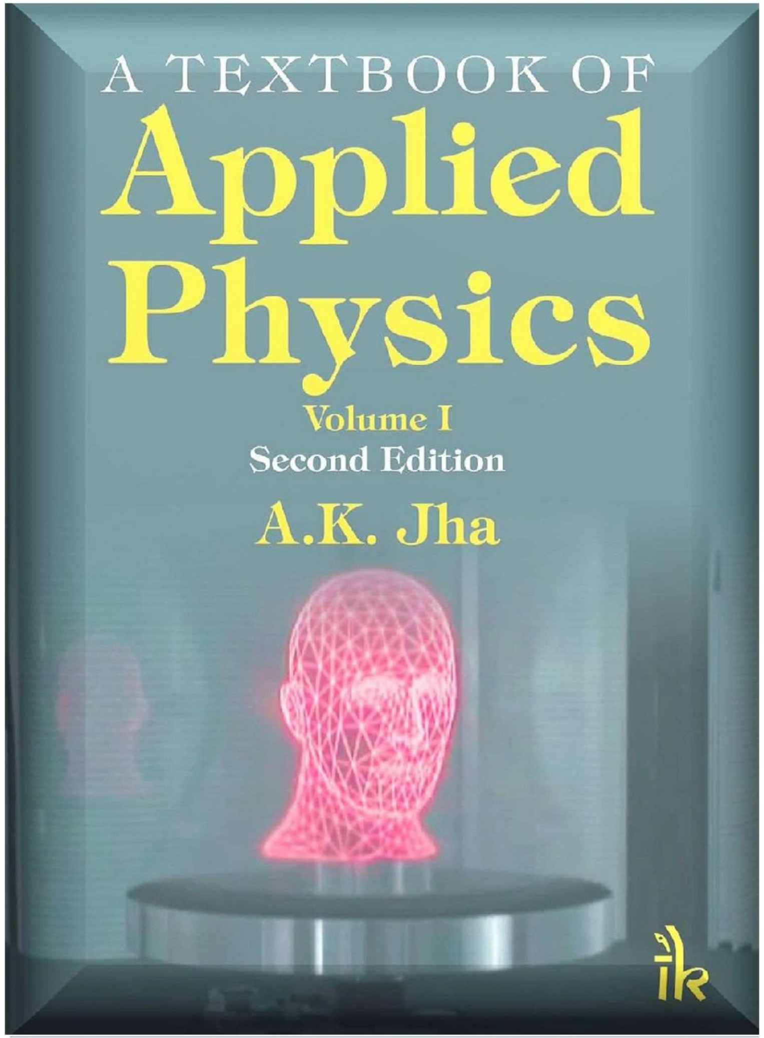
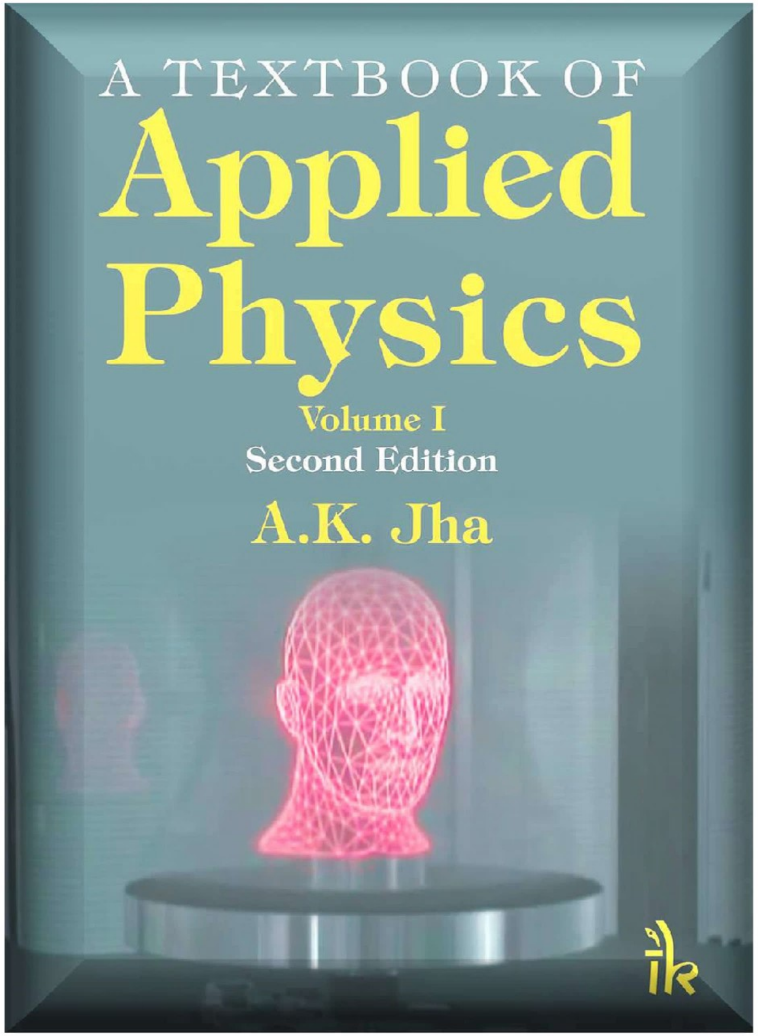


12/6/2020 Kindle Cloud Reader
A TEXTBOOK OF APPLIED PHYSICS VOLUME I -SE…



Second Edition
A TEXTBOOK OF APPLIED PHYSICS
Volume I

Dr. A.K. Jha

Ph.D. (IITD)

Professor

Ambedkar Institute of Advanced Communication

Technologies & Research

Delhi

I.K. International Publishing House Pvt. Ltd

NEW DELHI

Published by
I.K. International Publishing House Pvt. Ltd.
S-25, Green Park Extension
Uphaar Cinema Market
New Delhi–110 016 (India)
E-mail: info@ikinternational.com
Website: www.ikbooks.com

ISBN 978-93-84588-51-9

Published by Krishan Makhijani for I.K. International Publishing House Pvt. Ltd., S-25, Green Park Extension, Uphaar Cinema Market, New De
Printed by Rekha Printers Pvt. Ltd., Okhla Industrial Area, Phase II, New Delhi–110 020.

Preface to the Second Edition

My association with undergraduate students made me realize that in a semester of three months duration, it is really difficult for the students to look for right materials for different top different books prescribed in the syllabus. This prompted me to write this book. It is hoped that will find all the topics included in the first semester syllabus and utilize the time in learning th instead of wasting time in searching topics in different books. The topics in the book ha presented in a manner so that students of all background will find it easy to understand the principles.

I sincerely thank the scientists whose work constitute the subject matter and the large nu authors whose books I have consulted while preparing the manuscript of this book.

I thank the entire production team of I.K. International Publishing House Pvt. Ltd. spec Krishan Makhijani and Sh. Anand Singh Aswal for timely publication of this book.

In spite of our best efforts to make the book error free, some mistakes and errors might have due to oversight. I will be highly thankful to my teacher colleagues and the students who kindly b shortcomings to my notice and give useful suggestions so that they can be taken care in future of the book.

I wish reader students grand success and bright future.

Dr. A.K. Jha
Email: dr.akjha.dce@gmail.com

Contents

Unit I

Unit II

Unit III

12/6/2020 Kindle Cloud Reader
A TEXTBOOK OF APPLIED PHYSICS VOLUME I -SE…

1

Interference of Light

1.1 Introduction: Nature of light

We study light in the branch of physics called optics. *Optics is the branch of physics which deals with the study of the nature of light and the phenomenon exhibited by it.*

Our life is dependent on light and we cannot imagine our life without light. Naturally, from very early days, human beings must have wondered about the nature of light. Simple and common observations such as shape, size and sharpness of the image formed led to the concept of *rays of light* and its *rectilinear propagation*. Many theories have been put forward to explain the nature of light. We briefly discuss some of them here.

Newton in his *corpuscular theory* (1675 AD) assumed that light consists of very small particles called *corpuscles*, which are emitted by a luminous source and travel with large velocities. This theory explained rectilinear propagation of light, reflection and refraction, but could not explain interference, diffraction and polarization.

Huygen's wave theory of light (1678 AD) considered light to consist of *periodic disturbance transmitted through a medium in the form of longitudinal waves*. He proposed that a source of light creates periodic disturbances, which travel as waves in a way similar to that of sound travelling in air. This theory explained reflection, refraction, interference and diffraction, but rectilinear propagation and polarization could not be explained. In those days all the known waves required a material medium for their propagation. This led Huygen to assume an all-pervading homogeneous medium called *luminiferous ether*, which possessed inertia and elasticity. In order to explain the approximate rectilinear propagation of light, Fresnel and Young assumed the light to be transverse in nature.

To explain the light propagation as transverse waves with a high velocity, the ether was supposed to possess the properties of an elastic solid. The properties expected from the ether were contradictory and difficult to accept. It was the famous Michelson-Morley experiment which experimentally ruled out the existence of luminiferous ether. However, the concept of elastic ether was very useful in the theoretical development of the subject of light during the nineteenth century and it was discarded only after the advent of *Maxwell's electromagnetic theory* (1873), which assumed the light to be electromagnetic wave in nature and so, no material medium was required for its propagation.

Max Planck proposed *quantum theory of light*, according to which light consists of small particles in the form of discrete bundles of energy called *photons*. The energy of a photon is equal to $h\upsilon$, where υ is its frequency and h is the Planck's constant.

The present-day understanding about the nature of light is that it possesses *dual* (or double) *character*. Sometimes light manifests its wave nature (as in reflection, refraction, diffraction, polarization, etc.), while in some processes it manifests its particle nature as in photoelectric effect, Raman effect, Compton effect, etc.

We are going to discuss the phenomena like interference, diffraction and polarization manifesting the wave nature of light.

1.2 PRINCIPLE OF SUPERPOSITION

When two or more waves travelling through a medium superimpose each other, they lose their individual identity and a new wave is formed, whose amplitude is determined using the superposition principle. According to the principle of superposition, *the resultant amplitude at a point at any instant of time is the algebraic sum of the amplitudes due to the individual waves.* That is, if A is the resultant amplitude and A_1, A_2, A_3, ... are the amplitudes due to the individual waves, then

$$A = A_1 \pm A_2 \pm A_3 \pm \ldots \quad \ldots(1.1)$$

The +ve sign is taken when amplitudes of the waves are in the same direction and −ve sign when they are in opposite direction. The superposition principle is illustrated in Fig. 1.1. The resultant intensity is the square of the resultant amplitude.

$$I = A^2. \quad \ldots(1.2)$$

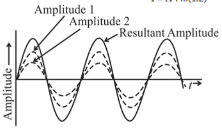


Fig. 1.1: Superposition of waves of the same phase and frequency.

1.3 COHERENT LIGHT SOURCES

Two light sources are said to be coherent when there is *no phase difference or a constant phase difference*, if the phase difference exists, between the light waves from the two sources.

Two independent sources of light cannot be coherent. We know that when an electron makes a transition from a higher energy state to a lower energy state, it gives out the balance energy in the form of radiation. When the frequency of these radiations lies within the visible spectrum, we get light, which consists of a broken chain of wave trains, accompanied by sudden and abrupt changes in phase, occurring at very short intervals of time (of the order of 10^{-8} sec). Consequently, with two independent sources, the phases of waves originating from them will be changing independently and, therefore, the phase difference between the two sources cannot be constant. Hence, two independent sources of light cannot be coherent.

The two sources to be coherent must emit light waves of the same frequency, nearly the same amplitude in addition to a constant phase difference. This means the wavelength (or colour) of the coherent sources is the same.

In general, the coherence or phase between two light waves can vary from point to point (in space) or change from instant to instant (or time) at a point. Thus, there are two types of coherence, which are discussed below:

Temporal Coherence: If the waves given by a light source maintain coherency at a point at two different instants of time, i.e., coherency is maintained with respect to time by a source, this type of coherency is referred to as temporal coherence.

Spatial Coherence: If the two waves maintain a constant phase relationship over any time at different points in space, the waves are said to be *spatially coherent*. This is possible even when two beams are individually incoherent, as long as any phase change in one beam is accompanied by a simultaneous equal phase change in the other beam (as in Young's double slit experiment).

Temporal (or time) coherency is a characteristic of a single beam of light, whereas spatial (or space) coherency concerns the relationship between two separate beams of light.

As discussed earlier, light waves given out by deexcitation of atoms are produced in the form of wave trains. These waves trains are of finite length, each wave train contains only a limited number of waves. The length of the wave train Δs is called the *coherence length* and is equal to the product of the number of waves N contained in the wave train and the wavelength λ, i.e., $\Delta s = N\lambda$. Since, velocity is defined as the distance covered per unit time, it takes a wave train of length Δs, certain length of time Δt, to pass a given point.

$$\Delta t = \frac{\Delta s}{c}$$

where c is the velocity of light. The length of time Δt is called the *coherence time.*

1.4 INTERFERENCE OF LIGHT

When two or more light waves propagate in a medium simultaneously, they superimpose each other. As a result, *redistribution* of energy takes place. This phenomenon is called *interference*. Thus, *interference of light is the phenomenon which takes place when two or more waves superimpose each other resulting in the redistribution of light energy.*

The points at which the waves meet in phase, they *reinforce* each other thereby increasing the amplitude and intensity of the resultant wave. This is called *constructive interference*. Thus, *at the points of constructive interference, the waves meet in the same phase or at a phase difference of* $2n\pi (n = 0, 1, 2, ...)$.

In terms of path difference, path difference for constructive interference

$$= 2n \cdot \frac{\lambda}{2},\ n = 0, 1, 2, .$$

That is, for the constructive interference, the path difference should be *even multiple of* $\frac{\lambda}{2}$. The points of constructive interference are the points of maximum brightness or *maxima*.

The points at which the waves meet in opposite phase, they cancel each other. This is called *destructive interference*. At the points of destructive interference, the waves meet in opposite phase or a phase difference of $(2n + 1)\pi$ $(n = 0, 1, 2, 3, ...)$. In terms of path difference, path difference for destructive interference $= (2n+1) \cdot \frac{\lambda}{2} (n = 0, 1, 2, 3,)..$ That is, the path difference should be an odd multiple of $\frac{\lambda}{2}$. These points are the points of *minimum brightness* or *darkness* or *minima*.

When interference pattern is obtained on a screen alternate bright and dark bands or fringes are obtained if the pattern is sustained or permanent. Alternatively, we say a fringe pattern is obtained. A fringe may be bright or dark. A fringe is defined as a *locus of points having a fixed path difference (and hence, phase difference)* from *two fixed points* (the location of the two sources).

1.5 conditions for permanent (Sustained) interference of light

Whenever two light waves meet, they interfere to produce a resultant interference pattern. However, in order to obtain the interference pattern continuously on a screen, the interference pattern must be *permanent* or *sustained*. To obtain a sustained interference pattern, the following conditions must be satisfied.

(*i*) *The two sources of light should be coherent.* There should not be a variation of phase difference with time. As independent sources of light cannot be coherent, the two interfering light waves must be obtained from a single light source. The necessity of constant phase difference can be understood from the following discussion: If φ is the phase difference at a point between the two interfering waves, then intensity at that point is:

$$I \propto \cos^2\varphi/2.$$

Hence, if φ changes with time then the *resultant intensity will change with time* and sustained interference pattern cannot be obtained.

(*ii*) *The two sources should emit light waves continuously of same wavelength and time period or frequency. That is, the two sources should be monochromatic.*

(*iii*) *The separation between the two sources should be small.* Due to this, the separation between the fringes becomes large so that they become visible.

(*iv*) *The separation between the sources and the screen should be large.* This also increases fringe width enhancing their visibility.

(*v*) *The amplitudes of the interfering waves should be equal or nearly equal.* This is necessary for good contrast between bright and dark fringes, i.e., to obtain nearly complete darkness at minima.

(*vi*) *The two sources should be extremely narrow.* If it is not so, the interference between different portions of the same source takes place, which ultimately reduces the visibility of the interference pattern.

(*vii*) *The two interfering waves should meet at a small angle.* If it is not so, a large path difference is introduced between the two interfering waves resulting in poor visibility and poor contrast in the interference pattern.

1.6 Production of coherent sources

For the production of a sustained interference pattern, at least two coherent sources are required. We know that independent sources of light cannot be coherent. Therefore, in all the interferometers, the interfering waves are obtained from a single light source. There are two methods of obtaining coherent interfering waves. These are as follows:

(*i*) **Division of Wavefront:** In this method, a wavefront is divided into two or more parts by reflection, refraction or diffraction. These portions then meet at a small angle to undergo interference and produce an interference pattern. This method is employed in Young's double slit experiment, Fresnel's biprism, Llyod'd mirror, etc.

(*ii*) **Division of Amplitude:** In this method, the amplitude of a beam is divided into two or more parts by a combination of reflection and refraction. These divided parts then reunite to undergo interference, after traversing different paths. In such cases, extended broad source of light is used so that bright fringes are obtained. This method is used in thin film devices, Newton's rings, Michelson interferometer, etc.

The interferometers employing the first method do not produce complementary interference pattern, while those employing the second method can produce complementary pattern with a few exceptions.

1.7 Conditions for constructive and Destructive interference

Let the two coherent light waves undergoing interference be represented by

$$y_1 = a\sin\omega t \quad ...(1.3)$$

$$\text{and } y_2 = b\sin(\omega t + \varphi) \quad ...(1.4)$$

where a and b are their respective amplitudes and φ is the phase difference between the two waves.

Applying the principle of superposition, the resultant displacement y is given by

$$y = y_1 + y_2$$

$$= a\sin\omega t + b\sin(\omega t + \varphi)$$

$$= a\sin\omega t + b\sin\omega t\cos\varphi + b\cos\omega t\sin\varphi$$

$$= \sin\omega t(a + b\cos\varphi) + \cos\omega t \cdot b\sin\varphi$$

$$\text{Putting } R\cos\theta = a + b\cos\varphi \quad ...(1.5)$$

$$\text{and, } R\sin\theta = b\sin\varphi \quad ...(1.6)$$

in the above expression, we get

$$y = \sin\omega t \cdot R\cos\theta + \cos\omega t \cdot R\sin\theta$$

$$= R(\sin\omega t\cos\theta + \cos\omega t \cdot \sin\theta)$$

$$= R\sin(\omega t + \theta) \quad ...(1.7)$$

Thus, R represents the amplitude of the resultant wave.

Squaring Eqs. (1.5) and (1.6) and adding, we get

$$R^2(\cos^2 + \sin^2\theta) = (a + b\cos\varphi)^2 + (b\sin\varphi)^2$$

$$\text{or, } R^2 = a^2 + b^2\cos^2\varphi + 2ab\cos\varphi + b^2\sin^2\varphi$$

$$\text{or, } = a^2 + b^2(\cos^2\varphi + \sin^2\varphi) + 2ab\cos\varphi$$

$$\text{or, } R^2 = a^2 + b^2 + 2ab\cos\varphi \quad ...(1.8)$$

$$\text{or, } R = \sqrt{a^2 + b^2 + 2ab\cos\phi} \quad ...(1.9)$$

Intensity being directly proportional to the square of the amplitude, the intensity of the resultant wave is given by:

$$I = R^2$$

$$\text{or, } I = a^2 + b^2 + 2ab\cos\varphi \quad ...(1.10)$$

For *constructive interference*, i.e., I to be maximum,

$$\cos\varphi = 1$$

$$\text{or, } \varphi = 0, 2\pi, 4\pi, ...$$

$$\text{or, } \varphi = 2n\pi, n = 0, 1, 2, ...$$

That is, *for constructive interference, the phase difference between the two waves should be an even integral multiple of* π.

Path difference for constructive interference

$$= \frac{\lambda}{2\pi} \cdot \phi = \frac{\lambda}{2\pi} \cdot 2n\pi = n\lambda.$$

Thus, *for constructive interference, the path difference between the two interfering waves should be an integral multiple of* λ.

The *maximum amplitude* R_{max}, using Eq. (1.9) and $\cos\varphi = 1$ is given by:

$$R_{max} = \sqrt{a^2 + b^2 + 2ab} = \sqrt{(a+b)^2} = a + b \quad ...(1.11)$$

Thus, when constructive interference takes place, the resultant amplitude is the sum of the amplitudes of the two waves, which is the maximum amplitude of the resultant wave.

$$\text{The maximum intensity, } I_{max} = R_{max}^2 = (a+b)^2 \quad ...(1.12)$$

This shows that the *maximum intensity is greater than the sum of the two individual intensities* $a^2 + b^2$.

For *destructive interference*, i.e., I to be minimum

$$\cos\varphi = -1$$

$$\text{or, } \varphi = \pi, 3\pi, 5\pi, ...$$

$$\text{or, } \varphi = (2n + 1)\pi, n = 0, 1, 2, 3, ...$$

Thus, for *destructive interference, the phase difference between the two waves should be an odd integral multiple of* π.

Path difference for destructive interference

$$= \frac{\lambda}{2\pi} \cdot \phi = \frac{\lambda}{2\pi} \cdot (2n+1)\pi = (2n+1) \cdot \frac{\lambda}{2}$$

Thus, *for destructive interference, the path difference between the two interfering waves should be an odd integral multiple of half the wavelength.*

The minimum amplitude R_{min}, using Eq. (1.9) and $\cos\varphi = -1$, is given by

$$R_{min} = \sqrt{a^2 + b^2 - 2ab} = \sqrt{(a-b)^2} = a - b \quad ...(1.13)$$

Thus, when destructive interference takes place, the resultant amplitude is equal to the difference of the amplitudes of the two waves, which is the minimum amplitude of the resultant wave.

$$\text{The minimum intensity } I_{min} = R_{min}^2 = (a-b)^2.$$

We see that resultant intensity is less than the sum of the two individual intensities $a^2 + b^2$.

In case, the amplitudes of two waves are equal, $a = b$

$$\therefore R_{max} = a + b = a + a = 2a$$

$$I_{max} = R_{max}^2 = 4a^2 \quad ...(1.14)$$

$$R_{min} = a - b = a - a = 0$$

$$\therefore I_{min} = R^2_{min} = 0 \quad ...(1.15)$$

Therefore, the points of destructive interference will be perfectly dark.

1.8 law of conservation of energy and interference

When the two waves $y_1 = a\sin\omega t$ and $y_2 = b\sin(\omega t + \varphi)$, having a phase difference φ interfere, the intensity of the resultant wave is given by Eq. (1.8),

$$I = R^2 = a^2 + b^2 + 2ab\cos\varphi$$

If $a = b$, then:

$$I = a^2 + a^2 + 2a^2\cos\varphi = 2a^2 + 2a^2\cos\varphi$$

$$= 2a^2(1 + \cos\varphi)$$

$$= 4a^2\cos^2\frac{\phi}{2} \quad ...(1.16)$$

Thus, the intensity distribution in the interference pattern is proportional to $\cos^2\left(\frac{\phi}{2}\right)$ as shown in Fig. 1.2. As evident in the figure,

$$\text{for, } \varphi = 0, \pm 2\pi, \pm 4\pi, ..., I = I_{max} = 4a^2.$$

Thus, the intensity varies between zero and $4a^2$ depending upon the phase difference φ between the interfering waves. One may think that the light energy disappears from the points of destructive interference in contradiction to the law of conservation of energy which states that the energy can neither be created nor destroyed. Also, at the points of constructive interference the intensity is $4a^2$, two times the sum of individual intensities, $I = a^2 + a^2 = 2a^2$. A close look at the *phenomenon of interference shows that it is in accordance with the law of conservation of energy*. The light energy is transferred to the points of constructive interference from the points of destructive interference. That is, the *energy is merely redistributed*. This becomes clear from the following discussion.

In a region where each beam acts separately, without undergoing interference, the intensity due to each beam is equal to a^2. So the region will have a uniform intensity equal $a^2 + a^2 = 2a^2$, as shown by the horizontal dotted line in Fig. 1.2.

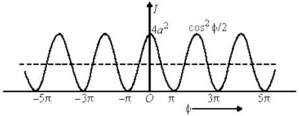


Fig. 1.2

When the beams undergo interference, the intensity is given by

$$I = 4a^2\cos^2\frac{\phi}{2}$$

Let us find the average of the intensity between 0 and π,

$$I_{av} = 4a^2 \cdot \frac{1}{\pi}\int_0^\pi \cos^2\frac{\phi}{2}$$

$$= 4a^2 \cdot \frac{1}{2\pi}\int_0^\pi (1+\cos\phi)d\phi$$

$$= 4a^2 \cdot \frac{1}{2\pi}\left(\int_0^\pi d\phi + \int_0^\pi \cos\phi d\phi\right)$$

$$= 4a^2 \cdot \frac{1}{2\pi}\cdot\pi \quad \left(\because \int_0^\pi \cos\phi\, d\phi = 0\right)$$

$$= 2a^2$$

which is the same as obtained earlier. In other words, the average intensity of the interfering beams remains the same. Thus, the *energy* is neither created nor destroyed but is merely *redistributed in the interference pattern*. Hence, the phenomenon of interference is in accordance with the law of conservation of energy.

1.9 young's double slit experiment

Interference was first observed by Thomas Young, an English scientist, in 1802, thus establishing the wave nature of light.

The experimental arrangement is shown in Fig. 1.3. S is a narrow slit illuminated by a monochromatic light of wavelength λ. There are two narrow slits S_1 and S_2, parallel and close to each other at a distance Δ apart. Light waves spread out from S and fall on both S_1 and S_2, which *serve* as the two coherent sources of light. Thus, the two coherent light waves are obtained from the same original source. Light waves spread out from the slits S_1 and S_2 and fall on the screen. These two waves interfere and equally spaced bright and dark bands, called *interference fringes*, are formed on the screen, at a distance Δ from the slits. To get widely spaced fringes, the two slits should be quite close to each other, about 0.2 mm apart.

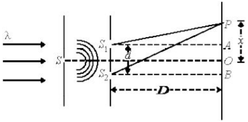


Fig. 1.3

To find the positions at which the bright and dark bands are located and the separation between them, let us consider a point P on the screen at a distance x from its centre O. The path difference p between the two waves on reaching P is given by:

$$p = S_2P - S_1P$$

From the right angled triangles S_2BP and S_1AP, using Pythagoras theorem, we have:

$$S_2P^2 - S_1P^2 = S_2B^2 + PB^2 - (S_1A^2 + PA^2)$$

$$= \left[D^2 + \left(x + \frac{d}{2}\right)^2\right] - \left[D^2 + \left(x - \frac{d}{2}\right)^2\right]$$

$$= D^2 + x^2 + \frac{d^2}{4} + xd - D^2 - x^2 - \frac{d^2}{4} + xd$$

$$= 2xd$$

$$\text{or, } (S_2P + S_1P)(S_2P - S_1P) = 2xd$$

$$\text{or, } S_2P - S_1P = \frac{2xd}{S_2P + S_1P}$$

In practice, points O and P are close enough so as to assume:

$$S_1P = S_2P \cong \Delta$$

$$\therefore \text{Path difference } p = S_2P - S_1P = \frac{2xd}{2D}$$

$$\text{or, } p = \frac{xd}{D} \quad ...(1.17)$$

Location of Bright Fringes: At the points where the path difference is equal to an integral multiple of λ, bright fringes or maxima are located, i.e., the location of nth bright fringe is

$$\frac{x_n d}{D} = n\lambda, \; n = 0, 1, 2, 3, ...$$

$$\text{or, } x_n = \frac{n\lambda D}{d} \quad ...(1.18)$$

For $n = 0$, $x_0 = 0$, this corresponds to the central bright fringe.

For $n = 1,\ x_1 = \dfrac{\lambda D}{d},$ this gives the location of first order bright fringe. Similarly, we can find the locations of higher order fringes.

Fringe Width is defined as the separation between any two consecutive fringe. Hence, fringe width β is given by:

$$\beta = x_n - x_{n-1}$$

$$= \frac{n\lambda D}{d} - \frac{(n-1)\lambda D}{d}$$

$$\text{or, } \beta = \frac{\lambda D}{d} \quad ...(1.19)$$

Thus, the fringe width β is directly proportional to Δ and λ and inversely proportional to Δ. It is evident from Eq. (1.19), that the *fringe width is independent of the order of fringe*. In other words, the separation between second and third fringes is the same as that between tenth and eleventh fringes.

Location of Dark Fringes: At the points where the path difference is equal to an odd multiple of $\dfrac{\lambda}{2},$ dark fringes or minima are located, i.e., the location of nth dark fringe is:

$$\frac{x_n d}{D} = (2n-1)\frac{\lambda}{2}, \quad n = 1, 2, 3$$

$$\text{or, } x_n = \frac{(2n-1)D\lambda}{2d} \quad ...(1.20)$$

When $n = 1,\ x_1 = \dfrac{D\lambda}{2d},$ this gives the location of the first dark fringe.

For $n = 2,\ x_2 = \dfrac{3D\lambda}{2d},$ this gives the location of the second dark fringe.

Similarly, we can find the locations of higher order fringes. Comparison of these locations with that of bright fringes shows that the dark fringes are located between the bright fringes.

Fringe width of the dark fringes is

$$\beta = x_n - x_{n-1}$$

$$= (2n-1)\frac{\lambda D}{2d} - \left[[2(n-1)-1]\frac{\lambda D}{2d}\right]$$

$$\text{or, } \beta = \frac{\lambda D}{d}$$

which is the same as that for bright fringes. Hence, we can conclude that *all bright and dark fringes are of equal width*. The intensity distribution in Young's double slit experiment is shown in Fig. 1.4.

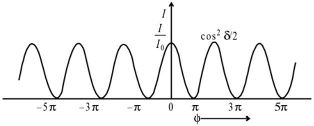


Fig. 1.4: Intensity variation in Young's experiment.

One important application of this experiment is that it can be used to determine the wavelength of the used light by measuring fringe width β and knowing Δ and Δ.

Interference Pattern with White Light: If instead of monochromatic light, white light is used, all the seven constituent colours produce their own interference pattern. At the centre of the screen, all the wavelengths meet in the same phase and, therefore, a white bright fringe is formed at the centre. On either side of this central bright (white) band, the first dark fringe formed will be due to the violet colour (ideally it should not be visible) as the wavelength is shortest for violet colour and here the two rays with path difference of $\frac{\lambda_v}{2}$ will meet. This will be followed by that due to indigo, blue, ... red colours. Then, the bright fringes due to all these colours will be formed in the same order. First few coloured fringes are visible, thereafter, due to the overlapping of colours, the fringes are not clearly visible.

Shape of the Fringe: A fringe is the locus of points for which the path (or phase) difference from the two sources S_1 and S_2 is a constant. Hyperbola is the geometrical structure which satisfies this condition. Therefore, the fringes in case of Young's double slit experiment is *hyperbolic* in shape. However, due to the small wavelength of light and large distance of the screen from the sources, the fringes appear to the *straight lines*. In other words, when Δ >> Δ, the observed fringes appear to be parallel, equidistant, bright and dark lines.

Example 1.1

In Young's double slit experiment, the width of the fringes obtained from a source of light of wavelength 5000 Å is 3.6 mm. Calculate the fringe width if the apparatus is immersed in a liquid of refractive index 1.2.

Solution:

Here, $\lambda = 5000\ \text{Å} = 5000 \times 10^{-10}\ \text{m} = 5.0 \times 10^{-7}\ \text{m}$, $\beta = 3.6\ \text{mm}$, $\mu = 1.2$.

Refractive index of liquid

$$\mu = \frac{\text{Speed of light in space}}{\text{Speed of light in liquid}} = \frac{c}{\nu} = \frac{\nu\lambda}{\nu\lambda'} = \frac{\lambda}{\lambda'}$$

$$\text{or, } \lambda' = \frac{\lambda}{\mu}$$

∴ Fringe width in liquid

$$\beta' = \frac{D\lambda'}{d} = \frac{D\lambda}{d\mu} = \frac{\beta}{\mu} = \frac{3.6\ \text{mm}}{1.2} = 3.0\ \text{mm}.$$

Example 1.2

Two slits 0.125 mm apart are illuminated by a light of wavelength 4500 Å. The screen one metre away from the plane of slits. Find the separation between the second bright fringe on both the sides of the central maximum.

Solution:

Here, $\Delta = 0.125\ \text{mm} = 1.25 \times 10^{-4}\ \text{m}$,

$\lambda = 4500\ \text{Å} = 4.5 \times 10^{-7}\ \text{m}$, $\Delta = 1\ \text{m}$

Distance of second bright fringe (x_2) from the central maximum

= 2 × Fringe width

$$\therefore x_2 = 2\beta = \frac{2D\lambda}{d} = \frac{2 \times 1 \times 4.5 \times 10^{-7}}{1.25 \times 10^{-4}} = 7.2 \times 10^{-3}\ \text{m}$$

$= 7.2 \times 10^{-3}\ \text{m}$

∴ Separation between second bright fringes

$= 2x_2 = 2 \times 7.2 \times 10^{-3} = 14.4 \times 10^{-3}\ \text{m}.$

Example 1.3

A Young's double slit apparatus is immersed in a liquid of refractive index 1.25. Find the ratio of fringe widths in air and liquid.

Solution:

Fringe width in air:

$$\beta_a = \frac{D\lambda_a}{d} \quad ...(i)$$

Fringe width in liquid:

$$\beta_l = \frac{D\lambda_l}{d} \quad ...(ii)$$

$$\text{Refractive index, } \mu = \frac{\text{Wavelength of light in air}}{\text{Wavelength of light in liquid}} = \frac{\lambda_a}{\lambda_l}$$

$$\therefore \lambda_l = \frac{\lambda_a}{\mu}$$

Dividing (*i*) by (*ii*), we get:

$$\frac{\beta_a}{\beta_l} = \frac{\lambda_a}{\lambda_l} = \mu\frac{\lambda_a}{\lambda_a} = 1.25 = \frac{5}{4}$$

or, $\beta_a : \beta_l = 5 : 4$.

Example 1.4

Two narrow and parallel slits 0.08 cm apart are illuminated by a light of frequency 6×10^{11} kHz. At what distance from the slits should the screen be placed to obtain fringe width of 0.6 mm?

Solution:

Here, $\Delta = 0.08$ cm $= 0.08 \times 10^{-2}$ m, $\beta = 0.6$ mm $= 0.6 \times 10^{-3}$ m

$$\lambda = \frac{c}{\nu} = \frac{3\times10^8}{8\times10^{11}\times10^3} = 3.75\times10^{-7} \text{ m}$$

$$\therefore D = \frac{\beta\, d}{\lambda} = \frac{0.6\times10^{-3}\times0.08\times10^{-2}}{3.75\times10^{-7}} = 1.28 \text{ m.}$$

Example 1.5

In Young's double slit experiment, the separation between the slits is 1 mm and that between the slits and the screen is 1 m. The wavelength of the used light is 5893 Å. Compare the intensity at a point 1 mm from the centre to that at its centre. Also find the minimum distance from the centre of a point where the intensity is half of that at the centre.

Solution:

We know that the path difference at a point x on the screen from the centre is $\frac{xd}{D}$.

Here, $\Delta = 1$ mm $= 10^{-3}$ m, $\Delta = 1$ m

Therefore, path difference at a distance 1 mm ($x = 1$ mm $= 10^{-3}$ m) from the centre is

$$= \frac{10^{-3}\times10^{-3}}{1} = 10^{-6} \text{ m}$$

Corresponding phase difference $\phi = \frac{2\pi}{\lambda}$ path difference.

$$= \frac{2\times3.14\times10^{-6}}{5893\times10^{-10}} = 3.394\,\pi \text{ radians}$$

$\therefore$ Ratio of intensity with the central maximum

$$= \cos^2\frac{\phi}{2} = \cos^2\left(\frac{3.394\pi}{2}\right) = \cos^2(1.697\pi) = 0.3372$$

When the intensity is half of the maximum, phase difference φ is such that:

$$\cos^2\frac{\phi}{2} = 0.5 \Rightarrow \frac{\phi}{2} = 45° \Rightarrow \phi = \frac{\pi}{2}$$

$\therefore$ Corresponding path difference:

$$p = \phi\cdot\frac{\lambda}{2\pi} = \frac{\pi}{2}\times\frac{\lambda}{2\pi} = \frac{\lambda}{4}$$

The corresponding distance from the centre on the screen:

$$x = p\cdot\frac{D}{d} = \frac{\lambda}{4}\times\frac{1\text{ m}}{10^{-3}} = \frac{5893\times10^{-10}}{4\times10^{-3}}$$

$= 1.473 \times 10^{-4}$ m.

Example 1.6

In Young's double slit experiment, source of light of wavelength 4200 Å is used to obtain interference, fringes of width 0.64×10^{-2} m are formed. What should be the wavelength of the light source to obtain fringes 0.46×10^{-2} m wide, if the distance between the screen and the slits is reduced to half the initial value?

Solution:

Here, $\lambda = 4200$ Å $= 4200 \times 10^{-10}$ m, $\beta = 0.64 \times 10^{-2}$ m

We know, $\beta = \frac{\lambda D}{d}$

$$\therefore 0.64 \times 10^{-2} = \frac{4200\times10^{-10}\times D}{d} \quad ...(i)$$

In the second case, $\beta = 0.46 \times 10^{-2}$ m, $\lambda = ?$

Now $\Delta = \frac{D}{2}$,

as the distance between the screen and the slits is reduced to half the initial value.

$$\therefore 0.46 \times 10^{-2} = \frac{\lambda \times \frac{D}{2}}{d} = \frac{\lambda D}{2d} \quad ...(ii)$$

Dividing the first expression by the second, we get:

$$\frac{0.64}{0.46} = \frac{4200 \times 10^{-10}}{\lambda} \times 2$$

$$\text{or, } \lambda = \frac{4200 \times 10^{-10} \times 0.46 \times 2}{0.64}$$

$$= 6037.5 \text{ Å}.$$

Example 1.7

In a Young's double slit experiment, the angular width of a fringe formed on a distant screen is 0.1°. If the light used has a wavelength of 6000 Å, calculate the spacing between the slits.

$$\beta_q = \theta_{n+1} - \theta_n = \frac{y_{n+1}}{D} - \frac{y_n}{D} = \frac{\beta}{D}$$

$$\text{or, } \beta_q = \frac{\lambda D}{Dd} = \frac{\lambda}{d}$$

$\therefore$ Spacing between the slits $d = \dfrac{\lambda}{\beta_\theta}$

Here, $\lambda = 6000 \text{ Å} = 6 \times 10^{-7}$ m,

$$\beta_q = 0.1° = \frac{0.1 \times \pi}{180} \text{ radians}$$

$$\therefore \Delta = \frac{6 \times 10^{-7} \times 180}{0.1 \times 3.14} = 3.44 \times 10^{-3} = 0.344 \text{ mm.}$$

1.10 fresnel's biprism experiment

Critics of the Young's experiment argued that the fringes observed by Young were probably due to some complicated modification of light by the edge of the slits and not due to the true interference. To remove any doubt about the fact that the fringes obtained in Young's experiment were indeed due to interference of light waves, Fresnel devised new experiments like biprism experiment, double mirror experiment, etc. We discuss here Fresnel's biprism experiment.

A biprism is a combination of two thin prisms with their bases joined and two faces making an obtuse angle of about 179° so that the two other angles are each of about 30′ or 0.5°. In actual practice, the biprism is ground from a single optically true glass plate. A schematic diagram of the biprism experiment is shown in Fig. 1.5.

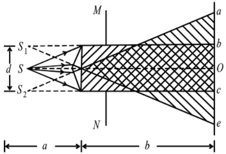


Fig. 1.5: Fresnel's biprism experiment.

S is a narrow vertical slit, perpendicular to the plane of paper. It is illustrated by a monochromatic source of light of wavelength λ. The light from the slit *S* falls on the prism. It refracts the light into two overlapping beams *be* and *ac* which seems coming from S_1 and S_2 respectively. Each half of the biprism produces a virtual image. In other words, two virtual images S_1 and S_2 are formed, which act as coherent sources. If screens µ and *N* are placed as shown in Fig. 1.5, the interference fringes are obtained only in region *bc*. In this region, the fringes are of equal length indicating that the fringes are due to interference. Beyond the region *bc*, fringes of large fringe width are obtained, which are due to diffraction of light. The wider bands are produced by the vertices of the two prisms, each of which acts as straight edge, giving a diffraction pattern. With this experiment, Fresnel was able to produce interference fringes without relying upon diffraction (as in Young's experiment). To observe the fringes, the screen is replaced by an eyepiece. When the point *O* is the principal focus of eyepiece, the fringes are observed in the field of view.

The point *O* being equidistant from S_1 and S_2, a bright fringe is formed at *O*. On both the sides of *O*, alternate bright and dark fringes are formed. If $S_1S_2 = \Delta$ and $SO = \Delta$, it can be shown, as in Young's double slit experiment, location of *n*th bright fringe $x_n = \dfrac{n\lambda D}{d} (n = 0, 2, 3, \ldots)$, location of *n*th dark fringe $x_n = \dfrac{(2n-1)D\lambda}{2d} (n = 1, 2, 3, \ldots)$, and fringe width $\beta = \dfrac{\lambda D}{d}$.

Determination of Wavelength Using Biprism: Fresnel's biprism can be used to determine the wavelength of the used monochromatic light. For this, the fringe width has to be determined experimentally. Then knowing Δ and Δ, the wavelength λ can be calculated using:

$$\lambda = \frac{d\beta}{D}$$

From Fig. 1.5, $\Delta = a + b$

$$\lambda = \frac{d\beta}{a+b} \quad ...(1.21)$$

Experimental Procedure: The experiment is performed on an optical bench having uprights for slit, biprism, lens and eyepiece. A specially designed variable vertical slit *S*, biprism *P*, micrometer eyepiece are mounted in the uprights capable of vertical and transverse adjustments and set on the optical bench with scale graduated in mm. The bed of the optical bench is levelled by spirit level. It is ensured that all the devices are at the same height. The eyepiece is focussed on its cross-wires either by moving the lens system or the cross-wires holder. The slit is made vertical and narrow. The slit is illuminated by monochromatic source of light, say, sodium light. The biprism is brought closer to the slit. On looking through the prism, two images of the slit are seen. At this stage, the edge of the biprism should be parallel to the slit. Now, the eyepiece is brought closer to the biprism and the overlapping regions are obtained in the biprism at right angle to the length of the optical bench. The interference fringes present in this overlapping region are obtained in the focal plane. For clear fringes, it is necessary that the edge of the biprism should be made perfectly parallel to the slit and the slit width should be made narrow. To make the slit parallel to the edge, the tangential screw provided with the biprism should be adjusted so that the clear fringes are seen.

Lateral Shift Removal: After getting the clear fringes, *lateral shift,* if present *should be removed,* i.e., fringes should not shift relative to cross-wires when the eyepiece is moved along the bench. When the axis of the optical bench and the axis of the observation, passing through slit, biprism edge and cross-wire, are inclined with each other, the fringe pattern laterally shifts to right or left on moving the eyepiece. It is shown in Fig. 1.6, when the axes are not inclined, i.e., they coincide, even on moving the eyepiece, no movement of fringe pattern is observed. If present, the lateral shift can be removed by the following process. Move the eyepiece backward and say, the fringes shift towards right. Adjust the biprism perpendicular to the bench in a direction so that the fringes move back to the left and the same fringe coincides with the cross-wires. Now, move the eyepiece towards the biprism, the fringes shift in the opposite direction. Repeat the same process to bring back the same fringe to coincide with the cross-wire. Repeat the process, till the lateral shift is removed completely.

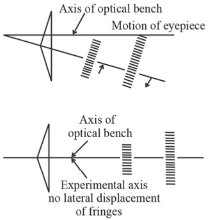


Fig. 1.6: Removal of lateral shift of fringes.

Measurements: In order to find wavelength λ, fringe width β, separation between virtual sources Δ and the separation between slit and the focal plane of the eyepiece Δ is determined as follows:

***(i)* Determination of β:** The centre of a bright fringe is brought on the cross-wires of the micrometer eyepiece by adjusting the slow motion screw attached to it and the position of the eyepiece is read on the scale. The micrometer is moved so that a number of fringes (say *n*) pass the cross-wires. Note down the distance (*x*) by which the micrometer has been displaced. Then, fringe width

$$\beta = \frac{x}{n}.$$

***(ii)* Determination of *D*:** The distance between the slit and the focal plane of the eyepiece is noted on the optical bench scale and the value of Δ is determined after the usual index correction. The index correction between the slit and the eyepiece upright is determined by making one end of a knitting needle touch the slit and focussing the eyepiece on the other end. Then the difference between length λ of the needle and the observed distance λ′ between the uprights, i.e., $l - l'$ gives the required index correction, which on being added to the observed Δ, gives the correct value of Δ.

***(iii)* Determination of *d*:** This is the most important part of the experiment and the method employed is often called the *displacement method*. A convex lens of the focal length less than one-fourth of the distance between the biprism and the eyepiece is mounted between the biprism and the eyepiece, as shown in Fig. 1.7, without disturbing the positions of other uprights. The position of the lens (λ) is adjusted till a clear view of two virtual sources S_1 and S_2 is seen through the eyepiece. The distance (say Δ_1) between S_1 and S_2 is measured with the help of the micrometer eyepiece.

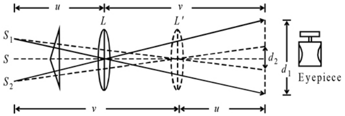


Fig. 1.7: Fresnel's biprism using lens.

Now, without disturbing other uprights, the lens is moved to the second position (λ'), so that again the two images are visible. Let Δ_2 be the separation between S_1 and S_2 for this position of lens, as noted using the micrometer.

If u and v are respectively the distances of the slit and the eyepiece from the lens, when at λ, then from magnification relation, we have:

$$\frac{v}{u} = \frac{d_1}{d}$$

At the conjugate position λ', the value of u and v are interchanged, so that,

$$\frac{u}{v} = \frac{d_2}{d}$$

Multiplying the above equations, we have

$$\frac{v}{u} \times \frac{u}{v} = \frac{d_1 d_2}{d^2}$$

$$\text{or, } \Delta^2 = \Delta_1\Delta_2$$

$$\text{or, } \Delta = \sqrt{d_1 d_2} \quad ...(1.22)$$

Alternatively, Δ can be measured by the following method. For a prism of very small refracting angle α, the deviation Δ produced in the ray is given by:

$$\Delta = (\mu - 1)\alpha \quad ...(1.23)$$

where, μ is the refractive index of the material of the prism. From Fig. 1.8, we have:

$$\Delta = 2\Delta \times a \quad ...(1.24)$$

where, a is distance between the slit and the biprism.

Substituting the value of Δ from Eq. (1.23), in Eq. (1.24), we get

$$\Delta = 2(\mu - 1)\alpha a \quad ...(1.25)$$

Here, α is in radians.

Thus, measuring β, Δ and Δ, the wavelength λ can be calculated using Eq. (1.21). Typically for Δ = 0.01 cm, λ = 6000 Å, Δ = 50 cm, β is approximately equal to 0.3 mm.

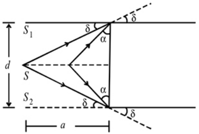


Fig. 1.8

Fringes with White Light: If instead of monochromatic light, white light is used, then central fringe at O will be white as all the seven wavelengths meet there in phase. All other fringes are coloured. As fringe width β is directly proportional to wavelength λ, on either side of the central white band, the first fringe will be of violet colour followed by indigo, blue,..., red. Because of the values of λ for violet and red, the fringe width of red fringe will be nearly double than that of violet. After the first few coloured fringes, the clarity of fringes diminishes and a general illumination fading off to white results due to their overlapping. *To locate central fringe, white light source is used,* which is not possible with monochromatic light as in this case, all fringes are alike.

Example 1.8

In a biprism experiment, a convex lens is placed in between the biprism and the eyepiece. In two different positions of the lens the distance between the images obtained in the eyepiece are 0.42 mm and 1.21 mm. In one position of the eyepiece, the fringe width is 0.4 mm; when the eyepiece is moved away 60 cm, the fringe width increases by 0.5 mm. Find the wavelength of the used light.

Solution:

$$\text{Here } \Delta_1 = 0.42 \text{ mm} = 0.42 \times 10^{-3} \text{ m}$$

$$\Delta_2 = 1.21 \text{ mm} = 1.21 \times 10^{-3} \text{ m}$$

∴ Separation between the slits

$$\Delta = \sqrt{d_1 d_2} = \sqrt{0.42 \times 1.21} \times 10^{-3} \text{ m}$$

$$= 0.7128 \times 10^{-3} \text{ m}$$

$$\text{Fringe widths: } \beta_1 = \frac{D_1\lambda}{d} \text{ and } \beta_2 = \frac{D_2\lambda}{d}$$

$$\therefore \beta_2 - \beta_1 = \frac{(D_2 - D_1)\lambda}{d} \Rightarrow \lambda = \frac{(\beta_2 - \beta_1)d}{(D_2 - D_1)}$$

$$\text{Given, } \Delta_2 - \Delta_1 = 60 \text{ cm} = 0.60 \text{ m},$$

$$\text{and, } \beta_2 - \beta_1 = 0.5 \text{ mm} = 0.5 \times 10^{-3} \text{ m}$$

$$\therefore \text{Wavelength } \lambda = \frac{0.5 \times 10^{-3} \times 0.7128 \times 10^{-3}}{0.60}$$

$$= 5940 \times 10^{-10} \text{ m}$$

$$= 5940 \text{ Å}.$$

Example 1.9

The refractive index of a biprism is 1.5. With a monochromatic source of light, the fringe width obtained is 0.2 mm. The whole set-up is now immersed in a liquid of refractive index 1.3. What is the new fringe width?

Solution:

The distance between the two virtual sources is given by:

$$\Delta = 2(\mu - 1)\alpha a$$

On immersing in the liquid of refractive index μ_l, Δ becomes:

$$\Delta_l = 2\left(\frac{\mu - \mu_l}{\mu_l}\right)\alpha a$$

$$\text{or, } \Delta_l = d\,\frac{\mu - \mu_l}{(\mu - 1)\mu_l}$$

$$\text{Fringe width in air, } \beta = \frac{D\lambda}{d}$$

Fringe width in liquid,

$$\beta_l = \frac{D\lambda_l}{d_l} = \frac{D}{d}\frac{\lambda(\mu - 1)\mu_l}{\mu_l(\mu - \mu_l)} \quad \left(\because \lambda_l = \frac{\lambda}{\mu_l}\right)$$

$$= \beta \cdot \frac{(\mu - 1)}{(\mu - \mu_l)}$$

$$= \frac{0.2 \times 10^{-3} \times 0.5}{0.2} = 5 \times 10^{-4} \text{ m}.$$

Example 1.10

In a biprism experiment using light of wavelength 5860 Å, the bandwidth obtained is 0.3 mm. A mica sheet of thickness 5 µm and refractive index 1.45 is kept in the path of one of the interfering waves. What is the shift of the central band? When another mica sheet is kept in the path of one of the beams, the central bright band is shifted to the position occupied by 5th dark band earlier. Find the thickness of the mica sheet.

Solution:

Let y be the shift in the position of the central band, then:

$$y = \frac{(\mu - 1)tD}{d} = \frac{(\mu - 1)t\beta}{\lambda}$$

Here, $\beta = 0.3 \text{ mm} = 0.3 \times 10^{-3}$ m, $\lambda = 5860$ Å,

$$t = 5 \times 10^{-6} \text{ m}$$

$$\text{m} = 1.45$$

$$\therefore y = \frac{0.45 \times 5 \times 10^{-6} \times 0.3 \times 10^{-3}}{5860 \times 10^{-10}} = 1.52 \times 10^{-3} \text{ m}$$

With the other mica sheet:

$$y_0 = 4.5\beta$$

$$\therefore \text{Thickness } t = \frac{y_0\lambda}{\beta(\mu - 1)}$$

$$= \frac{4.5 \times 5860 \times 10^{-10}}{0.45}$$

$$= 5.860 \times 10^{-6} \text{ m}.$$

Example 1.11

A Fresnel biprism set up is illuminated by sodium light (λ = 5893 Å) and 62 fringes are obtained in the field of view. How many fringes will be observed in the field of view, if we replace the source by a mercury lamp with (*a*) green filter (λ = 5461 Å) and (*b*) violet filter (λ = 4358 Å)?

Solution:

Fringe width with sodium light:

$$\beta_s = \frac{D\lambda_s}{d} \quad ...(i)$$

In the given problem, Δ and Δ are constant while wavelength changes.

(*a*) Mercury lamp produces white light. However, due to a green filter, light of wavelength λ_g = 5461 Å will only illuminate the set-up. Therefore, now the fringe width:

$$\beta_g = \frac{D\lambda_g}{d} \quad ...(ii)$$

Dividing (*i*) by (*ii*), we get:

$$\frac{\beta_s}{\beta_g} = \frac{\lambda_s}{\lambda_g} \quad ...(iii)$$

or, $\beta_g\lambda_s \Rightarrow \beta_g = \dfrac{B_s\lambda_g}{\lambda_s}$ ...(iv)

With sodium light, the number of fringes visible is 62.

∴width of the field of view = 62β

If n_g be the number of fringes visible with green light, then the width of field of view is $n_g\beta_g$

$$\therefore 62\beta_s = n_g\beta_g$$

or, $n_g = \dfrac{62\beta_s}{\beta_g}$ ...(v)

Substituting the value of β_g from (iv) in (v), we get:

$$n_g = \frac{62\beta_s\lambda_s}{\beta_s\lambda_s} = 62\frac{\lambda_s}{\lambda_g} = \frac{62\times 5893\times 10^{-10}}{5461\times 10^{-10}}$$

$$= 66.9 = 67$$

(b) Similarly, for violet filter:

$$n_v = 62\frac{\lambda_s}{\lambda_v} = \frac{62\times 5893\times 10^{-10}}{4358\times 10^{-10}} = 83.84 \simeq 84$$

Example 1.12

Fresnel's biprism having an angle of 1° and refractive index 1.5 forms fringes on a screen placed 0.8 m from biprism. If the distance between the source and the biprism is 0.02 m, find the fringe width, if the wavelength of the used light is 6900 Å.

Solution:

Here, $\mu = 1.5,\ \alpha = 1° = \dfrac{1}{180}\times\pi \text{ radian},\ a = 0.02 \text{ m}$

$\Delta = 0.8 + 0.02 = 0.82$ m, $\lambda = 6900$ Å

$= 6.9\times 10^{-7}$ m

We know, $\Delta = 2(\mu - 1)\alpha a$

$$= 2(1.5-1)\times\frac{\pi}{180}\times 0.02$$

$$= 2\times 0.5\times\frac{3.14}{180}\times 0.02$$

Fringe width, $\beta = \dfrac{D\lambda}{d} = \dfrac{0.82\times 6.9\times 10^{-7}\times 180}{2\times 0.5\times 3.14\times 0.02}$

$$= \frac{1018.44}{0.0628}\times 10^{-7} = 1.62\times 10^{-3}\text{ m} = 1.62\text{ mm.}$$

Example 1.13

Fringes are formed by a Fresnel's biprism in the focal plane of a reading microscope, which is 100 cm from the slit. A lens inserted between the biprism and the microscope gives two images of the slit in two positions. In one case, the two images of the slit are 4.05 mm apart and in the other, they are 2.90 mm apart. If sodium light (λ = 5893 Å) is used, find the fringe width.

Solution:

Here, $\Delta = 100$ cm = 1 m, $\Delta_1 = 4.05$ mm $= 4.05\times 10^{-3}$ m

$\Delta_2 = 2.90$ mm $= 2.90\times 10^{-3}$ m,

$\lambda = 5893$ Å $= 5893\times 10^{-10}$ m

Separation between the slits:

$$\Delta = \sqrt{d_1 d_2} = \sqrt{4.05\times 10^{-3}\times 2.90\times 10^{-3}}$$

$= 3.427\times 10^{-3}$ m

Fringe width, $\beta = \dfrac{D\lambda}{d} = \dfrac{1\times 5893\times 10^{-10}}{3.427\times 10^{-3}} = 1.72\times 10^{-4}\text{ m.}$

1.11 Interference in parallel thin films

The colours in the case of thin films such as soap bubbles, oil spread on a water surface, etc. can be explained on the basis of interference of light beams reflected from the top and bottom surfaces of the films. Such studies have many practical applications such as non-reflecting and anti-glare coatings on the camera lens and other surfaces.

Consider a thin parallel film of uniform thickness t and refractive index μ, as shown in Fig. 1.9. A ray of light SA is incident on the film surface at A at an angle i. The ray is partly reflected along AP and partly refracted along AB with the angle of refraction being r. At B, it is again partly reflected and partly transmitted or refracted. Similarly, reflections and refraction take place at C, Δ, etc. as shown in Fig. 1.9.

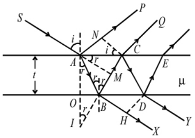


Fig. 1.9

We can study the interference in the reflected and transmitted light rays. This is the case of division of amplitude and the amplitude of any ray is less than the previous one. The rays are collected and focussed by a lens. We consider here interference in reflected and refracted rays separately.

Interference in Reflected Light: To find the condition for constructive and destructive interference, let us find the path difference between a pair of successive reflected rays *AP* and *CQ*. Draw perpendiculars *AM* and *CN* on *BC* and AP respectively. The optical path difference between the rays *AP* and *CQ* is given by:

$$\text{Path difference} = \mu(AB + BC) - AN$$

As shown in Fig. 1.9, ΔAOB is congruent to ΔBOI,

$$\therefore \text{Path difference} = \mu(IB + BC) - AN$$

$$= \mu IC - AN$$

From Snell's law; μ =

$$\frac{\sin i}{\sin r} = \frac{\frac{AN}{AC}}{\frac{CM}{AC}} = \frac{AN}{CM} \Rightarrow AN = \mu CM$$

$$\therefore \text{Path difference} = \mu IC - \mu CM$$

$$= \mu IM$$

$$= \mu AI\cos r = 2\mu AO\cos r$$

$$= 2\mu t\cos r \quad (\text{q } AO = t)$$

Since, the ray *AP* suffers reflection at *A* from an optically denser medium, it undergoes a phase change of π or a path difference of λ/2, where λ is the wavelength of light. While the ray *ABCQ* does not undergo any phase change, since it is reflected at *B*, the surface of a denser to rarer medium. Therefore,

$$\text{Total path difference} = 2\mu t\cos r + \frac{\lambda}{2} \quad \text{...(1.26)}$$

Sometimes, it is expressed as $2\mu t\cos r \pm \frac{\lambda}{2}$, however practically +ve or, –ve signs do not make any difference.

The film will appear *bright* (or the *condition for maxima*) if,

$$2\mu t\cos r + \frac{\lambda}{2} = n\lambda$$

$$\text{or, } 2\mu t\cos r = (2n-1)\frac{\lambda}{2},\ n = 0, 1, 2, 3, \ldots \quad \text{...(1.27)}$$

In such a case, the second ray will be in phase with the first, but third, fifth, seventh, etc., but it will be out of phase with the second, fourth, sixth, etc. Since the second ray is more intense than the third ray, the fourth more intense than the fifth, these pairs cannot cancel each other. The rays of stronger series combine with the first, which is the strongest of all, thus giving rise to maximum intensity.

The film will appear dark (or the *condition for minima*) if,

$$2\mu t\cos r + \frac{\lambda}{2} = (2n+1)\frac{\lambda}{2} \quad (n = 0, 1, 2, 3, \ldots)$$

$$\text{or } 2\mu t\cos r = n\lambda \quad \text{...(1.28)}$$

In this case, second ray is out of phase with the first, but the intensity of the first is much higher than that of the second. Therefore, they do not annul each other completely. The addition of the third, fourth, fifth, etc. which all are in phase with the second will be sufficient to produce nearly complete darkness at the minimum.

It follows from Eq. (1.26), that when the film is very thin compared to the wavelength of the light, i.e. $t << \lambda$, so that $2\mu t\cos r$ can be neglected in comparison to $\frac{\lambda}{2}$, then the film will appear dark as effective path difference reduces to $\frac{\lambda}{2}$.

Interference in Transmitted Light: As shown in Fig. 1.9, the rays emerging from the lower side of the film may be brought together by a lens and can be made to undergo interference. The path difference between the two rays *BX* and *DY* is given by:

$$\mu(BC + CD) - BH$$

which, as in the previous case, can be shown by making a similar construction, to be $2\mu t\cos r$.

In this case, there will not be any additional path difference as both the rays are transmitted from denser to rarer medium, i.e., reflection takes place from surface of rarer medium. Thus, the condition for brightness is:

$$2\mu t\cos r = n\lambda,\ n = 0, 1, 2, 3, \ldots \quad \text{...(1.29)}$$

While the condition for darkness is:

$$2\mu t\cos r = (2n+1)\frac{\lambda}{2},\ n = 0, 1, 2, 3, \ldots \quad \text{...(1.30)}$$

In this, case, if the film is very thin so that $2\mu t\cos r$ can be neglected, then the path difference between the two rays will be zero. Therefore, the two rays will reinforce each other and so the film will appear bright.

It is important to note that the conditions for brightness and darkness are opposite to each other in the reflected and transmitted light. In other words, the conditions for interference in reflected and transmitted light are *complementary*.

Colours in Thin Film: When monochromatic light is used to view a thin film, alternate bright and dark bands are observed. However, when white light is used, brilliant colours are seen due to the following reasons:

(*i*) The path difference between the interfering rays is $2\mu t\cos r$ and so it depends on μ, t and r. If t and r are constants, the path difference would change with μ or the wavelength of light used. Since, white light consists of seven colours, the path difference will be different for different colours, so with white light, the film shows various colours. It is common to see colours in soap bubbles or while standing on a bus stop on the water with some oil film spread over it.

(*ii*) If t and μ are constants and r is varied, the path difference changes, and the film shows different colours when observed from different directions.

(*iii*) If μ and r are constants and t varies, path difference changes for different colours. Therefore, different colours are seen even for the same angle of incidence.

Thus, different colours are seen even if one of the three parameters is varied, when seen in white light.

Conditions for Viewing Colours in Thin Film: In order to view the colours or fringes in thin film, the following two conditions should be satisfied:

(*i*) *A broad source of light is necessary*. The need of a broad source can be understood with the help of Fig. 1.10. If a point source is used, the reflected light from a very small region of the film enters the eye on account of finite size of the pupil of our eye (≈3 mm). Since, white light consists of seven colours and each colour gets reflected at a different angle, only some of the seven colours can be seen. However, if a broad source of light is used, light is incident on the film at a range of angle incidence and so, is reflected in a range of angles making it possible to view a broad region of the film and, therefore, all colours can be seen.

(*ii*) *The thickness of the film should not be too large*. If the film thickness is very large, then the path difference between the rays reflected from the bottom surfaces becomes quite large compared to the wavelength of the light used. This also worsens the mismatch between the amplitudes of the interfering rays, deteriorating the interference pattern. Further, if the thickness of the film is too large (a few millimeters), the coherency between the successive interfering rays is lost and there is no fixed interference pattern.

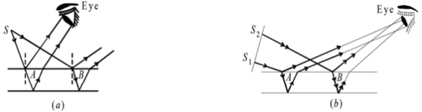


Fig. 1.10: Thin film viewed in (*a*) point source (*b*) extended source.

1.12 interference in wedge-shaped film

Let us consider two plane glass surfaces, PQ and PQ_1, inclined at an angle θ enclosing a wedge-shaped film (Fig. 1.11). Assume that the refractive index of the film is μ. Let the film be illuminated by a monochromatic light of wavelength λ from a slit held at S. Interference occurs between the rays of light reflected from the top and bottom surfaces of the film, producing interference fringes. The path difference p between the two interfering rays is:

$$p = \mu(AB + BC) - AN$$
$$= \mu(AM + MB + BC) - AN$$

Using Snell's law, $\mu = \dfrac{\sin i}{\sin r} = \dfrac{\frac{AN}{AC}}{\frac{AM}{AC}} = \dfrac{AN}{AM}$

or, $\mu AM = AN$

$$\therefore P = \mu(MB + BC) = \mu(MB + BC') = \mu MC'$$
$$= 2\mu t\cos(r + \theta)$$

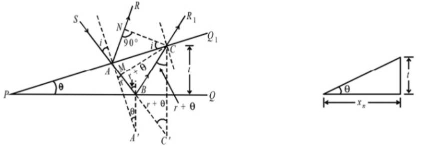


Fig. 1.11: Interference in a wedge-shaped film.

Due to reflection from the denser medium, an additional path difference of $\dfrac{\lambda}{2}$ is introduced, therefore,

$$\text{Total path difference} = 2\mu t\cos(r + \theta) + \frac{\lambda}{2} \quad \ldots (1.31)$$

For *constructive interference:*

$$2\mu t\cos(r + \theta) + \frac{\lambda}{2} = n\lambda$$

$$\text{or } 2\mu t\cos(r+\theta) = (2n-1)\frac{\lambda}{2}, \; n = 1, 2, 3, \ldots \quad \ldots(1.32)$$

For *destructive interference:*

$$2\mu t\cos(r+\theta) + \frac{\lambda}{2} = (2n+1)\frac{\lambda}{2}$$

$$\text{or, } 2\mu t\cos(r+\theta) = n\lambda, \; n = 0, 1, 2\ldots \quad \ldots(1.33)$$

For near normal incidence, r . 0, and in practice, θ is also very small, therefore, $\cos(r+\theta)$. 1. For air $\mu = 1$.

Thus, the condition for brightness reduces to:

$$2\mu t = (2n-1)\frac{\lambda}{2} \quad \ldots(1.34)$$

while that of darkness,

$$2\mu t = n\lambda \quad \ldots(1.35)$$

For small θ, the thickness of the film at location x (Fig. 1.11) can be approximated to $x\theta$. Therefore,

$$\text{For brightness } 2\mu x\theta = (2n-1)\frac{\lambda}{2} \quad \ldots(1.36)$$

$$\text{For darkness } 2\mu x\theta = n\lambda \quad \ldots(1.37)$$

If x_n is the distance of the nth dark fringe from the edge and x_{n+m} that of $(n+\mu)$th, dark fringe, then:

$$x_n = \frac{n\lambda}{2\mu\theta} \text{ and } x_{n+m} = \frac{(n+m)\lambda}{2\mu\theta}$$

$$\therefore \text{Fringe width } \beta = \frac{x_{n+m} - x_n}{m} = \frac{\lambda}{2\mu\theta} \quad \ldots(1.38)$$

This provides a method for calculating λ by measuring fringe width β and angle θ. Also, as $t = x\theta$, this is widely used to calculate the thickness of paper and other thin objects as an experiment in many laboratories.

This experiment can also be used to check whether the two surfaces forming wedge are *optically plane* or not (a plane is said to be optically plane if variation in its thickness is less than $\lambda/10$). *If the two surfaces are optically plane, then the fringes of equal thickness are straight. Each fringe is the locus of points, where the thickness of the film has a constant value.* If the surfaces are not plane, i.e., if thickness varies, there will be irregularity in the fringe, i.e., fringes will not be of equal thickness.

If white light is used to illuminate the film, each of the seven colours produces its own interference pattern. The fringe width of violet is minimum while that of red is maximum. At the edge of the film $t = 0$ and hence the path difference between the interfering rays will be $\lambda/2$ and so the edge of the film appears dark. In the direction of increasing thickness, few coloured fringes appear which increasingly overlap producing general illumination.

Example 1.14

A soap film, 5×10^{-5} cm thick is viewed at an angle of 35° to the normal. Find the wavelength of light in the visible spectrum which will be absent from the reflected light ($\mu = 1.33$).

Solution:

If i is the angle of incidence and r that of refraction, then from Snell's law:

$$m = \frac{\sin i}{\sin r}$$

$$\text{or, } 1.33 = \frac{\sin 35°}{\sin r} \Rightarrow \sin r = \frac{\sin 35°}{1.33} \Rightarrow r = 25.55°$$

$$\therefore \cos r = 0.90$$

$$\text{For darkness, } 2\mu t\cos r = n\lambda$$

$$\lambda = \frac{2\mu t\cos r}{n} = \frac{2 \times 1.333 \times 5 \times 10^{-5} \times 0.9}{n} \text{ cm}$$

Putting $n = 1$, $\lambda = 12.0 \times 10^{-5}$ cm outside visible range, in infrared region.

for $n = 2$, $\lambda = 6.0 \times 10^{-5}$ cm, which lies in the visible region.

for $n = 3$, $\lambda = 4.0 \times 10^{-5}$ cm, this also lies in the visible region.

for $n = 4$, $\lambda = 3.0 \times 10^{-5}$ cm, which lies in the ultraviolet region.

Thus, wavelength 6.0×10^{-5} cm and 4.0×10^{-5} cm will interfere destructively and will be absent in the reflected light.

Example 1.15

A thin layer of colourless oil is spread over water in a container ($\mu = 1.4$). If the light of wavelength 6400 Å is absent in the reflected light, what is the minimum thickness of the oil layer?

Solution:

Let t be the thickness of the oil layer spread on water.

$$\text{For minima, } 2\mu t\cos r = n\lambda$$

For minimum thickness, we take $n = 1$ and $\cos r = 1$.

$$\therefore t_{min} = \frac{\lambda}{2\mu} = \frac{6000 \text{ Å}}{2 \times 1.4} = 2143 \text{ Å}.$$

Example 1.16

Light of wavelength 5893 Å is reflected at nearly normal incidence from a soap film of refractive index = 1.42. What is the least thickness of the film that will appear (*a*) bright and (*b*) dark?

Solution:

(*a*) For brightness of the film in reflected light:

$$2\mu t\cos r = (2n-1)\frac{\lambda}{2}$$

For near normal incidence,

$$\cos r = 1$$

$$\therefore 2\mu t = (2n-1)\frac{\lambda}{2}$$

$$\text{or, } t = \frac{(2n-1)\lambda}{4\mu}$$

For minimum thickness, $n = 1$.

$$\therefore t_{min} = \frac{\lambda}{4\mu}$$

$$= \frac{5893\ \text{Å}}{4\times 1.42} = 1037.5\ \text{Å}$$

(*b*) For darkness of the film in reflected light,

$$2\mu t\cos r = n\lambda$$

For near normal incidence, $\cos r = 1$ and so minimum thickness ($n = 1$), is

$$t_{min} = \frac{\lambda}{2\mu}$$

$$= \frac{5893\ \text{Å}}{2\times 1.42} = 2075\ \text{Å}.$$

Example 1.17

A parallel film of sodium light (λ = 5893 Å) strikes a film of oil (μ = 1.46) floating on water. When viewed at an angle 30° from the normal, 8th dark band is seen. What is the thickness of the film?

Solution:

For darkness,

$$2\mu t\cos r = n\lambda$$

$$\text{or, } t = \frac{n\lambda}{2\mu\cos r}$$

$$\text{For Snell's law, } \mu = \frac{\sin i}{\sin r}$$

$$\text{or, } \sin r = \frac{\sin i}{\mu}$$

$$\therefore \cos r = \sqrt{1-\sin^2 r} = \sqrt{1-\frac{\sin^2 i}{\mu^2}}$$

Here, $\mu = 1.46$ and $i = 30°$

$$\therefore \cos r = \sqrt{1-\frac{\sin^2 30°}{(1.46)^2}}$$

$$= \sqrt{1-\frac{1}{4\times(1.46)^2}} = 0.94$$

$$\therefore t = \frac{8\times 5893\times 10^{-10}}{2\times 1.46\times 0.94} = \frac{8\times 5893\times 10^{-10}}{2.7448}$$

$$= 1.717\times 10^{-6}\ \text{m}$$

$$= 1.717\ \mu\text{m}.$$

Example 1.18

A non-reflecting coating of magnesium fluoride (MgF_2) of refractive index 1.38 is put on a glass of refractive index 1.52. What should be the minimum thickness of coating for a light of 5500 Å.

Solution:

Here, λ = 5500 Å, μ = 1.38

We know, for darkness or destructive interference (so as to work as non-reflecting coating and as two successive reflections from dense medium),

$$2\mu t\cos r = (2n-1)\frac{\lambda}{2}$$

For minimum thickness taking ($\cos r = 1$ and $n = 1$),

$$2\mu t_m = \frac{\lambda}{2}$$

$$\text{or, } t_m = \frac{\lambda}{4\mu} = \frac{5500\ \text{Å}}{4\times 1.38} = 996.34\ \text{Å} = 99.634\ \text{nm}.$$

Example 1.19

White light is incident on a soap film at an angle $\sin^{-1}(4/5)$ and the reflected light, on examination by a spectrometer shows dark bands. The consecutive dark bands correspond to wavelength 6.1 ×

10^{-5} cm and 6.0×10^{-5} cm. If $\mu = 1.333$ for the film, calculate the thickness of the film.

Solution:

Here, $i = \sin^{-1}\left(\frac{4}{5}\right)$ or $\sin i = \frac{4}{5}$, $\mu = 1.333$

$\lambda_1 = 6.1 \times 10^{-5}$ cm $\lambda_2 = 6.0 \times 10^{-5}$ cm

The condition for dark band: $2\mu t\cos r = n\lambda$

In case of two consecutive dark bands, corresponding to wavelength λ_1 and λ_2, we can write:

$$2\mu t\cos r = n\lambda_1 \quad ...(i)$$

$$\text{and, } 2\mu t\cos r = (n+1)\lambda_2$$

$$\therefore n\lambda_1 = (n+1)\lambda_2$$

$$\text{or, } n(\lambda_1 - \lambda_2) = \lambda_2$$

$$\text{or, } n = \frac{\lambda_2}{\lambda_1 - \lambda_2}$$

Putting the value of n in (i), we have:

$$2\mu t\cos r = \frac{\lambda_1\lambda_2}{\lambda_1 - \lambda_2}$$

$$\text{Also } \cos r = \sqrt{1-\sin^2 r} = \sqrt{1-\frac{\sin^2 i}{\mu^2}}$$

$$\left(\because \mu = \frac{\sin i}{\sin r} \text{ or } \sin r = \frac{\sin i}{\mu}\right)$$

$$= \sqrt{1-\frac{(4/5)^2}{(0.333)^2}} = \sqrt{1-\left(\frac{4}{5}\times\frac{3}{4}\right)^2} = \sqrt{1-\left(\frac{3}{5}\right)^2}$$

$$= \sqrt{1-\frac{9}{25}} = \sqrt{\frac{16}{25}} = \frac{4}{5}$$

Putting the value of $\cos r$, we have,

$$t = \frac{\lambda_1\lambda_2}{\lambda_1 - \lambda_2}\cdot\frac{1}{2\mu\cos r}$$

$$= \frac{6.1\times10^{-5}\times6.0\times10^{-5}}{(6.1-6.0)\times10^{-5}}\cdot\frac{5}{2\times1.333\times4}$$

$$= 0.00172 \text{ cm.}$$

Example 1.20

Two glass plates enclose a wedge-shaped air film, touching at one edge and separated by a wire of 0.05×10^{-3} m diameter at a distance of 0.15 m from the edge. Calculate the fringe width if light of wavelength 6000 Å from a broad source falls normally on the film.

Solution:

Here, $x = 0.15$ m, $\lambda = 6000$ Å $= 6.0 \times 10^{-7}$ m, $\mu = 1$

$$\text{Angle of the wedge, } \theta = \frac{0.05\times10^{-3}}{x} = \frac{0.05\times10^{-3}}{0.15} \text{ rad}$$

$$\therefore \text{Fringe width, } \beta = \frac{\lambda}{2\mu\theta} = \frac{6.0\times10^{-7}\times0.15}{2\times1\times0.05\times10^{-3}} = 9\times10^{-4} \text{ m.}$$

Example 1.21

Inference fringes are produced with monochromatic light falling normally on a wedge-shaped film containing a medium of refractive index 1.4. The angle of the wedge is 10 seconds of an arc and the distance between successive fringes is 0.5×10^{-2} m. What is the wavelength of the light used?

Solution:

$$\text{Here, } \theta = 10 \text{ second of an arc} = \frac{10\pi}{60\times60\times180} \text{ radian}$$

$\mu = 1.4$, $\beta = 0.5 \times 10^{-2}$ m

$$\text{We know } \beta = \frac{\lambda}{2\mu\theta}$$

$$\therefore \text{Wavelength } \lambda = 2\beta\mu\theta = \frac{2\times0.5\times10^{-2}\times1.4\times10\times3.14}{60\times60\times180}$$

$$= 6790 \times 10^{-10} \text{ m}$$

$$= 6790 \text{ Å.}$$

Example 1.22

Two pieces of plane glass are placed together with a piece of paper between the two at one edge, forming a wedge-shaped air film. If on viewing the film normally with a light of 6000 Å, ten fringes per cm are observed. What is the angle of the wedge?

Solution:

Let θ be the angle of the wedge and t the thickness of the air film at a distance x from the edge at which the two glasses are in contact.

So, for small angle:

$$t = x\theta$$

For normal incidence, the condition for darkness,

$$2t = n\lambda$$

$$\text{or, } 2x\theta = n\lambda$$

$$\text{or, } x = \frac{n\lambda}{2\theta}$$

Location of nth dark fringe,

$$x_n = \frac{n\lambda}{2\theta}$$

Location of $(n + 1)$th dark fringe,

$$x_{n+m} = \frac{(n+m)\lambda}{2\theta}$$

$$\therefore x_{n+m} - x_n = \frac{m\lambda}{2\theta} \Rightarrow \theta = \frac{m\lambda}{2(x_{n+m} - x_n)}$$

Given, $\lambda = 6000$ Å $= 6.0 \times 10^{-7}$ m.

$x_{n+m} - x_n = 1$ cm $= 10^{-2}$ m, $\mu = 10$

$$\theta = \frac{10 \times 6.0 \times 10^{-7}}{2 \times 10^{-2}} = 3 \times 10^{-4} \text{ radians}$$

$$= 3 \times 10^{-4} \times \frac{180}{\pi}$$

$$= 0.017^\circ$$

$$= 0.017 \times 60 \text{ minute} = 1.03 \text{ minute.}$$

1.13 Newton's rings

In this experiment a plano-convex lens is placed on a plane glass plate, with its convex surface on the plate such that an air film of gradually increasing thickness is formed. If monochromatic light is made to fall normally and viewed as shown in Fig. 1.12, alternate bright and dark circular fringes, in the form of rings, are observed. The interfering rays in this case are obtained from the reflection of the incident rays at the top and bottom surfaces of the air film between the glasses. This type of interference was first observed by Hooke in 1665. Newton studied the phenomenon carefully and measured the diameters of the rings and these rings are therefore known as Newton's rings. Later, Thomas Young interpreted the results and concluded that the rings are formed due to interference.

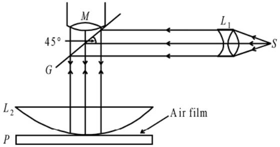


Fig. 1.12: Experimental set-up for Newton's rings.

Experimental arrangement for obtaining Newton's rings is shown in Fig. 1.12. Light from a monochromatic source, S, rendered parallel by a convex lens λ_1 and reflected by a glass plate G held at 45° to the incident rays is made to fall normally on a long focus plano-convex lens λ_2. Light transmitted through the lens λ_2 on reflection from the bottom surface of the air film interferes with the rays of light reflected from the upper surface of the film (Fig. 1.13). The rays reflected from the bottom surface is actually reflected by the glass plate (optically denser) and, therefore, suffer a phase change of π. The interfering rays produce circular fringes because of circular symmetry of the air film.

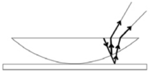


Fig. 1.13: Interference in Netwon's rings set-up.

The fringes obtained in this case are *localized fringes* and, therefore, the travelling microscope μ has to be focussed at the point of contact of plano-convex lens and glass plate. The fringes obtained in Young's double experiment, Fresnel's biprism, etc. *non-localized fringes*. This is because there is a wide region of space in which waves cross each other and so, it is not necessary to put the observing screen or focus the microscope at any particular location unlike for Newton's rings.

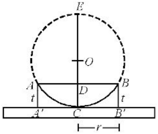


Fig. 1.14

Theory: Suppose R is the radius of curvature of the lens whose vertical section is ACB (Fig. 1.14).

The lens is in contact with plane glass plate at C in such a way that the film thickness at AA' and BB' is equal, i.e. $AA' = BB' = t$. The points A' and B' are equidistant from C and the thickness of the film is t at a distance r from C. From the geometry of the circle, we have:

$$ED \times DC = AD \times DB$$

$$\text{or, } (EC - DC) \times DC = AD \times DB$$

$$\text{or, } (2R - t) \times t = r^2$$

$$2Rt - t^2 = r^2$$

where r is the radius of the ring whose diameter passes through A' and B'.

Since the thickness of the air film is very small, t^2 can be neglected in comparison to $2Rt$. Therefore,

$$2Rt = r^2 = \left(\frac{A'B'}{2}\right)^2 = \left(\frac{d}{2}\right)^2 = \frac{d^2}{4}$$

$$2t = \frac{d^2}{4R} \quad ...(1.39)$$

where Δ is the diameter of the ring passing through A' and B'.

The path difference between the two rays, reflected from A and A' or B and B' is $2\mu t \cos r + \frac{\lambda}{2}$.

For normal incidence,

$$r = 0, \cos r = 1$$

$$\therefore \text{Path difference} = 2\mu t + \frac{\lambda}{2}$$

At the point of contact, $t = 0$, the path difference is $\frac{\lambda}{2}$. Hence, *the central spot is dark.*

For bright fringe,

$$2\mu t + \frac{\lambda}{2} = n\lambda$$

$$\text{or, } 2\mu t = (2n-1)\frac{\lambda}{2}, n = 1, 2, 3, \dots \quad ...(1.40)$$

In the case of air film, μ = 1, so that the condition for brightness.

$$2t = (2n-1)\frac{\lambda}{2} \quad ...(1.41)$$

For dark fringe:

$$2\mu t + \frac{\lambda}{2} = (2n-1)\frac{\lambda}{2}$$

$$\text{or, } 2\mu t = n\lambda, n = 1, 2, 3, \dots \quad ...(1.42)$$

In case of air film, μ = 1, so that the condition for darkness becomes

$$2t = n\lambda \quad ...(1.43)$$

The fringes obtained in this case are shown in Fig 1.15. Why the fringe is circular in this case can be understood on the basis of Eq. (1.42). For a monochromatic light of wavelength λ and given n, the locus of points showing the same path difference will have the same value of t which is possible for only a circle, i.e., a fringe must be circular in shape. In fact, this can be utilised to check whether a glass plate is plane or not. The fringes will be circular only if the glass plate is plane, if not, t will vary and so the fringes will not be exactly circular but irregular in shape.

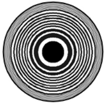


Fig. 1.15: Newton's rings in reflected light.

Putting the value of $2t$ from Eq. (1.39) in Eq. (1.40), we get for nth bright fringe:

$$\frac{d_n^2}{4R} = (2n-1)\frac{\lambda}{2}$$

$$\left(\text{for non-air film } \frac{d_n^2}{4R} = \frac{(2n-1)\lambda}{2\mu}\right)$$

$$\text{or, } \Delta_n = \sqrt{2(2n-1)\lambda R} \quad ...(1.44)$$

$$\text{or, } \Delta_n \propto \sqrt{(2n-1)}$$

where, $n = 1, 2, 3, \dots$ for first, second, third, ... rings respectively. Thus, the *diameter of the bright rings are proportional to the square root of odd natural numbers.*

Putting Eq. (1.38) in Eq. (1.42), we get for the dark rings.

$$\frac{d_n^2}{4R} = n\lambda \left(\text{for non-air film } \frac{d_n^2}{4R} = \frac{n\lambda}{\mu}\right)$$

$$\text{or, } \Delta_n = \sqrt{4n\lambda R} \quad ...(1.45)$$

$$\text{or, } \Delta_n = \sqrt{n}$$

Thus, *the diameters of the dark rings are proportional to the square of the natural numbers.*

While counting the order of the dark rings 1, 2, 3, ... the central ring, which is also dark, is not counted.

When viewed in white light, the central spot remains dark. It is followed by few coloured rings (in order or violet, indigo, ... red) and then due to overlapping of the rings of different colours, rings cannot be seen clearly.

Newton's Ring in Transmitted Light: Newton's rings can also be observed in transmitted light. The interference fringes are produced by the interference of transmitted rays, as shown in Fig. 1.16. The ray BC originates from AB by refraction at B, while ray $B'C'$ also originates from AB by two reflections, one at B and another at A' and then a refraction at B'. At each reflection, a phase change of π is introduced so that a net phase change of $\pi + \pi = 2\pi$ is introduced in $B'C'$ with respect to BC, which physically means no phase change. Thus, the only phase or path difference between the rays BC and $B'C'$ is on account of the difference in their optical path traversed, which for thickness t of the film is $2\mu t\cos r$. Therefore, for bright rings.

$$2\mu t\cos r = n\lambda \quad ...(1.46)$$

In case of air film, $\mu = 1$ and for $\cos r = 1$, the above condition becomes $2t = n\lambda$.

For dark rings.

$$2\mu t\cos r = (2n-1)\frac{\lambda}{2} \quad ...(1.47)$$

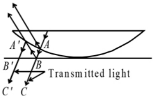


Fig. 1.16: Newton's ring formation by transmitted light.

In case of air film, $\mu = 1$, $\cos r = 1$, the above condition becomes

$$2t = (2n-1)\frac{\lambda}{2}$$

Proceeding in the same manner as in reflected system,

$$\text{for } n\text{th bright ring } d_n^2 = 4nR\lambda \Rightarrow d_n = \sqrt{n}$$

$$\text{For } n\text{th dark ring } d_n^2 = 2(2n-1)\lambda R \Rightarrow d_n \propto \sqrt{(2n-1)}$$

Thus, the reflected and transmitted systems are *complementry* to each other. In other words, in the transmitted system the central spot is bright or the dark rings in the reflected system become bright rings in the transmitted system and vice versa (Fig. 1.17). The contrast between bright and dark rings is poor in the transmitted system compared to that in the reflected system.

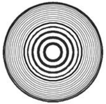


Fig. 1.17: Newton's rings in transmitted light.

Production of Rings with Bright Centre in Reflected Light: We have seen that the rings formed by reflected light have a dark centre when there is an air film between the glass surfaces. However, it is possible to have rings with bright centre in the reflected system with an arrangement shown in Fig. 1.18.

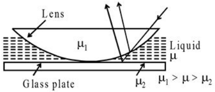


Fig. 1.18: Newton's rings with bright centre in reflected light.

For this, it is required that instead of air film, the space between the glass surfaces is filled with a transparent liquid of refractive index μ such that $\mu_1 > \mu > \mu_2$, where the refractive index of lens material is μ_1 and that of glass plate is μ_2. This is possible if a few drops of sassafras oil are put between a convex lens of crown glass and a glass plate of flint glass. If $\mu_1 > \mu > \mu_2$, then both the reflections at the top and bottom of the film are from denser to rarer medium and no additional path difference of $\lambda/2$ occurs in one of the reflected lights as in the case of air film, Therefore, the central spot ($t = 0$), is bright in the reflected system.

Applications: Newton's ring set-up is widely used to determine the wavelength of the monochromatic light used and the refractive index of the medium enclosed between the lens and the glass plate. The procedure for these is discussed below. In addition, it can also be used to check the flatness of glass plate used. Any deviation in the shape of rings formed by the glass plate and the lens indicates that either the glass plate is not perfectly flat or the lens surface is not smooth.

(i) Determination of λ: The set-up is illuminated by the monochromatic light whose wavelength is to be determined and the rings are obtained. Using the travelling microscope, measure the diameters of nth and $(n + \mu)$th dark rings by placing the cross-wire at the centre of respective rings.

The diameter Δ_n of nth dark ring,

$$\frac{d_n^2}{4R} = \frac{n\lambda}{\mu} \quad ...(1.48)$$

The diameter Δ_{n+m} of $(n + \mu)$th dark ring,

$$\frac{d_{n+m}^2}{4R} = \frac{(n+m)}{\mu}\lambda \quad ...(1.49)$$

Subtracting, we get:

$$\frac{d_{n+m}^2 - d_n^2}{4R} = \frac{m\lambda}{\mu}$$

$$\text{or, } \lambda = \frac{\mu(d_{n+m}^2 - d_n^2)}{4mR} \quad ...(1.50)$$

For air film $\mu = 1$,

$$\lambda = \frac{d_{n+m}^2 - d_n^2}{4mR} \quad ...(1.51)$$

Thus, knowing μ and measuring the radius of curvature R and the two diameters, wavelength λ can be determined. Students have often asked me in the class why not use, Eq. (1.48) only and find λ? Can the reader students answer this question? It is related to the accuracy of the results.

***(ii)* Determination of refractive index μ of a liquid:** For this purpose, Newton's rings are obtained with air film and the diameter of nth and $(n + \mu)$th rings are determined, so that:

$$d_{n+m}^2 - d_n^2 = 4\mu\lambda R \quad ...(1.52)$$

Now the liquid whose refractive index μ is to be determined is poured (few drops) onto the glass plate and then the lens is placed gently without disturbing the other arrangement such that the film encloses is of the liquid. Again, the diameter of nth and $(n + \mu)$th dark rings are determined so that:

$$d'_{n+m} - d_n'^2 = \frac{4m\lambda R}{\mu} \quad ...(1.53)$$

Dividing Eq. (1.52) by Eq. (1.53), we have:

$$\mu = \frac{d_{n+m}^2 - d_n^2}{d_{n+m}'^2 - d_n'^2} \quad ...(1.54)$$

When λ is known, the first part of the procedure, i.e., determination of diameters with air film can be avoided and the refractive index can be calculated using Eq. (1.53).

Also, with liquid film, the diameter of nth dark ring is:

$$d_n'^2 = \frac{4n\lambda R}{\mu}$$

that with air, $d_n^2 = 4n\lambda R$

$$\therefore \frac{d_n'^2}{d_n^2} = \frac{1}{\mu}$$

The above expression indicates that the diameter of a particular ring is smaller with the liquid compared to that in air.

1.14 Newton's rings by two curved surfaces

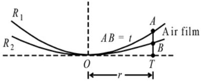


Fig. 1.19

Consider two curved surfaces of radii of curvature R_1 and R_2 in contact with each other at the point O (Fig. 1.19). A thin film of air is enclosed between the two surfaces and when illuminated by a monochromatic source of light of wavelength λ, dark and bright rings are formed which can be viewed with travelling microscope. Suppose the radius of the nth ring is r. The thickness of the air film at A is:

$$AB = AT - BT$$

$$\text{From geometry, } AT = \frac{r^2}{2R_1}$$

$$\text{and, } BT = \frac{r^2}{2R_2}$$

$$\therefore AB = \frac{r^2}{2R_1} - \frac{r^2}{2R_2}$$

As AB is thickness t,

$$\therefore t = \frac{r^2}{2R_1} - \frac{r^2}{2R_2} \quad ...(1.55)$$

We know, the path difference for the rays reflected at top and bottom surface of the film is $2\mu t \cos r + \frac{\lambda}{2}$.

From dark rings,

$$2\mu t \cos r = n\lambda$$

For air, $\mu = 1$ and for normal incidence $\cos r = 1$.

$$\therefore 2t = n\lambda$$

or, $$2\left(\frac{r_n^2}{2R_1} - \frac{r_n^2}{2R_2}\right) = n\lambda$$

or, $$r_n^2\left(\frac{1}{R_1} - \frac{1}{R_2}\right) = n\lambda,\ n = 0, 1, 2, 3, \ldots \quad \ldots(1.56)$$

For bright rings: $$2t = (2n-1)\frac{\lambda}{2}$$

or, $$2\left(\frac{r_n^2}{2R_1} - \frac{r_n^2}{2R_2}\right) = (2n-1)\frac{\lambda}{2}$$

or, $$r_n^2\left(\frac{1}{R_1} - \frac{1}{R_2}\right) = (2n-1)\frac{\lambda}{2},\ n = 1, 2, 3, \ldots \quad \ldots(1.57)$$

Special Case: Consider the case when the two surfaces are placed in contact as shown in Fig. 1.20, i.e., convex side oppositely faced.

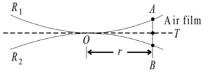


Fig. 1.20

Here, thickness of the film is:

$$AB = AT + BT$$

or, $$t = \frac{r^2}{2R_1} + \frac{r^2}{2R_2}$$

For nth dark ring in air:

$$2t = n\lambda$$

or, $$2\left(\frac{r_n^2}{2R_1} + \frac{r_n^2}{2R_2}\right) = n\lambda$$

or, $$r_n^2\left(\frac{1}{R_1} + \frac{1}{R_2}\right) = n\lambda,\ n = 0, 1, 2, 3, \ldots \quad \ldots(1.58)$$

For nth bright ring in air:

$$2t = (2n-1)\frac{\lambda}{2}$$

or, $$2\left(\frac{r_n^2}{2R_1} + \frac{r_n^2}{2R_2}\right) = (2n-1)\frac{\lambda}{2}$$

or, $$r_n^2\left(\frac{1}{R_1} + \frac{1}{R_2}\right) = (2n-1)\frac{\lambda}{2},\ n = 1, 2, 3, \ldots \quad \ldots(1.59)$$

Example 1.23

As shown in Fig. 1.21, two 10 cm long glass slides are in contact at one end and separated by a piece of paper of 0.02 mm thickness at the other end. It is illuminated by light of wavelength 5000Å (a) Find the separation between the interference fringes in reflected light. (b) Is the fringe at the line of contact bright or dark? (c) If white light is used to illuminate it, for which colour: red or yellow, the fringes will be farther apart? (Assume normal incidence of light.)

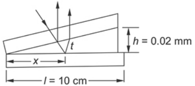


Fig. 1.21

Solution:

(a) We know, for destructive interference in a wedge shaped air film

$$2t = nl\ (n = 0, 1, 2, 3, \ldots)$$

It is obvious from Fig. 1.21, for the similar triangles, the thickness of the air wedge at each point is proportional to the distance x from the line of contact

$$\frac{t}{x} = \frac{h}{l}$$

Combining this with Eq. (1.37), we have

$$\frac{2xh}{l} = nl \quad \left(\mu = 1 \text{ for air and } \theta \approx \frac{h}{l}\right)$$

$$\therefore x = n\frac{l\lambda}{2h} = n\frac{0.1\,\text{m} \times 5.0 \times 10^{-7}\,\text{m}}{2 \times 0.02 \times 10^{-3}\,\text{m}} = n(1.25\,\text{mm})$$

Thus, the successive dark fringes corresponding to $n = 1, 2, 3, \ldots$ will be 1.25 mm apart.
(b) Substituting $n = 0$, in the above equation gives $x = 0$, which is the location of the line of contact. Hence, there will be dark fringe at the line of contact.
(c) It is evident from the equation that $x \mu l$, i.e. fringe spacing is proportional to the wavelength of the light used. Thus, fringes of red colour will be farther apart.

Example 1.24

Suppose in the above example the space between the glass slides of refractive index 1.5 is filled with an oil of refractive index 1.4. How the answers of (a) and (b) will be modified.

Solution

(a) In the oil film, the wavelength of light reduces to l_l = $\frac{\lambda_a}{1.4} = \frac{5000Å}{1.4}$ = 3571.43Å. Putting this value, we see that the fringe spacing reduces by a factor 1.4 and is equal to 0.89286 mm (or n(0.89286 mm)).
(b) At the line of contact the fringe will continue to be dark.

Example 1.25

In a Newton's ring experiment, the diameter of the 10th ring changes from 1.40 to 1.27 cm when a liquid is introduced between the lens and the glass plate. What is the refractive index of the liquid?

Solution:

We know, $$D_n^2 = \frac{4n\lambda R}{\mu} \quad \ldots(i)$$

Without liquid in air $\mu = 1$, $$D_{10}^2 = 4 \times 10 \times \lambda R$$

With liquid: $$D_{10}'^2 = \frac{4 \times 10 \times \lambda R}{\mu} \quad \ldots(ii)$$

Dividing (*i*) by (*ii*), we get:

$$\mu = \frac{D_{10}^2}{D_{10}'^2} = \frac{(1.40)^2}{(1.27)^2} = 1.2515.$$

Example 1.26

In a Newton's ring experiment, the diameter of the 5th dark ring was 0.3 cm and the diameter of the 25th ring was 0.8 cm. If the radius of curvature of the plano-convex lens is 100 cm, find the wavelength of the light used.

Solution:

Here, $\Delta_5 = 0.3$ cm, $\Delta_{25} = 0.8$ cm.

$\mu = (n + \mu) - n = 25 - 5 = 20$, $R = 100$ cm

We know, $$\lambda = \frac{D_{n+m}^2 - D_n^2}{4mR}$$

$$= \frac{(0.8)^2 - (0.3)^2}{4 \times 20 \times 100} = \frac{0.64 - 0.09}{8000} = 6.875 \times 10^{-5} \text{ cm}$$

Example 1.27

Newton's rings are formed by reflection in the air film between plane surface and spherical surface of radius 50 cm and it is noticed that the centre of the ring system is bright. If the diameter of the third bright ring is 0.181 cm and the diameter of the 23rd bright ring is 0.501 cm, what is the wavelength of the light used? What conclusion can be drawn from the fact that the centre of ring system is bright?

Solution:

We know, $$\lambda = \frac{D_{n+m}^2 - D_n^2}{4mR}$$

Here, $\Delta_3 = \Delta_n = 0.181$ cm, $\Delta_{23} = \Delta_{n+m} = 0.501$ cm

$R = 50$ cm, $\mu = 20$

∴ Wavelength $$\lambda = \frac{(0.501)^2 - (0.181)^2}{4 \times 20 \times 50} = \frac{0.251 - 0.033}{4000}$$

$= 5.450 \times 10^{-5}$ cm

= 5450 Å

Normally, in the reflected system, the centre of the ring system is dark. In this case, the centre is bright which implies that an additional path difference of $\lambda/2$ is introduced between the interfering rays due to the two glass surfaces not being exactly in contact. For normal incidence, the condition for brightness with air film ($\mu = 1$) is:

$$2t = (2n - 1)\frac{\lambda}{2}$$

For centre, $n = 1$

$$\therefore 2t = \frac{\lambda}{2}$$

$$\text{or, } t = \frac{\lambda}{4}$$

Thus, an air film of thickness equal to $\lambda/4$ exists at the centre between the two glass surfaces.

Example 1.28

Newton's rings are formed between a plane glass surface of a plano-convex lens. The diameter of the third dark ring is 10^{-2} m. A light of wavelength 5890 Å is used at such an angle that the light passes through the air film at an angle of 30° to the normal. Find the radius of curvature of the lens.

Solution:

Here, $\lambda = 5890$ Å $= 5890 \times 10^{-10}$ m,

$\Delta_3 = 10^{-2}$ m, $n = 3$, $r = 30°$, $\mu = 1$

We know, for the nth dark ring.

$$2\mu t\cos r = n\lambda \text{ and } \frac{d_n^2}{4R} = 2t$$

$$\text{or, } \frac{\mu d_n^2}{4R}\cos r = n\lambda \Rightarrow R = \frac{\mu d_n^2 \cos r}{4n\lambda}$$

$$\therefore R = \frac{1 \times (10^{-2})^2 \times \cos 30°}{4 \times 3 \times 5890 \times 10^{-10}} = 12.23 \text{ m.}$$

Example 1.29

A plano-convex lens of radius 3 m is placed on an optically flat glass plate and is illuminated by monochromatic light. The diameter of the 8th bright ring in the reflected system is 0.72×10^{-2} m. Calculate the wavelength of the used light.

Solution:

For the nth bright ring in reflected system,

$$\frac{d_n^2}{4R\mu} = (2n-1)\frac{\lambda}{2}$$

$$\text{or, } \lambda = \frac{d_n^2}{2R\mu(2n-1)}$$

Here, $R = 3$ m, $\Delta_8 = 0.72 \times 10^{-2}$ m, $\mu = 1$, $n = 8$

$$\therefore \lambda = \frac{(0.72 \times 10^{-2})^2}{2 \times 3 \times 1 \times 15} = 5760 \times 10^{-10} = 5760 \text{ Å.}$$

Example 1.30

A Newton's rings set-up (assume perfect point of contact) is viewed with a light of wavelength 600 nm. Now the plano-convex lens (of radius of curvature 1 m) is raised very slowly vertically above the plate. Discuss the variation in fringe pattern.

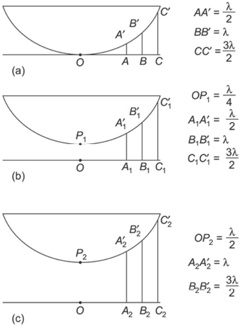


Fig. 1.22: As the lens moves up rings collapse to the centre.

Solution:

As the set-up has a perfect point of contact, the thickness of the film t at the point of contact is zero and hence the central spot will be dark with its boundary at $A¢$ where $AA¢ = \frac{\lambda}{2}$. As the radii of the dark rings is given by $\sqrt{n\lambda R}$, the radius of first dark ring $OA = \sqrt{\lambda R} = \sqrt{6 \times 10^{-9} \text{ m} \times 1 \text{m}}$ = 0.0774 cm and that of second ring $OB = \sqrt{2\lambda R}$ = 0.1094 cm.

Now as the plano-convex lens is raised and for the position such that $t = \frac{\lambda}{4}$ (= 1.5×10^{-7} cm) then the path difference (= $2t$, for air film m = 1) between the interfering rays corresponding to the central spot will be $\frac{\lambda}{2}$ and the central spot will be bright as the effective path difference is l

$\left(=\frac{\lambda}{2}+\frac{\lambda}{2}\right)$ and the constructive interference will take place.

As the plano-convex lens is raised further by $\frac{\lambda}{4}$ the central spot will be dark as the first dark ring will collapse to the centre. The ring which was initially at $B¢$ now shifts to $A¢_2$ and the ring initially at $C¢$ shifts to $B¢_2$. Thus, as the lens moves up the rings collapse to the centre and the central spot alternatively changes from dark to bright.

Example 1.31

A Newton ring arrangement is used with a sodium light source having two spectral lines Δ_1 having wavelength 5890 Å and Δ_2 of wavelength 5896 Å. If the *n*th dark ring due to Δ_1 line coincide with (*n* + 1)th dark ring due to the Δ_2 line, find the diameter of the *n*th dark ring. The radius of curvature of the plano-convex lens is 90 cm.

Solution:

Let l_1 = 5896 Å = 5896 3 10^{-10} m = 5896 3 10^{-8} cm

and l_2 = 5896 Å = 5890 3 10^{-10} m = 5890 3 10^{-8} cm

and R = 90 cm

Given $(D_n)_{\lambda_1} = (D_{n+1})_{\lambda_2}$

\ $\sqrt{4n\lambda_1 R} = \sqrt{4(n+1)\lambda_2 R}$ ⇒ $nl_1 = (n+1)\, l_2$

or n 3 5896 × 10^{-8} cm = (n + 1) 5890 3 10^{-8} cm

$$\frac{n+1}{n} = \frac{5896}{5890} = 1.00102$$

or $1 + \frac{1}{n} = 1.00102 \Rightarrow \frac{1}{n} = 1.00102 - 1 = 0.00102$

or $n = \frac{1}{0.00102} \simeq 980$

\ Δ_n =

$$\sqrt{4n\lambda_1 R} = \sqrt{4\times 980\times 5896\times 10^{-10}\,\text{m}\times 0.90\,\text{m}}$$

= 0.0456 m = 4.56 cm

Example 1.32

Newton's rings by reflection are formed between two biconvex lenses having equal radii of curvatures of 100 cm each. Calculate the distance between 5th and 15th dark rings if a light of wavelength 5400 Å illuminates the system.

Solution:

The radius of the *n*th dark ring formed between the biconvex lens is given by:

$$r_n^2 = \frac{n\lambda}{\left(\frac{1}{R_1}+\frac{1}{R_2}\right)}$$

Given, $R_1 = R_2$ = 100 cm, λ = 5400 Å = 5.4 × 10^{-5} cm

$r_{15} - r_5$ = ?

For the 5th ring, r_5 =

$$\sqrt{\frac{5\times 5.4\times 10^{-5}}{\frac{1}{100}+\frac{1}{100}}} = \sqrt{\frac{5\times 5.4\times 10^{-5}\times 100}{2}}$$

= 0.1161 cm

For the 15th ring, r_{15} =

$$\sqrt{\frac{15\times 5.4\times 10^{-5}}{\frac{1}{100}+\frac{1}{100}}} = \sqrt{\frac{15\times 5.4\times 10^{-5}\times 100}{2}}$$

= 0.20 cm

∴ Distance between 5th and 15th dark ring

= $r_{15} - r_5$ = 0.20 − 0.1161 = 0.0839 cm.

Example 1.33

A convex surface of radius 3 m of a plano-convex lens rests on a concave spherical surface of radius 4 m and Newton's rings are viewed with reflected light of wavelength 6000 Å. Find the diameter of the 10th bright ring.

Solution:

Here, R_1 = 3 m, R_2 = 4 m, λ = 6000 Å = 6.0 × 10^{-7} m

We know, $r_n^2\left(\frac{1}{R_1}-\frac{1}{R_2}\right) = (2n-1)\frac{\lambda}{2}$

For 10th ring: $r_{10}^2\left(\frac{1}{3}-\frac{1}{4}\right) = 19\times\frac{6.0\times 10^{-7}}{2}$

r_{10} = $\sqrt{19\times 3\times 10^{-7}\times 12} = 8.27\times 10^{-3}$ m

Diameter of 10 ring $\Delta_{10} = 2r_{10}$ = 2 × 8.27 × 10^{-3} m

= 1.654 × 10^{-2} m

= 16.54 × 10^{-3} m = 16.54 mm.

1.15 fabry-perot interferometer

Fabry-Perot interferometer consists of two glass plates, *A* and *B*, separated by a distance Δ. The inner surfaces of the plates are optically plane, exactly parallel and thinly silvered so that about 70% of incident light gets reflected. The outer faces of the plates are also parallel to each other but inclined to their respective inner faces. A schematic diagram of the interferometer is shown in Fig. 1.23.

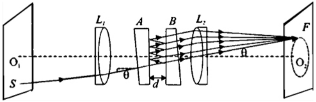


Fig. 1.23: Fabry-Perot interferometer showing formation of circular fringes.

Monochromatic light of wavelength l from a broad source *S* is incident on a collimating lens λ_1 which renders the incident rays parallel. A ray of light entering the air film between the plates undergoes multiple reflection between the silvered surfaces and emerges from the plate *B* as a parallel beam and they are made to interfere by converging them at φ, the focal plane of the lens λ_2. The *fringes are concentric circles* with O_2 as centre, having equal inclination. The fringes of constant inclination are called *Haidinger fringes*.

If q is the angle of incidence on the silvered face of *A*, then the path difference between successive rays is 2mΔ cos q or 2Δ cos q as m = 1 for air. Therefore, for a bright fringe:

2Δ cos q = *n*l, *n* = 0, 1, 2, 3, ...(1.60)

for all points on a circle passing through φ with centre at O_2 on the axis O_1O_2.

In practice, one of the plates is kept fixed and the other is movable with a rack and pinion arrangement, so that the film thickness Δ can be changed. The fringes can be obtained up to 10 cm of plate separation. Due to large separation, fringes are observed almost near the normal direction. If the distance Δ between the two plates is decreased, the value of q decreases for a given value of *n* and l. This means when Δ is decreased, the rings shrink and disappear at the centre. When Δ is decreased by l/2, one ring disappears at the centre.

In this interferometer, the radius of a ring is a function of the wavelength of light used and very sharp bright fringes are produced (better than Michelson interferometer), so this interferometer is *used for revolving two wavelengths and analysis of a single spectral line.*

Theory of intensity distribution: Consider a plane wave of unit amplitude incident at an angle q on the glass plate *AB*, as shown in Fig. 1.24. Although, from an extended source, light is incident on the plate at all possible angles. Due to multiple reflections, a set of parallel reflected rays A_1R_1, A_2R_2, A_3R_3 .... and a set of parallel transmitted rays B_1T_1, B_2T_2, B_3T_3, ... are produced. Let !*R* and !*T* be the reflection and transmission coefficients of amplitudes from the surfaces respectively. Thus, the amplitude of A_1B_1 is !*T* as the amplitude of incident ray SA_1 is unity. The amplitude of B_1T_1 is *T* (= !*T* 3 !*T*). The amplitude of B_1A_2 is !*RT* and that of A_2B_2 is *R*!*T* (= !*R* 3 !*RT*). The amplitude of B_2T_2 is *RT* (= !*T*3 *R*!*T*). In this way, we can see that the amplitudes of B_1T_1, B_2T_2, B_3T_3, B_4T_4,...... are *T*, *RT*, R^2T, R^3T,.... respectively. As these rays are obtained from the same incident ray, they are coherent and produce sustained interference pattern.

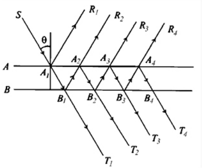


Fig. 1.24

Neglecting the small phase change due to reflection from silvered surfaces, the path difference between the two consecutive transmitted rays is 2Δ cos q (m = 1 for air), the corresponding phase difference:

$$\delta = \frac{2\pi}{\lambda} \times (2d\cos\theta) = \frac{4\pi d\cos\theta}{\lambda}$$

Let the incident wave be represented by:

y_i = *a* sin wt = sin wt (*a* is assumed to be unity).

As discussed above, the transmitted rays can be represented as:

y_1 = *T* sin wt

y_2 = *RT* sin (wt – d)

y_3 = R^2T sin (wt – 2d)

y_4 = R^3T sin (wt – 3d), and so on.

Let *A* be the resultant amplitude of these rays (upon interfering and f the phase difference, so that it can be represented as:

y = *A* sin (wt – f)

Also, from the superposition principle, we have:

y = y_1 + y_2 + y_3 + ...

\ *A* sin (wt – f) = *T* sin wt + *RT* sin (wt – d) + R^2T sin (wt – 2d) + ...

or, *A* sin wt cos f – *A* cos wt sin f

= *T* sin wt + *RT* sin wt cos d – *RT* cos wt sin d

+ R^2T sin wt cos 2d – R^2T cos wt sin 2d + ...

Equating the coefficients of sin wt and cos wt on both sides, we get:

A cos f = *T* + *RT* cos d + R^2T cos 2d + ...(1.61)

and, *A* sin f = *RT* sin d + R^2T sin 2d + ...(1.62)

The resultant intensity *I* is given by:

I = A^2 = (*A* cos f + *i A* sin f) 3 (*A* cos f – *i A* sin f) (*i* = !1)

From Eqs. (1.61) and (1.62), we have:

A cos f + *iA* sin f = *T* + *RT* (cos d + *i* sin d) + R^2T (cos 2d + *i* sin 2d) + ...

which in terms of exponential can be put as:
A cos f + iA sin f = $T + RTe^{id} + R^2Te^{2id} + ...$
= $T\,[1 + Re^{id} + R^2\varepsilon^{2id} + ...]$

$$= \frac{T}{1 - Re^{i\delta}}$$

$$A\cos\phi - iA\sin\phi = \frac{T}{1 - \mathrm{Re}^{-i\delta}}$$

Similarly,

The resultant intensity is therefore given by:

$$I = A^2 = \frac{T}{1 - \mathrm{Re}^{i\delta}} \times \frac{T}{1 - \mathrm{Re}^{-i\delta}}$$

$$= \frac{T^2}{1 + \mathrm{R}^2 - 2R\cos\delta}$$

(□ ε^{id} = cos d + i sin d, ε^{-id} = cos d – i sin d)

$$= \frac{T^2}{(1-R)^2 + 2R - 2R\cos\delta}$$

$$= \frac{T^2}{(1-R)^2 + 2R(1-\cos\delta)}$$

$$= \frac{T^2}{(1-R)^2 + 4R\sin^2\delta/2}$$

$$= \frac{T^2}{(1-R)^2\left[1 + \frac{4R}{(1-R)^2}\sin^2\frac{\delta}{2}\right]} \quad ...(1.63)$$

This expression, known as Airy's formula, shows that the resultant intensity depends upon the properties of the silver coating and d.

Condition for maxima: For maximum intensity of the fringe:

$$\sin^2 = \frac{\delta}{2} = 0 \Rightarrow \frac{\delta}{2} = n\pi$$

or, d = 2np, n = 0, 1, 2, 3,....

Maximum intensity

$$I_{\max} = \frac{T^2}{(1-R)^2} \quad ...(1.64)$$

Condition for minima: The intensity is minimum when

$$\sin^2\frac{\delta}{2} = 1$$

or, $$\frac{\delta}{2} = (2n+1)\frac{\pi}{2} \Rightarrow \delta = (2n+1)\pi,\ n = 0, 1, 2, 3$$

$$I_{\min} = \frac{T^2}{(1-R)^2\left[1 + \frac{4R}{(1-R)^2}\right]}$$

Minimum intensity,

$$= \frac{T^2}{(1+R)^2} \quad ...(1.65)$$

When there is no absorption at the reflecting surface, $T = 1 - R$,

$$I_{\max} = 1 \text{ and } I_{\min} = \frac{(1-R)^2}{(1+R)^2}$$

\

$$\frac{I_{\max}}{I_{\min}} = \frac{(1-R)}{(1+R)} \quad ...(1.66)$$

\

Also, from Eqs. (1.63) and (1.64), we have:

$$\frac{I}{I_{\max}} = \frac{1}{1 + \frac{4R}{(1-R)^2\ \sin^2\frac{\delta}{2}}}$$

$$= \frac{1}{[1 + \frac{4R}{(1-R)^2\ \sin^2\frac{\delta}{2}}]} \quad ...(1.67)$$

$$F = \frac{4R}{(1-R)^2}$$

where,

Circular shape of fringe: In terms of path difference,

For maxima: 2Δ cos q = nl

$$2d\cos\theta = (2n-1)\frac{\lambda}{2}$$

For minima:

For constant Δ, as in this interferometer, for a particular n (order of fringe) and l, q is a constant. The locus of points having the same value of q is a circle and hence *circular fringes are obtained*. The bright fringe in the interference pattern is very sharp as the intensity falls rapidly almost to zero on either side. The sharpness increases with reflecting power of the silver coating.

Visibility of fringes: The visibility of the fringe is defined as:

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \quad ...(1.68)$$

Substituting the values of $I_{\max}$ and $I_{\min}$ from Eqs. (1.64) and (1.65), we get

$$V = \frac{2R}{1+R^2} \quad ...(1.69)$$

Thus, the visibility of the fringes depends only on the reflection coefficient of the silver coating and is independent of its transmission coefficient.

Sharpness of fringes–Half fringe width: Figure 1.25 shows the variation of intensity I of a fringe with phase difference d for a given value of R. The *half fringe width of a fringe is defined as the width of fringe (in terms of phase difference) between the two points on either side of maxima where the intensity is half of its maximum value*, as shown in Fig. 1.25. From Eq. (1.67).

