 \text{ N m}^{-2})/(2.2 \times 10^9 \text{ N m}^{-2})$$
$$= 1.36 \times 10^{-2} \text{ or } 1.36 \text{ \%}$$

8.5.4 Poisson's Ratio

The strain perpendicular to the applied force is called **lateral strain**. Simon Poisson pointed out that within the elastic limit, lateral strain is directly proportional to the longitudinal strain. The ratio of the lateral strain to the longitudinal strain in a stretched wire is called **Poisson's ratio.** If the original diameter of the wire is d and the contraction of the diameter under stress is Δd, the lateral strain is $\Delta d/d$. If the original length of the wire is L and the elongation under stress is ΔL, the longitudinal strain is $\Delta L/L$. Poisson's ratio is then $(\Delta d/d)/(\Delta L/L)$ or $(\Delta d/\Delta L)$ (L/d). Poisson's ratio is a ratio of two strains; it is a pure number and has no dimensions or units. Its value depends only on the nature of material. For steels the value is between 0.28 and 0.30, and for aluminium alloys it is about 0.33.

8.5.5 Elastic Potential Energy in a Stretched Wire

When a wire is put under a tensile stress, work is done against the inter-atomic forces. This work is stored in the wire in the form of elastic potential energy. When a wire of original length L and area of cross-section A is subjected to a deforming force F along the length of the wire, let the length of the wire be elongated by l. Then from Eq. (8.8), we have $F = YA \times (l/L)$. Here Y is the Young's modulus of the material of the wire. Now for a further elongation of infinitesimal small length dl, work done dW is F dl or YAdl/L. Therefore, the amount of work done (W) in increasing the length of the wire from L to $L + l$, that is from $l = 0$ to $l = l$ is

$$W = \int_0^l \frac{YAl}{L} dl = \frac{YA}{2} \times \frac{l^2}{L}$$

$$W = \frac{1}{2} \times Y \times \left(\frac{l}{L}\right)^2 \times AL$$

$$= \frac{1}{2} \times \text{Young's modulus} \times \text{strain}^2 \times \text{volume of the wire}$$

$$= \frac{1}{2} \times \text{stress} \times \text{strain} \times \text{volume of the wire}$$

This work is stored in the wire in the form of elastic potential energy (U). Therefore the elastic potential energy per unit volume of the wire (u) is

$$u = \frac{1}{2} \sigma \varepsilon \qquad (8.14)$$

8.6 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 8.1, we find that mild steel has a yield strength (σ_y) of about 300×10^6 N m^{-2}. Thus, the area of cross-section (A) of the rope should at least be

$$A \geq W/\sigma_y = Mg/\sigma_y \tag{8.15}$$
$$= (10^4 \text{ kg} \times 9.8 \text{ m s}^{-2})/(300 \times 10^6 \text{ N m}^{-2})$$
$$= 3.3 \times 10^{-4} \text{ m}^2$$

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 the 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. 8.6. 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

$$\delta = W l^3/(4bd^3 Y) \tag{8.16}$$

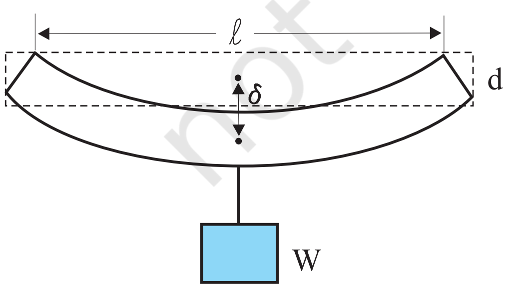


Fig. 8.6 *A beam supported at the ends and loaded at the centre.*

This relation can be derived using what you have already learnt and a little calculus. From Eq. (8.16), 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. 8.7(b). This is called buckling. To avoid this, a common compromise is the cross-sectional shape shown in Fig. 8.7(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.

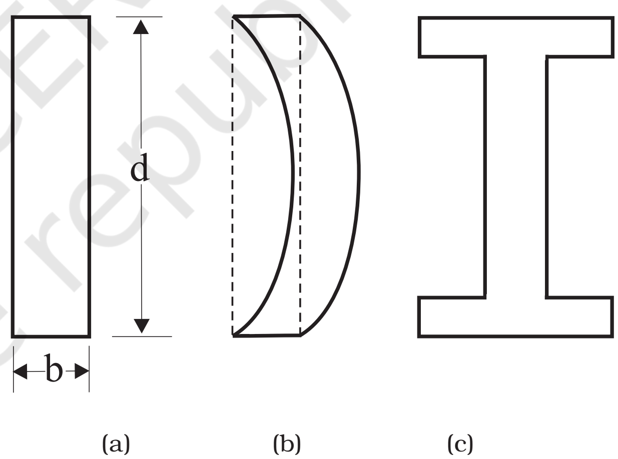


Fig. 8.7 *Different cross-sectional shapes of a beam. (a) Rectangular section of a bar; (b) A thin bar and how it can buckle; (c) Commonly used section for a load bearing bar.*

The use of pillars or columns is also very common in buildings and bridges. A pillar with rounded ends as shown in Fig. 8.9(a) supports less load than that with a distributed shape at the ends [Fig. 8.9(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 material, etc.

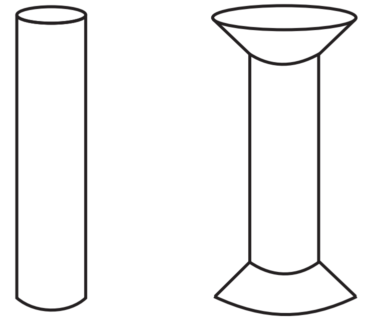

(a) (b)

Fig. 8.8 *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

$$h\rho g = 30 \times 10^7 \text{ N m}^{-2}.$$

$$h = 30 \times 10^7 \text{ N m}^{-2}/(3 \times 10^3 \text{ kg m}^{-3} \times 10 \text{ m s}^{-2})$$

$$= 10 \text{ km}$$

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

8.1 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?

8.2 Figure 8.9 shows the strain-stress curve for a given material. What are (a) Young's modulus and (b) approximate yield strength for this material?

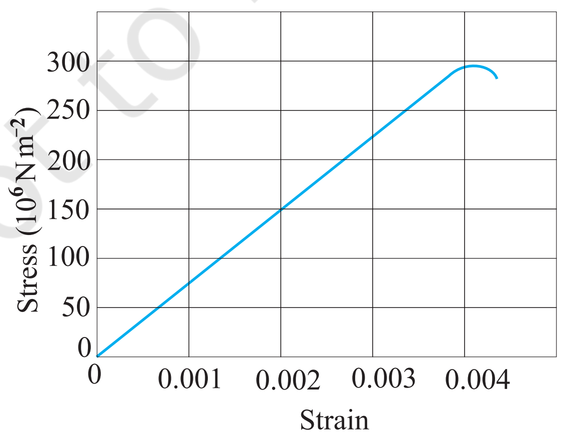


Fig. 8.9

8.3 The stress-strain graphs for materials *A* and *B* are shown in Fig. 8.10.

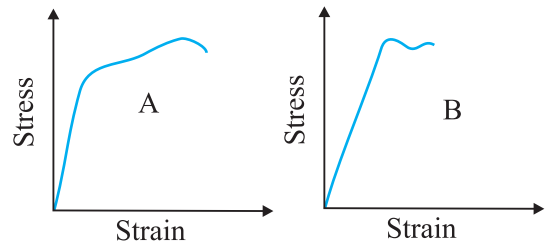


Fig. 8.10

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?

8.4 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.

8.5 Two wires of diameter 0.25 cm, one made of steel and the other made of brass are loaded as shown in Fig. 8.11. 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.

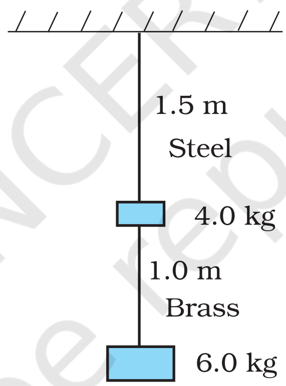


Fig. 8.11

8.6 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?

8.7 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.

8.8 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?

8.9 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 ?

8.10 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.

8.11 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.

8.12 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.

8.13 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}?

8.14 Compute the fractional change in volume of a glass slab, when subjected to a hydraulic pressure of 10 atm.

8.15 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.

8.16 How much should the pressure on a litre of water be changed to compress it by 0.10%? carry one quarter of the load.

11087CH10

CHAPTER NINE

MECHANICAL PROPERTIES OF FLUIDS

9.1 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.

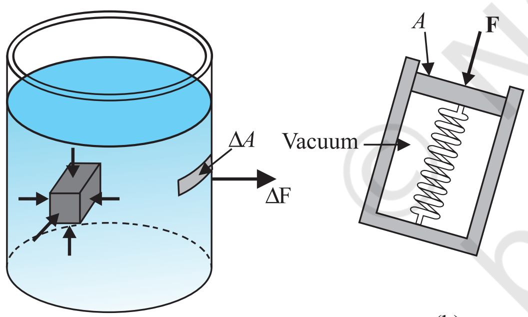
Shear stress can change the shape of a solid keeping its volume fixed. The key property of fluids is that they offer very little resistance to shear stress; their shape changes by application of very small shear stress. The shearing stress of fluids is about million times smaller than that of solids.

9.2 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 impact 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.9.1(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. 9.1(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.



(a) (b)

Fig. 9.1 *(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.

$$P_{av} = \frac{F}{A} \tag{9.1}$$

In principle, the piston area can be made arbitrarily small. The pressure is then defined in a limiting sense as

$$P = \lim_{\Delta A \to 0} \frac{\Delta F}{\Delta A} \tag{9.2}$$

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. (9.1) and (9.2). 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,

$$\rho = \frac{m}{V} \tag{9.3}$$

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^{o}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^{o}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 9.1.

Table 9.1 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^{o}C) and 1 atm pressure.*

Example 9.1 The two thigh bones (femurs), each of cross-sectional area 10 cm² 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 $A = 2 \times 10\ \text{cm}^2 = 20 \times 10^{-4}\ \text{m}^2$. The force acting on them is $F = 40$ kg wt = 400 N (taking $g = 10\ \text{m s}^{-2}$). This force is acting vertically down and hence, normally on the femurs. Thus, the average pressure is

$$P_{av} = \frac{F}{A} = 2 \times 10^5\ \text{N m}^{-2}$$

9.2.1 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.

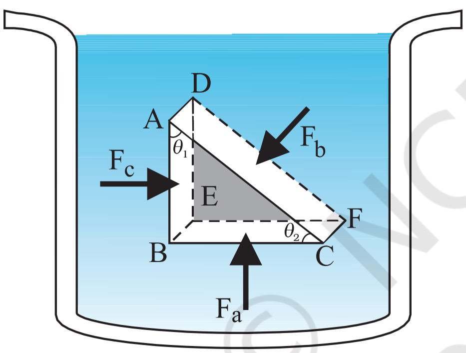


Fig. 9.2 *Proof of Pascal's law. ABC-DEF is an element of the interior of a fluid at rest. This element is in the form of a right-angled prism. The element is small so that the effect of gravity can be ignored, but it has been enlarged for the sake of clarity.*

Fig. 9.2 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. 9.2 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,

$$\frac{F_b}{A_b} = \frac{F_c}{A_c} = \frac{F_a}{A_a}; \qquad P_b = P_c = P_a \tag{9.4}$$

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 in a horizontal plane.

9.2.2 Variation of Pressure with Depth

Consider a fluid at rest in a container. In Fig. 9.3 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

$$(P_2 - P_1)\ A = mg \tag{9.5}$$

Now, if ρ is the mass density of the fluid, we have the mass of fluid to be $m = \rho V = \rho hA$ so that

$$P_2 - P_1 = \rho gh \tag{9.6}$$

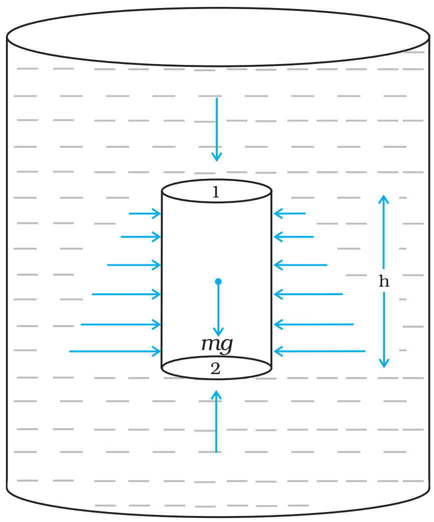


Fig.9.3 *Fluid under gravity. The effect of gravity is illustrated through pressure on a vertical cylindrical column.*

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. (9.6) gives

$$P = P_a + \rho g h \qquad (9.7)$$

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. (9.7). 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.9.4] 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.

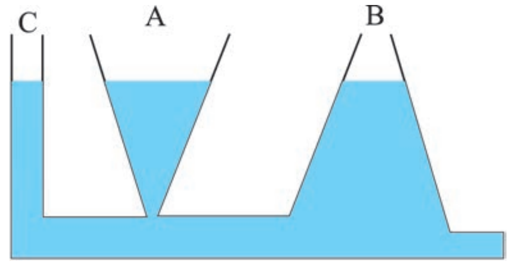


Fig 9.4 *Illustration of hydrostatic paradox. The three vessels A, B and C contain different amounts of liquids, all upto the same height.*

Example 9.2 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. (9.7)

$P = P_a + \rho gh$

= 1.01 × 10^5 Pa + 1000 kg m^{-3} × 10 m s^{-2} × 10 m

= 2.01 × 10^5 Pa

≈ 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.

9.2.3 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.9.5 (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. Thus, the pressure at Point A=0. The pressure inside the coloumn at Point B must be the same as the pressure at Point C, which is atmospheric pressure, P_a.

$$P_a = \rho g h \qquad (9.8)$$

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. (9.8). 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 the other end is connected to the system whose pressure we want to measure [see Fig. 9.5 (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. (9.8) and is proportional to manometer height h.

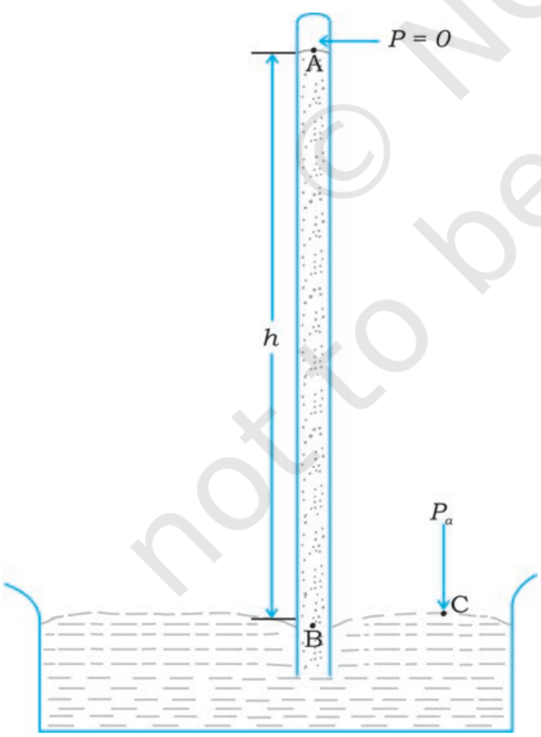
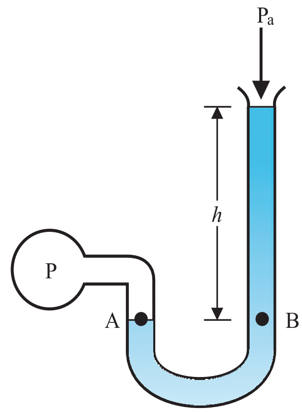


Fig 9.5 *(a) The mercury barometer.*



(b) The open tube manometer

Fig 9.5 *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 9.3 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. (9.7)

$$\rho g h = 1.29 \text{ kg m}^{-3} \times 9.8 \text{ m s}^2 \times h \text{ m} = 1.01 \times 10^5 \text{ Pa}$$

$$\therefore h = 7989 \text{ m} \approx 8 \text{ km}$$

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 9.4 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 = 10 m s^{-2}.)

Answer Here $h = 1000$ m and $\rho = 1.03 \times 10^3$ kg m^{-3}.

(a) From Eq. (9.6), absolute pressure

$P = P_a + \rho g h$
$= 1.01 \times 10^5$ Pa
$+ 1.03 \times 10^3$ kg m^{-3} $\times$ 10 m s^{-2} $\times$ 1000 m
$= 104.01 \times 10^5$ Pa
≈ 104 atm

(b) Gauge pressure is $P - P_a = \rho g h = P_g$

$P_g = 1.03 \times 10^3$ kg m^{-3} $\times$ 10 ms^2 $\times$ 1000 m
$= 103 \times 10^5$ Pa
≈ 103 atm

(c) The pressure outside the submarine is $P = P_a + \rho g h$ and the pressure inside it is P_a. Hence, the net pressure acting on the window is gauge pressure, $P_g = \rho g h$. 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

9.2.4 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 [Fig. 9.6 (a)]. 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.

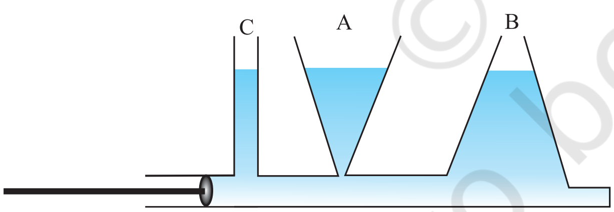


Fig 9.6 *(a) Whenever external pressure is applied on any part of a fluid in a vessel, it is equally transmitted in all directions.*

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 another form of the Pascal's law and it 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. 9.6 (b), 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 $P = \dfrac{F_1}{A_1}$ 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) $F_2 = PA_2 = \dfrac{F_1 A_2}{A_1}$. By changing the force at A_1, the platform can be moved up or down. Thus, the applied force has been increased by a factor of $\dfrac{A_2}{A_1}$ and this factor is the mechanical advantage of the device. The example below clarifies it.

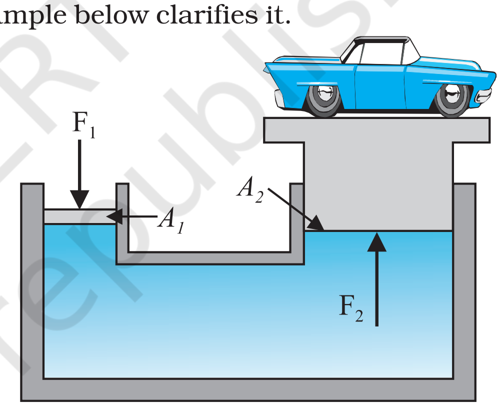


Fig 9.6 *(b) Schematic diagram illustrating the principle behind the hydraulic lift, a device used to lift heavy loads.*

> ***Example 9.5*** 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,

$$F_2 = \frac{A_2}{A_1}F_1 = \frac{\pi\left(3/2\times10^{-2}\,\text{m}\right)^2}{\pi\left(1/2\times10^{-2}\,\text{m}\right)^2}\times10\,\text{N}$$

$$= 90\ \text{N}$$

(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.

$$L_1A_1 = L_2A_2$$

$$L_2 = \frac{A_1}{A_2}L_1 = \frac{\pi\left(1/2\times10^{-2}\,\text{m}\right)^2}{\pi\left(3/2\times10^{-2}\,\text{m}\right)^2}\times6\times10^{-2}\,\text{m}$$

$$\simeq 0.67\times10^{-2}\,\text{m} = 0.67\ \text{cm}$$

Note, atmospheric pressure is common to both pistons and has been ignored.

Example 9.6 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 9.7). If the mass of the car to be lifted is 1350 kg, calculate F_1. What is the pressure necessary to accomplish this task? ($g = 9.8\ \text{ms}^{-2}$).

Answer Since pressure is transmitted undiminished throughout the fluid,

$$F_1 = \frac{A_1}{A_2}F_2 = \frac{\pi(5\times10^{-2}\,\text{m})^2}{\pi(15\times10^{-2}\,\text{m})^2}(1350\,\text{kg}\times9.8\,m\,s^{-2})$$

$$= 1470\ \text{N}$$

$$\approx 1.5\times10^3\ \text{N}$$

The air pressure that will produce this force is

$$P = \frac{F_1}{A_1} = \frac{1.5\times10^3\,\text{N}}{\pi\left(5\times10^{-2}\right)^2\,\text{m}} = 1.9\times10^5\,\text{Pa}$$

This is almost double the atmospheric pressure.

Hydraulic brakes in automobiles also work on the same principle. When we apply a little 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.

9.3 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.

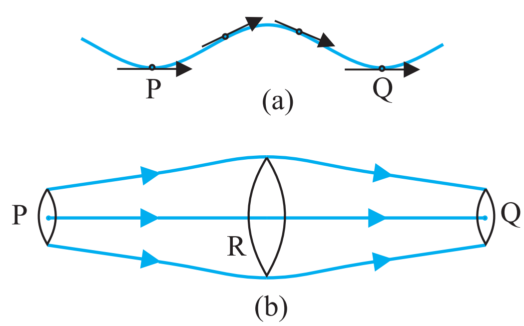


Fig. 9.7 *The meaning of streamlines. (a) A typical trajectory of a fluid particle. (b) A region of streamline flow.*

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.9.7 (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.9.7 (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,

$$\rho_P A_P v_P \Delta t = \rho_R A_R v_R \Delta t = \rho_Q A_Q v_Q \Delta t \tag{9.9}$$

For flow of incompressible fluids

$$\rho_P = \rho_R = \rho_Q$$

Equation (9.9) reduces to

$$A_P v_P = A_R v_R = A_Q v_Q \tag{9.10}$$

which is called the **equation of continuity** and it is a statement of conservation of mass in flow of incompressible fluids. In general

$$Av = \text{constant} \tag{9.11}$$

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 9.7b) 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 9.8 displays streamlines for some typical flows. For example, Fig. 9.8(a) describes a laminar flow where the velocities at different points in the fluid may have different magnitudes but their directions are parallel. Figure 9.8 (b) gives a sketch of turbulent flow.

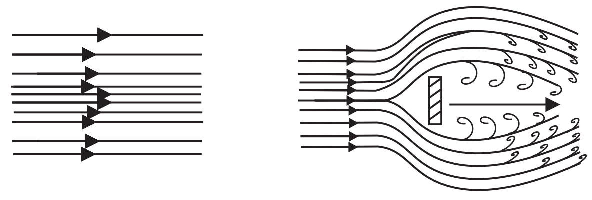


Fig. 9.8 *(a) Some streamlines for fluid flow. (b) A jet of air striking a flat plate placed perpendicular to it. This is an example of turbulent flow.*

9.4 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. 9.9. 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

$$\Delta K = \frac{1}{2}\,\rho\,\Delta V\,(v_2^2 - v_1^2)$$

We can employ the work – energy theorem (Chapter 6) to this volume of the fluid and this yields

$$(P_1 - P_2)\,\Delta V = \frac{1}{2}\,\rho\,\Delta V(v_2^2 - v_1^2) + \rho g \Delta V\,(h_2 - h_1)$$

We now divide each term by ΔV to obtain

$$(P_1 - P_2) = \frac{1}{2}\,\rho\,(v_2^2 - v_1^2) + \rho g\,(h_2 - h_1)$$

We can rearrange the above terms to obtain

$$P_1 + \frac{1}{2}\,\rho v_1^2 + \rho g h_1 = P_2 + \frac{1}{2}\,\rho v_2^2 + \rho g h_2 \qquad (9.12)$$

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

$$P + \frac{1}{2}\,\rho v^2 + \rho g h = \text{constant} \qquad (9.13)$$

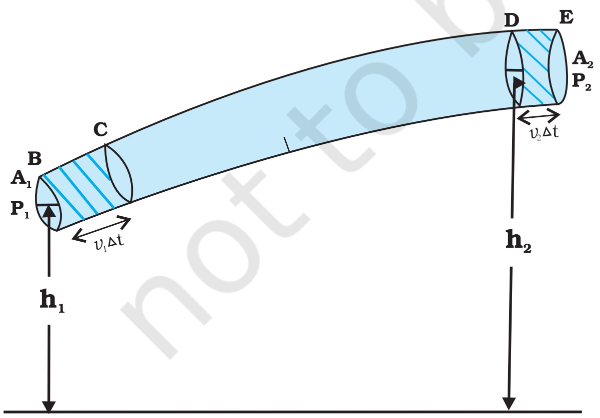


Fig. 9.9 *The flow of an ideal fluid in a pipe of varying cross section. The fluid in a section of length $v_1\Delta t$ moves to the section of length $v_2\Delta t$ in time Δt.*

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 $\left(\frac{\rho v^2}{2}\right)$ 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. (9.6).

9.4.1 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. 9.10). The air above the liquid, whose surface is at height y_2, is at pressure P. From the equation of continuity [Eq. (9.10)] we have

$$v_1 A_1 = v_2 A_2$$

$$v_2 = \frac{A_1}{A_2} v_1$$

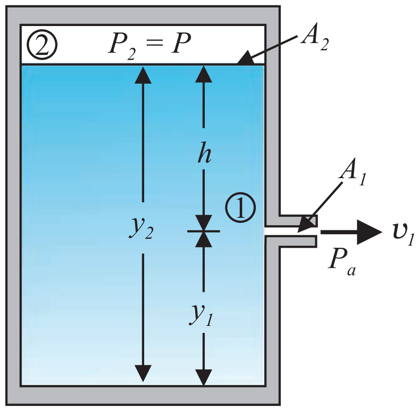


Fig. 9.10 *Torricelli's law. The speed of efflux, v_1, from the side of the container is given by the application of Bernoulli's equation. If the container is open at the top to the atmosphere then* $v_1 = \sqrt{2gh}$.

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. (9.12)

$$P_a + \frac{1}{2}\rho v_1^2 + \rho g y_1 = P + \rho g y_2$$

Taking $y_2 - y_1 = h$ we have

$$v_1 = \sqrt{2g\,h + \frac{2(P - P_a)}{\rho}} \tag{9.14}$$

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

$$v_1 = \sqrt{2gh} \tag{9.15}$$

This is also the speed of a freely falling body. Equation (9.15) represents **Torricelli's law**.

9.4.2 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. 9.11(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 9.11(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 the ball is larger and below it is smaller (see Section 9.3). 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 there is a net upward force on the ball. This dynamic lift due to spining is called **Magnus effect**.

Aerofoil or lift on aircraft wing: Figure 9.11 (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. 9.11 (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.

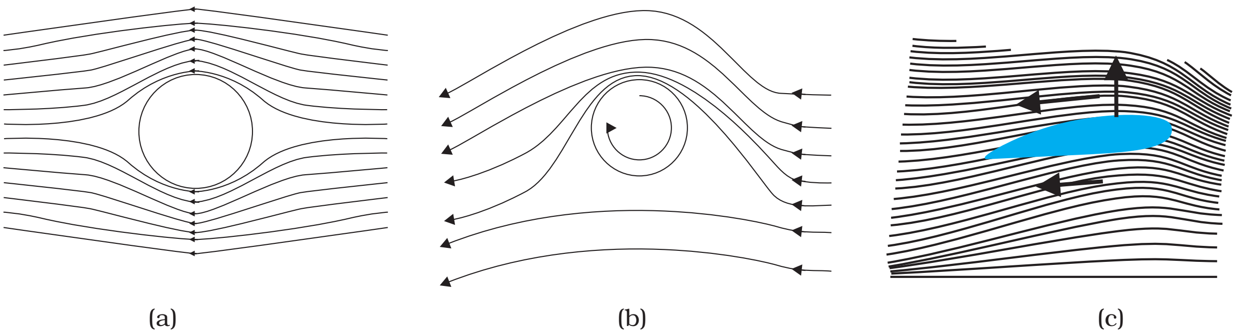


Fig 9.11 *(a) Fluid streaming past a static sphere. (b) Streamlines for a fluid around a sphere spinning clockwise. (c) Air flowing past an aerofoil.*

Example 9.7 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 $\rho = 1.2$ kg m^{-3}]

Answer (a) The weight of the Boeing aircraft is balanced by the upward force due to the pressure difference

$\Delta P \quad A = 3.3 \times 10^5 \text{kg} \times 9.8$

$\Delta P = (3.3 \times 10^5 \text{ kg} \times 9.8 \text{ m s}^{-2}) / 500 \text{ m}^2$

$= 6.5 \times 10^3 \text{ Nm}^{-2}$

(b) We ignore the small height difference between the top and bottom sides in Eq. (9.12). The pressure difference between them is then

$$\Delta P = \frac{\rho}{2}\left(v_2^2 - v_1^2\right)$$

where v_2 is the speed of air over the upper surface and v_1 is the speed under the bottom surface.

$$\left(v_2 - v_1\right) = \frac{2\Delta P}{\rho\left(v_2 + v_1\right)}$$

Taking the average speed

$v_{av} = (v_2 + v_1)/2 = 960 \text{ km/h} = 267 \text{ m s}^{-1}$,

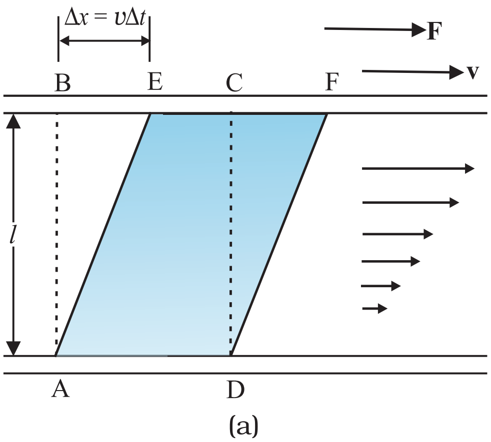
we have

$$\left(v_2 - v_1\right)/v_{av} = \frac{\Delta P}{\rho v_{av}^2} \approx 0.08$$

The speed above the wing needs to be only 8 % higher than that below.

9.5 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. 9.12 (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. 9.12 (b). The velocity on a cylindrical surface in a tube is constant.



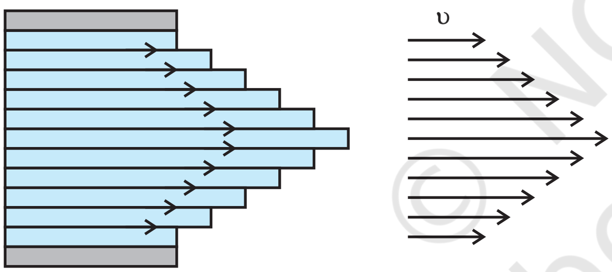


Fig 9.12 *(a) A layer of liquid sandwiched between two parallel glass plates, in which the lower plate is fixed and the upper one is moving to the right with velocity* **v**
(b) velocity distribution for viscous flow in a pipe.

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.

$$\eta = \frac{F/A}{v/l} = \frac{Fl}{vA} \tag{9.16}$$

The SI unit of viscosity is poiseiulle (Pl). Its other units are N s m^{-2} or Pa s. The dimensions of viscosity are [ML^{-1}T^{-1}]. Generally, thin liquids, like water, alcohol, etc., are less viscous than thick liquids, like coal tar, blood, glycerine, etc. The coefficients of viscosity for some common fluids are listed in Table 9.2. We point out two facts about blood and water that you may find interesting. As Table 9.2 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.

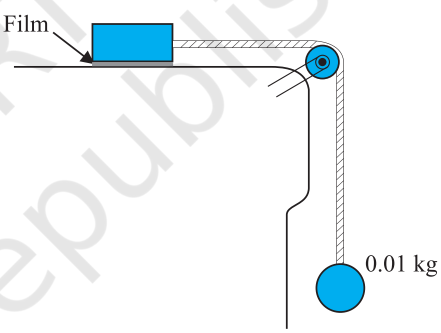


Fig. 9.13 *Measurement of the coefficient of viscosity of a liquid.*

The viscosity of liquids decreases with temperature, while it increases in the case of gases.

Example 9.8 A metal block of area 0.10 m^2 is connected to a 0.010 kg mass via a string that passes over an ideal pulley (considered massless and frictionless), as in Fig. 9.13. 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.

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

$$F = T = mg = 0.010 \text{ kg} \times 9.8 \text{ m s}^{-2} = 9.8 \times 10^{-2} \text{ N}$$

$$\text{Shear stress on the fluid} = F/A = \frac{9.8 \times 10^{-2}}{0.10} \text{N/m}^2$$

$$\text{Strain rate} = \frac{v}{l} = \frac{0.085}{0.30 \times 10^{-3}}$$

$$h = \frac{\text{stress}}{\text{strain rate}} \text{ s}^{-1}$$

$$= \frac{(9.8 \times 10^{-2} \text{ N})(0.30 \times 10^{-3} \text{ m})}{(0.085 \text{ m s}^{-1})(0.10 \text{ m}^2)}$$

$$= 3.46 \times 10^{-3} \text{ Pa s}$$

Table 9.2 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

9.5.1 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 are 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

$$F = 6\pi\eta a v \tag{9.17}$$

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 the 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

$$v_t = 2a^2 (\rho-\sigma) g / (9\eta) \tag{9.18}$$

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 9.9 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. (9.18)

$$\eta = \frac{2}{9} \times \frac{(2 \times 10^{-3})^2 \text{ m}^2 \times 9.8 \text{ m s}^{-2}}{6.5 \times 10^{-2} \text{ m s}^{-1}} \times 7.4 \times 10^3 \text{ kg m}^{-3}$$

$$= 9.9 \times 10^{-1} \text{ kg m}^{-1} \text{ s}^{-1}$$

9.6 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, hair 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.

9.6.1 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. 9.14(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. 9.14(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 9.14 (c) in a distance of the order of a few molecular sizes.

9.6.2 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 a

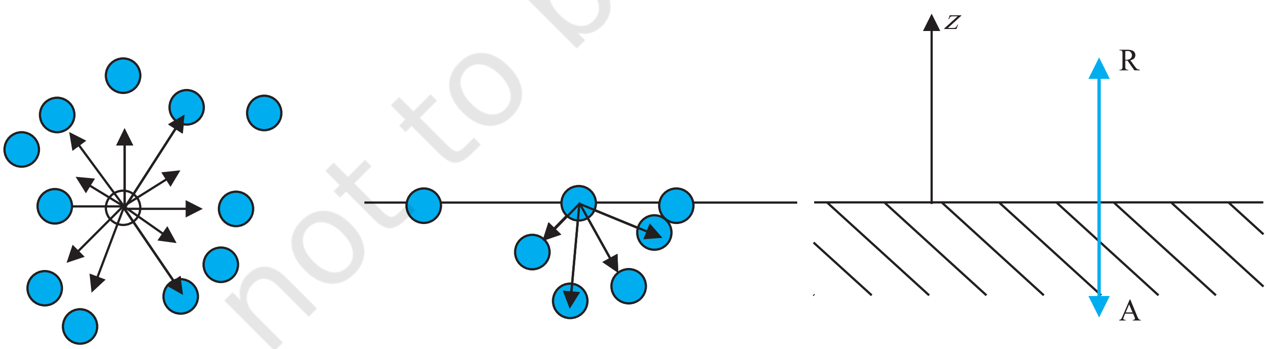


Fig. 9.14 Schematic picture of molecules in a liquid, at the surface and balance of forces. (a) Molecule inside a liquid. Forces on a molecule due to others are shown. Direction of arrows indicates attraction of repulsion. (b) Same, for a molecule at a surface. (c) Balance of attractive (A) and repulsive (R) forces.

horizontal liquid film ending in bar free to slide over parallel guides Fig (9.15).

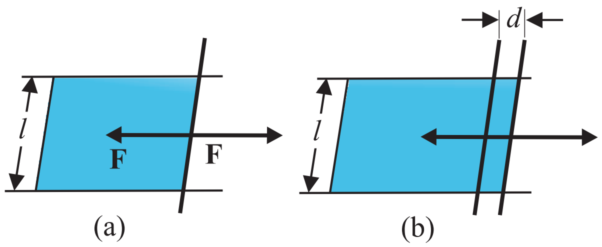


Fig. 9.15 *Stretching a film. (a) A film in equilibrium; (b) The film stretched an extra distance.*

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

$$S\,(2dl) = Fd \tag{9.19}$$

$$\text{Or, } S=Fd/2dl = F/2l \tag{9.20}$$

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 9.14 (a) and (b) there is only the force S per unit length acting inwards.

Table 9.3 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 9.3 **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 attraction between the solid surface and the liquid. It can be directly measured experimentaly as schematically shown in Fig. 9.16. 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.

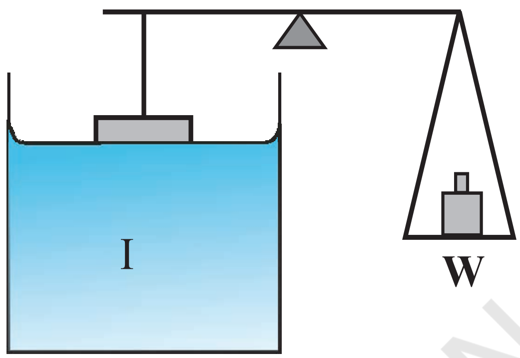


Fig. 9.16 *Measuring Surface Tension.*

Suppose the additional weight required is W. Then from Eq. 9.20 and the discussion given there, the surface tension of the liquid-air interface is

$$S_{la} = (W/2l) = (mg/2l) \tag{9.21}$$

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.

9.6.3 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. 9.17 (a) while spreads over a clean plastic plate as shown in Fig. 9.17(b).

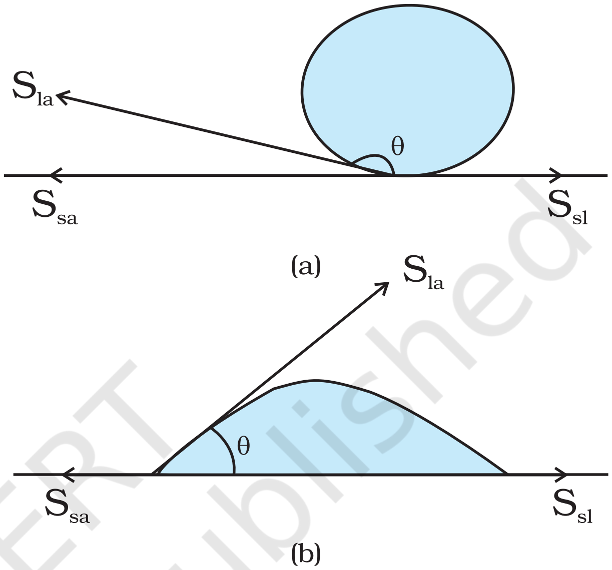


Fig. 9.17 *Different shapes of water drops with interfacial tensions (a) on a lotus leaf (b) on a clean plastic plate.*

We consider the three interfacial tensions at all the three interfaces, liquid-air, solid-air and solid-liquid denoted by S_{la}, S_{sa} and S_{sl}, respectively as given in Fig. 9.17 (a) and (b). At the line of contact, the surface forces between the three media must be in equilibrium. From the Fig. 9.17(b) the following relation is easily derived.

$$S_{la} \cos\theta + S_{sl} = S_{sa} \tag{9.22}$$

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.

9.6.4 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. 9.18(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

$$[4\pi(r+\Delta r)^2 - 4\pi r^2]\, S_{la} = 8\pi r\, \Delta r\, S_{la} \qquad (9.23)$$

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

$$W = (P_i - P_o)\, 4\pi r^2 \Delta r \qquad (9.24)$$

so that

$$(P_i - P_o) = (2\, S_{la} / r) \qquad (9.25)$$

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 9.18 (b).

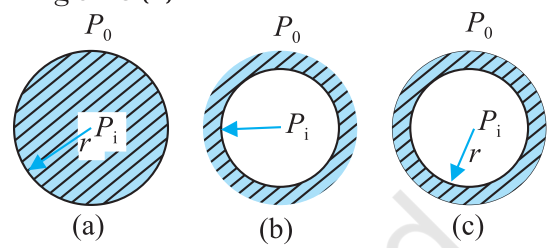


Fig. 9.18 *Drop, cavity and bubble of radius r.*

A bubble Fig 9.18 (c) differs from a drop and a cavity; in this it has two interfaces. Applying the above argument we have for a bubble

$$(P_i - P_o) = (4\, S_{la} / r) \qquad (9.26)$$

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!

9.6.5 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 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. 9.19). The contact angle between water and

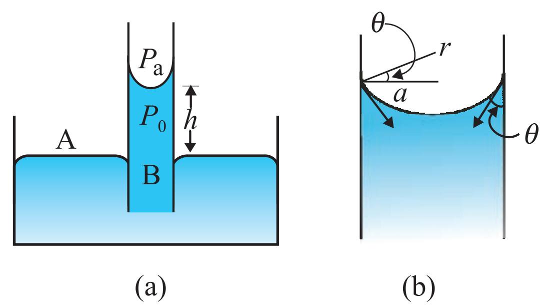


Fig. 9.19 *Capillary rise, (a) Schematic picture of a narrow tube immersed water. (b) Enlarged picture near interface.*

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**

$$(P_i - P_o) = (2S/r) = 2S/(a \sec\theta)$$

$$= (2S/a)\cos\theta \qquad (9.27)$$

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. 9.19(a). They must be at the same pressure, namely

$$P_0 + h\rho g = P_i = P_A \qquad (9.28)$$

where ρ is the density of water and h is called the capillary rise [Fig. 9.19(a)]. Using Eq. (9.27) and (9.28) we have

$$h\rho g = (P_i - P_0) = (2S\cos\theta)/a \qquad (9.29)$$

The discussion here, and the Eqs. (9.24) and (9.25) 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 9.3), we find that

$$h = 2S/(\rho g a)$$

$$= \frac{2 \times (0.073 \text{ N m}^{-1})}{(10^3 \text{ kg m}^{-3})(9.8 \text{ m s}^{-2})(5 \times 10^{-4} \text{ m})}$$

$$= 2.98 \times 10^{-2} \text{ m} = 2.98 \text{ cm}$$

Notice that if the liquid meniscus is convex, as for mercury, i.e., if $\cos\theta$ is negative then from Eq. (9.28) for example, it is clear that the liquid will be lower in the capillary !

Example 9.10 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
$P = P_a + \rho gh$
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 (ρgy) remains a constant.
$P + \rho v^2/2 + \rho gy$ = 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. 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 concept, 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. 9.4) is in equilibrium because the pressures exerted on the various faces are equal.

3. The expression for pressure
$P = P_a + \rho gh$
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 principle 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. 9.9] will be lower than the value given by Eq. (9.12).
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. 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	$[\mathrm{M\,L^{-1}T^{-2}}]$	pascal (Pa)	1 atm = 1.013×10^5 Pa, Scalar
Density	ρ	$[\mathrm{M\,L^{-3}}]$	kg $\mathrm{m^{-3}}$	Scalar
Specific Gravity		No	No	$\frac{\rho_{substance}}{\rho_{water}}$, Scalar
Co-efficient of viscosity	η	$[\mathrm{M\,L^{-1}\,T^{-1}}]$	Pa s or poiseiulles (Pl)	Scalar
Surface Tension	S	$[\mathrm{M\,T^{-2}}]$	N $\mathrm{m^{-1}}$	Scalar

EXERCISES

9.1 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.

9.2 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

9.3 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)

9.4 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

9.5 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 ?

9.6 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.

9.7 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.

9.8 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 ?

9.9 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 ?

9.10 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)

9.11 Can Bernoulli's equation be used to describe the flow of water through a rapid in a river ? Explain.

9.12 Does it matter if one uses gauge instead of absolute pressures in applying Bernoulli's equation ? Explain.

9.13 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].

9.14 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}.

9.15 Figures 9.20(a) and (b) refer to the steady flow of a (non-viscous) liquid. Which of the two figures is incorrect ? Why ?

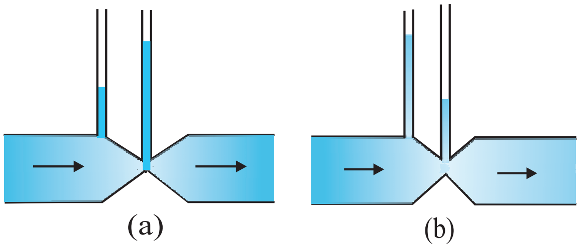


Fig. 9.20

9.16 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 ?

9.17 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 ?

9.18 Figure 9.21 (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.

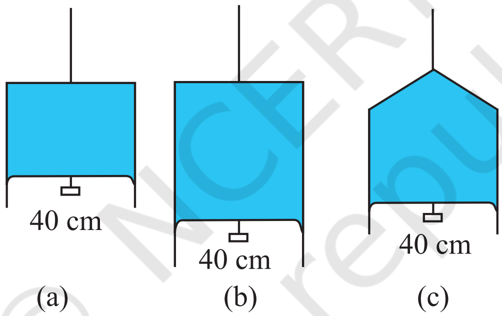


Fig. 9.21

9.19 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.

9.20 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).

11087CH11

CHAPTER TEN

THERMAL PROPERTIES OF MATTER

10.1 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 horse 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.

10.2 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 degree Celsius (°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.

10.3 MEASUREMENT OF TEMPERATURE

A measure of temperature is obtained using a thermometer. Many physical properties of materials change sufficiently with temperature. Some such properties are used as the basis for constructing thermometers. The commonly used property is variation of the volume of a liquid with temperature. For example, in common liquid–in–glass thermometers, mercury, alcohol etc., are used whose volume varies linearly with temperature over a wide range.

Thermometers are calibrated so that a numerical value may be assigned to a given temperature in an appropriate scale. 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 the 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, respectively. 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.

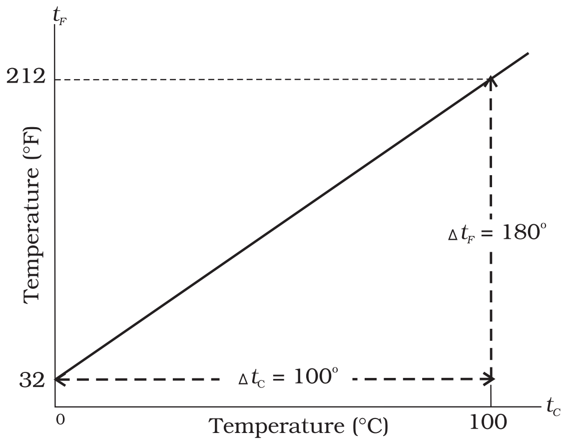


***Fig. 10.1** A plot of Fahrenheit temperature (t_F) versus Celsius temperature (t_C).*

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. 10.1), whose equation is

$$\frac{t_F - 32}{180} = \frac{t_C}{100} \qquad (10.1)$$

10.4 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 French scientist Jacques Charles (1747–1823). Low-density gases obey these laws, which may be combined into a single

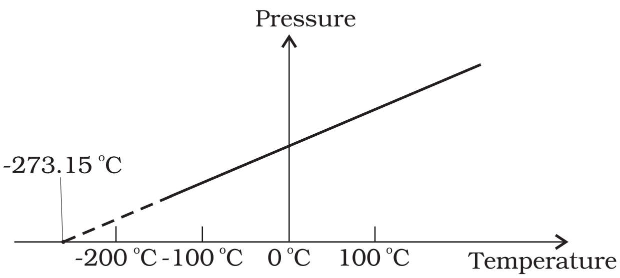


Fig. 10.2 *Pressure versus temperature of a low density gas kept at constant volume.*

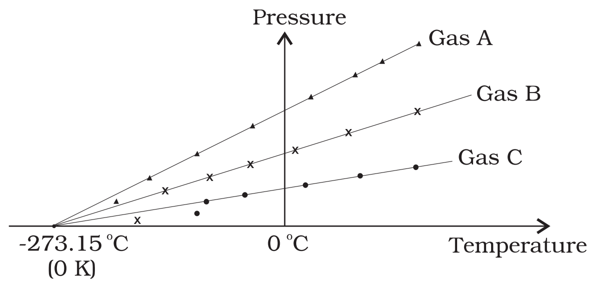


Fig. 10.3 *A plot of pressure versus temperature and extrapolation of lines for low density gases indicates the same absolute zero temperature.*

relationship. 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 low-density gas and is known as **ideal-gas equation**:

$$\frac{PV}{T} = \mu R$$

or $PV = \mu RT$ (10.2)

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. 10.2, 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. 10.2.

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. 10.3. 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. 10.4).

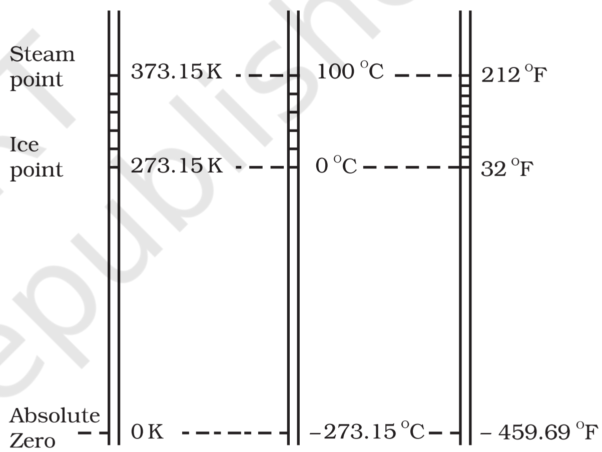


Fig. 10.4 *Comparision of the Kelvin, Celsius and Fahrenheit temperature scales.*

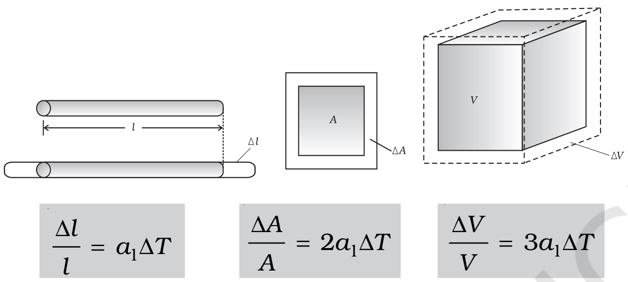
The size of unit in Kelvin and Celsius temperature scales is the same. So, temperature on these scales are related by

$$T = t_C + 273.15 \qquad (10.3)$$

10.5 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 some time to open it. This would allow the metallic lid 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 slightly warm water. If we take out the thermometer from the warm water the level of mercury falls again. Similarly, in 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. 10.5).



(a) Linear expansion (b) Area expansion (c) Volume expansion

Fig. 10.5 *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.

$$\frac{\Delta l}{l} = \alpha_l \, \Delta T \qquad (10.4)$$

where α_l is known as the **coefficient of linear expansion** (or linear expansivity) and is characteristic of the material of the rod. In Table 10.1, typical average values of the coefficient of linear expansion for some material in the temperature range 0 °C to 100 C are given. From this Table, compare the value of α_l 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_l.

Table 10.1 **Values of coefficient of linear expansion for some material**

Material	α_l (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 (or volume expansivity)**, α_V as

$$\alpha_V = \left(\frac{\Delta V}{V}\right)\frac{1}{\Delta T} \qquad (10.5)$$

Here α_V is also a characteristic of the substance but is not strictly a constant. It depends in general on temperature (Fig 10.6). It is seen that α_V becomes constant only at a high temperature.

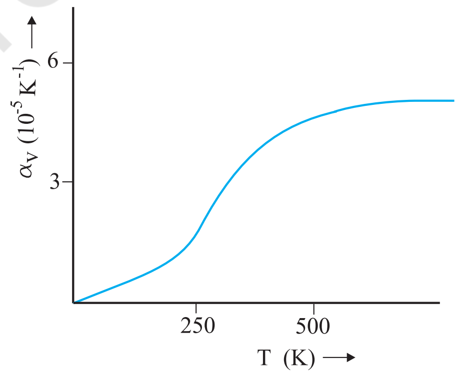


Fig. 10.6 *Coefficient of volume expansion of copper as a function of temperature.*

Table 10.2 gives the values of coefficient 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 material, 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 (ethanol) is more than mercury and expands more than mercury for the same rise in temperature.

Table 10.2 **Values of coefficient of volume expansion for some substances**

Material	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}
Mercury	18.2×10^{-5}
Water	20.7×10^{-5}
Alcohol (ethanol)	110×10^{-5}

Water exhibits an anomalous behaviour; 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. 10.7(a)]. Below 4 °C, the volume increases, and therefore, the density decreases [Fig. 10.7(b)].

This means that water has the 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

$P\Delta V = \mu R \Delta T$

$$\frac{\Delta V}{V} = \frac{\Delta T}{T}$$

$$\text{i.e., } \alpha_v = \frac{1}{T} \text{ for ideal gas} \qquad (10.6)$$

At 0 °C, α_v = 3.7×10^{-3} K^{-1}, which is much larger than that for solids and liquids. Equation (10.6) 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 3300×10^{-6} K^{-1}, as

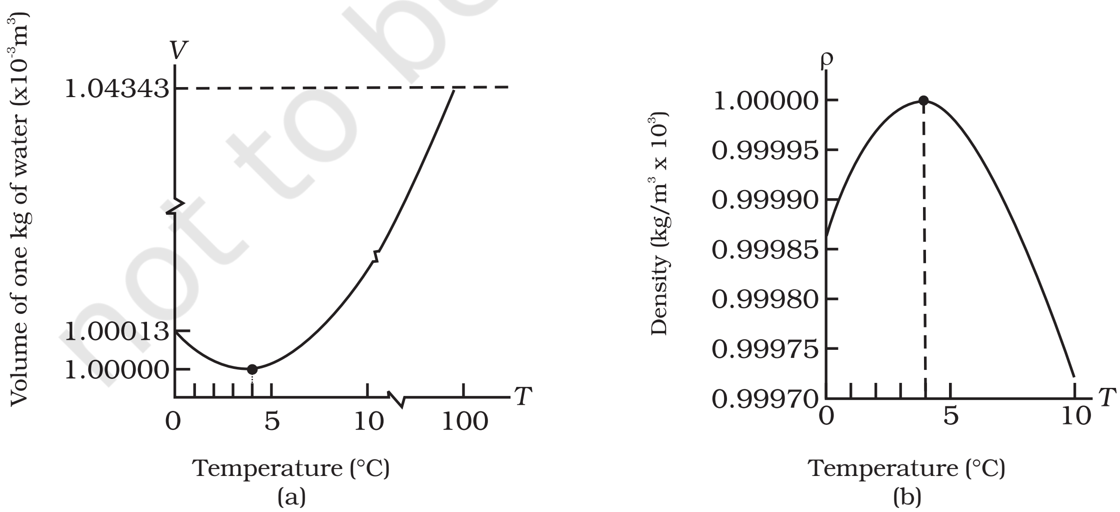


Fig. 10.7 *Thermal expansion of water.*

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, $\Delta V = (l+\Delta l)^3 - l^3 \simeq 3l^2 \, \Delta l$ (10.7)

In Equation (10.7), terms in $(\Delta l)^2$ and $(\Delta l)^3$ have been neglected since Δl is small compared to l. So

$$\Delta V = \frac{3V\,\Delta l}{l} = 3V\alpha_l \; \Delta T \qquad (10.8)$$

which gives

$$\alpha_v = 3\alpha_l \qquad (10.9)$$

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 cm² that is prevented from expanding while the temperature rises by 10 °C. The coefficient of linear expansion of steel is $\alpha_{l(steel)} = 1.2 \times 10^{-5}\ \mathrm{K}^{-1}$. Thus, the compressive strain is $\frac{\Delta l}{l} = \alpha_{l(steel)} \Delta T = 1.2 \times 10^{-5} \times 10 = 1.2 \times 10^{-4}$. Youngs modulus of steel is $Y_{(steel)} = 2 \times 10^{11}\ \mathrm{N\ m^{-2}}$. Therefore, the thermal stress developed is $\frac{\Delta F}{A} = Y_{steel}\left(\frac{\Delta l}{l}\right) = 2.4 \times 10^{7}\ \mathrm{N\ m^{-2}}$, which corresponds to an external force of

$$\Delta F = AY_{steel}\left(\frac{\Delta l}{l}\right) = 2.4 \times 10^{7} \times 40 \times 10^{-4} \simeq 10^{5}\mathrm{N}.$$ 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 10.1 Show that the coefficient of area expansion, $(\Delta A/A)/\Delta T$, of a rectangular sheet of the solid is twice its linear expansivity, α_l.

Answer

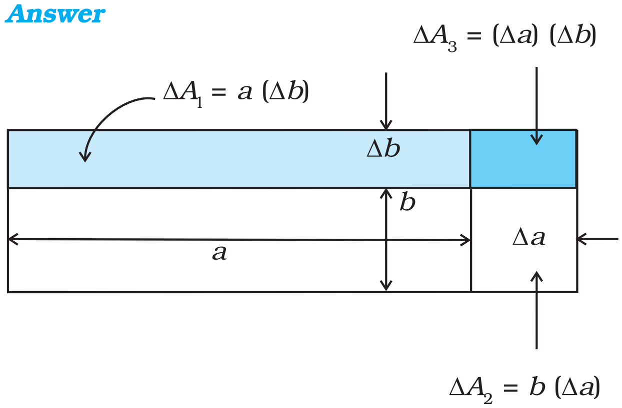


Fig. 10.8

Consider a rectangular sheet of the solid material of length a and breadth b (Fig. 10.8). 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. 10.8, 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 $\alpha_l \simeq 10^{-5}\ \mathrm{K}^{-1}$, from Table 10.1, 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 10.2 A blacksmith fixes iron ring on the rim of the wooden wheel of a horse 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?

Answer

Given, $T_1 = 27\ °\mathrm{C}$

$L_{T1} = 5.231\ \mathrm{m}$

$L_{T2} = 5.243\ \mathrm{m}$

So,

$$L_{T2} = L_{T1}\ [1+\alpha_l\,(T_2 - T_1)]$$

$5.243\ \mathrm{m} = 5.231\ \mathrm{m}\ [1 + 1.20\ 10^{-5}\ \mathrm{K}^{-1}\ (T_2 - 27\ °\mathrm{C})]$

or $T_2 = 218\ °\mathrm{C}$. ◀

10.6 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

$$S = \frac{\Delta Q}{\Delta T} \tag{10.10}$$

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 given off 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 given off by a substance of mass m when it undergoes a temperature change ΔT, then the specific heat capacity, of that substance is given by

$$s = \frac{S}{m} = \frac{1}{m}\frac{\Delta Q}{\Delta T} \tag{10.11}$$

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 given off) by it. It is defined as the amount of heat per unit mass absorbed or given off 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

$$C = \frac{S}{\mu} = \frac{1}{\mu}\frac{\Delta Q}{\Delta T} \tag{10.12}$$

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 11. Table 10.3 lists measured specific heat capacity of some substances at atmospheric pressure and ordinary temperature while Table 10.4 lists molar specific heat capacities of some gases. From Table 10.3 you can note that water

Table 10.3 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 also used as a coolant in automobile radiators, as well as, a heater in hot water bags. Owing to its high specific heat capacity, water warms up more slowly than 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 10.4 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

10.7 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 done is called a **calorimeter**. It consists of a metallic vessel and stirrer of the same material, like copper or aluminium. The vessel is kept inside a wooden jacket, which contains heat insulating material, 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 (Fig. 10.20). 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 10.3 A sphere of 0.047 kg aluminium is 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 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 temperature of aluminium sphere=100 °C
Final temperature = 23 °C
Change in temperature (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 = $m_1 s_{Al} \Delta T = 0.047 \text{kg} \times s_{Al} \times 77\ °C$

Mass of water (m_2) = 0.25 kg

Mass of calorimeter (m_3) = 0.14 kg

Initial temperature of water and calorimeter=20 °C

Final temperature of the mixture = 23 °C

Change in temperature (ΔT_2) = 23 °C – 20 °C = 3 °C

Specific heat capacity of water (s_w)

$= 4.18 \times 10^3$ J kg^{-1} K^{-1}

Specific heat capacity of copper calorimeter

$= 0.386 \times 10^3$ J kg^{-1} K^{-1}

The amount of heat gained by water and calorimeter = $m_2\, s_w\, \Delta T_2 + m_3 s_{cu} \Delta T_2$

$= (m_2 s_w + m_3 s_{cu})\, (\Delta T_2)$

$= (0.25 \text{ kg} \times 4.18 \times 10^3 \text{ J kg}^{-1} \text{ K}^{-1} + 0.14 \text{ kg} \times 0.386 \times 10^3 \text{ J kg}^{-1} \text{ K}^{-1})\ (23\ °C - 20\ °C)$

In the steady state heat lost by the aluminium sphere = heat gained by water + heat gained by calorimeter.

So, $0.047 \text{ kg} \times s_{Al} \times 77\ °C$

$= (0.25 \text{ kg} \times 4.18 \times 10^3 \text{ J kg}^{-1} \text{ K}^{-1} + 0.14 \text{ kg} \times 0.386 \times 10^3 \text{ J kg}^{-1} \text{ K}^{-1})(3\ °C)$

$s_{Al} = 0.911 \text{ kJ kg}^{-1} \text{K}^{-1}$ ◀

10.8 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. 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. 10.9). You will observe no change in the temperature as 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).

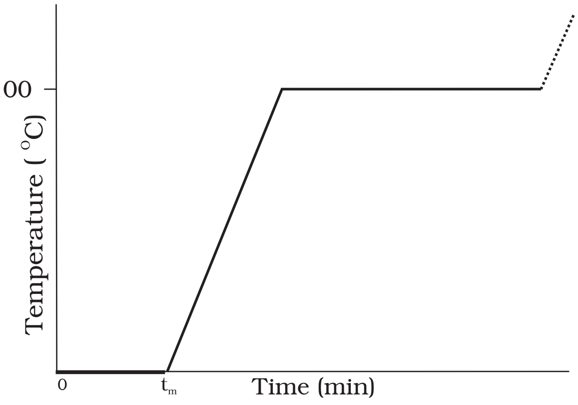


Fig. 10.9 *A plot of temperature versus time showing the changes in the state of ice on heating (not to scale).*

The change of state from solid to liquid is called **melting** or fusion and from liquid to solid is called **freezing**. It is observed that the temperature remains constant until the entire amount of the solid substance melts. That is, **both the solid and the 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 is 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. 10.10. 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 under the skates. Water is formed due to the increase of pressure and it acts as a lubricant.

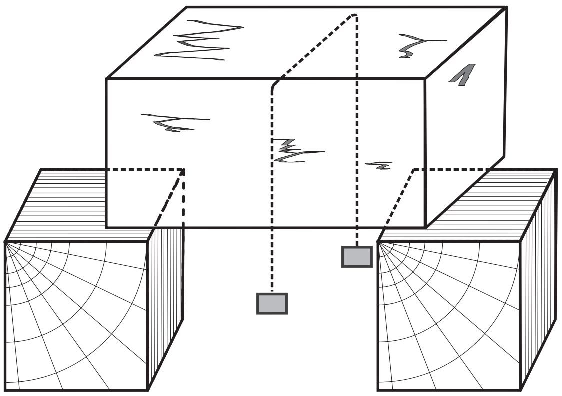


Fig. 10.10

After the whole of ice gets converted into water and as we continue further heating, we shall see that temperature begins to rise (Fig.10.9). 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.

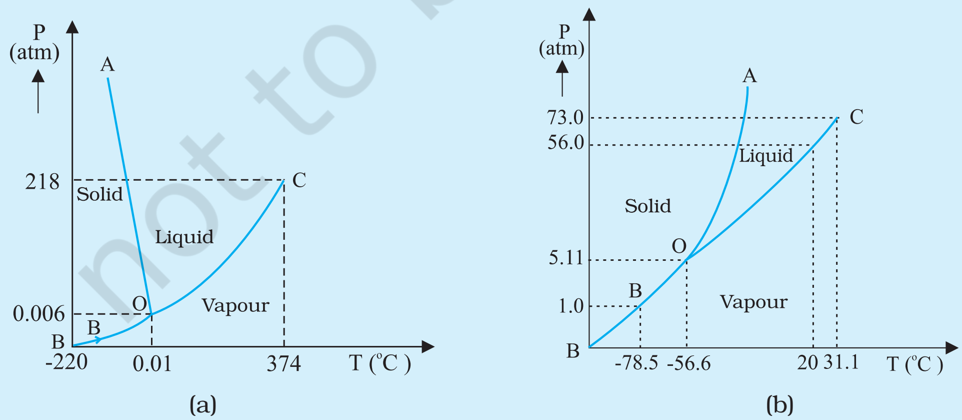


Figure : *Pressure-temperature phase diagrams for (a) water and (b) CO_2 (not to the scale).*

thermometer and steam outlet through the cork of the flask (Fig. 10.11). 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.

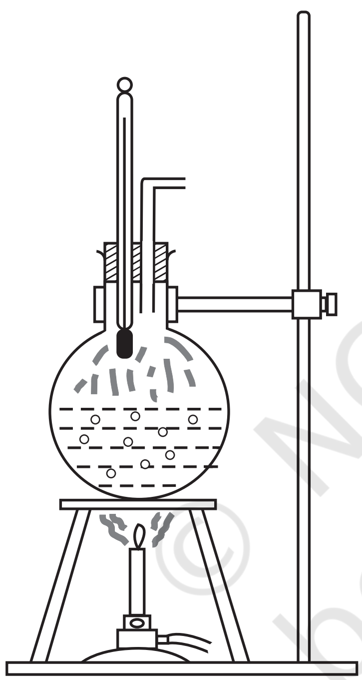


Fig. 10.11 *Boiling process.*

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.

10.8.1 Latent Heat

In Section 10.8, 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 10.5 **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})
Ethanol	–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 $$L = Q/m \qquad (10.13)$$

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 for a quantity of water is shown in Fig. 10.12. The latent heats of some substances, their freezing and boiling points, are given in Table 10.5.

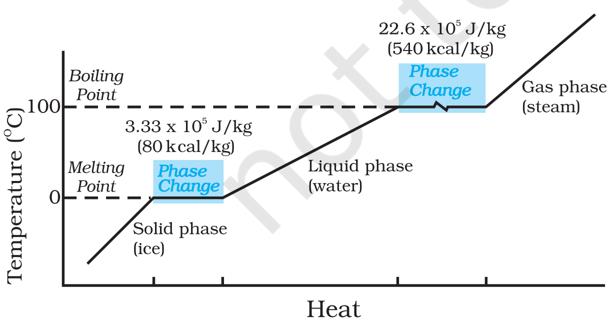


Fig. 10.12 *Temperature versus heat for water at 1 atm pressure (not to scale).*

Note that when heat is added (or removed) during a change of state, the temperature remains constant. Note in Fig. 10.12 that the slopes of the phase lines are not all the same, which indicate 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 is needed to melt 1 kg ice at 0 °C, and 22.6×10^5 J of heat is needed to convert 1 kg water into 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 10.4*** When 0.15 kg of ice at 0 °C is 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 = $m_I s_w\,(\theta_f - \theta_i)_I$

= (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

$L_f = 3.34\ 10^5$ J kg^{-1}.

> **Example 10.5** 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, $L_{f\,ice}$
$= 3.35 \times 10^5$ J kg^{-1}

latent heat of steam, L_{steam}
$= 2.256 \times 10^6$ 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.

$= m\, s_{ice}\, \Delta T_1 = (3 \text{ kg})\ (2100 \text{ J kg}^{-1} \text{ K}^{-1})\ [0-(-12)]°\text{C} = 75600 \text{ J}$

Q_2 = heat required to melt ice at 0 °C to water at 0 °C

$= m\, L_{f\,ice} = (3 \text{ kg})\ (3.35 \times 10^5 \text{ J kg}^{-1})$

$= 1005000$ J

Q_3 = heat required to convert water at 0 °C to water at 100 °C.

$= m s_w\, \Delta T_2 = (3\text{kg})\ (4186\text{J kg}^{-1} \text{ K}^{-1})\ (100\ °\text{C})$

$= 1255800$ J

Q_4 = heat required to convert water at 100 °C to steam at 100 °C.

$= m\, L_{steam} = (3 \text{ kg})\ (2.256 \times 10^6 \text{ J kg}^{-1})$

$= 6768000$ J

So, $Q = Q_1 + Q_2 + Q_3 + Q_4$

$= 75600\text{J} + 1005000 \text{ J} + 1255800 \text{ J} + 6768000 \text{ J}$

$= 9.1 \times 10^6$ J ◀

10.9 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. 10.13).

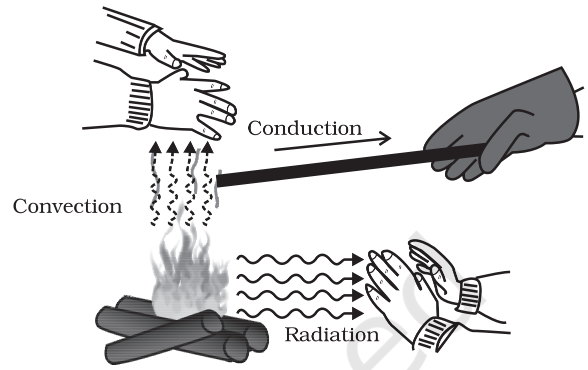


Fig. 10.13 *Heating by conduction, convection and radiation.*

10.9.1 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. 10.14). 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

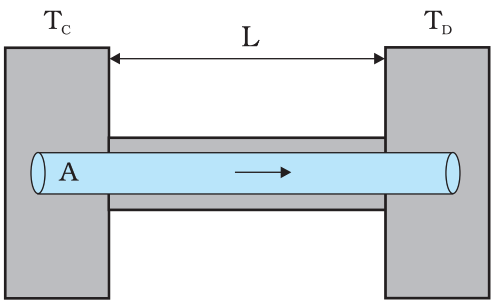


Fig. 10.14 *Steady state heat flow by conduction in a bar with its two ends maintained at temperatures T_C and T_D; ($T_C > T_D$).*

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 :

$$H = KA\frac{T_C - T_D}{L} \tag{10.14}$$

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 10.6. 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 good thermal conductors and, 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 10.5. 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 10.6 **Thermal conductivities of some material**

Material	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 10.6 What is the temperature of the steel-copper junction in the steady state of the system shown in Fig. 10.15. 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}).

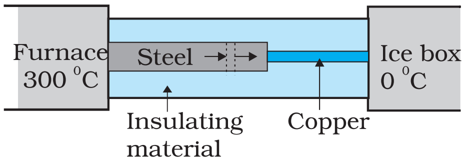


Fig. 10.15

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,

$$\frac{K_1 A_1 (300 - T)}{L_1} = \frac{K_2 A_2 (T - 0)}{L_2}$$

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 10.7*** 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. 10.16. 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.

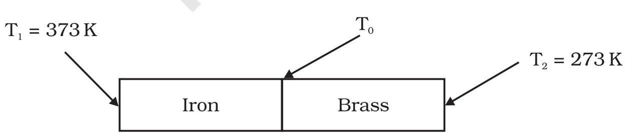


Fig 10.16

Answer

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, $H = H_1 = H_2$

$$= \frac{K_1 A_1 (T_1 - T_0)}{L_1} = \frac{K_2 A_2 (T_0 - T_2)}{L_2}$$

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

$$T_0 = \frac{(K_1 T_1 + K_2 T_2)}{(K_1 + K_2)}$$

Using this equation, the heat current H through either bar is

$$H = \frac{K_1 A (T_1 - T_0)}{L} = \frac{K_2 A (T_0 - T_2)}{L}$$

$$= \left(\frac{K_1 K_2}{K_1 + K_2}\right)\frac{A(T_1 - T_0)}{L} = \frac{A(T_1 - T_2)}{L\left(\frac{1}{K_1} + \frac{1}{K_2}\right)}$$

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

$$H' = \frac{K'\ A\ (T_1 - T_2)}{2\ L} = H$$

$$K' = \frac{2\ K_1\ K_2}{K_1 + K_2}$$

(i) $$T_0 = \frac{(K_1 T_1 + K_2 T_2)}{(K_1 + K_2)}$$

$$= \frac{(79\,\text{W m}^{-1}\text{K}^{-1})(373\,\text{K}) + (109\,\text{W m}^{-1}\text{K}^{-1})(273\,\text{K})}{79\,\text{W m}^{-1}\text{K}^{-1} + 109\,\text{W m}^{-1}\text{K}^{-1}}$$

$= 315$ K

(ii) $$K' = \frac{2K_1\ K_2}{K_1 + K_2}$$

$$= \frac{2 \times (79\,\text{W m}^{-1}\,\text{K}^{-1}) \times (109\,\text{W m}^{-1}\,\text{K}^{-1})}{79\,\text{W m}^{-1}\,\text{K}^{-1} + 109\,\text{W m}^{-1}\,\text{K}^{-1}}$$

$= 91.6$ W m^{-1} K^{-1}

(iii) $H' = H = \dfrac{K' \ A \ (T_1 - T_2)}{2 \ L}$

$$= \frac{(91.6 \text{ W m}^{-1} \text{K}^{-1}) \times (0.02 \text{ m}^2) \times (373 \text{K} - 273 \text{K})}{2 \times (0.1 \text{ m})}$$

$= 916.1$ W ◀

10.9.2 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 relatively 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 water has a greater specific heat capacity 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 the 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 revereseed (Fig. 10.17).

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**.

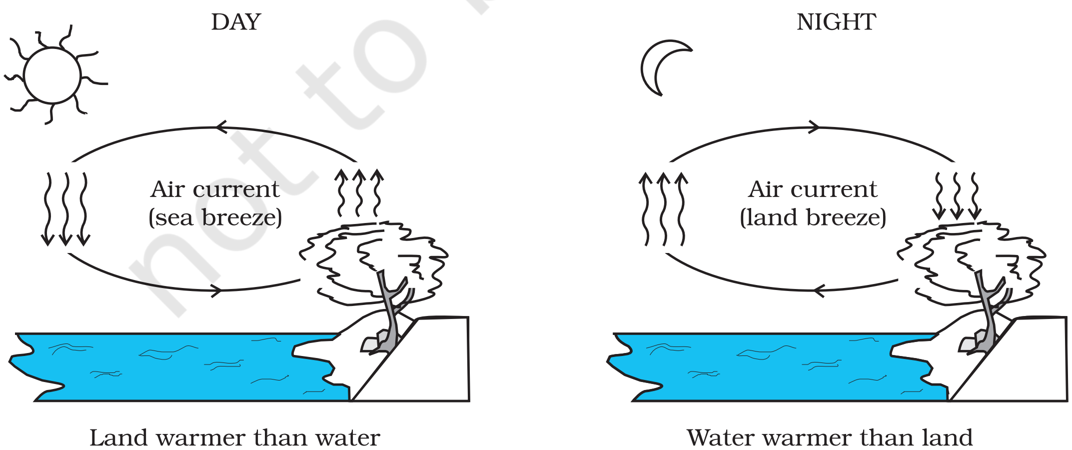


Fig. 10.17 *Convection cycles.*

10.9.3 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. Similarly, we quickly feel the warmth of the fire nearby even though air conducts poorly and before convection takes some time to set in. The third mechanism for heat transfer needs no medium; it is called radiation and the energy so transferred 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 detail 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 transferred to the earth from the Sun through empty space. All bodies emit radiant energy, whether they are solid, liquid or gas. The electromagnetic radiation emitted by a body by virtue of its temperature, like 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 utensils for cooking food are blackened so that they absorb maximum heat from fire and transfer 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 to the contents of the bottle. The outer wall similarly reflects back any incoming radiation. The space between the walls is evacuated 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).

