According to the quantum theory, light is emitted in the form of photons which are bundles of electromagnetic radiation that oscillate with a definite frequency and propagate through space with the speed of light. Individual photons behave like particles but when their number is large they exhibit the properties of a continuous wave.

6.3. WAVES

As light is considered to have wave-character, let us first acquaint ourselves with the wave motion. A wave is any disturbance, from an equilibrium condition, that propagates with time from one region of space to another.

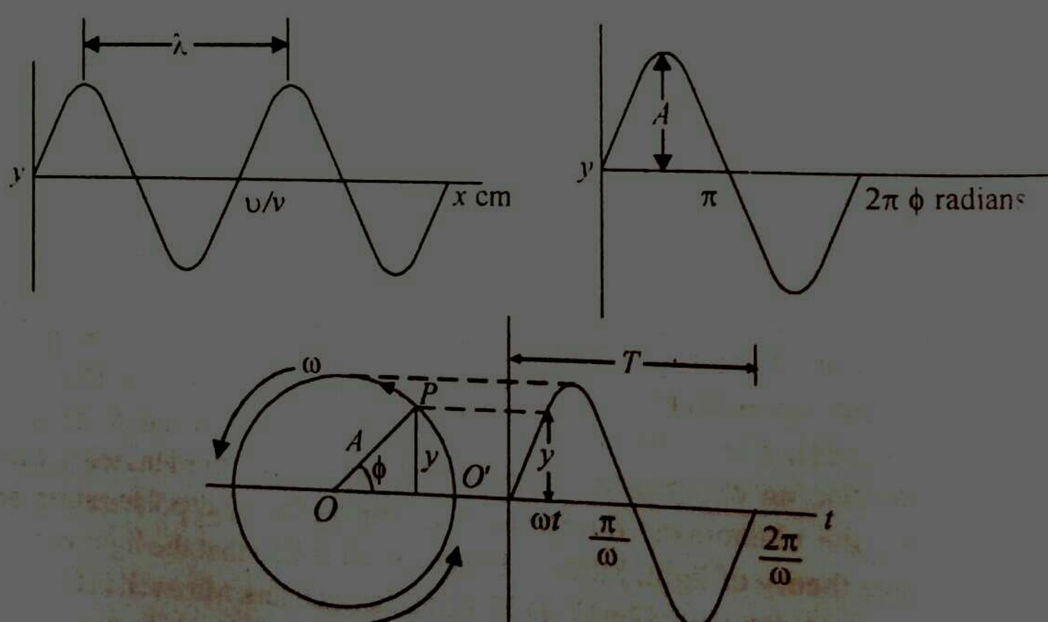
We are all familiar with what happens when a pebble is dropped into the still water of a well or a pond. The pebble generates ripples which expand in the form of circles. If the motion of a floating object such as a leaf is watched, we observe that the leaf moves up and down but does not travel outward with the wave. We, therefore, conclude that it is the energy supplied by the agent of disturbance which moves forward as successive waves. We identify the wave motion with the help of crests and troughs travelling away from the centre of disturbance. It is apparent from the motion of the floating leaf that every particle of the medium (water) oscillates about its mean position at right angles to the direction of wave propagation. Such wave motion is known as **transverse** wave motion. Water waves and electromagnetic waves are examples of transverse waves.

Waves can be of any form. Harmonic waves are simple and ideal. They are described mathematically by a sine or cosine function. A wave is described by the displacement of a particle, y. The value of y is a function of space coordinates and a function of time t.

Thus

$$y = f(x, y, z, t) \quad \ldots(6.1)$$

If this function is known, it can be used to predict the position of any particle at any time, and any other kinematic information. Such a function is called a **wave function** and contains a complete description of the motion. A sinusoidal wave propagating in one direction, say x-direction may be represented by the wave function



$$y(x, t) = A \sin\left[\frac{2\pi}{\lambda}(x - \upsilon t)\right] \qquad ...(6.2)$$

The wave form represented by equation (6.2) is shown in Fig. 6.1(a)

The meaning of equation (6.2) may be understood as follows :

At any given instant of time t, it gives the displacement y of a particle from the equilibrium position as a function of the coordinate x of the particle.

At a time $t = 0$

$$y(x, o) = A \sin\left(2\pi\frac{x}{\lambda}\right) \qquad ...(6.3)$$

Similarly, at any given coordinate, x, equation (6.2) gives the displacement y of the particle at that coordinate, as a function of time t. Thus at $x = 0$.

$$y(o, t) = -A \sin\frac{2\pi}{\lambda}\upsilon t \qquad ...(6.4)$$

The constant A in the above equations represents the maximum value of the displacement and is called the **amplitude** of the wave. The crest-to-crest distance or the distance between any two adjacent identical points is called the **wavelength** of the wave and is denoted by λ. It is the distance over which the waveform repeats itself. The time taken by the waveform to advance through a distance λ is called the **period**, T. The wave travels at a speed υ and covers a distance υT in time T.

$$\therefore \qquad \lambda = \upsilon T \qquad ...(6.5)$$

The reciprocal of the period is called the **frequency**. It may be defined as the number of crests that pass an observer per second.