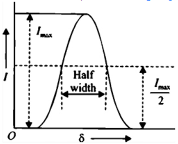


Fig. 1.25

$$\frac{I}{I_{\max}} = \frac{1}{1+F\sin^2\dfrac{\delta}{2}}$$

At half fringe width,

$$\frac{I}{I_{\max}} = \frac{1}{2}$$

$$\frac{1}{1+F\sin^2\dfrac{\delta}{2}} = \frac{1}{2}$$

\

$$1+F\sin^2\frac{\delta}{2} = 2$$

or,

$$F\sin^2\frac{\delta}{2} = 1$$

or,

$$\sin\frac{\delta}{2} = \frac{1}{\sqrt{F}} = \frac{1}{\sqrt{\left[\dfrac{4R}{(1-R)^2}\right]}} \frac{1-R}{2\sqrt{R}}$$

$$\delta = 2\sin^{-1}\left(\frac{1-R}{2\sqrt{R}}\right)$$

or,

For small value of d,

$$\sin\frac{\delta}{2} \approx \frac{\delta}{2}$$

$$\delta = \frac{1-R}{\sqrt{R}}$$

\

The sharpness of a bright fringe depends upon its half fringe width. Smaller is the half fringe width; sharper is the bright fringe. Fig. 1.26 shows the variation of $I/I_{\max}$ with phase difference d for R = 0.25, 0.5, 0.75 and 0.95. It is obvious from the figure that higher the value of R, sharper is the maxima, as half fringe width decreases as R increases. In this interferometer maxima are much sharper than those in Michelson's interferometer.

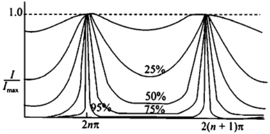


Fig. 1.26

Determination of l using F.P. Interferometer: The interferometer is adjusted so that circular fringes are obtained. Let n be the order of fringe at the centre (in this interferometer, order decreases from the centre), so that:

$2\Delta = n$l (\ q is zero at the centre)

Now, the movable plate is moved from a known distance, say from x_1 to x_2 and let N be the number of fringes that disappear at the centre. Thus,

$$N\frac{\lambda}{2} = x_2 - x_1$$

or,
$$\lambda = \frac{2(x_2 - x_1)}{N} \quad ...(1.70)$$

Thus, the wavelength l can be determined.

1.16 interference filter

The visible spectrum is spread over 3700 Å (3800 Å to 7500 Å) and there are occasions, when we need only a small portion of this spectrum. If the spectrum required is of bandwidth 500 Å, then coloured glasses or absorbing dyes can be used to filter out this spectrum from the visible spectrum. However, when the required bandwidth Dl is 100 Å or less (centred around a chosen λ), the coloured glasses and absorbing dyes cannot serve the purpose. Then we need to use an interference filter, a device developed by Walter Geffeken in 1939.

An interference filter, shown in Fig. 1.27, *A* and *B* are the two thin glass plates, on their inner surfaces, thin film of a reflecting silver film is deposited. In between the glass plates, a thin layer of dielectric material, like magnesium fluoride (MgF_2) or quartz having thickness of the order of wavelength of light is placed.

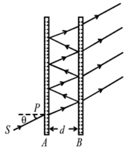


Fig. 1.27: An interference filter.

When a ray of light *SP* is incident, due to multiple internal reflections, a number of transmitted parallel rays are produced. The path difference between each successive pair of rays is 2μΔcosθ, where μ is refractive index of the sandwiched dielectric medium, Δ its thickness. For near normal incidence, cosθ = 1, the path difference is 2μΔ. The emergent rays are all in phase when 2μ*d* is equal to one wavelength λ or its integral multiple, i.e.,

$$2\mu\Delta = n\lambda,\ n = 1, 2, 3, \ldots \quad ...(1.71)$$

Therefore, the wavelength λ will get reinforced, i.e., constructive interference will take place.

When white light is incident, constructive interference will take place for a particular wavelength which satisfies Eq. (1.71). Therefore, in the transmitted beam, predominantly only a particular wavelength will be present. Thus, it is possible to filter only a particular wavelength from the incident white light.

If Δ = 5000 Å and supposing μ = 1, then:

$$\lambda = \frac{2 \times 1 \times 5000}{n}$$

or 10000, 5000, 3333 Å, ... as n = 1, 2, 3, ...

Clearly, only wavelength of 5000 Å lies in the visible spectrum and so if white light is incident, only the light of 5000 Å will be present in the transmitted beam. By making suitable choice of Δ and μ, any value of λ in the visible spectrum can be obtained.

Interference filters are used in spectroscopic work for studying the spectra is a narrow range of wavelengths.

1.17 michelson interferometer

This interferometer is based on the division of amplitude. Light from an extended source is divided into two parts of equal intensities by partial reflection and transmission. These beams are then sent in two directions, perpendicular to each other and are again brought together after reflection from plane mirrors to undergo interference and form circular bright and dark fringes. These fringes are observed and examined with the help of a telescope. This interferometer is used for standardisation of metre, resolution of spectral line, determination of thickness and refractive index of thin transparent sheets, determination of wavelength, etc. One important difference between this interferometer compared to Fresnel's biprism and Newton's rings is the production of interference between the rays differing in path by many wavelengths in this case, as compared to only a few wavelengths in the others.

Description of Apparatus: Michelson interferometer is shown in Fig. 1.28. It consists of two optically plane highly polished plane mirrors μ_1 and μ_2 held perpendicular to each other. There are two plane glass plates G_1 and G_2 of exactly the *same thickness* and of the same material placed parallel to each other. The plate G_1 is *half silvered at the back so that the incident beam is divided into a reflected and a transmitted beam of equal intensity*. The plates are inclined at an angle of 45° with the mirrors μ_1 and μ_2. The mirror μ_1 is mounted on a carriage capable of backward and forward movement and the motion is controlled by a micrometer up to ~10^{-5} cm (of the order of the wavelength of light). Both the mirrors are provided with three levelling screws at their back, so that they can be titled about horizontal and vertical axes and made exactly perpendicular to each other. Telescope *T* receives the reflected light from mirrors μ_1 and μ_2 and the fringes are observed through it. The lens λ makes the source of light an extended one.

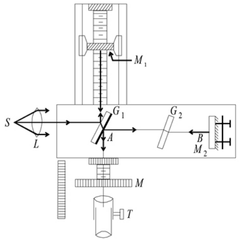


Fig. 1.28: Michelson interferometer.

Working: Light from a monochromatic source S after rendered parallel by a collimating lens λ fall on the glass plate G_1. It is partly reflected at the silvered face of G_1 and partly transmitted. The reflected ray travels normally towards mirror μ_1 while the transmitted ray, travels normally towards mirror μ_2. Both the rays are reflected back by the mirrors and retrace back their path. These rays again meet at the silvered face of the plate G_1 and enter a short focus telescope T. The two rays entering the telescope are originally derived from the same source, i.e., they are coherent, they undergo interference and produce interference fringes in the field of view of the telescope. When the two mirrors are exactly perpendicular, known as *normal adjustment,* concentric bright and dark circular fringes are obtained.

The ray travelling towards mirror μ_1, cross the plate G_1 twice, whereas the ray travelling towards mirror μ_2 does not cross G_1 even once. To make the path traversed inside the glass by both the rays equal, an identical glass plate G_2, called *compensating glass plate,* is provided which is parallel to G_1.

When looking in the direction of mirror μ_1, a virtual image (M'_2) of mirror μ_2, due to reflection in G_1 in also observed in addition to μ_1 (Fig. 1.29).

Thus, of the two interfering beams, one comes by reflection from mirror μ_1 and the other which is reflected from μ_2 behaves as if it is reflected from μ_2. Hence, the Michelson interferometer is optically equivalent to an air film between μ_1 and μ'_2. Depending upon path difference between the interfering rays and angle between μ_1 and μ'_2, circular, straight and parabolic fringes are formed.

Types of Fringes

***(i)* Circular Fringes:** When mirror μ_2 is exactly perpendicular to mirror μ_1 or the mirror μ_1 and the virtual mirror μ'_2 (image of mirror μ_2) are parallel, an air film of constant thickness is enclosed between μ_1 and μ'_2 (Fig. 1.29). The images S_1 and S_2 of the source S are formed in the mirrors μ_1 and μ'_2 respectively.

If Δ is the separation between μ_1 and μ'_2, then the separation between virtual images S_1 and S_2 is 2Δ. Therefore, the path difference between the rays is:

$$S_1S_2\cos\theta = 2\Delta\cos\theta$$

where θ is the angle, the ray from virtual source S_2 makes with the normal. That is, if we look obliquely into the interferometer, the line of sight makes an angle θ with the axis.

As the ray coming from mirror μ_2 suffers reflection at the silvered surface of the glass plate G_1, so an additional path difference of $\frac{\lambda}{2}$ is introduced.

∴ Total path difference

$$= 2d\cos\theta + \frac{\lambda}{2} \quad \text{...(1.72)}$$

For Bright Fringes:

$$2d\cos\theta + \frac{\lambda}{2} = n\lambda$$

$$\text{or, } 2\Delta\cos\theta = (2n-1)\frac{\lambda}{2}, \quad n = 1, 2, 3, \ldots \text{...(1.73)}$$

when viewed along the axis, θ = 0, therefore,

$$2\Delta = (2n-1)\frac{\lambda}{2}$$

For Dark Fringes:

$$2d\cos\theta + \frac{\lambda}{2} = (2n+1)\frac{\lambda}{2}$$

$$\text{or, } 2\Delta\cos\theta = n\lambda \quad \text{...(1.74)}$$

when viewed along the axis, θ = 0, therefore,

$$2\Delta = n\lambda.$$

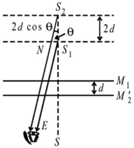


Fig. 1.29

(a) (b)

Fig. 1.30

Thus, for given value of Δ, n and λ, each fringe is a locus of constant θ. Since the loci of constant θ are concentric circles, circular fringes are obtained (Fig. 1.30). These fringes are of equal inclination and the *fringes of equal inclination are known as Haidinger's fringes*. The formation of fringes is shown in Fig. 1.31.

When μ_1 and μ'_2 coincide so that $\Delta = 0$, the field of view is perfectly dark [Fig.1.30 (*b*)]. As the separation between μ_1 and μ'_2 increases (i.e., Δ becomes larger), a ring of particular order n increases in size as the product $2\Delta\cos\theta$ must remain constant. *The rings, therefore expand and new rings appear at the centre, one ring appearing each time one mirror is moved*

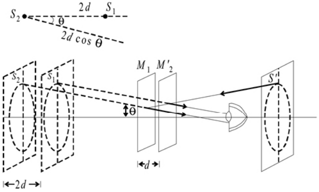
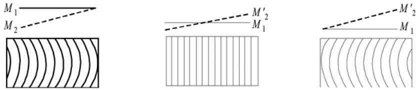


Fig. 1.31: Formation of circular fringes in Michelson's interferometer.

$\frac{\lambda}{2}$.

Also, as d increases, the rings in the periphery disappear slowly as compared to the appearance of new rings at the centre leading to crowding of the field of view by the fringes. Conversely, as d is reduced, the rings disappear at the centre and the pattern contracts.

***(ii)* Localised Fringes:** When the mirrors, μ_1 and μ_2 are not exactly perpendicular to each other or μ_1 and μ'_2 are not exactly parallel, the air film enclosed between them is wedge- shaped and the path difference is very small. The shape of the fringes for various types of intersections of μ_1 and μ'_2 are shown in Fig. 1.32. When μ_1 and μ'_2 intersect in the middle, the path difference is very small and the straight fringes are formed. When μ_1 and μ'_2 are inclined as shown in Fig. 1.32 (*a*) and (*c*), the path differnce increases and the curved fringes, with convexity towards the thin edge of the wedge are formed. Even in these cases, the path difference is much smaller compared to that in case of circular fringes.



(a) (b) (c)

Fig. 1.32

***(iii)* Fringe with White Light:** When white light is used and μ_1 is parallel to M'_2 the central fringe is dark and a few (8 to 10) coloured fringes are observed. To observe these fringes, the mirror μ_2 is titled so that a wedge-shaped film is formed giving rise to a distinct straight fringe bordered on either side by coloured fringes. These fringes are very useful to find zero path difference, especially in the standardisation of the meter.

Applications: Michelson interferometer can be used for the following applications.

***(i)* Determination of Wavelength of Monochromatic Light:** The interferometer is illuminated by the monochromatic source of light whose wavelength λ is to be determined. The mirrors are adjusted and the circular fringes are obtained. The cross-wire is adjusted at the centre of a particular fringe. When the mirror μ_1 is moved slowly, each fringe gets displaced parallel to itself in the field of view and the number of fringes which cross the centre of the field of view gives the measure of the distance the mirror has moved in terms of the wavelength of light. When the mirror moves through a distance $\lambda/2$, one fringe crosses the field of view as an additional path difference of λ is introduced between the two interfering rays.

Let n be the number of fringes which shift across the cross-wire, when the mirror μ_1 is displaced through a distance Δ, resulting in a path difference of 2Δ between the two rays, Hence,

$$2\Delta = n\lambda$$

$$\text{or, } \lambda = \frac{2d}{n} \quad ...(1.75)$$

Thus, counting n and recording Δ on a micrometer, wavelength λ can be calculated.

***(ii)* Determination of Thickness or Refractive Index of a Thin Sheet:** Let a thin transparent sheet of refractive index μ and thickness t is placed in the path of one of the interfering beams. Consequently, the optical path in this beam will increase as light travels slower in any medium compared to that in free space. The optical path will become μt in the medium and the increase in the optical path will be $(\mu - 1)t(= \mu t - t)$ on passing through the medium. Since the beam traverses the sheet twice, total increase in optical path is $2(\mu - 1)t$.

The procedure involves that the interferometer is set for parallel fringes using white light. The cross-wire is set on the central white fringe. Now the thin sheet is placed in the path of one of the interfering beams. As a result, a shift in fringe system occurs. Now the number of fringes n is counted between the cross-wire and the new position of central white fringe, Hence,

$$2(\mu - 1)t = n\lambda \quad ...(1.76)$$

If refractive index μ and wavelength λ are known, the thickness of the sheet can be calculated using the relation:

$$t = \frac{n\lambda}{2(\mu - 1)} \quad ...(1.77)$$

In case thickness t of the sheet is known, refractive index can be calculated using the relation:

$$\mu = \frac{n\lambda}{2t} + 1 \quad ...(1.78)$$

***(iii)* Resolution of Spectral Lines:** This interferometer is very useful in resolving two close spectral lines. Suppose the source S used to illuminate the interferometer consists of two wavelengths λ_1 and λ_2 (λ_1 being slightly higher than λ_2; for example, Δ_1 (l_1 = 5890 Å) and Δ_2 (l_2 = 5896 Å) lines of sodium light source.) Each wavelength produces its own interference pattern, which on superposition gives rise to resultant positions of maxima and minima. At positions, where a bright fringe of one pattern falls over a bright fringe of other pattern resultant maxima are produced. While on the location where a bright fringe of one system meets the dark fringe of the other system, resultant minima are formed. The position of mirror μ_1 is adjusted and noted so that the cross-wire is at the centre of a maximum. Now mirror μ_1 is moved so that the position of one maximum is replaced by the next maximum, i.e., cross-wire is at the centre of the next maximum. This happens when the *number of fringes contained in the interference pattern of smaller wavelength λ_2 is one more than those for higher wavelength λ_1*. Let Δ be the distance through which the mirror μ_1 has been moved so that the corresponding path difference between the interfering rays is 2Δ. Thus,

$$2\Delta = n\lambda_1 \Rightarrow n = \frac{2d}{\lambda_1}$$

$$\text{and, } 2\Delta = (n+1)\lambda_2 \Rightarrow n + 1 = \frac{2d}{\lambda_2}$$

$$\therefore \frac{2d}{\lambda_2} - \frac{2d}{\lambda_1} = 1$$

$$\text{or, } 2d\left(\frac{\lambda_1 - \lambda_2}{\lambda_1\lambda_2}\right) = 1$$

$$\text{or, } \lambda_1 - \lambda_2 = \frac{\lambda_1\lambda_2}{2d} \quad ...(1.79)$$

Since the difference between λ_1 and λ_2 is very small, putting $\lambda_1 - \lambda_2$ = Dl and $\lambda_1\lambda_2$. λ^2, the above expression can be put as:

$$\text{Dl} = \frac{\lambda^2}{2d}. \quad ...(1.80)$$

***(iv)* Standardisation of the Meter:** While learning the basics of measurements, you must have realised that smaller the unit of measurement, better is the accuracy of the measured quantity. The wavelength of light is one of the smallest accurately measurable quantity, which is also invariable. For these reasons, the standard meter is calibrated or standardised in terms of the wavelength of light using Michelson interferometer. The experiment to calibrate the standardised meter in terms of the wavelength of the cadmium red line was first performed by Michelson and Benoit in 1895.

The maximum path difference between the interfering rays for which fringes can be obtained in 24 cm and, therefore, the movable mirror cannot be displaced more than 12 cm. Hence, it is *not possible to count the number of fringes crossing the field of view, when the movable mirror is displaced by one meter*. In other words, *it is not possible to directly count the number of wavelengths contained in one meter*. To overcome this difficulty, Michelson devised *nine sub-standards known as etalons*, the length of each being approximately twice its predecessor. The longest etalon was 10 cm long and the shortest was $\frac{10}{256}$ cm (= 0.390625 mm). An etalon consists of a bronze bar B on which two front silvered plane mirrors μ_1 and μ_2 are fixed and the planes of these mirrors are so adjusted that they are exactly vertical and parallel to each other. The separation between the mirror surfaces is d (Fig. 1.33). The mirrors can be made perfectly parallel with the help of the screws attached to them. The experiment can be divided into two main parts:

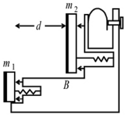


Fig. 1.33

(*a*) The number of wavelengths of the monochromatic cadmium light is counted for the shortest etalon.

(*b*) The length of the second etalon is compared with the shorter (previous) etalon and the process is continued until the number of wavelengths contained in a length of one meter is determined, which is equivalent to *a standard meter,* and can be reproduced whenever needed. The standard meter is represented in terms of wavelengths of red, green and blue lines of cadmium source.

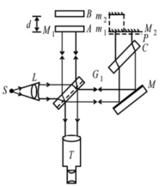
Experimental Procedure: As shown in Fig. 1.34, the basic Michelson interferometer is modified for this purpose. Light from a source *S* is incident on a lens λ which renders the rays parallel.

The parallel rays of light are incident on a glass plate G_1 with its bottom surface silvered. One portion of the beam is reflected towards mirror μ_1 while the other portion after reflection from μ_1 falls upon the mirrors μ_1 and μ_2 of the shortest etalon and the fixed mirror μ_2. The centre of μ_2 lies in a horizontal plane parallel and above the plane containing the centres of μ_1 and μ_2. *P* is the reference plane and is the image of μ_1. The mirrors μ_1 and μ_2 are adjusted such that their planes are parallel to the reference plane *P*. Circular fringes are visible in the field of view of both μ_1 and μ_2, when seen through the telescope *T*. Now the mirror μ_1 is adjusted such that the reference plane *P* makes a small angle with μ_1 and μ_2, when the reference plane *P* intersects the plane of μ_1 in the middle, straight line fringes are obtained with white light as shown in Fig. 1.35 (*a*). The fringes are due to the wedge-shaped film between *P* and μ_1 [Fig. 1.35 (*b*)].

Now, white light is replaced by monochromatic cadmium light and the fixed mirror μ_2 is adjusted to be perfectly parallel to μ_1 so that the circular fringes are visible in the field of view. With white light, when the mirror μ_1 is at *A*, straight line fringes are visible in μ_2 in the field of view of the telescope [Fig. 1.35(*c*)]. The circular fringes are formed at infinity.

The mirror μ_1 is moved and the number of circular fringes that cross the field of view are counted. When the plane of reference *P* intersects the middle of μ_2, straight line fringes are visible in μ_2 with white light [Fig. 1.35(Δ)]. White light is used to note the initial and final positions of the reference plane *P* intersecting μ_1 and μ_2, while the monochromatic light is used to count the number of fringes. Suppose Δ is distance covered by the mirror μ_1 in going from *A* to *B* and *n* is the number of fringes crossing the field of view, then:

$$d = \frac{n\lambda}{2}$$



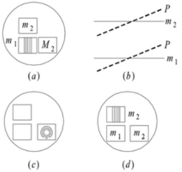


Fig. 1.34 Fig. 1.35

The next step is to compare the shortest etalon with the next etalon (with mirrors μ_3 and μ_4). The two etalons are arranged side by side (Fig. 1.36). First, the mirror μ_1 is adjusted such that the reference plane *P* intersects the middle of the mirrors μ_1 and μ_3, the lower mirrors visible in μ_1 and μ_3 [Fig. 1.36 (*a*)], indicating that μ_1 and μ_3 are coplanar. Now the mirror μ_1 is moved such that straight line fringes are visible in the field of view of the upper mirror μ_2 and the central dark fringe in the middle. Keeping μ_1 fixed (so that mirror plane *P* is fixed) the shorter etalon is moved backward till μ_1 intersects *P* [Fig. 1.36 (*c*)], and the central dark fringe is in the middle of μ_1. Thus, etalon has moved through a distance equal to its length.

If the etalon $\mu_3\mu_4$ is exactly twice length of the etalon $\mu_1\mu_2$, the mirror μ_2 and μ_4 should be coplanar. When the mirror μ_1 is moved such that the reference plane *P* just intersects the middle of μ_2 and μ_4, straight line fringes are produced [Fig. 1.36 (Δ)] and the number of fringes crossing the field of view is counted. Similarly, the next etalon is compared with the etalon $\mu_3\mu_4$, and so on. In this manner, the number of wavelength in 10 cm long etalon is counted.

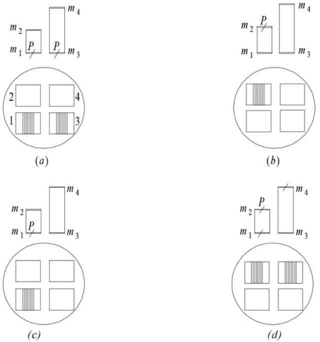


Fig. 1.36

The final result with three cadmium lines are:

Red line: 1 m = 1,553,163.5 λ_R, λ_R = 6438.4722 Å

Green line: 1 m = 1,966,249.7 λ_G, λ_G = 5085.8240 Å

Blue line: 1 m = 2,083,372.1 λg, λ_B = 4799.9107 Å

Example 1.34

When the movable mirror of a Michelson interferometer is shifted through 0.0589 mm, 200 fringes cross the field of view. What is the wavelength of the light which is illuminating the interferometer?

Solution:

Here, $\Delta = 0.0589 \times 10^{-3}$ m, $n = 200$

∴ Wavelength λ =

$$\frac{2d}{n} = \frac{2 \times 0.0589 \times 10^{-3}}{200} = 5890 \times 10^{-8} \text{ cm}$$

Example 1.35

When a thin plate of glass of refractive index 1.5 is placed in the path of one of the interfering beams of Michelson interferometer, a shift of 30 fringes across the field of view is observed. If the thickness of the plate is 0.018 mm, calculate the wavelength of the used light.

Solution:

Here, m = 1.5, $n = 30$, $t = 0.018$ mm $= 1.8 \times 10^{-5}$ m

Path difference due to glass plate $2(\mu - 1)t$

$$\therefore 2(\mu - 1)t = n\lambda$$

$$\text{or, } \lambda = \frac{2(\mu - 1)t}{n} = \frac{2 \times 0.5 \times 1.8 \times 10^{-5}}{30}$$

$= 6.0 \times 10^{-7}$ m = 6000 Å.

Example 1.36

An air cell of 5 cm length with transparent thin windows is introduced in one of the arms of a Michelson interferometer. Circular fringes are observed with light of wavelength 5460 Å. As the pressure of air in the cell is reduced from 760 mm to 380 mm, 20 fringes disappear. What is the refractive index of air at 760 mm pressure? By what distance and in which direction the mirror should be moved so that an equal number of fringes reappear in the field of view (μ − 1 is proportional to pressure)?

Solution:

Let μ be the refractive index of air at 760 mm pressure. As (μ − 1) is proportional to pressure, the value of (μ′ − 1) at 380 mm pressure will be:

$$\frac{380}{760} \times (\mu - 1) = \frac{1}{2}(\mu - 1)$$

Path difference introduced due to cell,

$$2(\mu' - 1)t = 2\frac{1}{2}(\mu - 1)t = (\mu - 1)t$$

If n is the number of fringes disappearing from the field of view, then,

$$(\mu - 1)t = n\lambda$$

Here, $t = 5$ cm, $n = 20$, $\lambda = 5460$ Å $= 5.460 \times 10^{-5}$ cm

$$\therefore (\mu - 1)5 = 20 \times 5.460 \times 10^{-5}$$

or, m = 1 + 0.0002184 = 1.0002184

Thus, the refractive index of air at 760 mm pressure is 1.0002184.

The order of central fringe is maximum in case of Michelson interferometer. The fringes appear when the path difference increases and disappear when path difference decreases. In the present case, the fringes disappear due to decrease in optical path due to decreased refractive index on the reduction of pressure. For the reappearance of the fringes, the optical path should be increased, i.e., the movable mirror should be moved in a direction so that the optical path difference increases. If x is the distance by which the mirror is displaced so that n fringes reappear, then:

$$x = \frac{n\lambda}{2}$$

Here, $n = 20$, $\lambda = 5460$ Å $= 5.46 \times 10^{-5}$ cm

$$\therefore x = \frac{20 \times 5.46 \times 10^{-5}}{2} = 5.46 \times 10^{-4} \text{ cm.}$$

Example 1.37

In Michelson interferometer, the positions of movable mirror read 0.6025 mm and 0.8970 mm for a pair of consecutive bright? Mean wavelength of the used light is 5393 Å. Find the difference between the two wavelengths.

Solution:

Here, $\Delta = 0.8970 - 0.6025 = 0.2945$ mm

$= 2.945 \times 10^{-4}$ m

$\lambda = 5893 \times 10^{-10}$ m

We know, $\lambda_2 - \lambda_1 = \frac{\lambda^2}{2d}$

$$\therefore d\lambda = \frac{5893 \times 5893 \times 10^{-20} \text{ m}}{2 \times 2.945 \times 10^{-4} \text{ m}}$$

$= 5.896 \times 10^{-10}$ m . 6 Å.

Exercises

Short Answer Type

1. Briefly introduce Huygen's wave theory. What was the need for the assumption of an all prevailing homogeneous medium called luminiferous ether?
2. State the principle of superposition of waves.
3. What do you mean by coherent sources? Why can't independent sources be coherent?
4. Differentiate between temporal and spatial coherence.
5. Define 'interference'. Give the conditions for constructive and destructive interference.
6. Why coherent sources are required for interference? How are they obtained?
7. Define 'fringe'. What is the actual shape of a fringe in Young's double slit experiment and why do they appear straight?
8. Define fringe width and show that in interference pattern all fringes are equally spaced.
9. When white light is used, what type of interference pattern is obtained in Young's double slit experiment?
10. What do you mean by lateral shift? How is it removed in Fresnel's biprism?
11. Explain the fringes obtained with white light in Fresnel's biprism.
12. Show that in a thin film interference patterns, in reflected and transmitted light are complimentary.
13. Discuss the various conditions required for viewing colours in thin film. Why should the thickness of the film not be too large?
14. What type of fringes are obtained in a thin wedge-shaped film? How can it be used to check the flatness of a glass surface?
15. What are Newton's rings? Why are they circular in shape?
16. Is the central band in Newton's ring bright or dark in (*i*) reflected light, and (*ii*) transmitted light? Why?
17. Explain how Newton's rings with bright central band can be obtained in reflected light?
18. What are Haidinger fringes?
19. Define visibility of fringes and obtain an expression for the same.
20. Why the fringes in normal condition in Michelson interferometer are circular in shape?
21. Discuss fringe pattern with white light in Michelson interferometer.
22. Explain how Michelson interferometer can be used for the determination of wavelength of monochromatic light source?

Long Answer type

1. What is light? Starting from the corpuscular theory, discuss the present understanding about the nature of light.
2. What is interference of light? Give the conditions for sustained interference. How coherent sources are obtained in interferometers?
3. Give the analytical treatment for interference and hence obtain the conditions for brightness and darkness. Derive the expressions for maximum and minimum intensity, when the two waves have the same amplitude.
4. Show that the phenomenon of interference is in accordance with the law of conservation of energy.
5. Discuss Young's double slit experiment. Obtain an expression for the fringe width and show that it is independent of the order of the fringe.
6. Explain the phenomenon of interference in Fresnel's biprism, mentioning how coherent sources are produced. Describe the method for the determination of wavelength of a monochromatic light.
7. Discuss the interference in a thin film and obtain the condition for brightness and darkness for both reflected and transmitted light.
8. Discuss interference of light in a thin wedge-shaped film and obtain an expression for the fringe width.
9. Discuss the formation of Newton's rings in reflected monochromatic light and show that the diameters of bright rings are proportional to the square root of the odd natural numbers and that of dark rings to the square root of the natural numbers.
10. Discuss the method for determining the refractive index of a liquid by Newton's rings method.

11. Discuss the formation of Newton's rings by two curved surfaces when they are placed: (*i*) one above the other and (*ii*) oppositely directed.

12. Describe the interference in Fabry-Perot interferometer and show that the circular fringes of equal inclination are formed. Obtain the conditions for maxima and minima.

13. Explain the construction and working of an interference filter.

14. Describe the construction and working of Michelson interferometer. Discuss the conditions for various types of fringes.

15. Discuss the procedure for determining the refractive index and thickness of a transparent sheet using Michelson interferometer.

16. What do you mean by resolution of spectral lines? How can it be done using Michelson interferometer?

17. What do you mean by standardisation of meter? Discuss the procedure involved.

Numericals

1. In Young's double slit experiment, light of wavelength 6400 Å is used. The slits are 2 mm apart and the fringes are observed on a screen placed 10 cm away from the slits and it is found the interference pattern shifts by 5 mm when a transparent sheet of thickness 0.5 mm is introduced in the path of one of the slits. Find the refractive index of the sheet. [**Ans.** 1.2]

2. Obtain an expression for the fringe with β in an interference pattern produced by two parallel slits using light of wavelength λ. If an additional variable path difference is introduced in the waves from one of the sources at a rate of p cm per second (may be by introducing a gas cell of variable pressure), deduce an expression for the rate at which the fringe system will be shifting.

[**Ans.** $\beta = \dfrac{D\lambda}{d}$, fringe shift $y_0 = \dfrac{\beta}{\lambda}(\mu - 1)e$ (ε = thickness), rate of fringe shift $\dfrac{dy_0}{dt} = \dfrac{\beta}{\lambda}(\mu - 1)p\,\text{cm/s}), \dfrac{de}{dt} = p$]

3. In Young's double slit experiment using a monochromatic light source, on introduction of a transparent sheet of refractive index 1.6 and thickness 1.964 μm in the path of one of the interfering waves, the fringe pattern shifts by a certain distance. The sheet is then removed and the separation between the slits and screen is doubled. It is observed that the distance between successive maxima (or minima) is the same as the observed fringe shifts in the first case. Determine the value of λ. (Fringe shift $\dfrac{D}{d}(\mu - 1)t$; 5892 Å)

4. In a double slit experiment, fringes are produced using light of wavelength 4800 Å. One slit is covered by a thin plate of glass of refractive index 1.4 and the other slit by another glass plate of refractive index 1.7 and same thickness. It is observed that now the central bright fringe shifts to the location originally occupied by the fifth bright fringe. What is the thickness of the glass plate?

(use $n\lambda = (\mu - \mu')t$, $n = 5$, $\mu - \mu' = 0.3$ to get $t = 8.0 \times 10^{-8}$ m)

5. A biprism of angle 1° and refractive index 1.5 forms interference fringes on a screen 1 m from the biprism. The distance between the slit and biprism 0.1 m. Find the bandwidth of the fringes when the biprism is illuminated by a light of wavelength 6000 Å. [**Ans.** 3.438×10^{-4} m]

6. The distance between the slit and the biprism and between the biprism and the screen are 50 cm each. The angle of prism is 179° and its refractive index is 1.5. If the distance between successive fringes is 0.0135 cm, find the wavelength of the used light. [**Ans.** 5893 Å]

7. The inclined faces of glass biprism (μ = 1.5) make an angle of 2° with the base of the prism. The slit is 10 cm away from the prism and is illuminated by a light of wavelength 5500 Å. Calculate (*a*) the separation between the coherent sources formed by the biprism and (*b*) the fringe width at a distance 1 m from the slit . [**Ans.** (*a*) 0.349 cm, (*b*) 0.0157 cm]

8. A biprism of obtuse angle 176° is made of glass having refractive index 1.5. A slit, illuminated with light of unknown wavelength, is placed 20 cm behind the biprism. The width of interference fringe formed on a screen placed 80 cm away of the biprism is found to be 8.25×10^{-3} cm. What is the wavelength of the light used. [**Ans.** 5757 Å]

9. In a biprism experiment with light of wavelength 5893 Å, 40 fringes are observed in the field of view, If light of wavelength 4358 Å is used, how many fringes will be seen in the same field of view?

[**Ans.** 54]

10. A plane wave of monochromatic light falls normally on a uniform thin film of oil which covers a glass plate. The wavelength of the source can be varied continuously. Complete destructive interference of the reflected light is observed only for wavelengths of 5000 Å and 7000 Å. If the refractive index for the oil and that of glass are 1.30 and 1.50 respectively, calculate the thickness of the film.

(If the order for λ_1 is n, then for $\lambda_2(\lambda_2 > \lambda_1)$, order will $n - 1$, t = 0.67 μm).

11. A drop of oil of volume 0.2 cm^3 is dropped on the surface of a tank of water of area 1 m^2. The film spreads uniformly over the whole surface and white light reflected normally is observed through a spectrometer. The spectrum is seen to contain first dark band, whose centre has wavelength of 5.5×10^{-7}m (in air). Find the refractive index of oil. [**Ans.** 1.375]

12. White light is reflected from an oil film of thickness 0.01 mm and refractive index 1.4 at an angle of 45° to the vertical. If the reflected light falls on the slit of spectrometer, calculate the number of dark bands seen between wavelengths 4000 and 5000 Å. [**Ans.** 12]

13. A monochromatic light of wavelengths 5893 Å falls normally on a wedge-shaped air film. If the length of the wedge is 0.05 m, calculate the distance at which the 12th dark and 12th bright fringes will from the line of contact of the glass plates forming the wedge. Thickness of the specimen = 154×10^{-5} m).

[**Ans.** 9.61×10^{-4} m, 9.21×10^{-4} m]

14. A wedge-shaped air film of angle 20 seconds is illuminated by light of wavelength 6000 Å. Find the fringe width. [**Ans.** 3.09×10^{-3} m]

15. In a Newton's ring experiment, the diameter of the 5th ring is 0.336×10^{-2} m and the diameter of the 15th ring is 0.59×10^{-2} m. Find the radius of the plano-convex lens if the wavelength of light used is 5890 Å. [**Ans.** 0.998 m]

16. A Newton's ring arrangement is used with a source emitting two wavelengths $\lambda_1 = 6 \times 10^{-5}$m and $\lambda_2 = 4.5 \times 10^{-5}$ m. It is found that nth dark ring due to λ_1, coincides with $(n + 1)$th dark ring for λ_2. If radius of curvature of the lens is 90 cm, find the diameter of the nth dark ring. [**Ans.** 0.254 cm]

17. Light containing two wavelengths λ_1 and λ_2 falls normally on a plano-convex lens of the curvature R resting on a glass plate. If the nth dark due to λ_1 coincides with $(n + 1)$th dark ring due λ_2, prove that the radius of the nth dark ring of λ_1 is $\sqrt{\dfrac{\lambda_1\lambda_2 R}{\lambda_1 - \lambda_2}}$.

$$\left[\begin{array}{ll} \textbf{Hint: } \text{Given; } r = \sqrt{n\lambda_1 R} = \sqrt{(n+1)\lambda_2 R} \Rightarrow & n\lambda_1 R = (n+1)\lambda_2 R \\ \Rightarrow & n = \dfrac{\lambda_2}{\lambda_1 - \lambda_2} \\ \therefore & r = \sqrt{n\lambda_1 R} = \sqrt{\dfrac{\lambda_1\lambda_2 R}{\lambda_1 - \lambda_2}} \end{array}\right]$$

18. Newton's rings formed by sodium light between a flat glass plate and a convex lens are viewed normally. What will be the order of the dark ring which will have double the diameter of 40th bright ring?
[**Ans.** 160]

19. When the space between the flat disc and convex surface in a Newton rings set up in filled with a liquid, the radius of fringes is reduced to 80% of the original value. Calculate the refractive index of the liquid.[**Ans.** 1.56]

20. A thin plate is introduced in the path of one of the beams of light in Michelson interferometer and it is found that 50 bands have crossed the line of observation. If the wavelength of light is 5896 Å and $\mu = 1.4$, determine the thickness of the plate.[**Ans.** 3685×10^{-8} m]

21. In a Michelson interferometer the readings of a pair maximum indistinctness were found to be 0.6939 mm and 0.9884 mm. If the mean wavelength of the two wavelength of the two components of light be 5893 Å, deduce the difference between the wavelength of the components. [**Ans.** 5.896 Å]

14. A wedge-shaped air film of angle 20 seconds is illuminated by light of wavelength 6000 Å. Find the fringe width. [**Ans.** 3.09×10^{-3} m]

15. In a Newton's ring experiment, the diameter of the 5th ring is 0.336×10^{-2} m and the diameter of the 15th ring is 0.59×10^{-2} m. Find the radius of the plano-convex lens if the wavelength of light used is 5890 Å. [**Ans.** 0.998 m]

16. A Newton's ring arrangement is used with a source emitting two wavelengths $\lambda_1 = 6 \times 10^{-5}$m and $\lambda_2 = 4.5 \times 10^{-5}$ m. It is found that nth dark ring due to λ_1, coincides with $(n + 1)$th dark ring for λ_2. If radius of curvature of the lens is 90 cm, find the diameter of the nth dark ring. [**Ans.** 0.254 cm]

17. Light containing two wavelengths λ_1 and λ_2 falls normally on a plano-convex lens of the curvature R resting on a glass plate. If the nth dark due to λ_1 coincides with $(n + 1)$th dark ring due λ_2, prove that the radius of the nth dark ring of λ_1 is $\sqrt{\dfrac{\lambda_1\lambda_2 R}{\lambda_1 - \lambda_2}}$.

$$\left[\begin{aligned} &\textbf{Hint: } \text{Given; } r = \sqrt{n\lambda_1 R} = \sqrt{(n+1)\lambda_2 R} \Rightarrow n\lambda_1 R = (n+1)\lambda_2 R \\ &\Rightarrow \qquad n = \frac{\lambda_2}{\lambda_1 - \lambda_2} \\ &\therefore \qquad r = \sqrt{n\lambda_1 R} = \sqrt{\frac{\lambda_1\lambda_2 R}{\lambda_1 - \lambda_2}} \end{aligned}\right]$$

18. Newton's rings formed by sodium light between a flat glass plate and a convex lens are viewed normally. What will be the order of the dark ring which will have double the diameter of 40th bright ring?
[**Ans.** 160]

19. When the space between the flat disc and convex surface in a Newton rings set up in filled with a liquid, the radius of fringes is reduced to 80% of the original value. Calculate the refractive index of the liquid.[**Ans.** 1.56]

20. A thin plate is introduced in the path of one of the beams of light in Michelson interferometer and it is found that 50 bands have crossed the line of observation. If the wavelength of light is 5896 Å and $\mu = 1.4$, determine the thickness of the plate.[**Ans.** 3685×10^{-8} m]

21. In a Michelson interferometer the readings of a pair maximum indistinctness were found to be 0.6939 mm and 0.9884 mm. If the mean wavelength of the two wavelength of the two components of light be 5893 Å, deduce the difference between the wavelength of the components. [**Ans.** 5.896 Å]

2
Diffraction of Light

2.1 INTRODUCTION

It is commonly observed that the path of the light entering a dark room through a hole in the window illuminated by sunlight is straight. If an opaque obstacle is placed in the path of light, a sharp shadow is cast. Both these examples show *rectilinear propagation of light*, i.e., light travels in a straight line. The rectilinear propagation of light could also be explained on the basis of Newton's corpuscular theory.

It is, however, very common to listen conversation from a room even when one happens to be outside the room towards a corner, indicating bending of sound waves around windows and doors. Other waves also bend around the edges of an obstacle. If light also propagates in the form of waves, it was expected that a beam of light bends around the edges of an obstacle.

It was Grimaldi (Italy, 1618–1663), who observed that with a small source of light, the shadow of an object was larger than its geometrical shadow. He further observed that the shadow was not sharp and well-defined (contrary to what was expected on the basis of rectilinear propagation of light). He concluded that this was due to the bending of light waves around the corners of an obstacle. He used the term diffraction to describe this phenomenon.

Thus, *the phenomenon of bending and spreading of light waves into the geometrical shadow around the corners of an obstacle or aperture in its path is known as diffraction* of light. In fact, any departure from rectilinear path may be called diffraction. This phenomenon is a manifestation of the wave nature of light. Thus, light can also show diffraction and, so deviates from its rectilinear path, if it encounters an aperture of the size comparable to its wavelength (of the order of 10^{-7} m). If the obstacle is of suitable size and the point source of light is sufficiently bright, bright and dark bands are also observed in the geometrical shadow. The intensity distribution, i.e., bright and dark bands, on the screen is known as diffraction pattern.

Newton also observed this phenomenon and tried to explain it in terms of his corpuscular theory. Thomas Young later tried to explain this phenomenon as an interference effect between the direct light near the edge of the obstacle and the light reflected from the edge at grazing incidence. Fresnel studied the phenomenon extensively and found that this phenomenon did not depend upon the nature or shape of the edge of the obstacle and he ruled out the Young's idea of the dependence of the phenomenon upon reflection.

Based on his experimental observations, Fresnel explained the phenomenon in 1815 satisfactorily. According to him, *diffraction is the mutual interference between the secondary wavelets from the same wavefront*. Alternatively, *it is interference between the light coming from various parts of a single aperture*. In fact, this theory is based on Huygen's principle of secondary wavelets, according to which, every point on any surface (wavefront) through which light is passing, may be treated as a source of secondary wavelets and all the subsequent effects may be explained in terms of the combined effect of the secondary wavelets alone. If this idea is correct, then light from a sufficiently small aperture should spread in all directions beyond it, which is indeed the case as observed experimentally and is explained in the following sections.

Validity of Ray Optics: Fresnel's Distance

A parallel beam of light on passing through slit of width a gets diffracted into a beam of an angle (called angular width) θ, given by :

$$\theta = \frac{\lambda}{a}$$

where λ is the wavelength of light and is being assumed that θ is small.

In traversing a distance Δ = the distance between slit and screen, this beam spreads over linear width x due to diffraction, given by:

$$x = D \cdot \frac{\lambda}{a} \quad \text{...(2.1)}$$

A parameter, known as Fresnel's distance is defined to decide the distance upto which the concepts of ray optics (i.e., rectilinear propagation of light) are valid and beyond which the spreading is large. *Fresnel's distance is defined as the distance of the screen from the slit at which the spreading of light due to diffraction becomes equal to the size of the slit.* It is denoted by D_F.

Thus, when $x = a$, $\Delta = \Delta_F$. so from Eq. (2.1)

$$a = \frac{D_F \lambda}{a}$$

$$\text{or, } \Delta_F = \frac{a^2}{\lambda} \quad \text{...(2.2)}$$

When $D < D_F$, the broadening by diffrection is not much and light travels almost along a straight path and the concepts of ray optics are valid.

$$\text{for, } D < D_F$$

$$\text{or, } D < \frac{a^2}{\lambda}$$

$$\text{or, } a > \sqrt{\lambda D}$$

The quantity $\sqrt{\lambda D}$ is called *size of Fresnel's zone* and is denoted by a_F.

$$\therefore a_F = \sqrt{\lambda D}$$

Thus, if $a > a_F$ the concepts of ray optics can be used without introducing any significant error.

2.2 TYPES OF DIFFRACTION

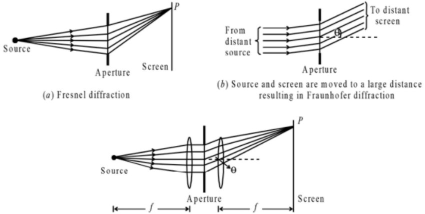
The phenomenon of diffraction is broadly divided into two categories:

(*i*) Fraunhofer diffraction, and

(*ii*) Fresnel diffraction.

(*i*) Fraunhofer Diffraction: In Fraunhofer diffraction, the *source of light and the screen are effectively at infinite distance* from the diffracting element. This condition is achieved by using two convex lenses, one for rendering the incident rays parallel before it falls on the aperture and another, to focus the diffracted light on the screen (Fig. 2.1). Thus, in Fraunhofer diffraction, the incident wavefront is plane and the secondary wavelets originating from the unblocked portion of the wavefront are in the same phase at every point in the plane of the aperture. So, we get the diffraction because of the interference between parallel rays which are brought to focus with a convex lens. This lens forms in its focal plane a reduced image of the source modified by diffraction at the aperture, which would appear at infinity in the absence of the lens.

(*ii*) Fresnel Diffraction: In Fresnel diffraction, the *source of light and the screen are at finite distances* from the diffracting element, i.e., aperture or obstacle. No lenses are used in this case. In this type of diffraction, interference takes place between light waves reaching a point (near or in the shadow of the obstacle) from different parts of the same wavefront at the aperture without modification by lenses (Fig. 2.1). In Fresnel diffraction, the incident wavefront is either *spherical or cylindrical.* The phase of the waves at all the points in the plane of the aperture of the obstacle is not the same.



(*a*) Fresnel diffraction

(*b*) Source and screen are moved to a large distance resulting in Fraunhofer diffraction

(*c*) Fraunhofer diffraction conditions produced by lenses, leaving source and screen in their original positions

Fig. 2.1

Fresnel Assumptions: In order to explain the observed diffraction pattern, Fresnel took into account the secondary wavelets originating from various points of wavefront. He showed that the rectilinear propagation of light is only approximate and even in the geometrical shadow region, wavelets undergo destructive and constructive interference. His arguments is based on the following assumptions:

(*i*) Each element of a wavefront continuously sends secondary waves, unlike a single spherical pulse in case of Huygen's theory.

(*ii*) A wavefront can be divided into a large number of zones or elements called *Fresnel's zone* of small area.

(*iii*) The resultant amplitude at any point is determined by the combined effect of all the secondary waves reaching there from various zones.

(*iv*) The effect at a point due to any particular zone depends on the distance of the point from the zone.

(*v*) The effect at any point Q (Fig. 2.2) also depends on the angle θ between the line PQ and normal PO. Actually, the effect along any direction is proportional to $(1 + \cos\theta)$ which is called the *obliquity factor*. Naturally, the effect is maximum along PO as $\theta = 0$ and $\cos\theta = 1$. The resultant effect along PA is one-half of that along PO because $\theta = 90°$ and $\cos 90° = 0$. Along PS, the resultant effect is zero as $\theta = 180°$, $\cos 180° = -1$ and hence reducing the factor $1 + \cos\theta$ to $1 - 1 = 0$. This property of secondary waves eliminated the problem of backward propagation of energy, faced by Huygen's principle assuming that waves spread in all directions. Thus, amplitude of the secondary waves along backward direction is zero and hence there is no wave travelling backward.

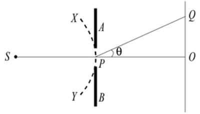


Fig. 2.2

2.3 Half period zones

Consider a point source S of monochromatic light which is at a sufficiently large distance so that for all practical purposes, we may consider the incident wavefront at the aperture to be plane (Fig. 2.3). We want to find the resultant amplitude at a distant point O due to the aperture limiting the exposed part of the incident wavefront XY. If O is joined with S, the point P on the wavefront is called the *pole of the wavefront with respect to O*.

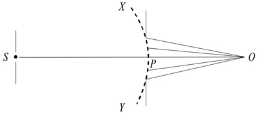


Fig. 2.3

Let $ABCD$ be a plane wavefront (perpendicular to the plane of the paper) of the monochromatic light of wavelength λ at any time, propagating towards right (Fig. 2.4). In order to find the effect at O, we consider every point on the wavefront $ABCD$ to become the centre of disturbances and the net effect can be found due to all these disturbances reaching the point O. All the points on the wavefront $ABCD$ will be in the same phase, but secondary waves originating from $ABCD$ will reach O with different phases. PO is perpendicular drawn from P to O and let $PO = b$. In Fresnel's method, the wavefront is divided into a number of concentric *half period zones*, also called *Fresnel's zones*.

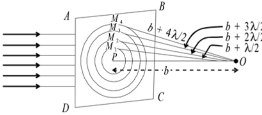


Fig. 2.4

In order to find the resultant effect at O, with P as centre and radii $OM_1 = b + \lambda/2$, $OM_2 = b + 2\cdot\lambda/2$, $OM_3 = b + 3\cdot\lambda/2$ ...$OM_{n-1} = b + (n-1)\cdot\lambda/2$ and $OM_n = b + n\cdot\lambda/2$, a series of concentric spheres are drawn which cut the wavefront $ABCD$ in circles of radii PM_1, PM_2, PM_3, ..., PM_n.The circular zones enclosed between PM_1, $\mu_1\mu_2$, $M_2\mu_3$, ... are known as *half period zones or elements*. The area corresponding to PM_1, $M_1\mu_2$, $M_2\mu_3$, ... are called first, second, third, ... half period zones respectively. The point P is the pole of the wavefront with respect to the point O. A Fresnel half period zone with respect to an external point O is a *thin annular zone* or a thin strip of the primary wavefront surrounding the pole P such that the distances of its outer and inner edges from P differ by $\lambda/2$. In other words, each zone differs from its neighbour (predecessor or successor) by a phase difference of π or a path difference of $\lambda/2$. Thus, *waves originating from P and M_1, on reaching O, will have a phase difference of π or a path differece of $\lambda/2$.*

The resultant effect at the point O due to the whole wavefront will depend on the following factors:

(*i*) Area of the each zone; higher the area, more will be amplitude at the point O.

(*ii*) The distance of the point O from the zones; more the distance, lower will be amplitude at the point O.

(*iii*) The obliquity factor; the amplitude depends on the angle which the considered point makes with the respective zones. The resultant amplitudes at the point O will be equal to the sum of amplitudes from the individual zones. Now we consider each factor separately:

Area of Half Period Zone

The radius of the first half period zone,

$$PM_1 = \sqrt{OM_1^2 - OP^2}$$

$$= \sqrt{\left(b + \frac{\lambda}{2}\right)^2 - b^2}$$

$$= \sqrt{b^2 + \frac{\lambda^2}{4} + b\lambda - b^2} = \sqrt{b\lambda}$$ ($\because \lambda^2 << b\lambda$)

The radius of the second half period zone,

$$PM_2 = \sqrt{OM_2^2 - OP^2}$$

$$= \sqrt{(b+\lambda)^2 - b^2}$$

$$= \sqrt{b^2 + \lambda^2 + 2b\lambda - b^2} = \sqrt{2b\lambda}$$ ($\because \lambda^2 << 2b\lambda$)

Similarly, radius of $(n - 1)$th half period zone, $PM_{n-1} = \sqrt{(n-1)b\lambda}$ and radius of nth half period zone, $PM_n = \sqrt{nb\lambda}$

Thus, we find that for the same λ and b, the *radii of the half period zones are proportional to* $\sqrt{n}$ (n = 1, 2, 3, ...), i.e., *square root of the natural numbers.*

Area of first half period zone = $\pi \cdot PM_n^2 = \pi b\lambda$

Area of half period zone = $\pi \cdot PM_2^2 - \pi PM_1^2$

$= 2\pi b\lambda - \pi b\lambda = \pi b\lambda$

Continuing similarly, area of nth half period zone

$$= \pi PM_n^2 - \pi PM_{n-1}^2$$

$$= \pi nb\lambda - (\pi(n-1)b\lambda) = \pi b\lambda$$

Thus, we find that the *area of each half period zone is approximately equal and is independent, of the order of the zone*. More precisely, if the approximation is not made, it is found that there is slight decrease in area with the order n, however, for all practical purposes, areas may be taken as equal. Also, area of half period zone is directly proportional to b (distance of the point O from the wavefront) and λ (wavelength of light).

Distance and Obliquity Factor

With the increase in the order of the half period zone, the contribution made by each successive zone will continue to decrease slowly, but regularly due to increasing distance and obliquity factor. Let $\mu_1, m_2, m_3, ..., m_{n-1}, m_n$ represent the amplitudes at the point O due to the first, second, third,..., $(n - 1)$th and nth half period zones, respectively. As shown in Fig. 2.5 and for the reasons mentioned above, μ_1 is slightly greater than μ_2, m_2 than μ_3, ..., m_{n-1} than μ_n. That is, the quantities, μ_1, m_2, μ_3, ... are of continuously decreasing order such that the amplitude of any zone may be taken as the arithmetic mean between the preceding and succeeding zones. Thus, to a close approximation:

$$m_2 = \frac{m_1 + m_3}{2}, \quad m_3 = \frac{m_2 + m_4}{2}$$

and so on.

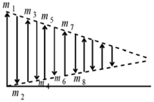


Fig. 2.5

Relative Phases of the Waves from the Zones

Since the distances of the half period zones from the point O are different, the secondary waves originating from these zones will reach O in different phases.

Let us take the phase of the wavelets reaching O from P to be zero. The wavelength from μ_1 will arrive after the time $T/2$ than those form P as the path OM_1 is longer by $\lambda/2$ compared to the path PO. The difference in phase, therefore, for wavelets coming from P and μ_1 is π radians. The phases of the wavelets from intermediate points between P and μ_1 will vary from 0 to π. Hence, the average phase of all the wavelets from the first half period zone is $\frac{0+\pi}{2} = \frac{\pi}{2}$ radians.

The phase difference between wavelets from μ_1 and μ_2 will again be π as they differ in path by $\lambda/2$. Hence, the phase difference between the wavelets reaching O from P and μ_2 will be 2π radians. So the phase of the wavelets from the sound half period zone will vary between π and 2π. Thus, the average phase of wavelets from the second half period zone is $\frac{\pi + 2\pi}{2} = \frac{3\pi}{2}$ radians.

This shows that *the wavelets from the second half period zone reach the point O in opposite phase compared to those from the first half zone.*

Similarly, we can see that the average phase of the wavelets from the third half period zone is $\frac{5\pi}{2}\left(= \frac{2\pi + 3\pi}{2}\right)$ radians. Proceeding in this manner, we find that the *resultant phase difference between the wavelets from any two successive zones is π radians*. In other words, *wavelets from the successive zones reach the point O in opposite phase.*

The resultant amplitude at O due to whole wavefront is given by

$$A = \mu_1 - \mu_2 + \mu_3 - \mu_4 + \mu_5 ... \pm \mu_n ... \quad (2.3)$$

(+ if n is odd, – if n is even)

We can also put the above expression as :

$A =$

$$\frac{m_1}{2}+\left(\frac{m_1}{2}-m_2+\frac{m_3}{2}\right)+\left(\frac{m_3}{2}-m_4+\frac{m_5}{2}\right)+\frac{m_5}{2}\ldots$$

The expression within brackets are each equal to zero. Therefore,

$$A = \frac{m_1}{2}+\frac{m_n}{2}, \text{ if } n \text{ is odd}$$

$$\text{and, } A = \frac{m_1}{2}+\frac{m_{n-1}}{2}-m_n, \text{ if } n \text{ is even}$$

Usually, n is quite large and so the effect due to the $(n-1)$th or nth zone compared to the first zone is almost negligible on account of the distance and obliquity factor. Therefore, the resultant amplitude at O due to the whole wavefront reduces to:

$$A = \frac{m_1}{2} \quad \ldots(2.4)$$

Thus, the resultant *amplitude due to the entire wavefront is one-half of that which will be produced by the first half period zone alone.*

Intensity at a point is proportional to the square of the amplitude.

$$\text{Therefore, } I = \frac{m_1^2}{4} \quad \ldots(2.5)$$

Thus, the intensity at O due to the entire wavefront is one-fourth of that due to the first half period zone alone.

We find that only first half period zone is effective in producing the illumination at O. Fig 2.6 show the variation of intensity at O with the number of exposed zones.

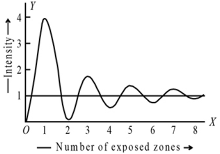


Fig. 2.6

A small obstacle of the size of the first half period zone at the aperture can effectively block the entire wavefront. Taking λ = 6000 Å and b = 50 cm, the area of the first half period zone comes but to be $\pi b\lambda = 3.14 \times 6.0 \times 10^{-5} \times 50 = 9.42 \times 10^{-3}$ cm² = 0.00942 cm². Thus, placing an obstacle of the size of 0.0942 cm², the effect of the entire wavefront can be blocked and practically, there will be no light at O, which is a manifestation of the *rectilinear propagation of light*. In fact, approximate rectilinear propagation of light is a consequence of extreme shortness of the wavelength of light. In case of sound waves, the wavelengths are far greater (almost million times) than that of light and hence the area of the first half period zone for a plane wavefront of sound is very large. Thus, to stop the effect of a sound wavefront, and obstacle of every larger size, about a million times larger than that for light, is required. Hence, for an obstacle of ordinary size, sound waves do not appear to travel in a straight line. *If the size of the obstacle placed in the path of light is comparable to the wavelength of light, then it is possible to observe illumination in the region of the geometrical shadow also. Thus, rectilinear progagation of light is only approximately true.*

Example 2.1

Light of 5000 Å is incident on a circle hole of radius (*i*) 1 cm and (*ii*) 1 mm. How many half period zones are contained in the circle if the screen is placed at a distance of 1 m to observe the diffraction?

Solution

Here, λ = 5000 Å = 5.0×10^{-5} cm, b = 1 m = 100 cm

(*i*) Area of the circular hole = $\pi r^2 = \pi$ cm²

Area of each half period zone = $\pi b\lambda = 5.0 \times 10^{-5} \times \pi$ cm²

$$\therefore \text{Number of half period zones} = \frac{\pi}{5.0\times10^{-5}\times\pi} = 0.2\times10^5 = 20000$$

(*ii*) Area of circular hole = $\pi \times (0.1)^2$ cm²

$$\therefore \text{Number of half period zones} = \frac{\pi\times(0.1)^2}{5.0\times10^{-5}\times\pi} = 0.2\times10^3 = 200$$

Example 2.2

Light of 6000 Å is incident on a circular hole and is received on a screen 50 cm away. What is the radius of the hole if the intensity of light on the screen is four times the intensity when light is incident without the hole?

Solution

The intensity will increase to four times in the presence of the hole than in its absence if the radius of the hole is equal to that of the first half period zone.

Radius of the first half period zone = $\sqrt{b\lambda}$

Here, b = 50 cm, λ = 6000 Å = 6.0×10^{-5} cm

$\therefore$ Radius of the hole =

$$\sqrt{50\times6.0\times10^{-5}} = 5.48\times10^{-2} \text{ cm}$$

= 0.0548 cm

2.4 zone plate

A zone plate is a specially constructed optical device which *blocks the light from every alternate zone*. It is an interesting example of diffraction which also provides experimental verification of the Fresnel's idea of division of a wavefront into half period zones. It is based on the fact that because the light waves from consecutive zones reach O in opposite phases reducing the intensity, it is possible to obtain a bright illumination at O by obstructing the light from alternate zones. In this way, all the waves reach at O in phase, causing reinforcement. We have seen that the resultant amplitude A to O is given by:

$$A = \mu_1 - \mu_2 + m_3 - \mu_4 + m_5 \ldots$$

In a *positive zone plate* (Fig. 2.6 (*a*)), odd zones are transparent and the even zones are opaque so that the resultant amplitude at O is given by:

$$A = \mu_1 + m_3 + m_5 + m_7 + \ldots$$

In a *negative zone plate* (Fig. 2.7), even zones are transparent and the odd zones are opaque so that the resultant amplitude is given by:

$$A = -(\mu_2 + \mu_4 + \mu_6 + \mu_8 + \ldots\ldots)$$

(*a*) Positive zone plate (*b*) Negative zone plate

Fig. 2.7

In both the cases, the amplitude and hence the intensity at O increases enormously. This action is called the *focussing action* of a zone plate. The construction of zone plate is based on the fact that radii of the zones are proportional to the square root of the natural numbers $(r_n \propto \sqrt{n}; r_1 \propto \sqrt{1}, r_2 \propto \sqrt{3}, r_3 \propto \sqrt{3}, \ldots)$ and the area of each zone is equal.

Construction: To construct a zone plate, large number of concentric circles are drawn on a white cardboard with radii of the circles being proportional to the square root of the natural numbers. Now, for the construction of a positive zone plate, the odd numbered zones (i.e., 1st, 3rd, 5th, ...) are painted black. Then a reduced photograph is taken on thin glass plate. In the developed negative, the odd zones are transparent to the incident light and the even zones will not pass the incident light. The resulting zone plate is known as positive zone plate.

For the construction of the negative zone plate, the even numbered zones (i.e., 2nd, 4th, 6th, ...) are painted black. The rest of the procedure remains the same. Now the even numbered zones are transparent and the odd numbered zones are opaque to the light incident light.

Thus, *a zone plate is a specially constructed screen with alternate transparent and opaque concentric zones.*

Theory: As shown in Fig. 2.8, consider a zone plate perpendicular to the plane of the paper. S is a point source of monochromatic light giving out spherical waves of wavelength λ, at a distance a from the centre P of the zone plate. Let O be the point on a screen placed at right angle to the plane of the paper, at a distance b from the centre of the zone plate, where the amplitude is to be determined.

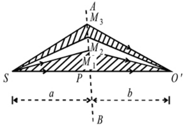


Fig. 2.8

With P as centre, imagine the plate AB to be divided into circular zones with radii equal to PM_1, PM_2, $PM3$,... *such* that,

$$SM_1 + \mu_1 O = SP + PO + \frac{\lambda}{2}$$

$$SM_2 + \mu_2 O = SP + PO + \frac{2 \cdot \lambda}{2}$$

....

$$SM_n + \mu_n O = SP + PO + n \cdot \frac{\lambda}{2} \quad \ldots(2.6)$$

We know $SP = a$ and $PO = b$, therefore, if r_n ($= PM_n$) is the radius of the nth circle, then,

$$SM_n = \sqrt{SP^2 + PM_n^2} = \sqrt{a^2 + r_n^2}$$

$$= a\left(1 + \frac{r_n}{a}\right)^{\frac{1}{2}}$$

Using binomial expansion and neglecting terms higher than r_n^2 we get:

$$SM_n = a + \frac{1}{2}\frac{r_n^2}{a} \quad \ldots(2.7)$$

$$\text{Similarly, } M_n O = \sqrt{PO^2 + PM_n^2} = \sqrt{b^2 + r_n^2}$$

$$\text{or, } \mu_n O = b + \frac{1}{2}\frac{r_n^2}{b} \quad \ldots(2.8)$$

Substituting the values of SM_n and $\mu_n O$ in Eq. (2.6), we get:

$$a+\frac{r_n^2}{2a}+b+\frac{r_n^2}{2b}=a+b+\frac{n\lambda}{2}$$

or, $$r_n^2\left(\frac{1}{a}+\frac{1}{b}\right)=n\lambda \quad ...(2.9)$$

or, $$r_n^2=n\lambda\left(\frac{ab}{a+b}\right) \quad ...(2.10)$$

Since for a given experiment set-up, a, b and λ are constants, it follows from Eq. (2.10),

$$r_n \propto \sqrt{n}$$

i.e. *the radii of zones are proportional to the square roots of natural numbers.*

The area of the nth zone = $\pi r_n^2-\pi r_{n-1}^2$

$$=\pi\left[\frac{n\lambda ab}{a+b}-\left(\frac{(n-1)\lambda ab}{ab}\right)\right]$$

$$=\frac{\pi\lambda ab}{a+b} \quad ...(2.11)$$

The above expression is independent of the order n, implying that the *area of each zone is the same* for a given a, b and λ. But the distance of the zone from P and the obliquity increases as the order increases due to which the magnitude of μ decreases with the order of the zone. Therefore, for a positive zone plate, the resultant amplitude at O is:

$$A=\mu_1+\mu_3+\mu_5+...$$

and that for a negative zone plate is:

$$A=-(\mu_2+\mu_4+\mu_6+...)$$

which is much larger than $\frac{m_1}{2}$ when the light from all the zones is allowed to reach O. Thus, a much brighter image of S is obtained at O. This explains the focussing action of a zone plate. Thus, *the function of a zone plate is similar to that of a convex (converging) lens.*

Eq. (2.10) can also be written as:

$$\frac{1}{a}+\frac{1}{b}=\frac{n\lambda}{r_n^2} \quad ...(2.12)$$

This is similar to the lens formula:

$$\frac{1}{u}+\frac{1}{v}=\frac{1}{f} \quad ...(2.13)$$

with a and b being the object and image distances respectively. Comparing Eqs. (2.12) and (2.13), we get:

$$\frac{1}{f}=\frac{n\lambda}{r_n^2}$$

or, $$\varphi=\frac{r_n^2}{n\lambda} \quad ...(2.14)$$

Here, φ is the primary or principal or first order focal length of the zone plate. Thus, the *zone plate behaves like a convex lens with multiple foci* for a light of particular wavelength λ, the focal length depending on n and r. A convex lens has only one focal length for a fixed λ, given by $\frac{1}{f}=(\mu-1)\left(\frac{1}{R_1}-\frac{1}{R_2}\right)$, the lens maker's formula.

Multiple Foci of Zone Plate: If the point source S is located at infinity, i.e., $a=\infty$, then from Eq. (2.11). it follows that the *area of each zone is* $\pi b\lambda$. In this case, the image of S is formed at the principal focal length $b=\frac{r_n^2}{n\lambda}$.

Now, consider a point O' along the axis of the zone plate at a distance $b/3$ from the zone plate, then the *area of each half period zone* will be $\frac{\pi b\lambda}{3}$, i.e, one-third of the original area. Hence, each zone on the zone plate will contain three area elements for the image at O'. Hence, the resultant amplitude at O' is:

$$A=(\mu_1-\mu_2+\mu_3)+(\mu_7-\mu_8+\mu_9)+(\mu_{13}-\mu_{14}+\mu_{15})+...$$

$$=\left[m_1-\left(\frac{m_1+m_3}{2}\right)+m_3\right]+\left[m_7-\left(\frac{m_7+m_9}{2}\right)+m_9\right]+\ldots$$

$$=\frac{1}{2}(m_1+m_3+m_7+m_9+m_{13}+\ldots)$$

Thus, O' will be sufficiently bright, although lesser than that at O and it represents *second focal point* and its distance from P is the *second focal length* given by:

$$f_3=\frac{r_n^2}{3n\lambda} \quad ...(2.15)$$

Similarly, other foci occur on the axis at distances $\frac{b}{5},\frac{b}{7}\ldots$ from the zone plate, although the brightness of the successive foci will decrease gradually. If *a zone is occupied by even number of half period zones, they cancel each other in pair and produce zero illumination. Thus, whenever a zone is occupied by an odd number of half period zones a brightness will be produced at the corresponding focal points.* Generalising the above discussion, we can state that the focal lengths are given by:

$$f_p = \frac{r_n^2}{(2p-1)n\lambda}, \quad p = 1, 2, 3, \ldots(2.16)$$

Thus, $f_1 = \frac{r_n^2}{n\lambda}, f_2 = \frac{r_n^2}{3n\lambda}, f_3 = \frac{r_n^2}{5n\lambda}, \ldots$ are the various focal points of a zone plate, i.e., it is a *multifoci device*.

Comparison: Zone Plate vs Convex Lens

Similarities

(*i*) Both zone plate and convex lens form real images of an object on the other side of the object and the distances of object and image are connected by similar relations.

(*ii*) Focal length in both the cases depends upon wavelength of light λ and, therefore, both suffer from chromatic aberrations.

Differences

(*i*) Refraction is a physical process, which the light undergoes in case of a lens while light undergoes diffraction in case of a zone plate.

(*ii*) Image formed by a convex lens is brighter than that due to a zone plate.

(*iii*) For convex lens $\varphi_r > \varphi_v$ while for zone plate, focal length of red colour is smaller than that for violet colour $\varphi_v > \varphi_r$.

(*iv*) A convex lens is a *unifocal* device for a given wavelength of light while a zone plate is a *multifoci* device of focal lengths $\frac{r_n^2}{n\lambda}, \frac{r_n^2}{3n\lambda}, \frac{r_n^2}{5n\lambda}, \ldots$ of decreasing intensity.

Example 2.3

Find the radius of the first zone in a zone plate of focal lengths 25 cm for light of wavelength 5000 Å.

Solution:

Here, $\varphi_1 = 25$ cm, $\lambda = 5000$ Å $= 5.0 \times 10^{-5}$ cm

$r_1 = ?$

we know, $\varphi_n = \frac{r_n^2}{n\lambda}$

for the first zone, $\varphi_1 = \frac{r_n^2}{\lambda} \Rightarrow r_1 = \sqrt{f_1\lambda}$

$r_1 = \sqrt{25 \times 5.0 \times 10^{-5}} = 3.54 \times 10^{-2}\text{cm} = 0.0354 \text{ cm}$

Example 2.4

With a zone plate, for a point source of light of 6000 Å on the axis, the brightest and the second most bright images are formed at 30 cm and 6 cm respectively from the zone plate. Both the images are on the same and on the other side of the source. Calculate:

(*a*) the distance of the source from the zone plate,
(*b*) the radius of the first zone, and
(*c*) principal focal length.

Solution:

If a and b are distances of the object and the image from the zone plate, then we have:

$$\frac{1}{a} + \frac{1}{b} = \frac{(2p-1)n\lambda}{r_n^2}$$

where r_n is the radius of the nth zone $p = 1, 2, 3, \ldots$

(*a*) For the brightest image $p = 1$ and $b = 30$ cm

$$= \frac{n\lambda}{r_n^2} \quad \ldots(i)$$

for the second most bright image, $p = 2$ and $b = 6$ cm

$$\therefore \frac{1}{a} + \frac{1}{6} = \frac{3n\lambda}{r_n^2} \quad \ldots(ii)$$

Dividing (*ii*) by (*i*), we get:

$$\frac{\frac{1}{a} + \frac{1}{6}}{\frac{1}{a} + \frac{1}{30}} = 3$$

$$\text{or, } \frac{1}{a} + \frac{1}{6} = \frac{3}{a} + \frac{3}{30}$$

$$\text{or, } \frac{3}{a} - \frac{1}{a} = \frac{1}{6} - \frac{3}{30}$$

$$\text{or, } \frac{2}{a} = \frac{2}{30} \Rightarrow a = 30 \text{ cm}$$

(*b*) From (*i*), the radius of the first zone, $n = 1$, is given by:

$$r_1^2 = \lambda\left(\frac{30a}{30+a}\right)$$

$$= 6.0\times10^{-5}\times\left(\frac{30\times30}{30+30}\right)$$

$$= \frac{6\times900}{60}\times10^{-5} = 9.0\times10^{-4}\text{ cm}$$

or $r_1 = 3\times10^{-2}$ cm = 0.03 cm

(*c*) Principal focal length, $f_1 = \dfrac{ab}{a+b}$

$$= \frac{30\times30}{30+30} = \frac{900}{60} = 15\text{ cm}$$

Example 2.5

The central circle of a zone plate has a radius of 0.07 cm. Light of wavelength 5000 Å coming from (*i*) an object at infinity, (*ii*) an object 147 cm away from the plate falls on it. Find the position of the principal image in each case.

Solution:

Here, r_1 = 0.07 cm, λ = 5000 Å = 5.0×10^{-5} cm

we know for a zone plate

$$\frac{1}{a}+\frac{1}{b}=\frac{1}{f_p},\ f_p=\frac{r_n^2}{(2p-1)n\lambda}$$

for principal image, $p = 1,\ f_p = f_1 = \dfrac{r_n^2}{n\lambda}$

for the central circle, $n = 1$

$$\therefore \varphi_1 = \frac{(0.07)^2}{5.0\times10^{-5}} = 98\text{ cm}$$

(*i*) $a = \infty$, b =?

$$\therefore \frac{1}{\infty}+\frac{1}{b} = \frac{1}{98}$$

or, b = 98 cm

(*ii*) a = 147 cm, b =?

$$\frac{1}{147}+\frac{1}{b} = \frac{1}{98}$$

$$\text{or, } \frac{1}{b} = \frac{1}{98}-\frac{1}{147}=\frac{3-2}{294}=\frac{1}{294}$$

or, b = 294 cm

2.5 Fresnel diffraction at a straight edge

Let A be a sharp and straight edge of an opaque obstacle AB (e.g. edge of a razor blade) and S is a narrow rectangular slit both parallel to each other and perpendicular to the plane of the paper (Fig. 2.9). Let the slit be illuminated by a monochromatic source of light of wavelength λ. XY is the incident cylindrical wavefront. O is a point on the screen perpendicular to the plane of paper and SAO is a straight line perpendicular to the screen. Thus, below the point O on the screen, is the geometrical shadow of the obstacle. We want to understand the distribution of light intensity on the screen.

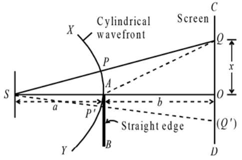


Fig. 2.9: Diffraction at a straight edge.

If the rectilinear propagation of light is true, i.e., according to the laws of geometrical optics, below the point O, there should be complete darkness and above O, the illumination should start abruptly or there should be complete illumination. But experimentally, it is observed (Fig. 2.11):

(*i*) Below O, the intensity falls continuously and rapidly as we move into the geometrical shadow and after a short distance, there is complete darkness. In other words, there exists a slight illumination in geometrical shadow, which decreases rapidly before complete darkness.

(*ii*) Just above O, alternate dark and bright bands of unequal width and varying intensity parallel to straight edge are observed. The bands become closer and less distinct as we move away from O and ultimately we get a region of uniform illumination.

We shall now try to understand the observed variation in illumination on the basis of Fresnel's half period zones.

Intensity Variation Inside the Geometrical Shadow: As shown in Fig. 2.9, incident cylindrical waveform at the straight edge. For the point O on the screen, the pole of the wavefront is at A and so only the upper half of the wavefront is exposed so that the displacement at *O*, is only one half of what it would have been if whole of the wavefront were effective. Consider this upper half wavefront to be divided into half period strips (Fig. 2.10). Let the amplitudes of the waves reaching at *O* from the respective half period strips be μ_1, μ_2, m_3, ... As the successive strips are in opposite phase, the resultant amplitude of the waves is:

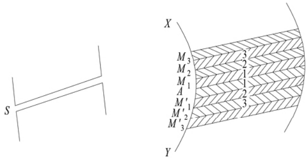


Fig. 2.10

$$A = \mu_1 - \mu_2 + \mu_3 - \mu_4 + \ldots$$

$$= \frac{m_1}{2}$$

$$\therefore \text{Intensity at } O = \left(\frac{m_1}{2}\right)^2 = \frac{m_1^2}{4}$$

As we move below the point *O*, inside the geometrical shadow, say at *Q*′, the pole of wavefront moves to μ_1, the boundary of the first half period strip and so the wavelets from the first half period are obstructed and the displacement at *Q*′ is given by:

$$A = -m_2 + \mu_3 - \mu_4 + \mu_5 + \ldots$$

$$-\frac{m_2}{2}$$

$$\therefore \text{Intensity at } Q' = \left(-\frac{m_2}{2}\right)^2 = \frac{m_2^2}{4}$$

which is definitely smaller than that at *O*.

Similarly, as we move downward in the geometrical shadow, the second, third,... half period strips are obstructed as the pole of the wavefront shifts downwards. Thus, the displacements reduce to $\frac{m_3^2}{4}, \frac{m_4^2}{4} \ldots$. Since $m_1 > \mu_2 > \mu_3 > \ldots$, therefore, as the point of consideration moves inside the geometrical shadow, the intensity of the light falls off rapidly and soon diminishes to zero.

Diffraction Bands Outside the Geometrical Shadow: As we move from *O* towards the point *Q* distant *x* from *O*, the pole of the wavefront also moves up towards *P* and one after another, the first, the first two, the first three, etc., half period strips or elements of the lower half also get exposed so that the illumination of the point *Q* will be due to the complete upper half of the wavefront above *P* together with that due to the half period elements contained in the part *PA* of the wavefront. If the number of half period elements in *PA* is even, the displacements from the alternate elements mutually cancel each other in pairs and the intensity at *Q* is due to the upper half of the wavefront above *P*. If the number of half period elements contained in *PA* is odd, one half period element will be left uncancelled and the effect of this at *Q*, will be added so that the intensity will be greater.

When only one half period element is contained in *PA* for a position *Q* little above *O*, then the displacement for this point is equal to $\frac{m_1}{2} + m_1 = \frac{3m_1}{2}, \frac{m_1}{2}$ being the displacement due to the upper exposed part of the wavefront. Similarly, for a point further up, when *PA* contains two half period elements, the displacement is $\frac{m_1}{2} + m_1 - m_2 = \frac{3m_1}{2} - m_2$, and so on. The corresponding intensities are $\left(\frac{m_1}{2}\right)^2, \left(\frac{3m_1}{2}\right)^2, \left(\frac{3m_1}{2} - m_2\right)^2,$ etc. It is obvious that the second intensity is higher than the first but the third intensity is lower than the second. Thus, as we move above *O*, the intensity varies, alternatively becoming higher and lower than its value of *O*, i.e., $\left(\frac{m_1}{2}\right)^2$.

It is higher if *PA* contains an odd number of half period elements and lower, if *PA* contains an even number of elements.

Positions of Maxima and Minima in the Diffraction Pattern: To find the positions of the diffraction bands and their widths, we consider a point *Q* outside the geometrical shadow. Join *SQ* passing through *P* at the wavefront and join *AQ*. The number of half period elements contained in *PA* is equal to the number of half wavelengths in the path difference *AQ* – *PQ*.

$$AQ = \sqrt{AO^2 + OQ^2}$$

$$= \sqrt{b^2 + x^2} \quad (AO = b, OQ = x)$$

$$= b\left(1 + \frac{x^2}{b^2}\right)^{\frac{1}{2}}$$

Using binomial theorem and neglecting the terms higher than x^2, we get:

$$AQ = b + \frac{1}{2} \cdot \frac{x^2}{b} \quad \ldots(2.16a)$$

$$PQ = SQ - SP$$

$$= \sqrt{(a+b)^2 + x^2} - a$$ (Assuming SP to be nearly equal to a)

$$= (a+b)\left(1 + \frac{x^2}{(a+b)^2}\right)^{\frac{1}{2}} - a$$

Using binomial theorem and neglecting terms of order higher than x^2, we get:

$$PQ = a + b + \frac{1}{2}\cdot\frac{x^2}{(a+b)} - a$$

$$= b + \frac{x^2}{2(a+b)} \quad \text{...(2.17)}$$

∴ Path difference, $AQ -$

$$PQ = \left(b + \frac{x^2}{2b}\right) - \left(b + \frac{x^2}{2(a+b)}\right)$$

$$= \frac{x^2}{2}\left(\frac{1}{b} - \frac{1}{a+b}\right) = \frac{ax^2}{2b(a+b)} \quad \text{...(2.18)}$$

For a maximum at Q, this path difference must be equal to an odd multiple of $\frac{\lambda}{2}$. Thus,
For brightness:

$$\frac{ax^2}{2b(a+b)} = (2n+1)\frac{\lambda}{2} \quad n = 0, 1, 2, \ldots$$

or, location of bright bands

$$x_n = \sqrt{\frac{b(a+b)(2n+1)\lambda}{a}} \quad \text{...(2.19)}$$

For a minimum at Q, the path difference should be equal to an even number multiple of $\frac{\lambda}{2}$. Thus,

for darkness: $$\frac{ax^2}{2b(a+b)} = 2n\cdot\frac{\lambda}{2} \quad n = 1, 2, \ldots$$

or, locations of dark bands:

$$x_n = \sqrt{\frac{2b(a+b)n\lambda}{a}} \quad \text{...(2.20)}$$

It is evident from the above expression that for bright bands, $x\alpha\sqrt{(2n+1)}$ while for dark bands, $x\alpha\sqrt{2n}$. This implies that the *width of the bands are not equal,* i.e., *bright and dark bands are not equally spaced,* unlike in the case of interference bands where the bands are of equal width. Also, the intensity of the bright bands are not the same, unlike the interference maxima.

Diffraction Pattern: The observed diffraction pattern on the screen due to a straight edge is shown in Fig. 2.11. It is clear from the pattern that inside the geometrical shadow, the intensity falls off rapidly without any maxima and minima and within a short distance of the shadow of the edge, there is complete darkness. At the edge of the geometrical show, i.e., at the point O, the intensity is $I_o/4$. Outside the geometrical shadow, i.e., above the point O, there is a system of alternate dark and bright bands of gradually diminishing intensity and contrast and finally a uniform illumination is obtained. The shadow cast by the edge is not sharp, showing that rectilinear propagation of light is not completely true.

By measuring the locations of first band corresponding to the location O and other bands with the help or a travelling microscope, wavelength of the light used can be determined using Eq. (2.20).

Diffraction Bands with White Light: If white light is used instead of monochromatic light, a few coloured bands running parallel to the shadow of the edge are observed in the illuminated region above O as the locations of bands depend upon the wavelength of the light. Due to overlapping, the clarity of the bands is lost subsequently.

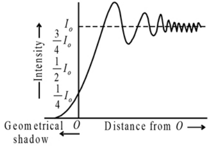


Fig. 2.11: Intensity variation in the diffraction pattern of a straight edge.

Example 2.6

A diffraction pattern is obtained with a straight edge using light of wavelength 5500 Å. The separation between the edge and source slit is 15 cm while that between the edge and the eyepiece is 3 m. Find the position of the first maximum.

Solution:

Here, $\lambda = 5500$ Å $= 5.5 \times 10^{-5}$ cm

a = 15 cm, b = 3 m = 300 cm

Locations of brightness for a straight edge,

$$x_n = \sqrt{\frac{b(a+b)(2n+1)\lambda}{a}}$$

For the first maximum

$$n = 1,\ x_1 = \sqrt{\frac{3b(a+b)\lambda}{a}}$$

$$x_1 = \sqrt{\frac{3\times 300\times 315\times 5.5\times 10^{-5}}{15}}$$

= 1.019 cm from the edge of the geometrical shadow.

Example 2.7

In an experiment with a straight edge, the narrow source slit is illuminated with light of 6000 Å. The separation between the source slit and edge and that between the edge and screen is 0.2 m and 2 m respectively. What is the separation between first and second dark bands?

Solution:

Here, λ = 6000 Å = 6.0 × 10^{-5} cm

a = 0.2 m = 20 cm

b = 2 m = 200 cm

Location for dark bands, $$x_n\sqrt{\frac{2b(a+b)n\lambda}{a}}$$ n = 1, 2, 3,...

Location of 1st dark band,

$$x_1 = \sqrt{\frac{2\times 200\times 220\times 6\times 10^{-5}}{20}}$$

= 0.5138 cm

Location of 2nd dark band,

$$x_2 = \sqrt{\frac{2\times 200\times 220\times 12\times 10^{-5}}{20}}$$

= 0.7266 cm

Separation between first and second dark bands

= 0.7266 – 0.5138

= 0.2128 cm.

2.6 Cornu's spiral

It is an elegant method for investigating intensity distribution in Fresnel's diffraction patterns, as you will realize in the following discussions. We know that to find the effect at a point due to an incident wavefront, *Fresnel's method consists in dividing the wavefront into half period zones* and the path difference between the secondary waves from any two corresponding points of two neighbouring zones is $\frac{\lambda}{2}$.

In Fig. 2.12. S is a point source of light and XY is the incident spherical wavefront. With reference to the point O, P is the pole of the wavefront. Let a and b be the separation of the points S and O respectively from the pole P of the wavefront. With O as centre, a sphere is drawn such that it touches the incident wavefront at P. The path difference between the waves taking the paths SAO and SPO is given by:

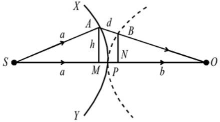


Fig. 2.12

$$\Delta = SA + AO - SPO$$
$$= SA + AO - (SP + PO)$$
$$= a + AB + b - (a + b) = AB$$

For large values of a and b, AM and BN can be taken to be approximately equal and the above expression for the path difference can be put as:

$$\Delta = AB = MP + PN$$

From the geometry of a circle, we have:

$$MP = \frac{AM^2}{2SP} = \frac{h^2}{2a},\ \text{and}$$

$$PN = \frac{BN^2}{2PO} = \frac{h^2}{2b}.$$

$$\therefore \text{Path difference } \Delta = \frac{h^2}{2a} + \frac{h^2}{2b} = \frac{h^2(a+b)}{2ab} \quad \text{... (2.21)}$$

Let AM be the radius of the nth half period zone so that this path difference is equal to $\frac{n\lambda}{2}$ in accordance with Fresnel's method of constructing the half period zones. Thus,

$$\frac{h^2(a+b)}{2ab} = \frac{n\lambda}{2} \quad \text{...(2.22)}$$

The resultant amplitude at an external point due to the wavefront can be obtained by the following method:

Let the first half period zone be divided into eight substrips and each of these substrips is represented by a vector from O to μ_1 as shown in Fig. 2.13. There is continuous phase change due to continuous increase in the obliquity factor from O to μ_1. The resultant amplitude at the external point due to the first half period zone is given by $OM_1(= \mu_1)$. Similarly, if the process is continued, we obtain the vibration curve $\mu_1\mu_2$, corresponding to the second half period zone. The resultant amplitude at the considered point due to the first two half period zones is given by $OM_2(= A)$.

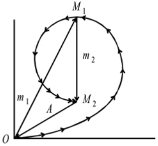


Fig. 2.13

If instead of eight substrips, each half period zone is divided into large number of substrips of infinitesimally small width for the whole of the wavefront, we obtain the vibration curve for the whole wavefront, which is a spiral as shown in Fig. 2.14. J_1 and J_2 correspond to the two extremities of the wavefront and μ_1, μ_2, etc. refer to the edges of the first, second, etc. of the respective half period zones.

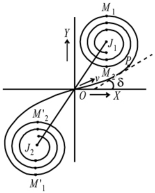


Fig. 2.14: Cornu's spiral.

Similarly, M'_1, M'_2, etc. refer to the edges of the first, second, etc. half period zones of the lower portion of the wavefront. The spiral, shown in Fig. 2.14, is called *Cornu's spiral which is basically the vibration curve for an incident wavefront. The characteristic of this curve is that for any point P on the curve, the phase lag* d *is directly proportional to the square of the distance (v) of the point from the origin and the distance is measured along the curve.* For a path difference Δ, the corresponding phase difference Δ is given by:

$$d = \frac{2\pi}{\lambda}\cdot d$$

$$= \frac{2\pi}{\lambda}\cdot\frac{h^2(a+b)}{2ab}$$

$$= \frac{\pi}{2}\cdot\left[\frac{2h^2(a+b)}{ab\lambda}\right]$$

$$\text{or, } \Delta = \frac{\pi}{2}\cdot v^2 \quad \text{...(2.23)}$$

where, v is a dimensionless variable given by

$$v^2 = \frac{2h^2(a+b)}{ab\lambda}$$

$$\text{or, } v = h\sqrt{\frac{2(a+b)}{ab\lambda}} \quad \text{...(2.24)}$$

We can use Cornu's spiral to approximate the effect at any point in any diffraction problem, irrespective of the value of a, b and λ.

Fresnel Integrals: For any point on the Cornu's spiral, the x and y coordinates can be represented by two integrals known as *Fresnel's integrals*. Let us consider a point P on the spiral. The distance of the point P along the curve from the origin is v, Let the tangent to the curve at P makes an angle Δ with the x-axis, where Δ corresponds to the phase change from O to P.

For a small displacement dv of the point along the curve, let the corresponding changes in the coordinates of the point be dx and dy.

Then,

$$dx = \Delta v\cos\Delta$$

$$\text{and, } dy = \Delta v\sin\Delta$$

Putting the value of Δ from Eq. (2.23) in the above expression, we get:

$$dx = \cos\left(\frac{\pi v^2}{2}\right)dv$$

$$\text{and, } dy = \sin\left(\frac{\pi v^2}{2}\right)dv$$

Therefore, the coordinates x and y of the Cornu's spiral are given by

$$x = \int dx = \int_0^v \cos\left(\frac{\pi v^2}{2}\right) dv \qquad ...(2.25)$$

$$\text{and, } y = \int dy = \int_0^v \sin\left(\frac{\pi v^2}{2}\right) dv \qquad ...(2.26)$$

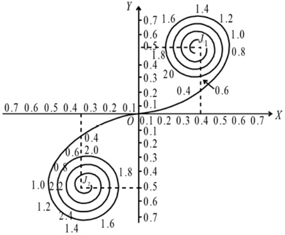


Fig. 2.15: Cornu's spiral

These two integrals are Fresnel's integrals, when for different value of v(0 to ∞), the integrals are evaluated (for v = 0, x = 0, y = 0 and for v = ∞, x = 0.5, y = 0.5) and a curve is plotted in the Cartesian coordinates, Cornu's spiral is obtained (Fig. 1.15). In fact, sometimes Cornu's spiral can also be introduced as a curve obtained by plotting Fresnel's integrals.

Properties of Cornu's Spiral

The variation of intensity in diffraction pattern can be explained elegantly with the help of Cornu's spiral.

In Fig. 2.14, OJ_1 represents the vibration curve for the upper half of the wavefront and OJ_2 that of the lower half of the wavefront. If no part of the wavefront is obstructed, i.e., the entire wavefront is exposed, the resultant amplitude at a point is represented by J_1J_2. The radius vector r drawn from the origin O to any point on the curve represents the resultant vibration due to the corresponding portion of the wavefront. As $r^2 = x^2 + y^2$, r^2 *represents the intensity of illumination and* r *represents the amplitude of the resultant vibration,* following are the important properties of the spiral:

(*i*) The intensity is proportional to the square of the resultant amplitude, that is,

$$I = k(x^2 + y^2)$$

When the entire wavefront is exposed, then $v \to \infty$ and the values of the integrals are:

$$x = \int_0^\infty \cos\left(\frac{\pi v^2}{2}\right) dv = \frac{1}{2}$$

$$\text{and, } y = \int_0^\infty \sin\left(\frac{\pi v^2}{2}\right) dv = \frac{1}{2}$$

So, for the point J_1 in Fig. 2.15, the x and y coordinates are $\left(\frac{1}{2}, \frac{1}{2}\right)$.

Similarly, for the point, J_2 on the lower half of the spiral, the coordinates are $\left(-\frac{1}{2}, -\frac{1}{2}\right)$.

(*ii*) The slope of the curve at any point (x, y) is:

$$\tan\Delta = \frac{dy}{dx} = \frac{dy}{dv} + \frac{dx}{dv} = \frac{\sin\left(\frac{\pi v^2}{2}\right)dv}{\cos\left(\frac{\pi v^2}{2}\right)dv} = \tan\left(\frac{\pi v^2}{2}\right) \qquad ...(2.27)$$

when v = 0, Δ = 0.

This implies that the curve is parallel to x-axis at the origin.

(*iii*) The distance between two neighbouring points (x, y) and ($x + dx$, $y + dy$) is given by:

$$dv = \sqrt{(dx)^2 + (dy)^2}$$

Differentiating Eq. (2.27), we get:

$$d\Delta = \frac{2\pi v}{2} \cdot dv$$

$$\text{or } \frac{dv}{d\delta} = \frac{1}{\pi v} \qquad ...(2.28)$$

where, $\frac{dv}{d\delta}$ measures the radius of curvature of the spiral at the considered point.

We find:

$$\frac{dv}{d\delta} \propto \frac{1}{v}$$

which shows that with the increase in the value of v, the radius of curvature of the curve gradually decreases and takes the shape of a spiral. Finally, when $v \to \infty$, the curve ends at a point J_1 or J_2

(*iv*) It may be noted that for the values of v on the curve equal to $1, \sqrt{2}, \sqrt{3}, \ldots$ correspond to points for which the phase differences are equal to $\frac{\pi}{2}, \pi, \frac{3\pi}{2}, \ldots$ respectively and at each of these points, the tangent is parallel to one or the other axes *OX* and *OY*.

Applications of Cornu's Spiral

As an illustration of the usefulness of Cornu's spiral in explaining the observed diffraction pattern, we take the following example. From the preceding discussions, it is clear that each half of the incident cylindrical wavefront gives rise to a spiral curve and these spirals are identical in shape and position. Thus, OJ_1 represents the effect due to the upper half of wavefront and OJ_2 that due to the lower half of wavefront, while J_1J_2 given the effect of the whole of the wavefront.

Straight Edge: We have discussed the variation of intensity due to diffraction at a straight edge in Section 2.5. At the edge of the geometrical shadow *O*, the amplitude is represented by OJ_1. If we consider a point *Q* in the illuminated (outside geometrical shadow) region, the resultant amplitude is due to the complete half wavefront *PX* and portion *PA* of the lower half of the wavefront. The effect of the upper half wavefront is represented by OJ_1 and the effect due to the portion *PA* by (say) *OM* (μ can be M'_1, M'_2, etc.) (Fig. 2.16). Hence, the resultant effect is represented by $J_1\mu$. The exact location of μ (at M'_1, M'_2, etc.) depends upon the number of half period zones contained in *PA*. Hence, as $J_1\mu$ passes through a maximum and minimum for every convolution of μ around it, corresponds a bright and dark band on the screen. In figure J_1b', J_1d' correspond to maxima and J_1c' corresponds to minima. Thus, in the illuminated region, alternate bright and dark bands parallel to the length of the edge are observed on the screen.

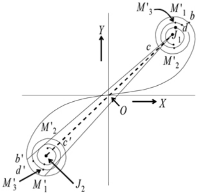


Fig. 2.16

For a point below *O* in the geometrical shadow, the lower half of the wavefront and a portion of the upper half of the wavefront are cut off and the tail of the amplitude vector moves to the right of *O*. The amplitude slowly decreases and becomes zero as the tail approaches J_2. Thus, we see that in the region of geometrical shadow, the intensity falls slowly and continuously.

Similarly, we can explain diffraction patterns in all the other discussed cases of Fresnel diffraction.

2.7 Fraunhofer diffraction at a single slit

Consider a narrow slit *AB* of width *a* (Fig 2.17). Let a plane wavefront of monochromatic light wavelength λ be incident, normally on this slit. According to Huygen's principle, each point of the slit becomes a source of secondary wavelets, all originating in the same phase. These wavelets spread out in all directions. As the width of the slit is of the order of the wavelength of light, therefore, diffraction of light waves takes place on emerging through the slit. The diffracted rays are focussed on the screen by a convex lens λ.

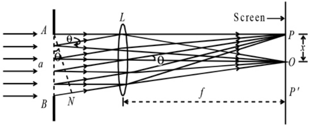


Fig. 2.17: Fraunhofer diffraction at a single slit.

The diffraction pattern obtained on the screen consists of a central bright band surrounded by alternate bright and dark bands of progressively reducing contrast. Here, it is important to note, that had the rectilinear propagation of light been completely true, we should have obtained a perfect image of the slit at *O*.

Central Maxima: The secondary waves reaching *O* from the equidistant points from the centre if the slit in the upper half of the slit and lower half of the slit, meet at *O* after travelling the same distance. That is, they meet at *O* in the same phase and so reinforce each other giving a central bright at *O*.

Dark Bands: The secondary waves diffracted at an angle θ, converge at point *P* (Fig. 2.17). The intensity at point *P* will depend on the path difference of waves reaching *P*. A perpendicular (*AN*) from the point *A* is drawn on the ray diffracted at an angle θ from the point *B*. The path difference between the waves reaching *P* from *A* and *B* is given by:

$$BN = AB\sin\theta = a\sin\theta$$

If path difference $a\sin\theta = \lambda$, the point *P* will be a point of minimum intensity. This is because if we consider the whole wavefront at the slit to be divided into two equal halves, then the path difference between the waves originating in the first half and the corresponding point in the second half will be $\frac{\lambda}{2}$ (or a phase difference of π) when they meet *P*. Thus, all these waves will interfere destructively and so the point P will be a point of minimum intensity. The points for which the path difference is:

$$a\sin\theta = \lambda$$

form the *first band or the first minimum*. The points for which the path difference is:

$$a\sin\theta = 2\lambda$$

form the *second dark band* or the second minimum. In general, the points on the screen for which the path difference is:

$$a\sin\theta_n = 2n \cdot \frac{\lambda}{2} = n\lambda \quad \text{...(2.29)}$$

form the *nth dark band or the* nth minimum, while q_n gives the direction of the nth minimum.

Secondary Maxima or Bright Bands: If for any point on the screen, θ is such that the path difference BN is:

$$a\sin\theta = \frac{3\lambda}{2}$$

then such a point will be the position of first secondary maximum. In this case, the wavefront at the slit can be considered to be divided into three equal parts. The waves from the corresponding points from the two successive parts will reach the point with a path difference of $\frac{\lambda}{2}$ and so, will interfere destructively cancelling out each other's effect. However, the waves from the third part produce a diminished illumination and so the secondary maximum is formed.

Similarly, if the path difference BN for a point on the screen is:

$$a\sin\theta = 5 \cdot \frac{\lambda}{2}$$

then the point will be the position of the *second secondary maximum*. In general, for the *nth secondary maximum*.

$$a\sin\theta_n = (2n+1) \cdot \frac{\lambda}{2}, \quad n = 1, 2, 3, \ldots \quad \text{...(2.30)}$$

Thus, the diffraction pattern due to a single slit consists of a central bright maximum surrounded by secondary maxima and minima. This intensity distribution in the diffraction pattern is depicted in Fig. 2.17. The central maximum is broad if the slit is narrow. The intensity of the secondary maxima is much less compared to that of central maximum. It can be shown that the intensity of the first, second and third secondary maxima is approximately $\frac{I_o}{22}, \frac{I_o}{61}$ and $\frac{I_o}{121}$ respectively, where I_o is the intensity of the central maximum.

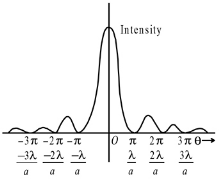


Fig. 2.18: Intensity distribution.

Width of the Central Maximum: Let P be the position of the first minimum and $OP = x$ (Fig. 2.17), then,

$$a\sin\theta = \lambda$$

$$\text{or, } \sin\theta = \frac{\lambda}{a}$$

If θ is small, then sinθ = θ so that:

$$\theta = \frac{\lambda}{a}$$

If the focal length of the lens λ is φ and the lens is held close to the slit, then,

$$\sin\theta = \frac{x}{f}$$

For small θ, sinθ = θ so that $\theta = \frac{x}{f}$

$$\therefore \frac{x}{f} = \frac{\lambda}{a}$$

$$\text{or, } x = \frac{f\lambda}{a} \quad \text{...(2.31)}$$

Thus, the width of the central maximum is proportional to the wavelength λ of the light used and inversely proportional to the width a of the slit. Hence, a narrow slit produces a broad central maximum, while if the slit is wide, the central maximum is narrow.

Diffraction Pattern with White Light: If instead of monochromatic light, white light is used, then the central maximum remains white as all seven wavelengths meet there in the same phase reinforcing each other. The first minimum and the first secondary maximum is first formed by the violet colour due to its shortest wavelength while the last is due to red colour due to its longest wavelength. Thus, a coloured pattern is formed. However, after the first few coloured bands due to overlapping, the clarity of the bands is lost.

Mathematical Treatment for the Intensity Distribution

We have discussed the physical reasons for the observed intensity variation in the diffraction pattern due to a single slit. Now, we investigate the same mathematically.

The incident wavefront on the slit AB (Fig. 2.17) can be imagined to be divided into large number of infinitesimally small strips. We have seen that the path difference between the secondary waves originating from the extreme points A and B is $a\sin\theta$, where a is the width of the slit and θ is the angle of diffraction. The corresponding phase difference is given by:

$$2\alpha = \frac{2\pi}{\lambda}\cdot a\sin\theta \quad \ldots(2.32)$$

As the incident wavefront on the slit AB is plane, the amplitude of the wave from each strip can be taken to be the same but their initial phase will differ by a constant amount. The resultant amplitude due to all the small strips can be obtained by the vector polygon method. The phase difference increases regularly by a small amount from strip to strip and the vibration polygon tends to be a circular arc OM (Fig. 2.19) OX gives the direction of the initial vector and NM the direction of the final vector due to the secondary waves. P is the centre of the circular arc.

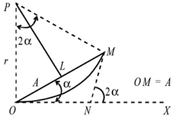


Fig. 2.19

$\angle MNX = 2\alpha$

$\therefore \angle OPM = 2\alpha$

In ΔOPL, $\sin\alpha = \dfrac{OL}{r}$, r being the radius of the circular arc.

or, $OL = r\sin\alpha$

$\therefore$ Chord $OM = 2OL = 2r\sin\alpha \quad \ldots(2.33)$

The length of the arc OM is directly proportional to the width of slit

$\therefore$ length of the arc $OM = ka$

where k is a proportionality constant, while a is the width of the slit,

$$\text{Also, } 2\alpha = \frac{\text{Arc } OM}{\text{radius}} = \frac{k\alpha}{r}$$

$$\text{or, } 2r = \frac{ka}{\alpha} \quad \ldots(2.34)$$

Putting this value of $2r$ in Eq. (2.33), we get,

$$\text{Chord } OM = \frac{ka}{\alpha}\cdot\sin\alpha$$

But $OM = A$, where A is the resultant amplitude.

$$\therefore A = (ka)\cdot\frac{\sin\alpha}{\alpha} \quad \ldots(2.35)$$

Thus, the resultant amplitude of vibration at a point on the screen is given by $A_o\dfrac{\sin\alpha}{\alpha}$ and the intensity I at the point is given by:

$$I = A^2 = A_0^2\frac{\sin^2\alpha}{\alpha^2} = I_o\cdot\left(\frac{\sin\alpha}{\alpha}\right)^2 \quad \ldots(2.36)$$

Thus, the intensity at any point on the screen is proportional to $\left(\dfrac{\sin\alpha}{\alpha}\right)^2$, where $\alpha = \dfrac{\pi}{\lambda}\cdot a\sin\theta.$

I_o is the intensity of the central maximum.

Diffraction Maxima and Minima: Since the intensity at any point on the screen is a function of α and hence of θ, it follows that a series of maximum and minimum values are obtained.

***(i)* Central Maximum:** For the point O on the screen (Fig. 2.17)

$q = 0$

and $\alpha = 0$

As $a \to 0$, the value $\dfrac{\sin\alpha}{\alpha}$ becomes equal to 1. Hence, the intensity at O is:

$$I = I_0 = \left(\frac{\sin\alpha}{\alpha}\right)^2 = I_0$$

Thus, at O the central maximum is formed with the maximum intensity I_o.

***(ii)* Minima:** The intensity is minimum (zero) when,

$\sin\alpha = 0$

or, $\alpha = \pm n\pi$, where n is an integer.

$$\text{But } \alpha = \frac{\pi}{\lambda}\cdot a\sin\theta$$

$$\frac{\pi}{\lambda} \cdot a \sin\theta = \pm n\pi$$

or, $a\sin\theta = \pm n\lambda$ ...(2.37)

where n = 1, 2, 3, ... gives the directions of first, second, third,... order minima. It is important to remember that n = 0 is not admissible In Eq. (2.37) as it gives θ = 0 which corresponds to the central maximum

Thus, for α = ±π, ±2π, ±3π, ..., the intensity

$$I = I_0\left(\frac{\sin\alpha}{\alpha}\right)^2 = 0$$

It also follows from Eq. (2.37) that for $\sin\theta = \pm\frac{\lambda}{2}, \pm\frac{2\lambda}{2}, \pm\frac{3\lambda}{a}, \ldots$ we obtain the first, second-third ... order minima.

***(iii)* Secondary Maxima:** In addition to the central maximum for θ = 0, there are secondary maxima of decreasing intensity between the minima. To determine the positions of the secondary maxima, let us differentiate Eq. (2.36 with respect to α and equate it to zero, i.e.,

$$\frac{dI}{d\alpha} = 2I_o \frac{\sin\alpha}{\alpha}\left(\frac{\alpha\cos\alpha - \sin\alpha}{\alpha^2}\right) = 0$$

Thus, either sinα = 0 or αcosα − sinα = 0. We know, except α = 0, sinα = 0 gives the positions of minima.

or, α = tanα

This equation can be solved graphically by plotting the curves for y = α and y = tanα. The points of intersection gives the roots of equation α = tanα and hence the values of α for maxima. As shown in Fig. 2.20, y = α is straight line making an angle of 45° with the axes and y = tanα is a discontinuous curve of infinite number of branches at intervals α = π with asymptotes at:

$$\alpha = \pm\frac{\pi}{2}, \pm\frac{3\pi}{2}, \pm\frac{5\pi}{2}, \ldots$$

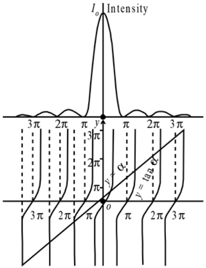


Fig. 2.20

After a careful observation of the points of intersection of the two curves; we find that except for the first root α = 0, the values of α for a maximum value approach to $(2n+1)\frac{\pi}{2}, n = 1, 2, 3, \ldots$. For all practical purposes, maxima occurs at:

$$\alpha = 0, \frac{3\pi}{2}, \frac{5\pi}{2}, \frac{7\pi}{2}, \ldots$$

or more exactly, α = 0, 1.430π, 2.459π, 3.417π, ...

Substituting these value of α in the expression

$$I = I_o \frac{\sin^2\alpha}{\alpha^2},$$ we get the following cases.

(*i*) for central maximum α = 0

$$\therefore I = I_o \frac{\sin^2\alpha}{\alpha^2} = I_o$$

(*ii*) for the first secondary maximum $\alpha = \frac{3\pi}{2}$

$\therefore I =$

$$I_o\left(\frac{\sin\frac{3\pi}{2}}{\frac{3\pi}{2}}\right)^2 = I_o\left(-\frac{1}{\frac{3\pi}{2}}\right)^2 = \frac{4}{9\pi^2}I_o = \frac{I_o}{22}$$

(*iii*) for the second secondary maximum, $\alpha = \frac{5\pi}{2}$

$I =$

$$I_o\left(\frac{\sin\frac{5\pi}{2}}{\frac{5\pi}{2}}\right)^2 = I_o\left(\frac{1}{\frac{5\pi}{2}}\right)^2 = \frac{4I_o}{25\pi^2} = \frac{I_o}{61}$$

and so on.

Thus, the intensities of maxima decrease rapidly and if the intensity of central maximum is taken as unity, the intensities of the first, second, third ... secondary maxima are $\frac{1}{22}:\frac{1}{61}:\frac{1}{121}:\ldots$ respectively.

Example 2.8

Determine the angular separation between the central maximum and first order minimum of the diffractions pattern due to a single of width 0.25 mm, when light of wavelength 5890 Å is incident normally on the slit.

Solution:

Here, a = 0.25 mm = 0.25 × 10^{-3} m, λ = 5890 Å = 5890 × 10^{-10} m.

∴ Angular separation θ between the central maximum and the first order minimum is given by:

$$\sin\theta = \frac{\lambda}{a} = \frac{5890\times10^{-10}\text{m}}{0.25\times10^{-10}\text{m}} = 0.02356$$

Since sinθ is very small, therefore, we can approximate

sinθ . θ

∴ θ = 0.02356 radian

or, θ = 0.135° = 8.1 minute

Example 2.9

A slit 4.0 cm wide is irradiated with microwaves of wavelength 2.0 cm. Find the angular spread of central maximum assuming normal incidence on the slit.

Solution:

Here, a = 4.0 cm = 4.0 × 10^{-2} m, λ = 2.0 cm = 2.0 × 10^{-2} m

for central maximum,

$$\sin\theta = \frac{\lambda}{a} = \frac{2.0\times10^{-2}}{4.0\times10^{-2}} = 0.5$$

∴ θ = 30°

Thus, the angular spread of central maximum is ± 30° on either side of the central point.

Example 2.10

Find the half angular width of the central bright maximum in the Fraunhofer diffraction pattern of a slit of width 12 × 10^{-7} m when the slit is illuminated by monochromatic light of wavelength 6000 Å.

Solution:

We know $\sin\theta = \frac{\lambda}{a}$, where θ is the half angular width of the central maximum.

Given: a = 12 × 10^{-7} m, l = 6000 Å = 6.0 × 10^{-7} m

$$\therefore \sin\theta = \frac{\lambda}{a} = \frac{6.0\times10^{-7}}{12\times10^{-7}} = 0.5 = 30°.$$

Thus, the half angular width of the central bright maximum is 30°.

Example 2.11

A screen is placed 2 m away from the lens to obtain the diffraction pattern in the focal plane of the lens in a single slit diffraction experiment. Find the slit width if the first minima lies 5 mm on either side of the central maximum, when plane light waves of wavelength 5000 Å are incident on the slit.

Solution:

Here, φ = 2 m, x = 5 × 10^{-3} m

l = 5000 Å = 5.0 × 10^{-7} m, n = 1

Using $\sin\theta = \frac{n\lambda}{a}$, we have

$$\left[\therefore \sin\theta = \frac{x}{f} = \frac{5\times10^{-3}}{2} = 2.5\times10^{-3}\right]$$

$$a = \frac{n\lambda}{\sin\theta}$$

$$= \frac{1\times5\times10^{-7}}{2.5\times10^{-3}} = 2\times10^{-4}\text{ m} = 0.2\times10^{-3} = 0.20\text{ mm}$$

Example 2.12

The spectral lines of sodium Δ_1 and Δ_2 have wavelengths of approximately 5890 Å and 5896 Å. A sodium lamp sends incident plane wave onto a slit width 2 μm. A screen is located 2 m from the slit. Find the spacing between the first maxima of the two sodium lines as measured on the screen.

Solution:

Here, $\lambda_1 = 5890$ Å $= 5890 \times 10^{-10}$ m,

$\lambda_2 = 5890$ Å $= 5896 \times 10^{-10}$ m

$a = 2$ μm $= 2 \times 10^{-6}$ m, $\Delta = 2$ m

Distance of the first secondary maximum from the centre of the screen is given by:

$$\sin\theta = \frac{x}{D} = \frac{3}{2}\cdot\frac{\lambda}{a}$$

$$\therefore x = \frac{3}{2}\cdot\frac{D\lambda}{a}$$

For the two wavelengths,

$$x_1 = \frac{3}{2}\frac{D\lambda_1}{a} \text{ and } x_2 = \frac{3}{2}\frac{D\lambda_2}{a}$$

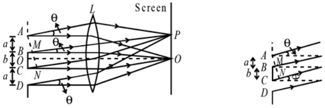
∴ Spacing between the first maximum due to the two sodium lines:

$$= x_2 - x_1 = \frac{3D}{2a}(\lambda_2 - \lambda_1) = \frac{3\times 2\times 6\times 10^{-10}}{2\times 2\times 10^{-6}} = 9\times 10^{-4} \text{ m}$$

$= 0.9 \times 10^{-3}$ m $= 0.9$ mm

2.8 Fraunhofer Diffraction at Double Slit

In Fig. 2.21, AB and CD are two parallel rectangular slits, each of with a and separated by an opaque portion width b. The slits and the screen are perpendicular to the plane of the paper. λ is a collecting lens in whose focal plane, the screen is placed.



(*a*) (*b*)

Fig. 2.21

Let a plane wavefront of light, of wavelength λ is incident on the slits. All the secondary waves which pass undeviated or undiffracted are brought at focal point O of the lens. Therefore, O corresponds to the position of the central bright maximum as all the waves at O in the same phase.

Now consider the rays diffracted at an angle θ. At any point on the screen, we have to consider: (*i*) the interference between the waves reaching from the corresponding points from the two slits and (*ii*) the interference between the waves from the different parts of the same slit or the diffraction due to single. The separation between the corresponding points of the two slits is $(a + b)$. Therefore, the path difference between the rays diffracted from the corresponding points A and C, or any corresponding points in the two slits, is given by:

$$CN = (a + b)\sin\theta.$$

and the corresponding phase difference 2β is given by

$$2\beta = \frac{2\pi}{\lambda}\cdot(a+b)\sin\theta$$

$$\therefore \beta = \frac{\pi}{\lambda}\cdot(a+b)\sin\theta \qquad ...(2.38)$$

The rays diffracted at the extreme points (upper end A and lower end B) from the individual slits, at an angle θ and reaching P, will have a phase difference 2α, given by:

$$2\alpha = \frac{2\pi}{\lambda}\cdot a\sin\theta$$

$$\alpha = \frac{\pi}{\lambda}\cdot a\sin\theta \qquad ...(2.39)$$

Analytically, the resultant displacement y_1 due to the rays from the first slit can be represented by:

$$y_1 = A\sin\omega t \qquad ...(2.40)$$

$$\text{where, } A = A_0\frac{\sin\alpha}{\alpha} \qquad ...(2.41)$$

A_0 is the resultant amplitude of the direct rays and A that of the rays diffracted at an angle θ. The rays from the second slit will have some additional phase difference, say 2β, compared to those from the first slit and so the resultant displacement y_2 due to the rays from the second slit can be represented by:

$$y_2 = A\sin(\omega t + 2\beta) \qquad ...(2.42)$$

Thus, the resultant displacement of the rays diffracted at an angle θ is given by $y = y_1 + y_2$

$$= A\sin\omega t + A\sin(\omega t + 2\beta)$$

$$= A[\sin\omega t + \sin(\omega t + 2\beta)]$$

which on expanding and solving, gives:

$$y = 2A\cos\beta\sin(\omega t + \beta) \qquad ...(2.43)$$

Hence, the amplitude of the resultant displacement is:

$$2A\cos\beta = 2A_0\frac{\sin\alpha}{\alpha}\cos\beta$$

∴ Resultant intensity is proportional to $4A_0^2 \dfrac{\sin^2\alpha}{\alpha^2}\cos^2\beta$

$$\text{or, } I \propto 4A_o^2 \frac{\sin^2\alpha}{\alpha^2}\cos^2\beta \qquad \ldots(2.44)$$

Thus, the resultant intensity in the diffraction pattern depends upon two variable factors: (*i*) $A_o^2 \dfrac{\sin^2\alpha}{\alpha^2}$, which gives diffraction bands similar to those due to single slit, as discussed in the previous section and (*ii*) $\cos^2\beta$, which gives a system of interference fringes due to the superposition of the wavelets from the corresponding parts of the two slits. The diffraction bands have greater dispersion while the interference has lesser dispersion and is superimposed on the diffraction pattern.