10.9.4 Blackbody Radiation

We have so far not mentioned the wavelength content of thermal radiation. The important thing about thermal radiation at any temperature is that it is not of one (or a few) wavelength(s) but has a continuous spectrum from the small to the long wavelengths. The energy content of radiation, however, varies for different wavelengths. Figure 10.18 gives the experimental curves for radiation energy per unit area per unit wavelength emitted by a blackbody versus wavelength for different temperatures.

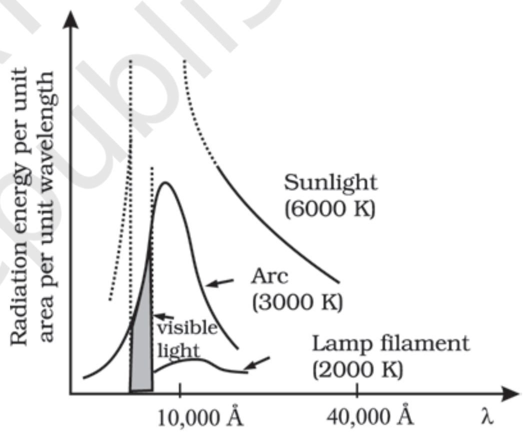


Fig. 10.18: *Energy emitted versus wavelength for a blackbody at different temperatures*

Notice that the wavelength λ_m for which energy is the maximum decreases with increasing temperature. The relation between λ_m and T is given by what is known as **Wien's Displacement Law**:

$$\lambda_m T = \text{constant} \tag{10.15}$$

The value of the constant (Wien's constant) is 2.9×10^{-3} m K. This law explains why the colour of a piece of iron heated in a hot flame first becomes dull red, then reddish yellow, and finally white hot. Wien's law is useful for estimating the surface temperatures of celestial bodies like, the moon, Sun and other stars. Light from the moon is found to have a maximum intensity near the wavelength 14 μm. By Wien's law, the surface of the moon is estimated to have a temperature of 200 K. Solar radiation has a maximum at λ_m = 4753 Å. This corresponds to T = 6060 K. Remember, this is the temperature of the surface of the sun, not its interior.

The most significant feature of the blackbody radiation curves in Fig. 10.18 is that they are *universal*. They depend only on the temperature and not on the size, shape or material of the blackbody. Attempts to explain blackbody radiation theoretically, at the beginning of the twentieth century, spurred the quantum revolution in physics, as you will learn in later courses.

Energy can be transferred by radiation over large distances, without a medium (i.e., in vacuum). The total electromagnetic energy radiated by a body at absolute temperature T is proportional to its size, its ability to radiate (called emissivity) and most importantly to its temperature. For a body, which is a perfect radiator, the energy emitted per unit time (H) is given by

$$H = A\sigma T^4 \qquad (10.16)$$

where A is the area and T is the absolute temperature of the body. This relation obtained experimentally by Stefan and later proved theoretically by Boltzmann is known as **Stefan-Boltzmann law** and the constant σ is called Stefan-Boltzmann constant. Its value in SI units is 5.67×10^{-8} W m^{-2} K^{-4}. Most bodies emit only a fraction of the rate given by Eq. 10.16. A substance like lamp black comes close to the limit. One, therefore, defines a dimensionless fraction e called *emissivity* and writes,

$$H = Ae\sigma T^4 \qquad (10.17)$$

Here, e = 1 for a perfect radiator. For a tungsten lamp, for example, e is about 0.4. Thus, a tungsten lamp at a temperature of 3000 K and a surface area of 0.3 cm^2 radiates at the rate $H = 0.3 \times 10^{-4} \times 0.4 \times 5.67 \times 10^{-8} \times (3000)^4 = 60$ W.

A body at temperature T, with surroundings at temperatures T_s, emits, as well as, receives energy. For a perfect radiator, the net rate of loss of radiant energy is

$$H = \sigma A\,(T^4 - T_s^4)$$

For a body with emissivity e, the relation modifies to

$$H = e\sigma A\,(T^4 - T_s^4) \qquad (10.18)$$

As an example, let us estimate the heat radiated by our bodies. Suppose the surface area of a person's body is about 1.9 m^2 and the room temperature is 22 °C. The internal body temperature, as we know, is about 37 °C. The skin temperature may be 28 °C (say). The emissivity of the skin is about 0.97 for the relevant region of electromagnetic radiation. The rate of heat loss is:

$$H = 5.67 \times 10^{-8} \times 1.9 \times 0.97 \times \{(301)^4 - (295)^4\}$$

$$= 66.4 \text{ W}$$

which is more than half the rate of energy production by the body at rest (120 W). To prevent this heat loss effectively (better than ordinary clothing), modern arctic clothing has an additional thin shiny metallic layer next to the skin, which reflects the body's radiation.

10.10 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 slow or fast 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 a two-holed lid. Fix the stirrer through one hole and fix a thermometer through another 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 a 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. 10.19).

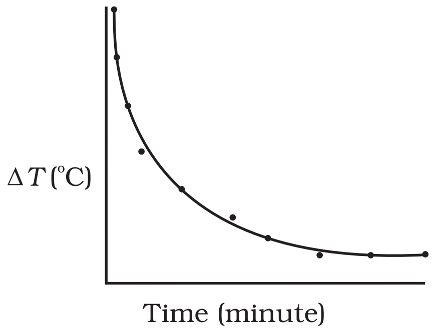


Fig. 10.19 *Curve showing cooling of hot water with time.*

From the graph you can 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

$$-\frac{dQ}{dt} = k(T_2 - T_1) \tag{10.19}$$

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

$$\frac{dQ}{dt} = ms\frac{dT_2}{dt} \tag{10.20}$$

From Eqs. (10.15) and (10.16) we have

$$-m\,s\frac{dT_2}{dt} = k(T_2 - T_1)$$

$$\frac{dT_2}{T_2 - T_1} = -\frac{k}{ms}dt = -K\,dt \tag{10.21}$$

where $K = k/m\,s$

On integrating,

$$\log_e (T_2 - T_1) = -K\,t + c \tag{10.22}$$

$$\text{or } T_2 = T_1 + C'\,\mathrm{e}^{-Kt};\ \text{where } C' = \mathrm{e}^c \tag{10.23}$$

Equation (10.23) 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.

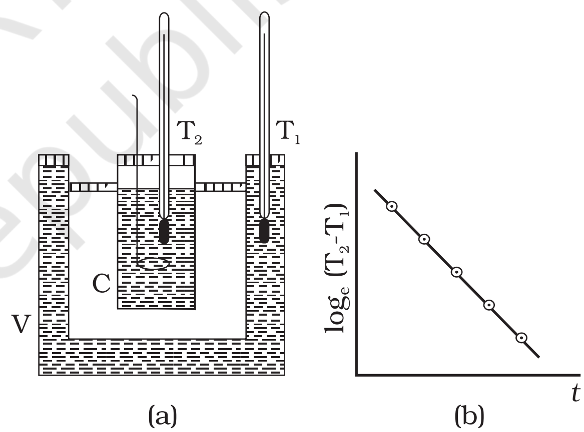


Fig. 10.20 *Verification of Newton's Law of cooling.*

Newton's law of cooling can be verified with the help of the experimental set-up shown in Fig. 10.20(a). The set-up consists of a double-walled vessel (V) containing water 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 $\log_e (T_2 - T_1)$ [or $\ln(T_2 - T_1)$] and time (t). The nature of the graph is observed to be a straight line having a negative slope as shown in Fig. 10.20(b). This is in support of Eq. 10.22.

Example 10.8 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. (10.21), we have

$$\frac{\text{Change in temperature}}{\text{Time}} = K\Delta T$$

$$\frac{8\,°\text{C}}{2\ \text{min}} = K(70\,°\text{C})$$

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.

$$\frac{2\,°\text{C}}{\text{Time}} = K(50\,°\text{C})$$

When we divide above two equations, we have

$$\frac{8\,°\text{C}/2\ \text{min}}{2\,°\text{C}/\text{time}} = \frac{K\,(70\,°\text{C})}{K\,(50\,°\text{C})}$$

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 corresponds to 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 :

$$\frac{\Delta l}{l} = \alpha_l \Delta T$$

$$\frac{\Delta V}{V} = \alpha_V \Delta T$$

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 capacity	s	$[L^2T^{-2}K^{-1}]$	$J\,kg^{-1}\,K^{-1}$	
Thermal Conductivity	K	$[MLT^{-3}K^{-1}]$	$J\,s^{-1}K^{-1}$	$H = -KA\frac{dT}{dx}$

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

10.1 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.

10.2 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 ?

10.3 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 ?

10.4 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 ?

10.5 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 ?

10.6 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} .

10.7 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}.

10.8 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}.

10.9 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.

10.10 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}).

10.11 The coefficient of volume expansion of glycerine is 49×10^{-5} K^{-1}. What is the fractional change in its density for a 30 °C rise in temperature ?

10.12 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}.

10.13 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}).

10.14 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 ?

10.15 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 ?

10.16 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 minutes, 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}.

10.17 A 'thermacole' icebox is a cheap and an 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}]

10.18 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}.

10.19 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

10.20 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 Eleven

11087CH12

Thermodynamics

11.1 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.

11.2 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.

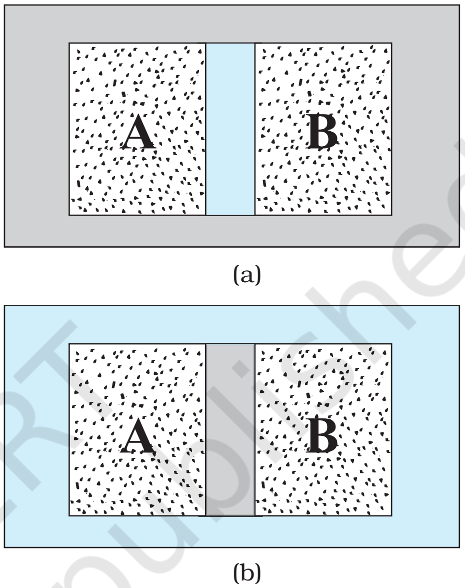


Fig. 11.1 *(a) Systems A and B (two gases) separated by an adiabatic wall – an insulating wall that does not allow flow of heat. (b) The same systems A and B separated by a diathermic wall – a conducting wall that allows heat to flow from one to another. In this case, thermal equilibrium is attained in due course.*

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

* *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. However, a thermodynamic state is specified by five state variables viz., pressure, volume, temperature, internal energy and entropy. Entropy is a measure of disorderness in the system. Enthalpy is a measure of total heat content of the system.*

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. 11.1 (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. 11.1(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 are equal. We shall see how does one arrive at the concept of temperature in thermodynamics? The Zeroth law of thermodynamics provides the clue.

11.3 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. 11.2(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.11.2(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.

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.

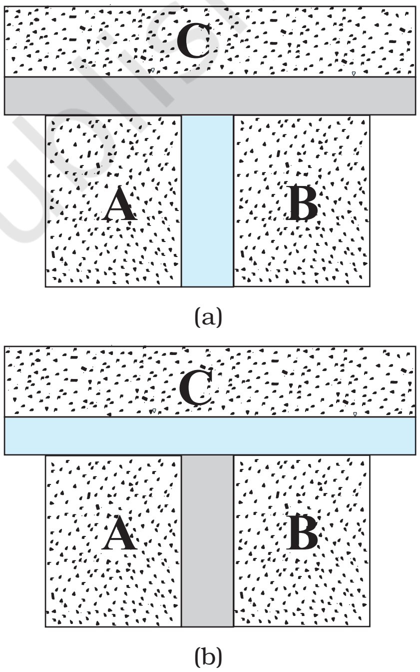


Fig. 11.2 *(a) Systems A and B are separated by an adiabatic wall, while each is in contact with a third system C via a conducting wall. (b) The adiabatic wall between A and B is replaced by a conducting wall, while C is insulated from A and B by an adiabatic wall.*

* *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.*

11.4 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 11.7). 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. 11.3).

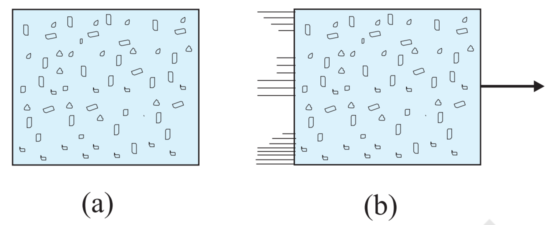


Fig. 11.3 *(a) Internal energy U of a gas is the sum of the kinetic and potential energies of its molecules when the box is at rest. Kinetic energy due to various types of motion (translational, rotational, vibrational) is to be included in U. (b) If the same box is moving as a whole with some velocity, the kinetic energy of the box is not to be included in U.*

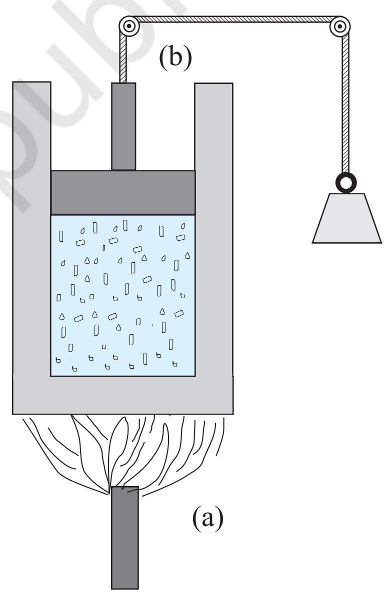


Fig. 11.4 *Heat and work are two distinct modes of energy transfer to a system that results in change in its internal energy. (a) Heat is energy transfer due to temperature difference between the system and the surroundings. (b) Work is energy transfer brought about by means (e.g. moving the piston by raising or lowering some weight connected to it) that do not involve such a temperature difference.*

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. 11.4. 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.

11.5 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

$$\Delta Q = \Delta U + \Delta W \tag{11.1}$$

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 (11.1) 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. (11.1) in the alternative form

$$\Delta Q - \Delta W = \Delta U \tag{11.2}$$

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. (11.2), 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 11.8),

$$\Delta Q = \Delta W$$

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. (11.1) gives

$$\Delta Q = \Delta U + P\Delta V \qquad (11.3)$$

As an application of Eq. (11.3), 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,

$$\Delta W = P(V_g - V_l) = 1.013 \times 10^5 \times (1671 \times 10^{-6}) = 169.2\ \text{J}$$

Equation (11.3) 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.

11.6 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 10) to be

$$S = \frac{\Delta Q}{\Delta T} \qquad (11.4)$$

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 :

$$s = \frac{S}{m} = \left(\frac{1}{m}\right)\frac{\Delta Q}{\Delta T} \qquad (11.5)$$

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

$$C = \frac{S}{\mu} = \frac{1}{\mu}\frac{\Delta Q}{\Delta T} \qquad (11.6)$$

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 11.1 lists measured specific and molar heat capacities of solids at atmospheric pressure and ordinary room temperature.

We will see in Chapter 12 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 (See Section 12.5 and 12.6). Consider a solid of N atoms, each vibrating about its mean position. An oscillator in one dimension has average energy of 2 × ½ $k_BT = k_BT$. In three dimensions, the average energy is 3 k_BT. For a mole of a solid, the total energy is

$$U = 3\,k_BT \times N_A = 3\,RT \quad (\because k_BT \times N_A = R)$$

Now, at constant pressure, $\Delta Q = \Delta U + P\,\Delta V \cong \Delta U$, since for a solid ΔV is negligible. Therefore,

$$C = \frac{\Delta Q}{\Delta T} = \frac{\Delta U}{\Delta T} = 3R \qquad (11.7)$$

Table 11.1 **Specific and molar heat capacities 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 11.1 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 11.5 shows this variation in the temperature range 0 to 100 °C.

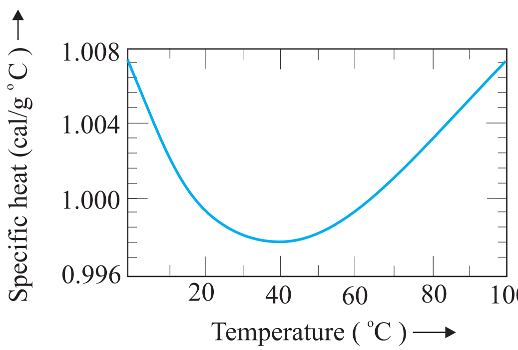


Fig. 11.5 *Variation of specific heat capacity of water with temperature.*

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.

$$C_p - C_v = R \qquad (11.8)$$

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. (11.3) for 1 mole of the gas :

$$\Delta Q = \Delta U + P\Delta V$$

If ΔQ is absorbed at constant volume, $\Delta V = 0$

$$C_v = \left(\frac{\Delta Q}{\Delta T}\right)_v = \left(\frac{\Delta U}{\Delta T}\right)_v = \left(\frac{\Delta U}{\Delta T}\right) \qquad (11.9)$$

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,

$$C_p = \left(\frac{\Delta Q}{\Delta T}\right)_p = \left(\frac{\Delta U}{\Delta T}\right)_p + P\left(\frac{\Delta V}{\Delta T}\right)_p \qquad (11.10)$$

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

$$P\left(\frac{\Delta V}{\Delta T}\right)_p = R \qquad (11.11)$$

Equations (11.9) to (11.11) give the desired relation, Eq. (11.8).

11.7 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. 11.6(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. 11.6(b)]. Eventually, the gas attains a uniform temperature and pressure and comes to thermal and mechanical equilibrium with its surroundings.

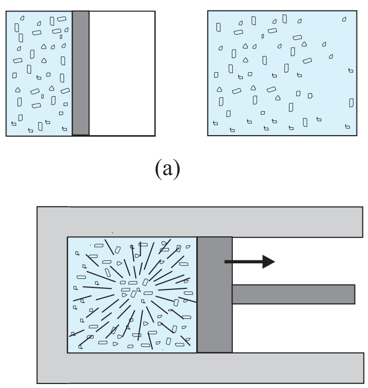


Fig. 11.6 *(a) The partition in the box is suddenly removed leading to free expansion of the gas. (b) A mixture of gases undergoing an explosive chemical reaction. In both situations, the gas is not in equilibrium and cannot be described by state variables.*

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.)

11.8 THERMODYNAMIC PROCESSES

11.8.1 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 11.7 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 (Fig.11.7). 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'.

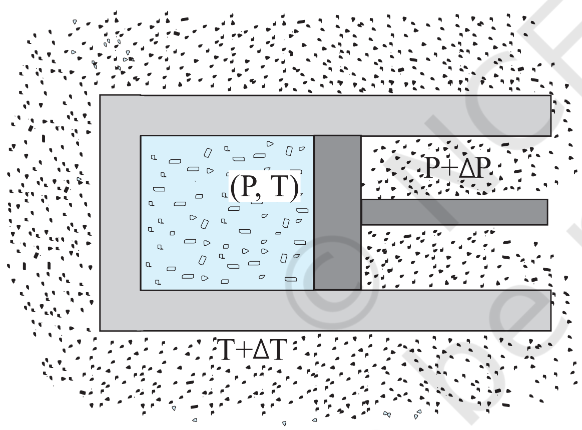


Fig. 11.7 *In a quasi-static process, the temperature of the surrounding reservoir and the external pressure differ only infinitesimally from the temperature and pressure of the system.*

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. 11.2

Table 11.2 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 :

11.8.2 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,

$$W = \int_{V_1}^{V_2} P \, \mathrm{d}V$$

$$= \mu RT \int_{V_1}^{V_2} \frac{\mathrm{d}V}{V} = \mu RT \ \ln \frac{V_2}{V_1} \qquad (11.12)$$

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. (11.12) 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.

11.8.3 Adiabatic process

In an adiabatic process, the system is insulated from the surroundings and heat absorbed or released is zero. From Eq. (11.1), 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.

$$P V^{\gamma} = \text{const} \tag{11.13}$$

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) :

$$P_1 V_1^{\gamma} = P_2 V_2^{\gamma} \tag{11.14}$$

Figure 11.8 shows the *P-V* curves of an ideal gas for two adiabatic processes connecting two isotherms.

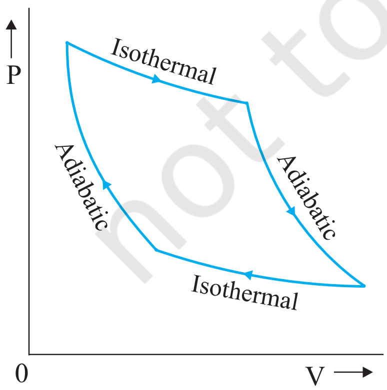


Fig. 11.8 *P-V curves for isothermal and adiabatic processes of an ideal gas.*

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).

$$W = \int_{V_1}^{V_2} P \, dV$$

$$= constant \times \int_{V_1}^{V_2} \frac{dV}{V^{\gamma}} = constant \times \left. \frac{V^{-\gamma+1}}{1-\gamma} \right|_{V_1}^{V_2}$$

$$\frac{constant}{(1-\gamma)} \times \left[\frac{1}{V_2^{\gamma-1}} - \frac{1}{V_1^{\gamma-1}} \right] \tag{11.15}$$

From Eq. (11.14), the constant is $P_1 V_1^{\gamma}$ or $P_2 V_2^{\gamma}$

$$W = \frac{1}{1-\gamma} \left[\frac{P_2 V_2^{\gamma}}{V_2^{\gamma-1}} - \frac{P_1 V_1^{\gamma}}{V_1^{\gamma-1}} \right]$$

$$= \frac{1}{1-\gamma} \left[P_2 V_2 - P_1 V_1 \right] = \frac{\mu R (T_1 - T_2)}{\gamma - 1} \tag{11.16}$$

As expected, if work is done by the gas in an adiabatic process ($W > 0$), from Eq. (11.16), $T_2 < T_1$. 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.

11.8.4 Isochoric process

In an isochoric process, V is constant. No work is done on or by the gas. From Eq. (11.1), 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.

11.8.5 Isobaric process

In an isobaric process, P is fixed. Work done by the gas is

$$W = P(V_2 - V_1) = \mu R (T_2 - T_1) \tag{11.17}$$

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.

11.8.6 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. (11.1), the total heat absorbed equals the work done by the system.

11.9 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. For a refrigerator, the Second Law says that the co-efficient of performance can never be infinite. 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.

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.

11.10 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.

11.11 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.

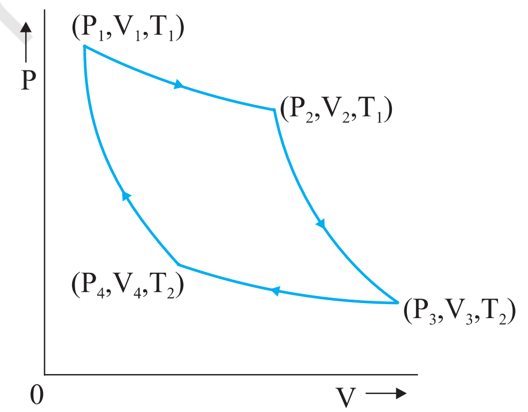


Fig. 11.9 *Carnot cycle for a heat engine with an ideal gas as the working substance.*

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. 11.9. 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. (11.12). This is also the work done $(W_{1 \rightarrow 2})$ by the gas on the environment.

$$W_{1 \rightarrow 2} = Q_1 = \mu R T_1 \ln\left(\frac{V_2}{V_1}\right) \quad (11.18)$$

(b) *Step* $2 \rightarrow 3$ 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. (11.16), is

$$W_{2\rightarrow 3} = \frac{\mu R\left(T_1 - T_2\right)}{\gamma - 1} \quad (11.19)$$

(c) *Step* $3 \rightarrow 4$ 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. (11.12). This is also the work done $(W_{3 \rightarrow 4})$ on the gas by the environment.

$$W_{3\rightarrow 4} = Q_2 = \mu R T_2 \ln\left(\frac{V_3}{V_4}\right) \quad (11.20)$$

(d) *Step* $4 \rightarrow 1$ 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.(11.16), is

$$W_{4\rightarrow 1} = \mu R\left(\frac{T_1 - T_2}{\gamma \text{ - } 1}\right) \quad (11.21)$$

From Eqs. (11.18) to (11.21) total work done by the gas in one complete cycle is

$$W = W_{1\rightarrow 2} + W_{2\rightarrow 3} - W_{3\rightarrow 4} - W_{4\rightarrow 1}$$

$$= \mu R T_1 \ \ln\left(\frac{V_2}{V_1}\right) - \mu R T_2 \ \ln\left(\frac{V_3}{V_4}\right) \quad (11.22)$$

The efficiency η of the Carnot engine is

$$\eta = \frac{W}{Q_1} = 1 - \frac{Q_2}{Q_1}$$

$$= 1 - \left(\frac{T_2}{T_1}\right)\frac{\mathrm{In}\left(\frac{V_3}{V_4}\right)}{\mathrm{In}\left(\frac{V_2}{V_1}\right)} \quad (11.23)$$

Now since step $2 \rightarrow 3$ is an adiabatic process,

$$T_1 V_2^{\gamma-1} = T_2 V_3^{\gamma-1}$$

i.e. $$\frac{V_2}{V_3} = \left(\frac{T_2}{T_1}\right)^{1/(\gamma-1)} \quad (11.24)$$

Similarly, since step $4 \rightarrow 1$ is an adiabatic process

$$T_2 V_4^{\gamma-1} = T_1 V_1^{\gamma-1}$$

i.e. $$\frac{V_1}{V_4} = \left(\frac{T_2}{T_1}\right)^{1/\gamma-1} \quad (11.25)$$

From Eqs. (11.24) and (11.25),

$$\frac{V_3}{V_4} = \frac{V_2}{V_1} \quad (11.26)$$

Using Eq. (11.26) in Eq. (11.23), we get

$$\eta = 1 - \frac{T_2}{T_1} \quad \text{(Carnot engine)} \quad (11.27)$$

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. 11.9 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

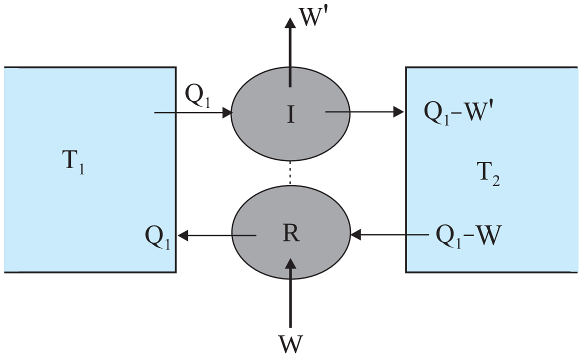


Fig. 11.10 *An irreversible engine (I) coupled to a reversible refrigerator (R). If W ' > W, this would amount to extraction of heat W' – W from the sink and its full conversion to work, in contradiction with the Second Law of Thermodynamics.*

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. (11.27) 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. (11.27)], is true for any Carnot engine.

This final remark shows that in a Carnot cycle,

$$\frac{Q_1}{Q_2} = \frac{T_1}{T_2} \qquad (11.28)$$

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 (11.28), 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 11.9.

SUMMARY

1. The zeroth law of thermodynamics states that '*two systems in thermal equilibrium with a third system separately 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

$$C = \frac{1}{\mu}\frac{\Delta Q}{\Delta T}$$

where μ is the number of moles of the substance. For a solid, the law of equipartition of energy gives

$$C = 3\,R$$

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

$$Q = W = \mu R T \ln\left(\frac{V_2}{V_1}\right)$$

9. In an adiabatic process of an ideal gas

PV^{γ} = constant

where $$\gamma = \frac{C_p}{C_v}$$

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

$$W = \frac{\mu R (T_1 - T_2)}{\gamma - 1}$$

10. 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.

11. 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.

12. 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.

13. 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}	$\alpha_v = 3\,\alpha_1$
Heat supplied to a system	ΔQ	$[ML^2 T^{-2}]$	J	Q is not a state variable
Specific heat capacity	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

11.1 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 ?

11.2 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}.)

11.3 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.

11.4 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 ?

11.5 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)

11.6 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 ?

11.7 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?

11.8 A thermodynamic system is taken from an original state to an intermediate state by the linear process shown in Fig. (11.13)

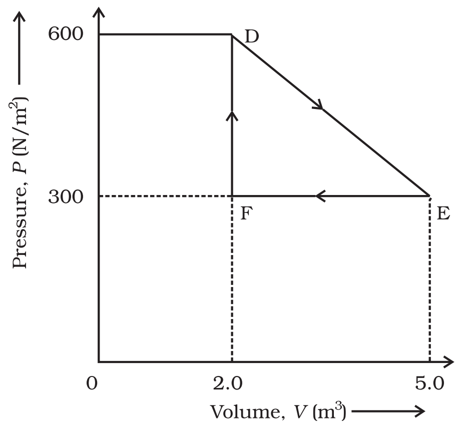


Fig. 11.11

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

CHAPTER TWELVE

11087CH13

KINETIC THEORY

12.1 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.

12.2 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 19^{th} 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.

12.3 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 Chapter 10)

$$PV = KT \tag{12.1}$$

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.

$$\text{As } \frac{P_1V_1}{N_1T_1} = \frac{P_2V_2}{N_2T_2} = \text{constant} = k_B \tag{12.2}$$

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 the 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 1 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

$$PV = \mu\, RT \tag{12.3}$$

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 absolute temperature, R = 8.314 J mol^{-1}K^{-1}. Here

$$\mu = \frac{M}{M_0} = \frac{N}{N_A} \tag{12.4}$$

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. (12.4) and (12.3) can also be written as

$$PV = k_B\, NT \quad \text{or} \quad P = k_B\, nT$$

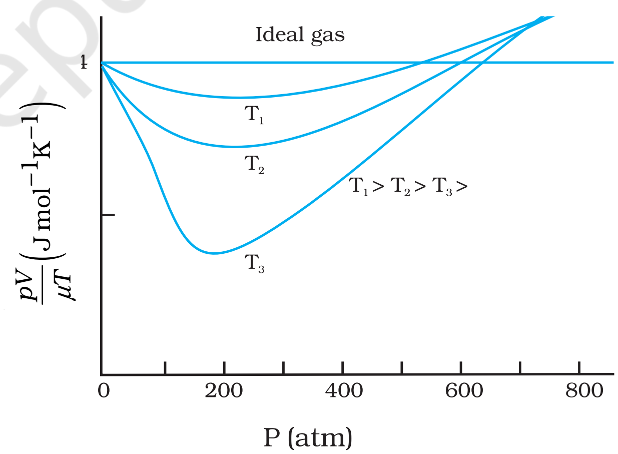


Fig. 12.1 *Real gases approach ideal gas behaviour at low pressures and high temperatures.*

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. (12.3) is

$$P = \frac{\rho RT}{M_0} \tag{12.5}$$

where ρ is the mass density of the gas.

A gas that satisfies Eq. (12.3) 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. 12.1 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. (12.3), we get

$$PV = \text{constant} \tag{12.6}$$

i.e., keeping temperature constant, pressure of a given mass of gas varies inversely with volume. This is the famous **Boyle's law**. Fig. 12.2 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. (12.1) 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. 12.3.

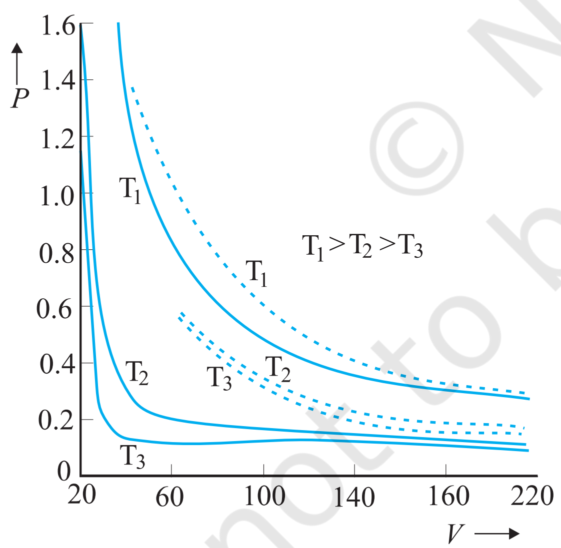


Fig.12.2 *Experimental P-V curves (solid lines) for steam at three temperatures compared with Boyle's law (dotted lines). P is in units of 22 atm and V in units of 0.09 litres.*

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 :

$$PV = (\mu_1 + \mu_2 + \dots)\, RT \tag{12.7}$$

$$\text{i.e. } P = \mu_1 \frac{RT}{V} + \mu_2 \frac{RT}{V} + \dots \tag{12.8}$$

$$= P_1 + P_2 + \dots \tag{12.9}$$

Clearly $P_1 = \mu_1 RT/V$ is the pressure that 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.

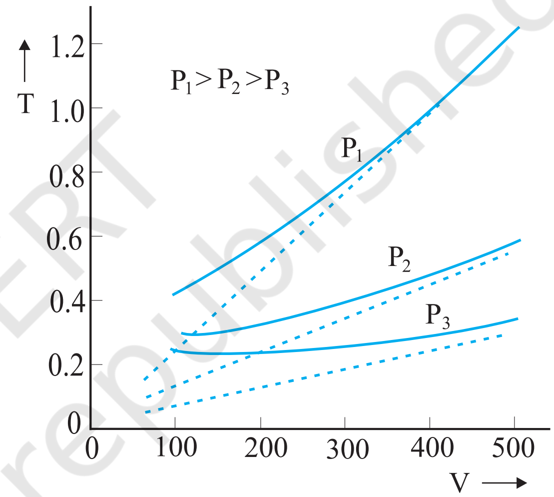


Fig. 12.3 *Experimental T-V curves (solid lines) for CO_2 at three pressures compared with Charles' law (dotted lines). T is in units of 300 K and V in units of 0.13 litres.*

We next consider some examples which give us information about the volume occupied by the molecules and the volume of a single molecule.

Example 12.1 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 $1000/0.6 = 1/(6 \times 10^{-4})$ 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 12.2 Estimate the volume of a water molecule using the data in Example 12.1.

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

$= (3 \times 10^{-26}\ \text{kg})/\ (1000\ \text{kg m}^{-3})$

$= 3 \times 10^{-29}\ \text{m}^3$

$= (4/3)\ \pi\ (\text{Radius})^3$

Hence, Radius $\approx 2 \times 10^{-10}$ m = 2 Å

Example 12.3 What is the average distance between atoms (interatomic distance) in water? Use the data given in Examples 12.1 and 12.2.

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. 12.1). 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 $2 \times 20 = 40$ Å.

Example 12.4 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

$$\frac{\rho_1}{\rho_2} = \frac{m_1/V}{m_2/V} = \frac{m_1}{m_2} = \frac{\mu_1}{\mu_2} \times \left(\frac{M_1}{M_2}\right)$$

$$= \frac{3}{2} \times \frac{20.2}{32.0} = 0.947$$

12.4 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, interaction between 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.

12.4.1 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. 12.4. A molecule with velocity (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 = $2mv_x$.

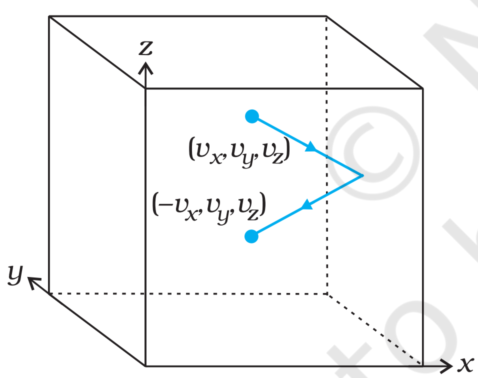


Fig. 12.4 *Elastic collision of a gas molecule with the wall of the container.*

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 ½$A\, v_x\, \Delta t\, n$, where n is the number of molecules per unit volume. The total momentum transferred to the wall by these molecules in time Δt is :

$$Q = (2mv_x)\,(½\, n\, A\, v_x\, \Delta t) \tag{12.10}$$

The force on the wall is the rate of momentum transfer $Q/\Delta t$ and pressure is force per unit area :

$$P = Q\,/(A\,\Delta t) = n\, m\, v_x^2 \tag{12.11}$$

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:

$$P = n\, m\, \overline{v_x^2} \tag{12.12}$$

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,

$$\overline{v_x^2} = \overline{v_y^2} = \overline{v_z^2}$$

$$= (1/3)\,[\overline{v_x^2} + \overline{v_y^2} + \overline{v_z^2}\,] = (1/3)\,\overline{v^2} \tag{12.13}$$

where v is the speed and $\overline{v^2}$ denotes the mean of the squared speed. Thus

$$P = (1/3)\, n\, m\, \overline{v^2} \tag{12.14}$$

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. 9, 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 ½ $n\, Av_x\, \Delta t$. 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.

12.4.2 Kinetic Interpretation of Temperature

Equation (13.14) can be written as

$$PV = (1/3)\, nV m\, \overline{v^2} \qquad (12.15a)$$

$$PV = (2/3)\, N \times ½\, m\, \overline{v^2} \qquad (12.15b)$$

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*,

$$E = N\ (1/2)\, m\, \overline{v^2} \qquad (12.16)$$

Equation (12.15) then gives :

$$PV = (2/3)\, E \qquad (12.17)$$

We are now ready for a kinetic interpretation of temperature. Combining Eq. (12.17) with the ideal gas Eq. (12.3), we get

$$E = (3/2)\ k_B\, NT \qquad (12.18)$$

or $E/N = ½\ m\, \overline{v^2} = (3/2)\, k_B T$ (12.19)

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. (12.18) 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 (12.14) becomes

$$P = (1/3)\, [n_1 m_1 \overline{v_1^2} + n_2 m_2 \overline{v_2^2} + \dots] \qquad (12.20)$$

In equilibrium, the average kinetic energy of the molecules of different gases will be equal. That is,

$$½\ m_1 \overline{v_1^2} = ½\, m_2 \overline{v_2^2} = (3/2)\, k_B\, T$$

so that

$$P = (n_1 + n_2 + \dots)\, k_B\, T \qquad (12.21)$$

which is Dalton's law of partial pressures.

From Eq. (12.19), 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 $< v^2 >$.)

$$v_{rms} = 516\ \text{m s}^{-1}$$

The speed is of the order of the speed of sound in air. It follows from Eq. (12.19) that at the same temperature, lighter molecules have greater rms speed.

> **Example 12.5** 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 ½ $m\, v_{rms}{}^2$ = average kinetic energy per molecule = (3/2)) $k_B T$ where m is the mass

* *E denotes the translational part of the internal energy U that may include energies due to other degrees of freedom also. See section 12.5.*

of a molecule of the gas. Therefore,

$$\frac{(\mathbf{v}^2_{rms})_{Ar}}{(\mathbf{v}^2_{rms})_{Cl}} = \frac{(m)_{Cl}}{(m)_{Ar}} = \frac{(M)_{Cl}}{(M)_{Ar}} = \frac{70.9}{39.9} = 1.77$$

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 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 12.6 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 $m<v^2>$ 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 $\frac{\Delta V}{V} = 0.44$ %.

[^{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. 12.5). 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 12.12). Can you guess the explanation from the above answer?

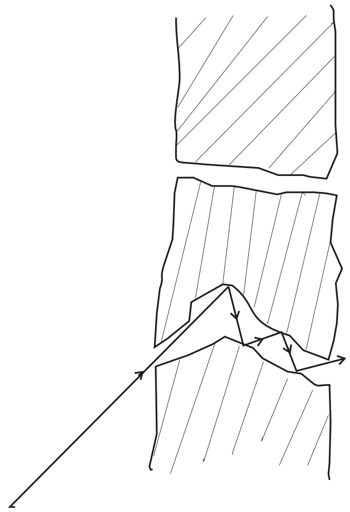


Fig. 12.5 *Molecules going through a porous wall.*

Example 12.7 (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.5 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 used a heavy cricket bat while playing. Did 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.) ◀

12.5 LAW OF EQUIPARTITION OF ENERGY

The kinetic energy of a single molecule is

$$\varepsilon_t = \frac{1}{2}mv_x^2 + \frac{1}{2}mv_y^2 + \frac{1}{2}mv_z^2 \qquad (12.22)$$

For a gas in thermal equilibrium at temperature T the average value of energy denoted by $<\varepsilon_t>$ is

$$\langle \varepsilon_t \rangle = \left\langle \frac{1}{2}mv_x^2 \right\rangle + \left\langle \frac{1}{2}mv_y^2 \right\rangle + \left\langle \frac{1}{2}mv_z^2 \right\rangle = \frac{3}{2}k_BT \qquad (12.23)$$

Since there is no preferred direction, Eq. (12.23) implies

$$\left\langle \frac{1}{2}\ mv_x^2 \right\rangle = \frac{1}{2}\ k_BT\ ,\left\langle \frac{1}{2}\ mv_y^2 \right\rangle = \frac{1}{2}\ k_BT\ ,$$

$$\left\langle \frac{1}{2}\ mv_z^2 \right\rangle = \frac{1}{2}\ k_BT \qquad (12.24)$$

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. (12.24) 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 12.6 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 ε_t and rotational energy ε_r.

$$\varepsilon_t + \varepsilon_r = \frac{1}{2}mv_x^2 + \frac{1}{2}mv_y^2 + \frac{1}{2}mv_z^2 + \frac{1}{2}I_1\omega_1^2 + \frac{1}{2}I_2\omega_2^2 \qquad (12.25)$$

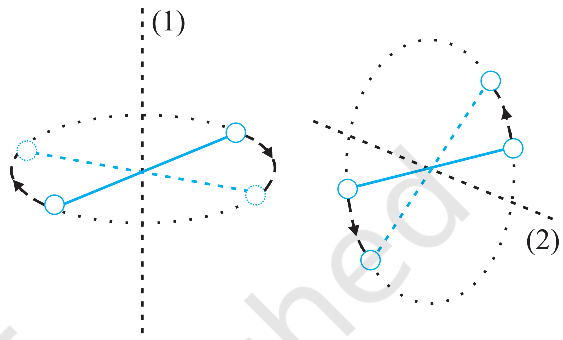


Fig. 12.6 *The two independent axes of rotation of a diatomic molecule*

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:

$$\varepsilon_v = \frac{1}{2}m\left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2 + \frac{1}{2}ky^2$$

$$\varepsilon = \varepsilon_t + \varepsilon_r + \varepsilon_v \qquad (12.26)$$

where k is the force constant of the oscillator and y the vibrational co-ordinate.

Once again the vibrational energy terms in Eq. (12.26) contain squared terms of vibrational variables of motion y and $\mathrm{d}y/\mathrm{d}t$.

* *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 12.6.*

At this point, notice an important feature in Eq.(12.26). While each translational and rotational degree of freedom has contributed only one 'squared term' in Eq.(12.26), 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. The 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.

12.6 SPECIFIC HEAT CAPACITY

12.6.1 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

$$U = \frac{3}{2}k_BT \times N_A = \frac{3}{2}RT \qquad (12.27)$$

The molar specific heat at constant volume, C_v, is

$$C_v \text{ (monatomic gas)} = \frac{dU}{dT} = \frac{3}{2}RT \qquad (12.28)$$

For an ideal gas,

$$C_p - C_v = R \qquad (12.29)$$

where C_p is the molar specific heat at constant pressure. Thus,

$$C_p = \frac{5}{2}R \qquad (12.30)$$

The ratio of specific heats $\gamma = \frac{C_p}{C_v} = \frac{5}{3}$ (12.31)

12.6.2 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

$$U = \frac{5}{2}k_BT \times N_A = \frac{5}{2}RT \qquad (12.32)$$

The molar specific heats are then given by

$$C_v \text{ (rigid diatomic)} = \frac{5}{2}R,\ C_p = \frac{7}{2}R \qquad (12.33)$$

$$\gamma \text{ (rigid diatomic)} = \frac{7}{5} \qquad (12.34)$$

If the diatomic molecule is not rigid but has in addition a vibrational mode

$$U = \left(\frac{5}{2}k_BT + k_BT\right)N_A = \frac{7}{2}RT$$

$$C_v = \frac{7}{2}R,\ C_p = \frac{9}{2}R,\ \gamma = \frac{9}{7}R \qquad (12.35)$$

12.6.3 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

$$U = \left(\frac{3}{2}k_BT + \frac{3}{2}k_BT + f k_BT\right)N_A$$

i.e., $C_v = (3 + f)\ R$, $C_p = (4 + f)\ R$,

$$\gamma = \frac{(4+f)}{(3+f)} \qquad (12.36)$$

Note that $C_p - C_v = R$ is true for any ideal gas, whether mono, di or polyatomic.

Table 12.1 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 12.2. 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 as given in Table12.1 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 verified experimentally at ordinary temperatures.

Table 12.1 **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

Table12.2 **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

Example 12.8 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 mo1^{-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.

12.6.4 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,

$$C = \frac{\Delta Q}{\Delta T} = \frac{\Delta U}{\Delta T} = 3R \qquad (12.37)$$

Table 12.3 **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 12.3 shows the prediction generally agrees with experimental values at ordinary temperature (Carbon is an exception).

12.7 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 move straight unhindered; their paths keep getting incessantly deflected.

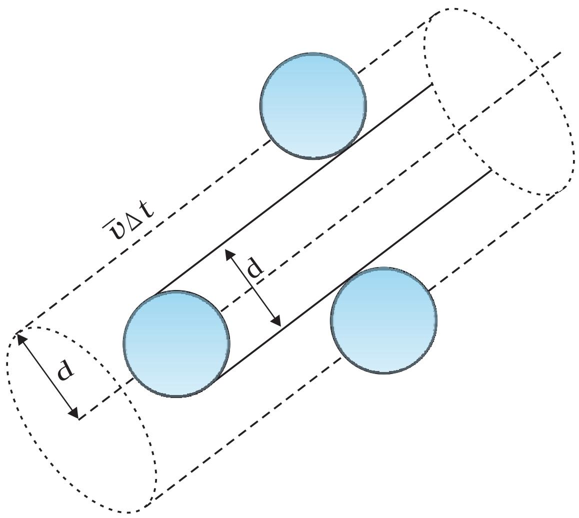


Fig. 12.7 *The volume swept by a molecule in time Δt in which any molecule will collide with it.*

Suppose the molecules of a gas are spheres of diameter *d*. Focus on a single molecule with the average speed <*v*>. It will suffer collision with any molecule that comes within a distance *d* between the centres. In time Δt, it sweeps a volume $\pi d^2 <v> \Delta t$ wherein any other molecule will collide with it (see Fig. 12.7). If *n* is the number of molecules per unit volume, the molecule suffers $n\pi d^2 <v> \Delta t$ collisions in time Δt. Thus the rate of collisions is $n\pi d^2 <v>$ or the time between two successive collisions is on the average,

$$\tau = 1/(n\pi <v> d^2) \qquad (12.38)$$

The average distance between two successive collisions, called the mean free path *l*, is :

$$l = <v> \tau = 1/(n\pi d^2) \qquad (12.39)$$

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 <*v*> by $<v_r>$ in Eq. (12.38). A more exact treatment gives

$$l = 1/\left(\sqrt{2}\; n\pi d^2\right) \qquad (12.40)$$

Let us estimate *l* and τ for air molecules with average speeds <*v*> = (485m/s). At STP

$$n = \frac{\left(0.02 \times 10^{23}\right)}{\left(22.4 \times 10^{-3}\right)}$$

$= 2.7 \times 10^{25}\ \mathrm{m}^{-3.}$

Taking, $d = 2 \times 10^{-10}$ m,

$\tau = 6.1 \times 10^{-10}$ s

and $l = 2.9 \times 10^{-7}\ \mathrm{m} \approx 1500\ d$ (12.41)

As expected, the mean free path given by Eq. (12.40) 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 12.9*** Estimate the mean free path for a water molecule in water vapour at 373 K. Use information from Exercises 12.1 and Eq. (12.41) above.

Answer The *d* for water vapour is same as that of air. The number density is inversely proportional to absolute temperature.

$$\text{So } n = 2.7 \times 10^{25} \times \frac{273}{373} = 2 \times 10^{25}\,\mathrm{m}^{-3}$$

Hence, mean free path $l = 4 \times 10^{-7}$ 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

$$PV = \mu RT = k_B NT$$

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

$$P = \frac{1}{3} n m \overline{v^2}$$

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.

$$\frac{1}{2} m \overline{v^2} = \frac{3}{2} k_B T, \quad v_{rms} = \left(\overline{v^2}\right)^{1/2} = \sqrt{\frac{3k_B T}{m}}$$

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 :

$$l = \frac{1}{\sqrt{2}\, n \pi d^2}$$

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 $\frac{1}{2}\, k_B T$. 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 $2 \times \frac{1}{2}\, k_B T = k_B T$.

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 $\frac{1}{2}\, mv^2$ 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

12.1 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.

12.2 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.

12.3 Figure 12.8 shows plot of PV/T versus P for 1.00×10^{-3} kg of oxygen gas at two different temperatures.

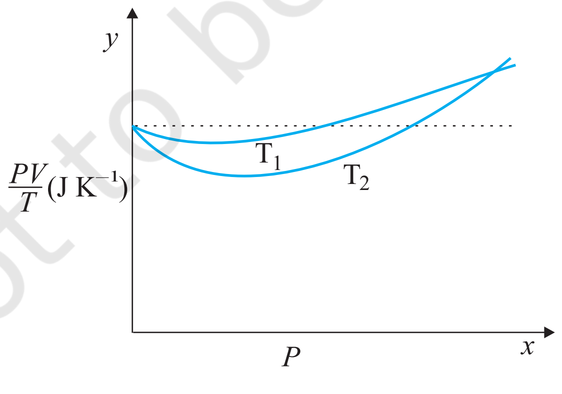


Fig. 12.8

(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 pressure high temperature region of the plot) ? (Molecular mass of H_2 = 2.02 u, of O_2 = 32.0 u, R = 8.31 J mo1^{-1} K^{-1}.)

12.4 An oxygen cylinder of volume 30 litre 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).

12.5 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 ?

12.6 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.

12.7 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).

12.8 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 ?

12.9 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).

12.10 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).

CHAPTER THIRTEEN

OSCILLATIONS

13.1 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. Examples of such periodic to and fro motion abound: a boat tossing up and down in a river, the piston in a steam engine going back and forth, etc. 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. In a solid, the atoms vibrate about their equilibrium positions, the average energy of vibrations being proportional to temperature. AC power supply give voltage that oscillates alternately going positive and negative about the mean value (zero).

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.

13.2 PERIODIC AND OSCILLATORY MOTIONS

Fig. 13.1 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. 13.1 (a). If a child climbs up a step, comes down, and repeats the process identically, its height above the ground would look like that in Fig. 13.1 (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. 13.1 (c). Note that both the curved parts in Fig. 13.1 (c) are sections of a parabola given by the Newton's equation of motion (see section 2.6),

$h = ut + \frac{1}{2}gt^2$ 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.**

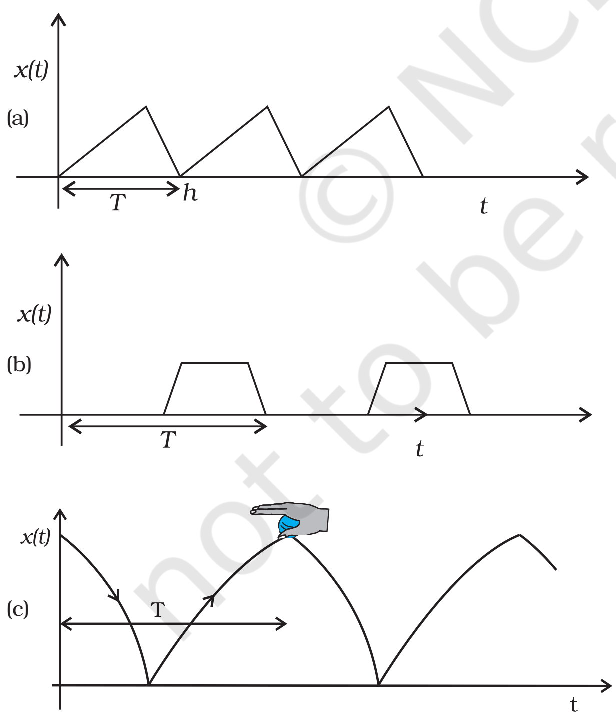


Fig. 13.1 Examples of periodic motion. The period T is shown in each case.

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.

13.2.1 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

$$\nu = 1/T \qquad (13.1)$$

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,

$$1 \text{ hertz} = 1 \text{ Hz} = 1 \text{ oscillation per second} = 1 \text{ s}^{-1} \qquad (13.2)$$

Note, that the frequency, ν, is not necessarily an integer.

Example 13.1 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

13.2.2 Displacement

In section 3.2, 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 the spring is fixed to a rigid wall [see Fig. 13.2(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. 13.2(b)]. The term displacement is not always to be referred

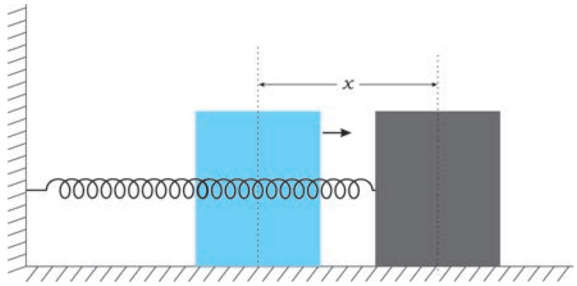


Fig. 13.2(a) *A block attached to a spring, the other end of which is fixed to a rigid wall. The block moves on a frictionless surface. The motion of the block can be described in terms of its distance or displacement x from the equilibrium position.*

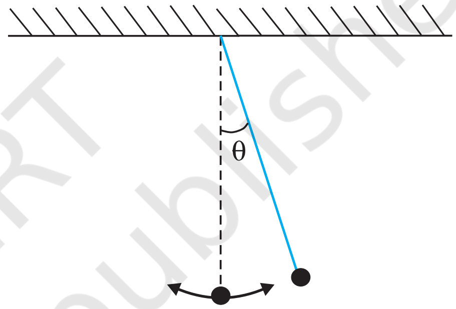


Fig.13.2(b) *An oscillating simple pendulum; its motion can be described in terms of angular displacement θ from the vertical.*

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 AC 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.

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

$$f(t) = A \cos \omega t \qquad (13.3a)$$

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

$$T = \frac{2\pi}{\omega} \tag{13.3b}$$

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,

$$f(t) = A \sin \omega t + B \cos \omega t \tag{13.3c}$$

is also a periodic function with the same period T. Taking,

$$A = D \cos \phi \text{ and } B = D \sin \phi$$

Eq. (13.3c) can be written as,

$$f(t) = D \sin (\omega t + \phi) \,, \tag{13.3d}$$

Here D and ϕ are constant given by

$$D = \sqrt{A^2 + B^2} \text{ and } \phi = \tan^{-1}\left(\frac{B}{A}\right)$$

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 13.2*** 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.

13.3 SIMPLE HARMONIC MOTION

Consider a particle oscillating back and forth about the origin of an x-axis between the limits $+A$ and $-A$ as shown in Fig. 13.3. This oscillatory motion is said to be simple harmonic if the displacement x of the particle from the origin varies with time as :

$$x(t) = A \cos (\omega t + \phi) \tag{13.4}$$

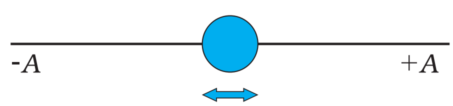


Fig. 13.3 *A particle vibrating back and forth about the origin of x-axis, between the limits +A and –A.*

where A, ω and ϕ are constants.

Thus, simple harmonic motion (SHM) is not any periodic motion but one in which displacement is a sinusoidal function of time. Fig. 13.4 shows the positions of a particle executing SHM at discrete value of time, each interval of time being $T/4$, where T is the period of motion. Fig. 13.5 plots the graph of x versus t, which gives the values of displacement as a continuous function of time. The quantities A,

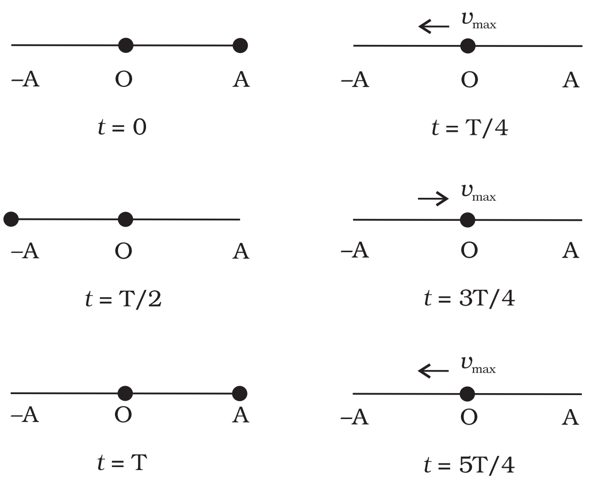


Fig. 13.4 *The location of the particle in SHM at the discrete values t = 0, T/4, T/2, 3T/4, T, 5T/4. The time after which motion repeats itself is T. T will remain fixed, no matter what location you choose as the initial (t = 0) location. The speed is maximum for zero displacement (at x = 0) and zero at the extremes of motion.*

ω and ϕ which characterize a given SHM have standard names, as summarised in Fig. 13.6. Let us understand these quantities.

The amplitutde A of SHM is the magnitude of maximum displacement of the particle. [Note, A can be taken to be positive without

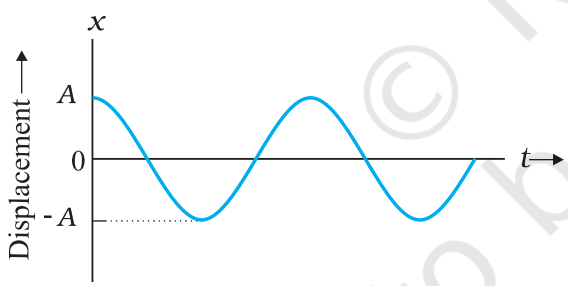


Fig. 13.5 *Displacement as a continuous function of time for simple harmonic motion.*

$x(t)$	:	displacement x as a function of time t
A	:	amplitude
ω	:	angular frequency
$\omega t + \phi$	:	phase (time-dependent)
ϕ	:	phase constant

Fig. 13.6 *The meaning of standard symbols in Eq. (13.4)*

any loss of generality]. As the cosine function of time varies from +1 to –1, the displacement varies between the extremes A and $-A$. Two simple harmonic motions may have same ω and ϕ but different amplitudes A and B, as shown in Fig. 13.7 (a).

While the amplitude A is fixed for a given SHM, the state of motion (position and velocity) of the particle at any time t is determined by the

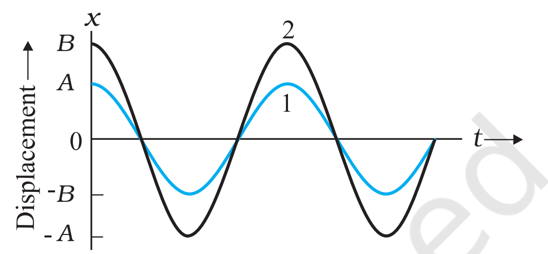


Fig. 13.7 (a) *A plot of displacement as a function of time as obtained from Eq. (14.4) with ϕ = 0. The curves 1 and 2 are for two different amplitudes A and B.*

argument $(\omega t + \phi)$ in the cosine function. This time-dependent quantity, $(\omega t + \phi)$ is called the *phase* of the motion. The value of plase at $t = 0$ is ϕ and is called the *phase constant* (or *phase angle*). If the amplitude is known, ϕ can be determined from the displacement at $t = 0$. Two simple harmonic motions may have the same A and ω but different phase angle ϕ, as shown in Fig. 13.7 (b).

Finally, the quantity ω can be seen to be related to the period of motion T. Taking, for simplicity, $\phi = 0$ in Eq. (13.4), we have

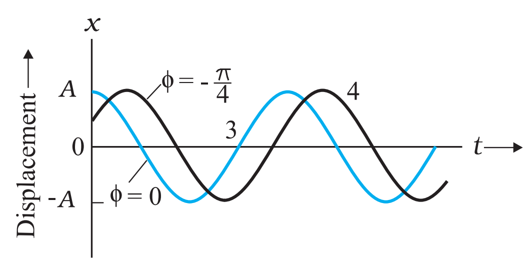


Fig. 13.7 (b) *A plot obtained from Eq. (13.4). The curves 3 and 4 are for ϕ = 0 and $-\pi/4$ respectively. The amplitude A is same for both the plots.*

$$x(t) = A \cos \omega t \quad (13.5)$$

Since the motion has a period T, $x(t)$ is equal to $x(t + T)$. That is,

$$A \cos \omega t = A \cos \omega (t + T) \quad (13.6)$$

Now the cosine function is periodic with period 2π, i.e., it first repeats itself when the argument changes by 2π. Therefore,

$$\omega(t + T) = \omega t + 2\pi$$

that is $\omega = 2\pi/ T$ (13.7)

ω is called the angular frequency of SHM. Its S.I. unit is radians per second. Since the frequency of oscillations is simply $1/T$, ω is 2π times the frequency of oscillation. Two simple harmonic motions may have the same A and ϕ, but different ω, as seen in Fig. 13.8. In this plot the curve (b) has half the period and twice the frequency of the curve (a).

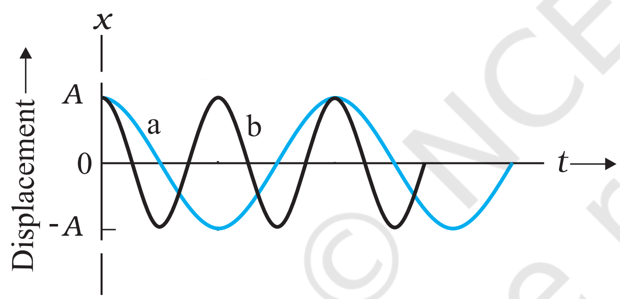


Fig. 13.8 *Plots of Eq. (13.4) for ϕ = 0 for two different periods.*

> **Example 13.3** 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) $\sin^2 \omega t$

$$= \tfrac{1}{2} - \tfrac{1}{2} \cos 2\omega t$$

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.

13.4 SIMPLE HARMONIC MOTION AND UNIFORM CIRCULAR MOTION

In this section, we show that the projection of uniform circular motion on a diameter of the circle follows simple harmonic motion. A simple experiment (Fig. 13.9) helps us visualise this connection. 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.

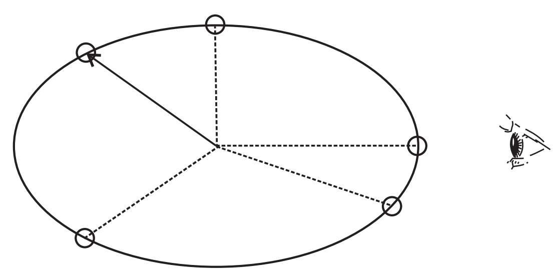


Fig. 13.9 *Circular motion of a ball in a plane viewed edge-on is SHM.*

Fig. 13.10 describes the same situation mathematically. Suppose a particle P is moving uniformly on a circle of radius A with angular speed ω. The sense of rotation is anticlockwise. The initial position vector of the particle, i.e., the vector $\overrightarrow{OP}$ at $t = 0$ makes an angle of ϕ with the positive direction of x-axis. In time t, it will cover a further angle ωt and its position vector

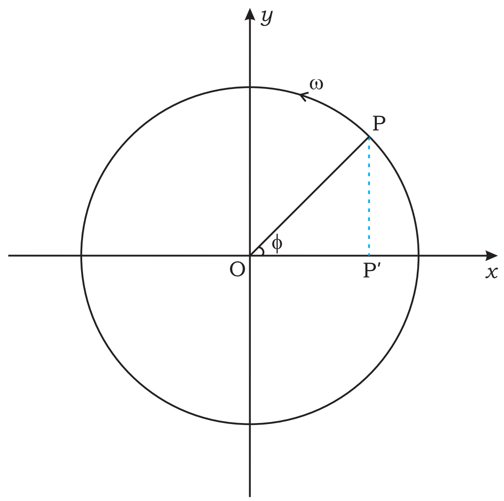


Fig. 13.10

will make an angle of $\omega t + \phi$ with the +ve x-axis. Next, consider the projection of the position vector OP on the x-axis. This will be OP′. The position of P′ on the x-axis, as the particle P moves on the circle, is given by

$$x(t) = A\cos(\omega t + \phi)$$

which is the defining equation of SHM. This shows that if P moves uniformly on a circle, its projection P′ on a diameter of the circle executes SHM. The particle P and the circle on which it moves are sometimes referred to as the *reference particle* and the *reference circle*, respectively.

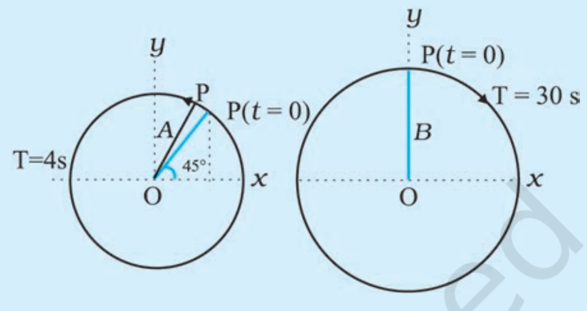
We can take projection of the motion of P on any diameter, say the y-axis. In that case, the displacement $y(t)$ of P′ on the y-axis is given by

$$y = A\sin(\omega t + \phi)$$

which is also an SHM of the same amplitude as that of the projection on x-axis, but differing by a phase of $\pi/2$.

In spite of this connection between circular motion and SHM, the force acting on a particle in linear simple harmonic motion is very different from the centripetal force needed to keep a particle in uniform circular motion.

Example 13.4 The figure given below depicts two circular motions. The radius of the circle, the period of revolution, the initial position and the sense of revolution are indicated in the figures. Obtain the simple harmonic motions of the x-projection of the radius vector of the rotating particle P in each case.



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,

$$x(t) = A\cos\left(\frac{2\pi}{T}t + \frac{\pi}{4}\right)$$

For $T = 4$ s,

$$x(t) = A\cos\left(\frac{2\pi}{4}t + \frac{\pi}{4}\right)$$

which is a SHM of amplitude A, period 4 s, and an initial phase* = $\frac{\pi}{4}$.

* *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(15°) 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.*

(b) In this case at $t = 0$, OP makes an angle of $90^o = \dfrac{\pi}{2}$ with the x-axis. After a time t, it covers an angle of $\dfrac{2\pi}{T}t$ in the clockwise sense and makes an angle of $\left(\dfrac{\pi}{2} - \dfrac{2\pi}{T}t\right)$ with the x-axis. The projection of OP on the x-axis at time t is given by

$$x(t) = B\cos\left(\frac{\pi}{2} - \frac{2\pi}{T}t\right)$$

$$= B\sin\left(\frac{2\pi}{T}t\right)$$

For $T = 30$ s,

$$x(t) = B\sin\left(\frac{\pi}{15}t\right)$$

Writing this as $x(t) = B\cos\left(\dfrac{\pi}{15}t - \dfrac{\pi}{2}\right)$, and comparing with Eq. (13.4). We find that this represents a SHM of amplitude B, period 30 s, and an initial phase of $-\dfrac{\pi}{2}$. ◀

13.5 VELOCITY AND ACCELERATION IN SIMPLE HARMONIC MOTION

The speed of a particle v in uniform circular motion is its angular speed ω times the radius of the circle A.

$$v = \omega A \tag{13.8}$$

The direction of velocity $\overline{\mathbf{v}}$ at a time t is along the tangent to the circle at the point where the particle is located at that instant. From the geometry of Fig. 13.11, it is clear that the velocity of the projection particle P′ at time t is

$$v(t) = -\omega A\sin(\omega t + \phi) \tag{13.9}$$

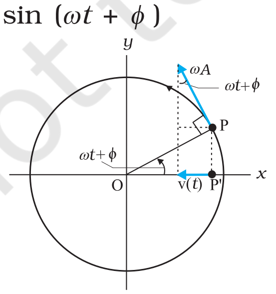


Fig. 13.11 *The velocity, v (t), of the particle P′ is the projection of the velocity* $\overline{\mathbf{v}}$ *of the reference particle, P.*

where the negative sign shows that $v(t)$ has a direction opposite to the positive direction of x-axis. Eq. (13.9) gives the instantaneous velocity of a particle executing SHM, where displacement is given by Eq. (13.4). We can, of course, obtain this equation without using geometrical argument, directly by differentiating (Eq. 13.4) with respect of t:

$$v(t) = \frac{\mathrm{d}}{\mathrm{d}t}x(t) \tag{13.10}$$

The method of reference circle can be similarly used for obtaining instantaneous acceleration of a particle undergoing SHM. We know that the centripetal acceleration of a particle P in uniform circular motion has a magnitude v^2/A or $\omega^2 A$, and it is directed towards the centre i.e., the direction is along PO. The instantaneous acceleration of the projection particle P′ is then (See Fig. 13.12)

$$a(t) = -\omega^2 A\cos(\omega t + \phi)$$

$$= -\omega^2 x(t) \tag{13.11}$$

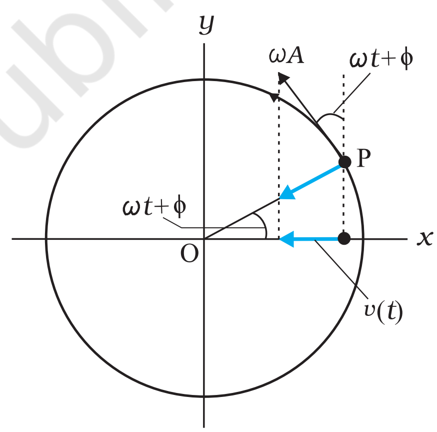


Fig. 13.12 *The acceleration, a(t), of the particle P′ is the projection of the acceleration* **a** *of the reference particle P.*

Eq. (13.11) gives the acceleration of a particle in SHM. The same equation can again be obtained directly by differentiating velocity $v(t)$ given by Eq. (13.9) with respect to time:

$$a(t) = \frac{\mathrm{d}}{\mathrm{d}t}v(t) \tag{13.12}$$

We note from Eq. (13.11) the important property that acceleration of a particle in SHM is proportional to displacement. For $x(t) > 0$, $a(t) < 0$ and for $x(t) < 0$, $a(t) > 0$. Thus, whatever the value of x between $-A$ and A, the acceleration $a(t)$ is always directed towards the centre.

For simplicity, let us put $\phi = 0$ and write the expression for $x(t)$, $v(t)$ and $a(t)$

$x(t) = A \cos \omega t$, $v(t) = -\omega A \sin \omega t$, $a(t) = -\omega^2 A \cos \omega t$

The corresponding plots are shown in Fig. 13.13. All quantities vary sinusoidally with time; only their maxima differ and the different plots differ in phase. x varies between $-A$ to A; $v(t)$ varies from $-\omega A$ to ωA and $a(t)$ from $-\omega^2 A$ to $\omega^2 A$. With respect to displacement plot, velocity plot has a phase difference of $\pi/2$ and acceleration plot has a phase difference of π.

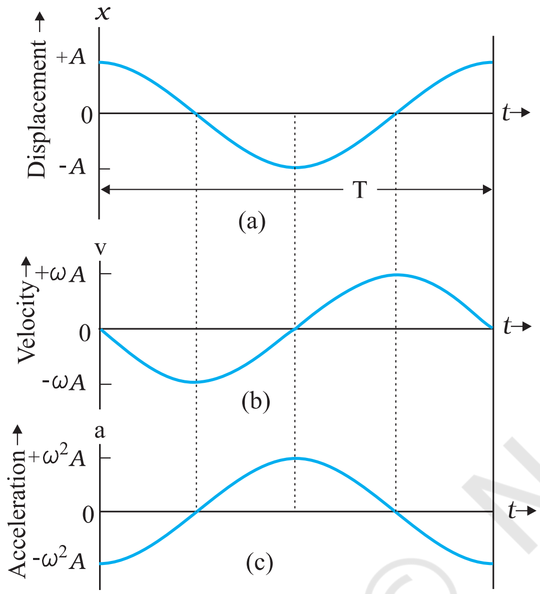


Fig. 13.13 *Displacement, velocity and acceleration of a particle in simple harmonic motion have the same period T, but they differ in phase*

> ***Example 13.5*** 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 $= (5.0 \text{ m}) \cos [(2\pi \text{ s}^{-1}) \times 1.5 \text{ s} + \pi/4]$
$= (5.0 \text{ m}) \cos [(3\pi + \pi/4)]$
$= -5.0 \times 0.707$ m
$= -3.535$ m

(b) Using Eq. (13.9), 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.(13.10), the acceleration of the body
$= -(2\pi \text{ s}^{-1})^2 \times$ displacement
$= -(2\pi \text{ s}^{-1})^2 \times (-3.535 \text{ m})$
$= 140 \text{ m s}^{-2}$ ◀

13.6 FORCE LAW FOR SIMPLE HARMONIC MOTION

Using Newton's second law of motion, and the expression for acceleration of a particle undergoing SHM (Eq. 13.11), the force acting on a particle of mass m in SHM is

$$F(t) = ma$$
$$= -m\omega^2 x(t)$$

i.e., $F(t) = -k\, x(t)$ (13.13)

where $k = m\omega^2$ (13.14a)

or $\omega = \sqrt{\dfrac{k}{m}}$ (13.14b)

Like acceleration, force is always directed towards the mean position—hence it is sometimes called the restoring force in SHM. To summarise the discussion so far, simple harmonic motion can be defined in two equivalent ways, either by Eq. (13.4) for displacement or by Eq. (13.13) that gives its force law. Going from Eq. (13.4) to Eq. (13.13) required us to differentiate two times. Likewise, by integrating the force law Eq. (13.13) two times, we can get back Eq. (13.4).

Note that the force in Eq. (13.13) is linearly proportional to $x(t)$. A particle oscillating under such a force is, therefore, calling a linear harmonic oscillator. In the real world, the force may contain small additional terms proportional to x^2, x^3, etc. These then are called non-linear oscillators.