$$\nu = \frac{1}{T} \qquad ...(6.6)$$

From (6.5) and (6.6), we get

$$\upsilon = \nu\lambda \qquad ...(6.7)$$

Equation (6.2) may be rewritten as

$$y(x, t) = A \sin\left[2\pi\left(\frac{x}{\lambda} - \frac{t}{T}\right)\right] \qquad ...(6.8)$$

Let

$$\omega = 2\pi/T \qquad ...(6.9)$$

and

$$k = 2\pi/\lambda \qquad ...(6.10)$$

Equation (6.8) is expressed in a more convenient form by substituting equations (6.9) and (6.10) into it. Thus

$$y(x, t) = A \sin(kx - \omega t) \qquad ...(6.11)$$

ω is called the **angular frequency** and represents the number of cycles contained in the time interval equal to 2π seconds. k is called the **wave number** and represents the number of wavelengths contained in a distance equal to 2π metres.

From equations (6.9), (6.10) and (6.5), we get

$$\upsilon = \frac{\omega}{k} \qquad ...(6.12)$$

The relations (6.5), (6.7) and (6.12) are the kinematic relations governing the motion of waves.

The argument of the sine function in equation (6.11) is

$$kx - \omega t = \phi \qquad ...(6.13)$$

... or briefly the phase. It can be expressed in radians or degrees. The maximum possible value for ϕ is 2π rad or 360°, as shown in Fig. 6.2 (*b*). Alternately, it may be expressed as either the ratio t/T or x/λ. Both are fractional numbers and the maximum value the ϕ can take in these two cases is 1. After 360°, or one T, or one λ the wave cycle repeats. It is seen from Fig. 6.1(*c*) that at any given time the phase ϕ is different for each point along the waveform.

The amplitude, frequency and phase are the three parameters that characterise a wave.

The wave function (6.11) assumes that the displacement $y = 0$ at $x = 0$ and $t = 0$. Such a situation is a special case and in general y may not be zero. The wave function in the most general case is written as

$$y(x, t) = A \sin(kx - \omega t - \theta) \quad \text{... (6.14)}$$

θ is called the **phase constant** which can be determined from a knowledge of the initial conditions.

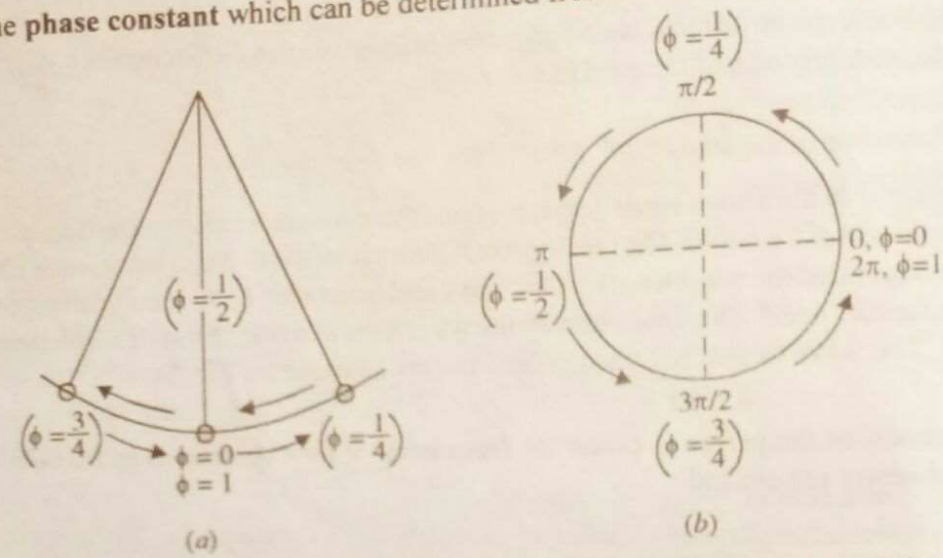


Fig. 6.2. Variation in phase. The phase is usually expressed as a proper fraction. The whole periods elapsed from the start are discarded because after each cycle, the motion is repeated in the same order. If the phase of the pendulum is said to be 3/4, it means that the pendulum is at left and is about to move toward the right.

The equation (6.11) can be made independent of the system of coordinates by converting it into a vector form. Let vector $\boldsymbol{k}$ have a magnitude equal to the wave number and a direction parallel to the positive direction of the x-axis. Such a vector is called the **wave vector**. Using k, Eq. (6.11) becomes

$$y(x, t) = A \sin(\boldsymbol{k}.x - \omega t)$$

In the most general case, where r is in any arbitary direction,

$$y(r, t) = A \sin(\boldsymbol{k}.\boldsymbol{r} - \omega t) \quad \text{...(6.15)}$$

The wave travelling on a rope is a one dimensional wave, whereas water waves are two dimensional waves. Sound waves, light waves, radio waves etc., are three dimensional waves that expand in the form of spheres.

6.4. WAVE FRONT AND THE RAY

Waves start from a source and spread out into new and new regions of space. In case of waves on a water surface, the ripples start from the point of disturbance and expand in the form of circles. The circles are in fact the crests of the waves. All the particles located at the crest will be in the same state of oscillation and hence, in the same phase. Therefore, a ripple is the locus of points having the same phase. The propagation of ripples is carried out by circles formed one after another at a distance of λ from each other, as shown in Fig. 6.3(*a*).

