NATIONAL OPEN UNIVERSITY OF NIGERIA

SCHOOL OF SCIENCE AND TECHNOLOGY

COURSE CODE: PHY306

COURSE TITLE: OPTICS I

Course Code	PHY 306
Course Title	OPTICS II
Course Developer	DR. A. B. Adeloye Dr. M. A. Chendo

PHYSICS DEPARTMENT
UNIVERSITY OF LAGOS

Programme Leader	Dr. Ajibola S. O. **National Open University of Nigeria** Lagos

OPTICS II

COURSE GUIDE

Course Introduction

In Optics I you studied the nature of light. There you studied that light is a wave motion. A very important characteristic of wave motion is the phenomenon of interference.

The term interference refers to the phenomenon that waves, under certain conditions, intensify or weaken each other. The phenomenon of interference is inseparably tied to that of diffraction. In fact, diffraction is more inclusive; it contains interference and, in a sense, even refraction and reflection. It is only because diffraction is mathematically more complex that we treat interference and diffraction in separate Blocks, and discuss interference first.

The prerequisite of all interference is the superposition of waves. If light from a source is divided by suitable apparatus into two beams and then superposed, the intensity in the region of superposition is found to vary from point to point between maxima, which exceed the sum of the intensities in the beams, and minima, which may be zero. This phenomenon is called interference. There are two general methods of obtaining beams from a single beam of light, and these provide a basis of classifying the arrangements used to produce interference. In one method, the beam is divided by passage through apertures placed side by side. This method, which is also called division of wave front, is useful only with sufficiently small sources. Alternatively, the beam is divided at one or more partially reflecting surfaces, at each of which, part of the light is reflected and part transmitted. This method is called division of amplitude. It can be used with extended sources, and so effects may be of greater intensity than with the division of the wavefront. In either case, it is convenient to consider separately the effects which result from the superposition of more than two beams (multiple beam interference).

The phenomenon of interference, is explained on the wave model of light. What may puzzle you is the fact that light casts shadows of objects, i.e., light appears to travel in straight lines rather than bending around obstacles. This apparent contradiction was explained by Fresnel. You will learn that the ease with which a wave bends around corners is determined by the size of the obstacle relative to wavelength of light. The wavelength of light is about 10^{-7} 10 m and the obstacles used in ordinary experiments are about 10^{5} times bigger. However, a large number of obstacles, whose sizes are comparable to the wavelength of light, do exhibit diffraction of light.

The phenomenon of diffraction was first observed by Grimaldi and a systematic explanation is due to Fresnel. According to him, in diffraction phenomenon, interference takes place between secondary wavelets from different parts of the same wavefront. Diffraction is classified in two categories: Fresnel diffraction and Fraunhofer diffraction.

Unit 1 begins with the study of wave motion. Being familiar to most students from their study of Oscillations and Waves, it will serve primarily as a review. With the help of the principle of superposition, we have explained the phenomenon of interference. In this unit, we discuss in detail the phenomenon of interference produced by the division of the wavefront of light wave.

In Unit 2, we will consider the formation of interference pattern by the division of amplitude. Such studies have many practical applications. Finally we briefly mention these applications.

Unit 3 is devoted to interferometry. It deals with Michelson interferometer, which is an example of two beam interference and Fabry-Perot interferometer which is an example of multiple beam interference. Finally, an appendix given at the end of the unit provides a brief introduction to complex amplitudes. You might like to read it to enrich your knowledge. However, you will not be examined on it.

For Fresnel diffraction, discussed in Unit 4, the experimental arrangement is fairly simple. The source or the observation screen or both are at a finite distance from the obstacle. But theoretical analysis of Fresnel diffraction, being essentially based on geometrical construction, is somewhat cumbersome. Nevertheless, Fresnel diffraction is more general; it includes Fraunhofer diffraction as a special case.

In Fraunhofer diffraction, the source of light and the observation screen (or human eye) are effectively at infinite distance from the obstacle. The Fraunhofer diffraction from a single slit is of particular interest in respect of the general theory of optical instruments. This is discussed in detail in Unit 5. You will learn that when a narrow vertical slit is illuminated by a distant point source, the diffraction pattern consists of a series of spots along a horizontal line and situated symmetrically about a central spot. For a circular aperture, the diffraction pattern consists of concentric rings with a bright central disc.

In Unit 6 you will learn about double slit and N-slit diffraction patterns. A distinct feature of double slit pattern is that it consists of bright and dark fringes similar to those observed in interference experiments. The N-slit diffraction pattern shows well-defined interference maximum. The sharpness of interference maximum increases as N increases. For a sufficiently large value of N, interference maxima become narrow lines. This is why diffraction gratings are an excellent tool in spectral analysis.

An important point to learn is that fringed (diffracted) image of a point source is not a geometrical point. And diffraction places an upper limit on the ability of optical devices to transmit perfect information about any object. That is, all optical systems are diffraction limited. In Unit 7 you will learn to characterise the ability of an optical instrument to distinguish two close but distinct diffraction images of two objects or wavelengths based on the Rayleigh criterion.

UNIT 1 INTERFERENCE BY DIVISION OF WAVEFRONT

Structure

1.1 INTRODUCTION

Anyone with a pan of water can see how the water surface is disturbed in a variety of characteristic patterns, which is due to interference between water waves. Similarly, interference occurs between sound waves as a result of which two people who hum fairly pure tones, slightly different in frequency, hear beats. But if we shine light from two torches or flashlights at the same place on a screen, there is no evidence of interference. The region of overlap is merely uniformly bright. Does it mean that there is no interference of light waves? The answer is 'No.'

The interference in light is as real an effect as interference in water or sound waves, and there is one example of it familiar to everybody — the bright colours of a thin film of oil spread out on a water surface. There are two reason why the interference of light is observed in some cases and not in others? Firstly, light waves have very short wavelengths — the visible part of the spectrum extends only from 400 mm for violet light to 700 mm for red light. Secondly, every natural source of light emits light waves only as short trains of random pulses, so that any interference that occurs is averaged out during the period of observation by the eye, unless special procedures are used.

Like standing waves and beats, the phenomenon of interference depends on the superposition of two or more individual waves under rather strict conditions that will soon be clarified. When interest lies primarily in the effects of enhancement or diminution of light waves, these effects are usually said to be due to the interference of light. When enhancement (or constructive interference) and diminution (or destructive interference) conditions alternate in a spatial display, the interference is said to produce a pattern of fringes as in the double slit interference pattern. The same condition may lead to enhancement of one colour at the expense of the other colour, producing interference colours as in the case of oil slicks and soap film about which you will study in next unit.

In this unit, we will consider the interference pattern produced by waves originating from two **point** sources. However, in the case of light waves, one cannot observe interference between the waves from two independent sources, although the interference does take place. Thus, one tries to derive the interfering waves from a single wave so that the constant phase difference is maintained between the interfering waves. This can be achieved by two methods. In the first method a beam is allowed to fall on two closely spaced holes, and the two beam emanating from the holes interfere. This method is known as division of wavefront and will be discussed in detail in this unit. In the other method, known as division of amplitude, a beam is divided at two or more reflecting surfaces, and the reflected beams interfere. This will be discussed in the next unit.

As the phenomenon of interference can be successfully explained by treating light as a wave motion, it is necessary to understand the fundamentals of wave motion. We shall therefore begin this unit with the study of wave motion which will serve as a recapitulation.

In the next unit we will study how interference takes place by division of amplitude of light wave.

Objectives
After studying this unit, you should be able to

- use the principle of superposition to interpret constructive and destructive interference,
- distinguish between coherent and incoherent sources of light,
- describe the origins of the interference pattern produced by double slit,
- describe the intensity distribution in interference pattern,
- express the fringe-width in terms of wavelength of light,
- describe various arrangements for producing interference by division of wavefront,
- appreciate the difference between Biprism and Lloyd's mirror fringes.

1.2 WAVE MOTION

Study Comment

You may find it useful to go through the Unit 6 of the course "Oscillations and Waves."

Simple Harmonic Motion
A simple harmonic motion is defined as the motion of a particle which moves back and forth along a straight line such that its acceleration is directly proportional to its displacement from a fixed point in the line, and is always directed towards that point.

The best and elementary way to represent a simple harmonic motion is to consider the motion of a particle along a reference circle (See Fig. 5.1). Suppose a particle *P* travels in a circular path, counterclockwise, at a uniform angular velocity ω . The point *N* is the perpendicular projection of *P* on the diameter *AOA'* of the circle. When the particle *P* is at point *B*, the perpendicular projection is at *0*. As the particle *P* starts from *B*, and moves round the circle, *N* moves from *0* to *A*, *A* to *A'* and then returns to *O*. This back and forth motion of *N* is simple harmonic. Let us obtain expressions for displacement, velocity and acceleration and define few terms.

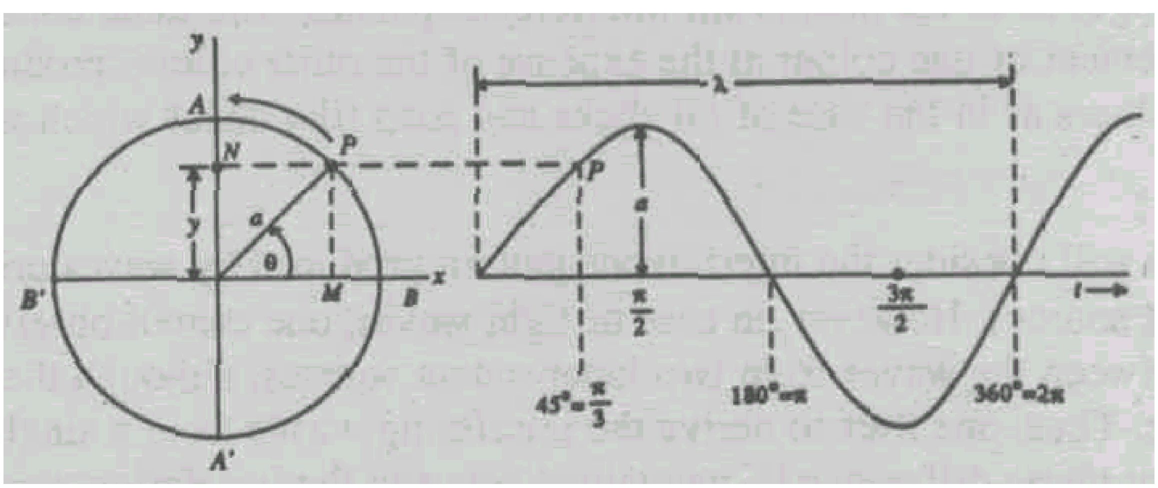


Fig. 5.1. Reference Circle (Left) and Simple Harmonic Motion (Right)

Displacement

Suppose the particle *P* starts from *B* and traces an angle θ in time *t*. Then its angular velocity ω is

$$\omega = \frac{\theta}{t}$$

where the angle θ is measured in radians. The displacement, *y*, of *N* from O at time *t*, is thus given by

$$y = ON = OP \sin NPO$$

$$= a \sin\theta \qquad [\because \angle NPO = \angle POB = \theta]$$

But $\omega = \frac{\theta}{t}$, so that $\theta = \omega t$

$$\therefore \quad y = a \sin\omega t \tag{1.1}$$

This is the equation of simple harmonic motion.

SAQ 1

See Fig. 1.1. If you have studied the motion of the point M, which is the foot of the perpendicular from the point *P* on the x-axis, then write down the equation of simple harmonic motion.

Velocity: The velocity of *N* is given by

$$\frac{dy}{dt} = a\omega \cos\omega t = \omega\sqrt{a^2 - y^2} \tag{1.2}$$

Acceleration: The acceleration of *N* is

$$\frac{d^2y}{dt^2} = -\omega^2 a \sin\omega t = -\omega^2 y \tag{1.3}$$

Periodic Time: The periodic time, *T, of N* is time taken by *N* to make one complete vibration. Thus

$$T = \frac{2\pi}{\omega} \tag{1.4}$$

Amplitude: Amplitude of vibration is equal to the radius of the reference circle i.e., *a*.

SAQ 2

A particle is executing simple harmonic motion, with a period of 3s and an amplitude of 6 cm. One-half second after the particle has passed through its equilibrium position, what is its (a) displacement, (b) velocity, and (c) acceleration?

Phase: The phase of a vibrating particle represents its state as regards

(i) the amount of displacement suffered by the particle with respect to its mean position, and

(ii) the direction in which the displacement has taken place.

In Fig. 1.1, we have conveniently chosen t = 0 as the time when P was on the x-axis. The choice of the time t = 0 is arbitrary, and we could have chosen time t = 0 to be the instant when P was at P' (see Fig. 1.2). If the angle $P'OX = \theta$, then the projection on the y-axis at any time t would be given by

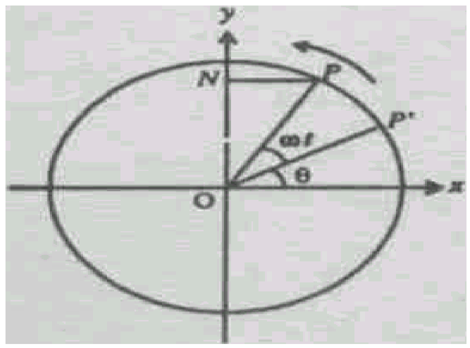


Fig. 1.2 At t = 0, the point P is at P' and, therefore, the initial phase is θ

$$y = a\sin(\omega t + \theta) \tag{1.5}$$

The quantity $(\omega t + \theta)$ is known as the phase of the motion and θ represents the initial phase. It is obvious from the discussion that the value of θ is quite arbitrary, and depends on the instant from which we start measuring time.

We next consider two particles, P and Q rotating on the circle with the same angular velocity ω and P' and Q' are their respective positions at t = 0. Let the angle $\angle P'OX$ and $\angle Q'OX$ be θ and ϕ respectively (see Fig. 1.3).

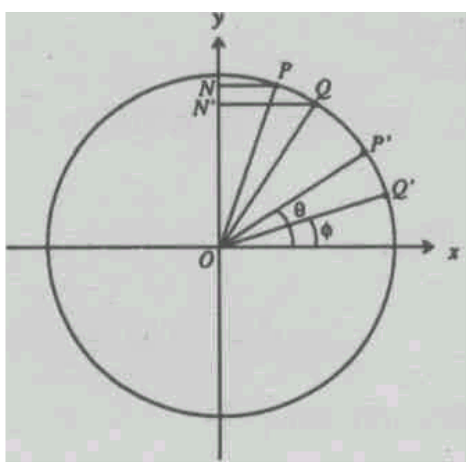


Fig.1.3: The points N and N' execute simple harmonic motion with the same frequency ω. The initial phases of N and N' are θ and ϕ respectively

Clearly at an arbitrary time t the distance of the foot of perpendiculars from the origin would be

$$y_P = a\sin(\omega t + \theta) \tag{1.6a}$$

$$y_Q = a\sin(\omega t + \phi) \tag{1.6b}$$

The quantity

$$(\omega t + \theta) - (\omega t + \phi) = \theta - \phi \tag{1.7}$$

represents the phase difference between the two simple harmonic motions and if $\theta - \phi = 0$ (or an even multiple of π) the motions are said to be **in phase,** and if $\theta - \phi = \pi$ (or an odd multiple of π), the motions are said to be **out of phase.** If we choose a different origin of time, the quantities θ and ϕ would change by the same additive constant; consequently, the phase difference $(\theta - \phi)$ is independent of the choice of the instant $t = 0$.

Energy: A particle performing simple harmonic motion possesses both types of energies: potential and kinetic. It possesses potential energy on account of its displacement from the equilibrium position and kinetic energy on account of its velocity. These energies vary during oscillation; however, their sum is conserved provided no dissipative forces are present. Since the acceleration of vibrating particle is $\omega^2 y$, the force needed to keep a particle of mass m at a distance y from O is $m\omega^2 y$. If the particle is to be displaced through a further distance dy, the work to be done will be $m\omega^2 y\,dy$. Now the potential energy of the particle at a displacement y is equal to the total work done to displace the particle from O through a distance y.

$$\therefore \qquad \text{P.E.} = \int_0^y \omega^2 y\, my\, dy = \frac{1}{2} m\omega^2 y^2 \tag{1.8}$$

Using Eq. (1.2), the kinetic energy of the particle is given by

$$\text{K. E.} = \frac{1}{2} m\left(\frac{dy}{dt}\right)^2 = \frac{1}{2} m\omega^2 (a^2 - y^2) \tag{1.9}$$

The total energy of the particle at any distance y from 0 is given by

$$\text{Total energy} = \text{K.E.} + \text{P.E.} \qquad = \frac{1}{2} m\omega^2 (a^2 - y^2) + \frac{1}{2} m\omega^2 y^2$$

$$= \frac{1}{2} m\omega^2 a^2 \tag{1.10}$$

Therefore, the total energy (intensity) is proportional to (amplitude)^2, and, since $\omega = 2\pi f$, f being the frequency, the energy is also proportional to (frequency)^2.

If I represents the intensity associated with a light wave then

$$I \propto a^2$$

where a represents the amplitude of the wave.

Wave-motion

So far we considered a single particle, *P*, executing simple harmonic motion. Let us consider a number of particles which make a continuous elastic medium. If any one particle is set in vibration, each successive particle begins a similar vibration, but a little later than the one before it, due to inertia. Thus, the phase of vibration changes from particle to particle until we reach a particle at which the disturbance arrives exactly at the moment when the first particle has completed one vibration. This particle then moves in the same phase as the first particle. This simultaneous vibrations of the particles of the medium together make a wave. Such a wave can be represented graphically by means of a displacement curve drawn with the position of the particles as abscissa and the corresponding displacement at that instant as ordinate. If the particles execute simple harmonic motion, we obtain a sine curve as shown in Fig. 1.4.

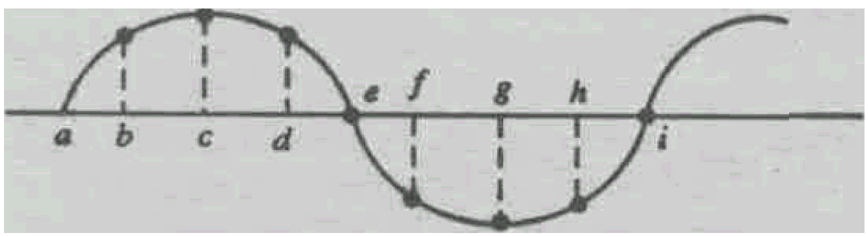


Fig. 1.4. Graphical representation of a wave

It will be seen that the wave originating at *a* repeats itself after reaching *i*. The distance *ai*, after travelling which the wave-form repeats itself, is called the wavelength and is denoted by λ. It is also evident that during the time *T*, while the particle at *a* makes one vibration, the wave travels a distance λ. Hence the velocity *v* of the wave is given by

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

If n is the frequency of vibration then f = 1/*T*. Hence, we have

$$v = n\lambda \tag{1.11}$$

Particles in Same Phase

Particles *a* and i have equal displacements (= zero) and both are tending to move upwards. They are said to be in the same phase. The distance between them is one wavelength. Hence, wavelength is the distance between two nearest particles vibrating in the same phase. Two vibrating particles will also be in the same phase if the distance between them is $n\lambda$, where *n* is an integer.

Particles in Opposite Phase

Particles a and e both have the same displacement (= zero), but while a is tending to go up, e is tending to move downwards. They are said to be in opposite phase. The distance between them is $\lambda/2$. The particles are out of phase if the distance between them is $(2n-1)\lambda/2$, where n is an integer.

Equation of a Simple Harmonic Wave

Fig. 1.5 shows the wave travelling in the positive x-direction. The displacement y of the particle at *0* at any time t is given by

$$y = a\sin\omega t \tag{5.1}$$

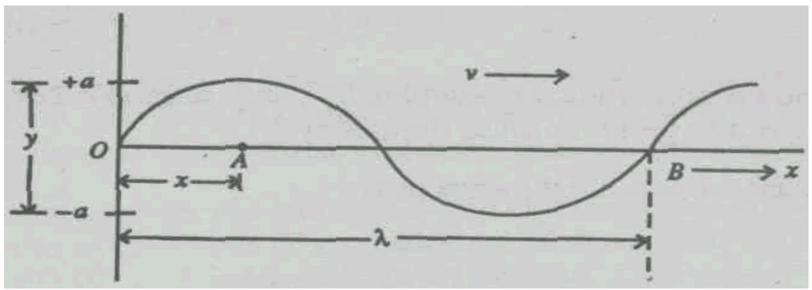


Fig. 1.5 A simple harmonic wave travelling towards right

Let v be the velocity of propagation of the wave. Then the wave starting from *0* would reach point *A*, distant *x* from *0* in x/v seconds. Hence, the particle at *A* must have started its vibration x/v seconds later than the particle at *0*, Consequently, the displacement at *A* at the time *t* would be same as was at *O* at time x/v seconds earlier, i.e., at time $t-\frac{x}{v}$. Substituting $t-\frac{x}{v}$ for t in Eq. (1.1) we obtain the displacement at *A* at time *t*, which is given by

$$y = a\sin\omega\left(t-\frac{x}{v}\right)$$

Using the relation $\omega = 2\pi/T$ and $v = \lambda/T$, we get

$$y = a\sin\frac{2\pi}{\lambda}(vt-x) \tag{1.12}$$

This equation represents the displacement of a particle at a distance *x* from a fixed point at a time *t*. This is, therefore, the equation of the wave. The wave shown in Fig. 1.5 is generated along a stretched string and in a rope. Such types of waves are called transverse waves. From Unit 4 of Optics I, you already know that light travels in the form of transverse waves, therefore Eq. (1.12) represents a light wave.

Relation between Phase Difference and Path Difference

The equation of simple harmonic wave is given by Eq. (1.12). If there are two particles P_1 and P_2 at distance x_1 and x_2 from the origin, then, the phase angle of P_1 at a time t equals $\frac{2\pi}{\lambda}(vt - x_1)$

the phase angle of P_2 at a time t equals $\frac{2\pi}{\lambda}(vt - x_2)$

$\therefore$ phase difference between P_1 and P_2 equals

$$\frac{2\pi}{\lambda}(vt - x_1) - \frac{2\pi}{\lambda}(vt - x_2) = \frac{2\pi}{\lambda}(x_2 - x_1)$$

But $(x_2 - x_1)$ is the path difference between P_2 and P_1.

$$\therefore \text{ Phase difference} = \frac{2\pi}{\lambda} \times (\text{path difference})$$

> The expression $\frac{2\pi}{\lambda} \times$ (path difference) can be obtained in a less formal manner by remembering that a difference in phase of 2π corresponds to a path difference of one wavelength and calculating the required phase difference by proportion.

When two or more sets of waves are made to overlap in some region of space, interesting effects are observed. For example, when two stones are dropped simultaneously in a quiet pool, two sets of waves are created. In the region of crossing, there are places where the disturbance is almost zero, and others, where it is greater than that given by either wave alone. These effects can be explained using a very simple law known as the principle of superposition. We will use this principle in investigating the disturbance in regions where two or more light waves are superimposed. Let us now briefly study this principle.

1.3 PRINCIPLE OF SUPERPOSITION

In any medium, two or more waves can travel simultaneously without affecting the motion of each other. Therefore, at any instant the resultant displacement of each particle of the medium is merely the vector sum of displacements due to each wave separately. This principle is known as "the principle of superposition." It has been observed that when two sets of waves are made to cross each other, then after the waves have passed out of the region of crossing, they appear to have been entirely uninfluenced by the other set of waves. Amplitude, frequency and all other characteristics of the waves are as if they had crossed an undisturbed space.

As a simple example, we consider a long stretched string *AB* (see Fig. 1.6). The end *A* of the string is made to vibrate up and down. This vibration is handed down from particle to particle of the string. Let the string be vibrating in the form of a triangular pulse, which propagates to the right with a certain speed *v*. We next assume that from the end *B* an identical pulse is generated which starts moving to the left with the same speed *v*.

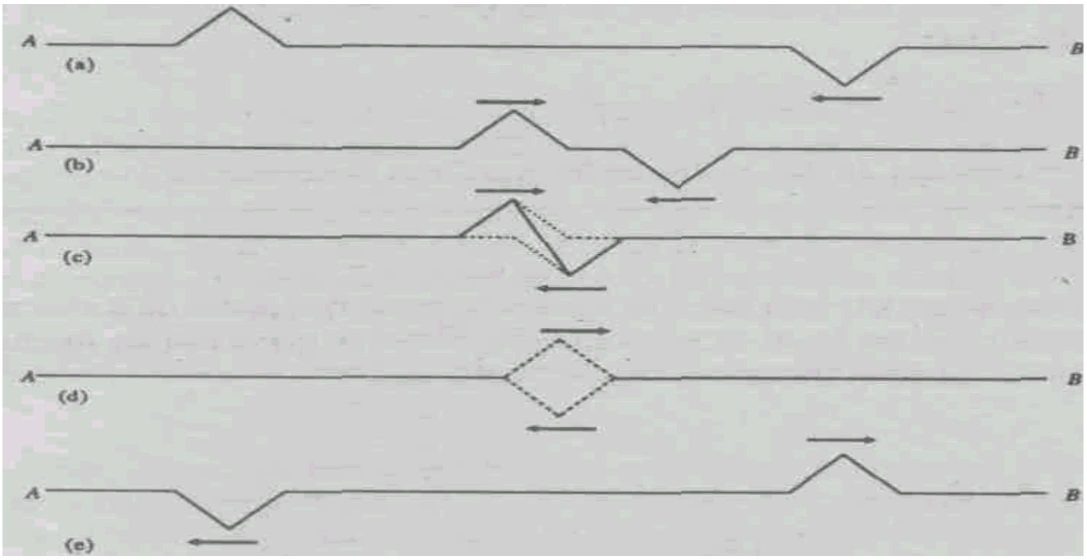


Fig. 1.6 The propagation of two triangular pulses in opposite directions in a stretched string. The solid line gives the actual shape of the string; (a), (b), (c), (d) and (e) correspond to different instants of tune.

Fig. 1.6(a) shows the position of pulse at $t = 0$. At a little later time, each pulse moves close to the other as shown in Fig. 1.6(b), without any interference. Fig. 1.6(c) represents the position at an instant when the two pulses interfere; the dashed curves represent the profile of the string, if each of the impulses were moving all by itself, whereas the solid curve shows the resultant displacement obtained by the algebraic addition of each displacement. Shortly later in Fig. 1.6(d) the two pulses overlap each other and the resultant displacement is zero everywhere. At a much later time, the impulses cross each other (Fig. 1.6(e)) and move as if nothing had happened. This could hold provided the principle of superposition is true.

Let us consider the following case of superposition of waves.

Superposition of Two Waves of Same Frequency but having Constant Phase Difference
Consider two waves of same frequency but having constant phase difference, say δ. Since they have same frequency, i.e., the same angular velocity, we write

$$y_1 = a_1 \sin \omega t$$

and

$$y_2 = a_2 \sin(\omega t + \delta)$$

where a_1 and a_2 are two different amplitudes, and ω is the common angular frequency of the two waves. By the principle of superposition, the resultant displacement is

$$\begin{aligned} y &= y_1 + y_2 \\ &= a_1 \sin \omega t + a_2 \sin(\omega t + \delta) \\ &= a_1 \sin \omega t + a_2 \sin \omega t \cos\delta + a_2 \cos\omega t \sin \delta \\ &= \sin \omega t (a_1 + a_2 \cos\delta) + \cos\omega t (a_2 \sin \delta) \end{aligned}$$

Let us write

$$a_1 + a_2 \cos\delta = A\cos\theta \tag{1.14a}$$

and

$$a_2 \sin\delta = A\sin\theta \tag{1.14b}$$

where A and θ are new constants. This gives

$$y = \sin\omega t A\cos\theta + \cos\omega t A\sin\theta$$

or

$$y = A\sin(\omega t + \theta)$$

Hence the resultant displacement is simple harmonic and of amplitude A. Squaring and adding Eq. 1.14a and 1.14b, we get

$$A^2\cos^2\theta + A^2\sin^2\theta = (a_1 + a_2\cos\delta)^2 + (a_2\sin\delta)^2$$

or

$$A^2 = a_1^2 + a_2^2 + 2a_1a_2\cos\delta$$

Thus, the resultant intensity I which is proportional to the square of the resultant amplitude, is given as

$$I = A^2 = a_1^2 + a_2^2 + 2a_1a_2\cos\delta \tag{1.15}$$

(Here we have taken the constant of proportionality as 1, for simplicity).

Thus, we find that the resultant intensity is not equal to the sum of the intensities due to separate waves i.e., $(a_1^2 + a_2^2)$. Since the intensity of wave is proportional to square of amplitude, $I_1 \propto a_1^2$ and $I_2 \propto a_2^2$. As before, taking the proportionality constant as 1 , we can rewrite Eq. (1.15) as

$$I = I_1 + I_2 + 2\sqrt{I_1I_2}\cos\delta \tag{1.16}$$

In Example 1, see how Eq. (1.16) has been used to find the resultant intensity.

Example 1
Consider interference due to two coherent waves of the same frequency and constant phase difference having intensities I and $4I$, respectively. What is the resultant intensity when the phase difference between these two waves is $\pi/2$ and π?

Solution
According to Eq. (1.16)

$$I \quad = I_1 + I_2 + 2\sqrt{I_1I_2}\cos\delta$$

Given:

$I_1 = I$ and $I_2 = 4I$, so

$$I = 5I + 2I\sqrt{4}\cos\delta$$
$$= 5I + 4I\cos\delta$$

Hence

$$I_{\pi/2} = 5I + 4I\cos 90^0 = 5I$$

$$I_{\pi} = 5I + 4I\cos\pi = I$$

Thus there is a variation of intensity due to interference phenomenon.

Refer again to Eq. (1.16). The intensity I is maximum when $\cos\delta = +1$, that is, when phase difference is given by

$$\delta = 2n\pi \qquad \text{(even multiple of } \pi\text{)}$$

From Eq. (1.16)

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

The resultant intensity is, thus, greater than the sum of the two separate intensities. If

$I_1 = I_2$, then, $I_{\max} = 4I_1$

The intensity I is minimum when $\cos\delta = -1$, i.e., when δ is given by

$$\delta = (2n+1)\pi \quad \text{(odd multiple of } \pi\text{)}$$

We have from Eq. (1.16)

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

The resultant intensity is thus less than the sum of two separate intensities. If $I_1 = I_2$, then $I_{\min}$ = 0, which means that there is no light.

SAQ 3

Two waves of same frequency and constant phases difference have intensities in the ratio 81:1. They produce interference fringes. Deduce the ratio of the maximum to minimum intensity.

In general, for the two waves of same intensity and having a constant phase difference of δ, the resultant intensity is given by

$$I = 2I_1 + 2I_1\cos\delta$$
$$= 2I_1(1+\cos\delta)$$

$$= 4I_1 \cos^2 \frac{\delta}{2}$$

Therefore, we find that when two waves of the same frequency travel in approximately the same direction and have a phase difference that remains constant with the passage of time, the resultant intensity of light in not distributed uniformly in space. The non-uniform distribution of the light intensity due to the superposition of two waves is called **interference.** At some points the intensity is maximum and the interference at these points is called constructive interference. At some other points the intensity is minimum and the interference at these points is called destructive interference.

> Usually, when two light waves are made to interfere, we get alternate dark and bright bands of a regular or irregular shape. These are called interference fringes.

SAQ 4

Fig. 1.7 shows two situations where waves emanating from two sources, *A* and *B*, arrive at point C and interfere. Which of the two situations indicate constructive interference and destructive interference? Give reasons. (Eq. (1.13) will help you in answering this question.)

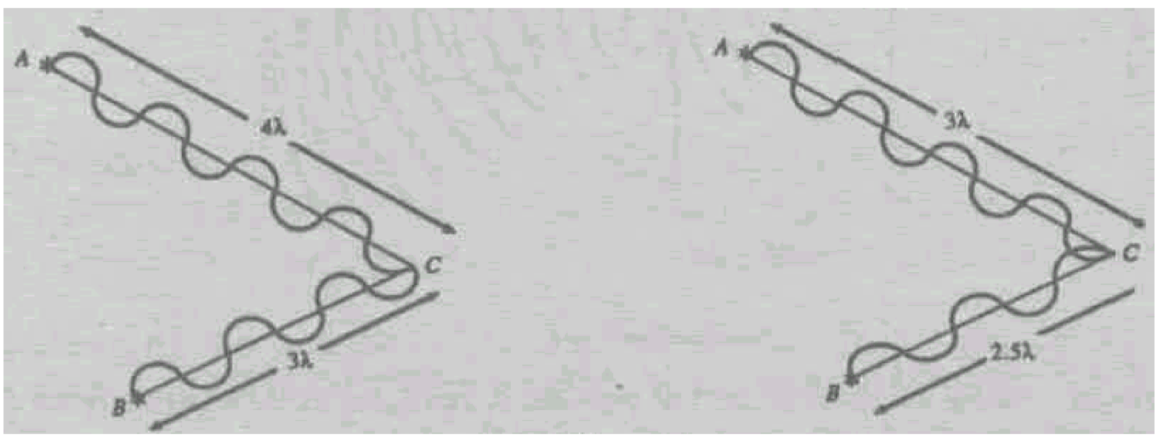


Fig. 5.7

After solving the above SAQ one can infer that: for constructive interference,

$$\boxed{\text{path difference} = n\lambda\text{, where } n = 0, 1, 2, 3, \ldots} \qquad (1.18)$$

for destructive interference,

$$\boxed{\text{path difference} = m\frac{\lambda}{2}\text{, where } m = 1, 3, 5, 7, \ldots} \qquad (1.19)$$

For the production of stationary interference patterns, i.e. definite regions of constructive and destructive interference, the interfering waves must have (1) the same frequency, and (2) a constant phase difference (and they must be travelling in the same or nearly the same direction).

If these conditions are satisfied, we say the wave sources and the waves are **coherent.** Sources can readily be found with the same vibrating frequency; however, the phase relationship between the waves may vary with time. In the case of light, the waves are radiated by the atoms of a source. Each atom contributes only a small part to the light emitted from the source and the waves bear no particular phase relationship to each other; the atoms randomly emit light, so the phase "constant" of the total light wave varies with time. Hence, light waves brought together from different light sources are coherent over very short periods of time and do not produce stationary interference patterns. Light from two lasers (about this you will study in Optics III) can be made to form stationary interference patterns, but the lasers must be phase-locked by some means. How, then, was the wave nature of light originally investigated, since lasers are a relatively recent development? In the following sections we will discuss the various arrangements, which provide coherent sources and enable us to observe interference phenomenon. Thomas Young had first demonstrated the interference of light. In the next section we will describe the experiment done by him.

1.4 YOUNG'S DOUBLE-SLIT EXPERIMENT

One of the earliest demonstrations of such interference effect was first done by Young in 1801, establishing the wave character of light. Young allowed sunlight to fall on a pinhole S_0, punched in a screen *A* as shown in Fig. 1.8. The emergent light spreads out and falls on pinholes S_1 and S_2, punched in the screen *B*. Pinholes S_1 and S_2 act as coherent sources. Again, two overlapping spherical waves expand into space to the right of screen *B*, Fig. 1.8 shows how Young produced an interference pattern by allowing the waves from pinholes S_1 and S_2 to overlap on screen C.

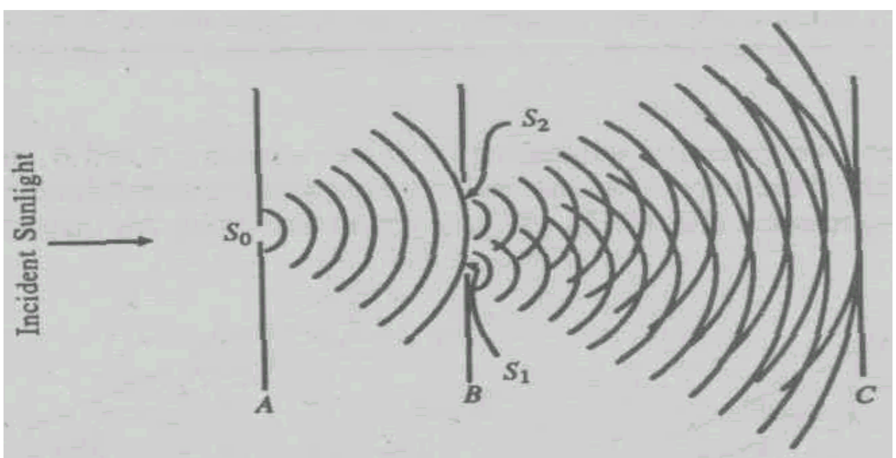


Fig.1.8 Young's double slit experiment. The pinholes S, and S_2 act as coherent sources and an interference pattern is observed on the screen *C.*

Fig. 1.9 shows the section of the wavefront on the plane containing S_0, S_1 and S_2. Since the waves emanating from S_1 and S_2 are coherent, we will see alternate bright and dark curves of fringes, called interference fringes. The interference pattern is symmetrical about a bright central fringe (also called maximum), and the bright fringes decrease in intensity, the farther they are from the central fringe.

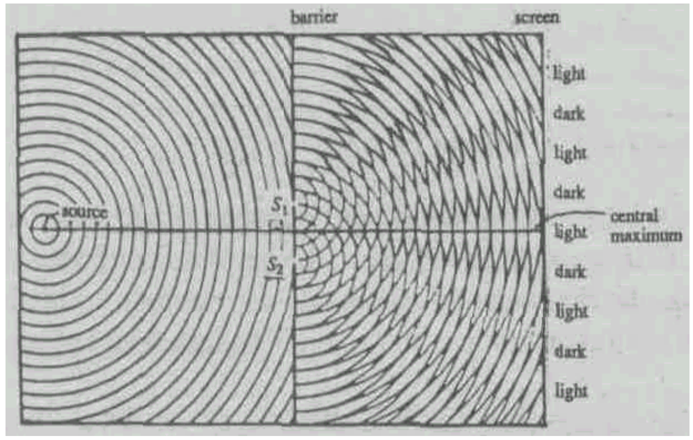


Fig. 1.9 Sections of the spherical wavefronts emanating from S_0, S_1 and S_2

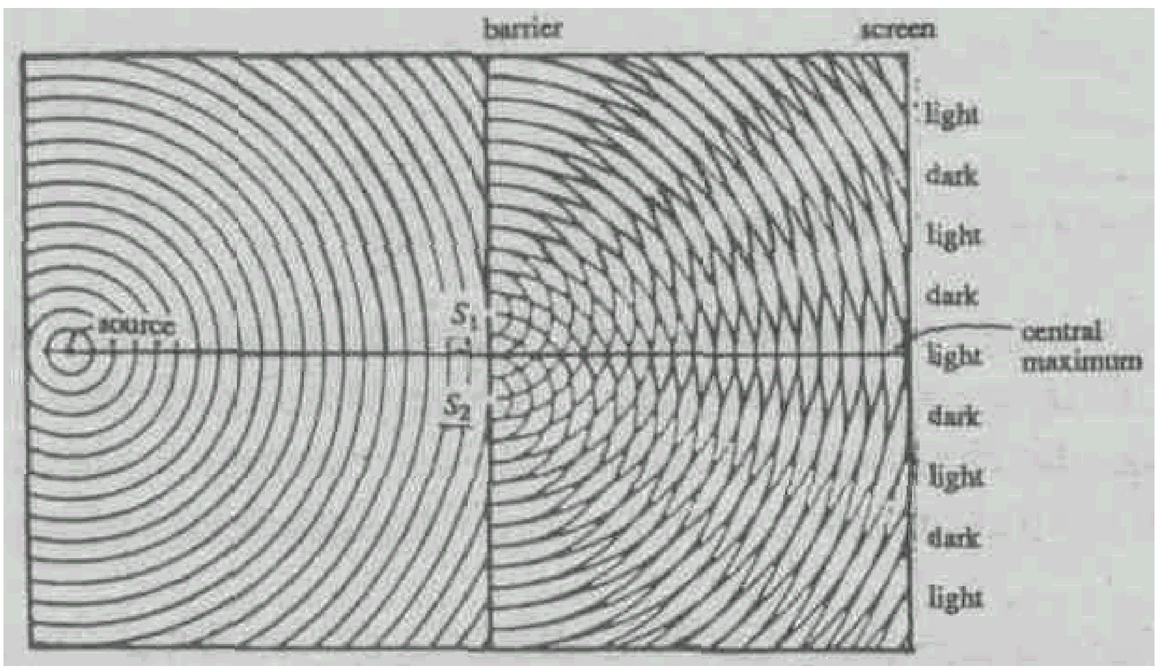


Fig. 1.9: Sections of the spherical wavefronts emanating from S_0, S_1 and S_2

To analyse the interference pattern and investigate the spacing of the interference fringes, consider the geometry in Fig. 1.10. Let *S* be a narrow slit illuminated by monochromatic light, and S_1 and S_2 two parallel narrow slits very close to each other and equidistant from *S*. The light waves from *S* arrive at S_1 and S_2 in the same phase. Beyond S_1 and S_2, the waves proceed as if they started from S_1 and S_2 with the same phase because the two slits are equidistant from *S*.

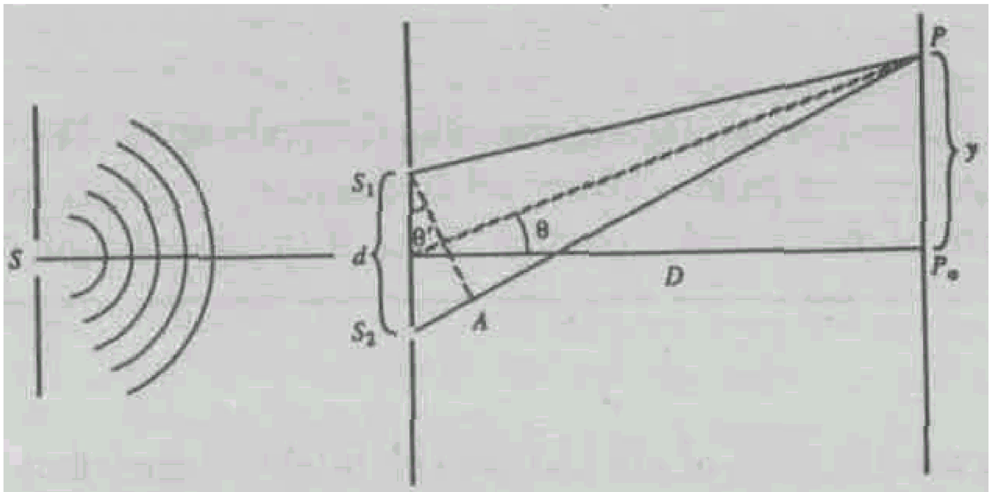


Fig. 1.10 The geometry of Young's experiment: The path difference of the light from the slits arriving at *P on* the screen is $d \sin\theta$

It is assumed that the waves start out at the same phase, because the two slits S_1 and S_2 are equidistant from *S*. Furthermore, the amplitudes are the same, because S_1 and S_2 are the same size slits and very close to each other. (So the amplitude does not vary very much.) Hence these waves produce an interference pattern on a screen placed parallel to S_1 and S_2.

To find the intensity at a point *P* on the screen, we join S_1P and S_2P. The two waves arrive at *P* from S_1 and S_2 having traversed different paths S_1P and S_2P. Let us calculate this path difference $S_2P - S_1P$. Let

y = distance of *P* from P_0, the central point on the screen

d = separation of two slits S_1 and S_2

D = distance of slits from the screen

The corresponding path difference is the distance S_2A in Fig. 1.10, where the line S_1A has been drawn to make S_1 and *A* equidistant from *P*. As Young's experiment is usually done with $D >> d$, the angles θ and θ'' are nearly same and they are small.

Hence, we may assume triangle $S_1A\,S_2$ as a right-angled triangle and $S_2A = d \sin\theta' = d \sin\theta = d \tan\theta$, as for small θ, $\sin\theta = \tan\theta$. As can be seen from the Fig. 1.10, $\tan\theta = y/D$.

$$\therefore \qquad S_2P - S_1P = S_2A = d\,\frac{y}{D} \tag{1.20}$$

Now the intensity at the point *P* is a maximum or minimum according as the path difference $S_2P - S_1P$ is an integral multiple of wavelength or an odd multiple of half wavelength (See Eq. 1.18 and Eq. 1.19). Hence, for bright fringes (maxima),

$$S_2P - S_1P = \frac{yd}{D} = 0\,,\ \lambda\,,\ 2\lambda\,,\ 3\lambda\,,\ \ldots = m\lambda$$

where m = 0, 1, 2, ...

$$\therefore \qquad y = mD\lambda / d \qquad \text{(bright fringes)} \tag{1.21}$$

The number *m* is called the order of the fringe. Thus the fringes with m = 0, 1, 2,.... etc. are called the zero, first, second....etc. orders. The zeroth order fringe corresponds to the central maximum, the first order fringe (*m* = 1) corresponds to the first bright fringe on either side of the central maximum, and so on. For dark fringes (minima),

$$S_2P - S_1P = \frac{yd}{D} = \frac{\lambda}{2}, \frac{3\lambda}{2}, \frac{5\lambda}{2}, \ldots, \left(m + \frac{1}{2}\right)\lambda$$

where m = 0, 1, 2, …

$$\therefore \qquad y = \left(m + \frac{1}{2}\right)\frac{\lambda D}{d} \qquad \text{(dark fringes)} \tag{1.22}$$

Eq. (1.21) or Eq. (1.22) can be used to find out the distance v« of the nth order bright (or dark) fringe. Try to solve the following SAQ.

SAQ 5
Monochromatic light passes through two narrow slits 0.40 mm apart. The third-order bright fringe of the interference pattern, observed on a screen 1.0 meter from the slits, is 3.6 mm from the centre of the central maximum. What is the wavelength of the light ?

Fringe Width

If y_n and y_{n+1} denote the distances of the *n*th and the $(n+1)$th bright fringes, then

$$y_n = \frac{D}{d} n\lambda$$

and

$$y_{n+1} = \frac{D}{d}(n+1)d$$

The spacing between the *n*th and the $(n+1)$th fringes (bright) is given by

$$y_{n+1} - y_n = \frac{D}{d}(n+1)\lambda - \frac{D}{d}n\lambda = D\lambda / d$$

It is independent of *n*. Hence, the spacing between any two consecutive bright fringes is the same. Similarly, it can be shown that the spacing between two dark fringes is also $\frac{D}{d}\lambda$. The spacing between any two consecutive bright or dark fringes is called the **fringe-width,** which is denoted by β. Thus

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

One also finds, by experiment, that fringe-width

(i) varies directly as *D*,

(ii) varies directly as the wave-length of the light used, and

(iii) inversely as the distance *d* between the slits.

The fringe-widths are so fine that to see them, one usually uses a magnifier or eyepiece.

To make certain that you really understand the meaning of the fringe width, try the following SAQs.

SAQ 6
In a two-slit interference pattern with λ = 6000 Å, the zero order and tenth order maxima fall at 12.34 mm and 14.73 mm respectively. Find the fringe width.

SAQ7
If in the SAQ 6, A is changed to 5000 A, deduce the positions of the zero order and twentieth order fringes, other arrangements remaining the same.

Shape of the Interference Fringes
In Fig. 1.11, suppose S_1 and S_2 represent the two coherent sources. At the point *P*, there is maximum or minimum intensity according as

$$S_2P - S_1P = n\lambda$$

or

$$S_2P - S_1P = \left(n + \frac{1}{2}\right)\frac{\lambda}{2}$$

Thus for a given value of *n*, the locus of points of maximum or minimum intensity is given by

$$S_2P - S_1P = \text{constant}$$

which is the equation of a hyperbola with S_1 and S_2 as foci. In space, the locus of points of maximum or minimum intensity for a particular value of *n* will be a hyperboloid of revolution, obtained by revolving the hyperbola about the line S_1S_2.

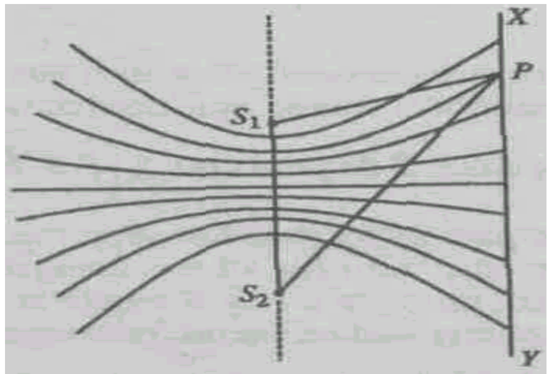


Fig. 1.11 Shape of the fringes

In practice, fringes are observed on a screen *XY* in a plane normal to the plane of the figure and parallel to the line joining S_1S_2. Hence the fringes that are observed are simply the sections of the hyperboloids by this plane, i.e., they are hyperbolae. Since the wavelength of light is extremely small (of the order of 10^{-5} cm), the value of $(S_2P - S_1P)$ is also of that order. Hence these hyperbolae appear, more or less, as straight lines.

Intensity Distribution in the Fringe-System

To find the intensity, we rewrite Eq. (1.15), taking $a_1 = a_2$, as follows

$$I = A^2 = 2a^2(1+\cos\delta)$$

$$= 4a^2\cos^2\frac{\delta}{2}$$

If the phase difference is such that $\delta = 0, 2\pi, 4\pi, \ldots$, this gives $4a^2$ or 4 times the intensity of either beam. If $\delta = \pi, 3\pi, 5\pi, \ldots$, the intensity is zero.

In-between, the intensity varies as $\cos^2\delta/2$. Fig. 1.12 shows a plot of intensity against the phase difference. When the two beams of light arrive at a point on the screen, exactly out of phase, they interfere destructively, and the resultant intensity is zero. One may well ask what becomes of the energy of the two beams, since the law of conservation of energy tells us that it cannot be destroyed. The answer to this question is that the energy, which apparently disappears at the minima, is actually still present at the maxima, where the intensity is greater than would be produced by the two beams acting separately. In other words, the energy is not destroyed, but merely redistributed in the interference pattern. The average intensity on the screen is exactly what would exist in the absence of interference. Thus, as shown in Fig. 1.12, the intensity in the interference pattern varies between $4A^2$ and zero. Now each beam, acting separately, would contribute A^2, and so, without interference, we would have a uniform intensity of $2A^2$, as indicated by the broken line. Let us obtain the average intensity on the screen for *n* fringes. We have

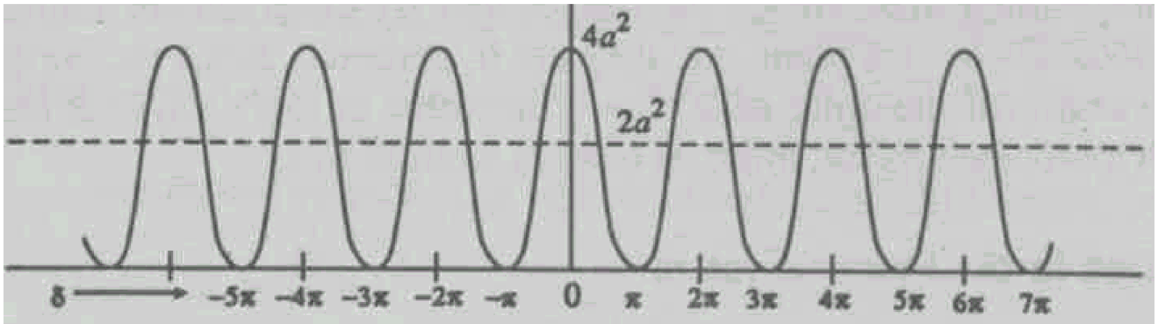


Fig. 1.12 Intensity distribution for the interference fringes from two waves of the same frequency

$$I_{average} = \frac{\int_0^\pi I d\delta}{\int_0^\pi d\delta}$$

$$= \frac{\int_0^\pi \left(4A^2 \cos^2 \frac{\delta}{2}\right) d\delta}{\int_0^\pi d\delta}$$

$$= \frac{\int_0^\pi (2A^2 + 2A^2 \cos\delta) d\delta}{\int_0^\pi d\delta} \qquad \left(\because\ 1 + \cos\theta = \cos^2 \frac{\theta}{2}\right)$$

$$= [2A^2\delta + 2A^2 \sin\delta]_0^\pi$$

$$= \frac{2A^2\pi}{\pi}$$

$$= 2A^2$$

Thus, the average intensity is equal to the sum of the separate intensities. That is whatever energy apparently disappears at the minima is actually present at the maxima. There is no violation of the law of conservation of energy in the phenomenon of interference.

Till now we have considered interference pattern produced when a monochromatic light from a narrow slit falls on two parallel slits. What happens if white light is used to illuminate slits? Read the following sub-section.

1.4.1 White-Light Fringes

If white light is used to illuminate the slits we obtain an interference pattern consisting of a central 'white' fringe, having on both sides a few coloured fringes and then a general illumination.