> ***Example 13.6*** Two identical springs of spring constant k are attached to a block of mass m and to fixed supports as shown in Fig. 13.14. 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.

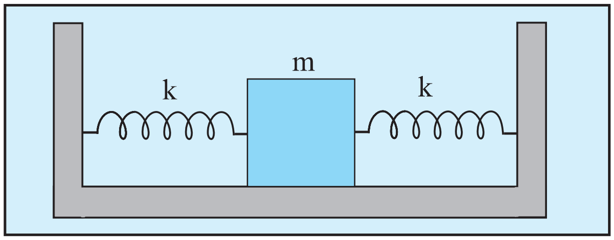


Fig. 13.14

Answer Let the mass be displaced by a small distance x to the right side of the equilibrium position, as shown in Fig. 13.15. Under this situation the spring on the left side gets

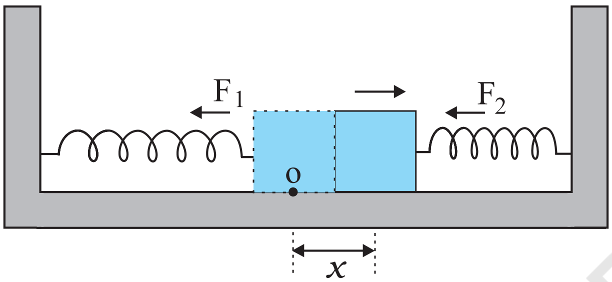


Fig. 13.15

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,

$$T = 2\pi\sqrt{\frac{m}{2k}}$$

13.7 ENERGY IN SIMPLE HARMONIC MOTION

Both kinetic and potential energies of a particle in SHM vary between zero and their maximum values.

In section 13.5 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

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

$$= \frac{1}{2}m\,\omega^2 A^2 \sin^2(\omega t + \phi)$$

$$= \frac{1}{2}k\,A^2 \sin^2(\omega t + \phi) \qquad (13.15)$$

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 (U) 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

$$U = \frac{1}{2}k\,x^2 \qquad (13.16)$$

Hence the potential energy of a particle executing simple harmonic motion is,

$$U(x) = \frac{1}{2}k\,x^2$$

$$= \frac{1}{2}k\,A^2 \cos^2(\omega t + \phi) \qquad (13.17)$$

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. (13.15) and (13.17) that the total energy, E, of the system is,

$$E = U + K$$

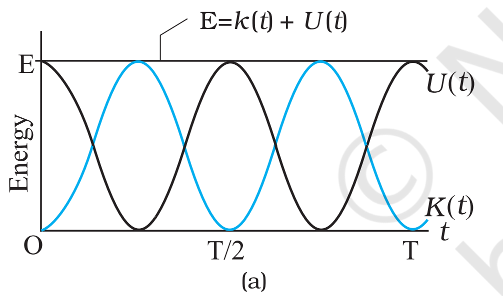
$$= \frac{1}{2} k\, A^2 \cos^2(\omega t + \phi) + \frac{1}{2} k\, A^2 \sin^2(\omega t + \phi)$$

$$= \frac{1}{2} k\, A^2 \left[\cos^2(\omega t + \phi) + \sin^2(\omega t + \phi)\right]$$

Using the familiar trigonometric identity, the value of the expression in the brackets is unity. Thus,

$$E = \frac{1}{2} k\, A^2 \tag{13.18}$$

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. 13.16.



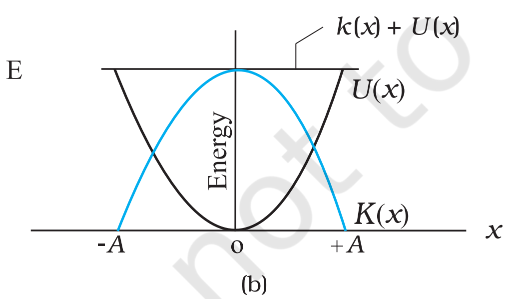


Fig. 13.16 *Kinetic energy, potential energy and total energy as a function of time [shown in (a)] and displacement [shown in (b)] of a particle in SHM. The kinetic energy and potential energy both repeat after a period T/2. The total energy remains constant at all t or x.*

Observe that both kinetic energy and potential energy in SHM are seen to be always positive in Fig. 13.16. Kinetic energy can, of course, be never negative, since it is proportional to the square of speed. Potential energy is positive by choice of the undermined constant in potential energy. Both kinetic energy and potential energy peak twice during each period of SHM. For $x = 0$, the energy is kinetic; at the extremes $x = \pm A$, it is all potential energy. In the course of motion between these limits, kinetic energy increases at the expense of potential energy or vice-versa.

▶ ***Example 13.7*** 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. (13.14b), 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

$$= 0.1 \times 7.07 \quad 0.866 \text{ m s}^{-1}$$

$$= 0.61 \text{ m s}^{-1}$$

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. ◀

13.8 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.

We shall show that this periodic motion is simple harmonic for small displacements from

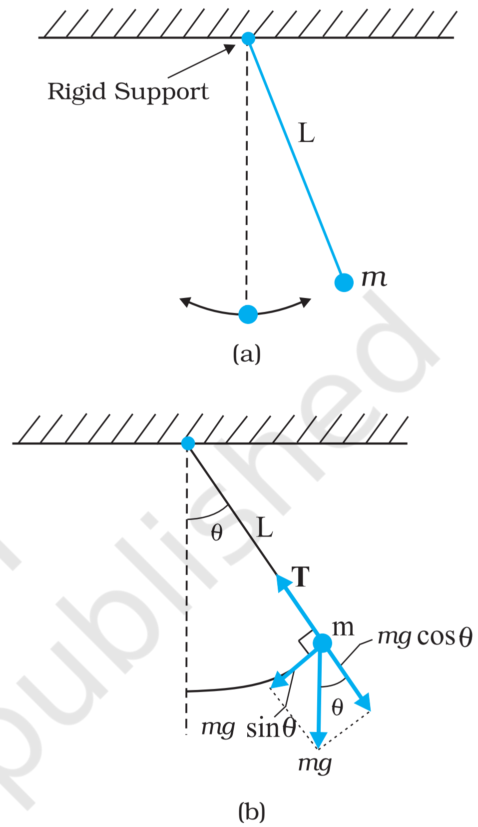


Fig. 13.17 *(a) A bob oscillating about its mean position. (b) The radial force T-mg cosθ provides centripetal force but no torque about the support. The tangential force mg sinθ provides the restoring torque.*

the mean position. Consider simple pendulum — a small bob of mass m tied to an inextensible massless string of length L. The other end of the string is fixed to a rigid support. The bob oscillates in a plane about the vertical line through the support. Fig. 13.17(a) shows this system. Fig. 13.17(b) is a kind of 'free-body' diagram of the simple pendulum showing the forces acting on the bob.

Let θ be the angle made by the string with the vertical. When the bob is at the mean position, $\theta = 0$

There are only two forces acting on the bob; the tension T along the string and the vertical force due to gravity (=mg). The force mg can be resolved into the component $mg\cos\theta$ along the string and $mg\sin\theta$ perpendicular to it. Since the motion of the bob is along a circle of length L and centre at the support point, the bob has a radial acceleration ($\omega^2 L$) and also a tangential acceleration; the latter arises since motion along the arc of the circle is not uniform. The radial acceleration is provided by the net radial force $T - mg\cos\theta$, while the tangential acceleration is provided by $mg\sin\theta$. It is more convenient to work with torque about the support since the radial force gives zero torque. Torque τ about the support is entirely provided by the tangential component of force

$$\tau = -L\,(mg\sin\theta) \tag{13.19}$$

This is the restoring torque that tends to reduce angular displacement — hence the negative sign. By Newton's law of rotational motion,

$$\tau = I\alpha \tag{13.20}$$

where I is the moment of inertia of the system about the support and α is the angular acceleration. Thus,

$$I\alpha = -m\,g\sin\theta\;\;L \tag{13.21}$$

Or,

$$\alpha = -\frac{m\,g\,L}{I}\sin\theta \tag{13.22}$$

We can simplify Eq. (13.22) if we assume that the displacement θ is small. We know that $\sin\theta$ can be expressed as,

$$\sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} \pm \ldots \tag{13.23}$$

where θ is in radians.

Now if θ is small, $\sin\theta$ can be approximated by θ and Eq. (13.22) can then be written as,

$$\alpha = -\frac{mgL}{I}\theta \tag{13.24}$$

In Table 13.1, we have listed the angle θ in degrees, its equivalent in radians, and the value of the function $\sin\theta$. From this table it can be seen that for θ as large as 20 degrees, $\sin\theta$ is nearly the same as θ **expressed in radians**.

Table 13.1 **sin θ as ma function of angle θ**

θ (degrees)	θ (radians)	$\sin\theta$
0	0	0
5	0.087	0.087
10	0.174	0.174
15	0.262	0.259
20	0.349	0.342

Equation (13.24) is mathematically, identical to Eq. (13.11) except that the variable is angular displacement. Hence we have proved that for small q, the motion of the bob is simple harmonic. From Eqs. (13.24) and (13.11),

$$\omega = \sqrt{\frac{mgL}{I}}$$

and

$$T = 2\pi\sqrt{\frac{I}{mgL}} \tag{13.25}$$

Now since the string of the simple pendulum is massless, the moment of inertia I is simply mL^2. Eq. (13.25) then gives the well-known formula for time period of a simple pendulum.

$$T = 2\pi\sqrt{\frac{L}{g}} \tag{13.26}$$

Example 13.8 What is the length of a simple pendulum, which ticks seconds ?

Answer From Eq. (13.26), the time period of a simple pendulum is given by,

$$T = 2\pi\sqrt{\frac{L}{g}}$$

From this relation one gets,

$$L = \frac{gT^2}{4\pi^2}$$

The time period of a simple pendulum, which ticks seconds, is 2 s. Therefore, for $g = 9.8\ \text{m s}^{-2}$ and $T = 2$ s, L is

$$= \frac{9.8(\text{m s}^{-2}) \times 4(\text{s}^2)}{4\pi^2}$$

$$= 1\ \text{m}$$

SUMMARY

1. The motion that repeats itself is called *periodlic motion.*
2. The *period T* is the time reequired for one complete oscillation, or cycle. It is related to the frequency v by,

$$T = \frac{1}{v}$$

The *frequency v* of periodic or oscillatory motion is the number of oscillations per unit time. In the SI, it is measured in hertz :

1 hertz = 1 Hz = 1 oscillation per second = 1s^{-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,

$$\omega = \frac{2\pi}{T} = 2\pi v \qquad \text{(angular frequency)}.$$

4. Simple harmonic motion can also be viewed as 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 in a 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 $K = ½\, mv^2$ and potential energy $U = ½\, kx^2$. 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 Hooke's law restoring force given by $F = -k\,x$ exhibits simple harmonic motion with

$$\omega = \sqrt{\frac{k}{m}} \qquad \text{(angular frequency)}$$

$$T = 2\pi\sqrt{\frac{m}{k}} \qquad \text{(period)}$$

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,

$$T = 2\pi\sqrt{\frac{L}{g}}$$

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 $F = -k\,x$

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 $F = -k\,x$ 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 $\pi/2$. Thus, for example, a particle subject to a force $-m\omega^2 r$ with initial position (0, A) and velocity (ωA, 0) will move uniformly in a circle of radius A.
4. For linear simple harmonic motion with a given ω, two initial conditions are necessary and sufficient to determine the motion completely. The initial conditions 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 :

$$x = A\cos\omega t + B\sin\omega t$$

$$x = A\cos(\omega t + \alpha),\quad x = B\sin(\omega t + \beta)$$

The three forms are completely equivalent (any one can be expressed in terms of any other two forms).

Thus, damped simple harmonic motion is not strictly simple harmonic. It is approximately so only for time intervals much less than $2m/b$ where b is the damping constant.

Exercises

13.1 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 centre of mass.

(d) An arrow released from a bow.

13.2 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.

13.3 Fig. 13.18 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) ?

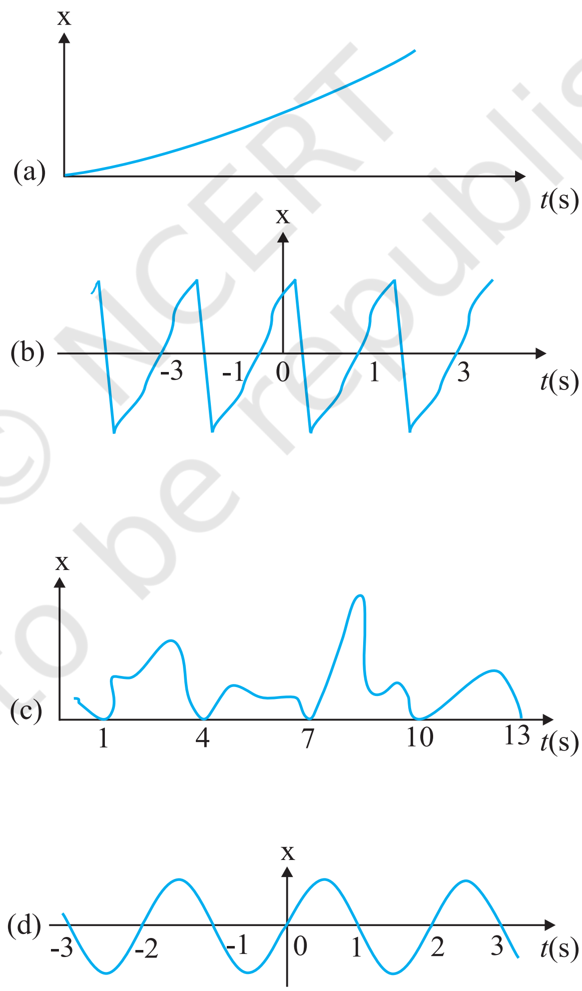


Fig. 18.18

13.4 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) $1 + \omega t + \omega^2 t^2$

13.5 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.

13.6 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) $a = 100x^3$

13.7 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 ω cm/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.

13.8 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 ?

13.9 A spring having with a spring constant 1200 N m^{-1} is mounted on a horizontal table as shown in Fig. 13.19. 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.

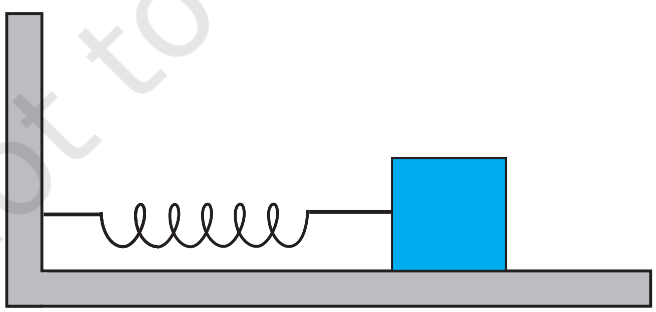


Fig. 13.19

Determine (i) the frequency of oscillations, (ii) maximum acceleration of the mass, and (iii) the maximum speed of the mass.

13.10 In Exercise 13.9, 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?

13.11 Figures 13.20 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.

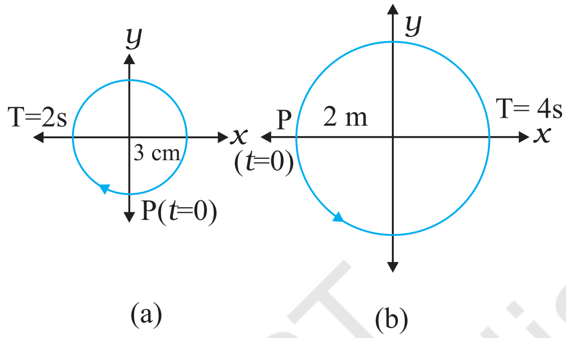


Fig. 13.20

Obtain the corresponding simple harmonic motions of the x-projection of the radius vector of the revolving particle P, in each case.

13.12 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) $x = 2 \cos \pi t$

13.13 Figure 13.21(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 13.21 (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. 13.21(b) is stretched by the same force **F**.

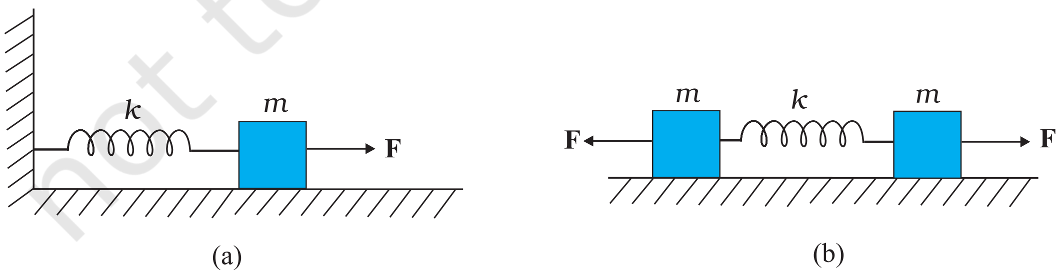


Fig. 13.21

(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 ?

13.14 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 ?

13.15 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})

13.16 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 ?

13.17 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).

13.18 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.

CHAPTER FOURTEEN

11087CH15

WAVES

14.1 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.

Waves transport energy and the pattern of disturbance has information that propagate from one point to another. All our communications essentially depend on transmission of signals through waves. Speech means production of sound waves in air and hearing amounts to their detection. Often, communication involves different kinds of waves. For example, sound waves may be first converted into an electric current signal which in turn may generate an electromagnetic wave that may be transmitted by an optical cable or via a satellite. Detection of the original signal will usually involve these steps in reverse order.

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 most familiar type of waves such as waves on a string, water waves, sound waves, seismic waves, etc. is the so-called mechanical waves. These waves require a medium for propagation, they cannot propagate through vacuum. They involve oscillations of constituent particles and depend on the elastic properties of the medium. The electromagnetic waves that you will learn in Class XII are a different type of wave. Electromagnetic waves do not necessarily require a medium - they can travel through vacuum. Light, radiowaves, X-rays, are all electromagnetic waves. In vacuum, all electromagnetic waves have the same speed c, whose value is :

$$c = 299,792,458 \text{ ms}^{-1}. \qquad (14.1)$$

A third kind of wave is the so-called Matter waves. They are associated with constituents of matter : electrons, protons, neutrons, atoms and molecules. They arise in quantum mechanical description of nature that you will learn in your later studies. Though conceptually more abstract than mechanical or electro-magnetic 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. 14.1. If the spring at one end is pulled suddenly and released, the disturbance travels to the other end. What has

Fig. 14.1 *A collection of springs connected to each other. The end A is pulled suddenly generating a disturbance, which then propagates to the other end.*

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 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.

14.2 TRANSVERSE AND LONGITUDINAL WAVES

We have seen that motion of mechanical waves involves oscillations of constituents of the medium. If the constituents of the medium oscillate perpendicular to the direction of wave propagation, we call the wave a transverse wave. If they oscillate along the direction of wave propagation, we call the wave a longitudinal wave.

Fig.14.2 shows the propagation of a single pulse along a string, resulting from a single up and down jerk. If the string is very long compared

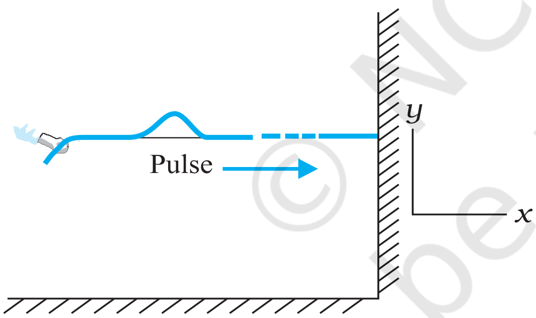


Fig. 14.2 *When a pulse travels along the length of a stretched string (x-direction), the elements of the string oscillate up and down (y-direction)*

to the size of the pulse, the pulse will damp out before it reaches the other end and reflection from that end may be ignored. Fig. 14.3 shows a similar situation, but this time the external agent gives a continuous periodic sinusoidal up and down jerk to one end of the string. The resulting disturbance on the string is then a sinusoidal wave. In either case the elements of the string oscillate about their equilibrium mean

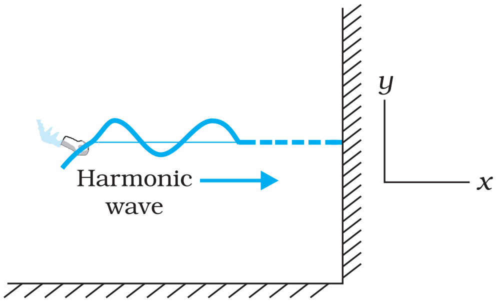


Fig. 14.3 *A harmonic (sinusoidal) wave travelling along a stretched string is an example of a transverse wave. An element of the string in the region of the wave oscillates about its equilibrium position perpendicular to the direction of wave propagation.*

position as the pulse or wave passes through them. The oscillations are normal to the direction of wave motion along the string, so this is an example of transverse wave.

We can look at a wave in two ways. We can fix an instant of time and picture the wave in space. This will give us the shape of the wave as a whole in space at a given instant. Another way is to fix a location i.e. fix our attention on a particular element of string and see its oscillatory motion in time.

Fig. 14.4 describes the situation for longitudinal waves in the most familiar example of the propagation of sound waves. A long pipe filled with air has a piston at one end. A single sudden push forward and pull back of the piston will generate a pulse of condensations (higher density) and rarefactions (lower density) in the medium (air). If the push-pull of the piston is continuous and periodic (sinusoidal), a

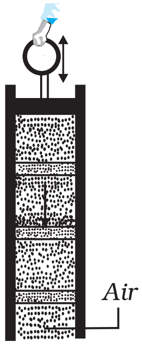


Fig. 14.4 *Longitudinal waves (sound) generated in a pipe filled with air by moving the piston up and down. A volume element of air oscillates in the direction parallel to the direction of wave propagation.*

sinusoidal wave will be generated propagating in air along the length of the pipe. This is clearly an example of longitudinal waves.

The waves considered above, transverse or longitudinal, are travelling or progressive waves since they travel from one part of the medium to another. The material medium as a whole does not move, as already noted. A stream, for example, constitutes motion of water as a whole. In a water wave, it is the disturbance that moves, not water as a whole. Likewise a wind (motion of air as a whole) should not be confused with a sound wave which is a propagation of disturbance (in pressure density) in air, without the motion of air medium as a whole.

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 not in fluids. Fluids, as well as, solids can sustain compressive strain; therefore, longitudinal waves can be propagated in all elastic media. For example, in medium like steel, 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—not more than a few centimetre—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 meters. 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 water waves involves a complicated motion—they not only move up and down but also back and forth. The waves in an ocean are the combination of both longitudinal and transverse waves.

It is found that, generally, transverse and longitudinal waves travel with different speed in the same medium.

Example 14.1 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

14.3 DISPLACEMENT RELATION IN A PROGRESSIVE WAVE

For mathematical description of a travelling wave, we need a function of both position x and time t. Such a function at every instant should give the shape of the wave at that instant. Also, at every given location, it should describe the motion of the constituent of the medium at that location. If we wish to describe a sinusoidal travelling wave (such as the one shown in Fig. 14.3) the corresponding function must also be sinusoidal. For convenience, we shall take the wave to be transverse so that if the position of the constituents of the medium is denoted by x, the displacement from the equilibrium position may be denoted by y. A sinusoidal travelling wave is then described by:

$$y(x,t) = a\sin(kx - \omega t + \phi) \qquad (14.2)$$

The term ϕ in the argument of sine function means equivalently that we are considering a linear combination of sine and cosine functions:

$$y(x,t) = A\sin(kx - \omega t) + B\cos(kx - \omega t) \qquad (14.3)$$

From Equations (14.2) and (14.3),

$$a = \sqrt{A^2 + B^2} \quad \text{and} \quad \phi = \tan^{-1}\left(\frac{B}{A}\right)$$

To understand why Equation (14.2) represents a sinusoidal travelling wave, take a fixed instant, say $t = t_0$. Then, the argument of the sine function in Equation (14.2) is simply

kx + constant. Thus, the shape of the wave (at any fixed instant) as a function of x is a sine wave. Similarly, take a fixed location, say $x = x_0$. Then, the argument of the sine function in Equation (14.2) is constant $-\omega t$. The displacement y, at a fixed location, thus, varies sinusoidally with time. That is, the constituents of the medium at different positions execute simple harmonic motion. Finally, as t increases, x must increase in the positive direction to keep $kx - \omega t + \phi$ constant. Thus, Eq. (14.2) represents a sinusiodal (harmonic) wave travelling along the positive direction of the x-axis. On the other hand, a function

$$y(x,t) = a \sin(kx + \omega t + \phi) \qquad (14.4)$$

represents a wave travelling in the negative direction of x-axis. Fig. (14.5) gives the names of the various physical quantities appearing in Eq. (14.2) that we now interpret.

$y(x,t)$	:	displacement as a function of position x and time t
a	:	amplitude of a wave
ω	:	angular frequency of the wave
k	:	angular wave number
$kx-\omega t+\phi$	:	initial phase angle ($a{+}x = 0,\ t = 0$)

Fig. 14.5 *The meaning of standard symbols in Eq. (14.2)*

Fig. 14.6 shows the plots of Eq. (14.2) for different values of time differing by equal intervals of time. In a wave, the crest is the point of maximum positive displacement, the trough is the point of maximum negative displacement. To see how a wave travels, we can fix attention on a crest and see how it progresses with time. In the figure, this is shown by a cross ( ) on the crest. In the same manner, we can see the motion of a particular constituent of the medium at a fixed location, say at the origin of the x-axis. This is shown by a solid dot (•). The plots of Fig. 14.6 show that with time, the solid dot (•) at the origin moves periodically, i.e., the particle at the origin oscillates about its mean position as the wave progresses. This is true for any other location also. We also see that during the time the solid dot (•) has completed one full oscillation, the crest has moved further by a certain distance.

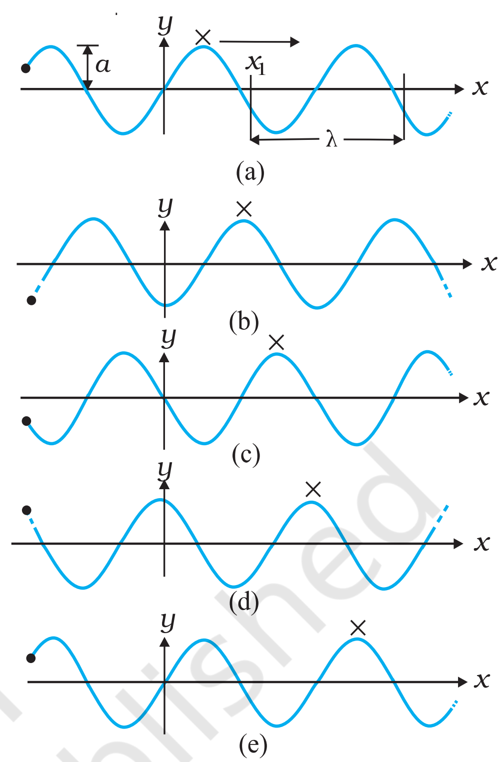


Fig. 14.6 *A harmonic wave progressing along the positive direction of x-axis at different times.*

Using the plots of Fig. 14.6, we now define the various quantities of Eq. (14.2).

14.3.1 Amplitude and Phase

In Eq. (14.2), since the sine function varies between 1 and –1, the displacement $y(x,t)$ varies between a and $-a$. We can take a to be a positive constant, without any loss of generality. Then, a represents the maximum displacement of the constituents of the medium from their equilibrium position. Note that the displacement y may be positive or negative, but a is positive. It is called the **amplitude** of the wave.

The quantity $(kx - \omega t + \phi)$ appearing as the argument of the sine function in Eq. (14.2) is called the phase of the wave. Given the amplitude a, the phase determines the displacement of the wave at any position and at any instant. Clearly ϕ is the phase at $x = 0$ and $t = 0$. Hence, ϕ is called the initial phase angle. By suitable choice of origin on the x-axis and the intial time, it is possible to have $\phi = 0$. Thus there is no loss of generality in dropping ϕ, i.e., in taking Eq. (14.2) with $\phi = 0$.

14.3.2 Wavelength and Angular Wave Number

The minimum distance between two points having the same phase is called the wavelength of the wave, usually denoted by λ. For simplicity, we can choose points of the same phase to be crests or troughs. The wavelength is then the distance between two consecutive crests or troughs in a wave. Taking $\phi = 0$ in Eq. (14.2), the displacement at $t = 0$ is given by

$$y(x,0) = a \sin kx \tag{14.5}$$

Since the sine function repeats its value after every 2π change in angle,

$$\sin kx = \sin(kx + 2n\pi) = \sin k\left(x + \frac{2n\pi}{k}\right)$$

That is the displacements at points x and at

$$x + \frac{2n\pi}{k}$$

are the same, where n=1,2,3,... The least distance between points with the same displacement (at any given instant of time) is obtained by taking $n = 1$. λ is then given by

$$\lambda = \frac{2\pi}{k} \quad \text{or} \quad k = \frac{2\pi}{\lambda} \tag{14.6}$$

k is the angular wave number or propagation constant; its SI unit is radian per metre or $rad\ m^{-1}$ *

14.3.3 Period, Angular Frequency and Frequency

Fig. 14.7 shows again a sinusoidal plot. It describes not the shape of the wave at a certain instant but the displacement of an element (at any fixed location) of the medium as a function of time. We may for, simplicity, take Eq. (14.2) with $\phi = 0$ and monitor the motion of the element say at $x = 0$. We then get

$$y(0,t) = a\sin(-\omega t)$$
$$= -a \sin \omega t$$

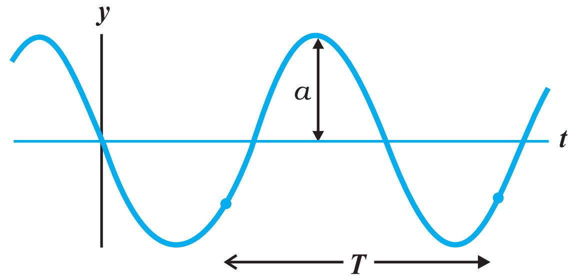


Fig. 14.7 *An element of a string at a fixed location oscillates in time with amplitude a and period T, as the wave passes over it.*

Now, the period of oscillation of the wave is the time it takes for an element to complete one full oscillation. That is

$$-a \sin \omega t = -a \sin \omega(t + T)$$
$$= -a \sin(\omega t + \omega T)$$

Since sine function repeats after every 2π,

$$\omega T = 2\pi \quad \text{or} \quad \omega = \frac{2\pi}{T} \tag{14.7}$$

ω is called the angular frequency of the wave. Its SI unit is rad s^{-1}. The frequency ν is the number of oscillations per second. Therefore,

$$\nu = \frac{1}{T} = \frac{\omega}{2\pi} \tag{14.8}$$

ν 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. (14.2), the displacement function for a longitudinal wave is written as,

$$s(x, t) = a \sin (kx - \omega t + \phi) \tag{14.9}$$

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. (14.9), 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 $s(x, t)$.

* *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}.*

Example 14.2 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. (14.2),

$$y(x, t) = a \sin (kx - \omega t),$$

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

$$k = 80.0 \text{ m}^{-1} \text{ and } \omega = 3.0 \text{ s}^{-1}$$

We, then, relate the wavelength λ to k through Eq. (14.6),

$$\lambda = 2\pi/k$$

$$= \frac{2\pi}{80.0 \text{ m}^{-1}}$$

$$= 7.85 \text{ cm}$$

(c) Now, we relate T to ω by the relation

$$T = 2\pi/\omega$$

$$= \frac{2\pi}{3.0 \text{ s}^{-1}}$$

$$= 2.09 \text{ s}$$

and frequency, $\nu = 1/T = 0.48$ Hz

The displacement y at $x = 30.0$ cm and time $t = 20$ s is given by

$$y = (0.005 \text{ m}) \sin (80.0 \times 0.3 - 3.0 \times 20)$$
$$= (0.005 \text{ m}) \sin (-36 + 12\pi)$$
$$= (0.005 \text{ m}) \sin (1.699)$$
$$= (0.005 \text{ m}) \sin (97^0) \simeq 5 \text{ mm}$$

14.4 THE SPEED OF A TRAVELLING WAVE

To determine the speed of propagation of a travelling wave, we can fix our attention on any particular point on the wave (characterised by some value of the phase) and see how that point moves in time. It is convenient to look at the motion of the crest of the wave. Fig. 14.8 gives the shape of the wave at two instants of time, which differ by a small time internal Δt. The entire wave pattern is seen to shift to the right (positive direction of x-axis) by a distance Δx. In particular, the crest shown by a dot (•) moves a

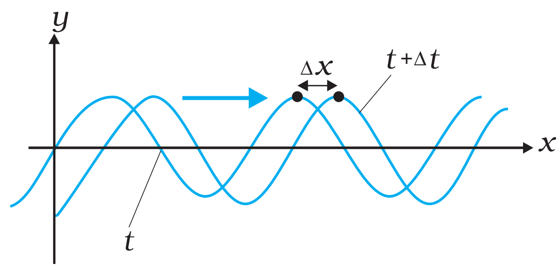


Fig. 14.8 *Progression of a harmonic wave from time t to $t + \Delta t$. where Δt is a small interval. The wave pattern as a whole shifts to the right. The crest of the wave (or a point with any fixed phase) moves right by the distance Δx in time Δt.*

distance Δx in time Δt. The speed of the wave is then $\Delta x/\Delta t$. We can put the dot (•) on a point with any other phase. It will move with the same speed v (otherwise the wave pattern will not remain fixed). The motion of a fixed phase point on the wave is given by

$$kx - \omega t = \text{constant} \quad (14.10)$$

Thus, as time t changes, the position x of the fixed phase point must change so that the phase remains constant. Thus,

$$kx - \omega t = k(x+\Delta x) - \omega(t+\Delta t)$$

or $k\,\Delta x - \omega\,\Delta t = 0$

Taking $\Delta x, \Delta t$ vanishingly small, this gives

$$\frac{dx}{dt} = \frac{\omega}{k} = v \quad (14.11)$$

Relating ω to T and k to λ, we get

$$v = \frac{2\pi\nu}{2\pi/\lambda} = \lambda\nu = \frac{\lambda}{T} \quad (14.12)$$

Eq. (14.12), a general relation for all progressive waves, shows that in the time required for one full oscillation by any constituent of the medium, the wave pattern travels a distance equal to the wavelength of the wave. It should be noted that the speed of a mechanical wave is determined by the inertial (linear mass density for strings, mass density in general) and elastic properties (Young's modulus for linear media/ shear modulus, bulk modulus) of the medium. The medium determines the speed; Eq. (14.12) then relates wavelength to frequency for the given speed. Of course, as remarked earlier, the medium can support both transverse and longitudinal waves, which will have different speeds in the same medium. Later in this chapter, we shall obtain specific expressions for the speed of mechanical waves in some media.

14.4.1 Speed of a Transverse Wave on Stretched String

The speed of a mechanical wave is determined by the restoring force setup in the medium when it is disturbed and the inertial properties (mass density) of the medium. The speed is expected to be directly related to the former and inversely to the latter. For waves on a string, the restoring force is provided by the tension T in the string. The inertial property will in this case be linear mass density μ, which is mass m of the string divided by its length L. Using Newton's Laws of Motion, an exact formula for the wave speed on a string can be derived, but this derivation is outside the scope of this book. We shall, therefore, use dimensional analysis. We already know that dimensional analysis alone can never yield the exact formula. The overall dimensionless constant is always left undetermined by dimensional analysis.

The dimension of μ is $[ML^{-1}]$ and that of T is like force, namely $[MLT^{-2}]$. We need to combine these dimensions to get the dimension of speed v $[LT^{-1}]$. Simple inspection shows that the quantity T/μ has the relevant dimension

$$\frac{[MLT^{-2}]}{[ML]} = [L^2T^{-2}]$$

Thus if T and μ are assumed to be the only relevant physical quantities,

$$v = C\sqrt{\frac{T}{\mu}} \qquad (14.13)$$

where C is the undetermined constant of dimensional analysis. In the exact formula, it turms out, C=1. The speed of transverse waves on a stretched string is given by

$$v = \sqrt{\frac{T}{\mu}} \qquad (14.14)$$

Note the important point that the speed v depends only on the properties of the medium T and μ (T is a property of the stretched string arising due to an external force). It does not depend on wavelength or frequency of the wave itself. In higher studies, you will come across waves whose speed is not independent of frequency of the wave. Of the two parameters λ and ν the source of disturbance determines the frequency of the wave generated. Given the speed of the wave in the medium and the frequency Eq. (14.12) then fixes the wavelength

$$\lambda = \frac{v}{\nu} \qquad (14.15)$$

Example 14.3 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,

$$\mu = \frac{5.0 \times 10^{-3}\ \text{kg}}{0.72\ \text{m}}$$

$$= 6.9 \times 10^{-3}\ \text{kg m}^{-1}$$

Tension, T = 60 N

The speed of wave on the wire is given by

$$v = \sqrt{\frac{T}{\mu}} = \sqrt{\frac{60\ \text{N}}{6.9 \times 10^{-3}\ \text{kg m}^{-1}}} = 93\ \text{m s}^{-1}$$

14.4.2 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 volume elements of air. The elastic property that determines the stress under compressional strain is the bulk modulus of the medium defined by (see Chapter 8)

$$B = -\frac{\Delta P}{\Delta V/V} \qquad (14.16)$$

Here, the change in pressure ΔP produces a volumetric strain $\frac{\Delta V}{V}$. B has the same dimension as pressure and given in SI units in terms of pascal *(Pa)*. The inertial property relevant for the propagation of wave is the mass density ρ, with dimensions $[ML^{-3}]$. Simple inspection reveals that quantity B/ρ has the relevant dimension:

$$\frac{\left[ML^{-2}T^{-2}\right]}{\left[ML^{-3}\right]} = \left[L^2T^{-2}\right] \quad (14.17)$$

Thus, if B and ρ are considered to be the only relevant physical quantities,

$$v = C\sqrt{\frac{B}{\rho}} \quad (14.18)$$

where, as before, C is the undetermined constant from dimensional analysis. The exact derivation shows that C=1. Thus, the general formula for longitudinal waves in a medium is:

$$v = \sqrt{\frac{B}{\rho}} \quad (14.19)$$

For a linear medium, like a solid bar, the lateral expansion of the bar is negligible and we may consider it to be only under longitudinal strain. In that case, the relevant modulus of elasticity is Young's modulus, which has the same dimension as the Bulk modulus. Dimensional analysis for this case is the same as before and yields a relation like Eq. (14.18), with an undetermined C, which the exact derivation shows to be unity. Thus, the speed of longitudinal waves in a solid bar is given by

$$v = \sqrt{\frac{Y}{\rho}} \quad (14.20)$$

where Y is the Young's modulus of the material of the bar. Table 14.1 gives the speed of sound in some media.

Table 14.1 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

Liquids and solids generally have higher speed of sound than gases. [Note for solids, the speed being referred to is the speed of longitudinal waves in the solid]. This happens because they are much more difficult to compress than gases and so have much higher values of bulk modulus. Now, see Eq. (14.19). Solids and liquids have higher mass densities (ρ) than gases. But the corresponding increase in both the modulus (B) of solids and liquids is much higher. This is the reason why the sound waves travel faster in solids and liquids.

We can estimate the speed of sound in a gas in the ideal gas approximation. For an ideal gas, the pressure P, volume V and temperature T are related by (see Chapter 10).

$$PV = Nk_BT \quad (14.21)$$

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.(14.21) that

$$V\Delta P + P\Delta V = 0$$

or $$-\frac{\Delta P}{\Delta V/V} = P$$

Hence, substituting in Eq. (14.16), we have

$$B = P$$

Therefore, from Eq. (14.19) the speed of a longitudinal wave in an ideal gas is given by,

$$v = \sqrt{\frac{P}{\rho}} \quad (14.22)$$

This relation was first given by Newton and is known as Newton's formula.

Example 14.4 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)

$$= \frac{29.0 \times 10^{-3}\ \text{kg}}{22.4 \times 10^{-3}\ \text{m}^3}$$

$$= 1.29\ \text{kg m}^{-3}$$

According to Newton's formula for the speed of sound in a medium, we get for the speed of sound in air at STP,

$$v = \left[\frac{1.01 \times 10^5 \text{ N m}^{-2}}{1.29 \text{ kg m}^{-3}}\right]^{1/2} = 280 \text{ m s}^{-1} \quad (14.23)$$

The result shown in Eq.(14.23) is about 15% smaller as compared to the experimental value of 331 m s^{-1} as given in Table 14.1. 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 (see Section 11.8),

$$PV^{\gamma} = \text{constant}$$

i.e. $\Delta(PV^{\gamma}) = 0$

or $P\gamma\, V^{\gamma - 1}\, \Delta V + V^{\gamma}\, \Delta P = 0$

where γ is the ratio of two specific heats, C_p/C_v.

Thus, for an ideal gas the adiabatic bulk modulus is given by,

$$B_{ad} = -\frac{\Delta P}{\Delta V/V}$$
$$= \gamma P$$

The speed of sound is, therefore, from Eq. (14.19), given by,

$$v = \sqrt{\frac{\gamma P}{\rho}} \quad (14.24)$$

This modification of Newton's formula is referred to as the **Laplace correction**. For air $\gamma = 7/5$. Now using Eq. (14.24) 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.

14.5 THE PRINCIPLE OF SUPERPOSITION OF WAVES

What happens when two wave pulses travelling in opposite directions cross each other (Fig. 14.9)? It turns out that wave pulses continue to retain their identities after they have crossed. However, during the time they overlap, the wave pattern is different from either of the

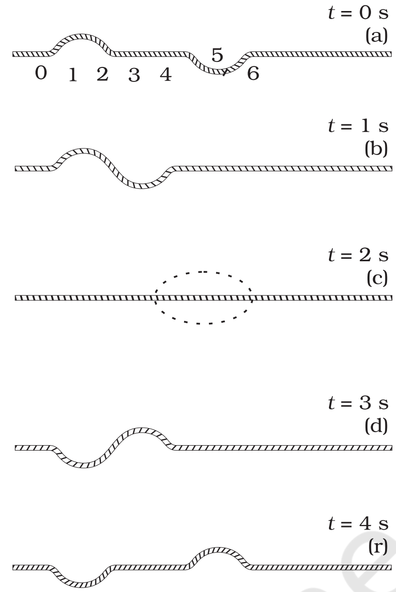


***Fig. 14.9** Two pulses having equal and opposite displacements moving in opposite directions. The overlapping pulses add up to zero displacement in curve (c).*

pulses. Figure 14.9 shows the situation when two pulses of equal and opposite shapes move towards each other. When the pulses overlap, the resultant displacement is the algebraic sum of the displacement due to each pulse. This is known as the principle of superposition of waves. According to this principle, each pulse moves as if others are not present. The constituents of the medium, therefore, suffer displacments due to both and since the displacements can be positive and negative, the net displacement is an algebraic sum of the two. Fig. 14.9 gives graphs of the wave shape at different times. Note the dramatic effect in the graph (c); the displacements due to the two pulses have exactly cancelled each other and there is zero displacement throughout.

To put the principle of superposition mathematically, let $y_1(x,t)$ and $y_2(x,t)$ be the displacements due to two wave disturbances in the medium. If the waves arrive in a region simultaneously, and therefore, overlap, the net displacement $y(x,t)$ is given by

$$y(x, t) = y_1(x, t) + y_2(x, t) \quad (14.25)$$

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

$$y = f_1(x - vt) + f_2(x - vt) + \ \dots + f_n(x - vt)$$

$$= \sum_{i=1}^{n} f_i\,(x - vt) \qquad (14.26)$$

The principle of superposition is basic to the phenomenon of interference.

For simplicity, consider two harmonic travelling waves on a stretched string, both with the same ω (angular frequency) and k (wave number), and, therefore, the same wavelength λ. Their wave speed will be identical. Let us further assume that their amplitudes are equal and they are both travelling in the positive direction of x-axis. The waves only differ in their initial phase. According to Eq. (14.2), the two waves are described by the functions:

$$y_1(x,\ t) = a \sin\,(kx - \omega t) \qquad (14.27)$$

and $y_2(x,\ t) = \ a \sin\,(kx - \omega t + \phi)$ (14.28)

The net displacement is then, by the principle of superposition, given by

$$y\ (x,\ t) = a \sin\,(kx - \omega t) + a \sin\,(kx - \omega t + \phi) \qquad (14.29)$$

$$= a\left[2\sin\left[\frac{(kx-\omega t)+(kx-\omega t+\phi)}{2}\right]\cos\frac{\phi}{2}\right] \qquad (14.30)$$

where we have used the familiar trignometric identity for $\sin A + \sin B$. We then have

$$y(x,t) = 2a\ \cos\frac{\phi}{2}\ \ \sin\left(kx-\omega t+\frac{\phi}{2}\right) \qquad (14.31)$$

Eq. (14.31) is also a harmonic travelling wave in the positive direction of x-axis, with the same frequency and wavelength. However, its initial phase angle is $\dfrac{\phi}{2}$. The significant thing is that its amplitude is a function of the phase difference ϕ between the constituent two waves:

$$A(\phi)\ =\ 2a \cos \tfrac{1}{2}\phi \qquad (14.32)$$

For $\phi = 0$, when the waves are in phase,

$$y(x,t) = 2a\ \sin(kx-\omega t) \qquad (14.33)$$

i.e., the resultant wave has amplitude 2a, the largest possible value for A. For $\phi = \pi$, the waves are completely, out of phase and the resultant wave has zero displacement everywhere at all times

$$y\ \ (x,\ t)\ \ = 0 \qquad (14.34)$$

Eq. (14.33) refers to the so-called constructive interference of the two waves where the amplitudes add up in the resultant wave. Eq. (14.34) is the case of destructive intereference where the amplitudes subtract out in the resultant wave. Fig. 14.10 shows these two cases of interference of waves arising from the principle of superposition.

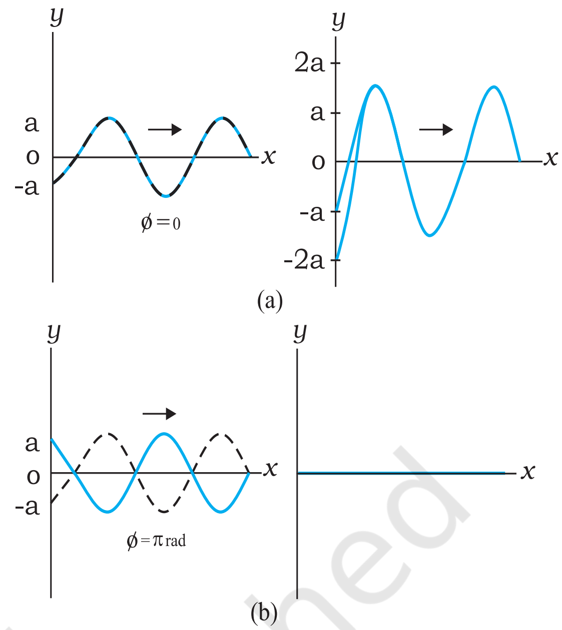


Fig. 14.10 *The resultant of two harmonic waves of equal amplitude and wavelength according to the principle of superposition. The amplitude of the resultant wave depends on the phase difference ϕ, which is zero for (a) and π for (b)*

14.6 REFLECTION OF WAVES

So far we considered waves propagating in an unbounded medium. What happens if a pulse or a wave meets a boundary? If the boundary is rigid, the pulse or wave gets reflected. The phenomenon of echo is an example of reflection by a rigid boundary. If the boundary is not completely rigid or is an interface between two different elastic media, the situation is some what complicated. A part of the incident 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.

Fig. 14.11 shows a pulse travelling along a stretched string and being reflected by the boundary. Assuming there is no absorption of energy by the boundary, the reflected wave has the same shape as the incident pulse but it suffers a phase change of π or 180^0 on reflection. This is because the boundary is rigid and the disturbance must have zero displacement at all times at the boundary. By the principle of superposition, this is possible only if the reflected and incident waves differ by a phase of π, so that the resultant displacement is zero. This reasoning is based on boundary condition on a rigid wall. We can arrive at the same conclusion dynamically also. As the pulse arrives at the wall, it exerts a force on the wall. By Newton's Third Law, the wall exerts an equal and opposite force on the string generating a reflected pulse that differs by a phase of π.

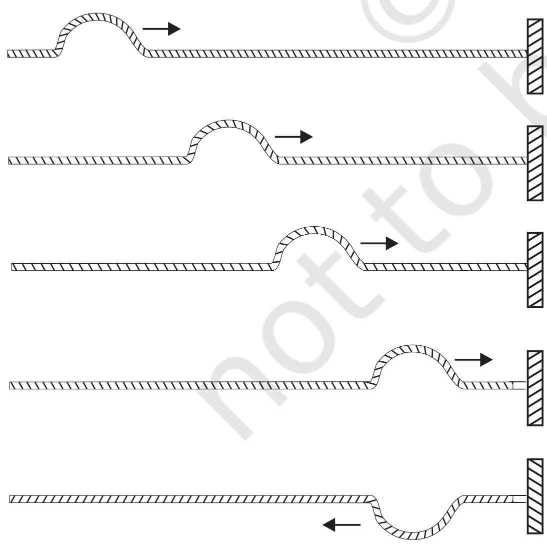


Fig. 14.11 Reflection of a pulse meeting a rigid boundary.

If on the other hand, the boundary point is not rigid but completely free to move (such as in the case of a string tied to a freely moving ring on a rod), the reflected pulse has the same phase and amplitude (assuming no energy dissipation) as the incident pulse. The net maximum displacement at the boundary is then twice the amplitude of each pulse. An example of non- rigid boundary is the open end of an organ pipe.

To summarise, a travelling wave or pulse suffers a phase change of π on reflection at a rigid boundary and no phase change on reflection at an open boundary. To put this mathematically, let the incident travelling wave be

$$y_2(x,t) = a\sin(kx - \omega t)$$

At a rigid boundary, the reflected wave is given by

$$y_r(x, t) = a \sin (kx - \omega t + \pi).$$
$$= -a \sin (kx - \omega t) \qquad (14.35)$$

At an open boundary, the reflected wave is given by

$$y_r(x, t) = a \sin (kx - \omega t + 0).$$
$$= a \sin (kx - \omega t) \qquad (14.36)$$

Clearly, at the rigid boundary, $y = y_2 + y_r = 0$ at all times.

14.6.1 Standing Waves and Normal Modes

We considered above reflection at one boundary. But there are familiar situations (a string fixed at either end or an air column in a pipe with either end closed) in which reflection takes place at two or more boundaries. In a string, for example, a wave travelling in one direction will get reflected at one end, which in turn will travel and get reflected from the other end. This will go on until there is a steady wave pattern set up on the string. Such wave patterns are called standing waves or stationary waves. To see this mathematically, consider a wave travelling along the positive direction of x-axis and a reflected wave of the same amplitude and wavelength in the negative direction of x-axis. From Eqs. (14.2) and (14.4), with $\phi = 0$, we get:

$$y_1(x, t) = a \sin (kx - \omega t)$$

$$y_2(x, t) = a \sin (kx + \omega t)$$

The resultant wave on the string is, according to the principle of superposition:

$$y(x, t) = y_1(x, t) + y_2(x, t)$$

$$= a\,[\sin(kx - \omega t) + \sin(kx + \omega t)]$$

Using the familiar trignometric identity Sin (A+B) + Sin (A–B) = 2 sin A cosB we get,

$$y(x,\ t) = 2a \sin kx \cos \omega t \qquad (14.37)$$

Note the important difference in the wave pattern described by Eq. (14.37) from that described by Eq. (14.2) or Eq. (14.4). The terms kx and ωt appear separately, not in the combination $kx - \omega t$. The amplitude of this wave is $2a \sin kx$. Thus, in this wave pattern, the amplitude varies from point-to-point, but each element of the string oscillates with the same angular frequency ω or time period. There is no phase difference between oscillations of different elements of the wave. The string as a whole vibrates in phase with differing amplitudes at different points. The wave pattern is neither moving to the right nor to the left. Hence, they are called standing or stationary waves. The amplitude is fixed at a given location but, as remarked earlier, it is different at different locations. The points at which the amplitude is zero (i.e., where there is no motion at all) are **nodes**; the points at which the amplitude is the largest are called **antinodes.** Fig. 14.12 shows a stationary wave pattern resulting from superposition of two travelling waves in opposite directions.

The most significant feature of stationary waves is that the boundary conditions constrain the possible wavelengths or frequencies of vibration of the system. The system cannot oscillate with any arbitrary frequency (contrast this with a harmonic travelling wave), but is characterised by a set of natural frequencies or **normal modes** of oscillation. Let us determine these normal modes for a stretched string fixed at both ends.

First, from Eq. (14.37), the positions of nodes (where the amplitude is zero) are given by

$\sin kx = 0$.

which implies

$$kx = n\pi; \quad n = 0, 1, 2, 3, \ldots$$

Since, $k = 2\pi/\lambda$, we get

$$x = \frac{n\lambda}{2}\ ; n = 0, 1, 2, 3, \ldots \qquad (14.38)$$

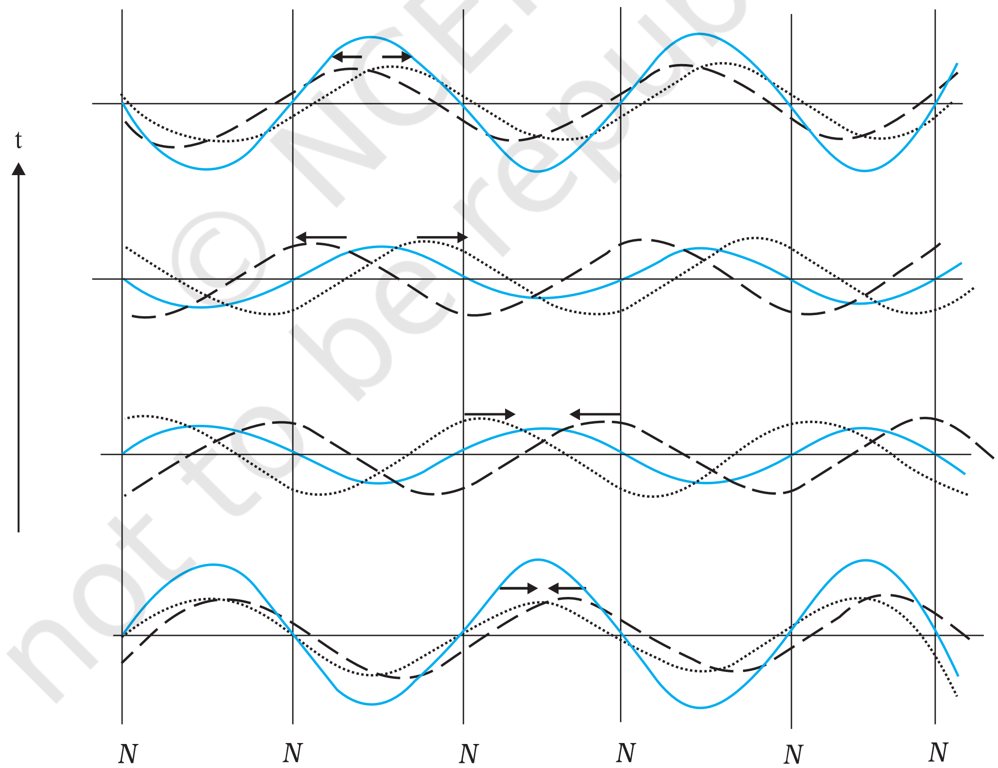


Fig. 14.12 *Stationary waves arising from superposition of two harmonic waves travelling in opposite directions. Note that the positions of zero displacement (nodes) remain fixed at all times.*

Clearly, the distance between any two successive nodes is $\frac{\lambda}{2}$. In the same way, the positions of antinodes (where the amplitude is the largest) are given by the largest value of sin kx:

$|\sin kx| = 1$

which implies

$kx = (n + ½)\ \pi\ ;\ n = 0,\ 1,\ 2,\ 3,\ ...$

With $k = 2\pi/\lambda$, we get

$$x = (n + ½)\frac{\lambda}{2}\ ;\ n = 0,\ 1,\ 2,\ 3,\ ... \tag{14.39}$$

Again the distance between any two consecutive antinodes is $\frac{\lambda}{2}$. Eq. (14.38) can be applied to the case of a stretched string of length L fixed at both ends. Taking one end to be at $x = 0$, the boundary conditions are that $x = 0$ and $x = L$ are positions of nodes. The $x = 0$ condition is already satisfied. The $x = L$ node condition requires that the length L is related to λ by

$$L = n\frac{\lambda}{2};\quad n = 1, 2, 3, ... \tag{14.40}$$

Thus, the possible wavelengths of stationary waves are constrained by the relation

$$\lambda = \frac{2L}{n};\quad n = 1, 2, 3, ... \tag{14.41}$$

with corresponding frequencies

$$v = \frac{nv}{2L},\ \text{for}\ n = 1, 2, 3, \tag{14.42}$$

We have thus obtained the natural frequencies - the normal modes of oscillation of the system. The lowest possible natural frequency of a system is called its **fundamental mode** or the **first harmonic**. For the stretched string fixed at either end it is given by $v = \frac{v}{2L}$, corresponding to $n = 1$ of Eq. (14.42). Here v is the speed of wave determined by the properties of the medium. The $n = 2$ frequency is called the second harmonic; $n = 3$ is the third harmonic and so on. We can label the various harmonics by the symbol v_n ($n = 1, 2, ...$).

Fig. 14.13 shows the first six harmonics of a stretched string fixed at either end. A string need not vibrate in one of these modes only. Generally, the vibration of a string will be a superposition of different modes; some modes may be more strongly excited and some less. Musical instruments like sitar or violin are based on this principle. Where the string is plucked or bowed, determines which modes are more prominent than others.

Let us next consider normal modes of oscillation of an air column with one end closed

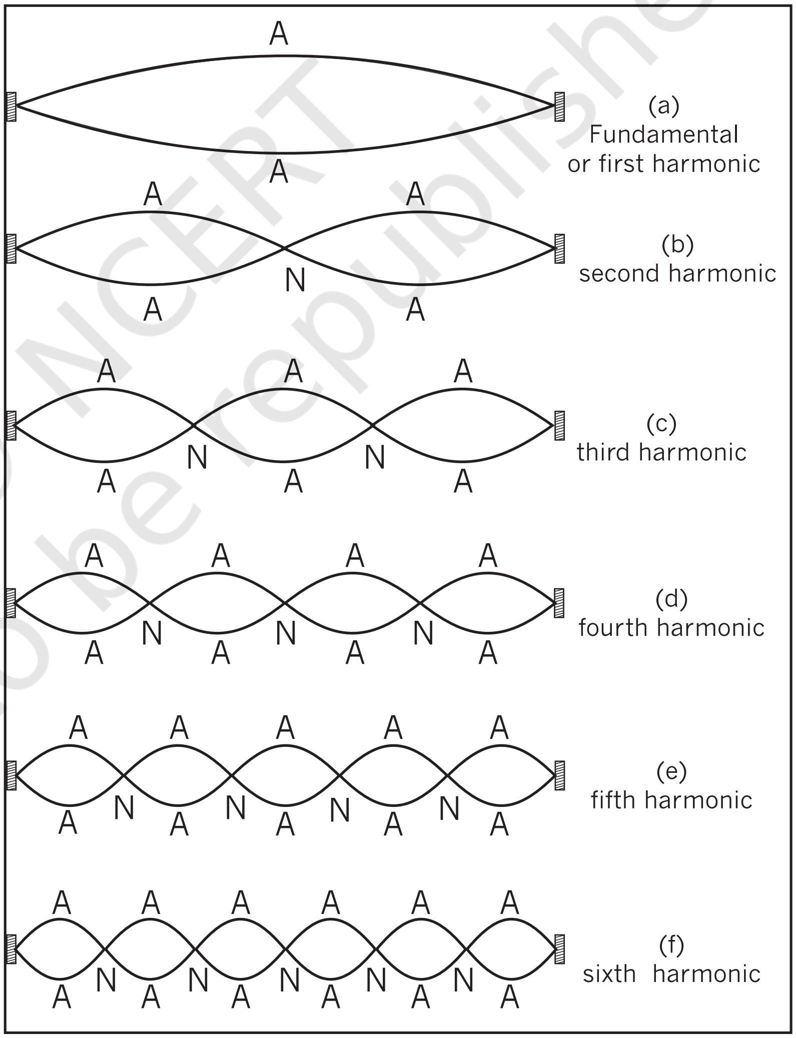


Fig. 14.13 *The first six harmonics of vibrations of a stretched string fixed at both ends.*

and the other open. A glass tube partially filled with water illustrates this system. The end in contact with water is a node, while the open end is an antinode. At the node the pressure changes are the largest, while the displacement is minimum (zero). At the open end - the antinode, it is just the other way - least pressure change and maximum amplitude of displacement. Taking the end in contact with water to be $x = 0$, the node condition (Eq. 14.38) is already satisfied. If the other end $x = L$ is an antinode, Eq. (14.39) gives

$$L = \left(n + \frac{1}{2}\right)\frac{\lambda}{2}, \text{ for } n = 0, 1, 2, 3, \ldots$$

The possible wavelengths are then restricted by the relation :

$$\lambda = \frac{2L}{(n + 1/2)}, \text{ for } n = 0, 1, 2, 3, \ldots \quad (14.43)$$

The normal modes – the natural frequencies – of the system are

$$\nu = \left(n + \frac{1}{2}\right)\frac{v}{2L}; \ n = 0, 1, 2, 3, \ldots \quad (14.44)$$

The fundamental frequency corresponds to $n = 0$, and is given by $\frac{v}{4L}$. The higher frequencies are **odd harmonics**, i.e., odd multiples of the fundamental frequency : $3\frac{v}{4L}$, $5\frac{v}{4L}$, etc. Fig. 14.14 shows the first six odd harmonics of air column with one end closed and the other open. For a pipe open at both ends, each end is an antinode. It is then easily seen that an open air column at both ends generates all harmonics (See Fig. 14.15).

The systems above, strings and air columns, can also undergo forced oscillations (Chapter 13). If the external frequency is close to one of the natural frequencies, the system shows **resonance**.

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.

Example 14.5 A pipe, 30.0 cm long, is open at both ends. Which harmonic mode of the pipe resonates 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

$$\nu_1 = \frac{v}{\lambda_1} = \frac{v}{2L} \quad \text{(open pipe)}$$

where L is the length of the pipe. The frequency of its nth harmonic is:

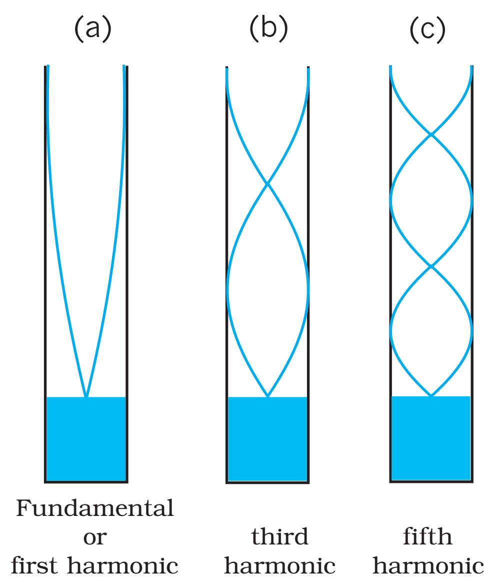
$$\nu_n = \frac{nv}{2L}, \text{ for } n = 1, 2, 3, \ldots \text{ (open pipe)}$$

First few modes of an open pipe are shown in Fig. 14.15.

For $L = 30.0$ cm, $v = 330$ m s^{-1},

$$\nu_n = \frac{n\ 330\ (\text{m s}^{-1})}{0.6\ (\text{m})} = 550 \text{ n s}^{-1}$$

Clearly, a source of frequency 1.1 kHz will resonate at ν_2, i.e. the **second harmonic**.



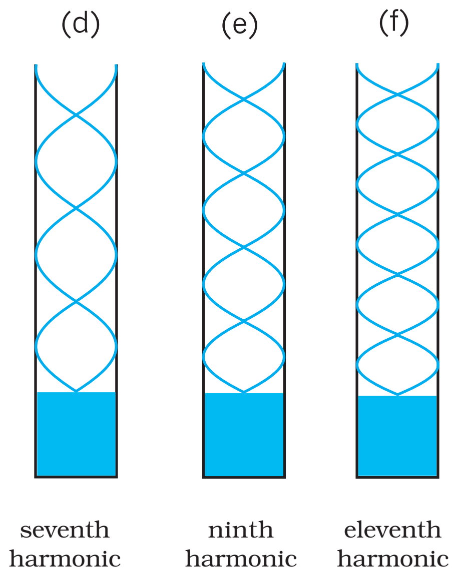


Fig. 14.14 *Normal modes of an air column open at one end and closed at the other end. Only the odd harmonics are seen to be possible.*

Now if one end of the pipe is closed (Fig. 14.15), it follows from Eq. (14.15) 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. ◀

14.7 BEATS

'Beats' is an interesting phenomenon arising from interference of waves. When two harmonic sound waves of close (but not equal) frequencies are heard at the same time, we hear a sound of similar frequency (the average of two close frequencies), but we hear something else also. We hear audibly distinct waxing and waning of the intensity of the sound, with a frequency equal to the difference in the two close frequencies. Artists use this phenomenon often

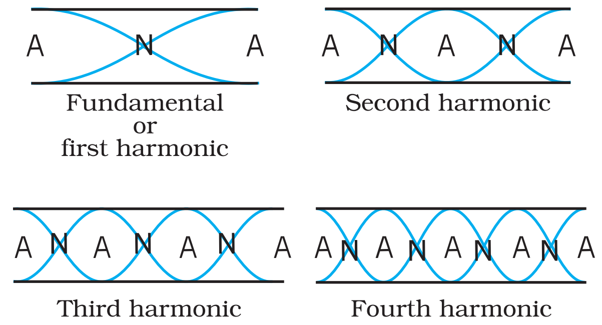


Fig. 14.15 *Standing waves in an open pipe, first four harmonics are depicted.*

while tuning their instruments with each other. They go on tuning until their sensitive ears do not detect any beats.

To see this mathematically, let us consider two harmonic sound waves of nearly equal angular frequency ω_1 and ω_2 and fix the location to be x = 0 for convenience. Eq. (14.2) with a suitable choice of phase ($\phi = \pi/2$ for each) and, assuming equal amplitudes, gives

$$s_1 = a \cos \omega_1 t \quad \text{and} \quad s_2 = a \cos \omega_2 t \qquad (14.45)$$

Here we have replaced the symbol y by s, since we are referring to longitudinal not transverse displacement. Let ω_1 be the (slightly) greater of the two frequencies. The resultant displacement is, by the principle of superposition,

$$s = s_1 + s_2 = a\,(\cos \omega_1 t + \cos \omega_2 t)$$

Using the familiar trignometric identity for cos A + cosB, we get

$$= 2\,a \cos \frac{(\omega_1 - \omega_2)t}{2} \cos \frac{(\omega_1 + \omega_2)t}{2} \qquad (14.46)$$

which may be written as :

$$s = [2\,a \cos \omega_b t] \cos \omega_a t \qquad (14.47)$$

If $|\omega_1 - \omega_2| << \omega_1, \omega_2, \omega_a >> \omega_b$, th where

$$\omega_b = \frac{(\omega_1 - \omega_2)}{2} \text{ and } \omega_a = \frac{(\omega_1 + \omega_2)}{2}$$

Now if we assume $|\omega_1 - \omega_2| << \omega_1$, which means $\omega_a >> \omega_b$, we can interpret Eq. (14.47) as follows. The resultant wave is oscillating with the average angular frequency ω_a; however its amplitude is **not** constant in time, unlike a pure harmonic wave. The amplitude is the largest when the term *cos* $\omega_b t$ takes its limit +1 or –1. In other words, the intensity of the resultant wave waxes and wanes with a frequency which is $2\omega_b = \omega_1 -$

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.

ω_2. Since $\omega = 2\pi\nu$, the beat frequency ν_{beat}, is given by

$$\nu_{beat} = \nu_1 - \nu_2 \qquad (14.48)$$

Fig. 14.16 illustrates the phenomenon of beats for two harmonic waves of frequencies 11 Hz and 9 Hz. The amplitude of the resultant wave shows beats at a frequency of 2 Hz.

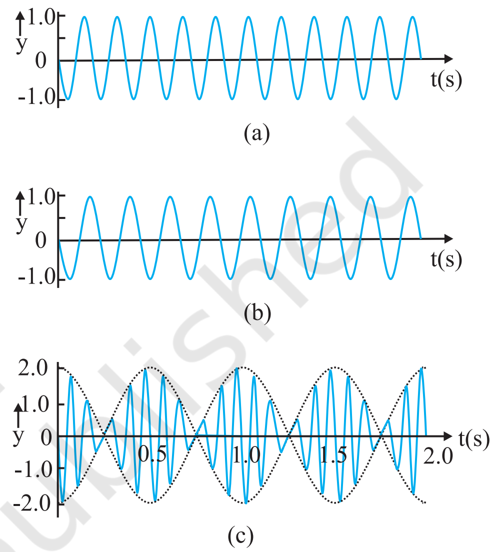


Fig. 14.16 *Superposition of two harmonic waves, one of frequency 11 Hz (a), and the other of frequency 9Hz (b), giving rise to beats of frequency 2 Hz, as shown in (c).*

▶ ***Example 14.6*** 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. ◀

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 antinodes.
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

   $$\nu = \frac{\omega}{2\pi}$$

9. *Speed* of a progressive wave is given by $v = \frac{\omega}{k} = \frac{\lambda}{T} = \lambda\nu$
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

    $$v = \sqrt{\frac{T}{\mu}}$$

11. *Sound waves* are longitudinal mechanical waves that can travel through solids, liquids, or gases. The speed v of sound wave in a fluid having *bulk modulus* B and density ρ is

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

    The speed of longitudinal waves in a metallic bar is

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

    For gases, since $B = \gamma P$, the speed of sound is

    $$v = \sqrt{\frac{\gamma P}{\rho}}$$

12. When two or more waves traverse simultaneously in 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 :

$$y(x, t) = \left[2a\cos\frac{1}{2}\phi\right]\sin\left(kx - \omega t + \frac{1}{2}\phi\right)$$

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

$$\nu = \frac{n\,v}{2\,L}, \qquad n = 1, 2, 3, \ldots$$

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

$$\nu = (n + ½)\frac{v}{2L}, \qquad n = 0, 1, 2, 3, \ldots$$

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 or open at both the ends, vibrates with certain frequencies called their 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

$$\nu_{beat} = \nu_1 \sim \nu_2$$

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	v	$[LT^{-1}]$	$m\ s^{-1}$	$v = \nu\lambda$
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. In a mechanical wave, 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 (v) depends only on elastic and other properties (such as mass density) of the medium. It does not depend on the velocity of the source.

EXERCISES

14.1 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?

14.2 A stone dropped from the top of a tower of height 300 m 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})

14.3 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}.

14.4 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.

14.5 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) $1/(x + vt)$

14.6 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}.

14.7 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.

14.8 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 ?

14.9 For the wave described in Exercise 14.8, 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 ?

14.10 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) $3\lambda/4$

14.11 The transverse displacement of a string (clamped at its both ends) is given by

$$y(x, t) = 0.06 \sin\left(\frac{2\pi}{3}x\right)\cos(120\ \pi t)$$

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.

14.12 (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?

14.13 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) $y = \cos x \sin t + \cos 2x \sin 2t$

14.14 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?

14.15 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.

14.16 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?

14.17 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}).

14.18 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?

14.19 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.

12089CH01

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. 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

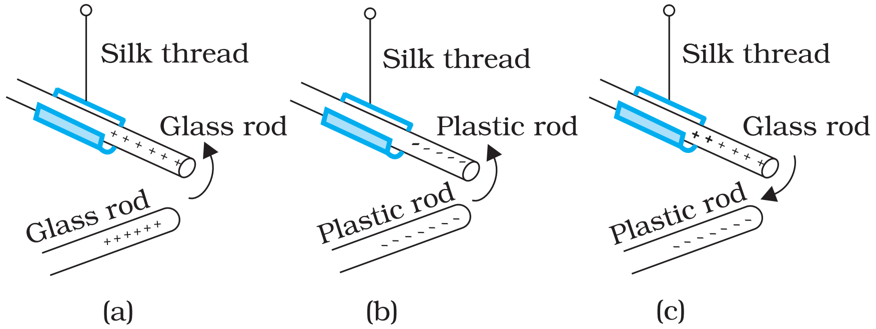


FIGURE 1.1 Rods: like charges repel and unlike charges attract each other.

elektron meaning *amber*. Many such pairs of materials were known which on rubbing could attract light objects like straw, pith balls and bits of papers.

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.

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 entry 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. There are two kinds of electrification and we find that (i) *like charges repel* and (ii) *unlike charges attract* each other. 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 neutralise or nullify each other's effect. Therefore, the charges were named as *positive* and *negative* by the American scientist Benjamin Franklin. 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 electrically neutral.

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.

Try to understand why material bodies acquire charge. 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.

1.3 Conductors and Insulators

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

* There is a third category called *semiconductors*, which offer resistance to the movement of charges which is intermediate between the conductors and insulators.

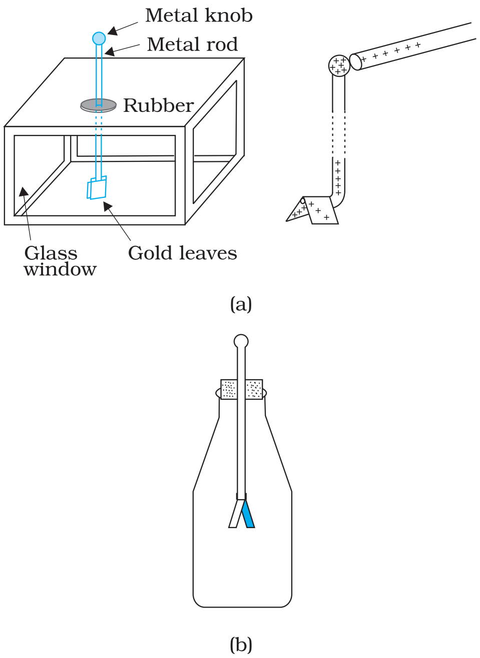


FIGURE 1.2 Electroscopes: (a) The gold leaf electroscope, (b) Schematics of a simple electroscope.

like spoon does not. The charges on metal leak through our body to the ground as both are conductors of electricity. However, if a metal rod with a wooden or plastic handle is rubbed without touching its metal part, it shows signs of charging.