In a similar manner, the propagation of a three dimensional wave may be visualized as being carried out by innumerable wave surfaces formed one after another at a distance of λ from each other. A **wave surface** is a surface over which the phase of the wave is constant. Generally, the wave

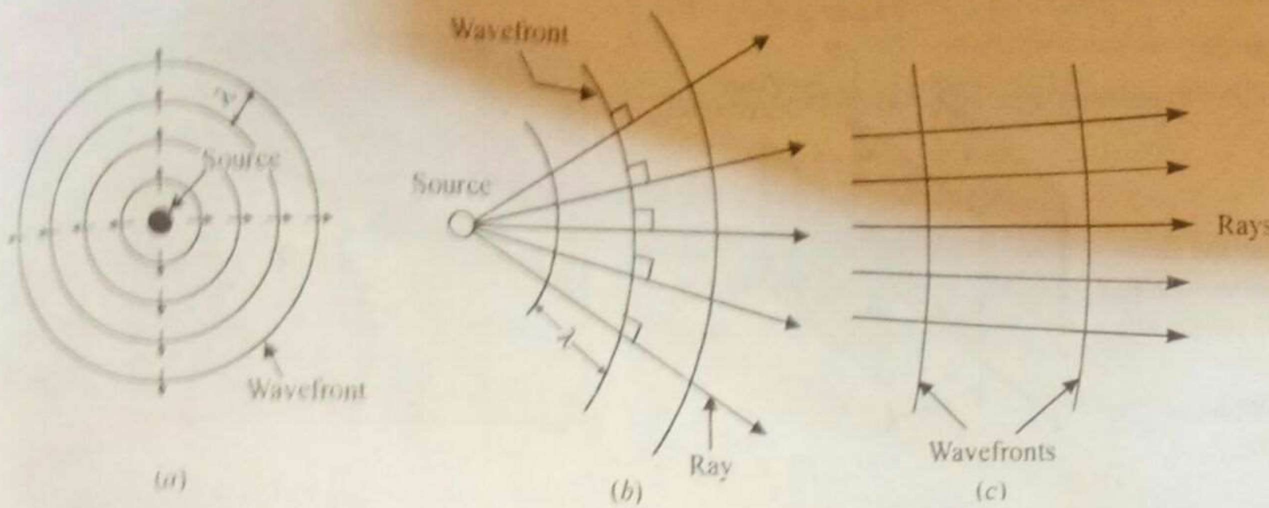


Fig. 6.3. Wavefront and rays—(*a*) a point source produces spherical waves. (*b*) the ray direction is prependicular to the wavefront. (*c*) at large distance from the source, the wavefronts tend to be parallel planes.

surfaces are drawn as passing through the crest or trough of a sine wave. Therefore, the wave surfaces are separated by one wavelength. The furthermost wave surface from the source is called a **wavefront**. The wavefront is dynamic and separates that region of space into which the wave is already spread from the region into which the wave is yet to enter. The propagation of the wave is visualised by the advancing wavefront and stationary wave surfaces behind it. Usually, all the wave surfaces are referred to as wavefronts.

The shape of the wavefront is determined by the source generating the waves and by the size of the object which the waves encounter. If the source is a point source, it radiates waves in all directions uniformly and the wavefronts will be a family of concentric spheres. On the other hand, a linear source produces a family of coaxial cylindrical wavefronts. At distances far from the source, both the spherical and cylindrical wavefronts may be treated as plane wavefronts. When the size of the slit, through which a plane wave passes is much larger than the wavelength, the wavefront remains plane. On the other hand, if the slit is of the order of wavelength, the wavefront becomes spherical.

It is convenient sometimes to describe the wave propagation in terms of rays instead of wavefronts. **Rays** are lines drawn normal to the wavefronts and represent the direction of energy flow. Rays are parallel lines in case of plane waves and radial lines in case of spherical waves. They are illustrated in Fig. 6.3.

6.5. MATHEMATICAL REPRESENTATION OF A PLANE WAVE

A plane wave propagating along the x-axis is shown in Fig. 6.4. The wavefronts are plane surfaces and they are parallel to the yz-plane. The displacement of any particle is a function of x and t only. Therefore, a plane wave moving along the x direction is represented by the wave function.

$$y(x, t) = A \sin (kx - \omega t) \qquad \text{...(6.16)}$$

The wave functions given by equations (6.11) and (6.16) are identical in form. One may observe from (6.16) that the amplitude A of the wave does not depend on x and is a constant. Such a wave is said to have **zero vergence** and is not attenuated as it propagates through the space. For this reason, we deal with only plane harmonic waves and plane wavefronts in our course.

6.8. LIGHT WAVE

A light wave is an electromagnetic wave consisting of periodically varying electric and magnetic fields oscillating at right angles to each other and to the direction of propagation of the wave. A portion of an electromagnetic wave is shown in Fig. 6.8.

The electric field in the wave is defined by the electric field strength vector ***E*** and the magnetic field by the vector of magnetic induction ***B***. Vectors ***E*** and ***B*** are of equal importance to the wave. However, a light wave is often represented by ***E*** wave since many of the effects of light are mainly due to its electric field. The magnetic field is implied to be oscillating in a plane normal to the plane of electric field oscillations and is not shown specifically in the diagrams.