Location of Diffraction Bands: We know, $\alpha = \dfrac{\pi}{\lambda}\cdot a\sin\theta$, as θ is actually very small, therefore,

$$\alpha = \frac{\pi}{\lambda}\cdot a\theta \qquad \ldots(2.45)$$

The amplitude of the diffraction bands,

$$A = A_o \frac{\sin\alpha}{\alpha}$$

is maximum when a → o, since $\lim_{\alpha\to 0}\dfrac{\sin\alpha}{\alpha} = 1$ and it is equal to zero for $\alpha = \pm\pi, \pm 2\pi, \pm 3\pi, \ldots \pm n\pi$. Using Eq. (2.39), we obtain the locations of dark diffraction bands for:

$$\theta = \pm\frac{\lambda}{a}, \pm\frac{2\lambda}{a}, \pm\frac{3\lambda}{a}, \ldots$$

for secondary maxima $\alpha = \pm\dfrac{3\pi}{2}, \pm\dfrac{2\pi}{2}, \ldots \pm(2n+1)\dfrac{\pi}{2}$, therefore, using Eq. (2.39), we obtain the locations of secondary maximum at:

$$\theta = \pm\frac{3\lambda}{2a}, \pm\frac{5\lambda}{2a}, \ldots \pm\frac{(2n+1)}{a}\frac{\lambda}{2}.$$

The variation of $A_o^2 \dfrac{\sin^2\alpha}{\alpha^2}$ is shown in Fig. 2.22(*a*).

Locations of Interference Fringes: The location of interference maximum and minimum depend on the factor $\cos^2\beta$. For maximum,

$$\cos^2\beta = 1 \Rightarrow \beta = \pm n\pi,\ n = 1, 2, 3, \ldots$$

From Eq. (2.38), we thus have:

$$\frac{\pi}{\lambda}(a+b)\sin\theta = \pm n\pi$$

$$\text{or, } \sin\theta = \pm\frac{n\lambda}{a+b} \qquad \ldots(2.46)$$

Since θ is small, we can replace sinθ by θ,

$$\therefore \theta = 0,\ \pm\frac{\lambda}{a+b},\ \pm\frac{2\lambda}{a+b},\ \pm\frac{3\lambda}{a+b},\ \ldots \qquad \ldots(2.47)$$

The location θ = o corresponds to the central maximum of the interference pattern in the direction of the incident rays. This is also the direction of central maximum due to diffraction pattern. In this case, all the rays meet in the same phase and reinforce each other forming the central maximum. The other values of q in Eq. (2.47) gives the direction of the principal maximum, e.g., $\pm\dfrac{\lambda}{a+b}$ gives the direction of the first principal maximum, $\pm\dfrac{2\lambda}{a+b}$ that of second principal maximum, and so on.

For minima, $\cos^2\beta = 0 \Rightarrow \beta = \pm(2n+1)\dfrac{\pi}{2}$, $n = 0, 1, 2, 3, \ldots$

From Eq. (2.39), we thus have,

$$\frac{\pi}{\lambda}\cdot(a+b)\sin\theta = \pm(2n+1)\cdot\frac{\pi}{2}$$

$$\sin\theta = \pm\frac{(2n+1)}{(a+b)}\cdot\frac{\lambda}{2}$$

Since θ is small, replacing sinθ by θ, we have:

$$\theta = \pm\frac{\lambda}{2(a+b)}, \pm\frac{3\lambda}{2(a+b)}, \pm\frac{5\lambda}{2(a+b)}, \ldots \qquad \ldots(2.48)$$

Thus, the interference minima occur along the directions for which the angle θ is odd multiple of $\pm\dfrac{\lambda}{2(a+b)}$. The intensity variation due to interference is shown in Fig. 2.22 (*b*). It is obvious from Eqs. (2.47) and (2.48) that the separation between neighbouring interference maximum or minimum is equal to $\frac{\lambda}{a+b}$. The resultant intensity is given by:

$$I = 4A_o^2 \frac{\sin^2 \alpha}{\alpha^2} \cos^2 \beta$$

and its variation with θ is shown in Fig. 2.22 (*c*), which is obtained by the product of ordinates of Fig. 2.22 (*a*) and (*b*). In Fig 2.22, only half of the figures are shown while the complete figure is shown in Fig. 2.23 for the case when $a = 3b$, which shows finer interference bands superimposed upon relatively broad diffraction bands.

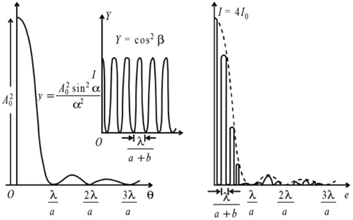


Fig. 2.22: Intensity variation due to double slit.

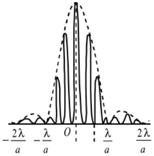


Fig. 2.23: Resultant pattern for $b = 3a$.

You can show, by extending the above discussions, that for three slits, the resultant amplitude is given by:

$$y = A\sin\omega t + A\sin(\omega t + 2\beta) + A\sin(\omega t + 4\beta)$$

which on solving yields $A(1 + 2\cos 2\beta)$

$$I = A^2(1 + 2\cos 2\beta)^2 = I_o(1 + 2\cos 2\beta)^2$$

($A^2 = I_o$ = intensity due to each slit). Since the maximum value of $(1 + 2\cos^2\beta)$ can be equal to 3, therefore, the resultant intensity is 9 times that due to a single slit, compared to 4 times in a double slit. Thus, as the number of slits increases, the central maximum becomes more and more intense and narrow and the intensity of subsidiary maxima falls rapidly. However, the principal maxima become sharper with the increase in the number of slits, we will see in the next section that with diffraction grating, which is an optical device with very large member of slits, a very intense maximum and principal maxima are obtained.

Missing Orders: Sometimes, depending on the relative values of slit width *a* and the separation between the slits *b*, cartain orders of interference maxima will be missing in the resultant pattern, as evident from the following discussion. We know, the condition for interference maxima is:

$$(a + b)\sin\theta = n\lambda \quad ...(2.49)$$

While the condition for diffraction minima is:

$$a\sin\theta = \mu\lambda \quad ...(2.50)$$

In Eqs. (2.49) and (2.50), *n* and μ are integers. If the value of *a* and *b* are such that both the above equations are satisfied simultaneously for the same value of θ in such cases, *n*th order interference maximum and μth order diffraction minimum will coincide on the screen. Due to diffraction minimum, there is nothing reinforcement and the corresponding interference maximum will be missing. Let us consider the following cases:

(*i*) Let $a = b$

$$\text{Then, } 2a\sin\theta = n\lambda,$$
$$\text{and } a\sin\theta = \mu\lambda$$
$$\therefore \frac{n}{m} = 2 \Rightarrow n = 2\mu$$

If μ = 1, 2, 3,..., then *n* = 2, 4, 6,...

Thus, 2nd order, 4th order, 6th order,... interference maxima will be missing in the resultant pattern. There will be three interference maxima in the central diffraction maximum. The position of the second interference maximum corresponds to the first diffraction minimum.

(*ii*) When $2a = b$

$$\text{then, } 3a\sin\theta = n\lambda$$
$$\text{and, } a\sin\theta = \mu\lambda$$
$$\therefore \frac{n}{m} = 3 \Rightarrow n = 3\mu$$

If μ = 1, 2, 3, ..., then *n* = 3, 6, 9, ...

Thus, 3rd order, 6th order, 9th order, ... interference maxima will be missing in the resultant pattern. On either side of the central maximum, the number of interference maxima is two and hence there are five interference maxima in the central diffraction maximum. The position of the third interference maximum corresponds to the first diffraction minimum.

Example 2.13

Find the missing orders for a double slit Fraunhofer diffraction pattern if the slit widths are 0.16 mm separated by 0.8 mm.

Solution:

We know, for interference maxima $(a + b)\sin\theta = n\lambda$ and for diffraction minima $a\sin\theta = \mu\lambda$

$$\therefore \frac{a+b}{a} = \frac{n}{m}$$

Here, a = 0.16 mm, b = 0.8 mm

$$\therefore \frac{0.16+0.8}{0.16} = \frac{n}{m}$$

$$\text{or, } \frac{n}{m} = \frac{0.96}{0.16} = 6 \quad \Rightarrow n = 6\,\text{m}$$

As μ = 1, 2, 3,... therefore, 6th, 12th, 18th, ... order interference maximum will be missing.

2.9 diffraction grating: introduction

A diffraction grating is an optical device having a large number of equidistant narrow rectangular slits of equal width, placed side by side and parallel to one another. The slits are separated by opaque spaces. When a plane wavefront is incident on a grating surface, light is transmitted though the slits are obstructed by the opaque regions. Such a grating is called *transmission grating*.

Fraunhofer was the first person who produced a grating in 1821 with 120 narrow slits. Later Rutherford (1880) and Rowland improved upon it. Finally, after long experimentation, R.W. Wood produced the improved grating on a luminised pyrex glass. It is constructed by ruling equidistant parallel lines with a fine diamond point with the help of a dividing engine on an optically plane glass plate. *The ruled lines are opaque to light while the space between any two lines (acting as slit) is transparent to light while the space between any two lines (acting as slit) is transparent to light*. The original ruled gratings are very expensive and in the laboratories, their photographic reproductions or replicas on celluloid are generally used. A thin layer of a solution of a celluloid dissolved in a volatile solvent is poured over the surface of a grating, whose replica is to made and is allowed to dry. On drying, the plastic film retains a faithful impression of the original rulings and is mounted on a glass plate which is then used as transmission grating. Sometimes, when reflection grating is desired, the replica on the film is placed on a silvered surface to form a reflection grating.

Nowadays, gratings with 12,000 to 30,000 lines per inch of size two to six inches are available. On a grating, the regions through which light can pass are called transparencies while the opaque regions are called *opacities*. If a is the width of a transparency and b that of an opacity, then the distance $(a + b)$ is called *grating constant* or *grating element,* which is also the separation between the corresponding points of the two adjacent transparencies.

Diffraction gratings are available in most of the physics laboratories. You can easily observe beautiful rainbow colours due to reflection of light from a CD which acts as a *reflection grating*. The microscopic pits (or grooves of size less than 0.1 mm deep) on the CD surface act as a reflection grating splitting white light into its constituent colours.

2.10 diffraction in a plane transmission grating

In Fig. 2.24, a parallel beam of monochromatic light of wavelength λ is incident on a plane transmission grating having N number of slits of equal width a separated by opaque regions each of width b. Most of the secondary waves pass undeviated or undiffracted and all such waves are made to converge by the convex lens at its focal point O on the screen. At this point all the secondary waves reinforce one another and corresponds to the position of the *central bright maximum* where the most intense image of the slits is formed.

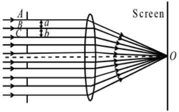


Fig. 2.24

A portion of the incident is diffracted at various angles on passing through the grating. Let us consider the secondary waves diffracted at an angle θ with the direction of the incident light (Fig. 2.25). The collecting lens is suitably rotated such that the axis of the lens is parallel to the direction of the secondary waves. These secondary waves converge at the point P_1 on the screen, having different phases. The intensity at P_1 will depend on the path difference between the corresponding points in two adjacent slits. If AK is the normal drawn to the direction of the diffracted rays, then CN is the path difference between the diffracted rays, CN is the path differece between the rays diffracted from the corresponding points A and C at an angle θ

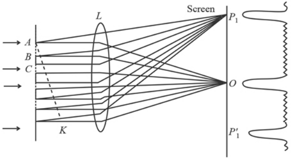


Fig 2.25: Principal maximum in diffraction grating.

$\therefore$ Path difference, $CN = AC\sin\theta = (a + b)\sin\theta$

If this path difference is an even multiple of $\frac{\lambda}{2}$, the point P_1 will be bright. Thus,

for reinforcement: $(a + b)\sin\theta =$

$$\pm 2n \cdot \frac{\lambda}{2} = \pm n\lambda \quad ...(2.51)$$

whereas when the path difference is odd multiple of $\frac{\lambda}{2}$, the point P_1 will be dark. Thus,

for annulment: $(a + b)$ $\sin\theta = \pm(2n+1)\cdot\frac{\lambda}{2}$ ...(2.52)

where n = 0, 1, 2, 3,.... Eq. (2.51) is often referred to as the *grating rule* and gives the direction of the nth order principal maximum which depends on λ and $(a + b)$. These conditions are valid for any pair of corresponding points in the adjoining slits in the grating.

For n = 0, θ = 0 which corresponds to the *central maximum* at O. Secondary waves from all the slits arrive in the same phase and reinforce each other to produce most intense central maximum.

For $n = \pm 1, \sin\theta_1 = \pm\frac{\lambda}{a+b}$ which corresponds to the *first order principal maximum* on either side of the central maximum at P_1 and P'_1.

For $n = \pm 2, \sin\theta_2 = \pm\frac{2\lambda}{a+b}$ which corresponds to the *second order principal maximum.* Similarly, we can find the locations of 3rd, 4th,..., nth order principal maximum. It may be observed that the directions of the principal maxima are the same as those for the maxima in the interference pattern of two slits. However, the difference in the two cases is the brightness which increases significantly with the increase in the number of slits.

When white light is used, rainbow coloured bands appear at each location of the principal maxima in the order of violet, indigo,..., red in the order of the wavelength of the different colours.

Secondary Maxima and Minima: Between any two consecutive principal maxima, there is a set of much smaller secondary maxima and minima. In fact, if there are N number of slits in the grating then there are $(N - 1)$ secondary minima altered by $(N - 2)$ secondary maxima between any two consecutive principal maxima.

We know that $(a + b)\sin\theta_n = n\lambda$ gives the direction of the nth order principal maximum. If there is small change in the angle of diffraction such that the path difference between the rays diffracted at the extreme ends of the grating λ, then the path difference between the rays diffracted at the corresponding points of the adjacent slits will be $\frac{\lambda}{N}$. If we imagine the grating surface to be divided into two halves, the path difference between the secondary waves diffracted in the first half and the corresponding points in the second half will be $\frac{\lambda}{2}$ cancelling out each other. Thus, this gives the position of the *first secondary minimum.* Similarly, we can see if the angle of diffraction is such that the path difference between the extreme ends of the grating is $2\lambda, 3\lambda,..., (N - 1)\lambda$(why not $N\lambda$?) so that the path difference between the corresponding points between the adjoining slits is $\frac{2\lambda}{N}, \frac{3\lambda}{N}, \ldots \frac{(N-1)\lambda}{N}$, corresponding directions will be the direction of second, third, ..., $(N - 1)$th secondary minimum respectively. In between any two secondary minima, there is a secondary maximum. If the path difference between the extreme diffracted rays are $\frac{3\lambda}{2}, \frac{5\lambda}{2}, \frac{7\lambda}{2}, \ldots$ the corresponding directions give the directions of the secondary maxima. In all, *there are* $(N - 1)$ *secondary minima separated by* $(N - 2)$ *secondary maxima between any two consecutive principal maxima.* The intensity of the secondary maxima and the angular separation of the secondary maxima and minima are so small compared to the principal maxima, that they are not observable and there seems to be uniform darkness between any two principal maxima.

Let the angular separation between the first secondary minimum after the nth principal maximum be $\Delta\theta$. We have seen that the additional path difference between the corresponding points of successive slits should be $\frac{\lambda}{N}$. Thus, for the first minimum after nth principal maximum:

$$(a + b)\sin(\theta_n + \Delta) = n\lambda + \frac{\lambda}{N} \quad ...(2.53)$$

Expanding, we have,

$$(a + b)[\sin\theta_n\cos\Delta\theta + \sin\Delta\theta\cos\theta_n] = n\lambda + \frac{\lambda}{N}$$

As $\Delta\theta$ is very small, $\cos\Delta\theta$ = 1 and $\sin\Delta\theta = \Delta\theta$. Thus,

$$(a + b)[\sin\theta_n + \Delta\theta\cos\theta_n] = n\lambda + \frac{\lambda}{N}$$

$$\text{or, } (a + b)\sin\theta_n + (a + b)\Delta\theta\cos\theta = n\lambda + \frac{\lambda}{N}$$

$$\text{As } (a + b)\sin\theta_n = n\lambda$$

$$\therefore (a + b)\Delta\theta\cos\theta_n = \frac{\lambda}{N}$$

$$\text{or, } \Delta\theta = \frac{\lambda}{N(a+b)\cos\theta_n} \quad ...(2.54)$$

$\therefore$ Angular width of nth principal maximum = $2\Delta\theta$

$$= \frac{2\lambda}{N(a+b)\cos\theta_n} \quad ...(2.55)$$

It is indicated from the above equation that the principal maximum becomes sharper with increasing N.

Mathematical Analysis of Diffraction in a Grating: We know that when a plane wavefront is incident on two slits, the phase difference between corresponding points is given by:

$$2\beta = \frac{2\pi}{\lambda}\cdot(a+b)\sin\theta$$

and the resultant amplitude of vibrations from each slit is:

$$A_o\frac{\sin\alpha}{\alpha} \quad \text{where} \quad \alpha = \frac{\pi}{\lambda}a\sin\theta$$

In a grating with N slits, we have to find the resultant amplitude of N vibrations in a direction θ, each of amplitude $A_o\frac{\sin\alpha}{\alpha}$ having a common phase difference $\frac{2\pi}{\lambda}(a+b)\sin\theta$ or 2β. By the method of vector addition, it can be shown that the resultant R is given by:

$$R = a\frac{\sin(n\delta/2)}{\sin\delta/2}$$

In the present case, $a = A_o\frac{\sin\alpha}{\alpha}, n = N, \delta = \frac{2\pi}{\lambda}(a+b)\sin\theta$

$$\therefore R = A_o\frac{\sin\alpha}{\alpha}\times\frac{\sin\left[N\frac{\pi}{\lambda}(a+b)\sin\theta\right]}{\sin\left[\frac{\pi}{\lambda}(a+b)\sin\theta\right]}$$

$$= A_o\frac{\sin\alpha}{\alpha}\cdot\frac{\sin N\beta}{\sin\beta} \quad \text{...(2.56)}$$

where, $\beta = \frac{\pi}{\lambda}(a+b)\sin\theta$

Hence, the intensity is given by:

$$I = R^2 = A_o^2\frac{\sin^2\alpha}{\alpha^2}\cdot\frac{\sin^2 N\beta}{\sin^2\beta} \quad \text{...(2.57)}$$

This equation gives the intensity distribution for a finite number of waves of equal amplitudes and phases increasing in arithmetic progression.

For a single slit, $N = 1, I = A_o^2\frac{\sin^2\alpha}{\alpha^2}$

For a double slit $N = 2, I = A_o^2\frac{\sin^2\alpha}{\alpha^2}\cos^2\beta$

In Eq. (2.57) the factor $A_o^2\frac{\sin^2\alpha}{\alpha^2}$ gives the intensity distribution due to a single slit, while the factor $\frac{\sin^2 N\beta}{\sin^2\beta}$ gives the intensity distribution due to all the slits, i.e., interference effect of the waves reaching from the individual slits. Let us denote $\frac{\sin^2 N\beta}{\sin^2\beta}$ by Z. The intensities of maxima and minima is obtained by equating $\frac{dZ}{d\beta}$ to zero. Thus,

$$Z = \frac{\sin^2 N\beta}{\sin^2\beta}$$

$$\therefore \frac{dZ}{d\beta} = \frac{\sin^2\beta\cdot 2N\sin N\beta\cdot\cos N\beta - 2\sin^2 N\beta\sin\beta\cos\beta}{\sin^4\beta}$$

$$= 2N\sin N\beta\cos N\beta\cdot\frac{1}{\sin^2\beta} - \sin^2 N\beta\left(\frac{\cos\beta}{\sin^3\beta}\right)$$

Equating this to zero, we get:

$$2N\sin N\beta\cos N\beta\cdot\frac{1}{\sin^2\beta} - 2\sin^2 N\beta\left(\frac{1}{\sin^2\beta}\frac{\cos\beta}{\sin\beta}\right) = 0$$

$$\text{or,} \quad 2\frac{\sin^2 N\beta}{\sin^2\beta}\cdot\frac{\cos N\beta}{\sin N\beta} - 2\frac{\sin^2 N\beta}{\sin^2\beta}\cdot\frac{\cos\beta}{\sin\beta} = 0$$

$$\text{or} \quad \frac{2\sin^2 N\beta}{\sin^2\beta}(N\cot N\beta - \cot\beta) = 0$$

$$\therefore \frac{\sin N\beta}{\sin\beta} = 0 \quad ...(2.58)$$

$$\text{or, } N\cot N\beta - \cot\beta = 0 \quad ...(2.59)$$

Principal Maxima: From Eq. (2.58), the intensity would be maximum when $\sin\beta = 0 \Rightarrow \beta = \pm n\pi$, $n = 0, 1, 2, 3, \ldots$. But, at the same time, $\sin N\beta = 0$, so that the factor $\frac{\sin N\beta}{\sin\beta}$ becomes indeterminant fraction $\frac{0}{0}$. To find its value, we employ the usual method of finding the quotient of the first derivative of both the numerator and the denominator (Hospital's rule). Thus,

$$\lim_{\beta\to\pm n\pi}\frac{\sin N\beta}{\sin\beta} = \lim_{\beta\to\pm n\pi}\frac{\frac{d}{d\beta}(\sin N\beta)}{\frac{d}{d\beta}(\sin\beta)}$$

$$= \lim_{\beta\to\pm n\pi}\frac{N\cos N\beta}{\cos\beta} = \pm N$$

Thus, the resultant amplitudes in these directions is proportional to N, i.e.,

$$R = A_o\frac{\sin\alpha}{\alpha}\cdot N$$

and resultant intensity

$$I = R^2 = A_o^2\frac{\sin^2\alpha}{\alpha^2}\cdot N^2$$

Hence, the intensities are proportional to N^2 and they correspond to maxima. These maxima are very intense and are called *principal maxima*. The positions of these principal maxima correspond to:

$$\beta = \pm n\pi$$

$$\text{or, } \frac{\pi}{\lambda}(a+b)\sin\theta = \pm n\pi$$

$$\text{or, } (a+b)\sin\theta = \pm n\lambda,\ n = 0, 1, 2, 3, \ldots \quad ...(2.60)$$

The above expression is known as grating law for the principal maxima. For $n = 0$, we get *central* or *zero maximum*. For $n = \pm1, \pm2, \pm3, \ldots$ we get the first, second, third, ... order principal maximum respectively on either side of the central maximum.

Minima: A series of minima occur when:

$$\sin N\beta = 0, \text{ provided } \sin b \,{}^1\, 0$$

otherwise $\frac{\sin N\beta}{\sin\beta}$ becomes zero. Thus, for minima $\sin N\beta = 0$

$$\text{or, } N\beta = \pm\mu\pi$$

$$\text{or, } N\frac{\pi}{\lambda}(a+b)\sin\theta = \pm\mu p \quad ...(2.61)$$

$$\text{or, } N(a+b)\sin\theta = \pm\mu\lambda$$

where, μ has any integral value except $0, N, 2N, \ldots nN$ because for these values, $\sin\beta$ becomes zero and these positions correspond to principal maxima. In other words, $\mu = 0, 1, 2, 3, \ldots (N-1)$ minima between any two principal maxima.

Secondary Maxima: As there are $(N-1)$ minima between two adjacent principal maxima, there must be $(N-2)$ other maxima separating these minima between two principal maxima. These are known as *secondary maxima*. Consider Eq. (12.59)

$$N\cot N\beta - \cot\beta = 0$$

Its roots, other than those for which $\beta = \pm n\pi$ (which correspond to principal maxima) give the positions of secondary maxima.

To find the intensity of secondary maxima, the fraction $\frac{\sin^2 N\beta}{\sin^2\beta}$ has to be evaluated subject to the condition, $N\cot N\beta - \cot\beta = 0.$

$$\frac{\sin^2 N\beta}{\sin^2\beta} = \frac{1}{\text{cosec}^2 N\beta}\cdot\frac{1}{\sin^2\beta} = \frac{1}{1+\cot^2 N\beta}\cdot\frac{1}{\sin^2\beta}$$

$$= \frac{N^2}{N^2+N^2\cot^2 N\beta}\cdot\frac{1}{\sin^2\beta}$$

$$= \frac{N^2}{N^2+\cot^2\beta}\cdot\frac{1}{\sin^2\beta} \quad (\square\ N\cot N\beta = \cot\beta)$$

$$= \frac{N^2}{N^2+\frac{\cos^2\beta}{\sin^2\beta}}\times\frac{1}{\sin^2\beta}$$

$$= \frac{N^2}{N^2\sin^2\beta+\cos^2\beta} = \frac{N^2}{N^2\sin^2\beta+1-\sin^2\beta}$$

$$= \frac{N^2}{1+(N^2-1)\sin^2\beta}$$

$\therefore$

$$\frac{\text{Intensity of secondary maxima}}{\text{Intensity of principal maxima}} = \frac{1}{1+(N^2-1)\sin^2\beta} \quad ...(2.62)$$

As N increases, intensity of secondary maxima relative to principal maxima decreases and becomes negligible when N is very large. Figure 2.26 (a) and (b) show the variation of intensity due to the factors $\frac{\sin^2\alpha}{\alpha}$ and $\frac{\sin^2 N\beta}{\sin^2\beta}$ respectively. The resultant intensity variation is shown in Fig. 2.26 (c)

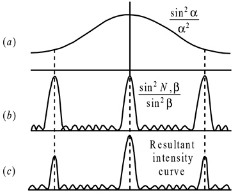


Fig. 2.26

2.11 absent spectra

According to the grating rule $(a + b)\sin\theta = \pm n\lambda$. Now consider, $(a + b)\sin\theta = \lambda$, if $(a + b) < \lambda$, then $\sin\theta > 1$ which is not possible. Hence, the first order spectrum will be absent. Similarly, second, third, ... order spectra will be absent if $(a + b) < 2\lambda$, $(a + b) < 3\lambda$,... . In general, if $(a + b) < n\lambda$, then the nth order spectrum will be absent. The condition for absent spectra can be obtained from the following consideration. We have seen that the condition for nth order principal maximum is given by:

$$(a + b)\sin\theta = n\lambda \quad ...(2.63)$$

While the condition for diffraction minimum from a single slit is given by:

$$a\sin\theta = \mu\lambda \quad ...(2.64)$$

In the above expressions, n and μ are integers. If the two conditions are satisfied simultaneously, then the nth order interference maximum and μth order diffraction minimum will coincide. Due to diffraction minimum for individual slits, there is nothing for reinforcement and the corresponding interference maximum will be missing. Dividing Eq. (2.63) by Eq. (2.64), we get:

$$\frac{a+b}{a} = \frac{n}{m}$$

or, $$\frac{b}{a} = \left(\frac{n}{m} - 1\right) \quad ...(2.65)$$

(i) When the width of the transparency is equal to the width of the opacity, i.e., $a = b$, then,

$$\frac{n}{m} = 2 \Rightarrow n = 2\mu$$

Since, $\mu = 1, 2, 3, ...$

$$n = 2, 4, 6, ...$$

In this case, second order, fourth order, ... spectrum will be missing in the resultant pattern.

(ii) When $\frac{b}{a} = 2$ i.e., the width of opacity is twice the width of the transparency, then:

$$\frac{n}{m} = 3 \Rightarrow n = 3\mu$$

Since, $\mu = 1, 2, 3,...$

$$n = 3, 6, 9,...$$

Thus, 3rd order, 6th order, ... spectra corresponding to interference maxima will be missing in the resultant diffraction pattern,

2.12 maximum number of orders in a grating spectra

According the Grating rule:

$$(a + b)\sin\theta = n\lambda$$

or, $$n = \frac{(a+b)}{\lambda}\sin\theta$$

As the maximum possible value of $\sin\theta = 1$, therefore, the maximum possible order of spectra is:

$$n_{max} = \frac{(a+b)}{\lambda}$$

If $(a + b)$ is between λ and 2λ, i.e., grating constant $(a + b) < 2\lambda$, then

$$n_{max} < \frac{2\lambda}{\lambda} < 2$$

and hence only the first order spectrum is possible. Similarly, if $(a + b)$ is between 2λ and 3λ, upto second order spectrum is possible. In the college laboratories, mostly gratings with N = 15,000 lines per inch are used. Taking $\lambda = 6000$ Å and using the relation $(a + b)\sin\theta = n\lambda$, we have:

$$\sin\theta_1 = \frac{1\cdot\lambda}{a+b} = \frac{6.0\times10^{-5}}{2.54}\times15000$$

$$\left(\because (a+b) = \frac{2.54}{N}\text{cm}\right)$$

$= 0.35403 \Rightarrow \theta_1 = 20.75°$

$$\sin\theta_2 = \frac{2\cdot\lambda}{a+b} = \frac{2\times6.0\times10^{-5}\times15000}{2.54} = 0.7086 \Rightarrow \theta_2 = 45.12°$$

$$\sin\theta_3 = \frac{3\cdot\lambda}{a+b} = \frac{3\times6.0\times10^{-5}\times15000}{2.54} = 1.063,$$

which is not possible as $\sin\theta$ cannot exceed 1. Therefore, *only the first and second order principal maxima on either side of central maxima can be observed.*

2.13 overlapping of spectral lines in a grating

When the light incident on a grating consists of many wavelengths, then in accordance with the grating rule $(a + b)\sin\theta = n\lambda$, a spectrum line (or its direction θ) of a shorter wavelength and of higher order may overlap with a spectral line of longer wavelength and of lower order.

Suppose, the angle of diffraction θ be the same for (*i*) the first order spectral line of wavelength λ, (*ii*) the second order spectral line of wavelength λ_2, (*iii*) the third order spectral line of wavelength λ_3, and so on. Then from the grating rule, we have $(a + b)\sin\theta = 1\cdot\lambda_1 = 2\cdot\lambda_2 = 3\lambda_3 = ...$ The third order spectral line of wavelength 7000 Å (red), the fourth order spectral line of wavelength 5250 Å (green and the fifth order spectral line of wavelength 4200 Å are all formed at the same location on the screen in the diffraction pattern, because,

$$(a + b)\sin\theta = 3\times7000\times10^{-10}$$
$$= 4\times5250\times10^{-10}$$
$$= 5\times4200\times10^{-10}$$

In the visible region of spectrum (3900 to 7600 Å), however, no overlapping of spectral lines takes place. The diffracting angle for the red end of the spectrum in the first order is less than the diffracting angle for the violet colour of the spectrum of the second order. However, if a photographic plate is used for observations, the spectrum is recorded from 2000 Å to ultraviolet region. Then, the first order spectral line corresponding to wavelength 4000 Å overlaps with the second order spectral line of wavelength 2000 Å. The overlapping of the spectral line can be avoided by using suitable filter which absorbs one of the overlapping wavelengths.

2.14 dispersive power of grating

The dispersive power of a grating is defined as the change in the angle of diffraction corresponding to a unit change in the wavelength of the light used. According to grating rule, the condition for the formation of nth order principal maximum is $(a + b)\sin\theta = n\lambda$. This expression indicates that for a given grating constant $(a + b)$ and order n, angle of diffraction θ changes with the change in wavelength. If the wavelength changes from λ to $\lambda + \Delta\lambda$ and the corresponding diffracting angle changes from θ to $\theta + \Delta\theta$, then the dispersive power of the grating is given by $\frac{d\theta}{d\lambda}$. Differentiating the grating rule with request to λ, we get:

$$(a + b)\cos\theta\Delta\theta = nd\lambda$$

or,
$$\frac{d\theta}{d\lambda} = \frac{n}{(a+b)\cos\theta} = \frac{m\times n}{\cos\theta} \quad ...(2.66)$$

where, $m = \frac{1}{a+b}$ is the number of lines per cm length of the grating. From Eq. (2.66). it is implied that:

(*i*) The dispersive power is directly proportional to the order n of the spectrum, i.e., higher the order, higher is the dispersive power.

(*ii*) The dispersive power is inversely proportional to the grating constant $(a + b)$, i.e., smaller the grating constant, more widely spread is the spectrum.

(*iii*) The dispersive power is inversely proportional to $\cos\theta$, i.e., larger the value of θ, smaller is the value of the $\cos\theta$ and higher is the dispersive power. For small value of θ, $\cos\theta$ is constant and so dispersion is directly proportional to $\Delta\lambda$.

Eq. (2.66) basically gives the angular separation between two spectral lines per unit wavelength for a given order in a grating. In photography of the spectra, we need the linear dispersion between the spectral lines. If φ is the focal length of the used lens to obtain the spectrum on the photographic plate and $\Delta\lambda$ the linear separation of two spectral lines of wavelength l and $\lambda + \Delta\lambda$, then $\Delta\lambda = fd\theta$.

∴ Linear separation per unit wavelength of the *linear dispersive power* is given by:

$$\frac{dl}{d\lambda} = f\frac{d\theta}{d\lambda} = \frac{nf}{(a+b)\cos\theta} \quad ...(2.67)$$

Thus, the linear separation between wavelengths λ and $\lambda + \Delta\lambda$ in the nth order is given by:

$$dl = \frac{nf}{(a+b)\cos\theta}d\lambda \quad ...(2.68)$$

Example 2.14

A parallel beam of monochromatic light is incident on plane transmission grating having 5000 lines/cm and the second order spectral line is found to be diffracted at 30°. Find the wavelength of the used light.

Solution:

Here, $N = 5000 \Rightarrow$
$$a+b = \frac{1}{N} = \frac{1}{5000}\text{cm}$$

Substituting the given values in the grating rule $(a + b)\sin\theta = n\lambda$, we have,

$$\frac{1}{5000}\cdot\sin 30^\circ = 2\cdot\lambda$$

or, $\lambda = \dfrac{\sin 30^\circ}{2\times 5000} = \dfrac{0.5}{10,000} = 0.5\times 10^{-4}\text{ cm} = 5.0\times 10^{-5}\text{cm}$

Example 2.15

A plane transmission grating having 6000 lines per cm is used to obtain a spectrum of light from a sodium light in the second order. Find the angular separation between the two spectral lines if the two wavelengths in sodium light is 5890 Å and 5896 Å.

Solution:

Here, $a + b = \dfrac{1}{N} = \dfrac{1}{6000}\text{cm}, \quad n = 2$

$\lambda_1 = 5890 \times 10^{-8}$ cm, $\lambda_2 = 5896$ Å $= 5896 \times 10^{-8}$ cm

Using the grating rule $(a + b)\sin\theta = n\lambda$, we have:

$\sin\theta_1 = \dfrac{n\lambda_1}{a+b} = \dfrac{2\times 5890\times 10^{-8}}{(1/6000)}$

$= 2 \times 5890 \times 10^{-8} \times 6000 = 0.7068$

or, $\theta_1 = \sin^{-1}(0.7068) = 44°59'$

$\sin\theta_2 = \dfrac{n\lambda_2}{a+b} = \dfrac{2\times 5896\times 10^{-8}}{(1/6000)}$

$= 2 \times 5896 \times 10^{-8} \times 6000 = 0.7075$

or, $\theta_2 = \sin^{-1}(0.7075) = 45°2'$

∴ Angular separation of the spectral lines $= \theta_2 - \theta_1 = 3'$.

Example 2.16

What is the highest order spectrum which may be seen with monochromatic light of wavelength 5000 Å by means of a diffraction grating with 5000 lines/cm?

Solution:

Here, $\lambda = 5000$ Å $= 5.0 \times 10^{-5}$ cm, $N = 5000$ lines/cm

or, $(a + b) = \dfrac{1}{N} = \dfrac{1}{5000}\text{cm}$

We know, $(a + b)\sin\theta_n = n\lambda$

Maximum value of $\sin\theta_n = 1$,

$\therefore n_{max} = \dfrac{a+b}{\lambda} = \dfrac{1}{5000\times 50\times 10^{-5}} = 4.$

Example 2.17

In a grating spectrum which spectral line in the 4th order will overlap with 3rd order line of the wavelength 5461 Å?

Solution:

Let λ be a spectral line (wavelength) whose 4th line overlaps with the 3rd order line of the wavelength 5461 Å.

$\therefore (\alpha + \beta)\sin\theta = 4\lambda = 3 \times 5461 \times 10^{-10}$

or, $\lambda = \dfrac{3\times 5461\times 10^{-10}}{4}$

$= 4095.75 \times 10^{-10}$ m $= 4095.75$ Å

Example 2.18

A diffraction grating used at normal incidence gives a line (5400 Å) in a certain order superposed on the violet line (wavelength 4050 Å) of the next higher order. If the angle of diffraction is 30°, calculate the number of lines per cm.

Solution:

Let n and $(n + 1)$ be the overlapping orders.

$\therefore (a + b)\sin\theta = n\lambda_1 = (n + 1)\lambda_2$

or, $n = \dfrac{(a+b)\sin\theta}{\lambda_1}$ and $(n+1) = \dfrac{(a+b)\sin\theta}{\lambda_2}$

Subtracting the first expression from the second, we get:

$$\left(\frac{1}{\lambda_2} - \frac{1}{\lambda_1}\right)(a+b)\sin\theta = 1$$

or, $(a + b) = \dfrac{\lambda_2\lambda_2}{\lambda_1 - \lambda_2}\cdot\dfrac{1}{\sin\theta}$

Given $\lambda_1 = 5400$ Å $= 5.4 \times 10^{-5}$ cm, $\lambda_2 = 4050$ Å $= 4.05 \times 10^{-5}$ cm

$\theta = 30°$

$\therefore (a + b) = \dfrac{5.4\times 10^{-5}\times 4.05\times 10^{-5}}{(5.40-4.05)\times 10^{-5}}\times\dfrac{1}{\sin 30^\circ} = \dfrac{21.87}{1.35}\times 10^{-5}\times\dfrac{1}{0.5}$

$= 32.4 \times 10^{-5}$ cm

No. of lines/cm =

$$\frac{1}{a+b} = \frac{1}{32.4} \times 10^5 = 3086.$$

Example 2.19

For a plane transmission grating with 5000 lines/cm:

(*i*) what is the highest order of spectrum observable with light of 6000 Å and

(*ii*) if width of the opacity is twice that of transparency, find the absent orders of spectra.

Solution:

(*i*) For a plane transmission grating, the maximum order observable with wavelength λ is given by:

$$n_{max} = \frac{(a+b)}{\lambda}$$

Here, $(a + b) =$

$$\frac{1}{5000} \text{ and } \lambda = 6000 \times 10^{-8}\text{cm} = 6.0 \times 10^{-5}\text{cm}$$

$$n_{max} = \frac{1}{5000 \times 6.0 \times 10^{-5}} = 3.33 \cong 3$$

Thus, the 3rd order is the highest order that can be observed.

(*ii*) Condition for the order n to be missing is:

$$\frac{a+b}{a} = \frac{n}{m} \Rightarrow \frac{n}{m} = 1 + \frac{b}{a}$$

Here, $b = 2a$

$$\frac{n}{m} = 1 + 2 = 3 \Rightarrow n = 3\mu, \mu = 1, 2, 3, \ldots$$

or, $n = 3, 6, 9, \ldots$

Thus, 3rd, 6th, 9th, ... order spectra will be absent.

Example 2.20

A diffraction grating which has 4000 lines per cm is used at normal incidence. Calculate the dispersive power of the grating in the third order spectrum in the wavelength region 5000 Å.

Solution:

Dispersive power of a grating is given by:

$$\frac{d\theta}{d\lambda} = \frac{n}{(a+b)\cos\theta}$$

Also, $(a + b)\sin\theta = n\lambda$

Here, $n = 3$, $\lambda = 5000$ Å $= 5.0 \times 10^{-5}$ cm

$$(a+b) = \frac{1}{N} = \frac{1}{4000}\text{cm}$$

$\therefore \sin\theta =$

$$\frac{n\lambda}{(a+b)} = 3 \times 5.0 \times 10^{-5} \times 4000 = 0.6$$

$$\therefore \cos\theta = \sqrt{1-\sin^2\theta} = \sqrt{1-(0.6)^2} = 0.8$$

$$\text{Dispersive power } \frac{d\theta}{d\lambda} = \frac{3 \times 4000}{0.8} = 15{,}000.$$

Example 2.21

In the second order spectrum of a plane diffraction grating a certain spectral line appears at an angle of 10°, while another line of wavelength 0.5 Å greater appears at an angle 3″ greater. Find the wavelength of the lines and the minimum grating width required to resolve them. (Given: sin10° = 0.1736 and cos10° = 0.9848)

Solution:

According to the grating rule:

$$(a + b)\sin\theta = nl \ldots(ii)$$

Differentiating, we get:

$$(a + b)\cos\theta\Delta\theta = ndl \ldots(ii)$$

Dividing Eq. (*i*) by Eq. (*ii*), we get:

$$\frac{\sin\theta}{\cos\theta d\theta} = \frac{\lambda}{d\lambda}$$

$$\text{or, } l = \frac{\sin\theta d\lambda}{\cos\theta d\theta}$$

Given: $\theta = 10°$, $\Delta\lambda = 0.5 \times 10^{-8}$ cm

$\Delta\theta =$

$$3'' = \left(\frac{1}{6 \times 60}\right) = \frac{3}{60 \times 60} \times \frac{\pi}{180}\text{radian}$$

$$\therefore \lambda = \frac{\sin 10°}{\cos 10°} \cdot \frac{0.5 \times 10^{-8}}{\frac{3}{60 \times 60} \times \frac{\pi}{180}}$$

$$= \frac{0.1736 \times 0.5 \times 10^{-8} \times 60 \times 60 \times 180}{0.9848 \times 3 \times 3.14}$$

$$= 6063 \times 10^{-8}\ \text{cm} = 6063\ \text{Å}$$

$$\therefore \lambda + \Delta\lambda = 6063 \times 10^{-8} + 0.5 \times 10^{-8} = 6063.5 \times 10^{-8}\ \text{cm}$$

Thus, the wavelengths of the lines are 6063 Å and 6063.5 Å.

Resolving power of a grating: $\frac{\lambda}{d\lambda} = nN$

$$\text{or, } N = \frac{1}{n} \cdot \frac{\lambda}{d\lambda} = \frac{1}{2} \cdot \frac{6063 \times 10^{-8}}{0.5 \times 10^{-8}} = 6063$$

$\therefore$ Minimum grating width required = $N(a + b)$

$$= 6063 \times \frac{n\lambda}{\sin\theta}$$

$$= \frac{6063 \times 2 \times 6063 \times 10^{-8}}{0.1736}$$

$$= 4.24\ \text{cm.}$$

2.15 resolving power

When we see two small objects, to view them clearly they must be apart and if they are close, it is difficult to distinguish between them and if they are very close, they appear as one. *Our eyes see two objects, as separate only if the angle subtended by them at the eye is greater than one minute.* In other words, *the minimum angle of resolution of a normal human eye is one minute* and we cannot see two nearby objects as separate if the angle subtended by them at the eye is less than one minute. To view two objects placed very closely we use optical instruments such as telescope, microscope, prism, grating, etc. The process of observing two nearby objects as separate by employing a suitable optical instrument is called *resolution. The reciprocal of minimum angle of resolution for the instrument gives its resolving power* which is a numerical measure of the ability of the instrument for resolving two nearby objects.

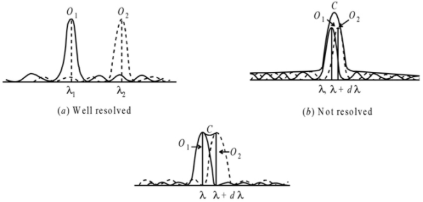
Rayleigh's Criterion for Resolution: We have seen in the preceding section that the image of a point source is not a point, but is a diffraction pattern. The diffraction pattern of a point source of lignt consists of a central bright disc surrounded by alternate dark and bright diffraction rings.

Lord Rayleigh, after an extensive study, proposed the condition for just resolution. According to Reyleigh, *two nearby images appear to be just resolved, if the position of the central maximum of one coincides with the first minimum of the other and vice versa.* This is known as *Rayleigh's Criterion of just (or limiting) resolution.* This condition is equivalent to the condition that the distance between the centres of the patterns should be equal to the radius of the central diffraction disc.

The above criterion for resolution is equally applicable when an object is viewed using light having two spectral lines of wavelengths λ_1 and λ_2. The two spectral lines appear to be just resolved if the central maximum of the diffraction pattern of one falls upon the first minimum of the diffraction pattern of the other and vice versa.

In Fig. 2.27(*a*), O_1 and O_2 are the central maxima of the diffraction patterns of two spectral lines wavelengths λ_1 and λ_2. In this case, the images are *well resolved*, i.e., they can be seen distinctly separate as the angular separation between their centres is much larger than that between the centres of the either and its first minimum. There is distinct point of zero intensity between the two centres which helps in viewing the central maxima separately and clearly.

In Fig. 2.27(*b*), the central maxima corresponding to the wavelengths λ_1 and λ_2 (= $\lambda_1 + \Delta\lambda$) are very close (due to very small difference in wavelengths). The angle of diffraction corresponding to the first minimum of O_1 is greater than the angle of diffraction corresponding to the central maximum of O_2. Hence, the two images overlap and they cannot be seen as separate. The resultant intensity curve has a maximum at *C* and the intensity of this maximum is higher than those for O_1 and O_2 and hence O_1 and O_2 cannot be distinguished, i.e., *two spectral lines cannot be resolved.*



(*a*) Well resolved

(*b*) Not resolved

(*c*) Just resolved

Fig. 2.27

In Fig 2.27(*c*), the position of the central maximum O_1 of wavelengths λ_1 coincides with the position of the first minimum. O_2 due to λ_2(= $\lambda_1 + \Delta\lambda$, $\Delta\lambda$ larger than the previous case) and vice versa. Thus, in accordance with Rayleigh's criterion, the *two are just resolved*. The resultant intensity curve shows as dip at *C*, in the middle of the central maxima of O_1 and O_2 (assuming the two spectral lines to be of the same intensity) due to which O_1 and O_2 can be distinguished and can just be seen separately.

Chromatic Resolving Power: For two spectral lines to appear just resolved, the central maximum of one must fall upon the first minimum of the other, i.e., the least angular separation between the central maxima of the diffraction pattern of the two spectral lines must be equal to half the angular width of the either central maximum. *The chromatic resolving power is the ratio of the mean wavelength of a pair of just resolvable spectral lines and* Dl *the least difference in wavelengths.* It is expressed as $\frac{\lambda}{\Delta\lambda}$ and gives the chromatic resolving power of the optical instrument at that wavelengths.

2.16 resolving power of a grating

One of the important applications of a diffraction grating is its ability to separate spectral lines having nearly the same wavelength. In other words, a diffraction grating can resolve the spectral lines. The resolving power of a grating is given by $\frac{\lambda}{d\lambda}$, where λ is the wavelength of a spectral line and Δλ is the difference between this line and a neighbouring line such that the two are just resolved by the grating.

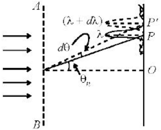


Fig. 2.28

In Fig 2.28, a beam of light consisting of wavelengths λ and λ + Δλ is incident normally on the surface of a plane transmission grating *AB*. The light of each wavelength forms a separate diffraction pattern on the spectroscopic slit. Let *P* be the *n*th principal maximum of a spectral line of wavelength λ at an angle of diffraction θ_n, while *P′* is the *n*th principal maximum of the second spectral line of wavelength l + Δλ at a diffracting angle $\theta_n + \Delta\theta$. According to Rayleigh's criterion, the two spectral lines will appear just resolved, if the position of *P′* coincides with the first minimum of *P* and vice versa, The condition for the *n*th principal maximum of wavelength λ is given by:

$$(a + b)\sin\theta_n = n\lambda$$

where $(a + b)$ is the grating constant.

The condition for the *n*th principal maximum of wavelength λ + Δλ is given by:

$$(a + b)\sin(\theta_n + \Delta\theta) = n(\lambda + \Delta\lambda) \quad ...(2.69)$$

The two lines will appear just resolved if the angle of diffraction $(\theta_n + \Delta\theta)$ also corresponds to the direction of the first minimum after the *n*th principal maximum *P* due to wavelength λ. We have seen earlier, for this an additional path difference of $\frac{\lambda}{N}$ is required between the corresponding points of the adjacent slits, where *N* is the total number of slits on the grating. Thus,

$$(a + b)\sin(\theta_n + \Delta\theta) = n\lambda + \frac{\lambda}{N} \quad ...(2.70)$$

Equating Eqs. (2.69) and (2.70) we have:

$$n(\lambda + \Delta\lambda) = n\lambda + \frac{\lambda}{N}$$

$$\text{or, } nd\lambda = \frac{\lambda}{N}$$

$$\text{or, resolving power, } \frac{\lambda}{d\lambda} = nN \quad ...(2.71)$$

Thus, the resolving power of a grating is directly proportional to (*i*) the order the spectrum and (*ii*) the total number of slits on the grating. The *resolving power is independent of the grating constant*. For a given grating, the separation between the spectral lines is double in the second order compared to that in the first order.

It is important to remember that dispersive power refers to wide separation of the spectral lines and is dependent on grating constant, while resolving power refers to the ability of the grating to show nearby spectral lines as separate ones.

2.17 resolving power of a telescope

A telescope is used to observe distant objects. In addition to magnify, a telescope should also resolve two nearby objects such as stars. The extent of resolution depends on the angle subtended at the objective of the telescope by the two point objects and not the linear separation between them. *The resolving power of a telescope is defined as the reciprocal of the minimum angle subtended at the objective by the two distant point objects which can just be seen as separate through the telescope.*

Suppose a parallel beam of light from a distant source, is incident normally on the objective of a telescope and is brought to focus at φ, as shown in Fig. 2.29. The objective lens in this case acts as a circular aperture and produces a diffraction pattern consisting of a bright circular disc with centre at φ surrounded alternatively by a series of dark and bright rings in the focal plane of the objective. Let the first minimum of the diffraction pattern be produced at *P*, then path difference *AP* – *BP* must be equal to λ.

$$\text{Also, } AP - BP = AE = AB \times \Delta\theta = \Delta \times \Delta\theta$$

where *AB* = Δ is the diameter of the lens aperture and Δθ is the angular separation between the central maximum and first minimum.

Since for first minimum, *AP* – *BP* = λ;

$$\lambda = \Delta \times \Delta\theta \Rightarrow \Delta\theta = \frac{\lambda}{D} \quad ...(2.72)$$

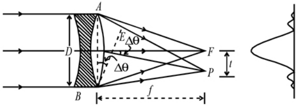


Fig. 2.29: Diffraction pattern of a distant source due to a telescope objective.

Now, consider two distant point sources of light, such as two stars S and S' having an angular separation of $\Delta\theta$. Suppose they produce two circular images (in the form of diffraction patterns) with centres are φ and P, lying in the focal plane of the telescope objective AB (Fig. 2.30), the central bright disc in each pattern will be surrounded by dark and bright bands.

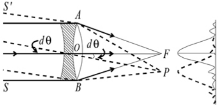


Fig. 2.30

In accordance with Rayleigh's criterion, the two stars S and S' will be just resolved, when the central maximum of one coincides with the first minimum of the diffraction pattern of the other, i.e., the angular separation of the central images of two stars is equal to the angular separation between the central maximum of one and its first minimum. Hence, for the limiting resolution.

$$d\theta = \Delta\theta = \frac{\lambda}{D}$$

where $\Delta\theta$ is the minimum angle of resolution. As discussed earlier, a more rigorous treatment of this condition for a circular aperture by Airy showed that instead of λ, the more correct value is 1.221, thus:

$$d\theta = \frac{1.22\lambda}{D}$$

is the minimum angle of resolution for a telescope with diameter of objective Δ.

$$\therefore \text{Resolving power} = \frac{1}{d\theta} = \frac{D}{1.22\lambda} \quad \text{...(2.73)}$$

Thus, *the resolving power of a telescope is proportional to the diameter of the objective and inversely proportional to the wavelength of the light used.*

The resolving power of a telescope increases with the increase of the aperture of the objective lens. Moreover, large aperture collects more light and makes it possible to view even faint stars. For this reason the diameters of the telescope objectives used for astronomical observations are very large.

If r is the radius of first dark ring surrounding the central disc and φ the focal length of the telescope objective,

Then,

$$\Delta\theta = \frac{r}{f} = \frac{1.22\lambda}{D}$$

$$\text{or, } r = \frac{1.22 f\lambda}{D} \quad \text{...(2.74)}$$

This indicates that for a given wavelength λ of incident light, the greater the diameter and smaller the focal of the objective, the smaller would be the radius of the central bright disc called Airy's disc. The diffraction pattern will, therefore, appear sharper and the value of minimum angular separation $\Delta\theta$ for the two point objects will be smaller, thus increasing the resolving power of the telescope.

To resolve two stars with an angular separation of one second (= 4.85×10^{-6} radian) using light of wavelength 5000 Å, the diameter of the objective should be equal to

$$D = \frac{1.22\lambda}{d\theta} = \frac{1.22 \times 5 \times 10^{-5}}{4.85 \times 10^{-6}} \cong 12.6 \text{ cm}$$

Dawe's Rule: In the above example, for a light of wavelength 5000 Å, the minimum resolvable angle $d\theta = \frac{1.22\lambda}{D}$ is equal to

$$\Delta\theta = \frac{1.22 \times 5 \times 10^{-5}}{D} \times \frac{180 \times 60 \times 60}{\pi} \text{ second} = \frac{12.4}{D} \text{second}$$

For a telescope with objective diameter D = 100 cm, we have

$$\Delta\theta = \frac{12.4}{100} \text{ second or } \frac{1}{8} \text{second}$$

Hence, the least angular separation (in seconds) which is resolvable by this telescope is:

$$d\theta = \frac{1.24}{\text{aperture in cm}} \quad \text{...(2.75)}$$

This is known as Dawe's rule.

2.18 resolving power of a prism

When a prism spectroscope is used, it forms an image of the slit for each wavelength or spectral line present in the incident light. Two close spectral lines can be resolved only if the angular separation of the images corresponding to them is higher than or equal to minimum angle of resolution of the telescope. If l and $\lambda + \Delta\lambda$ are the wavelengths of two close lines which are just resolved, then the ratio $\frac{\lambda}{d\lambda}$ gives the resolving power of prism.

If Fig. 2.31, ABC is section of a prism on which a parallel beam of light consisting wavelengths λ and $\lambda + \Delta\lambda$ is incident from a slit S. The collimating lens λ_1 makes the incident rays parallel. These rays are incident on the face AC of the prism which is placed in the minimum deviation position. Assuming the light to be passing through whole of the prism, let AM be the incident wavefront and BN and BN' the emergent wavefronts corresponding to the wavelengths λ and $\lambda + \Delta\lambda$ respectively. The telescope objective lens λ_2 of the spectrometer focusses the wavefront BN at I, which is the image or central maximum of the slit formed by the light of wavelength λ in the focal plane of the lens λ_2. Similarly, the wavefront BN' is focussed at I', which is the image or the central maximum of slit formed by the light of wavelength $\lambda + \Delta\lambda$. The two images or spectral lines will appear just resolved if the first minimum of one coincides with the central maximum of the other.

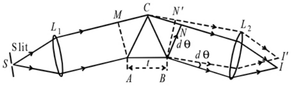


Fig. 2.31

Let μ be the refractive index of the material of the prism for the wavelength λ, then by Fermat's principle, the optical path between the two successive positions AM and BN of the wavefront must be the same, i.e.,

$$MC + CN = \mu AB \quad ...(2.76)$$

Now, say, for the light of wavelength $\lambda + \Delta\lambda$, the refractive index of the material of the prism $\mu - \Delta\mu$, therefore, for the emergent wavefront BN' we have,

$$MC + CN' = (\mu - \Delta\mu)AB \quad ...(2.77)$$

Subtracting Eq. (2.76) from Eq. (2.90), we get:

$$CN - CN' = AB \cdot \Delta\mu$$

$$\text{or, } NN' = AB \cdot \Delta\mu = td\mu \quad ...(2.78)$$

where $AB = t$ is the width of the base of the prism.

If θ and $\theta + \Delta\theta$ are the angles of deviation of light of wavelengths λ and $\lambda + \Delta\lambda$ respectively, then the angle between the two refracted wavefronts is $\Delta\theta$, from the geometry of Fig. 2.31.

$$NN' = BN \cdot \Delta\theta = Dd\theta \quad ...(2.79)$$

where Δ is the width of the emergent beam BN which is also equal to the diameter of the telescope objective through which the emergent beam is observed. Equating Eq. (2.79), we get

$$Dd\theta = td\mu$$

$$\text{or, } \Delta\theta = d \cdot \frac{d\mu}{D} \quad ...(2.80)$$

Also, for a telescope the limiting condition for the resolution is given by:

$$d\theta = \frac{\lambda}{D} \quad ...(2.81)$$

Hence, the two wavelengths λ and $\lambda + \Delta\lambda$ are just resolved in a prism spectroscope if,

$$\frac{\lambda}{D} = t \cdot \frac{d\mu}{D}$$

$$\text{or, } \lambda = t \cdot \Delta\mu$$

Thus, the resolving power of the prism can be obtained by dividing the above equation by $\Delta\lambda$, or,

$$\frac{\lambda}{d\lambda} = t \cdot \frac{d\mu}{d\lambda} \quad ...(2.82)$$

The quantity $\frac{d\mu}{d\lambda}$ gives the chromatic dispersion of the material of the prism and it can be obtained from Cauchy's relation:

$$\mu = A + \frac{B}{\lambda^2}$$

where A and B are constants. Differentiating w.r.t. λ, we get

$$\frac{d\mu}{d\lambda} = -\frac{2B}{\lambda^3}$$

$$\therefore \text{Resolving power } \frac{\lambda}{d\lambda} = t \cdot \frac{d\mu}{d\lambda} = \frac{-2Bt}{\lambda^3} \quad ...(2.83)$$

Thus, the resolving power of a prism is directly proportional to the width of the base of the prism and inversely proportional to the cube of the wavelength of light used. It is independent of the angle of the prism. For ordinary flint glass, the value of $\frac{d\mu}{d\lambda}$ is about 800 cm^{-1} in the yellow region and hence, to obtain a resolving power of 1000, the required width of the base of the prism is 1.25 cm.

Example 2.22

Light is incident normally on a grating of total ruled width 5×10^{-3} m with 2500 lines in all. Calculate the angular separation of the two sodium lines in the first order spectrum. Can they be seen distinctly?

Solution:

Here, $N = 2500$ over 5×10^{-3} m or 0.5 cm

$$\therefore (a + b) = \frac{\text{Grating width}}{\text{Total no. of lines on the grating}}$$

$$= \frac{0.5}{2500}\,\text{cm} = \frac{1}{5000}$$

$n = 1$, $\lambda_1 = 5890$ Å and $\lambda_2 = 5896$ Å

for λ_1(Δ_1 line): $(a + b)\sin\theta_1 = n\lambda_1$

$$\frac{1}{5000} \times \sin\theta_1 = 1 \times 5.890 \times 10^{-5}$$

$$\text{or, } \sin\theta_1 = 5000 \times 5.890 \times 10^{-5} = 0.2945$$

or, $\theta_1 = 17°8'$

Similarly, for λ_2: $\sin\theta_2 = 5000 \times 5.896 \times 10^{-5} = 0.2948$

or, $\theta_2 = 17°9'$

∴ Angular separation between the two lines = $\theta_2 - \theta_1 = 1'$

for just resolution: $\dfrac{\lambda}{d\lambda} = nN$

$$\therefore \frac{5893\times10^{-8}}{6\times10^{-8}} = 1 \times N$$

or, $N = 982$ lines

Thus, the minimum number of lines for just resolution is 982 while the grating has 2500 lines, so the lines can be seen distinctly.

Example 2.23

What is the minimum number of lines per cm in a 2.5 cm wide grating which will just resolve the two sodium lines (5890 Å and 5896 Å) in the second order spectrum.

Solution:

We know for a grating $\dfrac{\lambda}{d\lambda} = nN$

Here, $\lambda = \dfrac{5890\text{ Å} + 5896\text{ Å}}{2} = 5893\text{ Å} = 5893\times10^{-8}\text{ cm}$

$\Delta\lambda = 5896 - 5890\text{ Å} = 6\text{ Å} = 6.0 \times 10^{-8}$ cm

$n = 2, N = ?$

$$\therefore N = \frac{\lambda}{d\lambda}\cdot\frac{1}{n} = \frac{5893\times10^{-8}}{6\times10^{-8}}\cdot\frac{1}{2} \cong 491$$

Width of grating = 2.5 cm

∴ Number of lines per cm

$$= \frac{491}{2.5}\times196.4 = 196$$

Example 2.24

A plane transmission grating has 16,000 per inch over a length of 5 inch. Find (*a*) the resolving power of the grating in the second order and (*b*) the smallest wavelength difference that can be resolved for the light of wavelength 6000 Å.

Solution:

Here, $N = 16{,}000 \times 5 = 80{,}000$ $n = 2$

l = 6000 Å = 6.0×10^{-5} cm

(*a*) Resolving power of the grating $\dfrac{\lambda}{d\lambda} = nN = 2\times80{,}000 = 1{,}60{,}000$

(*b*) Smallest resolvable wavelength:

$$\Delta\lambda = \frac{\lambda}{nN} = \frac{6.0\times10^{-5}}{1{,}60{,}000}.$$

$= 3.75 \times 10^{-10}$ cm $= 0.0375 \times 10^{-8}$ cm

$= 0.0375$ Å.

Exercises

Short Answer Type

1. Does light travel in a straight line? Explain.

2. What do you mean by diffraction? Differentiate between interference and diffraction.

3. Distinguish between Fresnel and Fraunhofer diffraction.

4. Define Fresnel distance and explain its significance.

5. Explain Fresnel's assumptions.

6. What are Half Period Zones? Explain.

7. What is the phase relationship between the secondary waves from the (*i*) successive and (*ii*) alternate half period zones?

8. What is a Zone Plate? Differentiate between Positive and Negative Zone plates.

9. Compare a zone plate with a convex lens giving their similarities and dissimilarities.

10. Show the intensity variation due to diffraction of light at a straight edge and explain it. Hence, give the difference between interference and diffraction bands.

11. What is Cornu's spiral? Mention its important characteristics.

12. Write the expression for the resultant intensity due to the diffraction light at a double slit and show it graphically.

13. If the maximum intensity in the diffraction pattern of a single slit is I_0, what will be the maximum intensity in the diffraction pattern due to (*i*) a double slit and (*ii*) a triple slit.

14. What is a diffraction grating? Write grating rule.

15. If a diffraction grating has 10,000 lines, what is the number of secondary maxima and minima between any two principal maxima?

16. Obtain the expression for the angular width of the *n*th principal maximum.

17. Give the condition for missing spectra? If $b = 2a$, spectra of which orders will be missing?

18. What do you mean by overlapping of spectral lines? Explain.

19. Give the condition for maximum number of orders observable with a grating.

20. What do you mean by dispersive power of a grating? On what factors does it depend?

21. Define: (*i*) Resolution, (*ii*) Minimum Angle of Resolution, (*iii*) Resolving Power, (*iv*) Chromatic Resolving Power.

22. What is the limiting angle of resolution of our eye?

23. Explain Rayleigh's criterion for just resolution.

24. If you have the option of working with a grating of 5000 lines or 10,000 lines, with which should you work and why?

25. Give the advantages of a large aperture of the objective of a telescope.

26. How is the resolving power of a prism dependent on the width of the base? Explain.

Long Answer Type

1. What is diffraction? How does it differ from interference. Differentiate between Fresnel and Fraunhofer's class of diffraction.

2. What are half period zones? Show that the area of each half period zone is approximately equal and their radii are proportional to the square root of the natural numbers.

3. Show that the effect of the entire wavefront is approximately equal to half the amplitude of the first half period zone. What conclusion can you draw about the rectilinear propagation of light?

4. What is a Zone Plate? Explain its construction and theory.

5. Discuss Fresnel diffraction at a straight edge. Draw the intensity variation and explain it. Obtain the expression for the locations of bright and dark bands.

6. What is Cornu's Spiral? Explain. Mention its important characteristics. Using Cornu's Spiral, explain the observed diffraction due to a straight edge.

7. Discuss Fraunhofer Diffraction of light at a narrow slit. Deduce the conditions for maxima and minima and draw the intensity variation curve.

8. Show that the relative intensities of the maxima in the diffraction pattern of a single slit are nearly in the ration $1:\frac{1}{22}:\frac{1}{61}:\frac{1}{121}:\ldots$

9. Discuss Fraunhofer Diffraction at a circular aperture and obtain an expression for the radius of the Airy's Disc.

10. Discuss Fraunhofer Diffraction of light at a double slit and obtain the conditions for diffraction and interference maxima and minima.

11. Give the construction and theory of a plane transmission grating and explain the formation of spectra by it.

12. What do you understand by missing order spectrum? Obtain the conditions for the same.

13. Define dispersive power of a diffraction grating and deduce an expression for the same.

14. Define resolution and resolving power. Explain Rayleigh's criterion for just resolution with the help of suitable diagrams.

15. Obtain the expression for resolving power of a grating? On what factors does it depend?

16. Deduce the expression for resolving power of a telescope. How can you increase the resolving power of a telescope?

17. Obtain the expression for resolving power of a prism. To get between resolution, will you work with red or blue light? Why?

Numericals

1. A narrow slit illuminated by light of wavelength 6.4×10^{-5}cm is placed at a distance of 3 m from straight edge. If the distance between the straight edge and the screen is 6 metres, calculate the distance between the first and the fourth dark band. [Ans. 0.48 cm]

2. A narrow slit illuminated by light of wavelength 6000 Å is placed at a distance of 10 cm from a straight edge. If measurements are done at a distance of 100 cm from the straight edge, calculate the distance between the first and the second dark bands. [Ans. 0.1503 cm]

3. A zone plate is constructed in such a way that the radii of the circles which define the zones are the same as the radii of Newton's rings formed between a plane surface and a surface having radius of curvature of 1 m. Find the primary focal length of zone plate. [Ans. 1 m]

4. A plane transmission grating has 40,000 lines in all with grating element 12.5×10^{-8}cm. Calculate the maximum resolving power for which it can be used in the range of wavelength 5000 Å. [Ans. 80,000]

5. When light of wavelength 5690 Å falls normally on a grating, the deviation is found to be 11° in the first order spectrum. Calculate the number of lines per metre length of the grating. [Ans. 3.35×10^5]

6. How large a prism is needed to resolve sodium D-lines if the rate of change refractive index with wavelength is 6.2×10^{-5} per millimicron? [Ans. 1.6 cm]

7. Calculate the least width a grating must have to resolve two components of sodium *D* lines in the second order, the grating having 800 lines per cm. [Ans. 0.614 cm]

8. A grating is ruled over a width of 10 cm and the number of lines on the grating is 5000 lines per cm. Find the smallest wavelength difference that could be resolved in the region of 5000 Å in the first order. [Ans. 0.1°]

9. Diffraction pattern of single slit of width 0.5 cm is formed by a lens of focal length 40 cm. Calculate the distance between the first dark and next bright fringe from the axis, l = 4890 Å. [Ans. 5600 Å]

10. In Fraunhofer Diffraction pattern due to a single slit, the screen is at a distance of 100 cm from the slit and the slit is illuminated by a monochromatic light of wavelength 5893 Å.The slit width is 0.1 mm. Calculate the separation between the central maximum and the first secondary minimum. [Ans. 0.5893 cm]

11. A single slit of width 0.4 mm is illuminated by monochromatic light and diffraction bands are observed on a screen 2 m away. If the centre of the second dark is 1.6 cm from the middle of the central bright band, calculate the wavelength of the used light. [Ans. 5600 Å]

12. A parallel beam of monochromatic light of wavelength l = 5896 Å from a narrow slit is diffracted by a plane diffraction grating 12,000 lines per inch, placed with its plane normal to the beam. Calculate the angle at which the second order diffracted image of the slit will be observed. [Ans. 33°51′]

13. A diffraction grating used at normal incidence gives a line l_1 = 6000 Å in a certain order superimposed on another line l_2 = 4500 Å of the next higher order. If the angle of diffraction is 30° how many lines are there in one cm in the grating. [Ans. 2778 lines per cm]

14. A grating has 100 lines ruled on it. What is the difference between two wavelengths that appear just separated in the second order spectrum in the wavelength region 6000 Å? [Ans. 30 Å]

15. What other lines in the range 4000 – 7000 Å will coincide with the fifth order line of 6000Å in a grating spectrum? [Ans. 5000 Å, 4285.7 Å]

16. How many orders will be visible, if the wavelength of incident radiation be 5000 Å and the number of lines on the grating be 2620 to an inch? [Ans. 19]

17. A zone plate is illuminated by sodium light (l = 5896 Å) placed at a distance of 100 cm. If the image of the point source is obtained at a distance of 2 m on the other side, what will be the power of the equivalent lens, which may replace the zone plate without disturbing the set up? Also calculate the radius of the first zone of the plate. [Ans. +1.5D, 0.0627 cm]

18. Calculate the radii of first three clear elements of a zone plate which is designed to bring a parallel beam of light of wavelength 6000 Å to its first focus at a distance of 2 m. [Ans. 0.109 cm, 0.185 cm, 0.245 cm]

19. With respect to a point 50 cm distant, for wavelength 6000 Å, find the number of half period zones contained in a circular hole of radius (*i*) 1 mm and (*ii*) 1 cm. [Ans. (*i*) 3.33, (*ii*) 333]

20. The wavelength of the *C* line of hydrogen atom is 6263 Å. What is the highest order of the *C* line that can be observed with a spectroscope with 6000 lines per cm? Will this line appear to be of the same colour in each order? [Ans. Second, yes]

Unit II

3

Polarization of Light

3.1 INTRODUCTION

Light is electromagnetic wave, with the electric field vector ε and magnetic field vector **B** being mutually perpendicular to its direction of propagation. Ordinary light wave is said to be unpolarized, if it has vibrations in all directions in a plane perpendicular to the direction of propagation, i.e., its electric field vector (ε) vibrates in all directions in a plane perpendicular to the direction of propagation. Unpolarized light is represented in Fig. 3.1. Light waves are transverse waves and so vibrations are at right angles to the direction of propagation. In Fig. 3.1, the direction of propagation is perpendicular to the plane of the paper, while vibrations are in the plane of the paper.

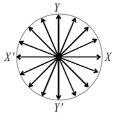


Fig. 3.1: Vibration in unpolarized light.

When a wave has only *y*-displacement, we say that the wave is *linearly polarized* in *y*-direction while a wave having only *z*-displacement is said to be linearly polarized in *z*-direction. Light being a transverse electromagnetic wave, its fluctuating electric and magnetic fields are perpendicular to each other and to the direction of propagation. The direction of polarization of an electromagnetic waves is taken as the direction of the electric field vector ε, not the magnetic field because many common electromagnetic wave detectors respond to the electric forces on electrons in materials not the magnetic forces.

A beam of ordinary light may be considered to consist of two sets of mutually perpendicular vibrations, one vibrating in one plane and the other at right angles to it. In Fig. 3.2 (*a*), the direction of propagation of the beam is in the plane of the paper and the two sets of waves vibrating at right angles to each other, one in the plane of the paper and other perpendicular to it, are respectively represented by arrows [Fig. 3.2 (*b*)] and dots [Fig. 3.2 (*c*)].

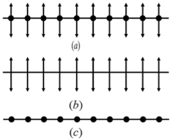


Fig. 3.2

When unpolarized light is passed through a tourmaline crystal cut with its face parallel to its crystallographic axis *AB* (Fig. 3.3), only those vibrations of the incident light pass through the crystal, which are parallel to *AB*. All other vibrations are absorbed. The emergent light consists of vibrations in only one plane, i.e., in the plane of the paper and is said to be *plane polarized*.

This phenomenon of restricting the vibrations (or electric vector ε) of light in a particular plane is called polarization of light. The light wave so obtained is called *polarized light* or *plane polarized light*. The tourmaline crystal here acts as a *polarizer*, as it is used to produce polarized light.

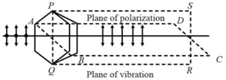


Fig. 3.3

Plane of vibration: The plane in which the vibrations of polarized light are confined is called the *plane of vibration* of the polarized light. In Fig. 3.3, the plane *PQRS* is the plane of vibration.

Plane of polarization: The plane which is perpendicular to the plane of vibration is called the *plane of polarization* of the polarized light. In Fig. 3.3, the plane *ABCD* is the plane of polarization.

3.2 EXPERIMENTAL DEMONSTRATION OF POLARIZATION: TRANSVERSE NATURE OF LIGHT

The phenomenon of polarization of light and the fact that light wave is transverse in nature can be demonstrated experimentally using a simple arrangement shown in Fig. 3.4.

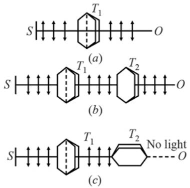


Fig. 3.4

Take a tourmaline crystal T_1 cut it with its face parallel to its crystallographic axis, shown by the vertical dotted line in Fig. 3.4 (*a*). Hold the crystal T_1, normally to the path of a beam of unpolarized light from the source *S* [Fig. 3.4(*a*)] and observe the emergent ray by the naked eye. The emergent beam will be observed to be slightly coloured.

It is observed that when crystal T_1 is rotated, there is no change in the intensity and the colour of the emergent beam of light. Take a second crystal T_2, cut similarly and hold it in the path of the beam so that its axis is parallel to the axis of the crystal T_1 [Fig. 3.4(*b*)]. The beam of light will be seen to pass through the crystals T_2 unaffected.

Now rotate the crystal T_2 about the beam as axis. It will be seen that the intensity of the emergent beam decreases and when the axes of both the crystals are at right angles to each other, no light comes out of crystal T_2, Fig. 3.4(*c*). If the crystal T_2 is further rotated, light reappears and its intensity becomes maximum when the axes of the two crystals become parallel. This occurs after a rotation of T_2 through 90° from its position in Fig. 3.4 (*c*).

The above observations can be explained only if light waves are transverse in nature and the light falling on T_1, being unpolarized, has transverse vibrations in all possible directions. However, T_1 transmits only those vibrations, which are parallel to its axis, i.e., the emergent beam is plane polarized. When this beam falls on T_2, the amount of transmitted light depends upon the relative orientation of T_2 w.r.t. T_1. When the axis of T_2 is parallel to that of T_1, the entire incident light on T_2 is transmitted because vibrations of light transmitted by T_1 are parallel to the axis of T_2. While if the axis of T_2 is perpendicular to that of T_1, no light is transmitted by T_2 as the vibrations of light transmitted by T_1 are normal to the axis of T_2.

Light transmitted by the crystal T_1 is plane polarized as its vibrations lie only in the plane containing the crystallographic axes of T_1. The crystal T_1 is known as *polarizer*, while the crystal T_2 is known as *analyzer*.

A longitudinal wave cannot be polarized as its vibrations are always along the direction of propagation. For example, sound wave (a longitudinal wave) cannot be polarized. Its intensity will not change on passing through a rotating analyzer.

Since the intensity of light changes on rotating the analyzer T_2, due to the change in the relative orientation of its crystallographic axis, *light wave must be transverse in nature*.

3.3 MALUS LAW

When the light is incident on a polarizer, the transmitted light is plane polarized and if this light is made to pass through the analyzer, the intensity varies with the angle between the planes of the polarizer and analyzer. It was experimentally shown by Etiennie Louis Malus in 1809 that the *intensity of the polarized light (I) transmitted through the analyzer varies as the square of cosine of the angle (cos θ) between the plane of transmission of the analyzer and the plane of polarizer.* That is:

$$I \propto \cos^2\theta.$$

Proof: The proof is based on the fact that any polarized light can be resolved into two rectangular components (Fig. 3.5): (*i*) parallel to the plane of transmission of the analyzer, and (*ii*) perpendicular to the plane of analyzer.

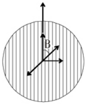


Fig. 3.5

Let a be the amplitude of the incident plane polarized light and θ be the angle that this vibration makes with the optic axis of the analyzer, the amplitude parallel to the plane of transmission of the analyzer is $a\cos\theta$. The perpendicular component $a\sin\theta$ will not be transmitted.

Thus, the transmitted component through the analyzer is $a\cos\theta$, whose intensity is given by:

$$I = (a\cos\theta)^2 = a^2\cos^2\theta$$

$$\text{or, } I = I_o\cos^2\theta \quad ...(3.1)$$

where $I_o = a^2$ = intensity of light transmitted by the polarizer.

$$\therefore I \propto \cos^2\theta$$

which is Malus law.

When polarizer and analyzer are parallel: θ = 0° or 180°:

$$\Rightarrow \cos\theta = \pm 1$$

$$\therefore I = I_o$$

When polarizer and analyzer are perpendicular:

$$\theta = 90° \Rightarrow \cos\theta = 0°$$

$\therefore I = 0$, *i.e.*, no light is transmitted across the analyzer.

3.4 POLARIZATION BY REFLECTION

Malus discovered in 1808 that the light can be polarized by reflection from the surface of a glass. He found that *polarized light is obtained, when an unpolarized monochromatic beam of light is reflected from a plane sheet of glass.* The extent of polarization depends upon the angle of incidence. The reflected light may be completely polarized, partially polarized or remains unpolarized depending upon the angle of incidence. For a particular angle of incidence, called *polarizing angle* (i_p) or *angle of polarization*, the reflected beam of light is completely plane polarized. For a glass surface, the value of polarizing angle is 57.5°.

Consider a beam of light incident along *AO* (Fig. 3.6) on the glass surface at polarizing angle i_p. The light reflected along *OB* is completely plane polarized. The light refracted along *OC* remains unpolarized. Light can also be polarized by reflection from water surface.

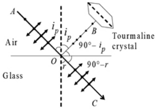


Fig. 3.6

The production of polarized light by reflection can be explained as follows. The vibrations of the incident light can be resolved into components, parallel to the glass surface and perpendicular to the glass surface. Only a portion of the component, parallel to glass surface gets reflected while the other components are transmitted. Therefore, the light reflected by glass is plane polarized which can be verified by rotating the tourmaline crystal, as discussed earlier.

3.5 BREWSTER'S LAW

In 1811, Brewster performed a number of experiments to study the polarization of light by reflection at the surfaces of different media. According to Brewster, *the tangent of the angle of polarization (i_p) is equal to the refractive index (μ) of the medium.* That is,

$$\mu = \tan i_p \quad ...(3.2)$$

This is known as Brewster's law. The reflected light in this case is completely plane polarized in the plane of incidence. The angle i_p is sometimes referred to as the Brewster angle.

The reflected and the refracted rays are found to be perpendicular to each other.

Proof: Let r_p be the angle of refraction corresponding to the polarizing angle i_p (Fig. 3.6). At polarizing angle of incidence, the reflected and refracted rays are perpendicular to each other.

$$\text{That is, } i_p + r_p = 90°$$

$$\text{or, } r_p = 90° - i_p$$

From Snell's law, refractive index of the transparent medium is given by:

$$\mu = \frac{\sin i}{\sin r}$$

$$\text{At } i = i_p,\ \mu = \frac{\sin i_p}{\sin r_p} = \frac{\sin i_p}{\sin(90° - i_p)} = \frac{\sin i_p}{\cos i_p}$$

$$\text{or, } \mu = \tan i_p$$

which proves Brewster's law.

For a crown glass of refractive index μ = 1.52, the polarizing angle can be obtained using Brewster's law. Thus:

$$\mu = 1.52$$

$\therefore \mu = 1.52 = \tan i_p$

or, $i_p = \tan^{-1}(1.52) = 56.6°$

For water surface, $\mu = \dfrac{4}{3} = 1.3333$

$i_p = \tan^{-1}(1.3333) = 53.1°$

Proof of perpendicularity of reflected and refracted light: From Snell's law of refraction, we have:

$$\mu = \frac{\sin i}{\sin r}$$

At $i = i_p$, $\mu = \dfrac{\sin i_p}{\sin r_p}$

From Brewster's law,

$$\mu = \tan i_p = \frac{\sin i_p}{\cos i_p}$$

From the above two expressions, we have:

$$\frac{\sin i_p}{\sin r_p} = \frac{\sin i_p}{\cos i_p}$$

$\sin r_p = \cos i_p$

$\sin r_p = \sin(90° - i_p)$ [□ $\cos\theta = \sin(90° - \theta)$]

$r_p = 90° - i_p$

$i_p + r_p = 90°$.

Brewster Window: One practical use of Brewster's law is in the design of a glass window, which can transmit all the incident light, i.e., `100% transmission of light or the glass window without any reflection. Such a window is called a *Brewster window* and is used in lasers.

When light is incident on a clean glass plate, most of the incident light is transmitted (» 92%) while a small fraction (» 8%) is reflected back. In a gas laser, light travels through the glass window about a hundred times, due to which the intensity falls to nearly:

$(0.92)^{100}$ » 3×10^{-4}.

Thus, the transmitted light beam has practically no intensity.

To overcome this problem, the window is tilted so that the light beam is incident at Brewster angle. After about hundred transmissions, the final beam is nearly plane polarized.

The light component having vibrations at right angles to the plane of incidence is reflected back. After about 100 reflections at the Brewster window, the transmitted light will have nearly 50% intensity of the incident light and will be completely plane polarized. Thus, nearly 50% of the incident light is retained while the other half is reflected back. Brewster's windows are used in gas lasers.

Example 3.1

An electromagnetic beam has an intensity of 28 Wm^{-2} and is linearly polarized in the vertical direction. Find the intensity of the transmitted beam by a polaroid, when its plane of transmission makes an angle of 45° with the vertical.

Solution:

Here, $I_0 = 28\ Wm^{-2}$, $\theta = 45°$

Using Malus law, $I = I_0\cos^2\theta$, we have

$I = 28 \times \cos^2 45°$

$$= 28 \times \left(\frac{1}{\sqrt{2}}\right)^2 = \frac{28}{2} = 14\ \mathrm{Wm}^{-1}$$

Example 3.2

Two Nicols are so oriented that the maximum amount of light is transmitted. To what fraction of its maximum value is the intensity of transmitted light reduced when the analyzer is rotated through (*i*) 30°, (*ii*) 60°?

Solution:

According to Malus law, intensity of transmitted light through the analyzer is given by:

$I = I_0\cos^2\theta$

(*i*) Here, $\theta = 30°$

$\therefore I = I_0(\cos 30°)^2$

$$= I_0\left(\frac{\sqrt{3}}{2}\right)^2 = \frac{3}{4}I_0 = 0.75I_0$$

That is, its maximum intensity is reduced to 75% of the maximum intensity.

(*ii*) Here, $\theta = 60°$

$\therefore I = I_0(\cos 60°)^2$

$$= I_0\left(\frac{1}{2}\right)^2 = \frac{I_0}{4} = 0.25I_0$$

This is, its maximum intensity is reduced to 25% of its maximum intensity.

Example 3.3

The velocity of light in water is 2.2×10^8 m/s. What is the polarizing angle of incidence for water surface? (Speed of light in free space = 3×10^8 m/s).

Solution:

Refractive index of water

$$= \frac{\text{Speed of light in space}}{\text{Speed of light in water}}$$

$$\text{or, } \mu = \frac{3 \times 10^8}{2.2 \times 10^8} = 1.3636$$

Using Brewster's law, $\mu = \tan i_p$, we have:

$$i_p = \tan^{-1}\mu$$

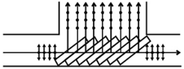
$$= \tan^{-1}(1.3636)$$

$$= 53.74°$$

3.6 PILE OF PLATES: POLARIZATION BY REFRACTION

When light is incident on a surface of glass or any other transparent medium, only a small fraction of the incident light is reflected and most of the incident light is refracted or transmitted through the medium. According to Brewster's law, when the light is incident on a glass plate at the polarizing angle, the light reflected from both the upper and lower surfaces of the plate is *completely plane polarized in the plane of the incidence*, i.e., having vibrations perpendicular to the plane of the incidence. The *intensity of the reflected light is very low, as 100% of light vibrations parallel to the plane of the incidence and 85% of the perpendicular vibrations are transmitted and only 15% of the perpendicular vibrations are reflected.* Thus, only about $\frac{15}{185} \cong 0.08$ or *8% of the incidence light is reflected at each reflection.*

In order to increase the intensity of the plane polarized light, the number of reflections can be increased by putting a number of glass plates together, i.e., we can use a pile of glass plates to increase the intensity of the totally plane polarized reflected light. Accordingly, as shown in Fig. 3.7, a set of about ten thin glass plates is put one upon another in a wooden tube with their planes at an angle of 32.5°, so that when a beam of light parallel to the axis falls on the glass plate the incidence angle is 57.5°, the polarizing angle for glass. As the process of reflection continues, more and more vibrations perpendicular to the plane of incidence, get reflected while the parallel vibrations are transmitted as the refracted beam. The larger the number of plates, the greater is the intensity of the reflected plane polarized light. Also, the transmitted light becomes free from the perpendicular vibrations and so the percentage of the plane polarized light increases. However, it is never completely plane polarized and always contains some perpendicular vibrations.



Pile of plates

Fig. 3.7

A similar pile can be used as an analyzer for detecting the plane polarized light. When the plane of incidence of the two piles are the same, the intensity of reflected light in the analyzer is maximum, while when their planes are perpendicular, the intensity is minimum.

3.7 Double Refraction

When a ray of light of unpolarized light is allowed to pass through calcite crystal ($CaCO_3$ or iceland spar) or quartz (SiO_2—both transparent to both visible and ultraviolet radiations), *two refracted rays* are produced, unlike in a glass in which there is only one refracted ray. This phenomenon is known as *double refraction* or *birefringence* and the crystals showing double refraction are called *doubly refracting crystals*.

The physical properties of isotropic medium such as glass, are the same in all directions. When the light is incident on such a medium refraction takes place in only one direction, in accordance with Snell's law. Thus, only one refracted ray is produced. However, for anisotropic medium such as calcite, quartz, tourmaline, etc., the physical properties are different in different directions. When the light is incident on such crystals, two polarized refracted rays are produced [Fig. 3.8 (*a*)].

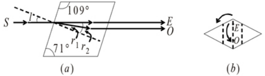


Fig. 3.8

This phenomenon can be demonstrated in a simple manner by putting a dot on a paper and viewing it through a calcite crystal. Two images of the dot will be observed [Fig. 3.8 (*b*)]. On rotating the crystal about the incident beam as axis, one image remains stationary, while the second image rotates around the first. The stationary image is known as *ordinary image* and the refracted rays, which produce this image are known as *ordinary rays* or *O-rays*, as they obey the ordinary laws of refraction. The second image, which rotates about the ordinary image, is known as *extraordinary image* and the refracted rays, which produce this image are known as *extraordinary rays* or *E-rays*, as they do not follow the ordinary laws of refraction.

Since the two opposite faces of a calcite crystal are always parallel so that the two refracted rays, ordinary and extraordinary, emerge parallel to the incident beam and, therefore, are parallel to each other. When the ray is incident normally, the ordinary ray passes straight and undeviated through the crystal, whereas the extraordinary ray is refracted at some angle but emerges out parallel to and displaced from the incident ray. As indicated in Fig. 3.8(*b*), on rotating the crystal about the beam as axis, the *E*-ray rotates about the fixed *O*-ray and the line joining the *O* and *E* images always remains parallel to the shorter diagonal of the emergent crystal face.

Both *O*- and *E*-rays are plane polarized and their planes of polarization are at right angles to one another, the vibrations of ordinary ray being normal to the plane of paper while those of the extraordinary ray in the plane of the paper. This can be verified by viewing the images through a tourmaline plate. When it is rotated, both images undergo a change in intensity; if one intensity is increasing, the other is decreasing and vice versa, i.e., if one is at its brightest, the other is not visible. Thus, the two images are at the brightest twice in one complete rotation of the tourmaline.

The refractive indices for *O*-ray is:

$$\mu_o = \frac{\sin i}{\sin r_1}$$

and that of *E*-ray is:

$$\mu_E = \frac{\sin i}{\sin r_2}$$

In calcite crystal, $r_1 < r_2$, therefore, $\mu_O > \mu_E$. As $\mu = \frac{v_{\text{free space}}}{v_{\text{medium}}}$, therefore,

$$v_O < v_E$$

i.e., inside a calcite crystal, the O-rays travel slower than the E-rays, in other words, in a calcite crystal the E-rays travel faster than the O-rays. The refractive index is the same for all angles of incidence for the O-rays, while it varies with the angle of incidence for the E-rays. Therefore, the O-rays travel with the same speed in all the directions within the crystal, while the E-rays travel with different velocities in different directions within the crystal.

3.8 Uniaxial and Biaxial Crystals

In the doubly refracting crystals such as calcite, quartz, etc., there is a direction known as *optic axis*, along which both *E*-rays and *O*-rays travel with the same velocity, i.e., no double refraction takes place if the unpolarized light is incident in the direction of the optic axis. In this regard, doubly refracting crystals are of two types: (*i*) uniaxial crystals and (*ii*) biaxial crystals.

Uniaxial Crystals: The crystals in which there is only one optic axis, are known as *uniaxial crystals*. Calcite, quartz, tourmaline, ice, nitrate of soda, etc. are uniaxial crystals.

Biaxial Crystals: The crystals which are characterized by the presence of two optic axes are known biaxial crystals. There are two directions of one uniform velocity. Borax, mica, selenite, aragonite, topaz, etc. are biaxial crystals.

We shall consider here the details of calcite crystal.

Calcite Crystal: Chemically, it is crystallized calcium carbonate ($CaCO_3$) and is also known as iceland spar because it is found in Iceland. It is transparent to visible as well as to ultraviolet light. It crystallizes in many forms and can be reduced by cleavage or breakage into a rhombohedron, as shown in Fig. 3.9. Each of the six faces of this crystal is a parallelogram having angles of 102° and 78° (precisely it is 101°5' and 78°5').

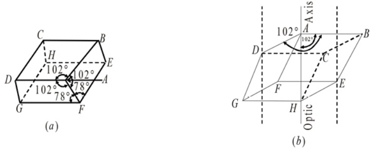


Fig. 3.9

Optic Axis: At the two opposite corners *A* and *H* of the rhombohedron, all the three angles of the faces are obtuse. These corners *A* and *H* are known as blunt corners of the crystal. *A line passing through one of the blunt corners (A or H), making equal angles with each of the three edges, gives the direction of the optic axis. Any line parallel to this line is also an optic axis. In fact, the optic axis is a direction and not a particular line.* In a special case, when the three edges of the crystal are equal, the line joining the two blunt corners *A* and *H* coincides with the crystallographic axis of the crystal and it gives the direction of the optic axis [Fig. 3.9 (*b*)]. *A ray of light incident along the optic axis or in a direction parallel to it will not split into two rays.* Thus, the phenomenon of double refraction is absent when the light is incident on the crystal along the optic axis.

Principal Section: A plane containing the optic axis of the crystal and perpendicular to the opposite faces of the crystal is called the *principal section* of the crystal. As a crystal has six faces, therefore, for every point inside the crystal there are three principal sections corresponding to each pair of opposite faces. A principal section always cuts the surface of a calcite crystal into a parallelogram with angles 109° and 71°.

Principal Plane: A plane in the crystal containing the optic axis and the ordinary ray is called the *principal plane of the ordinary ray*. Similarly, a plane containing the optic axis and the extraordinary ray is called the *principal plane of the extraordinary ray*. In general, the two planes do not coincide as the ordinary ray always lies in the plane of incidence, while it is generally not so for the extra-ordinary ray. In the particular case, when the plane of incidence is a principal section of the crystal, then the principal section of the crystal and the principal planes of the ordinary and the extraordinary rays coincide.

3.9 NICOL PRISM

In 1828, William Nicol invented an optical device, which is used for production and analysis of plane polarized light. This device is known as *Nicol prism*. It is widely used in optical instruments.

Principle: The working of a Nicol prism is based on the phenomenon of double refraction. When a beam of unpolarized light is passed through a doubly refracting crystal such as calcite, it splits into two plane polarized beams: (*i*) the ordinary ray with vibrations perpendicular to the principal section of the crystal and (*ii*) the extraordinary ray with vibrations parallel to the perpendicular to each other. Thus, the two rays are plane polarized, having vibrations perpendicular to each other. If somehow, one of the two rays is eliminated, then the ray emerging through the crystal will be *plane polarized*. In the Nicol prism, the *O*-ray is eliminated by total internal reflection and the *E*-ray is transmitted through the prism. Thus, plane polarized light with vibrations in the principal section of the crystal is obtained.

Construction: A calcite crystal whose length is three times its breadth is taken. The two end faces *AB* and *CD* of the crystal are cut in such a way that they make angles of 68° and 71° (Fig. 3.10). The resulting crystal is then cut into two parts along the plane A_1C_1 passing through the blunt corners along a plane perpendicular to the principal section such that A_1C_1 makes an angle of 90° with ends C_1D and A_1B.

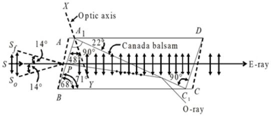


Fig. 3.10: Nicol prism.

The cut faces which are perpendicular to the end faces are ground, polished optically flat and then cemented together with a layer of Canada balsam. Canada balsam is a transparent substance, whose refractive index lies midway between the refractive indices of *O*-ray and *E*-ray for calcite. For sodium light with λ = 5893Å, refractive index of calcite for *O*-ray μ_o = 1.658, refractive index of Canada balsam μ_{CB} = 1.550 and refractive index of calcite for *E*-ray μ_E = 1.486.

Thus, *Canada balsam layer is optically denser than calcite for E-ray and rarer for O-ray*. The two ends of the prism are open for the incidence and emergence of light, while the top and bottom surfaces are coated with lamp black and are covered by a brass tube.

Working: When an unpolarized beam of light *SP* is incident on one end face of the prism parallel to BC_1, it splits up into ordinary and extraordinary rays. As discussed earlier, Canada balsam is optically denser than calcite for *E*-ray, while it is less dense than calcite for *O*-ray.

The refractive index for *O*-ray with respect to Canada balsam layer is:

$$\mu = \frac{1.658}{1.550}$$

If θ_c is critical angle, then:

$$\sin\theta_c = \frac{1.550}{1.658} = 0.935$$

$$\text{or, } \theta_c = \sin^{-1}(0.935) \cong 69°$$

As the length of the crystal is large, the angle of incidence of the *O*-ray at Canada-balsam layer is greater than critical angle of incidence, Hence, the *O*-ray is totally internally reflected and is absorbed by the lamp black coating on the side BC_1. The *E*-ray travels from an optically rarer medium, calcite to an optical denser medium, Canada balsam and, therefore, is not affected and is transmitted through the prism. Thus, a plane polarized light, with vibrations in the principal section parallel to the shorter diagonal of the end face of the crystal, is obtained. Thus, *the Nicol prism acts as a polarizer*.

Limitations: If the angle of incidence is less than the critical angle for the *O*-ray, it is not reflected and is transmitted through the prism and so the prism does not act as an polarizer. A Nicol prism can act as a polarizer only for slightly convergent and slightly divergent beam of light. It fails to be an effective polarizer with too convergent or too divergent incident beams. If the angle $S_o PS$ becomes greater than 14°, the *O*-ray will be incident at an angle less than the critical angle and so, it will also be transmitted, so that the transmitted light is not plane polarized.

The extraordinary ray also has a limit beyond which it is totally internally reflected by the Canada balsam layer. This is due to the variation of refractive index of calcite with the direction of propagation of *E*-ray. The refractive index for the *E*-ray is μ_E = 1.486, when it is travelling at right angles to the optic axis. Along the optical axis, both *E*- and *O*-rays travel with the same velocity and have the same refractive index, i.e., $\mu_E = \mu_O$ = 1.658. For the intermediate angles, the effective value of μ_E lies between 1.486 and 1.658. For a particular angle of incidence, μ_E may be more than μ_{CB} = 1.550 and at the same time, the angle of incidence on the balsam layer will be more than the critical angle. Therefore, the *E*-ray will also get totally internally reflected at the Canada balsam layer. No light will then be transmitted through the prism. The dimensions of the Nicol prism is such that the maximum value of the angle SPS_f for the *E*-ray to be transmitted by the prism is 14°.

A Nicol prism, therefore, cannot be used as a polarizer for highly convergent or divergent beams. To avoid the transmission of *O*-ray and the total internal reflection of *E*-ray and so, to get a pure plane polarized light, the angle between the extreme rays of the incident beam should be less than 28°.

Nicol Prism as an Analyzer: Nicol prism can also be used as an analyzer, i.e., it can be used for the detection of the plane polarized light.

Consider two Nicol prisms *P* and *A* arranged coaxially, one after another [Fig. 3.11 (*a*)]. When a beam of unpolarized light is incident on *P*, the emergent beam is plane polarized with its vibrations in the principal section of the prism. Thus, the prism *P* acts as a polarizer. The beam emerging from the polarizer is incident on the second prism *A* called the *analyzer*. When the principal section of the prism *A* is parallel to that of the prism *P*, the vibration of the plane polarized incident light will be parallel to the principal section of *A* and the incident ray is completely transmitted and the emergent beam has the maximum intensity. In this position, the prisms are referred to as *parallel Nicols*.

As the prism *A* is gradually rotated from this position, the intensity of the emergent beam decreases in accordance with Malus law and *when the two prisms are crossed* (i.e., the principal section of *A* is at right angle to the principal section of *P*), *no light comes out* of the analyzer *A*. In this position, the vibration of the incident *E*-ray on *A* is perpendicular to the principal section of *A* and, therefore, acts as O-ray and is totally internally reflected by Canada balsam layer. Thus, the prism *A* acts as an analyzer. The combination of *P* and *A* is together referred to as *polariscope*.

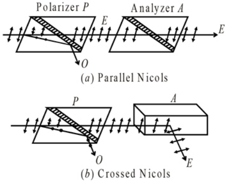


Fig. 3.11

For the intermediate angles between the parallel and crossed positions, the incident *E*-ray on *A* has two components of vibrations—parallel and perpendicular to the principal section of *A*. The parallel component of vibration is transmitted while the perpendicular component is totally internally reflected. Thus, a reduced brightness is seen through the analyzer *A*.

Huygen's Explanation of Double Refraction

Huygen explained double refraction using the principle of secondary wavelets. A point source of light in a doubly refracting crystal is the origin of the two wavefronts. *For the O-ray*, for which the velocity of light is the same in all directions, *the wavefront is spherical*. While *for the E-ray*, for which the velocity is different in different directions, *the wavefront is an ellipsoid of revolution*. However, in the direction of optic axis, the velocities of the *O*-ray and the *E*-ray are the same.

As shown in Fig. 3.12 (*a*), *S* is a point source of light in calcite crystal. The spherical wavefront represents the *O*-rays and the ellipsoid wave surface represents the *E*-rays at any instant of time. Such crystals in which *ordinary wave surface lies within the extraordinary wave surface are called negative crystals. In these crystals, E-rays travel faster than O-rays*. Calcite is a negative crystal. For negative crystals $\mu_O > \mu_E$. The velocity of the extraordinary ray varies as the radius vector of the ellipsoid. It is minimum and equal to the velocity of the *O*-ray along the optic axis and is maximum at right angles to the direction of the optic axis.

In crystals, such as quartz, the extraordinary wave surface, lies within the ordinary wave surface [Fig. 3.12 (*b*)]. Such crystals are *called positive crystals. In these crystals E-rays travel slower than O-rays. For positive crystals*, $\mu_O < \mu_E$. The velocity is maximum and equal to the velocity of the *O*-ray along optic axis and is minimum at right angles to the direction of optic axis. There are two points at which sphere and ellipsoid touch each other and the line joining these two points is the optic axis.

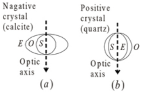


Fig. 3.12

3.10 ELLIPTICALLY AND CIRCULARLY POLARIZED LIGHT: INTERFERENCE OF POLARIZED LIGHT

According to electromagnetic theory of light, light consists of electric and magnetic vectors vibrating in mutually perpendicular directions, both being perpendicular to the direction of propagation of light. Most of the optical effects of the light wave can be explained by the electric vector ε and therefore, it is also called the *light vector*. In a plane or linearly polarized light, the light vector vibrates simple harmonically, along a straight line, perpendicular to the direction of propagation of light, i.e., the orientation of light vector remains unchanged but its magnitude undergoes a periodic change during vibrations. When two plane polarized light waves superimpose, then if the resultant light vector rotates such that it traces a circle or an ellipse, the light is said to be *circularly* or *elliptically polarized*. In the former case, the magnitude of the resultant vector remains constant and it changes in the later case.

Let a beam of plane polarized light of wavelength λ and amplitude *A* is incident normally upon a uniaxial doubly refracting crystal, with its faces cut parallel to the optic axis [Fig. 3.13 (*a*)]. Suppose the incident light is vibrating along *PQ* making an angle θ with the optic axis of the crystal as shown in Fig. 3.13 (*b*). On entering the crystal, vibration of the incident light will have component along *PO* equal to *A* sinθ and that along *PE* equal to *A* cosθ. Thus, the incident ray splits into extraordinary ray and ordinary ray, having amplitudes *A* cosθ and *A* sinθ respectively along the optic axis of the crystal and perpendicular to it.

(*a*) Uniaxial doubly refracting negative crystal

(*b*) Components of incident wave

Fig. 3.13

Both the rays travel along the same direction but with different velocities. Therefore, on emerging through the crystal, a phase difference is introduced between the two rays. The exact value of phase difference Δ depends upon the thickness '*t*' of the crystal, Thus, the incident plane polarized light, on emerging out of the crystal consists.of two simple harmonic vibrations, in two mutually perpendicular planes having the same period but different amplitudes and phases. They can be superimposed into a single motion which can be linear, circular or elliptical depending on their amplitude, phase and the value of θ.

Let Δ be the phase difference introduced between the two components in passing through the crystal. Then by a suitable choice of the origin of time, the displacement of the two rays in a negative crystal can be represented as

E-ray: $x = A\cos\theta \sin(\omega t + \Delta)$

O-ray: $y = A\sin\theta \sin \omega t$

Putting, $A \cos\theta = a$ and $A\sin\theta = b$, we have:

$$x = a\sin(\omega t + \Delta) \quad ...(3.3)$$

$$y = b\sin\omega t \quad ...(3.4)$$

From Eq. (3.4), we have:

$$\frac{y}{b} = \sin \omega t$$

$$\therefore \cos\omega t = \sqrt{1-\sin^2 \omega t} = \sqrt{1-\frac{y^2}{b^2}}$$

From Eq. (3.3), we have:

$$\frac{x}{a} = \sin\omega t\cos\Delta + \cos\omega t \sin\Delta$$

or, $$\frac{x}{a} = \frac{y}{b}\cos\delta + \sqrt{1-\frac{y^2}{b^2}}\cdot\sin\delta$$

or, $$\frac{x}{a} - \frac{y}{b}\cos\delta = \sqrt{1-\frac{y^2}{b^2}}\cdot\sin\delta$$

Squaring and rearranging, we get:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{2xy}{ab}\cos\delta = \sin^2\Delta \quad ...(3.5)$$

This is the general equation of ellipse with its major and minor axes inclined to the coordinate axes.

Special cases:

1. Linear Polarization: When $\Delta = 0$ or $\Delta = 2n\pi$ (n = 0, 1, 2, 3, ...),

$\sin\Delta = 0$ and $\cos\Delta = 1$, therefore, Eq. (1.5) becomes,

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{2xy}{ab} = 0$$

or, $$\left(\frac{x}{a} - \frac{y}{b}\right)^2 = 0$$

or, $$\pm\left(\frac{x}{a} - \frac{y}{b}\right) = 0$$

or, $$y = +\frac{b}{a}x$$

which is the equation of a straight line passing through the origin and having a slope $\left(\frac{b}{a}\right)$. This indicates that the emergent light is linearly or plane polarized [Figs. 3.14 (*a*) and (*f*)].

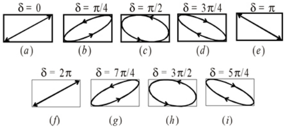


Fig. 3.14

For $\Delta = \pi, 3\pi, 5\pi, ...$ or $(2n + 1)\pi$ (n = 0, 1, 2, 3, ...), we have

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{2xy}{ab} = 0$$

or, $$\left(\frac{x}{a} + \frac{y}{b}\right)^2 = 0$$

or, $$y = -\frac{b}{a}x$$

which is also the equation of a straight line but with a slope $\left(-\frac{b}{a}\right)$ in the negative direction of x-axis. In this case also, the emergent light is linearly or plane polarized but the direction of vibration makes an angle $\left(2\tan^{-1}\frac{b}{a}\right)$ with that of the incident light [Fig. 3.14 (*e*)].

2. Elliptical Polarization: When the thickness of the plate is such that the phase difference between the two rays is:

$$\Delta = \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \ldots \text{ or}$$

$$(2n-1)\frac{\pi}{2}\,(n = 1, 2, 3, \ldots)$$

so that, $\sin\Delta = 1$ and $\cos\Delta = 0$

and, a ? b (or θ ? 45°), then Eq. (3.5) reduces to

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

which is the equation of an ellipse. Thus, the emergent light is *elliptically polarized*, the plane of the ellipse being normal to the direction of propagation [Figs. 3.14 (*c*) and (*h*)].

If the thickness of plate is such that $\delta = \frac{\pi}{4}, \frac{3\pi}{4}, \ldots,$ then

$$\sin\Delta = \cos\delta = \frac{1}{\sqrt{2}},$$ so that,

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\sqrt{2}\,\frac{xy}{ab}=\frac{1}{2}$$

which is again an equation of ellipse, but with major axis inclined in the positive and negative directions of the x-axis respectively for $\Delta = \pi/4$ and $3\pi/4$ as shown in Figs. 3.14 (*b*) and (*d*). So the emergent light is elliptically polarized.

3. Circular Polarization: When the thickness of the plate is such that the phase difference between the two rays is

$$\Delta = \frac{\pi}{2}, \frac{3\pi}{2}, \ldots \quad \text{or} \quad (2n-1)\frac{\pi}{2}\,(n = 1, 2, 3, \ldots)$$

and $\theta = 45°$ so that $a = b$, then Eq. (3.5) becomes

$$x^2 + y^2 = a^2$$

which is the equation of a circle with radius a. Thus, the emergent light is circularly polarized (Fig. 3.15).

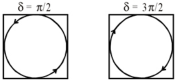


Fig. 3.15

In general, the resultant of two plane polarized lights is an elliptically polarized light. Under certain conditions, however, the resultant light is plane or circularly polarized. Thus, the plane polarized light and circularly polarized light are the special cases of elliptically polarized light. For $0 < \Delta < \pi$, the rotation of the electric vector is related to the direction of propagation in the same sense as the rotation of right-handed screw is rotated to the direction of translation. To the observer, the rotation is *anticlockwise* and the light is *left-handed elliptically polarized*. For $\pi < \Delta < 2\pi$, the rotation of electric vector appears *clockwise* and the light is said to the *right-handed elliptically polarized.*

3.11 QUARTER- AND HALF-WAVE PLATES

The phase difference introduced between E-ray and O-ray on emerging out from a crystal depends upon the thickness of the crystal. If the thickness of the plate is properly chosen, desired amount of phase difference can be introduced between the two components. Such a plate is known as *retardation plate* because they retard the motion of one of the rays. We discuss here two such plates:

1. Quarter-Wave Plate: *It is a uniaxial, doubly refracting crystal plate (e.g., calcite or quartz) of a suitable thickness cut with its optic axis parallel to the refracting faces, such that a phase difference of* $\frac{\pi}{2}$ *or a path difference of* $\frac{\lambda}{4}$ *is introduced between O-ray and E-ray on emerging through the plate.*

When a beam of plane polarized monochromatic light of wavelength λ is incident normally on such a crystal, it splits up into O and E components, which travel in the same direction but with different velocities.

In case of negative crystal, such as calcite, E-ray travels faster than O-ray so that $\mu_O > \mu_E$. If t is the thickness of the plate, then the path difference between the two components is:

$$= (\mu_O - \mu_E)t$$

For a quarter-wave plate, path difference between the two components:

$$= \frac{\lambda}{4}$$

$$\therefore (\mu_O - \mu_E)t = \frac{\lambda}{4}$$

$$\text{or, } t = \frac{\lambda}{4(\mu_O - \mu_E)} \qquad \ldots(3.6)$$

For positive crystals, like quartz, $\mu_E > \mu_O$, so that:

$$t = \frac{\lambda}{4(\mu_E - \mu_O)} \qquad \ldots(3.7)$$

If a plane polarized light with vibration inclined at an angle of 45° to the optic axis is incident on a quarter-wave plate, the emergent light is circularly polarized. If the angle is not 45°, the emergent light is elliptically polarized. A quarter-wave plate is the simplest device for producing and detecting circularly and elliptically polarized light.

2. Half-Wave Plate: *It is also a uniaxial doubly refracting crystal plate (e.g., calcite or quartz) of suitable thickness, cut with its optic axis parallel to the refracting faces such that a phase difference of* π *or a path difference of* $\frac{\lambda}{2}$ *is introduced between O-ray and E-ray on emerging through the crystal.*

When a beam of plane polarized monochromatic light of wavelength λ is incident normally on such a crystal, it splits up into O and E components which travel in the same direction but with different velocities. In case of negative crystal such as calcite, E-ray travels faster than O-ray so that $\mu_O > \mu_E$. If t is the thickness of the plate, then the path difference between the two components is:

$$= (\mu_O - \mu_E)t$$

For a half-wave plate, path difference between the two components $= \frac{\lambda}{2}$

$$\therefore (\mu_O - \mu_E)t = \frac{\lambda}{2}$$

$$\text{or, } t = \frac{\lambda}{2(\mu_O - \mu_E)} \quad ...(3.8)$$

For positive crystals like quartz, $\mu_E > \mu_O$, so that thickness of half wave plate

$$t = \frac{\lambda}{2(\mu_E - \mu_O)} \quad ...(3.9)$$

Since the path difference depends upon the wavelength of the light used, the refractive indices corresponding to yellow light of wavelength λ = 5893Å are generally used for calculating the required thickness of a quarter-wave or half-wave plate. A half-wave plate is used in construction of Laurent's half-wave device in polarimeter.

When a plane polarized light is passed through a half-wave plate, the emergent light is also plane polarized for all orientation of the plate with respect to the plane of vibration of the incident light, but now the direction of vibration is inclined at an angle 2q w.r.t. the incident light, where θ is the angle at which the direction of vibration in the incident light makes with the optic axis.

These plates are generally made of quartz by cutting it parallel to optic axis and then polishing to make them optically plane.

Example 3.4

Calculate the minimum thickness of a quarter-wave plate made of calcite crystal for the light of wavelength 5500 Å. The principal refractive indices for calcite are 1.652 and 1.488.

Solution:

For calcite crystal $\mu_O > \mu_E$.

$$\therefore \mu_O = 1.652, \mu_E = 1.488$$

$$\lambda = 5500\text{Å} = 5500 \times 10^{-8} \text{ cm}$$

$\therefore$ Thickness of quarter-wave plate:

$$t = \frac{\lambda}{4(\mu_O - \mu_E)}$$

$$= \frac{5500 \times 10^{-8}}{4 \times (1.652 - 1.488)}$$

$$= \frac{5500 \times 10^{-8}}{4 \times 0.164} = 8.38 \times 10^{-5} \text{ cm.}$$

Example 3.5

Calculate the thickness of half-wave plate for sodium light, if μ_O = 1.54 and the ratio of velocity of ordinary and extraordinary wave is 1.007. Is the crystal positive or negative?

Solution:

Here, $\frac{v_O}{v_E} = 1.007, \mu_O = 1.54, \lambda = 5893$ Å

$= 5893 \times 10^{-8}$ cm

Since the velocity of E-ray is less than that of O-ray ($v_O = 1.007 v_E$), the crystal is a positive crystal.

Also, $v \propto \frac{1}{\mu}$, therefore:

$$\frac{v_O}{v_E} = \frac{\mu_E}{\mu_O} = 1.007$$

$$\text{or, } \mu_E = 1.007\mu_O = 1.007 \times 1.54 = 1.55$$

Thickness of half-wave plate:

$$t = \frac{\lambda}{2(\mu_E - \mu_O)} = \frac{5893 \times 10^{-8}}{2(1.55 - 1.54)}$$

$$= \frac{5893 \times 10^{-8}}{2 \times 0.01} = 2.95 \times 10^{-3} \text{ cm.}$$

Example 3.6

The values of μ_E and μ_O for quartz are 1.5508 and 1.5418 respectively. Calculate the phase retardation for λ = 5000Å if the thickness of the plate is 0.032 mm.

Solution:

The path difference between E- and O-rays when they traverse a plate of thickness t is:

$$p = (\mu_E - \mu_O)t$$

$\therefore$ Phase difference or phase retardation

$$= \frac{2\pi}{\lambda} \times p = \frac{2\pi}{\lambda}(\mu_E - \mu_O)t$$

Here, $\mu_E = 1.5508, \mu_O = 1.5418$

$\lambda = 5000\text{Å} = 5000 \times 10^{-8}$ cm,

t = 0.032 mm = 0.0032 cm

$\therefore$ Phase retardation

$$= \frac{2 \times \pi}{5000 \times 10^{-8}}(1 - 5508 - 1.5418) \times 0.0032$$

$$= \frac{2\pi \times 0.009 \times 0.0032}{5000 \times 10^{-8}}$$

$= 1.152\pi$ radians.

Example 3.7

What should be the thickness of quarter-wave plate for a light of wavelength 5890Å if $\mu_E = 1.553$ and $\mu_o = 1.544$?

Solution:

We know

$$t = \frac{\lambda}{4(\mu_E - \mu_0)}$$

Here, $\lambda = 5890$ Å $= 5890 \times 10^{-8}$ cm

$m_E = 1.553$ and $\mu_o = 1.554$

$$\therefore t = \frac{5890 \times 10^{-8}}{4(1.553 - 1.544)}$$

$$= \frac{5890 \times 10^{-8}}{4 \times 0.009}$$

$= 1.636 \times 10^{-3}$ cm

Example 3.8

A given calcite plate behaves as a half-wave plate for a particular wavelength λ. Assuming variation of refractive index with λ to be negligible, how would the same plate behave for the light of wavelength 2λ.

Solution:

The thickness of a half-wave plate of calcite crystal for wavelength λ is given by:

$$t = \frac{\lambda}{2(\mu_O - \mu_E)} = \frac{2\lambda}{4(\mu_O - \mu_E)} = \frac{\lambda'}{4(\mu_O - \mu_E)}$$

Thus, for wavelength λ' = 2λ, it will behave as a quarter-wave plate.