A pair of white light coherent sources is equivalent to a number of pairs of monochromatic sources. Each monochromatic pair produces its own system of fringes with a different fringe-width β, since β depends on λ $\left(\beta = \frac{D\lambda}{d}\right)$.

At the centre of the pattern, the path difference between the interfering waves is zero. Therefore, the path difference is also zero for all wavelengths. Hence, all the different coloured waves of the white light produce a bright fringe at the centre. This superposition of the different colours makes the central fringe 'white'. This is the 'zero order fringe'.

As we move on either side of the centre, the path difference gradually increases from zero. At a certain point it becomes equal to half the wavelength of the component having the smallest wave-length, i.e., violet. This is the position of the first dark fringe of violet. Beyond this, we obtain the first minimum of blue, green, yellow and of red in the last. The inner edge of the first dark fringe, which is the first minimum for violet, receives sufficient intensity due to red, hence it is reddish. The outer edge of the first dark fringe, which is minimum for red, receives sufficient intensity due to violet, and is therefore, violet. The same applies to every other dark fringe. Hence, we obtain a few coloured fringes on both sides of the central fringe.

As we move further away from the centre, the path difference becomes quite large. Then, from the range 7500 - 4000 Å, a large number of wavelengths (colours) will produce maximum intensity at a given point, and an equally large number will produce minimum intensity at that point. For example, at any point *P*, we may have
path difference

$$\text{path difference} \begin{cases} 11\lambda_1 = 12\lambda_2 = 13\lambda_3 = \ldots, \text{etc, (maxima)} \\ = \left(11+\dfrac{1}{2}\right)\lambda_1' = \left(12+\dfrac{1}{2}\right)\lambda_2' = \left(13+\dfrac{1}{2}\right)\lambda_3', \ldots, \text{etc, (minima)} \end{cases}$$

Thus, at *P*, we shall have 11th, 12th, 13th, etc., bright fringes of λ_1, λ_2, λ_3, etc., and 11th, 12th, 13th, etc., dark fringes of λ_1', λ_2', λ_3', etc. Hence, the resultant colour at *P* is very nearly white. This happens at all points, for which the path difference is large. Hence, in the region of large path difference uniform white illumination is obtained.

For maxima, path difference = $n\lambda$, where n = 0, 1, 2, ….
For minima, path difference = $\left(n+\dfrac{1}{2}\right)\lambda$, where n = 0, 1, 2, …

SAQ 8

Let the path difference $S_1P - S_2P = 30 \times 10^{-5}$ cm What are the λ's for which the point *P* is a maximum?

In the usual interference pattern with a monochromatic source, a large number of interference fringes are obtained, and it becomes extremely difficult to determine the position of the central fringe. Hence, by using white light as a source the position of central fringe can be easily determined.

1.4.2 Displacement of Fringes

We will now discuss the change in the interference pattern produced when a thin transparent plate, say of glass or mica, is introduced in the path of one of the two interfering beams, as shown in Fig. 1.13. It is observed that the entire fringe-pattern is displaced to a point towards the beam in the path of which the plate is introduced. If the displacement is measured, the thickness of the plate can be obtained provided the refractive index of the plate and the wavelength of the light are known.

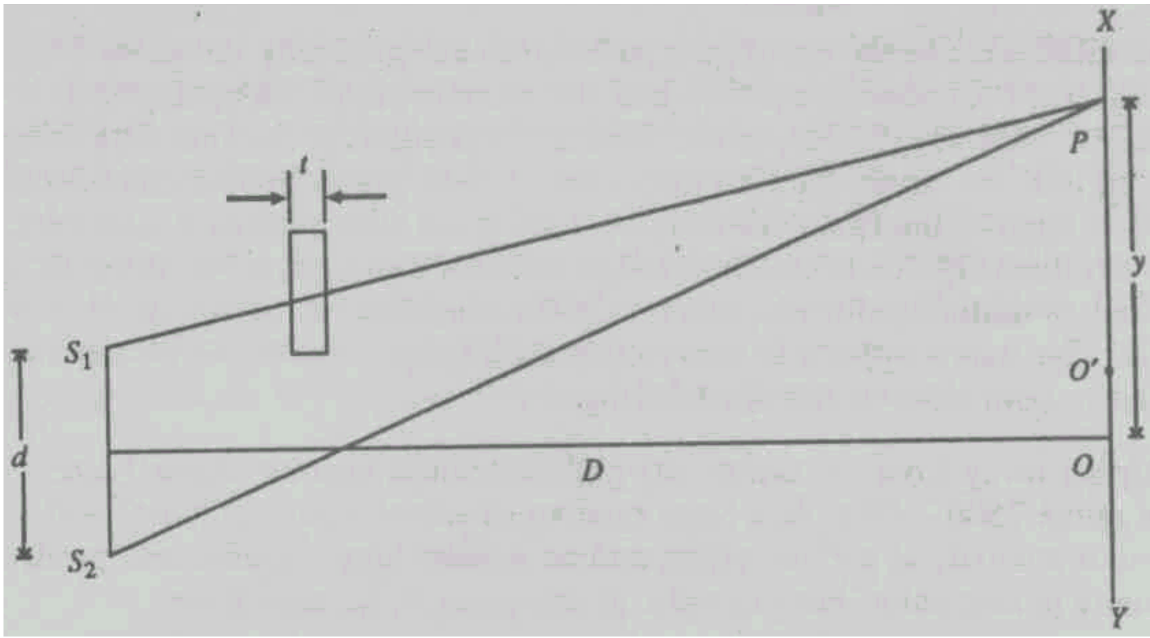


Fig. 1.13: If a thin transparent sheet (of thickness t) is introduced In one of the beams, the fringe pattern gets shifted by a distance $(\mu-1)tD/d$

Suppose a thin transparent plate of thickness t and refractive index μ is introduced in the path of one of the constituent interfering beams of light (say in the path of S_1P, shown in Fig. 1.13). Now, light from S_1 travels partly in air and partly in the plate. For the light path from S_1 to P, the distance travelled in air is (S_1P-t), and that in the plate is t. Let c and v be the velocities of light in the air and in the plate, respectively. If the time taken by light beam to reach from S_1 to P is, T, then

$$T=\frac{S_1P-t}{c}+\frac{t}{c}$$

or,

$$T=\frac{S_1P-t}{c}+\frac{\mu t}{c} \qquad \left(\because\ v=\frac{c}{\mu}\right)$$

$$=\frac{S_1P+(\mu-1)t}{c}$$

Thus the effective path in air from S_1 to P is $[S_1P+(\mu-1)t]$, i.e., the air path S_1P is increased by an amount $(\mu-1)t$, due to the introduction of the plate of material of refractive index, μ.

Let O be the position of the central bright fringe in the absence of the plate, the optical paths S_1O and S_2O being equal. On introducing the plate, the two optical paths become unequal. Therefore, the central fringe is shifted to O' such that at O' the two optical paths become equal. A similar argument applies to all the fringes. Now, at any point *P*, the effective path difference is given by

$$S_2P-[S_1P+(\mu-1)t] = (S_2P-S_1P)-(\mu-1)t$$

From Eq. (1.20), $S_2P-S_1P=\frac{d}{D}y$

$\therefore$ Effective path difference at $P=\frac{d}{D}y-(\mu-1)t$

If the point *P* is to be the centre of the nth bright fringe, the effective path difference should be equal to $n\lambda$, i.e.,

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

or

$$\boxed{y_n=\frac{D}{d}[n\lambda+(\mu-1)t]} \qquad (1.24)$$

In the absence of the plate (*t* = 0), the distance of the nth bright fringe from O is $\frac{D}{d}n\lambda$.

$\therefore$ Displacement y_0 of the nth bright fringe is given by

$$y_0=\frac{D}{d}[n\lambda+(\mu-1)t]-\frac{D}{d}n\lambda$$

$$\boxed{y_0=\frac{D}{d}(\mu-1)t} \qquad (1.25)$$

The shift is independent of the order of the fringe, showing that shift is the same for all the bright fringes. Similarly, it can be shown that the displacement of any dark fringe is also given by Eq. (1.25). Thus, the entire fringe-system is displaced through a distance $\frac{D}{d}(\mu-1)t$ towards the side on which the plate is placed. The fringe-width is given by

$$\beta = y_{n+1}-y_n$$

$$= \frac{D}{d}[(n+1)\lambda + (\mu - 1)t] - \frac{D}{d}[n\lambda + (\mu - 1)t] \quad \text{(see Eq. (1.24))}$$

which is the same as before the introduction of the plate.

Eq. (1.25) enables us to determine the thickness of extremely thin transparent sheets (like that of mica) by measuring the shift of the fringe system.

Now, apply this strategy yourself to SAQ 9.

SAQ 9
In a double slit interference arrangement one of the slits is covered by a thin mica sheet whose refractive index is 1.58. The distances S_1S_2 and AO (see Fig. 1.13) are 0.1 cm and 50 cm, respectively. Due to the introduction of the mica sheet, the central fringe gets shifted by 0.2 cm. Determine the thickness of the mica sheet.

1.5 FRESNEL'S BIPRISM

With regard to Young's double-slit experiment, objection was raised that the bright fringes observed by Young were probably due to some complicated modification of the light by the edges of the slits and not due to interference. Soon after, Fresnel devised a series of arrangements to produce the interference of two beams of light which was not subject to this criticism. One of the experimental arrangements, known as Fresnel's Biprism arrangement, is shown in Fig. 1.14.

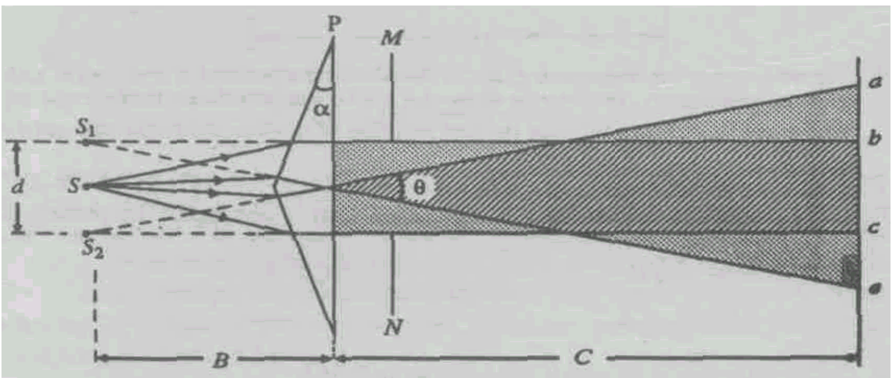


Fig. 1.14: Diagram of Fresnel's Biprism experiment

S is a narrow vertical slit illuminated by monochromatic light. The light from S is allowed to fall symmetrically on the Biprism *P*, placed at a small distance from *S* and having its refracting edges parallel to the slit. The light emerging from the upper and lower halves of the prism appears to start from two virtual images, S_1 and S_2 of S, which act as coherent sources. The cones of light bS_1e and aS_2c, diverging from S_1 and S_2, are superposed and the interference fringes are formed in the overlapping region *be*.

If screens *M* and *N* are placed as shown in the Fig. 1.14, interference fringes are observed only in the region *be*. When the screen *ae* is replaced by a photographic plate, a picture like the upper one, in Fig. 1.15, is obtained.

The closely spaced fringes in the centre of the photograph are due to interference, while the wide fringes at the edge of the photograph are due to diffraction. These wider bands are due to the vertices of the two prisms, each of which acts as a straight edge, giving a pattern of diffraction (about this you will learn in Optics III). When the screens *M* and *N* are removed from the light path, the two beams overlap over the whole region *ae*. The lower photograph in Fig. 5.15 shows for this case the equally spaced interference fringes superimposed on the diffraction pattern, of a wide aperture.

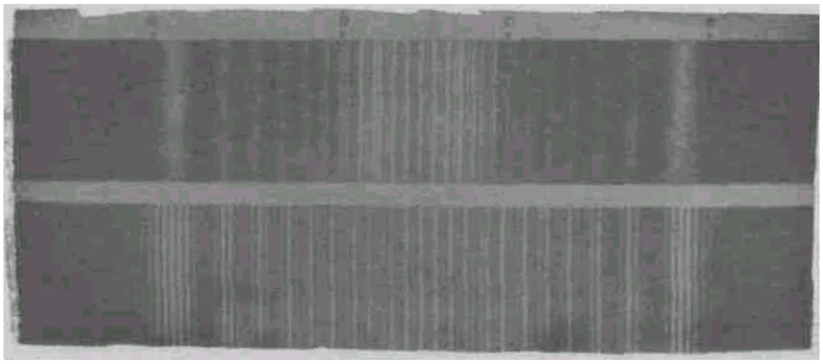


Fig. 1.15 Interference and diffraction fringes produced in the Fresnel Biprism experimental arrangement

With such an experiment, Fresnel was able to show the interference effect without the diffracted beams through the two slits. Just as in Young's double slit experiment, this arrangement can also be used to determine the wavelength of monochromatic light. The light illuminates the slit *S* and interference fringes can be easily viewed through the eyepiece. The fringe-width β can be determined by means of a micrometer attached to the eyepiece. If *D* is the distance between source and screen, and *d* the distance between the virtual images S_1 and S_2, the wave-length is given by

$$\lambda = \frac{\beta d}{D} \tag{5.26}$$

The distances *d* and *D* can easily be determined by placing a convex lens between the Biprism and the eyepiece. For a fixed position of the eyepiece, there will be two positions of the lens, shown as L_1 and L_2 in Fig. 1.16 where the images of S_1 and S_2 be seen at the eyepiece. Let d_1 be the distance between the two images, when the lens is at the position L_1 (at a distance b_1 from the eyepiece). Let d_2 and b_2 be the corresponding distances, when the lens is at L_2. Then it can easily be shown that

$$d = \sqrt{d_1 d_2} \tag{1.27a}$$

and

$$D = b_1 + b_2 \tag{1.27}$$

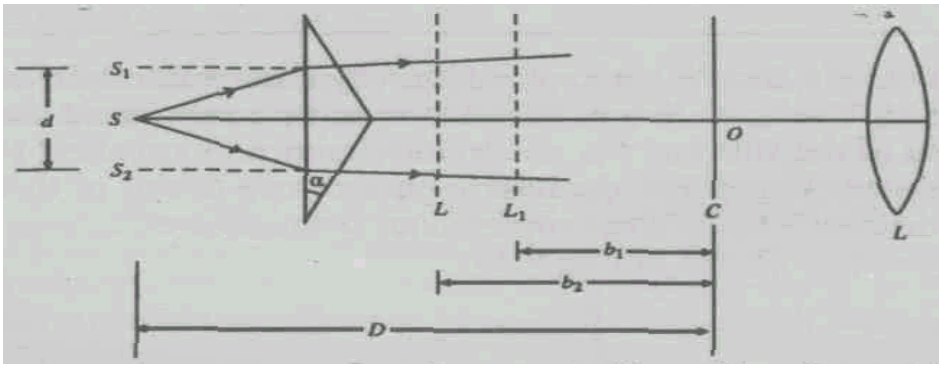


Fig. 1.16 Fresnel's biprism arrangement. *C* and *L* represent the position of cross wires and the eyepiece, respectively. In order to determine *d* a lens is introduced between the biprism and cross wires. L_1 and L_2 represent the two positions of the lens where the slits are clearly seen.

Use Eq. (1.26) and (1.27) to solve the following SAQ.

SAQ 10
In a Fresnel's Biprism experiment, the eyepiece is at a distance of 100 cm from the slit. A convex lens inserted between the Biprism and the eyepiece 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 case 2.10 mm apart. If sodium light of wavelength 5893 Å is used, find the thickness of the interference fringes.

1.6 SOME OTHER ARRANGEMENT FOR PRODUCING INTERFERENCE BY DIVISION OF WAVEFRONT

Two beams may be brought together in several other ways to produce interference. In **Fresnel's two-mirror** arrangement, light from a slit is reflected in two plane mirrors slightly inclined to each other. The mirror produces two virtual images of the slit, as shown in Fig. 1.17.

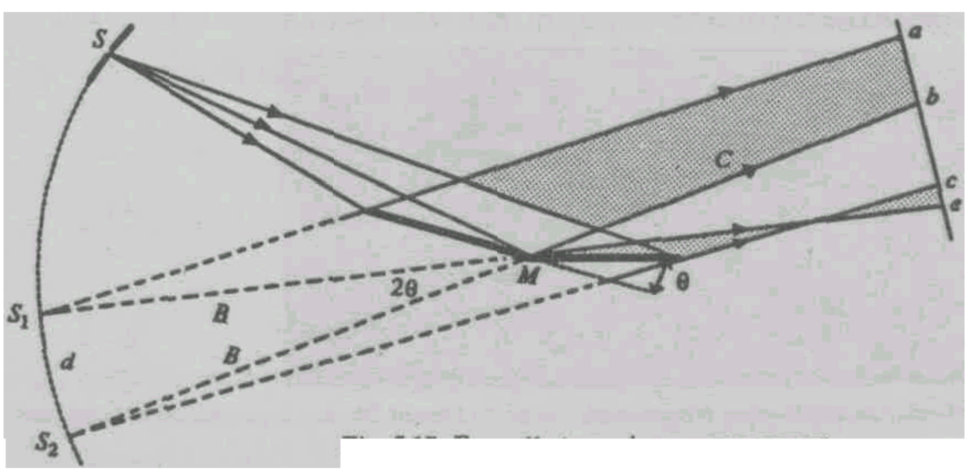


Fig. 1.17 Fresnel's two mirror arrangement.

They are like the images in Fresnel's biprism, and interference fringes are observed in the region *bc*, where the reflected beams overlap.

At grazing incidence, almost the entire incident light is reflected so that the direct and the reflected beam have nearly equal amplitudes. Hence the fringes have good contrast.

Even a simpler mirror method is available. This is known as **Lloyd's mirror**. Here the slit and its virtual image constitute the double source.

Lloyd's Mirror

It is a simple arrangement to obtain two coherent sources of light to produce a stationary interference pattern. It consists of a plane mirror *MN* (Fig. 1.18) polished on the front surface and blackened at the back (to avoid multiple reflection). S_1 is a narrow slit, illuminated by monochromatic light, and placed with its length parallel to the surface of the mirror. Light from S_1 falls on the mirror at nearly grazing incidence, and the reflected beam appears to diverge from S_2, which is the virtual image of S_1. Thus, S_1 and S_2 act as coherent sources. The direct cone of light AS_1E and the reflected cone of light BS_2C are superposed, and the interference fringes are obtained in the overlapping region *BC* on the screen.

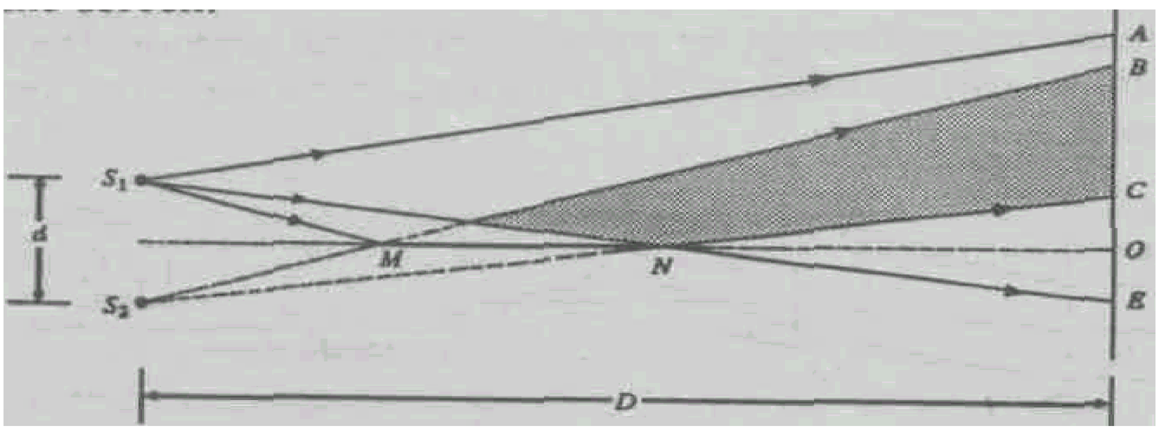


Fig. 5.18 Lloyd's mirror

Zero-Order Fringe

The central zero-order fringe, which is expected to lie at *0* (the perpendicular bisector of S_1S_2) is not usually seen since only the direct light, and not the reflected light, reaches *O*. It can be seen by introducing a thin sheet of mica in the path of light from S_1, when the entire fringe system is displaced in the upward direction. (You could see this yourself while solving SAQ 11.)

SAQ 11

Interference bands are obtained with a Lloyd's mirror with light of wavelength 5.45×10^{-5} cm. A thin plate of glass of refractive index 1.5 is then placed normally in the path of one of the interfering beams. The central dark band is found to move into the position previously occupied by the third dark band from the centre. Calculate the thickness of the glass plate.

With white light the central fringe is expected to be white, but actually it is found to be 'dark'. This is because the light suffers a phase change of π or a path-difference of $\frac{\lambda}{2}$ when reflected from the mirror. Therefore, the path difference between the interfering rays at the position of zero-order fringe becomes $\frac{\lambda}{2}$ (instead of zero), which is a condition for a minimum. Hence the fringe is dark.

Determination of Wavelength

Let *d* be the distance between the coherent sources S_1 and S_2, and *D* the distance of the screen from the sources. The fringe-width is then given by

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

Thus, knowing β, *D* and *d*, the- wavelength λ can be determined.

Acromatic Fringes and their Production by Lloyd's Mirror

A system of white and dark fringes, without any colours, obtained by white light are known as 'achromatic fringes'.

Ordinarily, with white light, we obtain a central white fringe, having on either side of it a few coloured fringes (as you have studied in subsection 1.4.1). This is because the fringe-width $\beta = \frac{D\lambda}{d}$ is different for different wavelengths (colours). If however, the fringe-width is made the same for all wavelengths, the maxima of each order for all wavelengths will coincide, resulting into achromatic fringes. That is, for achromatic fringes, we must have

$$\frac{D\lambda}{d} = \text{constant}$$

or $$\frac{\lambda}{d} = \text{constant}$$

We can easily realise this condition with a Lloyd's mirror by using a slit illuminated by a narrow spectrum of the white light as shown in Fig. 1.19. The narrow spectrum R_1V_1 is produced by a prism, or, preferably, by a plane diffraction grating. The Lloyd's mirror is placed with its surface close to the violet end of the spectrum and such that R_1V_1 is perpendicular to its plane.

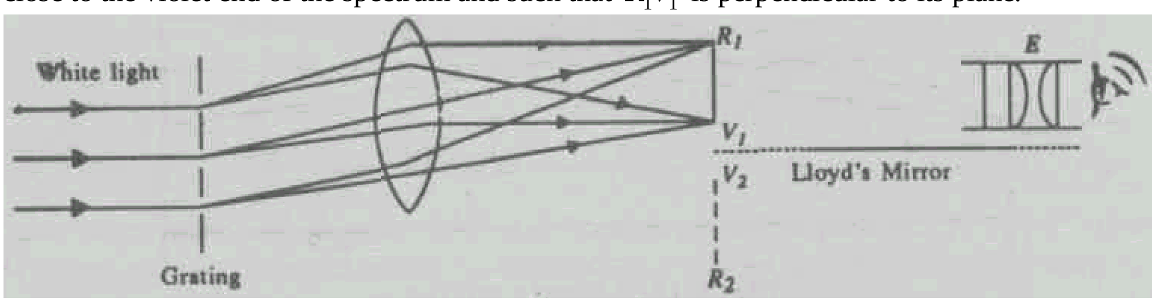


Fig. 1.19 Achromatic fringes produced by Lloyd's mirror

R_1V_1, and its virtual image, R_2V_2, formed by the mirror act as coherent sources. They are equivalent to a number of pairs of sources of different colours. Thus, the pair R_1R_2 produces a set of red fringes, and the pair V_1V_2 a set of violet fringes. The intermediate pairs produce the sets of fringes of intermediate colours. The red and violet fringes will be of the same width if

$$\frac{\lambda}{d} = \text{constant}$$

i.e., $$\frac{\lambda_r}{d_r} = \frac{\lambda_v}{d_v}$$

or $$\frac{d_r}{d_v} = \frac{\lambda_r}{\lambda_v}$$

where d_r is the distance R_1R_2, and d_v the distance V_1V_2.

Hence, the last expression gives

$$\frac{R_1R_2}{V_1V_2} = \frac{\lambda_r}{\lambda_v}$$

Therefore, if the distance of the violet end V_1 from the surface of the mirror is so adjusted by displacing the mirror laterally that the above condition is satisfied, the red and violet fringes will have the same width, and will exactly be superposed on each other. Since, in a grating spectrum, the dispersion is accurately proportional to the wavelength, the condition (λ / d) = constant is simultaneously satisfied for all the wavelengths. Thus, when this adjustment is made, fringes of all colours are superposed on one another. Hence, achromatic fringes are observed in the eyepiece *E* placed in the over-lapping region.

Difference between Biprism and Lloyd's Mirror Fringes

The following are the main points of difference between the biprism and Lloyd's mirror fringes.

(1) In biprism, the complete pattern of fringes is obtained. In Lloyd's mirror, ordinarily, only a few fringes on one side of the central fringe are visible, the central fringe itself being invisible.

(2) In biprism the central fringe is bright, while in Lloyd's mirror it is dark.

(3) The central fringe in biprism is less sharp than that in Lloyd's mirror.

The coherent sources in the biprism are A_1B_1 and A_2B_2 (Fig. 1.20a) the virtual images of a slit *AB*. In Lloyd's mirror, the coherent sources are a slit A_1B_1 itself and its virtual image B_2A_2 (Fig. 1.20 b). In both cases, A_1 and A_2 form one extreme pair of coherent point-sources, and B_1 and B_2 another extreme pair. In the biprism, the zero-order fringes corresponding to A_1B_1 and A_2B_2

are formed at A_0 and B_0, which lie on the right bisectors of A_1A_2 and B_1B_2, respectively. Hence, the zero-order fringe extends from A_0 to B_0. In Lloyd's mirror, on the other hand, all pair of coherent sources have a common perpendicular bisector, so that zero-order fringes due to all of these are formed in one and the same position. Hence the zero-order fringe is sharp in this case.

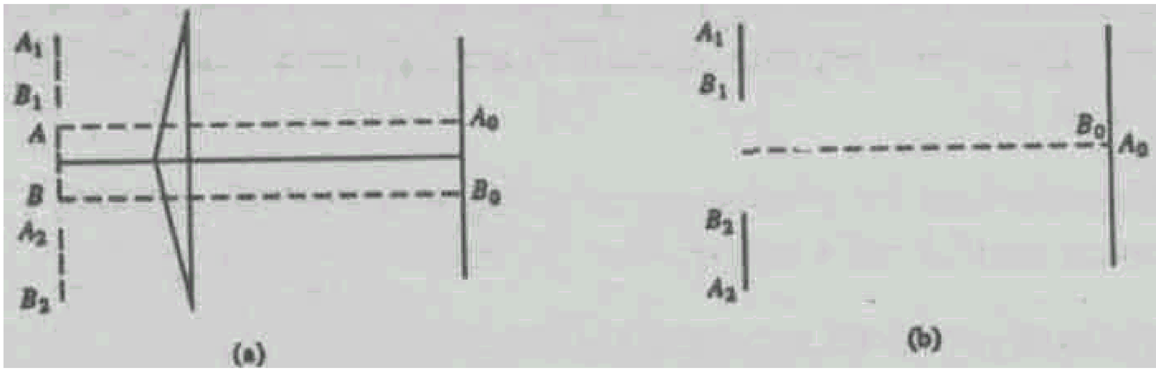


Fig. 1.20 Showing the difference between biprism and Lloyd's mirror fringes

(4) In biprism $A_1A_2 = B_1B_2 = d$. Hence, the fringe-width $\beta = \dfrac{D\lambda}{d}$ is the same for all pairs of coherent sources. In Lloyd's mirror arrangement d is different for different pairs of coherent sources, e.g., $A_1A_2 > B_1B_2$. Hence, the fringe-width is different for different pairs of coherent sources.

1.6 SUMMARY

- The relationship between phase difference and path difference is:

$$\text{phase difference} = \frac{2\pi}{\lambda}(\text{path difference})$$

- If two waves of same frequency and of amplitudes a_1 and a_2 and phase difference δ are superposed then, according to principle of superposition, the amplitude A of the resultant wave is given by

$$A^2 = a_1^2 + a_2^2 + 2a_1a_2\cos\delta$$

- Two sources are said to be coherent if they emit light waves with constant or no phase difference.
- When two waves of the same frequency travel in approximately the same direction and have a phase difference that remains constant with time, the resultant intensity of light is not distributed uniformly in space. This non-uniform distribution of the light intensity is due to the phenomenon of interference.

- For constructive interference

path difference = $n\lambda$, where n = 0, 1, 2, ...

and for destructive interference

path difference = $m\frac{\lambda}{2}$, where m = 1, 3, 5, 7, …

- In an interference pattern, the distance between any two consecutive maxima or minima is given by

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

  where β is called the fringe-width, A is the wavelength of light used, d is the distance between the two coherent sources, and D is the distance between the sources and the screen.

- When a thin transparent plate of thickness t and refractive index μ is introduced in the path of one of the constituent interfering beams of light, the entire fringe system is displaced through a distance $\frac{D}{d}(\mu - 1)t$.

- Just as in Young's double slit experiment, the wavelength of light can be determined from measurement of fringe-width produced by the biprism by the following relation:

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

  where $d = \sqrt{d_1 d_2}$ and $D = b_1 + b_2$

  d_i is the distance between the two images, when the lens is at the position L_1 at a distance b_1 from the eyepiece. d_2 and b_2 are the corresponding distances when the lens is at L_2.

- Some other devices for producing coherent sources are: Fresnel's two-mirror arrangement and Lloyd's mirror.

- Lloyd's mirror produces achromatic fringes.

5.7 TERMINAL QUESTIONS

(1) Young's experiments is performed with light of the green mercury line. If the fringes are measured with a micrometer eyepiece 80 cm behind the double slit, it is found that 20 of them occupy a distance of 10.92 mm. Find the distance between two slits. Given that the wavelength of green mercury line is 5460 Å.

(2) In a certain Young's experiment, the slits are 0.2 mm apart. An interference pattern is observed on a screen 0.5m away. The wavelength of light is 5000 Å. Calculate the distance between the central maxima and the third minima on the screen.

(3) A Lloyd's mirror, of length 5 cm, is illuminated with monochromatic light (λ = 5460 Å) from a narrow slit 0.1 cm from its plane, and 5 cm, measured in that plane, from its near edge. Find the separation of the fringes at a distance of 120 cm from the slit, and the total width of the pattern observed.

5.8 SOLUTIONS AND ANSWERS

SAQs

(1) The distance OM is given by $a\cos\theta$. Hence the equation is $x = a\cos\theta$ or $x = a\cos\omega t$.

(2) $y = a\sin\omega t = a\sin\dfrac{2\pi}{T}t$

If we replace π by 180°, and put a = 6 cm = 0.06 m, and T = 3 s, we get

$$y = (0.06)\sin\frac{2\times180^0}{3}t$$

(a) Thus displacement after 0.5 sec is,

$$\begin{aligned} y &= 0.06\sin\frac{2\times180^0}{3}t \\ &= 0.06\sin 60^0 \\ &= 0.052 \text{ m} \end{aligned}$$

(b) velocity v

$$\begin{aligned} v &= a\omega\cos\omega t \\ &= a\frac{2\pi}{T}\cos\frac{2\pi}{T}t \\ &= 0.06\times\frac{2\pi}{T}\cos\frac{2\times180^0}{3}\times0.5 \\ &= 0.06\times\frac{2\pi}{T}\times\cos 60^0 \\ &= 0.063 \text{ m/s} \end{aligned}$$

(c) acceleration = $\omega^2 y$

$$\begin{aligned} &= \left(\frac{2\pi}{T}\right)^2 a\sin\omega t \\ &= \left(\frac{2\pi}{T}\right)^2\times0.06\times\sin\frac{2\pi}{T}t \\ &= \left(\frac{2\pi}{3}\right)^2\times0.06\times\sin\frac{2\times180^0}{3}\times0.5 \\ &= 0.228\ ms^{-2} \end{aligned}$$

(3) We have

$$\frac{I_{max}}{I_{min}} = \frac{I_1 + I_2 + 2\sqrt{I_1 I_2}}{I_1 + I_2 - 2\sqrt{I_1 I_2}} = \frac{\left(\sqrt{I_1} + \sqrt{I_2}\right)^2}{\left(\sqrt{I_1} - \sqrt{I_2}\right)^2}$$

Now, $\frac{I_1}{I_2} = \frac{81}{1}$ or $\frac{\sqrt{I_1}}{\sqrt{I_2}} = \frac{9}{1}$ or $\sqrt{I_1} = 9\sqrt{I_2}$

Hence, $I_{max} / I_{min} = \frac{\left(9\sqrt{I_2} + \sqrt{I_2}\right)^2}{\left(9\sqrt{I_2} - \sqrt{I_2}\right)^2} = \frac{(10)^2 I_2}{(8)^2 I_2} = \frac{100}{64} = \frac{25}{16}$

(4) The phase difference is related to the path difference by Eq. (5.13) as follows

$$\begin{aligned}\text{phase difference} &= \frac{2\pi}{\lambda}(AC - BC) \\ &= \frac{2\pi}{\lambda}(4-3)\lambda \\ &= 2\pi\end{aligned}$$

This is the condition of maximum intensity. So the waves interfere, constructively, in Fig. 1.7(a).

In the case of Fig. 1.7(b)

$$\begin{aligned}\text{phase difference} &= \frac{2\pi}{\lambda}(AC - BC) \\ &= \frac{2\pi}{\lambda}(3-2.5)\lambda \\ &= \pi\end{aligned}$$

This is the condition of minimum intensity.

Here the waves are completely out of phase and destructive interference occurs.

(5) Given: d = 0.40 mm, D = 10^3 mm, y = 3.6 mm, and m = 3. Using Eq. (1.21), we get

$$\lambda = \frac{yd}{mD} = \frac{(3.6)(0.40)}{3\times 10^3} = 4.8\times 10^{-4}\text{ mm} = 4.8\times 10^{-5}\text{ cm}$$

Hence, the light is in the blue-green region of the visible spectrum.

6. With λ = 6000 Å, the distance between zero-order and tenth order fringe is 14.73 mm – 12.34 mm = 2.39mm, so that the fringe width is 2.39 mm/10 = 0.239 mm.

7. $\beta = \frac{D\lambda}{d}$. Therefore

$$\frac{(\beta)_{6000}}{(\beta)_{5000}} = \frac{5000}{5000} = \frac{6}{5}$$

$\therefore \quad (\beta)_{5000} = \frac{5}{6} \times (\beta)_{6000} = \frac{5}{6} \times 0.239 = 0.199$ mm

Thus, with $\lambda = 5000$ Å, the zero-order fringe will still be at 12.34 mm, while the twentieth order fringe will be at

$$12.34 \text{ mm} + (0.199 \text{ mm} \times 20) = 16.32 \text{ mm}$$

(8) For maxima, the path difference = $n\lambda$

or 30×10^{-5} cm

$\therefore \quad \lambda = \frac{30 \times 10^{-5}}{n}$ cm

where n = 1,2,3,4 ...

(9) $y_0 = 0.2$ cm; $d = 0.1$ cm; $D = 50$ cm

Hence $t = \frac{d\, y_0}{D(\mu - 1)} = \frac{0.1 \times 0.2}{50 \times 0.58}$

$= 6.7 \times 10^{-4}$ cm

(10) The fringe-width is given by

$$\beta = \frac{D\lambda}{2d}, \text{ where } d = \sqrt{d_1 \times d_2}$$

Here, $d_1 = 4.05$ mm = 0.405 cm and d_2 = 2.10 mm = 0.210 cm

$\therefore \quad d = \sqrt{0.405 \times 0.210} = 0.292$ cm

Also, D = 100 cm and λ = 5893 Å = 5893×10^{-8} cm.

$\therefore \quad \beta = \frac{100 \times 5893 \times 10^{-8}}{0.292} = 0.0202$ cm

(11) By introducing a glass plate of thickness t in one of the interfering beams, t cm of air (μ = 1) are replaced by t cm of glass (μ = 1.5). t cm of glass are optically equivalent to μt or 1.5 t cm of air. The, increase in the length of the path = $\mu t - t = 0.5\, t$. This produces a shift of 2 in the interference bands

$\therefore \quad 0.5t = 2\lambda = 2 \times 5.45 \times 10^{-5}$

and $\quad t = \frac{2 \times 5.45 \times 10^{-5}}{0.5} = 21.8 \times 10^{-5}$ cm

TQs

(1) The fringe width β in Young's experiment is $\beta = \lambda D / d$.

Since 20 fringes occupy a distance of 10.92 mm, the fringe width β is

$\beta = (10.92 / 20)$ mm = (10.92/20) mm « (10.92 $\times 10^{-3}$/20) m.

Also D = 80 cm = 0.8 m, and $\lambda = 5.460 \times 10^{-7}$ m

$$\therefore \quad d = \frac{5.460 \times 10^{-7} \times 0.8 \times 20}{10.92 \times 10^{-3}} \text{ m} = 0.7912 \times 10^{-4} \text{ m}$$

2. See Fig. (1.10). Suppose the required distance on the screen is y.

Here $d = 2 \times 10^{-4}$ m (slit separation)

$\lambda = 5 \times 10^{-7}$ m (wavelength)

$D = 5 \times 10^{-1}$ m (distance between slit to screen)

The minima is observed when the phase difference between the two waves is an odd multiple of π, i.e., when

$$\delta = \pi, 3\pi, 5\pi, 7\pi, \ldots$$

At the third minimum, $\delta = 5\pi$

From Eq. (1.13), path difference = $\dfrac{2\pi}{\lambda}\delta = \dfrac{2\pi}{\lambda}(5\pi)$

But from Fig. 1.10, the path difference between the waves arriving at P is $d \sin\theta$

Hence, $5\pi = \dfrac{2\pi}{\lambda}(d \sin\theta)$

$$\text{or} \quad \sin\theta = \frac{\lambda}{2\pi}\frac{1}{d}(5\pi) = \frac{5\pi \times 5 \times 10^{-7}}{2\pi \times 2 \times 10^{-4}}$$

$$= 6.25 \times 10^{-3}$$

From Fig. 1.10, the required distance on the screen $y = D\tan\theta$

$= D\sin\theta = 5 \times 10^{-1} \times 6.25 \times 10^{-3} \qquad \because \tan\theta \approx \sin\theta$

= 3.1 m

3) Let *MM´* (Fig. 1.21) the Lloyd's mirror be 5 cm long. The source S_1 is as shown in the figure. The interference pattern is observed in the region *AB*.

The fringe width β is given by $\beta = \dfrac{\lambda D}{d}$

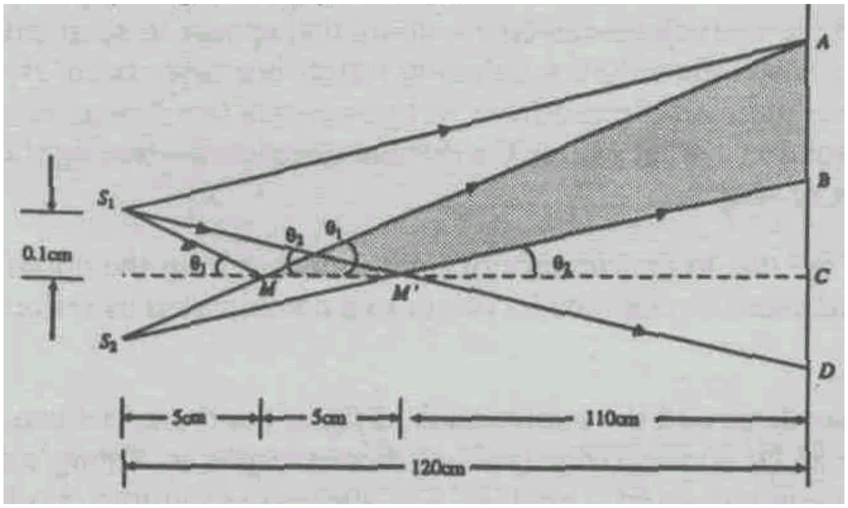


Fig. 1.21

Given λ = 5460 Å = 5.460 $\times 10^{-7}$ m; D = 120 cm = 1.20 m and d = 0.2 cm = 2 $\times 10^{-2}$ m

$$\beta = \frac{5.460 \times 10^{-7} \times 120}{2 \times 10^{-3}} \text{m} = 3.276 \times 10^{-4} \text{ m}$$

= 0.3276 mm.

The total width of interference pattern is obviously *AB*. From Fig. (1.21),

$$\tan\theta_1 = 0.1/5 \text{ and } \tan\theta_2 = 0.1/10$$

Also from right angled $\Delta BM'C$,

$BC / M'C = \tan\theta_2$ or $BC = 110 \times (0.1/10)$ = 1 cm

$\therefore$ $AB = AC - BC$ = 2.3 – 1.1 = 1.2 = 1.2×10^{-3} cm

UNIT 2 INTERFERENCE BY DIVISION OF AMPLITUDE

Structure

2.1 INTRODUCTION

We have all seen the marvellous rainbow colours that appear in soap bubbles and thin oil films. When a soapy plate drains, coloured reflections often occur from it. A similar effect occurs when light is reflected from a wet pavement that has an oil slick on it. Have you ever wondered what causes the display of colours when light is reflected from such thin oil film or soap bubble?

All these effects are due to interference of light reflected from the opposite surfaces of the film. Thus the phenomenon owes its origin to a combination of reflection and interference.

In the last unit, we discussed the interference of light, but there, the two interfering light waves were produced by division of wavefront. For example, in Young's double slit experiment, light coming out of a pin hole was allowed to fall into two holes, and the light waves emanating from these two holes interfered to produce the interference pattern. But the interference of light waves, which is responsible for the colour of thin films, involves two light beams derived from a single incident beam by division of amplitude of the incident wave. When a light wave falls on a thin film, the wave reflected from the upper surface interferes with the wave reflected from the lower surface. This gives rise to beautiful colours. However, one must initially consider how the phase of a light wave is affected when it is reflected.

In the last unit, you noted that in Lloyd's mirror, the interference takes place between waves coming direct from the source and those reflected from an optically denser medium. As a consequence of this, the central fringe is found to be 'dark' instead of 'bright'. This was explained by assuming the fact that a phase change of π takes place when light waves are reflected at the surface of a "denser" medium. We will begin this unit by giving proof of the statement made above; this proof will be based on the principle of reversibility of light.

It is also possible to observe interference using multiple beams. This is known as multiple beam interferometry, and it will be discussed in the next unit. It will be shown there that multiple beam interferometry offers some unique advantages over two beam interferometry.

Objectives

After studying this unit, you should be able to:

- prove that when a light wave is reflected at the surface of an optically denser medium, it suffers a phase change of *n*.

- describe the origin of the interference pattern produced by a thin film,

- describe the formation, shape and location of interference fringes obtained from a thin wedge-shaped film,
- describe how Newton's rings are used to determine the wavelength of light,
- explain why a thin coating of a suitable substance minimizes the reflection of light from a glass surface.
- distinguish between fringes of equal inclination and fringes of equal thickness.

2.2 STOKES' ANALYSIS OF PHASE CHANGE ON REFLECTION

To investigate the phase change in the reflection of light at an interface between two media, Sir G.C. Stokes used the principle of optical reversibility. This principle states that a light ray, that is reflected or refracted, will retrace its original path, if its direction is reversed, provided there is no absorption of light.

Fig. 2.1 (a) shows the surface *MN* separating media 1 and 2, the lower one being denser. Suppose medium 1 is air and medium 2 is glass.

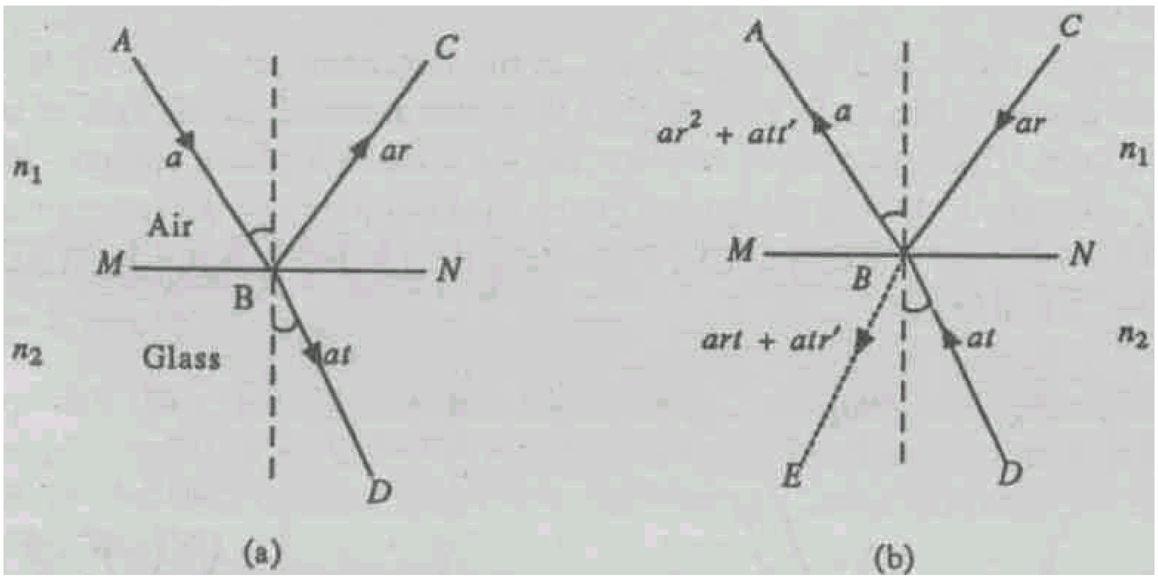


Fig. 2.1 (a) A ray is reflected and refracted at an air-glass interface, (b) The optically reversed situation; the two rays in the lower left must cancel. In both cases, $n_2 > n_1$ (n_1 and n_2 are the refractive indices of the media).

An incident light wave, *AB*, is partly reflected along *BC* and partly transmitted (refracted) along *BD*. Let a be the amplitude of the incident wave *AB*, *r* be the fraction of the amplitude reflected, and *t* be the fraction transmitted when the wave is travelling from medium 1 to 2. Then the amplitudes along *BC* and *BD* are *ar* and *at*, respectively.

Now, suppose the directions of the reflected and transmitted (refracted) waves are reversed. As shown in Fig. 2.1(b), the wave *BC*, on reversal, gives a reflected wave along *BA*, and a transmitted (refracted) wave along *BE*. The amplitude of the reflected wave along *BA* is $ar.r = ar^2$ and the amplitude of transmitted wave along *BE* is *art*. Similarly, the wave *BD*, on reversal, gives a transmitted wave along *BA* and a reflected beam along *BE*. Let r' and t' be the fractions of amplitude reflected and transmitted when the wave is travelling from medium 2 to medium 1. Then the amplitude of the transmitted wave along *BA* is att' and the amplitude of reflected wave along *BE* is atr'. But, according to the principle of reversibility of light, the reflected and transmitted waves *BC* and *BD*, when reversed, should give the original ray of amplitude *a* along *BA* only. Hence, the component along *BE* should be zero and that along *BA* should be equal to *a*. That is,

$$art + atr' = 0 \tag{2.1}$$

and

$$ar^2 + att' = a \tag{2.2}$$

From Eqs. (2.1) and (2.2), we get

$$r' = -r \tag{2.3}$$

and

$$tt' = 1 - r^2 \tag{2.4}$$

Eqs. (6.3) and (6.4) are known as **Stake's relations**.

You must be aware that a transverse wave in a spring undergoes a 180° phase change when reflected from a rigid support. A similar phase change occurs for the reflection of a light wave from the boundary of a medium having a greater index of refraction. The optically denser medium corresponds to a rigid support. A light wave reflected from the boundary of a medium whose index of refraction is greater than that of the medium in which the incident wave travels undergoes a 180^0 phase change.

Now, observe carefully Eq. (2.3). Here *r* is the fraction of amplitude reflected when the incident wave is travelling from a rarer to denser medium, and *r′* when incident wave is travelling from a denser to a rarer medium. The two fractions are numerically equal but have opposite signs. Hence, they are exactly out of phase with each other, i.e., their phase difference is 'π'. If no phase change occurs when a light wave is reflected by a denser medium then there must be a phase change of π when a light wave is reflected by a rarer medium—and conversely, if no phase change occurs when a light wave is reflected by a rarer medium then there must be a phase change of π when a light wave is reflected by a denser medium. Now, out of the two alternatives mentioned above, the second one is correct because it has been experimentally observed (See Section 1.6 in connection with Lloyd's mirror) that the phase change of π occurs when the light strikes the boundary from the side of rarer medium. Hence, light reflected by a material of higher refractive index than the medium in which the rays are travelling undergoes a 180° (or π) phase change.

Reflection by a material of lower refractive index than the medium in which the rays are travelling causes no phase change.

The following SAQ will provide a useful check of your understanding of this section.

SAQ 1
In Fig. 2.2, we have illustrated four situations. In the two examples on the left, the refractive index between the surfaces is higher than that outside; in the two examples on the right, it is lower. This determines whether or not there is a phase change. In Fig. 2.2(a) and (b), we have indicated the phase change taking place at the points marked by an arrow. Redraw the Fig. 2.2(c) and (d), indicating the phase change taking place at the points marked by an arrow.

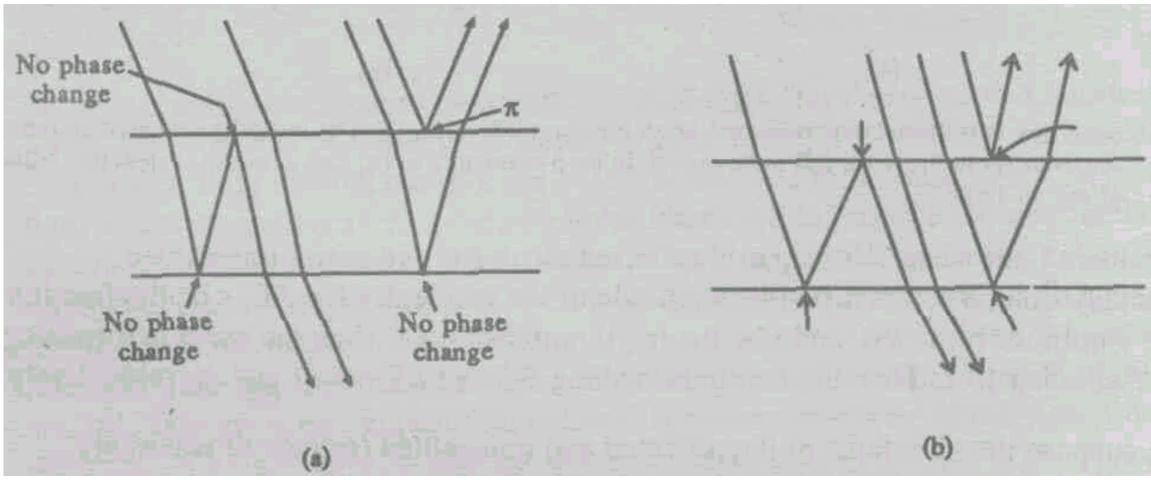


Fig. 2.2

2.3 INTERFERENCE IN THIN FILMS

Suppose a ray of light from a source *S* strikes a thin film of soapy water, at *A*, see Fig. 2.3(a). Part of this will be reflected as ray (1) and part refracted in the direction *AB.* Upon arrival at *B*, part of the latter will be reflected to *C,* and part refracted along BT_1. At *C,* the ray will again get partly reflected along *CD* and refracted as ray (2) along CR_2. A continuation of this process yields two sets of parallel rays, one on each side of the film. In each of these sets, of course, the amplitude decreases rapidly from one ray to the next. Considering only the first two reflected rays (1) and (2) we find that these two rays are in a position to interfere. This is because, if we assume *S* to be a monochromatic point source, the film serves as an amplitude-splitting device, so that ray (1) and (2) may be considered as arising from two coherent virtual sources S' and S'' lying behind the film, that is, the two images of *S* formed by reflection at the top and bottom surfaces of the film, as shown in Fig. 2.3 (b). If the set of parallel reflected rays is now collected by a lens, and focussed at *P,* each ray has travelled a different distance, and the phase relationship between them may be such as to produce destructive or constructive interference at *P.* It is such interference that produces the colours of this film when seen by naked eyes.