Mathematically, the light wave depicted in Fig. 6.8 is represented by the expressions

$$E = E_0 \sin (kx - \omega t) \quad \text{...(6.23)}$$

and

$$B = B_0 \sin (kx - \omega t) \quad \text{...(6.24)}$$

We have to note at this stage that the Fig. 6.8 and the equations (6.23) and (6.24) represent an **ideal** plane electromagnetic wave.

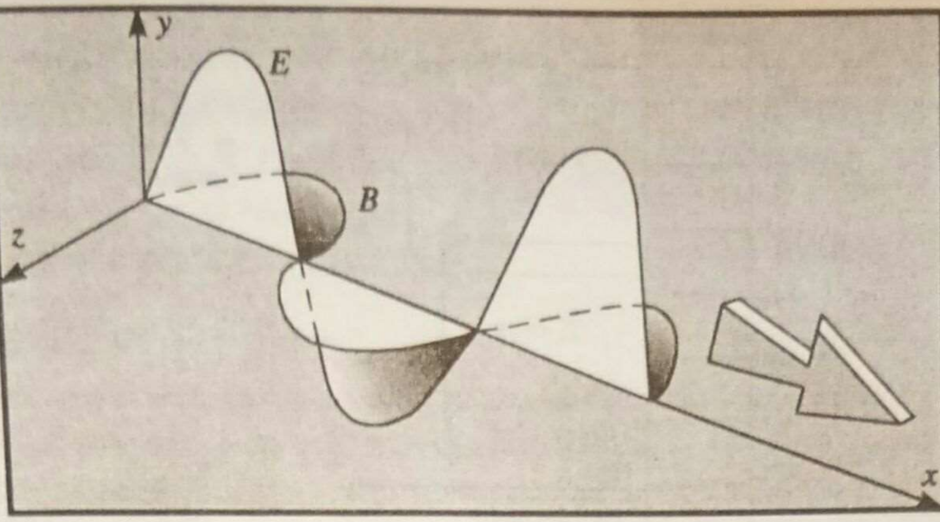


Fig.6.8. An electromagnetic wave.

(*i*) The wave described by Eqs. (6.23) and (6.24) is strictly monochromatic having a single definite frequency ν ($= \omega/2\pi$).

(*ii*) It is a **harmonic wave**.

(*iii*) It is a plane wave, the wave front being normal to x-axis.

(*iv*) The electric vector ***E*** of the wave always oscillates parallel to a fixed direction in space, *i.e.* y direction. In other words, the ***E*** vibrations are confined to xy plane. A wave having its vibrations confined to a single plane is called a **plane polarized** or a **linearly polarized** wave. The wave depicted in Fig. 6.8 is an ideally plane polarized wave.

(*v*) The wave has **zero vergence** and does not get attenuated as it propagates through space.

(*vi*) The variation of phase along the wave extension is predictable. Therefore, if two or more such waves travel along the same direction in space, their phase variations are synchronized and the phase difference stays constant. Thus the waves will be highly coherent.

We will soon learn in this chapter that actual light waves are far from ideal. It will be noted that ey are wave trains of limited extension, have a certain spread in frequency around a central value, tally unpolarized and incoherent. All the natural light sources such as the sun, a lamp or a flame mit only such wavetrains.

We will first proceed to understand the behaviour of light assuming it to be ideal and then study ow the results get modified when their real character is taken into account.

6.6. THE PRINCIPLE OF SUPERPOSITION

When two pebbles are dropped at different points in a pond, the expanding ripples pass through each other without mutual effect. Likewise, sound waves from different instruments in an orchestra propagate in space independent of each other and can be distinguished separately. There occur many such instances in which a number of waves meet and proceed unaffected. Since waves do not interact, each region of space where two or more waves meet, undergo the vibrations set up by each wave separately. The resultant displacement at a given point in space is determined by the principle of superposition.

The **principle of superposition** states that the net displacement of a given point in space at any time due to two or more waves is the algebraic sum of the displacements produced at that point by all waves.

Mathematically speaking, the principle of superposition states that if two or more waves are propagating through the space, the resultant wave function at any point is the algebraic sum of the wave functions of the individual waves. Thus, if $y_1(x, t)$ and $y_2(x, t)$ are the wave functions