Example 3.9

Calculate the thickness of a calcite plate which would convert plane polarized light into circularly polarized light. The principal refractive indices at the wavelength 5890Å of the used light are $m_O = 1.658$ and $\mu_E = 1.486$.

Solution:

The plane polarized light is converted into circularly polarized light if the plate introduces a path difference of $\frac{\lambda}{4}$ or an odd multiple of $\frac{\lambda}{4}$ between ordinary and extraordinary rays. Hence, if t is the thickness of the calcite plate, then:

$$(\mu_O - \mu_E)t = \frac{(2n-1)\lambda}{4}, n = 1, 2, 3, \ldots$$

$$\text{or, } t = \frac{(2n-1)\lambda}{4(\mu_O - \mu_E)}$$

Here, $\lambda = 5890\text{Å} = 5890 \times 10^{-8}$ cm, $\mu_O = 1.658$, $\mu_E = 1.486$

$$\therefore t = \frac{(2n-1)5890 \times 10^{-8}}{4(1.658 - 1.486)}$$

$= 8.56 \times 10^{-5} (2n - 1)$ cm

For $n = 1$: $t = 8.56 \times 10^{-5}$ cm

Thus, the least thickness is 8.56×10^{-5} cm and all the odd multiples of this will convert plane polarized light into circularly polarized light.

Example 3.10

A calcite plate of thickness 0.0020 cm is cut with optic axis parallel to the face. If $\mu_O = 1.648$ and $\mu_E = 1.481$ (ignoring variation with wavelength), what are the wavelengths in the visible range (take 4000Å to 7600Å) for which the plate behaves as a half-wave plate and also those for which the plate behaves as a quarter-wave plate.

Solution:

For a half-wave plate:

$$(\mu_O - \mu_E)t = (2n-1)\frac{\lambda}{2}, n = 1, 2, 3, \ldots$$

$$\text{or, } \lambda = \frac{2(\mu_O - \mu_E)t}{(2n-1)}$$

$$= \frac{2(1.648 - 1.481) \times 0.0020}{(2n-1)} \text{ cm}$$

$$= \frac{66800}{(2n-1)} \times 10^{-8} \text{ cm} = \frac{66800}{(2n-1)} \text{Å}$$

For $n = 5, 6, 7, 8$ we get λ = 7422Å, 6073Å, 4453Å in the visible region, for which the given calcite plate behaves as a half-wave plate.

For quarter-wave plate:

$$(\mu_O - \mu_E)t = (2n-1)\frac{\lambda}{4}, n = 1, 2, 3, \ldots$$

$$\text{or, } \lambda = \frac{4(\mu_0 - \mu_E)t}{(2n-1)}$$

$$= \frac{4(1.648 - 1.481) \times 0.0020}{(2n-1)} \text{ cm}$$

$$= \frac{133600}{(2n-1)} \times 10^{-8} \text{ cm} = \frac{133600}{(2n-1)} \text{Å}$$

On substituting, we find that for n = 10, 11, 12, 13, 14, 15, 16, 17, respectively, we get = 7032Å, 6362Å, 5807Å, 5344Å, 4948Å, 4607Å, 4310Å, 4048Å, in the visible region for which the given calcite plate behaves as a quarter-wave plate.

3.12 PRODUCTION OF PLANE, CIRCULARLY AND ELLIPTICALLY POLARIZED LIGHT

1. Plane Polarized Light: To produe plane polarized light, a beam of unpolarized light is passed through a Nicol prism. As the beam enters the Nicol prism, it splits up in ordinary and extraordinary components. The ordinary component is totally internally reflected at the Canada balsam layer and is absorbed, while the extraordinary component passes through the Nicol prism. The emergent beam is plane polarized, having its vibration parallel to the shorter diagonal of the end face of the prism.

2. Circularly Polarized Light: To produce circularly polarized light, *the two plane polarized light waves vibrating at right angles to each other, having the same amplitude and time period, having a phase difference of π/2 or a path difference of* $\frac{\lambda}{4}$ *should be made to interfere.*

For this purpose, a parallel beam of monochromatic light is allowed to fall on a Nicol prism N_1 (Fig. 3.16). The beam on passing through N_1 gets plane polarized. Another Nicol prism N_2 is placed at some distance from N_1 such that N_1 and N_2 are in crossed position. Therefore, the field of view, seen through N_2, will be dark. Now, a quarter-wave plate P is mounted on a tube T_1. The tube T_1 can rotate about the outer fixed tube T_2 placed between the prisms N_1 and N_2. Now, when light from N_1 passes through P to N_2, the field of view will be no longer dark and it turns bright. The quarter-wave plate P is rotated until the field of view is dark. Keeping P fixed, tube T_1 is rotated till the mark S on the plate P coincides with zero mark on T_1. Thereafter, by rotating the quarter-wave plate P, the mark S is made to coincide with 45° mark on the tube T_1.

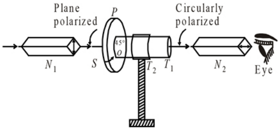


Fig. 3.16

The quarter-wave plate P is now in the proper position and the vibrations of the plane polarized light, falling on the plate P, make an angle of 45° with the direction of the optic axis of the quarter-wave plate P. In this condition, the plane polarized light on entering the plate P splits up into ordinary and extraordinary components having equal amplitude and time period and on emerging out of the plate P, the beam is circularly polarized. If the Nicol prism N_2 is rotated now, the field of view is uniform in intensity, similar to the ordinary light passing through a Nicol prism.

3. Elliptically Polarized Light: To produce elliptically polarized light, *the two plane polarized light waves vibrating at right angles to each other having unequal amplitudes, should have a phase difference of* $\frac{\pi}{2}$ *or a phase difference of* $\frac{\lambda}{4}$.

The same arrangement and procedure, shown in Fig. 3.16, and discussed earlier, can be used for the purpose. The only difference being that in this case, the vibrations of the plane polarized light falling on the quarter-wave plate should not make an angle of 45° with the optic axis and so, the mark S need not coincide with 45° mark on the tube T_1. Thus, to produce elliptically polarized light, a parallel beam of unpolarized light is allowed to fall on Nicol prism N_1. Another prism N_2 is placed such that N_1 and N_2 are crossed and so the field of view is dark. A quarter-wave plate P is introduced between N_1 and N_2 so that the plane polarized light from N_1 falls normally on the quarter-wave plate. The field of view gets illuminated as the light emerging from the plate P is elliptically polarized. It should be ensured that the vibrations of the plane polarized light falling on the plate P do not make an angle of 45° with the optic axis, otherwise light will be circularly polarized. When the prism N_2 is rotated, the intensity of illumination of the field of view change from maximum to minimum.

ANALYSIS OF POLARIZED LIGHT

(i) Plane Polarized Light: The beam of light under investigation is allowed to fall on a Nicol prism. The Nicol prism is rotated. If the light gets completely extinguished (complete darkness or intensity reduces to zero) twice in one complete rotation of the Nicol prism, the beam is plane polarized.

(ii) Circularly Polarized Light: The circularly polarized light, when observed through a rotating Nicol prism shows no change in the intensity of illumination.

The same is observed when ordinary unpolarized light is viewed through a rotating Nicol prism. In this respect there is no difference between circularly polarized and unpolarized light.

To distinguish between circularly polarized and unpolarized light, the beam is allowed to fall on a quarter-wave plate. On passing through a quarter-wave plate, circularly polarized light gets converted into plane polarized light, while unpolarized light remains unchanged. The light emerging out of the quarter-wave plate is viewed through a rotating Nicol prism. If the original beam is circularly polarized, light is completely extinguished twice in each rotation of the Nicol prism. While if the original beam is unpolarized, no variation in the intensity of illumination is observed when viewed through the rotating Nicol prism.

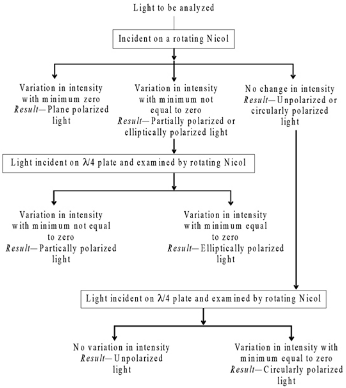
In brief, when a circularly polarized light is allowed to fall on a quarter-wave plate and then viewed through a rotating Nicol prism, the light gets completely extinguished twice in each rotation of the prism.

(iii) Elliptically Polarized Light: The beam of light is allowed to fall on a rotating Nicol prism. If the beam is elliptically polarized, the intensity of illumination changes from a maximum to a minimum (not zero). The intensity is maximum when the principal plane of the Nicol prism is parallel to the major axis of the ellipse and minimum, when parallel to the minor axis. However, the observed variation of intensity in this case is the same as what is observed when a mixture of unpolarized light and plane polarized light is incident on a rotating Nicol prism.

To differentiate elliptically polarized light from a mixture of unpolarized and plane polarized light, the beam under investigation is passed through a quarter-wave plate. The quarter-wave plate converts the elliptically polarized light into plane polarized light which, when viewed through a rotating Nicol prism, gets completely extinguished twice in each cycle. While the mixture of unpolarized and polarized light is not completely extinguished.

The above discussion is presented in Table 3.1 for the analysis of polarized light.

Table 3.1



3.13 OPTICAL ACTIVITY

Consider two Nicol prisms placed in crossed position, as shown in Fig. 3.17. When light falls on the Nicol N_1, no light emerges from Nicol N_2, Fig. 3.17 (*a*). However, if a quartz plate cut with faces parallel to the optic axis is introduced between N_1 and N_2 such that the plane polarized light from N_1 falls normally on the quartz, the light emerges out from N_2, Fig. 3.17 (*b*). In the first case (i.e., without quartz), the plane polarized light from N_1 is not able to pass through N_2 because of its vibration being perpendicular to the principal plane of N_2. However, in the latter case (i.e., with quartz), the quartz rotates the plane of polarization of the plane polarized light from N_1 and so the vibration is no longer perpendicular to the principal plane of N_2 and is therefore, transmitted. Also, if the Nicol N_2 is rotated slowly about the direction of propagation of light at some angle, there is again complete extinction of light. This shows that the *light emerging from quartz is still plane polarized but the plane of polarization has rotated through a certain angle*, as shown in Fig. 3.18. The amount of rotation through which the plane of vibration is turned, depends upon the *thickness* of the quartz plate and the *wavelength of light*.

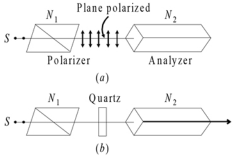


Fig. 3.17

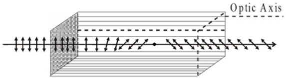



Fig. 3.18: Rotation of plane of polarization.

This property of rotating the plane of vibration of plane polarized light, above its direction of propagation by certain crystals or substances, is known as optical activity. The phenomenon is optical rotation and such substances are known as optically active substances. Crystals such as quartz, sugar crystals, sodium chlorate, cinnabar and solutions like turpentine, sugar solution, guanine sulphate solution, etc. are optically active substances. However, calcite is not an optically active substance.

There are two types of optically active substances:

(i) Right-Handed or Dextro Rotatory: These are optically active substances which *rotate the plane of polarization in a clockwise direction* w.r.t. an observer, when looking towards light travelling towards him. In other words, these substances rotate the plane of vibration to the right w.r.t. an observer when looking towards light travelling towards him. Quartz and cane sugar solution are such substances.

(ii) Left-Handed or Laevo-Rotatory: These are optically active substances which *rotate the plane of polarization in an anticlockwise direction* w.r.t. an observer looking towards light propagating towards him. In other words, these substances rotate the plane of vibration to the left w.r.t. an observer, when looking towards light travelling to him.

Biot Observations: Biot, in 1815, made an extensive study of the optical rotation and made the following observations:

1. The amount of rotation (θ) produced by an optically active substance is directly proportional to the length traversed in the medium by the light, i.e.,

$$\theta \propto l$$

2. In case of solution and vapours, the amount of rotation produced by a given length is proportional to the concentration (c) of the solution or vapour, i.e.,

$$\theta \propto c$$

3. The rotation produced is inversely proportional to the square of the wavelength λ of the light, i.e.,

$$\theta \propto \frac{1}{\lambda^2}$$

Thus, it is least for red colour light and the highest for violet colour light.

4. When there is more than one optically active substance, the total rotation produced is the algebraic sum of the rotations (θ_1, θ_2, θ_3,...) produced by individual substances, i.e.,

$$\theta = \theta_1 \pm \theta_2 \pm \theta_3 + \ldots$$

The rotation in anticlockwise direction is taken as positive while that in clockwise direction, is taken as negative.

Cause of Optical Activity: It is the atomic/molecular arrangement of the substance which is responsible for its optical activity. Rench suggested that the optically active crystals are made up of atomic layers which are slightly twisted from one another, so that the atoms have a spiral-like arrangement. In the right-handed or dextro-rotatory crystals, the layers are built up clockwise around the optic axis, whereas in the left-handed or laevo-rotatory crystals, the layers are built up anticlockwise.

In quartz (SiO_2) crystal, silicon (Si) and oxygen (O) atoms are arranged in the form of spiral around the optic axis (Fig. 3.19). The optical activity arises due to the fact that these spiral of atoms form planes which are slightly twisted from one another.

Fig. 3.19

In liquids, the optical activity is due to the molecular structure and they have certain structural symmetry in their molecules. The molecules of an optically active compound possess at least one tetravalent atom (e.g., carbon atom) surrounded by four different groups of atoms in tetrahedral arrangement. Fig. 3.20 represents the tetrahedral arrangement in atomic groups *P*, *Q*, *R*, *S* surround the *C*-atom. In Fig. 3.20 (*a*), *Q*, *R* and *S* are arranged anticlockwise while in Fig. 3.20 (*b*) they are arranged clockwise. One structure is mirror image of the other about *PC*. No rotation in space can make one structure coincide with the other. This suggests that the optical activity is due to the mirror image asymmetry. Thus, the essential requirement for such asymmetry is that a tetravalent atom such as *C* or *S* is attached to four different atomic groups.

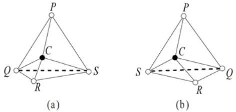


Fig. 3.20

FRESNEL'S THEORY OF OPTICAL ROTATION

Fresnel's theory is based on the fact that a linearly polarized light can be considered a resultant of two circularly polarized vibrations rotating in opposite directions with the same frequency.

Fresnel's Assumptions: *(i)* When a beam of plane polarized light enters in a crystal along the optic axis, it splits up into two circularly polarized vibrations, one clockwise and the other anticlockwise.

(ii) In an optically inactive substance like calcite, the two circularly polarized vibrations travel with the same angular velocity.

(iii) In an optically active substance like quartz, the two circularly polarized vibrations travel with different velocities. In a dextro-rotatory substance, the velocity of right-handed circularly polarized light is greater than the left-handed circularly polarized light. While in a laevo-rotatory substance the velocity of left-handed circularly polarized light travels faster than right-handed circularly polarised light. Hence, in an optically active substance, a phase difference is developed between the two beams while traversing through the substance.

(iv) After emerging from the crystal, the two circular vibrations recombine to produce a plane polarized light whose plane of vibration or polarization has been rotated w.r.t. the incident beam. The angle of rotation depends upon the phase difference developed between the two beams by the crystal.

Light Incident on Calcite: Let a plane polarized beam be incident on a calcite, crystal which is an optically inactive substance. According to Fresnel's theory, the incident plane polarized light, on entering the crystal, splits up into two opposite circularly polarized vibrations rotating with the same angular velocity. In Fig. 3.21, *OL* is circularly polarized vector rotating in anticlockwise direction and *OR* is circularly polarized vector rotating in clockwise direction. The resultant vector of *OL* and *OR* is *OA*. Circularly polarized light is the resultant of two rectangular components having a phase difference of $\pi/2$.

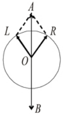


Fig. 3.21

For clockwise circular vibrations

$$x_1 = a\cos\omega t$$
$$y_1 = a\sin\omega t$$

For anticlockwise vibrations:

$$x_2 = -\,a\cos\omega t$$
$$y_2 = a\sin\omega t$$

Therefore, the resultant vibrations along the *x*-axis

$$x = x_1 + x_2 = a\cos\omega t - a\cos\omega t = 0$$

and that along the *y*-axis, $y = y_1 + y_2 = a\sin\omega t + a\sin\omega t$

$$= 2a\sin\omega t$$

Thus, the resultant vibration has an amplitude $2a$ and is plane polarized. The plane of vibration is the same as that of the incident vibration. The calcite crystal does not rotate the plane of vibration, i.e., it is optically inactive.

Light Incident on Quartz: In case, the linearly polarized light on entering the crystal, splits up into two circularly polarized vibrations rotating in opposite directions. For right-handed quartz crystal, the clockwise (or right-handed) vibrations travel faster than the anticlockwise (or left-handed) vibrations. As a result, on emerging out, the clockwise vibrations are rotated by a greater angle Δ than the anticlockwise vibrations. The resultant of these vibrations (vectors *OR* and *OL*) is along *OA'*. Therefore, the resultant vibrations are along *A'B'*. Before entering the crystal, the plane of vibrations is along *AB*. Therefore, the plane of vibration has rotated by an angle of $\Delta/2$. The value of this angle depends upon the thickness of the crystal. Hence, the phase difference between the two components is Δ. Thus, for clockwise vibrations:

$$x_1 = a\cos(\omega t + \Delta)$$
$$y_1 = a\sin(\omega t + \Delta)$$

for anticlockwise vibrations: $x_2 = -\,a\cos\omega t$

$$y_2 = a\sin\omega t$$

Therefore, the resultant displacement along the two axes are:

$$x = x_1 + x_2 = a\cos(\omega t + \Delta) - a\cos\omega t$$

$$= 2a\sin\frac{\delta}{2}\sin\left(\omega t + \frac{\delta}{2}\right) \quad ...(3.10)$$

$$y = y_1 + y_2 = a\sin(\omega t + \Delta) + a\sin\omega t$$

$$= 2a\cos\frac{\delta}{2}\sin\left(\omega t + \frac{\delta}{2}\right) \quad ...(3.11)$$

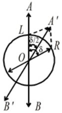


Fig. 3.22

It is evident that the resultant vibrations along the x-axis and y-axis have the same phase. Therefore, the resultant vibrations is plane polarized and it makes an angle $\frac{\delta}{2}$ with the original direction. Therefore, the plane of vibration has rotated through an angle $\frac{\delta}{2}$ on passing through the crystal.

Dividing Eq. (3.10) by Eq. (3.11), we get:

$$\tan\frac{\delta}{2} = \frac{x}{y}$$

If μ_L and μ_R are the refractive indices of the anticlockwise and clockwise vibrations respectively, then the path difference in passing through a thickness d of the crystal is equal to $(\mu_L - \mu_R)\, d$. If λ be the wavelength of the light, then,

$$\text{phase difference; } \Delta = \frac{2\pi}{\lambda} \times \text{path difference}$$

$$\text{or, } \Delta = \frac{2\pi}{\lambda} \times (\mu_L - \mu_R)d$$

Let the rotation in the plane of vibration be $\theta\left(=\frac{\delta}{2}\right)$, so that:

$$\theta = \frac{\delta}{2} = \frac{\pi}{\lambda}(\mu_L - \mu_R)d \qquad \text{...(3.12)}$$

If v_L and v_R are the velocities of the left-handed and right-handed circularly polarized lights respectively and c the speed of light, then,

$$\mu_L = \frac{c}{v_L} \text{ and } \mu_R = \frac{c}{v_R}$$

Substituting these values, we have:

$$\theta = \frac{\pi d}{\lambda}\left(\frac{c}{v_L} - \frac{c}{v_R}\right) \qquad \text{...(3.13)}$$

$$\text{or, } \theta = \frac{\pi d}{T}\left(\frac{1}{v_L} - \frac{1}{v_R}\right) \left(\because \frac{c}{\lambda} = \frac{1}{T}\right) \qquad \text{...(3.14)}$$

where T is the time period.

For the left-handed optically active crystals, $\mu_R > \mu_L$

$$\theta = \frac{\pi d}{\lambda}(\mu_R - \mu_L) \qquad \text{...(3.15)}$$

$$\text{or, } \theta = \frac{\pi d}{T}\left(\frac{1}{v_R} - \frac{1}{v_L}\right) \qquad \text{...(3.16)}$$

Fresnel's Experiment: Fresnel showed experimentally that when plane polarized light enters an optically active medium, it is resolved into two circularly polarized vibrations which travel with different velocities.

As shown in Fig. 3.23, Fresnel arranged a number of right-handed (R) and left-handed (L) quartz prisms, each having its optic axis parallel to its base. In order to verify Fresnel's theory, a beam of plane polarized light is allowed to fall normally on the face AB of the first prism. If the theory is correct, the incident plane polarized beam will split up into two circularly left-handed and right-handed polarized beams, which travel with the different velocities but in the same direction. When the beam is incident on the second prism (L), the beam which was faster in the first prism becomes slower in the second prism and vice versa. The second prism is optically denser for right-handed beam and hence bends towards normal, while the left-handed beam bends away from the normal. Thus, the two beams are separated as they pass through the prism L. Again at the boundary of the next prism R, the speeds are interchanged and the beam that is bent towards the normal in L, moves away from the normal. Thus, the two beams are separated more and more as they propagate through the arrangement and when the two beams emerge out, they are widely separated. When these beams are analyzed using a quarter-wave plate and a Nicol prism, both are found to be circularly polarized verifying the correctness of the theory.

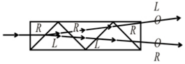


Fig. 3.23

3.14 SPECIFIC ROTATION

Optically active solutions (e.g., sugar solution, camphor in alcohol, etc.) rotate the plane of polarization of light. The angle of rotation depends upon: (*i*) the thickness of the medium, (*ii*) concentration of the solution or the density of optically active substance in the solvent, (*iii*) wavelength of light and (*iv*) temperature of the solution.

The optical activity of a substance is measured in terms of *specific rotation*. It is defined as *the rotation produced by a decimetre* (= 10 cm) *long column of liquid containing one gram of the active substance in one c.c. of the solution*. That is

$$S_\lambda^t = \frac{\theta}{lC} \quad \text{...(3.17)}$$

where S_λ^t represents the specific rotation at temperature t°C for the light of wavelength λ, θ is the angle of rotation in degrees, l is the length of the solution in decimetre and C the concentration of solution in gm per c.c. Its unit is degree (decimetre)$^{-1}$ (g/cm^3)$^{-1}$. If the length of the liquid is taken in cm, the specific rotation can expressed as:

$$S_\lambda^t = \frac{10\theta}{lC} \quad \text{...(3.18)}$$

3.15 POLARIMETErS

The device which is used to measure the angle through which the plane of polarization is rotated by an optically active medium is called *polarimeter*. When a polarimeter is calibrated to directly read the percentage of sugar in a solution, it is called a *saccharimeter*. We discuss here two types of polarimeter: (*i*) Laurent's half-shade polarimeter and (*ii*) bi-quartz polarimeter.

Laurent's Half Shade Polarimeter: The arrangement of various components of a Laurent's half-shade polarimeter is shown in Fig. 3.24. It consists of two Nicol prisms N_1 and N_2. N_1 acts as an polarizer, while N_2 acts as an analyzer. Next to N_1 is a half-shade device H, one half of which is a half-wave plate of quartz Q which covers one-half of the field of view while the other half is a glass plate G. T is a glass tube having a larger diameter in the middle. The optically active solution is filled in this tube. The larger diameter at the middle ensures that there is no bubble (even if it is present) in the path of light. The tube is closed at the ends by cover-slips. The Nicol N_2 is capable of rotation about a common axis of N_1 and N_2. The rotation of the analyzer can be read on a vernier scale attached to a telescope.

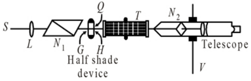
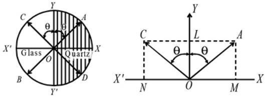


Fig. 3.24: Laurent's half shade polarimeter.

Light from a monochromatic source S is incident on a converging lens L which renders the incident light into a parallel beam. After passing through N_1, the beam gets plane polarized and is incident on the half-shade device.

Action of Half-Shade Device: A half-shade device, as shown in Fig. 3.35 (*a*) consists of a semi-circular plate of quartz, cut with faces parallel to the optic axis with thickness such that it acts as a half-wave plate, i.e., it introduces a path difference of l/2 or a phase difference of π between the *E*- and *O*-vibrations. The other half is made up of a semicircular plate of glass of suitable thickness such that it transmits and absorbs the same amount of light as the quartz plate.

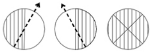


(a) (b)

Fig. 3.25

Let the plane of vibration of the plane polarized light be parallel to *AB*, i.e., the principal plane of polarizing Nicol N_1 is parallel to *AB*, inclined at an angle θ to the optic axis *YY'* of the quartz half. On passing through the glass half, the vibrations of light remain in the same plane *AB*, while it is not so in the quartz half. On passing through the quartz, the half, the beam is split up into the *E*-component along the optic axis (*YY'*) of the quartz and the *O*-component perpendicular to the optic axis, i.e., along *XX'*. On emerging from the quartz two components differ in phase by π, the *O*-component gains a phase of π over the *E*-component as the velocity of the *O*-component in quartz is higher than that of the *E*-component. Therefore, the direction of the *O*-component is reversed, i.e., from initial position *OM*, it is now represented by *ON*. Now, the resultant of *E*-component *OL* and *O*-component *ON* is *OC* making an angle θ with *Y*-axis, instead of the original resultant *OA*. *Thus, the effect of quartz plate is to rotate the plane of polarization by* 2θ. *Hence, the plane of vibrations of the beam emerging out from the quartz plate is along CD*. Whereas the plane of vibrations of the beam emerging from the glass plate remains along *AB*.

Thus, there are two plane polarized beams [Fig. 3.25 (*b*)], one emerging from the glass half with vibrations along *OA*, while the other emerging from the quartz half with vibrations along *OC*. If the principal plane of the analyzing Nicol is parallel to *AOB*, then the light from the glass half will pass unobstructed, while light from the quartz half will be partly obstructed by the analyzing Nicol and hence the *glass half will appear brighter than the quartz half*. On the other hand, if the principal plane of N_2 is parallel to *COD*, the light from the quartz half will pass unobstructed while light from glass half will be partly obstructed by the analyzing Nicol and hence the *quartz half will appear brighter than the glass half*. If, however, the principal plane of N_2 is parallel to *YY'*, it is equally inclined to the two plane polarized lights and hence the *field of view will be equally bright* (or more precisely, equally dark). The appearance of the circular plate, in three positions is shown in Fig. 3.26 (*a*), (*b*) and (*c*) respectively.



(a) (b) (c)

Fig. 3.26

It is obvious from the above discussion that the half shade serves the purpose of dividing the field of view into two halves. Even a small change in the position of analyzing Nicol N_2 from the position of equal brightness produces a marked change in the brightness of the two halves and so the angle of rotation of the plane of polarization can be accurately measured using this polarimeter. However, a particular half shade device works for the light of particular wavelength only.

Determination of Specific Rotation: To determine specific rotation given by Eq. (3.17), rotation θ produced by the length *l* of the solution of concentration *C*, needs to be measured.

For this purpose, the tube *T* is first filled with water and the analyzer N_2 is set in the position of equal brightness of the field of view and readings of the two verniers are noted on the two verniers are noted on the circular scale. Now the tube is filled with the optically active solution (say, sugar solution) of known concentration. The analyzer is rotated and the position of equal brightness is again obtained. The new positions of the two verniers are again noted on the circular scale. The difference in the two readings of the same vernier gives the angle of rotation produced by the solution. The process is repeated for different concentrations of the solution. A graph is plotted between θ and *C*, which comes out to be a straight line. The slope $\left(=\frac{\theta}{C}\right)$ is calculated and using Eq. (3.17) the specific rotation can be calculated.

Biquartz Polarimeter: A biquartz polarimeter is identical to a Laurent's half-shade polarimeter except the half-shade device is replaced by a biquartz plate and instead of monochromatic (sodium) light, white light is used. It is more accurate for the determination of the angle of rotation produced by an optically active substance than Laurent's half-shade polarimeter.

Action of Biquartz Plate: It consists of two semicircular plates of quartz, each of thickness 3.75 mm. One half consists of left-handed optically active quartz while the other half is right-handed optically active quartz. Both halves are cut perpendicular to the optic axis and joined together along the diameter *PQ*, as shown in Fig. 3.27. Both the halves rotate the plane of polarization for yellow light by 90°.

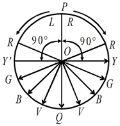


Fig. 3.27

When white light, rendered plane polarized by the polarizer, travels through the biquartz normally, travelling along the optic axis (since the plates are cut perpendicular to the optic axis), the phenomenon of *rotatory dispersion* takes place, *i.e., different colours suffer different rotations*. For longer wavelength, the dispersion is less while for the shorter wavelength, it is more. For yellow colour, the rotation is 90° and *YOY′* is a straight line.

If the principal plane of the analyzer is parallel to *POQ*, the yellow light will not be transmitted through the analyzer while the red and blue greyish-violet will be present in the same proportion in each half. Thus, the appearance of the two halves is identical. The two halves have a greyish-violet tint (colour), called the *tint of passage* or *sensitive tint*. When the analyzer is rotated to one side from this position, one half of the field of view appears blue, while the other half appears red. If the analyzer is rotated in the opposite direction, the colours are changed, i.e., the first half which was blue earlier, appears red and the second half which was red earlier, appears blue. The position of sensitive tint is highly sensitive and so the angle of optical rotation can be accurately determined.

Example 3.11

For quartz, the refractive indices for right-handed and left-handed vibrations are 1.55810 and 1.55821 respectively for λ = 4000Å. Find the amount of optical rotation produced at λ = 4000Å by a quartz plate of thickness 2 mm with its face perpendicular to the optic axis.

Solution:

The angle of rotation of the plane of polarization is given by:

$$\theta = \frac{\pi d}{\lambda}(\mu_L - \lambda_R)$$

Here, d = 2mm = 0.2 cm, λ = 4000Å = 4.0 × 10⁻⁵ cm

$\mu_L = 1.55821$, $\mu_R = 1.55810$

∴ Optical rotation,

$$\theta = \frac{3.14\times0.2}{4.0\times10^{-5}}(1.55821-1.55810)$$

= 1.726 radians

$$= \frac{1.726\times180}{3.14} = 98^\circ$$

Example 3.12

The rotation in the plane of polarization for the light of wavelength = 5893Å by a certain substance is 10°/cm. Calculate the difference between the refractive indices for right and left circularly polarized light in the substance.

Solution:

Here, θ = 10°, d = 1 cm, λ = 5893Å = 5893×10^{-8} cm

We know, $\theta = \frac{\pi d}{\lambda}(\mu_R - \mu_L)$

$$\therefore \mu_R - \mu_L = \frac{\theta\lambda}{\pi d} = \frac{10\times\pi\times5893\times10^{-8}}{\pi\times180\times1}$$

$$\left(10^\circ = \frac{10\times\pi}{180}\text{ radians}\right)$$

$= 3.27\times10^{-7}$.

Example 3.13

A tube of length 20 cm containing 10% sugar solution rotates the plane of polarization by 13.2°. What is the specific rotation of the sugar solution?

Solution:

The specific rotation is given by:

$$S = \frac{10\theta}{lC}$$

Here, $\theta = 13.2°$, $C = 0.1$ g/cc, $l = 20$ cm

$$\therefore S = \frac{10 \times 13.2}{20 \times 0.1} = 66°.$$

Example 3.14

A 20 cm long tube containing sugar solution is placed between two crossed Nicols and illuminated by light of wavelength 6000Å. If the specific rotation is 60° and optical rotation produced is 12°, what is the strength of the solution?

Solution:

Here, $l = 20$ cm, $\lambda = 6000Å = 6.0 \times 10^{-5}$ cm

$S = 60°, \theta = 12°$

We know, $S = \frac{10\theta}{lC}$

∴ Concentration or strength of the solution:

$$C = \frac{10\theta}{lS} = \frac{10 \times 12}{20 \times 60} = \frac{1}{10} \text{gm / cc} = 0.1 \text{ g / cc}.$$

Example 3.15

A length of 25 cm of a solution containing 50 gm of solute per litre causes a rotation of the plane of polarization of light by 5°. Find the rotation of the plane of polarization by a length 60 cm of solution containing 100 gm of solute per litre.

Solution:

Here, $l = 25$ cm,

$$C = \frac{50}{1000} = 0.05 \text{ g / cc}, \ \theta = 5°$$

∴ Specific rotation of the solution,

$$S = \frac{10\theta}{lC} = \frac{10 \times 5}{25 \times 0.05} = 40°$$

For the length, $l = 60$ cm and

$$C = \frac{100}{1000} \text{ gm / cc} = 0.1 \text{ g / cc}$$

∴ Rotation produced,

$$\theta = \frac{SlC}{10} = \frac{40 \times 60 \times 0.1}{10} = 24°.$$

Excercises

Short Answer Type

1. What do you mean by 'polarization'?
2. Why do we talk of restriction of electric field vector in the polarization phenomenon?
3. Define and depict (*i*) plane of vibration and (*ii*) plane of polarization of a plane polarized light.
4. Can sound waves be polarized? Why?
5. State Malus Law.
6. Can light be polarized by reflection? Define polarizing angle. What is its value for the glass surface?
7. State Brewster's Law.
8. What is a Brewster window? Explain.
9. What is Double Refraction? Why are ordinary and extraordinary rays so called?
10. Name a crystal each for which (*i*) $\mu_O > \mu_E$ and (*ii*) $\mu_E > \mu_O$.
11. Define Optic Axis and Principal Plane.
12. Giving two examples of each, distinguish between uniaxial and biaxial crystals.
13. Differentiate between positive and negative crystals. Give examples of both.
14. Briefly explain the working of a Nicol prism.
15. Why Canada balsam is a suitable cementing material in a Nicol prism?
16. What is the main limitation of a Nicol prism?
17. Point out the conditions for the production of elliptically and circularly polarized light.
18. What is a half-wave plate? What is its use?
19. What is the use of a quarter-wave plate?
20. How circularly polarized light can be produced and analyzed?
21. Explain the procedure for the production and detection of elliptically polarized light.
22. What do you mean by 'optical rotation'? Give one example each of dextro-rotatory and laevo-rotary substance.
23. Mention Biot's observations about optical rotation.
24. Define specific rotation. On what factors does it depend? Write its unit.

25. What is 'rotatory dispersion'?

26. What is 'polarimeter'? What is the difference between Laurent's half-shade and bi-quartz polarimeters?

27. What is a saccharimeter?

28. What do you mean by 'tint of passage'?

Long Answer Type

1. Distinguish between unpolarized and plane polarized light. Explain one method for obtaining plane polarized light.

2. Describe an experiment to show the transverse nature of light.

3. State and explain Brewster Law. Show that when light is incident on a transparent substance the polarizing angle, the reflected and refracted rays are at right angles to each other.

4. Describe the process of polarization using a pile of plates.

5. Why certain crystals like calcite show double refraction, while glass does not? In calcite crystal which ray travels faster? Why?

6. Distinguish between uniaxial and biaxial crystals. Explain the terms optic axis, principal section and principal plane.

7. What is Nicol prism? Explain its principle, construction and working. Discuss its role as a polarizer and an analyzer.

8. Discuss the usefulness and the limitations of a Nicol prism.

9. Give the Huygen's explanation of double refraction phenomenon.

10. Considering the interference of polarized light, obtain the condition for the production of (*a*) linearly, (*b*) elliptically and (*c*) circularly polarized light.

11. What is quarter-wave plate and half-wave plate? Obtain the expressions for required thickness both positive and negative crystals.

12. Discuss the method of production and analysis of (*i*) plane, (*ii*) circularly and (*iii*) elliptically polarized light.

13. What is optical activity? Differentiate between right-handed and left-handed rotations.

14. Describe Fresnel's theory of optical rotation and obtain an expression for the rotation produced in the plane of polarization

15. Describe the construction and working of Laurent's half-shade polarimeter.

16. In what way a bi-quartz polarimeter differs from Laurent's half shade polarimeter? Discuss its working.

Numericals

1. A ray of light is incident on the surface of a glass plate of refractive index 1.55 at the polarizing angle. Calculate the angle of refraction. [**Ans.** 32°50′]

2. A polarizer and an analyzer are oriented so that the amount of light transmitted is maximum. To what percentage of its maximum value is the intensity of the transmitted light reduced when the analyzer is rotated through (*a*) 30°, (*b*) 45°, (*c*) 60° and (*d*) 90°. [**Ans.** (*a*) 7.5%, (*b*) 50%, (*c*) 25%, (*d*) 0%]

3. A polarizer and analyzer are oriented so that the amount of light transmitted is maximum. How will you orient the analyzer so the transmitted light is reduced to (*a*) 0.5, (*b*) 0.25, (*c*) 0.75, (*d*) 0.125, (*e*) 0 of its maximum value? [**Ans.** (*a*) 45°, (*b*) 60°, (*c*) 30°, (*d*) 69°, (*e*) 90°]

4. Calculate the thickness of a quarter-wave plate for light of wavelength 5893Å, for which the *E*- and *O*-refractive indices are 1.553 and 1.544 respectively. [**Ans.** 1.673×10^{-5} cm]

5. The refractive index of calcite is 1.648 for *O*-ray and 1.486 for *E*-ray. A plane parallel plate of calcite is cut of thickness 0.01 mm. For what wavelengths in the visible region will this plate behave as a (*i*) quarter-wave plate, (*ii*) half-wave plate.

[**Ans.** (*i*) 720, 589.9, 498.5 and 432 nm, (*ii*) 648 and 462.9 nm]

6. Calculate the thickness of a doubly refracting crystal required to introduce a path difference of l/2 between the *O*-and *E*-rays, when λ = 6000 Å, m_O = 1.55 and m_E = 1.54. [**Ans.** 0.03 mm]

7. A beam of plane polarized light is converted into circularly polarized light by passing it through a crystal slice of thickness 3×10^{-5} m. Calculate the difference in the refractive indices of the two rays inside the crystal assuming the above thickness to be the minimum value required to produce the observed effect for the wavelength of light 600 nm. [**Ans.** 5×10^{-3} cm]

8. The plane of polarization of a plane polarized light is rotated through 6.5° in passing through a length of 20 cm of sugar solution of 5% concentration. Calculate the specific rotation of the sugar solution.

[**Ans.** 65°]

9. A polarimeter tube of 25 cm containing a sugar solution rotates the plane of polarization through 10°. If the specific rotation of sugar is 60°, calculate the strength of sugar solution. [**Ans.** 6.66%]

10. The refractive index of quartz for *O*-ray is 1.544 and the maximum refractive index for *E*-ray is 1.533, when measured with light of wavelength 5890Å. Find the minimum thickness of quartz placed between a polarizer and an analyzer to produce annulment of the light. [**Ans.** 0.00327 cm]

4
Optical Instruments

4.1 caRdinal points of coAxial lens systems

Most of the optical instruments consist of more than one lens. When a number of lenses having a common principal axis are used, the combination of such lenses is referred to as a *coaxial lens system*.

In case of a single thin lens, the thickness of the lens is neglected in the derivation of the lens formula and other expressions. It is immaterial from which point of the lens the distances are measured. However, in case of a thick lens, or a combination of lenses, the thickness of the lens cannot be neglected. Moreover, the method of evaluating the distance of the image, considering refraction at each surface of a lens, is extremely tedious. Gauss in 1841, to overcome this difficulty, proved that *any number of coaxial lenses can be treated as one unit and the usual formulae for thin lenses can be applied, provided all the distances are measured from two theoretical parallel planes fixed with respect to the refracting system.*

The points of intersection of these planes with the axis are called the *principal* or *Gauss points*. For any system of lenses, there are six points called *cardinal points,* knowing which the image formation in a lens system can be described completely even without knowing the structure of the system. The six cardinal points are: (*i*) *two principal points,* (*ii*) *two focal points* and (*iii*) *two nodal points.*

(*i*) **Principal Points:** Two conjugate planes characterized by unit, positive transverse magnification are known as *principal* (or unit) *planes*. The points, where the principal planes intersect the principal axis are the *principal points*. Thus, *we can define principal points as a pair of conjugate points on the principal axis having unit positive linear transverse magnification.* If an object is placed at one of the principal points, then an image of same size is formed at the other principal point.

As shown in Fig. 4.1, consider two lenses L_1 and L_2 placed coaxially with focal points F_1 and F_2. Let *AB* be a ray of light incident on the system parallel to the principal axis. This ray passes through the focal point F_2 of the second lens. The emergent ray intersects the incident ray at the point *B*. The perpendicular BP_2 intersects the principal axis at P_2 and this point gives the location of *second principal point*. A plane passing through P_2 and perpendicular to the principal axis is the second principal plane. Similarly, if we consider the ray F_1C, then the emergent ray will be parallel to the principal axis. The emergent and the incident rays thus intersect at *C*. The perpendicular CP_1 intersects the principal axis at P_1 and this gives the location of the *first principal point*. A plane passing through P_1 and perpendicular to the principal axis is the first principal plane.

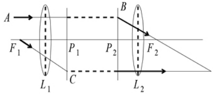
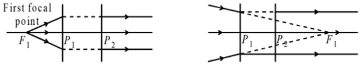


Fig. 4.1

(*ii*) **Focal Points:** The two focal points are a pair of points lying on the principal axis and conjugate to the points at infinity.

First Focal Point: *First focal point F_1 is an object point on the principal axis of coaxial lens system for which the image point lies at infinity.* For a convex (or converging) lens, the rays diverging from the first focal point F_1 become parallel to the principal axis after the refraction [Fig. 4.2 (*a*)]. Similarly, for a concave (or diverging) lens, the rays directed towards first focal point F_1 become parallel to the principal axis after the refraction [Fig. 4.2(*b*)].



(*a*) Convex (or converging) lens (*b*) Concave (or diverging) lens

Fig. 4.2

A plane normal to the principal axis and passing through the first focal point is the *first focal plane*.

Second Focal Point: *Second focal point F_2 is an image point of the principal axis of a coaxial lens system for which the object point lies at infinity.* For a convex lens, the rays incident parallel to the principal axis converge at this point after refraction through lens [Fig. 4.3 (*a*)]. Similarly, for a concave lens, the incident parallel rays after refraction through the lens appear to diverge from F_2 [Fig. 4.3 (*b*)].

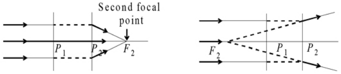


Fig. 4.3

A plane passing through the second focal point F_2 and perpendicular to the principal axis is the *second focal plane*.

(*iii*) Nodal Points: *These are a pair of conjugate points on the principal axis having unit positive angular magnification.* This means a ray of light directed towards one of the nodal points, after refraction through the optical system appears to be emerging from the second in a parallel direction. In Fig. 4.4, N_1 and N_2 are the two nodal points.

The planes passing through the nodal points and perpendicular to the principal axis are called the *nodal planes*.

The distance between the two nodal points is always equal to the distance between the two principal points. Also, the *principal points coincide with the nodal points* when the optical system is situated in the same medium, i.e., the medium on both sides of the optical system is the same. In such a case, the number of cardinal points reduces to four. This is clear from the following discussion.

In Fig. 4.4, H_1P_1 and H_2P_2 represent the first and second principal planes respectively of the lens system, while AF_1 and BF_2 represent the first and second focal planes respectively. A is a point on the first focal plane and AH_1 is a line drawn parallel to the principal axis. The conjugate ray will proceed from H_2, a point on the second principal plane such that $H_1P_1 = H_2P_2$ and passes through the second focal point F_2.

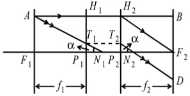


Fig. 4.4

Consider another ray AT_1 parallel to the emergent ray H_2F_2, it emerges out of T_2, a point on the second principal plane such that $T_2P_2 = T_1P_1$ and proceeds parallel to H_2F_2, both the rays originate at A, a point on the first focal plane. The points of intersections of the incident ray AT_1 and the conjugate emergent ray T_2D with the axis give the positions of the two nodal points. Thus, the two points N_1 and N_2 are a pair of conjugate points and the incident ray AN_1 is parallel to the conjugate ray T_2D.

$$\text{Also, } \tan\alpha_1 = \tan\alpha_2$$

$$\text{or, } \frac{\tan\alpha_2}{\tan\alpha_1} = 1$$

In the right-angled δs $T_1P_1N_1$ and $T_2P_2N_2$,

$$T_1P_1 = T_2P_2$$

$$\text{and, } \angle T_1N_1P_1 = \angle T_2N_2P_2 = \alpha$$

Hence, the two δs $T_1P_1N_1$ and $T_2P_2N_2$ are congruent.

$$\therefore P_1N_1 = P_2N_2$$

Adding N_1P_2 to both sides, we get

$$P_1N_1 + N_1P_2 = P_2N_2 + N_1P_2$$

$$\text{or, } P_1P_2 = N_1N_2$$

The distance between the nodal points N_1 and N_2, is equal to the distance between the principal points P_1 and P_2. Now, consider the two right-angled δs AF_1N_1 and $H_2P_2F_2$,

$$AF_1 = H_2P_2$$

$$\text{and, } \angle AN_1F_1 = \angle H_2F_2P_2$$

Hence, the two triangles AF_1N_1 and $H_2P_2F_2$ are congruent.

$$\therefore F_1N_1 = P_2F_2$$

$$\text{or, } F_1P_1 + P_1N_1 = P_2F_2$$

$$\text{or, } P_1N_1 = P_2F_2 - F_1P_1$$

$$\text{We know } P_1N_1 = P_2N_2$$

$$P_1F_1 = -f_1,\ P_2F_2 = +f_2$$

$$\therefore P_1N_1 = P_2N_2 = f_1 + f_2$$

If the medium on both sides of the system is optically the same, then

$$f_2 = -f_1$$

$$\therefore P_1N_1 = P_2N_2 = 0$$

which implies that the principal points coincide with the nodal points. Thus, when the medium on both sides of the system is optically the same, the total number of cardinal points reduces to four.

Deviation Produced by an Optical System

The angle of deviation (or simply deviation) for an optical system is defined as the *angle between the incident ray and the corresponding emergent ray*.

In Fig. 4.5 the cardinal points of a lens system with the same medium on both the sides are shown. Consider a ray AH_1 incident parallel to the principal axis at a height h. The ray H_2D passing through the second focal point is the emergent ray:

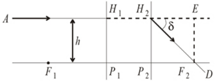


Fig. 4.5

The angle EH_2F_2 (;δ) is the angle of deviation.

$$\text{In D } EH_2F_2,\ \tan\delta = \frac{EF_2}{H_2E} = \frac{H_1P_1}{P_2F_2} = \frac{h}{f_2} = \frac{h}{f}$$

where $f_2 = f$ is the focal length of the system.

Normally, the angle of deviation is very small, therefore

$$\tan\delta = \delta = \frac{h}{f} \quad ...(4.1)$$

Thus, the deviation produced is equal to the ratio of the height at which the ray is incident on the first principal plane to the focal length of the optical system.

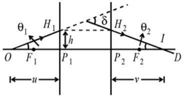


Fig. 4.6

Now consider the situation when the incident ray is not parallel to the principal axis, as shown in Fig. 4.6. From the geometry of the figure, deviation $\delta = \theta_1 + (-\theta_2)$ where, $\theta_1 = \angle P_1OH_1$ taken +ve as measured in anticlockwise direction while $\theta_2 = \angle P_2IH_2$ is taken as negative, as measured in clockwise direction. For the small values of θ_1 and θ_2, we have

$$\tan\theta_1 \cong \theta_1 = \frac{H_1P_1}{P_1O} = \frac{h}{-u}$$

$$\text{and, } \tan(-\theta_2) \cong -\theta_2 = \frac{H_2P_2}{P_2I} = \frac{h}{v}$$

$$\therefore \delta = \theta_1 + (-\theta_2) = \frac{h}{-u} + \frac{h}{v} = h\left(\frac{1}{v} - \frac{1}{u}\right) = \frac{h}{f}$$

which is the same as obtained earlier. Thus, whether the ray is incident parallel to the principal axis or not, the deviation suffered by it is the ratio of the height at which the ray is incident on the first principal plane to the focal length of the optical system.

Equivalent Focal Length and Cardinal Points of a Coaxial Lens System

Consider two coaxially placed convex lenses L_1 and L_2 in air at a distance d apart (Fig. 4.7). A ray of light AB is incident on the lens L_1, parallel to the principal axis at a height h_1. After refraction by the lens L_1 the ray gets deviated by an angle δ_1 and strikes the lens L_2 at a height h_2. Upon refraction from L_2, if it gets further deviated by an angle δ_2 and emerges through F_2, the second focal point, in the absence of the lens L_2, the ray would have passed through D.

The deviation produced by the first lens L_1, $\delta_1 = \dfrac{h_1}{f_1}$, where f_1 is the focal length of lens L_1.

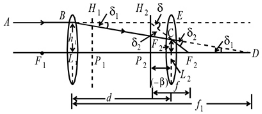


Fig. 4.7

The deviation produced by the second lens L_2, $\delta_2 = \dfrac{h_2}{f_2}$, where f_2 is the focal length of lens L_2.

In order to find the total deviation produced by the combination of lenses, AB is extended in the forward direction and CF_2 in the backward direction and they meet at H_2. The $\angle EH_2C = \delta$ is the total deviation produced by the combination. The deviation δ is also expressed by

$$\delta = \frac{h_1}{f}$$

where f is the equivalent focal length of the combination. The plane passing through H_2 and perpendicular to the principal axis meets the principal axis at P_2, the second principal point. The separation between P_2 and F_2 is the second principal focal length and is denoted by f.

Focal Length of Combination: The total deviation produced by the combination is given by:

$$\frac{h_1}{f} = \frac{h_1}{f_1} + \frac{h_2}{f_2} \quad \text{...(4.2)}$$

The triangles CL_2D and BL_1D are similar.

$$\therefore \frac{BL_1}{CL_2} = \frac{L_1D}{L_2D}$$

$$\text{or, } \frac{h_1}{h_2} = \frac{f_1}{(f_1 - d)} \quad \text{...(4.3)}$$

$$\text{or, } h_1f_1 - n_1d = h_2f_1$$

$$\text{or, } h_2 = h_1 - \frac{h_1d}{f_1} \quad \text{...(4.4)}$$

Substituting the above value of h_2 in Eq. (4.2), we get:

$$\frac{h_1}{f} = \frac{h_1}{f_1} + \frac{h_1}{f_2} - \frac{h_1d}{f_1f_2}$$

or, $$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} - \frac{d}{f_1 f_2} \quad ...(4.5)$$

The above expression gives the focal length f of the combination of two lenses of local length f_1 and f_2 separated by distance d.

When the two lenses are in contact, i.e., $d = 0$, the focal length of the combination is given by:

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} \quad ...(4.6)$$

Eq. (4.5) can also be put in the form:

$$f = \frac{f_1 f_2}{f_1 + f_2 - d} = -\frac{f_1 f_2}{\Delta} \quad ...(4.7)$$

where, $\delta = (d - f_1 - f_2)$.

Positions of Principal Points: Let the separation between the second lens L_2 and the second principal plane be β. The distance β, according to the sign convention, will be negative as it is towards the left of lens L_2.

The triangles $H_2P_2F_2$ and CL_2F_2 are similar.

$$\therefore \frac{H_2P_2}{CL_2} = \frac{P_2F_2}{L_2F_2}$$

$$\text{or, } \frac{h_1}{h_2} = \frac{f}{f - (-\beta)} \quad ...(4.8)$$

Comparing Eqs. (4.9) and (4.8), we get:

$$\frac{f}{f + \beta} = \frac{f_1}{f_1 - d}$$

or, $ff_1 + \beta f_1 = ff_1 - fd$

$$\text{or, } \beta = \frac{-fd}{f_1} \quad ...(4.9)$$

The negative sign indicates that the second principal point lies towards the left of the second lens.

Proceeding in a similar way, we can find the separation α between the first lens and the first principal plane and it is equal to:

$$\alpha = \frac{fd}{f_2} \quad ...(4.10)$$

The positive sign indicates that the first principal point lies on the left side of the lens L_1.

Position of the Focal Points: The distance L_2F_2, i.e., the distance between the second lens and the second focal point locates the position of second focal point.

$$L_2F_2 = P_2F_2 - P_2L_2 = f - (-\beta) = f + \beta$$

$$\therefore L_2F_2 = f + \beta = f - \frac{fd}{f_1} = f\left(1 - \frac{d}{f_1}\right) \quad ...(4.11)$$

Similarly, the location or the position of the first focal point is:

$$-L_1F_1 = P_1F_1 - L_1P_1 = f - \frac{fd}{f_2}$$

$$\text{or, } L_1F_1 = -f\left(1 - \frac{d}{f_2}\right). \quad ...(4.12)$$

4.2 Nodal Slide assembly: location of cardinal points

A nodal slide assembly is an apparatus which is used to locate the cardinal points. It has an optical bench, normally 1.5 m in length and is provided with four uprights. On one upright placed at one extreme end of the bench, is mounted a lamp enclosure with a circular opening for light to pass out and illuminate the cross-slits in a metal plate mounted on the adjacent upright. The next upright carries the *nodal slide and it is called so because by sliding the carriage, the position of the second nodal point of the lens or the lens system can be located.* It is capable of rotation about a vertical axis and its angle of rotation can be recorded from its graduated circular base. The lenses are placed in the holders on the carriage and the separation between the lenses can be read on the scale provided with the carriage. The carriage can be made to slide and as such, the axis of rotation of the nodal slide, which remains fixed, can be made to pass through any point on the axis of the lens system. To find the axis of the lens system, the axis of rotation of the nodal slide is passing through an index mark made on the slide through which the carriage scale moves. The index marks the axis of the rotation of the nodal slide.

The fourth upright carries a plane mirror which is capable of rotation about a horizontal axis perpendicular to the longer side of the bench.

The working of a nodal slide is based on the following facts:

(*i*) A beam of parallel rays, after refraction through the system, converges on the second focal point.

(*ii*) An incident ray, directed towards the first nodal point, after refraction from the system, emerges from the second nodal point in a parallel direction.

(*iii*) When the system is rotated slightly about a transverse axis passing through the second nodal point, the image which is situated on the second focal plane, remains stationary.

(*iv*) When the media on both the sides of the system are the same, the principal points and the nodal points coincide.

Suppose a beam of parallel rays is incident on an optical system in air whose nodal points, coinciding with the principal points, are N_1 and N_2 (Fig. 4.8). After refraction, the emergent beam converges at F_2, the second focal point and a real image is formed on the screen placed in the second focal plane. If the system is rotated about a transverse axis through O, (Fig. 4.9) located between N_2 and F_2, N_1 and N_2 take new position N_1' and N_2' respectively. A ray incident at N_1' emerges from N_2' taking the path $N_2'I_1'$ parallel to the incident ray. Since the incident beam is parallel, the image must lie in the second focal plane. Therefore, I_1' is the new location of the image as the ray N_2I_1' intersects the second focal plane at this point.

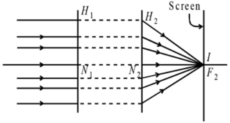


Fig. 4.8

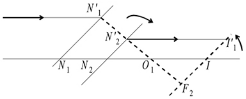


Fig. 4.9

Now, suppose the axis of rotation passes through O_2 (Fig. 4.10), located before N_1, then a small rotation of the system changes N_1 to N_1' and N_2 to N_2' as shown in the figure. A ray of light incident at N_1', emerges from N_2' in the direction I_2' and so I_2' is the new location of image.

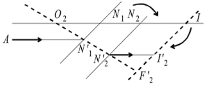


Fig. 4.10

Now, suppose the axis of rotation passes through N_2 (Fig. 4.11). Then on slight rotation of the system, the location of N_1 changes to N_1' while N_2 remains fixed. Since N_2 remains stationary, so the position of the parallel emergent ray N_2I is unchanged and the image remains stationary at I.

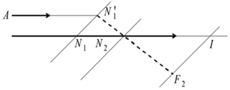


Fig. 4.11

From the above discussion, it is clear that when the axis of rotation is at any point other than N_2, a slight rotation of the system in any direction displaces the image in the same direction. However, *if the axis of rotation passes through the location of N_2, there is no displacement of the image and it remains stationary. In this way, the position of the second nodal point can be located.* In this position, the distance between the screen and the axis of rotation give the focal length of the system.

In the nodal slide experiment, the system is mounted properly and a well defined image of the slit is obtained on the screen. The carriage is now rotated through a small angle and it will be found that the image shifts towards the left or to the right. The carriage and the upright carrying the nodal slide are then adjusted till the location of no shift of the image is found, i.e., the nodal point is located.(Fig. 4.12) The distance between the screen and the axis of rotation for no shift in the image is equal to the first principal focal length of the lens system. By turning the lens system by 180° and repeating the process, the second focal length of the lens system can be determined.

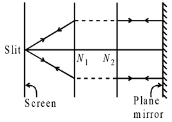


Fig. 4.12

Example 4.1

A coaxial lens system placed in air consists of two lenses of two focal lengths $3F$ and F separated by a distance $2F$. Find the positions of the cardinal points.

Solution:

Here, $f_1 = 3F, f_2 = F, d = 2F$

$$\therefore \frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} - \frac{d}{f_1 f_2}$$

or, $f =$

$$\frac{f_1 f_2}{f_1 + f_2 - d} = \frac{3F \cdot F}{4F - 2F} = \frac{3F^2}{2F} = \frac{3}{2}F$$

The separation α of the first principal point from the first lens:

$$\alpha = \frac{fd}{f_2}$$

$$= \frac{\frac{3}{2}F \cdot 2F}{F} = \frac{3F^2}{F} = +3F$$

The +ve sign indicates that the first principal point P_1 is on the right side of the first lens.

The second principal point β, $= \frac{-fd}{f_1}$

$$= \frac{-3/2F \cdot 2F}{3F} = -F$$

Thus, the second principal point P_2 is on the left side of the second lens.

As the medium on both sides of the lens system is same, the two nodal points N_1 and N_2 coincide with P_1 and P_2.

The first focal point $F_1 = -f\left(1 - \frac{d}{f_2}\right) = -\frac{3}{2}F\left(1 - \frac{2F}{F}\right) = +\frac{3}{2}F,$ from the first lens.

The second focal point $f_1 = f\left(1 - \frac{d}{f_1}\right) = \frac{3}{2}F\left(1 - \frac{2F}{3F}\right) = \frac{F}{2},$ from the second lens.

These cardinal points are shown in Fig. 4.13.

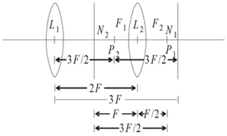


Fig. 4.13

Example 4.2

A thin converging lens and a thin diverging lens each of focal length 10 cm are placed coaxially 5 cm apart. Find the (*i*) focal length, (*ii*) power and (*iii*) principal points of the combination.

Solution:

Here, f_1 = 10 cm, f_2 = –10 cm, d = 5 cm

(*i*) $\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} - \frac{d}{f_1 f_2}$

$$= \frac{1}{10} - \frac{1}{10} + \frac{5}{100} = \frac{1}{20}$$

or, f = + 20 cm.

(*ii*) Power $P = \frac{1}{f}\text{(in metre)} = \frac{1}{0.2\text{ m}} = +5$ Dioptres

(*iii*) First principal point, $\alpha = \frac{fd}{f_2} = \frac{20 \times 5}{-10} = -10\text{ cm}$

Thus, the first principal point P_1 is on the left side of the first lens at a distance of 10 cm.

Second principal point, $\beta = \frac{-fd}{f_1} = \frac{-20 \times 5}{10} = -10\text{ cm}$

Thus, the second principal point P_2 is on the left of the second lens at a distance of 10 cm.

Example 4.3

Two thin converging lenses of focal lengths 3 cm and 4 cm respectively are placed coaxially in air and separated by a distance of 2 cm. An object is placed 4 cm in front of the first lens. Find the position and nature of the image and its lateral magnification.

Solution:

Here, f_1 = 3 cm, f_2 = 4 cm, d = 2 cm

$$\therefore \frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} - \frac{d}{f_1 f_2}$$

$$= \frac{1}{3} + \frac{1}{4} - \frac{2}{3 \times 4} = \frac{4+3-2}{12} = \frac{5}{12}$$

or, $f = \frac{12}{5} = 2.4\text{ cm}$

First principal point, a = $\frac{fd}{f_2} = \frac{2.4 \times 2}{4} = 1.20\text{ cm}$

Second principal point, β = $-\frac{fd}{f_1} = -\frac{2.4 \times 2}{3} = -1.60\text{ cm}$

These values are depicted in Fig. 4.14.

Object distance for the lens system;

$$u = -(4 + 1.20) = -5.20 \text{ cm}$$

$$\therefore \quad \frac{1}{v} - \frac{1}{u} = \frac{1}{f}$$

$$\text{or,} \quad \frac{1}{v} + \frac{1}{5.20} = \frac{1}{2.4}$$

$$\text{or,} \quad \frac{1}{v} = \frac{1}{2.4} - \frac{1}{5.2}$$

$$v = 4.46 \text{ cm}$$

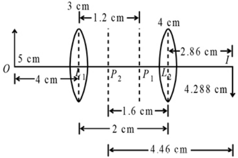


Fig. 4.14.

The distance of the image from lens L_2 = 4.46 − 1.6 = 2.86 cm. Thus, the image is real.

$$\text{Lateral magnification} = \frac{v}{u} = \frac{4.46}{5.20} = 0.8576$$

Hence, the image is diminished.

4.3 Defects in the images: aberrations

While deriving relations between objects and image distances, and focal length of a lens, it is assumed that (*i*) all the incident rays make small angles with the principal axis and (*ii*) the aperture of the lens is small. However, in practice, to have a bright image, we use lenses of large aperture. Also to have a larger field of view, rays having greater angular elevations are incident on the lenses. Moreover, due to finite size of the object, rays from different portion of the object are incident on the lens at different angle and height. It is well known that the deviation produced in any ray depends upon the height of the point of incidence and the angle of incidence. Therefore, paraxial and non-paraxial (or marginal) rays come to focus at different points.

Due to the above mentioned reasons, the image formed is very often, not as predicted by relations derived using simplifying assumptions. The image formed by the lenses can have many defects. Such optical defects in the formed images are called *aberrations*.

The deviation from the actual size, shape and position of an image compared to predicted using the thin lens formula, are called the aberrations produced by a lens. The aberrations can be divided broadly in two categories:

(*i*) Monochromatic (or Siedel) Aberrations: *The aberrations which are present even if the incident light contains a single wavelength* (i.e., *the incident light is monochromatic*) *are called monochromatic aberrations.* The causes of these aberrations have been discussed above. Primarily, monochromatic aberrations arise due to the participation of both paraxial and non-paraxial rays in image formation.

There are five types of monochromatic aberrations. These are: (*i*) spherical aberration, (*ii*) coma, (*iii*) astigmatism, (*iv*) curvature of the field, and (*v*) distortion.

(*ii*) Chromatic Aberrations: It is well known that the refractive index and hence the focal length of a lens is different for different wavelengths of the incident light. For a given lens, the refractive index of violet light is higher than that for red light. Thus, if the incident light on a lens is not monochromatic, a number of coloured images, corresponding to each wavelength, are formed. Even if the images are formed only by paraxial rays, they are formed at different positions and are of different sizes.

The aberrations which arise due to the presence of more than one wavelength in the incident light are called chromatic aberrations.

4.4 chromatic aberration

We have earlier defined the chromatic aberration. The same can also be stated as *the inability of a lens to form a white image of a white object*, i.e., it forms coloured images of an object with white light.

The chromatic aberration is due to the fact that the refractive index of the material of a lens is different for different wavelengths (colours) of light. Hence, the focal length is different for different wavelengths. According to lens maker's formula, the focal length (f) of a lens having radii of curvature R_1 and R_2 is given by:

$$\frac{1}{f} = (\mu - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right)$$

where μ is the refractive index, which is different for different wavelengths (in accordance with Cauchy's law: $\mu = A + \frac{B}{\lambda^2}$). Hence, the focal length f is also different for different wavelengths.

Due to the above mentioned reason, *the image of an object, when seen in white light, is coloured and blurred. This defect is known as chromatic aberration.* The chromatic aberrations are of two types: (*i*) longitudinal (or axial) chromatic aberration and (*ii*) lateral (or transverse) chromatic aberration.

Longitudinal Chromatic Aberration: The refractive index for violet colour μ_v is greater than that for red colour μ_r (*i.e.*, $\mu_v > \mu_r$). Therefore, the focal length for violet colour f_v is smaller than that for red colour f_r. The foci of the other intermediate colours lie in between these two. Thus, when a beam of white light is incident parallel to the principal axis on a convex lens [Fig. 4.15(*a*)], the violet rays converge first (i.e., closest to the lens) and the red rays converge last, i.e., farthest from the lens. If the screen is placed at F_v, the centre of the image will be violet while the outermost region will be red. If the screen is placed at F_r, the centre of the image will be red while the outermost region will be violet.

The separation between focal point of violet colour F_v and that of red colour F_r on the principal axis measures the longitudinal aberration. Thus, if f_v and f_r represent the focal lengths of violet and red colours respectively, then:

$$\text{Longitudinal aberration} = f_r - f_v$$

If an object is situated at the point O on the principal axis [Fig. 4.15(*b*)], its violet image is formed I_v, and the red image is located at I_r. *The separation between I_v and I_r measures the longitudinal chromatic aberration when the object is at finite distance.* That is,

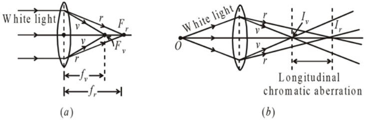


Fig. 4.15: Longitudinal chromatic aberration.

The longitudinal chromatic aberration is taken as positive when measured from I_v to I_r along the direction of the incidence. Thus, *in case of a convex lens, it is positive while in case of a concave lens, it is negative.*

Lateral Chromatic Aberration: Suppose an object is placed on the principal axis in front of a convex lens as shown in Fig. 4.16. When viewed in white light, the lens form seven images corresponding to each colour present in the incident light. The violet image is represented by B_vA_v and the red image is represented by B_rA_r in the figure. The images of other colours lie in between the violet and red images.

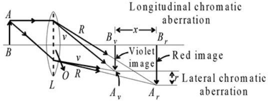


Fig. 4.16

As the magnification of the image is dependent on the focal length of a lens, which is different for different colours, the size of the images of different colours is different. The size of the red image is the largest while that of violet image is the smallest.

This defect, i.e., different size of images of different colours, is known as lateral (or transverse) chromatic aberration. We know,

$$\text{magnification, } m = \frac{\text{Size of the image}}{\text{Size of the object}} = \frac{\text{Distance of the image}}{\text{Distance of the object}}$$

As the distance of the images of different colours are different, the magnification and hence the size of the images of different colours are different leading to lateral chromatic aberration. The different $(B_rA_r - B_vA_v)$ is a measure of lateral chromatic aberration.

Calculation of Longitudinal Chromatic Aberration

We find below the expression for the longitudinal chromatic aberration in the two cases (*i*) when the object is at infinity and (*ii*) when the object is at a finite distance.

(*i*) Longitudinal chromatic aberration when the object is at infinity: The focal length of a lens with radii of curvature R_1 and R_2 is given by:

$$\frac{1}{f} = (\mu - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right)$$

where μ is refractive index of the material of the lens.

If f_v, f_r and f_y be the focal lengths of the lens for violet, red and yellow (mean) colours respectively and μ_v, μ_r and μ_y the corresponding refractive indices, then:

$$\frac{1}{f_v} = (\mu_v - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \quad \text{...(4.13)}$$

$$\frac{1}{f_r} = (\mu_r - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \quad \text{...(4.14)}$$

$$\frac{1}{f_y} = (\mu_y - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \quad \text{...(4.15)}$$

Subtracting Eq. (4.14) from Eq. (4.13), we get:

$$\frac{1}{f_v} - \frac{1}{f_r} = (\mu_v - \mu_r)\left(\frac{1}{R_1} - \frac{1}{R_2}\right)$$

or, $$\frac{f_r - f_v}{f_v f_r} = \frac{(\mu_v - \mu_r)}{(\mu_y - 1)}(\mu_y - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right)$$

or, $$\frac{f_r - f_v}{f_y^2} = \frac{(\mu_v - \mu_r)}{(\mu_y - 1)}\frac{1}{f_y} \quad (\because f_v f_r \cong f_y^2)$$

or,

$$\frac{f_r - f_v}{f_y^2} = \frac{\omega}{f_y}$$

$$\left(\therefore \omega = \frac{\mu_v - \mu_r}{\mu_y - 1} = \text{Dispersive power} \right)$$

or, longitudinal chromatic aberration

$$= f_r - f_v = \omega \cdot f_y \quad ...(4.16)$$

Thus, the longitudinal chromatic aberration of a thin lens, when the object is at infinity, is equal to the product of dipersive power of the lens and the focal length of the mean (yellow) ray.

***(ii)* Longitudinal chromatic aberration when the object is at a finite distance:** For a thin lens, the relation between u, v and f is given by:

$$\frac{1}{v} - \frac{1}{u} = \frac{1}{f}$$

When the object is at a finite distance, i.e., u is a constant, v and f changes with colour so differentiating the above expression, we get:

$$-\frac{dv}{v^2} = -\frac{df}{f^2}$$

or,

$$\frac{dv}{v^2} = \frac{df}{f^2} \quad ...(4.17)$$

If v_v and f_v represent the image distance and focal length for violet colour and v_r and f_r that for red colour, then:

$$dv = v_r - v_v \text{ and } df = f_r - f_v = \omega \cdot f_y$$

Putting these values in Eq. (4.17), we get:

$$\frac{v_r - v_v}{v_y^2} = \frac{\omega f_y}{f_y^2} = \frac{\omega}{f_y}$$

where we have used $v = v_y$ and $f = f_y$ (for mean or yellow colour)

$$\text{or, } v_r - v_v = \frac{\omega}{f_y} \cdot v_y^2 \quad ...(4.18)$$

From the Eqs. (4.17) and (4.18), we observe that the longitudinal chromatic aberration depends upon:

(*i*) The dispersive power of the lens material.
(*ii*) The mean (or yellow) focal length.
(*iii*) The image distance of the mean (or yellow) ray v_y, which in turn depends upon the subject distance u.

4.5 achromatism

It is possible to minimise chromatic aberration by a suitable combination of the lenses. Such a combination of lenses is referred to as *achromatic combination* and this phenomenon shown by such a combination is known as *achromatism*. The process of minimisation of chromatic aberration is known as *achromatisation*.

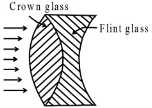


Fig. 4.17: Achromatic doublet.

It is possible to minimise chromatic aberration by using the combination of a suitable convex lens with a concave lens. Such a combination is known as *achromatic doublet*. To form an achromatic doublet, a crown glass convex lens of low focal length (or high power) and a flint glass concave lens of high focal length (or low power) are used. This condition is essential for the doublet to function overall as a convex lens. We discuss the following two cases:

(*i*) Two Lenses in Contact: We have discussed earlier that in a convex lens, violet image is closer to the lens that the red image. While if a convergent beam is incident on a concave lens, then the red image is closer to the lens that the violet image. Therefore, by using a combination of a concave and convex lens, an achromatic doublet can be formed, i.e., both the colours can be made to focus at one point. In other words, the focal length of achromatic doublet is independent of wavelength (or refractive index) of the incident light. An achromatic doublet is shown in Fig. 4.3. The condition we obtain for achromatism for such a doublet is given below. We know that the focal length of a lens is given by:

$$\frac{1}{f} = (\mu - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \quad ...(4.19)$$

where the symbols have their usual meanings. The change in focal length f is dependent on the change in refractive index μ. Differentiating the above equation, we have:

$$d\left(\frac{1}{f}\right) = d\mu\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \quad ...(4.20)$$

Dividing Eq. (4.20) by Eq. (4.19):

$$f \cdot d\left(\frac{1}{f}\right) = \frac{d\mu}{(\mu - 1)}$$

or, $$d\left(\frac{1}{f}\right) = \frac{d\mu}{(\mu-1)}\cdot\frac{1}{f} = \frac{\omega}{f} \quad ...(4.21)$$

where $\omega = \dfrac{d\mu}{(\mu-1)}$ is the dispersive power of the lens.

Let the focal lengths of the two lenses in contact be f_1 and f_2, and ω_1 and ω_2 their respective dispersive powers for the two wavelengths for which the combination is to function as an achromatic doublet. The equivalent focal length of the combination is given by:

$$\frac{1}{F} = \frac{1}{f_1} + \frac{1}{f_2}$$

If the combination is to function as achromatic, the focal length must be the same for all the wavelengths, i.e.,

$$d\left(\frac{1}{F}\right) = 0$$

or, $$d\left(\frac{1}{f_1}\right) + d\left(\frac{1}{f_2}\right) = 0$$

Using Eq. (4.21), we have:

$$d\left(\frac{1}{f_1}\right) = \frac{\omega_1}{f_1} \text{ and } d\left(\frac{1}{f_2}\right) = \frac{\omega_2}{f_2}$$

$$\therefore \frac{\omega_1}{f_1} + \frac{\omega_2}{f_2} = 0 \quad ...(4.22)$$

This is the *reθuired condition for achromatism*. For this condition to be satisfied f_1 and f_2 should be of opposite sign as ω_1 and ω_2 are always positive. In other words, if one lens is convex (or converging), the other lens must be concave (or diverging). Such doublets are widely used in microscopes, telescopes, photographic camera, etc. as objectives. Since the objective of most of the optical instruments should necessarily be convex, therefore, the converging (or convex) lens should be more powerful or should have smaller focal length, compared to the diverging (or concave) lens. Thus, if $f_1 < f_2$, it required from Eq. (4.22) that for achromatic condition to hold, $\omega_1 < \omega_2$. Therefore, the *convex lens of the doublet is a more powerful converging lens, made of crown glass while the concave is less powerful, made of flint glass.*

***(ii)* Two Lenses Separated by a Distance:** Let two convex lenses of focal lengths f_1 and f_2 be separated by a suitable distance x for which the combination behaves as an achromatic combination. We also suppose that the material of both the prisms is the same and hence their dispersive power is also the same.

If F is the equivalent focal length of the two lenses, then:

$$\frac{1}{F} = \frac{1}{f_1} + \frac{1}{f_2} - \frac{x}{f_1 f_2}$$

Now, since changes in f_1 and f_2 with wavelength (or refractive index) are responsible for the change in F, we differentiate the above equation partially and have:

$$d\left(\frac{1}{F}\right) = d\left(\frac{1}{f_1}\right) + d\left(\frac{1}{f_2}\right) - d\left(\frac{x}{f_1 f_2}\right)$$

$$= d\left(\frac{1}{f_1}\right) + d\left(\frac{1}{f_2}\right) - \frac{x}{f_1} d\left(\frac{1}{f_2}\right) - \frac{x}{f_2} d\left(\frac{1}{f_1}\right)$$

We have proved earlier, $d\left(\dfrac{1}{f_1}\right) = \dfrac{\omega_1}{f_1}$ and $d\left(\dfrac{1}{f_2}\right) = \dfrac{\omega_2}{f_2}$, therefore,

$$d\left(\frac{1}{F}\right) = \frac{\omega_1}{f_1} + \frac{\omega_2}{f_2} - \frac{x\omega_2}{f_1 f_2} - \frac{x\omega_1}{f_2 f_1}$$

For the combination to be achromatic, F or $\dfrac{1}{F}$ should not change with colour, i.e.,

$$d\left(\frac{1}{F}\right) = 0$$

or, $$\frac{\omega_1}{f_1} + \frac{\omega_2}{f_2} - \frac{x\omega_2}{f_1 f_2} - \frac{x\omega_1}{f_1 f_2} = 0$$

or, $$\frac{\omega_1}{f_1} + \frac{\omega_2}{f_2} = \frac{x\omega_2}{f_1 f_2} + \frac{x\omega_1}{f_1 f_2}$$

or, $$\frac{\omega_1 f_2 + \omega_2 f_1}{f_1 f_2} = \frac{x(\omega_1 + \omega_2)}{f_1 f_2}$$

$$\text{or, } x = \frac{\omega_1 f_2 + \omega_2 f_1}{\omega_1 + \omega_2} \quad ...(4.23)$$

In the case when the two lenses are made of the same material, so that $\omega_1 = \omega_2 = \omega$ (say), then:

$$x = \frac{\omega f_2 + \omega f_1}{\omega + \omega} = \frac{\omega(f_1 + f_2)}{2\omega} = \frac{f_1 + f_2}{2} \quad ...(4.24)$$

Thus, for the combination to be achromatic, *the separation between the two lenses must be equal to the mean focal lengths of the two lenses.* Huygen's eyepiece satisfies this condition. In this case, the foci of all colours do not lie at the same location, however, the size of all coloured images are the same and a sensation of whiteness is created in the eye. Hence, actually the lateral chromatic aberration is removed and not the longitudinal chromatic aberration.

Example 4.4

Two lenses of focal lengths 8 cm and 6 cm are placed at a certain distance apart. If they form an achromatic combination, find the separation between them assuming that they have the same dispersive power.

Solution:

For an achromatic combination, the separation x is given by:

$$x = \frac{f_1 + f_2}{2}$$

Here, $f_1 = 8$ cm, $f_2 = 6$ cm

$$\therefore \text{Separation, } x = \frac{8+6}{2} = 7 \text{ cm.}$$

Example 4.5

The objective lens of a telescope is an achromat of focal length 90 cm. If the dispersive powers of the two lenses are 0.024 and 0.036, find the focal lengths of the two lenses.

Solution:

For an achromat objective:

$$\frac{\omega_1}{f_1} + \frac{\omega_2}{f_2} = 0$$

$$\text{or, } \frac{1}{f_2} = -\frac{\omega_1}{\omega_2 f_1} = \frac{0.024}{0.036 f_1} = -\frac{2}{3f_1}$$

If F is the local length of the objective, then:

$$\frac{1}{F} = \frac{1}{f_1} + \frac{1}{f_2}$$

$$\text{or, } \frac{1}{90} = \frac{1}{f_1} - \frac{2}{3f_1} = \frac{1}{3f_1}$$

or, $f_1 = 30$ cm

$$\therefore f_2 = -\frac{3f_1}{2} = -\frac{3 \times 30}{2} = -45 \text{ cm}$$

Example 4.6

In an achromatic doublet, the converging lens is made of crown glass having dispersive power 0.012 and the diverging lens is made of flint glass having dispersive power 0.020. The two lenses are in contact and form an achromatic converging combination of focal length 30 cm. Calculate the focal lengths of the two lenses.

Solution:

For achromatism:

$$\frac{\omega_1}{f_1} + \frac{\omega_2}{f_2} = 0$$

Here, $\omega_1 = 0.012$ and $\omega_2 = 0.020$

$$\therefore \frac{0.012}{f_1} + \frac{0.020}{f_2} = 0$$

$$\text{or, } \frac{1}{f_2} = -\frac{0.012}{0.020 f_1} = -\frac{0.6}{f_1}$$

If F is the focal length of the achromatic combination, then:

$$\frac{1}{F} = \frac{1}{f_1} + \frac{1}{f_2}$$

$$\text{or, } \frac{1}{30} = \frac{1}{f_1} - \frac{0.6}{f_1} = \frac{0.4}{f_1}$$

or, $f_1 = 0.4 \times 30 = 12$ cm

$$f_2 = -\frac{f_2}{0.6} = -\frac{12}{0.6} = 20 \text{ cm}$$

Thus, the focal length of the converging (or convex) lens is 12 cm and that of diverging (or concave) lens is 20 cm.

Example 4.7

The refractive indices of crown glass for red and blue lights are 1.517 and 1.523 respectively and the corresponding value for dense flint glass are 1.650 and 1.664. Find the focal length of the component lenses in the design of a plano-convex achromatic doublet of focal length one metre.

Solution:

Here, $\mu_{cr} = 1.517$ and $\mu_{cb} = 1.523$

$$\omega_c = \frac{\mu_{cb} - \mu_{cr}}{\left(\frac{\mu_{cb} + \mu_{cr}}{2}\right) - 1} = \frac{0.006}{0.520} = 1.15 \times 10^{-2}$$

$\mu_{fr} = 1.650$ and $\mu_{fb} = 1.664$

$$\omega_f = \frac{\mu_{fb} - \mu_{fr}}{\left(\frac{\mu_{fb} + \mu_{fr}}{2}\right) - 1} = \frac{0.014}{0.657} = 2.13 \times 10^{-2}$$

For achromatic doublet:

$$\frac{\omega_1}{f_1} + \frac{\omega_2}{f_2} = 0$$

$$\therefore f_1 = -\left(\frac{\omega_1}{\omega_2}\right) f_2 \quad (\omega_1 = \omega_c,\ \omega_2 = \omega_f)$$

$$= -0.54 f_2$$

$$\text{Also, } \frac{1}{F} = \frac{1}{f_1} + \frac{1}{f_2}$$

$$1 = -\frac{1}{0.54 f_2} + \frac{1}{f_2} \quad (\because F = 1\ \text{m})$$

or, $f_2 = -0.85$ m

$\therefore f_1 = -0.54 \times -0.85 = 0.46$ m

Example 4.8

The refractive index of a plano-convex is 1.6 for violet and 1.5 for red light. The radius of curvature of the curved surface is 0.20 cm. Find the separation between violet and red foci of the lens.

Solution:

Here, $\mu_v = 1.6$, $\mu_r = 1.5$

$R_1 = 0.20$ m, $R_2 = \infty$

Using the lens maker's formula

$$\text{for violet colour: } \frac{1}{F_v} = (\mu_v - 1)\left(\frac{1}{R_1} + \frac{1}{R_2}\right)$$

$$= \frac{(0.6 - 1)}{0.6} = \frac{0.6}{0.20}$$

or, $F_v = 0.33$ m

$$\text{for red colour: } \frac{1}{F_r} = (\mu_r - 1)\left(\frac{1}{R_1} + \frac{1}{R_2}\right)$$

$$= \frac{(1.5 - 1)}{0.2} = \frac{0.5}{0.2}$$

or, $F_r = 0.40$ m

Separation between violet and red foci $F_r - F_v = 0.40 - 0.33 = 0.07$ m

4.6 spherical aberration

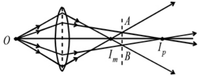


Fig. 4.18: Spherical aberration.

As shown in Fig. 4.18, consider a point source of monochromatic light placed on the principal axis of a large aperture convex lens. The rays from the source incident on the lens close to the axis, called *paraxial rays*, converge at the point I_p. While the rays from the source incident towards the ends of the lens, called *marginal* or *peripheral rays*, come to focus at I_m. Thus, the paraxial rays form the image at I_p while the marginal rays form the image at I_m. In other words, the paraxial rays form the image at a longer distance from the lens that the marginal rays. The image is not sharp at any point and is spread in the region I_m to I_p. *The failure or inability of a lens to converge the paraxial and marginal rays at a single point, after refraction through it, is called spherical aberration.* If the screen is placed perpendicular to the axis at *AB* (the location where paraxial and marginal rays cross), the image appears to be a circular patch of diameter *AB*. If the screen is placed at any position on the two sides of *AB*, the image patch has a larger diameter. For this reason, *the patch of the diameter AB is called the circle of least confusion and this is the position of the best image.*

The cause of spherical aberration can be understood as follows. A lens can be assumed to be made up of small circular zones. The focal lengths of the zones vary slightly with radius of the zones, i.e., different zones have different focal lengths. The focal length of the marginal zone is lesser than the paraxial zone and hence, the marginal rays are focussed earlier than paraxial rays. The spherical aberration can also be explained by saying that the marginal rays suffer greater deviation compared to paraxial rays because marginal rays are incident at a larger height than the paraxial rays. In short, we can say that the spherical aberration arises due to the fact that different annular zones have different focal lengths.

Longitudinal Spherical Aberration: The separation between I_m and I_p on the axis is equal to the *longitudinal spherical aberration*.

Lateral Spherical Aberration: The radius of the circle of least confusion is called the *lateral spherical aberration*.

The spherical aberration produced by a concave lens is shown in Fig. 4.19. The spherical aberration produced by a concave lens is negative while that produced by a convex lens is positive.

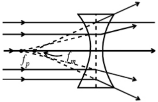


Fig. 4.19

Spherical Aberration Due to Refraction at a Spherical Surface

As shown in Fig. 4.20, a ray of light is incident at the point P of height h on a spherical surface AB with radius of curvature R. The ray gets refracted and intersects the axis at a point F_h, i.e., the focal point of the annular zone with height h is F_h. The distance OF_h is the focal length for the rays refracted in zone h.

From the figure, $f_h = R + CF_h$ ...(4.25)

From the triangle CPF_h, we have

$$\frac{CF_h}{\sin r} = \frac{R}{\sin(i-r)}$$

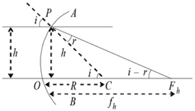


Fig. 4.20: Spherical aberration due to refraction at a spherical surface.

or, $$CF_h = \frac{R\sin r}{\sin(i-r)} = \frac{R\sin r}{\sin i\cos r - \cos i\sin r}$$

$$= \frac{R\sin r}{\mu\sin r\cos r - \cos i\sin r} \quad \left(\because \mu = \frac{\sin i}{\sin r}\right)$$

$$= \frac{R}{\mu\cos r - \cos i} \quad ...(4.26)$$

Substituting for CF_h in Eq. (4.25), we get:

$$f_h = R + \frac{R}{\mu\cos r - \cos i}$$

$$= R\left[1 + \frac{1}{\mu\cos r - \cos i}\right]$$

For spherical rays in the limiting case $h \to 0$ and cos r and cos$i$ tend to unity, so that the focal length of paraxial rays.

$$f_p = R\left[1 + \frac{1}{\mu - 1}\right] = \frac{\mu R}{\mu - 1} \quad ...(4.27)$$

Therefore, the change in focal length for h zone compared to the axial zone is given by:

$$\delta f_h = f_p - f_h = \frac{\mu R}{\mu - 1} - R\left[1 + \frac{1}{\mu\cos r - \cos i}\right]$$

$$= R\left[\frac{1}{\mu - 1} - \frac{1}{(\mu\cos r - \cos i)}\right] \quad ...(4.28)$$

This relation represents the spherical aberration present due to refraction at a spherical surface of radius of curvature R and refractive index μ. The above equation can be simplified by the following approximations.

From the figure, $\sin i = \dfrac{h}{R}$

or, $\sin r = \dfrac{h}{\mu R}$

Also, $\cos i = \sqrt{1 - \sin^2 i}$

$$= \left(1 - \frac{h^2}{R^2}\right)^{\frac{1}{2}} = \left(1 - \frac{h^2}{2R^2}\right)$$

and, $\cos r = \sqrt{1-\sin^2 r}$

$$= \left(1-\frac{h^2}{\mu^2R^2}\right)^{\frac{1}{2}} = \left(1-\frac{h^2}{2\mu^2R^2}\right)$$

Substituting the values of cos i and cos r in Eq. (4.28), we get

$$\Delta f_h = R\left[\frac{1}{\mu-1} - \frac{1}{\mu\left(1-\frac{h^2}{2\mu^2R^2}\right)-\left(1-\frac{h^2}{2R^2}\right)}\right]$$

which on simplifying gives:

$$\Delta f_h = \frac{h^2}{2(\mu-1)^2 f_p} \quad ...(4.29)$$

where, $f_p = \frac{\mu R}{\mu-1}$. Equation (4.29) is the simplified (and approximate) expression for the spherical aberration.

Spherical Aberration for a thin Lens: In Eq. (4.29), the term $(\mu-1)^2$ shows the dependence of spherical aberration on the radii of curvatures R_1 and R_2 by the lens maker's formula for thin lenses.

$$\frac{1}{f} = (\mu-1)\left(\frac{1}{R_1}-\frac{1}{R_2}\right)$$

Thus, the spherical aberration depends upon: (*i*) the refractive index of the lens medium and (*ii*) the shape factor $\beta\left(=\frac{R_1}{R_2}\right)$ of the lens.

We then express Eq. (4.2) as:

$$\Delta f_h = \frac{h^2}{f_p}[\phi(\mu,\beta)] \quad ...(4.30)$$

where f(μ, β) is a function of μ and β. This function is represented by:

$$f(\mu,\beta) = \frac{\beta^2\mu^3+\beta(\mu+2\mu^2-2\mu^3)+(\mu^3-2\mu^2+2)}{2\mu(\mu-1)^2(1-\beta)^2}$$

For spherical aberration to be minimum $\frac{\Delta f_h}{\Delta\beta} = 0,$ which yields:

$$\beta = \frac{R_1}{R_2} = \frac{2\mu^2-\mu-4}{\mu(2\mu+1)} \quad ...(4.31)$$

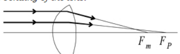
The value of δf_h is mainly dependent upon the shape factor β, hence by having proper ratio of R_1 and R_2, the spherical aberration can be minimised. A thin lens having $\beta = -\frac{1}{6}$ or $R_1 = -\frac{R_2}{6}$ has a minimum spherical aberration for a given aperture and focal length. Such a lens is called a *crossed lens*. For a plano-convex lens, β = 0 for which the spherical aberration is quite close to its minimum value and therefore, a plano-convex lens is usually used to minimise spherical aberration.

Minimisation of Spherical Aberration

It is not possible to eliminate spherical aberration completely. However, by using the following methods it can be minimised.

(*i*) By Using Stops: Spherical aberration can be reduced by reducing the aperture of the lens using stops. The stop used may be such that it either blocks the marginal rays or the paraxial rays. However, the disadvantage of using the stops is that it reduces the amount of light passing through the lens, diminishing the brightness of the image.

(*ii*) By Using Crossed Lens: For the lens with shape factor $\beta = \frac{R_1}{R_2} = -\frac{1}{6}$ (or the crossed lens) the spherical aberration is minimum. A crossed lens is shown in Fig. 4.21. *For spherical aberration, in general, the more curved surface should face the incident or emergent beam of light, whichever is more parallel to the axis.* By using crossed lens, the spherical aberration reduces but is not completely eliminated. The process by which the shape of the lens is changed, so as to minimise the spherical aberration without changing the focal length of the lens is often known as the *bending of the lens*.



Crossed lens

Fig. 4.21

(*iii*) By Using Plano-convex Lenses: Plano-convex lenses are used in optical instruments to reduce spherical aberration. It has been observed that the spherical aberration is proportional to the square of the total deviation produced by the lens. Let δ_1 and δ_2 be the deviations produced at the two surfaces of the lens and δ the total deviation produced by the lens so that $\delta = \delta_1 + \delta_2$. Thus, spherical aberration $\propto \delta^2$

$$\propto (\delta_1 + \delta_2)^2$$

$$\propto (\delta_1 - \delta_2)^2 + 4\delta_1\delta_2$$

Clearly, spherical aberration will be minimum when $\delta_1 = \delta_2$, i.e., the deviation is equally divided at the two surfaces. In *a plano-convex, the deviation is eθually shared between the two surfaces when the convex side faces the incident parallel beam and hence there is lesser spherical aberration*, as illustrated in Fig. 4.22.

The spherical aberration can be minimised by using crossed lens. Using Eq. (4.22) for the crossed lens, it is found that for:

Fig. 4.22: Spherical aberration in a plano-convex lens.

(a) $\mu = 1.5$ $\dfrac{R_1}{R_2} = -\dfrac{1}{6}$

(b) $\mu = 1.66$ $\dfrac{R_1}{R_2} = -\dfrac{1}{6.5}$

(c) $\mu = 1.686$ $\dfrac{R_1}{R_2} = 0$

The crossed lens is nearly plano-convex in case (*a*) and (*b*), and exactly in case (*c*). A plano-convex lens with a refractive index of 1.686 is ideal for minimum spherical aberration.

In general, spherical aberration can be minimised by choosing proper radii of curvature of the lens. The shape factor of a lens is given by

$$q = \frac{R_2 + R_1}{R_2 - R_1} \quad ...(4.32)$$

Figure 4.23 shows the variation of longitudinal (or axial) spherical aberration as a function of shape factor for various lens shapes having the same focal length and refractive index. The spherical aberration for a plano-convex lens (shape factor = 1.0) is minimum when the curved surface faces the incident light and is only slightly more than the double convex lens.

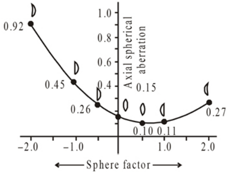


Fig. 4.23: Spherical aberration as a function of the shape factor.

(*iv*) By Using Two Plano-convex Lenses Separated by a Distance eθual to the Difference in their Focal Lengths: The spherical aberration can be minimised by using two plano-convex lenses of the same material (same μ) placed at a distance equal to the difference in their focal lengths. In this arrangement, the deviation produced is equally shared by the two lenses.

In Fig. 4.24, two plano-convex lenses L_1 and L_2 of focal lengths f_1 and f_2 are separated by a distance d. Let a ray of light AB, parallel to the axis, is incident on lens L_1 at the point B at height h_1 and suffers a deviation $\delta_1 = \dfrac{h_1}{f_1}$. This deviated ray is directed towards F_1, the second focal point of L_1. But it suffers another deviation at C, given by $\delta_2 = \dfrac{h_2}{f_2}$. The emergent ray meets the axis at F_2, the second focal point of the combination.

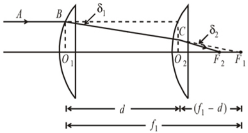


Fig. 4.24

We know, for minimum spherical aberration $\delta_1 = \delta_2$

$$\therefore \frac{h_1}{f_1} = \frac{h_2}{f_2}$$

$$\text{or, } \frac{h_1}{h_2} = \frac{f_1}{f_2} \quad ...(4.33)$$

From triangles BO_1F_1 and CO_2F_1, we have:

$$\frac{h_1}{h_2} = \frac{BO_1}{CO_2} = \frac{O_1F_1}{O_2F_1} = \frac{f_1}{f_1 - d} \quad ...(4.34)$$

From Eqs. (4.33) and (4.34), we get:

$$\frac{f_1}{f_2} = \frac{f_1}{f_1 - d}$$

$$\text{or, } f_2 = f_1 - d$$

$$\text{or, } d = f_1 - f_2 \quad ...(4.35)$$

This is the condition for minimum spherical aberration.

(*v*) By Using a Suitable Convex and Concave Lens Doublet: Spherical aberration for a concave lens is positive, while that for a concave lens is negative. By using a suitable combination of convex and concave lenses, the spherical aberration can be minimised.

Figure 4.25 (*a*) shows the spherical aberration due to a convex lens, the marginal image lies left of the paraxial image, while in case of a concave lens [Fig. 4.25 (*b*)] the marginal image is on the right side of paraxial image. By using a suitable combination [Fig. 4.25(c)] of the two lenses, the spherical aberration can be minimised. The problem with this combination is that it works only for a particular pair of object and image.

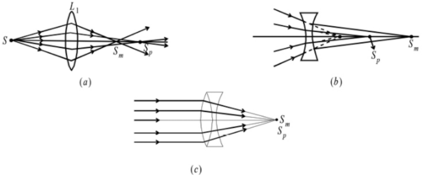


Fig. 4.25: Use of a doublet.

4.7 Other Monochromatic aberrations

In addition to spherical aberration, following are the other monochromatic aberrations:

(*i*) Coma: This is an optical defect due to which a comet-like image is formed instead of a point image, of a point object situated away from the principal axis of a lens (Fig. 4.26). Coma can be reduced or removed by using a *stop* or an *aplanatic lens*. A lens free from spherical aberration and coma is called an aplanatic lens.

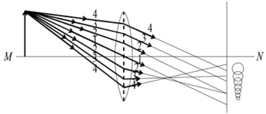


Fig. 4.26: Coma.

(*ii*) Astigmatism: In this defect, the horizontal and vertical lines of an object are focused in different planes. Thus, a sharp image cannot be obtained anywhere and the image gets tangentially or radially blurred, particularly in its outer parts (Fig. 4.27).

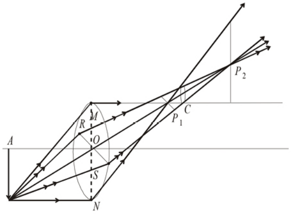


Fig. 4.27: Astigmatism.

Astigmatism can be corrected by using an astigmatic lens. Convex and concave lenses produce astigmatism of opposite nature and therefore it is possible to eliminate this defect by making use of a suitable convex and a concave lens, separated by a proper distance. Such a combination of lenses is called astigmatic lens and is used in photographic lens.

(*iii*) Curvature: Due to this defect, the image of a flat two-dimensional object becomes curved. This is shown in Fig. 4.28 where *AB* is a flat two-dimensional object and *A'B'* its image. The cause of this defect is that the focal length of the lens for para-axial beam becomes greater than marginal rays. This defect can be removed by using two lenses of opposite focal length in *Petzval condition* ($\mu f_1 = -\mu f_2$) and using a single meniscus lens with a proper stop at the right place.

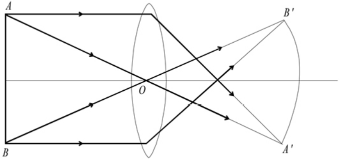


Fig. 4.28: Curvature

(iv) Distortion: The distortion arises in a lens if the magnification is dependent on the distance from the principal axis. *If the linear magnification produced is different from different parts of a straight image, then the images of eθual parts of an object will not be eθually long. This defect is called distortion* (Fig. 4.29).

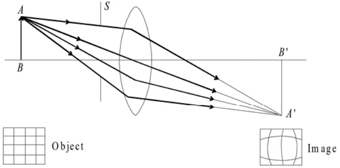


Fig. 4.29: Distortion.

Suppose, if a plane wire mesh is used as an object [Fig. 4.30(*a*)], then a lens with this defect produces its image as shown in Fig. 4.30(*b*) barrel (convex or positive) shaped distortion or as shown in Fig. 4.30(c) cushion (or concave or negative) shaped distortion. Distortion can be eliminated by using a lens combination called *orthoscopic doublet* or *rapid rectilinear lens* (description beyond the scope of this book).

Fig. 4.30

Example 4.9

Two thin lenses of focal lengths f_1 and f_2 separated by a distance d have an equivalent focal length 50 cm. The combination has no chromatic aberration and minimum spherical aberration. Assuming that both the lenses are made up of the same material, find the values of f_1 and f_2.

Solution:

If F is the focal length of the combination, then:

$$\frac{1}{F} = \frac{1}{f_1} + \frac{1}{f_2} - \frac{1}{f_1 f_2}$$

$$= \frac{f_1 + f_2 - d}{f_1 f_2} \quad ...(i)$$

Here, F = 50 cm

Also, for no chromatic aberration:

$$d = \frac{f_1 + f_2}{2} \quad ...(ii)$$

and for minimum spherical aberration

$$d = f_1 - f_2 \quad ...(iii)$$

$$\therefore f_1 - f_2 = \frac{f_1 + f_2}{2}$$

or, $2f_1 - 2f_2 = f_1 + f_2$

or, $2f_1 - f_1 = f_2 + 2f_2$

$f_1 = 3f_2$

Using (*iii*), $d = 3f_2 - f_2 = 2f_2 \Rightarrow f_2 = \dfrac{d}{2}$

$$\therefore f_1 = \frac{3d}{2}$$

Putting these values of f_1 and f_2 in Eq. (*i*), we have:

$$\frac{1}{50} = \frac{\frac{3d}{2} + \frac{d}{2} - d}{\left(\frac{3d}{2}\right) \times \frac{d}{2}}$$

Simplifying, we get

$$d = \frac{200}{3} = 66.67 \text{ cm}$$

$$\therefore f_1 = \frac{3d}{2} = \frac{3 \times 200}{2 \times 3} = 100 \text{ cm}$$

$$f_2 = \frac{d}{2} = \frac{200}{2 \times 3} = \frac{100}{2} = 33.33 \text{ cm.}$$

Example 4.10

Two thin lenses of focal lengths f_1 and f_2 separated by a distance x have an equivalent focal length 30 cm and both the lenses are of the same material. The combination satisfies the conditions of achromatism and spherical aberration. Find the values of f_1 and f_2.

Solution:

For no chromatic aberration:

$$x = \frac{f_1 + f_2}{2}$$

For no spherical aberration:

$$x = f_1 - f_2$$

$$\therefore \frac{f_1 + f_2}{2} = f_1 - f_2$$

$$\text{or, } f_1 = 3f_2$$

If F is the equivalent focal length of the combination, then:

$$\frac{1}{F} = \frac{1}{f_1} + \frac{1}{f_2} - \frac{x}{f_1 f_2}$$

$$\text{or, } \frac{1}{30} = \frac{1}{3f_2} + \frac{1}{f_2} - \frac{2f_2}{3f_2^2}$$

$$= \frac{1}{3f_2} + \frac{1}{f_2} - \frac{2}{3f_2} = \frac{1+3-2}{2f_2} = \frac{2}{3f_2}$$

$$\text{or, } f_2 = \frac{2 \times 30}{3} = 20 \text{ cm}$$

and, $f_1 = 3 \times f_2 = 3 \times 20 = 60$ cm

4.8 eyepiece

In optical instruments such as telescope and microscope, the *objective lens* faces the object and forms an image of the object and the *eye lens* receives the light from the objective lens and forms the magnified image. An image formed by a lens suffers from two main defects: (*i*) chromatic aberration and (*ii*) spherical aberration. To minimise chromatic aberration, an achromatic combination of lenses is used. To reduce spherical aberration, the aperture of the lens is reduced or stops are used which, however, reduces the field of view and the amount of light reaching the eye lens gets reduced, resulting in the diminishing of the brightness of the image.

To increase the brightness of the image and the field of view, an additional lens of comparatively larger aperture known as *field lens* is placed between the objective and the eye lens. *The field lens and the eye lens together constitute an eyepiece or occular*. Instead of a single eye lens, an eyepiece increases the field of view and brightness of the image. The two lenses of an eyepiece are so selected that the combination is achromatic and free from spherical aberration.

In Fig. 4.31, L_1, L_2 and L_3 are the objective lens, field lens and eye lens respectively. The lenses L_2 and L_3 together constitute an eyepiece. The extreme ray that can be collected by the eyelens L_3, in the absence of the field lens L_2, makes an angular object field α and the angular image field β. The best position for the eye or the centre of the exit pupil in this case is *E*. When the field lens L_2 is placed at the location where the image of the object due to the objective lens is formed, even the ray represented by dotted line, which was earlier not collected by the eye lens, is now collected by it, i.e., the field of view and the brightness of the image increase on placing the field lens. Now, the centre of the exit pupil shifts to *E'* (nearer than earlier) and the angular object field is α', where α' > α and the angular image field is β', where β' > β.

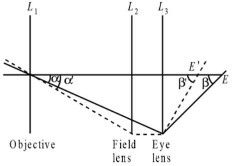


Fig. 4.31

From the above discussion, we can conclude that the field lens: (*i*) increases the angular object field (*ii*) brings the centre of the exit pupil near the eye lens and (*iii*) helps minimise aberrations. Commonly two types of eyepieces: (*i*) Huygen's eyepiece and (*ii*) Ramsden's eyepiece are used in optical instruments, which are discussed below.

4.9 huygen's eyepiece

Construction: *It consists of two plano-convex lenses having focal lengths in the ratio* 3 : 1 (3*f* and *f*) *and the distance between them is eθual to the difference in their focal lengths* (2*f*). The focal length and the positions of the two lenses are such that the eyepiece is achromatic and free from spherical aberration, i.e., chromatic and spherical aberrations are eliminated. Sometimes, two lenses of focal lengths 2*f* and *f* separated by 1.5*f* are also used in this eyepiece.

The arrangement of the lenses is shown in Fig. 4.32. The plano-convex lenses satisfy the conditions of minimum spherical and chromatic aberration. The convex side of the lenses, faces the incident rays. For the spherical aberration to be minimum, the separation between the two lenses must be equal to the difference of their focal lengths. For Huygen's eyepiece, the separation between the lenses $d = 3f - f = 2f$. Hence, this eyepiece satisfies the condition of minimum spherical aberration.

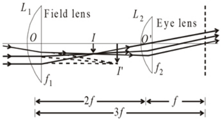


Fig. 4.32: Huygen's eyepiece.

For the chromatic aberration to be minimum, the separation between the two lenses must be equal to the mean of the two focal lengths. For Huygen's eyepiece, the separation between the lenses

$$d = \frac{3f + f}{2} = 2f.$$

Hence, this eyepiece also satisfies the condition for minimum chromatic aberration.

Working: As shown in Fig. 4.32. *I'* prime is the image of the distant object formed by the objective in the absence of the field lens. But in the presence of field lens, the rays get refracted on passing through it and the image *I* is formed. This image lies at the focus of the eye lens, so that the final image is seen at infinity, i.e., the final image is formed at infinity. As the focal length of the eye lens is *f* and the separation between the two lenses is 2*f*, hence image *I* lies in the middle of the two lenses, i.e., at a distance *f* from eye lens towards the left. Thus,

putting, $v = f$ and $f = 3f$

in the lens formula $\frac{1}{v} - \frac{1}{u} = \frac{1}{f}$, we have:

$$\frac{1}{f} - \frac{1}{u} = \frac{1}{3f}$$

$$\text{or, } \frac{1}{u} = \frac{1}{f} - \frac{1}{3f}$$

$$\text{or, } u = \frac{3f}{2}$$

Therefore, the eyepiece is so adjusted that the image formed by the objective of telescope or microscope is at a distance $\frac{3f}{2}$ from the field lens. This serves as a virtual object for the field lens and its image is formed at *I*, which acts as an object for eye lens and it forms the image at infinity.

Huygen's eyepiece is known as the *negative eyepiece* because the real inverted image *I'* formed by the objective of microscope or telescope lies behind the field lens and this image acts as a virtual object for the eye lens. This eyepiece cannot be used to examine directly an object or a real image formed by the objective. When the measurement of the final image is required, the cross-wires should be placed between the field lens and the eye lens. But the cross-wires are viewed through the eye lens only, while the distant object is viewed by rays refracted through the lens. Due to this reason, relative lengths of the cross-wires and the image are disproportionate. *Hence, cross-wires cannot be used in a Huygen's eyepiece and this is a disadvantage.* Therefore, this eyepiece is not used in telescopes and other optical instruments which are used for distance and angle measurements.

Cardinal Point: The focal length *F* of the combination of field lens and eye lens is:

$$\frac{1}{F} = \frac{1}{f} + \frac{1}{3f} - \frac{2f}{3f^2}$$

$$\text{or, } \frac{1}{F} = \frac{2}{3f}$$

$$\text{or, } F = \frac{3f}{2}$$

First Principal Point: P_1 is at distance α from the field lens:

$$\alpha = \frac{Fd}{f_2}$$

$$= \frac{\frac{3}{2}f \times 2f}{f} = 3f$$

Thus, the first principal point P_1 is at a distance 3*f* from the first (or field) lens towards the right.

Second Principal Point: P_2 is at a distance β from the eye lens:

$$\beta = -\frac{Fd}{f_1} = \frac{\frac{3}{2}f \times 2f}{3f} = -f$$

Thus, the second principal point P_2 is at a distance f from the eye lens towards the left.

Also, the nodal points coincide with the principal points as the system is in air.

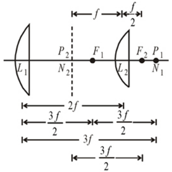


Fig. 4.33

First Focal Point: The location of the first focal point F_1 from the first lens is given by:

$$L_1F_1 = -F\left(1 - \frac{d}{f_2}\right) = -\frac{3f}{2}\left(1 - \frac{2f}{f}\right) = \frac{3f}{2}$$

Thus, the first focal point F_1 lies at a distance $\frac{3f}{2}$ from the first lens towards right.

Second Focal Point: The second focal point F_2 lies at a distance $\frac{3f}{2}$ from the second principal point towards the right. This point is at a distance $\frac{3f}{2} - f = \frac{f}{2}$ to the right of the eye lens. The cardinal points are shown in Fig. 4.33.

4.10 ramsden's eyepiece

Construction: It consists of two plano-convex lenses each of equal focal length f separated by a distance $\frac{2f}{3}$, i.e., equal to two-thirds of the focal length f. The convex faces are towards each other and the field lens is placed beyond the image formed by the objective (Fig. 4.34).

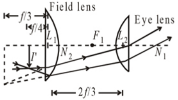


Fig. 4.34: Ramsden's eyepiece.

In this eyepiece, cross-wires are provided and it is used in telescopes and other optical instruments employed to measure the dimensions of the image.

The chromatic aberration in Ramsden's eyepiece is small but not completely absent, as the separation between the lenses is $\frac{2f}{3}$ and not f, the condition for achromatism is not quite satisfied. However, in some cases, both the lenses are made of a combination of crown and flint glass and chromatic aberration is almost eliminated. As both the lenses are plano-convex and their convex surfaces are facing each other, spherical aberration is very small. This is a positive eyepiece.

Working: The image I' formed by the objective serves as the object for the eyepiece. The eyepiece is so adjusted that the image I formed by the field lens lies in the first focal plane of the eye lens so that eye lens forms the final image at infinity. As the focal length of eye lens is f and the distance between the two lenses is $\frac{2f}{3}$, hence the image I lies at a distance $\frac{f}{3}$ from the field lens towards the left.

The image I' formed by the objective acts as the object for the field lens. Let it be at the location u. From the above discussion, it is clear that:

$$v = -\frac{f}{3}, \quad f = f$$

and, u = ?

Putting these values in the lens formula $\frac{1}{v} - \frac{1}{u} = \frac{1}{f}$, we have:

$$-\frac{1}{f/3} - \frac{1}{u} = \frac{1}{f}$$

$$\text{or,} \quad -\frac{1}{u} = \frac{1}{f} + \frac{3}{f} = \frac{4}{f}$$

or, $u = -\frac{f}{4}$

Thus, the image I' is formed at a distance $\frac{f}{4}$ from the field lens towards left. The eyepiece is therefore, so adjusted that the image formed by the objective of telescope or microscope lies at a distance $\frac{f}{4}$ from the field lens towards left. Then this image I' serves as an object for the field lens and its image is formed at I. This image I then acts as an object for the eye lens which forms the final image at infinity.

The cross-wire is placed at the location where the objective forms the image I', i.e., at a distance $\frac{f}{4}$ towards the left of field lens.

Focal length F of the equivalent lens is:

$$\frac{1}{F} = \frac{1}{f_1} + \frac{1}{f_2} - \frac{d}{f_1 f_2}$$

$$= \frac{1}{f} + \frac{1}{f} - \frac{2f}{3ff} = \frac{2}{f} - \frac{2}{3f} = \frac{4}{3f}$$

or, $F = \frac{3f}{4}$.

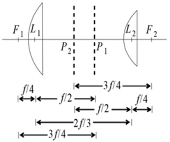


Fig. 4.35

Cardinal Point: The cardinal points are shown in Fig. 4.35 and their values are calculated as given below:

First Principal Point: P_1 is at a distance α from the field lens,

$$\alpha = \frac{Fd}{f_2} = \frac{\left(\frac{3f}{4}\right)\left(\frac{2f}{3}\right)}{f} = \frac{f}{2}$$

Thus, the first principal P_1 is at a distance $\frac{f}{2}$ from the lens towards the right side.

Second Principal Point: P_2 is at a distance β from the eye lens,

$$\beta = -\frac{Fd}{f_1} = -\frac{\left(\frac{3f}{4}\right)\left(\frac{2f}{3}\right)}{f} = -\frac{f}{2}$$

Thus, the second principal point P_2 is at a distance $\frac{f}{2}$ towards left of the eye lens. Since the lenses are in air, the nodal points coincide with the principal points.

First Focal Point F_1 is at a distance $\frac{3}{4}f$ (equal to the focal length of the combination) from the first principal point towards left. The distance of point F_1 from the first lens is given by

$$L_1F_1 = F\left(1 - \frac{d}{f_2}\right) = -\frac{3f}{4}\left(1 - \frac{2f}{3f}\right)$$

$$= -\frac{3f}{4} \times \frac{1}{3} = -\frac{f}{4}$$

Thus, the first focal point is at a distance $\frac{f}{4}$ from the field lens towards left.

Second Focal Point F_2 lies at a distance of $\frac{3}{4}f$ from the second principal point towards right. The distance of point F_2 from the second lens is given by

$$LF_2 = F\left(1-\frac{d}{f_1}\right)=\frac{3f}{4}\left(1-\frac{2f}{3f}\right)=\frac{f}{4}$$

Thus, the second focal point is at a distance $f/4$ from the eye lens towards the right side.

4.11 Comparison between huygen's and Ramsden's eyepieces

	Huygen's Eyepiece		Ramsden's Eyepiece
1.	It is a negative eyepiece. The image formed the object lies in between the field lens and the eye lens and so, no cross-wires can be used.	1.	It is a positive eyepiece. The image formed by the object lies in front of the field lens and so, cross-wires can be used.
2.	It satisfies the condition of minimum chromatic aberration (separation $d = (f_1 + f_2)/2$).	2.	It does not satisfy the condition of minimum chromatic aberration.
3.	The condition of minimum spherical aberration ($d = f_1 - f_2$) is satisfied.	3.	The condition of minimum spherical aberration is not satisfied.
4.	Its power is positive as both the lenses are convex.	4.	Its power is also positive as both lenses are convex.
5.	The eye clearance is small.	5.	The eye clearance is comparatively higher.
6.	It is generally used in the optical instruments used for biological observations, not involving any measurement.	6.	It is used in instruments used for measurements.
7.	Field of view is large.	7.	Field of view is large in this eyepiece also.
8.	The two principal planes are crossed.	8.	The two principal planes are crossed.
9.	It cannot be used as a simple microscope because the first focal plane lies to the right of the field lens and the focal plane is virtual.	9.	It can be used as a simple microscope because the first principal plane lies to the left of the field lens and the focal plane is real.
10.	Final image is convex, towards the eye.	10.	Final image is flat.

4.12 Gauss eyepiece

As shown in Fig. 4.36, Gauss eyepiece is a slight modification over Ramsden's eyepiece. As in Ramsden eyepiece, a Gauss eyepiece consists of two plano-convex lenses of equal focal length, separated by a distance equal to two-thirds of the focal length of either. In this eyepiece a glass plate G is placed at an angle of 45° to the axis of the lens system, to illuminate the field of view. The glass plate reflects the light incident on it from the source s, illuminating the field of view. The cross-wires are placed at C, at a distance $\frac{f}{4}$ in front of the field lens. It is generally used in the telescope of a spectrometer.

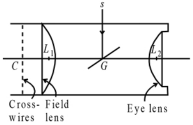


Fig. 4.36

Example 4.11

Light is incident on a Ramsden's eyepiece (focal length of each is lens is 3 cm) from the sun. Locate the position of the image formed and also find the point from which the distance is to be measured.