1.4 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.

1.4.1 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, ..., q_n, then the total charge of the system is $q_1 + q_2 + q_3 + ... + q_n$. Charge has magnitude but no direction, similar to 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.

1.4.2 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.

1.4.3 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 1 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. In this respect, 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 magnituOde, 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. However, 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 magnitude of scale involved that is very important.

Example 1.1 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 = $6.25 \times 10^9 \div (365 \times 24 \times 3600)$ 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 1.1

Example 1.2 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 $(250/18) \times 6.02 \times 10^{23}$.

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 $(250/18) \times 6.02 \times 10^{23} \times 10 \times 1.6 \times 10^{-19}$ C = 1.34×10^7 C.

EXAMPLE 1.2

1.5 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

$$F = k\ \frac{|q_1 \ q_2|}{r^2} \tag{1.1}$$

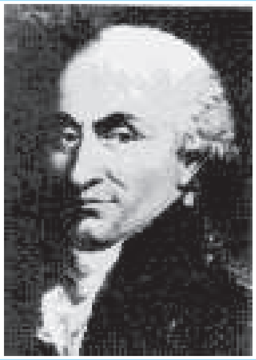
How did Coulomb arrive at this law from his experiments? Coulomb used a torsion balance* for measuring the force between two charged metallic spheres. When the separation between two spheres is much larger than the radius of each sphere, the charged spheres may be regarded as point charges. However, 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 $9 \times 10^9 \ \frac{\text{Nm}^2}{\text{C}^2}$. 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



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)

* 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.

* 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.

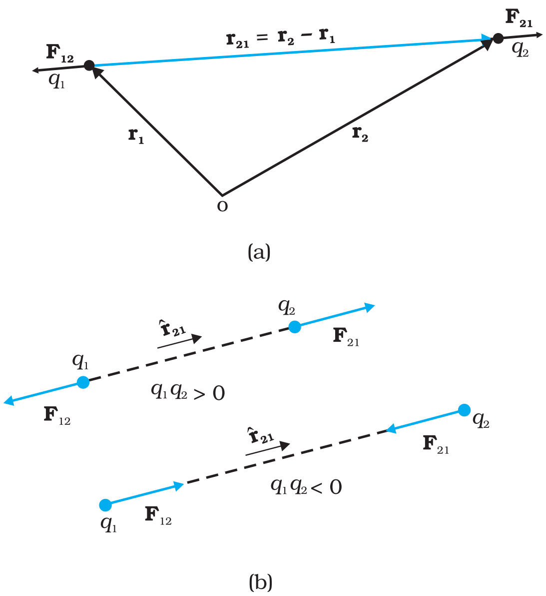


FIGURE 1.3 (a) Geometry and (b) Forces between charges.

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

$$F = \frac{1}{4\pi\varepsilon_0} \frac{|q_1 q_2|}{r^2} \tag{1.2}$$

ε_0 is called the *permittivity of free space* . The value of ε_0 in SI units is

$\varepsilon_0 = 8.854 \times 10^{-12}\ \mathrm{C^2\ N^{-1} m^{-2}}$

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. 1.3(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 $\mathbf{r}_2$, respectively 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. 1.3(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.

PHYSICS

Interactive animation on Coulomb's law:
http://webphysics.davidson.edu/physlet_resources/bu_semester2/menu_semester2.html

Example 1.3 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 and masses respectively. (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:

$$\left|\frac{F_e}{F_G}\right| = \frac{e^2}{4\pi\varepsilon_0 G m_p m_p} = 1.3\times10^{36}$$

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.

EXAMPLE 1.3

(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

$a = 2.3\times10^{-8}$ 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 1.3

Example 1.4 A charged metallic sphere A is suspended by a nylon thread. Another charged metallic sphere B held by an insulating

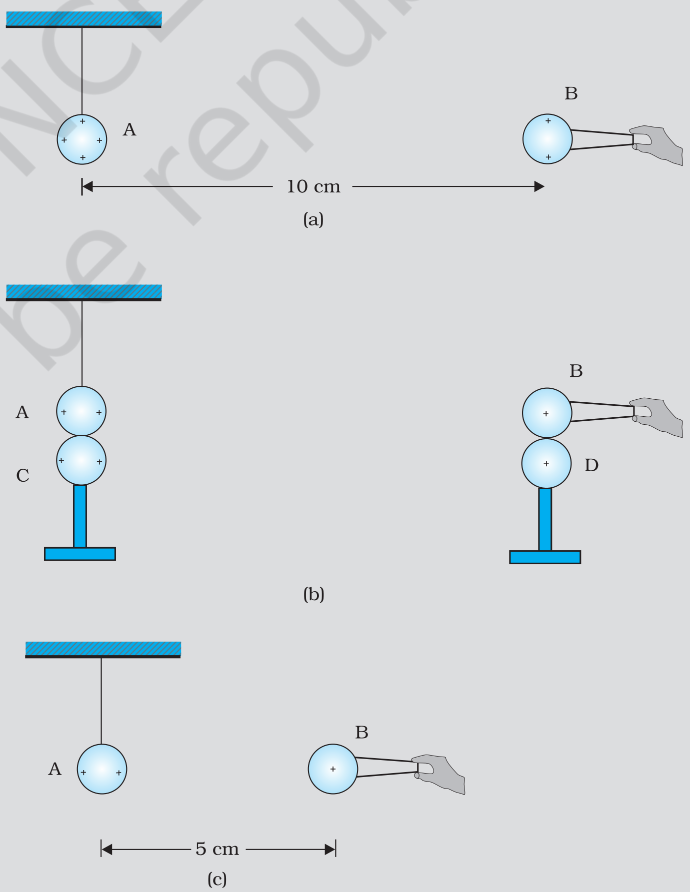


FIGURE 1.4

EXAMPLE 1.4

handle is brought close to A such that the distance between their centres is 10 cm, as shown in Fig. 1.4(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. 1.4(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. 1.4(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.

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 1.4

1.6 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, ..., q_n in vacuum. What is the force on q_1 due to q_2, q_3, ..., 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. 1.5(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.

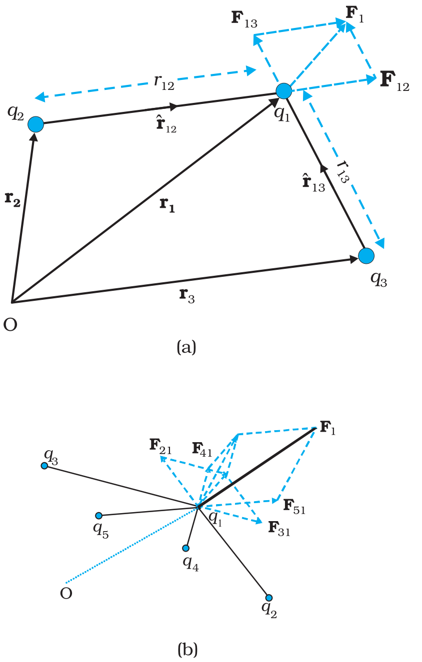


FIGURE 1.5 A system of (a) three charges (b) multiple charges.

Thus, $$\mathbf{F}_{12} = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r_{12}^2}\hat{\mathbf{r}}_{12}$$

In the same way, the force on q_1 due to q_3, denoted by $\mathbf{F}_{13}$, is given by

$$\mathbf{F}_{13} = \frac{1}{4\pi\varepsilon_0} \frac{q_1 q_3}{r_{13}^2} \hat{\mathbf{r}}_{13}$$

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

$$\mathbf{F}_1 = \mathbf{F}_{12} + \mathbf{F}_{13} = \frac{1}{4\pi\varepsilon_0} \frac{q_1 q_2}{r_{12}^2} \hat{\mathbf{r}}_{12} + \frac{1}{4\pi\varepsilon_0} \frac{q_1 q_3}{r_{13}^2} \hat{\mathbf{r}}_{13} \quad (1.4)$$

The above calculation of force can be generalised to a system of charges more than three, as shown in Fig. 1.5(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.,

$$\mathbf{F}_1 = \mathbf{F}_{12} + \mathbf{F}_{13} + ... + \mathbf{F}_{1n} = \frac{1}{4\pi\varepsilon_0}\left[\frac{q_1 q_2}{r_{12}^2}\hat{\mathbf{r}}_{12} + \frac{q_1 q_3}{r_{13}^2}\hat{\mathbf{r}}_{13} + ... + \frac{q_1 q_n}{r_{1n}^2}\hat{\mathbf{r}}_{1n}\right]$$

$$= \frac{q_1}{4\pi\varepsilon_0} \sum_{i=2}^{n} \frac{q_i}{r_{1i}^2} \hat{\mathbf{r}}_{1i} \quad (1.5)$$

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 1.5 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. 1.6?

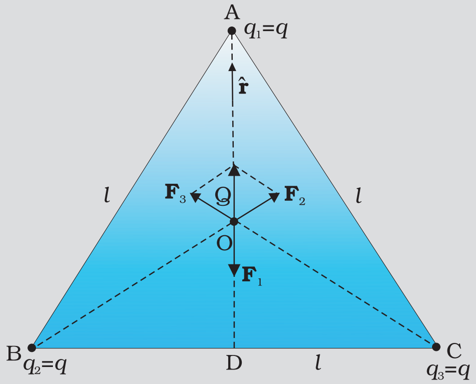


FIGURE 1.6

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 1.5

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 1.5

Example 1.6 Consider the charges q, q, and $-q$ placed at the vertices of an equilateral triangle, as shown in Fig. 1.7. What is the force on each charge?

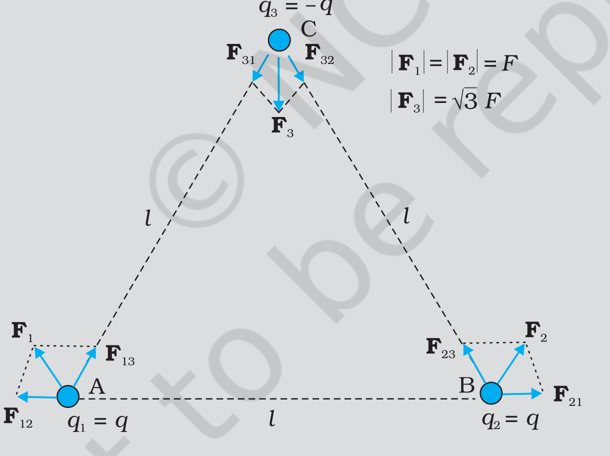


FIGURE 1.7

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. 1.7. 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 1.6

EXAMPLE 1.6

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 ∠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.

1.7 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

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0}\frac{Q}{r^2}\hat{\mathbf{r}} = \frac{1}{4\pi\varepsilon_0}\frac{Q}{r^2}\hat{\mathbf{r}} \qquad (1.6)$$

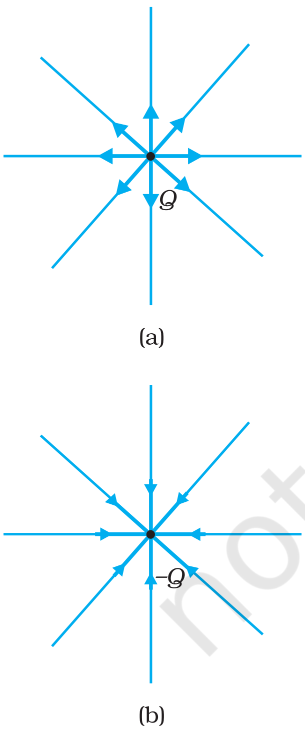


FIGURE 1.8 Electric field (a) due to a charge Q, (b) due to a charge $-Q$.

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

$$\mathbf{F} = \frac{1}{4\pi\varepsilon_0}\frac{Qq}{r^2}\hat{\mathbf{r}} \qquad (1.7)$$

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,

$$\mathbf{F}(\mathbf{r}) = q\,\mathbf{E}(\mathbf{r}) \qquad (1.8)$$

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*

* 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 1.14), 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.

1.7.1 Electric field 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 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, ..., q_n. We can use Coulomb's law and the superposition principle to determine this field at a point P denoted by position vector $\mathbf{r}$.

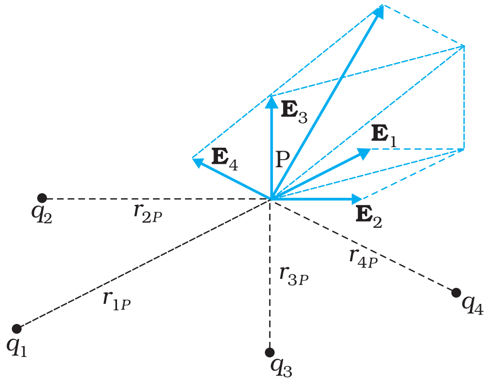


FIGURE 1.9 Electric field at a point due to a system of charges is the vector sum of the electric fields at the point due to individual charges.

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. 1.9)

$$\mathbf{E(r)} = \mathbf{E}_1(\mathbf{r}) + \mathbf{E}_2(\mathbf{r}) + \ldots + \mathbf{E}_n(\mathbf{r})$$

$$= \frac{1}{4\pi\varepsilon_0}\frac{q_1}{r_{1P}^2}\hat{\mathbf{r}}_{1P} + \frac{1}{4\pi\varepsilon_0}\frac{q_2}{r_{2P}^2}\hat{\mathbf{r}}_{2P} + \ldots + \frac{1}{4\pi\varepsilon_0}\frac{q_n}{r_{nP}^2}\hat{\mathbf{r}}_{nP}$$

$$\mathbf{E(r)} = \frac{1}{4\pi\varepsilon_0}\sum_{i=1}^{n}\frac{q_i}{r_{iP}^2}\hat{\mathbf{r}}_{iP} \qquad (1.10)$$

$\mathbf{E}$ is a vector quantity that varies from one point to another point in space and is determined from the positions of the source charges.

1.7.2 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 for a short interval of time and then 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 1.7 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. 1.10(a)]. The direction of the field is reversed keeping its magnitude unchanged and a proton falls through the same distance [Fig. 1.10(b)]. Compute the time of fall in each case. Contrast the situation with that of 'free fall under gravity'.

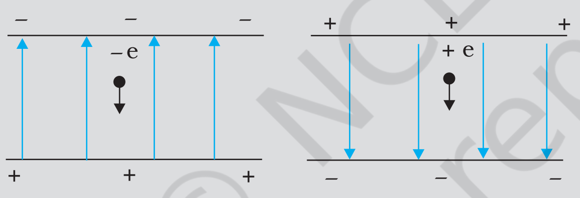


FIGURE 1.10

Solution In Fig. 1.10(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 $t_e = \sqrt{\dfrac{2h}{a_e}} = \sqrt{\dfrac{2hm_e}{eE}}$

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. 1.10 (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 1.7

$$t_p = \sqrt{\frac{2h}{a_p}} = \sqrt{\frac{2hm_p}{eE}} = 1.3 \times 10^{-7}\,\text{s}$$

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 1.7

Example 1.8 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. 1.11.

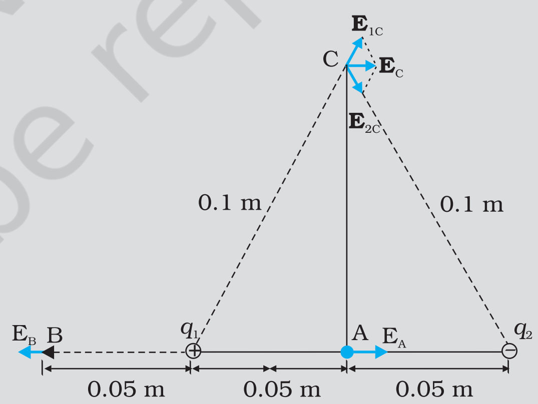


FIGURE 1.11

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

$$E_{1A} = \frac{(9 \times 10^{9}\,\text{Nm}^2\text{C}^{-2}) \times (10^{-8}\,\text{C})}{(0.05\,\text{m})^2} = 3.6 \times 10^{4}\ \text{N C}^{-1}$$

The electric field vector $\mathbf{E}_{2A}$ 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.

EXAMPLE 1.8

The electric field vector $\mathbf{E}_{1B}$ at B due to the positive charge q_1 points towards the left and has a magnitude

$$E_{1B} = \frac{(9\times10^9\,\mathrm{Nm^2C^{-2}})\times(10^{-8}\,\mathrm{C})}{(0.05\,\mathrm{m})^2} = 3.6 \times 10^4 \text{ N C}^{-1}$$

The electric field vector $\mathbf{E}_{2B}$ at B due to the negative charge q_2 points towards the right and has a magnitude

$$E_{2B} = \frac{(9\times10^9\,\mathrm{Nm^2C^{-2}})\times(10^{-8}\,\mathrm{C})}{(0.15\,\mathrm{m})^2} = 4 \times 10^3 \text{ N C}^{-1}$$

The magnitude of the total electric field at B is
$E_B = E_{1B} - E_{2B} = 3.2 \times 10^4$ 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

$$E_{1C} = E_{2C} = \frac{(9\times10^9\,\mathrm{Nm^2C^{-2}})\times(10^{-8}\,\mathrm{C})}{(0.10\,\mathrm{m})^2} = 9 \times 10^3 \text{ N C}^{-1}$$

The directions in which these two vectors point are indicated in Fig. 1.11. The resultant of these two vectors is

$$E_C = E_{1c}\cos\frac{\pi}{3} + E_{2c}\cos\frac{\pi}{3} = 9 \times 10^3 \text{ N C}^{-1}$$

$\mathbf{E}_C$ points towards the right.

EXAMPLE 1.8

1.8 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 1.12 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.

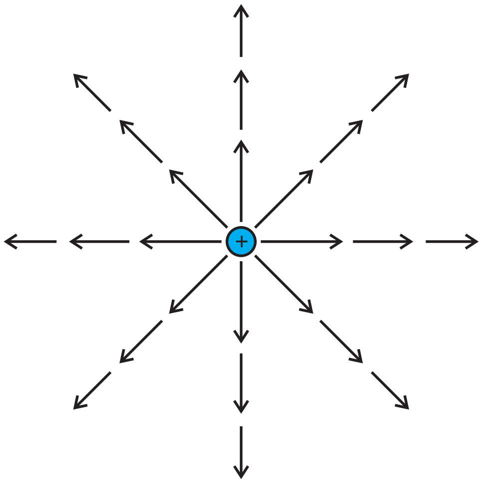


FIGURE 1.12 Field of a point charge.

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.

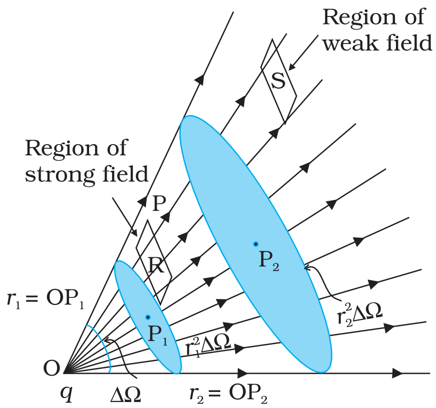


FIGURE 1.13 Dependence of electric field strength on the distance and its relation to the number of field lines.

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 1.13 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 generalisation 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. 1.13, 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 visualising 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 1.14 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) Electrostatic field lines do not form any closed loops. This follows from the conservative nature of electric field (Chapter 2).

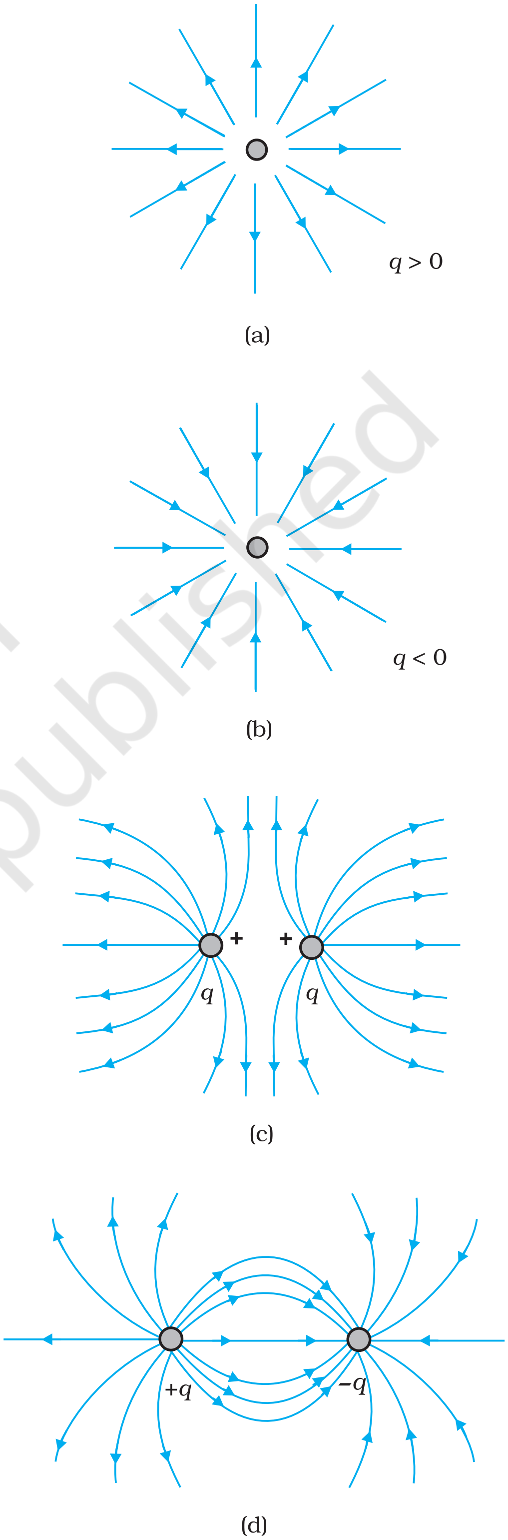


FIGURE 1.14 Field lines due to some simple charge configurations.

1.9 Electric Flux

Consider flow of a liquid with velocity $\mathbf{v}$, 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 $\mathbf{v}$, but makes an angle θ with it, the projected area in a plane perpendicular to $\mathbf{v}$ is δ 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

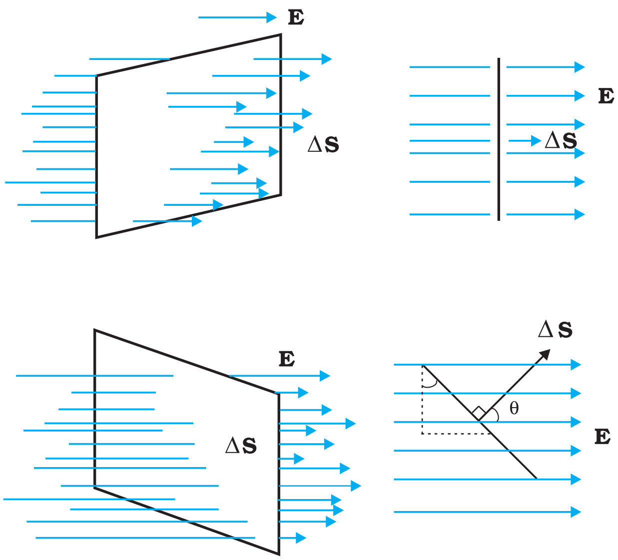


FIGURE 1.15 Dependence of flux on the inclination θ between $\mathbf{E}$ and $\hat{\mathbf{n}}$.

we place a small planar element of area ΔS normal to $\mathbf{E}$ 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. 1.15).

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. 1.16. 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.

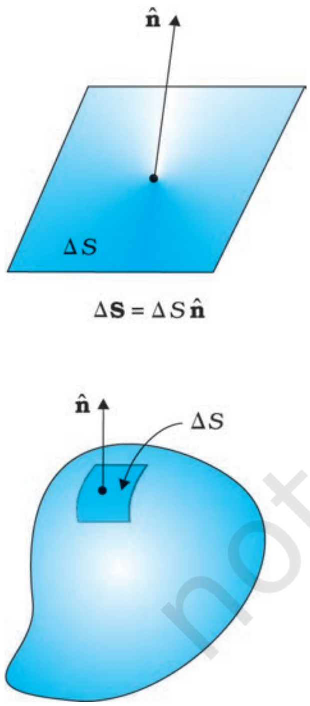


FIGURE 1.16 Convention for defining normal $\hat{\mathbf{n}}$ and ΔS.

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 $\mathbf{E}$ and $\Delta\mathbf{S}$. For a closed surface, with the convention stated already, θ is the angle between $\mathbf{E}$ 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 $\mathbf{E}$, or $E_\perp \Delta S$, i.e., component of $\mathbf{E}$ 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

$$\phi \simeq \Sigma\, \mathbf{E}\cdot\Delta\mathbf{S} \tag{1.12}$$

The approximation sign is put because the electric field $\mathbf{E}$ is taken to be constant over the small area element. This is mathematically exact only when you take the limit $\Delta S \to 0$ and the sum in Eq. (1.12) is written as an integral.

1.10 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:

1.10.1 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. 1.17(a). Then

$$\mathbf{E}_{-q} = -\frac{q}{4\pi\varepsilon_0 (r+a)^2}\,\hat{\mathbf{p}} \tag{1.13(a)}$$

where $\hat{\mathbf{p}}$ is the unit vector along the dipole axis (from $-q$ to q). Also

$$\mathbf{E}_{+q} = \frac{q}{4\pi\varepsilon_0 (r-a)^2}\,\hat{\mathbf{p}} \tag{1.13(b)}$$

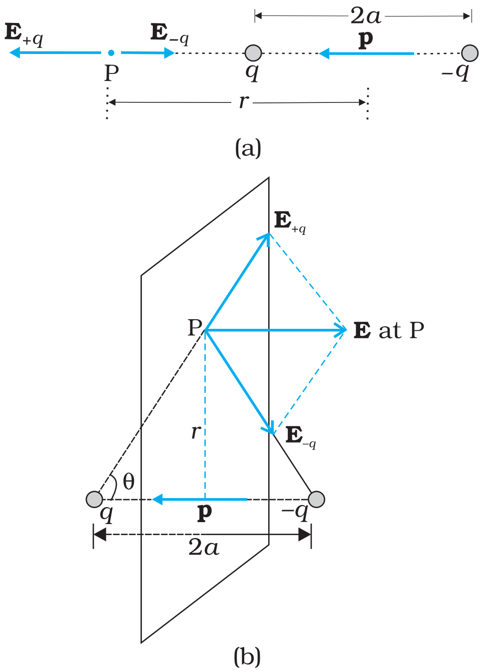


FIGURE 1.17 Electric field of a dipole at (a) a point on the axis, (b) a point on the equatorial plane of the dipole. $\mathbf{p}$ is the dipole moment vector of magnitude $p = q \times 2a$ and directed from $-q$ to q.

The total field at P is

$$\mathbf{E} = \mathbf{E}_{+q} + \mathbf{E}_{-q} = \frac{q}{4\pi\varepsilon_0}\left[\frac{1}{(r-a)^2} - \frac{1}{(r+a)^2}\right]\mathbf{p}$$

$$= \frac{q}{4\pi\varepsilon_o}\frac{4ar}{(r^2-a^2)^2}\mathbf{p} \qquad (1.14)$$

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. 1.17(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

$$\mathbf{E} = -(E_{+q} + E_{-q})\cos\theta\ \hat{\mathbf{p}}$$

$$= -\frac{2qa}{4\pi\varepsilon_o(r^2+a^2)^{3/2}}\mathbf{p} \qquad (1.17)$$

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* $\mathbf{p}$ 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 $\mathbf{p}$, 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 $\mathbf{r}$ and the dipole moment $\mathbf{p}$.

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.

1.10.2 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 1.9 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. 1.18(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. 1.18(b).

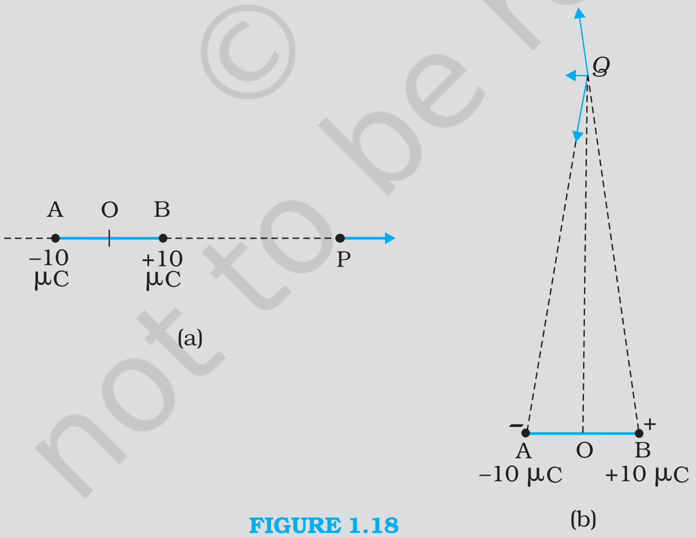


FIGURE 1.18

EXAMPLE 1.9

---

* Centre of a collection of positive point charges is defined much the same way as the centre of mass: $\mathbf{r}_{cm} = \dfrac{\sum_i q_i \mathbf{r}_i}{\sum_i q_i}$.

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 $\pm q$, $2a$ 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 1.9

$$E = \frac{p}{4\pi\varepsilon_0 r^3} \qquad (r/a >> 1)$$

$$= \frac{5\times10^{-8}\,\text{C m}}{4\pi(8.854\times10^{-12}\,\text{C}^2\,\text{N}^{-1}\,\text{m}^{-2})} \times \frac{1}{(15)^3\times10^{-6}\,\text{m}^3}$$

$= 1.33 \times 10^5\ \text{N C}^{-1}$.

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 1.9

1.11 Dipole in a Uniform External Field

Consider a permanent dipole of dipole moment **p** in a uniform external field **E**, as shown in Fig. 1.19. (By permanent dipole, we mean that **p** exists irrespective of **E**; it has not been induced by **E**.)

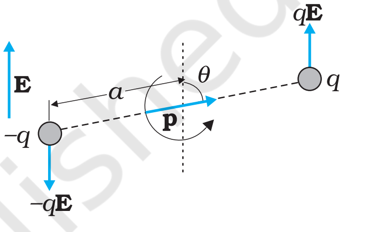


FIGURE 1.19 Dipole in a uniform electric field.

There is a force $q\mathbf{E}$ on q and a force $-q\mathbf{E}$ 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).

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,

$$\boldsymbol{\tau} = \mathbf{p}\times\mathbf{E} \qquad (1.22)$$

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.

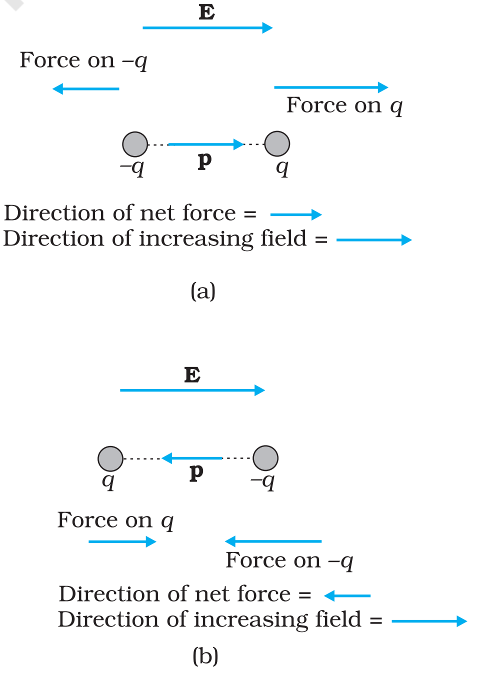


FIGURE 1.20 Electric force on a dipole: (a) **E** parallel to **p**, (b) **E** antiparallel to **p**.

Figure 1.20 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**.

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 'polarises' 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. This non-uniformity of the field makes a dipole to experience a net force on it. In this situation, it is easily seen that the paper should move in the direction of the comb!

1.12 Continuous Charge Distribution

We have so far dealt with charge configurations involving discrete charges $q_1, q_2, ..., q_n$. 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. 1.21) 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

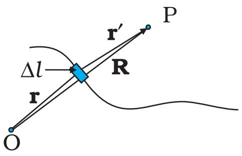
$$\lambda = \frac{\Delta Q}{\Delta l} \tag{1.24}$$

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:

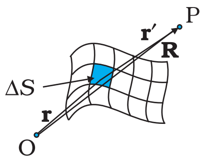
$$\rho = \frac{\Delta Q}{\Delta V} \tag{1.25}$$

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.

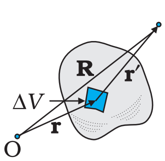
The notion of continuous charge distribution is similar to that we adopt for continuous mass distribution in mechanics. When we refer to



Line charge $\Delta Q = \lambda \Delta l$



Surface charge $\Delta Q = \sigma \Delta S$



Volume charge $\Delta Q = \rho \Delta V$

FIGURE 1.21 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 $\mathbf{r}$. The charge density ρ may vary from point to point, i.e., it is a function of $\mathbf{r}$. 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 $\mathbf{R}$ (Fig. 1.21). 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.

1.13 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. 1.22.

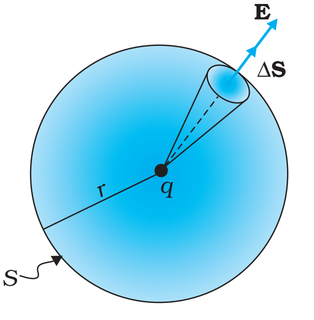


FIGURE 1.22 Flux through a sphere enclosing a point charge q at its centre.

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:

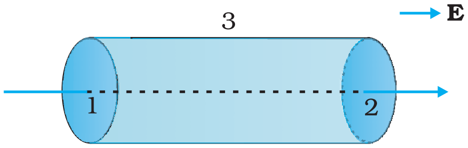


FIGURE 1.23 Calculation of the flux of uniform electric field through the surface of a cylinder.

$$\phi = \sum_{\text{all } \Delta S} \frac{q}{4\pi\varepsilon_0 r^2} \Delta S$$

Since each area element of the sphere is at the same distance r from the charge,

$$\phi = \frac{q}{4\pi\varepsilon_o r^2} \sum_{\text{all } \Delta S} \Delta S = \frac{q}{4\pi\varepsilon_0 r^2} S$$

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

$= q/\varepsilon_0$ (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. 1.23.

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 1.10 The electric field components in Fig. 1.24 are $E_x = \alpha x^{1/2}$, $E_y = E_z = 0$, in which α = 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.

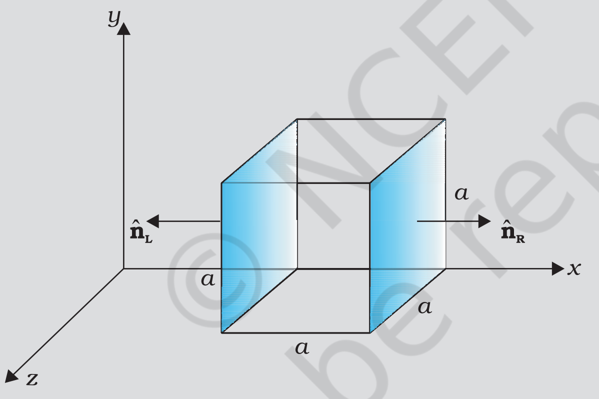


FIGURE 1.24

Solution

(a) Since the electric field has only an x component, for faces perpendicular to x direction, the angle between $\mathbf{E}$ and $\Delta\mathbf{S}$ 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 θ = 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 θ = 0°

$= E_R a^2$

Net flux through the cube

Example 1.10

$= \phi_R + \phi_L = E_R a^2 - E_L a^2 = a^2 (E_R - E_L) = \alpha a^2 [(2a)^{1/2} - a^{1/2}]$

$= \alpha a^{5/2} \left(\sqrt{2} - 1\right)$

$= 800 \ (0.1)^{5/2} \left(\sqrt{2} - 1\right)$

$= 1.05 \text{ N m}^2 \text{ C}^{-1}$

(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 1.10

Example 1.11 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. 1.25). (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

$\phi = 1.57 + 1.57 + 0 = 3.14 \text{ N m}^2 \text{ C}^{-1}$

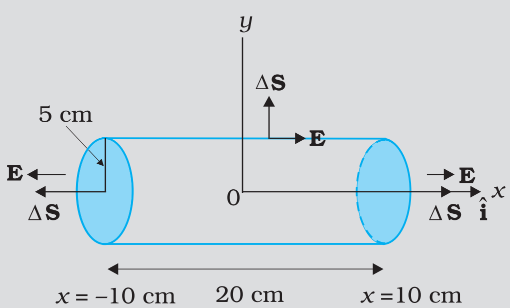


FIGURE 1.25

(d) The net charge within the cylinder can be found by using Gauss's law which gives

$q = \varepsilon_0 \phi$

$= 3.14 \times 8.854 \times 10^{-12}$ C

$= 2.78 \times 10^{-11}$ C

EXAMPLE 1.11

1.14 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.

1.14.1 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. 1.26.

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. 1.26(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.

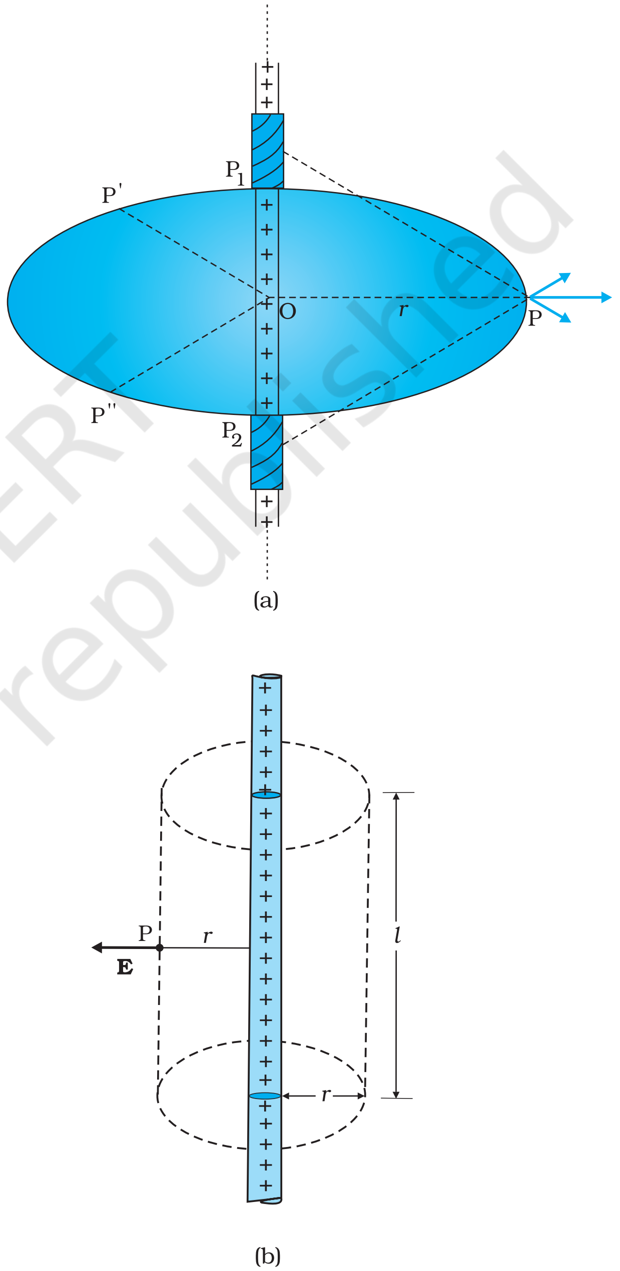


FIGURE 1.26 (a) Electric field due to an infinitely long thin straight wire is radial, (b) The Gaussian surface for a long thin wire of uniform linear charge density.

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., $E = \dfrac{\lambda}{2\pi\varepsilon_0 r}$

Vectorially, **E** at any point is given by

$$\mathbf{E} = \frac{\lambda}{2\pi\varepsilon_0 r}\,\hat{\mathbf{n}} \qquad (1.32)$$

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.

1.14.2 Field due to a uniformly charged infinite plane sheet

Let σ be the uniform surface charge density of an infinite plane sheet (Fig. 1.27). 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.

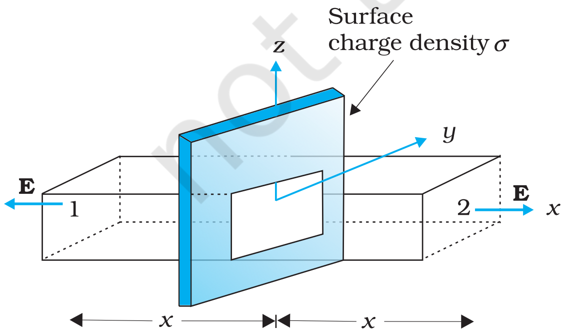


FIGURE 1.27 Gaussian surface for a uniformly charged infinite plane sheet.

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.

1.14.3 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. 1.28). 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 Δ**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.
Vectorically,

$$\mathbf{E} = \frac{q}{4\pi\varepsilon_0\,r^2}\hat{\mathbf{r}} \tag{1.34}$$

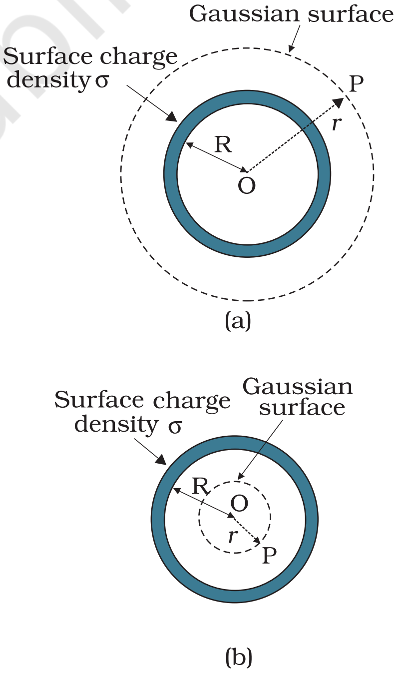


FIGURE 1.28 Gaussian surfaces for a point with (a) $r > R$, (b) $r < R$.

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. 1.28(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., $E = 0 \qquad (r < R)$ (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 1.12 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?

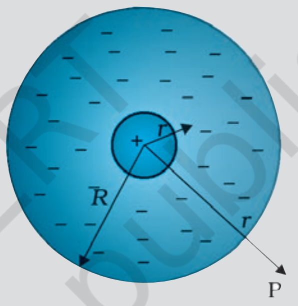


FIGURE 1.29

Solution The charge distribution for this model of the atom is as shown in Fig. 1.29. The total negative charge in the uniform spherical charge distribution of radius *R* must be $-Z\,e$, since the atom (nucleus of charge $Z\,e$ + 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 $$\rho = -\frac{3\,Ze}{4\pi R^3}$$

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

$$\phi = E(r) \times 4\pi r^2$$

EXAMPLE 1.12

* Compare this with a uniform mass shell discussed in Section 7.5 of Class XI Textbook of Physics.

where $E(r)$ is the magnitude of the electric field at r. This is because 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., $q = Ze + \dfrac{4\pi r^3}{3}\rho$

Substituting for the charge density ρ obtained earlier, we have

$$q = Ze - Ze\frac{r^3}{R^3}$$

Gauss's law then gives,

$$E(r) = \frac{Ze}{4\pi\varepsilon_0}\ \frac{1}{r^2} - \frac{r}{R^3}\ ; \qquad r < R$$

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 $E(r) = 0$; $\quad r > R$

At $r = R$, both cases give the same result: $E = 0$.

EXAMPLE 1.12

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_1q_2 and inversely proportional to the square of the distance r_{21} separating them. Mathematically,

$$\mathbf{F}_{21} = \text{force on } q_2 \text{ due to } q_1 = \frac{k\ (q_1q_2)}{r_{21}^2}\,\hat{\mathbf{r}}_{21}$$

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, ..., 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:

$$\mathbf{E} = \frac{-\mathbf{p}}{4\,\pi\varepsilon_o} \frac{1}{(a^2 + r^2)^{3/2}}$$

$$\cong \frac{-\mathbf{p}}{4\,\pi\varepsilon_o\, r^3}, \qquad \textit{for } r >> a$$

Dipole electric field on the axis at a distance r from the centre:

$$\mathbf{E} = \frac{2\,\mathbf{p}r}{4\,\pi\varepsilon_0 (r^2 - a^2)^2}$$

$$\cong \frac{2\mathbf{p}}{4\,\pi\varepsilon_0 r^3} \qquad \textit{for} \quad r >> a$$

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 $\mathbf{E}$, 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 $\mathbf{E}$ 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 $\mathbf{E}$, 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 Figure 1.30 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?

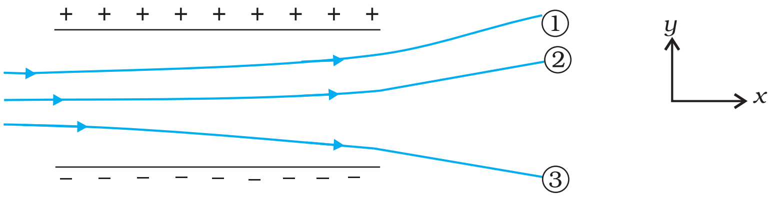


FIGURE 1.30

1.14 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?

1.15 What is the net flux of the uniform electric field of Exercise 1.14 through a cube of side 20 cm oriented so that its faces are parallel to the coordinate planes?

1.16 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?

1.17 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. 1.31. 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.)

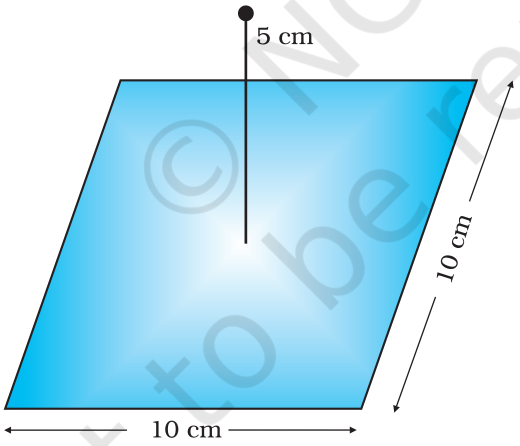


FIGURE 1.31

1.18 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?

1.19 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?

1.20 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?

1.21 A uniformly charged conducting sphere of 2.4 m diameter has a surface charge density of 80.0 $\mu C/m^2$. (a) Find the charge on the sphere. (b) What is the total electric flux leaving the surface of the sphere?

1.22 An infinite line charge produces a field of 9×10^4 N/C at a distance of 2 cm. Calculate the linear charge density.

1.23 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?

12089CH02

Chapter Two

ELECTROSTATIC POTENTIAL AND CAPACITANCE

2.1 Introduction

In Chapters 5 and 7 (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 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 **E** due to some charge configuration. First, for simplicity, consider the field **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

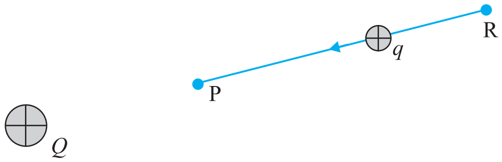


FIGURE 2.1 A test charge q (> 0) is moved from the point R to the point P against the repulsive force on it by the charge Q (> 0) placed at the origin.

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

$$W_{RP} = \int_R^P \mathbf{F}_{ext} \cdot d\mathbf{r}$$

$$= -\int_R^P \mathbf{F}_{ext} \cdot d\mathbf{r} \qquad (2.1)$$

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.

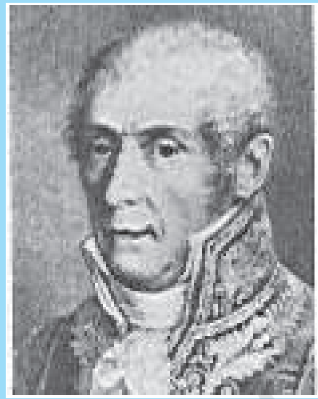
(ii) Equation (2.2) defines *potential energy difference* in terms of the physically meaningful quantity *work*. Clearly, potential energy so defined is undetermined to within an additive constant.What this means is that the actual value of potential energy is not physically significant; it is only the 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)

$$W_{\infty P} = U_P - U_\infty = U_P \tag{2.3}$$

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

$$= V_P - V_R \left(= \frac{U_P - U_R}{q} \right) \tag{2.4}$$

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.

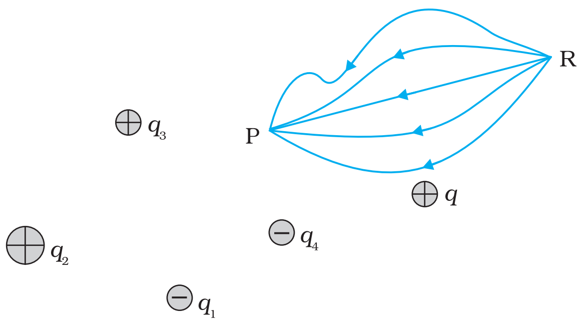


FIGURE 2.2 Work done on a test charge q by the electrostatic field due to any given charge configuration is independent of the path, and depends only on its initial and final positions.

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 $\mathbf{r}$ 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.

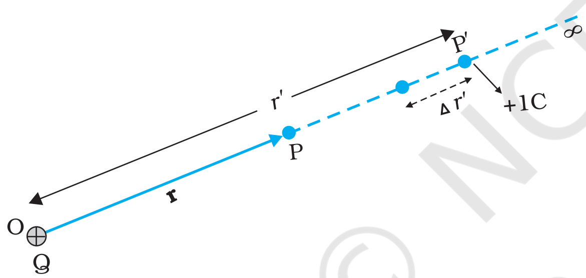


FIGURE 2.3 Work done in bringing a unit positive test charge from infinity to the point P, against the repulsive force of charge Q ($Q > 0$), is the potential at P due to the charge Q.

At some intermediate point P′ on the path, the electrostatic force on a unit positive charge is

$$\frac{Q \times 1}{4\pi\varepsilon_0 r'^2}\hat{\mathbf{r}}' \qquad (2.5)$$

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

$$\Delta W = -\frac{Q}{4\pi\varepsilon_0 r'^2}\Delta r' \qquad (2.6)$$

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 $r' = r$,

$$W = -\int_{\infty}^{r}\frac{Q}{4\pi\varepsilon_0 r'^2}dr' = \left.\frac{Q}{4\pi\varepsilon_0 r'}\right|_{\infty}^{r} = \frac{Q}{4\pi\varepsilon_0 r} \qquad (2.7)$$

This, by definition is the potential at P due to the charge Q

$$V(r) = \frac{Q}{4\pi\varepsilon_0 r} \qquad (2.8)$$

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.

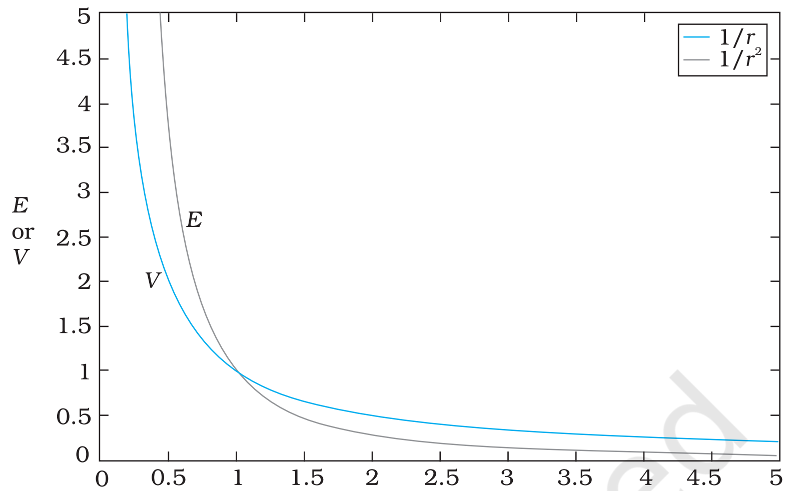


FIGURE 2.4 Variation of potential V with r [in units of $(Q/4\pi\varepsilon_0)$ m^{-1}] (blue curve) and field with r [in units of $(Q/4\pi\varepsilon_0)$ m^{-2}] (black curve) for a point charge Q.

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) $V = \dfrac{1}{4\pi\varepsilon_0}\dfrac{Q}{r} = 9 \times 10^9 \text{ Nm}^2 \text{ C}^{-2} \times \dfrac{4 \times 10^{-7}\text{C}}{0.09\text{ m}}$

$= 4 \times 10^4$ V

(b) $W = qV = 2 \times 10^{-9}\text{C} \times 4 \times 10^4\text{V}$

$= 8 \times 10^{-5}$ 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,

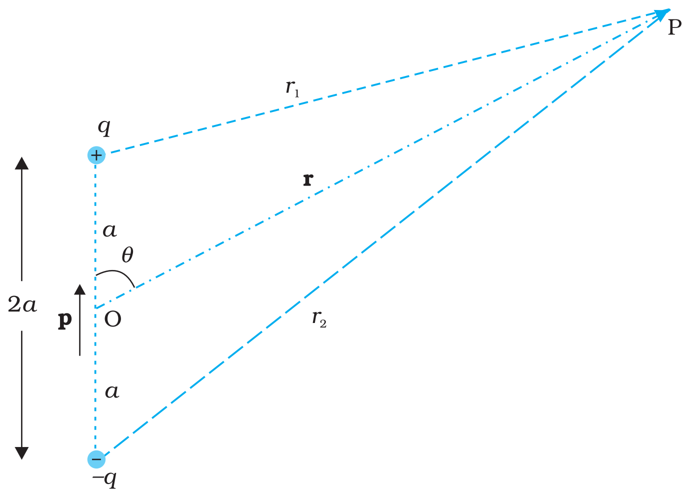


FIGURE 2.5 Quantities involved in the calculation of potential due to a dipole.

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,

$$\frac{1}{r_1} \cong \frac{1}{r}\left(1 - \frac{2a\cos\theta}{r}\right)^{-1/2} \cong \frac{1}{r}\left(1 + \frac{a}{r}\cos\theta\right) \qquad [2.13(a)]$$

$$\frac{1}{r_2} \cong \frac{1}{r}\left(1 + \frac{2a\cos\theta}{r}\right)^{-1/2} \cong \frac{1}{r}\left(1 - \frac{a}{r}\cos\theta\right) \qquad [2.13(b)]$$

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, $p\cos\theta = \mathbf{p}.\hat{\mathbf{r}}$

where $\hat{\mathbf{r}}$ is the unit vector along the position vector **OP**.

The electric potential of a dipole is then given by

$$V = \frac{1}{4\pi\varepsilon_0}\frac{\mathbf{p}.\hat{\mathbf{r}}}{r^2}; \qquad (r >> a) \tag{2.15}$$

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

$$V = V_1 + V_2 + ... + V_n \tag{2.17}$$

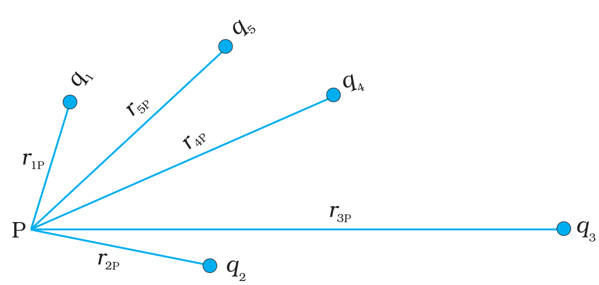


FIGURE 2.6 Potential at a point due to a system of charges is the sum of potentials due to individual charges.

$$= \frac{1}{4\pi\varepsilon_0}\left(\frac{q_1}{r_{1P}} + \frac{q_2}{r_{2P}} + \ldots\ldots + \frac{q_n}{r_{nP}}\right) \tag{2.18}$$

If we have a continuous charge distribution characterised by a charge density ρ (**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 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).

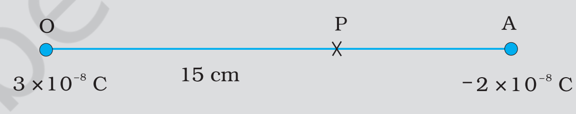


FIGURE 2.7

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

$$\frac{3}{x} - \frac{2}{x-15} = 0$$

EXAMPLE 2.2

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.

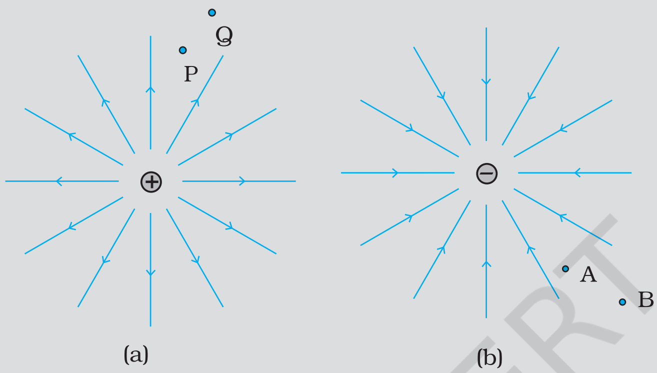


FIGURE 2.8

(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.

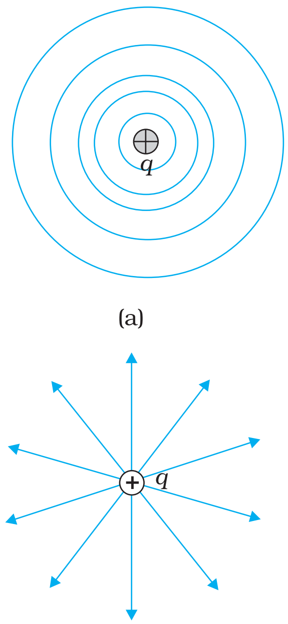
(d) In moving a small negative charge from B to A work has to be done by the external agency. It is positive.

(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://video.mit.edu/watch/4-electrostatic-potential-electric-energy-ev-conservative-field-equipotential-sufaces-12584/



(a)

(b)

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 $q > 0$.

2.6 Equipotential Surfaces

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.

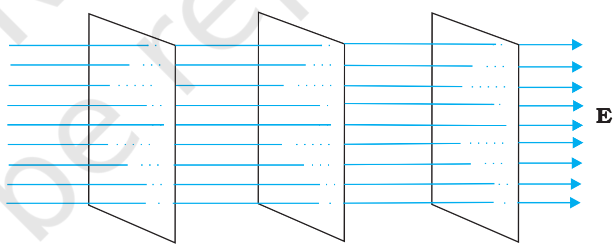
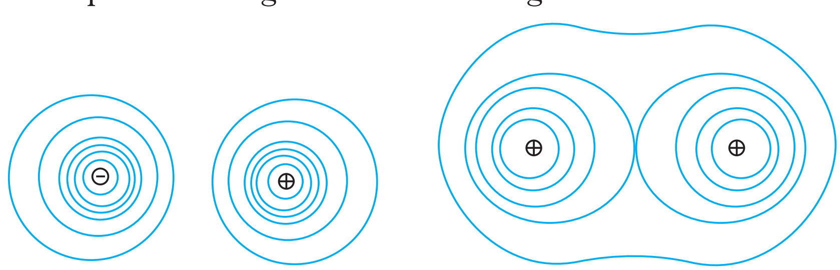


FIGURE 2.10 Equipotential surfaces for a uniform electric field.

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.



(a) (b)

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,

$$|\mathbf{E}|\,\delta l = V - (V + \delta V) = -\,\delta V$$

$$\text{i.e., } |\mathbf{E}| = -\frac{\delta V}{\delta l} \tag{2.20}$$

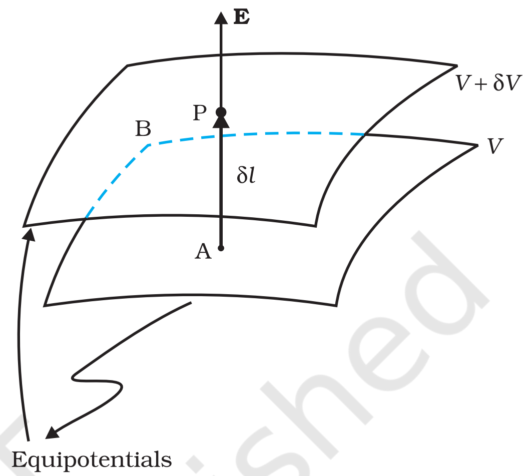


FIGURE 2.12 From the potential to the field.

Since δV is negative, $\delta V = -\,|\delta V|$. we can rewrite Eq (2.20) as

$$|\mathbf{E}| = -\frac{\delta V}{\delta l} = +\frac{|\delta V|}{\delta l} \tag{2.21}$$

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:

$$\text{work done on } q_2 = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r_{12}}$$

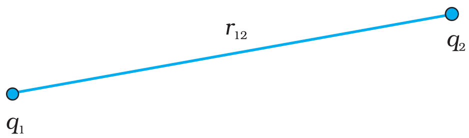


FIGURE 2.13 Potential energy of a system of charges q_1 and q_2 is directly proportional to the product of charges and inversely to the distance between them.

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

$$U = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r_{12}} \qquad (2.22)$$

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

$$q_2 V_1(\mathbf{r}_2) = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r_{12}} \qquad (2.23)$$

The charges q_1 and q_2 produce a potential, which at any point P is given by

$$V_{1,2} = \frac{1}{4\pi\varepsilon_0}\left(\frac{q_1}{r_{1\mathrm{P}}} + \frac{q_2}{r_{2\mathrm{P}}}\right) \qquad (2.24)$$

Work done next in bringing q_3 from infinity to the point $\mathbf{r}_3$ is q_3 times $V_{1,\,2}$ at $\mathbf{r}_3$

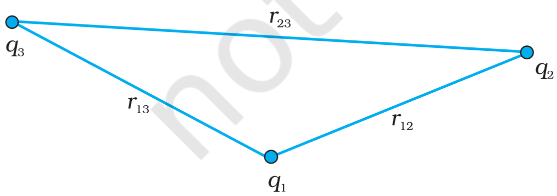


FIGURE 2.14 Potential energy of a system of three charges is given by Eq. (2.26), with the notation given in the figure.

$$q_3 V_{1,2}(\mathbf{r}_3) = \frac{1}{4\pi\varepsilon_0}\left(\frac{q_1 q_3}{r_{13}} + \frac{q_2 q_3}{r_{23}}\right) \qquad (2.25)$$

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)],

$$U = \frac{1}{4\pi\varepsilon_0}\left(\frac{q_1 q_2}{r_{12}} + \frac{q_1 q_3}{r_{13}} + \frac{q_2 q_3}{r_{23}}\right) \qquad (2.26)$$

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?

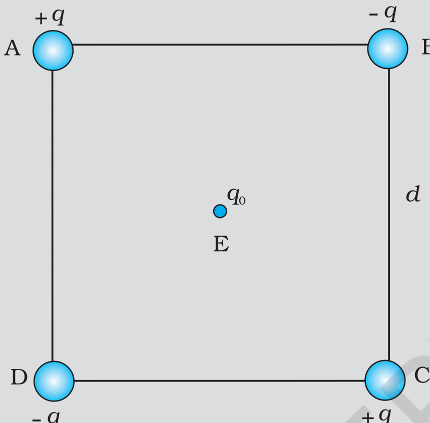


FIGURE 2.15

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 $\mathbf{r}$ relative to some origin, we can write:

Potential energy of q at $\mathbf{r}$ in an external field

$$= qV(\mathbf{r}) \tag{2.27}$$

where $V(\mathbf{r})$ is the external potential at the point $\mathbf{r}$.

Thus, if an electron with charge $q = e = 1.6\times10^{-19}$ C is accelerated by a potential difference of $\Delta V = 1$ 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 $\mathbf{E}$ 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 ($\mathbf{E}$ 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 $E = A\,(1/r^2)$; $A = 9 \times 10^5$ NC^{-1} m^2. What would the electrostatic energy of the configuration be?

Solution

(a) $U = \dfrac{1}{4\pi\varepsilon_0}\dfrac{q_1 q_2}{r} = 9 \times 10^9 \times \dfrac{7 \times (-2) \times 10^{-12}}{0.18} = -0.7$ 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.

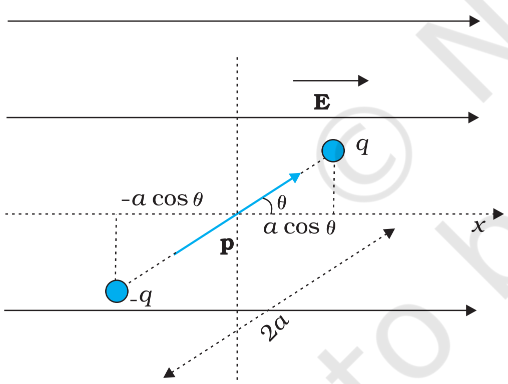


FIGURE 2.16 Potential energy of a dipole in a uniform external field.

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 $\boldsymbol{\tau}_{ext}$ 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

$$W = \int_{\theta_0}^{\theta_1} t_{\,ext}(\theta) d\theta = \int_{\theta_0}^{\theta_1} pE \sin\theta \, d\theta$$

$$= pE\left(\cos\theta_0 - \cos\theta_1\right) \tag{2.31}$$

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,

$$U(\theta) = pE\left(\cos\frac{\pi}{2} - \cos\theta\right) = \;\; pE\cos\theta = -\mathbf{p.E} \tag{2.32}$$

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,

$$U'(\theta) = -pE\cos\theta - \frac{q^2}{4\pi\varepsilon_0 \times 2a} = -\mathbf{p}.\mathbf{E} - \frac{q^2}{4\pi\varepsilon_0 \times 2a} \tag{2.34}$$

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 **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 $- (-6$J$) = 3$ 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

$$\mathbf{E} = \frac{\sigma}{\varepsilon_0}\hat{\mathbf{n}} \qquad (2.35)$$

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, $\mathbf{E}$ may be considered constant and $\mathbf{E}$ and δS are parallel or antiparallel. The charge enclosed by the pill box is $\sigma\delta S$.

By Gauss's law

$$E\delta S = \frac{|\sigma|\,\delta S}{\varepsilon_0}$$

$$E = \frac{|\sigma|}{\varepsilon_0} \qquad (2.36)$$

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.

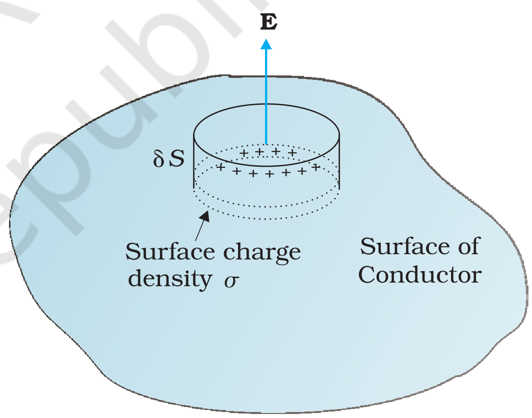


FIGURE 2.17 The Gaussian surface (a pill box) chosen to derive Eq. (2.35) for electric field at the surface of a charged conductor.

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

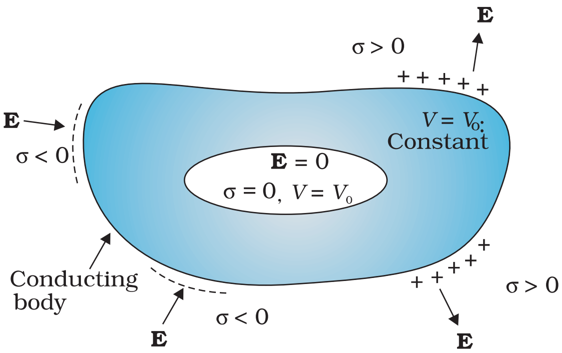


FIGURE 2.18 The electric field inside a cavity of any conductor is zero. All charges reside only on the outer surface of a conductor with cavity. (There are no charges placed in the cavity.)

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.

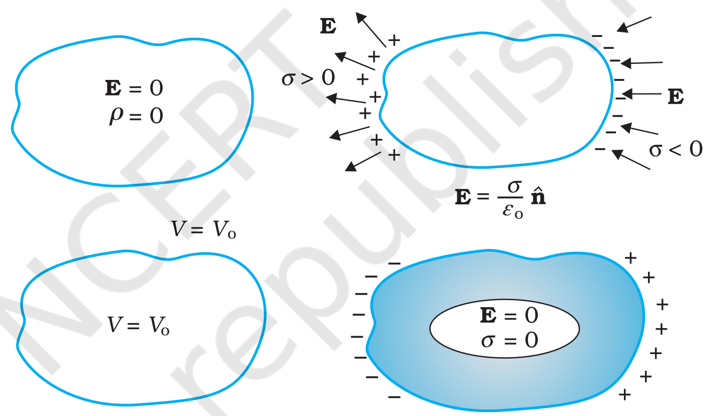


FIGURE 2.19 Some important electrostatic properties of a conductor.

Example 2.7

(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.

EXAMPLE 2.7

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.

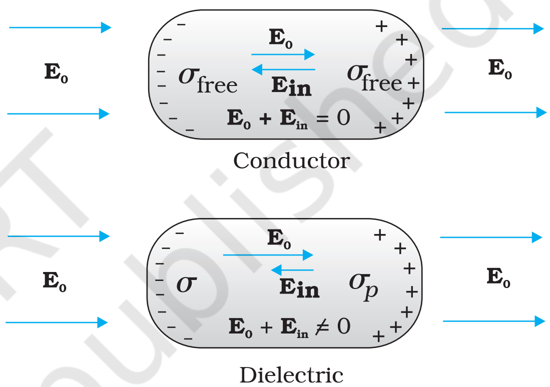


FIGURE 2.20 Difference in behaviour of a conductor and a dielectric in an external electric field.

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.

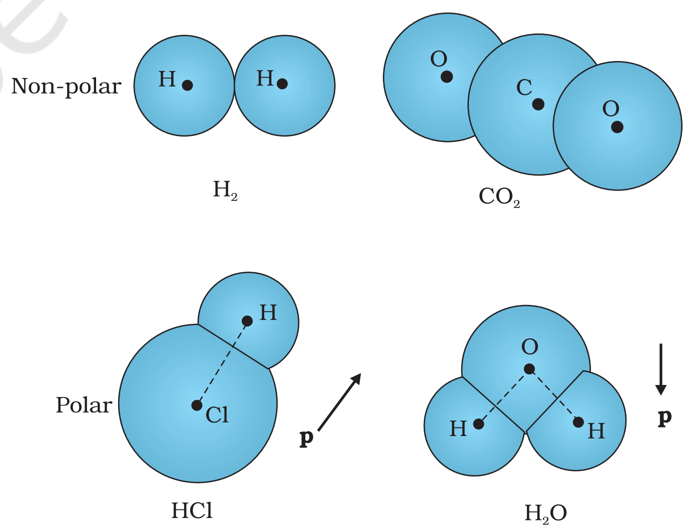
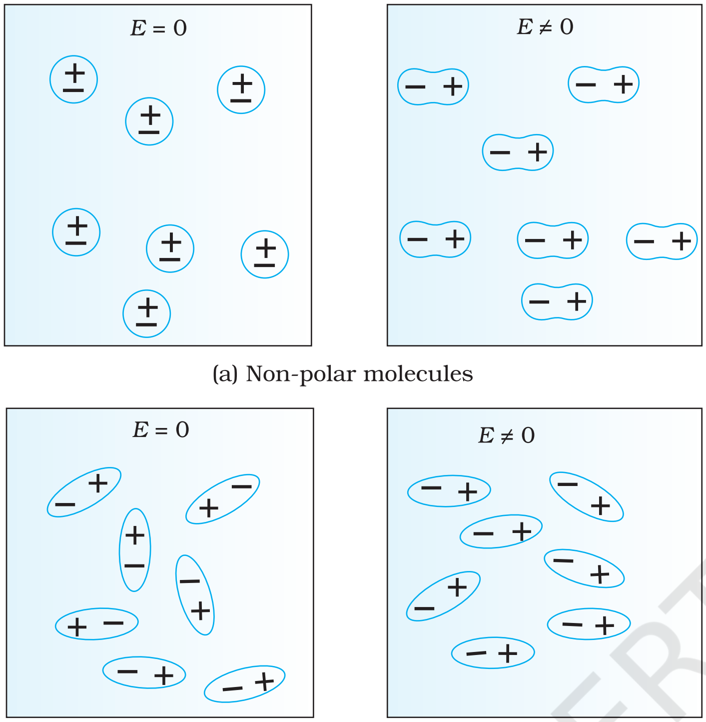


FIGURE 2.21 Some examples of polar and non-polar molecules.



(a) Non-polar molecules

(b) 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 overall 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,

$$\mathbf{P} = \varepsilon_0 \chi_e \, \mathbf{E} \tag{2.37}$$

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.

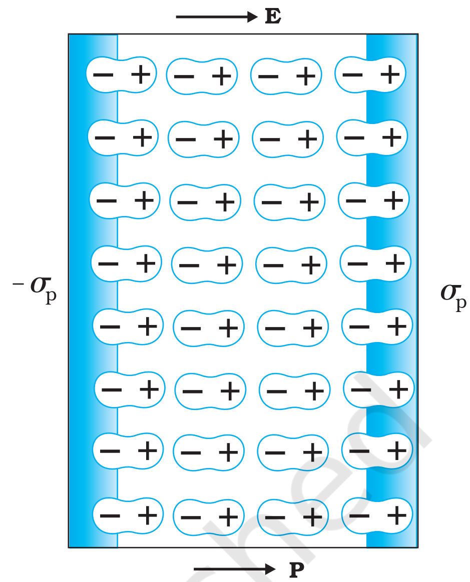


FIGURE 2.23 A uniformly polarised dielectric amounts to induced surface charge density, but no volume charge density.

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:

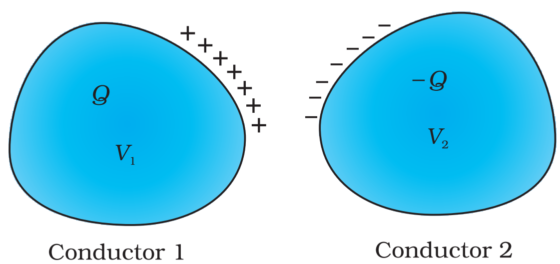


FIGURE 2.24 A system of two conductors separated by an insulator forms a capacitor.

$$C = \frac{Q}{V} \qquad (2.38)$$

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:

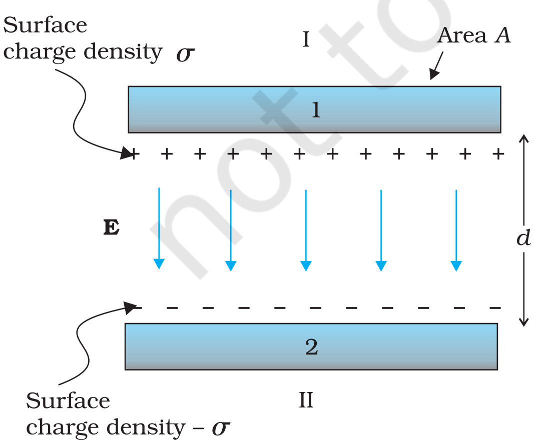


FIGURE 2.25 The parallel plate capacitor.