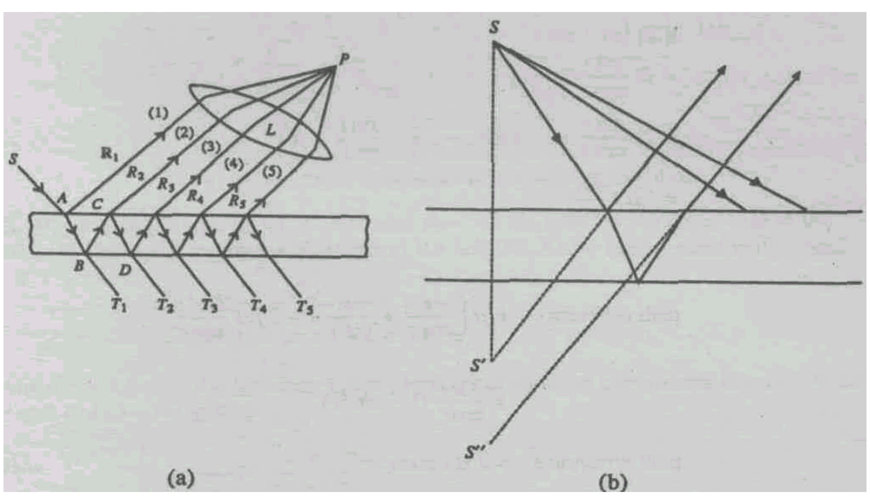


Fig. 2.3: (a) Multiple reflection in a soap film, (b) The interference pattern produced due to rays (1) and (2) is approximately the same as would have been produced by two coherent point sources S' and S''

Now, we know that the two rays reinforce each other, if the path difference between them is an integral multiple of λ, where λ is the wavelength of light, which is being used to illuminate the film. Hence, let us first find out the path difference between the reflected rays (1) and (2).

Path Difference in Reflected Light

Let the ray of light falling on the thin film of soapy water at *A* be incident at an angle *i,* as shown in Fig. 2.4. Let the thickness of the film be *t* and refractive index be $\mu(>1)$. At A it is partly reflected along AR_1 giving the ray (1) and partly refracted along *AB* at an angle *r.* At *B* it is again partly reflected along *BC* and partly refracted along BT_1, Similar reflections and refractions occur at *C.* Since, the rays AR_1 and CR_2, i.e. ray (1) and ray (2) have been derived from the same incident ray, they are coherent and in a position to interfere. Let *CN* and *BM* be perpendiculars to AR_1 and *AC.* As the paths of the rays AR_1 and CR_2 beyond *CN* are equal, the path difference between ray (1) and (2) is given by

(path *ABC* in film-path AN in air)

$$\therefore \text{ path difference} = \mu((AB + BC) - AN \tag{2.5}$$

Here $AB = BC = \dfrac{BM}{\cos r} = \dfrac{t}{\cos r}$.

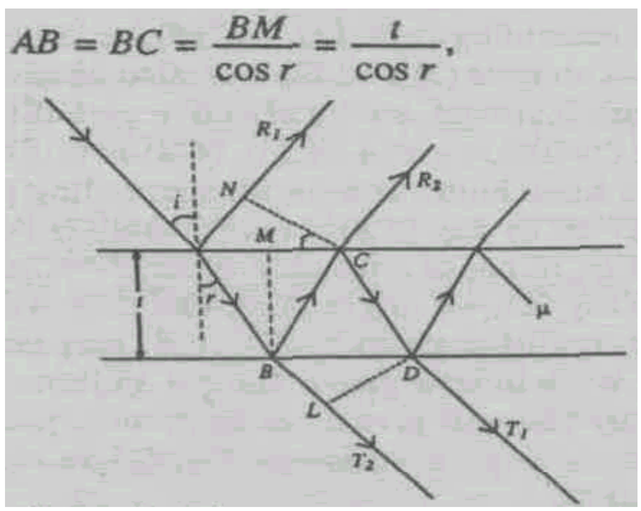


Fig. 2.4 Optical path difference between two consecutive rays in a multiple reflection

and $AN = AC \sin i$

Now, $AC = AM + MC$

$= BM \tan r + BM \tan r$

$= 2t \tan r$

$\therefore \quad AN = 2t \tan r \sin i$

$$= 2t\frac{\sin r}{\cos r}(\sin i)$$

$$= 2t\frac{\sin r}{\cos r}(\mu \sin r) \qquad \left[\because \frac{\sin i}{\sin r} = \mu\right]$$

$$= 2\mu t\frac{\sin^2 r}{\cos r}$$

Substituting these values of *AB*, *BC* and *AN* in Eq. (2.5) we get,

$$\text{path difference} = \mu\left(\frac{t}{\cos r} + \frac{t}{\cos r}\right) - 2\mu t\frac{\sin^2 r}{\cos r}$$

$$= \frac{2\mu t}{\cos r}(1 - \sin^2 r)$$

$\therefore$ path difference = $2\mu t \cos r$ (2.6)

At *A*, the ray is reflected while going from a rarer to a denser medium and suffers a phase change of π. At *B*, the reflection takes place when the ray is going from a denser to a rarer medium, and there is no phase change.

However, we must take account of the fact that ray (1) undergoes a phase change of π at reflection while ray (2) does not, since it is internally reflected (See SAQ 1). The phase change of π is equivalent to a path difference of $\frac{\lambda}{2}$. Hence, the effective path difference of between ray (1) and rays (2) is

$$2\mu t \cos r - \frac{\lambda}{2} \tag{2.7}$$

The sign of the phase change is immaterial. Here we have chosen the negative sign to make the equation a bit simpler in form.

As you know from Unit 1, if this path difference is an odd multiple of $\frac{\lambda}{2}$, we might expect rays (1) and (2) to be out of phase, and produce a minimum of intensity. Thus the condition

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

or $$2\mu t \cos r = n\lambda \tag{2.8}$$

becomes a condition for destructive interference as far as rays (1) and (2) are concerned.

Next, we examine the phases of the remaining rays, (3), (4), (5), ... Since the geometry is the same, the path difference between rays (3) and (2) will also be given by Eq. (2.6). But, here, only internal reflections are involved, so the effective path difference will still be given by Eq. (2.6). Hence, if the condition given by Eq. (2.8) is fulfilled, ray (3) will be in the same phase as ray (2). The same holds true for all succeeding pairs, and so we conclude that, under the condition given by Eq. (2.8), rays (1) and (2) will be out of phase, but rays (2), (3), (4),....., will be in phase with each other. Now, since ray (1) has considerably greater amplitude than ray (2), we might think that they will not completely annul each other, that is, the condition given by Eq. (6.8) may not produce complete darkness. But it is not so. We will now prove that the addition of rays (3), (4), (5), ... which are all in phase with ray (2), will give a net amplitude, just sufficient to make up the difference and to produce complete darkness. Fig. 2.5 shows the amplitude of successive rays in multiple reflection.

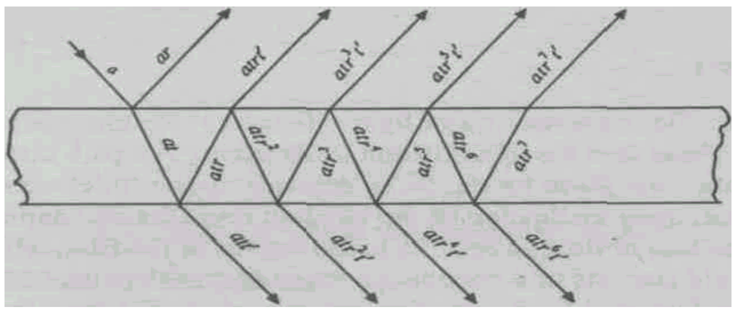


Fig. 2.5 Amplitude of successive rays in multiple reflection

Adding the amplitudes of all the reflected rays but the first, on the upper side of the film we obtain the resultant amplitude:

$$\begin{aligned} A &= atrt' + atr^3t' + atr^5t' + atr^7t' + ... \\ &= atrt'(1 + r^2 + r^4 + r^6 + ...) \end{aligned}$$

Since r is, necessarily, less than 1, the geometrical series in parentheses has a finite sum equal to $1/(1-r^2)$, giving

$$A = atrt'\frac{1}{1-r^2}$$

But from Stoke's treatment, Eq. (2.4), f$tt' = 1 - r^2$, we obtain

$$A = ar \qquad (2.9)$$

This is just equal to the amplitude of the first reflected ray, hence, we conclude that under the condition of Eq. (2.8), there will be complete destructive interference. On the other hand, if the path difference given by Eq. (2.7) is an integral multiple of λ, i.e., when

$2\mu t\cos r - \frac{\lambda}{2} = n\lambda$, where $n = 0, 1, 2$, etc.,

or

$$2\mu t \cos r = (2n+1)\frac{\lambda}{2} \tag{2.10}$$

then ray (1) and (2) will be in phase with each other and gives a condition of constructive interference. But rays (3), (5), (7), ... will be out of phase with rays (2), (4), (6), ... Since (2) is more intense than (3), (4) is more intense than (5), etc., these pairs cannot cancel each other. As the stronger series combines with ray (1), the strongest of all, there will be maximum of intensity.

Thus, when a thin film is illuminated by monochromatic light, and seen in reflected light, it appears bright or dark according as $2\mu t \cos r$ is an odd multiple of $\lambda / 2$ or an integral multiple of λ, respectively.

$2\mu t \cos r = (2n+1)\frac{\lambda}{2}$	(condition of maxima)
$2\mu t \cos r = n\lambda$	(condition of minima)

Before moving further, answer the following SAQ.

SAQ 2
Using Eq. (2.7), state whether the following statement is true or false. Give reasons. "An excessively thin film seen in reflected light appears perfectly black".

Now we are in a position to know the reason of the production of colours in thin film of soap water.

Colours in Thin Films
The eye looking at the film receives rays of light reflected at the top and bottom surfaces of the film. These rays are in a position to interfere. The path difference between the interfering rays, given by Eq. (2.7), depends upon *t* (thickness of the film) and upon *r*, and, hence, upon inclination of the incident rays (the inclination is determined by the position of the eye relative to the region of the film, which is being looked at). The sunlight consists of a continuous range of wavelengths (colours). At a particular point of the film, and for a particular position of the eye (i.e., for a particular *t* and a particular *r)*, the rays of only certain wavelengths will have a path difference satisfying the condition of maxima. Hence, only these wavelengths (colours) will be present with the maximum intensity. While some others, which satisfy the condition of the minima will be missing. Hence, the point of the film being viewed will appear coloured.

We are working out an example so that the phenomenon of production of colours in thin film is clear to you.

Example 1
A thin film of 4×10^{-5} cm thickness is illuminated by white light normal to its surface $(r = 0^0)$. Its refractive index is 1.5. Of what colour will the thin film appear in reflected light?

Solution
The condition for constructive interference of light reflected from a film is

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

Here, μ = 1-5; t = 4×10^{-5} cm and r = *0°* (since light falls normally) so that $\cos r = 1$.

$$\therefore \qquad 2\times1.5\times4\times10^{-5} = (2n+1)\frac{\lambda}{2}$$

or $$\lambda = \frac{2\times2\times1.5\times4\times10^{-5}}{2n+1}$$

$$\lambda = \frac{24\times10^{-5}\,cm}{2n+1} = \frac{24000}{2n+1}\,\text{Å}$$

Taking n = 0, 1, 2, 3, …, we get

λ = 24000 Å, 8000 Å, 4800 Å, 343l Å, ...

These are the wavelengths reflected most strongly. Of these, the wavelength lying in the visible region is 4806A (blue).

So far we nave considered viewing of thin film in reflected light. Suppose the eye is now situated on the lower side of the film, shown in Fig. 2.3 and Fig. 2.5. The rays emerging from the lower side of the film can also be brought together with a lens and made to interfere.

Let us find out what colours will arise, when the film is viewed in this position. For this, we have to first calculate the path difference between the rays in transmitted light.
The path difference between the transmitted rays BT_1 and DT_2 is given by Eq. (2.6), i.e.,

$$(BC + CD) - BL = 2\mu t \cos r$$

In this case, there is no phase change due to reflection at *B* or *C*, because in either case the light is travelling from denser to rarer medium (See SAQ 1). Hence, the effective path difference between BT_1 and DT_2 is also

$$2\mu t \cos r = n\lambda \qquad \text{(condition for maxima)} \qquad (2.12a)$$

where n = 1, 2, 3, …

In this case, the film will appear bright in the transmitted light.

The two rays BT_1 and DT_2 reinforce each other, if

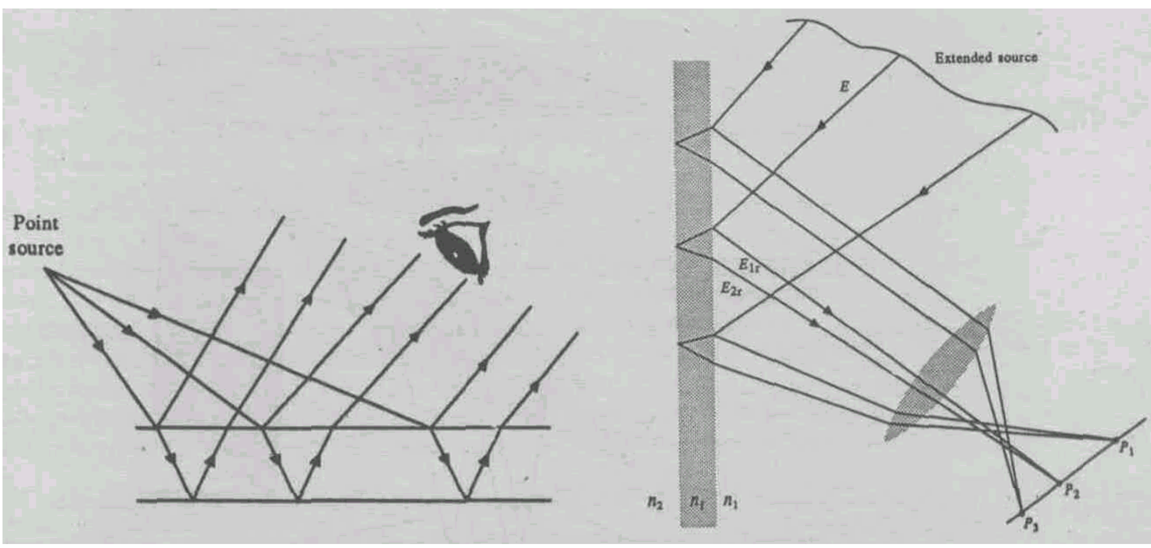
$$2\mu t \cos r = (2n+1)\frac{\lambda}{2} \text{ (condition of minima)} \qquad (2.12b)$$

where n = 0, 1, 2, ... and the film appears dark in transmitted light.

A comparison of Eqs. (6.11a), (6.11b), (6.12a) and (6.12b) shows that the conditions for the maxima and minima, in the reflected light are just the reverse of those in transmitted light. Therefore, only those colours will be visible in the transmitted light, which were missed in reflected light. Hence, the film which appears bright in reflected light will appear dark in transmitted light and vice versa. In other words, the appearances of colours in the two cases are complimentary to each other.

Interference fringes produced by thin films can be classified into two: Fringes of equal inclination and fringes of equal thickness.

Fringes of Equal Inclination
If the lens used in Fig. 2.3 to focus the rays has a small aperture, interference fringes will appear on a small portion of the film. Only the rays leaving the point source that are reflected directly into the lens will be seen (see Fig. 2.6a). For an extended source, light will reach the lens from various directions, and the fringe pattern will spread out over a large area of the film, as shown in Fig. 2.6b.



Flg. 2.6 (a) Fringes seen In a small portion of the film,
(b) Fringes seen on a large region of the film

The angle i or equivalently r, determined by the position P, will, in turn, control the path difference. The fringes appearing at points P_1 and P_2 in Fig. 2.7 are, accordingly, known as **fringes of equal inclination**.

Notice that as the film becomes thicker, the separation AC in Fig. 2.4 between ray (1) and (2) also increases, since $AC = 2t \tan r$. When only one of the two rays is able to enter the pupil of the eye, the interference pattern will disappear. The larger lens of a telescope could then, be used to gather in both rays, making the pattern visible. The separation can also be reduced by reducing r, and, therefore, i, i.e., by viewing the film at nearly normal incidence.

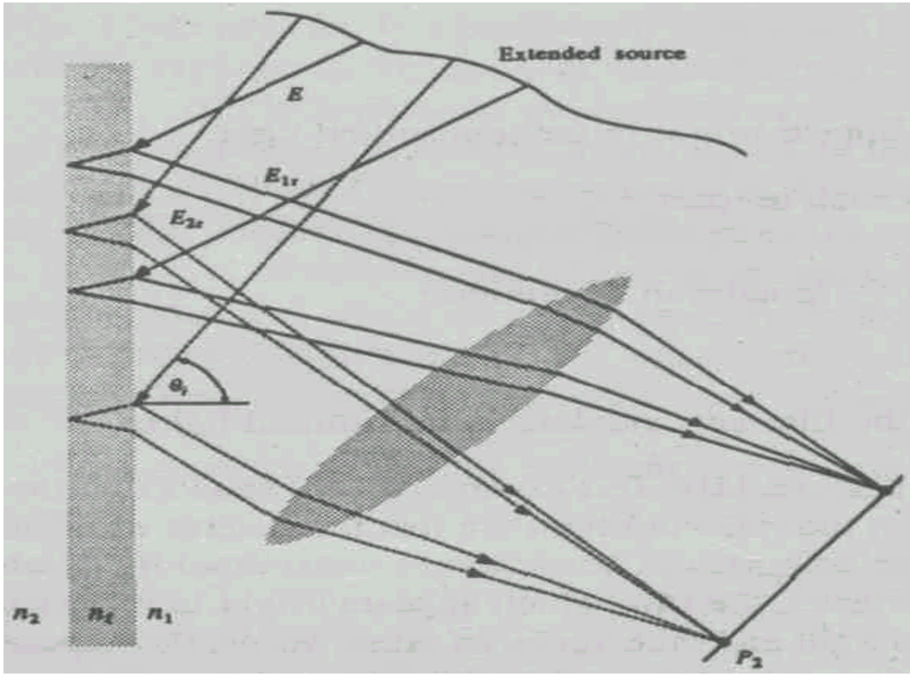


Fig. 2.7 All rays inclined at the same angle arrive at the same point

The equal inclination fringes that are seen in this manner for thick plates are known as **Haidinger fringes**. With an extended source, the symmetry of the set up requires that the interference pattern consists of a series of concentric circular bands centred on the perpendicular drawn from the eye to the film, as shown in Fig. 2.8.

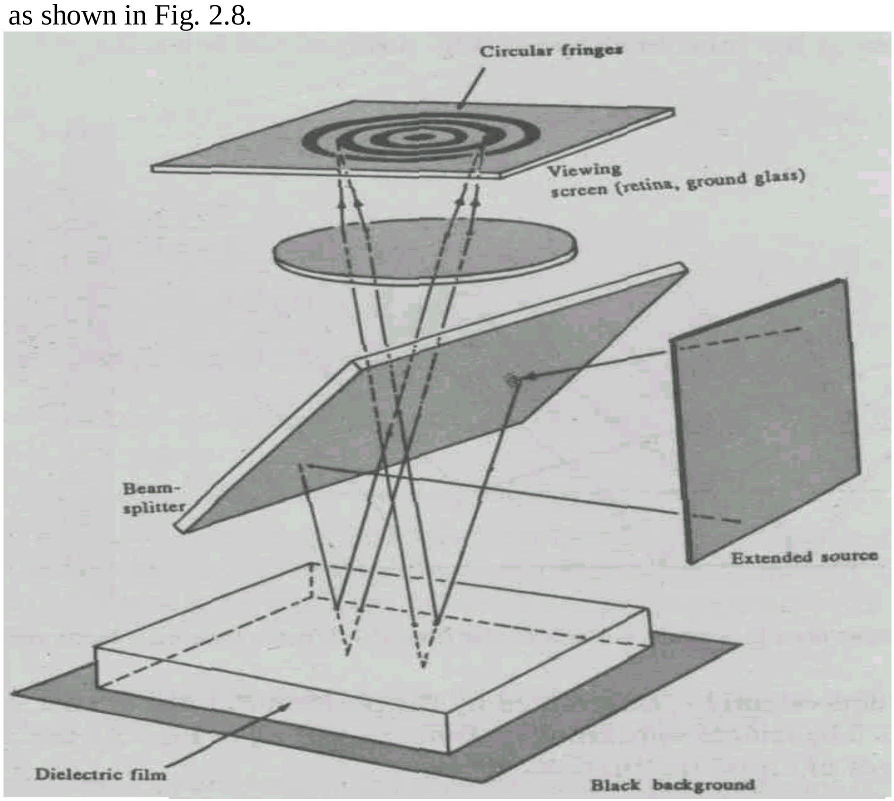


Fig. 2.8 Circular Haidinger fringes centred on the lens axis

Such fringes are formed at infinity, and are observed by a telescope focussed at infinity. These fringes are observed in Michelson interferometer, about which we will study in next unit.

Fringes of Equal Thickness

Interference fringes, for which thickness t is the dominant parameter rather than r, are referred to as fringes of equal thickness. Each fringe is the locus of all points in the film for which thickness is a constant. Such fringes are localised on the film itself, and are observed by a microscope focussed on the film. Fringes due to the wedge-shaped film belong to this class of fringes, which you will study in the next section.

Fringes of equal thickness can be distinguished from the circular pattern of Haidinger's fringes by the manner in which the diameters of the rings vary with order n. The central region in the Haidinger pattern corresponds to the maximum value of n, whereas just the opposite applies to fringes of equal inclination.

2.4 INTERFERENCE BY A WEDGE-SHAPED FILM

So far, we have assumed the film to be of uniform thickness. We will now discuss the interference pattern produced by a film of varying thickness, i.e., a film which is not plane-parallel. Such a film may be produced by a wedge, which consists of two non-parallel plane surfaces, as shown in Fig. 2.9a and 2.9b. Observe that the interfering rays do not enter the eye parallel to each other but appear to diverge from a point near the film.

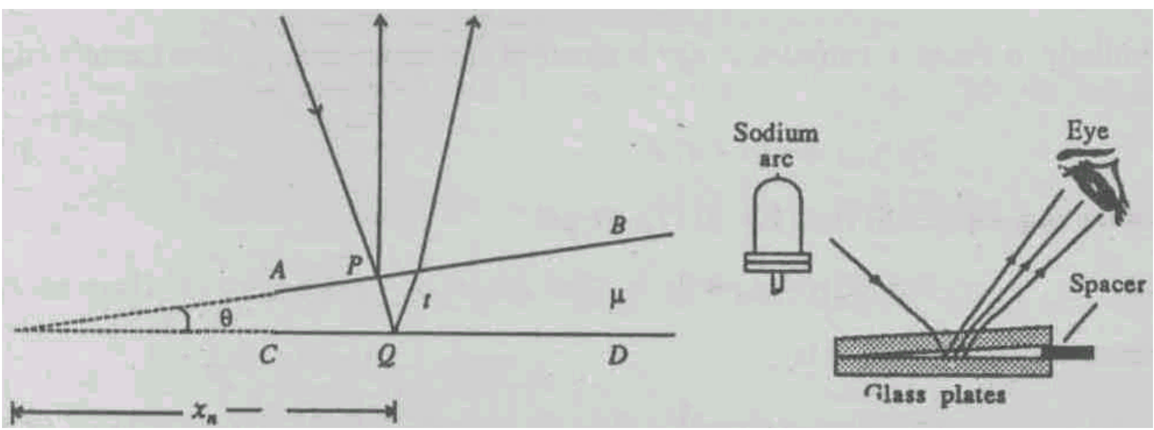


Fig. 2.9 Fringes of equal thickness: (a) method of visual observations,
(b) a parallel beam of light incident on a wedge

Let us consider a thin wedge-shaped film of refractive index μ, bounded by two plane surfaces AB and CD, inclined at an angle θ as shown in Fig. 2.9b. Let the film be illuminated by a monochromatic source of light from a slit held parallel to the edge of the wedge (the edge is the line passing through the point O and perpendicular to the plane of the paper). Interference occurs between the rays reflected at the upper and lower surfaces of the film. In this case the path difference for a given pair of rays is practically that given by Eq. (2.6). But, if it is assumed that light is incident almost normally at a point P on the film, the factor cos r may be considered equal to 1. Thus, the path difference between the rays reflected at the upper and lower surfaces is $2\mu t$, where t is the thickness of the film at P, An additional path difference of $\lambda/2$ is introduced in the ray reflected from the upper surface. The effective path difference between the two rays is

$$2\mu t - \frac{\lambda}{2} \tag{2.13}$$

Hence the condition for bright fringes becomes

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

or

$$2\mu t = (2n+1)\frac{\lambda}{2} \tag{2.14}$$

The condition for dark fringe is

$$2\mu t = n\lambda \tag{2.15}$$

It is clear that for a bright or dark fringe of a particular order, *t* must remain constant. Since in the case of a wedge-shaped film, *t* remains constant along lines parallel to the thin edge of the wedge, the bright and dark fringes are straight lines parallel to the thin edge of the wedge. Such fringes are commonly referred to as "fringes of equal thickness". At the thin edge, where *t* = 0, path difference = $\lambda/2$, which is a condition for minimum intensity. Hence, the edge of the film is dark. The resulting fringes resemble the localized fringes in the Michelson interferometer (this you will study in next unit) and appear to be formed in the film itself.

Spacing between Two Consecutive Bright (or Dark) Fringes

For the nth dark fringe, we have

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

Let this fringe be obtained at a distance x_n from the thin edge. Then $t = x_n \tan\theta = x_n\theta$ (when θ is small and measured in radians).

$$\therefore \qquad 2\mu x_n \tan\theta = (n+1)\lambda \tag{2.16}$$

Similarly, if the $(n+1)$th dark fringe is obtained at a distance x_{n+1} from the thin edge, then

$$2\mu x_n\theta = n\lambda \tag{2.17}$$

Subtracting Eq. (2.16) from Eq. (2.17), we get

$$2\mu\theta(x_{n+1} - x_n) = \lambda$$

Hence the fringe width β is

$$\beta = x_{n+1} - x_n = \frac{\lambda}{2\mu\theta} \tag{6.18}$$

where θ is measured in radians.

Similarly, it can be shown that the spacing between two consecutive bright fringes (fringe width) is $\dfrac{\lambda}{2\mu\theta}$.

SAQ 3

Using sodium light (λ = 5893 Å), interference fringes are formed by reflection from a thin air wedge. When viewed perpendicularly, 10 fringes are observed in a distance of 1 cm. Calculate the angle of the wedge.

If the fringes of equal thickness are produced in the air film between a convex surface of a long-focus lens and a plane glass surface, the fringes will be circular in shape because the thickness of the air film remains constant on the circumference of a circle. The ring-shaped fringes, thus produced, were studied by Newton. In the next section, we will study Newton's ring.

6.5 NEWTON'S RINGS

When a plano-convex lens of large radius of curvature is placed with its convex surface in contact with a plane glass plate, air-film is formed between the lower surface of the lens (LOL') and the upper surface of the plate (POQ), as shown in Fig. 2.10. The thickness of the air film is zero at the point of contact *O*, and it increases as one moves away from the point of contact. If monochromatic light is allowed to fall normally on this film, reflection lakes place at both the top and the bottom of the film. As a result of interference between the light waves reflected from the upper and lower surfaces of the air film, constructive or destructive interference takes place, depending upon the thickness of the film. The thickness of the air film increases with distance from the point of contact, therefore, the pattern of bright and dark fringes consists of concentric circles. In Fig. 2.10, 1 and 2 are the interfering rays corresponding to an incident ray *AB*. As the rings are observed in reflected light, the effective path difference between the interfering rays 1 and 2 is practically that given by Eq. (2.13).

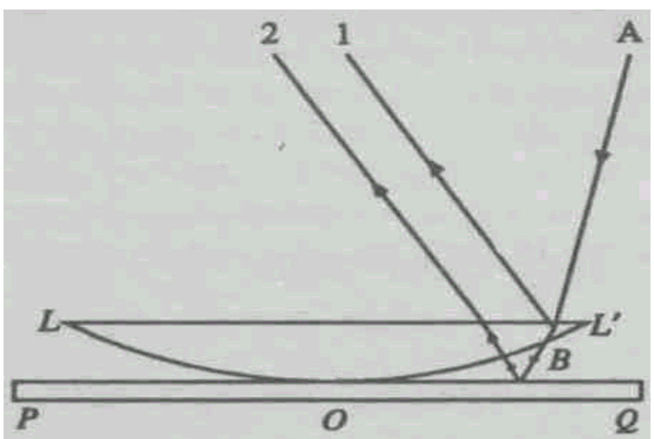


Fig. 2.10: An arrangement for observing Newton's rings

As we have considered an air-film, μ = 1. The condition for the bright ring which is given by Eq. (2.14), is

$$2t = (2n-1)\frac{\lambda}{2} \tag{2.19}$$

and the condition for the dark ring which is given by Eq. (2.15) is

$$2t = n\lambda \tag{2.20}$$

Let us find out the relationship between the radii of the rings and the wavelength of the light. Consider Fig. 6.11, where the lens *LOL' is* placed on the glass plate *POQ*. Let *R* be the radius of curvature of the curved surface of the lens. Let r_n be the radius of the *n*th Newton's ring corresponding to point *P*, where the film thickness is *t*. Draw perpendicular *PN*. Then, from the property of a circle, we have

$$PN^2 = ON \times NE$$

or

$$r_n^2 = t(2R-t)$$

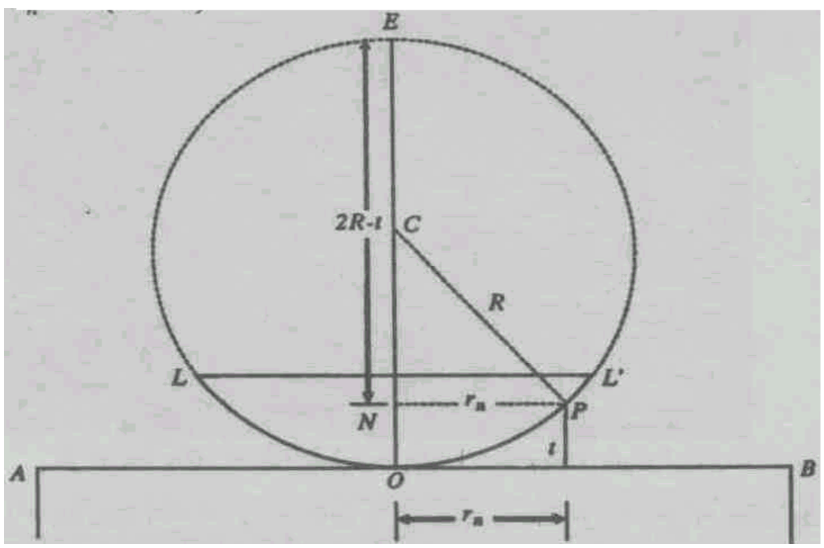


Fig. 2.11 r_n represents the radius of the nth dark ring, the thickness of air film (where the *n*th dark ring is formed) is *L*

Since *t* is small compared to *R*, we can neglect t^2.

Hence, $r_n^2 = 2Rt$

or $$2t = \frac{r_n^2}{R} \tag{2.21}$$

The condition for a bright ring is

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

But from Eq. (2.21), $2t = \frac{r_n^2}{R}$

$$\therefore \qquad \frac{r_n^2}{R} = (2n-1)\frac{\lambda}{2}$$

or

$$r_n^2 = (2n-1)\frac{\lambda R}{2} \qquad \text{(Bright ring)}$$

If D_n be the diameter of the nth bright ring, then $D_n = 2r_n$ or $r_n = \frac{D_n}{2}$. Substituting this in the last expression, we get

$$D_n^2 = 2(2n-1)\lambda R$$

or

$$D_n = \sqrt{2\lambda R}\sqrt{2n-1}$$

or

$$D_n \propto \sqrt{2n-1} \qquad (\lambda \text{ and } R \text{ being constant})$$

This shows that the radii of the rings vary as the square-root of odd natural numbers. Thus the rings will be close to each other as the radius increases, as shown in Fig. 2.12.

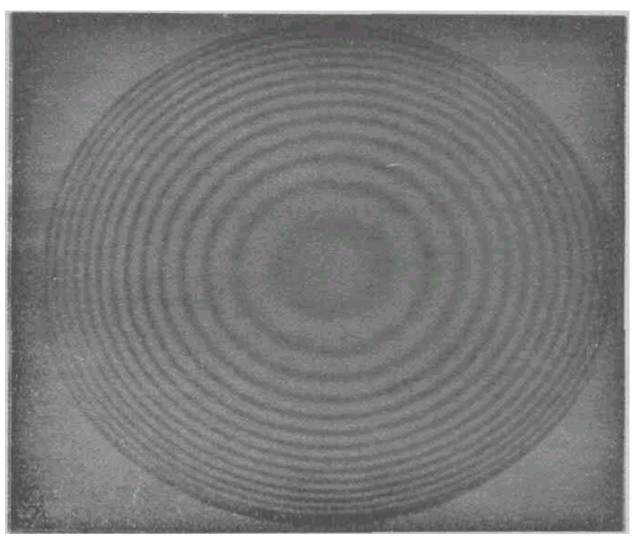


Fig. 2.12 Newton's rings as observed in reflected light

Between the two bright rings there will be a dark ring whose radius will be proportional to the square-root of the natural numbers. Attempt the following SAQ and prove the above statement yourself.

SAQ4
Using Eqs. (2.20) and (2.21), prove that the radius of the dark ring is proportional to the square-root of the natural numbers.

The ring diameters depend on wavelength, therefore, the monochromatic light will produce an extensive fringe system such as that shown in Fig. 2.12.

When the contact between lens and glass is perfect, the central spot is black. **This is direct evidence of the relative phase change of π between the two types of reflection, air-to-glass and glass-to-air, mentioned in Sec. 2.2. If there were no such phase change, the rays reflected from the two surfaces in contact should be in the same phase, and produce a bright spot at the centre.** However, the central spot can be made bright due to slight modification. In an interesting modification of the experiment, due to Thomas Young, if the lower plate is made to have a higher index of refraction than the lens, and the film in between is filled with an oil of intermediate index, then both reflections are at "rare-to-dense" surfaces. In this situation, no relative phase change occurs, and the central fringe of the reflected system is bright.

If D_n is the diameter of the nth bright ring, then

$$D_n^2 = 2(2n-1)\lambda R \tag{2.23}$$

If D_{n+p} is the diameter of the $(n+p)$th bright ring, then

$$D_{n+p}^2 = 2[2(n+p)-1]\lambda R \tag{2.24}$$

Subtracting Eq. (2.23) from Eq. (2.24), we get

$$D_{n+p}^2 - D_n^2 = 2[2(n+p)-1]\lambda R - 2(2n-1)\lambda R$$

$$= \quad 4p\lambda R$$

$$\therefore \qquad \lambda = \frac{D_{n+p}^2 - D_n^2}{4pR} \tag{2.25}$$

It may be mentioned here, that the point of contact may not be perfect. As such the nth ring may not be the *n*th fringe but Eq. (2.25) is almost always valid. On measuring the diameters of the rings and the radius of curvature *R*, the wavelength λ can be calculated with the help of the Eq. (2.25). In laboratory, the radius of curvature can be accurately measured with the help of a spherometer.

If a liquid of refractive index μ is introduced between the lens and the glass plate, then the expression for path difference between two interfering rays will also include μ. Then the radii of the dark rings would be given by

$$r_n = \left(\frac{n\lambda R}{\mu}\right)^{1/2} \tag{2.26}$$

Thus, when a little water is introduced between the lens and the plate, the rings contract according to the relation

$$\frac{\text{diameter of a ring in water - film}}{\text{diameter of the same ring in air - film}} = \frac{1}{\sqrt{\mu}} \tag{2.27}$$

where μ is the refractive index of water.

A ring system is also observed in the light transmitted by Newton's ring plates. There are two differences in the reflected and transmitted systems of rings, (i) The rings observed in transmitted light are exactly complementary to those seen in the reflected light, so that the central spot is now bright, (ii) The rings in transmitted light are much poorer in contrast than those in reflected light.

Before moving to the next section, solve the following SAQ.

SAQ5

If in a Newton's ring experiment, the air in the interspace is replaced by a liquid of refractive index 1.33, in what proportion would the diameters of the ring change?

6.6 APPLICATIONS OF THE PRINCIPLE OF INTERFERENCE IN THIN FILM

1. An important and simple application of the principle of interference within film is in the production of coated surfaces. To accomplish this, the glass lens is coated with the film of a transparent substance that has an index of refraction between the refraction indices for air and glass (See Fig. 2.13). The thickness of the film is one quarter of the wavelength of light in the film so that

$$t = \frac{\lambda}{4\mu_1}$$

If we assume normal incidence, then the path difference between the light wave reflected from the upper surface of the film and the light wave reflected from the lower surface of the film is $2\mu_1 t = 2\mu_1 \times \frac{\lambda}{4\mu_1} = \frac{\lambda}{2}$. Both waves undergo a phase change of 180° as reflections at both surfaces are from "rare-to-dense". Thus, the two reflected waves are out of phase because of path difference and, therefore, these interfere destructively. Such a film is known as non-reflecting film, because it gives zero reflection. However, this does not mean that a non-reflecting film destroys light, but it merely redistributes light so that a decrease of reflection is accompanied by a corresponding increase of transmission.

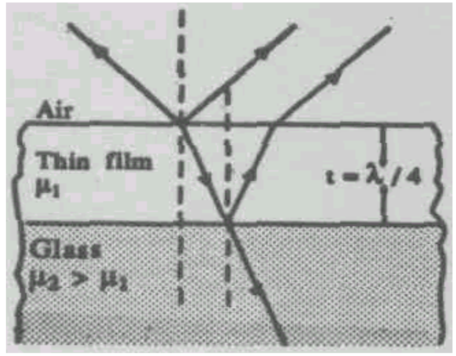


Fig. 2.13 A film coating on a glass lens makes the lens "non-reflecting" when the film thickness is $\lambda/4$ for normal incidence. The total path difference of the reflected rays is then $\lambda/2$, and the waves interfere destructively, i.e., the incident light is totally transmitted.

The practical importance of these films is that by their use one can greatly reduce loss of light by reflection at the various surfaces of lenses or prisms used in binoculars, cameras, etc. Usually, glass is coated with a very thin layer of magnesium fluoride, the refractive index of which (μ = 1.38) is intermediate between those of glass and air.

2. Another important application of thin film interference phenomenon is the converse of the procedure just discussed, viz., the glass surface is coated by a thin film of suitable material to increase the reflectivity. The film thickness is again $\lambda/4\mu_f$, where μ_f represents the refractive index of the film. The film is such that its refractive index is greater than that of the glass. This is because an abrupt phase change of π occurs only at the air-film interface and the beams reflected from the air-film interface and the film-glass interface constructively interfere.

3. The fringes obtained by a wedge-shaped film has important practical applications in the testing of optical surfaces for flatness. An air-film is formed between a perfectly plane surface and the surface under test. If the latter surface is plane, the fringes will be straight and parallel, and, if -not, these will be irregular in shape.

4. The accuracy of the grinding of a lens surface can be tested by observing the shape of Newton's rings formed between it and an accurately flat glass surface, using monochromatic light. If the rings are not perfectly circular, the grinding is imperfect.

2.7 SUMMARY

- When the light wave is reflected from a boundary, there is an abrupt change of phase. When the light ray is reflected while going from a rarer to a denser medium, it suffers a phase change of π. But there is no phase change when the light ray is reflected while going from a denser to a rarer medium.

- Length l in a medium of refractive index μ is optically equivalent to length μl in a vacuum. μl is called the optical path-length of distance l in the medium.

- For a thin film in reflected light, the conditions for constructive and destructive interference are:

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

$$2\mu t \cos r = n\lambda \qquad \text{(minima)}$$

where μ, is the refractive index of the film, t is its thickness and r is the angle of refraction in the film.

- For a thin film in transmitted light, the conditions for constructive and destructive interference are:

$$2\mu t \cos r = n\lambda \qquad \text{(maxima)}$$

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

- The basic formula for the path difference between the interfering rays, obtained due to division of amplitude by a film of thickness t and refractive index μ, is $2\mu t \cos r$, where r is the inclination of the ray inside the film. If the thickness of the film is uniform, the path difference $2\mu t \cos r$ varies only with inclination r, and gives rise to the "fringes of equal inclination". On the other hand, if the thickness of the film is rapidly varying, the path difference $2\mu t \cos r$ changes mainly due to changes in μ. This gives rise to the "fringes of equal thickness."

- The spacing p between two consecutive bright (or dark) fringes produced by wedge-shaped film is given by

$$\beta = \frac{\lambda}{2\mu\theta}$$

where λ is the wavelength of light being used for illuminating the film, μ the refractive index of the film, and θ (measured in radians) the angle between the two plane surfaces, which form the wedge-shaped film.

- The diameters of the bright rings are proportional to the square-roots of the odd natural numbers, whereas the diameters of dark rings are proportional to the square-roots of natural numbers, provided the contact is perfect.

- On measuring the diameters of Newton's rings and the radius of curvature R, the wavelength can be calculated with the help of the following relation:

$$\lambda = \frac{D_{n+p}^2 - D_n^2}{4pR}$$

- The phenomenon of interference is used in the testing of optical surfaces and producing non-reflecting glasses of reflective coatings.

2.8 TERMINAL QUESTIONS

(1) White light is reflected normally from a uniform oil film (μ = 1.33). An interference maximum for 6000 Å and a minimum for 4500 Å, with no minimum in between, are observed. Calculate the thickness of the film.

(2) Light (λ = 6000 Å) falls normally on a thin wedge-shaped film (μ = 1.5). There are ten bright and nine dark fringes over the length of the film. By how much does the film thickness change over this length?

(3) Two glass plates 12 cm long touch at one end, and are separated by a wire 0.048 mm in diameter at the other. How many bright fringes will be observed over the 12 cm distance in the light (λ = 6800 Å) reflected normally from the plates?

(4) Newton's rings are formed in reflected light of wavelength 5895×10^{-8} cm with a liquid between the plane and curved surfaces. The diameter of the fifth ring is 0.3 cm and the radius of curvature of the curved surface is 100 cm. Calculate the refractive index of the liquid, when the ring is (i) bright, (ii) dark.

(5) A Newton's rings arrangement is used with a source emitting two wave-lengths

$$\lambda_1 = 6.0\times10^{-5} \text{ cm and } \lambda_2 = 4.5\times10^{-5} \text{ cm}$$

and it is found that the *n*th dark ring due to λ_1 coincides with the $(n+1)$th dark ring due to λ_2. If the radius of curvature of the curved surface is 90 cm, find the diameter of the *n*th dark ring for λ_1.

2.9 SOLUTIONS/ANSWERS

SAQs

(1) See Fig. 2.14.

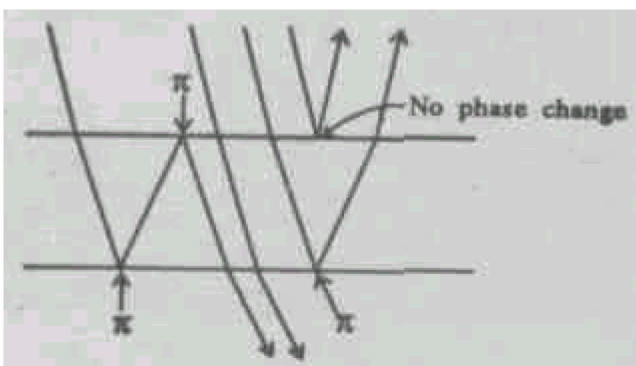


Fig. 2.14

(2) According to Eq. (2.7) the path difference between the interfering rays in reflected light is $2\mu t\cos r - \dfrac{\lambda}{2}$. When the film is excessively thin, *t* is very small, and $2\mu t\cos r$ is almost zero. Hence the path difference, in such a case becomes $\lambda/2$.

(3) Let θ radian be the angle of the air-wedge. For normal incidence, the fringe-width is given by

$$\beta = \frac{\lambda}{2\theta} \qquad (\because\ \mu = 1 \text{ for air})$$

Here $\lambda = 5893 \times 10^{-8}$ cm and β = 1/10 cm.

$$\therefore \qquad \theta = \frac{\lambda}{2\beta} = \frac{5893 \times 10^{-8}}{2 \times 1/10} = 2.95 \times 10^{-4} \text{ radian}$$

(4) According to Eq. (2.20), the condition for the dark ring is

$$2t = n\lambda$$

But from Eq. (2.19), $2t = \dfrac{r_n^2}{R}$

$$\therefore \qquad \frac{r_n^2}{R} = n\lambda$$

If D_n be the diameter of the nth dark ring, $r_n = \dfrac{D_n}{2}$

$$\therefore \qquad \frac{D_n^2}{4R} = n\lambda$$

or $\qquad D_n = \sqrt{4nR\lambda}$

or $\qquad D_n = \sqrt{4R\lambda}\sqrt{n}$

or $\qquad D_n \propto \sqrt{n}$

Thus, the diameters of the dark rings are proportional to the square root of the natural number.

(5) $$\frac{(D_n)^2_{air}}{(D_n)^2_{liquid}} = \frac{1}{\sqrt{\mu}} = \frac{1}{\sqrt{1.33}} = 0.867$$

The rings are contracted to 0.867 their previous diameters.

(6) In this case of interference in thin films, the situation is somewhat different. The reflections at both the upper and lower surfaces of the material (μ =1.25) film take place under similar conditions, i.e., when light is going from a rarer to a denser medium. Thus, there is a phase change of π at both reflections, which means no phase difference due to reflection between the two interfering beams.

The path difference between the two interfering beams is $2\mu t$ for normal incidence, where t is the thickness and μ the refractive index of the film.

The two beams will destroy each other, if the path difference is an odd multiple of $\lambda/2$, i.e., when

$2\mu t = (2n-1)\dfrac{\lambda}{2}$, where n = 1, 2, 3, ...

This is the condition of minima.
Here μ = 1.25 and λ = 6000 Å.

$$\therefore \qquad 2\times 1.25\times t = (2n-1)\times\frac{6000}{2}\ \text{Å}$$

Hence the required thickness is given by

$$t = (2n-1)\frac{6000}{2\times 2\times 1.25}\ \text{Å}$$

$$= (2n-1)1200\ \text{Å, where } n = 1, 2, 3, \ldots$$

TQs

(1) The condition for an interference maximum in the light reflected normally from an oil film of thickness t is

$$2\mu t = \left(n+\frac{1}{2}\right)\lambda\text{, where } n = 0, 1, 2, \ldots$$

and that for an interference minimum is

$$2\mu t = n\lambda\text{, where } n = 1, 2, 3, \ldots$$

Here μ = 1.33. Now there is a maximum for λ = 6000 Å

We can write

$$2\times 1.33\times t = \left(n+\frac{1}{2}\right)6000\ \text{Å} \qquad \text{(i)}$$

$$2\times 1.33\times t = (n+1)\,4500\ \text{Å} \qquad \text{(ii)}$$

In view of eq. (i) we have taken the integer $(n+1)$ rather than n in eq. (ii). Comparing eq. (i) and (ii), we get

$$\left(n+\frac{1}{2}\right)6000 = (n+1)4500$$

$$\therefore \qquad n = 1$$

Substituting n = 1 in eq. (i), we get

$$2 \times 1.33 \times t = \frac{3}{2} \times 6000 \text{ Å}$$

$$\therefore \quad t = \frac{3 \times 6000}{2 \times 2 \times 1.33} = 3383 \text{ Å}$$

(2) The condition of destructive interference in light reflected from a film is

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

Suppose the film thickness changes over this length by Δt. Let n be the order of the dark fringe appearing at one end of the film. The order of the dark fringe at the other end will be (n + 9). We, therefore, have

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

and $$2\mu(t + \Delta t)\cos r = (n + 9)\lambda$$

Subtracting, we get

$$2\mu(\Delta t)\cos r = 9\lambda$$

$$\therefore \quad t = \frac{9\lambda}{2\mu \cos r}$$

If the fringes are seen normally, then $\cos r = 1$.

$$\therefore \quad t = \frac{9}{2\mu} = \frac{9 \times 6300}{2 \times 1.5} = 18900 \text{ Å}$$

$$= 1.89 \times 10^{-4} \text{ cm}$$

(3) Let t be the thickness of the wire and l the length of the wedge, as shown in Fig. 2.15. The wedge angle is

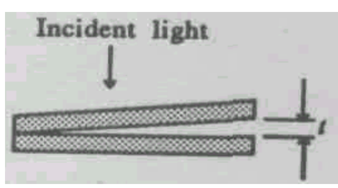


Fig. 2.15

$$\theta = \frac{t}{l} \text{ radian}$$

Now, fringe-width $\beta = \frac{\lambda}{2\theta}$.

Putting value of θ from above, we get

$$\beta = \frac{l\lambda}{2t}$$

Since N fringes are seen; $l = N\beta$. Thus

$$\beta = \frac{N\beta\lambda}{2t}$$

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

But λ = 6800 Å = 6800×10^{-8} cm and t = 0.048 mm = 0.0048 cm.

$$\therefore \qquad N = \frac{2 \times 0.0048}{6800 \times 10^{-8}} = 141$$

4.(i) The diameter D_n of the nth bright ring is given by

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

$$\therefore \qquad \mu = \frac{2(2n-1)\lambda R}{D_n^2}$$

Here n = 5, λ = 5895×10^{-8} cm, R = 100 cm and D_n = 0.3 cm

$$\mu = \frac{2(10-1) \times 5895 \times 10^{-8} \times 100}{(0.3)^2} = 1.18$$

(ii) The diameter of the nth dark ring is given by

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

$$\therefore \qquad \mu = \frac{4n\lambda R}{D_n^2} = \frac{4 \times 5 \times 5895 \times 10^{-8} \times 100}{(0.3)^2} = 1.31$$

(5) $D_n^2 = 4nR\lambda$

where D_n = diameter of nth ring, R = the radius of curved surface and λ = the wavelength of light.

If D_n and D_{n+1} be two diameters.

$$D_n^2 = 4nR\lambda_1 \qquad \text{(i)}$$

$$D_{n+1}^2 = 4(n+1)R\lambda_1 \qquad \text{(ii)}$$

But $\quad D_n = D_{n+1}$

$\therefore \quad 4nR\lambda_1 = 4(n+1)R\lambda_2$

or $\quad 4nR(\lambda_1 - \lambda_2) = 4R\lambda_2$

or
$$n = \frac{4R\lambda_2}{4R(\lambda_1 - \lambda_2)}$$

$$= \frac{\lambda_2}{\lambda_1 - \lambda_2}$$

$$= \frac{4.5\times10^{-5}}{(6-4.5)\times10^{-5}} = 3$$

Putting $n = 3$ in (i)

$$D_3 = 4\times3\times90\times6\times10^5$$

$$= 648\times10^{-4}$$

$$= 25.45\times10^{-2} \text{ cm}$$

UNIT 3 INTERFEROMETRY

Structure

3.1 INTRODUCTION

An instrument designed to exploit the interference of light and the fringe patterns that result from optical path differences, in any of a variety of ways, is called an optical interferometer. In this unit, we explain the functioning of the Michelson and the Fabry-Perot interferometers, and suggest only a few of their many applications.

In order to achieve interference between two coherent beams of light, an interferometer divides an initial beam into two or more parts that travel diverse optical paths and then superpose to produce an interference pattern. One criterion for broadly classifying interferometers distinguishes the manner in which the initial beam is separated. Wavefront division interferometers sample portions of the same wavefront of a coherent beam of light, as in the case of Young's double slit, Lloyd's mirror or Fresnel's biprism arrangement. Amplitude-division interferometers, instead, use some type of beamsplitter that divides the initial beam into two parts. The Michelson interferometer is of this type. Usually the beam splitting is managed by a semi-reflecting metallic film. In this interferometer, the two interfering beams are widely separated, and the path difference between them can be varied at will by moving the mirror or by introducing a refracting material in one of the beams. Corresponding to these two ways of changing the optical path, there are two important applications of this interferometer, which we will study in this unit.

There is yet another means of classification that distinguishes between those interferometers that function by the interference of two beams, as in the case of the Michelson interferometer, and those that operate with multiple beams, as in the Fabry-Perot interferometer. In this unit, we will show that the fringes so formed are sharper than those formed by two beam interference. Therefore, the interferometers involving multiple beam interference have a very high resolving power, and, hence, find applications in high-resolution spectroscopy.

Objectives

After studying this unit, you should be able to

- understand how Michelson interferometer produces different types of fringes, viz., circular, localised (or straight) and white light fringes,
- describe few applications of Michelson interferometer,
- relate the intensity of the transmitted light to the reflectance of the plate surface in Fabry-Perot interferometer, and

- understand the difference between Michelson interferometer and Fabry-Perot interferometer.

3.2 MICHELSON INTERFEROMETER

It is an excellent device to obtain interference fringes of various shapes, which have a number of applications in optics. It utilizes the arrangements of mirrors and a beam splitter.

Construction: Its configuration is illustrated in Fig. 3.1.