Solution:

Here, $f = 3$ cm

Distance between the lenses

$$d = \frac{2}{3}f = \frac{2}{3}\times 3 = 2.0 \text{ cm}$$

Focal length of the equivalent lens

$$= \frac{3}{4}f = \frac{3}{4}\times 3 = 2.25 \text{ cm}$$

Since the object is at infinity, the image will be formed at the second focal plane, i.e., at a distance of 2.25 cm from the second principal point.

Distance of the second principal point from the eye lens

$$= -\frac{Fd}{f_1} = -\frac{2.25\times 2}{3} = -1.5 \text{ cm.}$$

Example 4.12

The effective focal length of Ramsden's eyepiece is 3 cm. What is the focal length of a single lens? Also find the separation between the lenses.

Solution:

Here, $F = 3$ cm, $f = ?$ $d = ?$

We know that in a Ramsden's eyepiece, the focal length of both the lenses is the same, say, f and the separation between them is $\frac{2}{3}f.$ We know that:

$$\frac{1}{F} = \frac{1}{f_1}+\frac{1}{f_2}-\frac{d}{f_1 f_2}$$

$$d = \frac{2f}{3}$$

Here $f_1 = f_2 = f$ and

$$\therefore \frac{1}{F} = \frac{1}{f} + \frac{1}{f} - \frac{2f}{3 \cdot f \cdot f} = \frac{2}{f} - \frac{2}{3f} = \frac{4}{3f}$$

$$\text{or, } F = \frac{3f}{4}$$

∴ Focal length of a single lens,

$$f = \frac{4F}{3} = \frac{4 \times 3}{3} = 4 \text{ cm}$$

Separation, $d =$

$$\frac{2}{3} f = \frac{2}{3} \times 4 = \frac{8}{3} = 2.67 \text{ cm.}$$

4.13 Electron microscope

The human eye is a wonderful optical instrument having a wide field of view and a great range in distance. However, it has limitations too. We know that the limiting angle of resolution of a human eye is about 1′, i.e., if the angle subtended by two objects is less than 1′ it ceases to be distinguishable. In fact, whether an object is visible clearly or not, depends both on its size and distance. The human eye cannot clearly see anything which is closer than 25 cm and has diameter less than 0.01 cm. Optical microscope is employed to see the details of the objects having dimensions less than what can be seen clearly by a unaided eye. It is well known that lower the wavelength of light used, higher is the resolving power of the microscope.

In 1924, de Broglie argued (later experimentally verified) that when electrons are scattered or diffracted they behave as waves of wavelength $\lambda = \frac{h}{mv}$ (h = Planck constant, m = mass of the electron, v = velocity of the electron). An electron accelerated through a potential difference of 60,000 volts $\left(v = \sqrt{\frac{2eV}{m}}\right)$ the wavelength works out to be 5×10^{-10} cm (or 0.05 Å). This indicated that if a microscope could be constructed with electron beam, it would be possible to have much larger resolving power than that with an optical microscope. In 1927, a scientist named Busch showed that a beam of electrons can be focussed by suitable electric and magnetic fields as light is focussed using glass lens in optical microscopes. These developments paved the way for the construction of electron microscope and the first electron microscope was constructed in 1932. With continuous perfection of technique, an electron microscope can give a magnification upto 1,00,000 compared to the maximum magnification of an optical microscope upto 6,000.

Thus, the *working principle* of an electron microscope can be stated as: (i) a beam of electrons exhibits wave nature similar to light rays, but of much shorter wavelength, and (ii) a beam of electrons can be focussed by suitable electric and magnetic fields, very much like light rays are focussed by glass lenses.

Electrostatic focussing: As shown in Fig. 4.37, let P and q be the interface of two regions of uniform electrostatic potential V_1 and V_2 (assume $V_2 > V_1$) and v_1 and v_2 be the velocities of the electrons in the two respective regions near the interface. If m and e be the mass and charge of the electron respectively, then for the first region

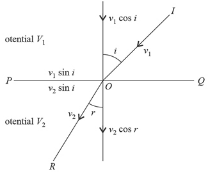


Fig. 4.37

$$\frac{1}{2} m v_1^2 = V_1 e$$

$$\text{or, } v_1 = \sqrt{\frac{2V_1 e}{m}} \quad ...(4.36)$$

Similarly, in the second region, velocity of the electron is given by:

$$v_2 = \sqrt{\frac{2V_2 e}{m}} \quad ...(4.37)$$

Dividing Eq. (4.36) by Eq. (4.37), we have:

$$\frac{v_1}{v_2} = \sqrt{\frac{V_1}{V_2}} \quad ...(4.38)$$

Resolving the velocities v_1 and v_2 into perpendicular and parallel components, if i is the angle of incidence and r the angle of refraction, the perpendicular components are $v_1 \cos i$ and $v_2 \cos r$ and the parallel components are $v_1 \sin i$ and $v_2 \sin r$. Since $V_2 > V_1$, the component $v_2 \cos r$ is greater than $v_1 \cos i$ and hence the electron is accelerated in the direction of the field. However, if $V_2 < V_1$, then the electron will be retarted. As Pq is an equipotential surface there is no change in the component parallel to Pq and hence:

$v_1 \sin i = v_2 \sin r =$ constant

$$\frac{\sin i}{\sin r} = \frac{v_2}{v_1} = \text{constant}$$

or, ...(4.39)

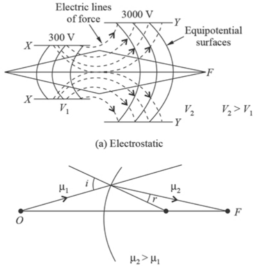
This expression is similar to Snell's law of refraction. From the above expressions, we have:

$$\frac{\sin i}{\sin r} = \frac{v_2}{v_1} = \sqrt{\frac{V_2}{V_1}} \quad \text{...(4.40)}$$

For $V_2 > V_1$, the electron beam is deflected towards the normal and for $V_2 < V_1$, the beam is deflected away from the normal, similar to the refraction of light.

The experimental arrangement used for electrostatic focussing of electrons is called *electrostatic electron lens* or simply *electron lens*. There are various types of electron lenses such as asymmetrical electron lens, symmetrical electron lens and double aperture lens. The simplest of them, the asymmetrical electron lens, is described below.

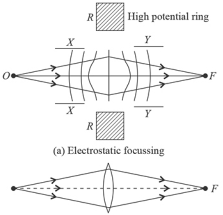
The simplest electrostatic electron lens consists of two cylindrical tubes X and Y, separated by a gap and charged to different positive potentials with the tube Y at a higher potential than X (Fig. 4.38). The resulting electric field is shown by dotted lines while the equipotential surfaces are represented by continuous curves. The curved equipotential surfaces form the electrostatic lens. The effect of this lens is that the beam passing through it gets focussed at F. By making Y larger than X, the electric lines of force can be made to spread out more and cause a corresponding change in the curvature of the equipotential surfaces and hence, the point of focus. In this set up, the electron beam travels from a region of low potential and velocity, to a region of high potential and velocity, similar to a light beam refracting at a convex spherical surface (denser medium) from air (rarer medium). To obtain the effect of a lens, a high potential ring is placed between the two cylinders. This is illustrated in Fig. 4.39, along with the optical analogue. It is important to note that the focal length of an electron lens and the velocity of the electrons can be changed by changing the potentials of the cylinders unlike an optical lens which has a fixed focal length.



(a) Electrostatic

(b) Optical analogue

Fig. 4.38



(a) Electrostatic focussing

Lens

(b) Optical analogue

Fig. 4.39

Most of the times, however, it is *magnetic focussing* that is used in electron microscopes. The principle of magnetic focussing is described below.

Magnetic focussing: Suppose an electron moving with velocity v enters a region of uniform magnetic field H making an angle q with the direction of the field (Fig. 4.40). Resolving the velocity v, represented by *OP*, into the two components, v cos q is the component in the direction of the field while v sin q is the component perpendicular to the direction of the field. If only the parallel component v cos q is present, the electron will move in the direction of the magnetic field. While in the presence of only the perpendicular component v sin q, the electron will move in a circular path, when both the components are present, the *resultant path of the electron is a helix*.

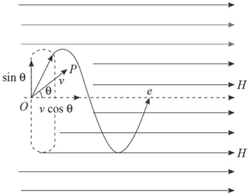


Fig. 4.40: Magnetic focussing.

The radius r of the circle due to the component v sin q is given by:

$$\frac{m(v\sin\theta)^2}{r} = Hev\sin\theta \quad ...(4.40a)$$

where, m is the mass of the electron and e the charge on it.

$$v\sin\theta = \frac{Her}{m}$$

\

Time taken by the electron to complete the circle:

$$t = \frac{2\pi r}{v\sin\theta} = \frac{2\pi rm}{Her} = \frac{2\pi m}{He} = \frac{2\pi}{H\left(\frac{e}{m}\right)} \quad ...(4.41)$$

The pitch of the helix l is the distance travelled by the electron parallel to the magnetic field in time t.

$l = v$ cos q × t

$$= v\cos\theta \times \frac{2\pi m}{He} = \frac{2\pi mv\cos\theta}{He} \quad ...(4.42)$$

It is clear from Eq. (4.41) that the time t is independent of angle q. Therefore, an electron starting from *O* in any direction will arrive at *R* at the same time. In the absence of the magnetic field, the electron would have moved along *OP*. Thus, it follows that if electrons emitted from a source at *O* and forming a divergent beam, is subjected to a uniform magnetic field, converge at a point *R* distant l, equal to the pitch of the helix. The value of l can be altered by changing the value of the applied magnetic field *H*. The action of the magnetic field is thus similar to that of a converging lens on a divergent beam of light rays. Hence, the magnetic field is referred to as *magnetic lens* of focal length l. If l is so adjusted that the electrons make two or more complete rotations in the helix, the focal point will occur at 2l, 3l... . Thus, unlike an optical lens which has a fixed focal length, a magnetic lens has a variable focal length determined by the value of the magnetic field.

There are many types of magnetic lenses. In Fig. 4.41 the most common type of the magnetic lens is shown. An electromagnet with sections of coils *A* and *B* is shown in the figure. The coil is surrounded by a soft iron shield with gaps *p* and q, in the form of rings, at diametrically opposite positions. A strong magnetic field exists across the gap and the field is symmetrical about the axis *OF*. If the electrons are produced by a source at the position *O*, the divergent beams of electrons converges at the point *F*. The focal length of a magnetic lens is small. Magnetic lens is usually employed in electron microscopes particularly when an intense and very fast beam of electrons is required.

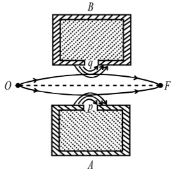


Fig. 4.41

Construction: Schematic diagram of an electron miscroscope is shown in Fig. 4.42(*a*). For comparison with an optical miscroscope, in Fig. 4.42(*b*), schematic diagram of an optical microscope is also shown. Following are the essential parts of an electron microscope:

> **(i) Electron Gun:** The electrons are produced by the electron gun. Electrons are emitted from the hot filament. The filament is surrounded by a metallic cylinder kept at a negative potential, which stops the electrons from spreading due to repulsion and a fine beam of electrons is produced. The beam is then accelerated through a high potential (upto 30 KV). Now, with these accelerated electrons, the waves of very short wavelength are associated.

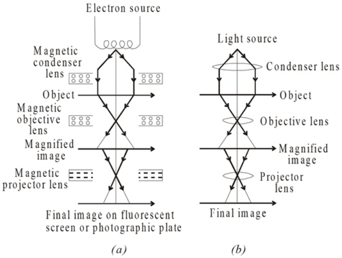


Fig. 4.42

These are three main types of electron microscope. These are: (*i*) scanning (SEM), (*ii*) transmission (TEM), and (*iii*) field emission electron microscopes. In the first two types, electrons are generated in an electron gun to act upon the atomic nuclei of the specimen. While in field emission type, the specimen itself is a source of radiation.

***(ii)* Magnetic (or Electrostatic) Lens:** The basic function of the condenser lens is to focus the electron beam from the electron gun on to the specimen to permit illustration and formation of the image. The different parts of the specimen absorb electrons differently and a corresponding image is formed. The objective lens forms an initial enlarged image of the illuminated specimen in a plane that is suitable for further enlargement by the projector lens. The projector lens, as the name indicates, is used to project the final magnified image on to the screen or photographic plate.

***(iii)* Fluorescent Screen or Photographic Plate:** The electron image is converted into visible light image on the fluorescent screen or photographic plate. To avoid the collision of electrons with air molecules, the entire system is enclosed inside a metallic frame and a high degree of a vacuum is maintained with a high speed diffusion pump.

Uses of Electron Microscope: Electron microscope is very useful tool for all types of researchers such as biologists, for studying viruses, bacteria and other microorganisms, materials scientists for studying crystal structures, textile technologists, for the investigation of fibres, paints, etc.

Exercises

Short Answer Type

1. Name the six cardinal points and represent them diagrammatically.
2. Define Principal Points and Principal Planes.
3. Define Nodal Points and Nodal Planes.
4. Define Focal Points and Focal Planes.
5. Briefly explain the properties of principal points and nodal points.
6. Define angle of deviation for an optical system and write an expression for it.
7. Write the expressions for the principal points and focal points of a lens system consisting of two thin lenses of focal lengths f_1 and f_2 placed coaxially at a distance *d* apart.
8. For what purpose is a nodal slide assembly used? How? Explain briefly.
9. What do you mean by Aberration? Name and mention the cause of the two types of aberrations.
10. What do you mean by Monochromatic and Chromatic Aberrations? Differentiate between the two types of aberrations.
11. Define: (*i*) Longitudinal Chromatic Aberration and (*ii*) Lateral Chromatic Aberration. Explain their causes.
12. Write suitable expressions for longitudinal chromatic aberration and hence mention the factors on which it depends.
13. What do you mean by Achromatism and Achromatisation?
14. Give the condition for achromatism when the two lenses are in contact.
15. Why in an achromatic doublet, the converging lens is more powerful and made of crown glass, while the diverging lens is less powerful and made of flint glass?
16. When the two lenses of same dispersive power are separated by a distance, what is the condition for their achromatism? Which one is actually removed, lateral or longitudinal chromatic aberration?
17. What is Spherical Aberration? How can it be minimised?
18. What is a crossed lens? How is it useful?
19. Mention various ways of minimisation of spherical aberration.
20. What is Coma? How can it be minimised?
21. What do mean by Astigmatism? How can it be removed?
22. What is meant by the defect 'curvature'? How can it be removed?
23. What do you mean by the term 'distortion'? How can it be eliminated?
24. What is Ramsden's eyepiece? Show the image formation in this eyepiece.
25. Are cross-wires used in Ramsden's eyepiece? Explain.
26. How does Gauss eyepiece differ from Ramsden's eyepiece? Explain.
27. Mention any three important differences between Huygen's and Ramsden's eyepieces.
28. Explain the working principle of an Electron Microscope.
29. Discuss the principle of Electrostatic Electron Lens.
30. Explain the concept of Magnetic Focussing. Why is the path of the electron

helical?

31. Mention the important uses of the electron microscope.

Long Answer Type

1. Explain the terms 'cardinal points' for a coaxial lens system.

2. Explain the formation of image using the cardinal points of a lens system.

3. Show that the principal points coincide with nodal points if the medium on both sides of the lens system is optically the same.

4. Derive an expression for the angle of deviation produced by an optical system.

5. Two thin lenses of focal lengths f_1 and f_2 are placed coaxially at a distance d apart. Find an expression for the equivalent focal length of the combination. Also, obtain the expressions for principal points and focal points for this lens system.

6. Describe the construction and working of a nodal slide and explain how the nodal points of a lens system can be located with its help.

7. Define 'aberrations'. Name various types of aberrations and briefly discuss their causes.

8. What is meant by chromatic aberration? Giving a suitable diagram, differentiate between longitudinal and lateral chromatic aberration.

9. What do you mean by longitudinal chromatic aberration? Obtain an expression for the same when the object is (*i*) at infinity and (*ii*) at finite distance. On what factor does it depend?

10. What is Achromatisation? Deduce the condition for achromatism of two lenses (*i*) in contact and (*ii*) separated by a distance.

11. What do you mean by spherical aberration? Differentiate between longitudinal and lateral spherical aberration. Briefly mention the various ways of minimisation of spherical aberration.

12. Obtain an expression for the spherical aberration due to refraction at a spherical surface.

13. Name and explain various types of monochromatic aberrations. How can they be minimised?

14. What is an eyepiece? Giving a suitable diagram, explain the advantages of using an eyepiece over a single eye lens.

15. Give the construction and theory of Huygen's Eyepiece. Why cannot a cross-wire be used with it?

16. Find the cardinal points of Huygen's Eyepiece and represent them diagrammatically.

17. Give the construction and theory of Ramsden's Eyepiece.

18. Find the cardinal points of Ramsden's Eyepiece and represent them diagrammatically.

19. Show that the Huygen's Eyepiece satisfies the condition of minimum spherical aberration and chromatism.

20. Give important differences and similarities between Huygen's and Ramsden's Eyepieces.

21. Explain the principle, construction and working of an Electron Microscope.

Numericals

1. Two thin convex lenses of focal lengths 6 cm and 2 cm are placed coaxially 4 cm apart. What is the power of the combination?

[**Ans.** 33.3 Dioptre]

2. Two thin convex lenses having focal lengths 5 cm and 2 cm are placed 3 cm apart coaxially. Find the equivalent focal length and the positions of principal points.

[**Ans.** 2.5 cm, 3.75 cm, 1.5 cm]

3. Two thin converging lenses of focal lengths 20 cm and 40 cm are placed 20 cm apart coaxially. An object is placed at a distance of 50 cm from the first lens. Find the positions of the principal points, focal points and the position of the image.

[**Ans.** a = 10 cm, b = −20 cm, P1F1 = −10 cm, P2F2 = 0 cm, image at 10 cm right of the second lens]

4. Calculate the separation between two thin lenses of focal lengths 12 cm and 4 cm respectively, if the combination is to function as a lens of effective length 6 cm.

[**Ans.** 8 cm]

5. A convex lens of 25 cm focal length is placed at 15 cm from a concave lens of 15 cm focal length. Find the cardinal points and represent them diagrammatically.

[**Ans.** Principal points: −75 cm, −45 cm, focal points: −150 cm, 30 cm]

6. Two thin converging lenses of focal lengths 15 cm and 20 cm are placed 10 cm apart coaxially. An object is placed at a distance of 24 cm from the first lens. Find (i) the positions of focal and principal points and (ii) the position of the image.

[**Ans.** f = 12 cm, a = 6 cm, b = −8 cm, v = 12 cm]

7. Two thin convex lenses of focal lengths 12 cm and 4 cm are placed 8 cm apart coaxially. Find the cardinal points of the system and represent them diagrammatically.

[**Ans.** Principal points: 12 cm, − 4 cm; focal points: 6 cm, 2 cm]

8. Two lenses of focal lengths 10 cm and 8 cm are placed a certain distance apart. Calculate the distance between the lenses if they form an achromatic combination.

[**Ans.** 9 cm]

9. The dispersive powers for crown and flint glass are 0.015 and 0.030 respectively. Calculate the focal lengths of the lenses if the combination is an achromatic doublet of focal length 60 cm, when placed in contact.

[**Ans.** 30 cm, −60 cm]

10. An achromatic doublet of focal length 20 cm is to be formed out of a combination of crown and flint glasses. The radius of curvature of the faces is contact is 15 cm. Calculate the radius of curvature of the other faces, given the

dispersive powers of the crown and flint glasses are 0.02 and 0.04 respectively and their respective refractive indices are 1.52 and 1.65.

[**Ans.** 7959 cm, 97.5 cm]

11. Calculate the equivalent focal length of an achromatic combination of two thin lenses of focal lengths 1 m and 0.4 m respectively separated by a distance. Assume that both the lenses are of the same material.

12. Two glasses having dispersive powers in the ratio of 2 : 3 are to be used in making an achromatic doublet of focal length 0.2 m. Calculate the focal lengths of the two lenses.

[**Ans.** 0.067 m, −0.1 m]

13. An eyepiece consists of two convex lenses of focal lengths 3 cm and 8 cm placed 6 cm apart. Calculate the focal length of the eyepiece. Also, calculate (*i*) the position of the object for which the image is formed at infinity and (*ii*) the position of the image, if object is placed at infinity.

[**Ans.** $\frac{24}{5}$ cm, (*i*) $\frac{6}{5}$ cm left of first lens, (*ii*) $\frac{6}{5}$ cm right of first lens].

Unit III

5

Laser

5.1 INTRODUCTION

The term LASER stands for *Light Amplification by Stimulated Emission of Radiation*. The discovery of laser is one of the most important discoveries of the last century. The first successful operation of a laser was demonstrated by T. Maiman in 1960 using a ruby crystal in USA. In the following year, the first gas laser, was fabricated by Ali Javan and coworkers. Since then, different types of lasers using solids, liquids and gases have been developed. The immense use of laser, from toys to warfare and from welding to surgery have made it very popular.

Characteristics of Laser Light

Like ordinary light, laser light is electromagnetic in nature. However, there are few characteristics of laser light not possessed by the normal light. Some of these characteristics of laser light are mentioned below:

***(i)* Directionality:** The laser beam is highly directional having almost no divergence (except the diffraction effect). The output beam of a laser has a well-defined wavefront and, therefore, it is highly directional. Due to its high directionality, a laser beam can be focussed on a point by passing it through a suitable convex lens. If a laser beam of wavelength λ = 6000 Å and beam radius 2 mm is passed through a convex lens of focal length 5 cm, then the area of the spot at the focal plane is of the order of:

$$\frac{\pi\lambda^2 f^2}{a^2} = \frac{\pi \times (6\times10^{-7}\text{ m})^2 \times (5\times10^{-2})^2}{(2\times10^{-3}\text{ m})^2}$$

$$\left(\frac{\lambda f}{a} = \text{radius of central maximum}\right)$$

$$= 7.1 \times 10^{-10}\text{ m}^2$$

which is extremely small.

For a typical laser beam, the beam divergence is less than 0.01 milliradian, i.e., for a metre of propagation, the spread is less than 0.01 mm. The normal light from a strong source spreads to about one kilometre for every kilometre of its propagation. Just imagine how much a normal beam of light would diverge when it reaches moon at a distance of 3,80,000 km while a laser beam spreads to just a few kilometres on reaching the moon.

***(ii)* Monochromaticity:** The laser light is nearly monochromatic. In reality, no light is perfectly monochromatic, i.e., it is not characterized by a single wavelength λ or frequency *υ* but instead, it is characterized by spread in frequency Δ*υ* about the central frequency (or Dl in case of wavelength λ). *The monochromaticity of a light* *is defined by* $\frac{\Delta v}{v}$. For perfect monochromaticity Δv = 0, which is not attainable in practice, but the value of Δv is much smaller for lasers compared to ordinary light. For an ordinary light Δv is of the order of 10^{10} Hz, whereas for a laser, it of the order of 500 Hz. Thus, for light of wavelength λ = 6000 Å or frequency $v = 5 \times 10^{14}$ Hz,

monochromaticity of ordinary light, $$\frac{\Delta v}{v} = \frac{10^{10}\ \text{Hz}}{5 \times 10^{14}\ \text{Hz}} = 2 \times 10^{-5}$$

monochromaticity of laser light, $$\frac{\Delta v}{v} = \frac{500\ \text{Hz}}{5 \times 10^{14}\ \text{Hz}} = 10^{-12}$$

Thus, laser light is highly monochromatic compared to ordinary light.

Monochromaticity is also taken as a measure of spectral purity. Smaller the value of monochromaticity, higher the purity of spectrum.

***(iii)* Coherence:** Laser radiation is characterized by *high degree of coherence, both spatial and temporal.* In other words, *a constant phase relationship exists in the radiation field of laser light source at different locations and times.* It is possible to observe interference effects from two independent laser beams. In fact, coherence is the main feature which distinguishes laser radiation from ordinary light and other characteristics (directionality, monochromaticity and intensity) are related to the high degree of coherence.

In general, the coherence or phase between two light waves can vary from point to point (in space) or change from instant to instant (or time). Thus, these are two types of coherence, which we are discussing below:

Temporal Coherence: If waves given by a light source maintain coherency at a point at two different points in time, i.e., coherency is maintained with respect to time by a source, this type of coherency is referred to as *temporal coherence.*

Spatial Coherence: If the two waves maintain a constant phase relationship over any time at different points in space, the waves are said to be *spatially coherent.* This is possible even when two beams are individually time incoherent, as long as any phase change in one beam is accompanied by a simultaneous equal phase change in the other beam (as in Young's double slit experiment).

Temporal (or time) coherency is characteristic of a single beam of light, whereas spatial (or space) coherency concerns the relationship between two separate beams of light.

As discussed earlier, light waves given out by de-excitation of atoms are produced in the form of wave trains. These wave trains are of finite length, each wave train contains only a limited number of waves. The length of the wave train Δs is called the *coherence length* and is equal to the product of the number of waves N contained in the wave train and the wavelength λ, i.e., $\Delta s = N\lambda$. Since, velocity is defined as the distance covered per unit time it takes a wave train of length Δs, a certain length of time Δt, to pass a given point,

$$\Delta t = \frac{\Delta s}{c}$$

where c is the velocity of light. The length of time Δt is called the *coherent time.*

Ordinary light consists of large number of separate waves without any phase relationship, cancelling and reinforcing each other randomly. The wavefront so produced changes from one point to the other and from instant to instant. Thus, there is neither spatial nor temporal coherence in ordinary light.

***(iv)* Intensity:** The laser beam is *highly intense* compared to ordinary light. Since, the laser power is concentrated in a beam of very small diameter (≈ few mm), even a small laser can deliver very high intensity at the focal plane of the lens. For example, if the power (P) of a laser beam is 1 watt, the intensity at focal point is given by:

$$I = \frac{P}{\text{Area}} = \frac{Pa^2}{\pi\lambda^2 f^2} = \frac{1\ \text{W}}{7.0 \times 10^{-10}\ \text{m}^2}$$

$$= 1.4 \times 10^{9}\ \text{Wm}^{-2}$$

Note that even a small power of 1 watt can give an intensity of 10^{9} W/m^2, which is extremely large.

The intensity of the laser beam is so high that to achieve equal intensity from a hot object it would have to be heated to 10^{30}K.

Example 5.1

The output of a laser has a pulse duration of 30 ms and average output power of 1 W per pulse. How much energy is released per pulse and how many photons does each pulse contain if the wavelength of the laser beam is 6600 Å?

Solution:

$$\text{Energy} = \text{Power} \times \text{Time}$$
$$= 1\ \text{W} \times 30 \times 10^{-3}\ \text{s}$$
$$= 0.003\ \text{Joule}$$

Number of photons in each pulse

$$= \frac{\text{Energy}}{hv} = \frac{\text{Energy}}{hc} \times \lambda$$

$$= \frac{0.003\ \text{J} \times 6.6 \times 10^{-7}\ \text{m}}{6.6 \times 10^{-34}\ \text{Js} \times 3 \times 10^{8}\ \text{m/s}} = 10^{16}.$$

Example 5.2

Laser light from a 2 mW source of aperture diameter 1.5 cm and wavelength 5000 Å is focussed by a length of focal length 20 cm. Calculate the area and intensity of the image.

Solution:

Area of the image

$$= \frac{\pi\lambda^2 f^2}{a^2}$$

Here, λ = 5000 Å = 5.0×10^{-7} m, f = 20 cm = 0.2 m

$$a = \frac{1.5}{2} = 0.75\ \text{cm} = 0.75 \times 10^{-2}\ \text{m}$$

Area of the image =

$$\frac{3.14 \times 5.0 \times 10^{-7} \times 5.0 \times 10^{-7} \times 0.2 \times 0.2}{0.75 \times 10^{-2} \times 0.75 \times 10^{-2}}$$

$= 5.6 \times 10^{-10}$ m²

$$\text{Intensity} = \frac{\text{Power}}{\text{Area}} = \frac{2 \times 10^{-3}\ \text{W}}{5.6 \times 10^{-10}\ \text{m}^2}$$

$= 0.357 \times 10^{7}$ Wm^{-2}

$= 3.57 \times 10^{6}$ Wm^{-2}.

Example 5.3

For an ordinary light source, the coherence time t_c = 10^{-10} sec. Obtain the degree of monochromaticity for λ_0 = 6000 Å.

Solution

Coherence time t_c = 10^{-10} sec.

$$\therefore \Delta v = \frac{1}{\tau_c} = \frac{1}{10^{-10}} = 10^{10}\ \text{Hz}$$

$$\text{For, } \lambda_0 = 6000\ \text{Å},\ v_0 = \frac{c}{\lambda_0} = \frac{3 \times 10^8}{6 \times 10^{-7}}$$

$= 0.5 \times 10^{15} = 5.0 \times 10^{14}$ Hz

$\therefore$ Monochromaticity

$$\frac{\Delta v}{v} = \frac{10^{10}}{5.0 \times 10^{14}} = 0.2 \times 10^{-4}.$$

5.2 Absorption and emission (Spontaneous and stimulated) of radiation

To understand the principle of working of a laser, it is necessary to understand the radiative absorption and emission of photons in an atomic system. According to quantum theory, atoms exist only in certain discrete energy states and absorption and emission of photons causes them to make a transition from one discrete energy state to another. Transition from one energy level to another can occur by *stimulated absorption* (or simply absorption), *spontaneous emission* and *stimulated* (or induced) *emission*. We shall see that the stimulated emission is particularly important for a laser.

Absorption: Figure 5.1 illustrates the absorption process in a two-energy level atom. The atoms are initially in the lower energy state E_1. When radiation is incident on the atom, the atom absorbs a photon of frequency v such that $hv = E_2 - E_1$ and makes a transition to higher energy level E_2. This process is known as *stimulated* (or *induced*) *obsorption* or simply *absorption*. The process is represented as:

$$\text{atom} + \text{photon} \rightarrow \text{atom}^*$$

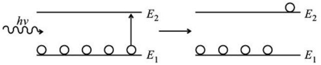


Fig. 5.1: Absorption of photon.

The asterisk indicates an excited state. In this process, only the photons of frequencies $v = \frac{E_2 - E_1}{h}$ are absorbed out of the incident photons of different frequencies.

In an atomic system, the absorption process strongly depends on the density of the radiation u of the incident radiation field and the number of atoms N_1 in the ground state E_1. Absorption decreases the number of ground state atoms (due to transition to higher energy state) with a rate of

$$R_{ab} = -\frac{dN_1}{dt} = B_{12} u N_1 \quad ...(5.1)$$

where, B_{12} is a constant of proportionality, also known as *Einstein's absorption coefficient*. R_{ab} also represents the upward transition rate of atoms.

Emission: The excited state is an unstable state. An atom which has been excited by absorption, remains in the excited state only for a certain time, called the *life time* t_{21} ($\approx 10^{-8}$ s) and afterwards, makes a transition to the lower energy state E_1, resulting in the emission of photon. This emission process can occur in the following two ways:

(i) Spontaneous Emission: When an atom undergoes transition to a lower energy state emitting a photon, *without any external stimulation*, the process is known as *spontaneous emission* (Fig. 5.2). The process is represented as:

$$\text{atom}^* \rightarrow \text{atom} + \text{photon}$$

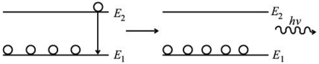


Fig. 5.2: Spontaneous emission.

Photons emitted by different atoms or molecules by spontaneous emission are *incoherent* and *non-monochromatic*. Light from an ordinary light source including light emitting diode (LED) is given out of this process and hence, it is incoherent. The spontaneous emission is not influenced by external radiation field. It depends on the population N_2 of excited atoms in state E_2. The rate of spontaneous emission is represented by:

$$R_{sp} = -\left(\frac{dN_2}{dt}\right)_{sp} = A_{21}N_2 \quad ...(5.2)$$

where A_{21} is a constant of proportionality, also known as *Einstein A coefficient*, which is the probability per unit time for spontaneous emission. The spontaneous emission rate gives the rate at which atoms are de-excited from the state E_2 to E_1.

(ii) Stimulated (or induced) Emission: This process occurs when a photon, from the incident radiation, having an energy equal to the difference between the two states $(E_2 - E_1)$, interacts with the atom in the excited state, inducing it to return to the lower energy state (Fig. 5.3). It can be represented as:

atom* + photon → atom + photon + photon

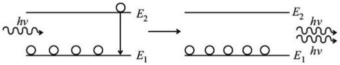


Fig. 5.3: Stimulated emission.

In this process, a new photon of the same frequency, phase, polarization and direction of propagation, as the stimulating photon, is generated. Thus, an incoming photon triggers the generation of an additional photon. Contrary to spontaneous emission, *coherent radiation is obtained in this mode of emission.* This implies that when an atom is stimulated to emit light energy by an incident wave, the liberated light energy adds upto the wave in a constructive manner providing *amplification of light. This process of light amplification is the basic principle of the device called LASER which stands for Light Amplification by Stimulated Emission of Radiation.* Unlike spontaneous emission, stimulated emission is directly proportional to the energy density u, of the external radiation field. Also, stimulated emission rate R_{st} increases with the increase in number N_2 of excited atoms. The rate of stimulated emission is given by:

$$R_{st} = -\left(\frac{dN_2}{dt}\right)_{st} = B_{21}uN_2 \quad ...(5.3)$$

where, B_{21} is a constant of proportionality also known as *Einstein's stimulated coefficient.*

5.3 Relation Between Einstein's Coefficients

In any transition process two energy states are involved. Let us consider a system having two energy states, a lower state of energy E_1 with population N_1 and a higher state of energy E_2 with population N_2.

The population N, i.e., number of atoms per unit volume, in the energy state E, according to Boltzmann distribution law, is given by:

$$N = N_0 e^{-E/kT} \quad ...(5.4)$$

where N_0 is the population in the ground state (E = 0), k is the Boltzmann's constant and T the absolute temperature. From the above equation it is clear that the population is maximum in the ground state and decreases exponentially as one goes to higher energy states. For our considered system, using Boltzmann distribution law, we have

$$\frac{N_2}{N_1} = \frac{e^{-E_2/kT}}{e^{-E_1/kT}} = e^{-(E_2-E_1)/kT} \quad ...(5.5)$$

$$\text{or } N_2 = N_1 e^{-(E_2-E_1)/kT} \quad ...(5.6)$$

Using the relation $E_2 - E_1 = hv$, we get:

$$N_2 = N_1 e^{-hv/kT} \quad ...(5.7)$$

In a closed system under thermal equilibrium, the net rate of emissions must equal the net rate of absorptions. That is,

$$R_{st} + R_{sp} = R_{ab}$$

Using Eqs. (5.1), (5.2) and (5.3), we get:

$$B_{21}uN_2 + A_{21}N_2 = B_{12}uN_1$$

$$\text{or } u = \frac{N_2 A_{21}}{N_1 B_{12} - N_2 B_{21}}$$

$$\text{or } u = \frac{N_2 A_{21}}{N_2 B_{21}\left(\dfrac{N_1 B_{12}}{N_2 B_{21}} - 1\right)}$$

Subtituting the value of $\dfrac{N_1}{N_2}$ from Eq. (5.7), we get:

$$u = \frac{A_{21}}{B_{21}\left(\dfrac{B_{12}}{B_{21}} e^{hv/kT} - 1\right)} \quad ...(5.8)$$

The energy density u, from Planck's radiation law, is given by:

$$u = \frac{8\pi h v^3}{c^3} \frac{1}{e^{hv/kT} - 1} \quad ...(5.9)$$

where c is the speed of light in free space.

Comparing Eqs. (5.8) and (5.9), we get:

$$B_{12} = B_{21} = B \quad ...(5.10)$$

and, $$\frac{A_{21}}{B_{21}} = \frac{8\pi h\nu^3}{c^3} \quad ...(5.11)$$

The above two expressions are called *Einstein's relations* and the coefficients A and B are called *Einstein's A and B coefficients.*

Absorption and stimulated emissions are mutually reverse processes and their probabilities are equal i.e., $B_{12} = B_{21}$.

5.4 Population Inversion

Consider an optical medium having two energy states E_1 and E_2 ($E_1 < E_2$) with population N_1 and N_2 respectively. Under the normal conditions, Boltzmann distribution law requires $N_1 > N_2$, i.e., the population of atoms in the lower energy state E_1 must be greater than that in the higher energy state E_2.

When radiation is passed through the medium, at any given time, numerous absorption and emission processes occur. *Each absorption process attenuates the incident light wave, whereas each stimulated emission amplifies it.* Unlike light waves generated by stimulated emission, light waves generated by spontaneous emission exhibit a random phase. The small number of spontaneously emitted photons, which accidently have the same phase as that of the photons generated by stimulated emission, aid the lasing action. *Laser operation, i.e., amplification of incident light wave, requires that stimulated emission is dominant compared to absorption.* We know

$$R_{st} = B_{21}uN_2$$

$$\text{and, } R_{ab} = B_{12}uN_1.$$

Consider the ratio

$$\frac{R_{st}}{R_{ab}} = \frac{B_{21}uN_2}{B_{12}uN_1} = \frac{N_2}{N_1} \quad ...(5.12)$$

It is evident from the above equation that the ratio depends only on the population of the atoms in the excited state and the ground state. On the basis of the above ratio, consider the following cases:

(i) $N_1 > N_2$: That is, population of the atoms in the ground state is higher than that in the excited state. In this case, *absorption is dominant and hence, the intensity of light wave decreases exponentially.*

(ii) $N_1 = N_2$: That is, population of atoms in the ground state is equal to that in the excited state. In this case, *absorption and stimulated emission are in equilibrium and hence, the intensity of light wave remains constant.*

(iii) $N_1 < N_2$: That is, population of atoms in the excited state is more than that in the ground state. *This condition, known as population inversion, is the most important case and must be achieved for the lasing action.* In this case, *stimulated emission is dominant and hence, the intensity of the incident light increases, i.e., the incident light wave is amplified.*

From Eq. (5.7), we have:

$$\frac{N_2}{N_1} = e^{-h\nu/kT}$$

which can also be expressed as:

$$\frac{N_1}{N_2} = e^{h\nu/kT} \quad ...(5.13)$$

In normal situation, $N_1 > N_2$ for $E_1 < E_2$. However, to achieve *optical amplification, i.e., lasing action,* it is essential to create a non-equilibrium distribution of atoms in such a way that the population of the upper energy level is greater than that of the lower energy level ($N_2 > N_1$ for $E_2 > E_1$).

Achieving Population Inversion: Pumping

The process of raising the atoms or molecules of the active medium to a higher energy state, so as to achieve population inversion, is called *pumping*. Commonly used methods of pumping are: (*i*) optical pumping, (*ii*) electric excitation, (*iii*) inelastic atom-atom collision, (*iv*) chemical reactions and (*v*) direct conversion.

(i) Optical Pumping: In this method, a highly intense radiation from an optical flash tube is incident on the active medium, raising the atoms to a higher energy state through stimulated absorption. Normally, the radiation is given out in the form of short flashes. Maiman used this method of pumping in ruby laser and it is still widely used in solid state lasers. For better efficiency, the active material is placed inside a helical flash lamp.

(ii) Electrical Excitation: In this method, an electric discharge is used to excite the atoms of the active material. The method is normally used in gas lasers, e.g., He – Ne, argon, etc. An extremely high electric field (≈ several kV/m) accelerates the electrons emitted by the cathode towards the anode. Collisions between these high energy electrons and the atoms of the active medium raise them to the higher energy level, producing population inversion.

(iii) Inelastic Atom-Atom Collision: In some cases, the active medium has two types of atoms (He – Ne). An electric discharge initially excites one type of atoms. These excited atoms then collide inelastically with the second type of atoms transferring energy from the excited atoms to the unexcited atoms. Thus, the second type of atoms also get excited, creating population inversion in them.

(iv) Chemical Pumping: In chemical laser, the energy required for excitation is obtained from a suitable chemical reaction. For example, in a chemical laser, hydrogen and fluorine combine to form hydrogen fluoride and this chemical laser produces an infrared laser beam of over 2 MW.

(v) Direct Conversion: In semiconductor lasers and light emitting diodes (LED) direct conversion of electrical energy into radiation takes place.

The two level atomic systems are, however, not suitable for achieving population inversion. Because if $B_{12} = B_{21}$, the probabilities of absorption and stimulated emission are equal, keeping the population of the two states unchanged.

For achieving population inversion, three or four energy level atomic systems are *used*. The energy level diagrams of such systems are shown in Fig. 5.4. These atomic systems are characterized by the presence of a central *metastable* state in which atoms spend an unusually long time. It is *the transition from this metastable state that the stimulated emission or lasing action takes place.*

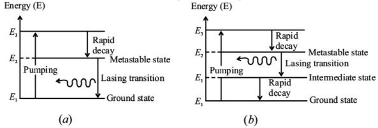


Fig. 5.4: Energy level diagrams showing population inversion and lasing for (a) three-level system (b) four-level system

Three-level System: As shown in Fig. 5.4 (*a*), a three-level atomic system consists of a ground level E_1, a metastable level E_2 and a third level E_3 above the metastable level. In equilibrium, the atomic distribution follows the Boltzmann's distribution law. However, with suitable pumping some of the atoms are excited from the ground state into the higher energy level E_3. The atoms rapidly decay by non-radiative processes to either E_2 or directly to E_1. Hence, empty states are always available in the level E_3. The metastable state E_2 is characterized by a longer lifetime ($\approx 10^{-3}$ s) of the atoms than that in E_3. Due to this, larger number of atoms accumulate in the metastable level E_2. *Over a period, the number of atoms in the metastable level E_2 becomes more than that in the ground state E_1, thereby, achieving population inversion between the level E_1 and E_2. Stimulated emission and lasing action can then occur by radiative transition from E_2 to E_1.*

Ruby laser is a three-level laser. The drawback with three-level laser is that it generally requires very high pumping power. It is so because ground state being the terminal state, more than half the ground state atoms must be pumped into the metastable state to achieve population inversion.

Four-level System: A four-level system, depicted in Fig. 5.4 (b) (e.g., He – Ne laser) is characterized by much lower pumping requirement. In such systems, pumping excites the atoms from the ground state E_1 to the level E_4. The excited atoms then decay to the metastable level E_3. Since the population of E_4 and E_2 remains practically unchanged, even a small increase in the number of atoms in the level E_3 creates population inversion. Thus, stimulated emission or lasing action takes place between level E_3 and level E_2.

Example 5.4

Find the relative population of atoms in the two states in a ruby laser that produces a light beam of wavelength 6943 Å at 300 K.

Solution

The relative population of atoms in two states with energy E_1 and E_2 is given by

$$\frac{N_2}{N_1} = e^{-(E_2-E_1)/kT}$$

$E_2 - E_1 = hn =$

$$\frac{hc}{\lambda} = \frac{6.6\times10^{-34}\times3\times10^{8}}{1.6\times10^{-19}\times\lambda} = \frac{12400\times10^{-10}}{\lambda}\text{eV}$$

$$= \frac{12400\times10^{-10}}{6493\times10^{-10}} = 1.79$$

$$\frac{N_2}{N_1} = \exp\left[-\frac{1.79}{8.61\times10^{-5}\times300}\right] = \exp(-69.3)$$

$= 8 \times 10^{-31}$

5.5 Components of a Laser System

A laser system requires three essential components for its operation. These are:

(*i*) *An active medium* in which the lasing action takes place.

(*ii*) *An energy source or the pumping source* which excites the atoms to higher energy state, achieving the population inversion.

(*iii*) A *resonant cavity* or *optical cavity* to provide the feedback for laser oscillations.

The active medium may be solid, liquid or gas. In most lasers, the active medium is enclosed in a cavity formed by two mirrors (plane or curved) facing each other. One of the mirrors of the cavity is 100% reflecting, while the other mirror is partially transparent to allow some of the radiations to pass through. The optical cavity is analogous to an oscillator as it provides positive feedback of the photons by reflection, at the mirrors, at either end of the cavity. Therefore, the area enclosed inside the optical cavity is also called an *optical resonator*. After multiple passes through the cavity due to reflection at the mirrors, light amplification becomes very large. One of the mirrors is partially transparent and the laser output is obtained through this mirror. A schematic diagram representing essential components and their arrangement is shown in Fig. 5.5.

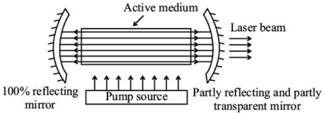


Fig. 5.5: Basic components of a laser.

Cavity Configurations

A single passage of light through the active medium in a laser is not enough for the amplification of light. To amplify light to a suitable level and get intense laser output, the light is made to pass through the active medium repeatedly. To do this, the active medium is placed inside an *optical cavity* called *resonator*. Such a system is also known as *laser oscillator*. A resonator consists of a pair of mirrors that reflect the light back and forth through the active medium. One of the mirrors is totally (100%) reflecting while the other is partially transparent through which the laser output is allowed to pass through.

In the beginning, the optical cavity consisted of two identical plane mirrors enclosing the active medium. However, these days, a *confocal cavity*, which has two identical concave mirrors separated by a distance equal to their radius R, is used. A *spherical resonator*, which has two concave mirrors of equal radius R separated by a distance $2R$ is also used. Another type of cavity, called *hemispherical cavity*, which consists of a concave mirror at one side and a plane mirror at the centre of curvature of the concave mirror, is sometimes also employed. It is easier to align this resonator, although, the output power is low. Nowadays, the most commonly used cavity is the *long radius cavity* which has two concave mirrors of equal and long radius of curvature, separated by a distance less than the radius of curvature.

Modes in the Cavity: Let us consider the simplest form of the cavity, the two parallel plane mirrors. A standing wave pattern is formed inside the cavity with nodes at the ends. If L is the length of the cavity, then the wavelength of the standing wave is given by:

$$\lambda = \frac{2L}{n}, \quad n = 1, 2, 3, \ldots \quad \ldots(5.14)$$

where n is an integer, called the *axial mode number* which decides the number of half wavelengths or *axial modes* that fit into the cavity.

For the wavelengths which satisfy Eq. (5.14), the cavity is said to be *resonant*. Since the cavity encloses the active medium whose refractive index is, say, μ, the length L has to replaced by the optical path $L\mu$. If the optical path $L\mu$ is denoted by S, then

$$\lambda = \frac{2L\mu}{n} = \frac{2S}{n} \quad \ldots(5.15)$$

$$\therefore \text{Frequency, n} = \frac{c}{\lambda} = \frac{cn}{2S} \quad \ldots(5.16)$$

The frequency difference between the consecutive modes is:

$$\text{Dn} = \nu - \nu_{n-1} = \frac{cn}{2S} - \frac{c(n-1)}{2S} = \frac{c}{2S} \quad \ldots(5.17)$$

Due to the presence of different modes, the output need not be monochromatic. In addition to these axial modes, we also have the transverse electromagnetic (TEM) modes. Usually, the TEM modes are few in number. Since the TEM_{00} has the same phase across the aperture, *it can be focussed on the least spot size giving the highest intensity*. Hence, TEM_{00} mode is widely used.

Gain: If the population inversion exists in the active medium of the laser, the amplification of radiation takes place, because in such a situation, the gain due to stimulated radiation is more than the loss due to absorption. The intensity of the beam I_n at frequency n is given by:

$$I_\nu = I_{0\nu} e^{\alpha_\nu x} \quad \ldots(5.18)$$

where I_{0n} is the intensity of the beam at $x = 0$ and α_n is the gain coefficient at frequency n.

As the radiation is reflected back and forth between the mirrors, a certain fraction of it is lost due to scattering and reflection losses. For the laser to oscillate in the cavity, the gain must exceed or at least equal the losses (Δ).

This is,

$$\text{Gain} = \frac{I_\nu - I_{0\nu}}{I_{0\nu}} \geq \delta$$

$$\text{or, } I_n - I_{0n} \geq I_{0n}\Delta$$

Substituting the value of I_n from Eq. (5.18)

$$e^{\alpha_\nu x} - 1 \geq \delta \quad \ldots(5.19)$$

Since the length of the cavity is L, the length covered in one cycle, from $x = 0$ to $x = L$ and then to $x = 0$, is $2L$. Hence, the oscillation condition given by Eq. (5.19) reduces to:

$$e^{\alpha_\nu 2L} - 1 \geq \delta \quad \ldots(5.20)$$

If $2\alpha_n L << 1$, then the above condition reduces to:

$$2\alpha_n L \geq d \quad \ldots(5.21)$$

If at a given frequency, the gain exceeds the loss, the lasing action will take place, i.e., intensity of the radiation builds up. However, over the time the population difference reduces and hence the gain drops until

$$2\alpha_n L = \Delta \quad \ldots(5.22)$$

The above condition is the *threshold condition* for the sustenance of the lasing action.

Active Medium

The active medium in a laser may either be a solid, liquid or gas. Very often the nature of active material is used to classify the lasers as solid state laser, gas laser, semiconductor laser, etc. In the first laser made, a ruby rod was used as the active medium. Ruby crystal is Al_2O_3 doped with Cr^{3+} ions. The Cr^{3+} ions serve as the active material and it has a concentration of nearly 0.05 percent by weight. Natural ruby is not a suitable active medium due to the presence of crystal defects and strains. Instead, synthetic rods of ruby grown from molten Al_2O_3 and Cr_2O_3 are used as active medium. Ruby laser gives out pulsed output (duration ~ 10 ps) of wavelength 6943 Å. The *YAG* laser ($Y_3Al_5O_{15}$ with Nd^{3+} active ion) is another common solid state laser which gives out both continuous and pulsed (duration ~10 to 150 ps) output of wavelength 10641 Å. Glass laser with Nd^{3+} active ion also gives out pulsed (duration ~ 1 ps) output of wavelength 1059 Å. Thus, the wavelengths of the lasers extend from infrared to ultraviolet region.

The lasers with gas as active media are also very popular. The active element in gas laser can be noble gas (He – Ne laser: continuous output of wavelengths 6328 Å, 3391 Å, 1152 Å), a positive ion (Ar^+ laser, K^+ laser), a neutral molecule (N_2 laser, CO_2 laser), a metal atom (He–Cd laser), and so on.

Semiconductor lasers such as GaAs, InP, InAs, InSb, PbSe, PbS, etc., are also used extensively these days. The advantages of semiconductor lasers are their high efficiency and the ease with which they can be employed in optical communication.

Dye lasers (organic dyes dissolved in suitable solvents) constitute another type of active medium. The laser output is mostly in the visible region.

5.6 Ruby Laser

Ruby laser is a three-level solid state laser and it was the first laser fabricated by Maiman in 1960. Ruby crystal which is a crystal of Al_2O_3 with some Al^{3+} ions replaced by Cr^{3+} ions, has been known to the mankind for hundreds of years, as a naturally occurring precious stone. For the laser, the ruby crystal is grown from a molten mixture of Al_2O_3 and Cr_2O_3. Typical Cr^{3+} concentration in the crystal is nearly 0.05 weight per cent. In the ruby crystal, Cr^{3+} ions are the active components. Ruby crystal is taken in the form of a cylindrical rod of 5 to 10 cm length and 5 mm in diameter. One end of the rod is completely polished while the other end is made partially reflecting. Hence, the ends work as the cavity mirrors.

The energy levels of Cr^{3+} ions responsible for the lasing action is shown in Fig. 5.6. It has three energy levels E_1, E_2 and E_3. The uppermost level E_3 consists of two bands ${}^4F_1$ and ${}^4F_2$. The states in these bands have extremely small lifetime of the order of 10^{-9} s. The second energy level E_2 is a metastable state and has a lifetime of the order of 10^{-3} s, several orders of magnitude longer than that of E_3. The level E_2, in fact, has two sublevels of difference of nearly 14 Å. The transition from E_3 to E_2 is non-radiative while transition from E_2 to E_1 is radiative and is responsible for the lasing action.

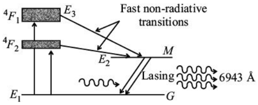


Fig. 5.6: Energy levels of chromium ion.

The ruby rod is placed inside a xenon flash lamp as shown in Fig. 5.7.

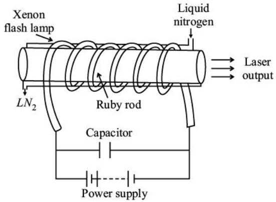


Fig. 5.7: Ruby laser.

The light from the flash lamp pumps the Cr^{3+} ions to the energy bands $4F_1$ and $4F_2$. The ions in these bands make a rapid non-radioactive decay to the metastable state E_2. Due to the longer lifetime in this level, the number of Cr^{3+} ions keeps on increassing. Subsequently, population inversion takes place between the levels E_1 and E_2. When the required population inversion is reached, lasing action is initially triggered by spontaneously emitted photons which are available in the system and spontaneous incoherent red fluorescence, typical of ruby, with a peak near 6940 Å, is produced. But as the pumping energy is increased and the *population inversion exceeds the threshold, stimulated emission starts and coherent laser light with a sharp peak at* 6943 Å is produced. Because of the sublevels in E_2, some other wavelengths are present, particularly for spontaneous transitions but for stimulated transitions, the wavelength 6943 Å dominates over other lines. The stimulated emission depopulates the level E_2 at a faster rate than the pump rate, stopping the lasing action momentarily. However, before the output falls to zero, the population again builds up in the level E_2, exceeds the threshold value and the lasing action restarts. This process occurs many times before the pumping flash ends. Hence, the output consists of a series of spikes as shown in Fig. 5.8 process is referred to as *spiking of the laser*. Since the flash lamp operation is pulsed, the laser output is also pulsed, i.e., the laser beam is emitted in the form of pulse, lasting a few milliseconds.

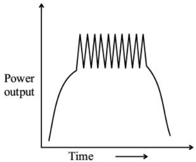


Fig. 5.8: Spikes in ruby laser output.

A large amount of energy is dissipated in the ruby rod and it has to be cooled for efficient continuous operation. Liquid nitrogen is circulated for this purpose.

5.7 Helium–Neon Laser

He–Ne laser was the first successfully operated gas laser, brought in operation in 1961 by Javan, Bennet and Harroit. Unlike ruby laser, *this laser gives a continuous laser beam*. By selecting proper resonator mirrors, laser beam of wavelengths 6328 Å, 33913 Å and 11523 Å can be produced. It is a low power laser with typical output power of a few milliwatt. It is a four-level laser.

Construction: He–Ne laser consists of a long (length 10 to 100 cm) and narrow (diameter 2 to 10 mm) discharge tube filled with helium and neon gases with typical partial pressures of 1 mm Hg (1 torr) and 0.1 mm Hg (0.1 torr) respectively. *The lasing action takes place due to the transitions* in neon atoms while the helium atoms help in the excitation of neon atoms. The ends of the cavity are enclosed by two concave mirrors. One of the mirrors is 100% reflecting at the lasing frequency while the other is partially reflecting. Earlier, the mirrors used to be sealed inside the glass but the tube does not last long since the seal gets eroded. Therefore, as shown in Fig. 5.9, now an external mirror arrangement is preferred. The glass tube is closed by windows which are tilted at the Brewster angle. Such windows, allow the light waves with electric field in the plane of the paper to pass through without any reflection. The light waves with electric field perpendicular to the plane of the paper, are reflected away from the cavity. With such windows, the output laser beam thus gets linearly polarized. Inside the gas cell, there are electrodes connected to the terminals of a dc power supply.

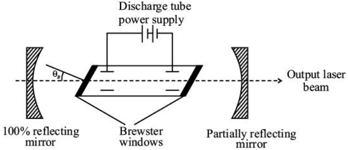


Fig. 5.9: He–Ne laser.

Lasing Action: The energy levels of *He* and *Ne* are shown in Fig. 5.10. When an electrical discharge is passed through the gas, high energy electrons produced in the tube collide with the gas atoms. As the concentration of helium atoms is higher, the probability of collision with *He* atoms is higher than that with the neon atoms. These collisions excite the helium atoms to the higher energy states. The helium atoms tend to accumulate in the metastable states E_{h2} and E_{h3}, with respective lifetimes of 10^{-4} and 5×10^{-6}s. The energy levels E_{n4} and E_{n6} of neon atoms have almost the same energy as the levels E_{h2} and E_{h3} of the helium atoms. Due to collisions between helium and neon atoms, the excited helium atoms excite the neon atoms to the levels, E_{n4} and E_{n6}. We can represent this process as:

$$\text{He}^* (E_{h2}) + \text{Ne}\,(E_{n1}) \rightarrow \text{He}\,(E_{h1}) + \text{Ne}^* (E_{n4})$$

$$\text{He}^* (E_{h3}) + \text{Ne}\,(E_{n1}) \rightarrow \text{He}\,(E_{h1}) + \text{Ne}^* (E_{n6})$$

where the letters in parentheses refer to the corresponding energy levels of gases.

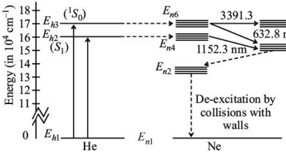


Fig. 5.10: Energy levels and the laser transitions in He–Ne laser.

The difference in the energy level of E_{h2} of He and E_{n4} of Ne is nearly 400 cm^{-1}. This difference in energy appears as the kinetic energy of the Ne atoms. Direct excitations of Ne atoms to the levels E_{n4} and E_{n6} are also possible. However, due to less number of Ne atoms in the tube, it is less likely. Depending upon the energy levels involved in the transition, the major transitions are as follows:

***(i)* The 6328 Å Transition:** When the lasing transition is from $E_{n6} \rightarrow E_{n3}$, the wavelength of the produced laser beam is 6328 Å. The level E_{n6} is $3S_2$ and E_{n3} is $2P_4$. This is the most commonly obtained laser beam in He–Ne laser. The lifetime of E_{n3} is of the order of 10^{-8} sec while that of E_{n6} is 10^{-7} sec, hence, population inversion build-up is possible between these two levels.

***(ii)* The 33913 Å Transition:** The transitions from $E_{n6} \rightarrow E_{n5}$ produce a laser beam of 33913 Å (or 3.3913 μm). The upper level is same in this case and in the 6328 Å transition.

***(iii)* The 11523 Å Transition:** This was the output wavelength of the first He–Ne laser. The transition from $E_{n4} \rightarrow E_{n3}$ produces photons beam of 11523 Å (or 1.1523 mm) wavelength.

He–Ne laser is widely used in laboratories where highly coherent and monochromatic sources are needed. It is also used in supermarket scanners, printers, Fourier transform spectrometers, holography, etc.

5.8 Applications of Lasers

Lasers are being used extensively nowadays in every field of life, from toys to warfare. Some of the important applications are mentioned below:

1. Communication: In *optical communication*, laser is used as optical carrier signal. Because of the large bandwidth, the information carrying capacity of laser light is enormous. The rate at which the information can be transmitted is proportional to bandwidth and the bandwidth is proportional to the carrier frequency which is of the order of 10^{15} Hz for the optical signals. High bandwidth together with low transmission loss, immunity to interference, signal security, reliability, etc., make optical communication the most advantageous mode of information transmission.

Laser is used in LIDAR (*Light Detection and Ranging*). It is used as *range finder*. The transit time of transmitted and reflected pulse of laser light is recorded and the distance of the reflecting object is estimated. This method is widely used for *finding range and detecting obstacle* in fog, smoke and underwater. Large range finders are used for military surveillance. Optical radars are used in aeroplanes, ambulances, police cars, etc.

In *high-speed photography*, laser is useful as it can detect fast moving bullets and missiles.

Laser is also used in holography which can be said to be a kind of three-dimensional photography. A normal photograph represents a two-dimensional recording of a three-dimensional scene. A hologram produces a three-dimensional image of a three-dimensional scene. A hologram also records the phase of the reflected light in addition to the amplitude of the reflected light. A hologram reproduces the image using a process called *reconstruction*. The availability of coherent laser radiation has made the reconstruction possible. The holographic pattern recognition is used for identifying fingerprints, postal address, etc. Holograms are also used as data storage devices.

2. Medical Applications: A narrow and intense beam of laser is used for destroying cancerous tissues. It is also used for other kinds of surgeries. During these processes, the heat generated seals up capillaries and blood vessels preventing blood loss. For certain internal surgeries, the laser beam is carried to the spot through an optical fibre.

The most successful application of laser has been in the eye surgery for the treatment of detached retina by the ophthalmologists. Very low power laser of short pulse duration is used for this purpose.

3. Industrial Applications: Lasers are used for cutting, welding, drilling, etc. Because the laser beam can be focussed onto a fine spot, it is particularly suitable for welding of fine wires, contacts in miniature assemblies, drilling holes, etc. It is used in aircraft and automobile industries to cut sheets of metals and alloys of thickness about 5 mm or less. Using lasers, holes as small as 10 μm in diameter can be drilled. Drilling of fine holes using laser is done in diamonds, watch jewels, etc.

Lasers are used for many purposes in electronics industry—for the manufacture of electronic components and integrated circuits. Lasers are used to perforate, divide silicon slices, selective evaporation, production of masks for integrated circuits, trimming of thick and thin film resistors, etc.

4. Scientific Research Applications: Laser is used for the determination of chemical and crystalline structures of molecules, i.e., laser spectroscopy. It is also used extensively in Raman spectroscopy to investigate the molecular structure. It is used to produce irreversible chemical changes in laser photochemistry.

It is used in non-linear optics, when a strong beam of laser light interacts with a medium, it polarizes the medium. The polarization *P* is given by:

$$\pi = c_1\mathbf{E} + c_2\mathbf{E}^2 + c_3\mathbf{E}^3 + \dots \quad \dots(5.23)$$

where **E** is the electric field associated with the incident radiation, c_1 is the dielectric susceptibility of the medium and c_2, c_3,... are higher order susceptibilities. With the normal incident light, only the first term which is linear in **E**, can be activated. The *non-linear* (or higher order) terms which gives the *non-linear effects* in **E** can be activated only by the radiations having extremely high electric field E. Only a giant pulse of lasers are capable of producing such fields.

5. Laser Induced Nuclear Fusion: Nuclear fusion is the process by which stars produce their energy and the scientists have been attempting to carry out nuclear fusion in the laboratory. One of the difficulties for nuclear fusion to occur is the requirement of very high temperature (~ 10^8 K). This is necessary to overcome the coulombic repulsion between the nuclei so that they fuse together. With the availability of high power laser pulse, creation of such high temperatures seems achievable. Many countries are developing laser enabled nuclear fusion facilities.

6. Applications in Metrology: The characteristics of lasers such as coherence, low divergence and monochromatism have been used in metrology. He – Ne laser is the most widely used laser because of its visible output, low power and low cost. Lasers are used for precision distance measurement, calibration testing comparison with standards, etc. Optical interference techniques using laser as a parallel monochromatic source has been used for the determination of diameter of very thin wires, to detect small variations in surface smoothness, etc. Laser scanning gauges to measure the roundeness and diameter have been made.

Exercises

Short Answer Type

1. What does the term laser stand for? Who constructed the first laser and when?
2. Name the characteristics which differentiate laser light from the normal light.
3. Define the term Monochromaticity. Give the order of magnitude of monochromaticity for ordinary light and laser light.
4. Differentiate between Spatial and Temporal Coherence.
5. Explain Spontaneous and Stimulated Emissions.
6. What is population inversion and what is its significance for the lasing action?
7. Why it is easier to achieve the lasing action in a four-level system compared to

7. Why it is easier to achieve the lasing action in a four-level system compared to that in a three-level system?

8. What is pumping? Briefly introduce various modes of pumping.

9. What is a Laser Oscillator? Briefly explain its working.

10. What is active medium? Broadly classify the lasers on the basis active medium.

11. The upper energy level for a laser system must be a short-lived one. Why?

12. Why is a Ruby Laser called so? What is the frequency of the laser light given out by a ruby laser?

13. What is the active material in a He – Ne laser? How is the population inversion achieved in a He – Ne laser?

14. What is LIDAR? Why is laser light suitable for this?

15. Mention a few important appplications of laser.

Long Answer Type

1. What do you mean by 'Laser'? Explain the characteristics which differentiate and make laser light useful compared to normal light.

2. Describe the following processes: (*a*) Absorption, (*b*) Spontaneous Emission, (*c*) Stimulated Emission, (*d*) Population Inversion, and (*e*) Pumping

3. Discuss Einstein's coefficients. Derive the relation between them.

4. Discuss cavity configurations and explain the role of cavity in a laser.

5. Name the basic components of a laser system. Explain the function of each component.

6. State and explain the threshold condition for lasering action.

7. Describe the working principle and construction of a ruby action.

8. Discuss how the lasing action is achieved in He–Ne laser. What are the major transitions in this laser?

9. Discuss the applications of lasers in the field of communication.

10. Explain medical applications of lasers.

11. Mention few industrial applications of lasers.

Numericals

1. If the pulse width of a laser of wavelength 1064 nm is 25 ms and the average output power per pulse is 0.8 W, calculate the number of photons in each pulse.

[**Ans.** 1.07×10^{17}]

2. The coherence time of an ordinary light source is 10^{-10} sec. Calculate the monochromaticity for the light of wavelength 5400 Å.

[**Ans.** 1.8×10^{-5}]

3. Calculate the number of photons emitted for a 2.5 mW He–Ne laser (λ = 6328 Å).

[**Ans.** 7.959×10^{15}]

4. A solar constant is radiant flux density at a spherical surface centred on the sun having a radius equal to that of the Earth's mean orbital radius, it has a value of 0.14 W cm^{-2}. If we assume an average wavelength of about 7000 Å, how many photons are incident on each square metre per second of a solar panel just above the atmosphere?

[**Ans.** 4.99×10^{21}]

6
Fibre Optics

6.1 INTRODUCTION

In the middle of the last century, experiments were carried out to find the viability of propagating information carrying light beam through the open atmosphere. It was realized that besides a huge loss of intensity, variable factors like rain, fog, dust, etc., are hindrances and a guiding medium is a must through which light can propagate. After intensive investigations, optical fibre was found to be the most viable option. Using optical fibre, information carrying light can be transmitted over a long distance without any significant loss. Optical fibre also enables us to peep into otherwise inaccessible locations like the interior of human body or the jet engines.

John Tyndall, in 1870, at Royal Society in England, for the first time showed that light could follow a curved path. In early 1950s, transmission of pictures along an aligned bundle of flexible glass fibres was attempted with remarkable success. This led to the development of the first *flexible fibrescope* by Harry Hopkins and N. Kapany. Kapany later developed the first practical glass-coated fibre and the term *fibre optics* was coined by him in 1956.

With the advent of lasers, the field of fibre optics got an added impetus. Nowadays, optical fibre is being used for the transmission of voice, television and digital signals. Optical fibres are replacing conventional coaxial cables for all forms of transmission.

6.2 Optical Fibre

An optical fibre is a very thin, flexible thread of transparent glass or plastic in which light is transmitted through multiple total internal reflection. It consists of a *central cylinder, called core, surrounded by a layer of material called cladding, which in turn is covered by a protective jacket.* It is the core through which the light actually propagates and is usually made up of glass or plastic of refractive index μ_1. Cladding keeps the light waves within the core because the refractive index μ_2 of the cladding material is less than that of the core ($\mu_2 < \mu_1$). It is also made up of glass or plastic and provides some strength to the core. The jacket protects the fibre from abrasion, crushing, moisture, etc. A schematic representation of an optical fibre is shown in Fig. 6.1. Typically, the core diameter varies from 5 µm to 100 µm while the cladding diameter is usually about 125 µm.

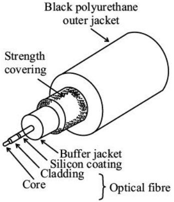


Fig. 6.1: Typical optical fibre.

Mainly glasses, silica and plastics are used in the manufacture of optical fibres. The basic requirement is the possibility to vary the refractive index. A typical fibre consists of an SiO_2: GeO_2 core with SiO_2 cladding. Addition of GeO_2 raises the refractive index of SiO_2. Plastic fibres have the advantage of higher flexibility over glass fibres and are preferred in systems that require higher flexibility. However, attenuation (or loss) of signal is higher in the plastic fibre than that in the glass fibre. It is essential that the core and cladding are free from microscopic cracks or flaws in the interior. In case of plastics, the core can be polysterene or polymethyl metaacrylate while the cladding is generally silicon or teflon.

Plastic fibres are made either entirely from plastic or with silica cores and plastic claddings. Full plastic fibres suffer from very high attenuation and hence are suitable for short distances and low bandwidth systems preferably in 600-700 nm wavelength range. These fibres can be made with large diameters due to higher flexibility.

Light Propagation in an Optical Fibre

The propagation of light through an optical fibre (step-index) is depicted in Fig. 6.2. It is the phenomenon of *total internal reflection* at the core-cladding interface which guides the light from one end of the fibre to the other end through the core of the fibre. We know that the two necessary conditions for the total internal reflection to take place are:

(i) The light wave should be propagating from an optically denser medium to a rarer medium.
(ii) The angle of incidence should be greater than the critical angle of incidence.

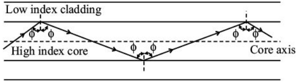


Fig. 6.2: Propagation of light in optical fibre.

As the refractive index μ_1 of core is higher than that of cladding μ_2, the first condition is satisfied for the light incident from the core on core-cladding interface.

If the angle of incidence is greater than critical angle of incidence for the core-cladding interface $\left(\phi > \phi_c = \sin^{-1}\frac{\mu_2}{\mu_1}\right)$ the light ray undergoes repeated total internal reflection at core-cladding interface.

For a core of refractive index μ_1 = 1.50 and a cladding of refractive index μ_2 = 1.47, the critical angle of incidence φ_c is equal to

$$\varphi_c = \sin^{-1}\left(\frac{1.47}{1.50}\right) = 78.5°$$

In Fig. 6.2, it is assumed that the angle of incidence φ at the interface is greater than the critical angle of incidence φ_c for the core-cladding interface and is, therefore, totally internally reflected, according to the laws of reflection. A light ray such as the one shown in Fig. 6.2 is known as meridional ray. A meridional ray is one which crosses the axis of the fibre.

Here, it is important to mention that we have assumed a perfect fibre, i.e., without any imperfection. Any imperfection and discontinuity at the core-clading interface can cause refraction, instead of total internal reflection causing loss of light ray into the cladding.

Acceptance Angle: From the above discussion, it is clear that not all the light rays incident on the core shall be transmitted through the fibre. Only the rays which are incident such that their angle of incidence at the interface is greater than φ_c, shall be transmitted. Other rays are lost even upon entering the fibre. This is illustrated in Fig. 6.3.

As shown in Fig. 6.3, any ray of light incident on the fibre core at an angle greater than θα (such as the ray *B*) will not be transmitted through the fibre because such a ray will be incident at the interface at an angle less than φ_c. Such a ray is refracted into the cladding and is eventually lost. In other words, *only the rays incident on the core at an angle less than a specific angle called acceptance angle* θα, *is transmitted through the fibre.*

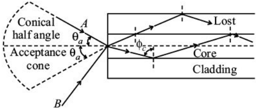


Fig. 6.3: The acceptance angle θα in an optical fibre.

The maximum (conical) angle (to the fibre axis) at which light entering the core is transmitted through the fibre, is called the *acceptance angle*. All the rays incident at the fibre at an angle greater than θα are eventually lost. The acceptance angle is unique for a particular fibre and different for different fibres and depends on the materials of core and cladding and core diameter.

From symmetry considerations, it can be seen that the emerging angle of the outgoing ray from the fibre is equal to the incidence angle of the entering ray into the fibre, provided the medium at both the ends are the same or they have the same refractive index.

6.3 Numerical Aperture

It is not always practical to talk about acceptance angle and refractive indices of core and cladding using the ray theory of light. A parameter, called the *numerical aperture* (NA) of the fibre, is more useful and is introduced below.

Figure 6.4 shows a light ray incident on the fibre core at an angle θ_1 to the fibre axis, which is less than the acceptance angle θα of the fibre. The ray enters the fibre from a medium (air) of refractive index μ_0. The core and cladding has refractive indices μ_1 and μ_2 ($\mu_1 > \mu_2$) respectively. Assuming the entrance face to be normal to the axis and considering the refraction at the air-core interface at *A*, from Snell's law, we have

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{\mu_1}{\mu_0} \quad ...(6.1)$$

From the right angled Δ*ABC*, we have:

$$\varphi = \frac{\pi}{2} - \theta_2$$

$$\theta_2 = \frac{\pi}{2} - \phi$$

$$\therefore \sin\theta_2 = \sin\left(\frac{\pi}{2} - \phi\right) = \cos\phi$$

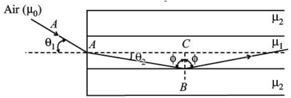


Fig. 6.4

Substituting for $\sin\theta_2$ in Eq. (6.1), we get

$$\frac{\sin\theta_1}{\cos\phi} = \frac{\mu_1}{\mu_0} \quad ...(6.2)$$

Using the identity $\sin^2\varphi + \cos^2\varphi = 1$, we have $\cos\phi = \sqrt{(1 - \sin^2\phi)}$. Thus, Eq. (6.2), can be written as

$$\frac{\sin\theta_1}{\sqrt{1 - \sin^2\phi}} = \frac{\mu_1}{\mu_0} \quad ...(6.3)$$

$$\phi = \phi_c = \sin^{-1}\left(\frac{\mu_2}{\mu_1}\right).$$

For the limiting case of total internal reflection,

Also, in the limiting case, θ_i becomes equal to acceptance angle for the fibre, $\theta = \theta\alpha$. Putting these limiting values in Eq. (6.3), we get

$$\mu_0 \sin\theta\alpha = \sqrt{\mu_1^2 - \mu_2^2} \quad \text{...(6.4)}$$

The parameter $\mu_0\sin\theta\alpha$ is called the *numerical aperture* (NA). Hence, the numerical aperture, sometimes also called the *figure of merit* of the optical fibre, is given by

$$NA = \mu_0 \sin\theta_a = \sqrt{\mu_1^2 - \mu_2^2} \quad \text{...(6.5)}$$

If the ends of the fibre are in air, then $\mu_0 = 1$. In that case,

$$NA = \sin\theta_a = \sqrt{\mu_1^2 - \mu_2^2} \quad \text{...(6.6)}$$

Normally, μ_2 is only a few percentage lower than μ_1 and, therefore, the following approximation can be made in order to express *NA* in terms of the relative refractive index difference Δ. From Eq. (6.6), we have

$$NA = \sqrt{(\mu_1+\mu_2)(\mu_1-\mu_2)} = \sqrt{2\mu_1(\mu_1-\mu_2)}$$

$$= \sqrt{2\mu_1^2\frac{(\mu_1-\mu_2)}{\mu_1}} = \mu_1\sqrt{2\Delta} \quad \text{...(6.7)}$$

where, $\mu_1 + \mu_2 \cong 2\mu_1$ and

$$\Delta = \frac{\mu_1 - \mu_2}{\mu_1}$$

Figure 6.5 shows the variation of numerical aperture with acceptance angle $\theta\alpha$.

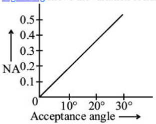


Fig. 6.5

As obvious from the above equations, a larger value of *NA* implies a larger acceptance angle. *NA is a measure of the light gathering ability of a fibre.* It is independent of the fibre core diameter and depends upon the refractive indices of the core and cladding.

Normally, the numerical aperture for fibres used in short distance communications are in the range of 0.4 to 0.5, while for the fibres used in long distance communications, it is kept between 0.1 and 0.3.

6.4 Types of Optical Fibres

There are two types of optical fibres: (*i*) step-index optical fibre and (*ii*) graded-index optical fibre.

(*i*) Step-Index Optical Fibre (SI Fibre): In this type of fibre, the core is homogeneous with constant refractive index μ_1 and the cladding has also a constant refractive index $\mu_2(\mu_1 > \mu_2)$, as shown in Fig. 6.26. (*a*). As there is an abrupt change of refractive index at the core-cladding interface, hence it is called a *step index fibre*. In Fig. 6.6 (*b*) the path of rays in a step-index fibre are shown. The two rays enter the fibre at different angles of incidence with the axis, travel along two different paths and emerge out at different times. An input pulse gets widened as it propagates through the fibre. The mechanism of light propagation in SI fibre is explained in Section 6.2.

Step-index fibre can be single-mode or multimode. Single-mode or mono-mode SI fibre allows only one mode of wavelength to propagate through it. The core diameter is very small, 2 to 8 μm. The numerical aperture and acceptance angle is small, making insertion of the light ray difficult. However, these fibres are most efficient and are used for high speed, large bandwidth, long distance transmission.

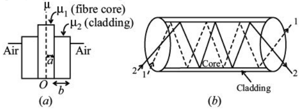


Fig. 6.6: Step-index fibre.

The multimode SI fibre has large diameter in the range 50 μm to 200 μm. As shown in Fig. 6.7, it supports large number of modes.

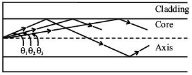


Fig. 6.7

The *marginal rays* (or *meridional rays*) take longer path for propagation than the *axial rays* and get delayed. This time delay known as *modal dispersion* (15 to 30 ns/km), causes distortion in the pulse. Due to this, short light pulse gets broadened, decreasing the transmission speed (or bandwidth). Also, there may be interaction between axial and marginal rays, causing mode mixing. These factors make these fibres less efficient. However, these fibres are cheap and have high numerical aperture and are used for short distance.

(*ii*) Graded-Index Optical Fibre (GRIN Fibre): In this type of fibre, the core has non-uniform refractive index that gradually decreases from the centre towards the core-cladding interface, while the cladding has a constant refractive index, as shown in Fig. 6.8 (*a*). The path of the rays are shown in Fig. 6.8 (*b*). These fibres are refractive type, i.e., light travels in a continuous curve undergoing refraction. A ray of light upon entering the fibre, bends and takes a periodic path along the axis. The rays entering the fibre at different angles follow different paths with the same period, both in space and time. Thus, there is a periodic self-focussing of the rays. The pulse dispersion is less in this type of fibre.

Light Propagation. In such fibres, as the ray goes from a region of higher refractive index to a region of lower-refractive index, it is bent away from the normal progressively till the condition of total internal reflection is met and the ray travels towards the core axis again being continuously refracted. The change of direction (or total internal reflection) may occur even before the ray reaches the core-cladding interface. The process repeats itself again and again. In these fibres, though the rays making larger angles with the axis traverse a longer path, they do so in a region of lower refractive index and hence, at a higher speed of propagation. Consequently, all rays have the same optical path and reach the other end at the same time giving a better transmission.

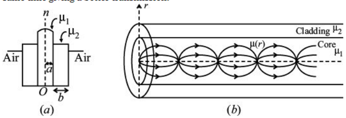


Fig. 6.8: Graded-index fibre.

The number of modes in a graded index fibre is only about half, compared to a similar step index fibre. The lower number of modes means less dispersion (≈ 2 ns/km). Hence, graded index fibres have larger data transmission capabilities. The manufacture of GRIN fibre is comparatively difficult owing to the problems in controlling the refractive index variation. The acceptance angle of these type of fibre being less than a similar step index fibre, coupling of the source to the fibre is difficult. The diameter is approximately 125 μm and costs more than SI fibres.

The size of the optical fibre is conventionally represented as the ratio of core and cladding diameters. Sizes of step-index fibres are $\frac{50}{125}, \frac{100}{140}, \frac{200}{230},$ etc. while the sizes of GRIN fibres are $\frac{50}{125}, \frac{62.5}{125}, \frac{80}{125},$ etc. A fibre with $\frac{50}{125}$ means the core of fibre is 50 μm in diameter, while the cladding is 125 μm in diameter.

Normalized Freθuency (*V*-number) of a Fibre: The *V*-number, also known as *normalized frequency* or *mode volume number* determines the number of modes a fibre can support or propagate.

It is given by, $$V = \frac{2\pi a}{\lambda}\sqrt{(\mu_1^2 - \mu_2^2)} \quad \text{...(6.8)}$$

where, a is radius of the core and λ is the wavelength (in free space).

Also, as the numerical aperture is given by:

$$NA = \sqrt{\mu_1^2 - \mu_2^2} = \mu_1\sqrt{2\Delta}$$

$$\therefore V = \frac{2\pi a}{\lambda}(NA) = \frac{2\pi a}{\lambda}\mu_1\sqrt{2\Delta} \quad \text{...(6.9)}$$

The maximum number of modes N_m for SI fibre is:

$$N_m = \frac{V^2}{2} \quad \text{...(6.10)}$$

If V = 20, the maximum number of mode is given by N_m = 200. In case $V < 2.405$, only one mode can be supported by the fibre and such a fibre is called single mode fibre (SMF). For multimode fibre (MMF), $V > 2.405$. *The wavelength corresponding to $V = 2.405$ is called the cut-off wavelength λ_c.*

The maximum number of modes in a graded index (GRIN) fibre is about half of that in SI fibre and is given by:

$$N_m = \frac{V^2}{4} \quad \text{...(6.11)}$$

6.5 Modes of Propagation in optical Fibre

An optical fibre is analogous to an optical waveguide. In optical fibre, light propagates as electromagnetic waves. The propagation of electromagnetic waves can be analyzed using Maxwell equations with appropriate boundary conditions. It is found that the propagating energy has distinct sets of solutions known as *modes. In simple words, we can express the modes of propagation in an optical fibre as the different distinct paths light can take while propagating through the fibre.*

In an optical fibre, all light rays entering within the acceptance angle will be trapped in the fibre due to total internal reflection. However, all the rays do not propagate through the fibre and only certain ray directions are allowed for propagations. *These allowed directions correspond to the modes of the fibre.* The path of the rays inside the fibre is zigzag (known as zigzag rays) except those propagating along the axial direction (also called *axial rays*). The zigzag rays undergo repeated reflection, undergoing some phase shift at each reflection. So light rays along some zigzag paths may be in phase and get intensified, while some others may be in antiphase, undergoing destructive interference and thus fading out. *The light ray paths along which the waves are in phase, constitute the modes.* The number of modes, a fibre can support depends on the ratio d / l, where d is diameter of the fibre core and λ is the wavelength of the propagating waves. Axial ray travels as the zero order ray while the higher order propagates at a smaller angle than the lower order modes. Figure 6.7 shows three modes of propagation. Each mode is a pattern of electric and magnetic fields distribution that is repeated along the fibre at equal interval. For lower order modes, the fields are concentrated near the centre of the fibre. For higher order modes, the fields are distributed towards the edge of the core and tend to send the light energy into cladding.

Optical fibres, on the basis of the modes of propagation, are classified as: (*i*) Single-mode fibre (SMF) and (*ii*) Multimode fibre (MMF). A single-mode fibre has a smaller core diameter while a multi-mode fibre has a large core diameter. A multimode step-index fibre has a diameter of 50 to 200 μm.

There are three basic types of fibre: *single-mode step-index fibre, multimode step-index fibre* and *multimode graded-index fibre*. The main features of these fibres are summarized in Table 6.1.

Table 6.1: Types of optical fibre

	Single-Mode Step-index Fibre	Multimode Step-index Fibre	Multimode Graded-index Fibre
1. Bandwidth	Very-very large; > 3 GHz-km	Large; < 200 MHz-km	Very large; 200 MHz to 3 GHz-km
2. Required Source	Laser	Laser or LED	Laser or LED
3. Splicing	Very difficult due to small core	Difficult	Difficult
4. Applications	Submarine cable system	Data links	Telephone trunk line
5. Cost	Less expensive	Least expensive	Most expensive

Example 6.1

In a step-index fibre, the refractive indices of core and cladding are 1.43 and 1.4 respectively. Calculate the critical angle, fractional refractive index, the acceptance angle and numerical aperture.

Solution:

Here, μ_1 = 1.43 and μ_2 = 1.4

Critical angle, θ_c = $$\sin^{-1}\frac{\mu_2}{\mu_1} = \sin^{-1}\left(\frac{1.4}{1.43}\right)$$

= $\sin^{-1}(0.9790)$ = 78.24°

Fractional refractive index

$$= \frac{\mu_1 - \mu_2}{\mu_1} = \frac{1.43 - 1.4}{1.43} = 0.020$$

Acceptance angle, $\theta\alpha$ =
$$\sin^{-1}(\sqrt{\mu_1^2 - \mu_2^2}) = \sin^{-1}(\sqrt{(1.43)^2 - (1.4)^2})$$
$$= \sin^{-1}(\sqrt{0.0849}) = \sin^{-1}(0.2914) = 16.94°$$

Numerical aperture = $\sin\theta_a$ = sin16.94 = 0.2914

Example 6.2

An optical fibre in air has a numerical aperture of 0.2, a core refractive index of 1.52 and a core diameter of 25 μm. It is operated at a wavelength of 1.3 μm. Find (*i*) the fibre *V*-number and (*ii*) the number of modes the fibre will support.

Solution:

Here, *NA* =
$$0.2,\ \mu_1 = 1.52,\ a = \frac{25}{2}\ \mu\text{m} = 12.5 \times 10^{-6}\ \text{m}$$

λ = 1.3 μm = 1.3 × 10^{-6} m

(*i*) *V* = $$\frac{2\pi a}{\lambda} NA = \frac{2 \times 3.14 \times 12.5 \times 10^{-6} \times 0.2}{1.3 \times 10^{-6}} = 12$$

(*ii*) Number of modes N_{max} = $$\frac{V^2}{2} = \frac{(12)^2}{2} = 72.$$

Example 6.3

An optical fibre with refractive indices of 1.47 and 1.46 of core and cladding respectively supports only one mode at a wavelength of 1300 nm. What is the maximum possible radius of the fibre?

Solution:

For single mode fibre, *V* < 2.405

or, $$\frac{2\pi a}{\lambda}\sqrt{\mu_1^2 - \mu_2^2} < 2.405$$

or, $$a < \frac{2.405\lambda}{2\pi\sqrt{\mu_1^2 - \mu_2^2}}$$

$$< \frac{2.405 \times 1300 \times 10^{-9}}{2 \times 3.14 \times \sqrt{(1.47)^2 - (1.46)^2}}$$

$$< \frac{2.405 \times 1.3 \times 10^{-6}}{6.28 \times \sqrt{(0.0293)}}$$

< 2.91 μm

Thus, the maximum radius = 2.91 μm.

Example 6.4

Find the maximum value of Δ for a single mode fibre with core of diameter 10 μm and refractive index 1.5. The signal transmitted through the fibre is of wavelength 1.3 μm while the *V* number of the fibre is 2.405. Also, find the refractive index of the cladding.

Solution:

We know, $V = \dfrac{\pi d}{\lambda} NA$ $(d = 2a)$

$\therefore NA =$

$$\frac{V\lambda}{\pi d} = \frac{2.405 \times 1.3 \times 10^{-6}}{3.14 \times 10 \times 10^{-6}} = 0.09957$$

$\therefore$ Acceptance angle

$\theta\alpha = \sin^{-1}(0.09957) = 5.71°$

Also, $NA = \mu_1\sqrt{2\Delta}$

or, $\Delta =$

$$\frac{NA^2}{2\mu_1^2} = \frac{(0.09957)^2}{2 \times (1.5)^2} = 0.0022 = 2.2 \times 10^{-3}$$

Also, $\Delta = \dfrac{\mu_1 - \mu_2}{\mu_1}$

$\therefore \mu_2 = \mu_1 - \Delta\mu_1 = \mu_1(1 - \Delta) = 1.5(1 - 0.0022) = 1.496$

6.6 Attenuation in Optical Fibre

Many types of losses occur during the propagation of light in an optical fibre, i.e., an optical signal gets progressively weaker as it propagates through the fibre. This reduction of signal strength is called *attenuation of signal*. The attenuation of signal is represented by a parameter, a, called fibre attenuation and is expressed as:

$$a = \frac{10}{L}\log\frac{P_i}{P_o} \quad ...(6.12)$$

where P_i is the input optical power, P_o the output optical power and L the length of the optical fibre in km.

In the ideal case, $P_i = P_o$ and a is equal to 0 dB/km. In practical cases, an attenuation upto 3 dB/km is within permissible limit. The attenuation is dependent upon the wavelength of the signal.

The losses in optical fibres occur due to the following three reasons:

(*i*) Absorption of signal
(*ii*) Scattering of signal
(*iii*) Geometrical effects.

(*i*) Absorption Loss: Absorption of light by glass (of which the fibre is made) and the impurities present is the major cause of loss, also called *material loss*, in optical fibres.

Even the purest glass absorbs radiation in certain wavelength ranges. This natural property of the substance is called the *intrinsic absorption*. Intrinsic absorption, particularly, the electronic absorption, is very strong in UV region while the vibrational absorption is strong in IR region (λ ~ 7 to 12 μm). However, the *wavelength range used in fibre communication is 0.1 to 1.6. μm*, so the above-mentioned losses are not important.

It is the loss due to *extrinsic absorption*, i.e., *absorption due to impurities*, which is the major source of loss in optical fibre. Two important types of impurities are: the transition metal ions and the hydroxyl (OH^-) ions. The transition metal impurities such as Fe, Cu, V, Co, Ni, Mn and Cr absorb strongly in the region of interest. Hence, the fibre should be free from these impurities is electronic absorption, i.e., low-lying electrons absorb photons to go to the higher energy state. The trapped OH^- ions in the material, absorb at 0.95, 1.23 and 1.39 μm, all in the region of interest (*A*, *B* and *C* respectively in Fig. 6.9). Therefore, the presence of OH^- ions should be minimized, must be kept below 0.01 part per million (ppm).

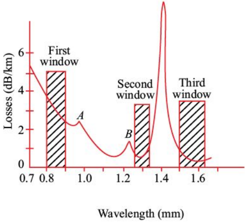


Fig. 6.9: Fibre losses over 0.7–1.6 μm.

Figure 6.9 shows the losses in optical fibre vs. wavelength for a silica based fibre. There are certain wavelength ranges at which attenuation is minimum. Such a range of wavelength is referred to as *optical window* or *transmission window* due to minimum attenuation in the wavelength range. These windows are suitable for transmission of information. The first, second and third window extends from 0.8–0.9 μm, 1.25–1.35 μm and 1.50–1.65 μm respectively.

(*ii*) Scattering Loss: During fibre manufacture, despite all precautions, localized microscopic variation in density and doping impurities cannot be removed completely due to which local variations in refractive index set in. These variations act as small scattering centres embedded in otherwise homogeneous medium. The sizes of these scattering centres are often smaller than the wavelength. A beam of light propagating through the fibre suffers losses due to Rayleigh scattering. Since, for Rayleigh scattering $\propto \dfrac{1}{\lambda^4}$, Rayleigh scattering sets a lower limit on wavelength that can be transmitted through a glass fibre to nearly 0.8 μm. Below this wavelength, scattering loss is appreciably high.

***(iii)* Loss Due to Geometrical Effects:** Bending (microscopic and macroscopic) of fibres, during manufacture and/or installation causes loss of power. Microbending is caused by small cracks in the glass. Macrobending occurs when the cracks extend to a large distance along the length of the fibre. Due to microbendig mode, coupling can occur while macrobending obstructs modes propagation.

6.7 Dispersion in Optical Fibres

Broadening (or spreading) of pulse during its propagation through an optical fibre is referred to as dispersion. Due to dispersion, a pulse is broader at the output of an optical fibre compared to that at the input. As a result, the pulse gets distorted (Fig. 6.10). The distortion of pulse due to dispersion effects is measured in terms of nanoseconds per km (ns/km).

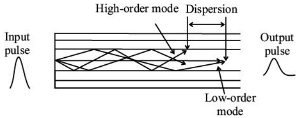


Fig. 6.10: Dispersion in multimode step-index fibre

There are three types of dispersion. These are:
(i) Intermodal dispersion, *(ii)* Material dispersion, and *(iii)* Wave guide dispersion.

***(i)* Intermodal Dispersion:** Intermodal dispersion is present in multimode fibres. A ray of light follows a zigzag path inside the fibre core. When a number of modes are available in a fibre, the axial and meridional rays travel at different speeds. The meridional rays take a longer path than the axial rays. Therefore, the axial rays reach at the output earlier than the meridional rays, i.e., there is a time delay between different rays. This causes a spread in the pulse or a dispersion is produced. This type of dispersion does not depend on the spectral width of the source and even a pulse from a pure monochromatic source shows intermodal dispersion. Naturally, single-mode fibre (SMF) does not show this dispersion. Figure 6.11 shows the paths taken by the two rays of different modes.

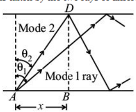


Fig. 6.11

Let $AB = x$, so that:

$$AD = \frac{x}{\sin\theta_2}$$

If L is the length of the optical fibre, then from the above equation, it follows that the total zigzag path L_t, is given by:

$$L_t = \frac{L}{\sin\theta_2}$$

$$\text{In general, } L_t = \frac{L}{\sin\theta} \quad ...(6.13)$$

where q is the angle of incidence in the relevant mode, with the normal to the wall of the fibre. In the highest order mode, the angle of incidence is minimum, equal to the critical angle of incidence θ_c,

$$\theta_{min} = \theta_c = \sin^{-1}\frac{\mu_2}{\mu_1}$$

where μ_2 is the refractive index of the cladding material and μ_1 that of core material. In the lowest order mode, the path length is given by:

$$L_{min} = \frac{L}{\sin 90^\circ} = L \quad (\{\ \theta_{max} = 90^\circ)$$

The highest mode path length is given by:

$$L_{max} = \frac{L}{\sin\theta_{min}}$$

$$= \frac{L}{\sin\theta_c} = L\cdot\frac{\mu_1}{\mu_2}\left(\because \sin\theta_c = \frac{\mu_2}{\mu_1}\right)$$

Let, ΔL be the path difference between the two rays,

$$\Delta L = L_{max} - L_{min}$$

$$= L\left(\frac{\mu_1}{\mu_2} - 1\right)$$

As, relative refractive index difference

$$\Delta = \frac{\mu_1 - \mu_2}{\mu_1}$$

$$\Delta L = L\left(\frac{\Delta}{1-\Delta}\right) \quad ...(6.14)$$

The fibre acts as a dielectric medium of dielectric constant (say) ε_r (> 1), so the speed of light inside the fibre is less than the speed of light in free space. The phase velocity v_p is given by:

$$v_p = \frac{1}{\sqrt{\mu_r \mu_o \varepsilon_r \varepsilon_o}} \quad ...(6.15)$$

where μ_o is the permeability in free space and ε_o is the permittivity in free space. The speed of light in free space is given by

$$c = \frac{1}{\sqrt{\mu_o \varepsilon_o}}$$

$$\therefore v_p = \frac{c}{\sqrt{\mu_r \varepsilon_r}}$$

Since, $\mu_r = 1$,

$$\therefore v_p = \frac{c}{\sqrt{\varepsilon_r}} \quad ...(6.16)$$

This phase velocity is the same as that in a transmission line for which the relation between the refractive index and dielectric constant is given by:

$$\mu^2 = \varepsilon_r$$

Using the above expression, Eq. (6.16), becomes:

$$v_p = \frac{c}{\mu_1} \quad ...(6.17)$$

If Δt is the time delay between the highest and lowest order mode, then

$$\Delta t = \frac{\Delta x}{v_p}$$

$$= L \cdot \left(\frac{\Delta}{1-\Delta}\right) \cdot \frac{1}{v_p} = L \cdot \left(\frac{\Delta}{1-\Delta}\right) \cdot \frac{\mu_1}{c}$$

$$= \frac{L}{c} \cdot \frac{\Delta}{1-\Delta} \cdot \mu_1 \quad ...(6.18)$$

This time delay is the characteristic of the fibre and is of the order of nanosecond per kilometre.

The graded-index (GRIN) fibres have a parabolic refractive index profile and the light takes zigzag path from a higher to a lower refractive index and back again. On averaging, the time delay due to intermodal dispersion is given by

$$\Delta t = \frac{\mu_1 L \Delta^2}{8c} \quad ...(6.19)$$

where μ_1 is the refractive index at the core centre and Δ is the fractional refractive index change from core centre to cladding. The GRIN fibres have lower intermodal dispersion (less than 1 ns/km) compared to SI fibres. This is because light travels from the centre towards cladding, taking a zigzag path, through regions of decreasing refractive index. We know, phase velocity $v_p = \frac{c}{\mu}$, so the phase velocity increases with distance and is maximum at the outer edge. Reverse phenomenon occurs as the ray travels towards the core. Thus, there is an alternate rise and fall of group velocity with a higher average phase velocity in higher modes than that in the lower modes which propagates along the core axis, which results in lower intermodal dispersion.

(ii) Material Dispersion: Light of different wavelengths have different velocities in any medium. Even a short pulse of light is not strictly monochromatic but centred about certain wavelength. So, it consists of more than one wavelength. Light of shorter wavelengths travels slower than those of longer wavelengths. This causes dispersion at the output end and even a short pulse of light gets broadened as it moves through an optical fibre. This type of dispersion is called *material dispersion*. Obviously, the spectral width ($\Delta\lambda$) determines the extent of dispersion. The material dispersion in an optical fibre of length L is given by:

$$Dm = \frac{\lambda(\Delta\lambda)}{c} \cdot L \frac{d^2\mu}{d\lambda^2} \quad ...(6.20)$$

where λ is the peak wavelength, μ is refractive index of core and c is the speed of light.

An LED giving out light of wavelength 820 nm with $\Delta\lambda$ = 38 nm gives a material dispersion of 3 ns/km, while a laser with peak wavelength of 1140 nm and $\Delta\lambda$ = 3 nm gives a material dispersion of just 0.3 ns/km. Clearly, lower is the spectral width of the pulse lower is the material dispersion.

(iii) Waveguide Dispersion: Waveguide dispersion arises due to the guiding properties of the fibre and is important in single-mode fibres. Like material dispersion, waveguide dispersion is also wavelength dependent. Effective refractive index for a mode of propagation varies with wavelength. Higher the wavelength, more is the fundamental mode spread from the core into cladding which causes the fundamental mode to propagate faster. This phenomenon is wavelength dependent and causes a time delay at the output. Waveguide dispersion can be reduced by the profile structure of the refractive index.

The total dispersion is equal to the root mean square of all the dispersions. In MMF, all the dispersions are observed simultaneously while in SMF, only material and wave guide dispersion are observed. In the fibres with low numerical aperture, smaller dispersion is observed while in fibres with high numerical aperture, large dispersion is observed. Dispersion can be reduced considerably with source of narrow spectral width and low numerical aperture fibre.

Dispersion limits the bandwidth of the fibre. Bandwidth is basically the range of frequency that can be transmitted without any significant distortion. Bandwidth for an optical fibre is given by:

$$\text{Bandwidth (MHz, km)} = \frac{310}{\text{Dispersion (ns/km)}} \quad ...(6.21)$$

Bandwidth is large for fibre with low numerical aperture. The product of bandwidth-distance specifies the information carrying capacity of an optical fibre.

Bits per second gives the rate at which information can be transferred through the fibre. Maximum bit rate B_{max} is given by:

$$B_{max} = \frac{1}{5\,(\text{Dispersion})} \quad ...(6.22)$$

Example 6.5

An optical fibre has an attenuation of 3.5 dB/km at 850 nm. If 0.5 MW of optical power is the input power of the fibre, calculate the power level after 4 km of propagation through the fibre.

Solution:

Here, attenuation, a = 3.5 dB/km

P_i = 0.5 MW, L = 4 km, P_o =?

We know, $a = \frac{10}{L}\log\frac{P_i}{P_o}$

$$\therefore \quad 10\log\frac{P_i}{P_o} = a.L = 3.5 \times 4 = 14 \text{ dB}$$

$$\log_{10}\frac{P_i}{P_o} = 1.4$$

$$\log_{10}\frac{0.5}{P_o} = 1.4$$

Taking antilog on both sides,

$$\frac{0.5}{P_o} = 25.11$$

$$\text{or, } P_o = \frac{0.5}{25.11} = 19.9\,\mu\text{W}.$$

Example 6.6

Calculate the dispersion/km of length and total dispersion in a 10 km length of a SI fibre having core of refractive index 1.558 and fractional refractive index 0.026.

Solution:

Maximum dispersion

$$= \frac{\mu_1}{c}\cdot\frac{\Delta}{1-\Delta}$$

$\therefore$ Maximum dispersion/km

$$= \frac{1.558}{3\times10^8}\times\frac{0.026}{1-0.026}\times1000 = 13.8\times10^{-9}\text{ sec}$$

= 13.8 ns.

Dispersion in 10 km = 13.8 × 10 = 138 ns.

Example 6.7

A SI multimode fibre has a core refractive index of 1.5 and cladding refractive index of 1.498. Find (*i*) intermodal factor of the fibre, (*ii*) dispersion in a 18 km length, and (*iii*) maximum bit rate allowed.

Solution:

Here, $\mu_1 = 1.5$, $\mu_2 = 1.498$

$$\Delta = \frac{1.5-1.498}{1.5} = 0.00133$$

(*i*) For 1 km length of fibre,

$$\Delta t = \frac{\mu_1 L}{c}\left(\frac{\Delta}{1-\Delta}\right)$$

$$= \frac{1.5\times1000\text{ m}}{3\times10^8}\times\frac{0.00133}{1-0.00133}$$

$= 0.6659 \times 10^{-8}$ sec = 6.659 ns/km

(*ii*) Total dispersion for 18 km length

Δt_{total} = 6.659 × 18 ns = 119.86 ns

$$(iii)\ B_{max} = \frac{1}{5\Delta t_{total}} = \frac{1}{5\times119.86} = 0.00166\times10^9\text{ bits / sec}$$

$= 1.66 \times 10^6$ bits/sec = 1.66 M bits/sec.

6.8 Optical Communications

Optical communication refers to the transfer of information carrying light from one place to another, using light wave as a carrier. Most often, optical fibre is used as the medium which guides light waves. Therefore, often the term optical fibre communication is used. The enormous information carrying capacity of the optical system makes optical communication unique. The frequencies of speech signal lie from 20 to 4000 Hz, i.e., a bandwidth of 4 kHz is required for the transmission of speech signal. The bandwidth of 20 kHz and 6 MHz are required for the transmission of music and television signals respectively. The frequency of a microwave channel extends from 10^9 to 5×10^9 Hz. Therefore, in the microwave frequency range, nearly 10^6 speech signals, or 10^5 music signals or 700 TV signals can be transmitted. The frequency of the visible region extends from 4.2×10^{14} Hz to 7.5×10^{14} Hz and, therefore, it is possible to transmit about 10^8 TV signals simultaneously. In other words, the information carrying capacity of light is extremely high.

Optical Fibre Communication System

A block diagram representation of optical fibre communication system is shown in Fig. 6.12. It basically consists of three blocks: transmitter, optical fibre and the receiver. A transducer is used to convert non-electrical message into electrical signals. During the transmission process, electrical signal is converted into light by the process called modulation. The light waves act as carrier waves. Light waves are generated by a suitable semiconductor laser or a light emitting diode (LED). The message signal in the form of an electrical signal is supplied to the optical source and the light output follows the variation of the message signal, i.e., modulated output is generated. Digital modulation is preferred over analogue modulation due to better efficiency of digital modulation. Digitally modulated signals are used for long distance transmission while analogue modulated signals are restricted to shorter distances.

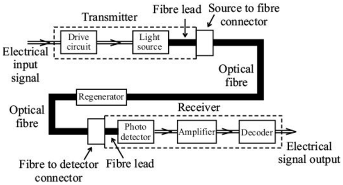


Fig. 6.12: Block diagram representation of an optical communication system.

Due to attenuation of optical signal over large distance due to various losses and dispersion in the optical fibre, the signal must be regenerated. For this, regenerators (or repeaters) are placed at regular intervals. A regenerator consists of a light receiver, photon to electron converter, amplifier, pulse shaper and finally the electron to photon converter. Together, these units strengthen the signal to the original level and transmit it onwards.

At the receiving end, an optical coupler is used to direct the light coming from the fibre to a photodetector which converts the received light into an electrical signal. The message is extracted from this electrical signal using a suitable decoder.

The information transmission capacity is increased by multiplexing several signals on to a single fibre. *Multiplexing* is the process in which more than one signal is transmitted over a single fibre.

Advantages of Optical Communication

Following are the advantages of communication using optical fibre:

(i) Extremely Large Bandwidth: More than 100 GHz of bandwidth is available for communication using optical fibre. Such a large bandwidth makes it possible to transmit a large volume of information or messages simultaneously over the same fibre.

(ii) Smaller Diameter and Light Weight: Optical fibres are very thin. The diameter of an optical fibre is less than that of a human hair. Even with the protective jacket, optical fibres are smaller in diameter and lighter in weight compared to conventional copper cables. This makes optical fibre specially suitable for use in satellites, ships, missiles, high-rise buildings, etc.

(iii) Immunity to Inductive Interference—Electrical Isolation: Because optical fibre is made up of dielectric materials, it is free from induced interference in the form of electrical noises, unlike in metallic cables. This makes it possible to use optical fibre even in a noisy electrical environment.

(iv) No Emission of Signal: Optical fibre cables do not radiate signals and so these do not interfere with other services. In other words, signal confinement is excellent.

(v) Negligible Crosstalk: In conventional communication circuits, signals often stray from one circuit to another, resulting in crosstalk, i.e., other calls being heard in the background. Even when large number of fibres are cabled together, there is negligible crosstalk.

(vi) Long Life: Metallic cables corrode and have a lifespan of around 12 to 15 years on an average. As glass or plastic, of which optical fibres are made, do not corrode, there is no corrosion problem with optical fibre and the estimated life of an optical fibre is around 25 to 30 years.

(vii) Greater Safety: In conventional cables, fear of short circuit prevents their use in areas such as chemical plants and coal mines with voltatile gases. There is no such problem with optical fibre and fibre can be repaired even when transmission is on.

(viii) High Tolerance to Environmental Factor: As optical fibres are made of glass or plastic, they have high tolerance to temperature extremes as well as to liquids and corrosive gases.

(ix) Conservation of Natural Resources: As commonly available sand and plastic, rather than scarce resources like copper and iron, is used to make optical fibres, it helps conserve the depleting natural resources.

(x) Low Cost: The glass and plastic, of which optical fibre is made, is much cheaper than the conventional conductors made from copper. Although, the cost of semiconductor laser, photodiodes, connector and couplers are high but considering the information carrying capacity, reliability, lifespan, security, etc., the overall cost is relatively lower.

Owing to the above-mentioned advantages, optical communication is progressively becoming the preferred mode of communication. More and more telecom companies are replacing their network by fibre communication based network.

6.9 Applications of Optical Fibres

Optical fibre is being extensively used in almost all fields of human activity. Some of the important applications are mentioned below:

***(i)* Optical Communication:** It is one of the most important applications of optical fibre. The advantages of optical communication has been discussed already. Optical fibres are replacing metallic telephone lines. Long distance optical fibre links have been established connecting various cities and countries. Compared to about 200 conversations at any instant through metallic cable, thousands of conversations are possible in an optical fibre cable. About 6500 km optical fibre cable link, the TAT-8 (transatlantic telephone), was established between USA, UK and France in 1988.

***(ii)* Medical Applications:** Optical fibres are used extensively for medical applications. Endoscopy uses optical fibre to inspect internal organs such as lungs, intestines, colon, etc. Doctors use optical fibre for simultaneous observation of the organs while performing microsurgery. Ophthalmologists, dermatologists, radiotherapists, etc. use optical fibre based instruments for diagnosis.

***(iii)* Millitary Applications:** Optical fibres are being used in various warfare and millitary applications. Due to their lightweight, high capacity, secrecy, etc. optical fibres are more convenient than copper cables in submarines, ships, satellites, earth stations and transmission lines. Optical sensors are mounted on the missiles to collect video information which is transmitted to ground control, from where further command is sent to the missiles. This helps in choosing the target.

***(iv)* Entertainment/Television Applications:** Optical fibres are used for short transmission links from the studio to the transmitter or live event to the equipment van. A coherent optical fibre bundle can be used to enhance the size of an image on a television screen.

***(v)* Industrial Applications:** Optical fibre is used in fabrication of the fibroscope, which is used to examine welds, nozzles and combustion chambers inside jet aircraft engines, which is otherwise inaccessible. In automobile industry, optical fibre cables are used for light display panels and to indicate lamp failure.

***(vi)* Computer Networking:** Optical fibres are used for networking of computers and other equipment in local area networks (LANs), generally used for linking systems in a building or distance less than 1 km. It is also used for metropolitan area networks (MANs) used for linking systems in a big city and wide area networks (WANs) used for linking systems in a wide area of a city. The optical links carry voice, video, data linking computers, TV and telephone including video telephone. There are many more applications like in holography, sensors, etc.

Exercises

Short Answer Type

1. Who developed the first optical fibre and when?

2. Explain the basic structure of an optical fibre.

3. Why refractive index of core is higher than that of cladding?

4. What is the use of a jacket in an optical fibre?

5. Give the relative merits and demerits of plastic fibre and glass fibre.

6. What do you mean by acceptance angle of an optical fibre? Explain.

7. What is numerical aperture of an optical fibre? What is its physical meaning?

8. Differentiate between step-index and graded-index fibres.

9. What is V-number? What is its significance?

10. What do you mean by modes of propagation? Explain.

11. What are meridional and axial rays?

12. What is optical communication? What do you mean by large bandwidth?

13. Mention any three important advantages of optical communication.

14. What are optical fibre sensors? Give the principle of working of a phase sensor and an intensity modulated sensor.

15. Briefly explain the medical applications of optical fibres.

16. Name the coupling components of optical fibres and briefly explain their role.

17. What do you mean by attenuation of signal? Explain.

18. Name the three different types of losses in an optical fibre.

19. What do you mean by 'dispersion' in optical fibre? Name various types of dispersion.

20. What is intermodal dispersion? Does a SMF show intermodal dispersion?

21. Explain the cause of material dispersion.

Long Answer Type

1. What is an optical fibre? Giving suitable diagram, explain the construction and role of each layer in an optical fibre.

2. Explain the terms: *(i)* acceptance cone, *(ii)* acceptance angle, and *(iii)* numerical aperture for an optical fibre. Show that numerical aperture is equal to $\sqrt{\mu_1^2 - \mu_2^2}$.

3. What is the difference between step-index and graded-index fibres. Explain with suitable diagrams. Also, discuss the mechanism of light propagation in both the types of fibres.

4. Differentiate between SMF and MMF fibres and explain their characteristics.

5. What are the sources of attenuation of signal in optical fibres? Explain.

6. What is dispersion? Explain various types of dispersion in optical fibres.

7. Explain the advantages of optical communication.

8. Explain different applications of optical fibres.

Numericals

1. A light ray enters from air (μ = 1) into an optical fibre. The fibre has core of refractive index 1.5 and cladding of refractive index 1.48. Calculate: (*i*) the critical angle, (*ii*) the fractional refractive index, (*iii*) the acceptance angle, and (*iv*) numerical aperture of the fibre.

[**Ans.** (*i*) 80.63°, (*ii*) 0.013, (*iii*) 14.13°, (*iv*) 0.244]

2. A step-index optical fibre has a NA of 0.16 and its core, of diameter of 60 μm, has refractive index 1.45. It is operated with a signal of wavelength 0.9 μm. What is the normalized frequency of the fibre and how many modes it can support?

[**Ans.** 2.093, 2]

3. A step-index fibre has core of refractive index 1.52, diameter of 29 μm and fractional refractive index of 0.007. Signal of wavelength 1.3 μm is transmitted through it. Calculate (*i*) the fibre *V*-number and (*ii*) the number of modes the fibre will support.

[**Ans.** 4.049, 8]

4. Find the core diameter necessary for single mode operation at 850 μm in SI fibre with μ_1 = 1.480 and μ_2 = 1.47. What is the NA and the maximum acceptance angle of the fibre?

[**Ans.** 0.1717, 9.89°]

7

Holography

7.1 Basic Principle of Holography

We are familiar with photograph and photography. In conventional photography, a photograph is a two-dimensional recording of a three-dimensional scene or object. What is recorded is merely the intensity distribution in the original scene on a light sensitive medium which is sensitive only to the intensity variations of light. Consequently, the information about the relative phases of the light waves from different parts of the scene is lost. In other words, information about the phase or the relative optical paths from different parts of the scene is lost. Therefore, a photograph gives a two-dimensional view of a three-dimensional object. Thus, the three-dimensional character (e.g. parallax) of the object is lost.

This limitation of photography is overcome in holography. In holography the complete wave field, both the intensity and the phase of the light scattered by the object is recorded. The holography was developed by Dennis Gabor 1947 (awarded Nobel Prize in 1971) in which he conceived the idea of recording the phase and amplitude of the light scattered by the object. Since all recording media respond only to the intensity, it is necessary to convert the phase information of the scattered light into intensity variation. This is done by coherent illumination discussed in the next section. A hologram is a two-dimensional recording but produces a three-dimensional image (Holos in Greek means, 'whole'). The wide availability of laser light has made the use of holography more widespread.

7.2 construction and reconstruction of image on hologram

The process of recording the image on a hologram or the construction of hologram is shown in Fig. 7.1. In this process, also called coherent illumination, the scattered wave (emanating from the object) is superimposed with another coherent wave, called the reference wave (usually a plane wave) and the photographic plate is made to record the resulting interference pattern. Thus, what is recorded on the photographic plate is the interference pattern produced by the two interfering waves. The intensity at any point in this pattern depends on the phase as well as the amplitude of the wave scattered by the object. The processed photographic plate, called a hologram, contains information on both the phase and the amplitude of the object wave. Unlike a photograph, a hologram bears no resemblance to the object and the information in the hologram exists in coded form.

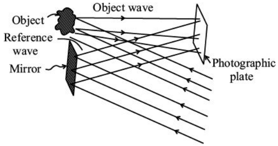


Fig. 7.1: Recording of a hologram.

The technique by which image is obtained from a hologram is called *reconstruction*. In the reconstruction process, the hologram is illuminated by a wave called the *reconstruction wave*. The reconstruction wave is identical to the reference wave used during the formation of the hologram. When the hologram is illuminated by the reconstruction wave, two waves are produced (Fig. 7.2). One wave appears to diverge from the object and provides the virtual image of the object. The second wave converges to form a real image of the object. The virtual image has all the characteristics of the object like 3D view, parallax, etc. Therefore, one can have different views of the object by changing the position of the eye. The real image can be photographed directly by placing a light sensitive medium in the location of the real image.

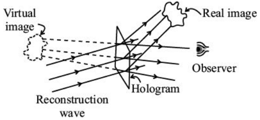


Fig. 7.2: Reconstruction process.

In a hologram each point contains light from the whole of the original scene. Therefore, whole scene can be reconstructed from an arbitrarily small part of the hologram. It is possible to break a hologram in small pieces and reconstruct the entire object from each small piece. However, the resolution decreases as the size of the hologram is reduced, i.e, the image becomes fuzzier.

7.3 applications of holography

The principle of holography is being used for a variety of applications. Some of them are mentioned below.