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),

$$E = \frac{\sigma}{2\varepsilon_0} - \frac{\sigma}{2\varepsilon_0} = 0 \qquad (2.40)$$

In the inner region between the plates 1 and 2, the electric fields due to the two charged plates add up, giving

$$E = \frac{\sigma}{2\varepsilon_0} + \frac{\sigma}{2\varepsilon_0} = \frac{\sigma}{\varepsilon_0} = \frac{Q}{\varepsilon_0 A} \qquad (2.41)$$

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,

$$V = E\,d = \frac{1}{\varepsilon_0}\frac{Qd}{A} \qquad (2.42)$$

The capacitance C of the parallel plate capacitor is then

$$C = \frac{Q}{V} = = \frac{\varepsilon_0 A}{d} \qquad (2.43)$$

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

$$C = \frac{8.85 \times 10^{-12}\,\text{C}^2\text{N}^{-1}\text{m}^{-2} \times 1\,\text{m}^2}{10^{-3}\,\text{m}} = 8.85 \times 10^{-9}\,\text{F} \qquad (2.44)$$

(You can check that if 1F= 1C V^{-1} = 1C $(\text{NC}^{-1}\text{m})^{-1}$ = 1 $\text{C}^2\,\text{N}^{-1}\text{m}^{-1}$.) 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:

$$A = \frac{Cd}{\varepsilon_0} = \frac{1\text{F} \times 10^{-2}\,\text{m}}{8.85 \times 10^{-12}\,\text{C}^2\text{N}^{-1}\text{m}^{-2}} = 10^9\,\text{m}^2 \qquad (2.45)$$

which is a plate about 30 km in length and breadth!

2.13 Effect of Dielectric on Capacitance

With the understanding of the behaviour 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.fsu.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.

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

$$V = E_0\left(\frac{1}{4}d\right) + \frac{E_0}{K}\left(\frac{3}{4}d\right)$$

$$= E_0 d\left(\frac{1}{4} + \frac{3}{4K}\right) = V_0\frac{K+3}{4K}$$

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

$$C = \frac{Q_0}{V} = \frac{4K}{K+3}\frac{Q_0}{V_0} = \frac{4K}{K+3}C_0$$

EXAMPLE 2.8

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

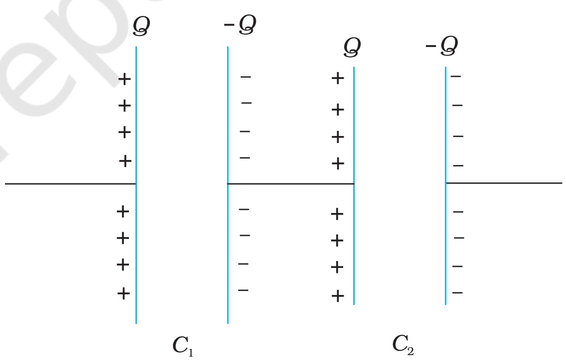


FIGURE 2.26 Combination of two capacitors in series.

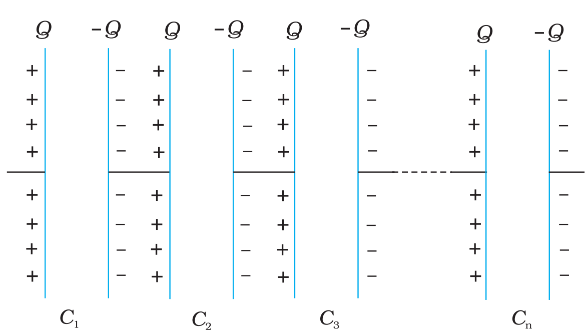


FIGURE 2.27 Combination of n capacitors in series.

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.

$$V = V_1 + V_2 = \frac{Q}{C_1} + \frac{Q}{C_2} \tag{2.55}$$

$$\text{i.e.,} \quad \frac{V}{Q} = \frac{1}{C_1} + \frac{1}{C_2}, \tag{2.56}$$

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

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} \tag{2.58}$$

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

$$V = V_1 + V_2 + \ldots + V_n = \frac{Q}{C_1} + \frac{Q}{C_2} + \ldots + \frac{Q}{C_n} \tag{2.59}$$

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:

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \ldots + \frac{1}{C_n} \tag{2.60}$$

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:

$$Q_1 = C_1 V, \; Q_2 = C_2 V \tag{2.61}$$

The equivalent capacitor is one with charge

$$Q = Q_1 + Q_2 \tag{2.62}$$

and potential difference V.

$$Q = CV = C_1 V + C_2 V \tag{2.63}$$

The effective capacitance C is, from Eq. (2.63),

$$C = C_1 + C_2 \tag{2.64}$$

The general formula for effective capacitance C for parallel combination of n capacitors [Fig. 2.28 (b)] follows similarly,

$$Q = Q_1 + Q_2 + \ldots + Q_n \tag{2.65}$$

$$\text{i.e.,} \quad CV = C_1 V + C_2 V + \ldots C_n V \tag{2.66}$$

which gives

$$C = C_1 + C_2 + \ldots C_n \tag{2.67}$$

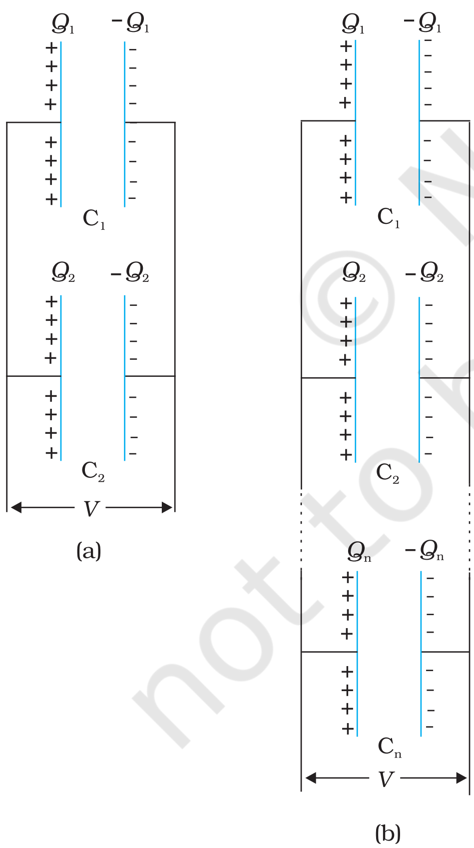


FIGURE 2.28 Parallel combination of (a) two capacitors, (b) n capacitors.

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.)

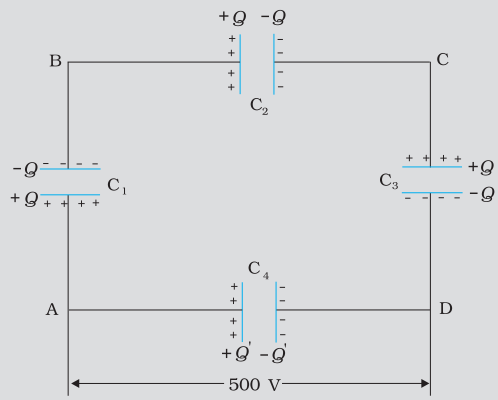


FIGURE 2.29

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 $C_1 = C_2 = C_3 = 10$ μF, $C' = (10/3)$ μF. 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

$$\frac{Q}{C_1} + \frac{Q}{C_2} + \frac{Q}{C_3} = 500\text{ V}.$$

Also, $Q'/C_4 = 500$ V.

This gives for the given value of the capacitances,

$$Q = 500\,V \times \frac{10}{3} \mu F = 1.7 \times 10^{-3}\text{C} \text{ and}$$

$$Q' = 500\,V \times 10 \mu F = 5.0 \times 10^{-3}\text{ C}$$

EXAMPLE 2.9

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

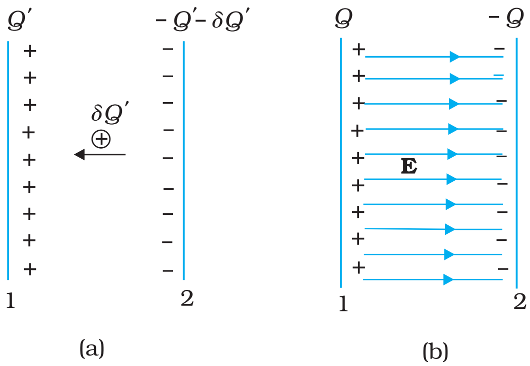


FIGURE 2.30 (a) Work done in a small step of building charge on conductor 1 from Q' to $Q' + \delta Q'$. (b) Total work done in charging the capacitor may be viewed as stored in the energy of electric field between the plates.

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).

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 (δW), resulting in charge Q' on conductor 1 increasing to $Q' + \delta Q'$, is given by

$$\delta W = V'\delta Q' = \frac{Q'}{C}\delta Q' \tag{2.68}$$

Integrating eq. (2.68)

$$W = \int_0^Q \frac{Q'}{C}\delta Q' = \frac{1}{C}\frac{Q'^2}{2}\bigg|_0^Q = \frac{Q^2}{2C}$$

We can write the final result, in different ways

$$W = \frac{Q^2}{2C} = \frac{1}{2}CV^2 = \frac{1}{2}QV \tag{2.69}$$

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. (2.69)] 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

$$= \frac{1}{2}\frac{Q^2}{C} = \frac{(A\sigma)^2}{2} \times \frac{d}{\varepsilon_0 A} \tag{2.70}$$

The surface charge density σ is related to the electric field E between the plates,

$$E = \frac{\sigma}{\varepsilon_0} \tag{2.71}$$

From Eqs. (2.70) and (2.71) , we get

Energy stored in the capacitor

$$U = (1/2)\varepsilon_0 E^2 \times A\,d \tag{2.72}$$

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 (2.72) shows that

Energy density of electric field,

$$u = (1/2)\varepsilon_0 E^2 \tag{2.73}$$

Though we derived Eq. (2.73) 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?

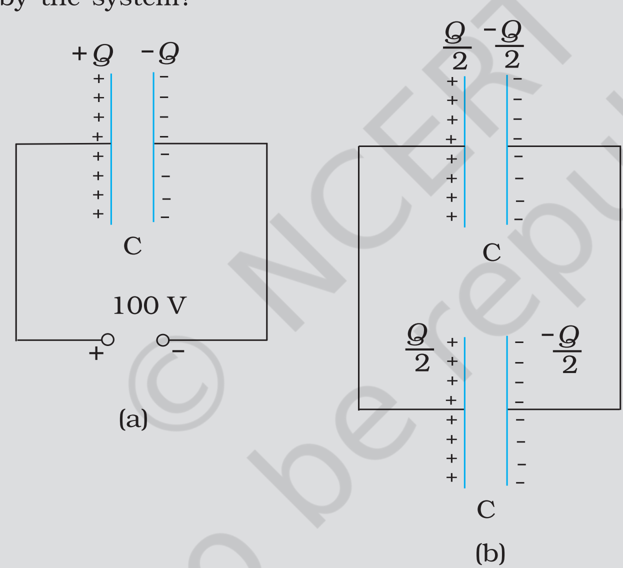


FIGURE 2.31

Solution

(a) The charge on the capacitor is

$$Q = CV = 900 \times 10^{-12}\ \text{F} \times 100\ \text{V} = 9 \times 10^{-8}\ \text{C}$$

The energy stored by the capacitor is

$$= (1/2)\ CV^2 = (1/2)\ QV$$

$$= (1/2) \times 9 \times 10^{-8}\text{C} \times 100\ \text{V} = 4.5 \times 10^{-6}\ \text{J}$$

(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

EXAMPLE 2.10

EXAMPLE 2.10

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.

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 $q(V_P - V_R)$, 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
$$V(\mathbf{r}) = \frac{1}{4\pi\varepsilon_o} \frac{\mathbf{p}.\hat{\mathbf{r}}}{r^2}$$
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 centred 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 $\mathbf{p}$ in a uniform electric field $\mathbf{E}$ is $-\mathbf{p}\cdot\mathbf{E}$.

8. Electrostatics field $\mathbf{E}$ is zero in the interior of a conductor; just outside the surface of a charged conductor, $\mathbf{E}$ is normal to the surface given by

$\mathbf{E} = \dfrac{\sigma}{\varepsilon_0}\hat{\mathbf{n}}$ 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 $(1/2)\varepsilon_0 E^2$.

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	**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?

12089CH03

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.

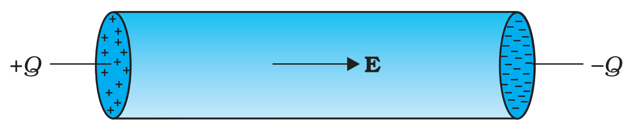


FIGURE 3.1 Charges $+Q$ and $-Q$ put at the ends of a metallic cylinder. The electrons will drift because of the electric field created to neutralise the charges. The current thus will stop after a while unless the charges $+Q$ and $-Q$ are continuously replenished.

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.

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

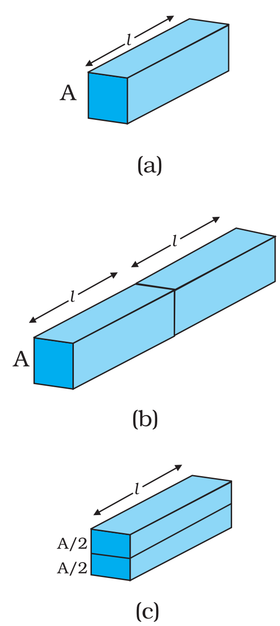


FIGURE 3.2 Illustrating the relation $R = \rho l/A$ for a rectangular slab of length l and area of cross-section A.

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

$$V = I \times R = \frac{I\rho l}{A} \tag{3.10}$$

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, $$E = j\,\rho \qquad (3.11)$$

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 $\mathbf{j}$ $(\equiv j\,\mathbf{E}/E)$. Thus, the last equation can be written as,

$$\mathbf{E} = \mathbf{j}\rho \qquad (3.12)$$

or, $$\mathbf{j} = \sigma\,\mathbf{E} \qquad (3.13)$$

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 i^{th} electron (i = 1, 2, 3, ... N) at a given time is $\mathbf{v}_i$, then

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{v}_i = 0 \qquad (3.14)$$

Consider now the situation when an electric field is present. Electrons will be accelerated due to this field by

$$\mathbf{a} = \frac{-e\,\mathbf{E}}{m} \qquad (3.15)$$

where $-e$ is the charge and $\boldsymbol{m}$ is the mass of an electron. Consider again the $\boldsymbol{i}^{th}$ electron at a given time $\boldsymbol{t}$. 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

$$\mathbf{V}_i = \mathbf{v}_i + \frac{-e\mathbf{E}}{m}\,t_i \qquad (3.16)$$

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

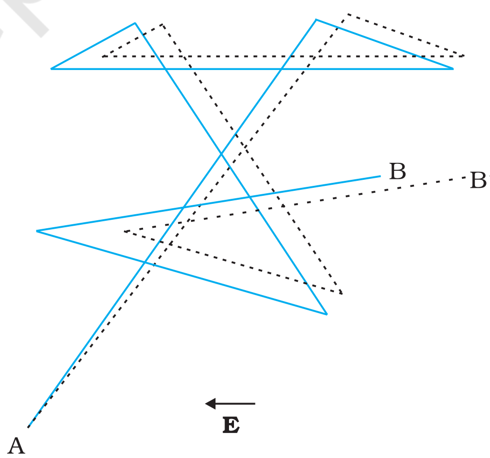


FIGURE 3.3 A schematic picture of an electron moving from a point A to another point B through repeated collisions, and straight line travel between collisions (full lines). If an electric field is applied as shown, the electron ends up at point B′ (dotted lines). A slight drift in a direction opposite 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 $\mathbf{v}_d$

$$\mathbf{v}_d \equiv (\mathbf{V}_i)_{average} = (\mathbf{v}_i)_{average} - \frac{e\mathbf{E}}{m}(t_i)_{average}$$

$$= 0 - \frac{e\mathbf{E}}{m}\tau = -\frac{e\mathbf{E}}{m}\tau \tag{3.17}$$

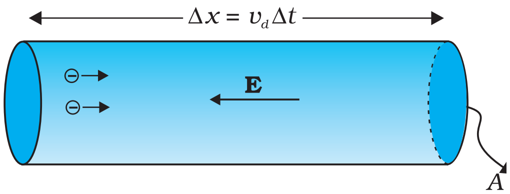


FIGURE 3.4 Current in a metallic conductor. The magnitude of current density in a metal is the magnitude of charge contained in a cylinder of unit area and length v_d.

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***.

Because of the drift, there will be net transport of charges across any area perpendicular to **E**. Consider a planar area A, located inside the conductor such that the normal to the area is parallel to **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 $-neA|\mathrm{v}_d|\Delta t$. **E** is directed towards the left and hence the total charge transported along **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,

$$I\,\Delta t = +neA\,|\mathrm{v}_d|\,\Delta t \tag{3.18}$$

Substituting the value of $|\mathbf{v}_d|$ from Eq. (3.17)

$$I\Delta t = \frac{e^2 A}{m}\tau n\,\Delta t\,|\mathrm{E}| \tag{3.19}$$

By definition I is related to the magnitude $|\mathrm{j}|$ of the current density by

$$I = |\mathrm{j}|A \tag{3.20}$$

Hence, from Eqs.(3.19) and (3.20),

$$|\mathrm{j}| = \frac{ne^2}{m}\tau|\mathrm{E}| \tag{3.21}$$

The vector **j** is parallel to **E** and hence we can write Eq. (3.21) in the vector form

$$\mathbf{j} = \frac{ne^2}{m}\tau\mathbf{E} \tag{3.22}$$

Comparison with Eq. (3.13) shows that Eq. (3.22) is exactly the Ohm's law, if we identify the conductivity σ as

$$\sigma = \frac{ne^2}{m}\tau \tag{3.23}$$

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,

$$n = \frac{6.0 \times 10^{23}}{63.5} \times 9.0 \times 10^6$$

$$= 8.5 \times 10^{28} \text{ m}^{-3}$$

which gives,

$$v_d = \frac{1.5}{8.5 \times 10^{28} \times 1.6 \times 10^{-19} \times 1.0 \times 10^{-7}}$$

$$= 1.1 \times 10^{-3} \text{ m s}^{-1} = 1.1 \text{ mm s}^{-1}$$

(b) (i) At a temperature T, the thermal speed* of a copper atom of mass M is obtained from [$<(1/2)\, Mv^2> = (3/2)\, k_B T$] 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. (12.23) of Chapter 12 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* μ defined as the magnitude of the drift velocity per unit electric field:

$$\mu = \frac{|\mathbf{v}_d|}{E} \qquad (3.24)$$

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,

$$\mu = \frac{v_d}{E} = \frac{e\tau}{m} \tag{3.25}$$

where τ is the average collision time for electrons.

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.

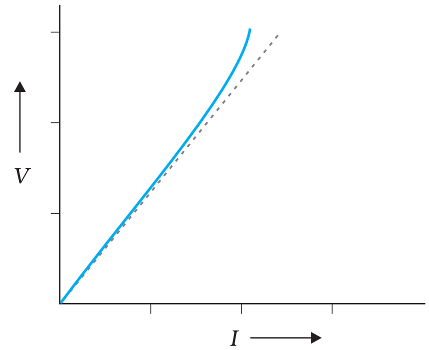


FIGURE 3.5 The dashed line represents the linear Ohm's law. The solid line is the voltage V versus current I for a good conductor.

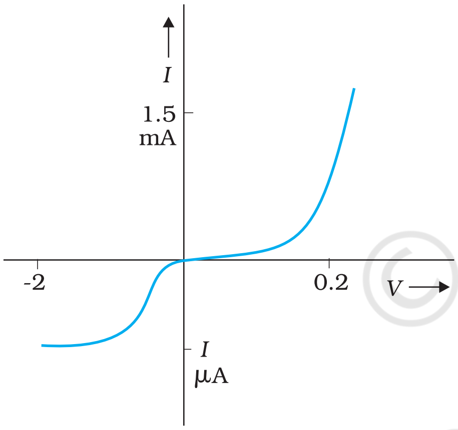


FIGURE 3.6 Characteristic curve of a diode. Note the different scales for negative and positive values of the voltage and current.

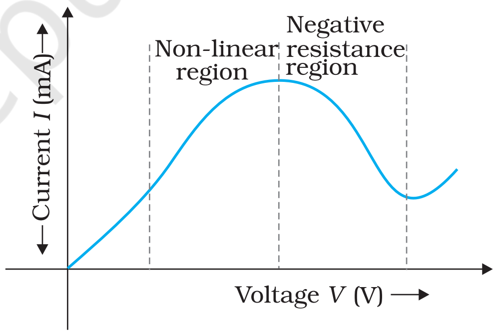


FIGURE 3.7 Variation of current versus voltage for GaAs.

(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 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 can be decreased by adding small amount of suitable impurities. This last feature is exploited in use of semiconductors for electronic devices.

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,

$$\rho_T = \rho_0 \left[1 + \alpha \left(T - T_0\right)\right] \tag{3.26}$$

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.

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. 3.8).

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.

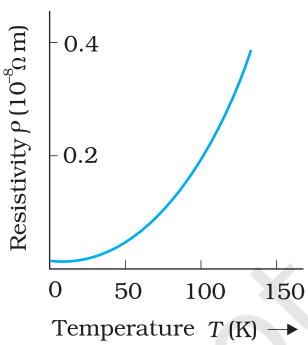


FIGURE 3.8 Resistivity ρ_T of copper as a function of temperature T.

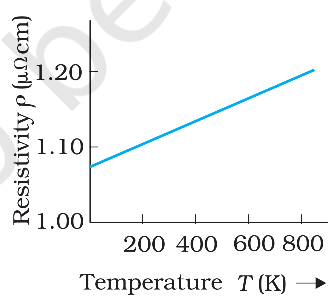


FIGURE 3.9 Resistivity ρ_T of nichrome as a function of absolute temperature T.

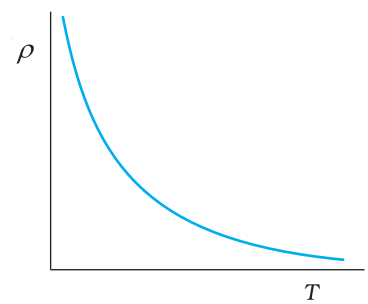


FIGURE 3.10 Temperature dependence of resistivity for a typical semiconductor.

Some materials like Nichrome (which is an alloy of nickel, iron and chromium) exhibit a very weak dependence of resistivity with temperature (Fig. 3.9). 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. 3.10.

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

$$T_2 - T_1 = \frac{(85.8 - 75.3)}{(75.3) \times 1.70 \times 10^{-4}} = 820\ ^\circ\text{C}$$

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 Ω, R_{100} = 5.23 Ω and R_t = 5.795 Ω

Now, $$t = \frac{R_t - R_0}{R_{100} - R_0} \times 100, \qquad R_t = R_0 (1 + \alpha t)$$

$$= \frac{5.795 - 5}{5.23 - 5} \times 100$$

$$= \frac{0.795}{0.23} \times 100 = 345.65\ °\text{C}$$

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, $V(\text{A}) > V(\text{B})$ and the potential difference across AB is $V = V(\text{A}) - V(\text{B}) > 0$.

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

$$\Delta U_{pot} = \text{Final potential energy} - \text{Initial potential energy}$$
$$= \Delta Q[(V(\text{B}) - V(\text{A})] = -\Delta Q\ V$$
$$= -I\ V\Delta t < 0 \tag{3.28}$$

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,

$$\Delta K = -\Delta U_{pot} \tag{3.29}$$

that is,

$$\Delta K = I\ V\Delta t > 0 \tag{3.30}$$

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,

$$\Delta W = I\ V\Delta t \tag{3.31}$$

The energy dissipated per unit time is the power dissipated $P = \Delta W/\Delta t$ and we have,

$$P = I\ V \tag{3.32}$$

Using Ohm's law $V = IR$, we get

$$P = I^2 R = V^2/R \tag{3.33}$$

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.

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.3.11), it is the chemical energy of the cell which supplies this power for as long as it can.

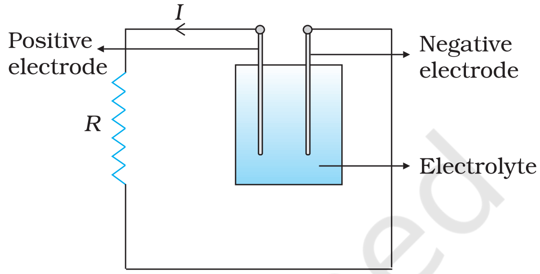


FIGURE 3.11 Heat is produced in the resistor R which is connected across the terminals of a cell. The energy dissipated in the resistor R comes from the chemical energy of the electrolyte.

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

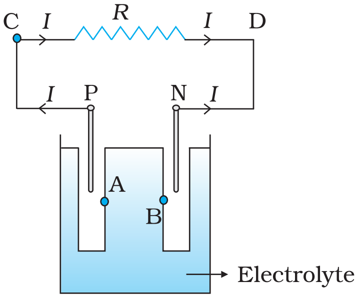
$$P_c = I^2 R_c$$

$$= \frac{P^2 R_c}{V^2} \tag{3.35}$$

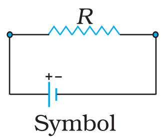
from Eq. (3.32). Thus, to drive a device of power P, the power wasted in the connecting wires is inversely proportional to V^2. The transmission cables from power stations are hundreds of miles long and their resistance R_c is considerable. To reduce P_c, these wires carry current at enormous values 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 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



(a)



(b)

FIGURE 3.12 (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.

Fig. 3.12. 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

$$\varepsilon = V_+ + V_- > 0 \tag{3.36}$$

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. 3.12). 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,

$$\begin{aligned} V &= \text{Potential difference between P and A} \\ &\quad + \text{Potential difference between A and B} \\ &\quad + \text{Potential difference between B and N} \\ &= \varepsilon \end{aligned} \tag{3.37}$$

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

$$\begin{aligned} V &= V_+ + V_- - I\,r \\ &= \varepsilon - I\,r \end{aligned} \tag{3.38}$$

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

$$V = I\,R \tag{3.39}$$

Combining Eqs. (3.38) and (3.39), we get

$$I\,R = \varepsilon - I\,r$$

$$\text{Or,}\ I = \frac{\varepsilon}{R + r} \qquad (3.40)$$

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.

3.11 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.

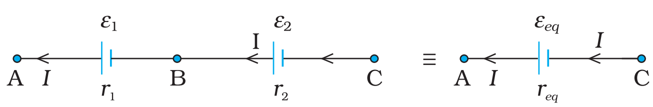


FIGURE 3.13 Two cells of emf's ε_1 and ε_2 in the series. r_1, r_2 are their internal resistances. For connections across A and C, the combination can be considered as one cell of emf ε_{eq} and an internal resistance r_{eq}.

Consider first two cells in series (Fig. 3.13), 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. 3.13. 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. (3.38) and hence,

$$V_{AB} \equiv V(A) \quad V(B) = \varepsilon_1 \quad I r_1 \qquad (3.41)$$

Similarly,

$$V_{BC} \equiv V(B) \quad V(C) = \varepsilon_2 \quad I r_2 \qquad (3.42)$$

Hence, the potential difference between the terminals A and C of the combination is

$$V_{AC} \equiv V(A) - V(C) = \left[V(A) - V(B)\right] + \left[V(B) - V(C)\right]$$

$$= (\varepsilon_1 + \varepsilon_2) - I(r_1 + r_2) \qquad (3.43)$$

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

$$V_{AC} = \varepsilon_{eq} - I\,r_{eq} \qquad (3.44)$$

Comparing the last two equations, we get

$$\varepsilon_{eq} = \varepsilon_1 + \varepsilon_2 \qquad (3.45)$$

$$\text{and}\quad r_{eq} = r_1 + r_2 \qquad (3.46)$$

In Fig.3.13, we had connected the negative electrode of the first to the positive electrode of the second. If instead we connect the two negatives, Eq. (3.42) would change to $V_{BC} = -\varepsilon_2 - Ir_2$ and we will get

$$\varepsilon_{eq} = \varepsilon_1 - \varepsilon_2 \qquad (\varepsilon_1 > \varepsilon_2) \tag{3.47}$$

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.

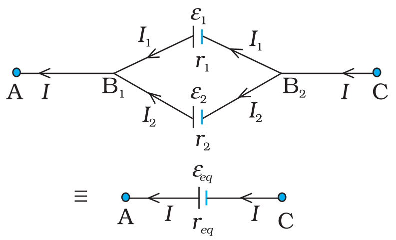


FIGURE 3.14 Two cells in parallel. For connections across A and C, the combination can be replaced by one cell of emf ε_{eq} and internal resistances r_{eq} whose values are given in Eqs. (3.54) and (3.55).

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. (3.47).

Next, consider a parallel combination of the cells (Fig. 3.14). 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

$$I = I_1 + I_2 \tag{3.48}$$

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. (3.38)

$$V \equiv V(B_1) - V(B_2) = \varepsilon_1 - I_1 r_1 \tag{3.49}$$

Points B_1 and B_2 are connected exactly similarly to the second cell. Hence considering the second cell, we also have

$$V \equiv V(B_1) - V(B_2) = \varepsilon_2 - I_2 r_2 \tag{3.50}$$

Combining the last three equations

$$I = I_1 + I_2$$

$$= \frac{\varepsilon_1 - V}{r_1} + \frac{\varepsilon_2 - V}{r_2} = \left(\frac{\varepsilon_1}{r_1} + \frac{\varepsilon_2}{r_2}\right) - V\left(\frac{1}{r_1} + \frac{1}{r_2}\right) \tag{3.51}$$

Hence, V is given by,

$$V = \frac{\varepsilon_1 r_2 + \varepsilon_2 r_1}{r_1 + r_2} - I\frac{r_1 r_2}{r_1 + r_2} \tag{3.52}$$

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

$$V = \varepsilon_{eq} - I\, r_{eq} \tag{3.53}$$

The last two equations should be the same and hence

$$\varepsilon_{eq} = \frac{\varepsilon_1 r_2 + \varepsilon_2 r_1}{r_1 + r_2} \tag{3.54}$$

$$r_{eq} = \frac{r_1 r_2}{r_1 + r_2} \tag{3.55}$$

We can put these equations in a simpler way,

$$\frac{1}{r_{eq}} = \frac{1}{r_1} + \frac{1}{r_2} \tag{3.56}$$

$$\frac{\varepsilon_{eq}}{r_{eq}} = \frac{\varepsilon_1}{r_1} + \frac{\varepsilon_2}{r_2} \tag{3.57}$$

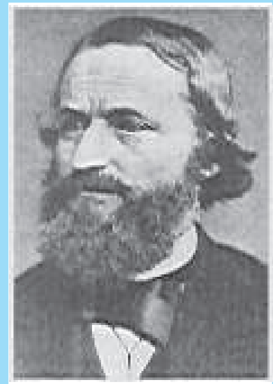
In Fig. (3.14), we had joined the positive terminals together and similarly the two negative ones, so that the currents I_1, I_2 flow out of positive terminals. If the negative terminal of the second is connected to positive terminal of the first, Eqs. (3.56) and (3.57) would still be valid with $\varepsilon_2 \to -\varepsilon_2$

Equations (3.56) and (3.57) can be extended easily. If there are n cells of emf ε_1, . . . ε_n and of internal resistances r_1,... 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

$$\frac{1}{r_{eq}} = \frac{1}{r_1} + ... + \frac{1}{r_n} \tag{3.58}$$

$$\frac{\varepsilon_{eq}}{r_{eq}} = \frac{\varepsilon_1}{r_1} + ... + \frac{\varepsilon_n}{r_n} \tag{3.59}$$

GUSTAV ROBERT KIRCHHOFF (1824 –



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.

3.12 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. (3.38) 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

$$V = \varepsilon + I\,r \tag{3.60}$$

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. 3.15).

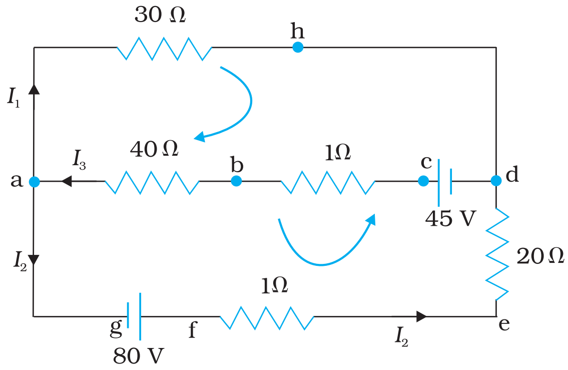


FIGURE 3.15 At junction a the current leaving is $I_1 + I_2$ and current entering is I_3. The junction rule says $I_3 = I_1 + I_2$. At point h current entering is I_1. There is only one current leaving h and by junction rule that will also be I_1. For the loops 'ahdcba' and 'ahdefga', the loop rules give $-30I_1 - 41\ I_3 + 45 = 0$ and $-30I_1 + 21\ I_2 - 80 = 0$.

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. 3.15).

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 3.5 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. 3.16). Determine the equivalent resistance of the network and the current along each edge of the cube.

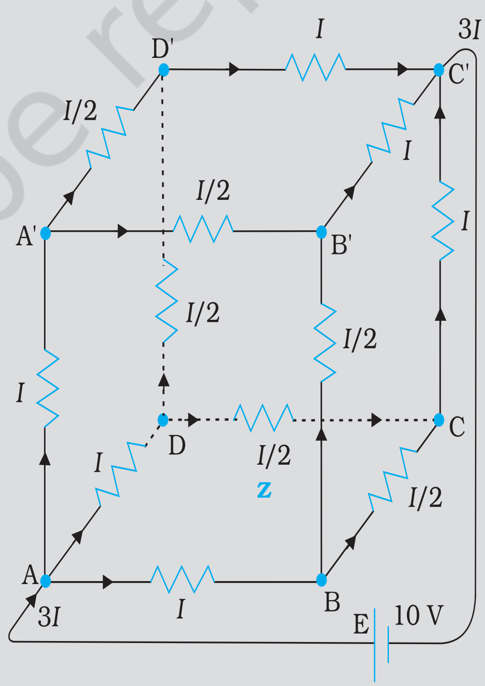


FIGURE 3.16

EXAMPLE 3.5

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 Ω, R_{eq} = (5/6) Ω and for ε = 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. 3.16.

EXAMPLE 3.5

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 3.5. 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 3.6.

Example 3.6 Determine the current in each branch of the network shown in Fig. 3.17.

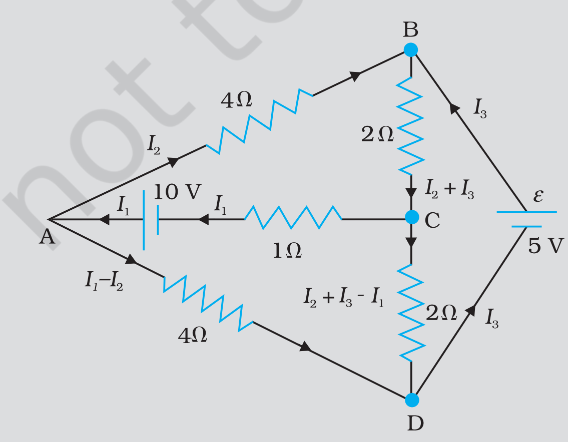


FIGURE 3.17

EXAMPLE 3.6

PHYSICS

Similation for application of Kirchhoff's rules:
http://www.phys.hawaii.edu/~teb/optics/java/kirch3/

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,

$$10 - 4(I_1 - I_2) + 2(I_2 + I_3 - I_1) - I_1 = 0 \qquad [3.61(a)]$$

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, $I_1 + 6I_2 + 2I_3 = 10$ [3.61(b)]

For the closed loop BCDEB, we get

$$5 - 2\,(I_2 + I_3) - 2\,(I_2 + I_3 - I_1) = 0$$

that is, $2I_1 - 4I_2 - 4I_3 = -5$ [3.61(c)]

Equations (3.61 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

$$5\ \text{V} + \left(\frac{5}{8}\times 4\right)\text{V} - \left(\frac{15}{8}\times 4\right)\text{V}$$

equal to zero, as required by Kirchhoff's second rule.

EXAMPLE 3.6

3.13 WHEATSTONE BRIDGE

As an application of Kirchhoff's rules consider the circuit shown in Fig. 3.18, 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

$$-I_1 R_1 + 0 + I_2 R_2 = 0 \qquad (I_g = 0) \tag{3.62}$$

and the second loop gives, upon using $I_3 = I_1$, $I_4 = I_2$

$$I_2 R_4 + 0 - I_1 R_3 = 0 \tag{3.63}$$

From Eq. (3.62), we obtain,

$$\frac{I_1}{I_2} = \frac{R_2}{R_1}$$

whereas from Eq. (3.63), we obtain,

$$\frac{I_1}{I_2} = \frac{R_4}{R_3}$$

Hence, we obtain the condition

$$\frac{R_2}{R_1} = \frac{R_4}{R_3} \qquad [3.64(a)]$$

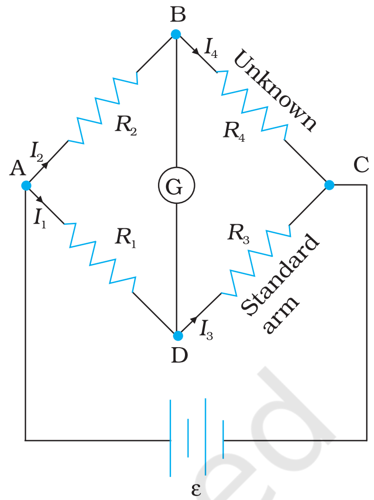


FIGURE 3.18

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,

$$R_4 = R_3 \frac{R_2}{R_1} \qquad [3.64(b)]$$

A practical device using this principle is called the *meter bridge*.

Example 3.7 The four arms of a Wheatstone bridge (Fig. 3.19) have the following resistances:

AB = 100Ω, BC = 10Ω, CD = 5Ω, and DA = 60Ω.

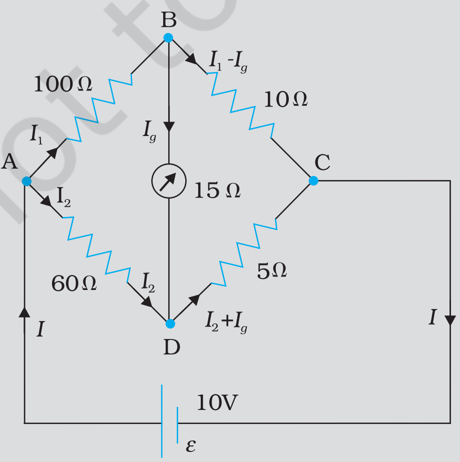


FIGURE 3.19

EXAMPLE 3.7

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 $20I_1 + 3I_g - 12I_2 = 0$ [3.65(a)]

Considering the mesh BCDB, we have

$$10\,(I_1 - I_g) - 15I_g - 5\,(I_2 + I_g) = 0$$

$$10I_1 - 30I_g - 5I_2 = 0$$

$$2I_1 - 6I_g - I_2 = 0 \qquad [3.65(b)]$$

Considering the mesh ADCEA,

$$60I_2 + 5\,(I_2 + I_g) = 10$$

$$65I_2 + 5I_g = 10$$

$$13I_2 + I_g = 2 \qquad [3.65(c)]$$

Multiplying Eq. (3.65b) by 10

$$20I_1 - 60I_g - 10I_2 = 0 \qquad [3.65(d)]$$

From Eqs. (3.65d) and (3.65a) we have

$$63I_g - 2I_2 = 0$$

$$I_2 = 31.5I_g \qquad [3.65(e)]$$

Substituting the value of I_2 into Eq. [3.65(c)], we get

$$13\,(31.5I_g) + I_g = 2$$

$$410.5\,I_g = 2$$

$$I_g = 4.87 \text{ mA.}$$

EXAMPLE 3.7

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 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} Ω m to 10^{-6} Ω 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* **j** gives the amount of charge flowing per second per unit area normal to the flow,

$$\mathbf{j} = nq\,\boldsymbol{v}_d$$

where n is the number density (number per unit volume) of charge carriers each of charge q, and $\boldsymbol{v}_d$ is the *drift velocity* of the charge carriers. For electrons $q = -e$. If **j** is normal to a cross-sectional area **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. *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.
13. 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.

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	$[\mathrm{M\,L^2\,T^{-3}\,A^{-1}}]$	V	Work/charge
Electromotive force	ε	$[\mathrm{M\,L^2\,T^{-3}\,A^{-1}}]$	V	Work/charge
Resistance	R	$[\mathrm{M\,L^2\,T^{-3}\,A^{-2}}]$	Ω	$R = V/I$
Resistivity	ρ	$[\mathrm{M\,L^3\,T^{-3}\,A^{-2}}]$	Ω m	$R = \rho l/A$
Electrical conductivity	σ	$[\mathrm{M^{-1}\,L^{-3}\,T^3\,A^2}]$	S	$\sigma = 1/\rho$
Electric field	$\mathbf{E}$	$[\mathrm{M\,L\,T^{-3}\,A^{-1}}]$	$\mathrm{V\,m^{-1}}$	$\dfrac{\text{Electric force}}{\text{charge}}$
Drift speed	v_d	$[\mathrm{L\,T^{-1}}]$	$\mathrm{m\,s^{-1}}$	$v_d = \dfrac{e\,E\,\tau}{m}$
Relaxation time	τ	[T]	s	
Current density	$\mathbf{j}$	$[\mathrm{L^{-2}\,A}]$	$\mathrm{A\,m^{-2}}$	current/area
Mobility	μ	$[\mathrm{M\,L^3\,T^{-4}\,A^{-1}}]$	$\mathrm{m^2\,V^{-1}s^{-1}}$	$v_d\,/E$

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 $\mathbf{j}$ and $\Delta\mathbf{S}$ 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 $\mathbf{E}$ is the sum of (i) motion due to random collisions and (ii) that due to $\mathbf{E}$. The motion due to random collisions averages to zero and does not contribute to v_d (Chapter 10, 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 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}.

3.4 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?

3.5 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.

3.6 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}.

3.7 Determine the current in each branch of the network shown in Fig. 3.20:

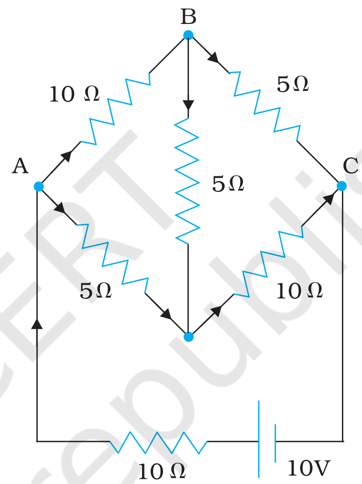


FIGURE 3.20

3.8 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?

3.9 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.

12089CH04

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, 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

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 took place in the 20th century. This was due to our increased understanding of electromagnetism and the invention of devices for production, amplification, transmission and detection of electromagnetic waves.

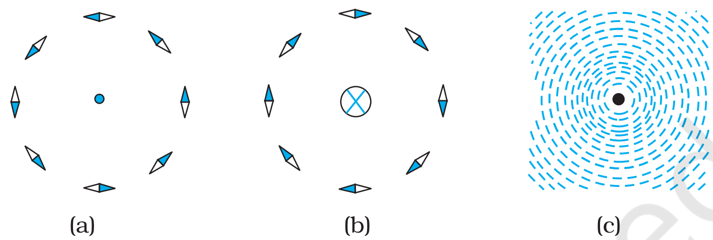


FIGURE 4.1 The magnetic field due to a straight long current-carrying wire. The wire is perpendicular to the plane of the paper. A ring of compass needles surrounds the wire. The orientation of the needles is shown when (a) the current emerges out of the plane of the paper, (b) the current moves into 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.

HANS CHRISTIAN OERSTED (1777–1851)

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.

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.

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.

$$\mathbf{E} = Q\ \hat{\mathbf{r}} / (4\pi\varepsilon_0)r^2 \qquad (4.1)$$

where $\hat{\mathbf{r}}$ is unit vector along **r**, and the field **E** is a vector field. A charge q interacts with this field and experiences a force **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 **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 **B (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.*

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.

HENDRIK ANTOON LORENTZ (1853 – 1928)

4.2.2 Magnetic Field, Lorentz Force

Let us suppose that there is a point charge q (moving with a velocity **v** and, located at **r** at a given time t) in presence of both the electric field **E (r)** and the magnetic field **B (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, **v** and **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.

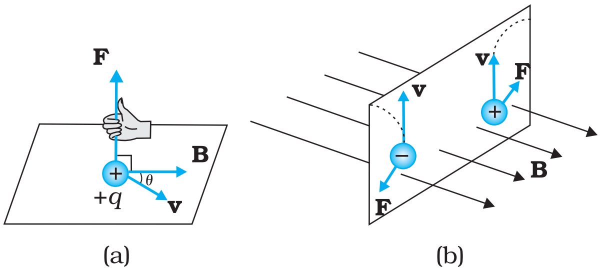


FIGURE 4.2 The direction of the magnetic force acting on a charged particle. (a) The force on a positively charged particle with velocity **v** and making an angle θ with the magnetic field **B** is given by the right-hand rule. (b) A moving charged particle q is deflected in an opposite sense to $-q$ in the presence of magnetic field.

(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.

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 nlA. 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:

$$\mathbf{F} = (nlA)q\,\mathbf{v}_d \times \mathbf{B}$$

where q is the value of the charge on a carrier. Now $nq\,\mathbf{v}_d$ is the current density **j** and $|(nq\,\mathbf{v}_d)|A$ is the current I (see Chapter 3 for the discussion of current and current density). Thus,

$$\mathbf{F} = [(nq\,\mathbf{v}_d)lA] \times \mathbf{B} = [\,\mathbf{j}Al\,] \times \mathbf{B}$$
$$= I\boldsymbol{l} \times \mathbf{B} \qquad (4.4)$$

where $\boldsymbol{l}$ 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 **j** to $\boldsymbol{l}$.

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 $d\boldsymbol{l}_j$ and summing

$$\mathbf{F} = \sum_j \mathrm{I}d\boldsymbol{l}_j \times \mathbf{B}$$

This summation can be converted to an integral in most cases.

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?

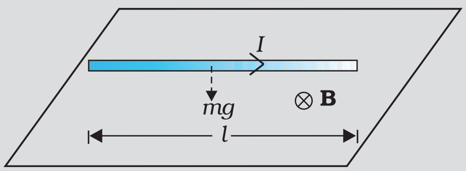


FIGURE 4.3

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:

$$m\,g = I\,lB$$

$$B = \frac{m\,g}{I\,l}$$

$$= \frac{0.2 \times 9.8}{2 \times 1.5} = 0.65 \text{ T}$$

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).

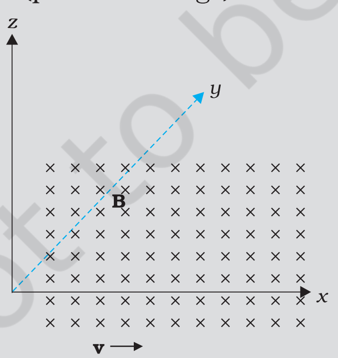


FIGURE 4.4

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 5) 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 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 $\mathbf{v}$ perpendicular to $\mathbf{B}$. 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* $\mathbf{v}$ *and* $\mathbf{B}$ *are perpendicular to each other* (Fig. 4.5).

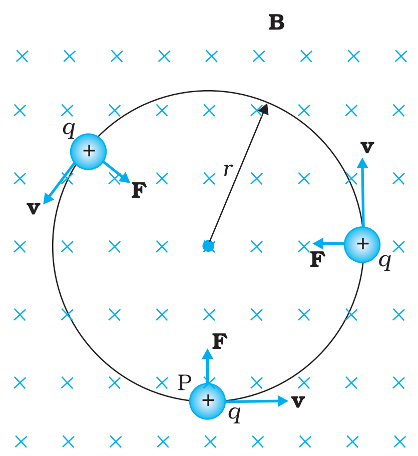


FIGURE 4.5 Circular motion

If velocity has a component along $\mathbf{B}$, 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 $\mathbf{B}$ is as before a circular one, thereby producing a *helical motion* (Fig. 4.6).

You have already learnt in earlier classes (See Class XI, Chapter 3) 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 $\mathbf{v}$ is perpendicular to the magnetic field $\mathbf{B}$, the magnetic force is perpendicular to both $\mathbf{v}$ and $\mathbf{B}$ and acts like a centripetal force. It has a magnitude $q\,v\,B$. Equating the two expressions for centripetal force,

$$m\,v^2/r = q\,v\,B, \text{ which gives}$$

$$r = m\,v / qB \tag{4.5}$$

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).

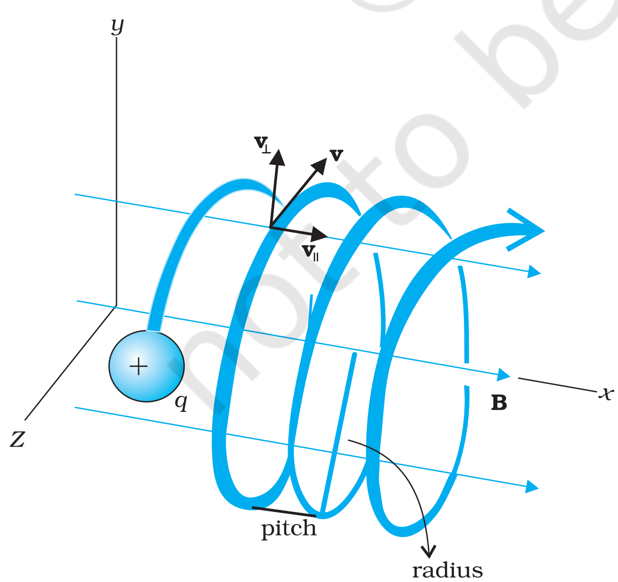


FIGURE 4.6 Helical motion

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

$$p = v_{||} T = 2\pi m v_{||} / qB \qquad [4.6(b)]$$

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

$r = mv / (qB) = 9 \times 10^{-31}$ kg $\times$ **3** $\times 10^{7}$ m s^{-1} / (1.6×10^{-19} C $\times 6 \times 10^{-4}$ T)

$= 28 \times 10^{-2}$ m = 28 cm

$\nu = v / (2\pi r) = 17 \times 10^{6}$ s^{-1} = 17×10^{6} Hz = 17 MHz.

$E = (½)mv^2 = (½)\, 9 \times 10^{-31}$ kg $\times 9 \times 10^{14}$ m^2/s^2 = 40.5×10^{-17} J

$\approx 4 \times 10^{-16}$ J = 2.5 keV.

EXAMPLE 4.3

4.4 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. Fig. 4.7 shows a finite conductor XY carrying current I. Consider an infinitesimal element dl of the conductor. The magnetic field dB 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$\boldsymbol{l}$ and the displacement vector **r**. According to Biot-Savart's law, the magnitude of the magnetic field dB is proportional to the current I, the element length |d$\boldsymbol{l}$|, and inversely proportional to the square of the distance r. Its direction* is perpendicular to the plane containing d$\boldsymbol{l}$ and **r**. Thus, in vector notation,

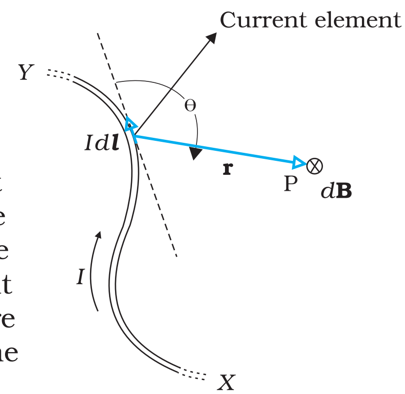


FIGURE 4.7 Illustration of the Biot-Savart law. The current element I d$\boldsymbol{l}$ produces a field dB at a distance r. The ⊗ sign indicates that the field is perpendicular to the plane of this page and directed into it.

$$d\mathbf{B} \propto \frac{I\, d\boldsymbol{l} \times \mathbf{r}}{r^3}$$

$$= \frac{\mu_0}{4\pi} \frac{I\, d\boldsymbol{l} \times \mathbf{r}}{r^3} \qquad [4.7(a)]$$

where $\mu_0/4\pi$ is a constant of proportionality. The above expression holds when the medium is vacuum.

* The sense of d$\boldsymbol{l}$ × **r** is also given by the *Right Hand Screw rule* : Look at the plane containing vectors d$\boldsymbol{l}$ 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.

The magnitude of this field is,

$$|d\mathbf{B}| = \frac{\mu_0}{4\pi}\frac{I\,dl\sin\theta}{r^2} \qquad [4.7(b)]$$

where we have used the property of cross-product. Equation [4.7 (a)] constitutes our basic equation for the magnetic field. The proportionality constant in SI units has the exact value,

$$\frac{\mu_0}{4\pi} = 10^{-7}\ \mathrm{Tm/A} \qquad [4.7(c)]$$

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* $I\,d\boldsymbol{l}$ 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 $I\,d\boldsymbol{l}$.

(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 $\mathbf{r}$ and the current element $I\,d\boldsymbol{l}$.

(iv) There is an angle dependence in the Biot-Savart law which is not present in the electrostatic case. In Fig. 4.7, the magnetic field at any point in the direction of $d\boldsymbol{l}$ (the dashed line) is zero. Along this line, $\theta = 0$, $\sin\theta = 0$ and from Eq. [4.7(a)], $|d\mathbf{B}| = 0$.

There is an interesting relation between ε_0, the permittivity of free space; μ_0, the permeability of free space; and c, the speed of light in vacuum:

$$\varepsilon_0\mu_0 = (4\pi\varepsilon_0)\frac{\mu_0}{4\pi} = \frac{1}{9\times10^9}\left(10^{-7}\right) = \frac{1}{(3\times10^8)^2} = \frac{1}{c^2}$$

We will discuss this connection further in Chapter 8 on the electromagnetic waves. Since the speed of light in vacuum is constant, the product $\mu_0\varepsilon_0$ is fixed in magnitude. Choosing the value of either ε_0 or μ_0, fixes the value of the other. In SI units, μ_0 is fixed to be equal to $4\pi \times 10^{-7}$ in magnitude.

EXAMPLE 4.4

Example 4.4 An element $\Delta\boldsymbol{l} = \Delta x\,\hat{\mathbf{i}}$ is placed at the origin and carries a large current I = 10 A (Fig. 4.8). What is the magnetic field on the y-axis at a distance of 0.5 m. Δx = 1 cm.

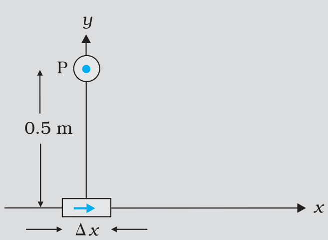


FIGURE 4.8

Solution

$$|\,\mathrm{d}\mathbf{B}\,| = \frac{\mu_0}{4\pi}\frac{I\,\mathrm{d}l \sin\theta}{r^2} \quad \text{[using Eq. (4.7)]}$$

$\mathrm{d}l = \Delta x = 10^{-2}\,\mathrm{m}$, $I = 10$ A, $r = 0.5$ m $= y$, $\mu_0/4\pi = 10^{-7}\,\dfrac{\mathrm{T\,m}}{\mathrm{A}}$

$\theta = 90°$; $\sin\theta = 1$

$$|\mathrm{d}\mathbf{B}| = \frac{10^{-7} \times 10 \times 10^{-2}}{25 \times 10^{-2}} = 4 \times 10^{-8}\,\mathrm{T}$$

The direction of the field is in the +z-direction. This is so since,

$$\mathrm{d}\boldsymbol{l} \times \mathbf{r} = \Delta x\,\hat{\mathbf{i}} \times y\,\hat{\mathbf{j}} = y\,\Delta x\left(\hat{\mathbf{i}} \times \hat{\mathbf{j}}\right) = y\,\Delta x\,\hat{\mathbf{k}}$$

We remind you of the following cyclic property of cross-products,

$$\hat{\mathbf{i}} \times \hat{\mathbf{j}} = \hat{\mathbf{k}};\ \hat{\mathbf{j}} \times \hat{\mathbf{k}} = \hat{\mathbf{i}};\ \hat{\mathbf{k}} \times \hat{\mathbf{i}} = \hat{\mathbf{j}}$$

Note that the field is small in magnitude.

EXAMPLE 4.4

In the next section, we shall use the Biot-Savart law to calculate the magnetic field due to a circular loop.

4.5 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 ($I\,\mathrm{d}\boldsymbol{l}$) 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).

Fig. 4.9 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.

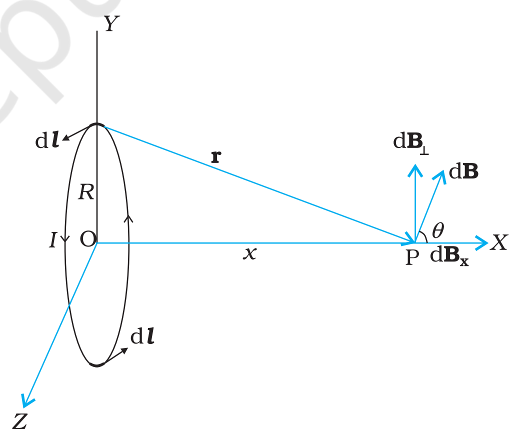


FIGURE 4.9 Magnetic field on the axis of a current carrying circular loop of radius R. Shown are the magnetic field $d\mathbf{B}$ (due to a line element $d\boldsymbol{l}$) and its components along and perpendicular to the axis.

Consider a conducting element $\mathrm{d}\boldsymbol{l}$ of the loop. This is shown in Fig. 4.9. The magnitude $\mathrm{d}B$ of the magnetic field due to $\mathrm{d}\boldsymbol{l}$ is given by the Biot-Savart law [Eq. 4.7(a)],

$$dB = \frac{\mu_0}{4\pi}\frac{I\,|\mathrm{d}\boldsymbol{l} \times \mathbf{r}|}{r^3} \tag{4.8}$$

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 $\mathrm{d}\boldsymbol{l}$ in Fig. 4.9 is in the y-z plane, whereas, the displacement vector $\mathbf{r}$ from $\mathrm{d}\boldsymbol{l}$ to the axial point P is in the x–y plane. Hence $|\,d\boldsymbol{l} \times \mathbf{r}\,| = r\,dl$. Thus,

$$\mathrm{d}B = \frac{\mu_0}{4\pi}\frac{I\mathrm{d}l}{\left(x^2 + R^2\right)} \tag{4.9}$$

The direction of dB is shown in Fig. 4.9. It is perpendicular to the plane formed by d$\boldsymbol{l}$ and **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$\boldsymbol{l}$ is cancelled by the contribution due to the diametrically opposite d$\boldsymbol{l}$ element, shown in Fig. 4.9. 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. 4.9,

$$\cos\theta = \frac{R}{(x^2+R^2)^{1/2}} \tag{4.10}$$

From Eqs. (4.9) and (4.10),

$$dB_x = \frac{\mu_0 I dl}{4\pi}\frac{R}{\left(x^2+R^2\right)^{3/2}}$$

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

$$\mathbf{B} = B_x\hat{\mathbf{i}} = \frac{\mu_0 I R^2}{2\left(x^2+R^2\right)^{3/2}}\hat{\mathbf{i}} \tag{4.11}$$

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,

$$\mathbf{B}_0 = \frac{\mu_0 I}{2R}\hat{\mathbf{i}} \tag{4.12}$$

The magnetic field lines due to a circular wire form closed loops and are shown in Fig. 4.10. 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.

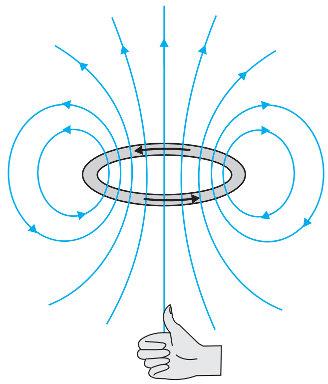


FIGURE 4.10 The magnetic field lines for a current loop. The direction of the field is given by the right-hand thumb rule described in the text. The upper side of the loop may be thought of as the north pole and the lower side as the south pole of a magnet.

Example 4.5 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. 4.11(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. 4.11(b)?

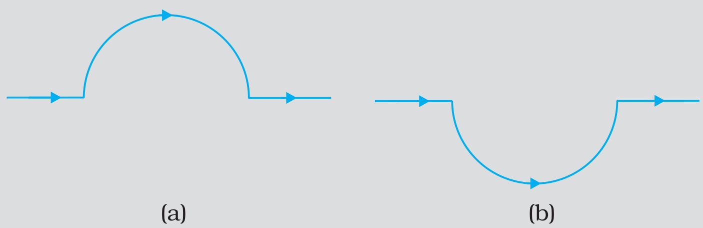


FIGURE 4.11

Solution

(a) d**l** and **r** for each element of the straight segments are parallel. Therefore, d$\boldsymbol{l}$ × **r** = 0. Straight segments do not contribute to |**B**|.

(b) For all segments of the semicircular arc, d$\boldsymbol{l}$ × **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 4.5

Example 4.6 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,

$$B = \frac{\mu_0 NI}{2R} = \frac{4\pi \times 10^{-7} \times 10^2 \times 1}{2 \times 10^{-1}} = 2\pi \times 10^{-4} = 6.28 \times 10^{-4} \text{ T}$$

EXAMPLE 4.6

4.6 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. 4.12). 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

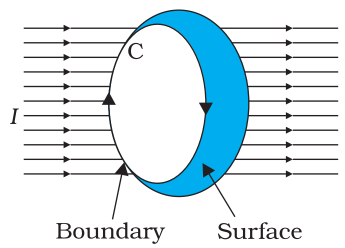


FIGURE 4.12

ANDRE AMPERE (1775 –1836)

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).

length of that element dl [Note: $B_t dl = \mathbf{B} \cdot d\boldsymbol{l}$]. 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 μ_0 times the total current passing through the surface, i.e.,

$$\oint \mathbf{B} \cdot d\boldsymbol{l} = \mu_0 I \qquad [4.13(a)]$$

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 $\oint \mathbf{B} \cdot d\boldsymbol{l}$. 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. [4.13(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*

(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. (4.13) reduces to,

$$BL = \mu_0 I_e \qquad [4.13(b)]$$

When there is a system with a symmetry such as for a *straight infinite current-carrying wire* in Fig. 4.13, 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. [4.13 (b)], B. $2\pi r$. We find that the magnetic field at a distance r outside the wire is *tangential* and given by

$$B \times 2\pi r = \mu_0 I,$$

$$B = \mu_0 I/(2\pi r) \qquad (4.14)$$

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 non-zero 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 4.7 Figure 4.13 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 $r > a$.

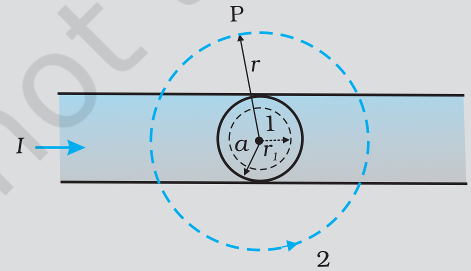


FIGURE 4.13

EXAMPLE 4.7

* 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.

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

$B(2\pi r) = \mu_0 I$

$$B = \frac{\mu_0 I}{2\pi r} \qquad [4.15(a)]$$

$$B \propto \frac{1}{r} \qquad (r > a)$$

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, $B(2\pi r) = \mu_0 \dfrac{I r^2}{a^2}$

$$B = \left(\frac{\mu_0 I}{2\pi a^2}\right) r \qquad [4.15(b)]$$

$$B \propto r \qquad (r < a)$$

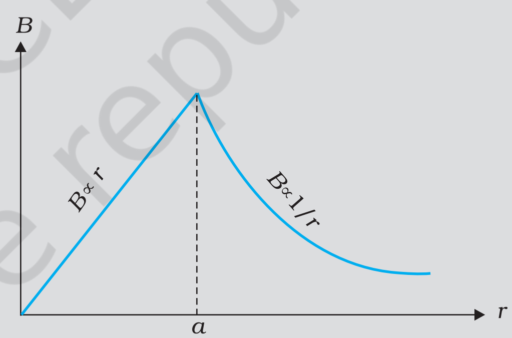


FIGURE 4.14

Figure (4.14) 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.

EXAMPLE 4.7

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 4.5, it cannot be applied to extract the simple expression $B = \mu_0 I/2R$ [Eq. (4.12)] 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*.

4.7 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.

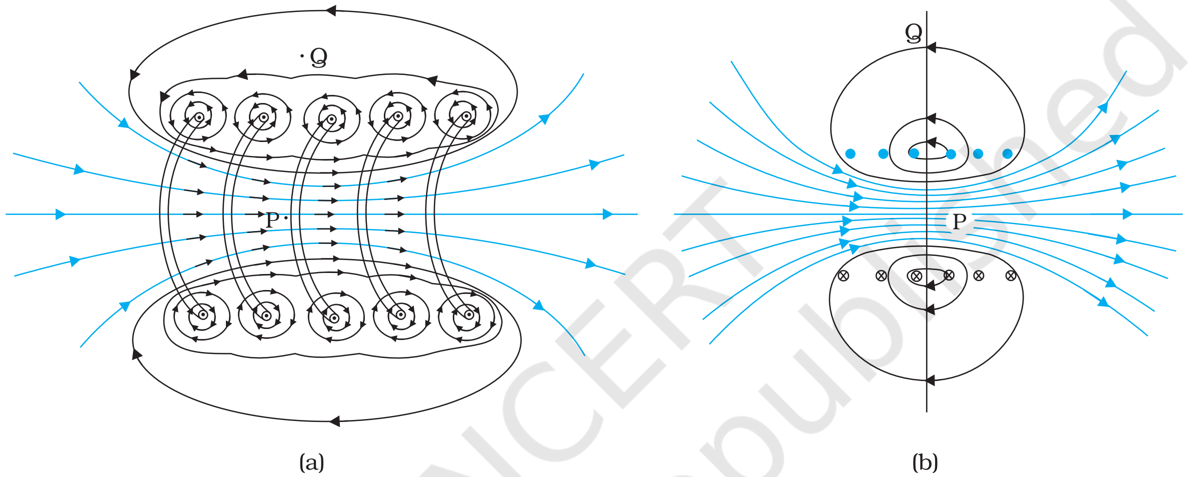


FIGURE 4.15 (a) The magnetic field due to a section of the solenoid which has been stretched out for clarity. Only the exterior semi-circular part is shown. Notice how the circular loops between neighbouring turns tend to cancel. (b) The magnetic field of a finite solenoid.

Figure 4.15 displays the magnetic field lines for a finite solenoid. We show a section of this solenoid in an enlarged manner in Fig. 4.15(a). Figure 4.15(b) shows the entire finite solenoid with its magnetic field. In Fig. 4.15(a), it is clear from the circular loops that the field between two neighbouring turns vanishes. In Fig. 4.15(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

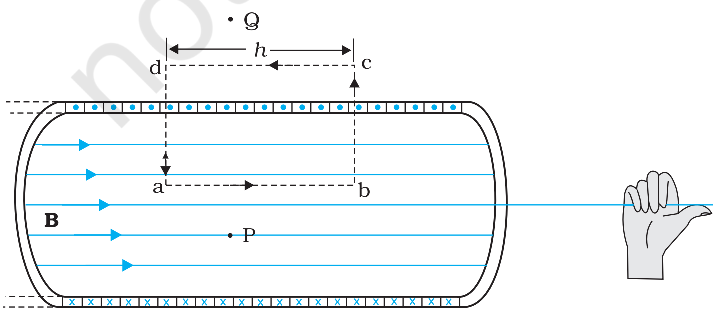


FIGURE 4.16 The magnetic field of a very long solenoid. We consider a rectangular Amperian loop abcd to determine the field.

solenoid is made longer it appears like a long cylindrical metal sheet. Figure 4.16 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.

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. 4.13 (b)]

$$BL = \mu_0 I_e, \quad B\,h = \mu_0 I\,(n\,h)$$

$$B = \mu_0\, n\, I \tag{4.16}$$

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.

EXAMPLE 4.8

Example 4.8 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)

$$B = \mu_0 n\, I$$
$$= 4\pi \times 10^{-7} \times 10^{3} \times 5$$
$$= 6.28 \times 10^{-3}\ \text{T}$$

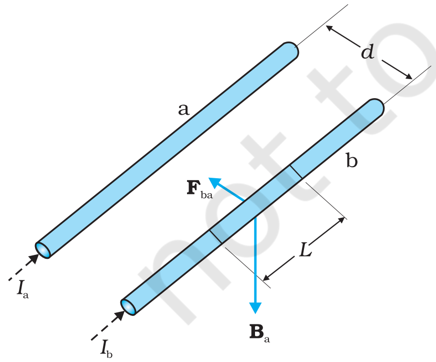


FIGURE 4.17 Two long straight parallel conductors carrying steady currents I_a and I_b and separated by a distance d. $\mathbf{B}_a$ is the magnetic field set up by conductor 'a' at conductor 'b'.

4.8 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 4.17 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. [4.15(a)] or from Ampere's circuital law,

$$B_a = \frac{\mu_0 I_a}{2\pi d}$$

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),

$$F_{ba} = I_b L B_a$$

$$= \frac{\mu_0 I_a I_b}{2\pi d} L \qquad (4.17)$$

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,

$$\mathbf{F}_{ba} = -\mathbf{F}_{ab} \qquad (4.18)$$

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. (4.17),

$$f_{ba} = \frac{\mu_0 I_a I_b}{2\pi d} \qquad (4.19)$$

The above expression is used to define the ampere (A), which is one of the seven SI base units.

---

* 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.

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).

Example 4.9 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,

$f = I\,B$

$= 1 \times 3 \times 10^{-5} = 3 \times 10^{-5}\ \mathrm{N\,m^{-1}}$

This is larger than the value $2\times10^{-7}\ \mathrm{Nm^{-1}}$ 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,

$\theta = 0°$

$f = 0$

Hence there is no force on the conductor.

EXAMPLE 4.9

4.9 Torque on Current Loop, Magnetic Dipole

4.9.1 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 1.11).

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. 4.18(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.

$$F_2 = I\,b\,B = F_1$$

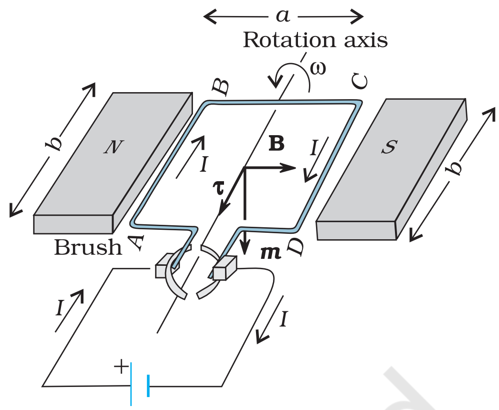
Thus, the *net force* on the loop *is zero*. There is a torque on the loop due to the pair of forces $\mathbf{F}_1$ and $\mathbf{F}_2$. Figure 4.18(b) shows a view of the loop from the AD end. It shows that the torque on the loop tends to rotate it anticlockwise. This torque is (in magnitude),

$$\tau = F_1\frac{a}{2} + F_2\frac{a}{2}$$

$$= IbB\frac{a}{2} + IbB\frac{a}{2} = I(ab)B$$

$$= I\,A\,B \qquad (4.20)$$

where $A = ab$ is the area of the rectangle.



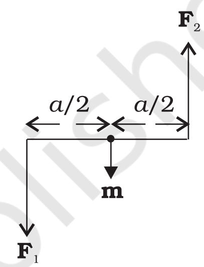


FIGURE 4.18 (a) A rectangular current-carrying coil in uniform magnetic field. The magnetic moment **m** points downwards. The torque **τ** is along the axis and tends to rotate the coil anticlockwise. (b) The couple acting on the coil.

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 the coil to be angle θ (The previous case corresponds to $\theta = \pi/2$). Figure 4.19 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 4.19(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,

$$\tau = F_1\frac{a}{2}\sin\theta + F_2\frac{a}{2}\sin\theta$$

$$= I\,ab\,B\sin\theta$$

$$= I\,A\,B\sin\theta \qquad (4.21)$$

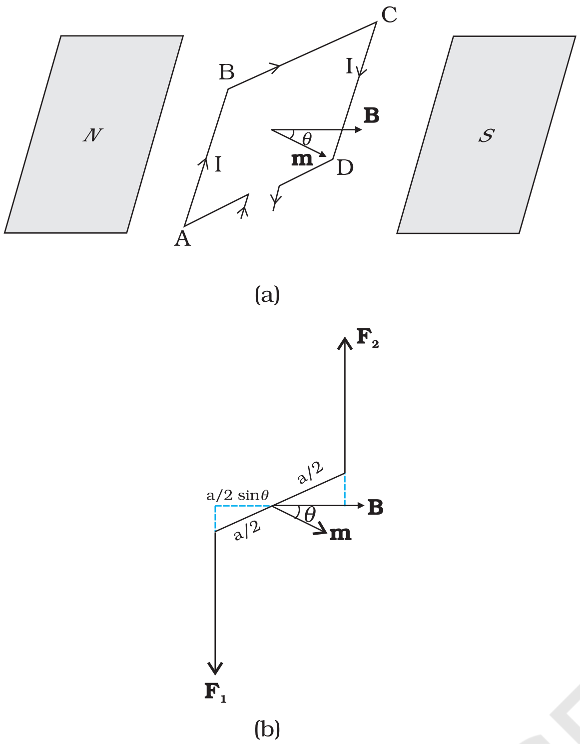


FIGURE 4.19 (a) The area vector of the loop ABCD makes an arbitrary angle θ with the magnetic field. (b) Top view of the loop. The forces $\mathbf{F_1}$ and $\mathbf{F_2}$ acting on the arms AB and CD are indicated.