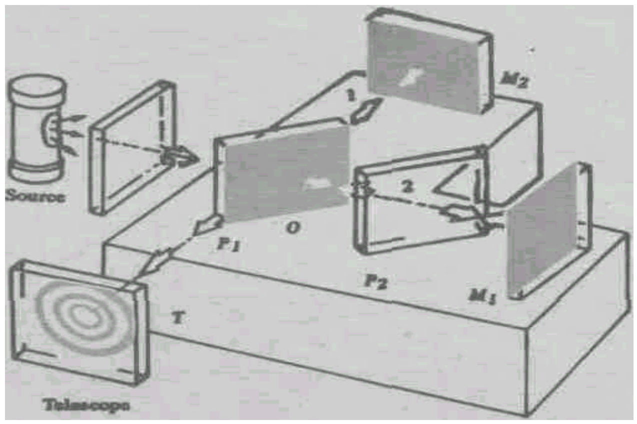


Fig. 3.1: Michelson interferometer

Its main optical parts are two plane mirrors M_1 and M_2 and two similar optically-plane parallel glass plates P_1 and P_2. The plane mirrors M_1 and M_2 are silvered on their front surfaces and are mounted vertically on two arms at right angles to each other. To obtain fringes, the mirrors M_1 and M_2 are made exactly perpendicular to each other by means of screws shown on mirror M_1. The mirror M_2 is mounted on a carriage which can be moved in the direction of the arrows. The plates P_1 and P_2 are mounted exactly parallel to each other, and inclined at 45° to M_1 and M_2. The surface of P_1 and P_2 is partially silvered. The plate P_1 is called beam splitter.

Working: An extended source (e.g., a diffusing ground glass plate illuminated by a discharge lamp) emits light waves in different directions, part of which travel to the right and fall on P_1. The light wave incident on P_1 is partly reflected and partly transmitted. Thus, the incident wave gets divided into two waves, viz., the transmitted wave 1 and the reflected wave 2. These two waves travel to M_1 and M_2 respectively. After reflection at M_1 and M_2 the two waves return to P_1. Part of the wave coming from M_2 passes through P_1 going downward towards the telescope, and part of the wave coming from M_1 gets reflected by P_1 toward the telescope. Since the waves entering the telescope are derived from the same incident wave, they are coherent, and, hence, in a position to interfere. The interference fringes can be seen in the telescope.

You must be eager to know the purpose of the plate P_2, because till now we have not mentioned anything about it.

Function of the plate $\mathbf{P_2}$: Note that if reflection at P_1 occurs at the rear surface at point O, as shown in Fig. 3.1, the light reflected at M_2 will pass through P_1 three times while the light reflected at M_1 will pass through only once. Thus, the paths of waves 1 and 2 in glass are not equal. Consequently, each wave will pass through the same thickness of glass only when a compensator plate P_2, of the same thickness and inclination at P_1, is inserted in the path of wave 1. The compensator plate is an exact duplicate of P_1 with the exception that it is not partially silvered. With the compensator in place, any optical path difference arises from the actual path difference.

In contrast to the Young double slit experiment, which uses light from two very narrow sources, the Michelson interferometer uses light from a broad spread out source.

Form of fringes: The form of the fringes depends on the inclination of M_1 and M_2. To understand how fringes are formed, refer to the Fig. 3.2, where the physical components are represented somewhat differently. An observer at the position of the telescope will, simultaneously, see both mirrors M_1 and M_2 along with the source L formed by reflection m the partially silvered surface of the glass plate P_1. Accordingly, we can redraw the interferometer as if all the elements were in a straight line. Here $M_1^{'}$ corresponds to the image of mirror M_1, formed by reflection at the silvered surface of the glass plate P_1, so that $OM = OM_1^{'}$. Depending on the positions of the mirrors, image $M_1^{'}$ may be in front of, behind or exactly coincident with mirror M_2. The surfaces L_1 and L_2 are images of the source L in mirrors M_1 and M_2 respectively. If we consider a single point S on the source L, emitting light in all directions, then on reaching O, it gets split, and thereafter its segments get reflected by M_1 and M_2. In Fig. 3.2 we represent this by reflecting the ray off both M_1 and M_2. Thus, the interference fringes may be regarded to be formed by light reflected from the surface of $M_1^{'}$ and M_2. Here S_1 and S_2 act as coherent point sources, because to an observer at D the two reflected rays will appear to have come from the image points S_1 and S_2. The mirror M_2 and the virtual image of M_1 play the same roles as the two surfaces of the thin film, discussed in Unit 2, and the same sort of interference fringes result from the light reflected by these surfaces.

Now, let us discuss the various types of fringes, viz., circular fringes, localised fringes and white light fringes.

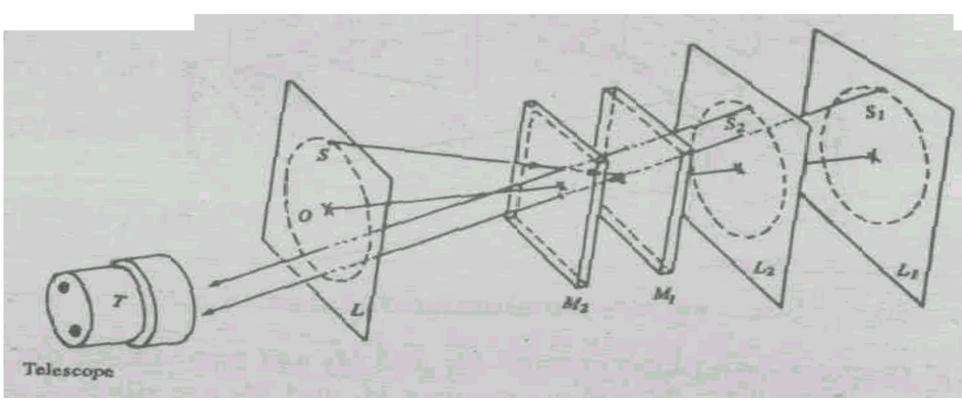


Fig. 3.2 A conceptual rearrangement of the Michelson Interferometer

3.2.1 Circular Fringes

These fringes are observed when M_1 is exactly perpendicular to M_2. In this situation the distance of the minors M_1 and M_2 from the plate P_1 can be varied.

Let us consider the various possible positions of the mirrors M_1 and M_2, and eventually see how it gives rise to circular fringes.

(i) If the two mirrors have the same axial distance from the rear face of P_1, and if they are perpendicular to each other, the image $M_1^{'}$ is coincident with M_2. At the coincidence position, the two paths are of equal length. Thus, we expect the waves to reinforce each other and to form a maximum. But this is not so, because of the π phase change, which occurs on external (air-to-glass) reflection only. No phase change occurs on internal (glass-to-air) reflection, and none occurs on transmission or refraction. Look again at Fig. 3.1 and note that it is the light that comes from M_1 and goes to the observer that is reflected, air-to-glass, at *O*, and undergoes the π change. This means that at the coincidence position, there will be a minimum: **the centre of the field will be dark**.

(ii) Now, we move one of the mirrors. If the mirror is moved through a quarter of wavelength, $d = \lambda/4$, the path length (because if d is separation between M_1 and M_2, then $2d$ is the separation between S_1 and S_2) changes by $\lambda/2$, the two waves getting out of phase by 180°, the phase change compensates, and we have a maximum. Moving the mirror by another $\lambda/4$, gives minimum, another $\lambda/4$ another maximum and so on. Thus,

$$2d = m\lambda, \text{ where } m = 0, 1, 2, \ldots \qquad (3.1)$$

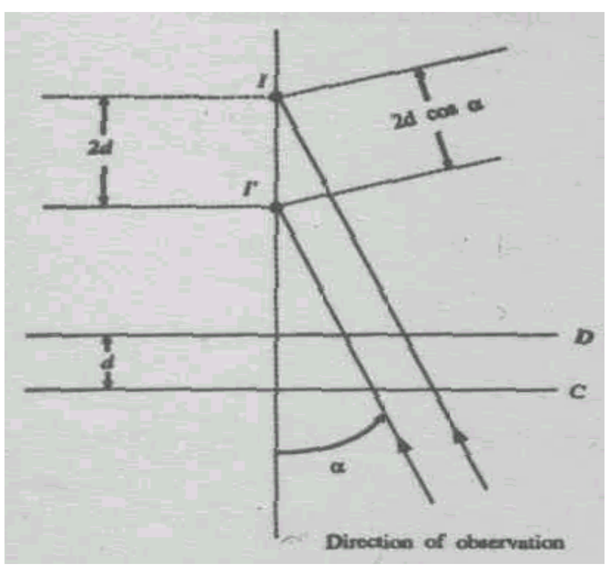


Fig. 3.3 Looking off-axis into the Michelson interferometer

(iii) Next, we assume that we look obliquely into the interferometer and that our line of sight makes an angle α with the axis. Ordinarily, the two planes M_1 and M_2 are at a distance *d* apart, and the two virtual images, I and I' separated by *2d*. But for oblique incidence, as we see from Fig. 3.3, the path difference between the two lines of sight becomes less and instead of Eq. (3.1), we get

$$2d\cos\alpha = m\lambda \text{, where } m = 0, 1, \ldots \tag{7.2}$$

For a given mirror separation *d*, and a given order *m*, wavelength λ and angle α is constant. The maxima will lie in the form of circles about the foot of the perpendicular from the eye to the mirrors. These circular fringes will look like the ones shown in Fig. 3.4. Fringes of this kind, where parallel beams are brought to interfere with a phase difference determined by the angle of inclination θ, are referred to as fringes of equal inclination. These fringes are also known as Haidinger fringes. They differ from the fringes of equal inclination considered in Unit 2, only in that, here there are no multiple reflections so that the intensity distribution is in accordance with Eq. (1.17)

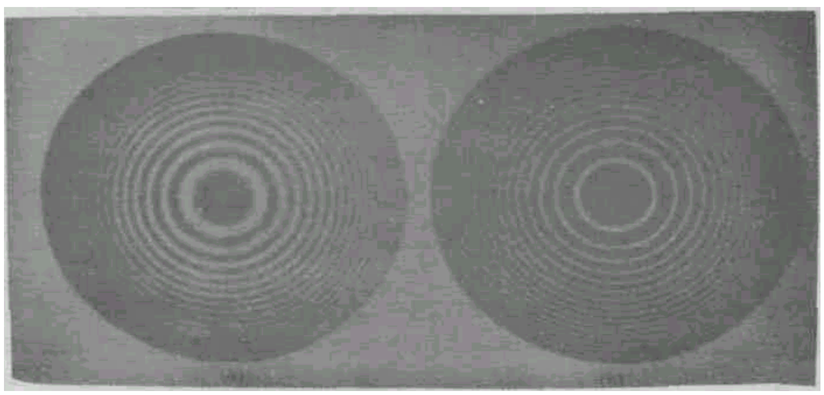


Fig. 3.4 Fringes observed using (a) Michelson Interferometer, (b) Fabry-Perot Interferometer

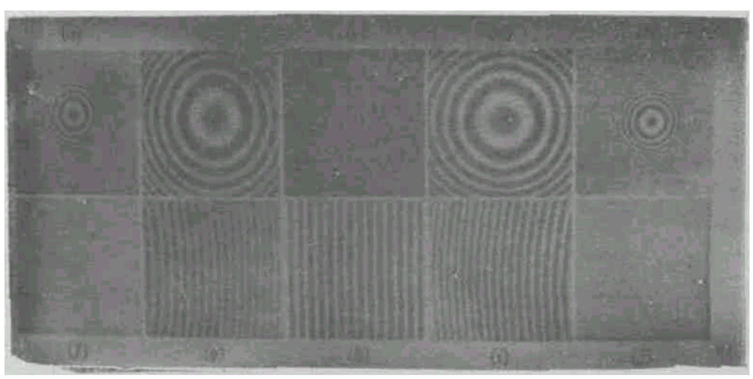


Fig. 3.5 Appearance of the various types of fringes observed in the Michelson Interferometer. Upper row shows circular fringes whereas lower row shows, localized fringes. Path difference increases outward, hi both directions, from the centre

The upper part of the Fig. 3.5 shows how the circular fringes look under different conditions. When M_2 is few centimetres beyond M_1, the fringe system will have the general appearance shown in (a) with the rings very closely spaced. If M_2 is now moved slowly toward M_1, so that *d* is decreased, Eq. (3.2) shows that a given ring, characterized by a given value of the order *m*, must decrease its radius, because the product $2d\cos\alpha$ must remain constant. The rings, therefore, shrink and vanish at the centre; a ring disappearing each time $2d$ decreases by λ, or *d* by $\lambda/2$. This follows from the fact at the centre $\cos\theta$ = 1, so that Eq. (3.2) becomes

$$2d = m\lambda$$

which is Eq. (3.1).

To change *m* by unity, *d* must change by $\lambda/2$. Now as M_2 approaches M_1 the rings become more widely spaced as indicated in Fig. (3.5b), until we reach a critical position, where the central fringe has spread out to cover the whole field of view, as shown in Fig. 3.5 (c). This happens when M_2 and M_1 are exactly coincident, for it is clear that under these conditions the path difference is zero for all angles of incidence. If the mirror is moved still farther, it effectively passes through M_1, and new widely spaced fringes appear, growing out from the centre. These will gradually become more closely spaced, when the path difference increases, as indicated in (d) and (e) of the Fig. 3.5.

3.2.2 Localized Fringes (Straight Fringes)

If the mirrors M_1 and M_2 are not exactly parallel, the air film between the mirrors is wedge-shaped, as indicated in Fig. 3.6.

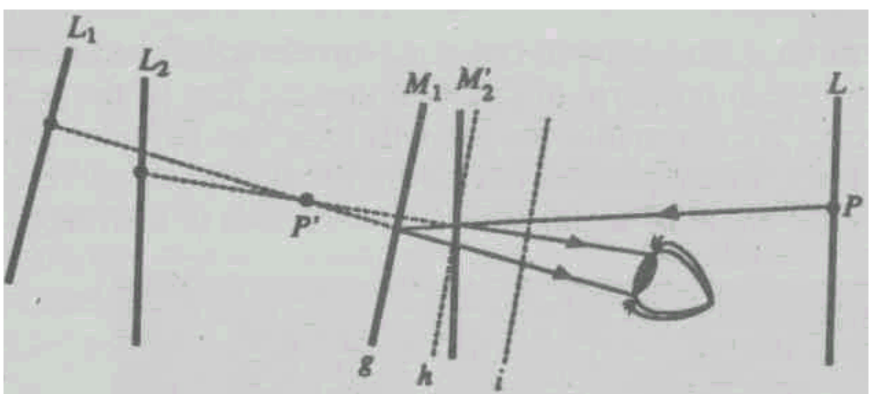


Fig. 3.6 The formation of fringes with inclined mirrors in the Michelson interferometer

The two rays reaching the eye from point *P* on the source are now no longer parallel, but appear to diverge from point *P'* near the mirrors. For various positions of *P* on the extended source, the path difference between the two rays remains constant, but the distance of *P'* from mirrors changes. If the angle between the mirrors is not too small, the latter distance is never great, and hence, in order to see these fringes clearly, the eye must be focused on or near the rear mirror M_2. The localized fringes are, practically, straight, because the variation of the path difference across the field of view is now due primarily to the variation of the thickness of the "air film" between the mirrors. With a wedge-shaped film, the locus of point of equal thickness is a straight line, parallel to the edge of the wedge. The fringes are not exactly straight, if *d* has an appreciable value, because there is also some variation of the path difference with angle. They are, in general, curved and are always convex toward the thin edge of the wedge. Thus, with a certain value of *d*, we might observe fringes shaped like those of Fig. 3.5(g). M_2 could then be in position such as *g* of Fig. 3.6. If the separation of the mirrors is decreased, the fringes will move to the left across the field, a new fringe crossing the centre each time *d* changes by $\lambda / 2$. As we approach the zero path difference, the fringes become straighter until the point is reached where M_2 actually intersects M_1, when they are perfectly straight, as in Fig. 3.5(h). Beyond this point, they begin to curve in the opposite direction, as shown in Fig. 3.5 (i). The blank fields shown in Fig. 3.5 (f) and (j) indicate that this type of fringe cannot be observed for large path differences. As the principle variation of path difference results from a change of the thickness *d*, these fringes have been termed fringes of equal thickness.

3.2.3 White Light Fringes

If a source of white light is used, no fringes will be seen at all except for a path difference so small that it does not exceed a few wavelengths. In observing these fringes, the mirrors are tilted slightly as for localized fringes, and the position of M_2 is found where it intersects M_1. With white light there will then be observed a central dark fringe, bordered on either side by 8 or 10 coloured fringes. This position is often rather troublesome to find, using white light only. It is best located approximately before hand by finding the place where the localized fringes in monochromatic light become straight. Then, a very slow motion of M_1 through this region, using white light, will bring these fringes into view.

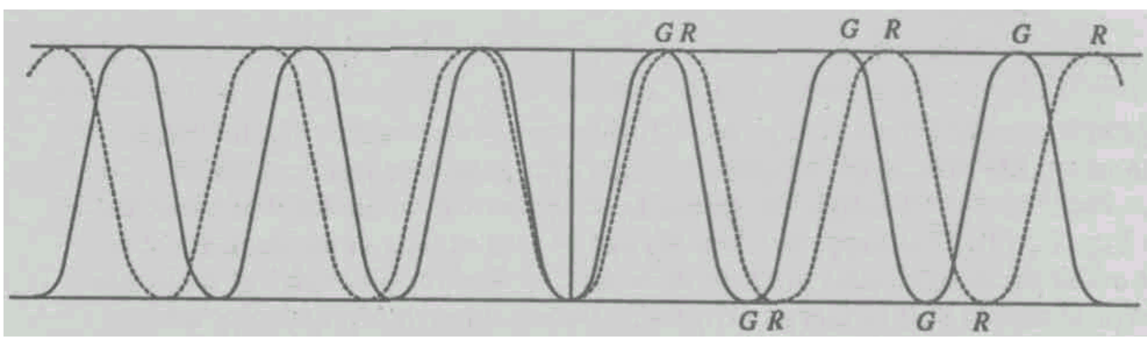


Fig. 3.7: The formation *of* white light fringes with a dark fringe at the centre

The fact that only a few fringes are observed with white light, is easily accounted for when we remember that such light contains all wavelengths between 400 and 750 mm. The fringes for a given colour are more widely spaced, the greater the wavelength. Thus, the fringes in different colours will only coincide for *d* = 0, as indicated in Fig. 3.7. The solid curve represents the intensity distribution in the fringes for the green light, and the broken curve for the red light. Clearly, only the central fringe will be uncoloured, and the fringes of different colours will begin to separate at once on either side. After 8 or 10 fringes, so many colours are present at a given point that the resultant colour is essentially white. While light fringes are, particularly, important in the Michelson interferometer, where they may be used to locate the position of zero path difference, as we shall see later.

3.2.4 Adjustment of the Michelson Interferometer

(i) **For Localised fringes**: The distance of the mirrors M_1 and M_2 from the silvered surface of P_1 are first made as nearly equal as possible by moving the movable mirror M_2. A pinhole is placed between the lens and the plate P_1 (Fig. 3.8). If M_1 is not perpendicular to M_2, four images of the pinhole are obtained, two by reflection at the semi-silvered surface of P_1 and the other two by reflection at the other surface of P_1.

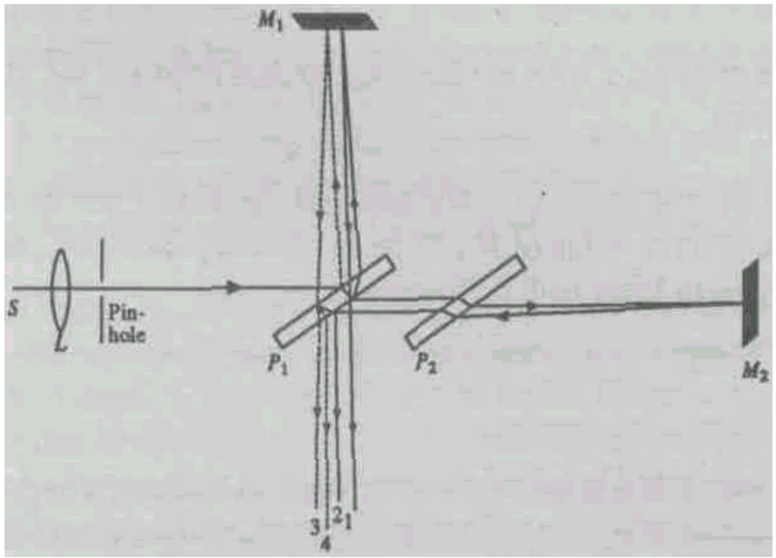


Fig. 3.8: Adjustment of Michelson interferometer

The former pair is, naturally, brighter than the latter. The small screws at the back of the mirror, M_1 are then adjusted until the two bright images appear to coincide. The pinhole is now removed. If the coincidence of the images was apparent, the air-film between M_1 and M_2 would be wedge-shaped, and the localised fringes would appear.

(ii) **For White light Localised Fringes**: First, the localised fringes with monochromatic light are obtained. The mirror M_2 is then moved until the fringes become straight. Monochromatic light is replaced by white light. M_2 is further moved in the same direction until the central achromatic fringe is obtained in the field of view.

(iii) **For Circular Fringes**: After localised fringes are obtained, the screws of M_1 are adjusted so that the spacing between these fringes increases. This happens when the angle of the wedge decreases. If this adjustment be continued, at one stage, the angle of the wedge will become zero, and the film will be of constant thickness. At this stage, circular fringes will appear. Finer adjustment is made until on moving the eye sideways or up and down, the fringes do not expand or contract.

3.2.5 Applications

There are three principal types of measurement that can be made with Michelson interferometer: (i) wavelengths of light (ii) width, and fine structure of spectrum lines (iii) refractive indices. As explained in the sub-section 3.2.3, when a certain spread of wavelengths is present in the light source, the fringes become indistinct and, eventually, disappear as the path difference is increased. With white light they become invisible when d is only a few wavelengths, whereas the circular fringes obtained with the light of single spectrum line can still be seen after the mirror has been moved several centimetres. Therefore, for making these measurements with this interferometer, it is adjusted for circular fringes.

(a) Determination of Wavelength of Monochromatic Light

After having adjusted interferometer for circular fringes, adjust the position of M_2 to obtain a bright spot at the centre of the field of view. If *d* be the thickness of the film and *n* the order of the spot obtained, we have

$$2d\cos\alpha = n\lambda \qquad (3.3)$$

But at the centre $\alpha = 0$, so that $\cos\alpha = 1$. Therefore

$$2d = n\lambda \qquad (3.4)$$

If now M_2 be moved away form M_1 by $\lambda/2$, 2*d* increases by λ. Therefore *n* + 1 replaces *n* in Eq. (3.4). Hence, (*n* + l)th bright spot now appears at the centre (see sec. 3.2.1). Thus, each time M_2 moves through a distance $\lambda/2$, the next bright spot appears at the centre. Suppose, during the movement of M_2 through a distance *x, N* new fringes appear at the centre of the field. Then we have

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

$$\therefore \qquad \lambda = \frac{2x}{N} \qquad (3.5)$$

Thus, by measuring the distance *x* with the micrometer and counting the number *N*, the value of λ can be obtained.

The determination of λ by this method is very accurate, because *x* can be measured to an accuracy of 10^{-4} mm, and the value of *N* can be sufficiently increased, as the circular fringes can be obtained up to large path differences.

SAQ 1

When the movable mirror of Michelson's interferometer is shifted through 0.0589 mm, a shift of 200 fringes is observed. What is the wavelength of light used? Give the answer in Angstrom units.

(b) **Determination of difference in Wavelength:** When the source of light has two wavelengths λ_1 and λ_2 very close together (like D_1 and D_2 lines of sodium), each wavelength produces its own system of rings. Let $\lambda_1 > \lambda_2$. When the thickness of the film is small, the rings due to λ_1 and λ_2 almost coincide, since λ_1 and λ_2 are nearly equal. The mirror M_2 is moved away. Then, due to different spacing between the rings of λ_1 and λ_2, the rings of λ_1 are gradually separated from those of λ_2. When the thickness of the air-film becomes such that dark rings of λ_1 coincide with bright rings of λ_2 (due to the closeness of λ_1 and λ_2, the dark rings due to λ_1 will practically coincide with bright rings due to λ_2 in the entire field of view), the rings have maximum indistinctness.

The mirror M_2 is moved further away through a distance, say, x until the rings, after becoming most distinct, once again become most indistinct. Clearly, during this movement, n fringes of λ_1 and (n + 1) fringes of λ_2 have appeared at the centre (because then the dark rings of λ_1 will again coincide with the bright rings of λ_2). Now, since the movement of the mirror M_2 by λ_2 results in the appearance of one new fringe at the centre, we have

$$x = n\frac{\lambda 1_1}{2} = (n+1)\frac{\lambda_2}{2}$$

or

$$n = \frac{2x}{\lambda_1} \text{ and } (n+1) = \frac{2x}{\lambda_2}$$

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

or $$\frac{2x(\lambda_1 - \lambda_2)}{\lambda_1 + \lambda_2} = 1$$

or $$\lambda_1 - \lambda_2 = \frac{\lambda_1\lambda_2}{2x}$$

Since λ_1 and λ_2 are close together, $\lambda_1\lambda_2$ can be replaced by λ^2, where λ is the mean of λ_1 and λ_2

$$\therefore \qquad \lambda_1 - \lambda_2 = \frac{\lambda^2}{2x} \tag{7.6}$$

Thus if we measure the distance moved by M_2 between two consecutive positions of disappearance of the fringe pattern and the mean wavelength is known, we can determine the difference $(\lambda_1 - \lambda_2)$.

SAQ 2

In Michelson's interferometer, the reading for a pair of maximum indistinctness were found to be 0.6939 mm and 0.9884 mm. If the mean wavelength of the two components of light be 5893Å, deduce the difference between the wavelengths of the components.

(c) Determination of Refractive Index of a Thin Plate

If a thickness t of a substance having an index of refraction μ is introduced into the path of one of the interfering beams in the interferometer, the optical path in this beam is increased because of the fact that light travels more slowly in the substance, and consequently, has a shorter wavelength. The optical path is now μt through the medium, whereas it was practically t

through the corresponding thickness of air (μ = 1). Thus, the increase in the optical path due to insertion of the substance is $(\mu - 1)t$.

In practice, the insertion of a plate of glass in one of the beams produces a discontinuous shift of the fringes so that the number of fringes cannot be counted. With monochromatic fringes, it is impossible to tell which fringe in the displaced set corresponds to one in the original set. With white light, the displacement in the fringes of different colours is very different. This illustrates the necessity of adjusting the interferometer to produce straight white light fringes. After having adjusted so, the cross-wire is set on the achromatic fringe, which is perfectly straight. The given plate is now inserted in the path of one of the interfering waves. This increases the optical path of the beam by $(\mu - 1)t$. Since the beam traverses the plate twice, an extra path difference of $2(\mu - 1)t$ is introduced between the two interfering beams. The fringes get shifted. The movable mirror M_2 is moved till the fringes are brought back to their initial positions so that the achromatic fringe again coincides with the cross wire. If the displacement of M_2 is *x*. then

$$2x = 2(\mu - 1)t$$

or $$x = (\mu - 1)t \tag{3.7}$$

Alternatively, if *N* be the number of fringes shifted then

$$2(\mu - 1)t = N\lambda \tag{3.8}$$

Thus, after measuring *x, t,* we may calculate λ may be calculated if μ is known, or μ may be calculated if *t* is known.

This method can be used to find the refractive index of a gas. The gas is introduced into an evacuated tube placed along the axis of one of the interfering beams, and the experiment is carried out as described above.

SAQ3
A transparent film of glass of refractive index 1.50 is introduced normally in the path of one of the interfering beams of a Michelson's interferometer, which is illuminated with light of wavelength 4800Å. This causes 500 dark fringes to sweep across this field. Determine the thickness of the film.

There is yet another type of interferometer, called the Fabry-Perot interferometer, which produces fringes much sharper than those produced by Michelson interferometer. In the next section, let us study this interferometer and see how it is used as a powerful spectrometer.

3.3 FABRY-PEROT INTERFEROMETER

It is based on the principle of multiple beam interference. It is a high resolving power instrument, which makes use of the 'fringes of constant inclination' produced by the transmitted light after multiple reflections between two parallel and highly-reflecting glass plates.

It consists of two optically-plane glass plates *A* and *B* (Fig. 3.9) with plane surfaces. The inner surfaces are coated with partially transparent films of high reflectivity and placed accurately parallel to each other. Screws are provided to secure parallelism if disturbed. The two uncoated surfaces of each plate are made to have a slight angle between them in order to avoid unwanted fringes formed due to multiple reflections in the plate itself.

One of the two plates is kept fixed, while the other can be moved to vary the separation of the two plates. In this configuration, the instrument is called a Fabry-Perot interferometer. Sometimes both the plates are at a fixed separation with the help of spacers. The system with fixed spacing is known as Fabry-Perot etalon. The Fabry-Perot interferometer (or etalon) is used to determine wavelengths precisely, to compare two wavelengths, to calibrate the standard metre in terms of wavelength, etc.

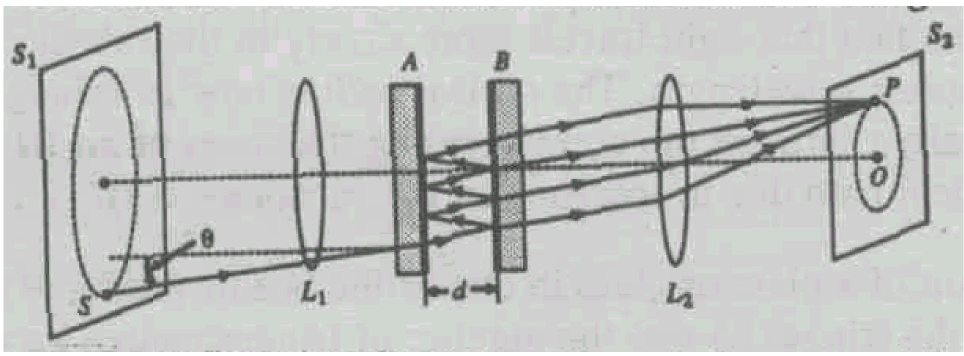


Fig. 3.9 Fabry-Perot interferometer. *S* is part of an external light source

S_1 is a broad source of monochromatic light and L_1 a convex lens which makes the beam more collimated. An incident ray suffers a large number of internal reflections successively at the two silvered surfaces, as shown. At each reflection a small fractional part of the light is also transmitted. Thus, each incident ray produces a group of coherent and parallel transmitted rays with a constant path difference between any two successive rays. A second convex lens, L_2, brings these rays together at a point *P* in its focal plane, where they interfere. Hence, the rays from all points of the source produce an interference pattern on a screen S_2 placed in the focal plane of L_2.

Formation of the Fringes: Let *d* be the separation between the two silvered surfaces, and θ the inclination of a particular ray with the normal to the plates. Then the path difference between any two successive transmitted rays corresponding to the incident ray is $2d\cos\theta$. The medium between the two silvered surfaces is usually air. As you saw, while solving SAQ 1 in Unit 2, that π phase changes occur on both of these (air-to-glass) surfaces, hence, the condition

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

holds for maximum intensity.

Here, *n* is an integer, called the order of interference, and λ the wavelength of light. The locus of points in the source which give rays of a constant inclination θ is a circle. Hence, with an extended source, the interference pattern consists of a system of bright concentric rings on a dark background, each ring corresponding to a particular value of θ. Fig. 3.4(b) shows the fringes obtained using a Fabry-Perot interferometer. Also shown, in the figure for comparison, are fringes obtained by using Michelson interferometer (see Fig. 3.4a). It can readily be seen that the Fabry-Perot interferometer, which employs the principle of multiple beam interference, produces much sharper fringes, and could, hence, be used to study hyperfine structure of spectral lines. The intensity distribution of the circular fringes of Fig. 3.4b is not in accordance with Eq. (1.17). To determine how much light is reflected and transmitted at the two surfaces, let us read the following section.

3.3.1 Intensity Distribution

Comment: You are advised to go through the Appendix given at the end of this unit carefully.

We return now to the problem of reflections from a parallel plate, already considered in a two-beam approximation in Unit 2. Fig. 3.10 shows the multiple reflections and transmissions through a plane parallel plate of "air" enclosed between two glass plates of Fabry-Perot interferometer. Here, *n'* is the refractive index of the glass plate and *n* the refractive index of the air enclosed. Suppose a wave is incident at an angle θ, as shown in Fig. 3.10. This incident wave will suffer multiple reflections. Let the reflection and transmission amplitude co-efficient be *r* and *t* at an external reflection and *r'* and *t'* at an internal reflection.

If the amplitude of the incident ray is expressed as $ae^{i\omega t}$, the successive transmitted rays can be expressed by appropriately modifying both the amplitude and phase of the initial wave. Referring to Fig. 3.10, these are

$$A_1 = (tt'a)e^{i\omega t}$$
$$A_2 = (tt'r'^2\, a)e^{i(\omega t-\delta)}$$
$$A_3 = (tt'r'^4\, a)e^{i(\omega t-2\delta)}$$

A little inspection of these equations shows that

$$A_N = tt'r'^{2(N-1)}\, ae^{i\omega t}e^{-i(N-1)\delta}$$

The quantities r, r', t, t', are given in terms of *n, n'*, θ, θ' by the Fresnel formulae. For our present purpose we do not need these explicit expressions but only relations between them. We have

$$tt' = T \tag{3.9a}$$

and $$r^2 = r'^2 = R \tag{3.9b}$$

where *R* and *T*, respectively are the reflectivity and transmissivity of the plate surfaces. Then, using Eq. 3.9, we have

$$A_1 = aTe^{i\omega t}$$
$$A_2 = aT\,\mathrm{Re}^{i(\omega t-\delta)}$$
$$A_3 = aTR^2e^{i(\omega t-2\delta)}\text{, and so on.}$$

By the principle of superposition, the resultant amplitude is given by

$$A = aT + aTRe^{i\delta} + aTR^2e^{2i\delta} + aTR^3e^{3i\delta} + \ldots$$

Here, we have ignored $e^{i\omega t}$, as it is of no importance in combining waves of the same frequency. Hence,

$$A = aT(1 + Re^{-i\delta} + R^2 e^{-2i\delta} + R^3 e^{-3i\delta} + ...)$$

The infinite geometric series in the parentheses has the common ratio $\mathrm{Re}^{i\delta}$ and has a finite sum because $r^2 < 1$. Summing up the series, we obtain

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

The complex conjugate of A is therefore

$$A^* = aT \frac{1}{1 - Re^{i\delta}}$$

Hence the resultant intensity I is given by

$$I = AA^* = \frac{a^2 T^2}{(1 - Re^{-i\delta})(1 - Re^{i\delta})}$$

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

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

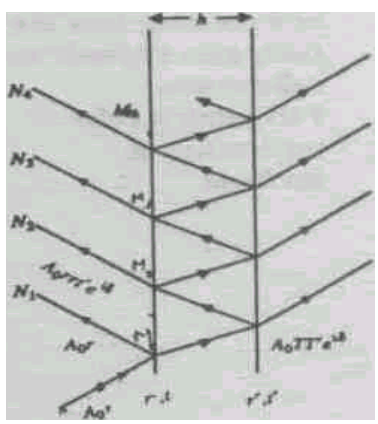


Fig. 3.10 Multiple reflection in a parallel "air" plate enclosed between the two plates of Fabry-Perot interferometer

$$\frac{a^2T^2}{(1-R)^2}\frac{1}{1+\frac{4R}{(1-R)^2}\sin^2\frac{\delta}{2}} \tag{3.10}$$

The intensity will be a maximum when $\sin^2\frac{\delta}{2}=0$, i.e., $\delta = 2n\pi$, where n = 0, 1, 2, … Thus

$$I_{\max} = \frac{a^2T^2}{(1-R)^2} \tag{3.11}$$

Similarly, the intensity will be a minimum when $\sin^2\frac{\delta}{2}=1$, i.e., $\delta = (2n+1)\pi$, where n = 0, 1, 2, … Thus

$$I_{\min} = \frac{a^2T^2}{(1-R)^2}\frac{1}{1+\frac{4R}{(1-R)^2}} = \frac{a^2T^2}{(1+R)^2} \tag{3.12}$$

Eq. (3.10) can now be written as

$$I = \frac{I_{\max}}{1+\frac{4R}{(1-R)^2}\sin^2\frac{\delta}{2}} \tag{3.13}$$

or

$$I = \frac{I_{\max}}{1+F\sin^2\frac{\delta}{2}} \tag{3.14}$$

Here, $F = \frac{4R}{(1-R)^2}$ is called the coefficient of finesse. Eq. (3.14) is the intensity expression for the Fabry-Perot fringes.

If we plot I against δ for different values of R (the reflectivity of the plates), a set of curves is obtained (Fig. 3.11). They show that the larger the value of R, the more rapid is the fall of intensity on either side of a maximum. (That is, the higher the reflectivity of the plates, the sharper the interference bright fringes.) Further, as Eq. (3.11) and (3.12) show, the larger the value of R, the greater is the difference between $I_{\max}$ and $I_{\min}$. In fact, we obtain a system of sharp and bright rings against a wide dark background.

As mentioned in the beginning of the sec. 3.3, Fabry-Perot interferometer is a high resolving power instrument. Its resolving power $\frac{\lambda}{\Delta\lambda}$ is given by

$$\frac{\lambda}{\Delta\lambda} = \frac{4\pi\, h \cos r \sqrt{F}}{4.147\lambda}$$

where h is the thickness of the film enclosed between the two silvered surfaces, r is the angle of refraction inside the film, λ the wavelength of the incident light and F is the coefficient of finesse.

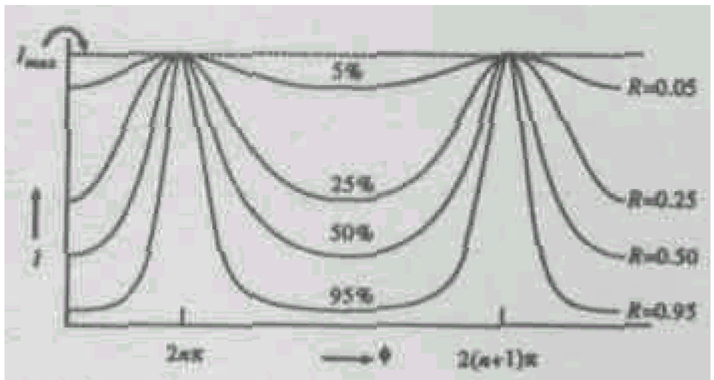


Fig. 3.11: The transmitted intensity as a function of δ showing how the sharpness depends on reflectance. Percentages refer to reflectance of surfaces

To have an idea of the numerical value of resolving power, let us consider a Fabry-Perot etalon with h = 1 cm and F = 80. The resolving power for normal incidence in the wavelength region around λ = 5000 Å would be

$$\frac{\lambda}{\Delta\lambda} = \frac{4\pi\sqrt{80}}{5\times10^{-5}\times4.147} = 5.42\times10^{5}$$

That is, two wavelengths separated by 0.0092 Å can be resolved at λ = 5000 Å

3.3.2 Superiority over Michelson's Interferometer

When the light consists of two or more close wavelengths (such as the D_1 and D_2 lines of sodium), then in a Fabry-Perot interferometer each wavelength produces its own pattern, and the rings of one pattern are clearly separated from the corresponding rings of the other pattern. Hence the instrument is very suitable for the study of the fine structure of spectral lines. In Michelson's instrument separate patterns are not produced. The presence of two close wavelengths is judged by the alternate distinctness and indistinctness of the rings when the optical path difference is increased.

3.4 SUMMARY

- The Michelson interferometer uses an extended monochromatic source.

- When M_1 and M_2 are perpendicular to each other, i.e., when M_1 and M_2 are parallel, the fringes given by a monochromatic source are circular and localized at infinity.

- When the mirrors of the interferometer are inclined with respect to each other, i.e., when M_1 and M_2 are not perpendicular to each other, a pattern of straight parallel fringes is obtained.

- Whether M_1 and M_2 are parallel or inclined, any fringe shift seen in an interferometer may be due to either a change in thickness or a change in refractive index.

- As the movable mirror is displaced by $\frac{\lambda}{2}$, each fringe will move to the position previously occupied by an adjacent fringe. If N is the number of fringes that have moved past a reference point, when the mirror is moved a distance x, then

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

- Michelson interferometer can be used in the measurement of two closely spaced wavelengths.

- Fabry-Perot interferometer, which employs the principle of multiple beam interference, produces much sharper fringes than those produced by Michelson interferometer.

- In the Fabry-Perol interferometer it is the fringe pattern formed by transmitted light that is observed and as such that intensity distribution would be given by

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

- The resolving power of Fabry-Perot interferometer is given by

$$\frac{\lambda}{\Delta\lambda} = \frac{4\pi h \cos r\sqrt{F}}{4.147\lambda}$$

3.5 TERMINAL QUESTIONS

(1) When one leg of a Michelson interferometer is lengthened slightly, 150 dark fringes sweep through the field of view. If the light used has λ = 480 mm, how far was the mirror in that leg moved?

(2) Circular fringes are observed in a Michelson interferometer illuminated with light of wavelength 5896 Å. When the path difference between the mirrors M_1 and M_2 is 0.3 cm, the central fringe is bright. Calculate the angular diameter of the 7th bright fringe.

3.6 SOLUTIONS AND ANSWERS

SAQs

(1) The distance, x, moved by the mirror when N fringes cross the field of view is given by

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

$$\therefore \qquad \lambda = \frac{2x}{N}$$

Here, x = 0.00589 cm, and N =200.

$$\therefore \qquad \lambda = \frac{2 \times 0.00589}{200} = 0.0000589 = 5890 \text{ Å}$$

(2) If x be the distance moved by the movable mirror between two consecutive positions of maximum indistinctness (or distinctness), we have

$$\Delta\lambda = \frac{\lambda_1 \times \lambda_2}{2x} = \frac{\lambda^2}{2x}$$

where λ is the average of λ_1 and λ_2.

Here λ = 5893 Å = 5893×10^{-8} cm and x = 0.9884 – 0.6939 = 0.2945 mm = 0.02945 cm.

$$\therefore \qquad \Delta\lambda = \frac{(5893 \times 10^{-8})^2}{2 \times 0.02945} = 5.896 \times 10^{-8} \text{ cm} = 5.896 \text{ Å}$$

(3) Let t be the thickness of the film. When it is put in the path of one of the interfering beams of the Michelson's interferometer, an additional path difference of $2(\mu - 1)t$ is introduced. If N be the number of fringes shifted, we have

$$2(\mu - 1)t = N\lambda$$

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

Here, N = 500, λ = 4800×10^{-8} cm, μ = 1.50.

$$\therefore \qquad t = \frac{500 \times 4800 \times 10^{-8}}{2(1.50 - 1)}$$

$$= \frac{500 \times 4800 \times 10^{-8}}{2 \times 0.50}$$

$$= 0.024 \text{ cm}$$

TQs

(1) Darkness is observed when the light beams from the two legs are 180° out of phase. As the length of one leg is increased by $\frac{\lambda}{2}$, the path length increases by λ, and the field of view changes from dark to bright to dark. When 150 fringes pass, the leg is lengthened by an amount

$$(150)\left(\frac{\lambda}{2}\right) = (150)\ (240) = 36{,}000 \text{ nm} = 0.036 \text{ mm}$$

(2) The expression for the bright circular fringe is

$$2d\cos r = n\lambda$$

At the centre r = 0, so that

$$2d = n\lambda \qquad \text{(i)}$$

n now stands for the order of the central bright fringe. The order of fringes decreases as we move outwards from the centre. Thus the second bright fringe is of $(n-1)$ th order, ..., the seventh bright fringe is of $(n-6)$ th order. Hence if θ be the angular radius of the 7th bright fringe, we have

$$2d\cos\theta = (n-6)\lambda \qquad \text{(ii)}$$

Eq. (i) and (ii) give

$$2d(1-\cos\theta) = 6\lambda$$

or

$$\cos\theta = 1 - \frac{6\lambda}{2d}$$

Putting the given values:

$$\cos\theta = 1 - \frac{6\times(5896\times10^{-8}\,cm)}{2\times 0.3\,cm}$$

$$= 1 - 0.0005896 = 0.9994$$

$\therefore \quad \theta = \cos^{-1}(0.9994) = 2^0$

$\therefore \quad$ angular diameter = 4^0

3.7 APPENDIX

Method of Complex Amplitudes

In place of using the sine or the cosine to represent a simple harmonic wave, one may write the equation in the exponential form as

$$y = ae^{i(\omega t - kx)} = ae^{i\omega t}e^{-i\delta}$$

where $\delta = kx$ is constant at a particular point in space and represents the phase of the wave. The presence of $i = \sqrt{-1}$ in this equation makes the quantities complex. We can nevertheless use this representation, and at the end of the problem take either the real (cosine) or the imaginary (sine) part of the resulting expression. The time-varying factor $\exp(i\omega t)$ is of no importance in combining waves of the same frequency, since the amplitudes and relative phases are independent of time. The other factor, $a\exp(-i\delta)$ is called the complex amplitude. It is a complex number whose modulus *a* is the real amplitude, and whose argument δ gives the phase relative to some standard phase. Negative sign merely indicates that the phase is behind the standard phase. In general, the vector *a* is given by

$$a = ae^{i\delta} = x + iy = a(\cos\delta + i\sin\delta)$$

Then it will be seen that

$$a = \sqrt{x^2 + y^2}\,,\ \tan\delta = \frac{y}{x}$$

Thus, if *a* is represented as in Fig. (3.12), plotting horizontally its real part and vertically its imaginary part, it will have the magnitude *a* and will make the angle δ with the *x* axis, as we require for vector addition.

The advantage of using complex amplitudes lies in the fact that the vector addition of real amplitudes can be written more easily in the form of an algebraic addition of complex amplitudes. For example, consider the real parts of two waves that follow the equations

$$\mathbf{A}_1 = A_1 e^{i(\omega t + \delta_1)}$$

and

$$\mathbf{A}_2 = A_2 e^{i(\omega t + \delta_2)} \tag{3.15}$$

Adding these two equations gives

$$\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2 = A_1 e^{i(\omega t + \delta_1)} + A_2 e^{i(\omega t + \delta_2)} \tag{3.16}$$

We can now take out the common exponent $i\omega t$:

$$\mathbf{A} = e^{i\omega t}(A_1 e^{i\delta_1} + A_2 e^{i\delta_2}) \tag{3.17}$$

The square of the resultant, A^2, is found by multiplying the complex terms by their complex conjugates:

$$\begin{aligned} A^2 &= (A_1 e^{i\delta_1} + A_2 e^{i\delta_2})(A_1 e^{-i\delta_1} + A_2 e^{-i\delta_2}) \\ &= A_1^2 + A_2^2 + A_1 A_2 (e^{i(\delta_1 - \delta_2)} - e^{-i(\delta_1 - \delta_2}) \end{aligned} \tag{3.18}$$

Then, from Euler's formula,

$$e^{i\delta} + e^{-i\delta} = \cos\delta + i\sin\delta + \cos\delta - i\sin\delta = 2\cos\delta \tag{3.19}$$

and therefore, Eq. (3.18) becomes

$$A^2 = A_1^2 + A_2^2 + 2A_1A_2\cos(\delta_1 - \delta_2) \tag{3.20}$$

the same as Eq. (1.15).

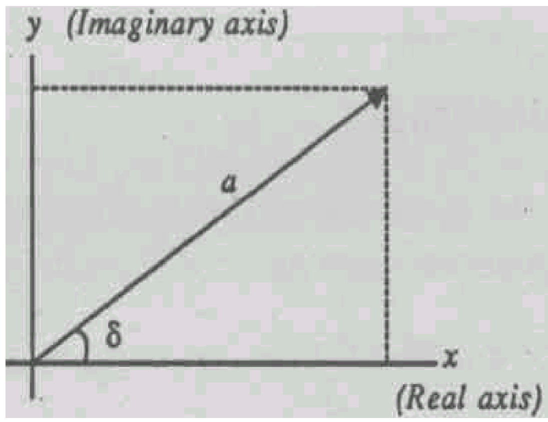


Fig. 3.12 Representation of a vector in the complex plane

Thus, in obtaining the resultant intensity as proportional to the square of the real amplitude, we multiply the resultant complex amplitude by its complex conjugate, which is the same expression with i replaced by $-i$ throughout.

UNIT 4 FRESNEL DIFFRACTION

Structure

4.1 INTRODUCTION

We know from our day-to-day experience that we can hear persons talking in an adjoining room the door of which is open. This is due to the ability of sound waves to bend around the corners of obstacles in their way. You are also familiar with the ability of water waves to propagate around obstacles. You may now ask: Does light, which is an electromagnetic wave, also bend around corners of obstacles in its path? Earlier in this course you have learnt the manifestation of wave nature of light in the form of interference: Light from two coherent sources interferes to form fringed pattern. But what may puzzle you is the fact that light casts shadows of objects, i.e. appears to travel in straight lines rather than bending around corners. This apparent contradiction was explained by Fresnel who showed that the ease with which a wave bends around corners is strongly influenced by the size of the obstacle (aperture) relative to its wavelength. Music and speech wavelengths lie in the range 1.7 cm to 17m. A door is about 1 m aperture so that long wavelength waves bend more readily around the door way. On the other hand, wavelength of light is about 10^{-7} m and the obstacles used in ordinary experiments are about 10^{5} times bigger. For this reason, light appears to travel along straight lines and casts shadows of objects instead of bending around their corners. However, it does not mean that light shows no bending, it does so under suitable conditions where size of obstacles is comparable with the wavelength of light. You can get a feel for this by closely examining shadows cast by objects. You will observe that the edges of shadows are not sharp. **The deviation of waves from their original direction due to an obstruction in their path is called diffraction.**

The phenomenon of diffraction finds great use in our daily life. You will learn that diffraction places a fundamental restriction on optical instruments, including the human eye, in respect of resolution of objects.

The phenomenon of diffraction was first observed by Grimaldi, an Italian mathematician. And a systematic explanation of diffraction was given by Fresnel on the basis of Huygens' principle. According to him, diffraction is attributed to mutual interference of secondary wavelets from a single wave. (The interference phenomenon involves two coherent wave trains.) This means that in diffraction phenomenon, interference takes place between secondary wavelets from different parts of the same wavefront.

Fraunhofer diffraction and Fresnel diffraction are also called far field diffraction and near field diffraction, respectively.

For mathematical convenience and ease of understanding, diffraction is classified in two categories: Fraunhofer diffraction and Fresnel diffraction. In Fraunhofer class of diffraction, the source of light and the observation screen (or human eye) are effectively at infinite distance from the obstacle. This can be done most conveniently using suitable lenses. It is of particular practical importance in respect of the general theory of optical instruments. You will learn about it in the next unit.

In Fresnel class of diffraction, the source or the observation screen or both are at finite distance from the obstacle. You will recognise that for Fresnel diffraction, the experimental arrangement is fairly simple. But its theoretical analysis is more difficult than that of Fraunhofer diffraction. Also, Fresnel diffraction is more general; it includes Fraunhofer diffraction as a special case. Moreover, it has importance in historical perspective in that it led to the development of the wave model of light. You will learn some of these details in this unit.

Objectives

After studying this unit you will be able to:

- state simple experiments which illustrate diffraction phenomenon
- describe an experimental set-up for diffraction at a circular aperture
- explain that Fraunhofer diffraction is a special case of Fresnel diffraction
- discuss the concept of Fresnel half-period zones and apply it to zone plate
- discuss diffraction pattern due to a circular aperture and a straight edge, and
- solve numerical problems.

4.2 OBSERVING DIFFRACTION: SOME SIMPLE EXPERIMENTS

As you know, the wavelength of visible light is very small (about 10^{-7} m). And to see diffraction, careful observations have to be made. We will now familiarise you with some simple situations and experiments to observe diffraction of light. The prerequisites for these are: (i) a source of light, preferably narrow and monochromatic, (ii) a sharp edged obstacle and (iii) an observation screen, which could be the human retina as well.

1. Look at a distant street light at night and squint. The light appears to streak out from the bulb. This is because light has bent around the corners of your eyelids.

2. Stand in a dark room and look at a distant light bulb in another room. Now move slowly until the doorway blocks half of the light bulb. The light appears to streak out into the umbra region of the dark room due to diffraction around the doorway.

3. Take a piece of fine cloth, say fine handkerchief or muslin cloth. Stretch it flat and keep it close to the eye. Now focus your eye on a distant lamp (at least 100 m away) through it. Do you observe an enlarged disc surrounded by a regular pattern of spots arranged along a rectangle? On careful examination you will note that the spots on the outer part of the pattern appear coloured. Now rotate the handkerchief in its own plane. Does the pattern rotate? You will he excited to see that the pattern rotates about the central disc. Moreover, the speed of rotation of the pattern is same as that of the handkerchief.

We are now tempted to ask: Do you know why this pattern of spots is obtained? You will agree that the handkerchief is a mesh (criss-cross) of fine threads in mutually perpendicular directions. Obviously, the observed pattern is formed by the diffraction of light from the lamp.