(*i*) Interferometry: Interferometry is one of the most important applications of the holography. The ability of the holographic process to release the object wave when reconstructed with a reconstruction wave makes it possible to perform interference between different waves which exist at different times. Therefore, in a technique often called double exposure holographic interferometry, the photographic plate is first partially exposed to the object wave and the reference wave. Then the object is stressed and the photographic plate is again exposed along with the same reference wave. The photographic plate is developed to obtain a hologram. When this hologram is illuminated with a reconstruction wave, two object waves emerge from the hologram—one corresponding to the unstressed object and the other corresponding to the stressed object. These two object waves interfere to produce interference fringes. These fringes are characteristic of the strain suffered by the body. When viewed through the hologram one finds a reconstruction of the object superimposed with fringes. Analysis (shape and number) of the fringe pattern produced in the body gives the distribution of strain in the object. This technique is used in *non-destructive testing* (NDT).

(*ii*) Data Storage: Holography is used in holographic data storage. Using this technique one can store information at high density inside crystals or photopolymers. This is highly important for electronic data storage devices.

(*iii*) Transient Microscopy: The ability of the holograms to record the information about the depth of the recorded scene makes holography useful in studying transient microscopic events. If one intends to study a transient phenomenon occurring in certain volume. Then using ordinary microscopy it becomes difficult to locate the exact position and make observation. However, if one records the hologram of the scene, then the event gets 'frozen in' into the hologram. The observer can focus through the depth of the reconstructed image from the hologram and can make observations at his own pace.

(*iv*) Character Recognition: Holography is being used to identify fingerprints, postal address, etc. The complicated wavefront from an object is generated from a hologram by the simple wavefront of the reference beam. The process is reversible and, therefore, the reference wave can be generated by the object wave. This principle is the basis of holographic pattern recognition.

(*v*) Production of Diffraction Gratings: Holography techniques have been used to produce diffraction gratings. The conventional diffraction gratings are prone to defects such as errors in ruling engine. The holographic gratings are error free.

(*vi*) Acoustic Holography: Holograms can be made with acoustic waves (also with microwaves) and visible light can be used for reconstruction. This has made it possible to obtain pictures which are otherwise not visible.

In addition to the above-mentioned uses, holography is used in the production of photographic masks used in the production of microelectric circuit, artwork, security hologram, etc. Security holograms are difficult to forge as they are replicated from a master hologram which requires expensive, specialized and technologically advanced equipment. Large number of companies put security hologram on their product to avoid forgery. This has become a widespread application. The uses of holography is in fact endless.

Exercises

1. What is holography? Differentiate between photography and holography.
2. Mention the advantages of holograms over photographs.
3. Discuss the process of construction and reconstruction of image on a

hologram with suitable diagrams.

4. What do you mean by reference wave? Explain.

5. Discuss the applications of holography.

Unit IV

8

Mechanics

8.1 FUNDAMENTAL INTERACTIONS

We know about different types of forces like weight, friction, tension, fluid resistance, etc. But all the forces or interactions in nature can be classified into the following four types of interaction. In other words, there are following *four* types of *fundamental interaction* in nature.

(*i*) Gravitational interaction,
(*ii*) Electromagnetic interaction,
(*iii*) Strong (unclear) interaction, and
(*iv*) Weak interaction.

The first two are familiar forces in our everyday life, the other two involve interaction between subatomic particles and cannot be observed with unaided senses.

All the interactions are effected or mediated with the help of certain particles, called *gauge particles,* which are discussed below.

Gravitational Interaction: Gravitational interaction, the mutual attraction between two masses is the *weakest* force. It acts along a line joining the centres of the two bodies. It is a long range (in fact, it can be felt at any distance and can be called infinite range force) inverse square (i.e., it decreases as $1 / r^2$) force. The gravitational interaction is mediated by a massless particle called the *graviton*. The graviton travels at the speed of light and has a spin of 2. This particle has not been observed, because of its extremely weak interaction with the measuring devices. The scientists expect that by detecting gravitational waves from supernova explosions, the existence of the particle graviton can be proved.

Electromagnetic Interaction: Electromagnetic interaction exists between charged particles and their interaction with electric and magnetic fields. It is mediated by exchange of *photons,* a massless particle of spin 1. The electromagnetic interaction is responsible for the processes like creation (or emission) of photon from the energy released in atomic (downward) transitions and absorption of photons in upward transitions. This is also a *long range force*.

This interaction is responsible for holding an electron in an atomic orbit, formation of molecules, generation and detection of electrical and optical signals and occurrence of many chemical and biological processes. In fact, almost all the interactions that we observe, which are not due to gravitational interaction, are due to electromagnetic interaction.

Strong Interaction: The strong interaction is responsible for binding the nucleons (protons and neutrons) together in a nucleus. It is the strongest force of all the fundamental forces. It keeps the protons inside the nucleus despite their repulsive force. It is responsible for nuclear reactions and decay of most of the elementary particles. It is a *short range force*.

The particles that do feel the strong interaction are called *hadrons* e.g., proton, neutron, pion, etc. In fact, none of these particles are truly elementary and are made up of six different elementary structureless particles called θ*uarks*. The particles that mediate the strong force between quarks are called *gluons*. A gloun is a massless particle of spin 1.

Weak Interaction: Weak interaction causes β-decays (in which electron and neutrino are produced) and is called the weak force because β-decays occur very slowly compared to the processes involving the strong force. Weak interaction is useful in explaining the properties of certain elementary particles. The particles responsible for the weak interaction are called the **W** (for weak) and **Z** particles. These particles are of spin one. The *electroweak theory,* unified the electromagnetic and weak interaction and its success gave strong evidence for **W** and **Z** particles. The electromagnetic and weak forces act independently except at extremely high particle energies.

The relative strength of an interaction determines the time scale over which it acts. A short time of the order of 10^{-23} second is sufficient for a strong force to cause a decay or a reaction. But for a weak force, a longer time of the order of 10^{-10} second is required. Hence, the mean lifetime of a decay process is an index of the type of interaction responsible for the process. Table 8.1 lists the relative strengths, range, gauge particles, etc. of the four fundamental forces.

Table 8.1

Interaction	Particles affected	Relative strength	Range	Gauge particle	Characteristic time	Role
Strong	quarks Hardons	1	10^{-15} m	Gluon Meson	$<10^{-22}$ s	Binds quarks to form nucleons, holds nucleons together to form nucleus
Electro-magnetic	Charged particles	10^{-2}	∞	Photon	$10^{-14}-10^{-20}$ s	Determines structure of atoms, molecules, solids and liquids.
Weak	Quarks	10^{-5}	10^{-18} m	W, Z	$10^{-8}-10^{-13}$ s	Determines composition of atomic nucleus
Graviton	All	10^{-39}	∞	Graviton	Years	Holds matters assembled.

The electroweak theory in a sense reduces the number of classes of interactions from four to three. Similar attempts are being made to understand strong, electromagnetic and weak interaction on the bases of a single unified theory called a *grand unified theory* (GUT) and further attempts are being made towards a possible unification of all interaction into a *theory of everything* (TOE).

8.2 law of conservation of energy

When only conservative forces are acting on a particle or a system of particles, *the mechanical energy i.e., the sum of its kinetic energy and potential energy remains conserved.* That is:

$$E = K + U = \text{constant}$$

The above statement is referred to as the *law of conservation of energy.*

The same can also be stated as, the change in total mechanical energy (ΔE) of a system is zero, i.e.,

$$\Delta K + \Delta U = \Delta E = 0$$

Consider a freely falling body, ignoring the air resistance. Let y_1 and v_1 be its height and velocity at one instant of time and y_2 and v_2 that at another. According to work-energy theorem, the total work done on the body equals the change in its kinetic energy, $W = DK = K_2 - K_1$. If the gravitational force is the only force acting on the body $W = -\Delta U = U_1 - U_2$ then equating the two expressions:

$$\Delta K = -\Delta U \text{ or } K_2 - K_1 = U_1 - U_2$$

$$\text{or, } K_1 + U_1 = K_2 + U_2$$

$$\text{or, } \frac{1}{2}mv_1^2 + mgy_1 = \frac{1}{2}mv_2^2 + mgy_2$$

$$\text{or, } E = K + U \text{ constant.}$$

In other words, when only the gravitational force is acting on a body, *the total mechanical energy of the system is conserved.*

When a ball is thrown up, its speed decreases with increasing height so its kinetic energy decreases ($\Delta K < 0$) while the potential energy increases ($\Delta U > 0$). The reverse is true when the ball is falling down. However total mechanical energy remains conserved.

When the body is acted upon by non-conservative force (e.g., frictional or viscous force) some amount of work appears in the form of heat, sound, light, etc. and is dissipated. Naturally, the amount of dissipation will depend on the actual path taken. When both conservative and a non-conservative force are acting on a body and W_c and W_N are the work done by conservative and non-conservative force respectively and ΔK is loss in its kinetic energy, then:

$$W_c + W_N = \Delta K$$

We know, $W_c = -\Delta U$, therefore

$$WN = -\Delta K + \Delta U$$

As, $\Delta K + \Delta U = \Delta E$, thus

$$\Delta E = W_N$$

That is, the change in the total energy of the particle is equal to the work done by the non-conservative force. When the non-conservative force is a frictional force, the work done by which appears as heat, $W_N = -Q$, then

$$\Delta E = -Q \text{ or } \Delta E + Q = 0$$

That is, the change in the total energy of the particle is zero or its total energy remains conserved. Thus, *the law of conservation of energy holds good in the case of both conservative and non-conservative forces.*

Example 8.1

In Fig 8.1, a curved track is shown on which a point of mass μ is moving. At the point *B* the horizontal velocity of the mass is v_0. Assume the track to be frictionless, except in the region *E* to *F* of length Δ. (*a*) Find the velocity of the mass at the points *C* and Δ. (*b*) Calculate the constant deceleration required to stop the mass at *F*, if the retarding frictional force starts operating at point *E*?

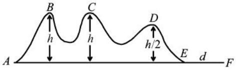


Fig. 8.1

Solution:

(*a*) Kinetic energy of the point mass at *B*,

$$K = \frac{1}{2}mv_0^2$$

Potential energy of the point mass at *B*,

$$U = mgh$$

∴ Total energy of the point mass at *B*,

$$E = K + U$$

$$= \frac{1}{2}mv_0^2 + mgh$$

Since at C the potential energy of the particle is the same as that at B, therefore, the principle of conservation of energy requires that its kinetic energy is also the same so as to keep the total energy conserved. Hence, the velocity of the mass at C is the same as that at B.

At D, the potential energy of the mass is $mgh/2$. Therefore, the principle of conservation of energy implies that the remaining potential energy $mgh/2$ is converted to kinetic energy at Δ. Thus,

Kinetic energy of the point mass at Δ,

$$K_D = \frac{1}{2}mv_0^2 + \frac{mgh}{2}$$

If v_D is the velocity at Δ, then

$$K_D = \frac{1}{2}mv_D^2$$

$$\therefore \frac{1}{2}mv_D^2 = \frac{1}{2}mv_0^2 + \frac{mgh}{2}$$

$$\Rightarrow v_D = \sqrt{v_0^2 + gh}$$

(*b*) At E, the potential energy is reduced to zero, so whole of its energy is now kinetic. If v_E is the velocity of the mass at E, then

$$\frac{1}{2}mv_E^2 = \frac{1}{2}mv_0^2 + mgh$$

$$\Rightarrow v_E = \sqrt{v_0^2 + 2gh} \quad ...(i)$$

Let a be the constant deceleration required to the stop the mass at F.

Then using $v^2 - u^2 = 2as$, we have

$$0 - v_E^2 = 2(-a)\Delta \Rightarrow -a = \frac{v_E^2}{2d}$$

Using Eq. (*i*), the required deceleration is

$$a = \frac{v_0^2 + 2gh}{2d}.$$

8.3 force and potential energy

In physics, we come across many situations in which we have the expression for the potential energy as a function of position (gravitational potential energy $U(y) = mgy$, elastic potential energy $U(x) = 1/2\, kx^2$) and have to find the corresponding force (gravitational force $F(y) = -mg$, elastic force $F(x) = -kx$). To develop the appropriate relation, consider the following example.

Consider the motion of a particle along x-axis so that the force can be denoted by $F(x)$ and the potential energy $U(x)$. Any displacement that the particle undergoes, the work done W by a conservative force equals, the negative of the change ΔU in potential energy.

$$W = -\Delta U \quad ...(8.1)$$

Now consider that the particle undergoes a small displacement Δx. The work done by the force $F(x)$ to be constant over the distance Δx. Using Eq. (8.1), we have

$$F(x)\,\Delta x = -\Delta U$$

$$\therefore F(x) = -\frac{\Delta U}{\Delta x}$$

In the limit $\Delta x \to 0$, $F(x) = -\dfrac{dU(x)}{dx}$ ...(8.2)

That is, a conservative *force is the negative of the rate of change of potential energy*. The above expression is understandable as where $U(x)$ changes rapidly with x, i.e., dU / dx is large, larger amount of work is done during a given period, during a given displacement and this requires a larger force magnitude. Also, when $F(x)$ is in the positive x-direction, $U(x)$ decreases with increase in x. In other words, $F(x)$ and dU / dx *have opposite sign*. Physically, Eq. (8.2) implies that *a conservative force always acts to push a particle towards a point of lower potential energy.*

For example, consider gravitational potential energy, $U(y) = mgy$

$$\therefore \text{Force } F(y) = \frac{dU(y)}{dy} = -\frac{d(mgy)}{dy} = -mg \quad ...(8.3)$$

which we know is the actual expression for the gravitational force. The variation of $U(y)$ and $F(y)$ is depicted in Fig. 8.2. Consider another example of elastic potential energy

$$U(x) = \frac{1}{2}kx^2.$$

$$\therefore \text{Force } F(x) = -\frac{dU(x)}{dx} = -kx \quad ...(8.4)$$

which we know is the actual expression for the force exerted by an ideal spring (Hook's law). The variations of $U(x)$ and $F(x)$ are shown in Fig. 8.3.

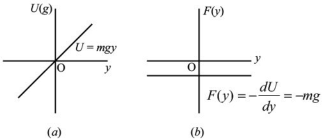


Fig. 8.2

Conservative Force as the Negative Gradient of U **(= ×** F **= 0)**

We can extend the previous discussions to a three-dimensional situation, where the particle may move in the *x*, *y* or *z*-directions under the action of a conservative force having components F_x, F_y, and F_z. Then the potential energy U is also a function of three coordinates. We can use Eq. (8.2) to find each component of force. The three components of the force are given by

$$F_x = \frac{\Delta U}{\Delta x},\ F_y = \frac{\Delta U}{\Delta y},\ F_z = -\frac{\Delta U}{\Delta z}$$

In the limits $\Delta x \to 0$, $\Delta y \to 0$, $\Delta z \to 0$ and expressing the derivative as partial derivative (using $\frac{\partial U}{\partial x}$ instead of $\frac{dU}{dx}$), we have

$$F_x = \frac{\partial U}{dx},\ F_y = \frac{\partial U}{\partial Y},\ F_z = \frac{\partial U}{dz}$$

As, $F = F_x\mathbf{i} + F_y\mathbf{j} + F_z\mathbf{k}$

$\therefore F =$

$$-\left(\frac{\partial U}{\partial x}\mathbf{i} + \frac{\partial U}{\partial y}\mathbf{j} + \frac{\partial U}{\partial z}\mathbf{k}\right) = -\left(\frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{dz}\mathbf{k}\right)U$$

$= -=U = -\text{grad}U$ ...(8.5)

$$\left(\therefore \text{Gradient } \nabla = -\frac{\partial}{dx}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{\partial z}\mathbf{k}\right)$$

Thus, a *conservative force is equal to the negative gradient of potential energy U*.

Now. curl **F** = = × F = −= × =U

$$-\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ F_x & F_y & F_z \end{vmatrix} = -\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ \frac{\partial U}{\partial x} & \frac{\partial U}{\partial y} & \frac{\partial U}{\partial z} \end{vmatrix}$$

=

=

$$-\left[\mathbf{i}\left(\frac{\partial^2 U}{\partial y\partial z} - \frac{\partial^2 U}{\partial z dy}\right) + \mathbf{j}\left(\frac{\partial^2 U}{\partial z\partial x} - \frac{\partial^2 U}{\partial x\partial z}\right) + \mathbf{k}\left(\frac{\partial^2 U}{\partial x\partial y} - \frac{\partial^2 U}{\partial y\partial x}\right)\right]$$

Since U is a perfect differential.

$$\frac{\partial^2 U}{\partial y\partial z} =$$

$$\frac{\partial^2 U}{\partial z\partial y}, \frac{\partial^2 U}{\partial z\partial x} = \frac{\partial^2 U}{\partial x\partial z}, \frac{\partial^2 U}{\partial x\partial y} = \frac{\partial^2 U}{\partial y\partial x},$$

we have

curl **F** = 0 ...(8.6)

The condition that curl **F** = 0 implies the existence of a potential function U such that **F** = −grad **U**. It thus follows that curl **F** = 0 is a necessary and sufficient condition for a conservative field of force **F.**

Example 8.2

A particle is moving in a potential energy field $U = A = Bx + Cx^2$, where all constants are positive.

(*a*) Obtain expression for the restoring force.

(*b*) At what point will this force vanish?

(*c*) Is this a point of stable equilibrium? If yes, obtain the value of the force constant.

Solution:

(*a*) The restoring force

$$F = -\frac{dU}{dx}$$

$$= -\frac{d}{dx}(A - Bx + Cx^2) = B - 2Cx$$

Thus, the linear restoring force,

$$F = B - 2Cx$$

(*b*) The force will vanish when

$$\frac{dU}{dx} = 0$$

or, $B - 2Cx = 0 \Rightarrow B = 2Cx$

or, $x = \frac{B}{2C}$

(*c*) For stable equilibrium $\frac{d^2U}{dx^2}$ must be +ve.

$$\frac{d^2U}{dx^2} = \frac{d}{dx}F = \frac{d}{dx}(B - 2Cx) = -2C$$

Since C is given to be positive, the point $x = \frac{B}{2C}$ will be a point of minimum potential energy and will, therefore, be a point of stable equilibrium.

From the relation $F = B - 2Cx$ ($F = kx$), it is clear that $2C$ is the force constant.

The variation of potential energy and force with x are shown in Fig. 8.3.

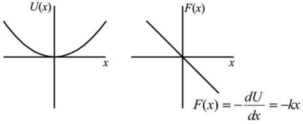


Fig. 8.3

8.4 conservative and non-conservative forces

A force, the work done by which in displacing a particle from one point to another, depends only on the initial and final positions and not on the actual path taken in reaching from the initial to final position is called a *conservative force*. While, *in the case of a non-conservative force, the work done depends on the actual path taken.*

All the central forces are conservative forces, e.g., gravitational and electrostatic forces. Such forces are *position-dependent forces,* since they depend only on the instantaneous position of the particle with respect to the fixed centre and nothing else. The work done by a conservative force is *reversible*. The work done by a conservative force in a cyclic process (same initial and terminal point) is zero.

In general, *velocity-dependent force, i.e., the forces* whose magnitude depends upon the magnitude and direction of velocity, like kinetic friction and fluid viscosity are non-conservative forces. Such forces cause mechanical energy to be lost or dissipated or *dissipative forces are non-conservative forces*. There are some non-conservative forces that increase mechanical energy. The fragments of an exploding firecracker fly off with very-large kinetic energy due to chemical reaction of gunpowder with oxygen. The forces generated by the reaction are non-conservative because the process in not reversible.

Lorentz force [q(**v**× **B**)], although velocity dependent, is conservative because it acts in a direction perpendicular to both the magnetic field **B** and the direction of the particle velocity **v.** Therefore, the work done by it, irrespective of the path taken is always zero and thus independent of the path.

A region of space in which a particle experiences a conservative force at every point is referred to as *conservative force field*. If the force happens to be the central force, the field is also referred to as *central-force field*.

Central Force—A Conservative Force: Central or position dependent force is expressed as $\mathbf{F} = F(r)\,\hat{\mathbf{r}}$, where $\mathbf{F}(r)$ is a function of r only (scalar) and $\hat{\mathbf{r}}$ is a unit vector along $\mathbf{r}$. The work done by such a force in displacing a particle from a point A to a point B is given by:

$$W_{AB} = \int_A^B \mathbf{F} \times d\cdot\mathbf{r} = \int_A^B F(r) \times \hat{\mathbf{r}}\, d\mathbf{r}$$

As, $\hat{\mathbf{r}} \cdot dr = \Delta r$, we have

$$W_{AB} = \int_A^B F(r)\,dr \quad \text{...(8.7)}$$

Thus, W_{AB} depends only on function form of $F(r)$ and the initial and terminal points. That is, the force **F** is a conservative force.

It is obvious from the above that the work done by the force **F** in moving the particle from B to A along any path (Fig. 8.4) is given by:

$$W_{BA} = \int_B^A \mathbf{F}\cdot d\mathbf{r} = -\int_A^B \mathbf{F}\cdot d\mathbf{r} = -W_{AB}$$

$$\Rightarrow W_{BA} - W_{AB} = 0 \quad \text{...(8.8)}$$

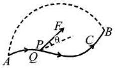


Fig. 8.4

The above expression implies that *the work done by a conservative force along a closed path is zero*. This fact is expressed as:

$$\oint \mathbf{F}\cdot d\mathbf{r} = 0 \quad \text{...(8.9)}$$

This is a *characteristic property of a conservative force*. In fact, *a conservative force is best defined as a force by which the work done in a round trip is zero and a non-conservative force as a force by which the work done in round trip is non-zero*. In a conservative force field, when the particle comes back to its initial point, it regains the same kinetic energy and potential energy as that before it started, so that its ability to do work is conserved. While it is not so in the case of a non-conservative force field.

Example 8.3

A particle lying in the $x - y$ plane acted upon by a force of magnitude kr($r = \sqrt{x^2 + y^2}$, distance from the origin) directed towards the origin. (*a*) Calculate the work needed to be done to move the particle from the origin to the point (1,1) along the radius vector. (*b*) What is the work done if the particle is first taken to (1, 0) and from there to (1, 1)? Is the force conservative?

Solution:

(*a*) Work done in moving the particle from the origin to *r* is,

$$W = \int_A^B \mathbf{F} \cdot d\mathbf{r} = \int_A^B \mathbf{F}\cos\theta \cdot dr = \int_0^r F dr$$

($\because$ Here, $F\cos\theta = F$)

Given, $F = kr$

$$\therefore W = \int_0^r kr dr = \left[\frac{kr^2}{2}\right]_0^2 = \frac{kr^2}{2} \quad ...(i)$$

At (1, 1), i.e., $x = 1$, $y = 1$, $r = \sqrt{1^2 + 1^2} = \sqrt{2}$

$$\therefore W = \frac{k(\sqrt{2})^2}{2} = k$$

Thus, the work done is *k* units.

(*b*) Work done in moving the particle from (0, 0) to (1, 0) is,

$$W = \int_0^1 k \cdot x dx = \left[\frac{kx^2}{2}\right]_0^1 = \frac{k}{2} - 0 = \frac{k}{2} \text{ units}$$

Work done in moving the particle from (1, 0) to (1, 1) is,

$$W = \int_0^1 ky\, dy = \left[\frac{ky^2}{2}\right]_0^1 = \frac{k}{2} - 0 = \frac{k}{2} \text{ units}$$

Total work done = $\frac{k}{2} + \frac{k}{2} = k$ units

Since the work done is path independent, the force is a conservative force.

Example 8.4

Show that the following two forces are conservative:

(*i*) $\mathbf{F} = (y^2 - x^2)\mathbf{i} + 2xy\mathbf{j}$

(*ii*) $\mathbf{F} = (2xy - z^2)\mathbf{i} + x^2\mathbf{j} + 2xz\mathbf{k}$

Also, calculate the work done on a body in case of the second force from the point *A* (0, 1, 2) to *B* (2, 3, 4).

Solution:

For a force to be conservative, the curl **F** should be zero.

$$(i) = \times \mathbf{F} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ y^2 - x^2 & 2xy & 0 \end{vmatrix}$$

$$= \left[\mathbf{i}\left(0 - \frac{\partial}{\partial z}(2xy)\right)\right] + \left[\mathbf{j}\left(\frac{\partial}{\partial z}(y^2 - x^2) - 0\right)\right] + \left[\mathbf{k}\left(\frac{\partial}{\partial x}(2xy) - \frac{\partial}{\partial y}(y^2 - x^2)\right)\right]$$

$= 0i + 0j + (2y - 2y)\mathbf{k} = 0$

Thus, the force $\mathbf{F} = (y^2 - x^2)\mathbf{i} + 2xy\mathbf{j}$ is a conservative force since $= \times \mathbf{F} = 0$.

$$(ii) = \times \mathbf{F} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ (2xy + z^2) & x^2 & 2xz \end{vmatrix}$$

$$= \mathbf{i}\left[\frac{\partial}{\partial y}(2xy) - \frac{\partial}{\partial z}x^2\right] + \mathbf{j}\left[\frac{\partial}{\partial z}(2xy + z^2) - \frac{\partial}{\partial x}(2xz)\right] + \mathbf{k}\left[\frac{\partial}{dx}x^2 - \frac{\partial}{\partial y}(2xy + z^2)\right]$$

$= (0 - 0)\mathbf{i} + (2z - 2z)\mathbf{j} + (2x - 2x)\mathbf{k} = 0$

Hence, this force $\mathbf{F} = (2xy + z^2)\mathbf{i} + x^2\mathbf{j} + 2xz\mathbf{k}$ is also a conservative force.

The work done, *W* by a force **F** in displacing a body from *A* to *B* is given by:

$$W = \int_A^B \mathbf{F} \cdot d\mathbf{r}$$

$\therefore W =$

$$\int_A^B [(2xy+z^2)\mathbf{i} + x^2\mathbf{j} + 2xz\,\mathbf{k}]\cdot[dx\,\mathbf{i} + dy\,\mathbf{j} + dz\,\mathbf{k}]$$

$$= \int_A^B (2xy+z^2)dx + x^2 dy + 2xz\,dz$$

$$= \int_A^B (2xy\,dx + x^2 dy) + (z^2 dx + 2xz\,dx)$$

$$= \int_A^B d(x^2 y) + d(z^2 x)$$

$$= \int_A^B d(x^2 y + z^2 x) = [x^2 y + z^2 x]_{0.1.2}^{2.3.4}$$

$= (4 \times 3 + 16 \times 2) - 0 = 44$ units

Thus, the work done by the force is 44 units.

Example 8.5

At any instant of time, position of a particle given by $\mathbf{r} = A\cos\theta\mathbf{i} + A\sin\theta\mathbf{j}$. Show that the force acting on the particle is a conservative force. Also find the expression for total energy.

Solution:

Position $\mathbf{r} = A\cos\theta\mathbf{i} + A\sin\theta\mathbf{j}$.

If ω is the angular velocity, then the position at instant t is,

$$\mathbf{r} = A\cos\omega t\mathbf{i} + A\sin\omega t\mathbf{j}$$

$\therefore$ Velocity $\mathbf{v}$ =

$$\frac{d\mathbf{r}}{dt} = A(-\sin\omega t)\cdot\omega\,\mathbf{i} + A\cos\omega t\cdot\omega\,\mathbf{j}$$

Acceleration a = $$\frac{d^2r}{dt^2} = \frac{dv}{dt} = -\omega^2(A\cos\omega t\,\mathbf{i} + A\sin\omega t\,\mathbf{j}) = -\omega^2 r$$

If μ is mass of the particle, force acting on it is:

$$F = \mu\mathbf{a} = -\mu\omega^2\mathbf{r}$$

The force to be conservative $= \times\,\mathbf{F} = 0$.

Here, $= \times\,\mathbf{F} =$ $$\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ x & y & 0 \end{vmatrix}$$ (q No component of $\mathbf{r}$ along $\mathbf{k}$)

Thus, the force acting on the particle is a conservative force. Therefore, $\mathbf{F} = -\text{grad}U$, where U is the potential energy of the particle.

$\mathbf{F} = -\text{grad}U =$

$$-\left(\frac{\partial U}{\partial x}\mathbf{i} + \frac{\partial U}{\partial y}\mathbf{j} + \frac{\partial U}{\partial z}\mathbf{k}\right) = -\frac{dU}{d\mathbf{r}}$$

$\therefore U =$ $$-\int \mathbf{F}\cdot dr = -\int(-m\omega^2\mathbf{r})\cdot d\mathbf{r} = m\omega^2 r^2$$

Since v is the linear velocity of the particle, its kinetic energy is given by:

$K =$ $$\frac{1}{2}mv^2 = \frac{1}{2}m\,(r\omega)^2\,\frac{1}{2}m\omega^2 r^2$$ (q $v = r\omega$)

Total energy, $K =$

$$K + U = \frac{1}{2}m\omega^2 r^2 + \frac{1}{2}m\omega^2 r^2 = m\omega^2 r^2$$

Thus, the expression for total energy is $E = \mu\omega^2 r^2$.

8.5 law of conservation of momentum

Consider a particle of mass m, moving with a linear velocity $\mathbf{v}$, *the product of its mass and linear velocity is called its momentum or linear momentum* $\mathbf{p}$.

Linear momentum = mass × velocity, $\mathbf{p} = \mu\mathbf{v}$.

According to Newton's second law of motion, the net force acting on a particle is equal to the time rate of change of momentum of the particle.

$\mathbf{F} =$ $$\frac{d\mathbf{p}}{dt} = \frac{d(m\mathbf{v})}{dt} \Rightarrow \mathbf{F}dt = d(m\mathbf{v})$$

Integrating with respect to time, we get

$$\int_{t_1}^{t_2}\mathbf{F}\,dt =$$

$$\int_{mv_1}^{mv_2} d\,(m\mathbf{v}) = m\mathbf{v}_2 - m\mathbf{v}_1 = \mathbf{p}_2 - \mathbf{p}_1 \qquad ...(8.10)$$

where $\mathbf{p}_2 - \mathbf{p}_1$ is the change in momentum, in the time interval $t_2 - t_1$ and $\int_{t_1}^{t_2}\mathbf{F}\,dt$ is the *impulse* of force during the interval $t_2 - t_1$. In other words, *the change in momentum of a particle during a time interval equals the impulse of the net force that acts on the particle during that interval.*

When the net force acting on a body is zero, so that,

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} = \frac{d(m\mathbf{v})}{dt} = 0$$

$$\Rightarrow \mathbf{p} = \mu\mathbf{v} = \text{a constant} \quad ...(8.11)$$

Thus, in the absence of an external force, the momentum of a body remains constant. Extending the above result to a system of two interacting bodies, the law of conservation of momentum states, *if the net external force on a system is zero, the total momentum of the system remains constant.* That is,

$$\mathbf{p}_1 + \mathbf{p}_2 = \mu_1\mathbf{v}_1 + \mu_2\mathbf{v}_2 = \text{constant} \quad ...(8.12)$$

We can generalize this principle. Let μ_1, μ_2, ..., m_n be the masses of n-interacting particles in a system having velocity $\mathbf{v}_1$, $\mathbf{v}_2$, ..., $\mathbf{v}_n$ respectively. The total momentum of the system is,

$$\mathbf{P} = \mathbf{P}_1 + \mathbf{p}_2 + ... + \mathbf{p}_n = \mu_1\mathbf{v}_1 + \mu_2\mathbf{v}_2 + ... + m_n\mathbf{v}_n$$

Differentiating with respect to time.

$$\frac{d\mathbf{P}}{dt} = \frac{d\mathbf{p}_1}{dt} + \frac{d\mathbf{p}_2}{dt} + \ldots = \frac{d}{dt}(\mathbf{p}_1 + \mathbf{p}_2 + \ldots + \mathbf{p}_n)$$

$$\Rightarrow \mathbf{F}_1 = \mathbf{F}_1 + \mathbf{F}_2 + ... + \mathbf{F}_n \quad ...(8.13)$$

where $\mathbf{F}_1$, $\mathbf{F}_2$, ..., $\mathbf{F}_n$ are the forces acting on the particles of masses μ_1, μ_2, ..., m_n respectively. The internal forces may be acting, but they cannot bring any change in the momentum, because forming pairs of equal and opposite forces they give rise to equal and opposite changes of momentum which cancel out.

If, $\mathbf{F} = 0 \Rightarrow \frac{d\mathbf{p}}{dt} = 0$ and therefore,

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

$$= \mu_1\mathbf{v}_1 + \mu_2\mathbf{v}_2 + \ldots + m_n\mathbf{v}_n \ldots = \text{constant} \quad ...(8.14)$$

That is, the sum of momentum $\mathbf{P}$ remains constant. Although, the individual values of $\mathbf{p}_1$, $\mathbf{p}_2$, ... may change but the sum remains unaltered. This is the *law of conservation of momentum* and may be stated as: *when the vector sum of the external forces acting upon a system of particles, equals zero, the total linear momentum of the system remains constant.* In a three-dimensional case, momentum being a vector quantity, each component (x-, y-, and z-) is conserved. The conservation of linear momentum of a system is a direct consequence of the translational invariance of the potential energy of the system.

8.6 collision

The common perception of the term *collision* is a body striking against another, like what happens when two vehicles moving in opposite direction undergo an accident or collide. However, in physics the term collision does not necessarily mean a particle striking against another, but any interaction of even a short duration in which energy and momentum of the system in affected. The essential feature of a collision is that due to it, a *redistribution of momentum* of the system takes place.

(i) Elastic Collision: When the forces between the bodies are conservative so that no mechanical energy is lost or gained in the collision, the total kinetic energy of the system is the same after the collision as before. Such *a collision in which the kinetic energy of the particles is conserved is called elastic collision.* In such a collision between two masses μ_1 and μ_2,

$$m_1\mathbf{u}_1 + \mu_2\mathbf{u}_2 = \mu_1\mathbf{v}_1 + m_2\mathbf{v}_2 \text{ and}$$

$$\frac{1}{2}m_1\mathbf{u}_1^2 + \frac{1}{2}m_2\mathbf{u}_2^2 = \frac{1}{2}m_1\mathbf{v}_1^2 + \frac{1}{2}m_2\mathbf{v}_2^2$$

where $\mathbf{u}_1$ and $\mathbf{u}_2$ are the velocities before collision and $\mathbf{v}_1$ and $\mathbf{v}_2$, that after the collision. Collision between some atomic fundamental particles are elastic collisions. Also, collision between two marbles or two billiard balls is almost elastic.

(ii) Inelastic Collision: A collision in which total *kinetic energy after the collision is less than that before the collision, is called on inelastic collisions.* A bullet embedded in a block of wood is an example of inelastic collision.

In such a collision, the loss of kinetic energy occurs mostly in the form of heat energy due to dissipation, increased vibration, etc. On atomic scale, an atom may be *excited* to higher energy state. If this loss is q, then,

$$\frac{1}{2}m_1\mathbf{u}_1^2 + \frac{1}{2}m_2 + \mathbf{u}_2^2 = \frac{1}{2}m_1\mathbf{v}_1^2 + \frac{1}{2}m_2\mathbf{v}_2^2 + Q$$

Normally, in an inelastic collision, the two particles stick together after the collision, thus the law of conservation of momentum requires.

$$\mu_1\mathbf{u}_1 + \mu_2\mathbf{u}_2 = (\mu_1 + \mu_2)\mathbf{v}$$

where $\mathbf{v}$ is the common velocity after the collision.

Example 8.6

A car of mass 1200 kg moving east with 12 m/s collides with a car of 800 kg mass moving north with 15 m/s. The cars stick together after the collision. With what velocity and in which direction wreckage will begin to move?

Solution:

The linear momentum must be conserved in both east-west and north-south directions, i.e., x and y directions respectively, as shown in Fig. 8.6. In accordance with the law of conservation of momentum,

before collision = momentum after collision

along x-direction

$$\mu_A v_{Ax} + \mu_B v_{Bx} = \mu_{AB} v_{ABx} \quad (i)$$

along y-direction:

$$\mu_A v_{Ay} + \mu_B v_{By} = A_B v_{ABy}$$

Here, $\mu_a = 1200$ kg $v_{Ax} = 12$ m/s $v_{Ay} = 0$

$\mu_B = 800$ kg $v_{Bx} = 0$ m/s $v_{By} = 15$ m/s

$\mu_{AB} = \mu_A + \mu_B = 2000$ kg $v_{ABx} = ?$ $v_{ABy} = ?$

From Eq. (*i*), we have

$$v_{ABx} = \frac{m_A v_{Ax} + m_B v_{Bx}}{m_{AB}} = \frac{1200 \times 12 + 0}{2000} = \frac{14400}{2000}$$

= 7.2 m/s.

From Eq. (*ii*), we have

$$v_{ABy} = \frac{m_A v_{Ay} + m_B v_{By}}{m_{AB}} = \frac{0 + 800 \times 15}{2000} = \frac{12000}{2000} = 6 \text{ m/s}$$

Therefore, the magnitude of the velocity v_{AB} is

$$v_{AB} = \sqrt{v_x^2 + v_y^2} = \sqrt{(7.2)^2 + (6)^2} = 9.37 \text{ m/s}$$

The direction of v_{AB} may be specified by the angle q between +*y (north)* and v_{AB}. From Fig. 8.5, we have,

$$\tan q = \frac{v_{ABx}}{v_{ABy}} = \frac{7.2}{6} = 1.2 \Rightarrow \theta = 50.19°$$

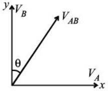


Fig. 8.5

Thus, after the collision the wreckage begins to move at 9.37 m/s in a direction 50.19 east of north.

Example 8.7

A 50 gm golf ball is stuck by a club and flies off at 80 m/s. If the head of the club was in contact with the ball for 0.4 m/s, how much average force was exerted on the ball during the impact?

Solution:

Since the ball was at rest before impact $v_1 = 0$.

Here, v_1 = m/s v_2 = 80 m/s

μ = 50 gm = 0.05 kg

∴ Change in momentum

$= \mu(v_2 - v_1) = 0.05(80 - 0)$

= 4 kg/s

Time of contact = 0.4 ms = 4×10^{-4} s

Force exerted

$$= \frac{\text{Change in momentum}}{\text{time}} = \frac{4 \text{ kg m/s}}{4 \times 10^{-4} \text{ s}}$$

= 10^4 Newton = 10 kN.

Example 8.8

An aeroplane weighing 60,000 kg can fly at a maximum velocity of 800 km/h. Its engine develops a total thrust of 75 kN. Ignoring air resistance, fuel consumption and change in altitude, how long does the plane take to reach the maximum velocity, starting from rest?

Solution:

Here, v_1 = 0 m/s

$$v_2 = 800 \text{ km/h} = \frac{800 \times 1000}{60 \times 60} \text{ m/s} = 222.22 \text{ m/s}$$

Impulse = change in momentum

$F \times t = \mu(v_2 - v_1) = mv_2$

$$t = \frac{60,000 \times 222.22}{75,000} = 178 \text{ seconds}$$

= 2 minute 58 seconds.

8.7 Central Forces: inverse square law forces

The forces which are directed towards or away from a fixed centre and act along the line joining the two bodies and whose magnitude is a function of the distance *r* between the centres of the two bodies are called *central forces*. Gravitational and electrostatic forces are central forces.

Let *r* be the distance between the two bodies, the central force operating between them is given by,

$$\mathbf{F} = F(r)\hat{\mathbf{r}}$$

where, $F(r)$ is a function of distance *r* and $\hat{\mathbf{r}}\left(\frac{\mathbf{r}}{r}\right)$ is a unit vector along the vector **r**. If one of the two bodies is fixed, the force on the other can be represented by,

$$\mathbf{F} = F(r)\frac{\hat{\mathbf{r}}}{r} \quad ...(8.15)$$

If the central force is attractive $F(r) < 0$ or –ve, while, if it is repulsive $F(r) > 0$ or +ve.

When a particle is moving under the influence of a central force, its angular momentum **J** *remains conserved.* This can be proved as follows. If τ is the torque acting on the body, then it is given by,

$$\tau = \frac{d\mathbf{J}}{dt} = \mathbf{r} \times \mathbf{F}$$

$$= \mathbf{r} \times F(r)\frac{\mathbf{r}}{r}$$

$$= 0$$

$$\Rightarrow \frac{d\mathbf{J}}{dt} = 0 \text{ or } \mathbf{J} = \text{constant.}$$

It can be seen that the *constancy of angular momentum implies that the motion of the particle remains confined in a plane.* If **p** is the linear momentum of the particle, then,

$$\mathbf{J} = \mathbf{r} \times \mathbf{p}$$

$$\text{or, } \mathbf{r} \cdot \mathbf{J} = \mathbf{r} \cdot (\mathbf{r} \times \mathbf{p})$$

Using triple product rule, $\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = (\mathbf{A} \times \mathbf{B}) \cdot \mathbf{C}$, we have,

$$\mathbf{r} \cdot \mathbf{J} = (\mathbf{r} \times \mathbf{r}) \cdot \mathbf{p} = 0 \; (\mathbf{A} \times \mathbf{A} = 0)$$

Thus, $\mathbf{r} \cdot \mathbf{J} = 0$, i.e., **r** and **J** are perpendicular to each other or the motion of the particle is confined to a plane.

The general form of central force is represented by the following expression, called *inverse nth power law,*

$$\mathbf{F} = \frac{C}{r^n}\hat{\mathbf{r}} \quad ...(8.16)$$

where C is a constant and is −ve if the force is attractive and +ve if the force is repulsive.

A central force is a conservative force, therefore, if U is the potential energy of a particle, then,

$$F(r) = -\text{grad } U = -\frac{dU}{dr} \quad ...(8.17)$$

Using Eq. (8.16) and (8.17), we get,

$$dU = \frac{C}{r^n}dr$$

Upon integrating, we have,

$$U(r) = \frac{C}{(n-1)r^{n-1}} + k,$$

where k is constant of integration.

At $r = \infty$, potential energy $U = 0$. Using this in the above equation, we find $k = 0$.

$$\therefore U(r) = \frac{C}{(n-1)\mathbf{r}^{n-1}} \quad ...(8.18)$$

Now, let us consider the following examples:

(*i*) If $n = -1$ and C is negative, from Eq. (8.17), we get

$$F(r) = -Cr$$

Therefore, from Eq. (8.18), we have,

$$U(r) = \frac{C}{2}r^2$$

which is a well-known expression for the potential energy of a simple harmonic oscillator. We know in a simple harmonic oscillator, the force is always directed towards a fixed point. Therefore, the force is a central force.

(*ii*) If $n = -2$ then from Eqs. (8.17) and (8.18), we get,

$$F(r) = \frac{C}{r^2} \text{ and } U(r) = \frac{C}{r}$$

where C may be +ve or −ve, if the force is electrostatic, depending on the sign of interacting charges. If the force is gravitational, C is always negative.

We find that the *force between two interacting particles* (or point charges) *is inversely proportional to the square of the distance between them.* Such force is known *as inverse square law force* and the law pertaining to it is called the *inverse square law, i.e., the force between two particles (or point charges) is inversely proportional to the square of the distance between them.*

The inverse square law occurs quite often in Physics. The two common laws are; the Newton's law of gravitation $\left(\mathbf{F} = -G\frac{m_1 m_2}{r^2}\hat{\mathbf{r}} \cdot C = -Gm_1 m_2\right)$ between two bodies of mass μ, and m_2 separated by r and Coulomb's law of electrostatic attractions repulsion

$$\left(\mathbf{F} = \pm k\frac{q_1 q_2}{r^2}\hat{\mathbf{r}},\ C = \pm k\, q_1 q_2,\ \text{−ve for attraction and +ve for repulsion}\right)$$

between two charge θ, and θ, separated by distance r.

Motion Under Inverse Square Law Force Field and Stability: When a large number of masses or point charges are acted upon by an Inverse Square Law Force, they cannot all remain at rest. This important conclusion can be drawn from the following discussion:

Consider a group of masses or charges and the equipotential lines or surfaces due to them. We know in the case of a point charge, the equipotential lines are concentric circles of increasing radii and decreasing values of potential with the point charge as centre of these circles. For a charged sphere, the charge behaves as if, it were concentrated at its centre and the equipotential surfaces are all concentric spheres of decreasing values of potential about the charged sphere. *For a test charge to be in stable equilibrium at any point in an electrostatic field. The potential of all other neighbouring points must be higher or lower than that at the point where the charge under consideration is located in accordance with the test charge being positive or negative. This is obviously impossible in the field of charges whether point charges or charged sphere.*

Now consider two fixed equal positive (or negative) charges (Fig. 8.6). The equipotential surfaces of these charges are shown in the figure. If a test charge is placed at any point in such an electrostatic field of the two charges, the potential there will neither be lower nor higher than that at all other neighbouring points and the test charge will not, therefore, remain in static equilibrium at that point or any other point in the considered electrostatic field. The same is true if the number of charges is more than two.

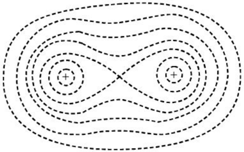


Fig. 8.6

However, if a charged particle, positive or negative, is placed at a point somewhere between two similar charges, where the electric field is zero (natural point), the test particle may remain in equilibrium. Although, it will not be in *stable equilibrium,* as a slight displacement from the equilibrium position will alter the field and the test charge will not be able to regain its previous position. Thus, we conclude that *no static equilibrium is possible in an Inverse Square Law Force Field.*

The above discussion is also applicable for any two bodies whether electron in an atom, the atoms in a star or the stars in a galaxy. We can, therefore, conclude that the *entire known universe must consist of moving bodies, in which a static state of the bodies is impossible.*

8.8 Symmetry, invariance and conservation laws

We have come across many conservation laws for momentum, energy, angular momentum, electric charge, baryon and lepton numbers, strangeness, parity, statistics, etc. Have you given a thought that if there is any relation between these conservation laws and some symmetry operation? But whether you have done this or not, long back, a German lady mathematician Emmy Noether worked on the relation between symmetry and conservation principle and concluded that *every conservation principle corresponds to a symmetry in nature.*

Symmetry implies some sort of invariance. A symmetry of a particular kind exists when a certain operation leaves something unchanged or invariant in appearance or any other feature. For example, rotation of a cylinder about its axis leaves the appearance of the cylinder unchanged. It is also symmetric with respect to its mirror image.

In physics, translation in space is the simplest symmetry operation, which implies that the laws of physics do not depend on where we choose the origin of the coordinate system. *The law of conservation of linear momentum is a consequence of the symmetry of the operation of translation in space. Another simple symmetry operation is translation in time,* i.e. it is immaterial when one chooses $t = 0$ or when you start measuring time. *This invariance gives rise to the principle of conservation of energy. Invariance under rotation in space, i.e.,* laws of physics do not depend on the orientation of the coordinate system in which they are expressed, *gives to the law of conservation of angular momentum.*

Conservation of electric charge is related to gauge transformations, which result in shifts of zero of the scalar potential $V(\boldsymbol{E} = -=V)$ and vector potential $\boldsymbol{A}(\boldsymbol{B} = -= \times \mathbf{A})$. Gauge transformations leave $\boldsymbol{E}$ and $\boldsymbol{B}$ unaffected by differentiating the potential, this invariance leads to charge conservation.

The interchange of identical particles in a system is a kind of symmetry operation which keeps the character of the wave function of the system invariant. In Bose-Einstein statistics, wave function is symmetric under such interchange of identical particles (Y(2, 1) = Y(1, 2)). While in Fermi-Dirac statistics, the wave function is antisymmetric under such interchange of particles (Y(2, 1) = −Y(1, 2)). Conservation of statistics, i.e., invariance of symmetric and antisymmetric behaviour of wave function, implies that no process occurring within an isolated system can change the statistical behaviour of that system. A system obeying Bose-Einstein statistics cannot spontaneously exhibit Fermi-Dirac statistics or vice versa. This conservation principle has applications in nuclear physics where it is observed that nuclei containing even number of nucleons (even mass number A) obey Bose-Einstein statistics while those with odd mass number A obey Fermi-Dirac statistics. Conservation of statistics is thus an additional condition a nuclear reaction must observe.

Those conservation laws which are obeyed in all interactions are called *absolute conservation laws,* e.g., conservation laws for momentum, energy, angular momentum, electric charge, statistics, baryon number (baryons are fermions consisting of 3 quarks, e.g., proton, neutron, etc.) lepton number, etc.

There are quantities which are conserved in some but not in all interactions, e.g., isospin and parity. These are useful in classifying elementary particles and their interaction. *Isospin* is a quantity used to describe charge independence of the strong interactions. It is conserved in strong interactions which are charge independent but not in electromagnetic and weak interactions. *Parity, which* describes the comparative behaviour of the two systems that are mirror images of each other, is conserved in strong and electromagnetic interactions but not in weak interaction. T.D. Lee and C.N. Yang received Noble Prize in 1957 for laying the theoretical foundation of non-conservation of parity in weak interactions.

So ultimately, the conservation laws and symmetry are related in a way that when a conservation law is violated in an interaction, the interaction is referred to as a symmetry breaking interaction.

8.9 oscillatory motion

A motion that repeats itself after a fixed interval of time, called the *time period,* is referred to as a *periodic motion* If the periodic motion is such that it traces the same path back and forth about a mean position, then the periodic motion is called an *oscillatory* (or *vibratory) motion* or an oscillation (or vibration). For example, (Fig. 8.7), a mass attached by a spring to a fixed point sliding to and fro on a frictionless surface, is an example of an oscillatory and a periodic motion. In an oscillatory motion, the displacement of the body on either side of its mean position remains confined within a well-defined limit.

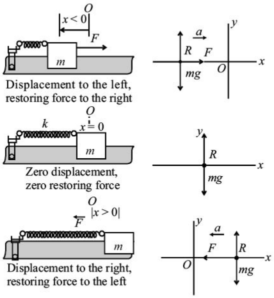


Fig. 8.7

Whenever the body is displaced from its equilibrium position O, a restoring force is developed in the spring, which tends to restore the body. Oscillations occur only if a restoring force develops, which tends to restore the body to its equilibrium position.

A particle undergoing periodic motion has always a stable equilibrium position. When it is moved from its stable equilibrium position, a restoring force (or torque) develops, which takes it back to its equilibrium position. However, by the time, the body reaches its equilibrium position, it acquires some kinetic energy due to which it overshoots, stopping somewhere on the other side and again the restoring force develops which pulls it back towards the equilibrium position. In this way, it keeps moving back and forth.

All oscillatory motions are also periodic motions, but all periodic motions need not be oscillatory. For example, the motion of earth around the sun is periodic, but not oscillatory. Whereas the motion of a simple pendulum is both periodic and oscillatory.

Some of the common terms associated with oscillatory and periodic motion are defined below.

Amplitude (A): The amplitude is the maximum magnitude of displacement from the equilibrium position. It is denoted by A. When the body moves from equilibrium O to A, from A to $-A$ and comes back to equilibrium position O, one cycle is said to have been completed. The SI unit of A is meter.

Time Period (T): The time taken by the body to complete one cycle is called its *time period*. It is denoted by T and its SI unit is second, but sometimes, it is also expressed as second per cycle.

Frequency (f): The number of cycles completed in one second is called the Frequency *of the oscillating body*. It is denoted by F (or n). Its SI unit is Hertz (Hz) named after German physicist Heinrich Hertz. Sometimes instead of Hertz other unit *cycle per second* (cps) is also used. The Frequency is the reciprocal of the time period:

$$f = \frac{1}{T}$$

Angular Frequency (ω): It is the angle subtended per unit time. Its unit is radian per second:

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

8.10 simple harmonic motion

The simplest kind of oscillatory motion occurs, as discussed in the previous section, when the restoring force developed is directly proportional to the displacement from equilibrium. This is possible if the spring used is ideal and obeys Hooke's law. If k is the force constant of the spring used, then the restoring froce developed may be represented as:

$$F = -kx \quad ...(8.19)$$

The force constant k is always positive and the above expression gives correct magnitude and sign of the force, whether x is positive or negative or zero. It has been assumed that there is no frictional force on the oscillating mass.

As represented by Eq. (8.19), when the restoring force is directly proportional to the displacement from the equilibrium position and oppositely directed (towards equilibrium position), the oscillatory motion is called *simple harmonic motion*. The maximum displacement or amplitude of the body on either side of mean position is equal. A body undergoing simple harmonic motion is called a *harmonic oscillator*. There are periodic motions in which restoring force is not proportional to displacement and they are not simple harmonic motions.

Using Newton's second law of motion, $F = \mu a$ and Eq. (8.19), we have:

$$a = -\frac{k}{m}x \quad ...(8.20)$$

The −ve sign signifies that the acceleration and displacement always have an opposite sign. Also, we can express Eq. (8.20) as:

$$a = \frac{d^2x}{dt^2} = -\frac{k}{m}x$$

$$\text{or,} \quad \frac{d^2x}{dt^2} + \frac{k}{m}x = 0 \quad (8.21)$$

This equation is the differential equation of motion of a simple harmonic oscillator and by solving this equation, we can find the dependence of the displacement upon time.

It follows from Eq. (8.20) that the acceleration is directly proportional to the displacement x and always has the opposite sign. This is the *basic characteristic* of simple harmonic motion.

Simple Harmonic Motion along a Circle

Any expression representing a simple harmonic motion, must express the displacement x of the body as a function of time x (t) such that the acceleration $\frac{d^2x}{dt^2}$ is equal to $-\frac{k}{m}$ times, the displacement itself. It is found that the simple harmonic motion is the projection of uniform cirular motion onto a diameter. As shown in Fig. 8.8, suppose a body at the point q at any instant is moving in a circle with constant angular speed ω. A circle in which the body is moving so that its projection is equivalent to the motion of a oscillating body, is called the *circle of reference*, while the point q is called the *reference point*. The point O being the centre of the circle, the vector Oq also rotates with angular velocity ω. Such a rotating vector is called a *phasor*. The x-component of the phasor at time t or the x-component of the point q is given by:

$$x = A\cos\theta = A\cos\omega t \quad ...(8.22)$$

where $A = |OQ|$ and $\theta = \omega t$. The point P, also called the *shadow of the point Q*, is the projection of the point Q on the x-axis. Since the point Q is in uniform circular motion, its acceleration is always directed towards the point O.

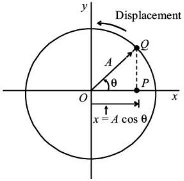


Fig. 8.8

$$\therefore \text{Velocity } \frac{dx}{dt} = -\omega A\sin\omega t \quad ...(8.23)$$

$$\text{and, acceleration } \frac{d^2x}{dt^2} = -\omega^2 x \quad ...(8.24)$$

Thus, the acceleration is directly proportional to the displacement x and always has −ve sign, the essential condition for a motion to be simple harmonic.

Comparing Eqs. (8.20) and (8.24), the expression for angular Frequency ω is given by

$$w^2 = \frac{k}{m} \Rightarrow \omega = \sqrt{\frac{k}{m}} \quad ...(8.25)$$

$$\therefore \text{Frequency, } f = \frac{\omega}{2\pi} = \frac{1}{2\pi}\sqrt{\frac{k}{m}} \quad ...(8.26)$$

$$\text{Time period, } T = \frac{1}{f} = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{m}{k}} \quad ...(8.27)$$

It is obvious from Eqs. (8.25) and (8.26) that if a body has large μ, its Frequency of oscillation will be low and its time will be high. This is important to note that *f and T of a body are determined by its mass and force constant k and are independent of the amplitude of oscillation.* This independence of time period on the amplitude may seem to be puzzling. However, this can be easily understood as follows. When the amplitude is large, the body has high $|x|$ and consequently a large restoring force. Due to large restoring force, the average speed of the body is large and it completes the oscillation quickly, which compensates for the body having to travel a larger distance. Thus, the body covers a larger distance in the same time interval.

Displacement, Velocity and Acceleration is SHM

As shown in Fig. 8.9, suppose that at $t = 0$, the particle is at θ so that $x = A\cos\omega t$. At any other time t, the phasor makes an angle q = ωt + f, so that,

$$x = A\cos(\omega t + f) \quad ...(8.28)$$

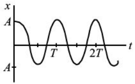


Fig. 8.9

The constant φ is called the *initial phase angle* and indicates that at $t = 0$, where the particle was located. The ωt + φ is the phase angle at any instant t. The displacement x with time t of a particle undergoing SHM is depicted in Fig. 8.9. It becomes clear that in SHM the motion is *periodic* and *sinusoidal* function of time. Simple harmonic motion represented by Eq. (8.21) can also be expressed using *sine* or any other function using appropriate trigonometric identity.

It is obvious from Fig. 8.9 as well as Eq. (8.28) that the displacement of the particle varies from A to $-A$ with time and the cycle repeats itself after 2π radian or 360°.

At $t = 0$, the position x is denoted by x_0.

$$\therefore x_0 = A\cos\varphi$$

when, $\varphi = 0$, $x_0 = A\cos\theta = A$, i.e., body starts at maximum positive displacement

$\varphi = \pi/2$, $x = A\cos\pi/2 = 0$, i.e., body starts from equilibrium position

$\varphi = \pi$, $x = A\cos\pi = -A$, i.e., body starts at maximum negative displacement

Using Eq. (8.28), the velocity of the body undergoing simple harmonic motion is,

$$v = \frac{dx}{dt} = -\omega A \sin(\omega t + \phi) \quad ...(8.29)$$

As the value of sin(ωt + φ) is between −1 to +1, the velocity oscillates between +ωA and −ωA, i.e., velocity amplitude $|v| = \pm\omega A$.

Differentiating Eq. (8.11) again, we have,

$$a = \frac{d^2x}{dt^2} = -\omega^2 A\,(\cos\omega t + \omega) \quad ...(8.30)$$

Thus, the acceleration oscillates between $+\omega^2 A$ to $-\omega^2 A$, i.e., acceleration amplitude $|a| = \pm\omega^2 A$.

It is clear from the above that when a body in simple harmonic motion passes through the *Equilibrium position, the velocity is maximum and the acceleration is zero.* At either end of the motion, the *velocity is zero and the body is instantaneously at rest. At these points, the restoring force and acceleration have their maximum magnitudes.*

8.11 energy of a harmonic oscillator

A body executing simple harmonic motion, possesses both kinetic energy and potential energy. We know that the acceleration of a body executing simple harmonic motion is always directed towards the equilibrium position, i.e., opposite to the direction in which its displacement increases. Work is, therefore, done during the displacement of the body and the body, therefore, possesses potential energy. Also, the body has velocity and, therefore, possesses kinetic energy.

If there is no dissipative force (e.g., friction, air resistance, etc.) or the force is conservative, the *sum of the kinetic energy and potential energy of the body remains constant, i.e., the mechanical energy of the oscillator remains conserved.* Consider a body of mass μ executing simple harmonic motion connected by an ideal spring, its kinetic energy is $K = \frac{1}{2}mv^2$ and its potential energy is $U = \frac{1}{2}kx^2$.

$$\therefore \text{ Total energy, } E = \frac{1}{2}mv^2 + \frac{1}{2}kx^2 = \text{constant} \quad ...(8.31)$$

At the extreme ends, i.e., when the amplitude is maximum, the body comes to an instantaneous rest, the kinetic energy becomes zero and whole of its energy is potential. At the mean position, the velocity is maximum and, therefore, its kinetic energy is maximum and the potential energy is zero. In between these two extreme positions, the energy of the oscillator is partly kinetic and partly potential. This can also be seen mathematically as given below:

The acceleration of the particle $a = \frac{d^2x}{dt^2} = -\omega^2 x.$

∴ Force required to cause a displacement x, $F = \mu\omega^2 x$.

∴ Work done in causing this displacement, $dw = \mu\omega^2 x\, dx$.

Potential energy at displacement x, $U = \int_0^x m\omega^2 x\, dx.$

$$= \frac{1}{2}m\left(\frac{k}{m}\right)x^2 = \frac{1}{2}kx^2 \quad ...(8.32)$$

We see that the potential energy of the oscillator is directly proportional to the square of the displacement.

At maximum displacement,

$$x = A, \quad U = \frac{1}{2}kA^2$$

As $x = \cos(\omega t + \varphi)$.

Potential energy

$$U = \frac{1}{2}k[A\cos(\omega t + \phi)]^2 = \frac{1}{2}kA^2\cos^2(\omega t + \phi) \quad ...(8.33)$$

Velocity of the particle,

$$v = \frac{dx}{dt} = \frac{d}{dt}[A\cos(\omega t + \phi)] = -\omega A\sin(\omega t + \phi)$$

∴ Kinetic energy,

$$K = \frac{1}{2}mv^2 = \frac{1}{2}m[-\omega A\sin(\omega t + \phi)]^2$$

$$= \frac{1}{2}m\omega^2 A^2\sin^2(\omega t + \phi)$$

$$= \frac{1}{2}kA^2\sin^2(\omega t + \phi) \quad ...(8.34)$$

$$\text{Total energy, } E = \frac{1}{2}mv^2 + \frac{1}{2}kx^2$$

$$= \frac{1}{2}m\left[-\omega A\sin(\omega t + \phi)^2 + \frac{1}{2}k\,[A\cos(\omega t + \phi)^2]\right]$$

$$= \frac{1}{2}kA^2\sin^2(\omega t + \phi) + \frac{1}{2}kA^2\cos^2(\omega t + \phi)$$

$$= \frac{1}{2}kA^2[\sin^2(\omega t+\phi)+\cos^2(\omega t+\phi)] = \frac{1}{2}kA^2 \quad ...(8.35)$$

Thus, the total energy is independent of the displacement x and remains constant throughout the motion. Since total energy is conserved, one form of energy can increase only at the expense of the other and kinetic energy is maximum when potential energy is zero and vice versa. This also implies that the maximum value of any one form of energy must be equal to the total energy of the oscillator. The variation of energy with displacement is shown in Fig. 8.10. The potential energy curve is parabolic in shape with its vertex at x = 0. The kinetic energy is also inverted parabolic with vertex of the parabola touching the upper horizontal line or total energy curve at x = 0. The total energy is represented by the upper horizontal line at their maximum values. Thus, we conclude that the *total energy remains constant throughout and is independent of both displacement and time.*

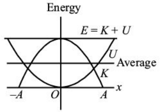


Fig. 8.10

Average Values of Kinetic and Potential Energies of a Harmonic Oscillator: A careful observation of Fig. 8.10 makes it clear that the average value of kinetic energy and potential energy is the same and is equal to $\frac{1}{4}kA^2 = \frac{1}{4}m\omega^2A^2.$ We can also obtain this result mathematically as follows:

Potential energy of particle at displacement x,

$$U = \frac{1}{2}m\omega^2x^2 = \frac{1}{2}m\omega^2A^2\cos^2(\omega t+\phi)$$

Average value of potential energy,

$$U_{av} = \frac{1}{T}\int_0^T \frac{1}{2}m\omega^2A^2\cos^2(\omega t+\phi)$$

$$= \frac{m\omega^2A^2}{4T}\int_0^T[1+\cos 2(\omega t+\phi)]\,dt$$

As the average of a *cosine or a sine* function over a complete cycle is zero,

$$U_{av} = \frac{m\omega^2A^2}{4T}\,|\,t\,|_0^T = \frac{m\omega^2A^2}{4T}T = \frac{1}{4}m\omega^2A^2 \quad ...(8.36)$$

Kinetic energy at displacement x,

$$K = \frac{1}{2}m\left(\frac{dx}{dt}\right)^2$$

$$= \frac{1}{2}m\left[\frac{d}{dt}[A]\cos(\omega t+\phi)\right]^2 = \frac{m\omega^2A^2}{2}\sin^2(\omega t+\phi)$$

Average kinetic energy over a cycle,

$$K_{av} = \frac{1}{T}\int_0^T\left(\frac{m\omega^2A^2}{2}\right)\sin^2+(\omega t+\phi)dt$$

$$= \frac{m\omega^2A^2}{4T}\int_0^T 2\sin^2(\omega t+\phi)dt$$

$$= \frac{m\omega^2A^2}{4T}\int_0^T[1-\cos 2(\omega t+\phi)]dt$$

$$= \frac{m\omega^2A^2}{4T}\,|\,t\,|_0^T = \frac{m\omega^2A^2}{4T}T = \frac{1}{4}m\omega^2A^2 \quad ...(8.37)$$

$$\left(\therefore \frac{1}{T}\int_0^T\cos 2(\omega t+\phi)\,dt = 0\right)$$

Thus, we find that the average value of kinetic and potential energies is the same and is equal to $\frac{1}{4}kA^2$.

8.12 examples of simple harmonic motion

1. Simple Pendulum: A simple pendulum consists of a point mass suspended from one end of an unstretchable and massless string whose other end is fixed at a rigid support, this fixed point is referred to as the *point of suspension*. Such a pendulum is an idealized simple pendulum. A simple pendulum is shown in Fig. 8.11.

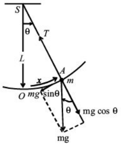


Fig. 8.11 A simple pendulum.

Let S be the point of suspension of the pendulum and O the equilibrium or mean position of the mass. Let λ be the length of the pendulum which is the distance from the point of suspension to the centre of mass of the bob. On displacing the bob a little from its equilibrium position and then gently releasing it, the pendulum starts oscillating about its mean position. The path of the bob is not a straight line, but the arc of a circle of radius λ is equal to the length of the string.

Let at any instant t, the bob of the pendulum, be at A, at an angle q with respect to the equilibrium position. The weight mg of the bob can be resolved into mg cos q and mg sin q. The component mg cos q is cancelled by the tension T. The restoring force F is the tangential component of the net force,

$F = -mg\sin\theta$ ...(8.38)

which tends to bring back the bob to its equilibrium position. The –ve sign indicates that the restoring force is oppositely directed to the angular displacement θ. It is important to note that the restoring force is proportional to sin θ and not θ and so the motion is not simple harmonic. However, if displacement θ is small, then sin θ = θ. $\left(\sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!}\ldots,\right.$ for low θ, the term θ^3 and higher order terms can be neglected). Then Eq. (8.38) can be approximated to:

$$F = -\mu g q$$

$$= -mg\frac{x}{L} \quad (x = \lambda q)$$

$$\text{or, } F = -\frac{mg}{L}x \quad \text{...(8.39)}$$

Thus, the restoring force is proportional to x (of course, for small x) and directed towards the equilibrium position. Equation (8.38) can also be put in terms of moment of inertia. If $-\frac{d^2\theta}{dt^2}$ is the acceleration of the bob towards the mean position and I its moment of inertia about the point suspension S, and $-\mu g\lambda\sin\theta$ the torque about the point of suspension, then:

$$I\frac{d^2\theta}{dt^2} = -mgL\sin\theta$$

For small θ, sinθ = θ

$$\therefore I\frac{d^2\theta}{dt^2} = -mgL\theta \Rightarrow \frac{d^2\theta}{dt^2} = -\frac{mgL}{I}\theta$$

Since moment of inertia of the bob about the point of suspension is mL^2, therefore,

$$\frac{d^2\theta}{dt^2} = -\frac{mgL}{mL^2}\theta = -\frac{g}{L}\theta = -\mu\theta \quad \text{...(8.40)}$$

The quantity $\mu = \frac{g}{L}$ is the acceleration per unit displacement. From Eq. (8.40) it is evident that the acceleration is proportional to angular displacement q and is directed towards mean position. Thus, the pendulum executes simple harmonic motion and its time period is given by:

$$T = 2\pi\sqrt{\frac{1}{\mu}} = 2\pi\sqrt{\frac{L}{g}} \quad \text{...(8.41)}$$

$$\therefore \text{Frequency, } f = \frac{1}{T} = \frac{1}{2\pi}\sqrt{\frac{8}{g}} \quad \text{...(8.42)}$$

The displacement in this case being angular, it is obvious that this is an example of angular simple harmonic motion. The time period is dependent on only the length of the pendulum and is proportional to square root of the length.

It is important to remember that the above discussion is based on the assumption that the displacement is very small. But in practice, it is impossible to have a point mass and a weightless string and the resistance buoyancy of the air appreciably affect the motion of the pendulum. The motion of bob is not strictly linear, but it also has a rotatory motion about the axis of the suspension. So, the motion of a pendulum is only approximately simple harmonic. Due to these practical aspects, the value of g calculated using a simple pendulum is not very accurate. For this reason, a physical (or compound) pendulum, discussed below, is a better option for the calculation of g.

2. Physical (Compound) Pendulum: A physical or a compound pendulum is any rigid pendulum of any shape and size capable of oscillating about a horizontal axis passing through it, unlike a simple pendulum in which all mass is supposed to be concentrated at a single point.

Figure 8.12 shows a physical pendulum capable of oscillating about S, the point of suspension in the equilibrium position of rest, the centre of gravity G lies vertically below S. The distance between point (or centre) of the suspension and the centre of gravity of the pendulum measures the length I of the pendulum.

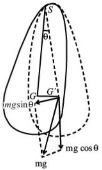


Fig. 8.12

Suppose the pendulum is displaced from its equilibrium position by a small angle θ, so that its new centre of gravity is G', *clearly SG'* = *I*. The weight *mg* of the pendulum acting at G' acts vertically downward and its reaction at the point of suspension causes a torque, which tends to restore the pendulum to its equilibrium position. The torque is given by:

$$\tau = -(mg)(I\sin\theta) \quad ...(8.43)$$

The −ve sign shows that the restoring torque is clockwise and the displacement is anticlockwise and vice versa. When θ is small, sinθ = θ. If *I* be the moment of inertia of the pendulum about the axis of suspension and $\frac{d^2\theta}{dt^2}$ its angular acceleration then the torque $\tau = I\frac{d^2\theta}{dt^2}$. Therefore, from Eq. (8.43), we have:

$$I\frac{d^2\theta}{dt^2} = mgI\theta$$

$$\text{or, } \frac{d^2\theta}{dt^2} = -\left(\frac{mgI}{I}\right)\theta = -\mu\theta \quad ...(8.44)$$

The quantity $\mu = -\frac{mgI}{I}$, acceleration per unit displacement, is what $\frac{k}{m}$ is in the spring mass system.

$$\therefore w = \sqrt{\frac{mgI}{I}} \Rightarrow \frac{2\pi}{T} = \frac{mgI}{I}$$

$$\therefore \text{Time period, } T = 2\pi\sqrt{\frac{I}{mgI}} \quad ...(8.45)$$

$$\text{Frequency, } f = \frac{1}{T} = \frac{1}{2\pi}\sqrt{\frac{mgI}{I}} \quad ...(8.46)$$

Equation (8.45) is commonly used to experimentally determine the moment of inertia of a body of an irregular and complicated shape. For this, the centre of gravity is located by balancing it and then the body is suspended. By measuring *T* and λ and knowing *g* at the place, *I* can be calculated. If I_0 (= mk^2, *k* the radius of gyration) be the moment of inertia of the pendulum about an axis through its centre of gravity *G*, parallel to the axis through *S*, then from the theorem of parallel axes, $I = I_0 + \mu\lambda^2 = mk^2 + ml^2 = \mu(k^2 + \lambda^2)$. From Eq. (8.45)

$$T = 2\pi\sqrt{\frac{m(k^2+l^2)}{mg\,l}} = 2\pi\sqrt{\frac{k^2+l^2}{g\,l}} = 2\pi\sqrt{\frac{k^2/l+l}{g}} = 2\pi\sqrt{\frac{L}{g}}$$

where $L = \frac{k^2}{l} + l.$ Thus, the time period of the compound pendulum is the same as that of a simple pendulum of length, λ, called the *length of an equivalent simple pendulum* or the *reduced length of the compound pendulum.*

Centre of Oscillation of a Compound Pendulum: A point *O* on the other side of *G* (Fig. 8.13) of the pendulum in the line *SG* and a distance $\frac{k^2}{l}$ from *G* is called the *centre of oscillation* of the pendulum and a horizontal axis passing through it and parallel to the horizontal axis through *S* is called the axis of oscillation of the pendulum. Clearly, $L = \frac{k^2}{l} + l$ is the *length of the equivalent simple pendulum.* It can be seen that the centres of suspension and oscillation are interchangeable as in either case, the time period of the pendulum is the same. Also there are two other points on either side of *G* about which the time period of the pendulum is the same as about *S* and *O*. They are at distance k^2/l and λ on either sides of *G*.

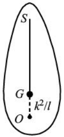


Fig. 8.13

Example 8.9

(*a*) Find the time period and Frequency of a simple pendulum 1.5 m long at a place where g = 9.8 m/s².

(b) A uniform rod of length 1.5 m pivoted at one end undergoes simple harmonic motion at the same place. Is the time period and Frequency of the rod less than or greater to that of the simple pendulum?

Solution:

(*a*) Time period

$$T = 2\pi\frac{1}{g} = 2\times 3.14\times\sqrt{\frac{1.5\text{ m}}{9.8\text{ m/s}^2}}$$

= 2.45 sec

$$\text{Frequency, } F = \frac{1}{T} = \frac{1}{2.45} = 0.41\text{ Hz}$$

(*b*) Moment of Inertia of a rod:

$$= \frac{1}{3}ML^2$$

Distance from the pivot to the centre of gravity:

$$= \frac{L}{2}$$

Time period, T =

$$2\pi\sqrt{\frac{I}{mgI}} = 2\pi\sqrt{\frac{\frac{1}{3}mL^2}{mgL/2}} = 2\pi\sqrt{\frac{2L}{3g}}$$

$$= 2\times 3.14\times\frac{2\times 1.5}{3\times 9.8} = 2.0\text{ sec}$$

$$\text{Frequency, } f = \frac{1}{T} = \frac{1}{2.0} = 0.5\text{ Hz}$$

Thus, while the time period of the rod is less, its Frequency is higher.

8.13 damped oscillations

Till now we have discussed idealized frictionless systems in which there was no non-conservative force. In such systems, the total energy remains constant and a system set into motion continues oscillating forever with no decrease in amplitude. However, all practical systems always encounter some resistive or viscous forces which cause dissipation of energy and amplitude of the oscillations keeps on decreasing and finally the oscillations die out.

The decrease in amplitude caused by dissipative forces is called *damping* and this decreasing amplitude oscillations are called *damped oscillations*.

The *simplest example is a simple harmonic oscillator with a frictional damping force* which is directly proportional to the velocity of the oscillating body.

Consider a body mass m attached to a spring of force constant k. Let at any instant, the displacement of the mass m be x and its instantaneous velocity is $\frac{dx}{dt}$.

The forces acting on the body are:

(*i*) A restoring force, $-kx$

(*ii*) A damping force (due to friction) proportional to the velocity, directed oppositely to the direction of motion, $-b\frac{dx}{dt}$, where b is a positive constant.

∴ Total instantaneous force acting on the body,

$$F = -kx - b\frac{dx}{dt}$$

According to Newton's second law of motion,

$$F = ma = m\frac{d^2x}{dt^2}$$

$$\therefore \frac{d^2x}{dt^2} + \frac{b}{m}\frac{dx}{dt} + \frac{k}{m}x = 0 \quad ...(8.47)$$

Putting $\frac{b}{m} = 2r$ (r is the damping constant) and $\frac{k}{m} = \omega_0^2$ (ω_0 is the natural Frequency of the oscillator), we get

$$\frac{d^2x}{dt^2} + 2r\frac{dx}{dt} + \omega_0^2 x = 0 \quad ...(8.48)$$

This is the differential equation of a damped harmonic oscillator.

To solve this homogeneous linear, second order differential equation, we try:

$$x = Ae^{\alpha t} \quad ...(8.49)$$

where A and α are arbitrary constants. Differentiating Eq. (8.49) with respect to t, we get

$$\frac{dx}{dt} = A\alpha e^{\alpha t} \text{ and } \frac{d^2x}{dt^2} = A\alpha^2 e^{\alpha t}$$

Substituting these in Eq. (8.48), we get

$$A\alpha^2 E^{\alpha t} + 2rA\alpha E^{\alpha t} + \omega_0^2 A e^{\alpha t} = 0$$

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

or $$\alpha^2 + 2r\alpha + \omega_0^2 = 0$$

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

Thus, α has two values

$$\alpha_1 = -r + \sqrt{r^2 - \omega_0^2}$$

and $$\alpha_2 = -r - \sqrt{r^2 - \omega_0^2}$$

General solution of Eq. (8.48) is the linear combination of the two solutions obtained from these two values of α.

Thus, $$x = A_1 e^{\left[-r+\sqrt{(r^2-\omega_0^2)}\right]t} + A_2 e^{\left[-r-\sqrt{(r^2-\omega_0^2)}\right]t} \quad ...(8.50)$$

where A_1 and A_2 are two arbitrary constants. The actual form of the above equation depends on the relative magnitude of r^2 and ω_0^2, as discussed below:

Case I: When $r^2 > \omega_0^2$ (heavy damping). In such cases $\sqrt{r^2 - \omega_0^2}$ is real and less than r. Hence, both the exponential terms $\left[-r + \sqrt{(r^2 - \omega_0^2)}\right]$ and $\left[-r - \sqrt{(r^2 - \omega_0^2)}\right]$ are negative. This implies that the displacement x of the particle continuously decreases with time. The displacement after attaining its amplitude dies out exponentially with time without changing direction. Thus, there is no oscillation and the motion is said to be *heavily damped* or *overdamped* or *aperiodic* or *dead beat*. Examples of this type of motion are a pendulum oscillating in a viscous fluid-like thick oil and by a moving-coil galvanometer shunted by a very low resistance. This is represented by curve (*i*) in Fig. 8.14.

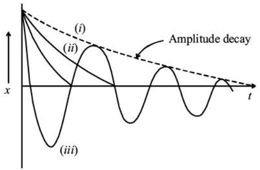


Fig. 8.14

Case II: When $r^2 = \omega_0^2$ (critical damping case). If we put $r^2 = \omega_0^2$ Eq. (8.50), each of the two terms become infinite and the solution breaks down. Hence, let us consider $\sqrt{r^2 = \omega_0^2}$ is not zero but *a very small quality* β. Then, we have

$$x = A_1 E^{(-r+b)t} + A_2 E^{(-r-b)t}$$

$$= e^{-rt}(A_1 E^{bt} + A_2 E^{-bt})$$

$$= e^{-rt}\left\{A_1\left(1 + \beta t + \frac{\beta^2 t^2}{2!}\ldots\right) + A_2\left(1 - \beta t + \frac{\beta^2 t^2}{2} - \ldots\right)\right\}$$

Neglecting the second and higher order terms of β, we have:

$$x = E^{-rt}[(A_1 + A_2) + \beta t(A_1 - A_2)]$$

$$= E^{-rt}(A + Bt) \quad ...(8.51)$$

where $A = A_1 + A_2$ and $B = \beta(A_1 - A_2)$.

The curve (*ii*) representing the above equation is shown in Fig 8.14. In this case, the displacement of the oscillator increases and then returns back quickly to its equilibrium position. The motion of the oscillator thus becomes *non-periodic* or *non-oscillatory*, i.e., it ceases to oscillate. This is called the case of *critical damping* and the motion is said to be *critically damped, the* condition for which $r \to \omega_0$. This knowledge is used in the design of many pointer type instruments like galvanometer in which the pointer moves to the correct position and returns back to zero position in a very short time.

Case III: $r^2 < \omega_0^2$ (low or underdamping case). This is the actual case or an underdamped harmonic oscillator. In this case, $\sqrt{r^2 < \omega_0^2}$ becomes imaginary. Let $\sqrt{r^2 - \omega_0^2} = j\sqrt{\omega_0^2 - r^2} = j\omega$, where $j = \sqrt{-1}$ $\omega = \sqrt{\omega_0^2 - r^2}$. Now Eq. (8.50) becomes:

$$x = A_1 E^{(-r+j\omega)t} + A_2 E^{(-r-j\omega)t}$$

$$= e^{-rt}(A_1E^{j\omega t} + A_2E^{-j\omega t})$$
$$= E^{-rt}\{A_1(\cos\omega t + j\sin\omega t) + A_2(\cos\omega t - j\sin\omega t)\}$$
$$= E^{-rt}\{(A_1 + A_2)\cos\omega t + j(A_1 - A_2)\sin\omega t)\}$$

Let, $A_1 + A_2 = a\sin\varphi$ and $j(A_1 - A_2) = a\cos\varphi$, where a and φ are two constants.

$$\therefore x = E^{-rt}(a\sin\varphi\cos\omega t + a\cos\varphi\sin\omega t)$$
$$= ae^{-rt}\sin(\omega t + \varphi) \quad \text{...(8.52)}$$

This is the *Equation of a damped harmonic oscillator*, with the following parameters:

Amplitude: The amplitude of the oscillation is ae^{-rt} which obviously is not constant but decays with time to zero, in accordance with the term E^{-rt}, called the *damping factor*. This is represented by the curve (*iii*) in Fig. 8.14. Although, the amplitude decreases exponentially with time, the under-damped harmonic oscillator does perform a sort of oscillatory motion. Of course, the motion is not repeated as such and thus the motion is not periodic terms of the usual sense of periodic motion.

The time interval in which the amplitude of the oscillation falls to $\frac{1}{e}$ of its initial is called the *mean lifetime* (τ_m) of the oscillator.

Thus, $ae^{-rt_m} = \frac{1}{e}a \Rightarrow e^{-rt_m} = e^{-1}$ taking $\log_e$ of the both sides.

$$-rt_m = -1 \Rightarrow \tau_m = \frac{1}{r}$$

Thus, the mean lifetime is equal to $\frac{1}{r}$ or $\frac{2m}{b}$. In the absence of any damping ($r = 0$), the mean lifetime would be infinite, i.e., the amplitude would have remained constant.

Time Period: From Eq. (8.51), it is clear that if we replace t by $t + \frac{2\pi}{\omega}$, x remains unchanged, i.e., the time period of the damped oscillations is $\frac{2\pi}{\omega}$, or,

$$T = \frac{2\pi}{\omega} = \frac{2\pi}{\sqrt{\omega_0^2 - r^2}} \quad \text{...(8.53)}$$

Obviously, this is little higher than $\frac{2\pi}{\omega_0}$, the time period of the undamped oscillation.

$$\therefore \text{Frequency, } F = \frac{1}{T} = \frac{1}{2\pi}\sqrt{\omega_0^2 - r^2}$$

where, $\omega_0 = \sqrt{\frac{k}{m}}$ and $r = \sqrt{\frac{b}{2m}}$

$$\therefore F = \frac{1}{2\pi}\sqrt{\frac{k}{m} - \frac{b^2}{4m^2}} \quad \text{...(8.54)}$$

It is clear that while the time period of the oscillation increases a little, the Frequency decreases a little in case of damped harmonic oscillator compared to that in undamped oscillator.

Logarithmic Decrement: The amplitude in the case of a damped harmonic osculatory goes on decreasing continuously. We have:

$$x = ae^{-rt}\sin(\omega t + \varphi) \quad \text{...(8.55)}$$

If, $\varphi = \frac{\pi}{2}$, $x = ae^{-rt}\cos\omega t$

At $t = 0$, let $x = a_0$. Let a_1, a_2, a_3, ... be the amplitudes at time T, $2T$, $3T$, ..., where T is the time period of oscillation then,

$$a_1 = a_0E^{-rT},\ a_2 = a_0E^{-2rT}$$
$$a_3 = a_0E{-}3rT, \ldots$$

....

We thus have:

$$\frac{a_0}{a_1} = \frac{a_1}{a_2} = \frac{a_2}{a_3} = \ldots a^{rt} = e^{\lambda_d}$$

where $\lambda_d\left(= rT = \frac{bT}{2m}\right)$ is the *logarithmic decrement*. Taking $\log_e$ of the above expression:

$$\lambda_d = \log_e\frac{a_0}{a_1} = \log_e\frac{a_1}{a_2} = \log_e = \frac{a_2}{a_3} = \ldots$$

Thus, *we define logarithmic decrement as the logarithm of the ratio of two successive amplitudes* (separated by time period) of the damped oscillations.

Power Dissipation: The amplitude of damped harmonic oscillator keeps on decreasing exponentially with time due to resistive force such as air friction. In other words, the energy of the oscillator is continuously dissipated. We obtain below an expression for the power dissipation of a damped harmonic oscillator. The displacement x of the oscillator at any time t is given by:

$$x = ae^{-rt}\sin(\omega_0 t + \varphi)$$

For weak damping, $r^2 << \omega_0^2$, i.e., $\frac{b^2}{4m^2} << \frac{k}{m}$. For convenience, let $\varphi = 0$. so that,

$$x = ae^{-rt}\sin\omega_0 t$$

where, $\omega_0 = \sqrt{\frac{k}{m}}$ and $r = \sqrt{\frac{b}{2m}}$

$$\therefore F = \frac{1}{2\pi}\sqrt{\frac{k}{m} - \frac{b^2}{4m^2}} \quad ...(8.54)$$

It is clear that while the time period of the oscillation increases a little, the Frequency decreases a little in case of damped harmonic oscillator compared to that in undamped oscillator.

Logarithmic Decrement: The amplitude in the case of a damped harmonic osculatory goes on decreasing continuously. We have:

$$x = ae^{-rt}\sin(\omega t + \varphi) \quad ...(8.55)$$

If, $\varphi = \frac{\pi}{2}$, $x = ae^{-rt}\cos\omega t$

At $t = 0$, let $x = a_0$. Let $a_1, a_2, a_3, \ldots$ be the amplitudes at time $T, 2T, 3T, \ldots$, where T is the time period of oscillation then,

$$a_1 = a_0E^{-rT},\ a_2 = a_0E^{-2rT}$$

$$a_3 = a_0E{-}3rT, \ldots$$

....

We thus have:

$$\frac{a_0}{a_1} = \frac{a_1}{a_2} = \frac{a_2}{a_3} = \ldots a^{rt} = e^{\lambda_d}$$

where $\lambda_d \left(= rT = \frac{bT}{2m} \right)$ is the *logarithmic decrement*. Taking, $\log_e$ of the above expression:

$$\lambda_d = \log_e \frac{a_0}{a_1} = \log_e \frac{a_1}{a_2} = \log_e = \frac{a_2}{a_3} = \ldots$$

Thus, *we define logarithmic decrement as the logarithm of the ratio of two successive amplitudes* (separated by time period) of the damped oscillations.

Power Dissipation: The amplitude of damped harmonic oscillator keeps on decreasing exponentially with time due to resistive force such as air friction. In other words, the energy of the oscillator is continuously dissipated. We obtain below an expression for the power dissipation of a damped harmonic oscillator. The displacement x of the oscillator at any time t is given by:

$$x = ae^{-rt}\sin(\omega_0 t + \varphi)$$

For weak damping, $r^2 << \omega_0^2$, i.e., $\frac{b^2}{4m^2} << \frac{k}{m}$. For convenience, let $\varphi = 0$. so that,

$$x = ae^{-rt}\sin\omega_0 t$$

$$\therefore \frac{dx}{dt} = ar{\cdot}E^{-rt}\sin\omega_0 t + ae^{-rt}\omega_0\cos\omega_0 t$$

Kinetic energy $= \frac{1}{2}m\left(\frac{dx}{dt}\right)^2 = \frac{1}{2}ma^2$ $[\omega_0 E^{-rt}\cos\omega_0 t - re^{-rt}\sin\omega_0 t]^2$

$$= \frac{1}{2}ma^2e^{-2rt}[\omega_0^2\cos^2\omega_0 t + r^2\sin^2\omega_0 t - 2r\omega_0\sin\omega_0 t\cos\omega_0 t]$$

For average kinetic energy, we take the average time of the above expression:

$$K_{av} = \frac{\int_0^T \frac{1}{2}ma^2e^{-2rt}[\omega_0^2\cos^2\omega_0 t + r^2\sin^2\omega_0 t - r\omega_0\sin 2\omega_0 t]}{T}$$

$$= \frac{1}{2}ma^2e^{-2rt}\left[\frac{1}{2}\omega_0^2 + \frac{r^2}{2}\right]$$

Since average of $\sin^2\omega_0 t$ and $\cos^2\omega_0 t$ is $\frac{1}{2}$ and that of $\sin^2\omega t$ is zero. Also E^{-2rt} is taken outside the integral assuming that the amplitude of oscillation ae^{-2rt} remains unchanged in one cycle of motion

$$\therefore K_{av} = \frac{1}{4}ma^2e^{-2rt}(\omega_0^2 + r^2) \quad ...(8.56)$$

For low damping, $\omega_0^2 >> r^2$,

$$\therefore K_{av} = \frac{1}{4}ma^2\omega_0^2e^{-2rt} \quad ...(8.57)$$

Potential energy of the oscillator,

$$U = \frac{1}{2}kx^2 = \frac{1}{2}k(ae^{-rt}\sin\omega_0 t)^2$$

$$=$$

$$\frac{1}{2}ka^2e^{-2rt}\sin^2\omega_0 t = \frac{1}{2}m\omega_0^2a^2e^{-2rt}\sin^2\omega_0 t$$

$$\left(\because \frac{k}{m} = \omega_0^2\right)$$

Average U over a cycle,

$$U_{av} = \frac{1}{2}m\omega_0^2a^2e^{-2rt}\cdot\frac{\int_0^T \sin^2\omega_0 t\,dt}{T}$$

$$= \frac{1}{4}m\omega_0^2a^2e^{-2rt} \quad ...(8.58)$$

Total average energy,

$$E = K_{av} + U_{av}$$

$$= \frac{1}{4}m\omega_0^2a^2e^{-2rt} + \frac{1}{4}m\omega_0^2a^2e^{-2rt} = \frac{1}{2}m\omega_0^2a^2e^{-2rt}$$

...(8.59)

At $t = 0$, $E_0 = \frac{1}{2}ma^2\omega_0^2,$ the same as that of undamped oscillator

$\therefore E = E_0E^{-2rt}$ ...(8.60)

This loss of energy is due to the work done against the damping or dissipative force and actually appears in the form of heat.

Relaxation Time: *The time required for the decay of mechanical energy to* $\frac{1}{e}$ *times its initial value is called the relaxation time* (τ) *of the oscillator*. We have:

$$E = E_0E^{-2rt}$$

At $t = \tau$,

$$E = \frac{E_0}{e}$$

$$\therefore \frac{E_0}{e} = E_0e^{-2r\tau} \Rightarrow -1 = -2r\tau \Rightarrow \tau = \frac{1}{2r} \quad ...(8.61)$$

Energy in terms of τ is given by

$$E = E_0E^{-2rt} = E_0E^{-2rt} \quad ...(8.62)$$

$$\text{Power, } P = \frac{2rE = \frac{E}{\tau}.}{} \quad ...(8.63)$$

8.14 Quality Factor

Quality factor is a measure of damping or the rate of energy decay of the oscillator. Lesser the damping, better is the quality of the harmonic oscillator as an oscillator and hence higher is its quality factor. Sometimes, it is also referred to as the *figure of merit* of the harmonic oscillator.

The quality factor of a harmonic oscillator is defined as 2π *times the ratio of the energy stored to the energy lost per period*. That is,

$$Q = 2\pi\frac{\text{energy stored}}{\text{energy lost per period}} = \frac{2\pi E}{PT}$$

where P is average loss of energy over a period. From Eq. (8.63), $P = \frac{E}{\tau}$

$$\therefore Q = \frac{E\omega}{P} = \frac{E\omega}{E/\tau} = \omega\tau$$

For a damped oscillator τ is infinite and, therefore, its Q is infinite.

Example 8.10

A body of mass 5 g is subjected to an elastic force of 40 dynes/cm and a frictional force of 5 dynes/cm^{-1}. If it is displaced through 2 cm and then released, find whether the resulting motion is oscillatory or not? Also, find the time period, if it is oscillatory.

Solution:

The equation of motion can be represented as:

$$\frac{d^2x}{dt^2} + 2r\frac{dx}{dt} + \omega_0^2x = 0$$

$$\text{where } 2r = \frac{b}{m} \text{ and } \omega_0^2 = \frac{k}{m}.$$

The motion is oscillatory if $r < \omega_0$ and the time period is:

$$T = \frac{2\pi}{\sqrt{\omega_0^2 - r^2}}$$

Here, μ = 5 g, k = 40 dynes/cm and b = 5 dynes/cm-sec^{-1}

$$\therefore \text{Damping constant, } r = \frac{b}{2m} = \frac{5}{2\times 5} = 0.50$$

Angular Frequency, w_o =

$$\sqrt{\frac{k}{m}} = \sqrt{\frac{40}{5}} - \sqrt{8} = 2.82$$

Thus, $r < \omega_0$ and hence the motion is oscillatory.

Time period, $T =$

$$\frac{2\pi}{\sqrt{\omega_0^2 - r^2}} = \frac{2\times 3.14}{\sqrt{8-(0.5)^4}} = \frac{2\times 3.14}{2.78} = 2.26 \text{ sec}$$

Example 8.11

A particle of mass 50 g experiences only a damping force proportional to its velocity. If its velocity is decreased from 100 cm/s to 10 cm/s in 23 seconds, calculate (*a*) relaxation time, (*b*) the damping force when its velocity is 50 cm/s. (*c*) the time in which its kinetic energy is reduced to one-tenth of its initial value and (Δ) the total distance travelled of its initial velocity is 10 cm/s (use $\log_e 10 = 2.30$).

Solution:

Let at any instant t, the velocity of the particle be v. Then the instantaneous damping force is $-bv$, where b is constant. If no other force is acting on the particle, then by using Newton's second law of motion, we have,

$$m\frac{dv}{dt} = -bv$$

$$\text{or, } \frac{dv}{dt} = -\frac{b}{m}v = -2rv$$

where $\frac{b}{m} = 2r.$ The above expression can be written as:

$$\frac{dv}{dt} = -2rdt$$

Integrating, $\log_e v = -2rt + C$ (constant of integration)

At $t = 0$, $v = v_0$ then $C = \log_e v_0$

$\therefore \log_e v = -2rt + \log_e v_0$

$$\text{or, } \log_e \frac{v}{v_0} = -2rt$$

$$\text{or, } \frac{v}{v_0} = E^{-2rt}$$

or, $v = v_0 E^{-2rt}$

This is the expression for the (damped) velocity, which decays exponentially with time.

Given: $v = 100$ cm/s decreases to $v = 10$ cm/s in $t = 23$ seconds. Putting these in the above expression we get:

$$10 = 100E^{-2r(23)}$$

$$\text{or, } \frac{1}{10} = E^{-2r(23)} \Rightarrow 10 = E^{2r(23)}$$

Taking $\log_e$ on both sides:

$$2r(23) = \log_e 10 = 2.30$$

$$\text{or, } 2r = \frac{2.3}{23} = \frac{1}{10}$$

Eq. (*i*) therefore becomes:

$$v = v_0 e^{-\frac{t}{10}} \quad ...(ii)$$

(*a*) Let τ be the relaxation time, i.e., the time interval in which velocity reduces to $\frac{1}{e}$ times its initial value then from Eq. (*ii*) we have:

$$\frac{1}{e} = E^{-t/10}$$

or, $E^{-1} = E^{-t/10}$

or, $\tau/10 = 1$

or, $\tau = 10$ sconds.

(*b*) The damping force $F = -bv = -2rmv$

$$= \frac{1}{10}\times 50\times 50 = -250 \text{ dynes}$$

(*c*) Kinetic energy $K = \frac{1}{2}mv^2 = \frac{1}{2}mv_0^2 e^{-t/5}$

At $t = 0$, $K = K_0$, therefore,

$K = K_0 E^{-t/5}$

If t' is the time in which kinetic energy reduces to $\frac{1}{10}$ of its initial value, then:

$$\frac{K}{K_0} = \frac{1}{10} = e^{-t'/5} \Rightarrow e^{-t'/5} = 10$$

Taking $\log_e$ both sides:

$t'/5 = \log_e 10 = 2.30$

$\therefore t' = 5 \times 2.30 = 11.5$ second.

Given: v = 100 cm/s decreases to v = 10 cm/s in t = 23 seconds. Putting these in the above expression we get:

$$10 = 100E^{-2r(23)}$$

$$\text{or, } \frac{1}{10} = E^{-2r(23)} \Rightarrow 10 = E^{2r(23)}$$

Taking $\log_e$ on both sides:

$$2r(23) = \log_e 10 = 2.30$$

$$\text{or, } 2r = \frac{2.3}{23} = \frac{1}{10}$$

Eq. (i) therefore becomes:

$$v = v_0 e^{-\frac{t}{10}} \quad ...(ii)$$

(a) Let τ be the relaxation time, i.e., the time interval in which velocity reduces to $\frac{1}{e}$ times its initial value then from Eq. (ii) we have:

$$\frac{1}{e} = E^{-\tau/10}$$

$$\text{or, } E^{-1} = E^{-\tau/10}$$

$$\text{or, } \tau/10 = 1$$

or, τ = 10 sconds.

(b) The damping force $F = -bv = -2rmv$

$$= \frac{1}{10} \times 50 \times 50 = -250 \text{ dynes}$$

(c) Kinetic energy $K = \frac{1}{2}mv^2 = \frac{1}{2}mv_0^2 e^{-t/5}$

At t = 0, $K = K_0$, therefore,

$$K = K_0 E^{-t/5}$$

If t' is the time in which kinetic energy reduces to $\frac{1}{10}$ of its initial value, then:

$$\frac{K}{K_0} = \frac{1}{10} = e^{-t'/5} \Rightarrow e^{-t'/5} = 10$$

Taking $\log_e$ both sides:

$$t'/5 = \log_e 10 = 2.30$$

$$\therefore t' = 5 \times 2.30 = 11.5 \text{ second.}$$

(Δ) Using $v = \frac{dx}{dt}$ in Eq. (ii) and integrating, we get

$$x = v_0 E^{-t/10}(-10) + C \text{ (constant of integration)}$$

At t = 0, x = 0.

$$\therefore C = 10v_0$$

Putting the value in C, we get:

$$x = -10v_0 E^{-t/10} + 10v_0 = 10v_0(1 - E^{-t/10})$$

At $t \to \infty$, $E^{-t/10} \to 0$ and $x = 10v_0$.

Total distance travelled if its initial speed is 100 cm/s is equal to:

$$x = 10 \times 100 \text{ cm/s} = 1000 \text{ cm} = 10 \text{ m.}$$

Example 8.12

The differential equation of an oscillator is:

$$\frac{d^2x}{dt^2} + 2r\frac{dx}{dt} + \omega_0^2 x = 0$$

If $\omega_0 >> r$, then calculate the time in which (i) amplitude becomes $\left(\frac{1}{e}\right)$ of its initial value, (ii) energy becomes $\left(\frac{1}{e}\right)$ of the initial value and (iii) energy becomes $\frac{1}{e^6}$ of its initial value.

Solution:

The given condition of $\omega_0 >> r$ along with the differential equation indicates that the motion of the oscillator is low or underdamped whose solution is:

$$x = ae^{-rt}\sin(\omega t - \varphi)$$

where a and φ are arbitrary constants and $\omega = \sqrt{\omega_0^2 - r^2}$.

The amplitude of oscillation is ae^{-rt}.

(i) Let a_0 be the amplitude at t = 0 and at the instant t, it becomes $\left(\frac{1}{e}\right)$ of its initial value, i.e., $a(t) = \frac{a_0}{e}$. Then we have:

$$\frac{a_0}{e} = a_0 E^{-rt}$$

$$\text{or, } E^{-1} = E^{-rt} \Rightarrow 1 = rt$$

or, $t = \dfrac{1}{r}$ second

which corresponds to mean lifetime.

(*ii*) The instantaneous energy of damped oscillations is given by:

$$E = E_0 E^{-2rt}$$

when energy becomes $\left(\dfrac{1}{e}\right)$ of the initial value, i.e., $E = \dfrac{E_0}{e}$, then:

$$\frac{E_0}{e} = E_0 E^{-2rt}$$

$$\text{or, } E^{-1} = E^{-2rt} \Rightarrow 2rt = 1$$

$$\text{or, } t = \frac{1}{2r}$$

which is the relaxation time of the oscillator.

(*iii*) When energy becomes $\dfrac{1}{e^6}$ times its initial value, then:

$$\frac{E_0}{e^6} = E_0 E^{-2rt}$$

$$\text{or, } E^{-6} = E^{-2rt} \Rightarrow 2rt = 6$$

$$\text{or, } t = \frac{3}{r} \text{ second.}$$

Example 8.13

The quality factor in a damped oscillator is 5.0×10^{-4}. After what time, its energy will fall to $\dfrac{1}{10}$ of its initial value? The Frequency of oscillator is 300 Hz and use $\log_e 10 = 2.30$.

Solution:

The quality factor of a damped oscillator is q = wt.

$$\therefore \text{ Relaxation time } \tau = \frac{Q}{\omega}$$

Here, $q = 5.0 \times 10^4$ and $\omega = 2\pi\upsilon = 600\,\pi \text{ sec}^{-1}$

$$\therefore \tau = \frac{5.0 \times 10^4}{600\pi}$$

The energy of a damped oscillator is $E = E_0 e^{-2rt} = E_0 e^{-t/\tau}$, where $\tau = \dfrac{1}{2r}$. Let t' be the time in which energy falls to $\dfrac{1}{10}$ of its initial value. Therefore,

$$\frac{E_0}{10} = E_0 E^{-t'/\tau}$$

$$\text{or, } 10 = E^{t'/\tau}$$

$$\text{or, } t' = \tau \log_e 10$$

$$= \frac{5.0 \times 10^4 \times 2.30}{600 \times 3.14} = 61 \text{ seconds.}$$

Example 8.14

The amplitude of a damped harmonic oscillator reduces from 25 cm to 2.5 after 100 complete oscillations each of period 2.3 seconds. Calculate logarithmic decrement of the system (use $\log_e 10 = 2.3$).

Solution:

Here, amplitude ratio of oscillation separated by 100 oscillations is:

$$= \frac{25 \text{ cm}}{2.5 \text{ cm}} = 10$$

$\therefore$ Logarithmic decrement:

$$\lambda_d = \frac{1}{100} \log_e 10 = \frac{2.3}{100} = 0.023.$$

Example 8.15

quality factor of vibrator is 1500. On starting the vibrations, it executes 250 vibrations per second. Calculate the time in which the amplitude decreases to $\dfrac{1}{e^3}$ of the initial value.

Solution:

quality factor,

$$q = wt$$

Here, $Q = 1500$, $\omega = 2\pi\upsilon = 2 \times 3.14 \times 250$

$$= 1570 \text{ sec}^{-1}$$

$$\therefore \text{ Relaxation time, } \tau = \frac{Q}{\omega} = \frac{1500}{1570} = 0.96 \text{ second}$$

If a_0 is the initial amplitude at $t = 0$ and $\frac{a_0}{e^3}$ at the time t, then we have:

$$\frac{a_0}{e^3} = ae^{-rt}$$

or, $E^{-3} = E^{-rt}$

or, $3 = rt$

$$\text{or, } t = \frac{3}{r} = 6\tau = 6 \times 0.96 \text{ seconds}$$

$$\left(\because \tau = \frac{1}{2r}\right)$$

= 5.76 seconds

Example 8.16

Show that in the presence of damping, the Frequency of an oscillator reduces by $12.5\theta^{-2}$%, where Q is its quality factor.

Solution:

Let v_0 and v be the Frequency of the oscillator in the absence and presence of damping respectively. We know that:

$$\omega^2 = \omega_0^2 - \left(\frac{1}{4\tau^2}\right)$$

$$\text{or, } \frac{\omega^2}{\omega_0^2} = 1 - \frac{1}{4(\omega_0\tau)^2}$$

$$= 1 - \frac{1}{4Q^2} \quad (\square\ q = \omega_0\tau)$$

$$\text{or, } \frac{\omega}{\omega_0} = \left(1 - \frac{1}{4Q}\right)^{1/2}$$

$$= 1 - \frac{1}{8Q^2}$$ (Using binomial expansion neglecting higher order terms)

$$\text{or, } \frac{2\pi v}{2\pi v_0} = 1 - \frac{1}{8Q^2}$$

$$\text{or, } \frac{v}{v_0} = 1 - \frac{1}{8Q^2} \Rightarrow v_0 - v = \frac{1}{8Q^2} \times v_0$$

∴ Change in Frequency,

$$\Delta v = v_0 - v = \frac{1}{8Q^2} \times v_0$$

Percentage change in Frequency,

$$= \frac{\Delta v}{v_0} \times 100 = \frac{1}{8Q^2} \times 100 = \frac{12.5}{Q^2}.$$

8.15 Forced oscillation: Resonance

When a body oscillates in any medium, other than free space, its oscillations get damped, i.e., its amplitude falls exponentially with time to zero. However, if we apply an external periodic force to the oscillator, not necessarily of the same Frequency as that of the natural Frequency of the oscillator, then the applied force tends to keep the oscillator oscillating against damping tendency of the oscillator. After some initial erratic movements, the oscillator starts oscillating with the Frequency of the applied force (also called the *driving force*) keeping a constant amplitude and phase as long as the applied or driving force remains operative. Such oscillations of the body are called the *forced* (or *driven*) *harmonic oscillations* and the body making such oscillations is called a *forced harmonic oscillator*.

The particular case, when the Frequency of the applied force is the same as the natural Frequency of the oscillator, the phenomenon of *resonance* or *resonant absorption* occurs. In such a case, as we shall see later, the amplitude of oscillation increases enormously.

Consider a body of mass m oscillating about an equilibrium position (Fig. 8.15) undergoing a damped harmonic motion. Suppose a periodic driving force $F = F_0 \sin pt$ is applied on it. The driving force is a sinusoidal force of amplitude F_0 and Frequency $\frac{p}{2\pi}$.

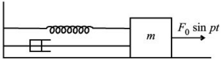


Fig. 8.15

The forces acting on the oscillator are:

(*i*) Restoring force = $-kx$

(*ii*) Damping force $= -b\frac{dx}{dt}$

(*iii*) Applied driving force = $F_0 \sin pt$.

∴ Net force on the oscillator is:

$$F = F_0 \sin pt - b\frac{dx}{dt} - kx$$

As $F = m\dfrac{d^2x}{dt^2}$, the equation of motion becomes:

$$m\frac{d^2x}{dt^2} = F_0 \sin pt - b\frac{dx}{dt} - kx$$

or

$$m\frac{d^2x}{dt^2} + b\frac{dx}{dt} + kx = F_0 \sin pt$$

or

$$\frac{d^2x}{dt^2} + \frac{b}{m}\frac{dx}{dt} + \frac{k}{m}x = \frac{F_0}{m}\sin pt \quad ...(8.64)$$

Let $\dfrac{b}{m} = 2r, \dfrac{k}{m} = \omega_0^2$ and $\dfrac{F_0}{m} = f_0,$ so that the above equation takes the form:

$$\frac{d^2x}{dt^2} + 2r\frac{dx}{dt} + \omega_0^2 x = f_0 \sin pt \quad ...(8.65)$$

The complete solution of this equation has *two* parts:

(*i*) A Complementary Function: When R.H.S. is zero i.e., $F_0 \sin pt = 0$ so that:

$$\frac{d^2x}{dt^2} + 2r\frac{dx}{dt} + \omega_0^2 x = 0$$

The solution or this equation represents transient part which dies away with time and is in fact an equation of damped harmonic oscillator (discussed earlier).

(*ii*) Steady State Function: When the steady state is attained, i.e., when the tussle between the damping and the applied force is settled and the oscillator oscillates with the Frequency of the applied force $p/2\pi$ and a constant amplitude. Let $x = A\sin(pt - \varphi)$ be a particular solution of Eq. (8.65), where φ is the phase difference between the applied force and the displacement of the oscillator. Differentiating with respect to time, we get:

$$\frac{dx}{dt} = pA\cos(pt - \varphi)$$

$$\frac{d^2x}{dt^2} = -p^2A\sin(pt - \varphi)$$

Substituting the values of $x, \dfrac{dx}{dt}$ and $\dfrac{d^2x}{dt^2}$ in Eq. (8.64), we get:

$$-p^2A\sin(pt - \varphi) + 2rpA\cos(pt - \varphi) + \omega_0^2A\sin(pt - \varphi)$$
$$= F_0\sin\{(pt - \varphi) + \varphi\}$$

or, $A(\omega_0^2 - p^2)\sin(pt - \varphi) + 2rpA\cos(pt - \varphi)$

$$= F_0\sin(pt - \varphi)\cos\varphi + F_0\cos(pt - \varphi)\sin\varphi$$

If this equation is to be satisfied for all values of t, then the coefficients of $\sin(pt - \varphi)$ and $\cos(pt - \varphi)$ on the two sides must be equal. Equating them, we obtain

$$A(\omega_0^2 - p^2) = F_0\cos f \quad ...(8.66)$$

$$\text{and } 2rpA = F_0\sin\varphi \quad ...(8.67)$$

Squaring and adding Eqs. (8.66) and (8.67), we get:

$$A^2\{(\omega_0^2 - p^2)^2 + 4r^2p^2\} = F_0^2$$

∴ Amplitude of the forced or driven oscillator,

$$A = \frac{f_0}{\sqrt{(\omega_0^2 - p^2)^2 + 4r^2p^2}} \quad ...(8.68)$$

Taking only positive value of the square root, as its negative value will mean opposite phase, there would be no effect on the value of A.

Dividing Eq. (8.67), by Eq. (8.66), we get

$$\tan\varphi = \frac{2rp}{\omega_0^2 - p^2} \quad ...(8.69)$$

∴ Phase differences between forced oscillator and the applied force is:

$$\varphi = \tan^{-1}\left(\frac{2rp}{\omega_0^2 - p^2}\right) \quad ...(8.70)$$

Substituting the value of A from Eq. (8.68), we get

$$x = \frac{f_0}{\sqrt{(\omega_0^2 - p^2)^2 + 4r^2p^2}}\sin(pt - \phi) \quad ...(8.71)$$

This is the solution of the differential Eq. (8.65) of a forced harmonic oscillator and represents a harmonic oscillator of Frequency $p/2\pi$, the same as that of the driving force but lagging behind in phase by φ given by Eq. (8.68).

Amplitude Resonance: Equation (8.68) implies that the amplitude of the forced oscillations, depends upon $\omega_0 - p$, i.e., difference in frequencies between the natural Frequency ω_0 of the oscillator and Frequency p of the applied force. It is clear that smaller this difference, larger will be the amplitude.

At Very Low Driving Frequency ($p < < \omega_0$), we have:

$$A ; \frac{f_0}{\omega_0} = \frac{F_0 / m}{k / m} = \frac{F_0}{k}$$

which indicates that the amplitude depends only on the force constant. Thus, *at very low driving* Frequency, the amplitude is independent of mass of the oscillator, the damping, the driving force and the driving Frequency.

At Very High Driving Frequency ($p >> \omega_o$), we have:

$$A ; \frac{f_0}{p^2} = \frac{F_0 / m}{p^2} = \frac{F_0}{mp^2}$$

Thus, at a very high driving Frequency, the amplitude depends upon the mass and decreases continuously as the driving Frequency increases.

Resonant Driving Frequency: The particular driving Frequency at which the amplitude of the driven oscillator is maximum is known as *resonant* Frequency and the phenomenon of amplitude becoming maximum is known as *amplitude resonance*. For this, the denominator in Eq. (8.68) should be a minimum, that is:

$$\frac{d}{dp}[(\omega_0^2 - p^2)^2 + 4r^2 p^2] = 0$$

or, $2(\omega^2_o - p^2)(-2p) + 4r^2(2p) = 0$

or, $\omega^2_o - p^2 = 2r^2$

or, $p = \sqrt{\omega_0^2 - 2r^2}$

For replacing p by p_R at resonance:

$$p_R = \sqrt{\omega_0^2 - 2r^2} \quad ...(8.72)$$

Thus, the amplitude of the driven oscillator is maximum when the driving Frequency is:

$$\frac{p_R}{2\pi} = \frac{\sqrt{\omega_0^2 - 2r^2}}{2\pi} \quad ...(8.73)$$

which is slightly less than the natural or undamped Frequency $\frac{\omega_0}{2\pi}$ as well as the damped Frequency $\frac{\sqrt{\omega_0^2 - r^2}}{2\pi}$ of the oscillator.

Substituting p_R for p in Eq. (8.68), we have

maximum amplitude $A_{\max} = \frac{f_0}{2r(\omega_0^2 - r^2)^{1/2}}$

Because, $\omega^2_o - 2r^2 = p^2$ or $\omega^2_o = p^2 + 2r^2$, we have:

$$A_{\max} = \frac{f_0}{2r(p^2 + r^2)^{1/2}} \quad ...(8.74)$$

Showing that the smaller the value r, the greater is the value of $A_{\max}$.

In the case when the damping is small ($r \to 0$), the resonant Frequency is nearly equal to the natural Frequency $\frac{\omega_0}{2\pi}$. In the ideal case of no damping, $r = 0$ and therefore,

$$p_R = \omega_o$$

Therefore, *in the absence of damping, the resonance takes place when the* Frequency *of the applied force is equal to the natural* Frequency *of the oscillator*. In the absence of damping ($r = 0$), the amplitude should become infinite. However, damping is never actually zero and so the amplitude is never infinite. Thus, we see that the damping controls the response even at resonance.

Phase Difference φ between displacement and driving force is given by:

$$\tan\varphi = \frac{2rp}{\omega_0^2 - p^2}$$

(*i*) At low driving Frequency ($p << \omega_o$), $\tan\varphi$ is a small positive quantity and the driven oscillator is nearly in phase with driving force.

(*ii*) At high driving Frequency ($p >> \omega_o$), $\tan\varphi$ is a small negative quantity and the driven oscillator is out of phase with the driving force, i.e., there is a phase difference of $\pi\left(\text{or } \frac{T}{2}\right)$ between the driven oscillator and the driving force.

(*iii*) At resonance $p = \omega_o$, $\tan\phi = \frac{2rp}{0} = \infty$ or $\phi = \frac{\pi}{2}$. That is, at resonance displacement lags behind the driving force by $\frac{\pi}{2}\left(\text{or } \frac{T}{4}\right)$.

From the above, it is clear that the phase angle φ changes from 0 to π, but remains positive throughout. The variation of phase lag with the increase of the driving Frequency is shown in Fig. 8.16.

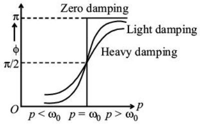


Fig. 8.16

In the curve, following points can be observed:

(*a*) When the damping is zero, the curve runs almost along the Frequency axis 0 to ω_0 and again parallel to it from ω_0 to $2\omega_0$ but differs from the first by π. In other words, for $p << \omega_0$, q = 0 and for $p >> \omega_0$, q = π.
(*b*) In the presence of damping, phase lag increases from 0 to $\pi/2$ as the driving Frequency increases from 0 to ω_0 and approaches π as p increases beyond ω_0.
(*c*) The rate of change of phase angle is rapid when the damping is lower than when it is high.