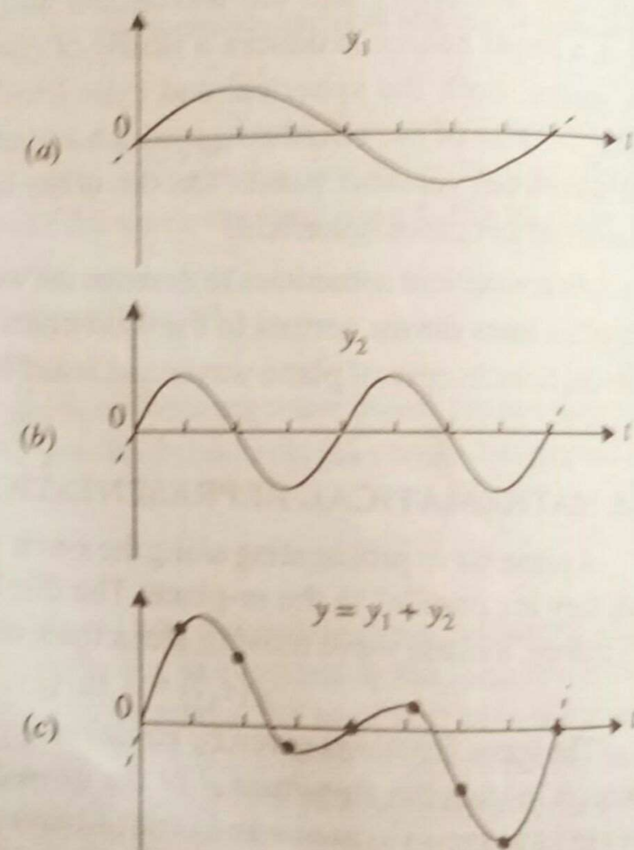


Fig. 6.5. Illustration of the principle of superposition.

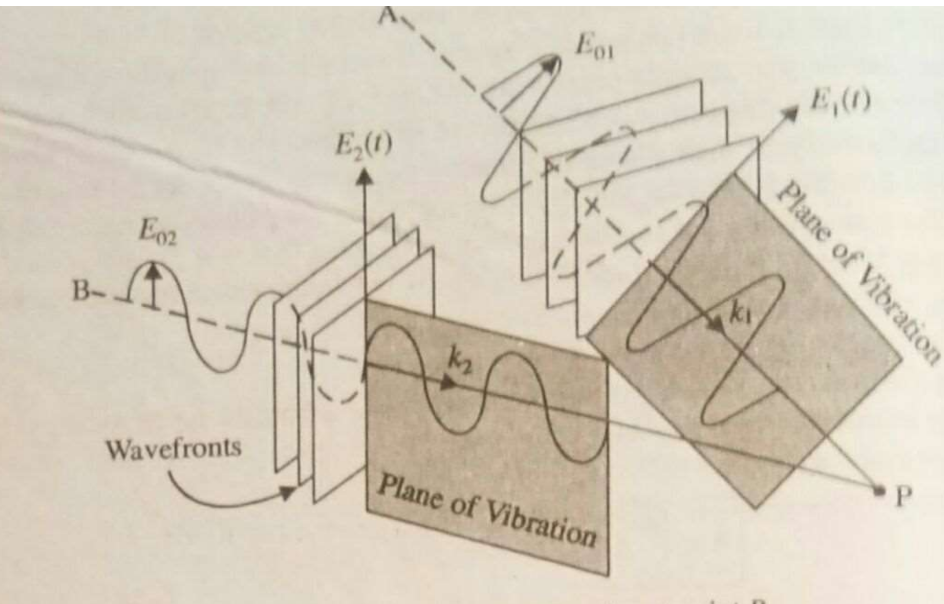


Fig. 6.18. Two light waves meeting at point P.

$$E_1 = E_{01} \sin(\omega t - kx_1 + \phi_1) \qquad ...(6.50)$$
$$E_2 = E_{02} \sin(\omega t - kx_2 + \phi_2) \qquad ...(6.51)$$

where x_1 and x_2 represent the paths travelled by waves 1 and 2. The principle of superposition applies here also. According to the superposition principle, the resultant electric field at a given place due to the simultaneous action of two or more harmonic waves is the algebraic sum of the electric fields of the separate constituent waves.

Thus if N waves meet at a place,

$$E = \sum_{i=1}^{N} E_i = E_1 + E_2 + E_3 + ... + E_N \qquad ...(6.52)$$

where $E_1, E_2, E_3, ...E_N$ are the electric fields of the individual waves. In the simple case of the superposition of two waves, we can write,

$$E = E_1 + E_2 \qquad ...(6.53)$$

$$\therefore \quad E = E_{01} \sin(\omega t - kx_1 + \phi_1) + E_{02} \sin(\omega t - kx_2 + \phi_2)$$
$$= E_{01} \sin(\omega t + \alpha_1) + E_{02} \sin(\omega t + \alpha_2) \qquad ...(6.54)$$

where $\quad \alpha_1 = (-kx_1 + \phi_1)$ and $\alpha_2 = (-kx_2 + \phi_2) \qquad ...(6.55)$

On expanding (6.54), we get

$$E = (E_{01} \sin \omega t \cos \alpha_1 + E_{01} \cos \omega t \sin \alpha_1) + (E_{02} \sin \omega t \cos \alpha_2 + E_{02} \cos \omega t \sin \alpha_2)$$
$$= (E_{01} \cos \alpha_1 + E_{02} \cos \alpha_2) \sin \omega t + (E_{01} \sin \alpha_1 + E_{02} \sin \alpha_2) \cos \omega t \qquad ...(6.56)$$

As the terms in parentheses are constant in time, we can set

$$E_{01} \cos \alpha_1 + E_{02} \cos \alpha_2 = E_0 \cos \alpha \qquad ...(6.57)$$

and

$$E_{01} \sin \alpha_1 + E_{02} \sin \alpha_2 = E_0 \sin \alpha \qquad ...(6.58)$$

where E_0 is the amplitude of the resultant wave and α the new initial phase angle. In order to solve for E_0 and α, we square the equations (6.57) and (6.58) and add them

$$E_0^2 \cos^2 \alpha = E_{01}^2 \cos^2 \alpha_1 + E_{02}^2 \cos^2 \alpha_2 + 2E_{01} E_{02} \cos \alpha_1 \cos \alpha_2$$
$$E_0^2 \sin^2 \alpha = E_{01}^2 \sin^2 \alpha_1 + E_{02}^2 \sin^2 \alpha_2 + 2E_{01} E_{02} \sin \alpha_1 \sin \alpha_2$$

or $$E_0^2 = E_{01}^2 + E_{02}^2 + 2E_{01} E_{02} \cos(\alpha_1 - \alpha_2) \qquad ...(6.59)$$

This is the equation for amplitude E_0.