4. Take a pair of razor blades and one clear glass electric bulb. Hold the blades so that the edges are parallel and have a narrow slit in between, as shown in Fig. 4.1. Keep the slit close to your eye and parallel to the filament. (Use spectacles if you normally do.) By slightly adjusting the width of the slit, you should observe a pattern of bright and dark bands, which show some colours. Now use a blue or red filter. What do you observe? Does the pattern become clearer?

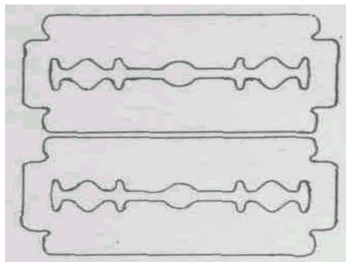


Fig. 4.1 Observing diffraction using a pair of razor blades

5. Mount a small ball bearing carefully on a plate of glass with a small amount of beeswax so that no wax spreads beyond the rim of the ball. Place this opaque obstacle in a strong beam of light (preferably monochromatic) diverging from a pinhole. Under suitable conditions, you will observe a bright spot, called **Poisson spot** at the centre of the shadow cast by the ball bearing. This exciting observation proved unchallengeable evidence for diffraction of light.

4.3 PRODUCING A DIFFRACTION PATTERN

In the Fresnel class of diffraction, the source of light or the screen or both are, in general, at a finite distance from the diffracting obstacle. On the other hand in Fraunhofer diffraction, this distance is effectively infinite. This condition is achieved by putting a suitable lens between the source and the screen. A large number of workers have observed and studied Fresnel and Fraunhofer diffraction patterns. Recently a systematic study of Fresnel diffraction pattern from obstacles of different shapes e.g. small spheres, discs and apertures of circular, elliptical, square, triangular or parallelograms etc of different sixes was done by Indian physicist Y.V. Kathvate under the guidance of Prof. C. V, Raman. Their experimental set up for photographing these patterns is shown in Fig. 4.2. It consists of a light tight box (nearly 5 m long) with a fine pinhole at one end. The light on the pinhole from a 100 W lamp was focussed using a convex lens, A red filter was used to obtain almost monochromatic light of wavelength 6320 Å.

The obstacle was placed at a suitable distance (about 2 m) from the pinhole. The photographic plate was mounted on a movable stand so that its distance from the obstacle could be varied. They used steel ball bearings of radii 1.58 mm, 1.98 mm, 2.37mm and 3.17 mm as spherical obstacles.

(As such, you should not attach much significance to the exactness of these sizes.) These four spheres were mounted on a glass plate, which was kept at a distance of about 2 m from the pinhole.

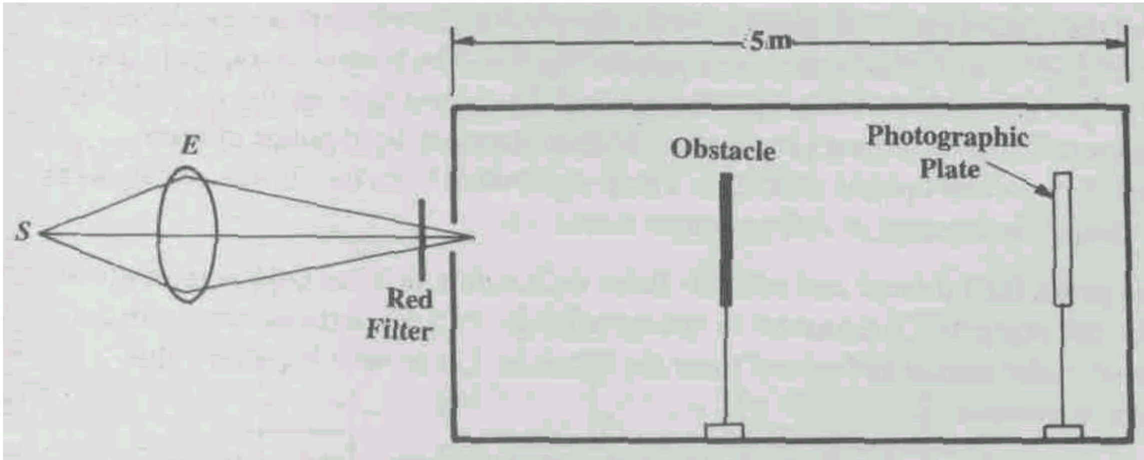
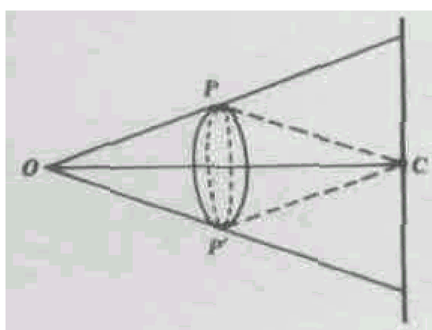


Fig. 4.2 Schematics of experimental arrangement used by Kathvate to observe Fresnel diffraction

Poisson was a member of the committee, which was appointed to judge Fresnel's dissertation. To disprove Fresnel, and hence the wave theory, Poisson argued that a central bright spot should appear in the shadow of a circular obstacle. His logic, called reductio ad absurdum goes as follows: Consider the shadow of a perfectly round object being cast by a point source (O) shown below. According to wave theory, all the waves at the periphery will be in phase. This is because they have covered the same distance from the source. So the waves starting from the rim PP' and reaching C should all be in phase at the centre of the shadow. This implies that there should be a bright spot at the centre of the shadow. This was considered absurd by Poisson; he was definitely not aware that the bright spot in question had already been discovered by Maraldi almost a century before. Soon after Poisson's objection, Arago carried out the experiment using a disk of 2mm diameter. To his surprise, he rediscovered the central bright spot.



The photographic plate was kept at distances of 5cm, 10cm, 20cm, 40cm and 180 cm from the mounted glass plate (obstacle). For the last case, the diffraction patterns obtained from these spheres are shown in Fig. 4.3 (a). These patterns essentially characterize the distribution of light intensity in the region of geometrical shadow of the obstacles.

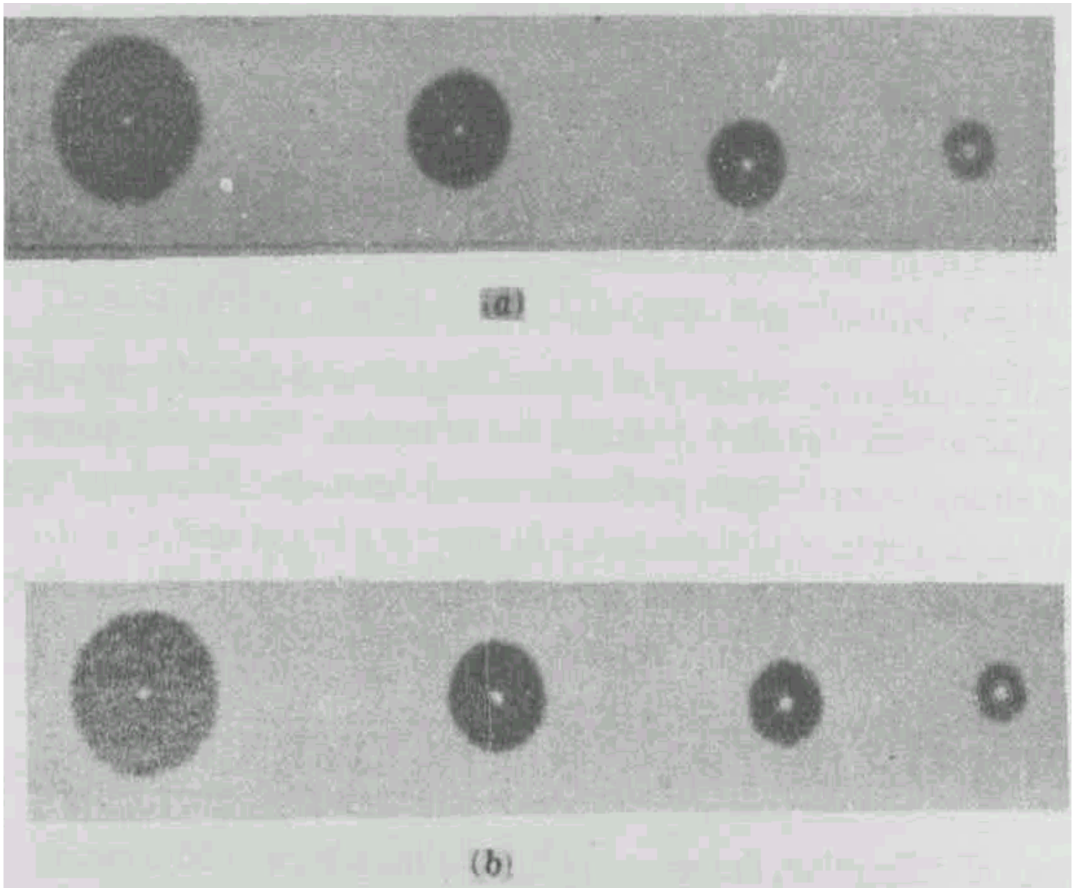


Fig. 4.3 Fresnel diffraction patterns: Kathvate experiments with (a) spheres and (b) circular discs of four sizes

The diffraction patterns for circular discs of the same size are illustrated in Fig. 4.3(b). You will find that these patterns are almost similar to those for spheres. Moreover, the diffraction patterns on the left half of this figure, which correspond to bigger spheres and discs (radii 3.17 mm and 2.37 mm), show the geometrical shadow and a central bright spot within it. On the other hand, in the diffraction pattern corresponding to the smaller sphere (or disc) of radius 1.98mm, the geometrical image is recognizable but has fringes appearing on its edges. The fringe pattern around the central spot becomes markedly clearer for the sphere of radius 1.58mm. An enlarged view of this pattern is shown in Fig. 4.4. The formation of the bright central spot in the shadow and the rings around the central spot are the most definite indicators of non-rectilinear propagation of light. Instead, light bends in some special way around opaque obstacles. These departures from rectilinear propagation come under the heading of diffraction phenomenon.

Let us pause for a minute and ask: Are these diffraction patterns unique for a given source and obstacle? The answer to this question is: Fresnel patterns vary with the distance of the source and screen from the obstacle. Let us now learn how this transition evolves.

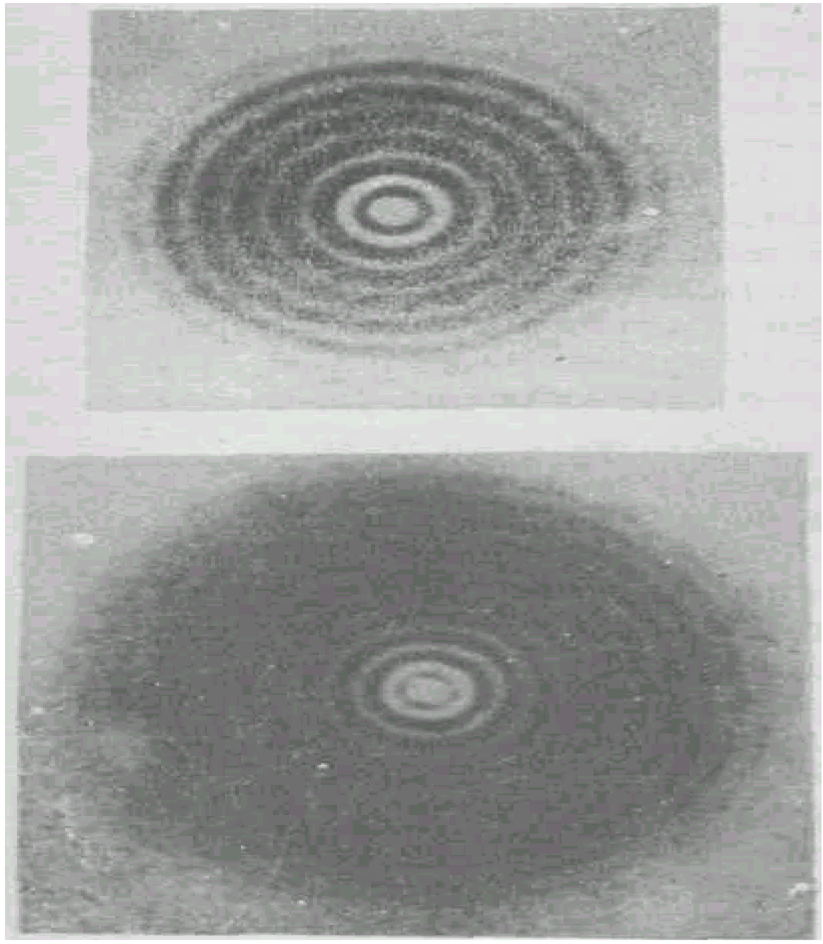


Fig. 4.4 Enlarged view of fringe pattern for the sphere of radius 1.58mm

4.3.1 Spatial Evolution of a Diffraction Pattern: Transition from Fresnel to Fraunhofer Class

To observe transition in the Fresnel diffraction pattern with distance, we have to introduce a small modification in Kathvate's experimental arrangement, as shown in Fig. 8.5 (a). The point source is now located at the focal point of a converging lens *L*. The spherical waves originating from the source *O* are changed into plane waves by this lens and the wavefront is now parallel to the diffracting screen with a narrow opening in the form of a long narrow slit (Fig.8.5 (b)). These waves pass through the slit. The diffracted waves are also plane and may have an angular spread. You may now like to know the shape, size and intensity distribution in the diffraction pattern on a distant screen.

1. When the incident wavefront is strictly parallel to the diffracting screen, we get a vertical patch of light when the screen is immediately behind the aperture. That is, a region $A'B'$ of uniform illumination on the screen. The size of this region is equal to the size of the slit both in width and height. The remaining portion of the screen is absolutely dark. A plot of this intensity distribution is shown in Fig. 4.6 (a). From P to A', the intensity is zero. At A', it abruptly rises to I_0, and remains constant from A' to B'. At B', it again drops to zero. We can say that $A'B'$ represents the edges of the geometrical shadow (and the law of rectilinear propagation holds).

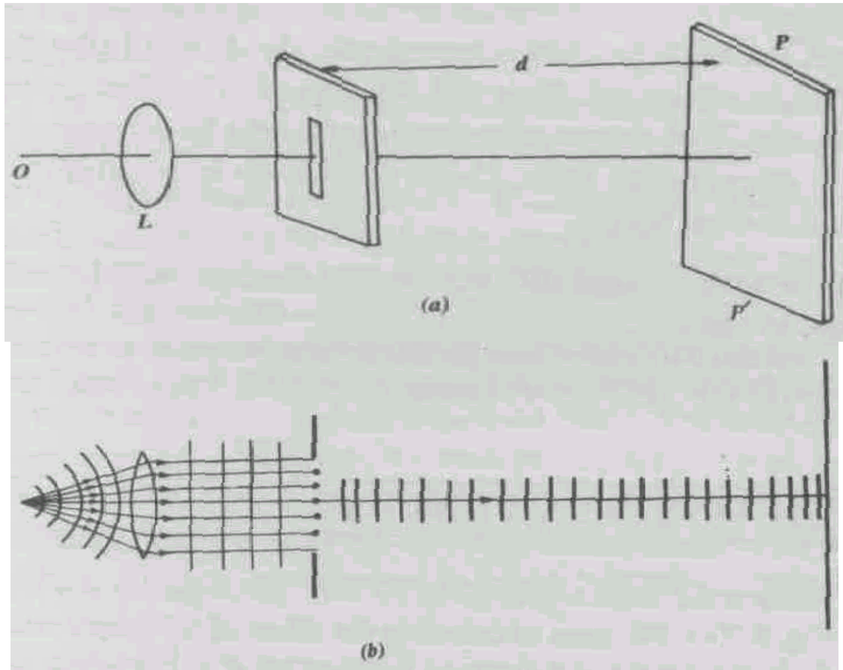


Fig. 4.5(a) Arrangement to observe transition In Kernel diffraction pattern (b) Cross-sectional view of the geometry shown in (a) above

2. As the screen is moved away from the aperture, a careful observation shows that the patch of light seen in (1) above begins to lose sharpness. If the distance between the obstacle and the observation screen is large compared to the width of the slit, some fringes start appearing at the edges of the patch of light. But this patch resembles the shape of the slit. The intensity distribution shows diffraction rippling effect somewhat like that shown in Fig. 8.6(b). From this we can say that the intensity distribution in the pattern depends on the distance at which the observation screen is placed.

A slit is a rectangular opening whose width (0.l mm or so) is much smaller than its length 1 cm or more.

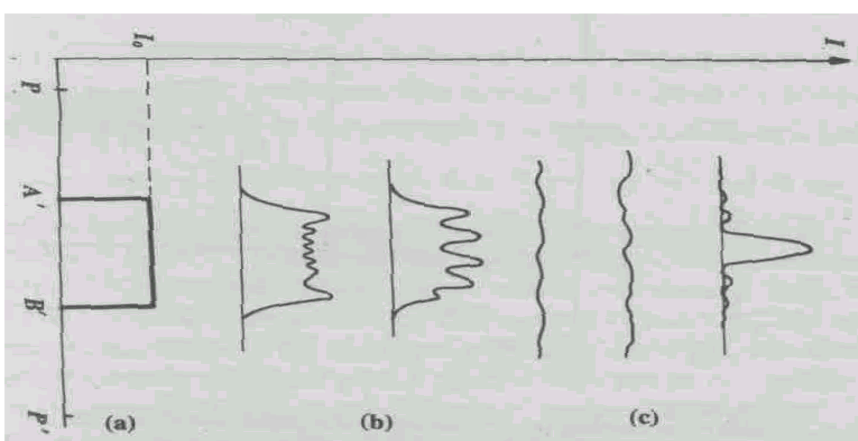


Fig. 4.6 Spatial evolution of a diffraction pattern

3. When *d* (~ 1m) is much greater than the width of the slit (~ 0.1 mm), the fringes seen in (2) above – close to edge of the patch – now spread out and the geometrical image of the slit can no longer be recognized. As distance is increased further, diffraction effects become progressively more pronounced.

4. When *d* is very large, i.e., once we have moved into the Fraunhofer region, ripples no longer change character. You can observe this pattern by putting a convex lens after the slit. The observation screen should be arranged so that it is at the second focal plane of the lens. These variations in Fraunhofer diffraction are shown in Fig. 4.6(c).

From this we may conclude that Fresnel diffraction can change significantly as the distance from the aperture is varied.

You must now be interested in understanding the physical basis of these observations . The first systematic effort in this direction was made by Fresnel. Let us learn about it now.

4.4 FRESNEL CONSTRUCTION

Let us consider a plane wave front represented by *WW* propagating towards the right, as shown in Fig. 4.7(a). We want to calculate the effect of this plane wavefront at an external point P_0 on the screen at a distance *d*. Then we will introduce an obstacle like a straight edge and see how intensity at P_0 changes.

We know that every point on the plane wavefront may be thought of as a source of secondary wavelets. We wish to compute the resultant effect at P_0 by applying Huygens-Fresnel principle. One way would be to write down the equations of vibrations at P_0 due to each wavelet and then add them together. This is a cumbersome proposition. The difficulty in mathematical calculation arises on two counts; (i) There are an infinite number of points which act as sources of secondary wavelets and (ii) Since the distance travelled by the secondary wavelets arriving at P_0 *is* different, they reach the point P_0 with different phases. To get over these difficulties, Fresnel devised a simple geometrical method which provided useful insight and beautiful explanation of diffraction phenomenon from small obstacles. He argued that it is possible to locate a series of points situated at the same distance from P_0 so that all the secondary wavelets originating from them travel the same distance. We can, in particular, find the locus of those points from where the wavelets travel a distance $b+\frac{\lambda}{2}$, $b+\frac{2\lambda}{2}$, $b+\frac{3\lambda}{2}$, and so on.

The Fresnel construction consists of dividing the wavefront into annular spaces enclosed by concentric circles (Fig. 4.7(b)). The net effect at P_0 will be obtained by summing contributions of wavelets from these annular spaces, called half period elements. When an obstacle is inserted in between the wavefront *WW'* and the point P_0, some of these half period elements will be obstructed depending upon the size and shape of the obstacle. The wavelets from the unobstructed parts only will reach P_0 and their resultant can be calculated easily by Fresnel's method. Let us now learn about Fresnel's construction, half period elements and the method of summation of the contributions of secondary wavelets.

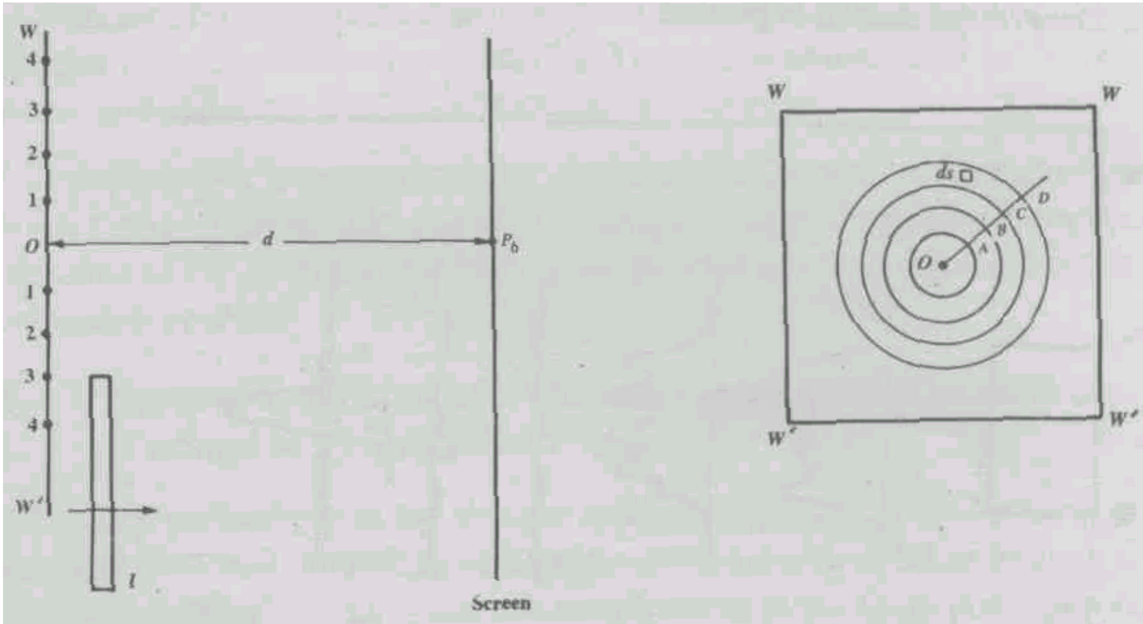


Fig. 4.7 Fresnel construction (a) Propagation of a plane wavefront and (b) division of wavefront into annular spaces enclosed by concentric circles

4.4.1 Half Period Elements

To discuss the concept of Fresnel's half-period elements we assume, for simplicity, that light comes from infinity so that the wavefront passing through the aperture is plane. Refer to Fig. 4.8. It shows a plane wavefront *WWF'F* of monochromatic light propagating along the z-direction. We wish to calculate the resultant amplitude of the field at an arbitrary point P_0 due to superposition of all the secondary Huygen's wavelets originating from the wavefront at the aperture. To do so, we divide the wavefront into half-period zones using the following construction: From the point P_0 we drop a perpendicular $P_0 O$ on the wavefront, which cuts it at *O*. The point *O* is called the pole of the wavefront with respect to the point P_0. Suppose that *b* is the distance between the foot of the perpendicular to P_0, i.e., $OP_0 = b$.

Now with P_0 as centre, we draw spheres of radii $b+\frac{\lambda}{2}$, $b+\frac{2\lambda}{2}$, $b+\frac{3\lambda}{2}$, and so on. You can easily visualise that these spheres will intersect the plane wavefront in a series of concentric circles with centre *O* and radii OQ_1, OQ_2, OQ_3, … as shown in Fig. 4.8. This geometrical construction divides the wavefront into circular strips called zones. The first zone is the space enclosed by the circle of radius OQ_{1t} the second zone is the annular space between the circles of radii OQ_2 and OQ_1. The third zone is annular space between the circles of radii OQ_3 and OQ_2, and so on. These concentric circles or annular rings are called **Fresnel zones** or **half period elements**. This nomenclature has genesis in the fact that the path difference between the wavelets reaching P_0 from corresponding points in successive zones is $\lambda / 2$.

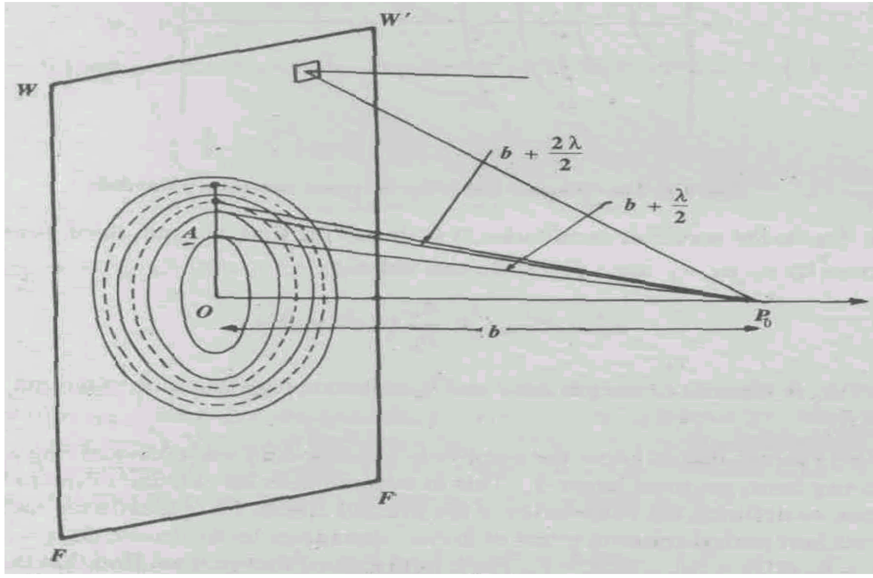


Fig. 4.8 Half-period zones on a plane wavefront: A schematic construction

To compute the resultant amplitude at P_0 due to all the secondary wavelets emanating from the entire wavefront, we first consider an infinitesimal area *dS* of the wavefront. We assure that the amplitude at P_0 due to *dS* is (i) directly proportional to the area *dS* since it determines the number of secondary wavelets, (ii) inversely proportional to the distance of *dS* from P_0 and (iii) directly proportional to the obliquity factor $(1+\cos\theta)$, where θ is the angle between the normal drawn to the wavefront at *dS* and the line joining *dS* to P_0. θ is zero for the central point *O*. As we go away from O, the value of θ increases until it becomes 90° for a point at infinite distance on the wavefront (Fig. 4.9). Physically, it ensures that wavefront moves forward. That is, there is no reverse (or backward) wave.

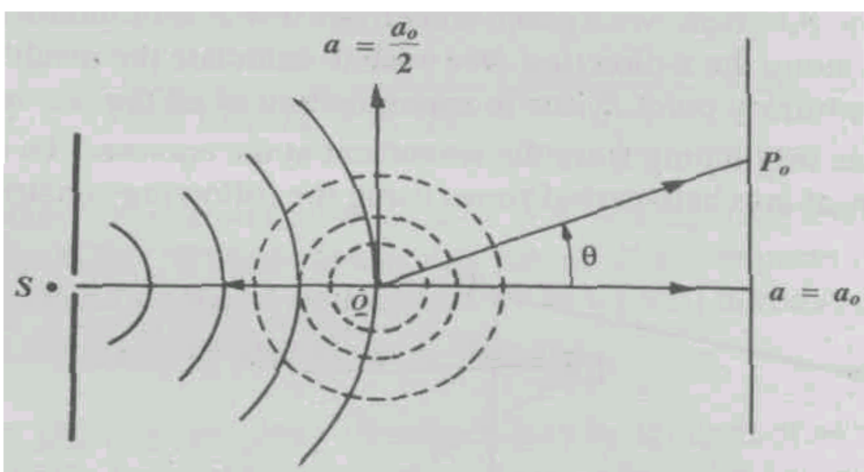


Fig. 4.9 The obliquity factor for Huygens' secondary wavelets

If we denote the resultant amplitudes at P_0 due to the first, second, third, fourth, ..., nth zone by a_1, a_2, a_3, a_4, …, then we can write

$$a_n = \text{const} \times \frac{A_n}{b_n}(1+\cos\theta) \tag{4.1}$$

where A_n is the area of the nth zone and b_n is the average distance of the nth zone from P_0. Eq. (4.1) shows that to know the amplitude of secondary wavelets arriving at P_0 from any zone, we must know A_n. This in turn requires knowledge of the radii of the circles defining the boundaries of the Fresnel zones. To calculate the radii of various half period zones in terms of known distances, let us denote $OQ_1 = r_1$, $OQ_2 = r_2$, $OQ_3 = r_3$, …, $OQ_n = r_n$. From Pythagoras' theorem we find that the radius of the first circle (zone) is given by

$$r_1 = \left[\left(b+\frac{\lambda}{2}\right)^2 - b^2\right]^{1/2} = \sqrt{b\lambda + \frac{1}{4}\lambda^2}$$
$$= \sqrt{b\lambda}$$

The approximation $\lambda << b$ holds for practical systems using visible light. Similarly, the radius of the nth circle (zone) is given by

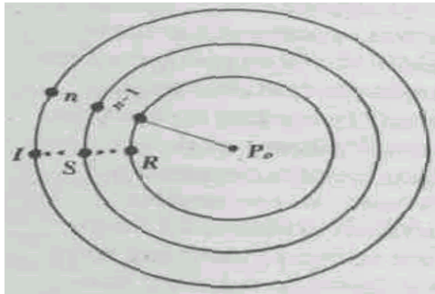
$$\begin{aligned} r_n &= \left[\left(b+\frac{n}{2}\lambda\right)^2 - b^2\right] \\ &= \left[nb\lambda + \frac{n^2\lambda^2}{4}\right]^{1/2} \\ &= \sqrt{nb\lambda} \end{aligned} \tag{4.2}$$

where we have neglected the term $\dfrac{n^2\lambda^2}{4}$ in comparison to $nb\lambda$. This approximation holds for all diffraction problems of interest to us here.

It readily follows from Eqs. (4.1) and (4.2) that the radii of the circles are proportional to the square root of natural numbers, i.e., $\sqrt{1}$, $\sqrt{2}$, $\sqrt{3}$, $\sqrt{4}$, … Therefore, if the first zone has radius r_1, the successive zones have radii 1.41 r_1, 1.73 r_1, 2 r_1, etc. For He-Ne laser light (λ = 6328 Å). If we take P_0 to be 30 cm away (b = 30 cm), the radius of the first zone will 0.436 mm.

Let us now calculate the area of each of the half-period zones. For the first zone

$$A_1 = \pi r_1^2 = \pi\left[\left(b+\frac{\lambda}{2}\right)^2 - b^2\right]$$



Refer to figure above and consider the contributions from the (*n* - 1)th and *n*th zones. Firstly, the areas of the two annular regions are approximately equal, i.e., the secondary wavelets starting from both the zones are equal. Secondly, the points on the innermost circle of the (*n* - 1)th zone, e.g., points like R are situated at a distance of $d+(n-2)\lambda/2$ from P_0, whereas the points on the innermost circle of the nth zone e.g. points like *S* are situated at a distance of $d+(n-1)\lambda/2$ from P_0. The difference in path between the secondary wavelets to reach P_0 from *R* and *S* is $\lambda/2$. This means that the waves reaching P_0 are out of phase by π and cancel each other. Similarly for every point between *R* and *S* in the (*n* - 1)th zone we have a corresponding point between *S* and *T* in the *n*th zone with a path difference of $\lambda/2$ or phase difference of π and hence cancel each other. Since the areas of the two zones are approximately equal, we arrive at the result that **for every point in the (*n* -1)th zone we have a point in the nth zone which is out of phase by π or half of a period.**

$$= \pi b\lambda + \frac{\pi}{4}\lambda^2$$

$$= \pi b\lambda \qquad (4.3a)$$

The area of the second zone, i.e. the annular region between the first and the second circles is

$$\pi(r_2^2 - r_1^2) = \pi[(b+\lambda)^2 - b^2] - \pi b\lambda$$

$$= 2\pi b\lambda - \pi b\lambda = \pi b\lambda \qquad (4.3b)$$

Similarly, you can readily verify that the area of the *n*th zone

$$A_n = \pi(r_n^2 - r_{n-1}^2) = \pi b\lambda \qquad (4.3c)$$

That is, all individual zones have the same area. The physical implication of the equality of zone areas is that the secondary wavelets starting from every zone will be very nearly equal. You must however remember that the result contained in Eq.(4.3) is approximate and is valid for cases where $b >> n\lambda$. A more rigorous calculation shows that the area of a zone gradually increases with *n*:

$$A_n = \pi\lambda\left[b + (n-1)\frac{\lambda}{2}\right] \tag{4.3d}$$

However, the effect of this increase is almost balanced by the increase in the average distance of the *n*th zone from P_0. That is, the ratio A_n / b_n in Eq. (4.1) remains $\pi\ \lambda$, which is a constant, independent of *n*. This means that the amplitude due to any zone will be influenced by the obliquity factor, which is actually responsible for monotonic decrease in the amplitudes of higher zones $(a_1 > a_2 > a_3 > ... > a_n)$. Also, it is important for our computation to note that consecutive zones differ by one-half of a wavelength. Therefore, the secondary waves from any two corresponding points in successive zones [*n*th and (*n* - l)th or (*n* + l)th] reach P_0 out of phase by π or half of a period.

Suppose that the contribution of all the secondary wavelets in the *n*th zone at P_0 is denoted by a_n. Then, the contribution of (*n* - l)th zone a_{n-1} will tend to annihilate the effect of the *n*th zone. Mathematically we write the resultant amplitude at P_0 due to the whole wavefront as a sum of an infinite series whose terms are alternately positive and negative but the magnitude of successive terms gradually diminishes:

$$\xi = a_1 + a_2 e^{i\pi} + a_3 e^{i2\pi} + a_4 e^{i3\pi} + ...$$

$$= a_1 - a_2 + a_3 - a_4 + ... + (-1)^{n+1} a_n + ... = \frac{a_1}{2} \tag{4.4}$$

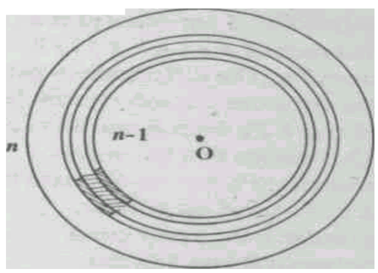
There are several methods of arriving at this result. Here we will describe a simple graphical construction. (The mathematical method is given as TQ).

Let us denote the amplitudes of resultant vectors **AB, CD, EF**, **GH**, ... respectively by a_1, a_2, a_3, a_4, ... due to the first, second, third, fourth, ... zone. (We know that, a_1, a_2, a_3, a_4, ..., a_n are alternately positive and negative). These vectors are shown separately in Fig. 4.10(a) to show their magnitudes and positions. But their true positions are along the same line, as shown in Fig. 4.10 (b). The resultant of the first two zones as will be the small vector **AD**. But the resultant of the first three zones is the large vector **AF**; of the four zones the smaller vector **AH** and so on. Refer to Fig. 4.10(a) again. You will note that the resultant of infinitely large number of zones is equal to $a_1 / 2$.
If we consider a finite number of zones, say *n*, the resultant is given by

$$\xi(n) = \frac{a_1}{2} + \frac{a_n}{2} \tag{4.5}$$

where n is any number (odd or even).

So far we have considered the effect of a whole number of half period elements at a given point. The sum of the amplitudes due to all the secondary wavelets starting from the *n*th zone was represented by a_n. But so far we have not computed the magnitude and phase of the amplitude vector a_n. An obvious related problem is to calculate the effect at P_0 due to a fraction of a given half period element. We can do this easily by the following vector summation method. We divide a Fresnel zone into a series of *n* sub-zones of equal areas. Refer to figure below. It shows such a division for the annular space between the (*n* - 1)th and the *n*th circles. *O* is taken as centre and circles of slightly differing radii have been drawn such that the annular space between two consecutive circles encloses equal area. Now within the area covered by a sub-zone, we can neglect variation in inclination factor. Since all these sub-zones have been drawn so that they have equal areas, the amplitude at P_0 due to these small equal areas will be the same. But the phases will change continuously from one sub-zone to the next sub-zone by $\lambda / 2n$ since the phase difference between the secondary wavelets starting from the innermost sub-zone of any one Fresnel half period zone is $\lambda / 2$ or π. If we make *n* very large, we will have infinitesimally small but equal areas and phases of wavelets from these may be taken to vary continuously and uniformly.



Thus we have a set of disturbances of equal amplitude but uniformly changing phase such that the phase difference between the two extreme disturbances is π. These extreme vectors are represented by *AA′* and *BB*' in the figure shown. We know that in such a case the vector diagram is a semicircle and the resultant of the summation of amplitudes is the diameter *AB*.

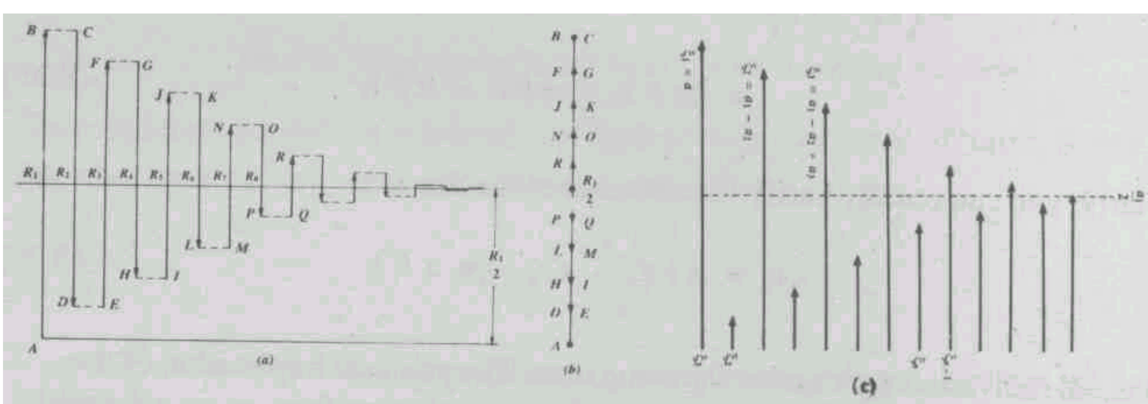


Fig. 4.10: Phasor diagram for Fresnel (half-period) zones. Individual amplitudes are shown in (a).

Actually all vectors are along a line. This is shown in (b). The resultant amplitude due to n (= 2, 3, ...) zones is shown in (c)

To see this, you closely re-examine Fig. 4.10(b). You will note that all vectors representing a_1, a_2, a_3, a_4, ... are line segments whose midpoints coincides with the midpoint of a_1 (marked as —). (You must convince yourself about this.) In other words, the vector representing a_n is a line, half of which is above the horizontal line passing through the midpoint of a_1 and the other half is below this line. The resultant of n zones is a vector joining A to the end of the vector representing a_n. When n is odd, the end point of the vector representing a_n will be above the horizontal line by $a_n/2$, which proves the required result.

If n is even, the end point will be below this horizontal line by $a_n/2$. Added vectorially, we have the same result. We thus see that the resultant amplitude at P_0 due to n zones is half the sum of amplitudes contributed by the first and the last zone. ξ will be numerically greater than $a_1/2$ when n is odd and smaller than $a_1/2$ when n is even. For example, the resultant contribution due to 7 zones is AO, which is equal to $\frac{a_1}{2}+\frac{NO}{2}$. On the other hand, for 8 zones, the resultant is $AQ=\frac{a_1}{2}-\frac{PQ}{2}$.

It may be emphasized that in this graphical method of summation of the series, we have used three properties: (i) vectors representing a_1, a_2, ... are all along the same straight line (ii) alternate vectors are oppositely directed and (iii) the magnitudes of a_1, a_2, ... decrease gradually.

We now consider a simple example to illustrate these concepts.

Example 1

Consider a series with n = 100 in which each term is equal to the arithmetic mean of the preceding and the following terms. Calculate the resultant.

Solution

As a special case, we can take the terms of the series as 100, 99, 98, ..., 3, 2, 1.

$$\therefore \quad \xi = (100-99)+(98-97)+(96-95)+\ldots+(4-3)+(2-1)$$

$$= 1+1+1+\ldots \qquad \text{(50 terms)}$$

$$= 50$$

which is half of the first term. Now consider the relation

$$\xi=\frac{a_1}{2}+\frac{a_n}{2}$$

and take different number of terms in this arithmetic series. If we have only one term, $a_1 = 100$ we take the first term as 100 and also the last term as 100. Then we get

$$\xi = \frac{a_1}{2} + \frac{a_n}{2} = 100$$

Next we take two terms. Then

$$\xi = (100 - 99) = 1$$

Also

$$\frac{a_1}{2} + \frac{a_n}{2} \qquad = \frac{100}{2} - \frac{99}{2}$$

$$= 50 - 49.5 = 0.5$$

For three terms, $\xi = (100 - 99) + 98 = 99$

and $\frac{a_1}{2} + \frac{a_3}{2} = 50 + 49 = 99$

For four terms, $\xi = (100 - 99) + (98 - 97) = 2$

and $\frac{a_1}{2} + \frac{a_4}{2} = 50 - 48.5 = 1.5$

For five terms $\xi = (100 - 99) + (98 - 97) + 96 = 98$

and $\frac{a_1}{2} + \frac{a_5}{2} = 50 + 48 = 98$

For six terms $\xi = (100 - 99) + (98 - 97) + (96 - 95) = 3$

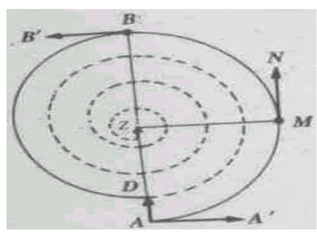
and $\frac{a_1}{2} + \frac{a_6}{2} = 50 - 47.5 = 2.5$

and so on. Thus we see that ξ is given by $\frac{a_1}{2} + \frac{a_n}{2}$ to a fairly good degree of accuracy.

4.4.2 Rectilinear Propagation

Refer to Fig. 4.11. It shows several collinear apertures *A, B, C, ...* Light originates from a point source and propagates towards the right. Suppose that the source is 1m away. We may take the spherical wave falling on the obstacle as nearly a plane wave. (The radius of curvature of the incident spherical wave will not qualitatively change the argument.) Let us work out the sizes of Fresnel half period elements for the typical case where the screen is 30 cm away from the aperture.

Now we will compute magnitude and phase of the resultant at AB. If all the equal disturbances from the sub-zones were in the same phase, the resultant would have been a line along *AA'* and equal lo the length of the arc of the semicircle *AB* ($= \pi r$) of radius *r*. But we find that the actual resultant amplitude is *AB* = *2r*. Thus the resultant amplitude is $\frac{2r}{\pi r} = \frac{2}{\pi}$ times the value which would be obtained if all the wavelets within a Fresnel half period element had the same phase. Since the line *AB* is parallel to the line *MN*, we see that the resultant phase of vector **AB** is the same as that of the vector **MN** representing the disturbance starting from the middle point (*M*) of the zone. In other words, AB is perpendicular to *AA'*. That is, it is a quarter-period behind the wavelet starting from the innermost sub-zone. We can find, in a similar manner, the resultant contribution due to the next half-period zone. It is given by *CD* and differs from *AB* by π. The resultant of the sum of these two zones is the small vector **AD**. The magnitudes of vectors and their phases for succeeding zones are shown in the figure below. The resultant curve is the vibration spiral with gradually smaller and smaller semicircles until eventually it coincides with Z. The resultant when all the half-period elements are considered is *AZ* which is half of that which would be produced by the first zone alone. It is equal to $\frac{1}{2} \times \frac{2}{\pi} = \frac{1}{\pi}$ times that which would be produced by all the wavelets from the first zone acting together in the same phase.



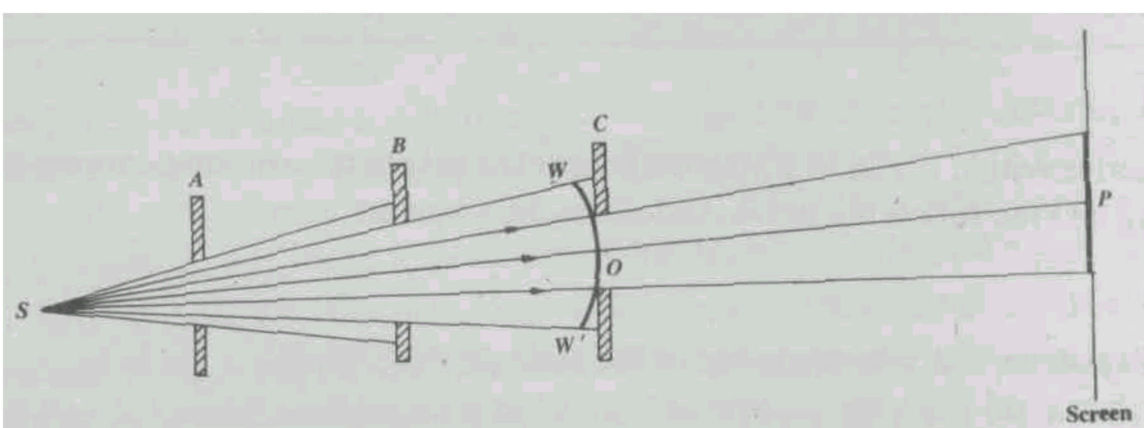


Fig. 4.11 Fresnel construction and rectilinear propagation of light

Taking $\lambda = 5\times10^{-5}$ cm, we get $r_1 = \sqrt{(30cm)\times(5\times10^{-5}cm)} = 3.87\times10^{-2}$ cm. This means that the diameter of the first zone is less than 1 mm. Let us consider the 100th zone. Its radius $r_{100} = \sqrt{30cm\times100\times5\times10^{-5}cm} = 3.87\times10^{-1}$ cm, so that the diameter will be a little less than 1 cm. Therefore, if the aperture is about 1 cm in diameter, the amplitude at P_0 due to the whole wavefront is $\frac{a_1}{2}+\frac{a_{100}}{2}$. a_{100} will be fairly small, so that the intensity is essentially half of that due to the first half period zone, which is the intensity expected at P_0 when the aperture is completely removed. We may say that light travels to P_0 from a region nearly 0.4 mm in radius around *O*. That is, **light travels in a straight line.**

Let us now understand the formation of shadows and illuminated regions due to an obstacle (Fig. 4.12). Consider the point P_2 whose pole is O_2. If the distance between O_2 and the edge *A* of the obstacle is nearly 1 cm, over 100 half period elements will be accommodated in it. And as seen above, the intensity at P_2 will be nearly equal to $\frac{a_1}{2}$. In other words, the obstacle T will have no effect at the point P_2- Similarly, at P_1, which is taken 1 cm inside the geometrical edge of the shadow, over 100 half period elements around O1 are obstructed and the intensity at P_1 will be less than $\frac{a_{100}}{2}$, which is almost negligible. This implies almost complete darkness at P_1. In other words, the obstacle has completely obstructed the light from the source and the region around point P_1 is in the shadow. Only around P_0, which signifies the geometrical edge of the shadow, we find fluctuations in intensity depending upon how many half period elements have been allowed to pass or have been obstructed. This explains the observed rectilinear propagation of light since Fresnel zones are obstructed or allowed through by obstacles of the size of a few mm for these typical distances.

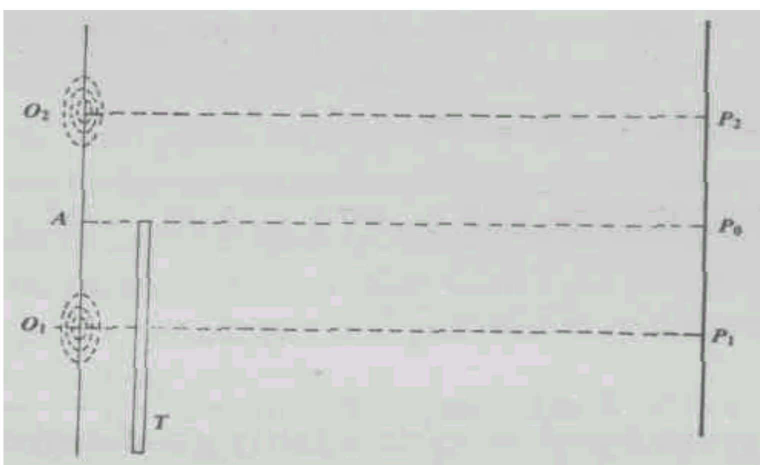


Fig. 4.12: Fresnel construction and formation of shadows/Illuminated regions

A special optical device, designed to obstruct light from alternate half-period elements is known as **Zone plate.** It provides experimental evidence in favour of Fresnel's theory. Let us learn about it now.

4.4.3 The Zone Plate

The zone plate is a special optical device designed to block light from every other half-period zone. You can easily make a zone plate by drawing concentric circles on a white paper, with their radii proportional to the square roots of natural numbers and shading alternate zones. Fig. 4.13 shows two zone plates of several Fresnel zones, where all even numbered or odd numbered zones are blackened out. Now photograph these pictures. The photographic transparency (negative) in reduced size acts as a Fresnel zone plate. (Recently, Gabor has proposed a zone plate in which zones change transmission according to a sinusoidal wave.) Lord Rayleigh made the first zone plate in 1871. Today zone plates are used to form images using X-rays and microwaves for which conventional lenses do not work.

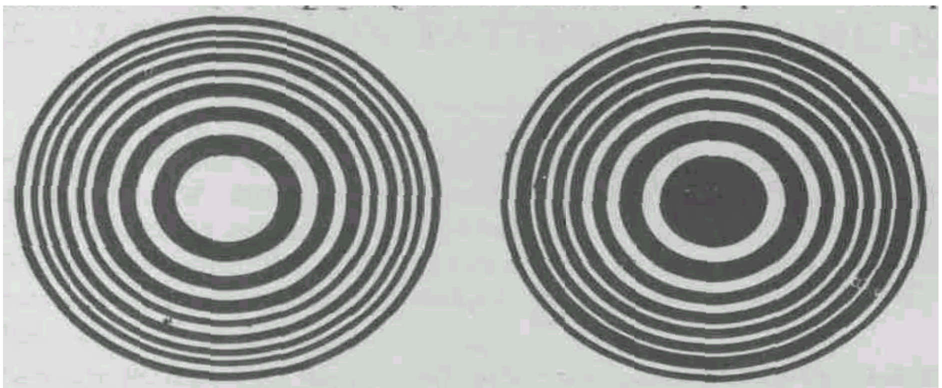


Fig. 4.13 Zone plates: (a) positive (b) negative

If you now pause for a while and logically reflect upon the possible properties of a Fresnel zone plate, you will arrive at the following conclusions:

1. A zone plate acts like a converging lens (see Example 2) and produces a very bright spot. To understand the formation of the spot let us suppose that the first ten odd zones are exposed to light. Then, Eq.(8.4) tells us that the resultant amplitude at P_0 is given by

$$\xi_{20} = a_1 + a_3 + a_5 + ... + a_{19} \qquad (8.5)$$

If the obliquity factor is not important, we may write $\xi_{20} = 10a_1$, which means that the amplitude for an aperture containing 20 zones is twenty times and intensity is 400 times that due to a completely unobstructed wavefront.

Example 2

Show that a zone plate acts like a converging lens.

Solution

Refer to Fig. 4.14. It shows the section of the zone plate perpendicular to the plane of the paper. *S* is a point source of light at a distance *u* from the zone plate. A bright image is formed at P_0 at a distance *v* from the plane of the zone plate.

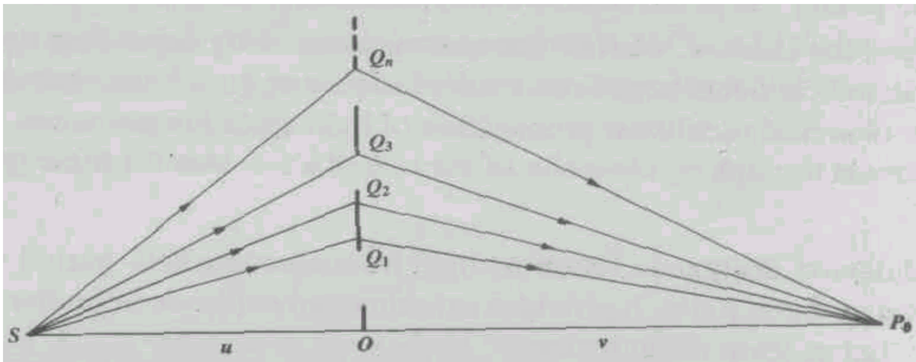


Fig. 4.14: Action of a Zone Plate as a converging lens

You can easily write

$$SQ_1 + Q_1P_0 = u + v + \frac{\lambda}{2}$$

$$SQ_2 + Q_2P_0 = u + v + \frac{2\lambda}{2}$$

$$SQ_n + Q_nP_0 = u + v + \frac{n\lambda}{2}$$

By Pythagoras' theorem we can write

$$SQ_n = \sqrt{SO^2 + OQ_n^2}$$

$$= \sqrt{u^2 + r_n^2} = u + \frac{r_n^2}{2v} + \ldots$$

where r_n is the radius of the *n*th zone.