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. (4.20) and (4.21) 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,

$$\mathbf{m} = I\,\mathbf{A} \tag{4.22}$$

where the direction of the area vector **A** is given by the right-hand thumb rule and is directed into the plane of the paper in Fig. 4.18. Then as the angle between **m** and **B** is θ, Eqs. (4.20) and (4.21) can be expressed by one expression

$$\boldsymbol{\tau} = \mathbf{m} \times \mathbf{B} \tag{4.23}$$

This is analogous to the electrostatic case (Electric dipole of dipole moment $\mathbf{p_e}$ in an electric field **E**).

$$\tau = \mathbf{p_e} \times \mathbf{E}$$

As is clear from Eq. (4.22), the dimensions of the magnetic moment are [A][L^2] and its unit is Am^2.

From Eq. (4.23), we see that the torque $\boldsymbol{\tau}$ vanishes when **m** is either parallel or antiparallel to the magnetic field **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 **m**). When **m** and **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. (4.23), still holds, with

$$\mathbf{m} = N\,I\,\mathbf{A} \tag{4.24}$$

EXAMPLE 4.10

Example 4.10 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. (4.12)

$$B = \frac{\mu_0 NI}{2R}$$

Here, N = 100; I = 3.2 A, and R = 0.1 m. Hence,

$$B = \frac{4\pi\times10^{-7}\times3.2}{2\times10^{-1}} = \frac{4\times10^{-5}\times10}{2\times10^{-1}} \quad \text{(using } \pi\times3.2 = 10\text{)}$$

$= 2 \times 10^{-3}$ T

The direction is given by the right-hand thumb rule.

(b) The magnetic moment is given by Eq. (4.24),

$m = N\,I\,A = N\,I\,\pi\,r^2 = 100 \times 3.2 \times 3.14 \times 10^{-2} = 10$ A m^2

The direction is once again given by the right-hand thumb rule.

(c) $\tau = |\mathbf{m} \times \mathbf{B}|$ [from Eq. (4.23)]

$= m\,B \sin\theta$

Initially, $\theta = 0$. Thus, initial torque $\tau_i = 0$. Finally, $\theta = \pi/2$ (or 90°). Thus, final torque $\tau_f = m\,B = 10 \times 2 = 20$ N m.

(d) From Newton's second law,

$$\mathcal{I}\frac{d\omega}{dt} = mB\sin\theta$$

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,

$$\mathcal{I}\,\omega\,d\omega = m\,B\sin\theta\,d\theta$$

Integrating from $\theta = 0$ to $\theta = \pi/2$,

$$\mathcal{I}\int_0^{\omega_f}\omega\,d\omega = mB\int_0^{\pi/2}\sin\theta\,d\theta$$

$$\mathcal{I}\frac{\omega_f^2}{2} = -m\,B\cos\theta\,\big|_0^{\pi/2} = m\,B$$

$$\omega_f = \left(\frac{2\,mB}{\mathcal{I}}\right)^{1/2} = \left(\frac{2\times20}{10^{-1}}\right)^{1/2} = 20\,\text{s}^{-1}$$

EXAMPLE 4.10

Example 4.11

(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

EXAMPLE 4.11

EXAMPLE 4.11

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 **A** of the horizontal loop is in the vertical direction, τ would be in the plane of the loop for any **B**.

(b) Orientation of stable equilibrium is one where the area vector **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 maximise flux, since for a given perimeter, a circle encloses greater area than any other shape.

4.9.2 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 behaviour to the electric field of an electric dipole. In Section 4.5, 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. (4.11)],

$$B = \frac{\mu_0 I R^2}{2\left(x^2 + R^2\right)^{3/2}}$$

and its direction is along the axis and given by the right-hand thumb rule (Fig. 4.10). 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,

$$B = \frac{\mu_0 I R^2}{2x^3}$$

Note that the area of the loop $A = \pi R^2$. Thus,

$$B = \frac{\mu_0 I A}{2\pi x^3}$$

As earlier, we define the magnetic moment **m** to have a magnitude IA, $\mathbf{m} = I\mathbf{A}$. Hence,

$$\mathrm{B} \simeq \frac{\mu_0 m}{2\pi x^3}$$

$$= \frac{\mu_0}{4\pi}\frac{2\mathbf{m}}{x^3} \qquad [4.25(a)]$$

The expression of Eq. [4.25(a)] is very similar to an expression obtained earlier for the electric field of a dipole. The similarity may be seen if we substitute,

$$\mu_0 \to 1/\varepsilon_0$$

$\mathbf{m} \rightarrow \mathbf{p}_e$ (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 1.9 [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)],

$$E \simeq \frac{\mathbf{p}_e}{4\pi\varepsilon_0 x^3}$$

where x is the distance from the dipole. If we replace $\mathbf{p} \rightarrow \mathbf{m}$ and $\mu_0 \rightarrow 1/\varepsilon_0$ in the above expression, we obtain the result for $\mathbf{B}$ for a point *in the plane of the loop* at a distance x from the centre. For $x >> R$,

$$\mathbf{B} \simeq \frac{\mu_0}{4\pi} \frac{\mathbf{m}}{x^3}; \qquad x >> R \qquad [4.25(b)]$$

The results given by Eqs. [4.25(a)] and [4.25(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 $\mathbf{p}$. 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. 4.10) 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 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 4.20 exhibits a very useful instrument for this purpose: the *moving coil galvanometer* (MCG). It is a device whose principle can be understood on the basis of our discussion in Section 4.9.

The galvanometer consists of a coil, with many turns, free to rotate about a fixed axis (Fig. 4.20), 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. (4.20) to be

$$\tau = NI\,AB$$

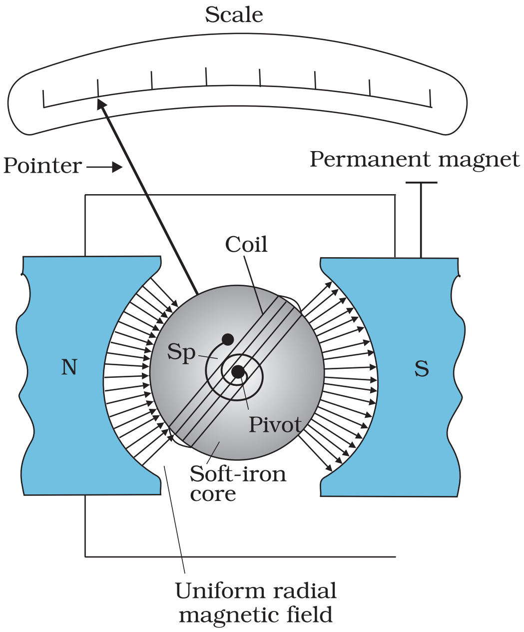


FIGURE 4.20 The moving coil galvanometer. Its elements are described in the text. Depending on the requirement, this device can be used as a current detector or for measuring the value of the current (ammeter) or voltage (voltmeter).

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

$$\phi = \left(\frac{NAB}{k}\right)I \qquad (4.26)$$

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.4.20. Depending on the direction of the current, the pointer's 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 μA. (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,

$$R_G\, r_s / (R_G + r_s) \simeq r_s \qquad \text{if} \quad R_G >> r_s$$

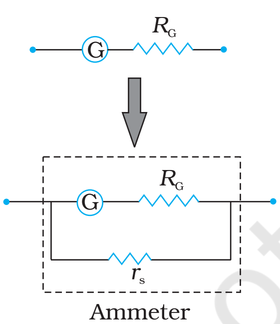


FIGURE 4.21 Conversion of a galvanometer (G) to an ammeter by the introduction of a shunt resistance r_s of very small value in parallel.

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. 4.21. 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. (4.26) this current sensitivity is,

$$\frac{\phi}{I} = \frac{NAB}{k} \qquad (4.27)$$

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.

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.4.22. Note that the resistance of the voltmeter is now,

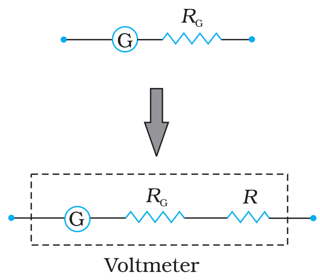


FIGURE 4.22 Conversion of a galvanometer (G) to a voltmeter by the introduction of a resistance R of large value in series.

$R_G + R \simeq R$: 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. (4.26),

$$\frac{\phi}{V} = \left(\frac{NAB}{k}\right)\frac{I}{V} = \left(\frac{NAB}{k}\right)\frac{1}{R} \qquad (4.28)$$

An interesting point to note is that increasing the current sensitivity may not necessarily increase the voltage sensitivity. Let us take Eq. (4.27) 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. (4.28), $N \rightarrow 2N$, 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 4.12 In the circuit (Fig. 4.23) 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?

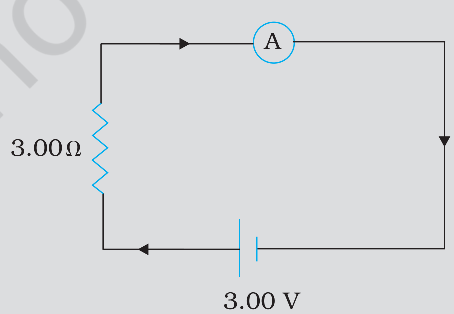


FIGURE 4.23

EXAMPLE 4.12

EXAMPLE 4.12

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,

$$\frac{R_G\, r_s}{R_G + r_s} = \frac{60\ \ \Omega \times 0.02\Omega}{(60 + 0.02)\Omega} \simeq 0.02\Omega$$

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

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:
   $$d\mathbf{B} = \frac{\mu_0}{4\pi} I \frac{d\mathbf{l} \times \mathbf{r}}{r^3}$$
   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

$$B = \frac{\mu_0 I R^2}{2(x^2 + R^2)^{3/2}}$$

At the centre this reduces to

$$B = \frac{\mu_0 I}{2R}$$

6. *Ampere's Circuital Law:* Let an open surface S be bounded by a loop C. Then the Ampere's law states that $\oint_C \mathbf{B}.d\mathbf{l} = \mu_0 I$ 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,

   $BL = \mu_0 I_e$

   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:

$$B = \frac{\mu_0 I}{2\pi R}$$

   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

   $B = \mu_0 n I$

   where n is the number of turns per unit length.
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,

    $\mathbf{m} = N\, I\, \mathbf{A}$

    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. 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	μ_0	Scalar	$[MLT^{-2}A^{-2}]$	$T\ m\ A^{-1}$	$4\pi \times 10^{-7}\ T\ m\ A^{-1}$
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:

$R_1 = 10\ \Omega,\ N_1 = 30,$

$A_1 = 3.6 \times 10^{-3}\ \text{m}^2,\ B_1 = 0.25\ \text{T}$

$R_2 = 14\ \Omega,\ N_2 = 42,$

$A_2 = 1.8 \times 10^{-3}\ \text{m}^2,\ B_2 = 0.50\ \text{T}$

(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 $4.8 \times 10^6\ \text{m s}^{-1}$ normal to the field. Explain why the path of the electron is a circle. Determine the radius of the circular orbit. ($e = 1.5 \times 10^{-19}$ 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.)

12089CH05

Chapter Five

MAGNETISM AND MATTER

5.1 Introduction

Magnetic phenomena are universal in nature. Vast, distant galaxies, the tiny invisible atoms, humans 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.

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 next describe how materials can be classified on the basis of their magnetic properties. We describe para-, dia-, and ferromagnetism.

5.2 The Bar Magnet

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. 5.1.

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.

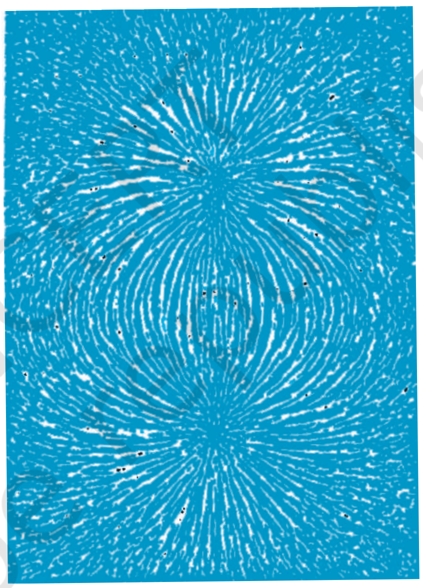


FIGURE 5.1 The arrangement of iron filings surrounding a bar magnet. The pattern mimics magnetic field lines. The pattern suggests that the bar magnet is a magnetic dipole.

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. 5.2. For comparison refer to the Chapter 1, Figure 1.14(d). Electric field lines of an electric dipole are also displayed in Fig. 5.2(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.

* 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.

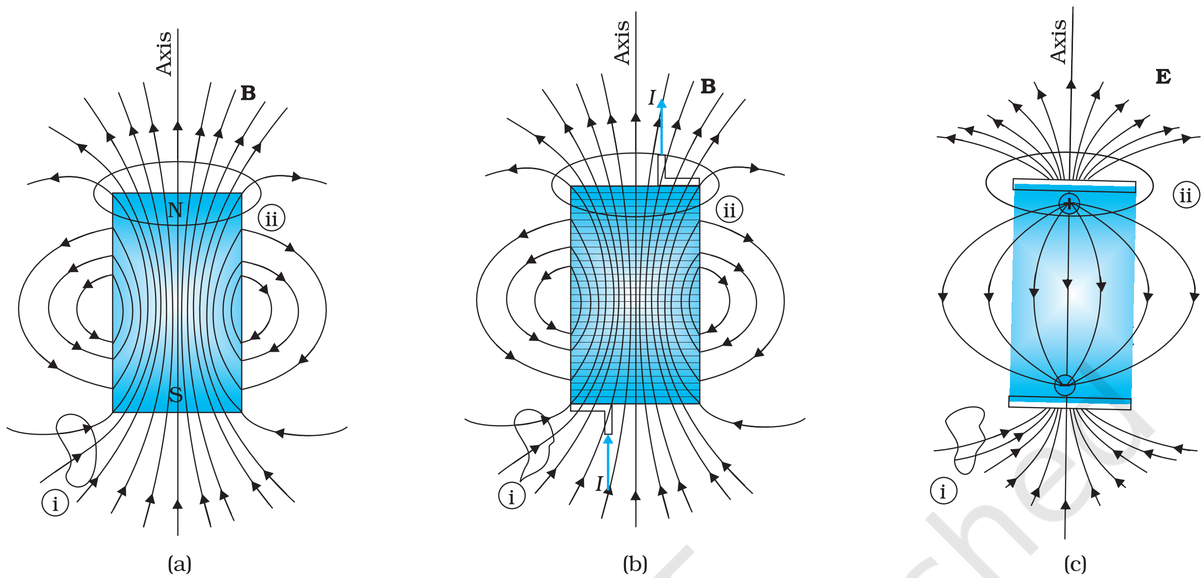


FIGURE 5.2 The field lines of (a) a bar magnet, (b) a current-carrying finite solenoid and (c) electric dipole. At large distances, the field lines are very similar. The curves labelled (i) and (ii) are closed Gaussian surfaces.

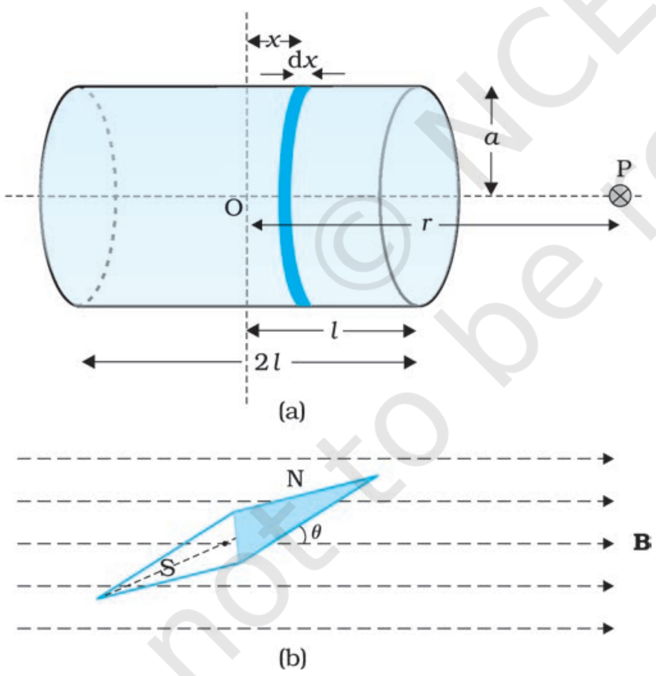


FIGURE 5.3 Calculation of (a) The axial field of a finite solenoid in order to demonstrate its similarity to that of a bar magnet. (b) A magnetic needle in a uniform magnetic field **B**. The arrangement may be used to determine either B or the magnetic moment **m** of the needle.

(ii) The tangent to the field line at a given point represents the direction of the net magnetic field **B** at that point.

(iii) The larger the number of field lines crossing per unit area, the stronger is the magnitude of the magnetic field **B**. In Fig. 5.2(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 4.9). We mentioned Ampere's hypothesis that all magnetic phenomena can be explained in terms of circulating currents.

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 may calculate the axial field of a finite solenoid depicted in Fig. 5.3 (a). We can demonstrate that at large distances this axial field resembles that of a bar magnet.

The magnitude of the field at point P due to the solenoid is

$$B = \frac{\mu_0}{4\pi}\frac{2m}{r^3} \tag{5.1}$$

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.

5.2.3 The dipole in a uniform magnetic field

Let's place a small compass needle of known magnetic moment **m** allowing it to oscillate in the magnetic field. This arrangement is shown in Fig. 5.3(b).

The torque on the needle is [see Eq. (4.23)],

$$\boldsymbol{\tau} = \mathbf{m} \times \mathbf{B} \tag{5.2}$$

In magnitude $\tau = mB\sin\theta$

Here $\boldsymbol{\tau}$ is restoring torque and θ is the angle between **m** and **B**.

An expression for magnetic potential energy can be obtained on lines similar to electrostatic potential energy.

The magnetic potential energy U_m is given by

$$U_m = \int \tau(\theta)d\theta$$

$$= \int mB\sin\theta\, d\theta = -mB\cos\theta$$

$$= -\mathbf{m}.\mathbf{B} \tag{5.3}$$

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 (5.3) 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

(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?

EXAMPLE 5.1

(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 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.

EXAMPLE 5.1

5.2.4 The electrostatic analog

Comparison of Eqs. (5.1), (5.2) and (5.3) 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:

$$\mathbf{E} \to \mathbf{B}\,,\ \mathbf{p} \to \mathbf{m}\,,\ \frac{1}{4\pi\varepsilon_0} \to \frac{\mu_0}{4\pi}$$

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:

$$\mathbf{B}_E = -\frac{\mu_0 \mathbf{m}}{4\pi r^3} \tag{5.4}$$

Likewise, the axial field ($\mathbf{B}_A$) of a bar magnet for $r >> l$ is:

$$\mathbf{B}_A = \frac{\mu_0}{4\pi}\frac{2\mathbf{m}}{r^3} \tag{5.5}$$

Equation (5.5) is just Eq. (5.1) in the vector form. Table 5.1 summarises the analogy between electric and magnetic dipoles.

TABLE 5.1 THE DIPOLE ANALOGY

	Electrostatics	**Magnetism**
	$1/\varepsilon_0$	μ_0
Dipole moment	$\mathbf{p}$	$\mathbf{m}$
Equatorial Field for a short dipole	$-\mathbf{p}/4\pi\varepsilon_0 r^3$	$-\mu_0\,\mathbf{m}\,/\,4\pi\, r^3$
Axial Field for a short dipole	$2\mathbf{p}/4\pi\varepsilon_0 r^3$	$\mu_0\,2\mathbf{m}\,/\,4\pi\, r^3$
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 5.2 Figure 5.4 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?

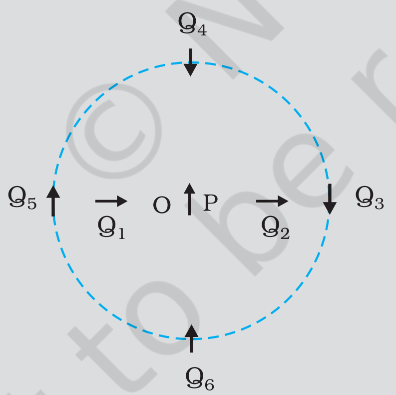


FIGURE 5.4

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. (5.4) and (5.5)]:

$$\mathbf{B}_P = -\frac{\mu_0}{4\pi}\frac{\mathbf{m}_P}{r^3} \qquad \text{(on the normal bisector)}$$

$$\mathbf{B}_P = \frac{\mu_0 2}{4\pi}\frac{\mathbf{m}_P}{r^3} \qquad \text{(on the axis)}$$

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 $\mathbf{B}_P$.

EXAMPLE 5.2

EXAMPLE 5.2

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

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.

5.3 MAGNETISM AND GAUSS'S LAW

In Chapter 1, we studied Gauss's law for electrostatics. In Fig 5.2(c), we see that for a closed surface represented by (i), 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 (ii), there is a net outward flux, since it does include a net (positive) charge.

The situation is radically different for magnetic fields which are continuous and form closed loops. Examine the Gaussian surfaces represented by (i) or (ii) in Fig 5.2(a) or Fig. 5.2(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.

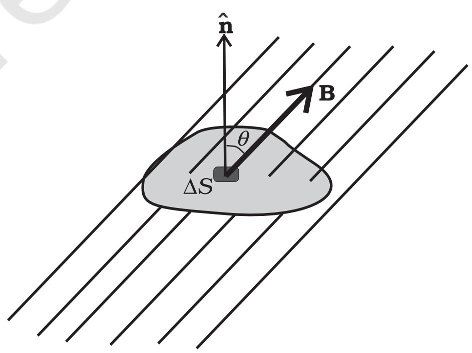


FIGURE 5.5

Consider a small vector area element $\Delta\mathbf{S}$ of a closed surface S as in Fig. 5.5. 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,

$$\phi_B = \sum_{\text{'all'}} \Delta\phi_B = \sum_{\text{'all'}} \mathbf{B}.\Delta\mathbf{S} = 0 \qquad (5.6)$$

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

$$\sum \mathbf{E}.\Delta\mathbf{S} = \frac{q}{\varepsilon_0}$$

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 5.3 Many of the diagrams given in Fig. 5.6 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.

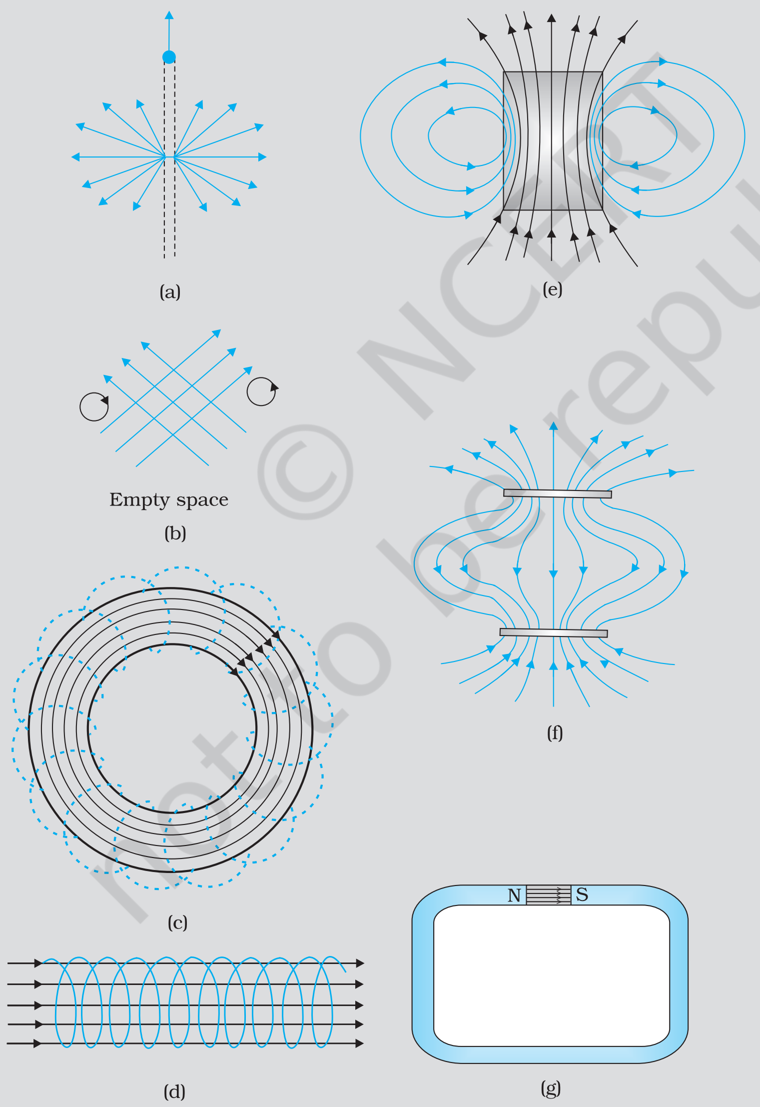


FIGURE 5.6

EXAMPLE 5.3

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.

(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. [5.6(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 5.4

Example 5.4

(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) If magnetic monopoles existed, how would the Gauss's law of magnetism be modified?

(c) 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*?

(d) 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) Gauss's law of magnetism states that the flux of **B** through any closed surface is always zero $\int_S \mathbf{B}.\Delta\mathbf{s} = 0$.

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, $\int_S \mathbf{B}.\Delta\mathbf{s} = \mu_0 q_m$ where q_m is the (monopole) magnetic charge enclosed by S.]

(c) *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.)

(d) *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 5.4

5.4 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:

$$\mathbf{M} = \frac{\mathbf{m}_{net}}{V} \tag{5.7}$$

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

$$\mathbf{B}_0 = \mu_0 nI \tag{5.8}$$

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

$$\mathbf{B} = \mathbf{B}_0 + \mathbf{B}_m \tag{5.9}$$

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 $\mathbf{M}$ of the material and is expressed as

$$\mathbf{B}_m = \mu_0 \mathbf{M} \tag{5.10}$$

where μ_0 is the same constant (permittivity of vacuum) that appears in Biot-Savart's law.

It is convenient to introduce another vector field $\mathbf{H}$, called the *magnetic intensity*, which is defined by

$$\mathbf{H} = \frac{\mathbf{B}}{\mu_0} - \mathbf{M} \tag{5.11}$$

where $\mathbf{H}$ has the same dimensions as $\mathbf{M}$ and is measured in units of A m^{-1}. Thus, the total magnetic field $\mathbf{B}$ is written as

$$\mathbf{B} = \mu_0 (\mathbf{H} + \mathbf{M}) \tag{5.12}$$

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 $\mathbf{H}$. The *other* is due to the specific nature of the magnetic material, namely $\mathbf{M}$. The latter quantity can be influenced by external factors. This influence is mathematically expressed as

$$\mathbf{M} = \chi \mathbf{H} \tag{5.13}$$

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. χ 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 $\mathbf{M}$ and $\mathbf{H}$ are opposite in direction. From Eqs. (5.12) and (5.13) we obtain,

$$\mathbf{B} = \mu_0 (1 + \chi)\mathbf{H} \tag{5.14}$$

$$= \mu_0 \mu_r \mathbf{H}$$

$$= \mu \mathbf{H} \tag{5.15}$$

where $\mu_r = 1 + \chi$, 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 μ and it has the same dimensions and units as μ_0;

$$\mu = \mu_0 \mu_r = \mu_0 (1+\chi).$$

The three quantities χ, μ_r and μ are interrelated and only one of them is independent. Given one, the other two may be easily determined.

EXAMPLE 5.5

Example 5.5 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

$$B = \mu_r \mu_0 H$$
$$= 400 \times 4\pi \times 10^{-7} \text{ (N/A}^2) \times 2 \times 10^3 \text{ (A/m)}$$
$$= 1.0 \text{ T}$$

(c) Magnetisation is given by

$$M = (B - \mu_0 H)/\mu_0$$
$$= (\mu_r \mu_0 H - \mu_0 H)/\mu_0 = (\mu_r - 1)H = 399 \times H$$
$$\cong 8 \times 10^5 \text{ A/m}$$

(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 $B = \mu_r n (I + I_M)$. Using $I = 2$A, $B = 1$ T, we get $I_M = 794$ A.

EXAMPLE 5.5

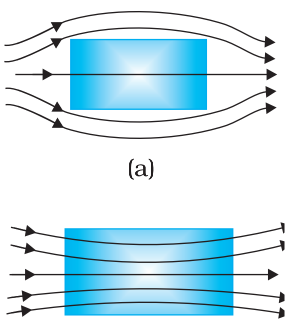
5.5 Magnetic Properties of Materials

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 5.2 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 5.2

Diamagnetic	Paramagnetic	Ferromagnetic
$-1 \le \chi < 0$	$0 < \chi < \varepsilon$	$\chi >> 1$
$0 \le \mu_r < 1$	$1 < \mu_r < 1 + \varepsilon$	$\mu_r >> 1$
$\mu < \mu_0$	$\mu > \mu_0$	$\mu >> \mu_0$



(a)

(b)

FIGURE 5.7 Behaviour of magnetic field lines near a (a) diamagnetic, (b) paramagnetic substance.

5.5.1 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 5.7(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, 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.

2024-25

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 $\mu_r = 0$. 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.

5.5.2 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 5.7(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, 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. For a paramagnetic material both χ and μ_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 at which point all the dipoles are perfectly aligned with the field.

5.5.3 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. 5.8(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 ferromagnetic substance of samples. This motion of suspension can be observed. Fig. 5.8(b) shows the situation when the domains have aligned and amalgamated to form a single 'giant' domain.

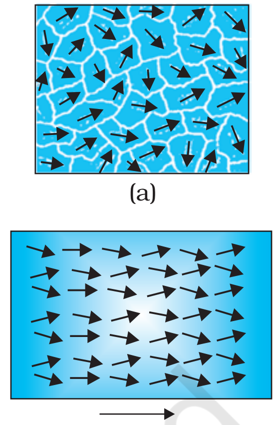


FIGURE 5.8 (a) Randomly oriented domains, (b) Aligned domains.

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.

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 $\mathbf{m} \times \mathbf{B}$,
   (c) its potential energy is $-\mathbf{m}\cdot\mathbf{B}$, 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. Consider a material placed in an external magnetic field $\mathbf{B}_0$. The magnetic intensity is defined as,

$$\mathbf{H} = \frac{\mathbf{B}_0}{\mu_0}$$

The magnetisation **M** of the material is its dipole moment per unit volume. The magnetic field **B** in the material is,

$$\mathbf{B} = \mu_0 (\mathbf{H} + \mathbf{M})$$

6. For a linear material $\mathbf{M} = \chi \mathbf{H}$. So that $\mathbf{B} = \mu \mathbf{H}$ and χ is called the magnetic susceptibility of the material. The three quantities, χ, the relative magnetic permeability μ_r, and the magnetic permeability μ are related as follows:

$$\mu = \mu_0 \mu_r$$

$$\mu_r = 1 + \chi$$

7. 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**.
8. 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	μ_0	Scalar	$[MLT^{-2}\ A^{-2}]$	$T\ m\ A^{-1}$	$\mu_0/4\pi = 10^{-7}$
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	$[L^{-1}\ A]$	$A\ m^{-1}$	$\dfrac{\text{Magnetic moment}}{\text{Volume}}$
Magnetic intensity Magnetic field strength	**H**	Vector	$[L^{-1}\ A]$	$A\ m^{-1}$	$\mathbf{B} = \mu_0\ (\mathbf{H} + \mathbf{M})$
Magnetic susceptibility	χ	Scalar	-	-	$\mathbf{M} = \chi \mathbf{H}$
Relative magnetic permeability	μ_r	Scalar	-	-	$\mathbf{B} = \mu_0 \mu_r \mathbf{H}$
Magnetic permeability	μ	Scalar	$[MLT^{-2}\ A^{-2}]$	$T\ m\ A^{-1}$ $N\ A^{-2}$	$\mu = \mu_0 \mu_r$ $\mathbf{B} = \mu \mathbf{H}$

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. 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.

5. There exists *a perfect diamagnet*, namely, a superconductor. This is a metal at very low temperatures. In this case $\chi = -1$, $\mu_r = 0$, $\mu = 0$. 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.
6. Diamagnetism is universal. It is present in all materials. But it is weak and hard to detect if the substance is para- or ferromagnetic.
7. 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 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?

5.2 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?

5.3 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?

5.4 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?

5.5 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)?

5.6 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?

5.7 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.

12089CH06

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.

JOSEPH HENRY (1797 – 1878)

Josheph Henry [1797 – 1878] American experimental physicist, professor at Princeton University and first director of the Smithsonian Institution. He made important improvements in electromagnets 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.

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.*

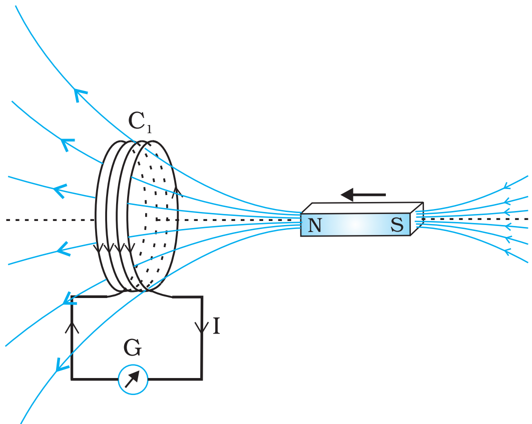


FIGURE 6.1 When the bar magnet is pushed towards the coil, the pointer in the galvanometer G deflects.

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.

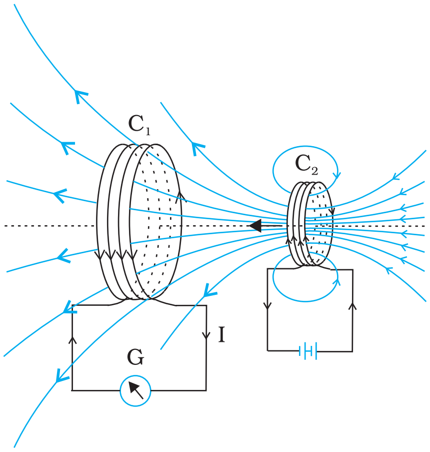


FIGURE 6.2 Current is induced in coil C_1 due to motion of the current carrying coil C_2.

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.

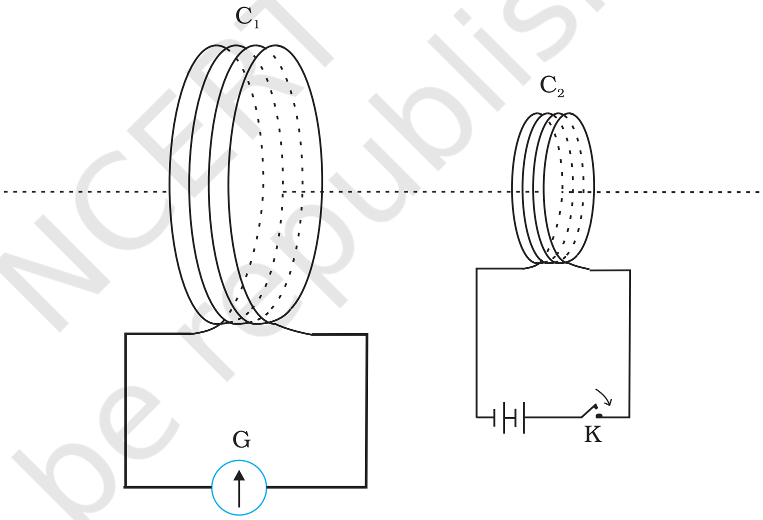


FIGURE 6.3 Experimental set-up for Experiment 6.3.

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.

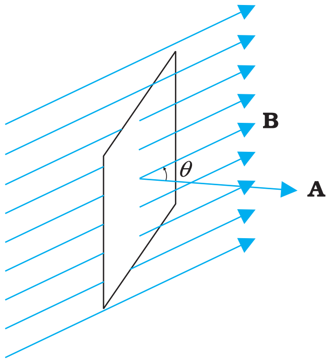


FIGURE 6.4 A plane of surface area **A** placed in a uniform magnetic field **B**.

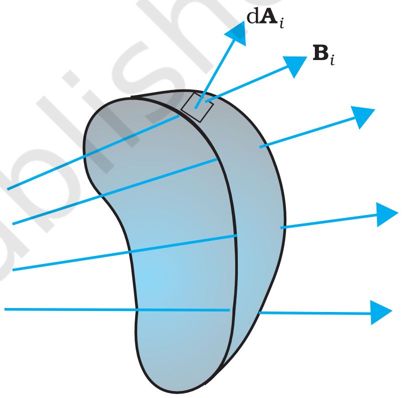


FIGURE 6.5 Magnetic field $\mathbf{B}_i$ at the i^{th} area element. $d\mathbf{A}_i$ represents area vector of the i^{th} area element.

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.

* 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

$$\varepsilon = -\frac{\mathrm{d}\Phi_B}{\mathrm{d}t} \tag{6.3}$$

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

$$\varepsilon = -N\frac{\mathrm{d}\Phi_B}{\mathrm{d}t} \tag{6.4}$$

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,

$$\Phi_{B\,(\text{initial})} = BA \cos\theta$$

$$= 3.0 \times 10^{-5} \times (\pi \times 10^{-2}) \times \cos 0°$$

$$= 3\pi \times 10^{-7}\ \text{Wb}$$

Final flux after the rotation,

$$\Phi_{B\,(\text{final})} = 3.0 \times 10^{-5} \times (\pi \times 10^{-2}) \times \cos 180°$$

$$= -3\pi \times 10^{-7}\ \text{Wb}$$

Therefore, estimated value of the induced emf is,

$$\varepsilon = N\frac{\Delta \Phi}{\Delta t}$$

$$= 500 \times (6\pi \times 10^{-7})/0.25$$

$$= 3.8 \times 10^{-3}\ \text{V}$$

$I = \varepsilon/R = 1.9 \times 10^{-3}$ 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.

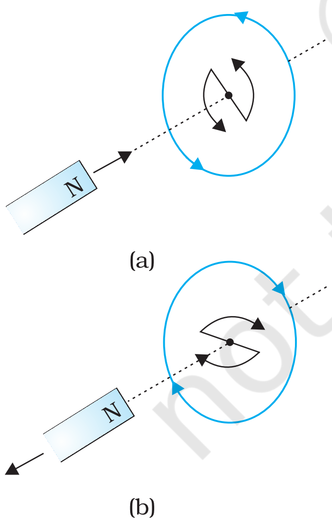


FIGURE 6.6 Illustration of Lenz's law.

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.

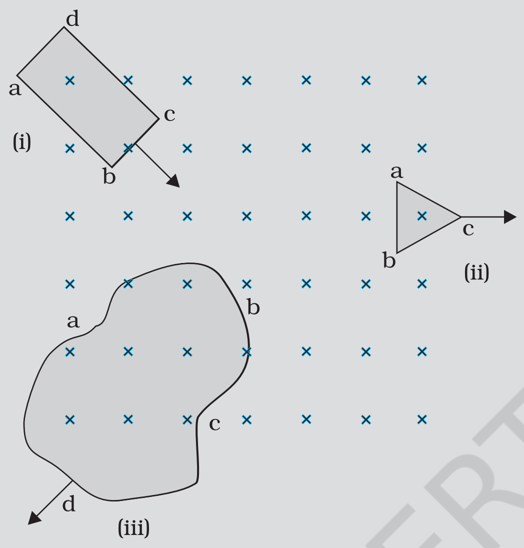


FIGURE 6.7

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

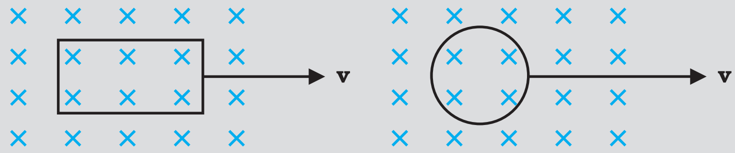


FIGURE 6.8

(d) Predict the polarity of the capacitor in the situation described by Fig. 6.9.

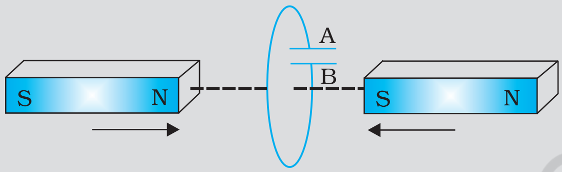


FIGURE 6.9

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

$$\Phi_B = Blx$$

Since x is changing with time, the rate of change of flux Φ_B will induce an emf given by:

$$\varepsilon = \frac{-\mathrm{d}\Phi_B}{\mathrm{d}t} = -\frac{\mathrm{d}}{\mathrm{d}t}(Blx)$$

$$= -Bl\frac{\mathrm{d}x}{\mathrm{d}t} = Blv \qquad (6.5)$$

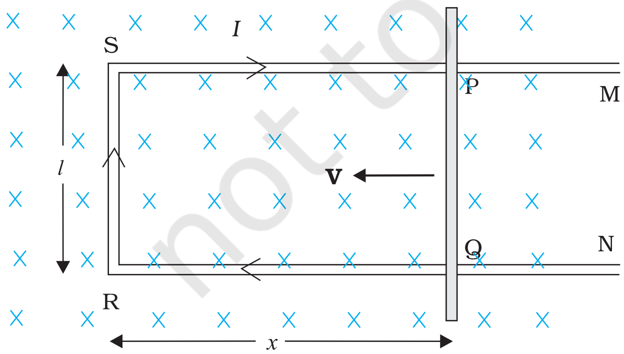


FIGURE 6.10 The arm PQ is moved to the left side, thus decreasing the area of the rectangular loop. This movement induces a current I as shown.

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

$$\mathbf{F} = q\,(\mathbf{E} + \mathbf{v} \times \mathbf{B}) = q\mathbf{E} \tag{6.6}$$

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.

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

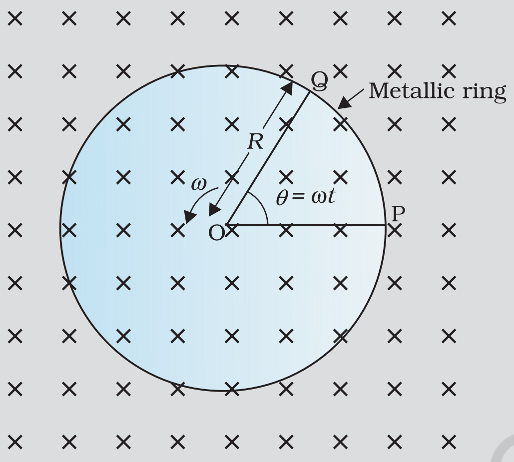


FIGURE 6.11

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 × (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

$$\varepsilon = B \times \frac{d}{dt}\left[\frac{1}{2} R^2 \theta\right] = \frac{1}{2} B R^2 \frac{d\theta}{dt} = \frac{B\omega R^2}{2}$$

[Note: $\dfrac{d\theta}{dt} = \omega = 2\pi\nu$]

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 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.

6.7.1 Mutual inductance

Consider Fig. 6.12 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.

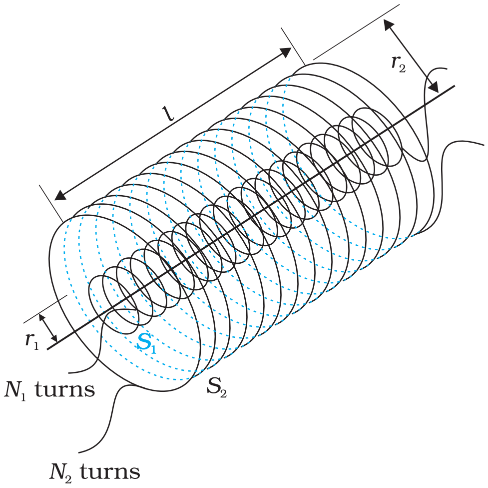


FIGURE 6.12 Two long co-axial solenoids of same length l.

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

$$N_1\Phi_1 = M_{12}I_2 \tag{6.7}$$

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,

$$N_1\Phi_1 = (n_1 l)(\pi r_1^2)(\mu_0 n_2 I_2)$$

$$= \mu_0 n_1 n_2 \pi r_1^2 l\, I_2 \tag{6.8}$$

where $n_1 l$ is the total number of turns in solenoid S_1. Thus, from Eq. (6.7) and Eq. (6.8),

$$M_{12} = \mu_0 n_1 n_2 \pi r_1^2 l \tag{6.9}$$

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,

$$N_2\Phi_2 = M_{21}\, I_1 \tag{6.10}$$

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

$$N_2\Phi_2 = (n_2 l)(\pi r_1^2)(\mu_0 n_1 I_1)$$

where $n_2 l$ is the total number of turns of S_2. From Eq. (6.10),

$$M_{21} = \mu_0 n_1 n_2 \pi r_1^2 l \tag{6.11}$$

Using Eq. (6.9) and Eq. (6.10), we get

$$M_{12} = M_{21} = M \text{ (say)} \tag{6.12}$$

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 6.8 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. (6.12)

$$M_{12} = M_{21} = \frac{\mu_0 \pi r_1^2}{2r_2}$$

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 6.8

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. (6.7), we have

$$N_1 \Phi_1 = M I_2$$

For currents varrying with time,

$$\frac{\mathrm{d}(N_1 \Phi_1)}{\mathrm{d}t} = \frac{\mathrm{d}(M I_2)}{\mathrm{d}t}$$

Since induced emf in coil C_1 is given by

$$\varepsilon_1 = -\frac{\mathrm{d}(N_1 \Phi_1)}{\mathrm{d}t}$$

We get,

$$\varepsilon_1 = -M \frac{\mathrm{d}I_2}{\mathrm{d}t}$$

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.

6.7.2 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

$$N\Phi_B \propto I$$

$$N\Phi_B = L\,I \qquad (6.13)$$

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. (6.13), the induced emf is given by

$$\varepsilon = -\frac{\mathrm{d}(N\Phi_B)}{\mathrm{d}t}$$

$$\varepsilon = -L\frac{\mathrm{d}I}{\mathrm{d}t} \qquad (6.14)$$

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,

$$L = \frac{N\Phi_B}{I}$$

$$= \mu_0 n^2 A l \qquad (6.15)$$

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 permeability), then,

$$L = \mu_r\, \mu_0\, n^2 A l \qquad (6.16)$$

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

$$\frac{\mathrm{d}W}{\mathrm{d}t} = |\varepsilon| I$$

If we ignore the resistive losses and consider only inductive effect, then using Eq. (6.14),

$$\frac{\mathrm{d}W}{\mathrm{d}t} = L\, I \frac{\mathrm{d}I}{\mathrm{d}t}$$

Total amount of work done in establishing the current I is

$$W = \int \mathrm{d}W = \int_0^I L\, I\, \mathrm{d}I$$

Thus, the energy required to build up the current I is,

$$W = \frac{1}{2} L I^2 \qquad (6.17)$$

This expression reminds us of $mv^2/2$ for the (mechanical) kinetic energy of a particle of mass m, and shows that L is analogous 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 (6.7) 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,

$$\varepsilon_1 = -M_{11} \frac{\mathrm{d}I_1}{\mathrm{d}t} - M_{12} \frac{\mathrm{d}I_2}{\mathrm{d}t}$$

M_{11} is the *self-inductance* and is written as L_1. Therefore,

$$\varepsilon_1 = -L_1 \frac{\mathrm{d}I_1}{\mathrm{d}t} - M_{12} \frac{\mathrm{d}I_2}{\mathrm{d}t}$$

Example 6.9 (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. (6.17), the magnetic energy is

$$U_B = \frac{1}{2} L I^2$$

$$= \frac{1}{2} L \left(\frac{B}{\mu_0 n}\right)^2 \qquad (\text{since } B = \mu_0 n I, \text{for a solenoid})$$

EXAMPLE 6.9

Interactive animation on ac generator:
http://micro.magnet.fsu.edu/electromag/java/generator/ac.html

PHYSICS

EXAMPLE 6.9

$$= \frac{1}{2}(\mu_0 n^2 Al)\left(\frac{B}{\mu_0 n}\right)^2 \qquad \text{[from Eq. (6.15)]}$$

$$= \frac{1}{2\mu_0} B^2 Al$$

(b) The magnetic energy per unit volume is,

$$u_B = \frac{U_B}{V} \qquad \text{(where } V \text{ is volume that contains flux)}$$

$$= \frac{U_B}{Al}$$

$$= \frac{B^2}{2\mu_0} \qquad (6.18)$$

We have already obtained the relation for the electrostatic energy stored per unit volume in a parallel plate capacitor (refer to Chapter 2, Eq. 2.73),

$$u_E = \frac{1}{2}\varepsilon_0 E^2 \qquad (2.73)$$

In both the cases energy is proportional to the square of the field strength. Equations (6.18) and (2.73) 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.

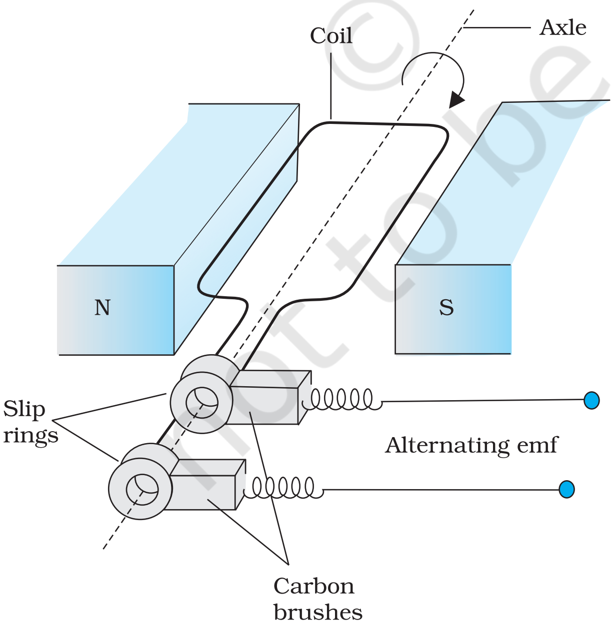


FIGURE 6.13 AC Generator

6.8 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. 6.13. 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.

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

$$\varepsilon = NBA\,\omega\sin\omega t \tag{6.19}$$

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

$$\varepsilon = \varepsilon_0 \sin\omega t \tag{6.20}$$

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. 6.14 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 (6.20) can be written as

$$\varepsilon = \varepsilon_0 \sin 2\pi\nu t \tag{6.21}$$

where ν is the frequency of revolution of the generator's coil.

Note that Eq. (6.20) and (6.21) 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.

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

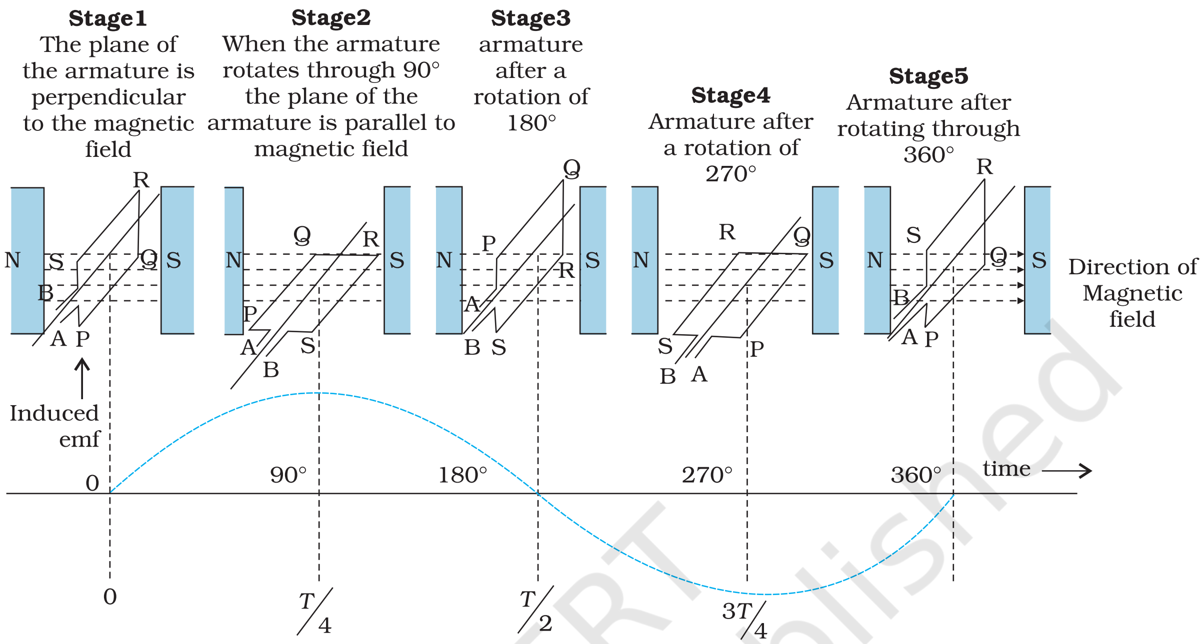


FIGURE 6.14 An alternating emf is generated by a loop of wire rotating in a magnetic field.

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 6.10 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 $\nu = 0.5$ Hz; $N = 100$, $A = 0.1$ m^2 and $B = 0.01$ T. Employing Eq. (6.19)

$\varepsilon_0 = NBA\,(2\,\pi\,\nu)$

$= 100 \times 0.01 \times 0.1 \times 2 \times 3.14 \times 0.5$

$= 0.314$ V

The maximum voltage is 0.314 V.

We urge you to explore such alternative possibilities for power generation.

EXAMPLE 6.10

SUMMARY

1. The magnetic flux through a surface of area **A** placed in a uniform magnetic field **B** is defined as,

   $\Phi_B = \mathbf{B.A} = BA \cos\theta$

   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. Inductance is the ratio of the flux-linkage to current. It is equal to $N\Phi/I$.
6. 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,

   $M_{12} = M_{21}$
7. 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.
8. The self-inductance of a long solenoid, the core of which consists of a magnetic material of relative 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.
9. 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.

EXERCISES

6.1 Predict the direction of induced current in the situations described by the following Figs. 6.15(a) to (f).

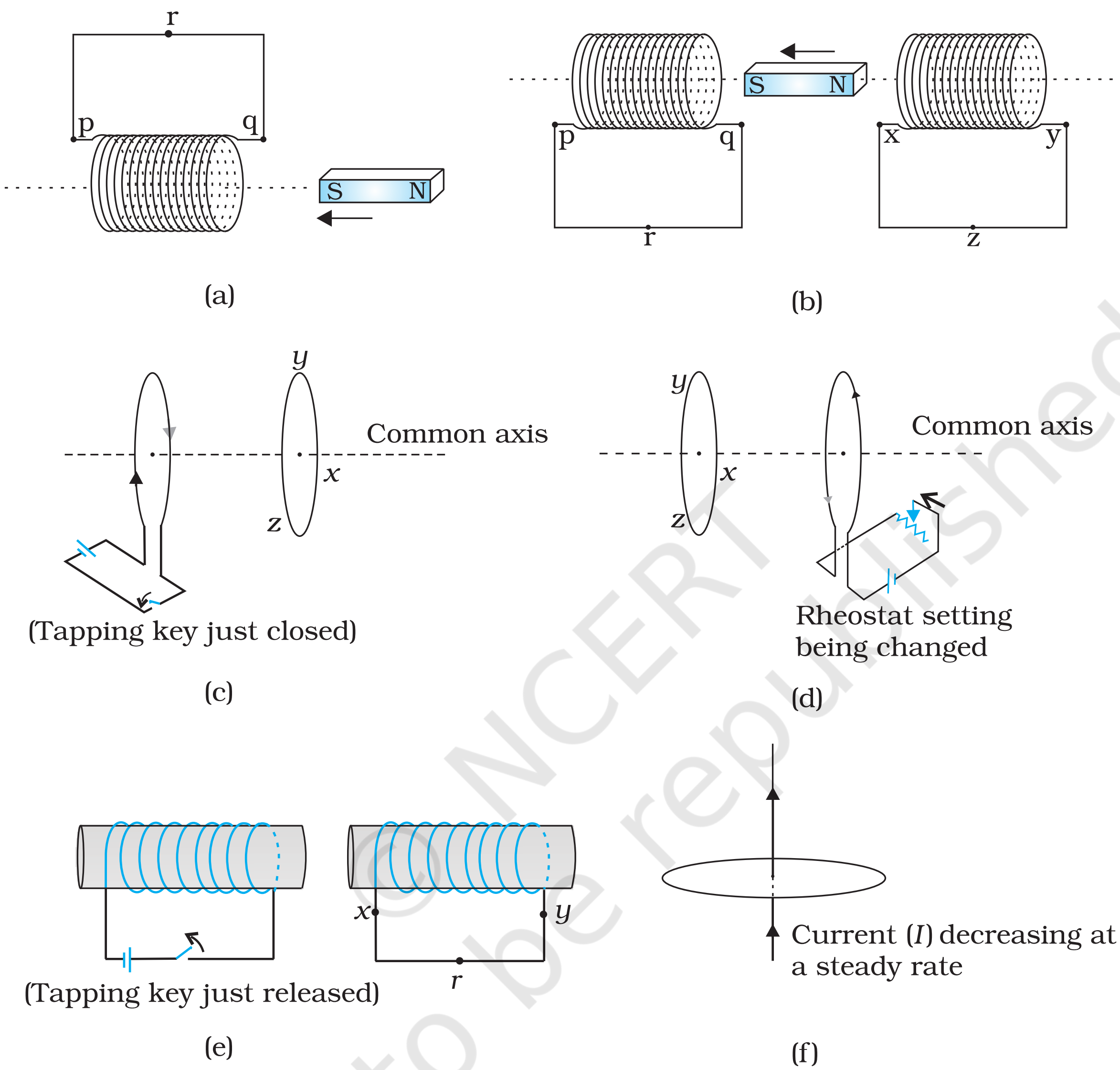


FIGURE 6.15

6.2 Use Lenz's law to determine the direction of induced current in the situations described by Fig. 6.16:

(a) A wire of irregular shape turning into a circular shape;

(b) A circular loop being deformed into a narrow straight wire.

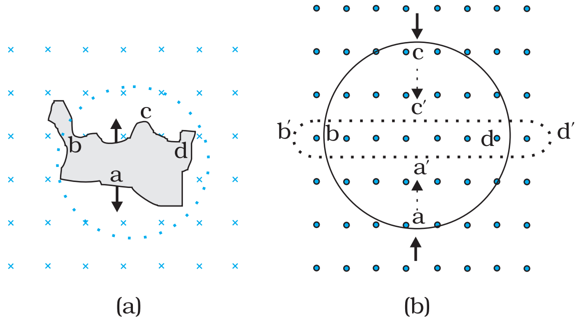


FIGURE 6.16

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 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?

6.7 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.

6.8 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?

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.

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 $\sum \varepsilon(t) = 0$ (refer to Section 3.12), 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

$$i = i_m \sin \omega t \tag{7.2}$$

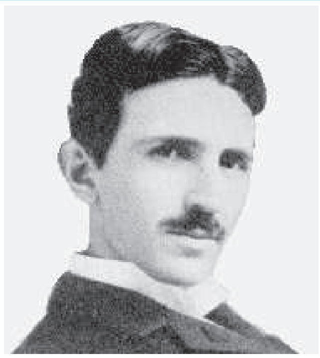
where the current amplitude i_m is given by

$$i_m = \frac{v_m}{R} \tag{7.3}$$

Equation (7.3) is 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

NICOLA TESLA (1856 – 1943)



Nicola Tesla (1856 – 1943) Serbian-American 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.

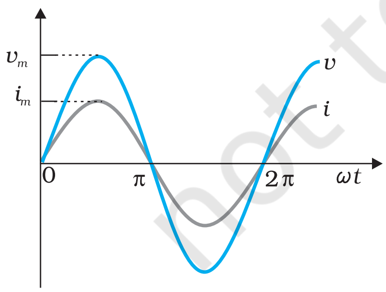


FIGURE 7.2 In a pure resistor, the voltage and current are in phase. The minima, zero and maxima occur at the same respective times.

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

$$p = i^2 R = i_m^2 R \sin^2 \omega t \qquad (7.4)$$

The average value of p over a cycle is*

$$\bar{p} = < i^2 R > = < i_m^2 R \sin^2 \omega t > \qquad [7.5(a)]$$

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,

$$\bar{p} = i_m^2 R < \sin^2 \omega t > \qquad [7.5(b)]$$

Using the trigonometric identity, $\sin^2 \omega t$ = $1/2\,(1 - \cos 2\omega t)$, we have $< \sin^2 \omega t > = (1/2)\,(1 - < \cos 2\omega t >)$ and since $< \cos 2\omega t > = 0$**, we have,

$$< \sin^2 \omega t > = \frac{1}{2}$$

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.

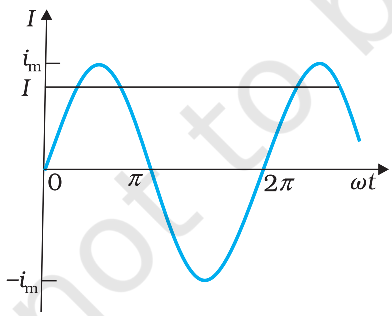
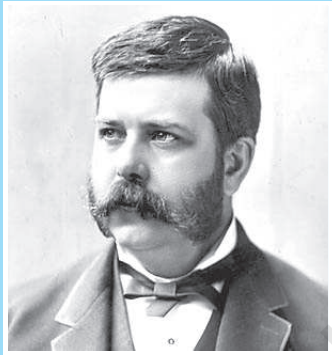


FIGURE 7.3 The rms current I is related to the peak current i_m by $I = i_m / \sqrt{2} = 0.707\, i_m$.



GEORGE WESTINGHOUSE (1846 – 1914)

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.

* The average value of a function $F(t)$ over a period T is given by $\langle F(t) \rangle = \frac{1}{T}\int_0^T F(t)\,dt$

** $< \cos 2\omega t > = \frac{1}{T}\int_0^T \cos 2\omega t\, dt = \frac{1}{T}\left[\frac{\sin 2\omega t}{2\omega}\right]_0^T = \frac{1}{2\omega T}[\sin 2\omega T - 0] = 0$

It is defined by

$$I = \sqrt{\overline{i^2}} = \sqrt{\frac{1}{2} i_m^2} = \frac{i_m}{\sqrt{2}}$$

$$= 0.707\ i_m \qquad (7.6)$$

In terms of I, the average power, denoted by P is

$$P = \overline{p} = \frac{1}{2} i_m^2 R = I^2 R \qquad (7.7)$$

Similarly, we define the *rms voltage* or *effective voltage* by

$$V = \frac{v_m}{\sqrt{2}} = 0.707\ v_m \qquad (7.8)$$

From Eq. (7.3), we have

$$v_m = i_m R$$

or, $$\frac{v_m}{\sqrt{2}} = \frac{i_m}{\sqrt{2}} R$$

or, $$V = IR \qquad (7.9)$$

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

$$v_m = \sqrt{2}\ V = (1.414)(220\ V) = 311\ V$$

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

$$P = V^2 / R = I V \quad \text{(since } V = I R\text{)}$$

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, $P = I\ V$

$$I = \frac{P}{V} = \frac{100\ \text{W}}{220\ \text{V}} = 0.454\text{A}$$

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.

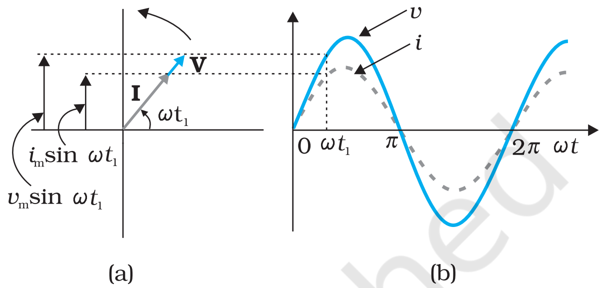


FIGURE 7.4 (a) A phasor diagram for the circuit in Fig 7.1. (b) Graph of v and i versus ωt.

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,

$$v - L\frac{\mathrm{d}i}{\mathrm{d}t} = 0 \qquad (7.10)$$

where the second term is the self-induced Faraday emf in the inductor; and L is the self-inductance of

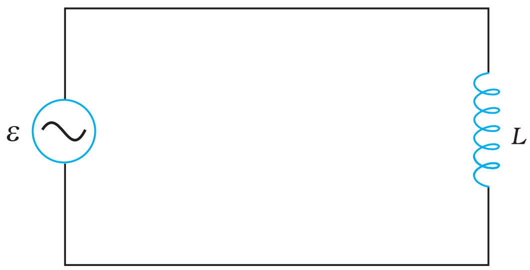


FIGURE 7.5 An ac source connected to an inductor.

* 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

$$\frac{di}{dt} = \frac{v}{L} = \frac{v_m}{L}\sin \omega t \qquad (7.11)$$

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

$$-\cos(\omega t) = \sin\left(\omega t - \frac{\pi}{2}\right), \text{ we have}$$

$$i = i_m \sin\left(\omega t - \frac{\pi}{2}\right) \qquad (7.12)$$

where $i_m = \dfrac{v_m}{\omega L}$ is the amplitude of the current. The quantity ωL is analogous to the resistance and is called *inductive reactance*, denoted by X_L:

$$X_L = \omega L \qquad (7.13)$$

The amplitude of the current is, then

$$i_m = \frac{v_m}{X_L} \qquad (7.14)$$

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).

PHYSICS

Interactive animation on Phasor diagrams of ac circuits containing, R, L, C and RLC series circuits:
http://www.animations.physics.unsw.edu.au/jw/AC.html

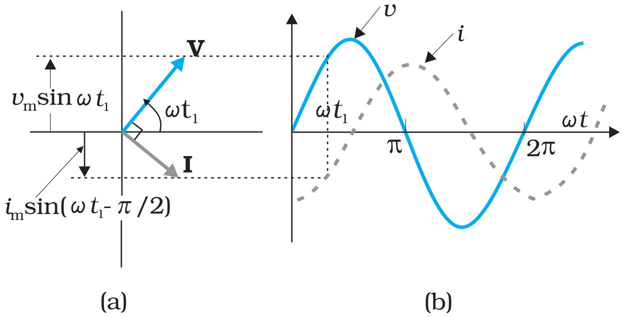


FIGURE 7.6 (a) A Phasor diagram for the circuit in Fig. 7.5.
(b) Graph of v and i versus ωt.

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

$$p_L = i\,v = i_m \sin\left(\omega t - \frac{\pi}{2}\right) \times v_m \sin(\omega t)$$

$$= -i_m v_m \cos(\omega t)\sin(\omega t)$$

$$= -\frac{i_m v_m}{2}\sin(2\omega t)$$

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.*

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,

$$X_L = 2\pi\nu L = 2 \times 3.14 \times 50 \times 25 \times 10^{-3}\ \Omega$$

$$= 7.85\Omega$$

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

7.5 AC Voltage Applied to a Capacitor

Figure 7.7 shows an ac source ε generating ac voltage $v = v_m \sin \omega t$ connected to a capacitor only, a purely capacitive ac circuit.

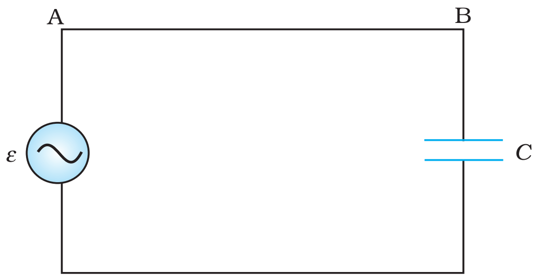


FIGURE 7.7 An ac source connected to a capacitor.

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.

When the capacitor is connected to an ac source, as in Fig. 7.7, 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 $i = \frac{\mathrm{d}q}{\mathrm{d}t}$

$$i = \frac{\mathrm{d}}{\mathrm{d}t}\left(v_m C \sin \omega t\right) = \omega C v_m \cos(\omega t)$$

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

$$i_m = \frac{v_m}{X_C} \tag{7.18}$$

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.

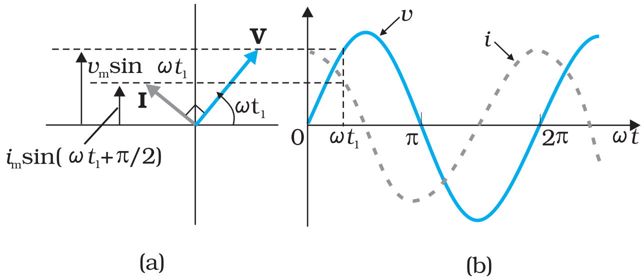


FIGURE 7.8 (a) A Phasor diagram for the circuit in Fig. 7.7. (b) Graph of v and i versus ωt.

A comparison of Eq. (7.16) with the equation of source voltage, Eq. (7.1) shows that the current is π/2 ahead of voltage. Figure 7.8(a) shows the phasor diagram at an instant t_1. Here the current phasor **I** is π/2 ahead of the voltage phasor **V** as they rotate counterclockwise. Figure 7.8(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 (2ωt)> = 0 over a complete cycle.

Thus, we see that in the case of an inductor, the current lags the voltage by π/2 and in the case of a capacitor, the current leads the voltage by π/2.

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 7.3

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

$$X_C = \frac{1}{2\pi\nu C} = \frac{1}{2\pi(50\text{Hz})(15.0\times10^{-6}\text{F})} = 212\,\Omega$$

The rms current is

EXAMPLE 7.4

EXAMPLE 7.4

$$I = \frac{V}{X_C} = \frac{220\text{ V}}{212\ \Omega} = 1.04\text{ A}$$

The peak current is

$$i_m = \sqrt{2}I = (1.41)(1.04\ A) = 1.47\,A$$

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. 7.9.

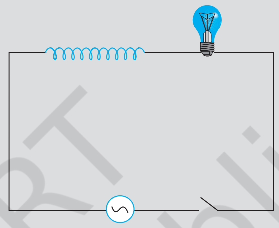


FIGURE 7.9

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 7.10 shows a series LCR circuit connected to an ac source ε. As usual, we take the voltage of the source to be $v = v_m \sin \omega t$.

If q is the charge on the capacitor and i the current, at time t, we have, from Kirchhoff's loop rule:

$$L\frac{di}{dt} + iR + \frac{q}{C} = v \tag{7.20}$$

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.

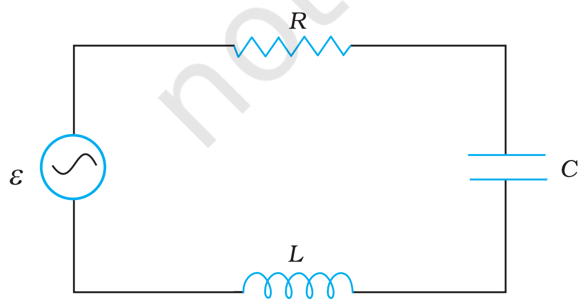


FIGURE 7.10 A series *LCR* circuit connected to an ac source.

7.6.1 Phasor-diagram solution

From the circuit shown in Fig. 7.10, 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. 7.11(a) with apppropriate phase-relations.

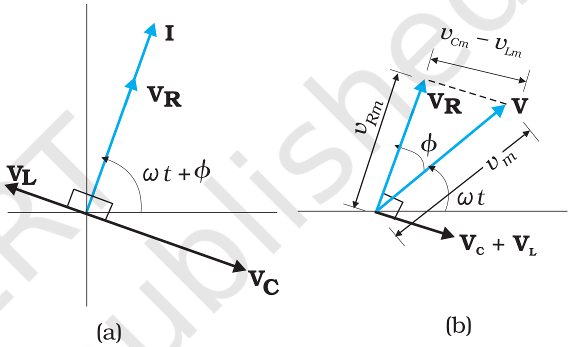


FIGURE 7.11 (a) Relation between the phasors $\mathbf{V}_L$, $\mathbf{V}_R$, $\mathbf{V}_C$, and *I*, (b) Relation between the phasors $\mathbf{V}_L$, $\mathbf{V}_R$, and ($\mathbf{V}_L + \mathbf{V}_C$) for the circuit in Fig. 7.10.

The length of these phasors or the amplitude of $\mathbf{V_R}$, $\mathbf{V_C}$ and $\mathbf{V_L}$ are:

$$\upsilon_{Rm} = i_m R,\ \upsilon_{Cm} = i_m X_C,\ \upsilon_{Lm} = i_m X_L \tag{7.22}$$

The voltage Equation (7.20) for the circuit can be written as

$$\upsilon_L + \upsilon_R + \upsilon_C = \upsilon \tag{7.23}$$

The phasor relation whose vertical component gives the above equation is

$$\mathbf{V_L} + \mathbf{V_R} + \mathbf{V_C} = \mathbf{V} \tag{7.24}$$

This relation is represented in Fig. 7.11(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 $|\upsilon_{Cm} - \upsilon_{Lm}|$. Since **V** is represented as the hypotenuse of a right-triangle whose sides are $\mathbf{V_R}$ and ($\mathbf{V_C} + \mathbf{V_L}$), the pythagorean theorem gives:

$$\upsilon_m^2 = \upsilon_{Rm}^2 + \left(\upsilon_{Cm} - \upsilon_{Lm}\right)^2$$

Substituting the values of υ_{Rm}, υ_{Cm}, and υ_{Lm} from Eq. (7.22) into the above equation, we have

$$\upsilon_m^2 = (i_m R)^2 + (i_m X_C - i_m X_L)^2$$

$$= i_m^2\left[R^2 + (X_C - X_L)^2\right]$$

$$\text{or,}\quad i_m = \frac{\upsilon_m}{\sqrt{R^2 + (X_C - X_L)^2}} \tag{7.25(a)}$$

By analogy to the resistance in a circuit, we introduce the *impedance Z* in an ac circuit:

$$i_m = \frac{\upsilon_m}{Z} \tag{7.25(b)}$$

$$\text{where } Z = \sqrt{R^2 + (X_C - X_L)^2} \tag{7.26}$$

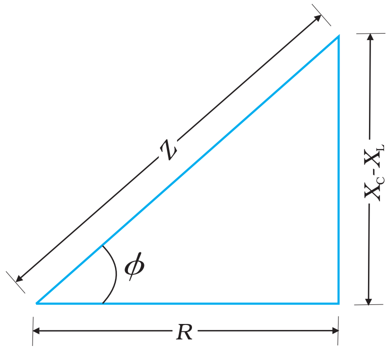


FIGURE 7.12 Impedance diagram.

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. 7.12:

$$\tan\phi = \frac{v_{Cm} - v_{Lm}}{v_{Rm}}$$

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. (7.12). 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 7.13 shows the phasor diagram and variation of v and i with ωt for the case $X_C > X_L$.