Dividing equation (6.58) by (6.57), we get

$$\tan \alpha = \frac{E_{01} \sin \alpha_1 + E_{02} \sin \alpha_2}{E_{01} \cos \alpha_1 + E_{02} \cos \alpha_2} \qquad ...(6.60)$$

Using equations (6.57) and Eq. (6.58) into (6.56), we obtain expression for the resultant disturbance.

$$E = E_0 \cos \alpha \sin \omega t + E_0 \sin \alpha \cos \omega t.$$

or $$E = E_0 \sin(\omega t + \alpha) \qquad ...(6.61)$$

Thus a single harmonic wave of the same frequency results from the superposition of two harmonic es of identical frequency. However, the amplitude and phase of the resultant wave are different n those of constituent waves.

The intensity of a wave is proportional to the square of the amplitude.

$$I \propto E_0^2 \qquad ...(6.62)$$

Using $I_1 = E_{01}^2$, $I_2 = E_{02}^2$ and $(\alpha_1 - \alpha_2) = \delta$ into (6.59), we obtain

$$\boxed{I = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos \delta} \qquad ...(6.63)$$

ollows that the resultant intensity is not simply the sum of the component intentsities and there is additional contribution $2\sqrt{I_1 I_2} \cos \delta$ known as the **interference term.**

(*i*) Whenever the phase difference $\delta = 0$, a maximum amount of light is obtained.

$$I_{max} = I_1 + I_2 + 2\sqrt{I_1 I_2}$$

Further, if $$I_1 = I_2$$
$$I_{ma} = 4I_1 \qquad ... (6.64)$$

(*ii*) Whenever the phase difference $\delta = 180°$, $\cos 180° = -1$ and a minimum intensity is observed.

$$I_{min} = I_1 + I_2 - 2\sqrt{I_1 I_2}$$

If $$I_1 = I_2, I_{min} = 0 \qquad .. (6.65)$$

Consequently, the region of space where the two waves overlap will have alternate bright and ark bands called **interference fringes.**

The phase difference δ in (6.63) is made up of two parts — one part arises from the difference in hases at the points of origin of the two waves while the second part is on account of the difference in ie paths traversed by them.

$$\delta = \alpha_1 - \alpha_2 = (-kx_1 + \phi_1) - (-kx_2 + \phi_2)$$
$$= k(x_2 - x_1) + (\phi_1 - \phi_2) = \frac{2\pi}{\lambda}\mu L + (\phi_1 - \phi_2)$$

or $$\delta = \frac{2\pi}{\lambda}\Delta + (\phi_1 - \phi_2) \qquad ...(6.66)$$

The term $(\phi_1 - \phi_2)$ is contribution of initial phase difference and the term $\frac{2\pi}{\lambda}\Delta$ is contribution lue to path difference. If the value of $(\phi_1 - \phi_2)$ stays constant in time, then the two waves are said to

6.21. YOUNG'S DOUBLE SLIT EXPERIMENT

The interference of light waves was first demonstrated by Thomas Young in 1801. A schematic of the apparatus used is shown in Fig. 6.22. A bright source illuminates a narrow slit S.

The light wave diffracts and illuminates two narrow, parallel slits S_1 and S_2, which are very close to each other. The slits S_1 and S_2 partition the incident wavefront. In accordance with Huygens principle, the two segments of the wavefront at slits S_1 and S_2 can be treated as point sources of waves. These sources are generated by the same primary wave and are, therefore, mutually coherent. The waves that spread from S_1 and S_2 partially overlap. The overlapping waves interfere and produce interference fringes on the screen (Fig. 6.22*b*).