Similarly, you can convince yourself that

$$Q_nP_0 = v + \frac{r_n^2}{2v} + \ldots$$

If $r_n << u$ or v, we can ignore terms higher than $\frac{r_n^2}{2u}$ or $\frac{r_n^2}{2v}$. Hence

$$SQ_n + Q_nP_0 = u + \frac{r_n^2}{2u} + v + \frac{r_n^2}{2v} = u + v + \frac{n\lambda}{2}$$

If we identify $f_n = \frac{r_n^2}{n\lambda}$ as the focal length of the zone plate, we find that

$$\frac{1}{u}+\frac{1}{v}=\frac{n\lambda}{r_n^2}=\frac{1}{f_n}$$

which is identical to the lens equation.

---

2. The zone plate has several foci. To understand this, we assume that the observation screen is at a distance of one focal length from the diffracting aperture. Then it readily follows from the above example that the most intense (first order) focal point is situated at $f_1 = r_1^2/\lambda$. To give you a feel for numerical values, let us calculate f_1 for a zone plate with radii $r_n = 0.1\sqrt{n}$ cm and illuminated by a monochromatic light of wavelength λ = 5500 Å. You can easily see that

$$f_1 = \frac{r_1^2}{\lambda} = \frac{(0.1cm)^2}{5500\times10^{-8}\ cm} = 182 \text{ cm}$$

To locate higher order focal points, we note from Eq. (4.2) that for r_n fixed, n increases as b decreases. Thus for $b = f_1/2$, n = 2. That is, as P_0 moves towards the zone plate along the axis, the same zonal area of radius r encompasses more half-period zones. At this point, each of the original zones covers two half-period zones and all zones cancel. When $b = f_1/3$, $n = 3$. That is, three zones contribute from the original zone of radius r_1. Of these, two cancel out but one is left to contribute. Thus other maximum intensity points along the axis are situated at

$$f_n = \frac{r_1^2}{n\lambda} \text{ for } n \text{ odd} \tag{4.9}$$

For the numerical example above, $f_3 = \frac{182}{3}$ cm, $f_5 = \frac{182}{5}$ cm , $f_7 = \frac{182}{7}$ cm and so on. Between any two consecutive foci, there will be dark points.

4.5 DIFFRACTION PATTERNS OF SIMPLE OBSTACLES

From Sec. 4.3 you will recall that by utilizing Kathvate's experimental arrangement, the Fresnel diffraction pattern of various apertures and obstacles could be photographed by varying distances between the source, the object and the photographic plate. We will now use results derived in Sec. 4.4 to explain the observed diffraction pattern of simple obstacles like circular aperture and straight edge.

We begin by studying the Fresnel diffraction pattern of a circular aperture.

4.5.1 A Circular Aperture

Refer to Fig. 4.15. It shows a sectional view of the experimental arrangement in which a plane wave is incident on a thin metallic sheet with a circular aperture. You will note that the plane of the wavefront is parallel to the plane of the metal plate; both being perpendicular to the plane of the paper.

Let us calculate the intensity at a point P_0 lying along the line passing through the centre of the circular aperture and perpendicular to the wavefront. Assume that the distance between the point P_0 and the circular aperture is *b*. As discussed earlier, the intensity at the observation point due to the entire uninterrupted plane wavefront is given by Eq. (4.4) where a_1, a_2, etc. give the contributions due to successive Fresnel zones.

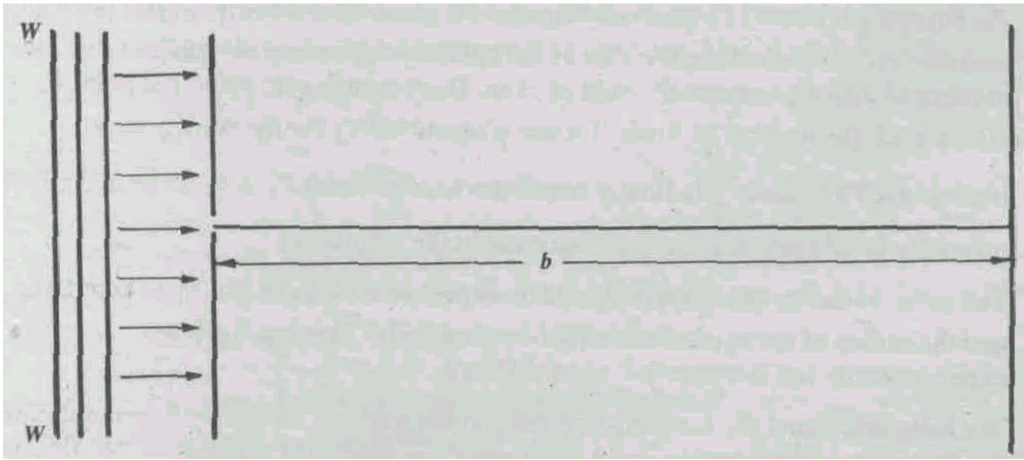


Fig. 4.15 Diffraction by a circular aperture: A cross-sectional view of the experimental arrangement

Our problem here can be solved by constructing appropriate Fresnel zones and finding out how many of these half period elements are transmitted by the aperture. However, it is important to note that for an aperture of a given size, the number of half period elements transmitted may not always be the same. This is because the radii of the Fresnel zones depend upon the distance of point P_0 from O $(r_n\sqrt{n\lambda b})$. You can easily convince yourself that if the point P_0 is far away from the aperture (*b* is very large), the radii of the first zone, equal to $\sqrt{\lambda b}$, may be larger than the radius of the aperture. In such a situation, all the secondary wavelets starting even from the entire first zone alone may not be transmitted. That is, the wavelets from a small portion of the first Fresnel zone only are transmitted.

The next question we have to address to is: How to calculate the amplitude at P_0 when the aperture has transmitted only a fraction of the first Fresnel zone? As a first approximation, we assume that the wavelets arrive at P_0 in phase. (This is quite justified because the path difference between the extreme wavelets within anyone half period elements is $\lambda/2$. If only a fraction of the first zone transmits here, the net phase difference will be correspondingly less.) Further, the inverse square law for intensity tells us that the amplitude at P_0 will be inversely proportional to *b*. Hence, the effect at P_0, which is at a large distance, will be small.

As the point P_0 moves towards the aperture (*b* becomes smaller), the zone size shrinks and a greater part of the central zone is transmitted. As a result, the intensity increases gradually. As the observation point comes closer and closer, with the shrinking of the sizes of zones, a stage may be reached when the first zone exactly fills the aperture. Then $\sqrt{b\lambda}$, the radius of the first zone is also the radius of the aperture. We know that the first zone contributes a_1 to the amplitude at P_0. Compare it with the situation where the obstacle with circular aperture is not present. The entire wave front contributes but the amplitude at P_0 is $\frac{a_1}{2}$. Since intensity is proportional to the square of amplitude, the intensities at P_0 with and without the aperture are respectively a_1^2 and $\frac{a_1^2}{4}$. That is, the intensity at a given point is four times as large when the aperture is inserted in the path than when it is completely removed. This surprising result is not apparent in the realm of everyday experience dominated by rectilinear propagation of light.

As the observation point P_0 comes still closer, the circular aperture may transmit the first two zones. The amplitude will then be $(a_1 - a_2)$ which is expected to be very small. The additional light produces practically zero amplitude, hence darkness, at P_0. Bringing the point P_0 gradually closer will cause the intensity to pass through maxima and minima along the axis of the aperture depending on whether the number of zones transmitted is odd or even. If we continue to bring the point P_0 closer to *O,* the number of Fresnel zones transmitted by the aperture goes on increasing. The value $\frac{a_1}{2}$ is finally reached when the point P_0 is so close that an infinitely large number of zones contribute to the amplitude.

The same variation in intensity should be experienced if the point P_0 is kept fixed
and the radius of the aperture is varied continuously. This can be done experimentally but is somewhat more difficult.

We have calculated the intensity at points on the axis but the above considerations do not give any information about the intensity at points off the axis. A detailed and complex mathematical analysis which we shall not discuss here, shows that P_0 is surrounded by a system of circular diffraction fringes. Photographs of these fringe patterns have been taken by several workers and we referred to Kathvate's experiments earlier in this unit.

We now illustrate the concepts developed here by solving an example.

Example 3
In an experiment a big plane metal sheet has a circular aperture of diameter 1 mm. A beam of parallel light of wavelength λ = 5000 Å is incident upon it normally. The shadow is cast on a screen whose distance can be varied continuously. Calculate the distance at which the aperture will transmit 1, 2, 3, ... Fresnel zones.

Solution
Let b_1, b_2, b_3, …, b_n be the distances at which 1, 2, 3, … zones are transmitted by an aperture of fixed radius *r*. From Eq. (4.2) we can write

$$nb_n\lambda = r_n^2$$

so that $b_n = \frac{r_n^2}{n\lambda}$

$$b_1 = \frac{r_1^2}{\lambda} = \frac{(0.05\,cm)^2}{5\times10^{-5}\,cm} = 50 \text{ cm}$$

Similarly, we find that

$$b_2 = \frac{r_2^2}{2\lambda} = \frac{50}{2} \text{ cm} = 25 \text{ cm},\ b_3 = \frac{50}{3} \text{ cm} = 16.7 \text{ cm},\ b_4 = \frac{50}{4} \text{ cm} = 12.5 \text{ cm},$$
$$b_5 = 10 \text{ cm},\ b_6 = 8.3 \text{ cm},\ b_7 = 7.1 \text{ cm},\ b_8 = 6.2 \text{ cm}.$$

The amplitudes corresponding to these distances are plotted in Fig. 4.16.

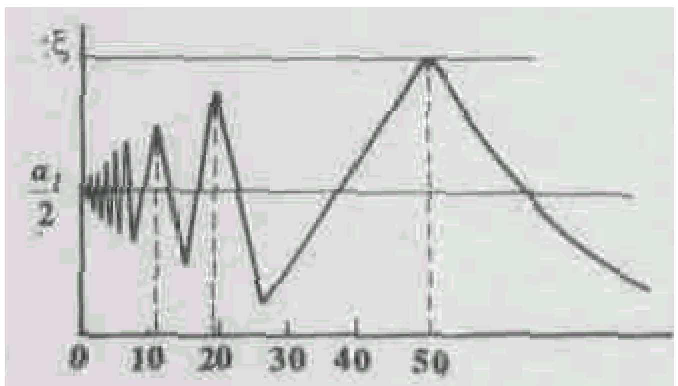


Fig. 4.16 Variation of amplitudes when a circular aperture transmits integral multiple Fresnel zones

---

Another conclusion of some historic interest follows if we substitute the aperture by a circular disc or a round obstacle just covering the first Fresnel zone. The light reaching the point of observation P_0 will be due to all zones except the first. The second zone is therefore the first contributing zone and the intensity of light spot at the centre of the shadow of the obstacle will be almost equally bright as when the first zone was unobstructed.

You may now ask: Why is the bright spot at the centre only? This is because there is no path difference and hence phase difference between waves reaching an axial point. At any other off-axis point, waves will reach with different phases and may tend to cancel mutually. The existence of this spot was demonstrated by Arago, though Poisson gave his theoretical arguments to disprove wave theory of light.

You may now like to answer an SAQ.

SAQ1
A 25 pence coin has a diameter of 2 cm. How many Fresnel zones does it cut off if the screen is 2 m away? Do you expect to see a bright spot at the centre? If we move the screen to a distance of 4

m, how many zones will it cut off? Will the bright spot now look brighter? Why? Take $\lambda = 5\times10^{-7}$ m.

So far we have discussed diffraction patterns which had axial symmetry: the object or aperture was circular and the plane wavefront originated from a point source. We now wish to consider the case wherein source is a slit source. This source will emit cylindrical waves with the slit as axis. Let us now study the diffraction pattern of a straightedge.

> The slit has a very small width compared to its length. Or we may say that in comparison to its width, it has an infinite length.

4.5.2 A Straight Edge

Let *S* be a slit source perpendicular to the plane of the paper. This sends a cylindrical wavefront towards the obstacle, which is a straight edge perpendicular to the paper. You can take a thin metal sheet or a razor blade with the sharp edge parallel to the slit. Fig. 4.17(a) shows a section perpendicular to the length of the slit.

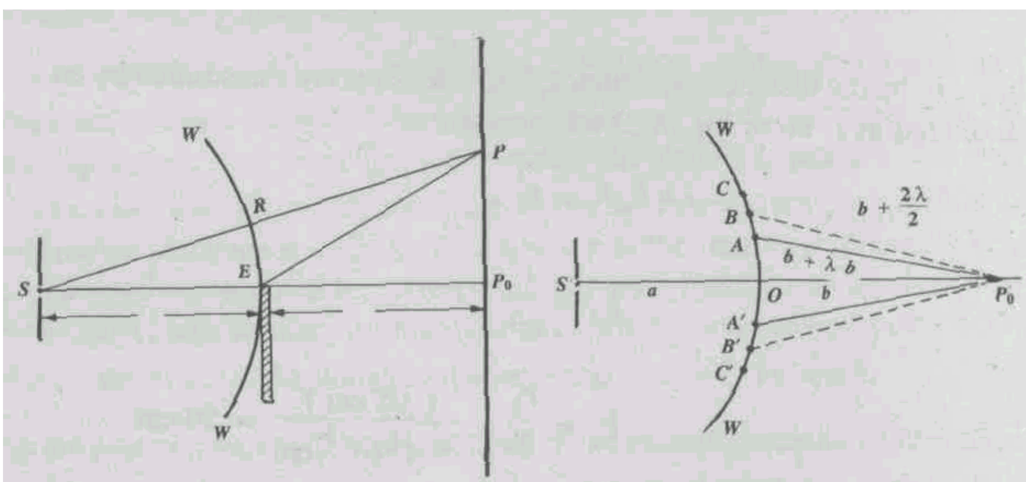


Fig. 4.17 (a) Cross sectional view of the geometry to observe diffraction due to a straight edge and (b) Fresnel construction divides the cylindrical wavefront in half period strips

The line joining *S* and *E,* the point on the wavefront, when produced meets the screen at P_0, which is the geometrical boundary of the shadow. Consider any point *P* on the screen. A line joining it to *S* cuts the wavefront at *R*. We wish to know how intensity varies on the screen. This calculation is somewhat complicated because we now have a cylindrical wavefront. Moreover, the obstacle does not have an axial symmetry.

For a plane wave and obstacles with axial symmetry you know how to construct Fresnel zones. To construct half period elements for a straight edge, we divide the cylindrical wavefront into strips. As before, we make sure in the construction that the amplitudes of the wavelets from these strips arrive at P_0 out of phase by π so that alternate terms are positive and negative. This is achieved by drawing a set of circles with P_0 as centre and radii b, $b+\frac{\lambda}{2}$, $b+\frac{2\lambda}{2}$, etc., cutting

the circular section of the cylindrical wave at points *O*, *AA'*, *BB'*, *CC'*, ... Fig. (4.17b). If lines are drawn through *A*, *A*', *B*, *B'* , etc. normal to the plane of the paper, the upper as well as the lower half of the wavefront gets divided into a set of **half-period strips.** These half period strips stretch along the wavefront perpendicular to the plane of the paper and have widths *OA*, *AB*, *BC* ...in the upper half and *OA'*, *A' B'*, *B' C'*, ... in the lower half. You may recall that Fresnel zones are of equal area. For half period strips, this does not hold. The areas of half-period strips are proportional to their widths and these decrease rapidly as we go out along the wavefront from *O*.
From the geometry of the arrangement it is obvious that on the screen there will be no intensity variation along the direction parallel to the length of the slit. Therefore, the bright and dark fringes will be straight lines parallel to the edge.

A plot of theoretically calculated intensity distribution on the screen is shown in Fig. 4.18. You will note the following salient features:

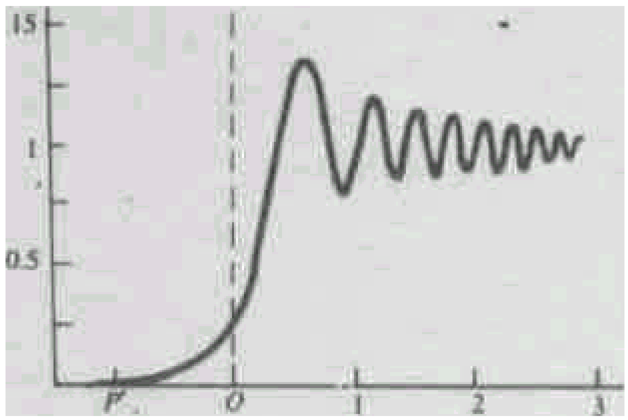


Fig. 4.18 Intensity distribution in the diffraction pattern due to a straight edge

(i) As we go from the point *P'* deep inside the shadow towards the point *O* defining the edge of the shadow, the intensity rises gradually. At *P'* the intensity is almost zero.

(ii) At O, the intensity is one-fourth of what would have been the intensity on the screen with the unobstructed wavefront.

(iii) On moving further towards *P*, the intensity rises sharply and goes through an alternating series of maxima and minima of gradually decreasing magnitude and approaches the value for the unobstructed wave. This is expected since the effect of the edge at far off distances will be almost negligible.

(iv) The intensity of first maxima is greater than the intensity of unobstructed wave, i.e. it is greater than 4 times the intensity at *O*. Beyond these alternate maxima and minima, there is uniform illumination.

(v) The diffraction fringes are not of equal spacing (as in interference experiments); the fringes gradually come closer together as we move away from the point *O*.

You may now like to know at least qualitative explanation of these results. To do so, we first consider the illumination at a point *P* outside the geometrical shadow. The line joining *P* and *S* cuts the wavefront at *R* so that the wavefront is divided in two parts. The amplitude of light at *P* is due to the part *WE* of the wavefront, which is completely unaffected by the straight edge. The amplitude at P will be maximum if RE contains odd number of half strips. This will happen if $EP - RP = (2n+1)\lambda/2$. (When $EP - RP = n\lambda$, the portion RE will contain even number of strips.) As pointed out earlier, the amplitudes due to strips are alternately positive and negative. Therefore, as point P moves away from O, the illumination on the screen will pass alternately through maxima and minima when the number of half period strips in RE is 1, 2, 3, 4, ...

It is worthwhile to ponder as to what pattern the geometry of the experimental configurations throws? We expect dark and bright bands parallel to the edge. However, the dark bands will not be completely dark, since the upper half of the wavefront RW always contributes light to this part of the screen.

Let us now consider the situation for the point P' inside the geometrical shadow. Refer to Fig. 4.19. You will note that the corresponding point R is shifted below the edge so that the illumination at P' is due entirely to the wavelets from the upper half of the wavefront; the lower portion having been blocked by the edge. Even the upper half is exposed only in part. If the edge cuts off r strips of the upper half of the wavefront, the effect at P' will be due to (r + 1), (r + 2), (r + 3) etc. strips which may be taken to be equal to one-half of that due to the (r + l)th strip. This will rapidly diminish to zero as shown in Fig. 4.18, because the effectiveness of higher order strips goes on decreasing.

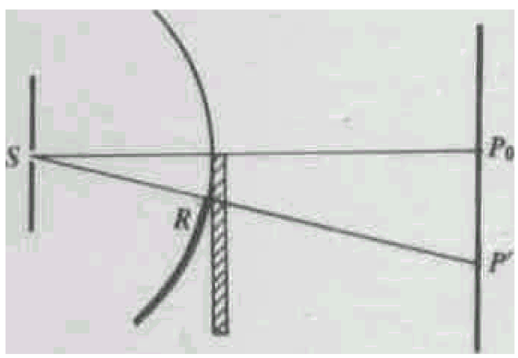


Fig. 4.19 The observation point is in the geometrical shadow of the straight edge

Let us now deduce the width of the diffraction bands. Again Refer to Fig. 4.17(a). Suppose that we have the nth dark band at P. Then

$$EP - RP = n\lambda \tag{4.6}$$

From the ΔEPO, we have

$$\begin{aligned} EP &= (b^2 + x^2)^{1/2} = b\left(1 + \frac{x^2}{b^2}\right)^{1/2} \\ &= b\left(1 + \frac{1}{2}\frac{x^2}{b^2}\right) = b + \frac{1}{2}\frac{x^2}{b} \end{aligned} \tag{4.7}$$

where we have retained only the first two terms in the binomial series. From the ΔSPO, we can similarly write

$$SP = (a+b) + \frac{1}{2}\frac{x^2}{(a+b)}$$

Hence,

$$RP = SP - SR = b + \frac{1}{2}\frac{x^2}{(a+b)} \tag{4.8}$$

and

$$EP - RP = \left(b + \frac{1}{2}\frac{x^2}{b}\right) - \left(b + \frac{1}{2}\frac{x^2}{a+b}\right)$$

$$= \frac{1}{2}\left(\frac{x^2}{b} - \frac{x^2}{a+b}\right) = \frac{x^2 a}{2b(a+b)} \tag{4.9}$$

For the nth dark band, we get

$$\frac{x^2 a}{2b(a+b)} = n\lambda$$

or

$$x = \sqrt{n\frac{2b(a+b)}{a}\lambda} \tag{4.10}$$

We therefore find that the distances of the dark bands from the edge of the geometrical shadow are proportional to the square root of natural numbers. Consequently the bands will get closer together as we go out from the shadow. This fact distinguishes the diffraction bands from the interference bands, which are equidistant.

To enable you to grasp these ideas, we now give a solved example.

Example 4

In the above experiment if a = 30 cm, b = 30 cm and $\lambda = 5\times10^{-5}$ cm, calculate the position of the 1st, 2nd, 3rd and 4th minima from the edge of the shadow.

Solution

From Eq. (4.10) we know that the distance of nth minima from the edge of the shadow is given by

$$x = \sqrt{n\frac{2b(a+b)}{a}\lambda}$$

If we substitute given values of a, b and λ and take n = 1, 2, 3, 4, we find that

$$x_1 = \left[\frac{2\times(30\,cm)\times(60\,cm)}{30\,cm}\times(5\times10^{-5}\,cm)\right]^{1/2}$$

$$= 7.75 \times 10^{-2} \text{ cm}$$

$$x_2 = \sqrt{2}x_1 = 1.09 \times 10^{-1} \text{cm}$$

$$x_3 = \sqrt{3}x_1 = 1.34 \times 10^{-1} \text{ cm}$$

$$x_4 = 2x_1 = 1.55 \times 10^{-1} \text{cm}$$

From these values we find that the distance between consecutive minima decreases continuously as we move away from the edge of the shadow.

You may now like to answer an SAQ.

SAQ 2

Instead of the straight edge, we keep a narrow obstacle, say a wire of diameter 1 mm. What will be the intensity on the screen?

Let us now summarise what you have learnt in this unit.

4.6 SUMMARY

- When the distance between the source of light and the observation screen or both from the diffracting aperture/obstacle is finite, the diffraction pattern belongs to Fresnel class.

- When the screen is very close to the slit, the illumination on the screen is governed by rectilinear propagation of light.

- The Fresnel diffraction pattern represents fringed images of the obstacle. Depending on the distance, there can be an infinite number of Fresnel diffraction patterns of a given obstacle/ aperture.

- When plane wavefronts are incident on a diffracting slit and the pattern is observed on a screen effectively at an infinite distance, the diffraction pattern belongs to Fraunhofer type. Unlike the Fresnel diffraction, there is only one Fraunhofer diffraction pattern.

- Fresnel construction for the diffraction pattern from any obstacle on which a plane wavefront is incident consists of dividing the wavefront into half period zones.

- The area of each Fresnel half-period zone is equal to $\pi b \lambda$.

- The resultant amplitude due to nth zone at any axial point is given by

$$a_n = \frac{A_n}{P_n}(1 + \cos\theta)$$

- The magnitude of resultant amplitude *AB* due to the first half period element is $\frac{2}{\pi}$ times the value which would be obtained if all the wavelets within the half- period element had the same phase.

- The phase of the resultant vector of the first half period zone is $\frac{\pi}{2}$ behind the phase of light from the centre of the zone.

- A zone plate is an optical device in which alternate half-period zones are blackened.

- The diffraction pattern due to a circular aperture consists of a central bright spot.

- The diffraction pattern of a straight edge consists of alternate bright and dark bands. The spacing between minima (or maxima) decreases as we move away from the edge of the shadow:

$$x = \sqrt{n\frac{2b(a+b)}{a}\lambda}$$

4.7 TERMINAL QUESTIONS

1. Starting from Eq. (4.4) establish Eqs. (4.6) and (4.7). Assume that the obliquity factor is such that each term in Eq. (4.4) is less than the arithmetic mean of its preceding and succeeding terms.)

2. The eighth boundary of a zone plate has a diameter of 6mm. Where is its principal focal point located for light of wavelength 5000 Å?

3. How many Fresnel zones will be obstructed by a sphere of radius 1 mm if the screen is 20cm away? Take = 5000 Å. If the distance of the screen is increased to 200 cm, what will be the size of the sphere which will cut off 10 zones.

4.8 SOLUTIONS AND ANSWERS

SAQs

1. The radius of the coin is equal to 1 cm. To know the number of zones being obstructed, we use the relation

$$n = \frac{r_n^2}{b\lambda}$$

where r_n = 1 cm, b = 200 cm and $\lambda = 5\times10^{-5}$ cm.

You should definitely expect to see a dim spot at the centre because the eleventh zone is the first contributing zone.

When the screen is 4 m away, the number of zones being obstructed is given by

$$n = \frac{(1\,cm)^2}{(400\,cm)\times(5\times10^{-5}\,cm)}$$

$= 5$

That is, only five zones are obstructed now and the first contributing term in Eq. (4.4) is a_6, which will contribute more than a_{11}. Therefore, the central spot is expected to be brighter. Does it not contradict the inverse square law?

2. Refer to Fig. 4.20. A point P_1 outside the geometrical shadow is similar to such a point in the straight edge. So we will have unequally spaced bright and dark fringes parallel to the wire on each side of the shadow. What is the intensity at Q inside the shadow? It is simply half the effect of the first half period strip on either side of the thin wire. It will show equally spaced fringes inside the shadow.

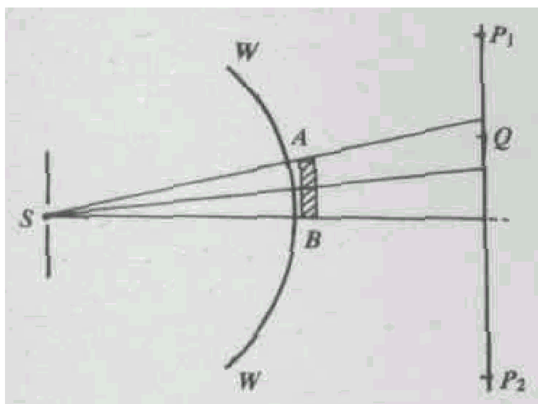


Fig. 8.20: A cross-sectional view of the arrangement for producing diffraction due to a narrow obstacle

TQs

1. We rewrite Eq. (4.4) as

$$a(P_0) = \frac{a_1}{2} + \frac{a_1}{2} - a_2 + \frac{a_3}{2} + \frac{a_3}{2} - a_4 + \frac{a_5}{2} + \ldots \quad \text{(i)}$$

When n is odd, the last term would be $\frac{a_n}{2}$. We are told that the obliquity is such that each term is less than the arithmetic mean of its preceding and succeeding terms i.e., $a_n > \frac{1}{2}(a_{n-1} + a_{n+1})$. Then, the quantities in the parentheses in (i) will be positive. So when n is odd, the minimum value of the amplitude of the fields produced by consecutive zones is given by

$$a(P_0) > \frac{1}{2}(a_1 + a_n) \quad \text{(ii)}$$

To obtain the upper limit, we rewrite Eq. (4.4) as

$$a(P_0) = a_1 - \frac{a_2}{2} - \frac{a_2}{2} - a_3 + \frac{a_4}{2} - \frac{a_4}{2} - a_5 + \frac{a_6}{2} - \ldots - \frac{a_{n-1}}{2} + a_n$$

Following the argument used in obtaining the lower limit on the amplitude, we find that the upper limit is

$$a(P_0) < a_1 - \frac{a_2}{2} - \frac{a_{n-1}}{2} + a_n \qquad \text{(iii)}$$

Since the amplitudes for any two adjacent zones are nearly equal, we can take $a_{n-1} = a_n$. Within this approximation

$$a(P_0) < \frac{a_1 + a_n}{2} \qquad \text{(iv)}$$

The results contained in (ii) and (iv) suggest that when n is odd, the resultant amplitude at P_0 is given by

$$a(P_0) = \frac{a_1 + a_2}{2} \qquad \text{(v)}$$

Following the same method, you can readily show that if n were even,

$$a(P_0) = \frac{a_1 - a_2}{2} \qquad \text{(vi)}$$

2. D_8 = 0.6 cm so that r_8 = 0.3 cm. We know that

$$f_n = \frac{r_n^2}{n\lambda}$$

$$\therefore \qquad f_8 = \frac{r_8^2}{8\lambda} = \frac{(0.3\,cm)^2}{8 \times (5 \times 10^{-5}\,cm)}$$

$= 2.25 \times 10^2$ cm

= 225 cm

3a. The radius of a Fresnel zone is given by

$$r_n = \sqrt{n\lambda b}$$

Here we are told that r_n = 0.1 cm, b = 20 cm and $\lambda = 5 \times 10^{-5}$ cm.

$$\therefore \qquad n = \frac{r_n^2}{b\lambda} = \frac{10^{-2}\,cm^2}{(20\,cm) \times (5 \times 10^{-5})} = 10$$

b. In this part we have to calculate r_n for given values of n = 10, b = 200 cm and $\lambda = 5\times10^{-5}$ cm:

$$r_n = \sqrt{10\times(200\,cm)\times5\times10^{-5}\,cm)}$$
$$= 0.32 \text{ cm}$$

UNIT 5 FRAUNHOFER DIFFRACTION

Structure

5.1 INTRODUCTION

In the previous unit you studied Fresnel diffraction and learnt that the diffraction pattern depends on the distance between aperture and screen as well as the source. As the observation screen is moved away from the aperture, the diffraction pattern passes from the forms predicted in turn by geometrical optics, Fresnel diffraction and Fraunhofer diffraction. When plane wavefront is incident at the diffracting aperture, the transition from Fresnel to Fraunhofer pattern is determined by the ratio of the size of the diffracting obstacle to its distance from the source and/or the observation screen. You will now learn about Fraunhofer diffraction in detail.

In Sec. 5.2 we have described the experimental arrangement and salient features of the observed Fraunhofer diffraction pattern from a single slit illuminated by a point source. This is followed by a simple discussion on theoretical analysis of the observed results. Since we deal with plane wavefronts, you will find that theoretical analysis is fairly simple. In Sec. 5.3 we have described Fraunhofer diffraction by a circular aperture because of its importance for optical devices. You will learn that the diffraction pattern consists of a central bright disc (called Airy disc) surrounded by concentric dark and bright rings. As a corollary, you will see that a random array of small and closely circular obstacles gives overlapping diffraction patterns called halos. You may have observed brilliant halos while deriving a car whose fogged window is illuminated by motorcycle at the back. We shall discuss the physical basis for diffraction halos at the end of this unit.

Objectives

After going through this unit you will be able to:

- describe experimental arrangement for observing Fraunhofer diffraction pattern from a narrow vertical slit and a circular aperture
- explain observed irradiance on the basis of simple theoretical analysis
- solve numerical problems, and
- explain formation of diffraction halos.

5.2 DIFFRACTION FROM A SINGLE SLIT: POINT SOURCE

From the previous unit, you may recall that to observe Fraunhofer diffraction pattern, we require a point source, which is far away (almost at infinity) from the diffracting aperture (a single slit in the present discussion). The wavefronts of light approaching the diffracting aperture can be assumed to be essentially plane. The observation screen should also be at infinite distance from the aperture. You may now like to ask: Is it practical to put the source of light and the observation screen at infinite distance from the diffracting aperture? This definitely is not practical because (i) the intensity of diffracted light reaching the observation screen would be reduced infinitesimally (inverse square law) and (ii) we will require infinitely big laboratory rooms. Do these limitations suggest that we cannot observe Fraunhofer diffraction? These difficulties are readily overcome by using converging lenses in an actual experiment.

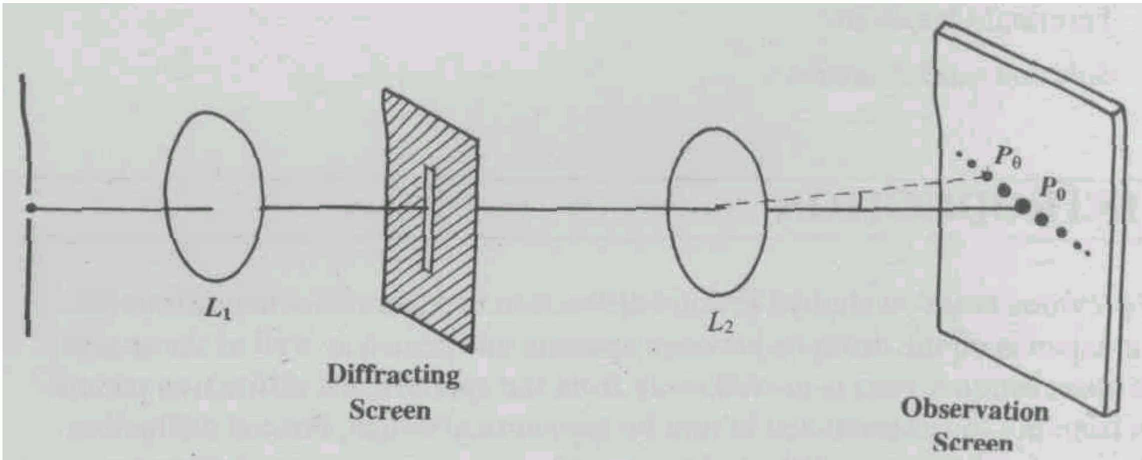


Fig. 5.1: Producing Fraunhofer diffraction pattern

The experimental arrangement for producing Fraunhofer diffraction pattern is shown in Fig. 5.1. The source of light is placed in the focal plane of a converging lens L_1, so that a plane wave is incident on a long narrow slit. Another convergent lens L_2 is placed on the other side of the slit. The observation screen is placed at the second focal point of this lens. Then light reaching any point on the observation screen is due to parallel diffracted wavelets from different portions of the wavefront at the slit. You must note that the observation screen and diffraction screen are kept parallel. Moreover, both the screens are perpendicular to the common axis of L_1 and L_2. The slit is so adjusted that the common axis of these lenses is perpendicular to the length of the slit and passes through the middle of the slit both in height and width.

In a physics laboratory this arrangement is easily achieved by using an ordinary spectrometer. We hope that you got an opportunity to work with a spectrometer in your second level laboratory course. To observe the diffraction from a point source, the slit of the collimator should be replaced by a fine pinhole, which should be carefully positioned at the focal point of the collimator lens. The observation screen can be placed at the second focal plane in the back focal plane of the telescope. Alternatively, we may observe the back focal plane of lens L_2 with an eyepiece. The diffracting screen with slit aperture is placed between the two lenses suitably on the turn table.

5.2.1 Observed Pattern

Let us pause for a minute and think how would diffraction pattern of the vertical slit appear? Or what would be the distribution of intensity in this pattern? You may think that the diffraction pattern would be a single vertical line or a series of vertical lines on the observation screen. This line of thought is wildly off-target. The actual diffraction pattern is astonishingly different; it consists of a horizontal streak of light composed of bright elongated spots connected by faint streaks. In other words, after passing through the vertical slit, light spreads along a horizontal line. This means that the diffraction pattern is along a line perpendicular to the length of the diffracting slit. You may interpret this horizontal diffraction as a spread out image of the point source. The extent of horizontal spreading is controlled by the width of the slit; as the width increases, the spreading decreases. And in the extreme case of a very wide slit, the (horizontal) diffraction streak reduces to a bright point. Physically, very wide slit means that the slit has effectively been removed.

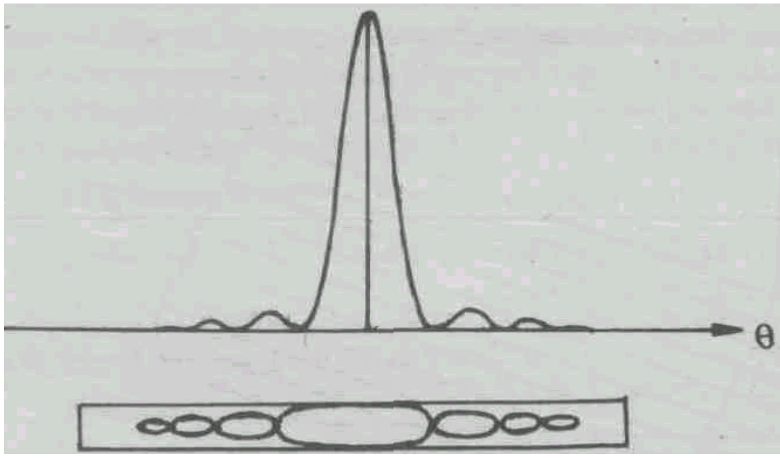


Fig. 5.2: Observed Fraunhofer diffraction pattern of a diffracting silt

The salient features of the observed Fraunhofer diffraction pattern of a single vertical slit from a point source are shown in Fig. 5.2. These are summarised below:

(i) The diffraction pattern consists of a horizontal streak of light along a line perpendicular to the length of the slit.

(ii) The horizontal pattern is a series of bright spots. The spot at the central point P_0, which lies at the intersection of the axis of L_1 and L_2 with the observation screen, is the brightest. On either side of the brightest spot we observe many more bright spots symmetrically situated with respect to P_0.

(iii) The intensity of the central spot is maximum. The peak intensities of other spots, on either side of the central spot, decrease rapidly as we move away from P_0. The central maximum is called principal maximum and the others as secondary maxima.

(iv) The width of the central spot is double of the width of other spots.

(v) A careful examination of the diffraction pattern shows that the central peak is symmetrical. But on either side of the central maximum, secondary maxima are asymmetrical. In fact, the positions of the maxima are slightly shifted towards the observation point P_0.

Let us now learn the theoretical basis of these results.

5.2.2 Calculation of Intensity Distribution

The first step in the calculation of intensity distribution is to realise that the observed diffraction pattern is focussed on the observation screen placed at the back focal plane of lens L_2. We know that only parallel rays are brought to focus in the back focal plane of the lens. The beam of rays parallel to the axis of the lens are focussed at the focal point. However, the beam inclined to the axis of the lens is brought to focus on the back focal plane but away from the focal point. We can as well describe this observation in terms of the wavefront, the two being perpendicular to each other. Since diffraction pattern lies on a horizontal line (which is at right angles to the common axis of L_1 and L_2, diffracted wavefronts will be vertical planes perpendicular to the plane of the paper. That is, after passing through the vertical slit, the incident plane waves are replaced by a system of vertical plane waves, which proceed in different directions. Therefore for our theoretical analysis it is sufficient to assume that when a plane wavefront falls on the diffracting slit, each point of the aperture such as $AA_1A_2A_3$... B (Fig. 5.3) becomes a source of secondary wavelets, which propagate in the direction of the point P_0 under consideration. These are diffracted plane waves. (You should realize that diffracted waves have no existence in the domain of geometrical optics. The diffracted waves arise due to interaction between light and matter. In the present case, the interaction is between light and the jaws of the slit.)

> The width of an image is specified by the distance between two consecutive minima.

> We take the plane of the paper as horizontal. The plane of the paper is defined by the diffraction streak and the axis of the lens L_2.

> Two sources are^ said to be coherent if they emit in-phase waves of the same frequency.

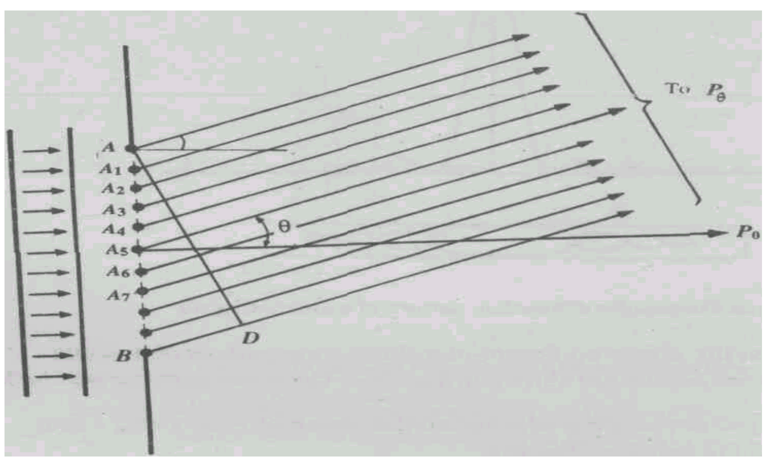


Fig. 5.3 Geometry of single silt diffraction

Refer to Fig. 5.3 which shows the geometry for the irradiance at the point P (on the distant screen) which makes an angle θ with the axis. In order to sum up the contributions of different wavelets at P, we must know their amplitudes and phases. The amplitudes of the disturbances from A, A_1, A_2, … will be very nearly equal. Do you know why? This is because the distance of point P from the diffracting screen is very large compared to the width (b) of the aperture.

Now let us consider the phases of the disturbances reaching the point P_0. You will agree that the points A, A_1, A_2, A_3, … B within the aperture form a series of coherent sources since they have originated from the same point source. Also points A, A_1, A_2, … B are in the same phase since they lie on the same plane wavefront The phase difference between different diffracted rays reaching P_0 arises due to the difference in path lengths travelled by them to reach this point. To know the phase difference, we draw a plane normal to the parallel diffracted rays. The trace of this plane in the plane of the paper is *AD* (Fig. 5.3). Though the disturbances are in phase at points *A*, A_1, A_2, … *B* when they start, they reach the trace AD in different phases because of the unequal path lengths travelled by them. The optical paths of diffracted waves from the plane *AD* to the focal point P_0 are equal. The optical paths of all rays between perpendicularly intersecting planes containing the parallel beam of light and the point where rays converge after traversing the lens are equal. Therefore, the wavelets arrive at P_0 with the same relative phase difference as the ones existing at the trace AD.

Let us consider the aperture *AB* to be divided into *n* equal parts so that $AA_1 = A_1A_2 = A_2A_3 = b/n = \Delta$. It means that the number of point sources is (*n* +1). Actually, the aperture contains a continuous distribution of points from *A* to *B*, and therefore in the limiting case, $n \to \infty$ and $\Delta \to 0$, such that $n\Delta \to b$. Consider two rays starting from two neighbouring points *A* and A_1. The path difference between them is $AA_1 \sin\theta$ where θ is the angle between the diffracted rays and the normal to the slit. Hence the corresponding phase difference is given by

$$\phi = \frac{2\pi}{\lambda}(AA_1 \sin\theta) = \frac{2\pi}{\lambda}\left(\frac{b}{n}\sin\theta\right) = \frac{2\pi}{\lambda}\Delta\sin\theta \tag{5.1}$$

Let the field at P_θ due to the disturbance originating from *A* be $a_0 \cos\omega t$. Then, the field due to the disturbance from A_1 is $a_0\cos(\omega t - \phi)$. Here we have assumed that the amplitudes of disturbances from different points are equal. The fields due to disturbances from successive points A_2, A_3, ... *B* are $a_0\cos(\omega t - 2\phi)$, $a_0\cos(\omega t - 3\phi)$, …, $a_0\cos(\omega t - n\phi)$, respectively. The magnitude of resultant field *E* at P_θ is equal to the sum of these disturbances. Hence

$$E = a_0\cos\omega t + a_0\cos(\omega t-\phi) + a_0\cos(\omega t - 2\phi) + \ldots + a_0\cos(\omega t - n\phi)$$

In Unit 2 of the course Oscillations and Waves, we summed up this series (Eq. (2.38)). We will just quote the result here:

$$\begin{aligned} E &= a_0\left[\frac{\sin\frac{n\phi}{2}}{\sin\left(\frac{\phi}{2}\right)}\right]\cos\left(\omega t - \frac{n\phi}{2}\right) \\ &= E_\theta\cos\left(\omega t - \frac{n\phi}{2}\right) \end{aligned} \tag{5.2}$$

where E_θ is the amplitude of the resultant field at P_θ:

$$E_\theta = a_0\frac{\sin\frac{n\phi}{2}}{\sin(\phi/2)} \tag{5.3}$$

In the limit $n \to \infty$ and $\Delta \to 0$, $n\Delta \to b$. Then from Eq. (5.1) we have

$$\frac{n\phi}{2} = \frac{n}{2}\frac{2\pi}{\lambda}\Delta \sin\theta = \frac{\pi}{\lambda}(n\Delta)\sin\theta = \frac{\pi}{\lambda} b \sin\theta$$

so that, $\phi = \dfrac{2\pi}{\lambda}\dfrac{b\sin\theta}{n}$ will be very small for $n \to \infty$. We may therefore write

$$\sin\left(\frac{\phi}{2}\right) \approx \frac{\phi}{2} = \frac{\pi b \sin\theta}{n\lambda}$$

Substitute this result in Eq. (5.3). On simplification you will find that

$$E_\theta = a_0 \frac{\sin\dfrac{n\phi}{2}}{\sin(\phi/2)} \approx a_0 \frac{\sin(n\phi/2)}{(\phi/2)} = na_0 \frac{\sin\left(\dfrac{\pi b \sin\theta}{\lambda}\right)}{\left(\dfrac{\pi b \sin\theta}{\lambda}\right)}$$

$$= na_0 \left(\frac{\sin\beta}{\beta}\right) = A\left(\frac{\sin\beta}{\beta}\right) \quad (5.4)$$

where we have written

$$A = na_0$$

and

$$\beta = \pi \frac{b\sin\theta}{\lambda} \quad 5.5$$

You will note that for a given wavelength, β signifies half of the phase difference between disturbances originating from the extreme points *A* and B. The expression for resultant field at P_θ takes the form

$$E_\theta = A\frac{\sin\beta}{\beta}\cos(\omega t - \beta) \quad (5.6)$$

The corresponding intensity distribution at P_θ is given by

$$I_\theta = A^2\left(\frac{\sin\beta}{\beta}\right)^2 \quad (5.7)$$

Let us pause for a while and ponder as to what have we achieved. This result suggests that the intensity is maximum at $\theta = 0$. This readily follows by noting that when we substitute $\theta = 0$ we have both β and $\sin\beta$ sin p equal to zero, but

$$\lim_{\beta \to 0} \frac{\sin \beta}{\beta} = 1$$

Therefore,

$$I_{\theta=0} = A^2$$

This result is expected on geometrical considerations. In the limits of a distant screen, the central point becomes equidistant from each point on the slit. All diffracted waves arrive in phase at P_0 and interfere constructively. A is then the value of the maximum intensity at the centre of the pattern. This maximum is also termed **principal maximum.**

For brevity we write $I_{\theta=0} = A^2 = I_0$. Then the intensity at any point at an angle θ with the horizontal axis, is given by

$$I = I_0 \left(\frac{\sin \beta}{\beta} \right)^2$$

Positions of maxima and minima

A plot of Eq. (5.7) for intensity distribution is shown in Fig. 5.4. You will note that the intensity is maximum for θ = 0: $I_{\theta=0} = I_0 = A^2$. The intensity gradually falls on either side of the principal maximum and becomes zero when $\beta = +\pi$ or $\beta = -\pi$, since $\sin \pm \pi$ is zero. This is the first minimum. So we can say that the angular half width of principal maximum is from 0 to π. The second minimum on either side occurs at $\beta = \pm 2\pi$. Thus we get the minima when

$$\begin{aligned} \beta &= \pm\pi,\ \pm 2\pi,\ \pm 3\pi,\ \ldots \\ &= m\pi,\ m = \pm 1,\ \pm 2,\ \pm 3,\ \ldots \end{aligned} \tag{5.8}$$

Note that the value m = 0 is excluded because it corresponds to the principal maximum (for β = 0). Substituting the value of β from Eq. (5.8) in Eq. (5.5) we find that the condition for minima is given by

$$\begin{aligned} b\sin\theta &= \pm\lambda,\ \pm 2\lambda,\ \pm 3\lambda,\ldots \\ &= m\lambda,\ m = \pm 1,\ \pm 2,\ \pm 3,\ \ldots \end{aligned} \tag{5.9}$$

You may now conclude that the angular width of the principal maximum (m = 1) is defined by $b\sin\theta = \lambda$ or $\theta = \dfrac{\lambda}{b}$. That is, θ depends upon the wavelength of light and the slit width. For a given slit width, the spread in diffraction pattern depends directly on the wavelength. Accordingly you should expect that red light would be diffracted through a larger angle than the blue or violet light.

You may now like to know: What will happen when white light illuminates a single slit? We expect that each wavelength will be diffracted independently. This gives rise to a white central spot surrounded by coloured fringes. The outer part of this pattern would tend to be reddish. You can easily observe this diffraction pattern by looking through the tines of a dinner fork at a candle in a dimly illuminated room. On twisting the fork about its handle, you will observe the diffraction pattern as soon as the cross-sectional area becomes small enough.

The expression $I_\theta = I_0\left(\frac{\sin\beta}{\beta}\right)^2$ gives the diffraction intensity in different directions. In order to determine the directions (and positions) of **secondary maxima,** we differentiate this equation with respect to β and equate the result to zero. This gives

$$\frac{dI_\theta}{d\beta} = 2I_0\left(\frac{\sin\beta}{\beta}\right)\left[\frac{\beta\cos\beta - \sin\beta}{\beta^2}\right]$$

$$= 2I_0 \sin\beta\left[\frac{\cos\beta}{\beta^2} - \frac{\sin\beta}{\beta^3}\right] = 0$$

or $\quad \sin\beta(\beta - \tan\beta) = 0$

From this we get the conditions $\sin\beta = 0$ and $\beta - \tan\beta = 0$. The condition $\sin\beta = 0$ implies that $\beta = \pm m\pi$, where m is any integer. This is a trivial condition as it signifies minima and is of no interest.

The condition $\beta = \tan\beta$ therefore gives the positions of secondary maxima. This is a transcendental equation. The roots of this equation can be found by a graphical method. All you have to do is to recall that an angle equals its tangent at intersections of the straight line

$$y = \beta$$

and the curve

$$y = \tan\beta \tag{5.10}$$

Plots of these curves are also shown in Fig. 5.4. The points of intersection excluding $\beta = 0$ (which corresponds to principal maximum) occur at β = 1.43π, 2.46π, 3.47π, etc. and give the position of the first, second, third maxima on either side of the central maximum. You should note that these maxjma do not fall midway between the two minima. For instance, the first maximum occurs at 1.43π rather than 1.50π. Similarly the second maxima occurs at 2.46π rather 2.50π and so on. This means that the intensity curves are asymmetrical. The plot clearly shows that the positions of maxima are slightly shifted towards the centre of the pattern. You may recall that this is observed experimentally as well.

Let us now calculate the intensities at these positions of maxima. The intensity of the first maximum is given by

$$\left(\frac{\sin 1.43\pi}{1.43\pi}\right)^2 = 0.0496$$

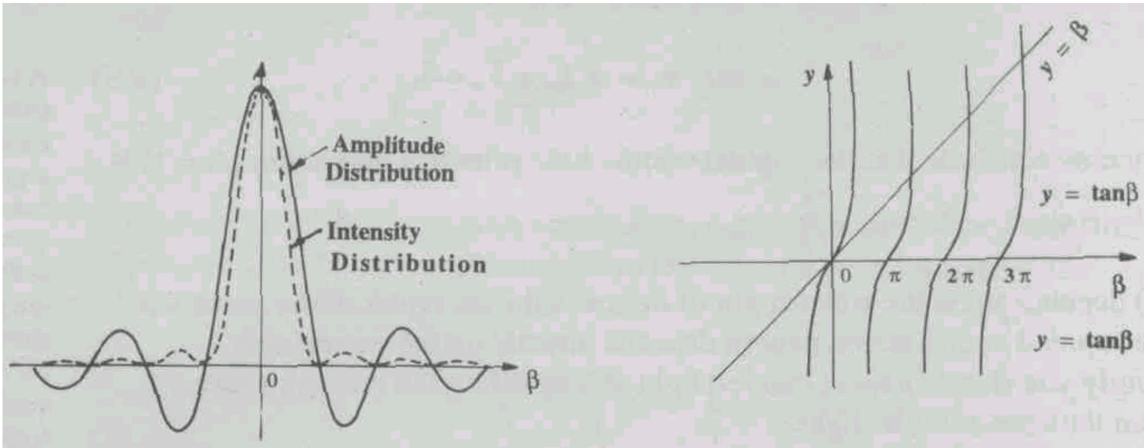


Fig. 5.4 Amplitude and intensity contours for Fraunhofer diffraction of a single slit showing positions of maxima and minima

This means that the intensity of the first secondary peak (nearest to the central peak) is about 4.96% of the central peak. Similarly, you can calculate and convince yourself that the intensities of the second and third maxima are about 1.68% and 0.83% of the central maximum. We call these maxima the secondary maxima.