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

(Δ) All curves pass through when $p = \omega_0$, i.e., amplitude resonance occurs when the driving Frequency is equal to the undamped natural Frequency of the oscillator or the resonance Frequency in all these cases remains ω_0.

8.16 Sharpness of resonance

The amplitude of forced oscillations attains peak value when Frequency of the applied force satisfies the resonance conditions. As soon as the Frequency changes from the resonance Frequency, the amplitude falls. The rate at which the amplitude falls as the Frequency is changed on either side of resonant Frequency is related with the sharpness of resonance. *If fall in amplitude for a small change in* Frequency *from the resonant value is high, then the resonance is said to be sharp, while if the fall in amplitude is small, then the resonance is said to be flat. In other words, sharper is the fall in amplitude on either side of the resonant* Frequency*, sharper is the resonance.*
Effect of Damping on the Sharpness of the Resonance: The amplitude of forced oscillations is given by:

$$A = \frac{f_0^2}{\sqrt{(\omega_0^2 - p^2)^2 + 4r^2p^2}}$$

Thus, the amplitude A depends upon the relative magnitude of the Frequency of the applied force p and the resonant Frequency ω_0 and the damping r. Fig. 8.17 shows the variation of amplitude A with the Frequency of the driving force for different range of damping r. Following points can be observed from the curves:

At low driving Frequency, for all values of damping, the amplitude is nearly the same. As the Frequency p increases, the amplitude increases. However, the extent of increase and the peak value depends upon the damping present. When there is no damping (r = 0), the amplitude becomes infinite at resonance. As the damping increases, the peak of the curve moves towards the left, i.e., the Frequency p of the driving force for which amplitude is maximum, i.e., the Frequency p of the driving force for which amplitude is maximum decreases. Also, with increasing damping, the peak moves downward, i.e., the maximum amplitude of the oscillator falls at resonance.

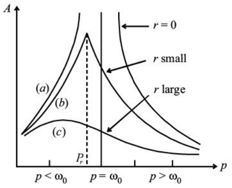


Fig. 8.17

As the driving Frequency p increases beyond resonant Frequency, the amplitude falls and tends towards zero for all values of damping. Lower the value of damping, more rapid is the fall. We can also observe that for the deviation from resonant Frequency, the amplitude of oscillation falls more rapidly for low damping, case compared to that with large damping.

Thus, *smaller the damping, sharper is the resonance and larger the damping, flatter is the resonance.*

8.17 Velocity resonance

We have seen that the displacement of a driven oscillator is given by:

$$x = A\sin(pt - \phi) = \frac{f_0}{\sqrt{(\omega_0^2 - p^2)^2 + 4r^2p^2}}\sin(pt - \phi)$$

The velocity at any instant t is:

$$v = \frac{dx}{dt} = \frac{f_0 p}{\sqrt{(\omega_0^2 - p^2)^2 + 4r^2p^2}}\cos(pt - \phi)$$

or, $v = pA\cos(pt - \varphi) = v_0\cos(pt - \varphi)$ ...(8.75)

where, $v_0 = pA = \dfrac{f_0 p}{\sqrt{(\omega_0^2 - p^2)^2 + 4r^2p^2}}$ ...(8.76)

is the expression for the velocity amplitude of the driven harmonic oscillator.

Alternatively, $v = v_0\sin(pt - \varphi + \pi/2)$ ...(8.77)

This shows that the *velocity leads the displacement by* $\pi/2$. From Eq. (8.76) it is clear that the velocity amplitude depends upon the values of p, ω_0 and r. It is zero when p = 0 and maximum when $p = p_R = \omega_0$ and the maximum value is:

$$v_0 = \frac{f_0}{2r} = \frac{f_0}{b} = f_0\tau \quad ...(8.78)$$

Attaining maximum velocity by a driven oscillator is known as *velocity resonance* and occurs when the driving Frequency is equal to the resonant Frequency $\omega_0 = \left(\frac{k}{m}\right)$, whatever be the damping. Naturally, for all other frequencies p greater than or smaller than p_R, the velocity amplitude will be smaller than that at resonance.

We have seen earlier that the *displacement at resonance lags in phase by π/2 behind force*. While, the velocity leads the displacement in phase by π/2. Thus, at resonance the *velocity of the driven oscillator is in phase with the driving force*. This is, therefore, the most favourable circumstance for the transfer of energy from the driving force to the driven oscillator as both F and v are in the same phase and v being at its maximum, the rate of transfer of energy Fv has its highest possible value.

8.18 power absorption by forced oscillator

We derive here an expression for the average power absorbed per cycle by a driven oscillator to compensate its loss of power in overcoming the frictional or resistive forces and thus to maintain its oscillations. When *the steady state is reached, the average power absorbed is equal to the average power dissipated.*

The displacement at any instant t of an oscillator subjected to driving force F_0 sin pt is given by

$$x = \frac{f_0}{(\omega_0^2 - p^2)^2 + 4r^2p^2}\sin(pt - \phi)$$

where $f_0 = \frac{F_0}{m}, r = \frac{b}{2b}$ (b = damping constant), $\omega_0 = \sqrt{\frac{k}{m}}$ (k = spring constant) and $\tan\phi = \frac{2rp}{\omega_0^2 - p^2}$. The velocity at any instant t is:

$$v = \frac{dx}{dt} = \frac{f_0 p}{(\omega_0^2 - p^2)^2 + 4r^2p^2}\cos(pt - \phi)$$

The power absorbed by the oscillator is:

$$P = Fv = (F_0 \sin pt)\cdot\frac{f_0 p}{(\omega_0^2 - p^2)^2 + 4r^2p^2}\cos(pt - \phi)$$

$$= \frac{mf_0^2 p}{(\omega_0^2 - p^2)^2 + 4r^2p^2}(\sin pt\cos pt\cos\phi + \sin^2 pt\sin\phi)$$

Average over a cycle of period $T\left(=\frac{2\pi}{p}\right)$ and using the identities,

$$\frac{1}{T}\int_0^T \sin pt\cos pt\, dt = 0 \quad\text{and}\quad \frac{1}{T}\int_0^T \sin^2 pt\,dt = \frac{1}{2},$$

we have

$$\therefore P_{av} = \frac{mf_0^2 p}{(\omega_0^2 - p^2)^2 + 4r^2p^2}\cdot\left(\frac{1}{2}\sin\phi\right)$$

Since $\tan\phi = \frac{2rp}{\omega_0^2 - p^2}$ and hence $\sin\phi = \frac{2rp}{\sqrt{\omega_0^2 - p^2 + 4r^2p^2}}$

$$\therefore p_{av} = \frac{mf_0^2 rp^2}{(\omega_0^2 - p^2)^2 + 4r^2p^2} \quad ...(8.79)$$

This is the expression for the average power absorbed by the oscillator.

Also, $\frac{f_0^2 p^2}{(\omega_0^2 - p^2) + 4r^2p^2} = v_0^2$, where v_0 is the velocity amplitude

$$\therefore P_{av} = mr\, v_0^2 = \frac{m v_0^2}{2\tau} \quad ...(8.80)$$

Power Dissipation by Driven Oscillator: The power absorbed from the driving force is dissipated in doing work against the damping force $-b\left(\frac{dx}{dt}\right)$ or $-2mr\left(\frac{dx}{dt}\right)$. The rate of doing work or instantaneous power P' against the damping force by the oscillator is given by:

$$P' = \left(2mr\frac{dx}{dt}\right)\cdot\frac{dx}{dt} = 2mr\left(\frac{dx}{dt}\right)^2$$

$$= 2mr\frac{f_0^2 p^2}{(\omega_0^2 - p^2)^2 + 4r^2p^2}\cos^2(pt - \phi)$$

Since $\frac{1}{T}\int_0^T \cos^2(pt - \phi) = \frac{1}{2}$, therefore, the average power dissipated is:

$$P'_{av} = \frac{mf_0^2\, rp^2}{(\omega_0^2 - p^2)^2 + 4r^2p^2} \quad ...(8.81)$$

Thus, we see that the average power absorbed is equal to the average power dissipated.

Maximum Power Absorption: The average power absorbed is given by:

$$P_{av} = \frac{mf_0^2\, r p^2}{(\omega_0^2 - p^2)^2 + 4r^2 p^2}$$

For P_{av} to be maximum, the denominator $(\omega_0^2 - p^2)^2 + 4r^2\ p^2$ should be minimum. Since the minimum value of a squared term is zero, therefore, $4r^2p^2$ is zero. Thus, for maximum power absorption,

$$(\omega_0^2 - p^2)^2 = 0$$

$$\Rightarrow p = \omega_0$$

which is the condition for *velocity resonance*. Hence, it can be concluded that the power transferred from the driving force to the oscillator is maximum at the Frequency of velocity resource and is equal to:

$$P_{av}\text{ (maximum)} = \frac{mf_0^2\, r p^2}{4r^2 p^2} = \frac{mf_0^2}{4r} = \frac{mf_0^2}{2b/m} = \frac{m^2 f_0^2}{2b} \quad ...(8.82)$$

where $f_0 = \frac{F_0}{m}$ and $r = \frac{b}{2m}$. We have also seen at $p = \omega_0$

$$P_{av}\text{(maximum)} = \frac{mf_0^2}{4r}$$

8.19 bandwidth of resonance curve

The power absorbed by a driven oscillator depends upon the Frequency of the driving force. It is maximum, when the Frequency of driving force is equal to the natural Frequency of the oscillator. A typical Frequency vs power absorption curve is shown in Fig. 8.18. It is symmetrical about the resonant Frequency, indicating that it is maximum at resonant Frequency and falls on either side of it. Therefore, there must be a Frequency on either side of ω_0 at which the absorbed power is half of the maximum power.

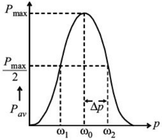


Fig. 8.18

If ω_1 and ω_2 are the values of p for which power absorbed is half of the maximum power, the Dw $(\omega_2 - \omega_1)$ is called the *bandwidth* of the resonance curve. The expression for the average power is:

$$P_{av} = \frac{mf_0^2\, rp^2}{(\omega_0^2 - p^2)^2 + 4r^2 p^2}$$

At ω_1 and ω_2, $P_{av} = \frac{1}{2} P_{av}$ (maximum)

$$\text{or,}\quad \frac{mf_0^2 rp^2}{(\omega_0^2 - p^2)^2 + 4r^2p^2} = \frac{1}{2}\frac{mf_0^2}{4r}$$

$$\text{or,}\quad (\omega_0^2 - p^2)^2 + 4r^2p^2 = 8r^2p^2$$

$$\text{or,}\quad (\omega_0^2 - p^2)^2 = 4r^2p^2$$

$$\text{or,}\quad \omega_0^2 - p^2 = \pm 2rp$$

$$\text{or,}\quad \omega_0 - p = \pm\frac{2rp}{\omega_0 + p}$$

If $\omega_2 - \omega_1$ is the bandwidth of the oscillator, then:

$$\omega_2 - p = \frac{2rp}{\omega_0 + p}$$

$$\text{and,}\quad \omega_1 - p = -\frac{2rp}{\omega_0 + p}$$

Subtracting, we get

$$\omega_2 - \omega_1 = \frac{4rp}{\omega_0 + p} = \frac{4r}{\frac{\omega_0}{p} + 1}$$

$$= 2r \quad ...(8.83)$$

$$\text{or,}\quad w_2 - \omega_1 = \frac{1}{\tau} \quad ...(8.84)$$

where $\tau\left(= \frac{1}{2r}\right)$ is the relaxation time.

Thus, the smaller the bandwidth, the sharper is the resonance and vice versa.

Using Eq. (8.83), we easily find that

$$|\omega_0 - \omega_1| = |\omega_2 - \omega_0| = \Delta p = r$$

This change Δp, in the value of driving Frequency p for which the average power absorbed by the driven oscillator falls from its maximum value P_{max} to half this value is the *half width of the average power absorbed*. From Eq. (8.84) if follows:

$$\Delta p = \frac{1}{2\tau} \quad ...(8.85)$$

quality Factor in Terms of Half-band Width

quality factor, q =

$$2\pi \frac{\text{average energy stored}}{\text{energy dissipated per cycle}} = 2\pi \frac{E}{PT}$$

Energy =

$$\text{K.E.} + \text{P.E.} = \frac{1}{2} mp^2 A^2 \cos^2(pt - \phi) + \frac{1}{2} m\omega_0^2 A^2 \sin^2(pt - \phi)$$

Since the average value of $\cos^2(pt - \varphi)$ and $\sin^2(pt - \varphi)$ over a cycle is $\frac{1}{2}$ and average power absorbed is $P_{av} = \frac{mv_0^2}{2\tau} = \frac{mA^2 p^2}{2\tau}$, we have,

q =

$$2\pi \frac{\frac{1}{4} mp^2 A^2 + \frac{1}{4} m\omega_0^2 A^2}{\left(\frac{mA^2 p^2}{2\tau}\right) \times T} = 2\pi \frac{\frac{1}{4} mA^2 (p^2 + \omega_0^2)}{\left(\frac{mA^2 p^2}{2\tau}\right) \cdot \frac{2\pi}{p}}$$

$$= \frac{1}{2} \frac{(p^2 + \omega_0^2)}{p} \cdot \tau = \frac{1}{2}\left(p + \frac{\omega_0^2}{p}\right) \cdot \tau$$

$$= \frac{1}{2}\left(1 + \frac{\omega_0^2}{p^2}\right) p\tau$$

At resonance, $p = p_R = \omega_0$

$$\therefore = \frac{1}{2}(2) p\tau = p\tau = \omega_0 \tau = \frac{\omega_0}{2\Delta p} \quad ...(8.86)$$

At low damping, τ is large so that the quality factor q = $\omega_0 \tau$ will be large making resonance sharp. Thus, the quality factor q is a measure of resonance sharpness in the case of a driven harmonic oscillator.

8.20 driven lcr series circuit

A series *LCR* circuit containing an inductance λ, a capacitance C and a resistance R is shown in Fig. 8.19. If it is connected to an external emf source to supply the necessary energy to excite and maintain oscillations, it is one of the most common examples of a driven oscillator.

Let $E = E_0 \sin pt$ be the external emf applied to the circuit. If I be the current in the circuit at a given instant and q the charge on the plates of the capacitor, then:

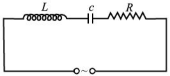


Fig. 8.19

potential drop across inductor = $-L\frac{dI}{dt}$,

potential drop across capacitor = q/C, and

potential drop across resistor = IR.

$$\therefore \quad E - L\frac{dI}{dt} = \frac{Q}{C} + IR$$

$$\text{or,} \quad L\frac{dI}{dt} + IR + \frac{Q}{C} = E = E_0 \sin pt \quad ...(8.87)$$

$$\text{or,} \quad \frac{dI}{dt} + \left(\frac{R}{L}\right) I + \frac{Q}{LC} = \left(\frac{E_0}{L}\right) \sin pt$$

As, $I = \frac{dQ}{dt}$,

$$\therefore \quad \frac{d^2Q}{dt^2} + \left(\frac{R}{L}\right)\frac{dQ}{dt} + \frac{Q}{LC} = \left(\frac{E_0}{L}\right) \sin pt \quad ...(8.88)$$

This is an equation identical in Eq. (8.65), with x being replaced by q, $2r$ $\frac{R}{L}$, ω_0^2 by $\frac{1}{LC}$ and F_0 by E_0/λ.

Its steady state solution is:

q = $$\frac{E_0/L}{\sqrt{\left(\frac{1}{LC}-p^2\right)+\left(\frac{pR}{L}\right)^2}}\sin(pt-\phi)$$ ...

(8.89)

where, φ = $$\tan^{-1}\frac{pR/L}{\frac{1}{LC}-p^2}$$

We can differentiate Eq. (8.89) with respect to time to obtain the expression for I.

Alternatively, we can obtain the solution of Eq. (8.87) by assuming that:

$I = I_0\sin(pt-\varphi)$

∴ $$\frac{dI}{dt}$$ $= I_0 p\cos(pt-\varphi)$

and q =

$$\int I dt = \int I_0\sin(pt-\phi)\,dt = -\frac{I_0}{p}\cos(pt-\phi)$$

Substituting these values in Eq. (8.89), we get:

$LI_0 p\cos(pt-\varphi) +$

$$RI_0\sin(pt-\phi)-\frac{I_0}{PC}\cos(pt-\phi)$$

$= E_0\sin pt$ ...(8.90)

or, $I_0[R\sin(pt-\varphi) +$ $$\left(Lp-\frac{1}{Cp}\right)\cos(pt-\phi)]$$

$= E_0\sin pt$

Putting $R = a\cos\varphi$ and $$Lp-\frac{1}{Cp}=a\sin\phi,$$ we have:

$a =$

$$\sqrt{R^2+\left(Lp-\frac{1}{Cp}\right)^2}\ \text{and}\ \tan\phi=\left(Lp-\frac{1}{Cp}\right)/R$$

Substituting these values of R and $Lp-\frac{1}{Cp}$ in Eq. (8.90), we get:

$I_0 a[\cos\varphi\sin(pt-\varphi)+\sin\varphi\cos(pt-\varphi)]$

$= E_0\sin pt$

or, $I_0 a\sin[(pt-\varphi)+\varphi] = I_0 a\sin pt = E_0\sin pt$

or, $$I_0\sqrt{R^2+\left(Lp-\frac{1}{Cp}\right)}\sin pt = E_0\sin pt$$

∴ $$I_0\sqrt{R^2+\left(Lp\frac{1}{Cp}\right)^2}$$ $= E_0$

or, I_0 = $$\frac{E_0}{\sqrt{R^2+\left(Lp-\frac{1}{Cp}\right)^2}}$$ ...(8.91)

∴ I =

$$I_0\sin(pt-\phi)=\frac{E_0}{\sqrt{R^2+\left(Lp-\frac{1}{Cp}\right)^2}}\sin(pt-\phi)$$

...(8.92)

where φ = $$\tan^{-1}\left(Lp-\frac{1}{CP}\right)/R$$ ...(8.93)

It the denominator of Eq. (8.91), the expression $\sqrt{R^2+\left(Lp-\frac{1}{Cp}\right)^2}$ is equivalent to the effective resistance of the series LCR circuit and it is called *impedance (Z) of the circuit*.

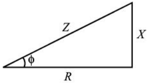


Fig. 8.20

The impedance term has two parts, Frequency independent ohmic resistance R and the Frequency dependent $\left(Lp-\frac{1}{Cp}\right)$ term called the *reactance* (*X*). The above relation between R, reactance and impedance is shown in Fig. 8.20. The reactance X also has two parts, Lp, the reactance due to the inductor λ called *inductive reactance* (X_L) and C_p, the reactance due to the capacitor C called *capacitive* (X_C). Thus,

$$X = X_L + X_C$$

Impedance, $Z =$

$$\sqrt{R^2 + X^2} = \sqrt{R^2 + (X_L - X_C)^2}$$

Peak value of current,

$$I_0 = \frac{E_0}{Z}$$

$$\therefore I = I_0 \sin(pt - \phi) = \frac{E_0}{Z} \sin(pt - \phi)$$

Clearly, the current and emf differ in the circuit by phase angle φ, where

$$\phi = \tan^{-1}\left(Lp - \frac{1}{Cp}\right) \Big/ R \;= \tan^{-1}\left(\frac{X_L - X_C}{R}\right).$$

The ohmic resistance R being independent of Frequency, has no effect on the phase angle. While, the inductor in a circuit makes the emf lead (in phase) the current and the capacitor in a circuit does the opposite, i.e., it makes the current lead the emf in phase. In an *LCR* circuit, containing both λ and C, the phase relation between current and emf depends upon the values of X_L and X_C. The following *three* cases arise:

(i) $X_L = X_C$: When the value of the Frequency p of the applied emf is such that $X_L = X_C$ then $X_L - X_C = 0$

or, $Z = R$.

That is, the impedance Z of the circuit is equal to resistance R and the current is equal to the maximum possible value and this is the condition of *resonance*.

$\therefore$ Amplitude or peak value of current $I_0 = \dfrac{E_0}{R}$, which approaches ∞ as $R \to 0$.

That is, there is no phase difference between current or emf or they are in the same phase

Thus, the condition for resonance is:

$$X_L = X_C \Rightarrow Lp = \frac{1}{Cp}$$

$$\text{or, } p^2 = \frac{1}{LC} \Rightarrow p = \frac{1}{\sqrt{LC}}$$

Resonance Frequency, $$\omega_0 = \frac{p}{2\pi} = \frac{1}{2\pi\sqrt{LC}} \quad \ldots (8.94)$$

This condition of *current resonance is analogous to velocity resonance in case of mechanical oscillator.*

(ii) $X_L > X_C$: When the value of the Frequency p is such that $X_L > X_C$ or $Lp > \dfrac{1}{Cp}$, then the *net reactance is inductive* and it follows from Eq. (8.93) that tan φ is positive, indicating that the current lags behind the emf. The value of p is greater than ω_0.

(iii) $X_L < X_C$: When the value of the Frequency p is such that $X_L < X_C$ or $Lp < \dfrac{1}{Cp}$, then the *net reactance is capacitive* and tanφ is −ve, indicating that the current leads the emf. The value of p is less than ω_0.

The above-mentioned variation of peak current I_0 and phase angle φ with Frequency p are shown in Fig. 8.21 (*a*) and (*b*) respectively, for different values of R which corresponds to damping in the case of mechanical oscillator. It is observed that lower value of R, sharper is the resonance and the current lags or leads the emf depending on whether p is greater or less than ω_0.

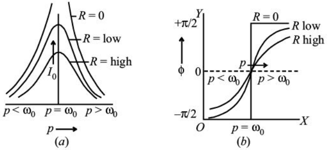


Fig. 8.21

Example 8.17

A harmonic oscillator of quality factor 10 is subjected to a sinusoidal applied force of Frequency one and a half times the natural Frequency of the oscillator. If the damping is small, obtain (*i*) the amplitude of the forced oscillation in terms of its maximum amplitude and (*ii*) the angle q by which it will be out of phase with the driving force.

Solution:

(*i*) The amplitude of the forced oscillation is given by:

$$A = \frac{f_0}{\sqrt{(\omega_0^2 - p^2) + 4r^2p^2}} = \frac{f_0}{\omega_0^2\sqrt{\left(1 - \dfrac{p^2}{\omega_0^2}\right)^2 + \dfrac{4r^2p^2}{\omega_0^4}}}$$

where the symbols have their usual meaning.

quality factor $Q = \omega_0^2\tau = \frac{\omega_0}{2r} = 10.$ Therefore, $2r = \frac{\omega_0}{10}.$

Given: $p = \frac{3}{2}\omega_0 \Rightarrow \frac{p}{\omega_0} = \frac{3}{2}$

$$\therefore A = \frac{f_0}{\omega_0^2\sqrt{\left(1-\left(\frac{3}{2}\right)^2\right)^2 - \left(\frac{1}{10}\right)^2\left(\frac{3}{2}\right)^2}}$$

$$= \frac{f_0}{\sqrt{\left(1-\frac{9}{4}\right)^2 + \frac{1}{100}\cdot\frac{9}{4}}}$$

$$= \frac{f_0}{\omega_0^2\sqrt{\left(\frac{25}{16}+\frac{9}{400}\right)}} = \frac{f_0}{\omega_0^2\sqrt{\left(\frac{634}{400}\right)}} = \frac{20f_0}{\omega_0^2\sqrt{634}}$$

For low damping, amplitude is maximum when $p = \omega_0$. So that:

$$A_{max} = \frac{f_0}{\omega_0^2\sqrt{(1-1)^2 + \frac{1}{100}\times 1}} = \frac{f_0}{\omega_0^2\sqrt{\left(\frac{1}{100}\right)}}$$

$$= \frac{f_0}{\omega_0^2\left(\frac{1}{10}\right)} \quad \frac{10f_0}{\omega_0^2}$$

$$\therefore \frac{A}{A_{max}} = \frac{20f_0}{\omega_0^2\sqrt{634}} \times \frac{\omega_0^2}{10f_0} = \frac{2}{\sqrt{634}} = \frac{2}{25.18} = 0.0794 \simeq 0.08$$

or, amplitude $A = 0.08A_{max}$

(ii)

$$\tan\phi \frac{2rp}{\omega_0^2 - p^2} = \frac{\frac{\omega_0 p}{10}}{\omega_0^2 - p^2} = \frac{\frac{\omega_0}{10}\times\left(\frac{3\omega_0}{2}\right)}{\omega_0^2 - \frac{9\omega_0^2}{4}} = \frac{3\omega_0^2}{20}\times\frac{4}{5\omega_0^2}$$

$$= \frac{3}{25} = 0.12$$

or, $\varphi = \tan^{-1}(0.12) = 6°51'$.

Example 8.18

A 0.5 kg mass is suspended from a linear spring of force constant 1000 N/m and a damping coefficient b = 0.05 Nsm^{-1}. An external force $F = F_0 \sin pt$ is applied where F_0 = 25 N and p is twice the natural Frequency of the system. Calculate (*i*) the amplitude of resulting motion. (ii) phase shift of displacement with respect to the driving force.

Solution:

(*i*) The amplitude of the displacement, neglecting the gravitational effect is given by:

$$A = \frac{f_0}{\sqrt{(\omega_0^2 - p^2) + 4r^2p^2}}$$

where, $F_0 = \frac{F_0}{m} = \frac{25}{0.5} = 50\,\text{N/kg}$

$$\omega_0^2 = \frac{k}{m} = \frac{1000}{0.5} = 2000\ \text{sec}^{-2}$$

$p = 2\omega_0$

q $2r = \frac{b}{m} \Rightarrow r = \frac{b}{2m} = \frac{0.05}{2\times 0.5} = 0.05\ \text{sec}^{-1}$

$$\therefore A = \frac{f_0}{\sqrt{\{\omega_0^2 - (2\omega_0)^2\}^2 + 4r^2(2\omega_0)^2}}$$

$$= \frac{f_0}{\sqrt{9\omega_0^4 + 16r^2\omega_0^2}}$$

$$= \frac{50}{\sqrt{9\times(2000)^2 + 16\times(0.05)^2\times(2000)^2}}$$

$$= \frac{50}{6 \times 10^3} = 8.33 \times 10^{-3} \text{ m} = 8.33 \text{ mm}$$

(*ii*) Phase shift

$$\phi = \tan^{-1}\left(\frac{2rp}{\omega_0^2 - p^2}\right)$$

$$= \tan^{-1}\left(\frac{2r(2\omega_0)}{\omega_0^2 - (2\omega_0)^2}\right) = \tan^{-1}\left(-\frac{4r}{3\omega_0}\right)$$

$$= \tan^{-1}\left(-\frac{4 \times 0.05}{3\sqrt{2000}}\right) = \tan^{-1}(-0.00149).$$

Exercises

1. What do you mean by fundamental interactions? Name them and discuss the nature of these interactions.

2. What are gauge particles? Name the gauge particles for all the fundamental forces.

3. What do you mean by Electroweak-theory and Grand Unified Theory?

4. Give a comparative description of the fundamental forces.

5. State and explain the Law of Conservation of Energy.

6. Show that a conservative force is the negative gradient of potential energy.

7. Show that for a conservation force, curl of the force is zero.

8. Differentiate between Conservative and Non-conservative Forces. Give examples of both types of forces.

9. Show that the work done by a conservative force in a round trip is zero.

10. State and explain the Law of Conservation of Momentum.

11. Define *collision* and differentiate between Elastic and Inelastic Collision.

12. What are *central forces*? Show that the angular momentum of a particle moving under the influence of a central force is constant.

13. No static equilibrium is possible in an inverse square law force field! Explain.

14. Justify the statement, 'the entire known universe must consist of moving bodies in which a static state of the bodies is impossible'.

15. Discuss the symmetry of interaction and conservation laws, giving suitable examples.

16. What is meant by damping? How does a damped harmonic motion differ from an ideal simple harmonic motion. Explain.

17. Obtain the differential equation representing a damped harmonic oscillator and solve it. Discuss different types of motion of a damped oscillator.

18. Obtain the expression for the velocity of a damped harmonic oscillator and obtain the expression and definition of relaxation time from it.

19. Derive an expression for the average total energy and average power dissipation by a damped harmonic oscillator.

20. Define a relaxation time and obtain the expression for power dissipation and relaxation time.

21. What is meant by terms (*i*) logarithmic decrement and (*ii*) quality factor of a damped harmonic oscillator? Derive the expression for them.

22. Justify the statement, 'the smaller the damping, the larger is the relaxation time and the higher is the quality factor.'

23. Show that the relaxation time of a damped harmonic oscillator has the dimensions of time. Also, show that (*a*) in this time (τ) the amplitude of the oscillation falls to 0.6084 of its undamped value and (*b*) it is reduced to half its undamped value in time 1.3846τ. ($\log_e 10 = 2.3$ and $E = 2.718$).

24. What do you mean by forced vibrations? Explain.

25. What is meant by the term resonance? Explain giving suitable example.

26. Obtain and solve the differential equation of a damped harmonic oscillator subjected to a sinusoidal force and obtain expression for maximum amplitude and quality factor.

27. Show that for an oscillator undergoing forced oscillations, the response is (*i*) independent of its mass if $p << \omega_0$ and (*ii*) independent of the spring constant if $p >> \omega_0$.

28. Show that in a driven oscillator, the maximum power is absorbed at the Frequency of velocity resonance and not at the Frequency of amplitude resonance.

29. Discuss the 'sharpness of resonance' in case of a forced oscillator and hence justify the statement that 'the quality factor q is measure of the sharpness of resonance in the case of a driven oscillator.

30. Discuss and LCR series circuit as an example of a forced oscillator and hence define reactance and impedance. Explain whether the current lags or leads the applied emf when the net reactance in the circuit is (*i*) inductive and (*ii*) capacitive.

Numericals

1. The four engines of an aeroplane develop a total power of 22 MW, when its velocity is 240 m/s. How much force do the engines exert?

[**Ans.** 92 kN]

2. A hammer with a 1.5 kg head is used to derive a nail into a wooden board. If hammerhead is moving at 3 m/s, when it strikes the nail and the nail moves 10 mm into the board, find the average force the hammerhead exerts on the nail

[**Ans.** 1.35 kN]

3. A 1200 kg car moving east at 30 km/h collides with an 1800 kg car moving north at 20 km/h. The car stick together after the collision. Find the velocity (magnitude and direction) of the wreckage.

[**Ans.** 17 km/h to the north-east]

4. A shot of mass μ penetrates thickness *s* of a fixed plate of mass μ. Prove that if μ is free to move, the thickness penetrated is μs(μ + μ).

5. A massless platform is kept on a light elastic spring. When a ball of 0.1 kg mass is dropped on the pan from a height of 0.24 m, the ball strikes the pan and the spring compresses by 0.01 m. From what height should the ball be dropped to cause compression of 0.04 m?

[**Ans.** 3.96 m]

6. The energy of a particle is given by $U = A - \frac{B}{x} + \frac{C}{x^2}$, where *A*, *B* and *C* are all positive constants. What is the position of stable equilibrium of the particle? What is the force constant for small oscillations of the particle about this position? For what values of the total energy can the motion of the particle be bounded?

[**Ans.** $2C/B$, $B^4/8C^3$, for total energy $E < A$]

7. Solve the equation: $\frac{d^2x}{dt^2} + 2\frac{dx}{dt} + 5x$ = 0 using the condition that at t = 0, x = 5 and $\frac{dx}{dt} = -5$. Also, interpret the result physically.

[**Ans.** $x = E^{-1}(5\cos 2t + \sin 2t)$]

8. A periodic external force acts on a 6 kg mass suspended from the lower end of a vertical spring having the spring constant 150 N/m. The damping force proportional to the instantaneous speed of mass is equal to 80 N when the speed is 2 m/s. What is the resonance Frequency?

[**Ans.** $\frac{5}{6}\pi$ vibr/sec]

9. A damped harmonic oscillator is subjected to a sinusoidal driving force whose Frequency is altered and amplitude is kept constant. It is found that the amplitude of the oscillator increases from 0.02 mm at very low driving Frequency to 8.0 mm at a Frequency of 100 cps. Obtain the value of (*a*) the quality factor, (*b*) the relaxation time, (*c*) the damping factor, and (Δ) the half-width of the resonance curve.

[**Ans.** (*a*) 400, (*b*) 0.4245 sec, (c) 1.178, (Δ) 2.04 rad/sec)]

10. In an *LCR* series circuit containing an inductance of 10 mH, a capacitance of 1 μ*F* and a resistance of 10 ω, an rms voltage of 100 V is applied. Calculate the Frequency at which the circuit will be in resonance with the current of the same Frequency and find the value of the current.

[**Ans.** 1592 Hz and 10 A]

11. A damped harmonic oscillator of quality factor 20 is subjected to a sinusoidal driving force of Frequency twice the natural Frequency of the oscillator. If the damping is small, what fraction will the amplitude of the oscillator be, of its maximum value and by what angle will it differ in phase from the driving force?

[**Ans.** $0.17A_{max}$, 178.5° out of phase]

9.
Wave Motion

9.1 wave motion: types of waves

Waves are the most common mode of energy transfer from one point to another. You must have observed waves on the water surface and in a long rope when one of its ends is fixed, while the other end is moved up and down quickly. Sound waves, light waves, radio waves all carry energy from its source point to other points. In all cases of wave motion, the matter of the intervening medium does not get transferred, while the electromagnetic waves do not require any material medium at all for their propagation. In this regard, the waves are classified in two broad categories. (*i*) The waves which require a material medium for their propagation, called *mechanical waves*, and (*ii*) the waves which do not require a material medium for their propagation are called *non-mechanical* or *electromagnetic waves*. We will confine ourselves to the study of mechanical waves in this chapter.

The medium (solid, liquid or gas) through which mechanical waves can propagate *should possess both inertia and elasticity*. The individual molecule of the medium does not travel with the wave but undergoes simple harmonic motion about their equilibrium position. The mechanical waves are of two types: (*i*) transverse and (*ii*) longitudinal.

(*i*) Transverse Wave: In a transverse wave motion, the particles of the medium move up and down about their equilibrium position perpendicular to the direction of propagation of the wave. Hence, the wave motion propagates forming *crests* and *troughs* (Fig. 9.1). A crest and an adjoining trough from one *wave* or *pulse*, and a succession or a series of pulse constitutes a wave train. A wave travelling along a stretched string is an example of transverse wave.

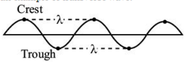


Fig. 9.1: Transverse waves.

Transverse wave is possible in a medium having the property of elasticity of shape or rigidity, e.g., a solid. Although, liquids, in general, do not possess the property of rigidity, however, they have the property of resisting any vertical displacement or keeping their level intact and, therefore, transverse waves are possible on liquid surface. While the gases neither have rigidity nor oppose any vertical displacement of the particle. Therefore, transverse waves are not possible in gaseous medium.

(*ii*) Longitudinal Waves: In longitudinal waves, the particles of the medium oscillate to and fro about their mean or equilibrium position along the direction of propagation of the waves. Such waves travel in the form of *compressions* (particles come closer) and *rarefactions* (particles move apart), as shown in Fig. 9.2. Therefore, the medium should possess elasticity of volume, i.e., longitudinal waves can propagate through solids and liquids as well as gases, Sound waves are the most common example of longitudinal waves. Waves in a spring or helix when one of its ends is suddenly compressed or pulled and then released is another example of longitudinal waves. In longitudinal waves, one compression (or rarefaction) and the adjoining rarefaction (compression), form one wave or pulse and their succession of forms, wave train.

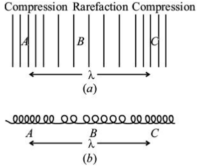


Fig. 9.2: Longitudinal waves.

As in the case of surface on water, in some cases, the waves are neither purely transverse or longitudinal. In such cases, the particles of the medium oscillate, along as well as perpendicular to the direction of propagation of wave, resulting in an elliptical path of the particles. However, we will confine our discussion to the pure transverse and longitudinal cases.

Characteristics of Wave Motion: The following important points concerning wave motion are important to remember:

(*a*) Whether the wave motion is transverse or longitudinal, it is the disturbance, a condition or state of motion that travels through the medium and there is no transfer of matter from one point to another. The particles of the medium simply undergo to and fro or up and down oscillating about their mean position and the particles do not travel with the wave.

(*b*) Each particle of the medium receives the disturbance or energy from its predecessor, undergoes the oscillatory motion and passes on the disturbance to the next succeeding particles. Thus, there is a definite time or phase lag between any two particles.

(*c*) The velocity of the particles of the medium (or particle velocity) is entirely different from the velocity of the wave motion or the wave velocity.

Wavefront: A plane or a surface on which all the particles of the medium are in the same phase or identical state of motion at a given instant is called a *wavefront*.

In a homogeneous and isotropic medium, the wavefront is always perpendicular to the direction of propagation of the wave. A line drawn perpendicular to the wavefront is called a *ray* and it also gives the direction of propagation of the wave.

Plane wave and *spherical waves* are the two most common types of waves. In case the waves are one-dimensional, such as those travelling along a stretched string or a spring. The wave produced by back and forth motion of a plane surface in a gaseous medium, plane wavefront is generated which is referred to as *plane waves*.

While in case the wave is produced from a point source in a homogeneous medium resulting in a spherical wave, stretched *a spherical wavefront* is obtained. Since the radius of a spherical wavefront goes on increasing as the waves move outward, its curvature goes on increasing progressively until at an infinite or large distance from the source, it increasingly takes the shape of a plane wavefront. In fact, at a large distance from the source, a portion of spherical wavefront can be treated as plane waves.

Wavelength: The *distance between the two nearest particles of the medium in the same phase of the wave motion is called the wavelength* (λ) *of the wave*. In the case of transverse wave (Fig. 9.1), separation between two successive crests of two successive troughs is equal to the wavelength. While in the case of longitudinal waves (Fig. 9.2), the distance between the centres of two nearest compressions or two nearest rarefactions is equal to the wavelength.

Alternatively, wavelength may also be defined as the distance covered by the disturbance or wave in one time period (T).

The SI unit of wavelengths is metre (m) and in CGS unit, is centimetre (cm).

Frequency: The number of waves produced per second is the frequency (v) of the wave motion. Naturally, it is also equal to the number of oscillations completed in one second by the particles of the medium. Frequency is equal to the reciprocal of time period. That is,

$$v = \frac{1}{T}$$

Its unit is Hertz (Hz) or cycle per second (cps). Angular frequency $\left(\omega = 2\pi v = \frac{2\pi}{T}\right)$ is equal to the time rate of change of the phase, of the wave motion. A phase change of 2π radian takes place in a time equal to the time period T of the wave motion.

Wave Number: The number of waves contained in a distance equal to the wavelength of the wave motion is called the *wave number*, i.e.,

$$k = \frac{2\pi}{\lambda}.$$

It is also defined as *space rate of change of phase*, as a phase change of 2π radian occurs in a distance of λ.

Phase Velocity: It is the velocity with which planes of equal phase (e.g., crests or troughs) propagate in the medium. It is also called the *wave velocity* and is given by:

$$v = v\lambda = \frac{\lambda}{T}.$$

Particle Velocity: The particle velocity is the simple harmonic velocity with which particles of the medium oscillate about its equilibrium position.

9.2 expression for plane progressive harmonic wave

A wave travelling onward in a given direction through a medium without any attenuation (so that its amplitude remains constant), is referred to as a *plane progressive wave*. Irrespective of the nature of the wave, transverse or longitudinal, there exists a regular phase difference between any two successive particles of the medium.

Suppose a wave originating at O (Fig. 9.3) is travelling along $+x$-axis. If y is the simple harmonic displacement of particle at position P, distant x from O at a time t after it originated at O, then y is expressed as:

$$y = a\sin(\omega t - f) \quad ...(9.1)$$

where a is the amplitude, ω the angular velocity and f the phase lag of the particle. Since the successive particles to the right of point O receive and repeat the harmonic motion after a definite interval of time resulting in a phase lag. This phase lag goes on increasing with increasing distance from O.

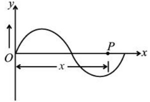


Fig. 9.3

If λ is the wavelength, then we know a distance λ corresponding to a phase difference of 2π. Thus, a path difference of x corresponds to a phase difference $\phi = \frac{2\pi x}{\lambda}$. Substituting $\omega = \frac{2\pi}{T}$ and $\phi = \frac{2\pi x}{\lambda}$ in Eq. (9.1), we get:

$$y = a\sin\left(\frac{2\pi t}{T} - \frac{2\pi x}{\lambda}\right) \quad ...(9.2)$$

If v is the wave or phase velocity, then $v = \lambda v = \frac{\lambda}{T}$.

$$\therefore T = \frac{\lambda}{v}$$

Substituting this value of T in the above equation, we have

$$y = a\sin\left(\frac{2\pi vt}{\lambda} - \frac{2\pi x}{\lambda}\right)$$

$$y = a\sin\frac{2\pi}{\lambda}(vt - x) \quad ...(9.3)$$

Equation. (9.3) can also be expressed in terms of propagation constant k. We know $\frac{2\pi}{\lambda} = k$, therefore,

$y = a\sin(\omega t - kx)$

or, $y = a\sin\frac{2\pi}{T}\left(t - \frac{x}{v}\right)$ (q $\lambda = vT$)

or, $y = a\sin\omega\left(t - \frac{x}{v}\right)$ ...(9.4)

When the wave is travelling towards the left, i.e., $-x$-axis, the wave equation expressed by Eq. (9.3), can be written as:

$$y = a\sin\frac{2\pi}{\lambda}[vt - (-x)] = a\sin\frac{2\pi}{\lambda}(vt + x) \quad ...(9.5)$$

or, $y = a\sin(\omega t + kx)$ ...(9.6)

The quantity $(\omega t \pm kx)$ represents the *phase* of the wave motion. We have assumed that at $t = 0$, $y = 0$, i.e., motion starts at O. If this is not the case and the particle has some initial phase θ_0, the wave equation is represented as:

$$y = a\sin(\omega t - kx + \theta_0). \quad ...(9.7)$$

9.3 wave (or phase) velocity and particle velocity

In the expression for the displacement $y = a\sin\frac{2\pi}{\lambda}(vt - x)$, v is the wave (or phase) velocity which is the rate at which the wave or disturbance moves across the medium.

Particle velocity, i.e., the velocity with which the particles of the medium oscillate is obtained by differentiating y with respect to t, that is,

Particle velocity $u = \frac{dy}{dt} = \frac{2\pi v}{\lambda} a\cos\frac{2\pi}{\lambda}(vt - x)$ ...(9.8)

Differentiating Eq. (9.3), w.r.t. x, slope of the displacement curve which physically represents the strain or compression is:

$$\frac{dy}{dx} = -\frac{2\pi}{\lambda} a\cos\frac{2\pi}{\lambda}(vt - x) \quad ...(9.9)$$

Particle velocity, $u = \frac{dy}{dt} = -v\frac{dy}{dx}$ ...(9.10)

or particle velocity at a point,

= −[(wave velocity) × (slope of displacement curve at that point)]

Sometimes, the −ve sign in the expression $\frac{dy}{dt} = -v\frac{dy}{dx}$ is ignored. Figure 9.4 shows the direction of displacement and particle velocity $\frac{dy}{dt}$ at any point x is shown by the arrow.

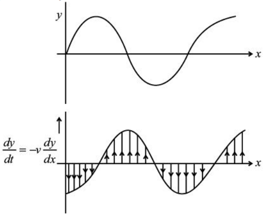


Fig. 9.4: The magnitude and direction of the particle velocity.

Differential Equation of Wave Motion

Differentiating Eq. (9.8) with respect to t, we have the acceleration of the particle.

$$\frac{d^2y}{dt^2} = -\frac{4\pi v^2}{\lambda^2} a\sin\frac{2\pi}{\lambda}(vt - x)$$

$$= -\frac{4\pi v^2}{\lambda^2} y \quad ...(9.11)$$

Also, differentiating Eq. (4.9), with respect to x, we have the rate of change of compression with distance:

$$\frac{d^2y}{dx^2} = -\frac{4\pi^2}{\lambda^2} a\sin\frac{2\pi}{\lambda}(vt - x) \quad ...(9.12)$$

From Eqs. (9.11) and (9.12), we get

$$\frac{d^2y}{dt^2} = v^2\frac{d^2y}{dx^2} \quad ...(9.13)$$

This equation is referred to as the differential equation of a plane one-dimensional progressive wave in which coefficient of $\frac{d^2y}{dx^2}$ (the curvature of displacement curve) gives the square of the wave velocity. Equation (9.13) can be written as

particle acceleration at a point

= (wave velocity)^2 × (curvature of displacement curve at that point).

9.4 variation of velocity and pressure in a progressive wave

A plane progressive wave is represented by the equation:

$$y = a\sin\frac{2\pi}{\lambda}(vt - x)$$

and the graphical representation of variation of displacement y is shown in Fig. 9.3.

The particle velocity is given by:

$$u = \frac{dy}{dt} = \frac{2\pi a v}{\lambda}\cos\frac{2\pi}{\lambda}(vt - x)$$

and its variation is shown in Fig. 9.4.

The (volume) strain in the medium is $\frac{dy}{dx}$, given by

$$\frac{dy}{dx} = -\frac{2\pi a}{\lambda}\cos\frac{2\pi}{\lambda}(vt - x) \quad ...(9.14)$$

The modulus of elasticity (K) of a medium is defined as

$$K = \frac{\text{Change in pressure}}{\text{Volume strain}} = \frac{-dP}{(dy/dx)}$$

$$\text{or, } dP = -K.\left(\frac{dy}{dx}\right)$$

$$\text{or, } dP = K\cdot\left(-\frac{dy}{dx}\right) \quad ...(9.15)$$

In a region where $\frac{dy}{dx}$ is –ve, so that dP is +ve, i.e., it is a *region of compression*. If $\frac{dy}{dx}$ is +ve, then dP is –ve, i.e., it is a *region of rarefaction*.

Using Eqs. (9.15) and (9.16), we have:

$$dP = \frac{2\pi K a}{\lambda}\cos\frac{2\pi}{\lambda}(vt - x) \quad ...(9.16)$$

The variation of dP is shown in Fig. 9.5, where P_0 is the normal pressure of the medium when the wave is not propagating.

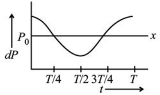


Fig. 9.5

9.5 energy of a plane progressive wave

In a plane progressive wave, the energy derived from a source is passed on from one particle of the medium to the next, and so on, resulting in regular transmission of energy across every section of the medium.

The *energy density (E)* of a plane progressive wave is the *total energy (kinetic +potential) per unit volume of the medium through which the wave is passing.*

The displacement y of a particle of the medium distant x from the wave originating point, at an instant t is given by:

$$y = a\sin\frac{2\pi}{\lambda}(vt - x)$$

Particle velocity, $u = \frac{dy}{dt} = \frac{2\pi v}{\lambda}a\cos\frac{2\pi}{\lambda}(vt - x)$

Particle acceleration,

$$\frac{du}{dt} = \frac{d^2y}{dt^2} = -\frac{4\pi^2v^2}{\lambda^2}a\sin\frac{2\pi}{\lambda}(vt - x)$$

$$= -\frac{4\pi^2v^2}{\lambda^2}y$$

Now, we consider unit volume of the medium in the form of extremely thin element of the medium parallel to the wavefront, so that,

mass of the element = r, the density (or mass per unit volume)

Considering the layer to be extremely thin, we can assume the velocity of all the particles to be the same in this layer. Therefore,

Kinetic energy per unit volume of the medium

$$= \frac{1}{2}\times(\text{mass})\times(\text{velocity})^2 = \frac{1}{2}\rho u^2$$

$$= \frac{1}{2}\rho\frac{4\pi^2v^2}{\lambda^2}a^2\cos^2\frac{2\pi}{\lambda}(vt-x)$$

$$= \frac{2\pi^2v^2\rho}{\lambda^2}a^2\cos^2\frac{2\pi}{\lambda}(vt-x) \quad ...(9.17)$$

Force acting on unit volume

= mass × acceleration

$$= \rho\times\frac{d^2y}{dt^2} = \frac{4\pi^2v^2\rho}{\lambda^2}y$$

We have ignored the –ve sign in acceleration as it only indicated that the acceleration and displacement is oppositely directed.

∴ Work done in displacement *dy* of the layer

= force × displacement

$$= \frac{4\pi^2v^2\rho y}{\lambda^2}dy$$

Total work done for displacement:

$$y = \int_0^y \frac{4\pi^2v^2\rho y\,dy}{\lambda^2} = \frac{4\pi^2v^2\rho}{\lambda^2}\int_0^y y\,dy$$

$$= \frac{4\pi^2v^2\rho}{\lambda^2}\cdot\frac{y^2}{2} = \frac{2\pi^2v^2\rho y^2}{\lambda^2}$$

$$= \frac{2\pi^2v^2\rho}{\lambda^2}a^2\sin^2\frac{2\pi}{\lambda}(vt-x)$$

This work done is stored in the medium in the form of potential energy.

Thus, potential energy per unit volume of the medium

$$= \frac{2\pi^2v^2\rho}{\lambda^2}a^2\sin^2\frac{2\pi}{\lambda}(vt-x)$$

Hence, total energy per unit volume of the medium or energy density of the plane progressive wave is given by:

E = K.E. + P.E.

$$= \frac{2\pi^2v^2\rho}{\lambda^2}a^2\cos^2\frac{2\pi}{\lambda}(vt-x) + \frac{2\pi^2v^2\rho}{\lambda^2}a^2\sin^2(vt-x)$$

$$= \frac{2\pi^2v^2\rho a^2}{\lambda^2}$$

$$= 2\pi^2\left(\frac{v}{\lambda}\right)^2\rho a^2 = 2\pi^2v^2\rho a^2$$

$$\left(\text{Frequency of } v = \frac{v}{\lambda}\right) \quad ...(9.18)$$

Here, it is important to note that both *kinetic and potential energies of the wave depend upon the x and t value while the total energy is independent of both x and t*.

Energy Current-Intensity of a Wave: If we take the cross-sectional area of the wave to be unity, we may regard the total energy per unit volume or every density $E = 2\pi^2v^2ra^2$ as the *total energy per unit length of the wave*. In a progressive wave, as new length equal to its velocity v is set into motion every second, the energy transferred per second is the energy contained in a length v.

The rate of flow of energy per unit area of cross-section of the wavefront along the direction wave propagation is called the energy current (C) or the energy flux of the wave is equal to Ev.

∴ Energy flux or energy current of a plane progressive wave is:

$C = 2\pi^2v^2ra^2v$ ...(9.19)

Now we define the *intensity of the wave (I)* as the *quantity of incident energy per unit area of the wavefront per unit time*. Thus, it is the same quantity as the energy flux or the energy current given by

$I = 2\pi^2v^2ra^2v$

showing that $I \propto \boldsymbol{a}^2$, i.e., the *intensity is proportional to the square of the amplitude of the wave.*

9.6 motion of transverse wave on a string: wave equation

Let us consider the motion of transverse waves on a uniform string, a small section of which is shown in Fig. 9.6. This section will perform vertical simple harmonic motions; our oscillator in consideration. The displacement y will depend on time t in addition to its variation with position x of observation. The wave equation, therefore, must relate the displacement y with time t and position x. We will discuss the motion in the plane of the paper so that we can treat the transverse waves on the string as plane polarized.

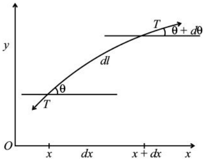


Fig. 9.6

Let a constant tension, T exist throughout the string and its linear density or mass per unit length is μ. We will neglect the effect of gravitation in our discussion. The force acting on the curved element of length dl is T at an angle θ at the left end of the element and T at an angle $\theta + d\theta$ at the right end. If the differential $\frac{dy}{dx}$ represents the *strain or compression*, then the length of the curved element is given by:

$$dl = \left[1 + \left(\frac{dy}{dx}\right)^2\right]^{1/2} dx \quad \text{...(9.20)}$$

When the magnitude of $\frac{dy}{dx}$ is very small so that the square term in the above expression can be neglected, then $dl = dx$. The mass of this element of the string is $\mu dl = r dx$.

The perpendicular force on the element dx is:

$$T\sin(\theta + d\theta) - T\sin\theta$$

in the positive y-direction. Since θ is very small, $\sin\theta \simeq \tan\theta = \frac{dy}{dx}$, so that the force is given by:

$$T\left[\left(\frac{dy}{dx}\right)_{x+dx} - \left(\frac{dy}{dx}\right)_x\right]$$

where the subscripts refer to the point where the derivative is evaluated. The difference between the two terms in the bracket defines the differential coefficient of the derivative $\frac{dy}{dx}$ times the separation between the two evaluation points so that the net force is

$$T\frac{d^2y}{dx^2}dx$$

Applying Newton's law, $F = ma$, we have

$$T\frac{d^2y}{dx^2}dx = \mu\, dx\frac{d^2y}{dt^2}$$

$$\text{or,} \quad \frac{d^2y}{dx^2} = \frac{\mu}{T}\frac{d^2y}{dt^2} \quad \text{...(9.21)}$$

The term $\frac{T}{\mu}$ has the dimension of velocity square. If v is the wave velocity, then

$$\frac{d^2y}{dx^2} = \frac{1}{v^2}\frac{d^2y}{dt^2} \quad \text{...(9.22)}$$

This is the *wave equation,* relating acceleration $\frac{d^2y}{dt^2}$ of a simple harmonic oscillator in a medium to the second derivative of its displacement $\frac{d^2y}{dx^2}$ with respect to position x.

Comparing Eq. (9.21) and (9.22), the velocity of transverse wave along the string is given by:

$$v = \sqrt{\frac{T}{\mu}} = \sqrt{\frac{\text{tension in the string}}{\text{mass per unit length of the string}}} \quad \text{...(9.23)}$$

Thus, the velocity of the transverse wave along the string depends only upon (*i*) tension applied to the string and (*ii*) mass per unit length of the string and is independent of the shape and amplitude or initial displacement produced in it.

Solution of the Wave Equation: Any solution of the wave equation (Eq. 9.22) will naturally be dependent on the variables x and t. A general solution is of the form $y = f_1(vt - x)$ or $y = f_2(vt + x)$. However, a more complete solution is a superposition of the form:

$$y = f_1(vt - x) + f_2(vt + x)$$

If f'_1 the differentiation of the function f_1, then:

$$\frac{dy}{dx} = -f'_1(vt - x)$$

$$\text{and,} \quad \frac{d^2y}{dx^2} = f''_1(vt - x)$$

Also, $\frac{dy}{dt} = vf'_1(vt - x)$

and, $\frac{d^2y}{dt^2} = v^2f''_1(vt - x)$

so that, $\frac{d^2y}{dx^2} = \frac{1}{v^2}\frac{d^2y}{dt^2}$

for $y = f_1(vt - x)$, verifying that $y = f_1(vt - x)$ is a solution of Eq. (9.22). Similarly, you can prove the same for $y = f_2(vt + x)$.

9.7 characteristic impedance of a string

When waves propagate through any medium, they encounter an impedance which is a characteristic of the medium. If the medium is lossless and possesses no resistive or dissipative mechanism, the impedance is determined by the two energy storing parameters, inertia and elasticity, and the impedance is real. However, if a loss mechanism is present, it results in a complex term in impedance.

The impedance offered to a progressive wave by the string is defined as,

impedance, $Z = \frac{\text{transverse force}}{\text{transverse velocity}} = \frac{F}{v_t}$ ... (9.24)

Because of the nature of the wave, the impedance is often referred to as *transverse impedance*. The following discussion will manifest the dual role of the string as a medium of propagation and as a forced oscillator.

We consider a progressive wave on a string (as shown in Fig. 9.7) generated at one end by an oscillating force represented by $F_0e^{i\omega t}$, which is confined to the direction transverse to the string and operates only in the plane of the paper. The tension T in the string has a constant value and at the end of the string, the balance of forces shows that the applied force is equal and opposite to $T\sin\theta$ at all time, so that,

$$F_0e^{i\omega t} = -T\sin\theta$$

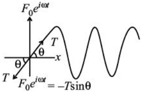


Fig. 9.7: A vertical force $F_0e^{i\omega t}$ driving one end of a string, behaving as a forced oscillator.

As θ is very small, sinθ . tanθ

$$F_0e^{i\omega t} = -T\sin\theta = T\tan\theta = -T\frac{dy}{dx}$$

The displacement of the progressive waves are represented experimentally by,

$$y = Ae^{i(\omega t - kx)}$$

where the amplitude A may be complex due to its phase relationship with force F. At left end of the string $x = 0$.

$$F_0e^{i\omega t} = -T\left(\frac{dy}{dx}\right)_{x=0} = ikTA\,e^{i(\omega t - k\cdot 0)}$$

which yields, $A = \frac{F_0}{ikT} = \frac{F_0}{i\omega}\left(\frac{v}{T}\right)$ $\left(\because k = \frac{\omega}{v}\right)$

and, $y = \frac{F_0}{i\omega}\left(\frac{v}{T}\right)e^{i(\omega t - kx)}$

Transverse velocity:

$$v_t = \dot{y} = F_0\left(\frac{v}{T}\right)e^{i(\omega t - kx)}$$

where the velocity amplitude $v_t = \frac{F_0}{Z}$, gives transverse impedance

$$Z = \frac{T}{v} = \rho v \quad (\square\ T = rv^2) \ ...(9.25)$$

or *characteristic impedance* of the string $Z = rv$ ... (9.26)

Since the wave velocity v is determined by inertia and the elasticity, the impedance is also dependent on these properties.

9.8 reflection and transmission of waves on a string at a boundary

A string offers a characteristic impedance rv to the waves propagating along it. This is of interest to know how the waves respond to a sudden change of impedance such as at a boundary.

Suppose a string consists of two sections smoothly at a point $x = 0$ (Fig. 9.8) with constant tension T along the whole string. Let the two sections have linear densities r_1 and r_2 and, therefore, the wave velocities in the two sections are $\frac{T}{\rho_1} = v_1^2$ and $\frac{T}{\rho_2} = v_2^2$ while the respective impedance of the two sections are r_1v_1 (= Z_1) and r_2v_2 (= Z_2).

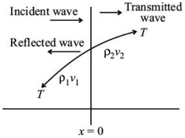


Fig. 9.8: Waves on a string reflected and transmitted at x = 0.

An incident wave travelling along the string meets the discontinuity in impedance at the position $x = 0$. Therefore, at this position, a part of the incident wave is reflected and the remaining part of it is transmitted into the region of impedance $r_2 v_2$.

The incident wave is represented as:

$$y_i = A_1 e^{i(\omega t - k_1 x)}$$

a wave of real amplitude A_1 (not a complex quantity) travelling in the positive x-direction with velocity v_1. The displacement of the reflected wave is represented as:

$$y_r = B_1 e^{i(\omega t + k_1 x)}$$

having amplitude B_1 and travelling in the negative x-direction with velocity v_1.

The transmitted wave displacement is represented as:

$$y_t = A_2 e^{i(\omega t - k_2 x)}$$

having amplitude A_2 and travelling in the positive x-direction with velocity v_2.

In order to find the reflection and transmission amplitude coefficients, i.e., the relative values of B_1 and A_2 with respect to A_1, we apply the conditions which must be satisfied at the impedance discontinuity, at $x = 0$. The boundary conditions applicable at $x = 0$ are:

(*i*) A geometrical condition that the displacement is the same to just left and right of $x = 0$ for all time, so that there is no discontinuity of displacement.

(*ii*) A dynamic condition that there is continuity of the transverse force $T\left(\frac{dy}{dx}\right)$ at $x = 0$, i.e., there is a continuous gradient. This is an essential condition. If it is not so, a finite difference in the force acts on an infinitesimally small mass of the string resulting in an infinite acceleration, which is not permissible.

The boundary condition (*i*) implies:

$$y_i + y_r = y_t$$

$$\text{or, } A_1 e^{i(wt-kx)} + B_1 e^{i(\omega t + k_1 x)} = A_2 e^{i(\omega t - k_2 x)}$$

At $x = 0$, the exponential terms cancel, giving us:

$$A_1 + B_1 = A_2 \quad \text{...(9.27)}$$

The boundary condition (*ii*) implies:

$$T\frac{d}{dx}(y_i + y_r) = T\frac{dyt}{dx}$$

Putting the value of y_i, y_r and y_t, differentiating and using $x = 0$ for all t, we obtain:

$$-k_1 T A_1 + k_1 T B_1 = -k_2 T A_2$$

$$\text{or, } -\frac{\omega T A_1}{v_1} + \frac{\omega T B_1}{v_1} = -\frac{\omega T A_2}{v_2}$$

$$\text{using } \frac{T_1}{v_1} = \rho_1 v_1 = Z_1 \text{ and } \frac{T_2}{v_2} = \rho_2 v_2 = Z_2,$$

we get

$$Z_1(A_1 - B_1) = Z_2 A_2 \quad \text{...(9.28)}$$

Putting Eq. (9.28) in Eq. (9.27), we get

$$Z_1(A_1 - B_1) = Z_2(A_1 + B_1)$$

$$\text{or, } A_1(Z_1 - Z_2) = B_1(Z_1 + Z_2)$$

∴ Reflection coefficient of amplitude,

$$\frac{B_1}{A_1} = \frac{Z_1 - Z_2}{Z_1 + Z_2} \quad \text{...(9.29)}$$

From Eq. (9.27), $B_1 = A_2 - A_1$. Substituting this in Eq. (9.28), we get:

$$Z_1(A_1 - A_2 + A_1) = Z_2 A_2$$

$$\text{or, } Z_1(2A_1 - A_2) = Z_2 A_2$$

$$\text{or, } 2A_1 Z_1 = A_2(Z_1 + Z_2)$$

∴ Transmission coefficient of amplitude,

$$\frac{A_2}{A_1} = \frac{2Z_1}{Z_1 + Z_2} \quad \text{...(9.30)}$$

It is clear that these coefficients are independent of ω, i.e., the relations hold for waves of all frequencies. They are real and, therefore, free from phase changes other than that of π radian which will change the sign of a term. The ratios depend on the values of impedances. If $Z_2 = \infty$, i.e., $x = 0$ is a fixed end of the string and no transmission of wave possible. So that,

$$\frac{B_1}{A_1} = -1$$

$$\text{or, } B_1 = -A_1 \Rightarrow A_1 = -B_1$$

implying that the wave is completely reflected with a phase change of π radian (phase reversal) which is essential for the formation of standing wave. As shown in Fig. 9.9, a group of waves having many component frequencies will retain shape upon reflection at $Z_2 = \infty$, but suffer phase inversion.

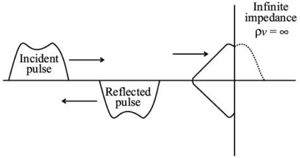


Fig. 9.9: A phase change of p radian in a pulse on getting reflected at an infinite impedance.

If $Z_2 = 0$, i.e., $x = 0$ is a free end of the string, then

$$\frac{B_1}{A_1} = 1 \text{ and } \frac{A_2}{A_1} = 2$$

This explains the flick at the end of a whip or free ended string, when a wave reaches it.

9.9 reflection and transmission of energy

In this section, we consider changes in energy when waves meet a boundary between two media having different impedance values. Let r be the mass per unit length of the string and the string is a simple harmonic oscillator of maximum amplitude A, so that its total energy is:

$$E = \frac{1}{2}\rho\,\omega^2 A^2$$

where, ω is the wave frequency.

The wave is propagating with velocity v so that each unit of length of the string takes up the oscillation with the passage of the wave. The rate at which energy is being transferred is:

$$\text{energy} \times \text{velocity} = \frac{1}{2}\rho\,\omega^2 A^2 v$$

Thus, the rate of energy arrival at the boundary x = 0 is the energy arriving with the incident wave, that is:

$$\frac{1}{2}\rho\omega^2 A_1^2 v \text{ or } \frac{1}{2} Z_1 \omega^2 A^2$$

The rate at which energy leaves the boundary through reflected and transmitted waves, is:

$$\frac{1}{2}\rho_1 v_1 \omega^2 B_1^2 + \frac{1}{2}\rho_2 v_2 \omega^2 A_2^2$$

$$\text{or, } \frac{1}{2} Z_1 \omega^2 B_1^2 + \frac{1}{2} Z_2 \omega^2 A_2^2$$

which from the ratio $\frac{B_1}{A_1}$ and $\frac{A_2}{A_1}$ becomes:

$$= \frac{1}{2}\omega^2 A_1^2 \frac{Z_1(Z_1 - Z_2)^2 + 4Z_1^2 Z_2}{(Z_1 + Z_2)^2}$$

$$= \frac{1}{2} Z_1 \omega^2 A_1^2$$

which is equal to the incident energy. Thus, the energy is conserved; energy arriving as incident waves is equal to the energy leaving as reflected and transmitted waves.

Reflected and Transmitted Intensity Coefficients: These are given by the following relations:

$$\frac{\text{Reflected energy}}{\text{Incident energy}} = \frac{Z_1 B_1^2}{Z_1 A_1^2} = \left(\frac{B_1}{A_1}\right)^2 = \left(\frac{Z_1 - Z_2}{Z_1 + Z_2}\right)^2 \quad \text{...(9.31)}$$

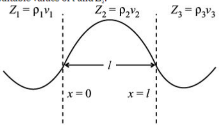
$$\frac{\text{Transmitted energy}}{\text{Incident energy}} = \frac{Z_2 A_2^2}{Z_1 A_1^2} = \frac{4Z_1 Z_2}{(Z_1 + Z_2)^2} \quad \text{...(9.32)}$$

It is clear from Eq. (9.31) that if $Z_1 = Z_2$, no energy is reflected, i.e., the entire incident energy is transmitted. As $Z_1 = Z_2$. The *impedances* are said to be matched.

9.10 Impedance matching

Impedance matching is an imported practical consideration in the process of energy transfer of any kind. Long distance cables carrying electrical energy must be perfectly (impedance) matched at all the joints to avoid wastage of energy from reflection. The power transfer from any generator is maximum when the load matched the generator impedance. A loudspeaker is matched to the impedance of an amplifier, by choosing the correct turn ratio of the coupling transformer. This method of using a suitable coupling element to match the impedance is of fundamental importance in engineering, physics, electronics and optics. We take here the example of waves on a string, but the results are valid for all wave systems.

We have discussed earlier that when a smooth joint exists between two strings of different impedances, energy is reflected at the boundary. We are now going to see that the insertion of a particular length of another string between the two mismatched strings matches the impedance and helps in eliminating energy reflection. In Fig. 9.10, we want to match the impedances $Z_1 = r_1 v_1$ and $Z_3 = r_3 v_3$ by inserting a string of length l having impedance $Z_2 = r_2 v_2$. Our problem here is to find the suitable values of l and Z_2.



$y_i = A_1 e^{i(\omega t - k_1 x)}$ $y_i = A_2 e^{i(\omega t - k_2 x)}$ $y_i = A_3 e^{i[\omega t - k_3 x - l]}$

$y_r = B_1 e^{i(\omega t + k_1 x)}$ $y_r = B_2 e^{i(\omega t + k_2 x)}$

Fig. 9.10

Figure 9.10 depicts the incident, reflected and transmitted displacements at the junctions $x = 0$ and $x = l$. Our objective is to make the ratio

$$\frac{\text{Transmitted energy}}{\text{Incident energy}} = \frac{Z_3 A_3^2}{Z_1 A_1^2}$$

equal to *unity*. The boundary conditions are that y and $T\left(\frac{dy}{dx}\right)$ are continuous across the junctions $x = 0$ and $x = l$. The continuity of the displacement y across Z_1 and Z_2 requires:

$$A_1 e^{i(\omega t - k_1 x)} + B_1 e^{i(\omega t + k_1 x)} = A_2 e^{i(\omega t - k_2 x)} + B_2 e^{i(\omega t + k_2 x)}$$

At $x = 0$, the exponential terms cancel giving:

$$A_1 + B_1 = A_2 + B_2 \quad ...(9.33)$$

Similarly, the continuity of $T\left(\frac{dy}{dx}\right)$ at $x = 0$, on differentiation, gives

$$T(-ik_1A_1 + ik_1B_1) = T(-ik_2A_2 + ik_2B_2)$$

Dividing the above equation throughout by ω and using the identity, $T\left(\frac{k}{\omega}\right) = \frac{T}{v} = \rho v = Z,$ we get:

$$-iZ_1A_1 + iZ_1B_1 = -iZ_2A_2 + iZ_2B_2$$

multiplying throughout by i, we get

$$Z_1(A_1 - B_1) = Z_2(A_2 - B_2) \quad ...(9.34)$$

Similarly, at $x = l$ the continuity of y requires:

$$A_2 e^{i(\omega t - k_2 x)} + B_2 e^{i(\omega t + k_2 x)} = A_3 e^{i[\omega t - k_3(x - l)]}$$

Putting $x = l$, we have:

$$A_2 e^{-ik_2 l} + B_2 e^{ik_2 l} = A_3 \quad ...(9.35)$$

The continuity of $T\left(\frac{dy}{dx}\right)$ at $x = l$ gives:

$$Z_2\left(A_2 e^{-ik_2 l} - B_2 e^{-ik_2 l}\right) = Z_3 A_3 \quad ...(9.36)$$

Substituting $B_1 = A_2 + B_2 - A_1$ from Eq. (9.33) in Eq. (9.34) we get:

$$Z_1(A_1 - A_2 - B_2 + A_1) = Z_2(A_2 - B_2)$$

$$\text{or, } A_1 = \frac{A_2(r_{12} + 1) + B_2(r_{12} - 1)}{2r_{12}} \quad ...(9.37)$$

$$\text{where, } r_{12} = \frac{Z_1}{Z_2}$$

Eqs. (9.35) and (9.36) give us

$$A_2 = \frac{r_{23} + 1}{2r_{23}} A_3 e^{ik_2 l} \quad ...(9.38)$$

$$\text{and, } B_2 = \frac{r_{23} - 1}{2r_{23}} A_3 e^{-ik_2 l} \quad ...(9.39)$$

$$\text{where } r_{23} = \frac{Z_2}{Z_3}$$

Equations (9.37) and (9.38) give

$$A_1 = \frac{A_3}{4r_{12}r_{23}}[(r_{12} + 1)(r_{23} + 1)e^{ik_2 l} + (r_{12} - 1)(r_{23} - 1)e^{-ik_2 l}$$

using $r_{12}r_{23} = \frac{Z_1Z_2}{Z_2Z_3} = \frac{Z_1}{Z_3} = r_{13},$ A_1 can be expressed as:

$$A_1 = \frac{A_3}{4r_{13}}[(r_{13} + 1)(e^{ik_2 l} + e^{-ik_2 l})(r_{12} + r_{23})(e^{ik_2 l} - e^{-ik_2 l})]$$

=

$$\frac{A_3}{2r_{13}}\left[(r_{13}+1)\cos k_2 l + i(r_{12}+r_{23})\sin k_2 l\right]$$

$$\therefore \left(\frac{A_3}{A_1}\right)^2 = \frac{4r_{13}^2}{(r_{13}+1)^2\cos^2 k_2 l + (r_{12}+r_{23})^2\sin^2 k_2 l}$$

or,

$$\frac{\text{Transmitted energy}}{\text{Incident energy}} = \frac{Z_3 A_3^2}{Z_1 A_1^2} = \frac{1}{r_{13}} = \frac{A_3^2}{A_1^2}$$

$$= \frac{4r_{13}}{(r_{13}+1)^2\cos^2 k_2 l + (r_{12}+r_{23})^2\sin^2 k_2 l}$$

Assuming, $l = \frac{\lambda_2}{4}, \cos k_2 l = 0$ and $\sin k_2 l = 1$, we get

$$\frac{Z_3 A_3^2}{Z_1 A_1^2} = \frac{4r_{13}}{(r_{12}+r_{23})^2}$$

When $r_{12} = r_{13}$, we have

$$\frac{Z_3 A_2^2}{Z_1 A_1^2} = 1$$

That is our objective of making $\frac{Z_3 A_3^2}{Z_1 A_1^2}$ equal to unity is satisfied if:

$$r_{12} = r_{23}$$

$$\text{or, } \frac{Z_1}{Z_2} = \frac{Z_2}{Z_3}$$

$$\text{or, } Z_2 = \sqrt{Z_1 Z_3} \quad ...(9.40)$$

That is, *if the impedance of the coupling medium is the harmonic mean of the two impedances to be matched and the thickness of the coupling medium is* $\frac{\lambda_2}{4}$ *where* $\lambda_2 = \frac{2\pi}{k_2}$, *then all the energy at frequency ω will be transmitted with zero reflection.* Optical lenses are coated with a thin dielectric layer of thickness one quarter of wavelength of the light to be used to eliminate reflections of the incident light in the optical instruments. Transmission lines are matched to loads by inserting quarter wavelength stubs of lines with the appropriate impedance.

9.11 superposition of waves: standing waves

According to the principle of superposition, *when two or more waves propagating through a medium simultaneously overlap, the resultant displacement at any point is the vector sum of the displacement due to individual waves.*

That is, $y(x, t) = y_1(x, t) + y_2(x, t) + y_3(x, t) + ...$

Here, $y(x, t)$ is the resultant displacement at the position x and instant t while $y_1(x, t)$, $y_2(x, t)$, $y_3(x, t)$,... are the displacements due to the waves 1, 2, 3,... respectively, at the same location x and the same instant of time t.

This principle depends upon the linearity of the wave equation and the corresponding linear-combination property of its solution, it is also called the *principle of linear superposition*. This principle is of central importance and applies to all types of waves, including electromagnetic waves.

Standing Waves on a String of Fixed Length: We have seen that a progressive wave is completely reflected at an infinite impedance and upon reflection, phase reversal or a phase change of π occurs. A string of fixed length l with both its ends being rigidly clamped is equivalent to both ends having infinite impedance.

Consider a monochromatic wave of frequency ω and amplitude a propagating along $+x$-direction and another wave of amplitude b and frequency ω travel in the $-x$-direction. Applying the principle of superposition, the displacement on the string at any instant t is given by:

$$y = ae^{i(wt - kx)} + be^{i(wt + kx)}$$

Here, the boundary condition is $y = 0$ at $x = 0$ and $x = l$ for all time. Using the boundary condition $y = 0$ at $x = 0$, gives:

$$0 = (a + b)e^{iwt}$$

$$\text{or, } a + b = 0 \Rightarrow a = -b$$

This implies the fact that a wave in either direction meeting the infinite impedance at either end is completely reflected with a π phase change in amplitude.

This is a general result for all wave shapes and frequencies.

Thus, using $a = -b$, we have:

$$y = ae^{iwt} = (e^{-ikx} - e^{ikx})$$

$$= (-2i)ae^{-iwt}\sin kx \quad ...(9.41)$$

This equation satisfies the standing (or stationary) wave time independent form of the wave equation:

$$\frac{d^2 y}{dx^2} + k^2 y = 0$$

Since $\frac{1}{v^2}\frac{d^2 y}{dt^2}$ can easily be seen to give $\left(-\frac{\omega^2}{v^2}\right)y$, i.e., $\frac{1}{v^2}\frac{d^2y}{dt^2} = \left(-\frac{\omega^2}{v^2}\right)y = -k^2y$

Using the boundary condition $y = 0$ at $x = l$ for all t, gives

$$\sin kl = \sin\frac{\omega l}{v} = 0$$

or, $\frac{\omega l}{v} = n\pi$, π = 1, 2, 3,...

This gives the values of *allowed frequencies,*

$$\omega_n = \frac{n\pi v}{l}$$

or, $2\pi\nu_n = \frac{n\pi v}{l}$

or, $n_n = \frac{nv}{2l} = \frac{v}{\lambda_n}$ (q v = nl)

$$\therefore l = \frac{n\lambda_n}{2} \quad ...(9.42)$$

Thus, $$\sin\frac{\omega_n x}{v} = \sin\frac{n\pi x}{l} \quad ...(9.43)$$

These frequencies (ω_n, n = 1, 2, 3,...) are called *normal frequencies* or *mode of vibration,* also referred to as eigen frequencies particularly in wave mechanics. Such allowed frequencies give the length of the string as an integral multiple of half wavelength.

For n = 1, the fundamental mode of vibration, from Eq. (9.42), we have:

$$l = \frac{\lambda_1}{2} \Rightarrow \lambda_1 = 2l$$

∴ Velocity of the wave $v = nl_1 = 2nl$

Similarly, for n = 2, i.e., *2nd* harmonic

$$l = \frac{2\lambda_2}{2} \Rightarrow l = \lambda_2$$

for n = 3, i.e., 3rd harmonic,

$$l = \frac{3\lambda_3}{2} \quad \lambda_3 = \frac{2}{3}l$$

for n = 4, i.e., 4th harmonic,

$$l = \frac{4\lambda_4}{2} \Rightarrow \lambda_4 = \frac{l}{2}$$

The string displacement for the first four harmonics are shown in Fig. 9.11. As observed in this figure, for harmonics $n > 1$, in addition to the two end points, there are other positions which are at rest. These positions are called *nodal points* or simply *nodes* of the standing waves. The nodal points occur when:

$$\sin\frac{\omega_n x}{v} = \sin\frac{n\pi x}{l} = 0$$

or, $\frac{n\pi x}{l} = rx$ (r = 0, 1, 2, 3, ...n)

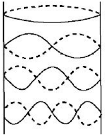


Fig. 9.11

When value r = 0 and r = n correspond to x = 0 and x = l respectively, and in between the two ends, there are n − 1 positions equally spaced along the string in the nth harmonic where the displacement is zero, then the standing wave is formed. Standing waves are formed, when a particular mode (or harmonic) is excited and the incident and reflected waves are superposed. If the amplitudes of these progressive waves are equal and opposite, nodal points come into existence. If the nodal points exist, it is confirmed that waves travelling in opposite directions are exactly the same in all respects, so that the energy carried in one direction is exactly equal to that carried in the other direction. That is, the net energy flux or the energy carried across unit area per second in a standing wave system is zero.

However, sometimes the reflection of waves may not be complete and the waves in the opposite directions may not cancel each other completely. Therefore, the formation of nodal points is not perfect. In such cases, a term called *standing wave ratio* is of relevance and is defined below:

Standing Wave Ratio: When a progressive wave is partially reflected, the reflection coefficient $\frac{B_1}{A_1}$, denoted by r, clearly $r < 1$. The standing waves formed in such situations will have maximum amplitude $A_1 + B_1$ at the reinforcement (or antinodal) points and minimum amplitude $A_1 - B_1$ at the cancellation (or nodal) points. In such cases, we define a parameter called *standing wave ratio* which is the ratio of maximum to minimum amplitudes in the standing wave. Thus,

Standing wave ratio =

$$\frac{A_1+B_1}{A_1-B_1}=\frac{1+B_1/A_1}{1-B_1/A_1}=\frac{1+r}{1-r}$$

Also, by measuring the maximum and minimum amplitudes, we can determine coefficient:

$$r=\frac{B_1}{A_1}=\frac{SWR-1}{SWR+1}$$

where, SWR stands for standing wave ratio.

9.12 different modes of standing waves in a stretched string

Suppose a string of length l is fixed at both ends so that it is held tightly along x-axis in undisturbed condition. If it is plucked at any point, waves are generated which travel to the end points, get reflected and form standing or stationary waves.

Depending upon the point at which the string is plucked, it can be set into different modes of vibration, i.e., stationary waves of different wavelength and frequencies may be produced. The two fixed end points, however, in all the modes remain nodal points or displacement nodes. If the string is *plucked in the middle, it vibrates in one segment giving the fundamental note* or *the lowest frequency* [Fig. 9.12(*a*)]. If the string is pressed in the middle and plucked at the mid-point of either half, it vibrates in two segments [Fig. 9.12(*b*)] producing a note of *twice the frequency of the fundamental note*. Similarly, by plucking at a suitable point a note of frequency multiple of fundamental note can be produced.

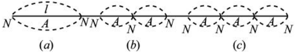


Fig. 9.12

We obtain these results analytically.

Let a simple harmonic wave $y_1=a\sin\frac{2\pi}{\lambda}(vt-x)$ be propagating along $+x$-direction which on getting reflected at the fixed end, starts travelling along $-x$-direction represented as $y_2=-a\sin\frac{2\pi}{\lambda}(vt+x)$.

The resulting stationary wave can be represented as:

$$y = y_1 + y_2$$

$$= a\sin\frac{2\pi}{\lambda}(vt-x)-a\sin\frac{2\pi}{\lambda}(vt+x)$$

$$\text{or, } y = -2a\sin\frac{2\pi x}{\lambda}\cos\frac{2\pi vt}{\lambda}$$

If the length of the string is l, then putting $y = 0$ at $x = -l$, we get:

$$\therefore 0 = 2a\sin\frac{2\pi l}{\lambda}\cos\frac{2\pi vt}{\lambda}$$

Since the displacement at the fixed end is always zero, the above relations are valid for all values of t. Therefore,

$$\sin\frac{2\pi l}{\lambda} = 0$$

$$\text{or, } \frac{2\pi l}{\lambda} =$$

$$n\lambda \Rightarrow \lambda=\frac{2l}{n},\quad n=1, 2, 3, 4,\ldots$$

∴ Frequency, of the vibrating string $\nu=\frac{v}{\lambda}=\frac{v}{2l/n}=\frac{nv}{2l}$

If m is the mass per unit length, then velocity $v=\sqrt{\frac{T}{m}}$.

∴ Frequency, n = $\frac{n}{2l}\sqrt{\frac{T}{m}}$

For *fundamental* mode or *first harmonic* $n = 1$, $\lambda = 2l$

\ Frequency n_1 = $\frac{1}{2l}\sqrt{\frac{T}{m}}$

which is the lowest frequency of the vibration.

For *second harmonic* of *first overtone*, $n = 2$, $\lambda = l$

∴ Frequency, n_2 = $\frac{1}{l}\sqrt{\frac{T}{m}}$

which is twice the frequency of the fundamental tone.

For *third harmonic* or *second overtone* $n=3,\ \lambda=\frac{2l}{3}$

∴ Frequency n_3 = $\frac{3}{2l}\sqrt{\frac{T}{m}}$

which is thrice the frequency of the fundamental tone. Similarly, we can find the frequencies of fourth harmonic or third overtone, and so on.

9.13 group and phase velocity

A plane progressive wave propagating along positive x-axis is represented as:

$$y = a\sin\frac{2\pi}{\lambda}(vt - x)$$

where v is the wave velocity. We can also write the above expression as:

$$y = a\sin\left(\frac{2\pi vt}{\lambda} - \frac{2\pi x}{\lambda}\right)$$

As, $\frac{v}{\lambda} = \nu,$ the frequency of the wave and $\frac{2\pi}{\lambda} = k,$ the propagation constant, also called *wave number*, we can write:

$$y = a\sin(2\pi\nu t - kx) \quad ...(9.44)$$

This equation implies that the constant phase $(\omega t - kx)$ of the wave travels along the positive direction of the x-axis with velocity $\frac{dx}{dt}$, i.e., the phase velocity of the wave is $\frac{dx}{dt}$.

Since the term $\omega t - kx$ is constant for a wave, on differentiating this with respect to time, we get

$$\omega - k\left(\frac{dx}{dr}\right) = 0 \Rightarrow k\frac{dx}{dt} = \omega$$

∴ Phase velocity of the wave $\frac{dx}{dt} = \frac{\omega}{k}$ ...(9.45)

Also, $\frac{\omega}{k} =$

$$\frac{2\pi\nu}{2\pi/\lambda} = \nu\lambda = \frac{\lambda}{T} = \text{wave velocity } v.$$

Thus, for a *single wave in a medium,*

$$\text{wave velocity} = \text{phase velocity} = v = \frac{\omega}{k} = \frac{\lambda}{T}$$

Let us now consider two (or more) wave trains propagating simultaneously in the same direction in a medium. Suppose, the two wave trains are of wavelengths λ and $\lambda + d\lambda$ and their amplitudes are the same, equal to a. We also assume that the medium is not dispersive so that the wave velocity v is independent of wavelength, i.e., both the waves travel with same velocity v. Since both the waves are travelling along the same path and in the same direction, they superpose each other.

The points where the positive (or negative) maximum displacements of the two waves meet, they reinforce and a maxima of resulting waves are formed there. While the points where the positive maximum of one meet the negative maximum of the other, zero displacement of the resulting wave results. In between these two extremes, the displacement of the resulting wave has intermediate values. The resulting wave is shown in Fig. 9.13, which is similar to beat wave, consists of a *group of waves* with the maximum displacements at the centre of the group and falling off gradually to zero on either side. The velocity with which the maxima of these groups of waves travel, is called the *group velocity* (u) of the resultant wave.

In a non-dispersive medium the resulting wave consisting of group of wave also travels with the same velocity, as that of each constituent wave and the shape or profile of the groups, therefore, remains intact. In other words, *in a non-dispersive medium, the group velocity is equal to the phase velocity of the wave.*

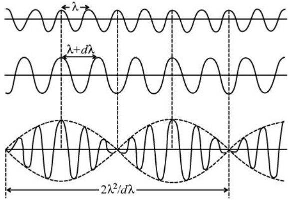


Fig. 9.13

However, for a dispersive medium, the group velocity u is not equal to the wave velocity v and they are related as:

$$u = v - \lambda\frac{dv}{d\lambda} \quad ...(9.46)$$

In a non-dispersive medium, $\frac{dv}{d\lambda} = 0,$ i.e., there is no change in velocity with wavelength, therefore $u = v$. Hence, in such a medium, wave groups propagate keeping their form of profile intact, e.g. sound waves in any homogeneous medium or light waves in vacuum. On the other hand, in a dispersive medium, with *normal dispersion* in general, wave velocity increases with wavelength, or $\frac{dv}{d\lambda}$ is positive. Therefore, *the group velocity in general, is less than the wave (or phase) velocity*. If the dispersive medium exhibits, *anomalous dispersion*, $\frac{dv}{d\lambda}$ is negative and $u > v$.

9.14 longitudinal (sound) waves

In the longitudinal waves the particles of the medium oscillate in the direction of the wave propagation. Sound wave is the most common example of longitudinal waves. These waves can propagate in all phases of matter–gases, liquids, solids and plasma. However, we will confine our discussion to their propagation in a gaseous medium.

Sound Wave in a Gaseous Medium: In a gaseous medium, only longitudinal waves can propagate as a gas or a liquid cannot sustain the transverse shear necessary for the propagation of transverse waves. We know, the sound waves propagate in the form of *compressions* (particles come closer) and rarefactions (particles move apart). Naturally, as the sound waves propagate, variation in pressure and density of the medium takes place.

Let us consider a given (constant) mass of a gas at equilibrium pressure P_o, volume V_o and density r_o. Suppose, when the waves propagate the pressure changes to $P = P_o + p$, the volume becomes equal to $V = V_o + v$ while the new density becomes $r = r_o + r_d$. Let the maximum pressure amplitude be denoted by P_m and dP be the fluctuating components superimposed on the equilibrium pressure P_o.

When the sound waves propagate, the changes in the medium are of an extremely small order and sets the limit within which the wave equation is appropriate. The fractional volume change

$$\frac{v}{V_0} = \delta,$$

is called the *dilation* while the fractional change in density

$$\frac{\rho_d}{\rho} = s$$

is called the *condensation*. For normal sound waves, the value of Δ and s are of the order of 10^{-3} and a value of $P_m = 2 \times 10^{-5}$ Nm^{-2} ($\sim 10^{-10}$ atmosphere) gives a sound wave which is audible at 1 kHz.

As the mass of the gas is constant.

$$r_o V_o = rV = r_o V_o (1 + \Delta)(1 + s)$$

$$\text{or, } (1 + \Delta)(1 + s) = 1$$

giving to a close approximation $d = -s$. ...(9.47)

That is, *dilation is equal and opposite to condensation.*

The elastic property of a gas is a measure of its compressibility and is defined in terms of its bulk modulus B, given by:

$$B = -\frac{dP}{dV/V} = -V\frac{dP}{dV} \quad ...(9.48)$$

Thus, the bulk modulus is the ratio of change in pressure for a fractional change in volume and the negative sign indicates that an increase in volume is accompanied by a fall in pressure. Laplace had correctly argued that when sound waves propagate through a gas, the compressions and rarefactions succeed each other rapidly. As gas (air) is a poor conductor of heat the equalisation of temperature is not possible across the region of compressions and rarefactions, i.e., the condition is not isothermal. However, the total heat content of the system remains constant, i.e., *the condition is adiabatic* in the medium during the propagation of sound waves.

For adiabatic changes,

$$PV^g = \text{constant} \quad ...(9.49)$$

where g is the ratio of specific heats at constant pressure and constant volume.

Differentiating Eq. (9.49), we get

$$V^g dP + gPV^{g-1} dV = 0$$

$$\text{or, } V^g dP = -gPV^{g-1} dP \Rightarrow \frac{V^\gamma}{V^{\gamma-1}}\frac{dP}{dV} = -\gamma P$$

$$\text{or, } -V\frac{dP}{dV} = gP = B_a \quad ...(9.50)$$

The subscript a denotes adiabatic. As for a given gas, its bulk modulus is constant, the elastic property of the gas gP is also a constant. We know $P = P_o + p$ or $dP = p$, the excess pressure, giving:

$$B_a = -\frac{\rho}{(v/v_0)} \text{ or } p = -B_a\delta = B_a s \quad ...(9.51)$$

Wave Equation: In a sound wave, the particle displacements and velocities are along the direction of propagation, say along the direction of propagation, say along x-axis and we will denote the displacement by h(x, t).

Let us consider the motion of an element of the gaseous medium of thickness Δx and unit cross-section under the influence of the sound waves. The displacements of the element are depicted in Fig. 9.14. The particles in the layer x are displaced by a distance h while those at $x + \Delta x$ are displaced by a distance h + Dh, Thus, the increase in thickness by Δx of the element of unit cross-section, which equals the increase in volume, is:

$$\text{Dh} = \frac{d\eta}{dx}\Delta x$$

$$\text{and, } \Delta = \frac{v}{V_0} = \left(\frac{d\eta}{dx}\right)\frac{\Delta x}{\Delta x} = \frac{d\eta}{dx} = -s$$

where $\frac{d\eta}{dx}$ is the *strain*.

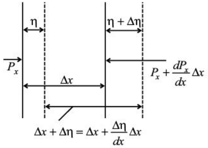


Fig. 9.14

Because the pressures along the x-axis on either side of the thin element are not balanced, the medium gets deformed. The net force acting on the element is given by:

$$P_x - P_{x+Dx} = \left[P_x - \left(P_x + \frac{dP_x}{dx}\Delta x\right)\right] = -\frac{dP_x}{dx}\Delta x$$

$$= -\frac{d}{dx}(P_0 + p)\Delta x = -\frac{dp}{dx}\Delta x$$

The mass of the considered element is equal to $r_0\Delta x$ and its acceleration is $\frac{d^2\eta}{dt^2}$. Using Newton's second law of motion, we have:

$$-\frac{dp}{dx}\Delta x = \rho_0 \Delta x \frac{d^2\eta}{dt^2} \quad \ldots(9.52)$$

where, $p = -B_a\delta = -\beta_a \frac{d\eta}{dx}$

So that, $-\frac{dp}{dx} = B_a \frac{d^2\eta}{dx^2}$ ...(9.53)

From Eqs. (9.52) and (9.53), we have:

$$B_a \frac{d^2\eta}{dx^2} = \rho_0 \frac{d^2\eta}{dt^2} \quad \ldots(9.54)$$

As, $\frac{B_a}{\rho_0} = \frac{\gamma\rho}{\rho_0}$ is the ratio of the elasticity to the inertia or density of the gas this ratio has the dimension:

$$\frac{\text{force}}{\text{area}}\frac{\text{volume}}{\text{mass}} = (\text{velocity})^2$$

$$\therefore \frac{\gamma\rho}{\rho_0} = v^2$$

or speed of sound wave,

$$v = \sqrt{\frac{\gamma \text{ pressure}}{\text{density}}} = \sqrt{\frac{\text{adiabatic elasticity}}{\text{density}}} \quad \ldots(9.55)$$

From Eq. (9.54), we have:

$$\frac{d^2\eta}{dx^2} = \frac{1}{v^2}\frac{d^2\eta}{dt^2} \quad \ldots(9.56)$$

is the *wave equation* of the propagating sound wave:

Denoting the maximum amplitude displacement by h_m, for a wave propagating in the position x-direction, we have

Displacement, $h = h_m e^{i(wt - kx)}$ ...(9.57)

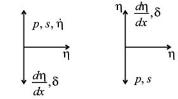
$\therefore$ Particle velocity, $\dot{\eta} = \frac{d\eta}{dt} = i\omega\eta_m e^{i(\omega t - kx)} = i\omega\eta$ ...(9.58)

Condensation, $\Delta = \frac{d\eta}{dx} = -ik\eta = -s$

or dilation, $s = ikh$

and excess pressure, $p = B_a s = iB_a kh$ ...(4.59)

The phase relationships between these parameters is shown in Fig 9.15, when the wave is propagating in $+x$ direction. The excess pressure p, the fractional density increase and the particle velocity $\dot{\eta}$ are all $\pi/2$ radian ahead in phase compared to displacement h. On the other hand, the volume change (π radian out of phase with density change) is $\pi/2$ radian behind the displacement.



(*a*) Wave in +ve x direction (*b*) Wave in –ve x direction

Fig. 9.15: Phase relationships between the particle displacemant h, particle velocity $\dot{\eta}$, excess pressure p and condensation $s = -\Delta$ (dilation).

In the case of wave propagating along negative x-direction, the above expression becomes:

displacement, $h = h_m e^{i(wt + kx)}$

particle velocity, $\dot{\eta} = \frac{d\eta}{dt} = -ik\eta = -s \quad \text{or} \quad s = ik\eta$ ...(9.60)

excess pressure, $p = B_a s = -iB_a kh$ ...(9.61)

The particle displacement in both the cases is measured in the x-direction and the thin element of the gaseous medium Δx, oscillates about the value h = 0 which defines its *equilibrium* or *central position*. For a wave in the positive x-direction, the value h = 0 with $\dot{\eta}$ a maximum in the positive x-direction gives a *maximum positive excess pressure* (or *compression*) with a maximum condensation s_m (maximum density) and minimum volume. For a wave propagating in negative x-

Displacement, h = $h_m e^{(wt-kx)}$ ...(9.57)

∴ Particle velocity, $\dot{\eta}$ =

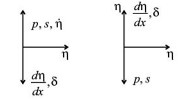
$$\frac{d\eta}{dt} = i\omega\eta_m e^{i(\omega t-kx)} = i\omega\eta \quad ...(9.58)$$

Condensation, Δ = $\dfrac{d\eta}{dx} = -ik\eta = -s$

or dilation, $s = ik$h

and excess pressure, $p = B_a s = iB_a k$h ...(4.59)

The phase relationships between these parameters is shown in Fig 9.15, when the wave is propagating in +x direction. The excess pressure p, the fractional density increase and the particle velocity $\dot{\eta}$ are all π/2 radian ahead in phase compared to displacement h. On the other hand, the volume change (π radian out of phase with density change) is π/2 radian behind the displacement.



(*a*) Wave in +ve x direction (*b*) Wave in –ve x direction

Fig. 9.15: Phase relationships between the particle displacemant h, particle velocity $\dot{\eta}$, excess pressure p and condensation $s = -\Delta$ (dilation).

In the case of wave propagating along negative x-direction, the above expression becomes:

displacement, h = $h_m e^{i(wt+kx)}$

particle velocity, $\dot{\eta}$ =

$$\frac{d\eta}{dt} = -ik\eta = -s \quad \text{or} \quad s = ik\eta \quad ...(9.60)$$

excess pressure, $p = B_a s = -iB_a k$h ...(9.61)

The particle displacement in both the cases is measured in the x-direction and the thin element of the gaseous medium Δx, oscillates about the value h = 0 which defines its *equilibrium* or *central position*. For a wave in the positive x-direction, the value h = 0 with η a maximum in the positive x-direction gives a *maximum positive excess pressure* (or *compression*) with a maximum condensation s_m (maximum density) and minimum volume. For a wave propagating in negative x-direction, the valve h = 0 with $\dot{\eta}$ maximum in the positive x-direction gives a *maximum negative excess pressure* (or *rarefaction*), a maximum volume and a minimum density.

9.15 acoustic impedance

The impedance offered by a medium to a wave through which it is propagated is given by:

specific acoustic impedance

$$= \frac{\text{excess pressure}}{\text{particle velocity}} = \frac{p}{\dot{\eta}} \quad ...(9.62)$$

which is equal to the ratio of a force per unit area to the velocity. We know, for a wave propagating along positive x direction,

$p = B_a s = iB_a k$h and $\dot{\eta} = i\omega\eta$

$$\therefore \frac{P}{\dot{\eta}} = \frac{iB_a k\eta}{i\omega\eta} = \frac{B_a k}{\omega} = \frac{B_a}{v} = \rho_0 v \quad ...(9.63)$$

Thus, *acoustic impedance* offered by a medium to the wave is *equal to the product of density and the wave velocity* and thus depends on the elasticity and inertia of the medium.

For a wave propagating in negative x-direction, the specific acoustic impedance is given by:

$$\frac{P}{\dot{\eta}} = -\frac{iB_a k\eta}{i\omega\eta} = -r_0 v \quad ...(9.64)$$

Thus, there is change of sign due to the changed phased relationship. The unit of $r_0 v$ is kg m^{-2} s^{-1} and for air, its value is nearly 400 kg m^{-2} s^{-1}, for water, 1.45 × 10^6 kg m^{-2} s^{-1} and steel 3.9 × 10^7 kg m^{-2} s^{-1}.

For plane sound waves, the specific acoustic impedance $r_0 v$ is a real quantity. However, for spherical waves, it has an added reactive component $\dfrac{ik}{r}$ for spherical waves, where r is the distance travelled by the wavefront. Naturally, as r becomes large, i.e., as the spherical waves travel and take the shape of plane wavefront, the contribution of this reactive term becomes insignificant.

9.16 reflection and transmission of sound waves

When a sound wave propagating in medium encounters a boundary separating two media of different accoustic impedances, the following two boundary conditions must be met with respect to the reflection and transmission of the wave. (*i*) The particle velocity $\dot{\eta}$ (*ii*) the acoustic excess pressure p, are continuous across the boundary. Physically, this ensures that the two media are in complete contact throughout the boundary.

As shown in Fig. 9.16, suppose a plane sound wave is propagating a medium of specific acoustic impedance $Z_1 = r_1 v_1$ and is incident normally at an infinite plane boundary, separating this medium from another medium of specific acoustic impedance $Z_2 = r_2 v_2$. *The* above-mentioned boundary conditions require.

$$\dot{\eta}_i + \dot{\eta}_r = \dot{\eta}_t \quad ...(9.65)$$

and, $p_i + p_r = p_t$ ...(9.66)

where the subscripts i, r ant t denote incident, reflected and transmitted waves.

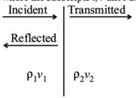


Fig. 9.16

For the incident wave $p_i = \rho_1 v_1 \dot{\eta}_i$ and for the reflected wave $p_r = -\rho_1 v_1 \dot{\eta}_r$ so Eq. (9.65) becomes

$$\rho_1 v_1 \dot{\eta}_i - \rho_1 v_1 \dot{\eta}_r = \rho_2 v_2 \dot{\eta}_t$$

$$\text{or, } Z_1 \dot{\eta}_i - Z_1 \dot{\eta}_r = Z_2 \dot{\eta}_t \quad ...(9.67)$$

Eliminating $\dot{\eta}_t$ from Eqs. (9.65) and (9.67), we get:

$$Z_1 \dot{\eta}_i - Z_1 \dot{\eta}_r = Z_2(\dot{\eta}_i + \dot{\eta}_r)$$

$$\text{or, } Z_1 \dot{\eta}_i - Z_2 \dot{\eta}_i = Z_1 \dot{\eta}_r + Z_2 \dot{\eta}_r$$

$$\text{or, } \dot{\eta}_i (Z_1 - Z_2) = \dot{\eta}_r (Z_1 + Z_2)$$

$$\text{or, } \frac{\dot{\eta}_r}{\dot{\eta}_i} = \frac{\omega \eta_r}{\omega \eta_i} = \frac{\eta_r}{\eta_i} = \frac{Z_1 - Z_2}{Z_1 + Z_2} \quad ...(9.68)$$

Similarly, eliminating $\dot{\eta}_r$ from Eqs. (9.65) and (9.67), we get:

$$\frac{\dot{\eta}_t}{\dot{\eta}_i} = \frac{\eta_t}{\eta_i} = \frac{2Z_1}{Z_1 + Z_2} \quad ...(9.69)$$

$$\text{Now, } \frac{p_r}{p_i} = -\frac{Z_1 \dot{\eta}_r}{Z_1 \dot{\eta}_i} = \frac{Z_2 - Z_1}{Z_1 + Z_2} = -\frac{\dot{\eta}_r}{\dot{\eta}_i} \quad ...(9.70)$$

$$\text{and, } \frac{p_t}{p_i} = \frac{Z_2 \dot{\eta}_t}{Z_1 \dot{\eta}_i} = \frac{2Z_2}{Z_1 + Z_2} \quad ...(9.71)$$

It is obvious from Eqs. (9.68) and (9.70) that $Z_1 > Z_2$, the incident and reflected particle velocities are in phase and the incident and reflected acoustic pressures are in opposite phase. The superposition of incident and reflected velocities which are in phase leads to the cancellation of pressure (a pressure node in the standing wave. If $Z_1 < Z_2$, the pressures are in one phase and the velocities are in opposite phase.

The transmitted particle velocity and acoustic pressure are always in phase with their incident counterparts.

At a rigid wall, where $Z_2 = \infty$, the velocity $\dot{\eta} = 0$. From Eqs. 9.65 $\dot{\eta}_i + \dot{\eta}_r = 0$, which leads to the doubling of pressure at the boundary.

9.17 reflection and transmission of sound intensity

Intensity of sound waves is a *measure of the energy flux, i.e., the rate at which energy crosses unit area* so that it is equal to the product of the energy density (kinetic + potential) and the velocity v. The intensity of sound waves is expressed as

$$I = \frac{1}{2}\rho_0 v \dot{\eta}_m^2 = \frac{1}{2}\rho_0 v \omega^2 \eta_m^2 = \rho_0 v \dot{\eta}_{rms}^2$$

$$= \frac{p_{rms}^2}{\rho_0 v} = p_{rms} \dot{\eta}_{rms} \quad ...(9.72)$$

Normal sound waves range in intensity between 10^{-12} Wm^{-2} to 1Wm^{-2}, which is very low and indicates the sensitivity of the ear. The roar of the people in a full capacity football stadium, greeting a goal generates heat energy just enough to heat a cup of tea.

As commonly used standard of sound intensity is I_0 and is given by:

$$I_0 = 10^{-2} \text{ Wm}^{-2}$$

which is the intensity of the average conversational sound wave between two persons standing nearby. Shouting by one person increases the intensity for the other by a factor of 100 and when the intensity of sound waves is in the range $100I_0$ to $1000I_0$ (10 Wm^{-2}) the sound wave is disturbing.

The response of the human ear to the sound intensity is not proportional to the intensity, so doubling the actual intensity of a certain sound does not lead to the sensation for a sound twice as loud, but only one of that is slightly louder than the original. For this reason, the *decibel* (dB) scale is used for loudness, or sound *intensity level*. An intensity of 10^{-12} W/m^2, which is just audible, is given the value 0 dB, a sound 10 times more intense is given the value of 10 dB; a sound 10^2 times more intense than 0 dB is given the value of 20 dB; a sound 10^3 times more intense than 0 dB is given the value 30 dB; and so forth. Formally, the sound intensity level β in dB, of a sound wave whose intensity is I, in W/m^2, is given by:

$$\beta(\text{dB}) = 10 \log \frac{I}{I_0} \quad ...(9.73)$$

where, $I_0 = 10^{-12}$ W/m^2. Normal conversation might be 60 dB, city traffic noise might be 90 dB, and a jet aircraft might produce as much as 140 dB (which may cause damage to the ear) at a distance of 30 m. Long-term exposure to intensity levels of 85dB usually leads to permanent hearing damage.

The intensity $(I = \rho_0 v \dot{\eta}_{\rm rms}^2 = Z \dot{\eta}_{\rm rms}^2)$ coefficients of reflection and transmission of sound waves at the boundary between the two media is given by

$$\frac{I_r}{I_i} = \frac{Z_1(\dot{\eta}_r^2)_{\rm rms}}{Z_1(\dot{\eta}_i^2)_{\rm rms}}$$

$$= \left(\frac{Z_1 - Z_2}{Z_1 + Z_2}\right)^2 \quad \text{...(9.74)}$$

and, $\frac{I_t}{I_i} =$

$$\frac{Z_2(\dot{\eta}_t^2)_{\rm rms}}{Z_1(\dot{\eta}_i)_{\rm rms}} = \frac{Z_2}{Z_1}\left(\frac{Z_1}{Z_1 + Z_2}\right)^2 = \frac{4Z_1Z_2}{(Z_1 + Z_2)^2}$$

...(9.75)

Using Eqs. (9.74) and (9.75)

$$\frac{I_r}{I_i} + \frac{I_t}{I_i} = \frac{(Z_1 - Z_2)^2}{(Z_1 + Z_2)^2} + \frac{4Z_1Z_2}{(Z_1 + Z_2)^2}$$

$$= \frac{Z_1^2 + Z_2^2 - 2Z_1Z_2 + 4Z_1Z_2}{(Z_1 + Z_2)^2} = \frac{Z_1^2 + Z_2^2 + 2Z_1Z_2}{(Z_1 + Z_2)^2}$$

$$= \frac{(Z_1 + Z_2)^2}{(Z_1 + Z_2)^2} = 1$$

Thus, $\frac{I_r}{I_i} + \frac{I_t}{I_i} = 1$

or, $I_i = I_r + I_t$ ...(9.76)

which is in accordance with the law of conservation of energy. As mentioned earlier, there is a huge difference or mismatch between the acoustic impedances of air on one hand and water or steel on the other. Due to which, there is an almost total reflection of sound wave energy at an air-water interface, independent of the side from which the wave approaches the boundary. While on a steel-water interface there is only 14% of acoustic energy transmission at the interface, a limitation which is a major problem for underwater transmission and detection devices relying on acoustics.

9.18 fourier's Theorem

Till now we have discussed the simple harmonic waves, which may be represented by a sine or a cosine curve. However, actual waves may not be so simple. For example, a musical sound, although it may be periodic in the sense that it repeats itself periodically, it is not a simple harmonic wave. Waves of this type, having an irregular profile, are called *complex waves* and the vibrations which are the source of these waves are called *complex vibrations.*

Fourier, a French mathematician, showed that such complex waves may be analysed into a set of simple harmonic waves. In fact, *any periodic motion, may be represented as a combination of sinusoidal or simple harmonic motions, whose frequencies are an integral multiple of the periodic motion under consideration.*

The following are the conditions, which the periodic function must satisfy so that it can be analysed:

(*i*) It must be *continuous* or the function should have only finite number of discontinuities of slope or magnitude in its time interval of one oscillation.

(*ii*) It should be a *single-valued function,* i.e., it should have only one value at any given instant *t*.

Most of the mechanical waves, including sound waves, satisfy the above-mentioned conditions.

The Fourier's theorem can, therefore, be stated as '*any finite, single-valued periodic function which is either continuous or which possesses only a finite number of discontinuities of slope or magnitude (within the interval of one time-period), may be regarded as a combination of simple harmonic vibrations whose frequencies are integral multiples of that of the given function.*

Suppose $y = f(t)$, is finite, *single-valued, continuous function of t* representing displacement (or pressure variation) with time having time period T such that $y = f(t) = f(t + T)$. That is, it is a periodic function of frequency $\frac{1}{T} = \frac{\omega}{2\pi}$. In accordance with the Fourier's theorem, this finite, single-valued, continuous period function $f(t)$ can be expanded into the following summation or series, also called *Fourier series,* given by:

$$y = f(t) = A_0 + A_1\sin\omega t + A_2\sin 2\omega t + ... + A_r\sin r\omega t$$

$$+ B_1\cos\omega t + B_2\cos 2\omega t + ... + B_r\cos\omega t$$

$$\text{or, } y = A_0 + SA_r\sin r\omega t + SB_r\cos r\omega t \quad \text{...(9.77)}$$

The constant A_0 represents the zero *frequency term* (very often it is equal to zero) and is a measure of the mean displacement of the time-axis. The *sine* and *cosine* terms represent the sinusoidal vibrations of frequencies which are integral multiples of frequency $\omega/2\pi$ of the periodic function and A_1, A_2, ...A_r and B_1, B_2,...B_r are amplitudes of these vibrations often referred to as *Fourier coefficients*. However, it is not necessary that all of them are present in a given series.

Evaluation of Fourier Coefficients: The value of the Fourier coefficients may be obtained easily. The process of this evaluation is also known as *Fourier* (or *harmonic*) *analysis*.

(*i*) Evaluation of A_0: Integrating both sides of Eq. (9.77), we get:

$$\int_0^T y\,dt = A_0\int_0^T dt + A_1\int_0^T \sin\omega t\,dt + \ldots + A_r\int_0^T \sin r\omega t\,dt$$

$$+ B_1\int_0^T \cos\omega t\,dt + \ldots + B_r\int_0^T \cos r\omega t\,dt$$

$$= A_0 T \quad \left(\because \int_0^T \sin\omega t\,dt = \int_0^T \cos\omega t\,dt = 0\right)$$

$$\text{or, } A_0 = \frac{1}{T}\int_0^T y\,dt \qquad \ldots(9.78)$$

From the above expression it is cleat that A_0 represents the mean value of the function (e.g. displacement) during a full time period. Thus, if A_0 is zero, then the axis of the function (displacement curve) would lie along the time-axis.

***(ii)* Evaluation of Constants $A_1, A_2, \ldots, A_r$:** Multiplying Eq. (9.77) by $\sin r\omega t$ and integrating in the limit 0 to T, we have:

$$\int_0^T y\sin r\omega t\,dt = A_0\int_0^T \sin r\omega t\,dt + A_1\int_0^T \sin\omega t\sin r\omega t\,dt + \ldots + A_r\int_0^T \sin^2 r\omega t\,dt$$

$$+ B_1\int_0^T \sin r\omega t\cos\omega t\,dt + \ldots + B_r\int_0^T \sin r\omega t\cos r\omega t\,dt$$

All integrals on R.H.S. except $A_r\int_0^T \sin^2 r\omega t\,dt$ vanish to zero in the chosen limits. Thus,

$$\int_0^T y\sin r\omega t\,dt = A_r\int_0^T \sin^2 r\omega t\,dt$$

$$= \frac{A_r}{2}\int_0^T (1-\cos 2r\omega t)dt = \frac{A_rT}{2}$$

$$\text{or, } A_r = \frac{2}{T}\int_0^T y\sin r\omega t\,dt = \frac{2}{T}\int_0^T y\sin\frac{2\pi rt}{T}\,dt \qquad \ldots(9.79)$$

Putting r = 1, 2, 3,... we can obtain the value of the constants $A_1, A_2, \ldots A_r$ using the above equation.

***(iii)* Evaluation of constants $B_1, B_2, \ldots B_r$:** Multiplying Eq. (9.77) by $\cos r\omega t$ and integrating between the limits t = 0 and $t = T$, we have

$$\int_0^T y\cos r\omega t\,dt = A_0\int_0^T \cos r\omega t\,dt + A_1\int_0^T \cos r\omega t\sin\omega t\,dt + \ldots + A_r\int_0^T \cos r\omega t\sin r\omega t\,dt$$

$$+ B_1\int_0^T \cos r\omega t\cos\omega t\,dt + \ldots + B_r\int_0^T \cos^2 r\omega t\,dt$$

All integrals on the R.H.S except $B_r\int_0^T \cos^2 r\omega t\,dt$ vanish to zero in the chosen limits. Thus,

$$\int_0^T y\cos r\omega t\,dt = B_r\int_0^T \cos^2 r\omega t\,dt$$

$$= \frac{B_r}{2}\int_0^T (1+\cos 2r\omega t)dt = \frac{B_rT}{2}$$

$$B_r = \frac{2}{T}\int_0^T y\cos r\omega t\,dt = \frac{2}{T}\int_0^T y\cos\frac{2\pi rt}{T}\,dt \qquad \ldots(9.80)$$

Putting r = 1, 2, 3,..., we can obtain the values of the constants B_1, B_2, B_r, using above equation.

Application of Fourier's Theorem: We consider below, one example of fourier analysis:

***(i)* The square wave:** As shown in Fig. 9.17, consider a square wave, sometimes also referred to a *top-hat wave* because of its resemblance with the shape of a top hat.

$$\left.\begin{array}{l}\text{Here, } y = f(t) = a \text{ from } t = 0 \text{ to } T/2\\ \text{and } y = f(t) = -a \text{ from } t = T/2 \text{ to } T\end{array}\right\} \qquad \ldots(9.81)$$

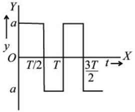


Fig. 9.17

Now according to Fourier series,

$y = f(t) = A_0 + A_1\sin\omega t + ... + A_r\sin\omega t + B_r\sin r\omega t + ... + B_r\cos r\omega t$

We know $A_0 = \frac{1}{T}\int_0^T y\,dt,$ using Eq. (9.78), we have:

$$A_0 = \frac{1}{T}\left[\int_0^{T/2} a\,dt - \int_{T/2}^{T} a\,dt\right] = 0$$

This implies that the *mean value of the displacement over a time period is zero, i.e., the axis of the displacement curve coincides with time-axis, about which the wave is symmetrical.*

Also, $A_r = \frac{2}{T}\int_0^T y\sin r\omega t\,dt,$ which in this case becomes:

$$A_r = \frac{2}{T}\left[\int_0^{T/2} a\sin r\omega t\,dt - \int_{T/2}^{T} a\sin r\omega t\,dt\right]$$

$$= \frac{2a}{T}\left[\left|\frac{-\cos r\omega t}{r\omega}\right|_0^{T/2} + \left|\frac{\cos r\omega t}{r\omega}\right|_{T/2}^{T}\right]$$

$$= \frac{2a}{r\omega T}\left[\left(-\cos\frac{r\omega T}{2} + 1\right) + \left(\cos r\omega T - \cos\frac{r\omega T}{2}\right)\right]$$

Since, $\omega = \frac{2\pi}{T}$ or $\omega T = 2\pi,$ we have:

$$A_r = \frac{a}{\pi r}(2 - 2\cos\pi r) = \frac{2a}{\pi r}(1 - \cos\pi r)$$

Thus, if r is even, $A_r = 0$

and if r is odd, $A_r = \frac{4a}{\pi r}$

That is, $A_0 = 0,\ A_1 = \frac{4a}{\pi},\ A_2 = 0,\ A_3 = \frac{4a}{3\pi},\ A_4 = 0, = \frac{4a}{5\pi},\dots$

The constants $B_r = \frac{2}{T}\left[\int_0^{T/2} a\cos r\omega t\,dt - \int_{T/2}^{T} a\cos r\omega t\,dt\right]$ which in this case becomes:

$$B_r = \frac{2}{T}\left[\int_0^{T/2} a\cos r\omega t\,dt - \int_{T/2}^{T} a\cos r\omega t\,dt\right]$$

$$= \frac{2a}{T}\left[\left|\frac{\sin r\omega t}{r\omega}\right|_0^{T/2} - \left|\frac{\sin r\omega t}{r\omega}\right|_{T/2}^{T}\right]$$

$$= \frac{2a}{r\omega t}\left[\left(\sin\frac{r\omega T}{2} - 0\right) - \left(\sin r\omega t - \sin\frac{r\omega t}{2}\right)\right]$$

Using $\omega T = 2\pi$, we get:

$$B_r = \frac{a}{\pi r}[(\sin\pi r - 0) - \sin(2\pi r - \sin\pi r)]$$

$= 0$

That is, $B_1 = B_2 = B_3 = ...B_r = 0$

Substituting, value of the constants, we get:

$$y = \frac{4a}{\pi}\left(\sin\omega t + \frac{1}{3}\sin 3\omega t + \frac{1}{5}\sin 5\omega t + \dots\right)$$

$$\text{or, } y = \frac{4a}{\pi}\left(\sin\frac{2\pi t}{T} + \frac{1}{3}\sin\frac{6\pi t}{T} + \frac{1}{5}\sin\frac{10\pi t}{T} + \dots\right)$$

...(9.82)

Thus, the *combination consists of the fundamental note and its odd harmonics*. In other words, the ratios of the frequency of the constituents are proportional to odd values of ω i.e., w, 3ω, 5ω,... and their amplitudes are in the ratio $1 : \frac{1}{3} : \frac{1}{5} : \dots$.