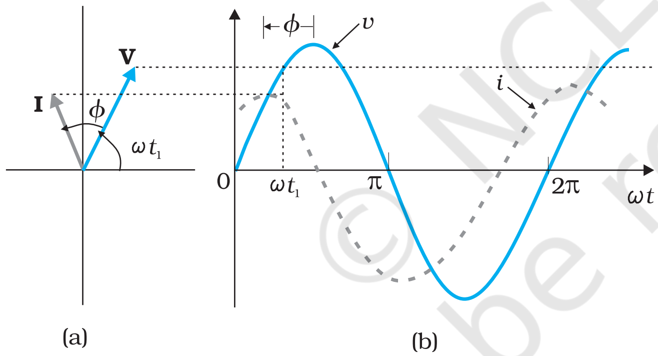


FIGURE 7.13 (a) Phasor diagram of **V** and **I**. (b) Graphs of v and i versus ωt for a series *LCR* circuit where $X_C > X_L$.

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 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 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 13, 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*:

$$X_c = X_L \text{ or } \frac{1}{\omega_0 C} = \omega_0 L$$

$$\text{or} \quad \omega_0 = \frac{1}{\sqrt{LC}} \qquad (7.28)$$

At resonant frequency, the current amplitude is maximum; $i_m = v_m/R$.

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 $\frac{1}{\sqrt{LC}}$ = 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).

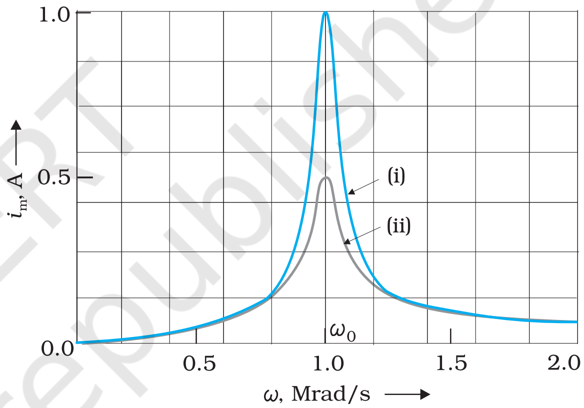


FIGURE 7.14 Variation of i_m with ω for two cases: (i) R = 100 Ω, (ii) R = 200 Ω, L = 1.00 mH.

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.

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

$$Z = \sqrt{R^2 + X_C^2} = \sqrt{R^2 + (2\pi\nu C)^{-2}}$$

$$= \sqrt{(200\,\Omega)^2 + (2 \times 3.14 \times 50 \times 15.0 \times 10^{-6}\,\text{F})^{-2}}$$

$$= \sqrt{(200\,\Omega)^2 + (212.3\,\Omega)^2}$$

$$= 291.67\,\Omega$$

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

$$p = v\,i = (v_m \sin \omega t) \times [i_m \sin(\omega t + \phi)]$$

$$= \frac{v_m i_m}{2}[\cos\phi - \cos(2\omega t + \phi)] \quad (7.29)$$

The average power over a cycle is given by the average of the two terms in R.H.S. of Eq. (7.29). 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,

$$P = \frac{v_m i_m}{2}\cos\phi = \frac{v_m}{\sqrt{2}}\frac{i_m}{\sqrt{2}}\cos\phi$$

$$= V\,I\cos\phi \quad [7.30(a)]$$

This can also be written as,

$$P = I^2\,Z\cos\phi \quad [7.30(b)]$$

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. (7.30) 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 (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. 7.15)

EXAMPLE 7.7

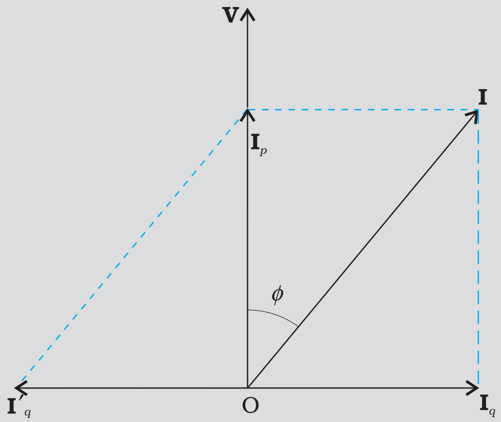


FIGURE 7.15

how this can be achieved. Let us resolve **I** into two components. $\mathbf{I}_p$ along 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.

$$X_L = 2\,\pi\nu L$$
$$= 2 \times 3.14 \times 50 \times 25.48 \times 10^{-3}\ \Omega = 8\ \Omega$$

$$X_C = \frac{1}{2\pi\nu C}$$

$$= \frac{1}{2 \times 3.14 \times 50 \times 796 \times 10^{-6}} = 4\Omega$$

Therefore,

$$Z = \sqrt{R^2 + (X_L - X_C)^2} = \sqrt{3^2 + (8-4)^2}$$
$$= 5\ \Omega$$

(b) Phase difference, $\phi = \tan^{-1}\dfrac{X_C - X_L}{R}$

$$= \tan^{-1}\left(\frac{4-8}{3}\right) = -53.1°$$

EXAMPLE 7.8

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, $I = \dfrac{i_m}{\sqrt{2}} = \dfrac{1}{\sqrt{2}}\left(\dfrac{283}{5}\right) = 40\text{A}$

Therefore, $P = (40\text{A})^2 \times 3\,\Omega = 4800\text{ W}$

(d) Power factor $= \cos\phi = \cos(-53.1°) = 0.6$

EXAMPLE 7.8

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 7.9

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 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. 7.16(a) or on separate limbs of the core as in Fig. 7.16(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.

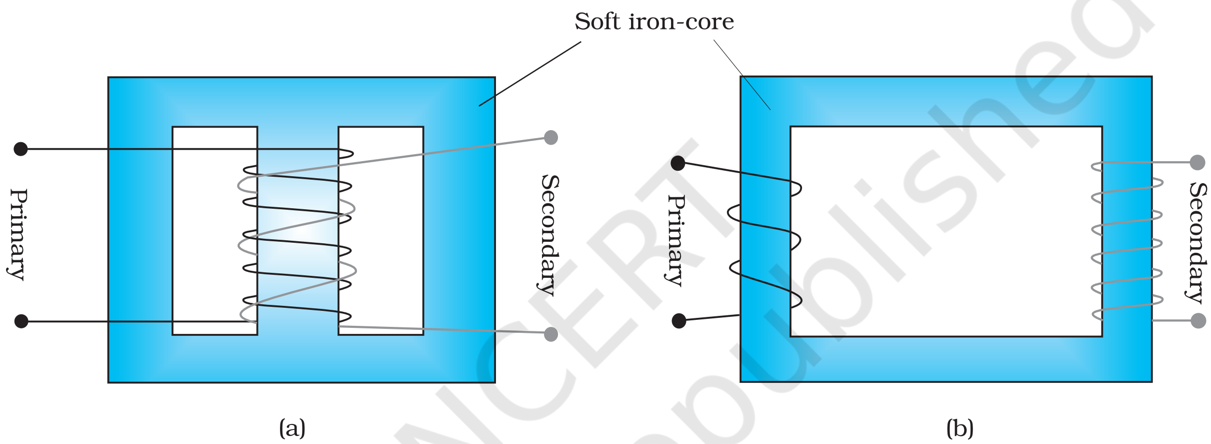


FIGURE 7.16 Two arrangements for winding of primary and secondary coil in a transformer: (a) two coils on top of each other, (b) two coils on separate limbs of the core.

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

$$\varepsilon_s = -N_s \frac{\mathrm{d}\phi}{\mathrm{d}t} \qquad (7.31)$$

The alternating flux ϕ also induces an emf, called back emf in the primary. This is

$$\varepsilon_p = -N_p \frac{\mathrm{d}\phi}{\mathrm{d}t} \qquad (7.32)$$

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. (7.31) and (7.32) can be written as

$$v_s = -N_s \frac{d\phi}{dt} \tag{7.31(a)}$$

$$v_p = -N_p \frac{d\phi}{dt} \tag{7.32(a)}$$

From Eqs. [7.31 (a)] and [7.32 (a)], we have

$$\frac{v_s}{v_p} = \frac{N_s}{N_p} \tag{7.33}$$

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 $p = i\,v$,

$$i_p v_p = i_s v_s \tag{7.34}$$

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. (7.33) and (7.34), we have

$$\frac{i_p}{i_s} = \frac{v_s}{v_p} = \frac{N_s}{N_p} \tag{7.35}$$

Since i and v both oscillate with the same frequency as the ac source, Eq. (7.35) 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:

$$V_s = \left(\frac{N_s}{N_p}\right) V_p \quad \text{and} \quad I_s = \left(\frac{N_p}{N_s}\right) I_p \tag{7.36}$$

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 using 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. 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	$[\mathrm{M\,L^2T^{-3}A^{-1}}]$	V	$V = \dfrac{v_m}{\sqrt{2}}$, v_m is the amplitude of the ac voltage.
rms current	I	[A]	A	$I = \dfrac{i_m}{\sqrt{2}}$, i_m is the amplitude of the ac current.
Reactance:				
Inductive	X_L	$[\mathrm{M\,L^2T^{-3}A^{-2}}]$		$X_L = \omega L$
Capacitive	X_C	$[\mathrm{M\,L^2T^{-3}A^{-2}}]$		$X_C = 1/\omega C$
Impedance	Z	$[\mathrm{M\,L^2T^{-3}A^{-2}}]$		Depends on elements present in the circuit.
Resonant frequency	ω_r or ω_0	$[\mathrm{T^{-1}}]$	Hz	$\omega_0 = \dfrac{1}{\sqrt{LC}}$ for a series RLC circuit
Quality factor	Q	Dimensionless		$Q = \dfrac{\omega_0 L}{R} = \dfrac{1}{\omega_0 C R}$ 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 ac 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.

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 A charged 30 μF capacitor is connected to a 27 mH inductor. What is the angular frequency of free oscillations of the circuit?

7.7 A series *LCR* circuit with $R = 20\ \Omega$, $L = 1.5$ H and $C = 35\ \mu$F 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?

7.8 Figure 7.17 shows a series *LCR* circuit connected to a variable frequency 230 V source. $L = 5.0$ H, $C = 80\mu$F, $R = 40\ \Omega$.

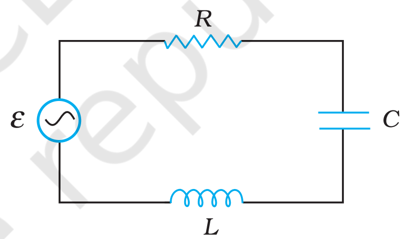


FIGURE 7.17

(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.

12089CH08

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)

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.

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 ~10^{-12} m) to long radio waves (wavelength ~10^6 m) is described.

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 $\mathbf{E}$ 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 $i = (dQ/dt)$, so that using Eq. (8.3), we have

$$\frac{d\Phi_E}{dt} = \frac{d}{dt}\left(\frac{Q}{\varepsilon_0}\right) = \frac{1}{\varepsilon_0}\frac{dQ}{dt}$$

This implies that for consistency,

$$\varepsilon_0\left(\frac{d\Phi_E}{dt}\right) = i \tag{8.4}$$

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

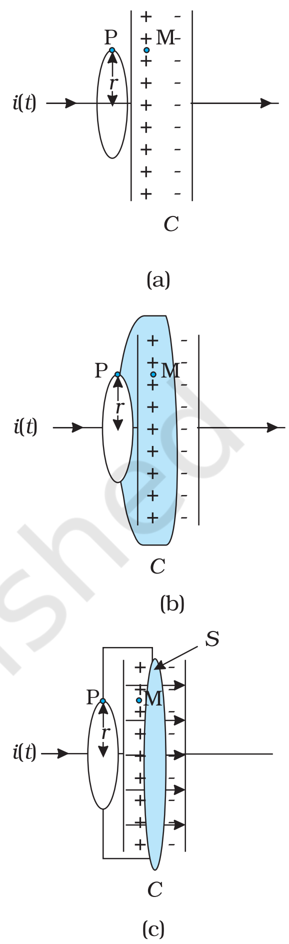
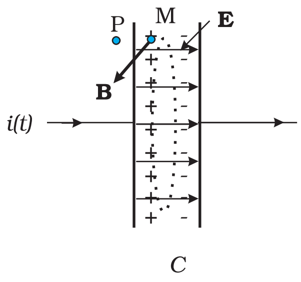
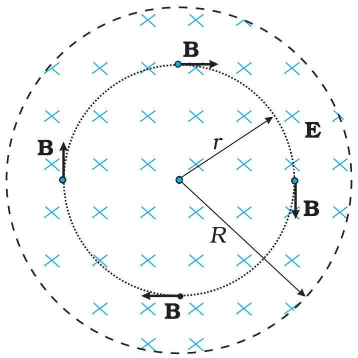


FIGURE 8.1 A parallel plate capacitor C, as part of a circuit through which a time dependent current i (t) flows, (a) a loop of radius r, to determine magnetic field at a point P on the loop; (b) a pot-shaped surface passing through the interior between the capacitor plates with the loop shown in (a) 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.



(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 $(= \varepsilon_0 \, (\mathrm{d}\Phi_E/\mathrm{d}t))$. So we have

$$i = i_c + i_d = i_c + \varepsilon_0 \frac{\mathrm{d}\Phi_E}{\mathrm{d}t} \tag{8.5}$$

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

$$\oint \mathbf{B} \cdot \mathrm{d}\mathbf{l} = \mu_0 \, i_c + \mu_0 \, \varepsilon_0 \frac{\mathrm{d}\Phi_E}{\mathrm{d}t} \tag{8.6}$$

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 IN VACUUM

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)

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

HEINRICH RUDOLF HERTZ (1857–1894)

Heinrich Rudolf Hertz (1857 – 1894) German physicist who was the first to broadcast and receive radio waves. He produced electromagnetic waves, sent them through space, and measured their wavelength 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.

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. 8.3, 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

$$k = \frac{2\pi}{\lambda} \qquad (8.8)$$

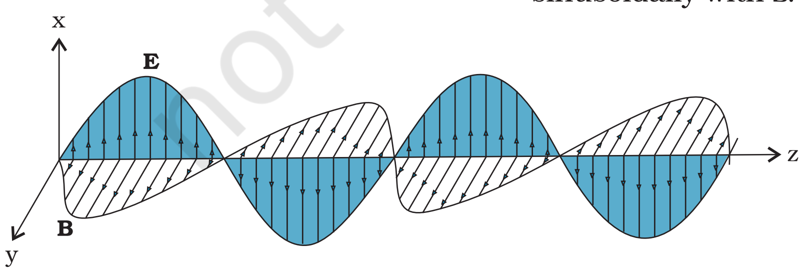


FIGURE 8.3 A linearly polarised electromagnetic wave, propagating in the *z*-direction with the oscillating electric field **E** along the *x*-direction and the oscillating magnetic field **B** along the *y*-direction.

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

$$\omega = ck, \text{ where, } c = 1/\sqrt{\mu_0 \varepsilon_0} \qquad [8.9(a)]$$

The relation $\omega = ck$ is the standard one for waves (see for example, Section 14.4 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.

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 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*.

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 8.1

Example 8.1 A plane electromagnetic wave of frequency 25 MHz travels in free space along the *x*-direction. At a particular point in space and time, **E** = 6.3 $\hat{\mathbf{j}}$ V/m. What is **B** at this point?

Solution Using Eq. (8.10), the magnitude of **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 **E** is along *y*-direction and the wave propagates along *x*-axis. Therefore, **B** should be in a direction perpendicular to both *x*- and *y*-axes. Using vector algebra, **E × B** should be along *x*-direction. Since, $(+\hat{\mathbf{j}}) \times (+\hat{\mathbf{k}}) = \hat{\mathbf{i}}$, **B** is along the *z*-direction.

Thus, $\mathbf{B} = 2.1 \times 10^{-8}\ \hat{\mathbf{k}}$ T

EXAMPLE 8.2

Example 8.2 The magnetic field in a plane electromagnetic wave is given by $B_y = (2 \times 10^{-7})$ T sin $(0.5\times10^3 x + 1.5\times10^{11} t)$.

(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) $E_0 = B_0 c = 2\times10^{-7}$ T × 3 × 10^8 m/s = 6 × 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

Electromagnetic spectrum
http://www.fnal.gov/pub/inquiring/more/light
http://imagine.gsfc.nasa.gov/docs/science/

PHYSICS

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. 8.4). *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.

We briefly describe these different types of electromagnetic waves, in order of decreasing wavelengths.

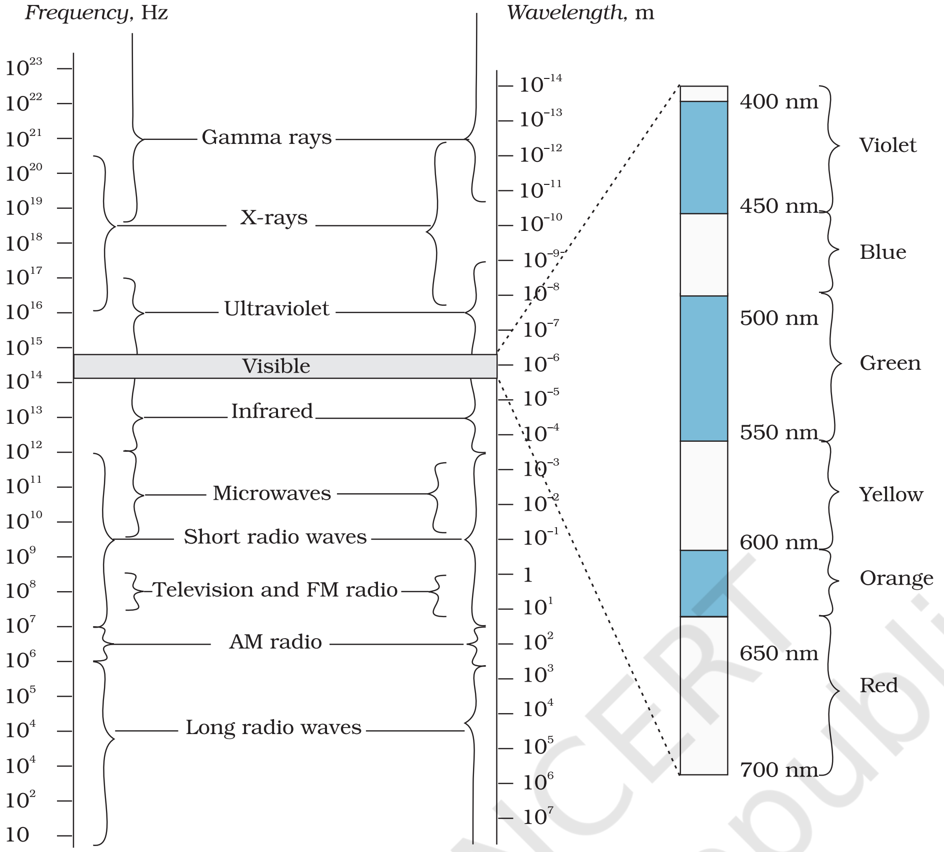


FIGURE 8.4 The electromagnetic spectrum, with common names for various part of it. The various regions do not have sharply defined boundaries.

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.

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 regions is not sharp and there are overlaps.

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. 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. 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*.
4. 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 8.5 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 change 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.

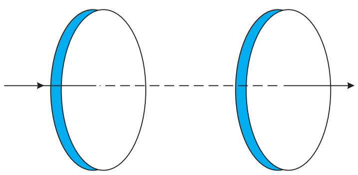


FIGURE 8.5

8.2 A parallel plate capacitor (Fig. 8.6) 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.

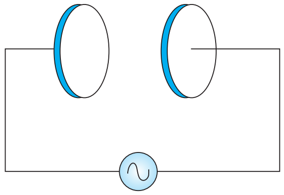


FIGURE 8.6

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}.]

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.

9.2 Reflection of Light by Spherical Mirrors

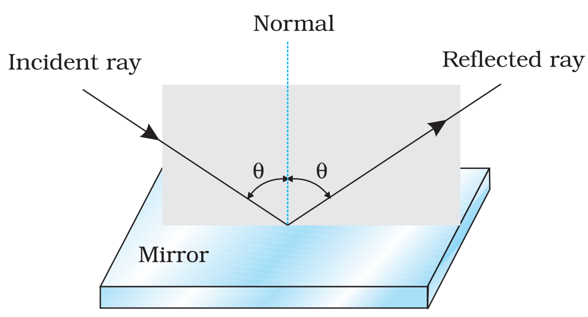


FIGURE 9.1 The incident ray, reflected ray and the normal to the reflecting surface lie in the same plane.

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.

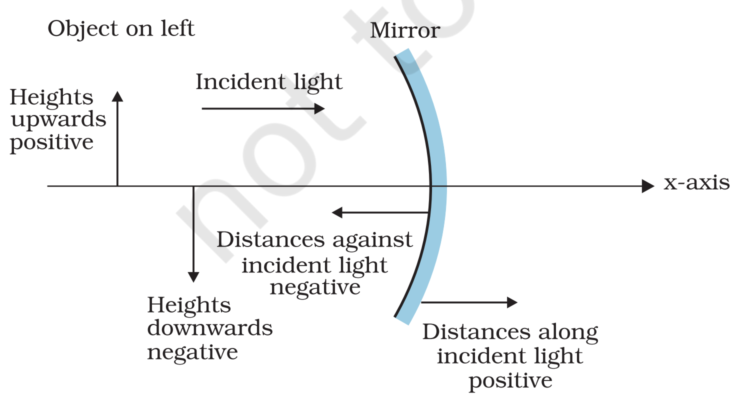


FIGURE 9.2 The Cartesian Sign Convention.

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.

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)].

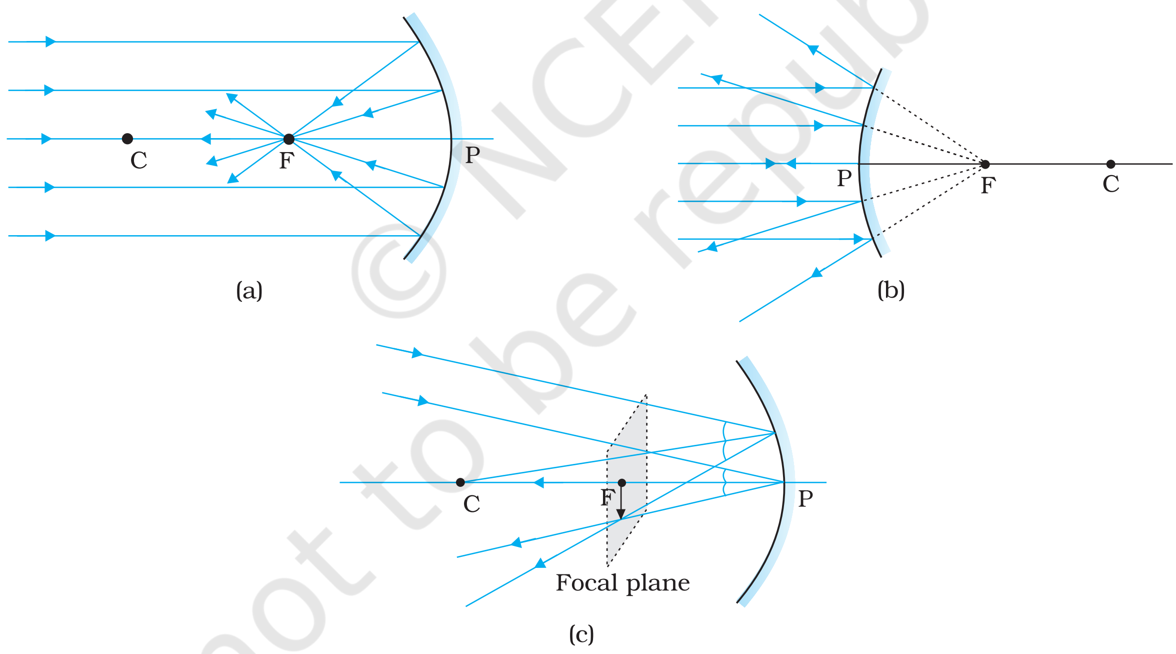


FIGURE 9.3 Focus of a concave and convex mirror.

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.

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

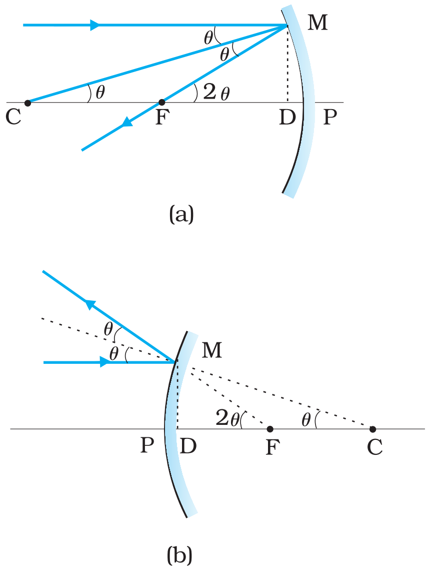


FIGURE 9.4 Geometry of reflection of an incident ray on (a) concave spherical mirror, and (b) convex spherical mirror.

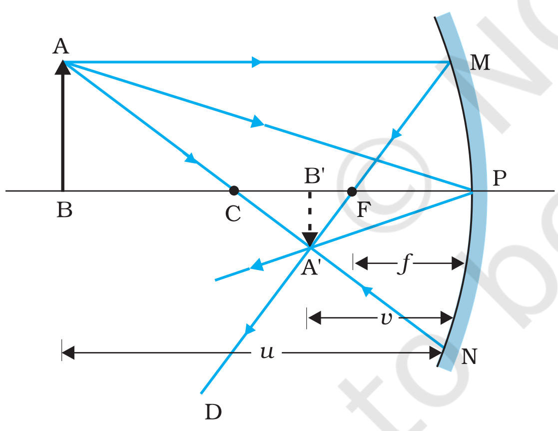


FIGURE 9.5 Ray diagram for image formation by a concave mirror.

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 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,

$$\frac{B'A'}{PM} = \frac{B'F}{FP}$$

$$\text{or} \quad \frac{B'A'}{BA} = \frac{B'F}{FP} \quad (\because PM = AB) \tag{9.4}$$

Since $\angle$ APB = $\angle$ 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

$$\frac{-v+f}{-f} = \frac{-v}{-u}$$

$$\text{or} \quad \frac{v-f}{f} = \frac{v}{u}$$

$$\frac{v}{f} = 1 + \frac{v}{u}$$

Dividing it by v, we get

$$\frac{1}{v} + \frac{1}{u} = \frac{1}{f} \tag{9.7}$$

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

$$m = \frac{h'}{h} = -\frac{v}{u} \qquad (9.9)$$

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.

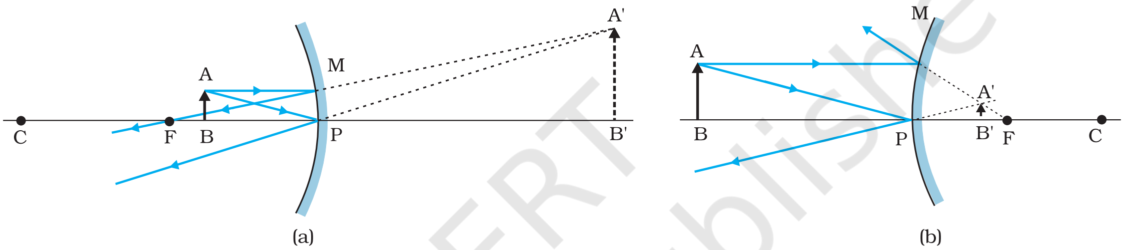


FIGURE 9.6 Image formation by (a) a concave mirror with object between P and F, and (b) a convex mirror.

EXAMPLE 9.1

Example 9.1 Suppose that the lower half of the concave mirror's reflecting surface in Fig. 9.6 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 9.2

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?

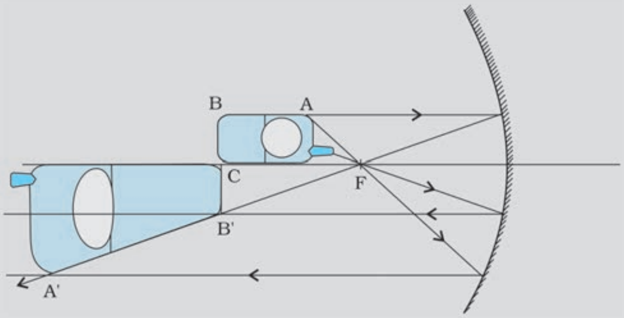


FIGURE 9.7

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.

(ii) The object distance u = –5 cm. Then from Eq. (9.7),

$$\frac{1}{v}+\frac{1}{-5}=\frac{1}{-7.5}$$

or $$v=\frac{5\times 7.5}{(7.5-5)}=15 \text{ cm}$$

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 9.3

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, $v = \frac{(-39)\times 1}{-39-1}=\frac{39}{40}$ 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.

EXAMPLE 9.4

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

$$\frac{1}{150}\,\text{m s}^{-1},\ \frac{1}{60}\,\text{m s}^{-1}\text{ and }\frac{1}{10}\,\text{m s}^{-1},\ \text{respectively.}$$

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.

EXAMPLE 9.4

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 (0°< i < 90°) 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:

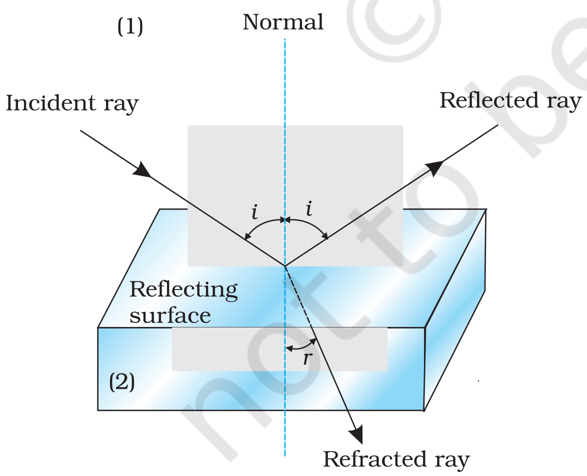


FIGURE 9.8 Refraction and reflection of light.

(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

$$\frac{\sin i}{\sin r} = n_{21} \qquad (9.10)$$

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.*

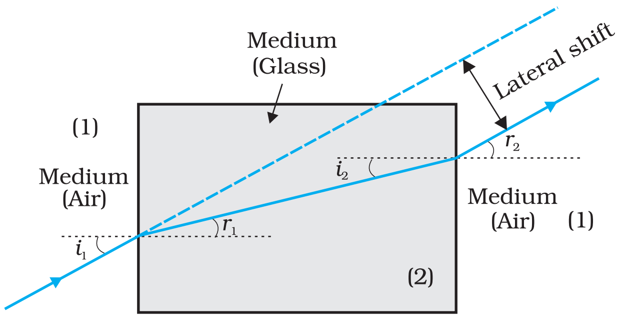


FIGURE 9.9 Lateral shift of a ray refracted through a parallel-sided slab.

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

$$n_{12} = \frac{1}{n_{21}} \tag{9.11}$$

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.

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 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).

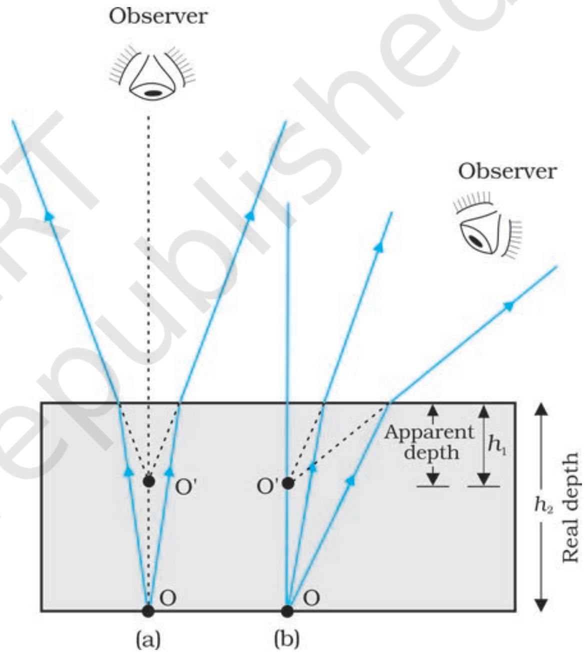


FIGURE 9.10 Apparent depth for (a) normal, and (b) oblique viewing.

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_1 B in Fig. 9.11. 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

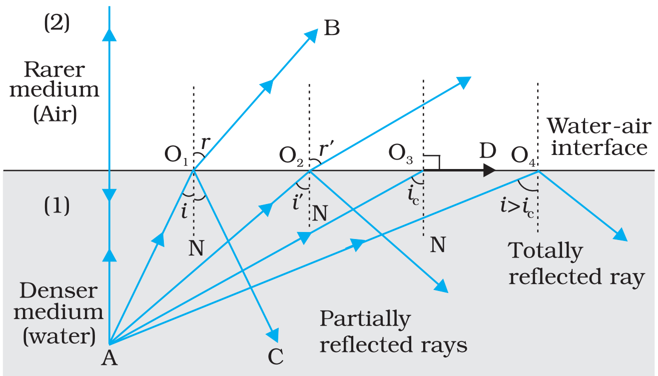


FIGURE 9.11 Refraction and internal reflection of rays from a point A in the denser medium (water) incident at different angles at the interface with a rarer medium (air).

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. 9.11. 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. 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 of the refracting medium 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 1 with respect to rarer medium 2 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 WITH RESPECT TO AIR

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. Add a few drops of milk or any other suspension to water and stir so that water 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. 9.12(a)]? Now direct the laser beam from one side of the beaker such that it strikes the upper surface of water more obliquely [Fig. 9.12(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. 9.12(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.

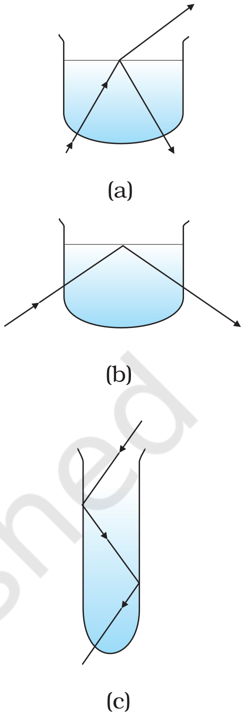


FIGURE 9.12 Observing total internal reflection in water with a laser beam (refraction due to glass of beaker neglected being very thin).

9.4.1 Total internal reflection in nature and its technelogical applications

(i) *Prism*: Prisms designed to bend light by 90° or by 180° make use of total internal reflection [Fig. 9.13(a) and (b)]. Such a prism is also used to invert images without chxanging their size [Fig. 9.13(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.

(ii) *Optical fibres*: Nowadays 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.

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. 9.14). 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

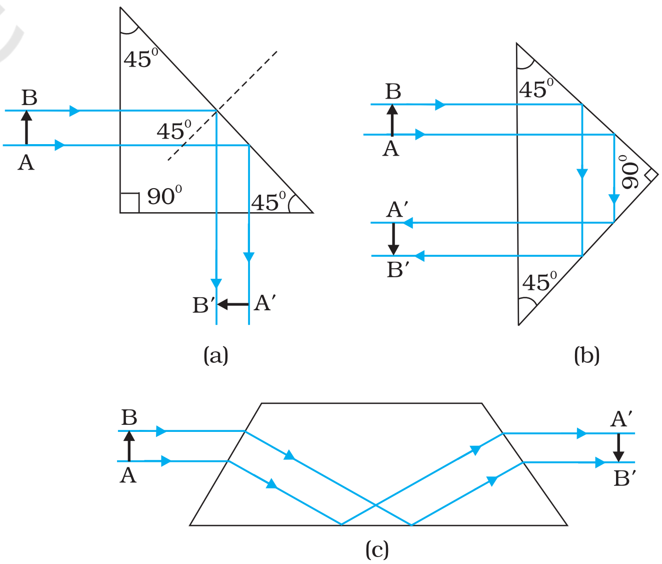


FIGURE 9.13 Prisms designed to bend rays by 90° and 180° or to invert image without changing its size make use of total internal reflection.

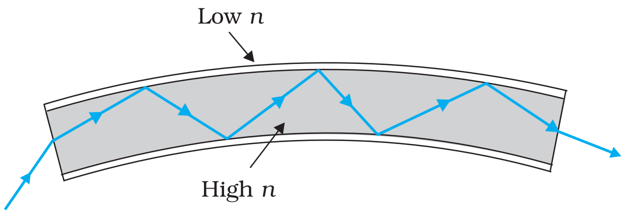


FIGURE 9.14 Light undergoes successive total internal reflections as it moves through an optical fibre.

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 9.15 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,

$$\tan \angle NOM = \frac{MN}{OM}$$

$$\tan \angle NCM = \frac{MN}{MC}$$

$$\tan \angle NIM = \frac{MN}{MI}$$

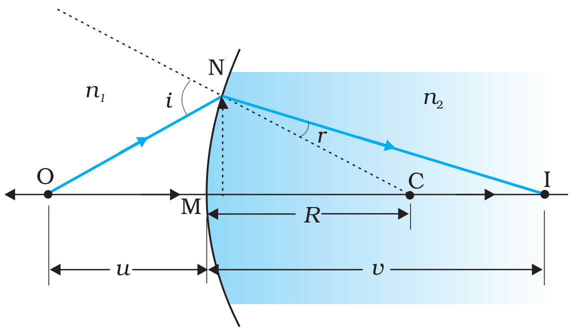


FIGURE 9.15 Refraction at a spherical surface separating two media.

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,

$$r = \angle NCM - \angle NIM$$

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

$$\frac{n_1}{OM} + \frac{n_2}{MI} = \frac{n_2 - n_1}{MC} \tag{9.15}$$

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 9.5 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 9.5

9.5.2 Refraction by a lens

Figure 9.16(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. 9.16(b)]. The image I_1 acts as a virtual object for the second surface that forms the image at I [Fig. 9.16(c)]. Applying Eq. (9.15) to the first interface ABC, we get

$$\frac{n_1}{OB}+\frac{n_2}{BI_1}=\frac{n_2-n_1}{BC_1} \qquad (9.17)$$

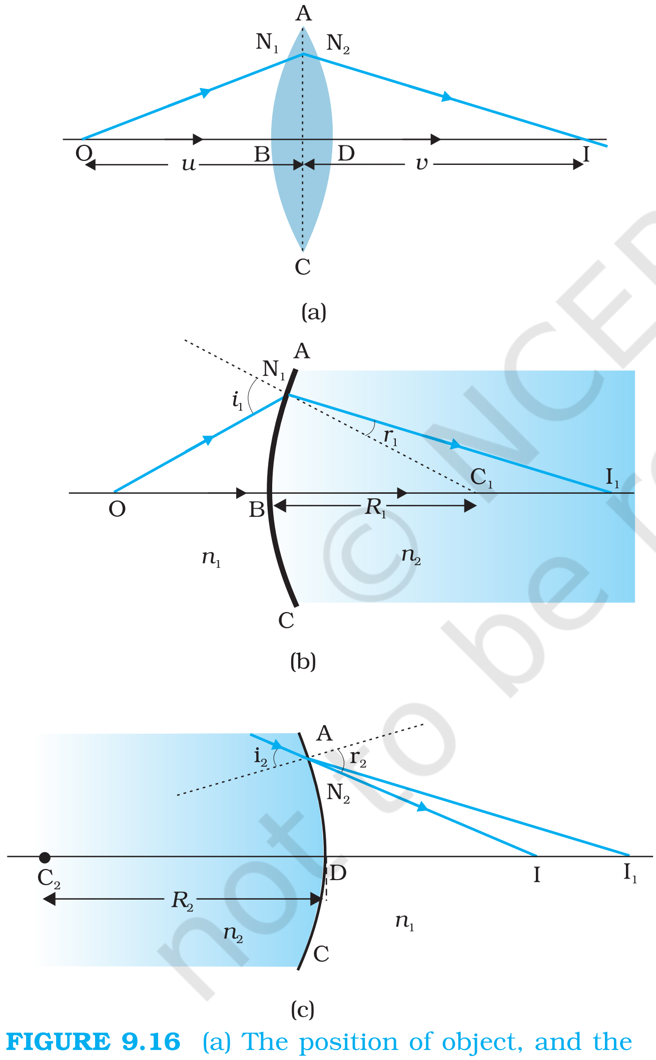


FIGURE 9.16 (a) The position of object, and the image formed by a double convex lens, (b) Refraction at the first spherical surface and (c) Refraction at the second spherical surface.

A similar procedure applied to the second interface* ADC gives,

$$-\frac{n_2}{DI_1}+\frac{n_1}{DI}=\frac{n_2-n_1}{DC_2} \qquad (9.18)$$

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. 9.17). 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.

* 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.

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) A ray of light passing through the first principal focus of a convex lens [Fig. 9.17(a)] emerges parallel to the principal axis after refraction.

(b) A ray of light incident on a concave lens appearing to meet the principal axis at second focus point emerges parallel to the principal axis after refraction [Fig. 9.17(b)].

Figures 9.17(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.

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

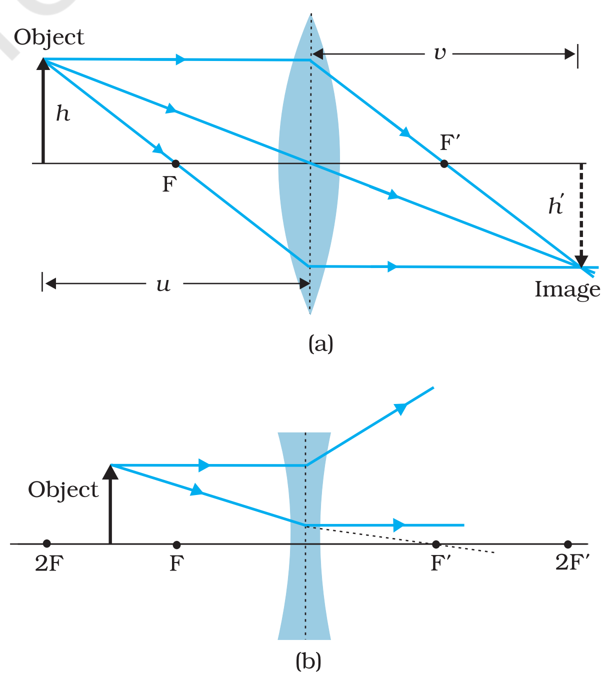


FIGURE 9.17 Tracing rays through (a) convex lens (b) concave lens.

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 9.6

Example 9.6 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.

9.5.3 Power of a lens

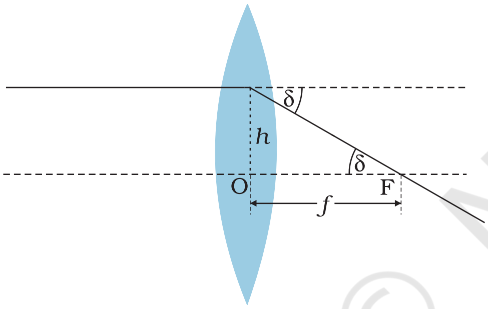


FIGURE 9.18 Power of a lens.

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 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 parallel to the principal axis falling at unit distance from the optical centre (Fig. 9.18).

$\tan\delta = \frac{h}{f}$; 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): $1\text{D} = 1\text{m}^{-1}$. 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 9.7

Example 9.7 (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.)

EXAMPLE 9.7

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,

$$\frac{1.33}{f} = (1.5 - 1.33)\left[\frac{1}{R_1} - \frac{1}{R_2}\right] \qquad (9.26)$$

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. 9.19). 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.

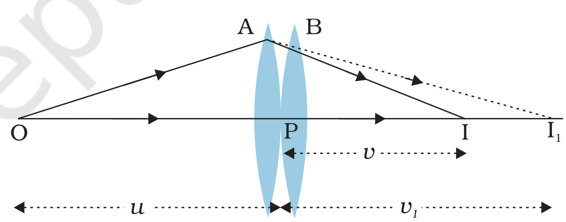


FIGURE 9.19 Image formation by a combination of two thin lenses in contact.

For the image formed by the first lens A, we get

$$\frac{1}{v_1} - \frac{1}{u} = \frac{1}{f_1} \qquad (9.27)$$

For the image formed by the second lens B, we get

$$\frac{1}{v} - \frac{1}{v_1} = \frac{1}{f_2} \qquad (9.28)$$

Adding Eqs. (9.27) and (9.28), we get

$$\frac{1}{v} - \frac{1}{u} = \frac{1}{f_1} + \frac{1}{f_2} \qquad (9.29)$$

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 $f_1, f_2, f_3,...$ 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 9.8

Example 9.8 Find the position of the image formed by the lens combination given in the Fig. 9.20.

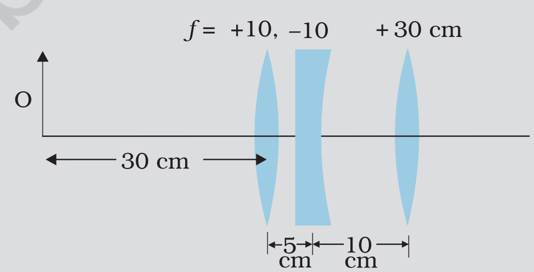


FIGURE 9.20

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 9.8

9.6 Refraction through a Prism

Figure 9.21 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.

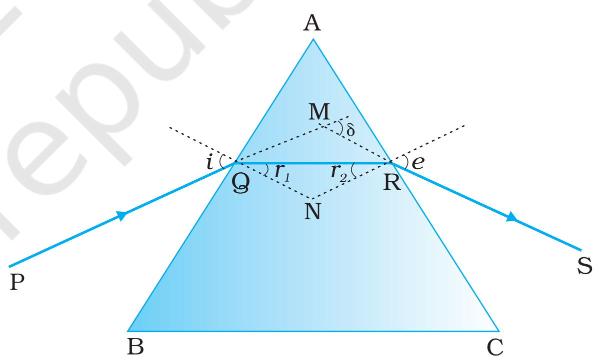


FIGURE 9.21 A ray of light passing through a triangular glass prism.

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

$$r_1 + r_2 = A \tag{9.34}$$

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. 9.22. 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

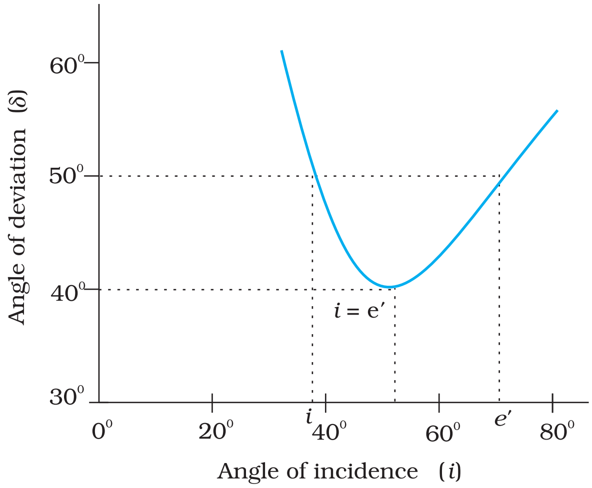


FIGURE 9.22 Plot of angle of deviation (δ) versus angle of incidence (i) for a triangular prism.

and e are interchanged. Physically, this is related to the fact that the path of ray in Fig. 9.21 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

$$n_{21} = \frac{n_2}{n_1} = \frac{\sin[(A + D_m)/2]}{\sin[A/2]} \tag{9.38}$$

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

$$n_{21} = \frac{\sin[(A + D_m)/2]}{\sin[A/2]} \simeq \frac{(A + D_m)/2}{A/2}$$

$$D_m = (n_{21} - 1)A$$

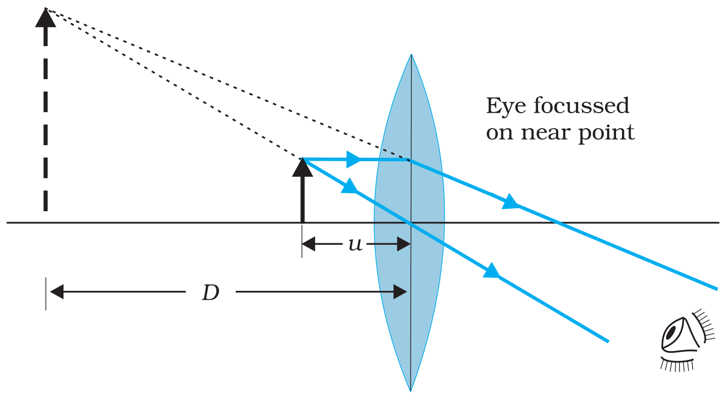
It implies that, thin prisms do not deviate light much.

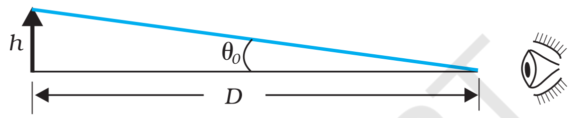
9.7 Optical Instruments

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. We have already studied about the human eye in Class X. We now go on to describe the principles of working of the microscope and the telescope.

9.7.1 The microscope

A simple magnifier or microscope is a converging lens of small focal length (Fig. 9.23). 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





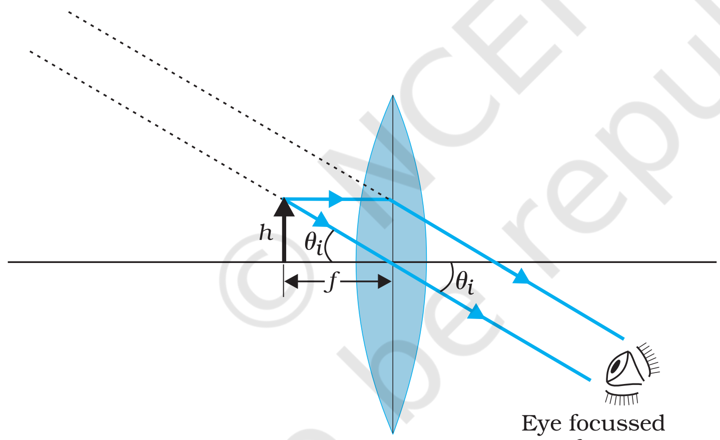


FIGURE 9.23 A simple microscope; (a) the magnifying lens is located such that the image is at the near point, (b) the angle subtanded by the object, is the same as that at the near point, and (c) the object near the focal point of the lens; the image is far off but closer than infinity.

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. 9.23(a), and the second in Fig. 9.23(b) and (c).

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

$$m = \left(1 + \frac{D}{f}\right) \tag{9.39}$$

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

$$\tan \theta_o = \left(\frac{h}{D}\right) \approx \theta_o \tag{9.40}$$

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. 9.23(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.

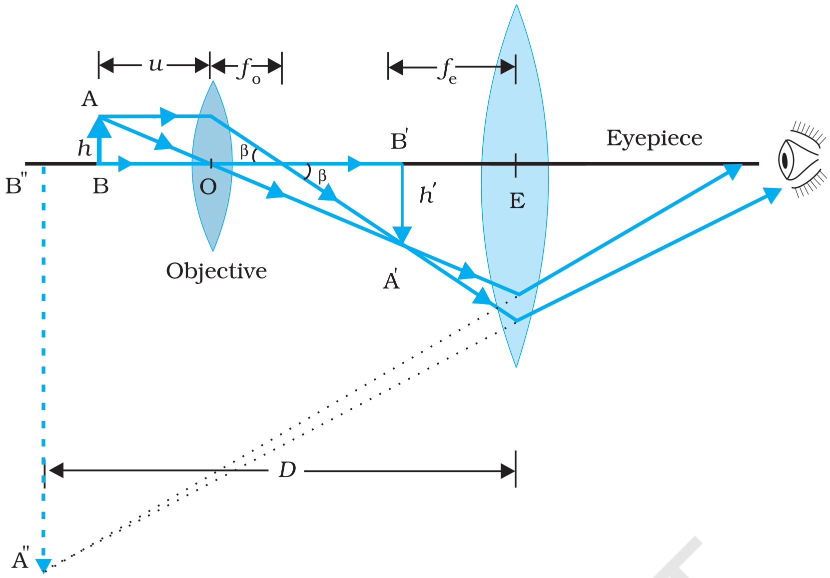


FIGURE 9.24 Ray diagram for the formation of image by a compound microscope.

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. 9.24. 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.

We now obtain the magnification due to a compound microscope. The ray diagram of Fig. 9.24 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.

9.7.2 Telescope

The telescope is used to provide angular magnification of distant objects (Fig. 9.25). 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

$$m \approx \frac{\beta}{\alpha} \approx \frac{h}{f_e} . \frac{f_o}{h} = \frac{f_o}{f_e} \qquad (9.46)$$

In this case, the length of the telescope tube is $f_o + f_e$.

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°.

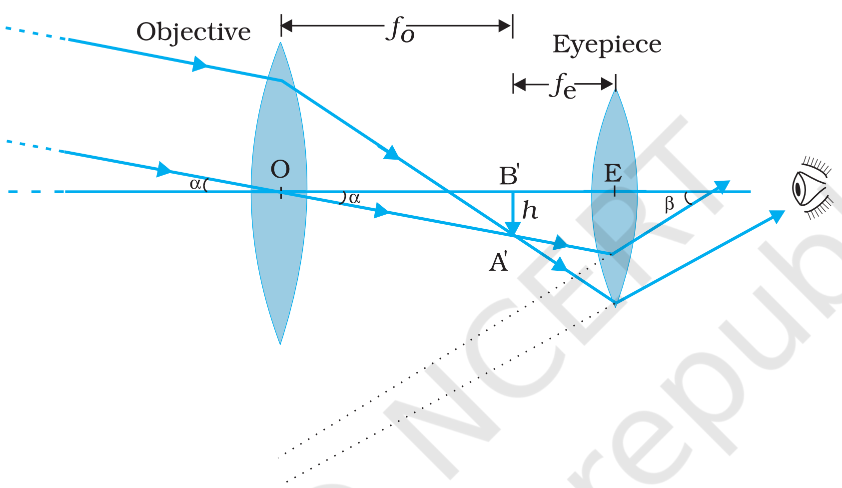


FIGURE 9.25 A refracting telescope.

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. There is no chromatic aberration in a mirror. 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

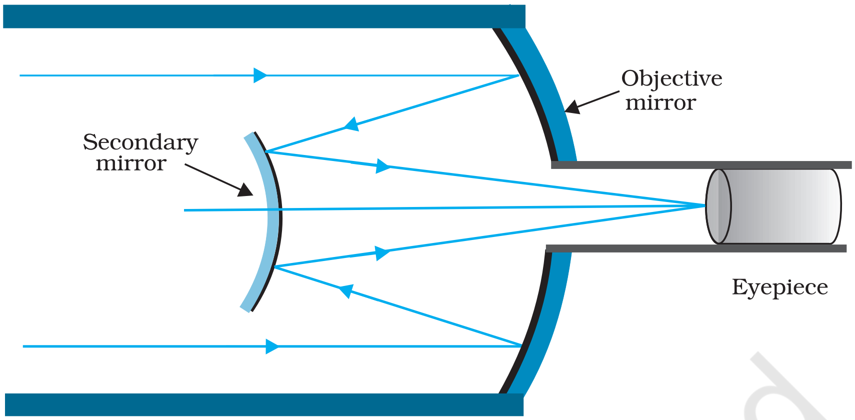


FIGURE 9.26 Schematic diagram of a reflecting telescope (Cassegrain).

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 in Fig. 9.26. 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

$$\frac{1}{f}=\frac{\left(n_2-n_1\right)}{n_1}\left(\frac{1}{R_1}-\frac{1}{R_2}\right)$$

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 colour.

8. *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.

9. *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 9.27(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. 9.27(c)].

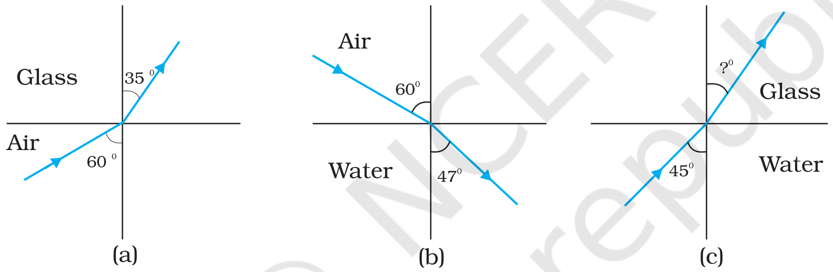


FIGURE 9.27

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 30 cm in contact with a concave lens of focal length 20 cm? 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 2.0 cm and an eyepiece of focal length 6.25 cm separated by a distance of 15 cm. 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 (25 cm), and (b) at infinity? What is the magnifying power of the microscope in each case?

9.12 A person with a normal near point (25 cm) using a compound microscope with objective of focal length 8.0 mm and an eyepiece of focal length 2.5 cm can bring an object placed at 9.0 mm 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 144 cm and an eyepiece of focal length 6.0 cm. 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 15 m. If an eyepiece of focal length 1.0 cm 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 3.48×10^6 m, and the radius of lunar orbit is 3.8×10^8 m.

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 9.28 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.

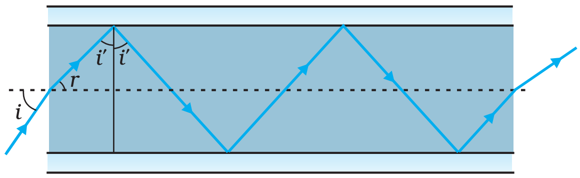


FIGURE 9.28

(b) What is the answer if there is no outer covering of the pipe?

9.18 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?

9.19 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.

9.20 (a) Determine the 'effective focal length' of the combination of the two lenses in Exercise 9.10, if they are placed 8.0cm 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.

9.21 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.

9.22 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.

9.23 (a) At what distance should the lens be held from the card sheet in Exercise 9.22 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.

9.24 What should be the distance between the object in Exercise 9.23 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 9.22 to 9.24 will help you clearly understand the difference between magnification in absolute size and the angular magnification (or magnifying power) of an instrument.]

9.25 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?

9.26 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?

9.27 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)?

9.28 (a) For the telescope described in Exercise 9.27 (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?

9.29 A Cassegrain telescope uses two mirrors as shown in Fig. 9.26. 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?

9.30 Light incident normally on a plane mirror attached to a galvanometer coil retraces backwards as shown in Fig. 9.29. 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?

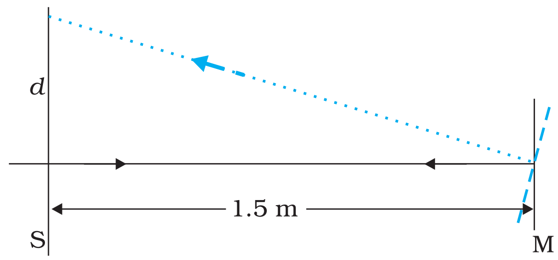


FIGURE 9.29

9.31 Figure 9.30 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?

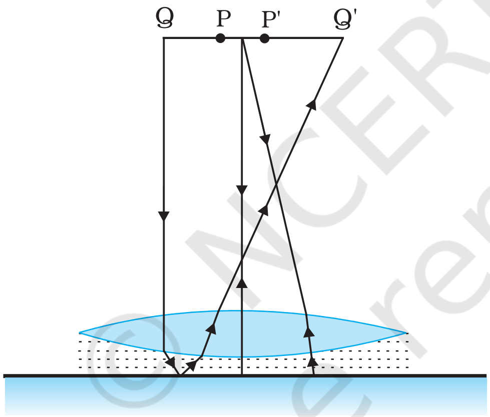


FIGURE 9.30

Notes

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 0.6 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*

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.

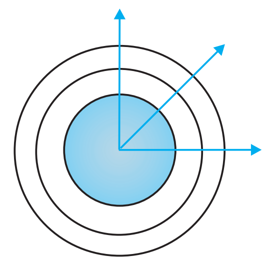


FIGURE 10.1 (a) A diverging spherical wave emanating from a point source. The wavefronts are spherical.

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 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)].

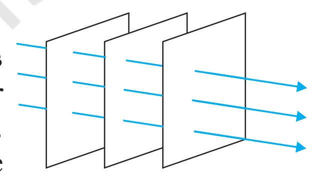


FIGURE 10.1 (b) At a large distance from the source, a small portion of the spherical wave can be approximated by a plane wave.

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.*

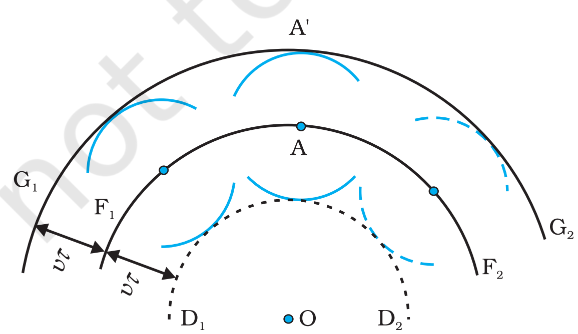


FIGURE 10.2 F_1F_2 represents the spherical wavefront (with O as centre) at $t = 0$. The envelope of the secondary wavelets emanating from F_1F_2 produces the forward moving wavefront G_1G_2. The backwave D_1D_2 does not exist.

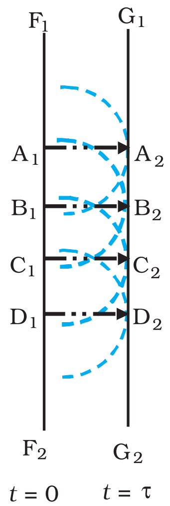


FIGURE 10.3 Huygens geometrical construction for a plane wave propagating to the right. F_1 F_2 is the plane wavefront at $t = 0$ and G_1G_2 is the wavefront at a later time τ. The lines A_1A_2, B_1B_2 … etc., are normal to both F_1F_2 and G_1G_2 and represent rays.

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,

$$BC = v_1 \tau$$

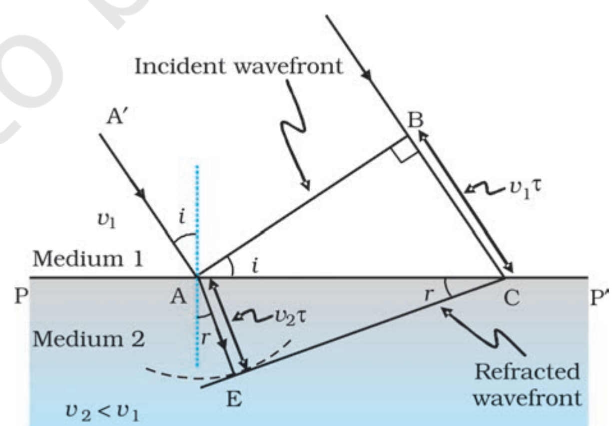


FIGURE 10.4 *A plane wave* AB is incident at an angle i on the surface PP′ separating medium 1 and medium 2. The plane wave undergoes refraction and CE represents the refracted wavefront. The figure corresponds to $v_2 < v_1$ so that the refracted waves bends towards the normal.

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

$$\sin i = \frac{\text{BC}}{\text{AC}} = \frac{v_1\tau}{\text{AC}} \tag{10.1}$$

and

$$\sin r = \frac{\text{AE}}{\text{AC}} = \frac{v_2\tau}{\text{AC}} \tag{10.2}$$

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

$$\frac{v_1}{\lambda_1} = \frac{v_2}{\lambda_2} \tag{10.7}$$

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)

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 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.

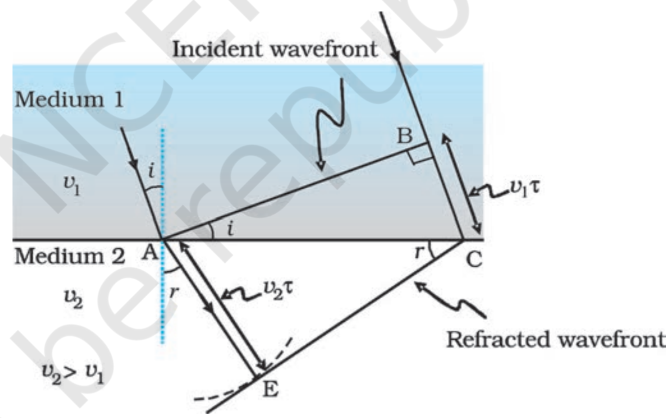


FIGURE 10.5 Refraction of a plane wave incident on a rarer medium for which $v_2 > v_1$. The plane wave bends away from the normal.

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 to 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

$$AE = BC = v\tau$$

Demonstration of interference, diffraction, refraction, resonance and Doppler effect
http://www.falstad.com/ripple/

PHYSICS

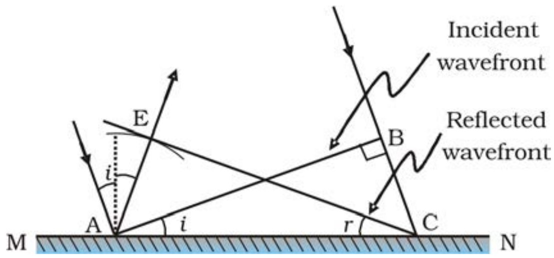


FIGURE 10.6 Reflection of a plane wave AB by the reflecting surface MN. AB and CE represent incident and reflected wavefronts.

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.

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.

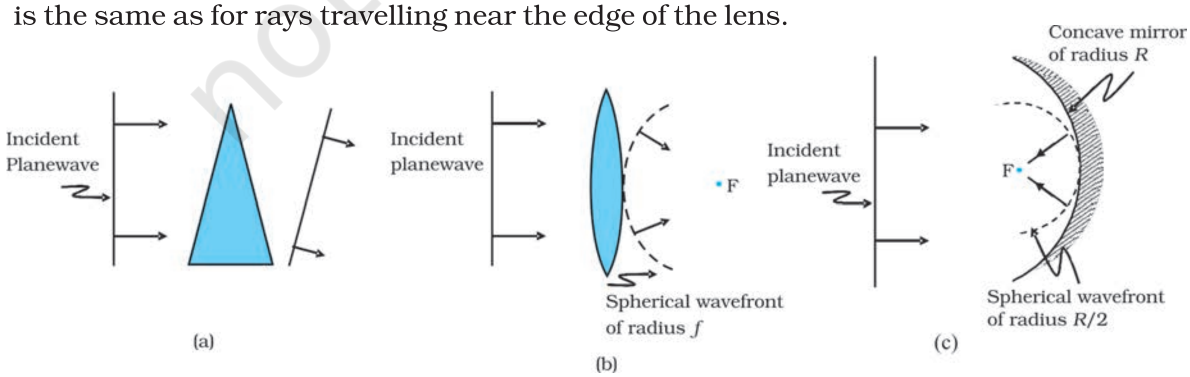


FIGURE 10.7 Refraction of a plane wave by (a) a thin prism, (b) a convex lens. (c) Reflection of a plane wave by a concave mirror.

EXAMPLE 10.1

Example 10.1

(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?

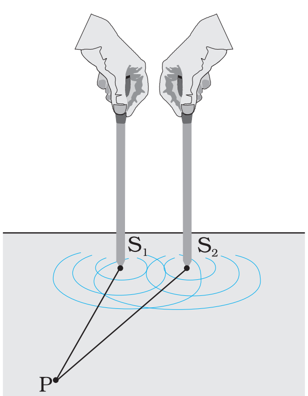
(c) In the wave picture of light, intensity of light is determined by the square of the amplitude of the wave. What determines the intensity of light in the photon picture of light.

Solution

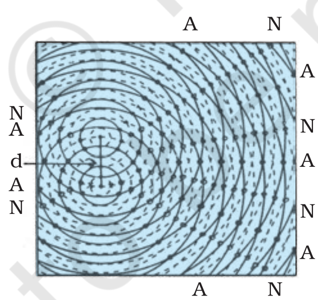
(a) Reflection and refraction arise through interaction of incident light with the atomic constituents of matter. Atoms may be viewed as oscillators, which take up the frequency of the external agency (light) causing forced oscillations. The frequency of light emitted by a charged oscillator equals its 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.



(a)



(b)

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 14 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_1P and S_2P 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 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

$$y_2 = a \cos (\omega t + 5\pi) = -\, a \cos \omega t$$

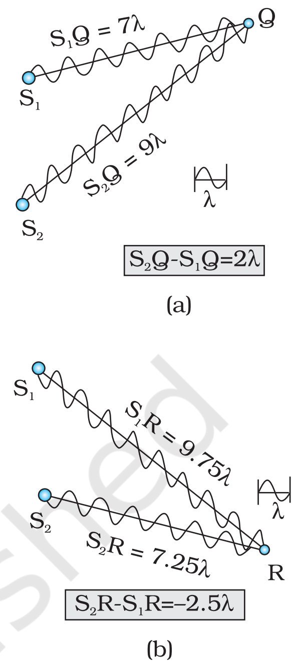


FIGURE 10.9 (a) Constructive interference at a point Q for which the path difference is 2λ. (b) Destructive interference at a point R for which the path difference is $2.5\,\lambda$.

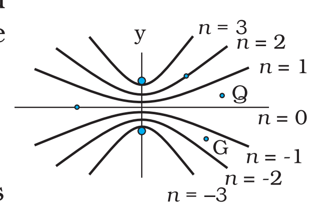


FIGURE 10.10 Locus of points for which $S_1P - S_2P$ is equal to zero, $\pm\lambda$, $\pm 2\lambda$, $\pm 3\lambda$.

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,

$$S_1P \sim S_2P = n\lambda \quad (n = 0, 1, 2, 3, ...) \tag{10.9}$$

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,

$$S_1P \sim S_2P = (n + \frac{1}{2})\lambda \quad (n = 0, 1, 2, 3, ...) \tag{10.10}$$

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

$$y_1 = a \cos \omega t$$

then, the displacement produced by S_2 would be

$$y_2 = a \cos(\omega t + \phi)$$

and the resultant displacement will be given by

$$\begin{aligned} y &= y_1 + y_2 \\ &= a [\cos \omega t + \cos(\omega t + \phi)] \\ &= 2\, a \cos(\phi/2) \cos(\omega t + \phi/2) \end{aligned}$$

$$\left[\because \cos A + \cos B = 2\cos\left(\frac{A+B}{2}\right)\cos\left(\frac{A-B}{2}\right)\right]$$

The amplitude of the resultant displacement is $2a \cos(\phi/2)$ and therefore the intensity at that point will be

$$I = 4\, I_0 \cos^2(\phi/2) \tag{10.11}$$

If $\phi = 0, \pm 2\pi, \pm 4\pi, ...$ which corresponds to the condition given by Eq. (10.9) we will have constructive interference leading to maximum intensity. On the other hand, if $\phi = \pm\pi, \pm 3\pi, \pm 5\pi$ ... [which corresponds to the condition given by Eq. (10.10)] 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

$$I = 2\, I_0 \tag{10.12}$$

at all points.

Ripple Tank experiments on wave interference
http://phet.colorado.edu/en/simulation/legacy/wave-interference

PHYSICS

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.

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.

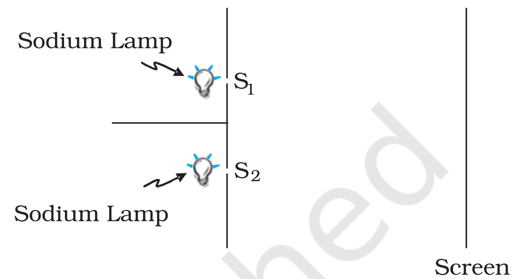
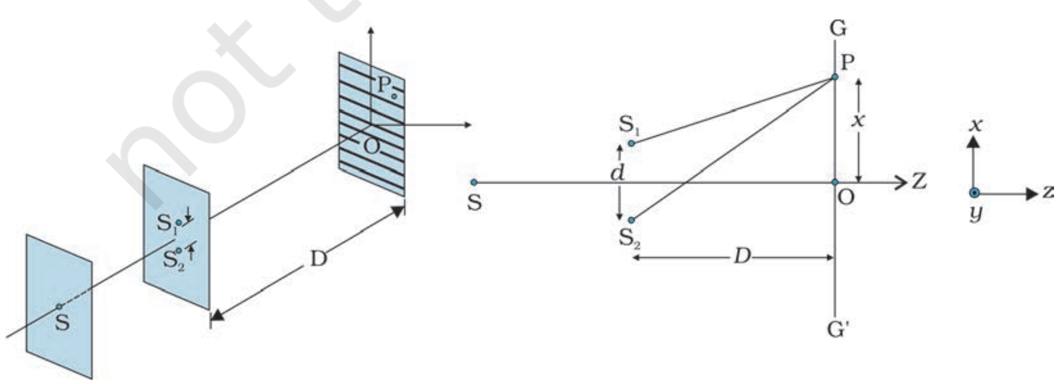


FIGURE 10.11 If two sodium lamps illuminate two pinholes S_1 and S_2, the intensities will add up and no interference fringes will be observed on the screen.

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)].

The 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.



(a) (b)

FIGURE 10.12 Young's arrangement to produce interference pattern.

2024-25

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.

We will have constructive interference resulting in a bright region when $\frac{xd}{D} = n\lambda$. That is,

$$x = x_n = \frac{n\lambda D}{d};\ n = 0, \pm 1, \pm 2, \ldots \qquad (10.13)$$

On the other hand, we will have destructive interference resulting in a dark region when $\frac{xd}{D} = (n + \frac{1}{2})\,\lambda$ that is

$$x = x_n = (n + \frac{1}{2})\,\frac{\lambda D}{d};\quad n = 0, \pm 1, \pm 2 \qquad (10.14)$$

Thus dark and bright bands appear on the screen, as shown in Fig. 10.13. Such bands are called *fringes*. Equations (10.13) and (10.14) show that dark and bright fringes are equally spaced.

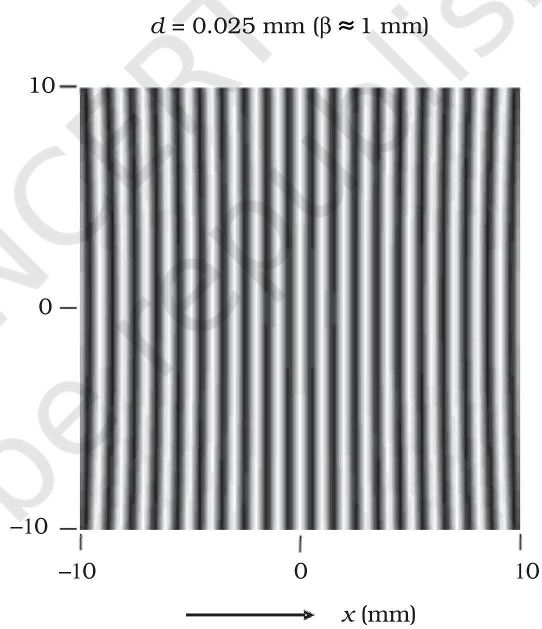


FIGURE 10.13 Computer generated fringe pattern produced by two point source S_1 and S_2 on the screen GG′ (Fig. 10.12); d = 0.025 mm, D = 5 cm and $\lambda = 5 \times 10^{-5}$ cm.) (Adopted from OPTICS by A. Ghatak, Tata McGraw Hill Publishing Co. Ltd., New Delhi, 2000.)