The intensities of the secondary maxima can be calculated to a fairly close approximation by finding the values of β at halfway positions, i.e., at

$$\beta = \frac{\pi + 2\pi}{2}, \frac{2\pi + 3\pi}{2}, \frac{3\pi + 4\pi}{2}, \text{ etc.}$$

The intensities at these positions are

$$7\frac{4}{9\pi^2}, \frac{4}{25\pi^2}, \frac{4}{49\pi^2}, \ldots, \text{or } \frac{1}{22.1}, \frac{1}{61.7}, \frac{1}{121}, \ldots$$ of the central maximum which are

very close to the above calculated values. From this you may conclude that most of the light is concentrated in the central maximum.

Another important characteristic of the principal maximum is that its width is double of the width of secondary maximum. We have left its mathematical proof as an exercise for you. Before you proceed, you should solve SAQ 1.

SAQ 1
Show that the principal maximum is twice as wide as the secondary maxima.

To give you a feel for numerical values and fix up the ideas developed in this section, we now give a few solved examples. You should go through these carefully.

Example 1
In the experimental set up used to observe Fraunhofers diffraction of a vertical slit (width 0.3mm), the focal length of lens L_2 is 30 cm. Calculate (a) the diffraction angles and positions of

the first, second and third minima, and (b) the positions of the first, second and third maxima on either side of the central spot. The slit is illuminated with yellow sodium light which is a doublet. You may take λ = 6000 Å.

Solution

You have seen that the conditions for minima are given by $b\sin\theta = m\lambda$; m •= ± 1, ± 2, ± 3, ... For small values of θ, we may write $\sin\theta \approx \theta$. Then

$$\theta = m\frac{\lambda}{b}$$

and the distance P_0P_θ is $f\theta$, where f is the focal length. Therefore, the diffraction angles θ_1, θ_2, θ_3 for the first, second and third minima are $\frac{\lambda}{b}$, $2\frac{\lambda}{b}$ and $3\frac{\lambda}{b}$, respectively.

On substituting the numerical values of λ and b we find that

$$\theta_1 = \frac{6000\times10^{-8}\,cm}{0.3\times10^{-1}\,cm} = 2\times10^{-3}\,\text{rad}$$

$$\theta_2 = 2\theta_1 = 4\times10^{-3}\,\text{rad}$$

$$\theta_3 = 3\theta_1 = 6\times10^{-3}\ \text{rad}$$

The distances d_1, d_2, d_3 of these minima from the central spot are

$$d_1 = f\theta_1 = (30\,cm)\times2\times10^{-3} = 60\times10^{-2}\,cm = 0.06\text{ cm}$$

$$d_2 = 2f\theta_1 = 2\times0.06\text{ cm} = 0.12\text{ cm}$$

$$d_3 = 3f\theta_1 = 3\times0.06\text{ cm} = 0.18\text{ cm}$$

You will note that these minima are separated by a distance of 0.06 cm on the focal plane of the lens. We know that the first three secondary maxima occur at $\beta = 1.43\pi$, 2.46π and 3.47π, respectively. The corresponding diffraction angles for these three maxima are

$$(\theta_1)_{\max} = 1.43\frac{\lambda}{b},\ (\theta_2)_{\max} = 2.46\frac{\lambda}{b} \text{ and}$$

$\therefore$ $$(\theta_1)_{\max} = (1.43)(2\times10^{-3}),$$

$$(\theta_2)_{\max} = (2.46)(2\times10^{-3})$$

and

$$(\theta_3)_{\max} = (3.47)(2\times10^{-3})$$

and the corresponding distances from the central point (P_0) are

$$d_1 = f(\theta_1)_{\max} = (30\,cm) \times 1.43 \times 2 \times 10^{-3} = 0.09 \text{ cm}$$

$$d_2 = f(\theta_2)_{\max} = (30\,cm) \times 2.46 \times 2 \times 10^{-3} = 0.15 \text{ cm}$$

$$d_3 = f(\theta_3)_{\max} = (30\,cm) \times 3.47 \times 2 \times 10^{-3} = 0.21 \text{ cm}$$

Example 2

In the above experiment, we change slit widths to 0.2mm, 0.lmm, and 0.6mm. Calculate the positions of the first and second minima.

Solution

For slit width b = 0.2 mm, we have

$$d_1 = f\theta_1 = (30\,cm) \times \frac{6000 \times 10^{-8}\,cm}{0.2 \times 10^{-1}\,cm} = 0.09 \text{ cm}$$

Similarly

$$d_2 = f\theta_2 = 2 \times 0.09 \text{ cm} = 0.18 \text{ cm}$$

These minima are separated by 0.09 cm. Recall that the corresponding value for a slit of width 0.03 cm was 0.06 cm. This means that for a given wavelength, the spread of secondary maximum increases as slit width decreases. This conclusion is brought out in the following calculations as well.

For a slit of width b = 0.1 mm, we have

$$d_1 = (30\,cm) \times \frac{6000 \times 10^{-8}\,cm}{0.1 \times 10^{-1}\,cm} = 0.18 \text{ cm}$$

and

$$d_2 = 2 \times 0.18 \text{ cm} = 0.36 \text{ cm}$$

For slit width b = 0.06 mm, we have

$$d_1 = (30\,cm) \times \frac{6000 \times 10^{-8}\,cm}{0.6 \times 10^{-1}\,cm} = 0.3 \text{ cm}$$

and

$$d_2 = 2 \times 0.3 \text{ cm} = 0.6 \text{ cm}$$

We thus find that for slits of widths 0.3mm, 0.2mm 0.1mm, and 0.06mm, the first minimum on either side of the principal maximum occurs at distances of 0.06 cm, 0.09 cm, 0.18 cm, and 0.3 cm. In these four cases, the corresponding principal maximum extends over 0.12 cm, 0.18 cm, 0.36 cm, and 0.6 cm.

This shows that as the slit becomes narrower, the spread of central maximum increases. Conversely, the wider the slit width, the narrower is the central diffraction maximum.

We now consider an interesting case where the width of the slit is varied in comparison to the wavelength of light.

Example 3

Consider a slit of width $b = 10\lambda$, 5λ, and λ. Calculate the spread of the central maximum.

Solution

From Eq. (5.9), we note that for a slit of width $b = 10\lambda$, the first minimum is located at

$$10\lambda \sin\theta = \lambda$$

or $\quad \sin\theta = 0.10$

and $\quad \theta = 5.7°$

For a slit of width 5λ, we have

$$5\lambda \sin\theta = \lambda$$

or $\quad \theta = 90°$

That is, as the aperture of the slit changes from 10λ to 5λ, the diffraction pattern spreads out about twice as far. For $b = \lambda$,

$$\sin\theta = 1$$

or $\quad \theta = 90°$

The first minimum falls at 90°. That is, the central maximum spreads out and the diffraction pattern shows no ripple. These features are shown in Fig. 5.5.

You may now like to answer an SAQ.

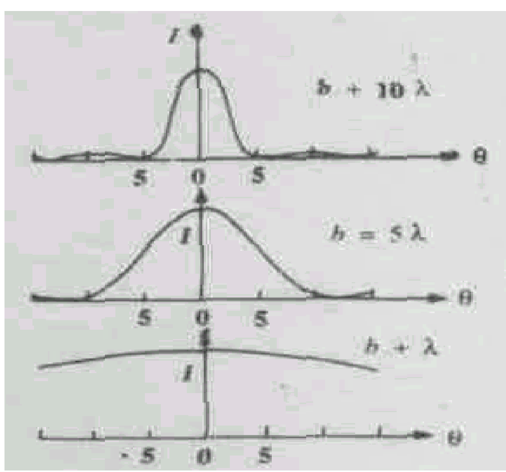


Fig. 5.5 Single- slit diffraction irradiances as the silt width varies

SAQ 2

We illuminate the slit of Example 1 with violet light of wavelength 4358 Å from a mercury lamp. Show that the diffraction pattern shrinks correspondingly.

Diffraction Pattern of a Rectangular Aperture

So far we have described Fraunhofer diffraction pattern of a slit aperture. Let us now consider what will happen if both dimensions of the slit are made comparable. We now have a rectangular aperture of width *b* and height *a* as shown in Fig 5.6 (a). We expect that the emergent wave will spread along the length as well as the width of the slit. Can you depict the diffraction pattern? It is shown in Fig. 5.6 (b). Mathematically, the intensity is given by $I = \dfrac{I_0 \sin^2 \alpha \sin^2 \beta}{\alpha^2 \beta^2}$

where, $\beta = b\pi \sin\theta / \lambda$ and $\alpha = \pi a \sin\theta / \lambda$.

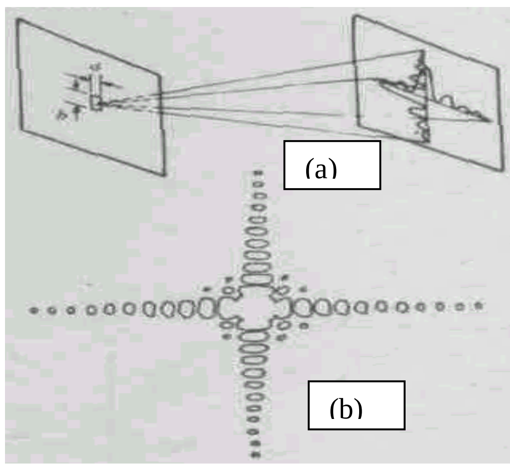


Fig. 5.6 Single-slit diffraction. Both dimensions of the rectangular aperture are small and a two-dimensional diffraction pattern is discernible on the screen (b) Diffraction Image of a single square aperture.

Slit Source

The experimental arrangement shown in Fig. 5.1 is modified as shown in Fig. 5.7. Here instead of the point source we use a slit source (Fig. 5.7(a)).

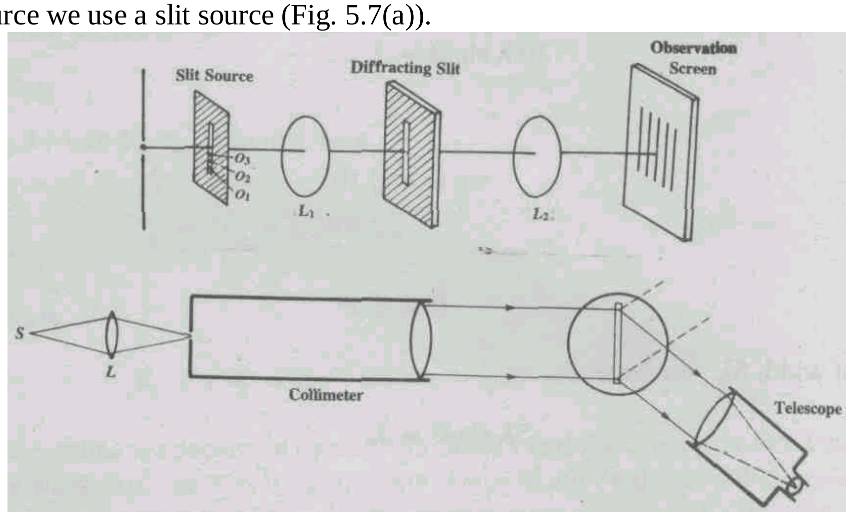


Fig. 5.7 (a) Experimental arrangement for diffraction from a vertical narrow single silt Illuminated by a silt source (b) Experimental arrangement in a physics laboratory.

As a matter of fact, the experimental arrangement, which is commonly employed in most experiments, uses a spectrometer (Fig. 5.7(b)). The slit of the collimator arm is illuminated so that each point of the slit source acts as an independent source. You know that a point source gives a horizontal streak of light as the diffraction pattern of a vertical slit. Now when we substitute a slit as a source, we can imagine a series of point sources O_1, O_2 , O_3, etc, one above the other to form the slit source (Fig. 5.7(a)). Each point source will give its own diffraction pattern since each point is to be regarded as an independent point source. With the same diffracting slit and the same lenses L_1 and L_2, the central diffraction maximum due to all point sources will lie above one another and give a central bright vertical fringe. Similarly from secondary maxima and minima points, we will obtain a series of vertical fringes, which will be situated at equal intervals on either side of the central fringe. The resulting pattern arises by superposition of a series of horizontal diffraction streaks stacked on each other in a vertical direction. The intensity along any horizontal line will be the same as in Fig. 5.2. We should note that each point of the slit source acts as an independent and effectively as a non-coherent source.

You will observe that clear fringes are obtained only when the width of the source slit is small. Suppose that the width of the source slit is gradually increased. This will lead to an increase in the width of its image on the observation screen. A stage will come when the width of the image, i.e., the fringe width, becomes comparable with the distances between successive vertical fringes. This will gradually make the vertical fringes less clear and indistinct. For a similar reason, we obtain clear fringes only when the source slit is parallel to the diffraction slit.

9.3 DIFFRACTION BY A CIRCULAR APERTURE

Fraunhofer diffraction by a circular aperture is of particular interest because a lens in an optical device (microscope, telescope, the eye) can be regarded as a circular aperture. For this case, the experimental arrangement is shown in Fig. 5.8(a). A plane wave is incident normally on the aperture and a lens whose diameter is much larger than that of the aperture is placed close to it. The Fraunhofer diffraction pattern is observed on the back focal plane of the lens. Because of the rotational symmetry of the system, we expect that the diffraction pattern will consist of concentric dark and bright rings. Fig. 5.8(b) shows the diffraction pattern, which is known as the Airy pattern.

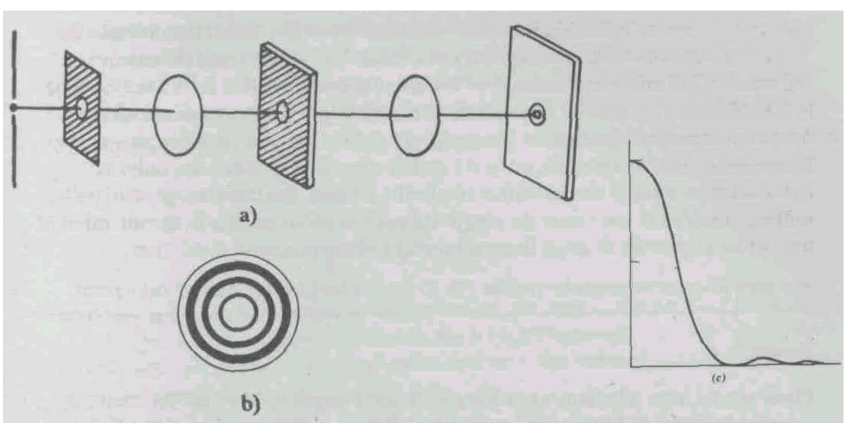


Fig. 5.8 (a) Experimental arrangement for observing the Fraunhofer diffraction pattern by a circular aperture, (b) The Airy pattern: The circle of light at the centre corresponds to the zeroth order, (c) The corresponding Intensity distribution

The detailed derivation of the diffraction pattern for a circular aperture involves complicated mathematics. So we just quote the final result for the intensity distribution:

$$I = I_0 \left[\frac{2J_1(\gamma)}{\gamma} \right]^2 \tag{5.11}$$

where

$$\gamma = \frac{\pi D}{\lambda} \sin\theta \tag{5.12}$$

Here D is the diameter of the aperture, λ is the wavelength of light and θ is the angle of diffraction, I_0 is the intensity at θ = 0 (which represents the central maximum) and $J_1(\gamma)$ is the Bessel function of the first order. (We know that you are not very familiar with Bessel functions.) We may just mention that the variation of $J_1(\gamma)$ is somewhat like a damped sine curve. Moreover, the intensity is maximum at the centre of the pattern since

$$\lim_{\gamma \to 0} \frac{2J_1(\gamma)}{\gamma} \to 1$$

similar to the relation

$$\lim_{\beta \to 0} \frac{\sin\beta}{\beta} \to 1$$

Other zeros of $J_1(\gamma)$ occur at γ = 3.832, 5.136, 7.016, ... which correspond to the successive dark circles in the Airy pattern. Thus the first dark ring appears when

$$\sin\theta = \frac{3.832\lambda}{\pi D} \approx \frac{1.22\lambda}{D} \tag{5.13}$$

Let us compare this result with the analogous equation for the narrow slit. We find that the angular half-width of the central disc, i.e., the angle between the central maximum and the first minimum of the circular aperture, differs from that for the slit pattern through the weird number 1.22. The intensity distribution of Eq. (5.11) is plotted in Fig. 5.8(c). The pattern is similar to that for a slit, except that the pattern for circular apertures now has rotational symmetry about the optical axis. The central maximum is consequently a circular disc of light, which may be regarded as the diffracted "image" of the circular aperture. It is called the **Airy disc.** It is surrounded by a series of alternate dark and bright fringes of decreasing intensity. However, the pattern is not sharply defined. If you consider any section through the circular aperture, intensity distribution is very much the same as obtained from a point source with a single slit. Indeed, the circular aperture pattern will be obtained if you rotate the single slit pattern about an axis in the direction of the light and passing through the central point of the principal maximum.

We now give an example to enable you to have a feel for the numerical values.

Example 4
Plane waves from a helium-neon laser with wavelength 6300 Å are incident on a circular aperture of diameter 0.5 mm. What is the angular location of the first minimum in the diffraction pattern? Also calculate the diameter of Airy disc on a screen 10m behind the aperture.

Solution
We know from Eq. (5.13) that

$$D \sin\theta = 1.22\lambda$$

On substituting the given values, we get

$$(0.5\times10^{-3}\, m)\sin\theta = 1.22\times630\times10^{-9}\text{ m}$$

or

$$\sin\theta = \frac{1.22\times630\times10^{-9}}{0.5\times10^{-3}} = 1.54\times10^{-3}$$

In the small angle approximation, $\sin\theta \approx \theta$, so that

$$\theta = 1.54\times10^{-3}\text{ rad} = 0.087°$$

On the screen placed 10m away, the linear location of the first minimum is

$$x = D\tan\theta \approx D\sin\theta \approx D\theta$$

Hence

$$x = (10\, m)\times(1.54\times10^{-3}\, rad) = 15.4\times10^{-3}\text{ m} = 1.54\text{ m}$$

This value of *x* signifies the radius of the Airy disc so that the diameter is about 3 cm.

You can observe a white light circular diffraction pattern by making a small pinhole in a sheet of aluminium foil. Then look through it at a distant light bulb or a candle standing in a poorly illuminated (dark) room.

Another important result of the above analysis is that the angular width of a beam is diffraction-limited. When a perfectly plane wave from a distant point source is incident on a diffracting aperture (of width or diameter *b)*, the angular width of the diffracted beam is λ / b. This is illustrated in Fig. 5.9. The angular width can be zero if *b* is infinite (1mm or so). At large distances from the diffracting aperture, beam width $W = L(\lambda / b)$. It has important implications for laser beams, which are known to be highly directional. To have an idea about it, let us consider a diffraction-limited laser beam (λ = 6000 Å) of 2 mm diameter. The angular spread of the beam is

$$\theta = \frac{\lambda}{b} = \frac{6\times10^{-5}\,cm}{0.2\,cm} = 3\times10^{-4} \text{ rad}$$

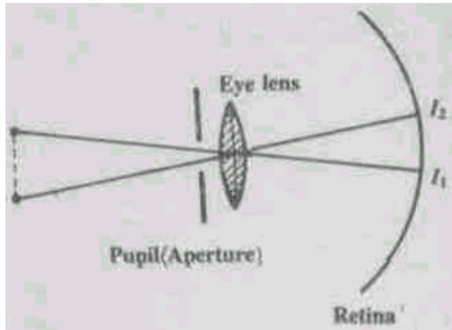


Fig. 5.9 Schematics of a diffraction limited system

It means that in an auditorium (of length 15 m), the spatial spread $W = (1500\,cm)\times(3\times10^{-4}) = 5$ mm, which is very small. For a typical penlight type flashlight, the transverse dimensions of the filament should be of the order of a micrometer, which is really hard to make.

Imagine that a random array of small circular apertures is illuminated by plane waves from a white point source. We know that each aperture will generate an Airy type diffraction pattern. If the apertures are small and close together, the diffraction patterns are large and overlap. The overlapping diffraction patterns produce a readily visible halo, namely, a central white disc surrounded by circular coloured rings. Which colour do you expect to be at the outermost rim? Should it not be red? Similar halos are also observed when the diffraction is due to a random array of circular obstacles.

Suspended water (n = 1.33) droplets in air (n = 1.00) give rise to diffraction halos. When observed through a light cloud cover around the sun or moon, these diffraction halos are referred to as coronas. We can distinguish between diffraction halos and ice crystal halos. Ice crystal halos are due to refraction and dispersion by the ice crystals; they have red on the inside of the rings.
While driving a car at night, you may have seen brilliant halos through fogged up car windows on which light of a motorcycle following you is incident. These are diffraction halos. You can easily produce such halos by breathing on the side of a clear glass and then looking through the fogged area at a small source (e.g., match, penlight, or distant bulb).

When the cornea swells (becomes oedematous), small droplets of fluid form randomly between the stromal fibres. These random droplets produce a diffraction halo that the person sees when looking at light. Such halos are one of the warning signs of high ocular pressures. These halos can also be produced by epithelial damage due to poorly fitting contact lenses.

5.4 SUMMARY

- To observe Fraunhofer diffraction pattern, the distance of the diffracting screen from the source and/or observation screen should be almost infinite. Experimentally this condition is achieved by using convergent lenses.

- The diffraction pattern of a vertical slit consists of a horizontal streak of light. This horizontal diffraction pattern may be regarded as a spread out image of the point source and consists of a series of diffraction spots symmetrically situated with respect to the central point.

- The central spot has a maximum intensity and its width is twice compared to other spots which are of equal width. Their intensities decrease rapidly. In fact, most of the light is concentrated in the central maximum.

- The plane wavefront incident on the slit gives rise to a system of vertical plane wavefronts which originate from each point of the diffracting aperture.

- The intensity at any point P_θ on the screen is computed by taking the phase difference between the successive diffracted waves into account. The intensity at a point P_θ is given by

$$I = I_0\left(\frac{\sin\beta}{\beta}\right)^2$$

where $\beta = \pi\dfrac{b\sin\theta}{\lambda}$ and b is the width of the slit.

- If the path difference *b* sin6 between waves diffracted by extreme ends of the slit is an integral multiple of λ, we obtain zero intensity.

- The diffraction pattern of a thin slit source consists of a series of vertical fringes. In this pattern, the central vertical fringe is the brightest and the intensity of other fringes decreases rapidly. The width of central fringe is double of that for other fringes.

- The diffraction pattern of a circular aperture consists of concentric rings with a central bright disc. The first dark ring appears when $\sin\theta = 1.22\lambda / D$.

5.5 TERMINAL QUESTIONS

1. A single slit has a width of 0.03 mm. A parallel beam of light of wavelength 5500 Å, is incident normally on it. A lens is mounted behind the slit and focussed on a screen located in its focal plane, 100 cm away. Calculate the distance of the third minimum from the centre of the diffraction pattern of the slit.

2. A helium-neon laser emits a diffraction-limited beam (λ = 6300 Å) of diameter 2 mm. What diameter of light patch would the beam produce on the surface of the moon at a distance of 376×10^3km from the earth? You may neglect scattering in the earth's atmosphere.

5.6 SOLUTIONS AND ANSWERS

SAQs

1. We know that angular spread of the central maximum is from

$$\theta = \sin^{-1}\left(\frac{\lambda}{b}\right) \text{ to } \theta = -\sin^{-1}\left(\frac{\lambda}{b}\right).$$

For small θ, we have $\sin\theta = \theta$ and we find that principal maximum is spread from $\theta = \frac{\lambda}{b}$ to $\theta = -\frac{\lambda}{b}$.

Similarly, you can show that the first secondary maximum on the positive side extends from $\theta_1 = \frac{\lambda}{b}$ to $\theta_1 = 2\frac{\lambda}{b}$ and on the negative side from $\theta = -\frac{\lambda}{b}$ to $\theta = -2\frac{\lambda}{b}$

Thus we see that the central maximum is twice as wide as a secondary maximum

2. We know that

$$d \sin\theta_1 = \lambda$$

$$\therefore \quad (0.3\times10^{-1}\ cm)\sin\theta_1 = 4358\times10^{-8}\ \text{cm}$$

In the small angle approximation we can lake

$$\theta_1 = 1.45\times10^{-3}\ \text{rad}$$

and

$$\theta_2 = 2.90\times10^{-3}\ \text{rad}$$

On comparing these values with those given in Example 1 for the first and second minima you will note that violet light is diffracted about 27% less.

TQs

1. From Eq. (5.9) we know that the conditions for minima are given by

$$b\sin\theta = n\lambda\ ;\ n = \pm 1,\ \pm 2,\ ...$$

Here $b = 0.03$ mm = 3×10^{-3} cm , n = 3 and λ = 5500 Å

$$\therefore \quad \sin\theta = \frac{n\lambda}{b} = \frac{3\times(5500\times10^{-8}\ cm)}{3\times10^{-3}\ cm} = 5.5\times10^{-4}$$

In the small angle approximation, $\sin\theta \approx \theta \approx \tan\theta$.

$$\therefore \quad x = 5.5\times10^{-4}\times(100\,cm)$$
$$= 5.5\times10^{-2}\ \text{cm}$$

2. Suppose that the light patch on the Moon is taken to be an Airy disc of diameter x of a diffraction limited beam of initial diameter 2 mm. Then using Eq. (5.13) we can write

$$\sin\theta = \frac{1.22\lambda}{D} = \frac{1.22\times(6300\times10^{-8}\,cm)}{(0.2\,cm)}$$

$$= 384.3\times10^{-6}$$

In the small angle approximation, $\sin\theta \approx \theta = 384\times10^{-6}$ rad. Since $x = 2r\theta$, we find on substituting the numerical values that

$$x = 2\times(376\times10^{3}\,km)\times(384.3\times10^{-6})$$

$$= 289 \text{ km}$$

UNIT 6 DIFFRACTION GRATING

Structure

6.1 INTRODUCTION

You have learnt about Fraunhofer diffraction produced by a single slit aperture. When a narrow vertical slit is illuminated by a distant point source, the Fraunhofer diffraction pattern consists of a series of spots situated symmetrically about a central spot, along a horizontal line. The intensity of the central spot is maximum and it decreases rapidly as we move away from the central spot. For a circular aperture, the diffraction pattern consists of concentric rings with a bright central disc. You also learnt that diffraction phenomenon limits the ability of optical devices to form sharp and distinct images of distinct objects. This restriction at one time hampered the spectroscopic work, particularly for substances whose spectrum consisted of doublets. (Sodium doublet wavelengths correspond to 5890Å and 5896Å. Because of their proximity, these wavelengths seem to overlap.) But you will recall that diffraction pattern is sensitive to wavelength of light as well as the slit width. To take advantage of these, it was thought that the problem could be overcome by increasing the number of diffracting slits. And the idea really worked. For simplicity, we have first discussed diffraction pattern by a double slit.

In Sec. 6.2 we have listed qualitative features of the observed double slit diffraction pattern and compared these with those of a single slit pattern. A distinct feature of double slit pattern is that it consists of bright and dark fringes similar to those observed in interference experiments. In Sec. 6.3 we have derived the equation for the resultant intensity distribution. This mathematical analysis is an extension of what you have already learnt for the single slit. You will learn that the intensity of the central maximum is four times the intensity due to either slit at that point. However, the interference maxima are diffused (broader). These results are generalised for the case of *N* equally spaced, identical slits in Sec. 6.4.

You will observe that as the number of slits increases, interference maxima get narrower (sharper). For sufficiently large value of *N*, interference maxima become narrow lines, For this reason, diffraction gratings are an excellent tool in spectral analysis. The occurrence of diffraction grating effects in nature is surprisingly common. Do you know that the green on the neck of a male mallard duck, blue appearance of wings of Morpho butterflies and the beautiful colours of the '*eye*' of the peacock's feathers are also due to diffraction grating effects? The layered structure in cat's retina acts as reflection grating and is responsible for metallic green reflection at night.

Objectives

After studying this unit, you should be able to

- state salient features of the double slit diffraction pattern
- qualitatively compare single-slit diffraction
- pattern with double and N-slit patterns
- derive equation for the intensity distribution for the double slit pattern
- extend the double-slit calculation for N equally spaced slits
- describe the use of a diffraction grating in spectral analysis, and
- solve numerical examples.

6.2 OBSERVING DIFFRACTION FROM TWO VERTICAL SLITS

Refer to Fig. 6.1. It shows the experimental arrangement for observing diffraction from two vertical parallel slit-apertures in an opaque screen. Both the slits have same width (b) and height (h). The width of the intervening opaque space between the two slits is a. Therefore, the distance between two similar points in these apertures $d = b + a$.

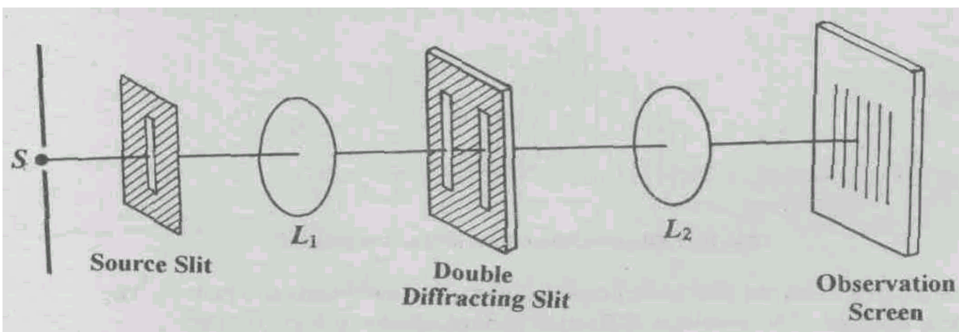


Fig. 6.1 Experimental arrangement for observing diffraction from two identical vertical slits

Have you noticed that diffracting apertures are illuminated by a slit source rather than a point source of light? We have used this arrangement because this corresponds more nearly to the actual conditions under which an experiment is performed. That is, the diffraction pattern from a slit source is of greater practical importance than that from a point source. The ray geometry of Fig. 6.1 for observing Fraunhofer diffraction from a double slit illuminated by a slit source is shown in Fig. 6.2. The length of the source slit in the arrangement should be adjusted to be parallel to the lengths of the diffracting slits.

Suppose we block one of the diffracting slits, say slit 1, shown in Fig. 6.1 and observe the diffraction pattern on the screen. Obviously, you should expect the single slit diffraction pattern (due to slit number 2 which has not been blocked). Next, uncover slit 1 and block the other. You should again expect single slit diffraction pattern with exactly the same intensity distribution. But what may surprise you at the first glance is that both diffraction patterns are not only identical, they are located at the same position.

> In a well-corrected lens consider parallel beams of light travelling in a direction inclined to the axis from different parts of the lens. They are all brought to focus on the back focal plane at a point which is located by the beam passing though the optical centre of the lens.

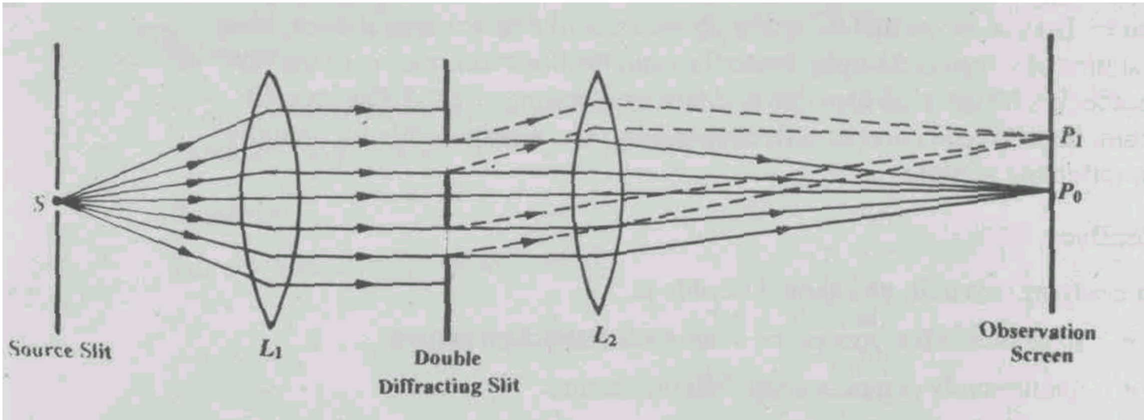


Fig. 6.2 Ray geometry or experimental arrangement shown in Fig. 6.1

Were you not expecting these diffraction patterns to be laterally displaced? These patterns are not laterally shifted with respect to one another because of the (well corrected) lens L_2. This is true even for N identical vertical slits. The diffracted wavefronts originating from any slit, and travelling along the axis of lens L_2 are focussed at P_0, which forms the peak of the central spot. The diffracted wavelets moving at an angle θ are focussed at P_θ.

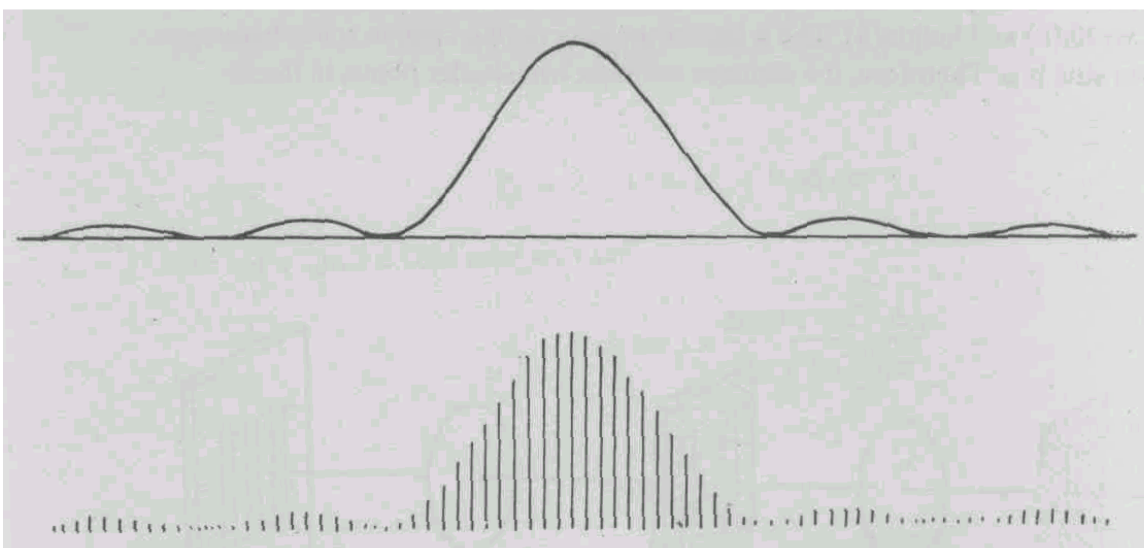


Fig. 6.3 Observed double slit diffraction pattern

Now uncover both the slits so that each slit gives its own diffraction pattern. The salient features of the resultant diffraction pattern, shown in Fig. 6.3, are summarised below:

(i) The double slit diffraction pattern consists of a number of equally spaced fringes similar to what is observed in interference experiments.

(ii) The intensities of all fringes are not equal. The fringes are the brightest in the central part of the pattern.

(iii) As we move away on either side of the central fringe, the intensity gradually falls off to zero.

(iv) The fringes reappear with reduced intensity three or four times and become too faint to observe thereafter.

(v) The intensity at the maximum of double slit pattern is greater than the intensity of principal maximum in single slit pattern.

What is responsible for this pattern? How bright are double slit fringes compared to those in the single slit pattern? You will discover answers to these and other related questions in the following section.

6.3 INTENSITY DISTRIBUTION IN DOUBLE SLIT PATTERN

For calculating the intensity distribution for the arrangement shown in Fig. 6.1 it is sufficient for us to consider a point source. This is because a point source gives the intensity distribution along a section perpendicular to the vertical fringes formed from a slit source. For deriving the equation for intensity of double slit pattern, we extend the procedure used for the single slit (Unit 5). Slit 1 acts as a source of diffracted plane wavefronts originating from points A_1, A_2, A_3, ... in it. We represent these by $a_0 \cos \omega t$, $a_0 \cos(\omega t - \phi)$, $a_0 \cos(\omega t - 2\phi)$, …, where ϕ is the constant phase difference. The magnitude of electric field E_1 produced by this slit at the point P_θ is given by (Eq. 5.6):

$$E_1 = A\left(\frac{\sin \beta}{\beta}\right)\cos(\omega t - \beta)$$

where $\beta = \dfrac{\pi b \sin \theta}{\lambda}$.

For every point like A_1 in slit 1, we have a corresponding point B_1 in slit 2 at a distance *d*. The phase difference between diffracted wavefronts reaching P_θ from A_1 and B_1 is given by

$$\delta = \frac{2\pi}{\lambda}(a+b)\sin\theta = \frac{2\pi}{\lambda} d \sin\theta \tag{6.2}$$

Therefore, the diffracted plane wavefronts starting from points B_1, B_2, B_3, ... may be represented as $a_0 \cos(\omega t - \delta)$, $a_0 \cos(\omega t - \delta - \phi)$, $a_0 \cos(\omega t - \delta - 2\phi)$, … And the field E_2 produced by slit 2 at P_θ is given by

$$E_2 = A\left(\frac{\sin \beta}{\beta}\right)\cos[(\omega t - \delta) - \beta] \tag{6.3}$$

Since the sources A_1, A_2, A_3, ... and B_1, B_2, B_3, ... are coherent, the magnitude of resultant field at P_θ due to the double-slit is obtained by the superposition of magnitudes of individual fields:

$$
\begin{aligned}
E &= E_1 + E_2 \\
&= A\left(\frac{\sin\beta}{\beta}\right)[\cos(\omega t - \beta) + \cos(\omega t - \beta - \delta)]
\end{aligned}
$$

Using the trigonometric identity $\cos A + \cos B = 2\cos\left(\frac{A+B}{2}\right)\cos\left(\frac{A-B}{2}\right)$, we can rewrite the above expression as

$$
\begin{aligned}
E &= 2A\left(\frac{\sin\beta}{\beta}\right)\cos\left[(\omega t - \beta) - \frac{\delta}{2}\right]\cos\left(\frac{\delta}{2}\right) \\
&= 2A\left(\frac{\sin\beta}{\beta}\right)\cos(\omega t - \beta - \gamma)\cos\gamma \qquad (6.4)
\end{aligned}
$$

where $\gamma = \frac{\delta}{2} = \frac{\pi}{\lambda} d \sin\theta$.

The intensity is proportional to the square of the amplitude. So

$$
I_\theta = 4A^2\left(\frac{\sin\beta}{\beta}\right)^2 \cos^2\gamma \qquad (6.5)
$$

For $\theta = 0$, both β and γ vanish so that

$$
I_{\theta=0} = 4A^2 = 4I_0
$$

and the expression for intensity of double slit diffraction pattern can be written as

$$
I_0 = 4I_0\left(\frac{\sin\beta}{\beta}\right)^2 \cos^2\gamma \qquad (6.6)
$$

Since the maximum value of I_θ is $4I_0$, we see that the double slit provides **four times** as much intensity in the central maximum as the single slit. This is exactly what you should have expected since the incident beams are in phase and amplitudes superpose.

If you closely examine Eq. (10.6) you will recognise that the term $(\sin^2\beta / \beta^2)$ represents the diffraction pattern produced by a single slit of width *b*. The $\cos^2\gamma$ term represents the interference pattern produced by two diffracted beams (of equal intensity) having phase difference δ. That is, the intensity of double slit diffraction pattern is product of the irradiances observed for the double-slit interference and single slit diffraction. For $a > b$, the $\cos^2\gamma$ factor will vary more rapidly than the $(\sin^2\beta / \beta^2)$ factor. Then **we obtain Young's interference**

pattern for slits of very small widths. In general, the product of sine and cosine factors may be considered as a modulation of the interference pattern by a single slit diffraction envelope. We shall discuss it in detail a little later.

Before we investigate the positions of maxima and minima, let us understand the physical phenomenon that takes place. Diffracted light emerging from these two slits constitutes two coherent beams. These interfere leading to the formation of fringes on the screen. But the intensity of a fringe depends upon the intensities of interfering beams and the phase difference between them when they reach the point under observation. We know that the intensities of diffracted beams are controlled by the diffraction conditions and the direction of observation. Consequently, the intensities of interference fringes are not the same at different points of the screen. In particular, in those directions in which the intensities of diffracted beams are large, the constructive interference will lead to brighter fringes whereas in directions where the two diffracted beams themselves have lower intensities, even their constructive interference will lead to faint fringes.

You should note that we have described the phenomenon as **interference between two diffracted beams.** How do we distinguish between the two words interference and diffraction, which we have used? When secondary wavelets originating from different parts of the same wavefront are made to superimpose, we call it diffraction. Such a case arises when we consider all the wavelets arising from the various points situated in the aperture between the two jaws of a slit. But when two separate beams coming from two different slits are superimposed, we call it interference. It should be clear that in all cases where we apply the principle of superposition, the wavelets have to be coherent in nature to produce an observable pattern.

Before you proceed, you may like to answer an SAQ.

SAQ1

If instead of a monochromatic source we use a source emitting two wavelengths, λ_1 and $\lambda_2\ (< \lambda_1)$, how will the double slit diffraction pattern get influenced?

6.3.1 Positions of Minima and Maxima

To study the position of minima and maxima in the double slit pattern, we use the equation

$$I_\theta = 4I_0\left(\frac{\sin\beta}{\beta}\right)^2 \cos^2\gamma$$

We note that the intensity I_θ will be zero when either $(\sin\beta/\beta)^2$ or $\cos^2\gamma$ is zero. From Unit 5 you will recall that the factor $(\sin\beta/\beta)^2$ will be zero for

$$\beta = \frac{\pi\, b\sin\theta}{\lambda} = \pi,\ 2\pi,\ 3\pi,\ \ldots,\ m\ \pi,\ (m \neq 0)$$

or

$$b\sin\theta = \lambda,\ 2\lambda,\ 3\lambda,\ \ldots,\ m\ \lambda \tag{6.7}$$

This equation specifies the directions along which the available intensity of either beam is zero by virtue of diffraction taking place at each slit.

The second factor ($\cos^2\gamma$) will be zero when

$$\gamma = \frac{\pi\, d\sin\theta}{\lambda} = \frac{\pi}{2},\ \frac{3\pi}{2},\ \frac{5\pi}{2},\ \ldots,\ \left(n+\frac{1}{2}\right)\pi$$

or

$$d\sin\theta = \frac{\lambda}{2},\ \frac{3\lambda}{2},\ \frac{5\lambda}{2},\ \ldots,\ \left(n+\frac{1}{2}\right)\lambda \qquad (6.8)$$

This gives the angles for the intensity to be zero by virtue of destructive interference between two beams. You may recall that this is the same as the condition for the minimum of the interference pattern between two point sources. Eqs. (6.7) and (6.8) specify the direction where the intensity is zero.

We cannot obtain the exact positions of the maxima by any simple relation. This is because we have to find the maximum of a function which is product of two terms. But we can find their approximate positions if we assume that $(\sin\beta/\beta)$ does not vary appreciably over a given region. We are quite justified in making this approximation if the slits are very narrow. Note that we observe the maxima near the centre of the pattern. Under these conditions, the positions of maxima are solely determined by the $\cos^2\gamma$ factor. You know that this factor defines maxima for

$$\gamma = 0,\ \pi,\ 2\pi,\ \ldots,\ n\,\pi$$

or

$$d\sin\theta = 0,\ \lambda,\ 2\lambda,\ \ldots,\ n\,\lambda \qquad (6.9)$$

We know that $d\sin\theta$ represents the path difference between the corresponding points in the two slits. When this path difference is a whole number of wavelengths, constructive interference occurs between the two beams. Then we get a maximum, which leads to the formation of a series of bright fringes. The central fringe corresponds to $d\sin 0 = 0$. The *n*th fringe (on either side) occurs when $d\sin\theta = n\lambda$. We therefore say that *n* represents the **order** of interference.

6.3.2 Missing Orders

In the intensity expression $I_\theta = 4I_0\left(\frac{\sin\beta}{\beta}\right)^2\cos^2\gamma$, we have $\beta = \frac{\pi\, b\sin\theta}{\lambda}$ and $\gamma = \frac{\pi\, d\sin\theta}{\lambda}$. Thus we see that β and λ are not independent. These are connected to each other through the relation

$$\frac{\gamma}{\beta} = \frac{\pi\, d\sin\theta}{\pi\, b\sin\theta} = \frac{d}{b} = \frac{a+b}{b} \qquad (6.10)$$

Cases of special interest arise when *d* is an integral multiple of *b,* say it is an integer p so that $d = pb$. This will happen when the opaque portion a is an integral multiple of the transparent part *b*. The possibilities are: *a* = *b*, *a* = *2b*, or *a* = *3b* etc, so that *d/b* = *p* = 2, 3, 4, ... etc in these cases. Under these conditions, the directions of diffraction minimum and interference maximum will necessarily coincide. To show this, let us assume that a direction of diffraction minimum is given by

$$b \sin\theta = m\lambda$$

We will automatically have the interference maximum in this direction since

$$d \sin\theta = (pb)\sin\theta = pb\sin\theta$$
$$= pm\lambda = n\lambda$$

where $n = pm$. The possible values of p are 2, 3, 4, ... and those of m are 1, 2, 3, ... Thus the nth order interference fringes for which $n = pm$ will have zero intensity since the intensity of both beams is zero by virtue of the diffraction condition. As a result, their constructive interference also leads to net zero intensity. These are usually known as missing orders. For example, when $p = 2$, we will have 2, 4, 6, 8, ... orders missing for m values of 1, 2, 3, ... etc. Similarly, when $p = 3$, we will have 3, 6, 9, ... orders missing and so on.

The special case when $d/b = 1$, means that the opaque part $a = 0$ and the two slits exactly join one another. Then we find that all the interference orders are missing. Actually this means that we now have a single slit of double width and what we get is a single slit diffraction pattern and (with no interference fringes).

These ideas are illustrated in the following example.

Example 1
Consider a double slit arrangement with $b = 7.0\times10^{-3}$ cm, $d = 3.5\times10^{-2}$ cm and λ = 6300 Å. How many interference minima will occur between the diffraction minima on either side of the central maximum? If a screen is placed at a distance of 5m from the diffracting aperture, what is the fringe width?

Solution
The first diffraction minima on either side will occur when $b\sin\theta = \pm\lambda$. That is, for $\sin\theta = \pm\lambda / b = 9\times10^{-3}$. The interference b minima will occur when Eq. (6.8) is satisfied, i.e. when

$$1\ d \sin d \sin\theta = \left(n + \frac{1}{2}\right)\lambda$$

On substituting the given values, we find that

$$\sin\theta = \left(n + \frac{1}{2}\right)\frac{\lambda}{d} = \left(n + \frac{1}{2}\right)1.8\times10^{-3},\ n = 0,\ 1,\ 2,\ \ldots$$

i.e.

$$\sin\theta = 0.9\times10^{-3},\ 2.7\times10^{-3},\ 4.5\times10^{-3},\ 6.3\times10^{-3}\ \text{and}\ 8.1\times10^{-3}$$

Thus, there will be ten minima between the two first order diffraction minima. If θ is small we may write $\theta_1 = 0.9\times10^{-3}$ rad, $\theta_2 = 2.7\times10^{-3}$ rad, $\theta_3 = 4.5\times10^{-3}$ rad, $\theta_4 = 6.3\times10^{-3}$ rad, $\theta_5 = 8.1\times10^{-3}$ and the angle between successive minima is 1.8×10^{-3} rad. The angular separation between two interference maxima is given by

$$\Delta\theta = \frac{\lambda}{d} = \frac{6.3\times10^{-5}\,cm}{3.5\times10^{-2}\,cm} = 1.8\times10^{-3} \text{ rad.}$$

Note that this is the same as the angle between successive minima. Thus the fringe width $f\cdot\Delta\theta\, d$ A8d is

$$(500 \text{ cm})\times1.8\times10^{-3} = 0.9 \text{ cm}$$

6.3.3 Graphical Representation

We will now plot $\cos^2\gamma$, $(\sin^2\beta/\beta^2)$, and their product separately to study the double slit pattern. Before doing that we must decide on the relative scale of the abscissas γ and β since the shape of the pattern will depend upon this choice. You already know that γ/β is equal *d/b*. Let us say that in a particular case γ/β = *d/b* = 4. We must then plot the proposed curves for $\gamma = 4\beta$. In Fig. 6.4, the curves (a) and (b) are plotted to the same scale of θ. Fig. 6.4(a) depicts the curve for $\cos^2\gamma$ which given a set of equidistant maxima of equal intensity located at $\beta = 0$, $\pm\pi$, $\pm2\pi$, $\pm3\pi$, ...

In Fig. 6.4(b) we have plotted $(\sin^2\beta/\beta^2)$ which gives a maximum at $\beta = 0$ and minima at $\beta = \pm\pi$, $\pm2\pi$, ... In Fig. 6.4(c) we have plotted their product. What do you observe? The intensity of the fringes in the resultant pattern is not the same as it was in Fig.6.4(a). It is modulated (reduced) by the factor $(\sin^2\beta/\beta^2)$. This means that the central fringe or the zeroth fringe is the brightest, and the successive three fringes are of decreasing intensity until we reach the point $\beta = \pi$, where the intensity is zero. Thus the fourth fringe corresponding to $\cos^2\gamma = \pm4\pi$ falls at $\beta = +\pi$ or $-\pi$ and their product is zero. Therefore, the fourth fringe on either side of the central maxima has zero intensity and its location at the angle satisfies simultaneously

$$B = \pm\pi \text{ and } \gamma = \pm4\pi$$

or

$$b\sin\theta = \pm\lambda \text{ and } d\sin\theta = \pm4\pi$$

This fourth fringe will therefore be missing. We will observe the 5th, 6th and 7th fringes. We can argue in a similar manner that for 8th fringe

$$\beta = \pm2\pi \text{ and } \gamma = \pm8\pi$$

which will therefore have zero intensity and thus be missing. You may now like to answer the following SAQ.

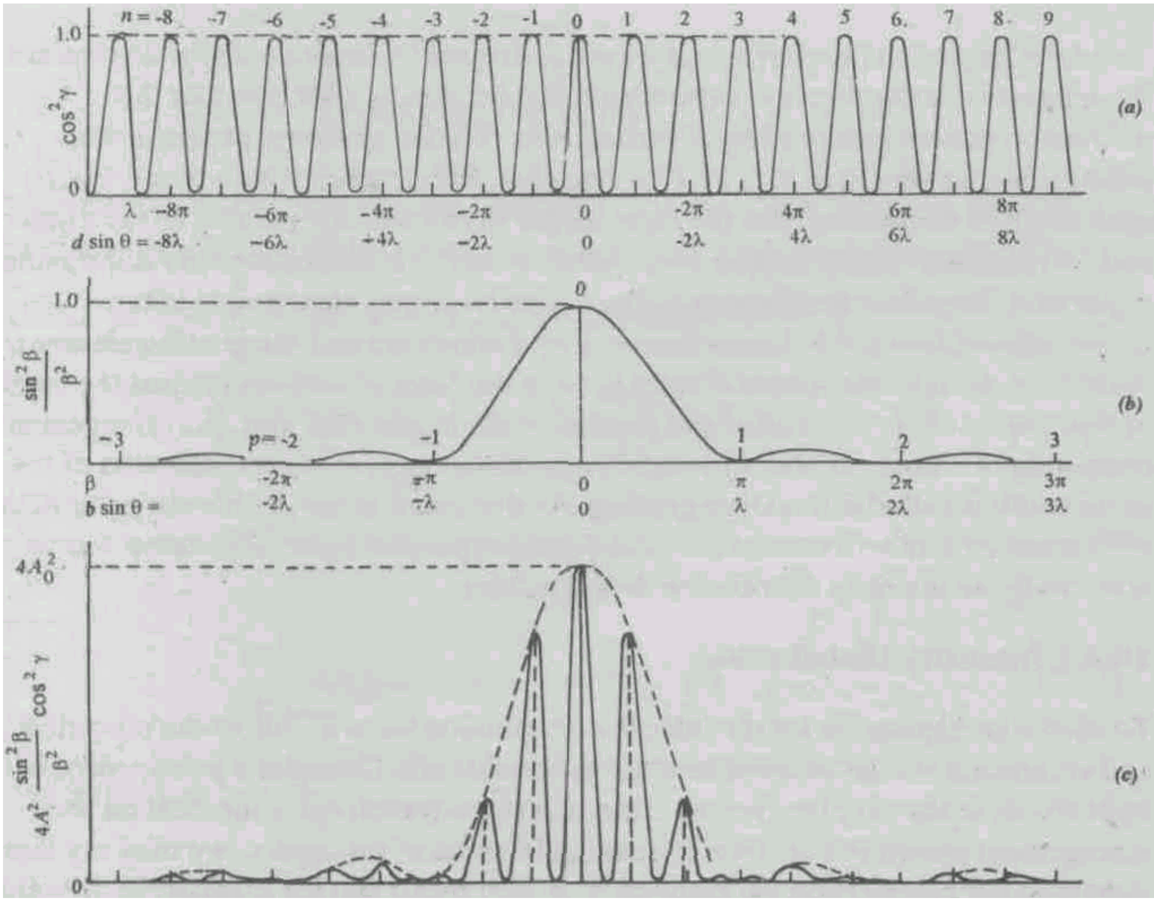


Fig. 6.4 Intensity curves for double slit. We have $\gamma = 4\beta$

SAQ2
Write down the general condition for missing orders in terms of the ratio *d/b*.