A superposition of the displacement curves (or harmonic waves) representing these constituents would result in a square wave. In Fig. 9.18 curve (*i*) represents the simple harmonic wave corresponding to $y_1 = \frac{4a}{\pi}\sin\omega t,$ curve (*ii*) represents the simple harmonic wave corresponding to $$y_3 = \frac{4a}{3\pi}\sin\omega t$$ having amplitude 1/3rd that of y_1 and the frequency thrice of y_1 and curve (*iii*) represents $$y_5 = \frac{4a}{5\pi}\sin\omega t$$ with amplitudes 1/5th that of y_1 and the frequency five times that of y_1. The superposition of these curves results in curve (*r*), which bears only a partial resemblance to the given square curve. However, if we also consider y_7, y_9, y_{11},... the resultant curve will become more and more closer to the given square curve.

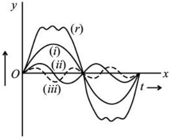


Fig. 9.18

Exercises

Short Answer Type

1. What do you mean by 'Wave Motion'? Differentiate between Mechanical and Non-mechanical waves.

2. List the properties of the medium required for the propagation of mechanical waves.

3. What are Transverse Waves? Briefly discuss their characteristics.

4. What are Longitudinal Waves? Explain their propagation mechanism.

5. Define 'Wavefront'. Briefly explain various types of wavefronts.

6. Define the terms: (*i*) Wavelength, (*ii*) Frequency, (*iii*) Wave Number (*iv*) Phase Velocity, and (*v*) Particle Velocity.

7. Obtain an expression relating Wave Velocity and Particle Velocity.

8. Define 'energy current' and obtain an expression for the same. Hence, define 'Intensity of Wave'.

9. What do you mean by 'impedance matching'? Discuss.

10. State and explain 'Principle of Superposition'.

11. Differentiate between Group Velocity and Phase Velocity.

12. Explain the terms: (*i*) Compressions, (*ii*) Rarefactions, (*iii*) Dilation, and (*iv*) Condensation.

13. Define acoustic impedance of sound wave and show it is equal to $r_0 n$.

14. Show $I_i = I_r + I_t$.

Long Answer Type

1. Discuss the various types of waves. Describe the propagation mechanism of transverse and longitudinal waves.

2. Obtain an expression for the displacement of a plane progressive wave. Hence, derive expression for wave velocity and differential equation of wave motion.

3. Discuss the variation of velocity and pressure for a plane progressive wave showing their variation graphically.

4. Show that for a plane progressive wave, both kinetic and potential energies depend on x and t but the total energy is independent of both x and t.

5. Discuss the motion of transverse wave on a string and obtain an expression for the wave equation.

6. Show that the characteristic impedance is $Z = rv$, where the symbols have their usual meaning.

7. Discuss reflection and transmission of waves on a string and derive the expression for reflection and transmission coefficients of amplitudes.

8. Discuss 'impedance matching' between two media coupled by an intervening medium. Hence, show that for the impedance matching Z_2 should be equal to $\sqrt{Z_1 Z_3}$.

9. Discuss the formation of standing waves on a string of fixed length. Hence, define 'standing wave ratio'.

10. State and explain Fourier's theorem and obtain the values of Fourier coefficients. Apply the theorem to a square wave.

11. Obtain the wave equation for sound wave.

12. Discuss reflection and transmission of sound waves on a boundary between the two media and hence show that if $Z_1 > Z_2$, the incident and reflected particle velocities are in phase and the incident and reflected acoustic pressures are in opposite phase.

Numericals

1. A plane progressive wave train of frequency 400 Hz has a phase velocity of 480 m/s. How far apart are two points 30° out of phase? (*b*) What is the phase difference between two displacements at a given point at times 10^{-3} sec apart?

[**Ans.** (*a*) 0.1 m, (*b*) 144°]

2. A plane progressive wave train of frequency along the +*x*-direction has the following characteristics: a = 0.2 cm, v = 360 cm/s and λ = 60 cm, (*a*) Write down the equation for it (*i*) when displacement is zero at x = 0 and t = 0, and (*ii*) when displacement is maximum at x = 0 and t = 0. (*b*) Obtain the displacement in both cases x = 120 cm and t = 2 sec.

[**Ans.** (*a*) (*i*) $y = 0.2 \sin 2\pi\,(6t - x/60)$, (*ii*) $y = 0.2 \cos 2\pi(6t - x/60)$, (*b*) (*i*) y = 0, (*ii*) y = 0.2 cm]

3. A plane progressive harmonic wave, travelling along the +*x*-direction has an amplitude of 5.0 cm and frequency, 100 cps. If its velocity is 6000 cm/s, obtain the value of (*i*) the wave number, (*ii*) displacement, (*iii*) particle velocity and (*iv*) particle acceleration at $x = 150$ cm and $t = 2$ sec.

[**Ans.** (*i*) 0.01661 cm, (*ii*) 0, (*iii*) 3142 cm/sec, (iv) 0]

4. A plane progressive harmonic wave is propagating with a velocity of 340 m/s in a medium of density 0.0015 gm/c.c If the amplitude of the wave is 10^{-4}cm and its frequency 300 Hz, obtain the value of (*i*) energy density and (*ii*) energy current for it.

[**Ans.** (*i*) 2.665×10^{-5}ergs/c.c., (*ii*) 0.9059 ergs/cm^2.sec]

5. A piece of wire 0.5 m long and mass 1.44 gm is stretched by a load of 25 kg. What is the frequency of the second harmonic?

[**Ans.** 583.2 Hz]

6. A string of mass 2 gm/m carries progressive waves of amplitude 1.5 cm, frequency 60 Hz and speed 200 m/s. Calculate: (*i*) the energy per unit length of the wave, (*ii*) the rate of energy propagation in the wave.

[**Ans.** (*i*) 3.2×10^{-2} J, (*ii*) 6.4 J/s]

Unit V

10

Relativity

10.1 relativity: INTRODUCTION and common terms

While sitting in a moving vehicle, if you look at the distant objects like trees or anything stationary, they appear to move in a direction opposite to the direction of the motion of the vehicle. It is easy to realize that for a person standing on the ground, you will appear to be moving in a particular direction, while to you, the standing person would appear to move in a direction opposite to your direction of motion. In other words, *what one observes is relative, it is not absolute and depends on the state of motion of the observer*.

We take another example, a boy sitting in a moving vehicle throws a ball upward and the ball returns to his hand after sometime and it is being watched by a person standing outside. Now, imagine what is observed by the boy and the person standing outside. For the boy, the ball has moved up straight and come down straight into his hand, as if he had been stationary. What does the person standing outside observes? He observes, that after the boy has thrown the ball and because the boy is continuously moving, the ball takes a parabolic path before finally coming to his hand. These two examples make it clear that what an observer observes, depends on his state or his *frame of reference*, as discussed below.

Newton, Galileo, Lorentz, Michelson, Morley, Einstein and others have made significant contributions in developing the subject of relativity, the understanding of which is necessary for knowing the mysteries of physics. We are going to see that length, mass, time, etc. are not absolute and their values depend on the state of the observer.

Some of the common terms used in relativity are defined below:

(*i*) Particle: *A particle is a small piece of matter, having practically no linear dimension, but only a position at a point*. It is characterized by its mass and charge.

(*ii*) Observer: *A person who locates, records, measures and interprets an event is called an observer*.

(*iii*) Event: In relativity, an event implies *anything that occurs suddenly or instantaneously at a point in space*. It involves a position and a time of occurrence.

(*iv*) Frame of Reference: Even the basic physical quantities like displacement, velocity, time, mass, etc. are not absolute and the measured values are relative, depending upon the reference and the state of observer. To locate the coordinates of a point, we assign a specific x-, y- and z-values of a coordinate system having its origin at O with $x = 0$, $y = 0$ and $z = 0$. If the origin of the coordinate system is changed, the coordinates of the point also change. We can state that the reference of the point has changed and the coordinate axes form a reference system.

A system of coordinate axes which defines the position of a particle or specifies the location of an event is called a frame of reference. The simplest frame of reference is the *Cartesian system of coordinates* in which the location of a point is specified by the three (x-, y- and z-) coordinates.

However, for complete specification of an event in a reference frame, i.e., for determination of its exact location as well as the exact time of its occurrence, in addition to the three space coordinates, another coordinate, time t of its occurrence should be specified. A frame of reference having four coordinates, x, y, z and t is referred to as a *space-time frame*—the four axes defining *a four-dimensional contirum* is called *space-time*.

10.2 Inertial and non-inertial frame of references

Mechanics is based on Newton's laws of motion, but the laws do not mention the frame of reference in which they are valid. *Newton's first law*, also called the *law of inertia*, states that a body at rest will continue to remain at rest while the one in motion will continue to move along a straight line. In other words, the body will have no acceleration. If this is viewed from the point of view of relativity, the previous sentence is meaningless because the frame of reference is not mentioned with respect to which the body is observed.

If we have a frame of reference for which the Newton's first law is true, then in another frame moving with respect to the former with an acceleration, the same body (not acted by external forces) will appear to have an accelerated motion. The former is called an *inertial frame of reference* while the latter is called a *non-inertial frame of reference*. Thus, there are two types of frames of reference which we are discussing below:

(*i*) Inertial Frame of Reference: Consider a system S of coordinates (x, y, z) with respect to which a body is in uniform motion. The velocity of the body is given by

$$\frac{dx}{dt} = u_x, \frac{dy}{dt} = u_y, \frac{dz}{dt} = u_z$$

where u_x, u_y and u_z are the velocity components along x-, y- and z-directions respectively.

Also, since the body is in uniform motion,

$$\frac{d^2x}{dt^2} = \frac{dv_x}{dt} = a_x = 0, \frac{d^2y}{dt^2} = \frac{dv_y}{dt} = a_y = 0,$$

$$\frac{d^2z}{dt^2} = \frac{dv_z}{dt} = a_z = 0$$

where a_x, a_y, a_z denote the acceleration along x-, y-, and z-directions.

Thus, with respect to this frame of reference, the body moves in a straight line with a uniform velocity when no external force is acting on it, in accordance with Newton's law. So, it is possible to choose a frame of reference with respect to which the body is at rest or is in uniform motion, moving with a constant velocity, i.e., with respect to this frame of reference, the body is unaccelerated. Such a frame of reference is called *inertial frame of reference*.

Hence, *an inertial frame of reference is the one in which bodies undergo uniform linear motion with constant velocity when no external force is acting on it*.

Alternatively, *an inertial frame of reference may be defined as a frame of reference in which the law of inertia and other laws of Newtonian mechanics are valid*. Such a frame of reference is also referred to as *Newtonian* or *Galilean* frame of reference. All frames of reference, moving with a constant velocity with respect to an inertial frame are also inertial frames of reference. *The inertial frames are non-accelerating frames*.

(*ii*) Non-inertial Frame of Reference: *A frame of reference in which a body, not acted by an external force, is accelerated and hence the law of inertia and other laws of Newtonian mechanics are not valid, is referred to as a non-inertial frame of reference*.

Let us consider two frames of reference S and S', where the frame S' is moving with an acceleration a with respect to the frame S. Now suppose a particle of mass μ is moving with a uniform velocity v with respect to an observer in frame S. The particle will appear to be moving with an acceleration $-a'$ to an observer in S'. Therefore, the observer in S', will measure a force $-ma'$ acting on the particle. This force is known as *fictitious force* as it is observed due to the frame S' accelerating with respect to the frame S. *The frame S' is a non-inertial frame as the Newton's law does not hold in it*. In case, the particle is moving with acceleration a in frame S, then the observed acceleration in S' will be $a - a'$ and the resulting force will be $\mu(a - a')$. Thus, $\varphi = ma - ma' = ma + (-ma')$ = *Real force in S + fictitious force in S'*. The force φ is known as *apparent force* in S'. *It is due to the fictitious force that a man inside the lift, when the lift is moving up, feels more weight than his real weight and feels lighter when the lift is moving down with uniform acceleration.*

10.3 Galilean Relativity

An event observed simultaneously in two separate reference frames has two different sets of coordinates. A set of equations relating to the two sets of coordinates of an event in the two frames of reference are called *transformation equations*, because they enable the observations made in one system to be transformed into those made in the other. If the two reference frames happen to be inertial or Galilean, the transformation is referred to as *Galilean transformation*.

Consider a frame of reference S which is inertial. There is another frame of reference S' with axes x', y', z' parallel to the axes x, y, z of the frame S Fig. 10.1. The frame S' is moving with a uniform velocity v along the x-axis and at t = 0 the two frames coincided, i.e., at t = 0, the axes of S and S' overlapped or $O = O'$. At any time t, the x-coordinate in S exceeds that in S' by vt, the distance covered by S' in time t in the + x direction. Therefore, the observed coordinates in the two frames are given by the transformation equations

$$\left.\begin{aligned} x' &= x - vt \\ y' &= y \\ z' &= z \end{aligned}\right\} \quad \text{...(10.1)}$$

Also, in Newtonian mechanics, $t' = t$.

The above four equations are known as *Galilean (position) transformations* from S to S'.

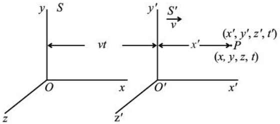


Fig. 10.1

The *inverse transformation* (from *S'* to *S*) is given by:

$$\left.\begin{aligned} x &= x' + vt \\ y &= y' \\ z &= z' \\ t &= t' \end{aligned}\right\} \quad ...(10.2)$$

The general form of transformation equations is $\mathbf{r}' = \mathbf{r} - \mathbf{v}t$ and that of inverse transformation equations is $\mathbf{r}' = \mathbf{r} + \mathbf{v}t$. *Differentiating the transformation Eq.* (10.1), *we obtain velocity transformation equations* from *S* to *S'*. These are:

$$\left.\begin{aligned} u'_x &= u_x - v \\ u'_y &= u_y \\ u'_z &= u_z \end{aligned}\right\} \quad ...(10.3)$$

Thus, *velocity is not invariant in Galilean transformation.* In general, the velocity transformation is given by

$$\frac{d\mathbf{r}'}{dt} = \frac{d\mathbf{r}}{dt} - \mathbf{v} \text{ or } \mathbf{u}' = \mathbf{u} - \mathbf{v}.$$

While the inverse velocity transformation is given by:

$$\mathbf{u} = \mathbf{u}' + \mathbf{v}$$

The *acceleration transformation equation* is obtained by differentiating Eq. (10.3) and is given by

$$\left.\begin{aligned} a'_x &= a_x \\ a'_y &= a_y \\ a'_z &= a_z \end{aligned}\right\} \quad ...(10.4)$$

Thus, the *acceleration is invariant in Galilean transformation.* According to Newton's second law $\mathbf{F} = \mu\mathbf{a}$,

Similarly,

$$\text{and,} \quad \left.\begin{aligned} F_x &= ma_x = ma'_x = F'_x \\ F'_y &= F_y \\ F'_z &= F_z \end{aligned}\right\} \quad ...(10.5)$$

Thus, the Newton's law is valid in both *S* and *S'*, i.e., *both S and S' are inertial frames. To generalize this statement, we may state that all frames of reference moving with a uniform velocity with respect to an inertial frame are also inertial.*

We can easily show that length or distance ($\lambda = \Delta x = x_2 - x_1 = x'_2 - x'_1 = \Delta x' = L'$) is *invariant in Galilean transformation.*

Invariance Principle: As Newton's laws are valid in both *S* and *S'*, no mechanical experiment will be able to distinguish between the two frames, one of which is in uniform relative motion with respect to the other. Galileo discovered this fact long before Newton and, therefore, the inertial frames are also known as *Galilean frames* (a term used by Einstein). The only physical laws known at the time of Galileo were the laws of mechanics, so the following principle is known as the *principle of Galilean relativity*.

The laws of physics are invariant in all Galilean (or inertial) frames of reference.

If the two observers, at rest in the two inertial frames *S* and *S'*, perform experiments in their respective frames independently, they will discover the same laws of physics.

The length, time and mass are invariant in Galilean transformation.

10.4 luminiferous ether hypothesis

The formulation of Maxwell's equations in the year 1862, showed that electromagnetic waves propagate through the free space with the speed of light *c*. Hertz later generated and detected electromagnetic waves in the laboratory. It was confirmed that light is an electromagnetic wave. Since light reaches us from distant stars, it traverses millions of kilometres through free space or vacuum. Up to that time, all the waves known to mankind were mechanical waves (e.g., sound wave) and so, required a material medium for its propagation and the scientists could not conceive that idea of wave propagating through free space of vacuum. It was assumed that an all pervading medium called *ether* filled the entire universe, due to which light waves could travel to us from the distant stars. Robert Hook, used the term *luminiferous ether as light waves propagated through it*. It was assumed that *all planets, stars and heavenly bodies move through this hypothetical medium, which offers no resistance* to their motion. Even the empty spaces within atoms and molecules was assumed to be filled with ether. All the efforts to measure the weight of the ether failed and so it was supposed to the weightless.

Scientists tried to detect and understand the relative motion of physical bodies with respect to ether. If the material bodies such as earth moving in space filled with ether carried ether with them, there would be no relative motion between the two. Hence, there will be no change in the velocity of light and the absolute motion of the moving bodies with respect to ether would not be detected. On the other hand, if the material bodies move in space and the ether remains at rest. there would be relative motion between the two. Hence, there would be a change in the velocity of light with respect to the moving bodies on account of their motion in ether and the absolute motion of the moving bodies could be detected.

In order to establish the existence of ether, many experiments were conducted, the most famous among them being Michelson-Morley experiment, which we are discussing below. The negative result of this experiment ruled out the existence of this hypothetical medium.

10.5 Michelson-Morley Experiment

Michelson devised an apparatus to detect the motion of the earth relative to ether at rest. He improved the accuracy of the set-up with his collaborator Morley. He was awarded the Nobel Prize for this experiment. It was expected that light would propagate with different speeds in different directions as viewed from the earth. The earth is moving with respect to ether at rest at a speed of 3×10^6 cm/s. Therefore, the time taken by light to travel equal distance in different directions would be different. The objective of the experiment was to find this time difference from which the relative velocity between ether and the earth could be estimated.

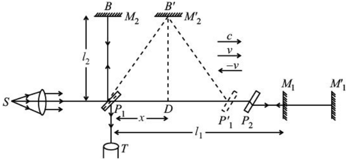


Fig. 10.2

The experimental arrangement is shown in Fig. 10.2. A beam of light from a monochromatic source S falls upon a half silvered glass plate P_1 inclined at 45° to the direction of propagation of light and is partly reflected and partly transmitted. The reflected ray travels towards mirror μ_2, gets reflected at B and retraces its path and finally enters the telescope T. Another portion of ray gets refracted and transmitted through P_1 and passes through a *compensating glass plate* P_2 which is used to equalize the total distance traversed inside the glass in the two paths. The ray gets reflected by mirror μ_1, retraces its path and then gets reflected at P_1 and finally meets the other ray in the telescope. Since the glass plates P_1 and P_2 are equally thick, so the path traversed by the two rays through the glass remains the same, traversing *thrice* through the glass. Depending upon the path (or phase) difference between the two rays, bright or dark interference fringes are formed.

Let us assume that the arm PM_1 is in the direction of the earth's velocity through the ether towards the right. If λ_1 is the distance of mirror μ_1 from P_1, then the time required by the light ray to traverse this distance is $\frac{l_1}{c+v}$, where c is the speed of light, v that of apparatus (i.e., the earth), relative to the ether and $c + v$ being the relative velocity of light with respect to the apparatus. The time required for traversing this distance in the opposite direction after reflecting from the mirror is $\frac{l_1}{c-v}$, $c - v$ being the relative velocity of light in this direction. So, the time taken in the complete journey is:

$$t_1 = \frac{l_1}{c+v} + \frac{l_1}{c-v} = \frac{l_1(c-v) + l_1(c+v)}{c^2 - v^2}$$

$$= \frac{2l_1c}{c^2 - v^2} = \frac{2\frac{l_1}{c}}{1 - \frac{v^2}{c^2}} \quad ...(10.6)$$

Now, let us find the time taken by light in the to and fro journey to the mirror μ_2. Since the earth is in motion, suppose the mirror μ_2 moves to M'_2 during the time the ray of light reaches μ_2 from P_1. Hence, the ray traverses the path $P_1 M'_2 P'_1$. $B'D$ is the median of the isosceles triangle $P_1 M'_2 P'_1$. Hence, $B'D$ is perpendicular to $P_1 P'_1$. If $P_1\Delta = BB' = x$ and λ_2 the distance of mirror μ_2 from P_1, then using Pythagoras's theorem, $B' P_1 = \sqrt{l_2^2 + x^2}$. It follows, in the period the point B moves to B' due to the earth's motion, the ray traverses the distance P_1B' in the same time. Hence,

$$\frac{x}{v} = \frac{\sqrt{l_2^2 + x^2}}{c} \Rightarrow \frac{x^2}{l_2^2 + x^2} = \frac{v^2}{c^2}$$

$$\text{or, } 1 - \frac{x^2}{l_2^2 + x^2} = 1 - \frac{v^2}{c^2}$$

$$\text{or, } \frac{l_2^2}{l_2^2 + x^2} = 1 - \frac{v^2}{c^2}$$

$$\text{or, } \sqrt{l_2^2 + x^2} = \frac{l_2}{\sqrt{1 - \frac{v^2}{c^2}}} \quad ...(10.7)$$

The time taken by the light ray to cover this distance $= \frac{\sqrt{l_2^2 + x^2}}{c}$

$$= \frac{l_2}{c\sqrt{1 - \frac{v^2}{c^2}}}$$

Hence, the time for the to and fro motion of the second ray:

$$t_2 = \frac{2l_2}{c\sqrt{1 - \frac{v^2}{c^2}}} \quad ...(10.8)$$

Interference fringes are produced due to path difference traversed by the two rays or difference between t_1 and t_2. If the mirrors μ_1 and μ_2 are exactly perpendicular to each other, the interference fringes will be concentric circles. Actually, in the Michelson-Morley experiment, the angle between the mirrors was kept slightly less than 90° so that the fringes were parallel lines as are obtained from a wedge. Now, from Eqs. (10.6) and (10.8)

$$\Delta t = t_1 - t_2 = \frac{2}{c\sqrt{1-\frac{v^2}{c^2}}}\left(\frac{l_1}{\sqrt{1-\frac{v^2}{c^2}}} - l_2\right) \quad \ldots(10.9)$$

The interferometer was rotated through 90° to interchange the two arms.

Naturally, λ_1 takes the place of λ_2 and vice versa and so we have:

$$t'_1 = \frac{2l_1}{c\sqrt{1-\frac{v^2}{c^2}}} \text{ and } t'_2 = \frac{2l_2}{c\left(1-\frac{v^2}{c^2}\right)}$$

$$\therefore \Delta t' = t'_1 - t'_2 = \frac{2}{c\sqrt{1-\frac{v^2}{c^2}}}\left(l_1 - \frac{l_2}{\sqrt{1-\frac{v^2}{c^2}}}\right) \quad \ldots(10.10)$$

The interference fringes shift only if the time difference changes. This change is given by

$$\Delta t' - \Delta t = \frac{2(l_1+l_2)}{c\sqrt{1-\frac{v^2}{c^2}}}\left(\frac{1}{\sqrt{1-\frac{v^2}{c^2}}} - 1\right)$$

Expanding the right-hand side in a Binomial series and retaining the terms only upto second power in $\frac{v}{c}$, we get:

$$\Delta t' - \Delta t = \frac{1}{c}(l_1+l_2)\frac{v^2}{c^2} \quad \ldots(10.11)$$

The corresponding path difference $c(\Delta t' - \Delta t) = -(l_1+l_2)\frac{v^2}{c^2}$. Hence, the shift in the number of fringes is:

$$\Delta n = \frac{c(\Delta t' - \Delta t)}{\lambda}$$

$$= -\frac{(l_1+l_2)}{\lambda}\frac{v^2}{c^2}$$

If $\lambda_1 = \lambda_2 = \lambda$, then

$$\Delta n = -\frac{2l}{\lambda}\frac{v^2}{c^2}$$

$$\text{or, } \Delta n = \frac{2l}{\lambda}\frac{v^2}{c^2} \quad \ldots(10.12)$$

Substituting the orbital velocity of the earth $v = 3 \times 10^4$ m/s, so that $\frac{v^2}{c^2} = 10^{-8}$, the arrangement should be able to detect the changes of the order of 10^{-8} or one part in 10^8 (0.00000001). In 1881, Michelson took $\lambda = 1.2$ m and for the light of wavelength $\lambda = 6000$ Å $= 6 \times 10^{-7}$ m, he obtained:

$$\frac{\Delta n}{n} = 0.04$$

Although, the instrument was capable of measuring such a small shift, *no fringe shift was actually observed.*

In 1887, Michelson improved the accuracy of measurement and repeated the experiment with his collaborator Morley. The length of the path was increased by allowing light rays to be reflected in their path so that $\lambda_1 + \lambda_2 = 22$ m, $\lambda = 5500$ Å, $\frac{v}{c} = 10^{-4}$. Then the fringe shift works out to be $\frac{\Delta n}{n} = 0.4$, i.e., a fringe shift of four-tenths. Although, the instrument was capable of detecting a fringe shift of $\frac{1}{100}$th of a fringe, *again no fringe shift was observed.*

To avoid the effect of vibrations on the instrument and the stress produced by rotating the arm, they fixed the instrument to a stone slab and floated it in mercury. In subsequent years, they repeated the experiment many times with more accurate instruments, but still no fringe shift could be detected.

The negative result of this experiment proved that the space or medium in which light propagated is not moving relative to the earth. In 1905, Einstein formulated the theory of relativity. He concluded that *the velocity of light is always the same in all directions and is independent of the relative motion of the observer, medium and source.* If it is so, we will have c instead of $c + v$ and c

$$\frac{\Delta n}{n} = 0.$$

– v giving *Hence, the null or negative result of Michelson-Morley experiment can be explained on the basis of the constancy of the speed of light.*

The following two important conclusions are drawn on the basis of the result of this experiment:

(*i*) Ether has no observable properties and there is nothing like absolute space or a fixed fundamental frame of reference with respect to which absolute motion of the bodies can be determined. In other words, absolute motion is meaningless.

(*ii*) The velocity of light is same in all directions and is independent of the relative motion of the source or the observer or both.

Example 10.1

What is the expected fringe shift in Michelson-Morley experiment, if λ = 11 m, v = 30 km/sec and λ = 6000 Å?

Solution:

The expected fringe shift Δn is $\frac{2lv^2}{c^2\lambda}$.

Here, λ = 11 m, v = 30 km/sec = 3×10^4 m/s

$c = 3 \times 10^8$ m/s and $\lambda = 6000 \times 10^{-10}$ m = 6×10^{-7} m

$\therefore \Delta n =$

$$\frac{2\times 11\times(3\times 10^4)^2}{(3\times 10^8)^2\times 6\times 10^{-7}} = \frac{2\times 11}{6\times 10} = 0.37.$$

10.6 Einstein's special theory of relativity

The special theory of relativity deals with solving the problem of absolute motion. According to Einstein, the concept of 'absolute motion' is meaningless. There are two postulates of Einstein's special theory of relativity, the *first postulate* (*the principle of equivalence*) is basically the principle of Galilean relativity and is stated as follows.

All laws of physics remain invariant in all inertial (or Galilean) frames: In other words, the fundamental laws of physics are the same when stated in all inertial frames of reference in uniform motion relative to each other. All inertial frames are equivalent and no scientific experiment will be able to find a difference between the two frames.

Michelson-Morley and other experiments proved that the velocity of light in empty space is the same in all Galilean frames and is independent of the direction of propagation. Einstein based his *second postulate (constancy of the velocity of light)* on this result, according to which ***the velocity of light in free space is constant and independent of not only the direction of propagation, but also the relative velocity between the source of light and the observer.***

Einstein had developed Lorentz transformation independently and showed that *time is not independent of the coordinate system*, i.e., t' ? t'. This drastically changed the age old concepts of space and time. Although, it was somewhat difficult to accept the frame independence of time, as classical physics tells us $t = t$, i.e., the time is absolute.

10.7 lorentz transformations

Occurrence of any physical event is specified by three space coordinates (x, y, z) and the time t, i.e., space time coordinate (x, y, z, t). Lorentz transformations relate the space and time coordinates of an event as observed in two inertial frames.

Let us consider two frames of reference S and S' (Figure 10.1), the frame S' moving with uniform velocity v along the x-axis. Suppose, there are two observers O and O' in the frames S and S' respectively. At some instant of time, origin of time measurements t and t', the origins of the coordinate frames were coincident. Let, at this instant of space and time, a light signal be emitted. For both the observers, signal develops into a *spherical wavefront* with its point of origin as centre. For the observer O in the frame S, the wavefront passes through a point P (x, y, z) at t, with equation of spherical wavefront being

$$x^2 + y^2 + z^2 - c^2t^2 = 0 \quad ...(10.13)$$

Similarly, for the observer O' in the frame S', the coordinates of P are (x', y', z') and the wavefront reaches there after time t' so that:

$$x'^2 + y'^2 + z'^2 - c^2t'^2 = 0 \quad ...(10.14)$$

According to the second postulate, the value of c being the same in all inertial frames. Here, we want to obtain the transformation connecting x, y, z, t and $x\times, y\times, z\times, t\times$ such that Eqs. (10.13) and (10.14) are satisfied, i.e.,

$$x'^2 + y'^2 + z'^2 - c^2t'^2 = g\,(x^2 + y^2 + z^2 - c^2t^2) \quad ...(10.15)$$

where g is any undetermined constant. As the relative motion is along x-axis, it is evident that:

$$y = y' \text{ and } z = z' \quad ...(10.16)$$

For g = 1 from Eq. (10.15), we have

$$x'^2 - c^2t'^2 = x^2 - c^2t^2 \quad ...(10.17)$$

The linear and simple transformation between x and x', for $v < < c$, can be written as

$$x' = \lambda(x - vt) \quad ...(10.18)$$

where constant λ is independent of x and t. The form of Eq. (10.18) is due to the fact that the transformation must reduce to Galilean transformation for low value of v, i.e., $v/c << 1$.

Since the motion is relative, we can well assume that the frame S is moving with velocity $- v$ along x-axis, therefore

$$x = \lambda(x' + vt') \quad ...(10.19)$$

Putting the value of x' from Eq. (10.18) in Eq. (10.19), we get

$$x = \lambda[\lambda(x - vt) + vt']$$

$$\text{or, } t' = \lambda\left[t - \frac{x}{v}\left(1 - \frac{1}{\lambda^2}\right)\right] \quad ...(10.20)$$

Substituting the values of x' and t' from Eqs. (10.18) and (10.20), in Eq. (10.12), we get

$$x^2 - c^2t^2 =$$

$$\lambda^2(x - vt)^2 - c^2\lambda^2\left[t - \frac{x}{v}\left(1 - \frac{1}{\lambda^2}\right)\right]^2$$

or,

$$x^2 - c^2t^2 - \lambda^2(x^2 - 2vxt + v^2t^2) + c^2\lambda^2\left[t - \frac{x}{v}\left(1 - \frac{1}{\lambda^2}\right)\right]^2 = 0$$

or,

$$x^2 - c^2t^2 - \lambda^2(x^2 - 2vxt + v^2t^2) + c^2\lambda^2\left[t^2 - \frac{2xt}{v}\left(1 - \frac{1}{\lambda^2}\right) + \frac{x^2}{v^2}\left(1 - \frac{t}{\lambda^2}\right)^2\right] = 0.$$

In order to satisfy this identity, the coefficient of various powers of x and t should vanish separately. Equating coefficient of xt to zero, we get

$$2\lambda^2 v - \frac{2c^2\lambda^2}{v}\left(1 - \frac{1}{\lambda^2}\right) = 0$$

or, $(v^2 - c^2)\lambda^2 + c^2 = 0$

$$\text{or, } \lambda = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \quad \ldots(10.21)$$

Putting the value of λ in Eq. (1018.), we get

$$x' = \frac{x - vt}{\sqrt{1 - \frac{v^2}{c^2}}} \quad \ldots(10.22)$$

Further, from Eq. (10.20), we have:

$$t' = \lambda\left[t - \frac{x}{v}\left(1 - \frac{1}{\lambda^2}\right)\right] = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}\left[t - \frac{x}{v}\left(1 - \frac{c^2 - v^2}{c^2}\right)\right]$$

$$= \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}\left[t - \frac{xv^2}{vc^2}\right]$$

$$\text{or, } t' = \frac{t - \frac{vx}{c^2}}{\sqrt{1 - \frac{v^2}{c^2}}} \quad \ldots(10.23)$$

Combining Eqs. (10.22), (10.16) and (10.23), we have

$$\left.\begin{aligned} x' &= \frac{x - vt}{\sqrt{1 - \frac{v^2}{c^2}}} \\ y' &= y \\ z' &= z \\ t' &= \frac{t - \frac{vx}{c^2}}{\sqrt{1 - v^2/c^2}} \end{aligned}\right\} \quad \ldots(10.24)$$

These are called *Lorentz transformation equations*.

Inverse Lorentz Transformations: If we consider the event to be observed in the frame S' rather than in S, than the corresponding transformation, called, *Inverse Lorentz transformation* can be obtained by changing v to $-v$ and primed quantities into unprimed quantities and vice versa.

Thus, the transformations, i.e., equations relating coordinates in the frame $S'(x', y', z', t')$ to that in the frame $S(x, y, z, t)$ are given by

$$\left.\begin{aligned} x &= \frac{x' + vt'}{\sqrt{1 - \frac{v^2}{c^2}}} \\ y &= y' \\ z &= z' \\ t &= \frac{t' + \frac{vx'}{c^2}}{\sqrt{1 - \frac{v^2}{c^2}}} \end{aligned}\right\} \quad \ldots(10.25)$$

If $v << c$, i.e., Lorentz transformations reduce to

$$x' = x - vt$$
$$y' = y$$
$$z' = z$$
$$t' = t$$

which are Galilean transformations. Thus, *for the low values of velocity v, the Lorentz transformations reduce to Galilean transformations.*

Example 10.2

Show the quantity $x^2 + y^2 + z^2 - c^2t^2$ is invariant under Lorentz transformation.

$$\left.\begin{aligned} x' &= \frac{x - vt}{\sqrt{1 - \frac{v^2}{c^2}}} \\ y' &= y \\ z' &= z \\ t' &= \frac{t - \frac{vx}{c^2}}{\sqrt{1 - v^2/c^2}} \end{aligned}\right\} \quad \text{...(10.24)}$$

These are called *Lorentz transformation equations.*

Inverse Lorentz Transformations: If we consider the event to be observed in the frame S' rather than in S, than the corresponding transformation, called, *Inverse Lorentz transformation* can be obtained by changing v to $-v$ and primed quantities into unprimed quantities and vice versa.

Thus, the transformations, i.e., equations relating coordinates in the frame $S'(x', y', z', t')$ to that in the frame S (x, y, z, t) are given by

$$\left.\begin{aligned} x &= \frac{x' + vt'}{\sqrt{1 - \frac{v^2}{c^2}}} \\ y &= y' \\ z &= z' \\ t &= \frac{t' + \frac{vx'}{c^2}}{\sqrt{1 - \frac{v^2}{c^2}}} \end{aligned}\right\} \quad \text{...(10.25)}$$

If $v << c$, i.e., Lorentz transformations reduce to

$$x' = x - vt$$
$$y' = y$$
$$z' = z$$
$$t' = t$$

which are Galilean transformations. Thus, *for the low values of velocity v, the Lorentz transformations reduce to Galilean transformations.*

Example 10.2

Show the quantity $x^2 + y^2 + z^2 - c^2t^2$ is invariant under Lorentz transformation.

Solution:

In order to prove that $x^2 + y^2 + z^2 - c^2t^2$ is invariant under Lorentz transformation, we need to show:

$$x^2 + y^2 + z^2 - c^2t^2 = x'^2 + y'^2 + z'^2 - c^2t'^2$$

where (x, y, z, t) and (x', y', z', t') are the coordinates of the same event observed by two observers in the frames S and S' respectively, with S' moving with velocity v relative to S. Let us consider the quantity $x^2 + y^2 + z^2 - c^2t^2$ and using Lorentz transformations, we have:

$$x^2 + y^2 + z^2 - c^2t^2 =$$

$$\frac{(x' + vt)^2}{1 - \frac{v^2}{c^2}} + y'^2 + z'^2 - c^2\left[\frac{\left(t' + \frac{vx'}{c^2}\right)^2}{1 - \frac{v^2}{c^2}}\right]$$

$$= \frac{1}{1 - \frac{v^2}{c^2}}\left[x'^2 + v^2t'^2 + 2x'vt' - c^2\left(t'^2 + \frac{v^2x'^2}{c^4} + \frac{2t'vx'}{c^2}\right)\right] + y'^2 + z'^2$$

$$= \frac{1}{1 - \frac{v^2}{c^2}}\left[x'^2 + v^2t'^2 + 2x'vt' - c^2t'^2 - \frac{v^2x'^2}{c^2} - 2t'vx'\right] + y'^2 + z'^2$$

$$= \frac{1}{1 - \frac{v^2}{c^2}}\left[\frac{c'^2x'^2 + c^2v^2t'^2 - c^4t'^2 - v^2x'^2}{c^2}\right] + y'^2 + z'^2$$

$$= \frac{1}{1 - \frac{v^2}{c^2}}\left[\frac{c^2(x'^2 - c^2t'^2) - v^2(x'^2 - c^2t'^2)}{c^2}\right] + y'^2 + z'^2$$

$$= \frac{1}{1 - \frac{v^2}{c^2}}\left[(x'^2 - c^2t'^2)\frac{(c^2 - v^2)}{c^2}\right] + y'^2 + z'^2$$

$$= \frac{1}{1 - \frac{v^2}{c^2}}\left[(x'^2 - c^2t'^2)\left(1 - \frac{v^2}{c^2}\right)\right] + y'^2 + z'^2$$

$$= x'^2 + y'^2 + z'^2 - c^2t'^2$$

Thus, the expression $x^2 + y^2 + z^2 - c^2t^2$ is invariant under Lorentz transformation.

10.8 consequences of lorentz transformation

We discuss below some of the interesting consequences or effects of relativistic or Lorentz transformation. A few of them appear to be quite strange and surprising. Due to the small relative motion between the frames of reference encountered in our daily life, we do not ordinarily come across any revealing relativistic phenomena.

10.8.1 Lorentz-Fitzgerald (Length) Contraction

Dutch physicist Lorentz and Irish physicist Fitzgerald independently put forward the hypothesis that *every body, moving at a velocity v, has its length contracted by the factor* $\sqrt{1-\frac{v^2}{c^2}}$ *in the direction of its motion,* while the dimensions in the direction perpendicular to the direction of motion remain unchanged. However, an observer at rest with respect to the body will not be aware of this contraction as the scale with which he measures will also get contracted by the same amount.

Let a rod be placed along the x-axis of the frame S (Fig. 10.3) with coordinates of its end points being $(x_1, 0, 0)$ and $(x_2, 0, 0)$. So the length of the rod measured by an observer stationary in the frame S is given by

$$\lambda_0 = x_2 - x_1.$$

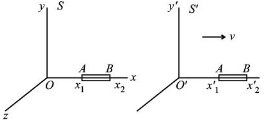


Fig. 10.3

The length measured by an observer in a frame, at rest relative to the rod is called its *proper length*, i.e., λ_0 *is the proper length of the rod.*

If an observer on the frame S', moving with uniform velocity v, tries to measure the length of the rod AB, he will measure the length to be $\lambda = x'_2 - x'_1$, where x'_1 and x'_2 are the end points $(x'_1, 0, 0)$ and $(x'_2, 0, 0)$ of the rod in the frame S'. Using Lorentz transformations, we have

$$x_1 = \frac{x'_1 + vt'}{\sqrt{1-\frac{v^2}{c^2}}} \text{ and } x_2 = \frac{x'_2 + vt'}{\sqrt{1-\frac{v^2}{c^2}}}$$

$$\therefore x_2 - x_1 = \frac{x'_2 - x'_1}{\sqrt{1-\frac{v^2}{c^2}}}$$

$$\text{or, } \lambda_0 = \frac{l}{\sqrt{l-\frac{v^2}{c^2}}}$$

$$\text{or, } \lambda = l_0\sqrt{1-\frac{v^2}{c^2}} \quad ...(10.26)$$

So, the observer in the frame S' finds the rod AB in the frame S to be contracted. This is called *Lorentz-Fitzgerald contraction.*

Reciprocity of Length Contraction: Imagine the rod to be placed at rest in the frame S' with space coordinates $(x'_1, 0, 0)$ and $(x'_2, 0, 0)$, so that the length of the rod measured by the observer O' in the frame S' is

$$\lambda_0 = x'_2 - x'_1$$

If an observer O in the frame S measures the length of the rod, he will observe

$$x'_1 = \frac{x_1 - vt}{\sqrt{1-\frac{v^2}{c^2}}} \text{ and } x'_2 = \frac{x_2 - vt}{\sqrt{1-\frac{v^2}{c^2}}}$$

$$\text{or, } \lambda_0 = x'_2 - x'_1 = \frac{x_2 - vt}{\sqrt{1-\frac{v^2}{c^2}}} - \left(\frac{x_1 - vt}{\sqrt{1-\frac{v^2}{c^2}}}\right)$$

$$= \frac{x_2 - x_1}{\sqrt{1-\frac{v^2}{c^2}}} = \frac{l}{\sqrt{1-\frac{v^2}{c^2}}}$$

$$\text{or, } \lambda = l_0\sqrt{1-\frac{v^2}{c^2}}$$

Thus, we find that every rigid body appears to be of maximum length in the coordinate system in which it is at rest, while the length appears to be contracted by $\sqrt{1-\frac{v^2}{c^2}}$, where v is the relative velocity.

Since the contraction depends upon the square of v it is unchanged when v is replaced by $-v$, i.e., contraction is reciprocal. The reciprocity is consistent with the first postulate. Because, if the contraction occurred only in one frame and not in the other, then we could determine the absolute velocity which is in contradiction with the postulates of relativity.

Further, due to the reciprocity, if there are two identical rods at rest, one in S and the other in S', each of the observers finds that the rod in the other frame is shorter. Also, as $v \rightarrow c$, from Eq. (10.26) we find that, $\lambda \rightarrow 0$ i.e., as v approaches the velocity of light, the measured length approaches zero. At $v = c$, $\lambda = 0$, i.e., if the rod moves with the velocity of light, its length reduces to zero, which is physically impossible.

10.8.2 Relativity of Time: Time Dilation

The time interval measured between two events that occur at the same point in a particular frame is a more fundamental quantity than the interval between events at different points. A term called ***proper time***, denoted by Dt_o, is used to describe the time interval between two events that occur at the same point in a frame of reference in which the clock is at rest.

Let us consider two frame of reference S and $S¢$, where $S¢$ is moving with uniform velocity v w.r.t. the frame S along $+x$-axis. Let two events take place at the same point in space. The time interval between these events as measured by an observer at rest in frame S, the rest frame, is Dt_o, the proper time. Thus, the proper time is given by

$$Dt_o = t_2 - t_1$$

where t_1 is the instant of the first event and t_2 that of the second.

For an observer in frame $S¢$, the corresponding time interval is given by

$$Dt = t¢_2 - t¢_1$$

where $t¢_1$ is the instant of the first event and $t¢_2$ that of the second.

Using Lorentz transformations, we have

$$t¢_1 = \frac{t_1 - \frac{vx}{c^2}}{\sqrt{1 - \frac{v^2}{c^2}}}$$

$$\text{and } t¢_2 = \frac{t_2 - \frac{vx}{c^2}}{\sqrt{1 - \frac{v^2}{c^2}}}$$

Substituting these in the expression for Dt, we get

$$Dt = \frac{t_2 - \frac{vx}{c^2}}{\sqrt{1 - \frac{v^2}{c^2}}} - \frac{t_1 - \frac{vx}{c^2}}{\sqrt{1 - \frac{v^2}{c^2}}}$$

$$= \frac{t_2 - t_1}{\sqrt{1 - \frac{v^2}{c^2}}}$$

$$\text{or } Dt = \frac{\Delta t_o}{\sqrt{1 - \frac{v^2}{c^2}}} \quad ...(10.27)$$

Since $v < c$, $g = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} > 1$, Dt is always larger than Dt_o, i.e., the time interval between two events occuring at a given point as observed by an observer in moving frame $S¢$ appears to be *longer* or *dilated* by the factor g compared to proper time. In other words, the *proper time interval is minimum*. This effect is known as ***time dilation*** and is equivalent to the slowing down of moving clocks.

A clock at rest in the frame S, will appear to run at a slower rate to an observer in the frame S'. Therefore, we can state a general rule: *every clock appears to run at its fastest rate when it is at rest, relative to the observer but appears to be slowed down by the factor* $\sqrt{1 - \frac{v^2}{c^2}}$ *relative to an observer moving at a velocity* v.

Related to the effect of time dilation is ***twin paradox***, i.e., of the two identical twins, if one goes at a high speed in a rocket and the other stays behind on the earth, the former, when he returns back to the earth, he will be younger than the latter.

If $v << c$, so that $\frac{v}{c} \rightarrow 0, \Delta t = \Delta t'$, i.e., the time intervals between the events in both the stationary and the moving frames are the same, as in classical physics.

In order to observe time dilation in the laboratory the following conditions must be satisfied.:

(*i*) The body should be moving with relativistic speed v,

(*ii*) The event occurring is independent of the speed v,

(*iii*) The time interval between the events should be sufficiently short so that the event occurs within a reasonably short travel of the body, otherwise the laboratory necessary will be inconveniently long.

Experimental Verification: Many fundamental particles disintegrate spontaneously after a very short time interval, called their *lifetime*. In a frame of reference S in which the particle is at rest, i.e., in its own frame of reference, the mean lifetime is denoted by t and is called the *proper lifetime* of the particle. In a frame of reference S' (say, laboratory frame) in which it is moving with velocity v relative to S, the lifetime gets dilated to:

$$t_1 = \frac{\tau}{\sqrt{1-\frac{v^2}{c^2}}} = \gamma\tau \left(\gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}\right) \quad ...(10.28)$$

The nuclear particles, π^+ mesons are produced with speed 0.99 c when high energy particles generated by a synchrocyclotron strikes a target. These mesons decay (break into μ mesons and neutrinos) such that in every time interval of 1.8×10^{-8} second, half of the particles disappear, i.e., its half life is 1.8×10^{-8} sec. If the flux of π^+ mesons is measured at two locations, 30 m apart the laboratory time interval Δt for travelling this distance is given by:

$$\Delta t = \frac{30\text{ m}}{0.99\times 3\times 10^8\text{ m/s}} = 10\times 10^{-8}\text{s}$$

Which is about 5.6 times its half lifetime 1.8×10^{-8} sec. If the half lifetime of the particles be T, then after time t, the fractional flux $\frac{N}{N_0}$ is given by:

$$\frac{N}{N_0} = \left(\frac{1}{2}\right)^{t/\tau} = \left(\frac{1}{2}\right)^{5.6} = 2^{-5.6} \approx 0.02 \text{ or } 2\%$$

Hence, the flux of π mesons should decrease to approximately 2% of the original flux in travelling 30 m. But the flux, actually observed was nearly 60%.

This discrepancy can be explained by calculating the proper time Δt by using the relation:

$$\Delta t = \frac{\Delta t_o}{\sqrt{1-\frac{v^2}{c^2}}}$$

$$\text{or, } \Delta t_o = \Delta t\sqrt{1-\frac{v^2}{c^2}}$$

$$= 10\times 10^{8}\sqrt{\left(1-\left(\frac{0.99c}{c}\right)^2\right)}$$

$= 1.4 \times 10^{-8}$ sec.

This is 0.78 times the half life. Hence, during this time the flux should reduce to $2^{-0.78} \approx 0.6$ or 60%, the same as observed.

From this experiment, it is clear that in the laboratory measurement, the time taken by π^+ mesons for 30 m travel is 10×10^{-8} second, while for the π^+ mesons themselves (imagine so) the time is 1.4×10^{-8} sec. Thus, a time dilation or elongation of 7.8 times occurs in this experiment.

10.8.3 Transformation of Velocities

We now want to obtain transformation equations connecting the velocities of a particle in the frames S and S'. Let the velocity in the frame S be $u(u_x, u_y, u_z)$ and that in S' be $u'(u'_x, u'_y, u'_z)$. Then:

$$u_x = \frac{dx}{dt} \text{ and}$$

$$u'_x = \frac{dx'}{dt'}$$

From Lorentz transformations Eq. (10.24), we have

$$dx' = \frac{dx - vdt}{\sqrt{1-\frac{v^2}{c^2}}}, dy' = dy, dz' = dz$$

$$\text{and, } dt' = \frac{dt - \frac{v}{c^2}dx}{\sqrt{1-\frac{v^2}{c^2}}}$$

Dividing dx', dy', dz' by dt', we get:

$$\frac{dx'}{dt'} = \frac{dx - vdt}{dt - \frac{v}{c^2}dt}$$

$$\frac{dy'}{dt'} = \frac{dy\sqrt{1-\frac{v^2}{c^2}}}{dt - \frac{v}{c^2}dx}$$

$$\frac{dz'}{dt'} = \frac{dz\sqrt{1-\frac{v^2}{c^2}dx}}{dt - \frac{v}{c^2}dx}$$

Further, dividing the numerator and denominator of the right-hand sides of all the above equations by dt, we get

$$\left.\begin{aligned} u'_x &= \frac{u_x - v}{1 - \frac{v}{c^2}u_x} \\ u'_y &= \frac{u_y\sqrt{1-\frac{v^2}{c^2}}}{1 - \frac{v}{c^2}u_x} \\ u'_z &= \frac{u_z\sqrt{1-\frac{v^2}{c^2}}}{1 - \frac{v}{c^2}u_x} \end{aligned}\right] \quad ...(10.29)$$

The inverse transformations are:

$$\left.\begin{aligned} u_x &= \frac{u'_x + v}{1 + \frac{v}{c^2}u'_x} \\ u_y &= \frac{u'_y\sqrt{1-\frac{v^2}{c^2}}}{1 + \frac{v}{c^2}u'_x} \\ u_z &= \frac{u'_z\sqrt{1-\frac{v^2}{c^2}}}{1 + \frac{v}{c^2}u'_x} \end{aligned}\right] \quad ...(10.30)$$

In the non-relativistic approximation $\frac{v}{c} << 1,$ Eq. (10.30) reduces to

$$u_x = u'_x + v$$
$$u_y = u'_y$$
$$u_z = u'_z$$

which are the classical results, i.e., the same as the Galilean transformations. When the particle is moving along x-axis,

$$u_x = u \text{ and}$$
$$u_y = u_z = 0.$$

Thus, from Eq. (10.29), we get:

$$u'_x = u' = \frac{u - v}{1 - \frac{v^2}{c^2}u} \quad ...(10.31)$$

$$u'_y = u'_z = 0$$

When, $u = c$,

$$u' = \frac{c - v}{1 - \frac{v}{c}} = c.$$

That is, *when a particle is moving with velocity of light* c (not possible for a particle having non-zero rest mass) relative to *S*, *its velocity as observed by an observer in the frame S′ is still* c. This illustrates that velocity transformations are consistent with the hypothesis of relativity, i.e., constancy of velocity of light. This is natural since the Lorentz transformations are based on the principle of consistency of velocity of light.

10.8.4 Relativity of Simultaneity

From Newtonian or Galilean relativity, it follows that if two events are simultaneous for one observer, it will be simultaneous for all other observers, irrespective of the state of motion of the source and the observer. It is not so in Lorentzian relativity.

Consider frames S and S' moving with uniform velocity v such that their coordinate systems coincide once, as in Fig. 10.1. If two events occur simultaneously at the instant t at $x = x_1$ and $x = x_2$ in the frames S, the two events are not simultaneous in the frame S', as can be seen below. The times of occurrences in the frame S' at the two locations x_1 and x_2 respectively are:

$$t'_1 = \frac{t - \frac{vx_1}{c^2}}{\sqrt{1-\frac{v^2}{c^2}}} \text{ and } t'_2 = \frac{t - \frac{vx_2}{c^2}}{\sqrt{1-\frac{v^2}{c^2}}}$$

$$\therefore t'_2 - t'_1 = \frac{\frac{v}{c^2}(x_1 - x_2)}{\sqrt{1-\frac{v^2}{c^2}}} \quad ...(10.32)$$

As, x_1 ? x_2, therefore, t'_1 ? t'_2. That is, the two events are not simultaneous, when observed from the frame S'. Thus, the concept of simultaneity is relative and not absolute.

10.8.5 Relativity of Mass

In classical physics, it is assumed that the mass of a body remains constant irrespective of the fact whether the body is at rest or is in motion. Using Lorentzian relativity, we shall see that it is not so. The change in mass with velocity will obviously change the momentum, energy and force.

Consider two identical and perfectly elastic particles of masses m'_1 and m'_2 in the frame S', moving with velocities $+ u'$ and $- u'$ parallel to the x'-axis, undergo a head-on collision. Due to the collision, the particle comes to rest momentarily and then rebounds under the elastic forces and moves back with velocities $-u'$ and $+u'$ respectively [Fig. 10.4 (*b*)].

Let us consider the same collision as viewed from the frame *S*, which is moving with velocity $-v$ relative to *S'* along the *x*-axis. Let the particles of masses μ_1 and μ_2 have velocities u_1 and u_2, before the collision with respect to the frame *S*. At the instant of collision, the colliding particles come to rest relative to each other. For the instant when the particles are together, let the combined mass be μ, as measured in the frame *S* [Fig. 10.4 (*a*)]. At the instant of collision, the colliding particles are at rest with respect to the frame *S'* but move with velocity $+v$ relative to the frame *S*. Applying the law of conservation of momentum, we have:

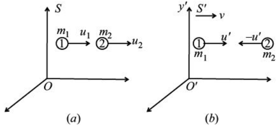


Fig. 10.4

$$\mu_1 u_1 + \mu_2 u_2 = Mv \quad ...(10.33)$$

where $\mu_1 + \mu_2 = \mu$.

Using the relation

$$u_x = \frac{u'_x + v}{1 + \dfrac{v u'_x}{c^2}} \text{ for}$$

$$\text{Particle 1: } u_1 = \frac{u' + v}{1 + \dfrac{v u'}{c^2}} \quad ...(10.34)$$

$$\text{Particle 2: } u_2 = \frac{-u' + v}{1 - \dfrac{v u'}{c^2}} \quad ...(10.35)$$

The direction or sign of u_2 depends on the relative magnitudes of u' and v. Substituting the above values of u_1 and u_2 in Eq. (10.33), we get

$$m_1\left(\frac{u' + v}{1 + \dfrac{u v}{c^2}}\right) + m_2\left(\frac{-u' + v}{1 - \dfrac{v u'}{c^2}}\right) = (m_1 + m_2)v$$

or

$$m_1\left(\frac{u' + v}{1 + \dfrac{u v'}{c^2}} - v\right) = m_2\left(v + \frac{u' - v}{1 - \dfrac{v u'}{c^2}}\right)$$

Simplifying and cancelling the common terms, we get

$$\frac{m_1}{m_2} = \frac{1 + \dfrac{v u'}{c^2}}{1 - \dfrac{v u'}{c^2}} \quad ...(10.36)$$

From Eqs. (10.34) and (10.35), it can be shown:

$$\frac{u' v}{c^2} = \frac{2c^2 - u_1^2 - u_2^2 - 2\sqrt{(c^2 - u_1^2)(c^2 - u_2^2)}}{u_1 - u_2} \quad ...(10.37)$$

Substituting the value of $\dfrac{u' v}{c^2}$ from Eq. (10.37) in Eq. (10.36) and simplifying further, we get:

$$\frac{m_1}{m_2} = \frac{\sqrt{1 - \dfrac{u_2^2}{c^2}}}{\sqrt{1 - \dfrac{u_1^2}{c^2}}} \quad ...(10.38)$$

If mass μ_2 in the system *S* is at rest before collision, i.e., $u_2 = 0$, then

$$\mu_1 = \frac{m_2}{\sqrt{1 - \dfrac{u_1^2}{c^2}}} \quad ...(10.39)$$

In Eq. (10.39), μ_1 is the mass of the particle moving with velocity u_1 and μ_2 is its mass when its velocity is zero. The mass of a particle when it is at rest is called its rest (or proper) mass, and is usually denoted by μ_0. Thus, using $\mu_1 = \mu$, $\mu_2 = \mu_0$ and $u_1 = v$, Eq. (10.39) can be expressed as:

$$\mu = \frac{m_0}{\sqrt{1-\frac{v^2}{c^2}}} \quad \text{...(10.40)}$$

This expression gives the variation of the mass of a particle with velocity. The mass of a particle, when it is moving with velocity v comparable to speed of light c is called its *relativistic* or *effective mass*. Thus, for the conservation of both mass and momentum to hold during collision, the mass of the particle moving with velocity v with respect to the frame S increases with the increase of velocity. Hence, *the rest mass of a particle is the lowest possible mass of a particle*.

It is implied from Eq. (10.40) that the mass of a particle goes on increasing with increasing velocity, the variation is shown in Fig. 10.5. When $v = c$, $\mu = \frac{m_0}{0} = \infty$, i.e., *when the velocity of a particle approaches the velocity of light, the mass of the particle becomes infinite. This indicates that the speed of light c is a limiting velocity, unattainable by a material body*.

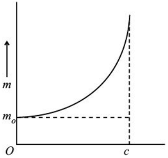


Fig. 10.5

When $v < c$, Eq. (10.40) can be expressed as

$$\mu = m_0\left[1+\frac{1}{2}\left(\frac{v^2}{c^2}\right)+\frac{3}{8}\left(\frac{v^2}{c^2}\right)^2+\ldots\right]$$

For $v << c$, neglecting $\frac{v^2}{c^2}$ and higher order terms, $\mu = \mu_0$, i.e., mass of the particle is equal to the rest mass of the particle, as in classical physics.

10.8.6 Relativistic Momentum and Force

Momentum of a particle of mass μ moving with velocity v is given by

$$p = mv = \frac{m_0 v}{\sqrt{1-\frac{v^2}{c^2}}}$$

Force acting on a body is equal to its rate of change of momentum,

$$\varphi = \frac{dp}{dt}$$

$$= \frac{d}{dt}\left(\frac{m_0 v}{\sqrt{1-\frac{v^2}{c^2}}}\right)$$

$$= \frac{m_0}{\left(1-\frac{v^2}{c^2}\right)^{\frac{3}{2}}}\frac{dv}{dt} \quad \text{...(10.41)}$$

Zero rest mass of a photon: The momentum of a particle is given by:

$$p = \frac{m_0 v}{\sqrt{1-\frac{v^2}{c^2}}}$$

According to de-Broglie hypothesis,

$$p = \frac{h}{\lambda}$$

A photon travels with the speed of light $v = c$, thus

$$\mu_0 = 0,$$

That is, the rest mass of a photon is zero.

10.9 Einstein's Mass-Energy Relation

According to work-energy theorem, the kinetic energy of a moving body is equal to the amount of work done by the external force on the body from rest. If φ is the force acting on the body, then work done by the force in increasing its velocity v from 0 to v is given by:

$$W = \int_0^v F \cdot ds$$

where ds is the small displacement in the direction of the force. Kinetic energy of the body ε_k is:

$$\varepsilon_k = W = \int_0^v F \cdot ds$$

$$= \int_0^v F\cdot\frac{ds}{dt}dt = \int_0^v F\cdot v\,dt$$

$$= \int_0^v \frac{dp}{dt}\cdot v\,dt \quad \left(\because F=\frac{dp}{dt}\right)$$

$$= \int_0^v v\frac{d}{dt}(mv)dt \quad (\because p = mv)$$

$$= \int_0^v v\cdot d(mv)$$

But, from Lorentzian relativity, we know $m = \frac{m_0}{\sqrt{1-\frac{v^2}{c^2}}}.$ Therefore,

$$\varepsilon_k = \int_0^v v\cdot d\left(\frac{m_0}{\sqrt{1-\frac{v^2}{c^2}}}v\right)$$

$$= m_0\int_0^v v\,d\left[v\left(1-\frac{v^2}{c^2}\right)^{-\frac{1}{2}}\right]$$

$$= m_0\int_0^v v\left[\left(1-\frac{v^2}{c^2}\right)^{-\frac{1}{2}}dv - v\cdot\frac{1}{2}\left(1-\frac{v^2}{c^2}\right)^{-\frac{3}{2}}\left(-\frac{2v}{c^2}\right)dv\right]$$

$$= m_0\int_0^v v\left(1-\frac{v^2}{c^2}\right)^{-\frac{3}{2}}\left[1-\frac{v^2}{c^2}+\frac{v^2}{c^2}\right]dv$$

$$= m_0\int_0^v \frac{v\,dv}{\left(1-\frac{v^2}{c^2}\right)^{\frac{3}{2}}}$$

Substituting $1-\frac{v^2}{c^2}=y,$ so that $-\frac{2v\,dv}{c^2}=dy,$ we get

$$\varepsilon_k = m_0\int_1^{1-\frac{v^2}{c^2}} \frac{1\left(\frac{c^2}{2}\right)dy}{y^{3/2}} \quad \left[\begin{array}{l}\because v=0,\ y=1\\ v=v,\ y=1-\frac{v^2}{c^2}\end{array}\right]$$

$$= \frac{m_0c^2}{2}\left[\frac{y^{-1/2}}{-\frac{1}{2}}\right]^{1-\frac{v^2}{c^2}}$$

$$= m_0c^2\left[\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}-1\right] = \frac{m_0c^2}{\sqrt{1-\frac{v^2}{c^2}}} - m_0c^2$$

or $\varepsilon_k = (\mu - \mu_0)c^2$ ...(10.42)

This is a *relativistic expression for kinetic energy*. Thus, *the kinetic energy for a moving body is equal to the gain in mass times the square of speed of light*. Since μ_0 is the rest mass of the body, the term $\mu_0 c^2$ is called the *rest energy* which may be regarded as *internal stored energy* in the body. The total energy of the body is the sum of the rest energy and the relativistic kinetic energy of the body, i.e.,

$$\varepsilon = \text{rest energy} + \text{relativistic kinetic energy}$$
$$= \mu_0 c^2 + (\mu - \mu_0)c^2$$
$$= mc^2 \quad ...(10.43)$$

$\varepsilon = mc^2$ is the famous *Einstein mass-energy relation*, which states a universal equivalence between mass and energy.

Non-relativistic case: The relativistic kinetic energy:

$$\varepsilon_k = (\mu - \mu_0)c^2$$

$$= \left(\frac{m_0}{\sqrt{1-\frac{v^2}{c^2}}} - m_0\right)c^2$$

$$= \left[\left(1-\frac{v^2}{c^2}\right)^{-1/2} - 1\right]m_0c^2.$$

In non-relativistic cases, $v < < c$, so expanding the above expression by bionomical theorem:

$$\varepsilon_k = \left[1+\frac{1}{2}\frac{v^2}{c^2}+\frac{3}{8}\frac{v^4}{c^4}+\ldots-1\right]m_0c^2$$

Neglecting the terms $\frac{v^4}{c^4}$ and higher orders, we have:

$$\varepsilon_k = \frac{1}{2}\cdot\frac{v^2}{c^2}\times m_0c^2 = \frac{1}{2}m_0v^2 \qquad ...(10.44)$$

which is the same usual expression for kinetic energy. Thus, we can conclude that for non-relativistic cases ($v < < c$), the kinetic energy is $\frac{1}{2}mv^2$ while in relativistic cases, the correct expression is $(\mu - \mu_0)c^2$.

Also, as $v \to c$ or $\frac{v}{c} \to 1,$ the mass $m = \frac{m_0}{\sqrt{1-\frac{v^2}{c^2}}}$ and kinetic energy tends to infinity which would require an infinite amount of external work to be done on the particle to accelerate it to the speed of light. This indicates the impossibility of attaining the speed of light by a material particle having finite rest mass.

Examples of Mass-Energy Interconversion

***(i)* Pair Production:** When a photon of energy, greater than or equal to 1.02 MeV, passes close to an atomic nucleus, it disappears and a pair of electron and positron is created. It is represented as:

g (gamma photon) $\to \varepsilon^-$ (electron) + ε^+ (positron).

This phenomenon is known as *pair-production*. In this process, the *energy* of g-ray photon (a particle of zero rest mass) is converted *into mass* in accordance with mass-energy relation. The rest energy of both electron and positron is 0.51 MeV and that is why the energy of the photon must be at least 1.02 eV. If the photon has energy higher than 1.02 MeV, the remaining energy appears as the kinetic energy of created particles and recoil nucleus.

***(ii)* Pair-Annihilation:** When an electron and a positron encounter each other, they annihilate each other and an equivalent amount of energy is produced in the form of g-ray photons. This is known as *pair-annihilation* and is represented as:

$$\varepsilon^- + \varepsilon^+ \to g.$$

Both the processes are in accordance with the laws of conservation of momentum, conservation of charge and mass-energy relation. The process of pair production and pair- annihilation also occurs in other particle-antiparticle pairs such as proton-anti-proton, etc.

10.10 Momentum-energy relation

Momentum of a particle of mass μ and moving with velocity v is given by:

$$p = mv = \frac{m_0v}{\sqrt{1-\frac{v^2}{c^2}}}$$

The total energy of this particle is

$$\varepsilon = mc^2 = \frac{m_0c^2}{\sqrt{1-\frac{v^2}{c^2}}}$$

$$\therefore \varepsilon^2 = \frac{m_0^2c^4}{1-\frac{v^2}{c^2}}$$

Putting the above values in the expression $p^2c^2 + \mu_0^2c^4$, we get

$$p^2c^2 + \mu_0^2c^4 = \frac{m_0^2v^2c^2}{1-\frac{v^2}{c^2}} + m_0^2c^4 = \frac{m_0^2v^2c^2 + m_0^2c^4 - m_0^2v^2c^2}{1-\frac{v^2}{c^2}}$$

$$= \frac{m_0^2c^4}{1-\frac{v^2}{c^2}} = E^2$$

$$\therefore \varepsilon^2 = p^2c^2 + \mu_0^2c^4$$

$$\text{or, } \varepsilon = \sqrt{p^2c^2 + m_0^2c^4} \qquad ...(10.45)$$

This is Einstein's well-known relation between momentum and energy.

Thus, for a particle at rest ($p = 0$), $\varepsilon = \mu_0c^2$. ...(10.46)

Equation (10.45) also implies that a particle may have energy and momentum even if its rest mass is zero, i.e., if $\mu_0 = 0$

$\varepsilon = pc$ ...(10.47)

For example, photon has zero rest mass but possess energy and momentum. Such particles always travel at speed of light in vacuum. Photons and neutrinos are the particles in the category.

Example 10.3

How fast does a rocket have to move relative to an observer for its length to be contracted to 95% of its length at rest?

Solution:

Here, $\lambda = 0.95\,\lambda_0$, v =?

$$\lambda = l_0\sqrt{1-\frac{v^2}{c^2}}$$

$$\frac{l}{l_0} = 0.95 = \sqrt{1-\frac{v^2}{c^2}}$$

$$\text{or, } 1-\frac{v^2}{c^2} = (0.95)^2$$

$$\text{or, } \frac{v^2}{c^2} = 1 - 0.9025 = 0.0975$$

$$\text{or, } v = \sqrt{0.0975}\cdot c$$

$$= 0.31c$$

Example 10.4

A cube with its sides of proper length λ_0 is viewed from a reference frame moving with uniform velocity v, parallel to an edge of the cube. Give the expression for the volume of the cube for the observer.

Solution:

The rest volume or the volume of the cube for an observer at rest with respect to the cube = $\lambda_0 \times \lambda_0 \times \lambda_0 = \lambda_0^3$. However, when viewed by an observer from a reference frame moving with uniform velocity v parallel to an edge of the cube, the length λ_0 of the edge of the cube will get contracted to $L_0\sqrt{1-v^2/c^2}$. However, no such contraction will be observed in the length of the other two edges. Therefore, the volume observed by the observer is:

$$= L_0\sqrt{1-v^2/c^2}\times L_0\times L_0$$

$$= L_0^3\sqrt{1-v^2/c^2}.$$

Example 10.5

A rod of length 1 m is moving along its length with a velocity 0.5c. Calculate the length of the rod as observed by an observer (a) on the earth, (b) moving with the rod itself.

Solution:

Here λ_0 = 1 m

v = 0.5c

(a) Let λ be the length as observed by an observer on the earth, a stationary frame.

$$\lambda = l_0\sqrt{1-\frac{v^2}{c^2}}$$

$$= 1\text{m}\times\sqrt{1-\frac{(0.5c)^2}{c^2}} = \sqrt{1-0.25} = \sqrt{0.75}$$

$$= 0.87 \text{ m.}$$

(b) As the observer is moving with the rod itself, i.e., the rod is stationary with respect to the observer, the observer will observe its proper length of 1 m.

Example 10.6

A rod placed in a frame of reference, is moving with a velocity 0.8c in a direction (a) parallel to its length and (b) at an angle 45° with its length. Calculate the percentage contraction in each case. What is the orientation of the rod in the moving frame of reference in the case (b)?

Solution:

(a) Let λ_0 be the length of the rod in a reference frame S at rest. Its length in frame S' moving with velocity 0.8c relative to S in a direction parallel to its length is

$$L' = L_0\sqrt{1-\frac{v^2}{c^2}}$$

$$= L_0\sqrt{1-\frac{(0.8c)^2}{c^2}}$$

$$= L_0\sqrt{1-0.64}$$

$$= L_0\sqrt{0.36}$$

$$= 0.6\,\lambda_0$$

Percentage contraction

$$= \frac{L_0 - 0.6L_0}{L_0}\times 100$$

$$= 40\%.$$

(b) Component of the proper length of the rod along the direction of motion = λ_0 cos

$45° = 0.71\,\lambda_o$.

Component of the proper length λ_o perpendicular to the direction of motion = λ_o sin $45° = 0.71\,\lambda_o$.

Only the former component, i.e., component along the direction of motion, undergoes length contraction in the frames S' and let it be denoted by λ'_{ox}.

$$\therefore \lambda'_{ox} = 0.71 L_0\sqrt{1-\frac{(0.8c)^2}{c^2}}$$

$$= 0.71\,\lambda_o \times 0.6$$

$$= 0.426\,\lambda_o.$$

The other component remains unchanged at $0.71\,\lambda_o$.

Total length in the frame S',

$$\lambda\times = \sqrt{(0.71L_0)^2 + (0.426L_0)^2}$$

$$= 0.82\,\lambda_o$$

Percentage contraction

$$= \frac{L_0 - 0.82L_0}{L_0} \times 100 = 18\%.$$

If θ is the angle that the length of the rod appears to make with the direction of velocity v of the frame S', then:

$$\tan\theta = \frac{\text{length component perpendicular to } v}{\text{length component parallel to } v}$$

$$= \frac{0.71L_0}{0.82L_0}$$

$$= 0.86$$

$$\text{or, } \theta = 40°\,53'$$

Thus, the rod makes an angle of $40°\,53'$ with its direction of motion.

Example 10.7

A rocket of length 100 m takes off from the ground. At what speed will an observer on the ground, observe its length to be 99 m?

Solution:

Here, $\lambda_o = 100$ m, $\lambda = 99$ m, v =?

$$99\text{ m} = 100\text{m}\sqrt{1-\frac{v^2}{c^2}}$$

$$\text{or, } 1-\frac{v^2}{c^2} = \frac{99\times 99}{100\times 100}$$

$$\text{or, } \frac{v^2}{c^2} = 1-\frac{9801}{10000} = \frac{199}{10000}$$

$$\text{or, } v = \frac{\sqrt{199}}{100}\times 3\times 10^8 \text{ m/s}$$

$$= 4.23 \times 10^7 \text{ m/s}.$$

Example 10.8

A particle with a mean proper lifetime of 2 μs moves through the laboratory with a speed of 0.9c. Calculate its lifetime as measured by an observer in the laboratory.

Solution:

Here: $\Delta t_o = 2\ \mu s = 2 \times 10^{-6}$ s, $v = 0.9c$, Δt =?

$$\therefore \Delta t = \frac{\Delta t_o}{\sqrt{1-\frac{v^2}{c^2}}}$$

$$= \frac{2\times 10^{-6}}{\sqrt{1-\frac{(0.9c)^2}{c^2}}}$$

$$= \frac{2\times 10^{-6}}{\sqrt{1-0.81}}$$

$$= \frac{2\times 10^{-6}}{\sqrt{0.19}}$$

$$= \frac{2}{0.43}\times 10^{-6}\text{ sec}$$

$$= 4.6 \times 10^{-6}\text{ sec}.$$

Example 10.9

The proper lifetime of a π meson is 2.5×10^{-8} sec. What is its velocity if its mean life is observed to be 2.5×10^{-7} sec?

Solution:

Here, $\Delta t_o = 2.5 \times 10^{-8}$ sec, $\Delta t = 2.5 \times 10^{-7}$ sec, v =?

$$\Delta t = \frac{\Delta t}{\sqrt{1-\frac{v^2}{c^2}}}$$

$$\text{or, } 1-\frac{v^2}{c^2} = \left(\frac{\Delta t_o}{\Delta t}\right)^2 = \left(\frac{2.5\times10^{-8}}{2.5\times10^{-7}}\right)^2 = 10^{-2}$$

$$\text{or, } \frac{v^2}{c^2} = 1 - 0.01 = 0.99$$

$$\text{or, } v = \sqrt{0.99}\,c = 0.995c.$$

Example 10.10

A burst of μ-mesons having mean proper lifetime of 2×10^{-6} sec are produced at a certain height in the atmosphere. These μ-mesons travel towards the earth at a speed of 0.99*c*. If only 1% of μ mesons produced reach the earth, estimate the height at which the burst takes place.

Solution:

The process is radioactive decay, which is governed by the decay law:

$$N = N_o e^{-t/\tau}$$

where N_o is the initial concentration, N is concentration at the instant t and t mean lifetime. Here,

$$N = \frac{N_0}{100}$$

$$\therefore \frac{1}{100} = e^{-t/2\times10^{-6}}$$

Taking log of both sides and simplifying, we get:

$$t = 2 \times 10^{-6} \times 2.302 \times \log_{10} 100$$

$$= 2 \times 10^{-6} \times 2.302 \times 2 = 9.208 \times 10^{-6} \text{ sec}$$

The time t' measured by an observer on earth:

$$t' = \frac{t}{\sqrt{1-\frac{v^2}{c^2}}}$$

$$\text{or, } t' = \frac{9.208\times10^{-6}}{\sqrt{1-\frac{(0.99c)^2}{c^2}}} = \frac{9.208\times10^{-6}}{\sqrt{1-0.98}} = \frac{9.208\times10^{-6}}{\sqrt{0.02}}$$

$$= \frac{9.208}{0.14}\times10^{-6}\text{ sec}$$

Height at which μ-meson is produced:

$$= 0.99c \times 65.77 \times 10^{-6}$$

$$= 0.99 \times 3 \times 10^{8} \times 65.77 \times 10^{-6} \text{ m} = 1.95 \times 10^{4} \text{ m}.$$

Example 10.11

There are two observers, Ram on the earth and Shyam is in a rocket, whose speed is 2×10^8 m/s, both set their watches at 10:00, when the rocket takes off (*a*) When Ram's watch reads 10:30, he looks at Shyam's watch through a telescope, (*b*) When Shyam's watch reads 10:30 he looks at Ram's watch through a telescope. What do the two watches read respectively?

Solution:

Here, proper time interval Δt is 30 minutes.

Therefore, the time interval $\Delta t'$ as recorded by Ram in Shyam's watch in the moving rocket ship is:

$$\Delta t = \frac{\Delta t_o}{\sqrt{1-\frac{v^2}{c^2}}} = \frac{30}{\sqrt{1-\left(\frac{2\times10^8}{3\times10^8}\right)^2}} = \frac{30}{\sqrt{1-\frac{4}{9}}} \frac{30}{\sqrt{\frac{5}{9}}} = 40.25 \text{ sec}$$

Thus, Shyam's watch, as seen by Ram, will read 10:40.25 when his own watch reads 10:30.

Since time dilation is a reciprocal effect, Ram's watch, as seen by Shyam through the telescope will also read 10:40.25 when his own watch reads 10:30.

Example 10.12

On her 20th birthday, a young lady decides that she would like to remain 20 for the next 10 years. In order to do that, she takes a journey into outer space with uniform velocity. What should be the speed with which she should move, relative to the laboratory, so that when she returns after 10 years (relative to the laboratory) she can still say that she is only 20. Take one year as 360 days.

Solution:

The speed should be such that the period of 10 years (in the laboratory) is equivalent to a day (recorded by a moving watch) which is approximately $\frac{1}{360}$ years.

Here, $\Delta t_o = 10$ years

$$\Delta t = \frac{1}{360} \text{ years, } v = ?$$

$$\frac{1}{360} = 10\sqrt{1-\frac{v^2}{c^2}}$$

$$\text{or, } \sqrt{1-\frac{v^2}{c^2}} = \frac{1}{3600} = \frac{1}{(60)^2}$$

$$\text{or, } \frac{v^2}{c^2} = 1-\frac{1}{(60)^4} = \frac{(60^2+1)(60^2-1)}{60^4}$$

$$\text{or, } \frac{v}{c} = 0.999$$

Hence, $v = 0.999\, c$.

With this example one can realize that if some one moves with $v = c$, his/her aging stops.

Example 10.13

Determine the time (as measured by a clock at rest on rocket) taken by the rocket to reach a distant star and return to earth with a constant velocity $v = \sqrt{0.9999}\, c$, if the distance of the star is 4 light years.

Solution:

Time taken by the rocket in the round trip, as measured by a clock on the earth is:

$$t = \frac{(4+4)c}{\sqrt{0.9999}c} \cong 8 \text{ years}$$

The time t_o, as measured by a clock on the rocket, in the round trip, which is the proper time, is:

$$t_o = t\sqrt{1-\frac{v^2}{c^2}}$$

Here, $v = \sqrt{0.9999}c$

$$\therefore v^2 = 0.9999c^2 \Rightarrow \frac{v^2}{c^2} = 0.9999$$

$$\therefore t = 8\sqrt{1-0.9999} = 8\sqrt{0.0001} = 0.01\times 8 = 0.08 \text{ years}$$

Thus, the time for the round trip, as measured on the clock by the rocket itself is 0.08 years.

Example 10.14

At what speed should a clock be moved so that it may appear to lose 1 minute in each hour?

Solution:

Here, $t' = 60$ m $t = 60 - 1 = 59$ m, $v = ?$

$$\therefore t = t'\sqrt{1-\frac{v^2}{c^2}}$$

$$59 = 60\sqrt{1-\frac{v^2}{c^2}}$$

$$\text{or, } \frac{v}{c} = \sqrt{1-\left(\frac{59}{60}\right)^2} = \sqrt{\frac{119}{3600}} = 0.18$$

or, $v = 0.18 \times 3 \times 10^8$ m/s $= 5.4 \times 10^7$ m/s.

Example 10.15

A wrist watch keeps correct time on the surface of earth. A pilot puts on this watch and leaves the earth in a spaceship which moves at a constant speed of 10^7 m/s. Calculate the number of seconds which the watch loses per day with respect to an observer on the earth.

Solution:

The time dilation expression is

$$T' = \frac{T}{\sqrt{1-\frac{v^2}{c^2}}}$$

If ΔT is the time lost by the wristwatch per day, then the proper time is

$$T = T' - \Delta T$$

$$\therefore T' = \frac{T'-\Delta T}{\sqrt{1-\frac{v^2}{c^2}}}$$

$$\therefore \quad = \frac{\Delta T}{\sqrt{1-\frac{v^2}{c^2}}}$$

$$T'\left(\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}-1\right) \Rightarrow \Delta T = T'\left[1-\sqrt{1-\frac{v^2}{c^2}}\right]$$

$T' = 1$ day $= 24 \times 60 \times 60 = 86400$ sec

Loss in time Δt =

$$86400\left[1-\sqrt{\left\{1-\left(\frac{10^7}{3\times10^8}\right)^2\right\}}\right]$$

$$= 86400\left[1-\sqrt{\left(1-\frac{1}{900}\right)}\right]$$

$$= 86400\left[1-\left(1-\frac{1}{900}\right)^{1/2}\right]$$

Using bionomial theorem and neglecting higher order terms, we get

$$\Delta T = 86400\left[1-\left(1-\frac{1}{2\times900}\right)\right]=\frac{86400}{1800}=48\text{ sec.}$$

Example 10.16

Two particles approach each other with a speed 0.8c with respect to the laboratory. What is their relative speed?

Solution:

The problem consists of two velocities – 0.8c and 0.8c observed from laboratory and therefore, to find the required velocity, we will have to apply the theorem of addition of velocities. For the solution of the problem, consider a system S in which the particle having velocity – 0.8c is at rest. So, we can suppose that the system S' i.e., laboratory is moving with velocity +0.8c relative to the system S, i.e., v = 0.8c. Now we have to find the velocity of the particle moving with velocity +0.8c relative to the system S' or laboratory in the system S. From the law of addition of velocities, we have:

$$u_x = \frac{u'+v}{1+\frac{u'v}{c^2}}=\frac{0.8c+0.8c}{1+\frac{(0.8c)(0.8c)}{c^2}}$$

$$= \frac{0.6c}{1.64}=0.9756c$$

Thus, the relative speed of the particles is 0.9756c.

Example 10.17

Two β-particles A and B approach each other with a velocity of 0.9c. What is their relative velocity (a) as observed by A and (b) as observed by a stationary observer?

Solution:

The problem is identical to the previous one.

Here u' = 0.9c, v = 0.9c

$$\therefore u_x = \frac{u'+v}{1+\frac{u'v}{c^2}}=\frac{0.9c}{1+\frac{(0.9c)(0.9c)}{c^2}}=\frac{1.8c}{1+0.81}=\frac{1.8c}{1.81}$$

$$= 0.9945c$$

The relative speed of particles in both cases (a) and (b) is 0.9945c.

Example 10.18

A man leaves the earth in a rocket ship and makes a round trip to the nearest star which is 4 light years away at a speed of 0.8c. How much younger will he be on his return compared to his twin brother who did not take journey?

Solution:

Due to length contraction, the apparent length of the travel is:

$$\lambda = l_0\sqrt{1-\frac{v^2}{c^2}}$$

$$= 4ly\times\sqrt{1-\frac{(0.8c)^2}{c^2}}$$

$$= 4\times0.6ly = 2.4ly$$

Total apparent length of the round trip is:

$$= 2\times2.4ly = 4.8ly$$

$$\therefore \text{Total time taken } t = \frac{\text{distance}}{\text{speed}}=\frac{4.8ly}{0.8c}=6\text{ years}$$

Due to time dilation, the dilated time interval is:

$$t' = \frac{t}{\sqrt{1-\frac{v^2}{c^2}}}$$

$$= \frac{6}{\sqrt{1-\left(\frac{0.8c}{c}\right)^2}}\text{ years}=\frac{6}{0.6}\text{ years}=10\text{ years}$$

So, he will be younger than his twin brother by (10 – 6) = 4 years, who was at rest on the earth.

Example 10.19

A nuclear scientist observes that a certain atom A, moving relative to him, with velocity 2 × 10^8 m/s emits a particle B which moves with a velocity of 2.8×10^8 m/s with respect to the atom. Calculate the velocity of the emitted particle relative to the scientist.

Solution:

Consider the particle A to be in system S' and the scientist to be in the system S, so that:

$$u_x = \frac{u' + v}{1 + \frac{u'v}{c^2}}$$

Here, u' = velocity of the particle B relative to $S' = 2.8 \times 10^8$ m/s

u = velocity of the system S' relative to $S = 2 \times 10^8$ m/s

u' = velocity of the particle B relative to S =?

$$\therefore u' = \frac{2.8 \times 10^8 + 2 \times 10^8}{1 + \frac{(2.8 \times 10^8) \times (2 \times 10^8)}{(3 \times 10^8)^2}} = \frac{4.8 \times 10^8}{1 + \frac{5.6}{9}}$$

$$= 2.95 \times 10^8 \text{ m/s.}$$

Example 10.20

Show that if in the S' frame we have $v'_y = c\sin\theta$ and $v'_x = c\cos\theta$, in the frame S, $v^2_x + v^2_y = c^2$. Frame $S\times$ moves with velocity V_i with respect to the frame S.

Solution:

Since frame S' is moving relative to S in the x-direction with velocity V, we have, using inverse Lorentz transformation:

$$v_x = \frac{v'_x + V}{1 + \frac{v'_x V'}{c^2}}, v_y = \frac{v'_y \sqrt{1 - \frac{v^2}{c^2}}}{1 + \frac{v'_x V}{c^2}}$$

Putting, $v'_x = c\cos\theta$ and $v'_y = c\sin\theta$, we get:

$$v^2_x + v^2_y = \left(\frac{c\cos\theta + V}{1 + \frac{V\cos\theta}{c}}\right)^2 + \frac{(c^2\sin^2\theta)\left(1 - \frac{V^2}{c^2}\right)}{\left(1 + \frac{V\cos\theta}{c}\right)^2}$$

$$= \frac{1}{\left(1 + \frac{V\cos\theta}{c}\right)^2}[c^2\cos^2\theta + 2cV\cos\theta + V^2 + c^2\sin^2\theta - V^2\sin^2\theta]$$

$$= \frac{1}{\left(1 + \frac{V\cos\theta}{c}\right)^2}[c^2 + 2cV\cos\theta + V^2(1 - \sin^2\theta)]$$

$$= \frac{1}{\left(1 + \frac{V\cos\theta}{c}\right)^2}[c^2 + 2cV\cos\theta + V^2\cos^2\theta]$$

$$= \frac{c^2}{\left(1 + \frac{V\cos\theta}{c}\right)^2}\left(1 + \frac{2V\cos\theta}{c} + \frac{V^2\cos^2\theta}{c^2}\right)$$

$$= \frac{c^2}{\left(1 + \frac{V\cos\theta}{c}\right)^2}\left(1 + \frac{V\cos\theta}{c}\right)^2 = c^2$$

or, $v^2_x + v^2_y = c^2$.

Example 10.21

The velocity of a particle is $3\mathbf{i} + 4\mathbf{j} + 12\mathbf{k}$ (m/s) in a frame of reference S', moving with velocity $0.8c$ along the x-axis, relative to a reference frame S at rest. What is the velocity of the particle in the frame S?

Solution:

The velocity $\mathbf{u}'$ of the particle in the moving frame is given by $\mathbf{u}' = 3\mathbf{i} + 4\mathbf{j} + 12\mathbf{k}$, so that $u'_x = 3$ m/s, $u'_y = 4$ m/s and $u'_z = 12$ m/s. The velocity of the frame S' relative to S is $0.8c$, i.e., $v = 0.8c$ using Lorentz transformation, we have:

$$u_x = \frac{u'_x + v}{1 + \frac{u'_x v}{c^2}} = \frac{3 + 0.8c}{1 + \frac{3 \times 0.8c}{c^2}} = \frac{3 + 0.8 \times 3 \times 10^8}{1 + \frac{3 \times 0.8}{3 \times 10^8}}$$

$$= \frac{3 + 2.4 \times 10^8}{10^8 + 0.8} \times 10^8 = 2.4 \times 10^8 \text{ m / s}$$

$$u_y =$$

$$\frac{u'_y\sqrt{1-\frac{v^2}{c^2}}}{1+\frac{u'_x v}{c^2}} = \frac{4\sqrt{1-\left(\frac{0.8c}{c}\right)^2}}{1+\frac{0.8c\times 3}{c^2}} = \frac{4\sqrt{1-0.64}}{1+\frac{3\times 0.8}{3\times 10^8}} = 2.4$$

$u'_z =$

$$\frac{u'_z\sqrt{\left(1-\frac{v^2}{c^2}\right)}}{1+\frac{vu'_x}{c^2}} = \frac{12\sqrt{1-\left(\frac{0.8c}{c}\right)^2}}{1+\frac{0.8c\times 3}{c^2}} = 7.2$$

Thus, the velocity of the particle in frame S is $\mathbf{u} = 2.4 \times 10^8\mathbf{i} + 2.4\mathbf{j} + 7.2\mathbf{k}$.

Example 10.22

Two velocities of 0.8c are inclined to each other at an angle of 30°. Find their resultant value.

Solution:

Imagine one of the velocities of 0.8c to be along the x-axis and the other inclined to it at an angle of 30°. This is equivalent to a reference frame S' moving along x'-axis with a velocity of 0.8c relative and parallel to a stationary frame S and a particle is moving with a velocity 0.8c inclined to it, at an angle of 30°. Then the resultant velocity of the particle, as appearing to an observer in frame S, can be determined as

Velocity of frame S', relative to S, along the x'-axis, v = 0.8c.

Velocity of the particle at an angle 30° with x'-axis (or the velocity v) = u'.

Resolving u' along x'-axis and y-axis,

component along x'-axis, u'_x:

=

$$u'\cos 30° = 0.8c\cos 30° = 0.8c\times\frac{\sqrt{3}}{2} = 0.4\sqrt{3}c$$

component along y'-axis:

$u'_y =$

$$u'\sin 30° = 0.8c\sin 30° = 0.8c\times\frac{1}{2} = 0.4c$$

Using Lorentz transformation, the values of these components, as observed in frame S are given by:

$u_x =$

$$\frac{u'_x+v}{1+\frac{u'_x v}{c^2}} = \frac{0.4\sqrt{3}c+0.8c}{1+\frac{0.4\sqrt{3}c(0.8c)}{c^2}} = \frac{c(0.4\sqrt{3}+0.8)}{1+0.4\sqrt{3}(0.8)}$$

$$= \frac{c(0.4\sqrt{3}c+08)}{1+0.32\sqrt{3}} = \frac{1.493c}{1.554} = 0.96c$$

$u_y =$

$$\frac{u'_y\sqrt{1-\frac{v^2}{c^2}}}{1+\frac{u'_x v}{c^2}} = \frac{0.4c\sqrt{1-\left(\frac{0.8c}{c}\right)^2}}{1+\frac{0.4\sqrt{3}c(0.8c)}{c^2}} = \frac{0.4c\sqrt{1-0.64}}{1+0.32\sqrt{3}}$$

$$= \frac{0.4c(0.6)}{1.554} = 0.15c$$

Resultant velocity u, as observed in S,

$$u = \sqrt{u_x^2+u_y^2}$$

$$= \sqrt{(0.96)^2+(0.15)^2} = \sqrt{0.9442}c = 0.97c$$

The angle φ this resultant velocity makes with the x-axis,

$$\varphi = \tan^{-1}\left(\frac{u_y}{u_x}\right) = \tan^{-1}\frac{0.15}{0.96} = \tan^{-1}0.1563 = 8°53'.$$

Example 10.23

At what velocity will the mass of a particle get doubled compared to its rest mass?

Solution:

We know,

$$\mu = \frac{m_0}{\sqrt{1-\frac{v^2}{c^2}}}$$

Here, $\mu = 2\mu_0$, v =?

$$\therefore \frac{m}{m_0} = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$$

or, $$1-\frac{v^2}{c^2} = \left(\frac{m_0}{m}\right)^2 = \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

or, $\frac{v^2}{c^2} = 1-\frac{1}{4}=\frac{3}{4}$

or, $v = \frac{\sqrt{3}}{2}c = 0.866\times 3\times 10^8\ \mathrm{m/s} = 2.598\times 10^8\ \mathrm{m/s}$.

Example 10.24

A man weighs 60 kg on the earth. When he is in a rocket ship in flight, his mass is 61 kg as measured by an observer on the earth. What is the speed of the rocket?

Solution:

Here, μ_0 = 60 kg, μ = 61 kg, v =?

Putting the values in $m = \frac{m_0}{\sqrt{1-\frac{v^2}{c^2}}}$, we get

$$\frac{61}{60} = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$$

or, $\frac{v^2}{c^2} = 1-\left(\frac{60}{61}\right)^2 = 1-\frac{3600}{3721}=\frac{121}{3600}$

or, $v = \frac{11}{60}\cdot c = \frac{11}{60}\times 3\times 10^8\ \mathrm{m/s} = 5.5\times 10^7\ \mathrm{m/s}$.

Example 10.25

(a) Calculate the rest energy of an electron (μ = 9.109 × 10^{-31} kg and $q = -\varepsilon = -1.602\times 10^{-19}$ C) in joules and electron volts.

(b) An electron is acclerated through (i) a potential of 20 kV in a TV picture tube and (ii) in high-voltage X-ray machine through a potential of 5 MV. Calculate the speed of electron in each case.

Solution

(a) We know, rest energy = mc^2

Here, $\mu = 9.109\times 10^{-31}$ kg, $c = 3\times 10^8$ m/s

\ $mc^2 = 9.109\times 10^{-31}\ \mathrm{kg}\times (3\times 10^8\ \mathrm{m/s})^2$

$= 8.198\times 10^{-14}$ J

$= 8.198\times 10^{-14}\ \mathrm{J}\times \frac{1\,\mathrm{eV}}{1.602\times 10^{-19}\ \mathrm{J}}$

$= 5.11\times 10^5$ eV = 0.511 MeV

(b) We know, $l = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$

Solving for v we get

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

The total energy ε of the accelerated electron is the sum of its rest energy mc^2 and the kinetic energy eV gained from the work done on it by the electric field, i.e.,

$E = l\,mc^2 = mc^2 + \mathrm{eV}$

Dividing throughtout by mc^2, we get

$l = 1 + \frac{\mathrm{eV}}{mc^2}$

(i) An electron accelerated through potential of 20 kV gains an amount of energy 20 keV. Thus, for this electron

$l = 1+\frac{20.0\ \mathrm{keV}}{0.511\ \mathrm{MeV}} = 1+\frac{20.0\times 10^3\ \mathrm{eV}}{0.511\times 10^6\ \mathrm{eV}}$

= 1.039

\ Speed of electron $v = c\sqrt{1-\left(\frac{1}{1.039}\right)^2} = 0.272\,c$

$= 0.272\times 3\times 10^8$ m/s

$= 8.16\times 10^7$ m/s

(ii) Here, V = 5 MV = 5×10^6 volt

\ $\frac{\mathrm{eV}}{mc^2} = \frac{5\times 10^6\ \mathrm{eV}}{0.511\times 10^6\ \mathrm{eV}} = 9.784$

and g = 1 + 9.784 = 10.784

\ $v = c\sqrt{1-\left(\frac{1}{10.784}\right)^2} = 0.996$ c

$= 0.996\times 3\times 10^8$ m/s

= 2.988 × 10⁸ m/c

Example 10.26

Two protons (each of mass μ = 1.67 × 10⁻²⁷ kg) are initially moving with equal speeds in opposite directions. They collide and continue to exist and produces a neutral pion (μ = 2.40 × 10⁻²⁸ kg). If the protons and the pions are at rest after the collision, find the initial speed of the protons assuming the energy is conserved in the process.

Solution

Since energy is conserved, total energy of two protons before the collision is equal to the rest energy of the two protons and the pion after the collision. Thus,

$$2(lMc^2) = 2(Mc^2) + mc^2$$

Dividing throughout by $2Mc^2$, we get

$$l = 1+\frac{m}{2M} = 1+\frac{2.40\times10^{-28}\text{ kg}}{2\times1.67\times10^{-27}\text{ kg}} = 1.072$$

\ Speed before collision $v = c\sqrt{1-\left(\frac{1}{\lambda}\right)^2}$

$$= c\sqrt{1-\left(\frac{1}{1.072}\right)^2}$$

= 0.36 c

= 0.36 × 3 × 10⁸ m/s = 1.08 × 10⁸ m/s

Example 10.27

A body at rest suddenly explodes into two fragments each of mass one kilogram that moves apart at speeds of 0.6 c relative to the original body. What was the mass of the body at rest?

Solution

The rest energy of the original body must be equal to the total energies of the two fragments. That is,

$$E_o = mc^2 = g\mu_1 c^2 + g\mu_2 c^2$$

$$= \frac{m_1c^2}{\sqrt{1-\frac{v^2}{c^2}}} + \frac{m_2c^2}{\sqrt{1-\frac{v^2}{c^2}}} = \frac{2mc^2}{\sqrt{1-\frac{v^2}{c^2}}}$$

$$\backslash\ \mu = \frac{E_o}{c^2} = \frac{2\times1}{\sqrt{1-(0.6)^2}} = 2.5\text{ kg}$$

Example 10.28

If the sun radiates energy at the rate of 4 × 10²⁶ Js⁻¹, find the rate at which its mass is decreasing.

Solution:

Here, ε = 4 × 10²⁶ Js⁻¹, c = 3 × 10⁸ m/s

\ Rate of decrease of mass Dμ = $\frac{E}{c^2} = \frac{4\times10^{26}\text{ Js}^{-1}}{9\times10^{16}\text{ m}^2/\text{s}^2}$

or Dμ = 4.44 × 10⁹ kg

Example 10.29

The binding energy of an electron to a proton (i.e., hydrogen atom) is 13.6 eV. Find the loss of mass in the formation of one atom of hydrogen.

Solution:

Here, ε = 13.6 eV = 13.6 × 1.6 × 10⁻¹⁹ J

Using Einstein's mass energy relation ε = Δμ·c², the loss of mass in the formation of 1 hydrogen atom is:

$$\Delta\mu = \frac{E}{c^2}$$

$$\text{or, } \Delta\mu = \frac{13.6\times1.6\times10^{-19}}{3\times10^8\times3\times10^8} = 2.42\times10^{-35}\text{ kg.}$$

Example 10.30

What is the speed of an electron having kinetic energy 1.02 MeV?

Solution:

Kinetic energy associated with a particle of mass μ is mc^2. If Δμ is the increase in the rest mass μ_o of the electron on attaining velocity v, then:

$$\Delta\mu = \frac{\text{K.E. acquired by the electron}}{c^2}$$

Here, K.E. = 1.02 MeV = 1.02 × 1.6 × 10⁻¹³ J

$$\therefore \Delta\mu = \frac{1.02\times1.6\times10^{-13}}{3\times10^8\times3\times10^8} = 1.813\times10^{-30}\text{ kg}$$

= 18.13 × 10⁻³¹ kg

Mass of the electron on acquiring the final velocity is

$$\mu = \mu_0 + \Delta\mu$$
$$= (9.11 + 18.13) \times 10^{-31} \text{ kg}$$
$$= 27.24 \times 10^{-31} \text{ kg}$$

Using the relation:

$$\mu = \frac{m_0}{\sqrt{1-\frac{v^2}{c^2}}}$$

or, $\frac{v^2}{c^2}$ =

$$1-\left(\frac{m_0}{m}\right)^2 = 1-\left(\frac{9.11\times 10^{-31}}{27.24\times 10^{-31}}\right) = 0.882$$

or, v =

$$\sqrt{0.882}c = 0.942\times 3\times 10^8 = 2.38\times 10^8 \text{ m/s}$$

Thus, the speed of the electron is 2.83×10^8 m/s.

Example 10.31

The velocity of a proton is such that $\beta\left(=\frac{v}{c}\right)$ for it is 0.995 in the laboratory. Find the corresponding relativistic energy and momentum.

Solution:

Rest mass of a proton

$$\mu_0 = 1.67 \times 10^{-27} \text{ kg}$$

$$\beta = \frac{v}{c} = 0.995 \Rightarrow v = 0.995c$$

Its mass at velocity v,

$$\mu = \frac{m_0}{\sqrt{1-\frac{v^2}{c^2}}} = \frac{1.67\times 10^{-27}}{\sqrt{1-(0.995)^2}} \text{ kg}$$

∴ Relativistic energy

$$= mc^2 = \frac{1.67\times 10^{-27}}{\sqrt{1-(0.995)^2}}\times(3\times 10^8)^2$$

$$= \frac{1.67\times 9\times 10^{-11}}{\sqrt{0.01}} = \frac{15.03\times 10^{-11}}{0.1}$$

$$= 1.5 \times 10^{-9} \text{ J}$$

We know, $\varepsilon^2 - p^2c^2 = \mu_0^2c^4$

∴ Relativistic momentum

$$p = \frac{\sqrt{E^2 - m_0^2c^4}}{c}$$

$$p = \frac{\sqrt{(1.5\times 10^{-9})^2 - (1.67\times 10^{-27})^2(3\times 10^8)^4}}{3\times 10^8}$$

$$= \frac{\sqrt{2.25\times 10^{-18} - 0.0225\times 10^{-18}}}{3\times 10^8}$$

$$= \frac{\sqrt{2.2275\times(10^{-18})}}{3\times 10^8} = \frac{1.49}{3}\times 10^{-17} \text{ kg m/sec}$$

$$= 4.97 \times 10^{-18} \text{ kg m/sec.}$$

Example 10.32

(*a*) Estimate the momentum of a photon having energy 1.00×10^{-17} Joule.

(*b*) Calculate the kinetic energy of an electron for which $\beta = 0.99$.

Solution:

(*a*) Rest mass of a photon is zero. Therefore, its momentum is:

$$p = \frac{E}{c} = \frac{1.00\times 10^{-17}}{3\times 10^8} = 3.33\times 10^{-26} \text{ kg m/s}$$

(*b*) Kinetic energy = total energy – rest energy

$$= mc^2 - m_0c^2 = \gamma m_0c^2 - m_0c^2 = m_0c^2(\gamma - 1)\left(\gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}\right)$$

For electron $\mu_0 = 9.11 \times 10^{-31}$ kg

Here, $\beta = \frac{v}{c} = 0.99$

∴ g =

$$\frac{1}{\sqrt{1-(0.99)^2}} = \frac{1}{\sqrt{1-0.98}} = \frac{1}{\sqrt{0.02}} = \frac{1}{0.14} = 7.07$$

\ Kinetic energy

$\varepsilon = 9.11 \times 10^{-31} \times (3 \times 10^8)2\ (7.07 - 1)$

$= 9.11 \times 9 \times 10^{-15} \times 6.07 = 4.98 \times 10^{-13}$ J

$= \dfrac{4.98\times10^{-13}}{1.6\times10^{-19}}$ eV $= 3.11 \times 10^6$ eV $= 3.11$ MeV.

Example 10.33

A gamma-ray photon of frequency υ_0 is emitted by a nucleus of mass μ. Show that the loss of internal energy suffered by the nucleus is $\left(1+\dfrac{h\nu_0}{2mc^2}\right)$ and not $h\upsilon_0$.

Solution:

The energy of the photon, $\varepsilon = h\upsilon_0$

∴ Momentum imparted to the photon:

$$p = \frac{E}{c} = \frac{h\nu_0}{c}$$

The nucleus also loses an equal amount of momentum and recoils with a velocity υ (say), thereby losing an additional amount of energy $= \dfrac{1}{2}m\nu^2$.

∴ Total energy lost by the nucleus is:

$$= h\nu_0 + \frac{1}{2}m\nu^2$$

$$= h\nu_0 + \frac{(m\nu)^2}{2m}$$

Since $m\upsilon$ is the loss of momentum of the nucleus, equal in magnitude to the gain in momentum p of the photon.

∴ Total energy lost is:

$$= h\nu_0 + \frac{p^2}{2m} = h\nu_0 + \frac{(h\nu_0)^2}{2mc^2} = h\nu_0\left(1+\frac{h\nu_0}{2mc^2}\right)$$

Example 10.34

Show that the momentum of a particle of rest mass μ_0 and kinetic energy K can be expressed as

$$p = \sqrt{\frac{K^2}{c^2} + 2m_0K}.$$

Solution:

We know,

$$\varepsilon^2 - p^2c^2 = \mu_0^2c^4$$

$$\therefore \varepsilon^2 = m_0^2c^4 + p^2c^2 \Rightarrow E = \sqrt{m_0^2c^4 + p^2c^2}$$

Also, the total energy is

$$\varepsilon = \text{rest energy} + \text{kinetic energy}$$

$$= \mu_0c^2 + K$$

Equating the above two expressions for ε, we get:

$$\sqrt{m_0^2c^4 + p^2c^2} = \mu_0c^2 + K$$

Squaring both sides, we get:

$$\mu_0^2c^4 + p^2c^2 = \mu_0^2c^4 + K^2 + 2\mu_0c^2K$$

$$\text{or, } p^2c^2 = K^2 + 2\mu_0c^2K$$

$$\text{or, } p^2 = \frac{K^2}{c^2} + 2m_0K$$

$$\text{or, } p = \sqrt{\frac{K^2}{c^2} + 2m_0K}.$$

Exercises

Short Answer Type

1. Define the terms: (*i*) Frame of Reference and (*ii*) Variant and Invariant Quantity.
2. Differentiate between Inertial and Non-inertial frame of references.
3. Is earth an inertial frame? Why?
4. Obtain Galilean Transformations for space and time.
5. Why was it assumed that the space is filled with luminiferous ether?
6. With what objective in mind, the Michelson-Morley experiment was designed and performed? State its conclusions.
7. State the postulates of the special Theory of Relativity.
8. Write Lorentz Transformation equations and inverse Lorentz Transformation equations.
9. Show that for $\upsilon << c$, Lorentz Transformation reduces to Galilean Transformation.
10. What is Lorentz-Fitzgerald contraction? Define proper length.

11. What do you mean by Time Dilation? Define proper time.

12. What is twin paradox?

13. Is the mass of a body constant? Explain.

14. What do you mean by mass-energy equivalence relation?

15. What is the rest mass of a photon?

Long Answer Type

1. What do you mean by an inertial frame of reference? Show that acceleration and force are invariant under Galilean Transformation.

2. Show that a frame of reference having a uniform translatory motion relative to an inertial frame is also inertial.

3. Use Galilean Transformations to show that the distance between the two points is invariant in the inertial frames.

4. Discuss Michelson-Morley experiment and explain the significance of the result obtained.

5. State and explain the postulates of the special theory of relativity and derive Lorentz transformations and inverse Lorentz transformation.

6. What is 'proper length'? Derive Lorentz-Fitzgerald contraction relation.

7. Deduce an expression for time dilation using Lorentz transformations. Give an example showing experimental verification of time dilation.

8. Show using addition of velocities that the velocity of light is an absolute constant, independent of the frame of reference and it is the maximum velocity attainable in nature.

9. Show that two events simultaneous in one frame of reference, will not be simultaneous in another frame, moving with uniform velocity relative to it.

10. Define 'rest mass' of a body and show $m = \frac{m_0}{\sqrt{1-\frac{v^2}{c^2}}}$, where the symbols have their usual meanings. Represent graphically the variation of mass with velocity.

11. Obtain the expression for relativistic momentum and force.

12. Obtain the relation $\varepsilon = mc^2$ and explain its significance.

13. Prove the relation $\varepsilon^2 - p^2c^2 = \mu^2c^4$, where p is the relativistic momentum.

Numericals

1. There are two systems S and S' such that $S\times$ has velocity $(3\mathbf{i} + 4\mathbf{j} + 6\mathbf{k})$ m/s relative to S. If the origins of the two systems coincide initially and coordinates at any point after 2 s are (4, 5, 6) metre relative to S, find the coordinates of the point relative to S'.

[**Ans.** (1, – 3, – 6)]

2. A particle has velocity $(5\mathbf{i} + 6\mathbf{j} - 4\mathbf{k})$ m/s relative to a cart moving with velocity $(100\mathbf{i} + 150\mathbf{j})$ m/s relative to an observer on the ground. Calculate the velocity of the particle relative to the observer on the ground.

[**Ans.** $(150\mathbf{i} + 156\mathbf{j} - 4\mathbf{k})$ m/s]

3. Calculate the expected fringe shift in the Michelson-Morley experiment if the effective length of each path be 6 metres, velocity of the earth 3×10^4 m/s and the wavelength of the monochromatic light used is 5000 Å.

[**Ans.** 0.24 fringe width]

4. The hypothetical speed of the earth through the ether, is its orbital speed of 3×10^4 m/s. If the light takes 3×10^{-7} sec to travel through the Michelson-Morley apparatus in the direction parallel to this motion, how long will it take to travel through it in the direction perpendicular to this motion?

[**Ans.** 3×10^{-7} sec]

5. The interval S_{12} between the two events is defined by the relation $S^2_{12} = x^2 + y^2 + z^2 + w^2$ where $w = ict$, $i = \sqrt{-1}$. Using Lorentz Transformation, show that S_{12} has the same value in all inertial frames, though distance and time may have different values.

6. Two events have the space-time coordinates (0, 0, 0, 0) and (4c, 0, 0, 3) in a given frame S. (*i*) What is the time-interval between them? (*ii*) Obtain the velocity of a frame in which (*a*) the two events occur simultaneously, and (*b*) the first event occurs two seconds earlier than the second.

[**Ans** (*i*) 2.645c, (*ii*) (a) 0.75c, (*b*) 0.3].

7. A rocket ship is 100 m long on the ground. When it is in flight, its length is 99 m to an observer on the ground. What is it speed?

[**Ans.** 4.23×10^7 m/s]

8. A rod has length 100 cm. When the rod is in a satellite moving with a velocity half the velocity of light, relative to laboratory, what is the length of the rod as determined by the observer (*a*) in the satellite, (*b*) in the laboratory.

[**Ans.** (*a*) 100 cm, (*b*) 86.6 cm]

9. An unstable particle has the mean proper lifetime of 2 ms. What will be its lifetime, when it is travelling with a speed of 0.9c.

[**Ans.** 4.58 ms]

10. Calculate the percentage contraction in the length of a rod moving with a speed of 0.8c in a direction at angle 60° with its own length.

[**Ans.** 13%]

11. A μ-meson particle has a lifetime 2×10^{-6} sec. (*a*) What is the mean lifetime, when the particle is travelling with a speed of 2.994×10^{10} cm/sec. (*b*) How far does it go during one mean life? (*c*) What distance it would have covered without relativistic effect?

12. A beam of particles of half-life 2.8×10^{-8} sec travels in the laboratory with speed 0.96 times the speed of light. How much distance does the beam travel before the flux falls to half the initial flux.

[**Ans.** 2.88×10^3 m]

13. The mass of a moving electron is 11 times its rest mass. Find its kinetic energy and momentum.

[**Ans.** 8.19×10^{-13} J, 2.978×10^{-21} kg m/s]

14. Compute the speed of the electron at which its mass becomes 4 times its mass at rest.

[**Ans.** 2.94×10^{8} m/s]

15. In the laboratory, the two particles are observed to travel in opposite directions with speed 2.8×10^{8} m/s. Find the relative speed of particles.

[**Ans.** 2.994×10^{8} m/s]

16. Calculate the mass and speed of 3 MeV electron. The rest mass of an electron is 9.1×10^{-31} kg.

[**Ans.** 5.33×10^{-30} kg. 0.985c]

17. If the kinetic energy of a particle is twice its rest mass energy, find its velocity.

[**Ans.** 0.943c]

18. Find the velocity that an electron must be given so that its momentum is 10 times its rest mass times the speed of light. What is the energy at this speed?

[**Ans.** 2.983×10^{8} m/s, 8.22×10^{-13} J]

19. What is the total energy of a 2.5 MeV electron?

[**Ans.** 3.011 MeV]

20. The velocity of a particle in a frame S', moving relative to a frame S with a velocity of 0.8c along the x-axis is represented by $7\mathbf{i} + 6\mathbf{j} + 5\mathbf{k}$ m/s. What will be the velocity of the particle in frame S'?

[**Ans.** $(2.4 \times 10^{8})\mathbf{i} + 3.6\mathbf{j} + 3\mathbf{k}$]

21. An aeroplane sets out to fly at 500 miles per hour. Show that it would have to fly for more than a thousand years in order to make a difference of 2 seconds between the times recorded by a clock in the aeroplane and a clock on the ground.

22. The length of the side of a square, as measured by an observer in a stationary frame of reference S, is l. What will be its apparent area, as observed by him in a reference frame S' moving relative to S with velocity v along one of the sides of the square?

23. How fast should a rocket ship move relative to an observer in order that one year on it corresponds to two years on the earth?

[**Ans.** 2.598×10^{8} m/s]

24. At what speed will a proton's mass exceed its rest mass ($1.67 \times 10{-}27$ kg) by 1%? What kinetic energy (in eV) does this speed correspond to?

[**Ans.** 4.2×10^{7} m/s, 9.3×10^{6} eV]

25. Calculate the wavelength of the radiation emitted by the annihilation of an electron with a positron, each of rest mass $9.1 \times 10{-}31$ kg. Is it visible?

[**Ans.** 0.024×10^{-10} m]

26. (a) What is the recoil momentum in the laboratory of a 57Fe nucleus recoiling due to the emission of 14 keV photon?

(b) What is the mass equivalent of the energy from an antenna radiating 1 kW of radio energy for 24 hours?

[Ans. (a) $7.5 \times 10{-}19$ gm-cm/s, (b) $9.6 \times 10{-}10$ kg]

Index