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. 10.15). To understand this, go to Fig. 10.14, 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.

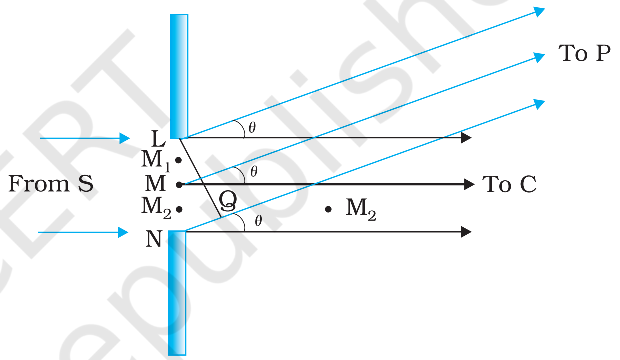


FIGURE 10.14 The geometry of path differences for diffraction by a single slit.

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.

It is observed that the intensity has a central maximum at $\theta = 0$ and other secondary maxima at $\theta \approx (n+1/2)\ \lambda/a$, which go on becoming weaker and weaker with increasing n. The minima (zero intensity) are at $\theta \approx n\lambda/a$, $n = \pm1, \pm2, \pm3, ...$

The photograph and intensity pattern corresponding to it is shown in Fig. 10.15.

There has been prolonged discussion about difference between intereference and diffraction among

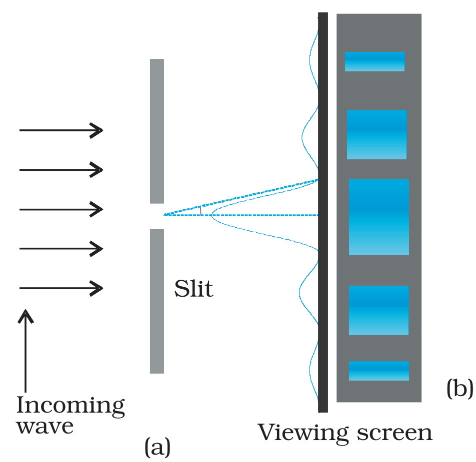


FIGURE 10.15 Intensity distribution and photograph of fringes due to diffraction at single slit.

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.

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. 10.16).

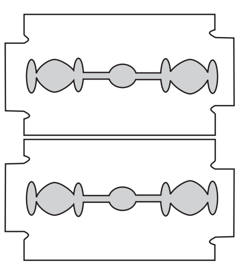


FIGURE 10.16 Holding two blades to form a single slit. A bulb filament viewed through this shows clear diffraction bands.

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. 10.15. 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.

* Richard Feynman was one of the recipients of the 1965 Nobel Prize in Physics for his fundamental work in quantum electrodynamics.

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. 10.17). Such a wave could be described by the following equation

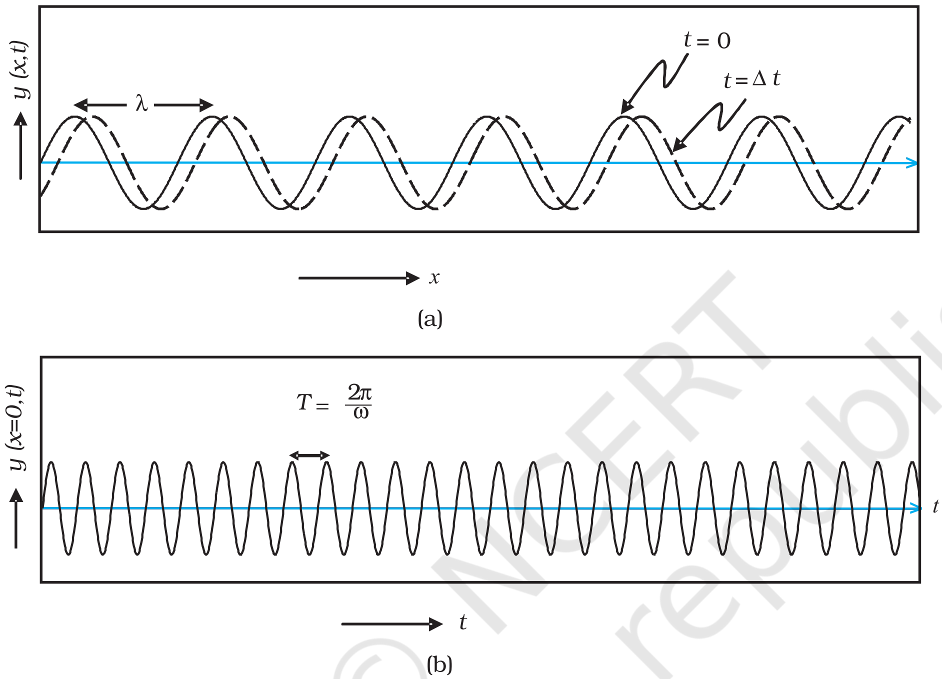


FIGURE 10.17 (a) The curves represent the displacement of a string at $t = 0$ and at $t = \Delta t$, respectively when a sinusoidal wave is propagating in the $+x$-direction. (b) The curve represents the time variation of the displacement at $x = 0$ when a sinusoidal wave is propagating in the $+x$-direction. At $x = \Delta x$, the time variation of the displacement will be slightly displaced to the right.

$$y(x,t) = a \sin (kx - \omega t) \tag{10.15}$$

where a and $\omega (= 2\pi\nu)$ represent the amplitude and the angular frequency of the wave, respectively; further,

$$\lambda = \frac{2\pi}{k} \tag{10.16}$$

represents the wavelength associated with the wave. We had discussed propagation of such waves in Chapter 14 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

$$z(x,t) = a \sin(kx - \omega t) \tag{10.17}$$

It should be mentioned that the linearly polarised waves [described by Eqs. (10.15) and (10.17)] 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. 10.18).

The experiment at figure 10.18 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:

$$I = I_0 \cos^2\theta \tag{10.18}$$

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

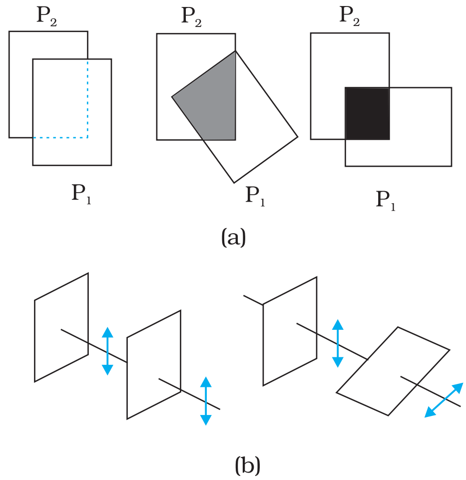


FIGURE 10.18 (a) Passage of light through two polaroids P_2 and P_1. The transmitted fraction falls from 1 to 0 as the angle between them varies from 0° to 90°. Notice that the light seen through a single polaroid P_1 does not vary with angle. (b) Behaviour of the electric vector when light passes through two polaroids. The transmitted polarisation is the component parallel to the polaroid axis. The double arrows show the oscillations of the electric vector.

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 10.2 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

$$I = I_0 \cos^2\theta,$$

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 $\theta = \pi/4$.

EXAMPLE 10.2

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 interference fringes.
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.
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.

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. 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.
4. 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?

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 eV = 1.602×10^{-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 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:

(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) *Photoelectric 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*.

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 thin 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 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.

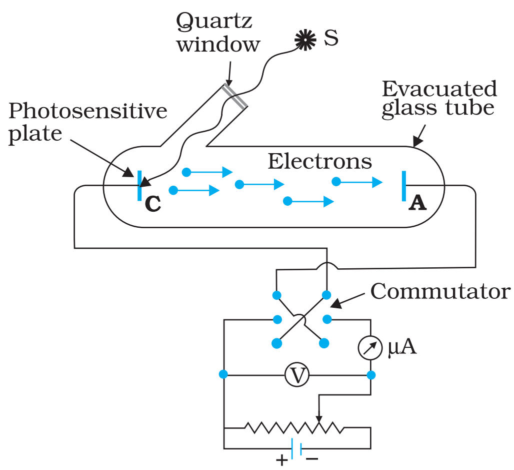


FIGURE 11.1 Experimental arrangement for study of photoelectric effect.

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

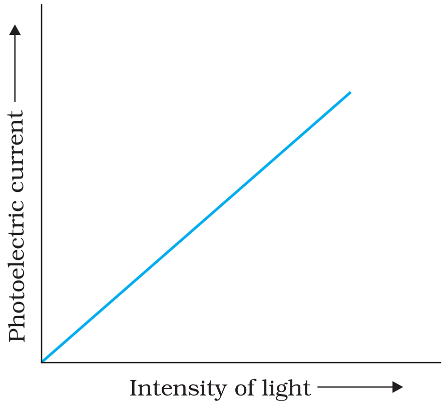


FIGURE 11.2 Variation of Photoelectric current with intensity of light.

on the emitter C. The intensity of light is varied by changing the distance of the light source from the emitter.

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 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 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 positive (accelerating) 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.

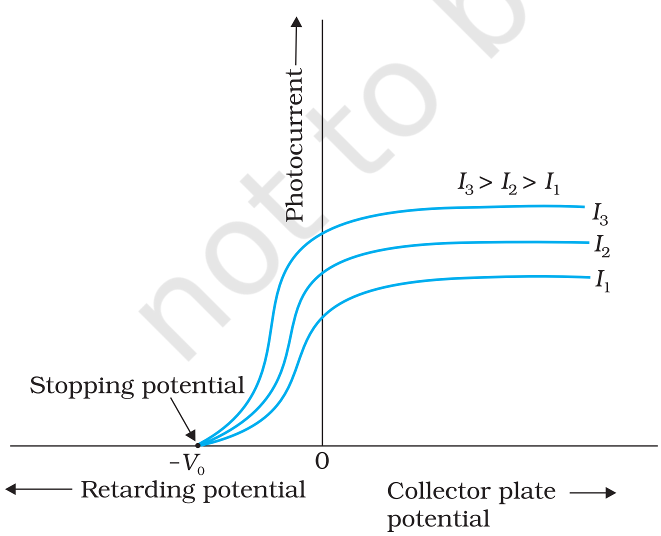


FIGURE 11.3 Variation of photocurrent with collector plate potential for different intensity of incident radiation.

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 sufficiently 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

$$K_{max} = e\,V_0 \qquad (11.1)$$

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 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.

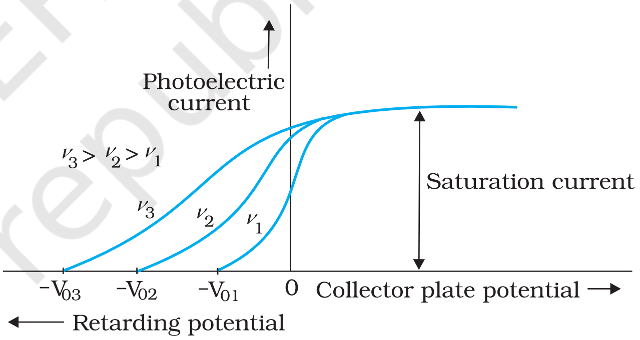


FIGURE 11.4 Variation of photoelectric current with collector plate potential for different frequencies of incident radiation.

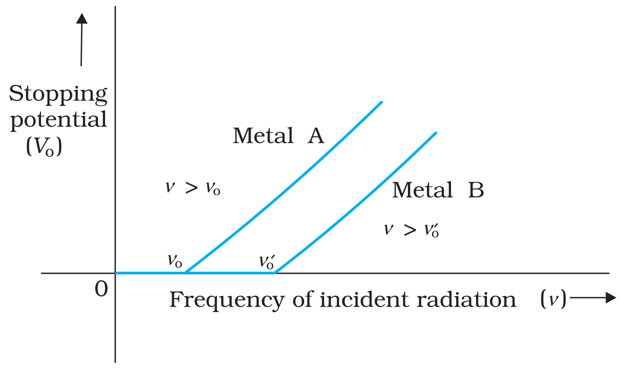


FIGURE 11.5 Variation of stopping potential V_0 with frequency ν of incident radiation for a given photosensitive material.

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

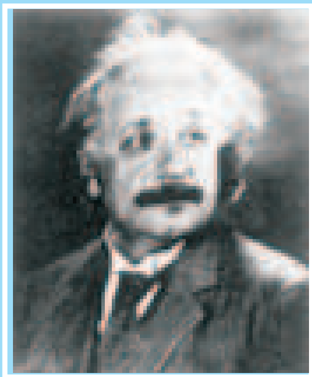
In 1905, Albert Einstein (1879-1955) proposed a radically new picture of electromagnetic radiation to explain photoelectric effect. In this picture, photoelectric emission does not take place by continuous absorption of energy from radiation. Radiation energy is built up of discrete units – the so called *quanta of energy of radiation*. Each quantum of radiant energy 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

$$K_{max} = h\nu - \phi_0 \quad (11.2)$$

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.

ALBERT EINSTEIN (1879 – 1955)



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.

- 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
  $$\nu_0 = \frac{\phi_0}{h} \tag{11.3}$$
  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

$$e V_0 = h\nu - \phi_0;\ \text{for } \nu \geq \nu_0$$

$$\text{or } V_0 = \frac{h}{e}\nu - \frac{\phi_0}{e} \tag{11.4}$$

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 ($= 6.626 \times 10^{-34}$ 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 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
$E = h\,\nu = (6.63 \times 10^{-34}$ J s$)\,(6.0 \times 10^{14}$ Hz$)$
$= 3.98 \times 10^{-19}$ 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

$$N = \frac{P}{E} = \frac{2.0 \times 10^{-3}\,\text{W}}{3.98 \times 10^{-19}\,\text{J}}$$

$= 5.0 \times 10^{15}$ photons per second.

Example 11.1

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 $e\,V_0$ 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

11.8 Wave Nature of Matter

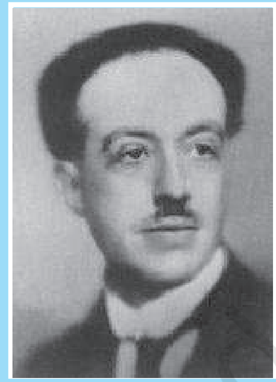
The dual (wave-particle) nature of light (electromagnetic radiation, in general) comes out clearly from what we have learnt in this and the preceding chapters. The wave nature of light shows up in the phenomena of interference, diffraction and polarisation. On the other hand, in photoelectric effect 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.

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.

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:

$$p = m\,v = 0.12 \text{ kg} \times 20 \text{ m s}^{-1} = 2.40 \text{ kg m s}^{-1}$$

$$\lambda = \frac{h}{p} = \frac{6.63 \times 10^{-34} \text{ J s}}{2.40 \text{ kg m s}^{-1}} = 2.76 \times 10^{-34} \text{ m}$$

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.

Example 11.3 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 $m = 9.11 \times 10^{-31}$ kg, speed $v = 5.4 \times 10^6$ m/s. Then, momentum

$p = m\,v = 9.11 \times 10^{-31}$ (kg) $\times$ 5.4 $\times 10^6$ (m/s)

$p = 4.92 \times 10^{-24}$ kg m/s

de Broglie wavelength, $\lambda = h/p$

$$= \frac{6.63 \times 10^{-34} \text{ J s}}{4.92 \times 10^{-24} \text{ kg m/s}}$$

$\lambda = 0.135$ 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 $\lambda' = h/p'$.

$$= \frac{6.63 \times 10^{-34} \text{ J s}}{4.50 \times \text{ kg m/s}}$$

$\lambda' = 1.47 \times 10^{-34}$ 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 11.3

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_o) 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.

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	$e V_0 = K_{max}$
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 $h\nu$*. This is not quite the same as saying that light consists of particles, each of energy $h\nu$.
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 $\lambda = \dfrac{h}{p}$ 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 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.

11.6 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.

11.7 The work function for a certain metal is 4.2 eV. Will this metal give hotoelectric emission for incident radiation of wavelength 330 nm?

11.8 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?

11.9 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.

11.10 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?

11.11 Show that the wavelength of electromagnetic radiation is equal to the de Broglie wavelength of its quantum (photon).

Chapter Twelve

ATOMS

12.1 Introduction

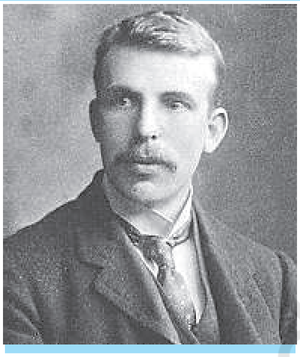
By the nineteenth century, enough evidence had accumulated in favour of atomic hypothesis of matter. In 1897, the experiments on electric discharge through gases carried out by the English physicist J. J. Thomson (1856 – 1940) revealed that atoms of different elements contain negatively charged constituents (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)

Ernst Rutherford (1871 – 1937) New Zealand born, 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.

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

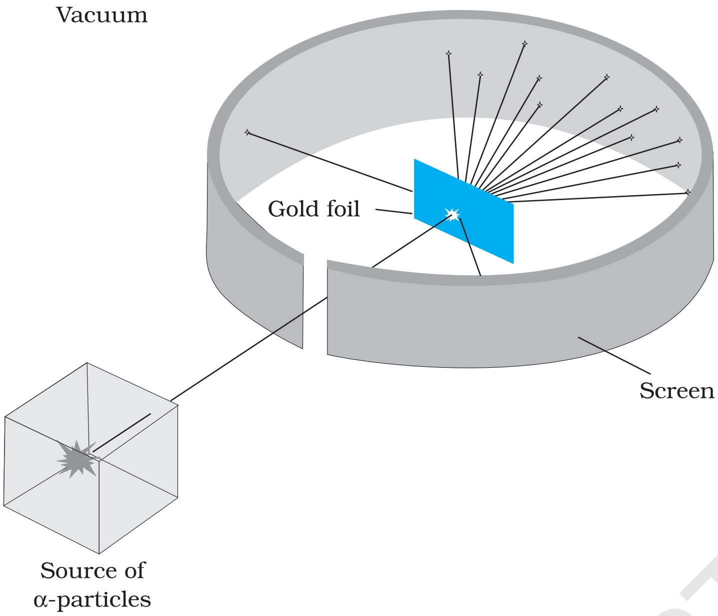


FIGURE 12.1 Geiger-Marsden scattering experiment. The entire apparatus is placed in a vacuum chamber (not shown in this figure).

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.

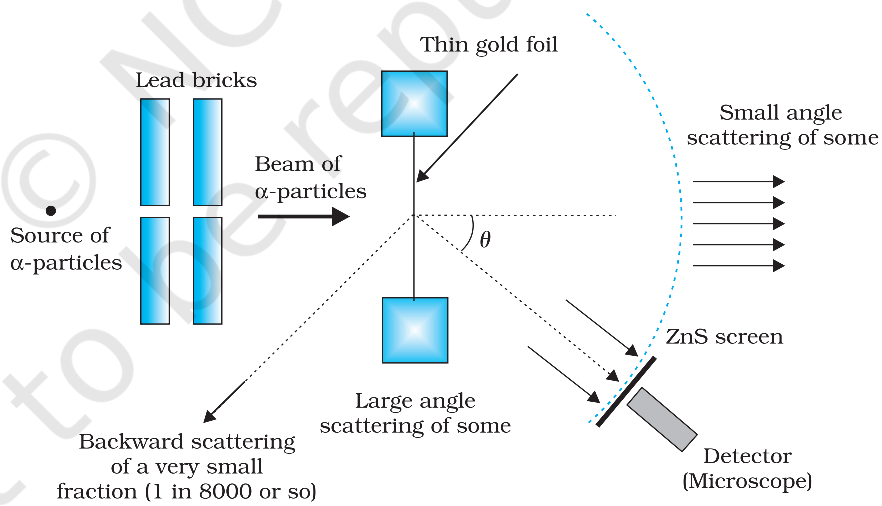


FIGURE 12.2 Schematic arrangement of the Geiger-Marsden experiment.

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.

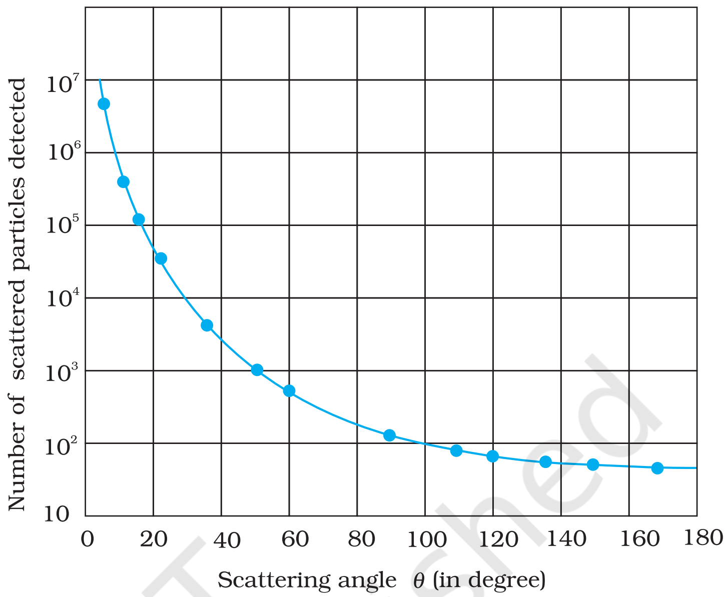


FIGURE 12.3 Experimental data points (shown by dots) on scattering of α-particles by a thin foil at different angles obtained by Geiger and Marsden using the setup shown in Figs. 12.1 and 12.2. Rutherford's nuclear model predicts the solid curve which is seen to be in good agreement with experiment.

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 10, Section 10.6 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, 2*e*, 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

$$F = \frac{1}{4\pi\varepsilon_0}\frac{(2e)(Ze)}{r^2} \qquad (12.1)$$

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 ($\theta \cong 0$).

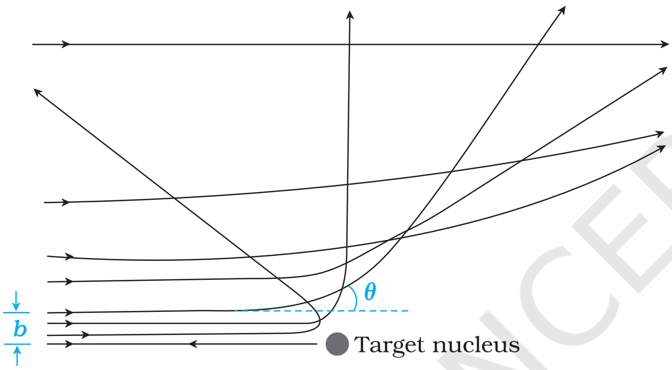


FIGURE 12.4 Trajectory of α-particles in the coulomb field of a target nucleus. The impact parameter, b and scattering angle θ are also depicted.

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 and positive charge 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 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 12.1

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

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

$$F_e = F_c$$

$$\frac{1}{4\pi\varepsilon_0}\frac{e^2}{r^2} = \frac{mv^2}{r} \qquad (12.2)$$

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 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

$$E = -\frac{e^2}{8\pi\varepsilon_0 r} = -2.2 \times 10^{-18} \text{ J}$$

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,

$$v = \frac{e}{\sqrt{4\pi\varepsilon_0 m r}} = 2.2 \times 10^6 \text{ m/s.}$$

EXAMPLE 12.3

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.

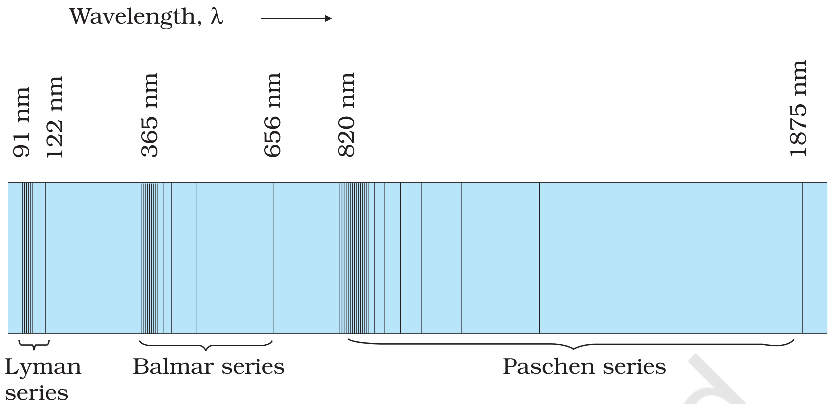
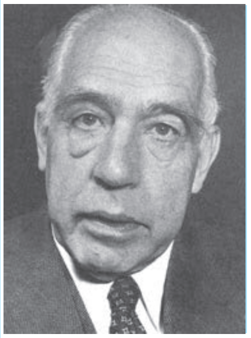


FIGURE 12.5 Emission lines in the spectrum of hydrogen.

12.4 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. 12.6). 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)



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.

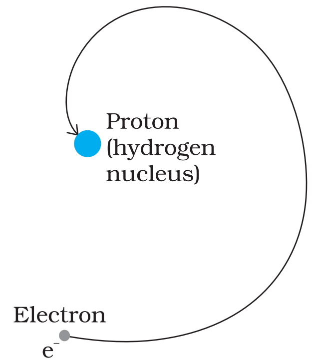


FIGURE 12.6 An accelerated atomic electron must spiral into the nucleus as it loses energy.

EXAMPLE 12.4

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.

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

$$L = nh/2\pi \tag{12.5}$$

(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*

$$h\nu = E_i - E_f \tag{12.6}$$

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 radius of nth possible orbit thus found is

$$r_n = \frac{n^2}{m}\left(\frac{h}{2\pi}\right)^2 \frac{4\pi\varepsilon_0}{e^2} \tag{12.7}$$

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 $$E_n = -\frac{me^4}{8n^2\varepsilon_0^2h^2} \tag{12.8}$$

Substituting values, Eq. (12.8) yields

$$E_n = -\frac{2.18\times10^{-18}}{n^2}\,\text{J} \tag{12.9}$$

Atomic energies are often expressed in electron volts (eV) rather than joules. Since 1 eV = 1.6×10^{-19} J, Eq. (12.9) can be rewritten as

$$E_n = -\frac{13.6}{n^2}\ \text{eV} \tag{12.10}$$

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).

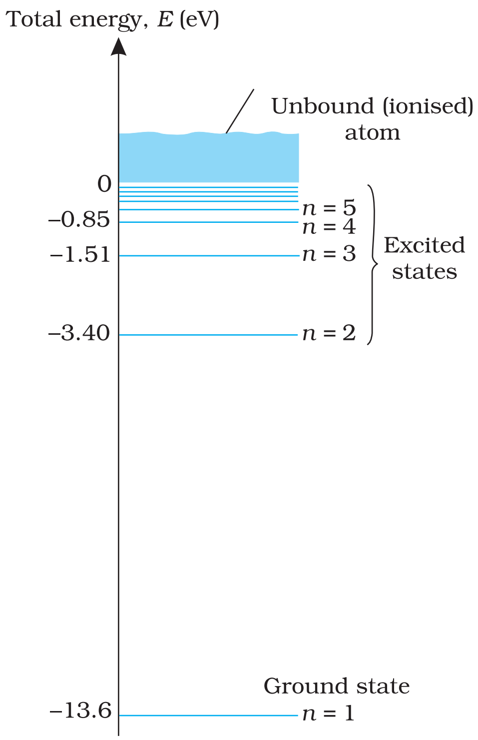


FIGURE 12.7 The energy level diagram for the hydrogen atom. The electron in a hydrogen atom at room temperature spends most of its time in the ground state. To ionise a hydrogen atom an electron from the ground state, 13.6 eV of energy must be supplied. (The horizontal lines specify the presence of allowed energy states.)

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. (12.10), 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 $E_2 - E_1 = -3.40 \text{ eV} - (-13.6) \text{ eV} = 10.2 \text{ eV}$. 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. (12.10), is given in Fig. 12.7. 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, (12.10) 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.

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

* 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. 12.7.

$h\nu_{if} = E_{ni} - E_{nf}$ (12.11)

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.

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.

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.

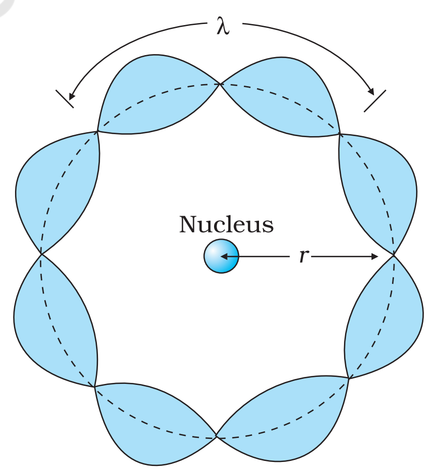


FIGURE 12.8 A standing wave is shown on a circular orbit where four de Broglie wavelengths fit into the circumference of the orbit.

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 14 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

$$2\pi\, r_n = n\lambda, \qquad n = 1, 2, 3... \qquad (12.12)$$

Figure 12.8 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. (12.12), we have

$$2\pi\, r_n = n\, h/mv_n \quad \text{or} \quad m\, v_n\, r_n = nh/2\pi$$

This is the quantum condition proposed by Bohr for the angular momentum of the electron [Eq. (12.15)]. 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.

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.

* 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.

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 most of the elements 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. 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 the principal 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.
   $$E_i + h\nu = E_f$$
7. 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:
   $$E_n = -\frac{me^4}{8n^2\varepsilon_0^2h^2}$$
   $$= -13.6 \text{ eV}/n^2$$
   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.

8. 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.
9. 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 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?

12.4 The ground state energy of hydrogen atom is –13.6 eV. What are the kinetic and potential energies of the electron in this state?

12.5 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.

12.6 (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.

12.7 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?

12.8 A 12.5 eV electron beam is used to bombard gaseous hydrogen at room temperature. What series of wavelengths will be emitted?

12.9 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.)

12090CH13

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

$$1u = \frac{\text{mass of one } ^{12}\text{C atom}}{12}$$

$$= \frac{1.992647 \times 10^{-26} \text{ kg}}{12}$$

$$= 1.660539 \times 10^{-27} \text{ kg} \tag{13.1}$$

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

$$m_p = 1.00727 \text{ u} = 1.67262 \times 10^{-27} \text{ kg} \tag{13.2}$$

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

$$m_n = 1.00866 \text{ u} = 1.6749\times10^{-27} \text{ kg} \tag{13.3}$$

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 ${}^{A}_{Z}\text{X}$ where X is the chemical symbol of the species. For example, the nucleus of gold is denoted by ${}^{197}_{79}\text{Au}$. 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, ${}_1^2\mathrm{H}$, which is an isotope of hydrogen, contains one proton and one neutron. Its other isotope tritium, ${}_1^3\mathrm{H}$, 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 ${}_1^3\mathrm{H}$ and ${}_2^3\mathrm{He}$ are isobars. Nuclides with same neutron number N but different atomic number Z, for example ${}_{80}^{198}\mathrm{Hg}$ and ${}_{79}^{197}\mathrm{Au}$, 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 (=1.2 fm; 1 fm = 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 like a 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

$$\text{Nuclear density} = \frac{\text{mass}}{\text{volume}} = \frac{9.27 \times 10^{-26}}{(4\pi/3)(1.2 \times 10^{-15})^3} \times \frac{1}{56}$$

$$= 2.29 \times 10^{17} \text{ kg m}^{-3}$$

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 MASS-ENERGY AND NUCLEAR BINDING ENERGY

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 × 10^{-27} × (2.9979 × 10^8)^2 kg m^2/s^2
= 1.4924 × 10^{-10} J

$$= \frac{1.4924 \times 10^{-10}}{1.602 \times 10^{-19}} \text{eV}$$

= 0.9315 × 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.

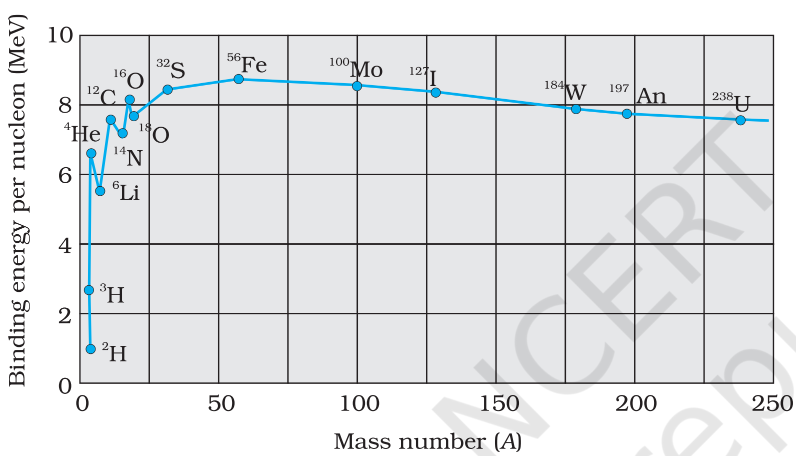


FIGURE 13.1 The binding energy per nucleon as a function of mass number.

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 13.7.2.

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.

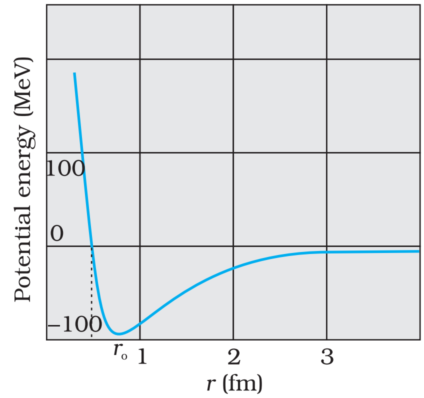


FIGURE 13.2 Potential energy of a pair of nucleons as a function of their separation. For a separation greater than r_0, the force is attractive and for separations less than r_0, the force is strongly repulsive.

(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.

(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.

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. Now, the greater the binding energy, the less is the total mass of a bound system, such as a nucleus. Consequently, if nuclei with less total binding energy transform to nuclei with greater binding energy, there will be a net energy release. This is what happens when a heavy nucleus decays into two or more intermediate mass fragments (*fission*) or when light nuclei fuse into a havier nucleus (*fusion*.)

Exothermic chemical reactions underlie conventional energy sources such as coal or petroleum. Here the energies involved are in the range of electron volts. On the other hand, in a nuclear reaction, the energy release is of the order of MeV. Thus for the same quantity of matter, nuclear sources produce a million times more energy than a chemical source. Fission of 1 kg of uranium, for example, generates 10^{14} J of energy; compare it with burning of 1 kg of coal that gives 10^7 J.

13.7.1 Fission

New possibilities emerge when we go beyond natural radioactive decays and study nuclear reactions by bombarding nuclei with other nuclear particles such as proton, neutron, α-particle, etc.

A most important neutron-induced nuclear reaction is fission. An example of fission is when a uranium isotope $^{235}_{92}U$ bombarded with a neutron breaks into two intermediate mass nuclear fragments

$$^{1}_{0}n + ^{235}_{92}U \rightarrow ^{236}_{92}U \rightarrow ^{144}_{56}Ba + ^{89}_{36}Kr + 3\,^{1}_{0}n \tag{13.10}$$

The same reaction can produce other pairs of intermediate mass fragments

$$^{1}_{0}n + ^{235}_{92}U \rightarrow ^{236}_{92}U \rightarrow ^{133}_{51}Sb + ^{99}_{41}Nb + 4\,^{1}_{0}n \tag{13.11}$$

Or, as another example,

$$^{1}_{0}n + ^{235}_{92}U \rightarrow ^{140}_{54}Xe + ^{94}_{38}Sr + 2\,^{1}_{0}n \tag{13.12}$$

The fragment products are radioactive nuclei; they emit β particles in succession to achieve stable end products.

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.

13.7.2 Nuclear fusion – energy generation in stars

When two light nuclei fuse to form a larger nucleus, energy is released, since the larger nucleus is more tightly bound, as seen from the binding energy curve in Fig.13.1. Some examples of such energy liberating nuclear fusion reactions are :

$$^{1}_{1}H + ^{1}_{1}H \rightarrow ^{2}_{1}H + e^{+} + \nu + 0.42\text{ MeV} \qquad [13.13(a)]$$

$$^{2}_{1}H + ^{2}_{1}H \rightarrow ^{3}_{2}He + n + 3.27\text{ MeV} \qquad [13.13(b)]$$

$$^{2}_{1}H + ^{2}_{1}H \rightarrow ^{3}_{1}H + ^{1}_{1}H + 4.03\text{ MeV} \qquad [13.13(c)]$$

In the first reaction, two protons combine to form a deuteron and a positron with a release of 0.42 MeV energy. In reaction [13.13(b)], two deuterons combine to form the light isotope of helium. In reaction (13.13c), two deuterons combine to form a triton and a proton. For fusion to take place, the two nuclei must come close enough so that attractive short-range nuclear force is able to affect them. However, since they are both positively charged particles, they experience coulomb repulsion. They, therefore, must have enough energy to overcome this coulomb barrier. The height of the barrier depends on the charges and radii of the two interacting nuclei. It can be shown, for example, that the barrier height for two protons is ~ 400 keV, and is higher for nuclei with higher charges. We can estimate the temperature at which two protons in a proton gas would (averagely) have enough energy to overcome the coulomb barrier:

$(3/2)k\,T = K \simeq$ 400 keV, which gives $T \sim 3 \times 10^9$ K.

When fusion is achieved by raising the temperature of the system so that particles have enough kinetic energy to overcome the coulomb repulsive behaviour, it is called *thermonuclear fusion*.

Thermonuclear fusion is the source of energy output in the interior of stars. The interior of the sun has a temperature of 1.5×10^7 K, which is considerably less than the estimated temperature required for fusion of particles of average energy. Clearly, fusion in the sun involves protons whose energies are much above the average energy.

The fusion reaction in the sun is a multi-step process in which the hydrogen is burned into helium. Thus, the fuel in the sun is the hydrogen in its core. The *proton-proton* (p, p) *cycle* by which this occurs is represented by the following sets of reactions:

$$ {}_1^1\mathrm{H} + {}_1^1\mathrm{H} \rightarrow {}_1^2\mathrm{H} + e^+ + \nu + 0.42 \text{ MeV} \qquad \text{(i)} $$

$$ e^+ + e^- \rightarrow \gamma + \gamma + 1.02 \text{ MeV} \qquad \text{(ii)} $$

$$ {}_1^2\mathrm{H} + {}_1^1\mathrm{H} \rightarrow {}_2^3\mathrm{He} + \gamma + 5.49 \text{ MeV} \qquad \text{(iii)} $$

$$ {}_2^3\mathrm{He} + {}_2^3\mathrm{He} \rightarrow {}_2^4\mathrm{He} + {}_1^1\mathrm{H} + {}_1^1\mathrm{H} + 12.86 \text{ MeV} \quad \text{(iv)} \qquad (13.14) $$

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 nucleus. If we consider the combination 2(i) + 2(ii) + 2(iii) +(iv), the net effect is

$$ 4\,{}_1^1\mathrm{H} + 2e^- \rightarrow {}_2^4\mathrm{He} + 2\nu + 6\gamma + 26.7\,\mathrm{MeV} $$

$$ \text{or} \quad (4\,{}_1^1\mathrm{H} + 4e^-) \rightarrow ({}_2^4\mathrm{He} + 2e^-) + 2\nu + 6\gamma + 26.7\,\mathrm{MeV} \qquad (13.15) $$

Thus, four hydrogen atoms combine to form an ${}_2^4\mathrm{He}$ atom with a release of 26.7 MeV of energy.

Helium is not the only element that can be synthesized in the interior of a star. As the hydrogen in the core gets depleted and becomes helium, the core starts to cool. The star begins to collapse under its own gravity which increases the temperature of the core. If this temperature increases to about 10^8 K, fusion takes place again, this time of helium nuclei into carbon. This kind of process can generate through fusion higher and higher mass number elements. But elements more massive than those near the peak of the binding energy curve in Fig. 13.1 cannot be so produced.

The age of the sun is about 5×10^9 y and it is estimated that there is enough hydrogen in the sun to keep it going for another 5 billion years. After that, the hydrogen burning will stop and the sun will begin to cool and will start to collapse under gravity, which will raise the core temperature. The outer envelope of the sun will expand, turning it into the so called *red giant*.

13.7.3 Controlled thermonuclear fusion

The natural thermonuclear fusion process in a star is replicated in a thermonuclear fusion device. In controlled fusion reactors, the aim is to generate steady power by heating the nuclear fuel to a temperature in the range of 10^8 K. At these temperatures, the fuel is a mixture of positive ions and electrons (plasma). The challenge is to confine this plasma, since no container can stand such a high temperature. Several countries around the world including India are developing techniques in this connection. If successful, fusion reactors will hopefully supply almost unlimited power to humanity.

Example 13.4 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. 13.10), 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

EXAMPLE 13.4

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.

EXAMPLE 13.4

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 1/12th mass of one atom of ^{12}C; 1u = 1.660563×10^{-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 ${}^{A}_{Z}\mathrm{X}$, 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 and calculated in accordance to 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. 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., $^{235}_{92}\mathrm{U} + ^{1}_{0}\mathrm{n} \rightarrow ^{133}_{51}\mathrm{Sb} + ^{99}_{41}\mathrm{Nb} + 4\,^{1}_{0}\mathrm{n}$
12. 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 1/12th 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. 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	$N = 6.023\times10^{23}$ per mole
$1/(4\pi\varepsilon_0) = 9\times10^{9}$ N m^2/C^2	$k = 1.381\times10^{-23}$J K^{-1}
1 MeV $= 1.6\times10^{-13}$J	1 u = 931.5 MeV/c^2
1 year $= 3.154\times10^{7}$ s	
m_H = 1.007825 u	m_n = 1.008665 u
$m(^4_2\text{He})$ = 4.002603 u	m_e = 0.000548 u

13.1 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

13.2 Obtain the binding energy of the nuclei $^{56}_{26}\text{Fe}$ and $^{209}_{83}\text{Bi}$ in units of MeV from the following data:

$m(^{56}_{26}\text{Fe})$ = 55.934939 u $m(^{209}_{83}\text{Bi})$ = 208.980388 u

13.3 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 $^{63}_{29}\text{Cu}$ atoms (of mass 62.92960 u).

13.4 Obtain approximately the ratio of the nuclear radii of the gold isotope $^{197}_{79}\text{Au}$ and the silver isotope $^{107}_{47}\text{Ag}$.

13.5 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

13.6 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.

13.7 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 $^{239}_{94}\text{Pu}$ undergo fission?

13.8 How long can an electric lamp of 100W be kept glowing by fusion of 2.0 kg of deuterium? Take the fusion reaction as

$$^{2}_{1}\text{H} + ^{2}_{1}\text{H} \rightarrow ^{3}_{2}\text{He} + \text{n} + 3.27 \text{ MeV}$$

13.9 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.)

13.10 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).

12090CH14

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.

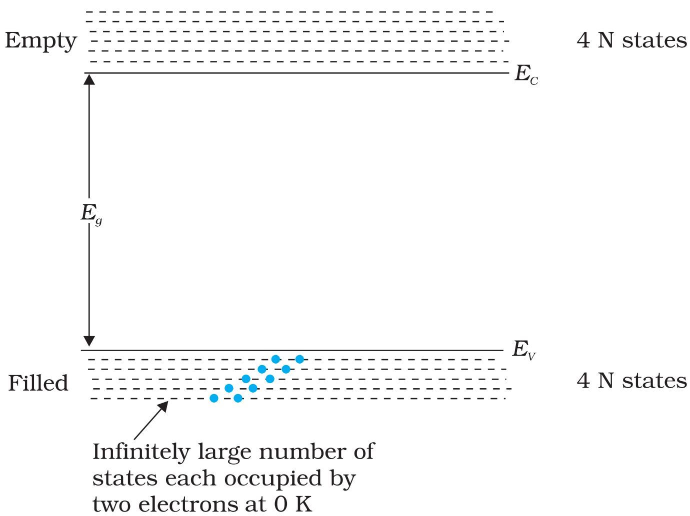


FIGURE 14.1 The energy band positions in a semiconductor at 0 K. The upper band, called the conduction band, consists of infinitely large number of closely spaced energy states. The lower band, called the valence band, consists of closely spaced completely filled energy states.

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:

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.

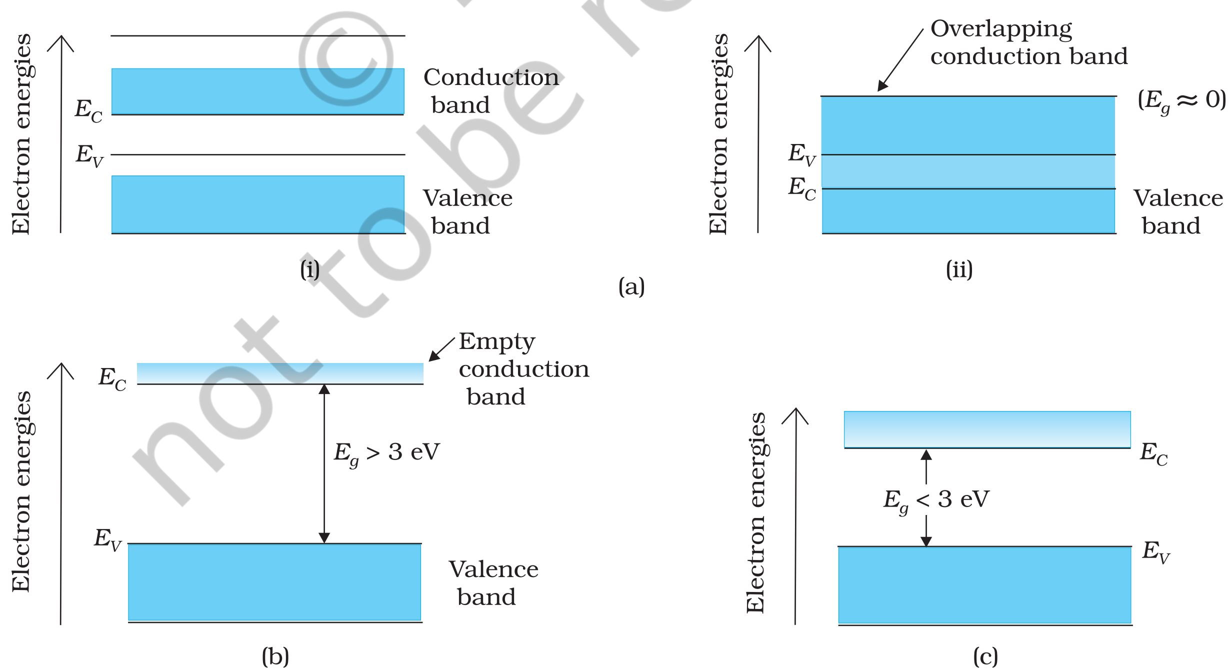


FIGURE 14.2 Difference between energy bands of (a) metals, (b) insulators and (c) semiconductors.

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.

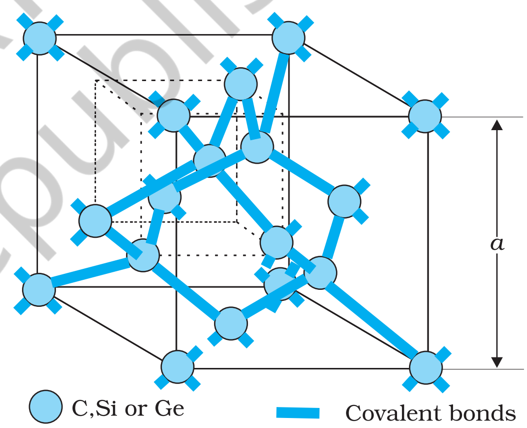


FIGURE 14.3 Three-dimensional diamond-like crystal structure for Carbon, Silicon or Germanium with respective lattice spacing a equal to 3.56, 5.43 and 5.66 Å.

In intrinsic semiconductors, the number of free electrons, n_e is equal to the number of holes, n_h. That is

$$n_e = n_h = n_i \tag{14.1}$$

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

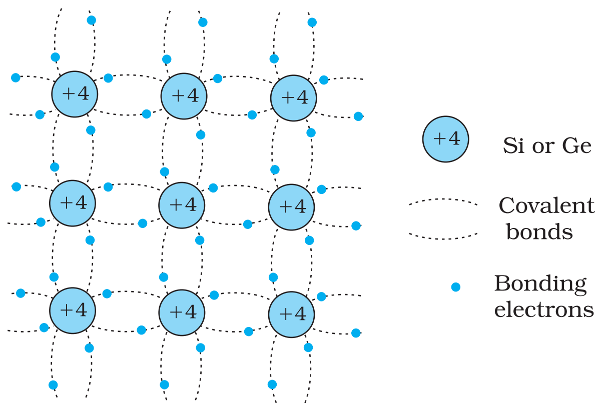
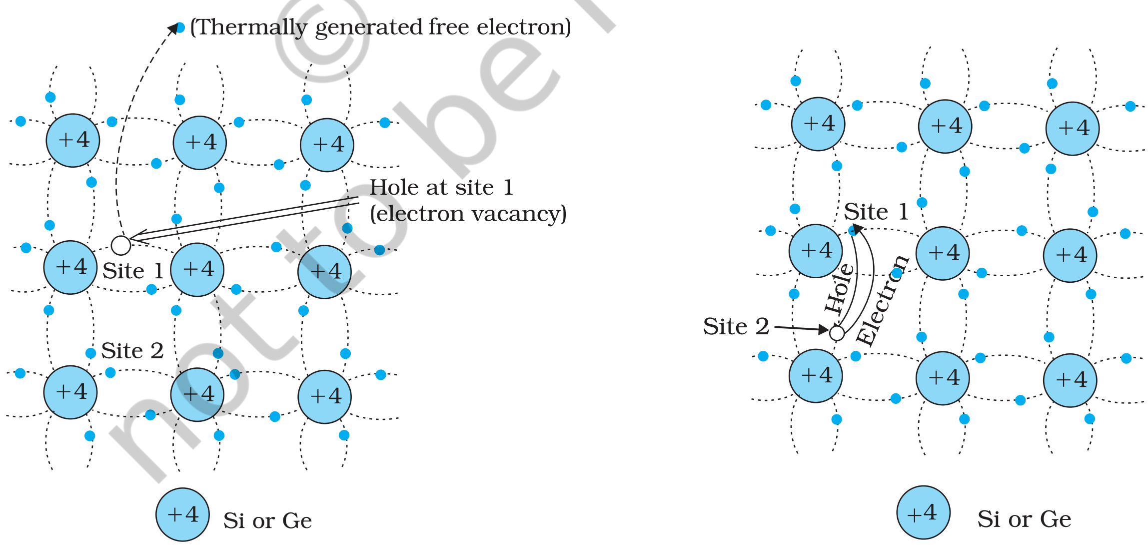


FIGURE 14.4 Schematic two-dimensional representation of Si or Ge structure showing covalent bonds at low temperature (all bonds intact). +4 symbol indicates inner cores of Si or Ge.

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:

$$I = I_e + I_h \tag{14.2}$$

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.



(a) (b)

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.

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.

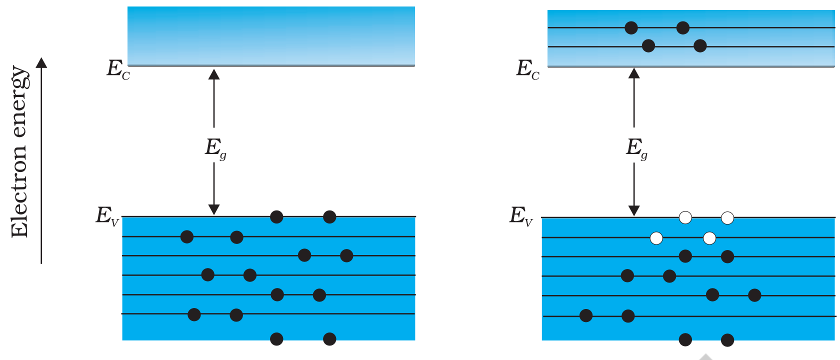


FIGURE 14.6 (a) An intrinsic semiconductor at $T = 0$ K behaves like insulator. (b) At $T > 0$ K, four thermally generated electron-hole pairs. The filled circles (•) represent electrons and empty circles (○) represent holes.

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.

EXAMPLE 14.1

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.

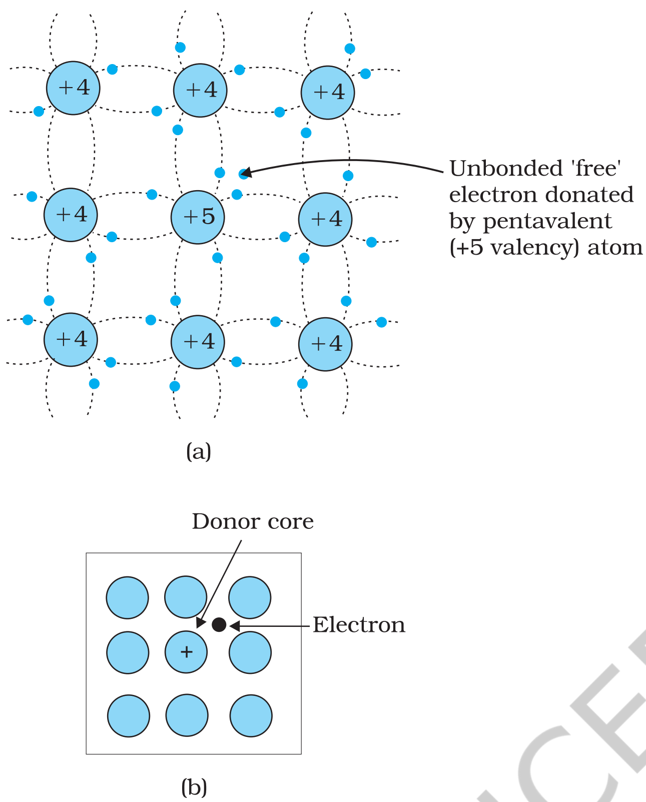


FIGURE 14.7 (a) Pentavalent donor atom (As, Sb, P, etc.) doped for tetravalent Si or Ge giving n-type semiconductor, and (b) Commonly used schematic representation of n-type material which shows only the fixed cores of the substituent donors with one additional effective positive charge and its associated extra electron.

(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.

(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 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.*

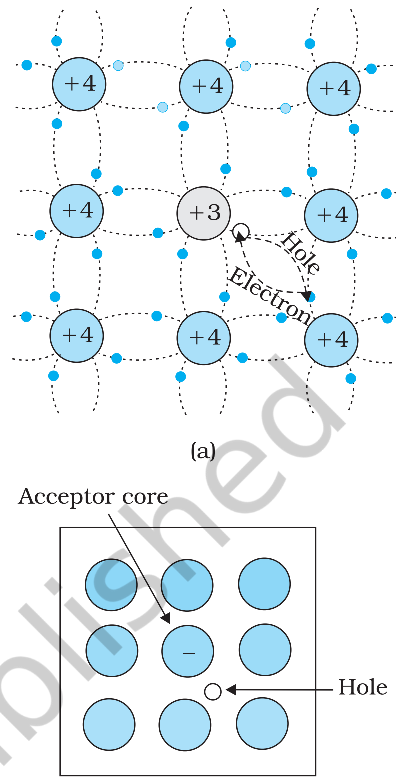


FIGURE 14.8 (a) Trivalent acceptor atom (In, Al, B etc.) doped in tetravalent Si or Ge lattice giving p-type semiconductor. (b) Commonly used schematic representation of p-type material which shows only the fixed core of the substituent acceptor with one effective additional negative charge and its associated hole.

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.

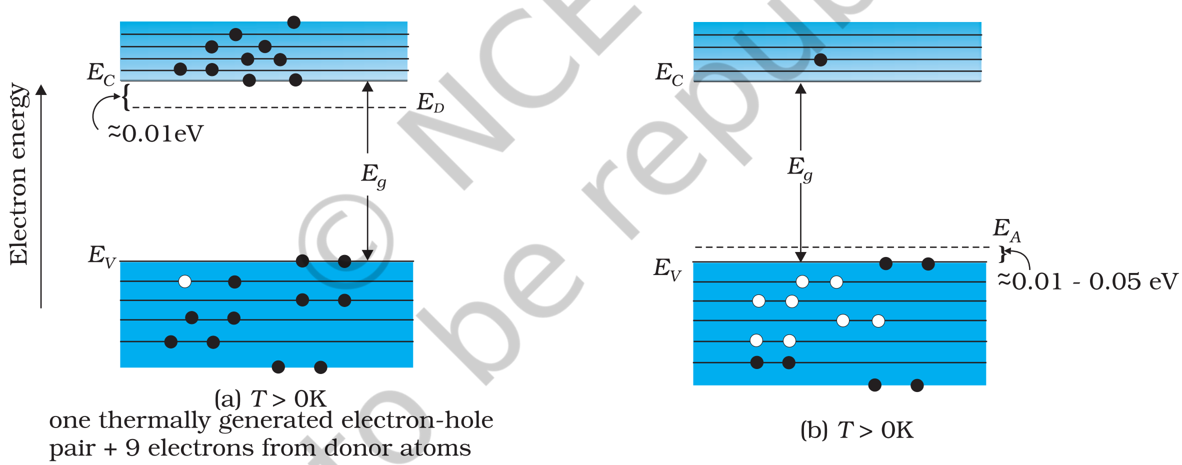


FIGURE 14.9 Energy bands of (a) n-type semiconductor at $T > 0$K, (b) p-type semiconductor at $T > 0$K.

EXAMPLE 14.2

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}

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

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 ($p \to n$) and electrons diffuse from n-side to p-side ($n \to p$). This motion of charge carries gives rise to diffusion current across the junction.

When an electron diffuses from $n \to p$, 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 $n \to p$, a layer of positive charge (or positive space-charge region) on n-side of the junction is developed.

Similarly, when a hole diffuses from $p \to n$ 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.

Formation and working of p-n junction diode
http://hyperphysics.phy-astr.gsu.edu/hbase/solids/pnjun.html

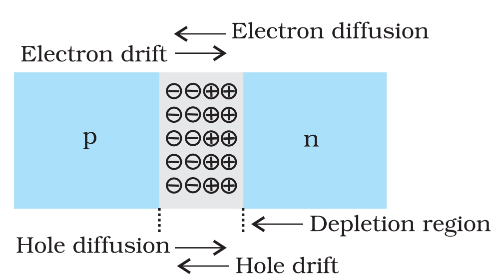


FIGURE 14.10 p-n junction formation process.

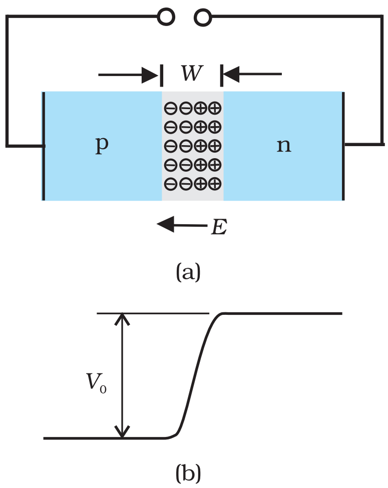


FIGURE 14.11 (a) Diode under equilibrium (V = 0), (b) Barrier potential under no bias.

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*.

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.

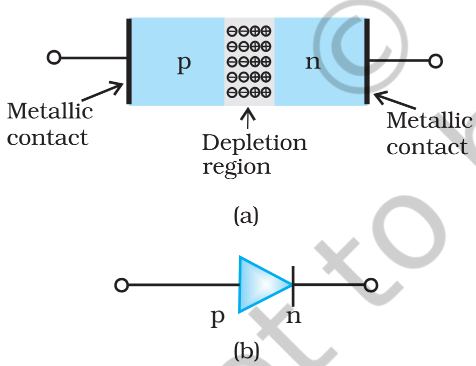


FIGURE 14.12 (a) Semiconductor diode, (b) Symbol for p-n junction diode.

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).

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 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.

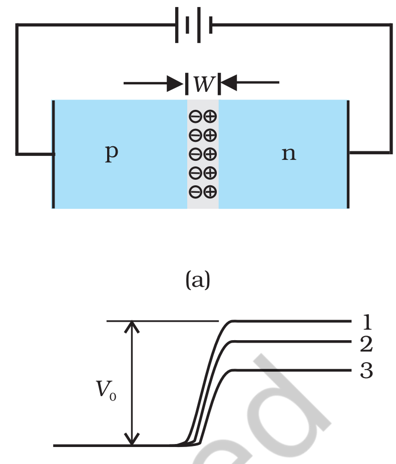


FIGURE 14.13 (a) p-n junction diode under forward bias, (b) Barrier potential (1) without battery, (2) Low battery voltage, and (3) High voltage battery.

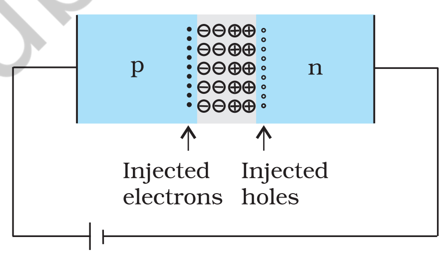


FIGURE 14.14 Forward bias minority carrier injection.

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

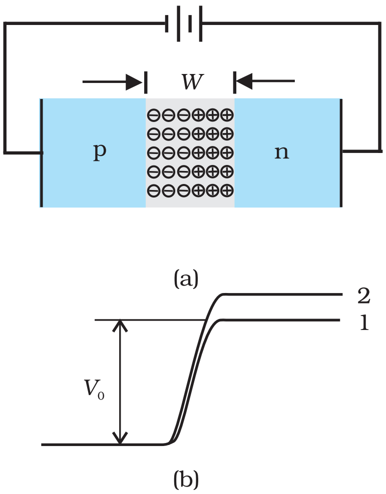


FIGURE 14.15 (a) Diode under reverse bias, (b) Barrier potential under reverse bias.

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.

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

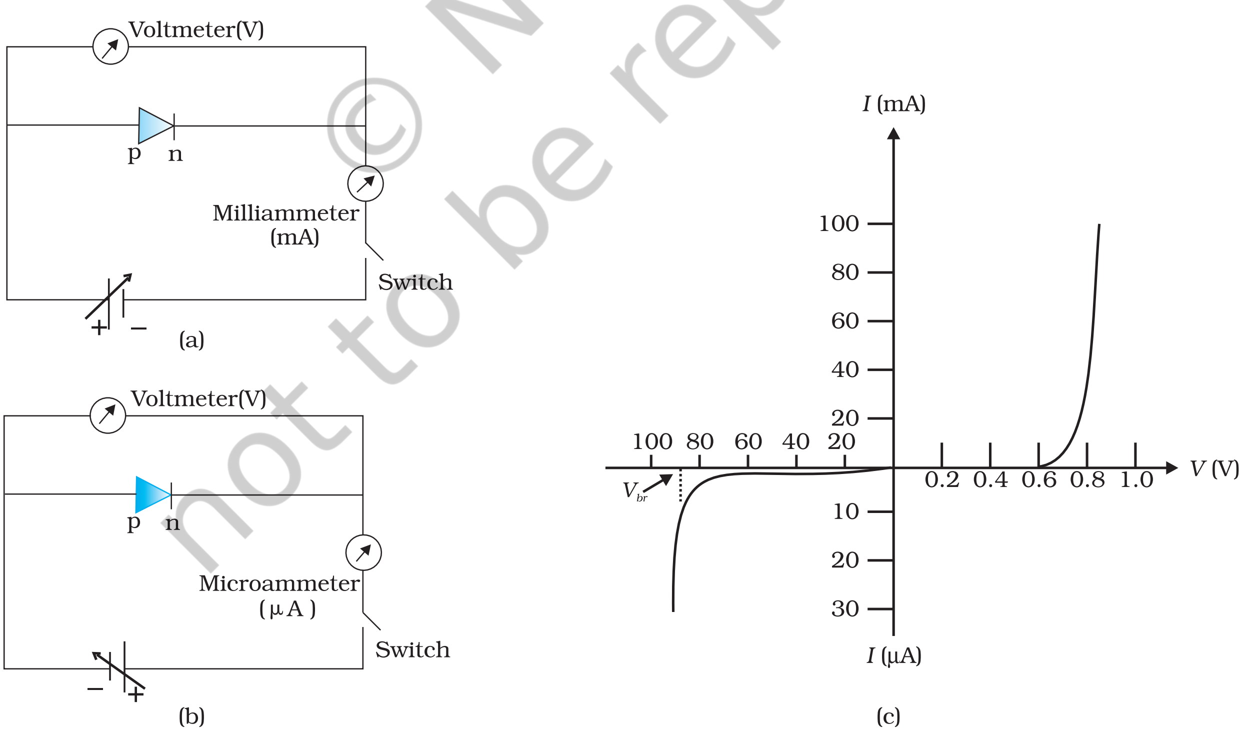


FIGURE 14.16 Experimental circuit arrangement for studying *V-I* characteristics of a p-n junction diode (a) in forward bias, (b) in reverse bias. (c) Typical *V-I* characteristics of a silicon diode.

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.

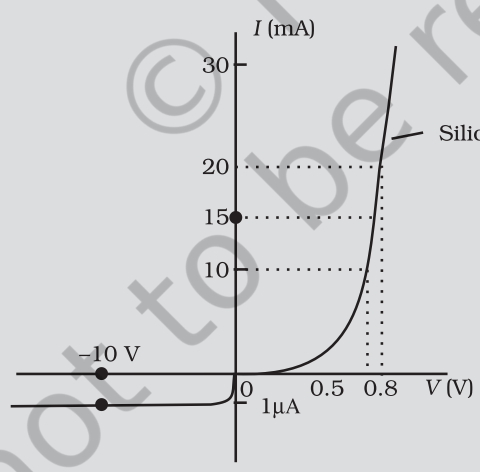


FIGURE 14.17

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.)

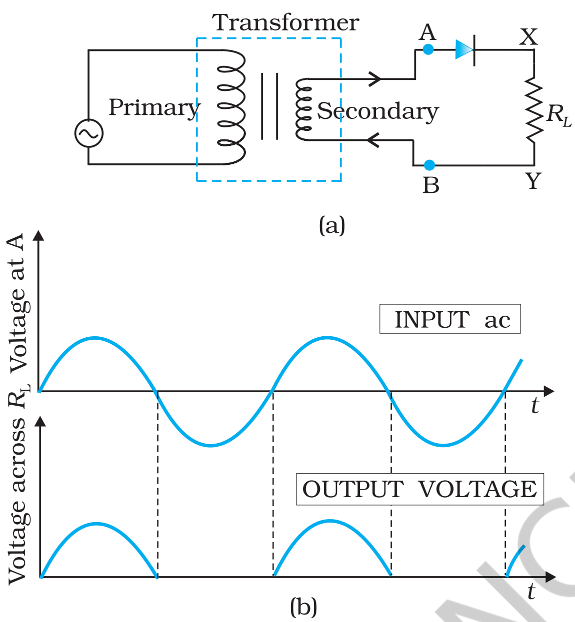


FIGURE 14.18 (a) Half-wave rectifier circuit, (b) Input ac voltage and output voltage waveforms from the rectifier circuit.

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 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.

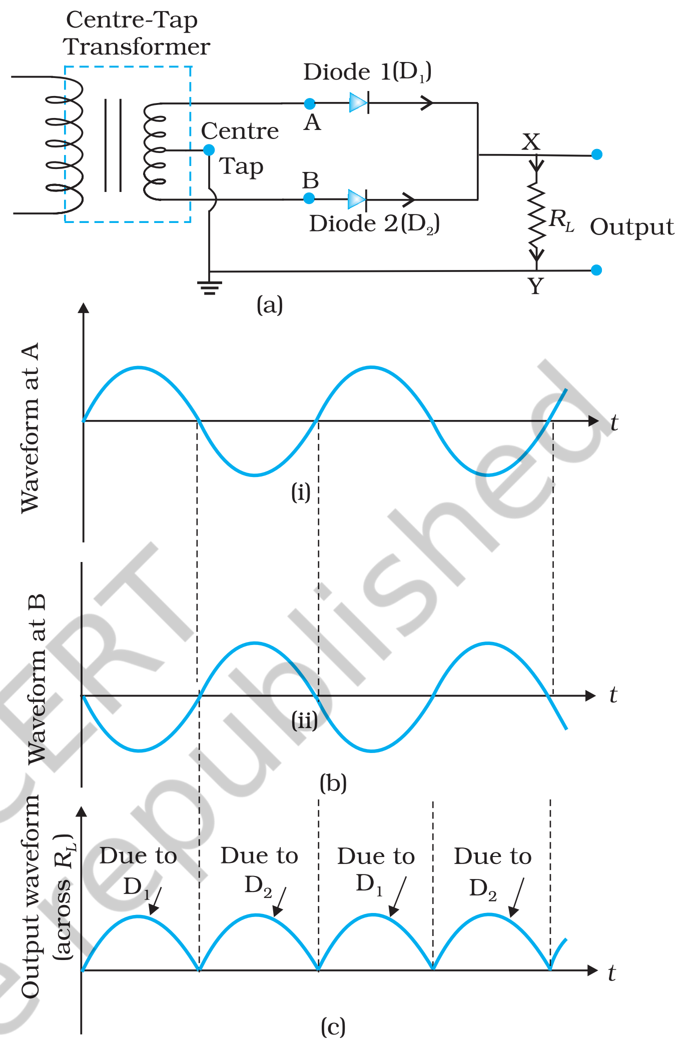


FIGURE 14.19 (a) A Full-wave rectifier circuit; (b) Input wave forms given to the diode D_1 at A and to the diode D_2 at B; (c) Output waveform across the load R_L connected in the full-wave rectifier circuit.

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 inversely upon the product of capacitance 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.

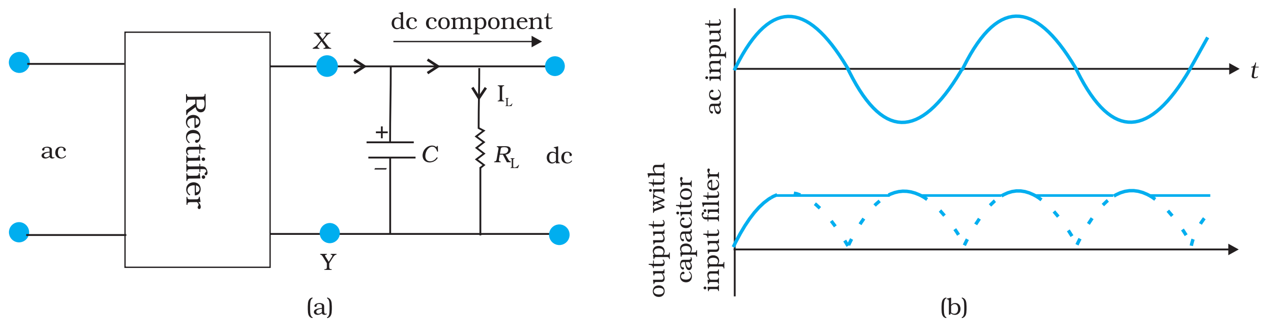


FIGURE 14.20 (a) A full-wave rectifier with capacitor filter, (b) Input and output voltage of rectifier in (a).

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 Ω 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.

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.

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 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.

Notes

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} Wm^{-2}K^{-4}
Wien's Constant	b	2.898×10^{-3} mK
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}

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

$$= \frac{25}{2.5} + 1 = 11; \ |u_e| = \frac{25}{11} \text{cm} = 2.27 \text{cm}; \ v_O = 7.2 \text{cm}$$

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 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.

9.19 21.4 cm

9.20 (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

9.21 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, $r = 19°$

$\sin i = 0.4962$; $i \simeq 30°$

9.22 (a) $\dfrac{1}{v} + \dfrac{1}{9} = \dfrac{1}{10}$

i.e., $v = -90$ cm,

Magnitude of magnification = 90/9 = 10.

Each square in the virtual image has an area $10 \times 10 \times 1\ \text{mm}^2 = 100\ \text{mm}^2 = 1\ \text{cm}^2$

(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.

9.23 (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.

9.24 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.

9.25 (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.

9.26 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.

9.27 (a) $m = (f_O/f_e) = 28$

(b) $m = \dfrac{f_O}{f_e}\left[1 + \dfrac{f_O}{25}\right] = 33.6$

9.28 (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.

9.29 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.

9.30 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.

9.31 $n = 1.33$

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 $$\lambda = \frac{1.2\times10^{-2}\times0.28\times10^{-3}}{4\times1.4} \text{ m} = 600 \text{ nm}$$

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 40m

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 6.59×10^{-34} J s
11.6 2.0 V
11.7 No, because $\nu < \nu_o$
11.8 4.73×10^{14} Hz
11.9 2.16 eV = 3.46×10^{-19}J
11.10 (a) 1.7×10^{-35} m (b) 1.1×10^{-32} m (c) 3.0×10^{-23} m
11.11 $\lambda = h/p = h/(h\nu/c) = c/\nu$

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 5.6×10^{14} Hz
12.4 13.6 eV; –27.2 eV
12.5 9.7×10^{-8}m; 3.1×10^{15}Hz.
12.6 (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.
12.7 2.12×10^{-10} m; 4.77×10^{-10} m
12.8 Lyman series: 103 nm and 122 nm; Balmer series: 656 nm.
12.9 2.6×10^{74}

Chapter 13

13.1 104.7 MeV
13.2 8.79 MeV, 7.84 MeV
13.3 1.584×10^{25} MeV or 2.535×10^{12}J
13.4 1.23

13.5 (i) $Q = -4.03$ MeV; endothermic

(ii) $Q = 4.62$ MeV; exothermic

13.6 $Q = m\left({}^{56}_{26}\mathrm{Fe}\right) - 2m\left({}^{28}_{13}\mathrm{Al}\right) = 26.90$ MeV; not possible.

13.7 4.536×10^{26} MeV

13.8 About 4.9×10^{4} y

13.9 360 KeV

Chapter 14

14.1 (c)

14.2 (d)

14.3 (c)

14.4 (c)

14.5 (c)

14.6 50 Hz for half-wave, 100 Hz for full-wave

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).