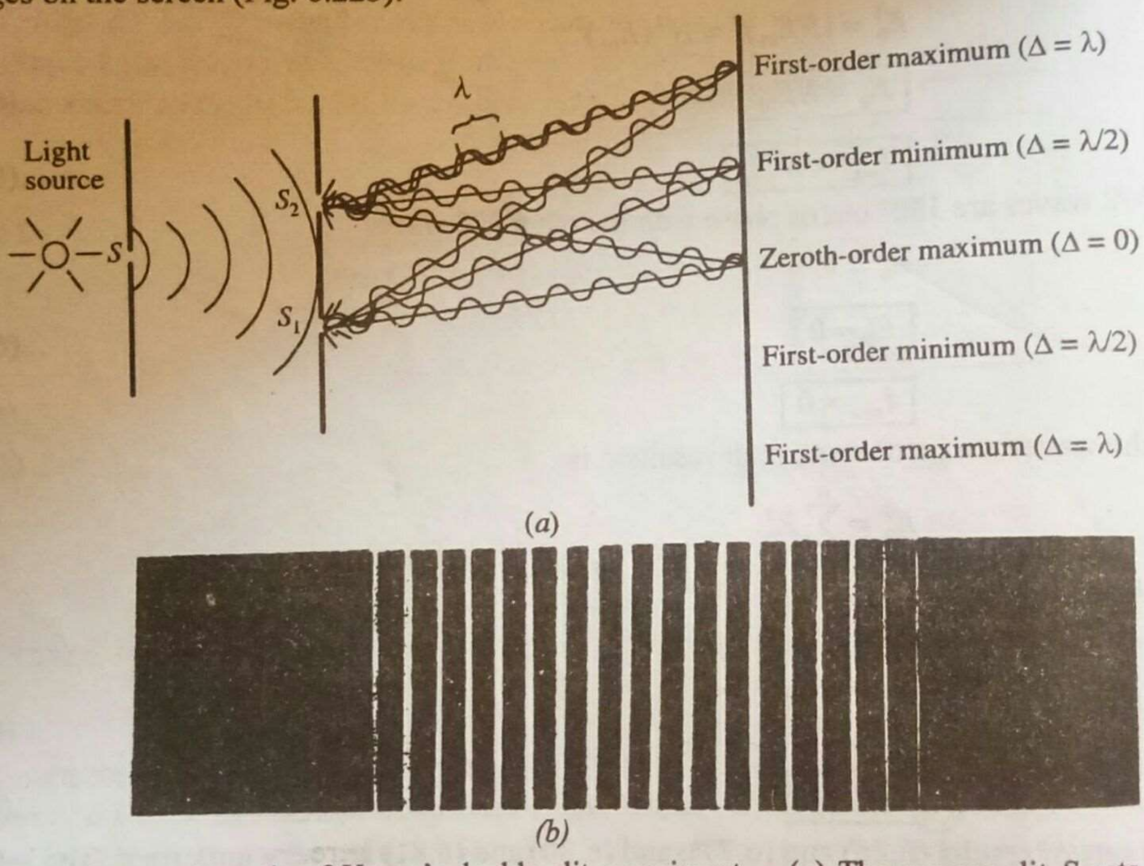


Fig. 6.22. Schematic of Young's double slit experiment — (*a*) The narrow slit S acts as a source of cylindrical waves which illuminate the slits S_1 and S_2. S_1 and S_2 behave as coherent sources and produce interference. (*b*) Interference fringes.

7.7. DIFFRACTION

The wave nature of light is further confirmed by the phenomenon of diffraction. The word diffraction is derived from the Latin word **diffractus** which means to break to pieces. When waves encounter obstacles (or openings), they bend round the edges of the obstacles if the dimensions of the obstacles are comparable to the wavelength of the waves. The bending of waves around the edges of an obstacle (or opening) is called **diffraction.**

Fig. 7.24 illustrates the passage of waves through an opening. When the opening is large compared to a wavelength, the waves do not bend round the edges. When the opening is small, the bending

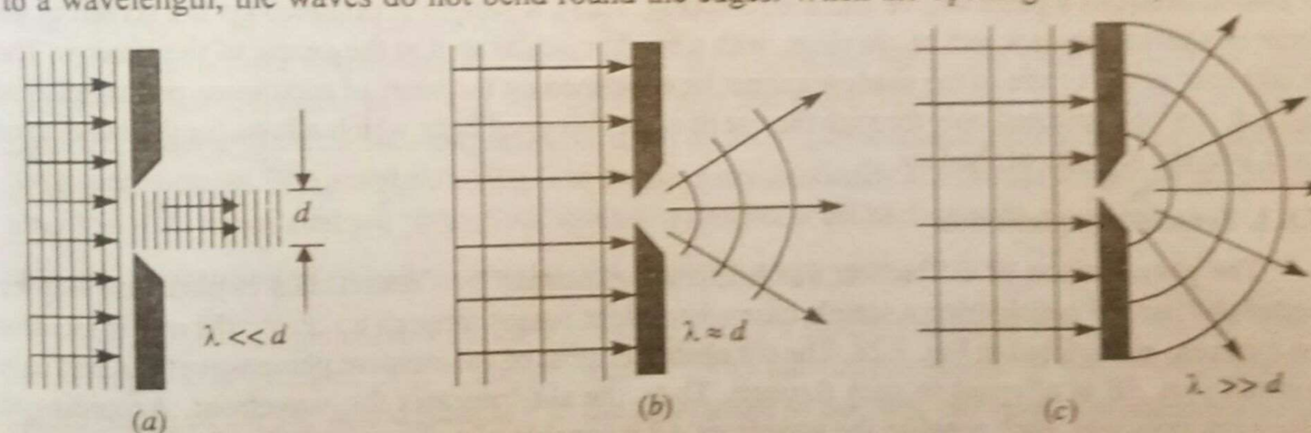


Fig. 7.24. Diffraction—(*a*) A plane wave does not bend at the slit if the opening $d \gg \lambda$. (*b*) Bending is perceptible when $\lambda \approx d$. (*c*) The bending takes place to such an extent that light can be perceived in a direction normal to the ray propagation, suggesting that the opening acts as a point source, when $\lambda > d$.

central bright region flanked by weaker maxima.

7.9.1 Formation of the Maxima and Minima

The wavefront incident on the slit is a plane wavefront from which the slit slices off a small part AB. Usually the size of d is of the order of 0.1 mm. Therefore, the size of the wavefront slice AB is also about 0.1 mm which is about 200 times the wavelength of incident light. According to Huygens principle, each point on AB acts as a source of secondary waves. It would then be appropriate to replace the wavefront AB with a string of point sources. As all the points on AB are **in phase,** the point sources will be **coherent**. Hence, the light from one portion of the slit can interfere with light from another portion and the resultant intensity on the screen will depend on the direction θ.

It can be seen from Fig. 7.27(a) that the waves travelling in a direction parallel to OP come to a focus at P. All these waves start at the slit in the same phase and after covering equal optical path lengths reach P in phase. Their superposition produces an intensity maximum at P. It is at the centre of the diffraction pattern and is called zero order **central maxima** (Fig. 7.27b).

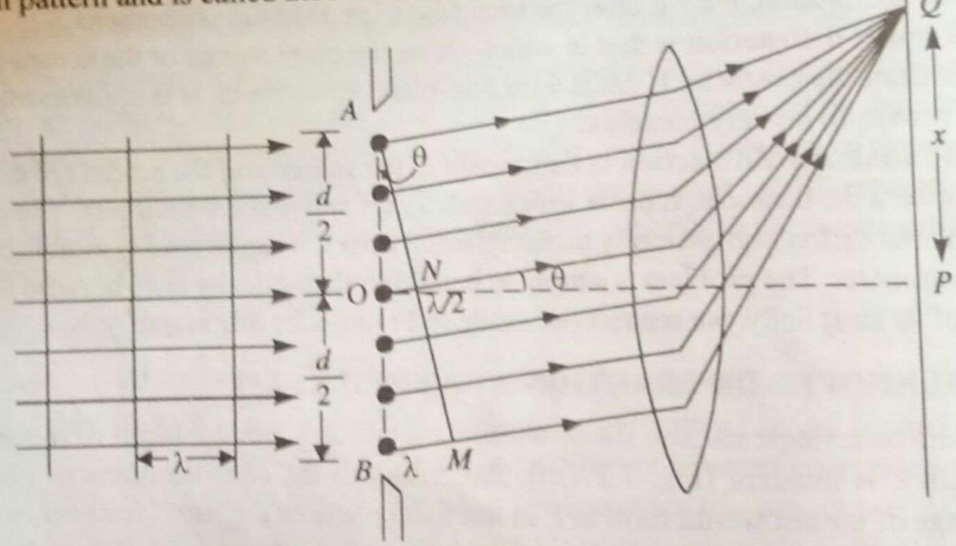


Fig..7.28. Conditions at the first minimum of diffraction pattern–Each point on the sliced portion AB of the wavefront acts as a point source. Any two waves that originate at points separated by $d/2$ distance are 180° out-of-phase and interfere destructively.

Let us consider another point Q on the screen (Fig. 7.28). The waves that leave the slit at an angle θ reach the point Q. The point Q will be bright or dark depending upon the path difference between the waves arriving at Q from different points on the wavefront AB. To find the resultant of the waves arriving at Q, it is convenient to divide the wavefront AB into two halves AO and OB. Waves ON and BM are in phase at the slit. Wave BM travels farther than ON by an amount of path difference equal to $(d/2) \sin \theta$. If this path difference is equal to $\lambda/2$, the two waves interfere destructively and produce darkness at Q. This is true for any two waves that originate at points separated by $(d/2)$ as the path difference between rays from two such points will be $\lambda/2$. Therefore, we may conclude that the waves from the upper portion AO interfere destructively with waves from the lower portion OB, if

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

or

$$\sin \theta = \lambda/d \qquad ...(7.52)$$

Therefore, the intensity at Q is zero and a dark band called the **first order minimum** is produced at Q. A similar dark band occurs at Q' below P at an angular distance θ governed by the equation (7.52). It is also known as the **first order minimum.**

We may also divide the slit into quarters, sixths and so on. It can be shown through arguments similar to above that a dark band occurs whenever

$$\sin \theta = 2\lambda/d,\ 3\lambda/d,\ 4\lambda/d, \quad ... \text{etc.}$$

They are known as second order minimum, third order minimum etc. The condition for the appearance of minima can be generalized as

$$\sin \theta = m\lambda/d$$

where $m = 1, 2, 3, ...$

It is more generally expressed as

$$\boxed{d \sin \theta = m\lambda} \qquad ...(7.53)$$

Since the first order minima occurs at an angle $\sin^{-1}(\lambda/d)$ and the second order minima at an angle $\sin^{-1}(2\lambda/d)$, it is natural to expect **first order maxima** half way between them, at $\theta = \sin^{-1}(3\lambda/2d)$. To arrive at this condition, we imagine the wavefront AB to be divided into three equal parts, as shown in Fig. 7.29. The waves from the extreme ends of the upper two parts will have a path difference $\lambda/2$. The waves from these two parts produce darkness at R. The waves from the third protion of AB are not cancelled and thus produce a weak maximum at R. As θ increases, still weaker maxima with rapidly falling of intensity are observed at $5\lambda/2d$, $7\lambda/2d$ ...etc.