6.4 FRAUNHOFER PATTERN FROM *N* IDENTICAL SLITS

You now know that interference of waves diffracted by individual slits determines the intensity distribution in the double slit pattern. Let us now consider the diffraction pattern produced by *N* vertical slits. We use the same experimental arrangement as shown in Fig. 6.1 for two slits. For simplicity we assume that (i) each slit is of width *b* and has the same length (ii) all slits are parallel to each other and (iii) the intervening opaque space between any two successive slits is the same, equal to *a*. Therefore the distance between any two equivalent points in two consecutive slits is $a + b$. Let us denote it by *d* which we call the **grating element.** As before, we take the source of light to be in the form of a slit and adjust the length of this source slit to be vertical and parallel to the length of *N* slits. As arrangement consisting of a large number of parallel, equidistant narrow rectangular slits of the same width is called a **diffraction grating.** As discussed in the double slit pattern, the diffraction pattern will consist of vertical fringes parallel to the slit source. Let us now study the intensity distribution in this pattern.

6.4.1 Intensity Distribution

To derive an expression for the intensity distribution we will follow the procedure and arguments similar to those used for the double slit. Consider a point source of light which sends out plane waves. That is, a plane wavefront is incident on the arrangement shown in Fig. 6.5. (Speaking in terms of ray-optics, we may say that light rays fall normally on the grating). You may recall that the intensity distribution along any section perpendicular to the vertical fringes formed from a slit source will be the same as obtained from a point source. Physically, light emerging from *N* slits after diffraction at each slit results in *N* diffracted beams. Since these are coherent, interference takes place between them resulting in the formation of fringes. It is important to note that diffraction controls the intensity from each slit in a given direction.

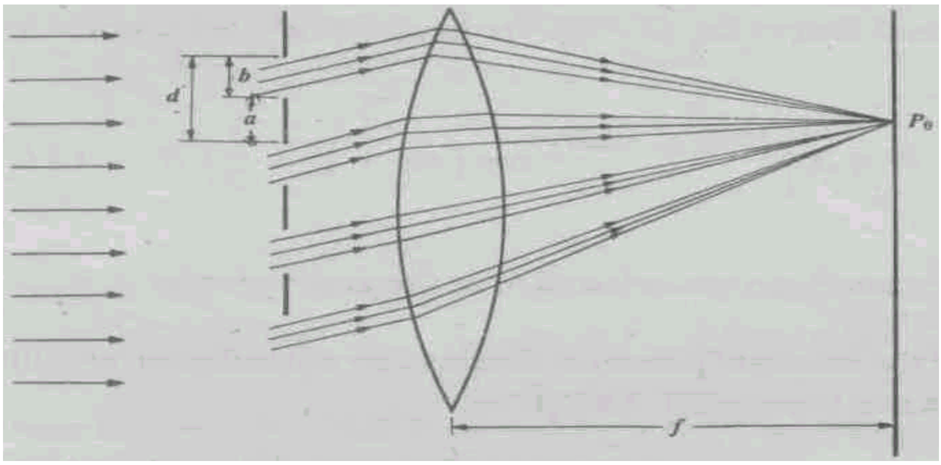


Fig. 6.5 Fraunhofer diffraction of a plane wave Incident normally on a multiple slit aperture

As before, we consider the diffracted rays proceeding towards P_θ, where θ is the angle between the diffracted rays and the normal to the grating. Let E_1, E_2, E_3, ..., E_n denote the fields produced by the first, the second, the third, ... and the *N*th slit at the point P_θ. Then we have

$$E_1 = A\frac{\sin\beta}{\beta}\cos(\omega t - \beta)$$

$$E_2 = A\frac{\sin\beta}{\beta}\cos(\omega t - \beta - \delta)$$

$$E_3 = A\frac{\sin\beta}{\beta}\cos(\omega t - \beta - 2\delta)$$

$$\vdots$$

$$E_N = A\frac{\sin\beta}{\beta}\cos[\omega t - \beta - (N-1)\delta]$$

Where various symbols have the same meaning as in Sec. 6.3. Also, we have assumed that the phase changes by equal amount 6 from one slit to the next.

The field E at P_θ is obtained by summing these *N* terms:

You have learnt to sum the series given here [Unit 2 of the course Oscillations and Waves Eq. (2.38)]:

In complex notation,

$$\exp(i\theta) = \cos\theta + i\sin\theta \qquad \text{(i)}$$

so that

$$\mathrm{Re}[\exp(i\theta)] = \cos\theta \qquad \text{(ii)}$$

It means that

$$\cos(\omega t - \beta) = \mathrm{Re}[e^{j(\omega t-\beta)}]$$

$$\cos(\omega t - \beta - \delta) = \mathrm{Re}[e^{j(\omega t-\beta-\delta)}]$$

.

.

.

$$\cos(\omega t - \beta - (N-1)\delta) = \mathrm{Re}[e^{j(\omega t-\beta-(N-1)\delta)}]$$

$$\therefore \quad \cos(\omega t - \beta) + \cos(\omega t - \beta - \delta) + \ldots$$

$$= \mathrm{Re}[e^{i(\omega t-\beta)} + e^{i(\omega t-\beta-\delta)} + \ldots + e^{i(\omega t-\beta-(N-1)\delta)}] \qquad \text{(iii)}$$

The RHS can be rewritten as

$$\text{RHS} = e^{i(\omega t-\beta)}[1 + e^{-i\delta} + e^{-2i\delta} + \ldots + e^{-i(N-1)\delta}] \qquad \text{(iv)}$$

This is a geometric series with common factor $e^{-i\delta}$ and can be summed easily using

$$\left(S = \frac{1-r^n}{1-r}\right):$$

$$\text{RHS} = e^{i(\omega t-\beta)}\,\frac{1-e^{-iN\delta}}{1-e^{-i\delta}}$$

$$= e^{i(\omega t-\beta)} \times \frac{e^{-iN\delta/2}}{e^{-i\delta/2}}\,\frac{e^{iN\delta/2} - e^{-iN\delta/2}}{e^{i\delta/2} - e^{-i\delta/2}}$$

$$= e^{i(\omega t-\beta-(N-1)\delta/2}\,\frac{\sin N\delta/2}{\sin\delta/2}$$

Hence, LHS of (iii) is recovered by the Real part, which is Eq. (6.12)

$$E = A\frac{\sin\beta}{\beta}\cos(\omega t - \beta) + A\frac{\sin\beta}{\beta}\cos(\omega t - \beta - \delta)$$

$$+ A\frac{\sin\beta}{\beta}\cos(\omega t - \beta - 2\delta) + \ldots$$

$$+ A\frac{\sin\beta}{\beta}[\omega t - \beta - (N-1)\delta]$$

You can write it as

$$E = A\frac{\sin\beta}{\beta}\{\cos(\omega t - \beta) + \cos(\omega t - \beta - \delta) + \ldots + \cos[\omega t - \beta - (N-1)\delta]\} \quad (6.11)$$

where $\gamma = \dfrac{\delta}{2} = \dfrac{\pi}{\lambda} d \sin\theta$. $\sin\gamma$ is referred to as the grating term.

The intensity of the resultant pattern is obtained by squaring the amplitude of the resultant field in this expression. Therefore,

$$I_\theta = A^2 \frac{\sin^2\beta \sin^2 N\gamma}{\sin^2\gamma} \qquad (6.13)$$

Let us pause for a while and ask: What have we achieved so far? We have obtained an expression for the resultant intensity of diffraction pattern from *N*-slits. We expect it to be true for any number of slits.

For a single slit, Eq. (6.13) reduces to

$$I_\theta = A^2 \frac{\sin^2\beta}{\beta^2}$$

which is the same as Eq. (5.7).

SAQ 3
Show that for $N = 2$, Eq. (6.13) reduces to Eq. (6.6) for the double slit.

6.4.2 Positions of Principal Maxima

For obtaining the positions of maxima (as well as minima), let us re-examine Eq. (6.13). We note that the intensity distribution is a product of two terms; the first term $(\sin^2\beta/\beta^2)$ represents the diffraction pattern produced by a single slit, whereas the second the term $(\sin^2 N\gamma/\sin^2\gamma)$ represents the interference pattern of N slits. The interference term controls the width of interference fringes, while the diffraction term governs their intensities.

As in case of the double slit, we cannot locate the exact positions of maxima; their approximate positions can however be obtained by neglecting the variation of $(\sin^2\beta/\beta^2)$. This is quite justified for very narrow slits. Therefore, for obtaining the positions of maxima we consider only the interference term.

The maximum value of $\dfrac{\sin^2 N\gamma}{\sin^2\gamma}(= N^2)$ occurs for $\gamma = 0, \pi, 2\pi, \ldots, n\pi$. At first glance, you will note that the quotient becomes indeterminate at these values. In such a situation, we compute the first derivative of the numerator as well as the denominator separately before inserting the value of argument. Following this procedure you will readily obtain

$$\lim_{\gamma \to n\pi} \frac{\sin N\gamma}{\sin\gamma} = \lim_{\gamma \to n\pi} \frac{N\cos N\gamma}{\cos\gamma} = \pm N$$

So that

$$\left(\frac{\sin N\gamma}{\sin\gamma}\right)^2 = N^2$$

The expression for intensity now takes the form

$$I_\theta = A^2 \frac{\sin^2\beta}{\beta^2} N^2 = N^2 A^2 \frac{\sin^2\beta}{\beta^2} \qquad (6.14)$$

where $\beta = \frac{\pi b \sin\theta}{\lambda}$

We therefore conclude that the positions of maxima are obtained when

$$\gamma = 0,\ \pi,\ 2\pi,\ \ldots,\ n\pi \ \text{ or } \ N\gamma = 0,\ N\pi,\ 2N\pi,\ \ldots,\ nN\pi \qquad (6.15)$$

Physically, at these maxima the fields produced by each of the slits are in phase and the resultant field is *N* times the field due to each of the slits.

When *N* is large, the intensity, being proportional to N^2, is very large and we will obtain intense maxima, if only $\sin^2\beta/\beta^2$ is not too small. Such maxima are known as **principal maxima.**

We can rewrite the condition of principal maxima as

$$d\sin\theta_{\max} = n\lambda \qquad (6.16)$$

which is identical to Eq. (6.9). It implies that

1. The principal maxima in *N*-slit pattern correspond in position to those of the double slit.
2. The relative intensities of different orders are modulated by the single slit diffraction envelope.
3. *n* cannot be greater than d/λ since $|\sin\theta| \le 1$. Can you imagine the implications of this condition? If you ponder for a while, you will realise that this condition suggests existence of only a finite number of principal maxima, which are designated as the first, second, third, . . . **order** of diffraction. Moreover, there will be as many first order principal maxima as the number of wavelengths in the incident wave.

4. The relation between β and γ obtained for double slit in terms of slit width and slit separation does not change. That is, Eq. (6.10) hold for *N*-slits as well.

6.4.3 Minima and Secondary Maxima

To be able to find the minima in the diffraction pattern, we locate the minima of the interference term. We note that the numerator in $\sin^2 N\gamma/\sin^2\gamma$ will become zero more often than the denominator. The numerator becomes zero for $N\gamma = 0,\ \pi,\ 2\pi,\ \ldots,\ p\pi$, or $\gamma = \frac{p\pi}{N}$.

Therefore, $\sin\gamma\left(= \sin\frac{p\pi}{N}\right)$ will not become zero for all integral values of p. It will become zero only for **special** cases when $p = 0$, N, $2N$, . . . and γ assumes values which are integral multiple of π. But you will recall that for these special values of γ, both $\sin N\gamma$ and $\sin\gamma$ vanish and the interference term defines the positions of principal maxima already discussed. However, for all other values of *p*, the numerator vanishes but not the denominator. That is, intensity vanishes when *p*, though an integer, is not an integral multiple of *N*. Hence, the condition for minimum is $\gamma = p\pi / N$ except when $p = nN$; *n* being the order. These values correspond to

$$\gamma = \left[\frac{\pi}{N}, \frac{2\pi}{N}, \ldots, \frac{(n-1)\pi}{n}, \frac{(N+1)\pi}{N}\right], \left[\frac{(N+2)\pi}{N}, \ldots, \frac{(2N-1)\pi}{N}\right],$$
$$\left[\frac{(2N+1)\pi}{N}, \ldots\right]$$

These values of γ correspond to path difference

$$d\sin\theta_{\min} = \left[\frac{\lambda}{N}, \frac{2\lambda}{N}, \frac{3\lambda}{N}, \ldots, \frac{(N-1)\lambda}{N}\right] \ldots \left[\frac{(N+1)\lambda}{N}, \ldots\right]$$

You should note that the values 0, $\frac{N\lambda}{N}$, $\frac{2N\lambda}{N}$, …, which correspond to $d\sin\theta_{\max} = n\lambda$ and represent principal maxima, are omitted.

Let us now summarise what you have learnt in this unit so far.

The condition for principal maxima:

$$\gamma = 0, \pi, 2\pi, \ldots, n\pi$$

and therefore

$$N\gamma = 0, N\pi, 2N\pi, \ldots, nN\pi$$

We may write

$$\gamma = \frac{\pi d}{\lambda}\sin\theta_{\max} = n\pi \quad \text{where } n = 0, 1, 2, \ldots$$

In terms of path difference,

$$d\sin\theta_{\max} = n\lambda$$

The conditions for minima:

$$N\gamma = nN\pi \pm \pi, \; nN\pi \pm 2\pi, \ldots, nN\pi \pm q\pi$$

where q is not an integral multiple of N. We can rewrite it as

$$N\gamma = n\pi \pm \frac{\pi}{N}, \; n\pi \pm \frac{2\pi}{N}, \ldots$$

In terms of path difference

$$d \sin\theta_{\min} = n\lambda \pm \frac{\lambda}{N},\ n\lambda \pm \frac{2\lambda}{N},\ \ldots,\ n\lambda \pm \frac{q\lambda}{N}$$

where $q \neq 0,\ N,\ 2N, \ldots$

If you write all possible values of $N\gamma$, you will find that we have (N - 1) positions of minima between any two successive principal maxima. Further, we know that between any two consecutive minima, there has to be a maxima. Such maxima are said to be **secondary** maxima. There will be (N - 2) positions of secondary maxima between two consecutive principal maxima. The secondary maxima are not symmetrical, as in the two-slit pattern. Moreover, the intensity of secondary maxima is very small and are therefore of little practical importance. Typical diffraction patterns and the corresponding intensity distributions predicted by Eq, (6.13) for N = 4 are shown in Fig. 6.6.

You may now like to answer the following SAQ.

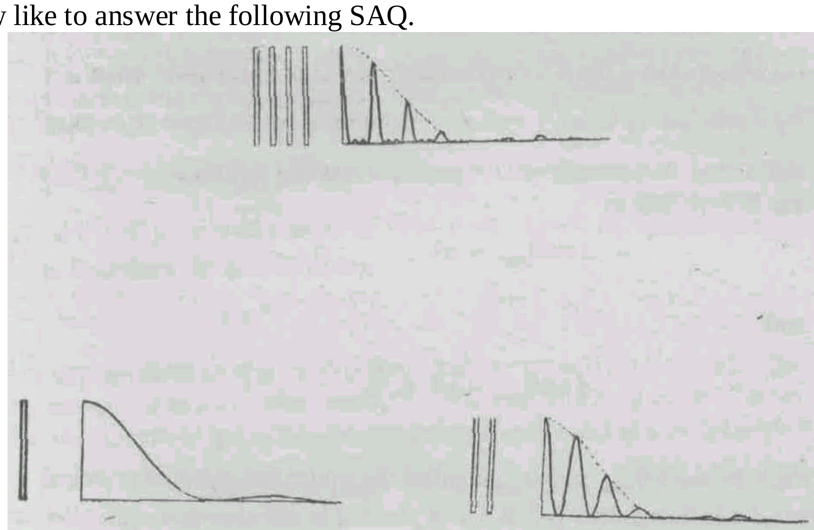


Fig. 6.6: Fraunhofer diffraction pattern for four slits. For comparison, patterns for one and double slits are also shown. The intensity distribution predicted by Eq. (6.13) is also shown.

SAQ 5

Show schematically the positions of principal maxima, secondary maxima and secondary minima for a diffraction grating with 6 slits.

Hint: We expect 5 minima between two consecutive principal maxima. Also we have 4 secondary maxima between the two principal maxima.

Example 2

Calculate the maximum number of principal maxima that can be formed with a grating 5000 lines per cm for light of wavelength 5000 Å.

$$\text{Grating element } d = \frac{1}{5000 \times 10^{-8}} = 2 \times 10^{-4} \text{ cm}$$

The condition for the formation of principal maxima is $d \sin\theta_{\max} = n\lambda$ since $|\sin\theta| \le 1$ we cannot have *n* greater than $\frac{d}{\lambda}$. In this specific case

$$n = \frac{2\times 10^{-4}\,cm}{5000\times 10^{-8}\,cm} = 4$$

Therefore, it will be able to show 1st, 2nd, 3rd and 4th orders of principal maxima.

If, on the other hand, we have a grating with 15000 lines cm^{-1}

$$n = \frac{(1/15000)}{5\times 10^{-5}} = \frac{6.6\times 10^{-5}}{5\times 10^{-5}}$$

which is less than 2. Such a grating will show only 1st order of spectrum with λ = 5000Å. You can verify this result while observing grating spectrum in your second level physics laboratory course.

6.4.4 Angular Half- Width of Principal Maxima

You now know that for *N* slits

1. The principal maxima occur when $\gamma = n\pi$ and therefore $N\gamma = Nn\pi$.

2. On either side of the principal maxima, we have a minimum when

   $N\gamma = nN\pi \pm \pi$ or when $\gamma = n\pi \pm \frac{\pi}{N}$. In terms of path difference and angle of diffraction, these conditions for principal maxima and the adjacent minimum can be rewritten as

$$d \sin\theta_{\max} = n\lambda \tag{6.16}$$

and

$$d \sin\theta_{\min} = n\lambda \pm \frac{\lambda}{N} \tag{6.17}$$

The angle between $\theta_{\max}$ and $Q_{\min}$ is called the annular half width of principal maxima, let us denote it by $\delta\theta$. We now proceed to calculate this angle. We can calculate $\delta\theta (=|\theta_{\max} - \theta_{\min}|)$ by computing $\theta_{\max}$ and $\theta_{\min}$ from Eqs. (6.16) and (6.17). Alternatively, by choosing $\theta_{\min} > \theta_{\max}$, we substitute $\theta_{\min} = \theta_{\max} + \delta\theta$ in Eq. (6.17) to obtain

You may now question as to why is $\delta\theta$ called **angular half width**. It is quite simple. You know that the principal maximum extends from minimum on one side to minimum on the other side and $\delta\theta$ is half of it. While solving SAQ 4 you have seen that for 6 slits the principal maximum extends from

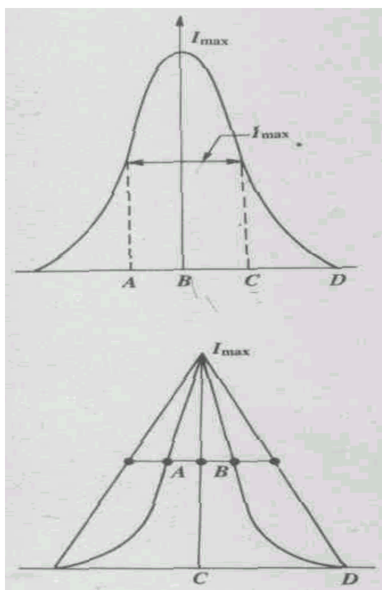
$N\gamma = 5\pi$ to $N\gamma = 7\pi$

or

$$d \sin\theta_{\max} = \frac{5\lambda}{6} \text{ to } \frac{7\lambda}{6}$$

You must note that the term **half width** of a spectrum line (or a diffraction curve) has a slightly different meaning. The diagram below represents the intensity vs θ curve. The half width gives

the width of the curve at $\frac{I_{\max}}{2}$. It is equal to *AB* in the diagram. The angular half width, on the other hand, is equal to *CD*. Obviously you can convince yourself that *AB* is not equal to *CD*. Only in the extreme case when the curve is a triangle, *AB* = *CD*.



$$d \sin(\theta_{\max} + \delta\theta) = n\lambda + \frac{\lambda}{N}$$

or

$$d \sin\theta_{\max} \cos\delta\theta + d \cos\theta_{\max} \sin\delta\theta = n\lambda + \frac{\lambda}{N}$$

For $\delta\theta \to 0$, $\cos\delta\theta \to 1$ and $\sin\delta\theta \to \delta\theta$. Hence,

$$d \sin\theta_{\max} + d \cos\theta_{\max}\delta\theta = n\lambda + \frac{\lambda}{N}$$

Using Eq. (6.16), we find that it takes a compact form:

$$d\cos\theta_{\max}\,\delta\theta = \frac{\lambda}{N}$$

so that

$$\delta\theta = \frac{\lambda}{Nd\cos\theta_{\max}} \tag{6.18}$$

which shows that the principal maximum becomes sharper as *N* increases. It is for this reason that grating spectrum is so sharp. You will now learn about it in detail.

6.5 DIFFRACTION GRATING

You have learnt about the diffraction pattern produced by a system of parallel equidistant slits. An arrangement of a large number of equidistant narrow vertical slits is known as diffraction grating. The first gratings were made by Fraunhofer. He stretched fine silver wire on a frame. His grating had nearly 200 wires to a centimetre. Afterwards, gratings were made by ruling fine lines with a diamond pen on a glass plate. The transparent part between the lines acted as a slit while the ruling itself acted effectively as the opaque part. Rowland was among the first to rule gratings on a metallic surface. He produced plane as well as concave gratings with nearly 5000 lines per centimetre. These gratings are difficult to make and are expensive but celluloid replicas can be made fairly cheaply and are commonly used in the physics laboratory for spectral analysis. You can make a simple coarse grating for demonstration purposes on a plate by drawing equidistant and parallel scratches on the photographic emulsion. Nowadays, it is possible to produce gratings holographically. Holographic gratings have greater rulings per cm and are definitely better than ruled gratings. You will get an opportunity to learn holographic details in Optics III.

6.5.1 Formation of Spectra

We have seen that for a monochromatic light of wavelength λ_1, the principal maxima are given by the grating equation

$$d\sin\theta_1 = n\lambda_1,\ n = 0, 1, 2, 3, \ldots$$

With the experimental arrangement described above we will get these principal maxima as one line in each order. Now if another source of light emits a longer wavelength λ_2, we will get a corresponding line in each order at a larger angle θ_2:

$$d\sin\theta_2 = n\lambda_2,\ n = 0, 1, 2, 3, \ldots$$

However if the same source of light emits both the colours corresponding to wavelengths λ_1 and λ_2, we will get two lines simultaneously in each order. These two lines will be seen as two spectrum lines separated from each other. This is because except the central maximum (zeroth order), the angles of diffraction for λ_1 and λ_2 are different in various other orders. In the central maxima θ = 0 for all wavelengths and this is why different colours are not separated from each other. What do you expect to observe when we have a white light source? The central image will be white while all other orders will show colours.

We note that in the grating equation, if we know *d*, θ and *n*, we can calculate the wavelength of light. Since the grating element *(d)* is known for a grating and θ can be measured, this arrangement provides a simple and accurate method of measuring λ. This is discussed in the following section.

6.5.2 Observing Grating Spectra

In your second level physics laboratory course, you must have observed grating spectra using a simple spectrometer. This arrangement is depicted in Fig. 6.7. The light from the given source is focussed (with the help of a lens) on the slit of the collimator, which sends out a parallel beam of light.

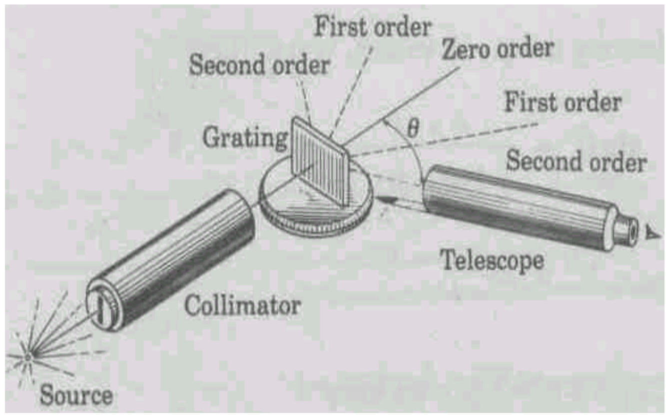


Fig. 6.7 A schematic diagram of experimental arrangement for observing grating spectra

> Light from a molecule gives a band like appearance and is often called band spectrum, while an incandescent lamp of similar sources will give a continuous spectrum, where various colours merge into one another.

The telescope arm is rotated and brought in line with the collimator. This ensures that the parallel beam of light falling on the objective of telescope is focussed at the crosswires, which is in the focal plane of the eyepiece. The position of the source of light should be adjusted to get the brightest image. We mount the diffraction grating on the turntable and adjust it so that the light is incident normally on the grating. Now we rotate the telescope arm to the left or right to get the first order spectrum in the field of view. If the source of light is a discharge tube containing sodium, mercury or argon the spectrum will consist of a series of spectrum lines. Each spectrum line is a diffracted image of the slit, formed by different wavelengths present in the source. To get sharp line images, we adjust the grating so that the diffracting slits are parallel to the collimator slit. This can be done by rotating the grating in its own plane.

To measure the wavelength of each line, we set the vertical crosswires at the centre of each spectrum line and note the position of the telescope in each case. The difference between the position of the telescope and the direct position gives the angle of diffraction for each of the lines. To reduce error, the position of the telescope is noted on both sides of the direct position and half of this angle gives the angle of diffraction.
You must have observed that

1. The spectrum exists on both sides of the direct beam.

2. Apart from the first order, the second or even third order spectrum (depending upon the grating element) are also present.

3. Different spectrum lines are not equally bright or sharp. This depends upon the energy levels and the transitions of the atom giving the spectrum. These concepts are further illustrated in the following example.

Example 3

Rowland ruled 14438 lines per inch in his grating, (i) Calculate the angles of diffraction for violet (λ = 4000 Å) and red (λ = 8000 Å) colours in the first order of spectrum. What is the largest wavelength which can be seen with this grating in the third order?

Solution

(i) The grating element $d = \dfrac{2.54}{14438}$ = 0.0001759 cm

$= 1.759 \times 10^{-4}$ cm

Suppose that the violet colour (λ = 4000 Å) is diffracted through angle θ_v. Recall the condition for maximum:

$$d \sin\theta_v = n\lambda$$

For first order on substituting the given values, you will get

$$\sin\theta_v = \frac{4\times10^{-5}}{1.759\times10^{-4}} = 0.2274$$

Therefore $\theta_v = 13^0$

Similarly, for red colour (λ = 8000 Å), we have

$$\sin\theta_r = \frac{8\times10^{-5}}{1.759\times10^{-4}} = 0.4548$$

so that

$$\theta_r = 27^0$$

This means that the entire visible spectrum in the first order extends from nearly θ = 13° to θ = 27°, i.e. covers an angle of about 14°.

(ii) $d \sin\theta = 3\lambda_{\max}$

According to the given condition, θ = 90° so that $\sin\theta = 1$ and $d = 3\lambda_{\max}$

or

$$\lambda_{\max} = \frac{d}{3} = \frac{1.759 \times 10^{-4}}{3} \text{ cm} = 5860 \text{ Å}$$

This calculation suggests that in the third order spectrum, the sodium doublet consisting of 5890 Å and 5896 Å will not be visible. Do you recall this from your observations on spectral analysis using a diffraction grating? If you have so far not opted for the second level physics, it will be worthwhile to verify this result.

If you calculate $\sin\theta_v$ and $\sin\theta_r$, for 1st, 2nd and 3rd orders, you will find that for

$$\text{1st order} \quad \left.\begin{array}{l} \sin\theta_v = 0.2274 \Rightarrow \theta_v \sim 13^0 \\ \sin\theta_r = 0.4548 \Rightarrow \theta_r \sim 27^0 \end{array}\right] \Rightarrow 14^0 \text{ spread}$$

$$\text{2nd order} \quad \left.\begin{array}{l} \sin\theta_v = 0.4548 \Rightarrow \theta_v \sim 27^0 \\ \sin\theta_r = 0.9096 \Rightarrow \theta_r \sim 65^0 \end{array}\right] \Rightarrow 38^0 \text{ spread}$$

$$\text{3rd order} \quad \left.\begin{array}{l} \sin\theta_v = 0.6822 \Rightarrow \theta_v \sim 43^0 \\ \sin\theta_r = 1 \quad \text{for } \lambda_{\max} \\ \qquad = 5860 \times 10^{-10} \text{ and} \theta_{\max} = 90^0 \end{array}\right] \Rightarrow 47^0 \text{ for 4000 Å - 6000 Å}$$

$$\sin\theta_r > 1 \text{ and cannot be observed} \qquad \Rightarrow \qquad \text{entire visible spectrum is not available}$$

Schematically it is shown below:

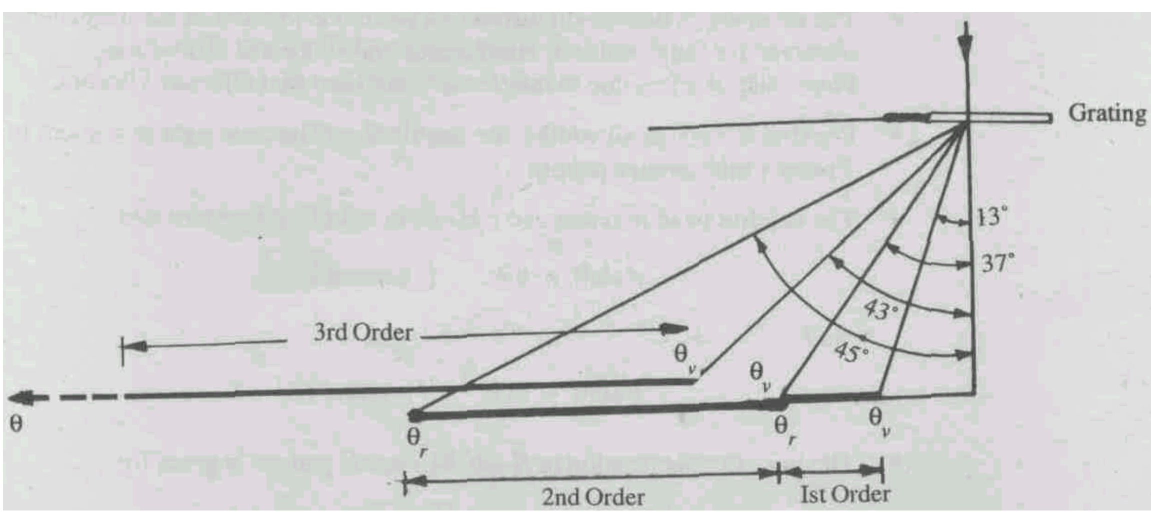


Fig. 6.8 Schematics of angles for overall spread of various orders of spectrum

Thus we find that in 1st order, red just touches second order violet. (This is because we have selected λ = 4000 Å and λ = 8000 Å). It means that there is essentially no overlapping of first and second order spectra. The third order λ_v, begins at $\theta \approx 43^0$. If you calculate wavelength λ_x of 2nd order present there you will find that

$$d \sin 43^0 = 3\lambda_v = 2\lambda_v \Rightarrow \lambda_x = \frac{3 \times 4000 \times 10^{-10}}{2} = 6000 \text{ Å}$$

Therefore λ = 6000 Å of the 2nd order occurs at the same place as λ = 4000 Å of third order. Therefore, from 6000 Å to 8000 Å will have overlapping colours. This difficulty is usually avoided by using suitable colour filters.

We now summarise what you have learnt in this unit.

6.6 SUMMARY

- The double slit diffraction pattern consists of a number of equally spaced fringes similar to what is observed in interference experiments. These fringes are the brightest in the central part of the pattern.

- In double slit pattern fringes reappear three or four times before they become too faint to observe.

- The central maximum in double slit pattern is four times brighter than that in single slit pattern.

- The intensity of double slit diffraction pattern at an angle 6 is given by

$$I_\theta = 4I_0 \frac{\sin^2 \beta}{\beta^2} \cos^2 \gamma$$

- Here, $I_0 = A^2$, $\beta = \dfrac{\pi b \sin\theta}{\lambda}$ and $\gamma = \dfrac{\pi}{\lambda} d \sin\theta$, where b is slit width and d is the distance between two similar points in these apertures. It is equal to $a + b$, where a is the width of the intervening opaque space between two slits.

- The intensity of double slit diffraction pattern is product of the irradiances observed for the double slit interference and single slit diffraction. Physically, it arises due to interference between two diffracted beams.

- For slits of very small widths, the double slit diffraction pattern reduces to Young's interference pattern.

- The conditions of maxima and minima in double slit pattern are:

$$d \sin\theta = n\lambda \quad \text{(maxima)}$$

and

$$b \sin\theta = m\lambda \quad \text{(minima)}$$

- The intensity distribution in *N*-slit diffraction pattern is given by

$$I_\theta = A^2 \frac{\sin^2\beta}{\beta^2}\frac{\sin^2 N\gamma}{\sin^2\gamma}$$

- The term $\sin\gamma$ is referred to as the grating term.
- As the number of slits increases, the maxima get narrower and for sufficiently large values of *N*, they become sharp lines. The angular half width of principal maximum $\delta\theta$ is given by

$$\delta\theta = \frac{\lambda}{Nd\cos\theta_{\max}}$$

- The principal maximum is sharp for large values of *N*.

6.7 TERMINAL QUESTIONS

1. If we use a white light source in the arrangement shown in Fig. 6.6, how will if affect the fringes?

2. Can there be principal maxima of zero intensity because of diffraction at each slit? If yes, discuss.

6.8 SOLUTIONS AND ANSWERS

SAQs

1. λ_1 will give its diffraction pattern within which we will get the interference fringes. The pattern for λ_2 will be smaller if $\lambda_2 < \lambda_1$. They will both be superimposed on one another, coinciding at θ = 0.

2. The general conditions for missing orders in terms of γ and β are $\gamma = \pm m\pi$ or $d\sin\theta = \pm m\lambda$ and $\beta = \pm p\pi$ or $b\sin\theta = p\pi$. Therefore

$$\frac{d}{b} = \frac{m}{p}$$

both *m* and *p* are integers, the missing orders occur when *d/b* is a ratio of two integers. When *d/b* = 1, i.e. the two slits exactly join, all the interference orders are missing. Physically it means that we have a single slit of double width and consequently no interference.

For $\frac{d}{b}$ = 2, second, fourth, sixth,... orders will be missing. What do you say about $\frac{d}{b}$ = 3?

3. For *N* = 2, Eq. (6.13) takes the form

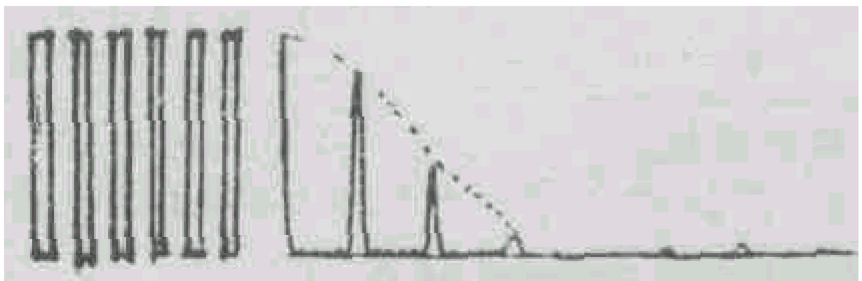
$$I_\theta = A^2 \frac{\sin^2\beta}{\beta^2}\frac{\sin^2 2\gamma}{\sin^2\gamma}$$

$$= A^2 \frac{\sin^2\beta}{\beta^2}\frac{(2\sin\gamma\cos\gamma)^2}{\sin^2\gamma}$$

$$= 4A^2 \frac{\sin^2\beta}{\beta^2}\cos^2\gamma$$

which is the required result for the double slit.

4. See figure given below:



TQs

1. As before, each wavelength will give its interference fringes. The central fringe for all wavelengths will coincide and hence the central fringe will be white. Fringes of order n = 1, 2, 3, ... located on either side of the central fringe, at different θ values given by $d\sin\theta = n\lambda$ for different wavelengths will be coloured.

2. There can be a principal maximum whose intensity is zero because of the diffraction at each slit. There are called **missing orders** or **absent spectra.** We know that the relationship between β and γ in terms of slit width and slit separation for N slits is the same as for the double slit. Therefore, the conditions for missing orders remain unaltered. And a particular maximum will be absent if it is formed at the same angle as the minimum of single slit diffraction pattern. This occurs at an angle which satisfies Eqs. (6.16) and (6.17).

UNIT 7 DIFFRACTION AND RESOLUTION

Structure

7.1 INTRODUCTION

In units 5 and 6 of this course you have learnt that due to diffraction, the image of an object is fringed even if an aberration-free converging lens is used. That is, the image of a point object is spread over a small area on the observation screen. Does this mean that no optical device can form a perfect image? The answer to this question is: The image of a point source is not a geometrical point. And diffraction does place a limit on the ability of optical devices to transmit perfect information (quality image) about any object. Such optical systems are said to be diffraction limited.

Broadly speaking, diffraction limited systems can be classified into two categories: (i) human eye, microscope and telescope which enable us to see two objects (near or distant) distinct and (ii) Grating and prism which form a spectrum and enable us to see two distinct wavelengths (colours). In principle, in both types of instruments two close fringed (diffraction) images are formed on the screen. The question that should logically come to your mind is: How to characterise the ability of an optical instrument to distinguish two close but distinct diffraction images of two objects or two wavelengths? This ability is measured in terms of resolving power. You may now like to know: What criterion enables us to compute resolving power? The most widely used criterion is due to Rayleigh. According to this, two diffraction images are said to be just resolved when the first minimum of diffraction pattern of one object falls at the same position where the central maximum of the diffraction pattern of the other lies. When the patterns come closer than this, the objects are not resolvable. When the patterns overlap less than this, the images are distinct and hence objects are resolvable. It is also important to know whether the same criterion applies to both types of optical devices? How can we improve resolution and see deeper in space even during the day? We have addressed all these aspects in this unit.

Objectives

After studying this unit, you should be able to:

- explain how diffraction limits image forming ability of optical devices
- use Rayleigh criterion to compute expressions for resolving power of a telescope, a microscope and a diffraction grating
- solve numerical problems based on resolution, and
- describe how Michelson stellar interferometer helps in improving resolution.

7.2 DIFFRACTION AND IMAGE FORMATION

You may recall from Unit 5 that when the size of pupil is greater than 2.4 mm, the human eye does not form a perfect point image (due to aberrations). However, for pupil sizes smaller than 2.4 mm, the human eye appears to be a diffraction-limited system. To gain some quantitative measure of visible acuity, let us estimate the size of image formed on our retina. If we approximate the pupil in human eye by a circular aperture, we have to consider how it influences the image formed by eye-lens on the retina (Fig. 7.1). From Unit 5 you may recall that the diffraction image of a point source due to a circular aperture is a bright central disc surrounded by a series of alternate dark and bright rings of decreasing intensity.

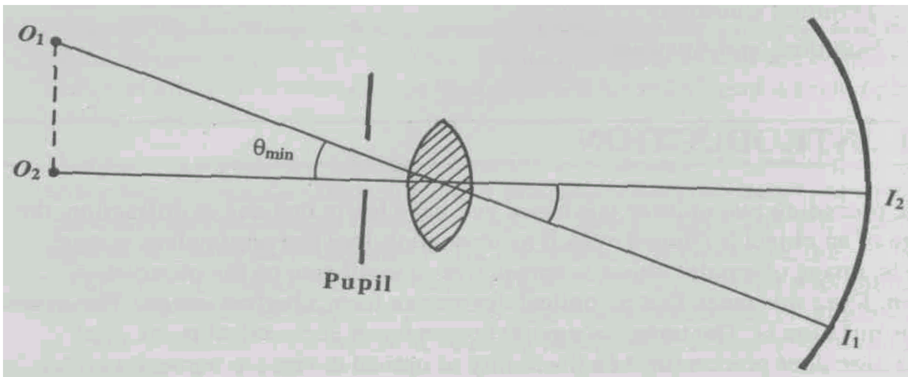


Fig. 11.1 Visible acuity and image formation on retina

The angular half-width of the central disc is given by $\theta = 1.22\lambda / D$, where D is the diameter of the aperture. And the lateral width of this image will be $f\theta$, where f is the focal length of eye-lens. This means that the size of an image formed on the retina depends on the wavelength of light and diameter of the aperture. If we take the pupil diameter to be 2 mm, then for middle of visible spectrum (λ = 5500 Å)

$$\theta = \frac{1.22}{\lambda} = \frac{1.22 \times (5.5 \times 10^{-5}\, cm)}{2 \times 10^{-1}\, cm} = 3.35 \times 10^{-4} \text{ rad} \approx 1 \text{ minute of arc}$$

Thus if the object is at a distance of 2 m, the size of image formed in a normal unaided human eye should be $(2 \times 3.35 \times 10^{-4}\ \text{rad}) \times 2$ m = 1.34×10^{-3} m.

Now refer to Fig. 7.2. It shows the image of a point source, luminous star say, formed by an astronomical telescope whose objective acts as a circular aperture and produces Airy pattern. The image essentially is a bright circular disc of angular diameter $2\theta \left(= \frac{2.44\lambda}{D} \right)$, which depends on λ and *D*. The larger the aperture, the truer the image, i.e., the smaller the Airy disc. On the other hand, if the aperture size is small, the size of the Airy disc increases. That is, no matter how free from aberrations an astronomical telescope objective is, what is observed at best is not a point image of a star. For similar reasons we find that the image of a point object formed by a microscope is of finite size. We may therefore conclude that **diffraction constrains an optical device in the formation of a sharp point-like image of a point source due to the finite sizes of its components.**

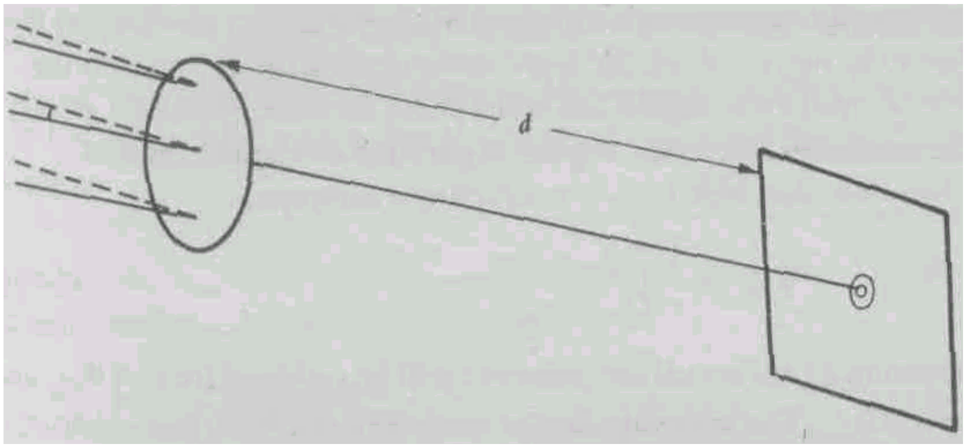


Fig. 7.2: Image of a luminous star formed by an astronomical telescope

An actual manifestation of this restriction arises in imaging when we observe two point sources or two spectrum lines. Since the objective of every optical instrument acts as a circular aperture and the point sources are mutually incoherent, the image consists of two independent Airy patterns. When the Airy discs are small and distinct, the two sources are said to be well resolved. The question now is how close can we bring these two discs so that they are just resolved. You will learn the answer to this question now.

7.3 RESOLVING POWER OF OPTICAL INSTRUMENTS

There are several criteria for the resolution limit. But we will confine ourselves to the conventional specification, the Rayleigh criterion, which however arbitrary, has the virtue of being particularly simple. According to this, the two patterns are resolved when the first minimum of the diffraction pattern of one coincides with the central maximum of the diffraction pattern of the other. In Rayleigh's own words:

> This rule is convenient on account of its simplicity and it is sufficiently accurate in view of the necessary uncertainty as to what exactly is meant by resolution.

We will now consider the specific cases of an astronomical telescope, a microscope and a diffraction grating.

7.3.1. Astronomical Telescope

Imagine that a telescope points towards two close luminous stars, which subtend an angle α on the objective. The plane waves from these stars reach the objective and give rise to Airy diffraction patterns (Fig.7.3). Since the stars are effectively at an infinite distance from us, the diffraction patterns (images) are formed in the back focal plane of the telescope objective, where it is examined with the aid of the eyepiece. The angle between mid points of central discs is equal to the angle subtended by the stars at the objective. For these stars to be just resolved, Rayleigh's criterion demands that maximum (centre) of the Airy disc due to one star should fall on the minimum (periphery) of the disc due to the other star, as shown in Fig. 7.4.

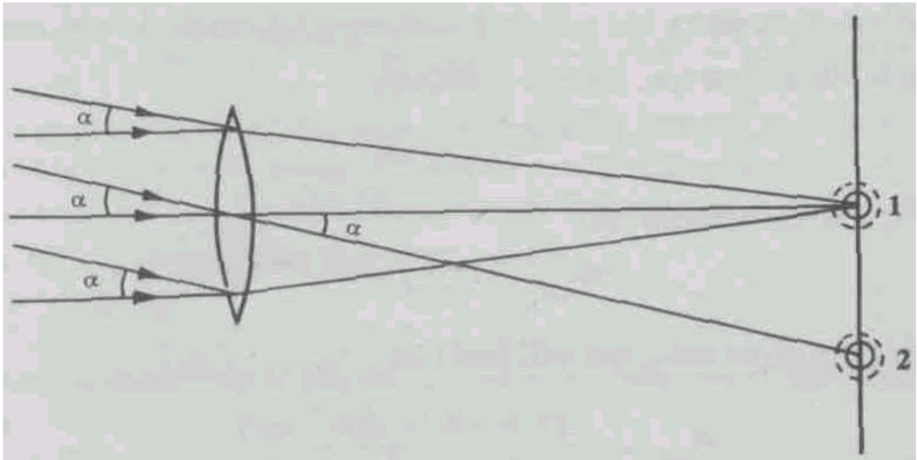


Fig. 7.3 Formation of Airy patterns In imaging of two stars by a telescope

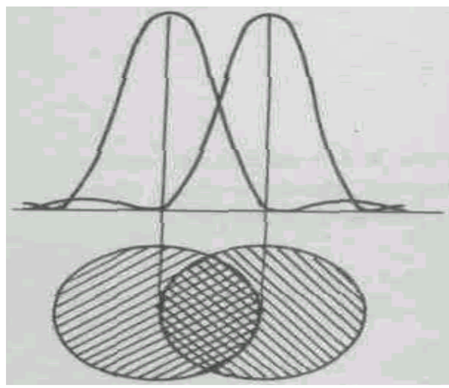


Fig. 7.4 Rayleigh criterion for Imaging of two stars of small angular separation

(The corresponding intensity curves are also shown.) Mathematically, we demand that for the two stars to be just resolved, the angle subtended by the two stars at the objective should be equal to the angular half width of the Airy disc. Recall Eq. (5.13). It suggests that the **minimum resolvable angular separation or angular limit of resolution** for two close stars which can be resolved by a telescope is

$$\theta_{\min} = \frac{1.22\lambda}{D} \qquad (7.1)$$

Two stars subtending an angle α at the objective will be resolved for $\alpha > \theta_{\min}$ and unresolved for $\alpha < \theta_{\min}$. The intensity plot for more than resolved, just resolved (Rayleigh limit), and unresolved stars are shown in Fig. 7.5.

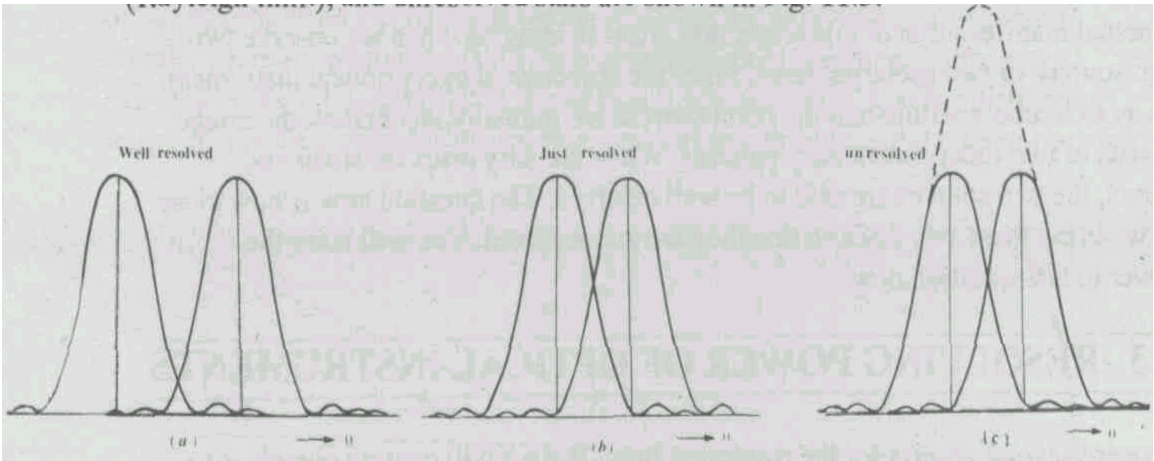


Fig. 7.5: Plot of Intensities of two resolved, just resolved and unresolved stars

The centre-to-centre linear separation of two just resolved stars, **the limit of resolution,** is given by

$$s = f\theta_{\min} = \frac{1.22\lambda f}{D} \tag{7.2}$$

where f is the focal length.

Before **the S.I. system** of units was adopted, the objective **sizes** were expressed in inches.

The **resolving power** of an optical device is generally defined as the reciprocal of the resolving limit, i.e., as $Q_{\min}^{-1}$ or s^{-1}. This means that resolution ability of diffraction-limited systems depends on the size of the aperture and the wavelength. For a given wavelength, the resolving power of a telescope can be increased by using objectives of larger diameter. To give you some appreciation of numerical values, we now give a solved example. You should go through it carefully.

Example 1
An astronomical observatory has a 40 inch telescope. Calculate the minimum angle of resolution for this telescope. Take λ = 6000 Å.

Solution
From Eq. (7.1) we recall that

$$\theta_{\min} = 1.22\lambda / D$$

On substituting the given data, you will find that

$$\theta_{\min} = \frac{1.22 \times (6 \times 10^{-5}\ cm)}{40 \times 2.54\,cm}$$

$= 7.2 \times 10^{-7}$ rad

= 0.15 seconds of arc

The diameter of the largest telescope is about 80 inch (~2m) and the corresponding angular separation of the objects it can resolve is 0.07 seconds of arc. This very low limit is not achieved in ground-based telescopes due to turbulence in the lower atmosphere.

For the human eye, $\theta_{\min} = 3.35 \times 10^{-4}$ rad. Therefore, the actual lateral width of the image of a distant point formed on your retina is

$$s = f\theta_{\min}$$

If we take f = 3 cm, we find that

$$s = (3 \text{ cm}) \times 3.35 \times 10^{-4}$$

$$= 10.05 \times 10^{-4} \text{ cm}$$

$$= 10 \text{ micron}$$

This is roughly three times the mean spacing between photoreceptors (cones) at the centre of the retina. Therefore, for a normal unaided human eye, the linear separation between two point objects at a distance of 3m subtending this angle will be equal to $(3.35 \times 10^{-4} \times 3m = 1 \times 10^{-3}$ m) = 1 mm. This means that the unaided eye will resolve two point objects 1mm apart at a distance of about 3 m.

You can easily verify this result at least qualitatively. You should just draw two lines one millimetre apart and view these from a distance. (Alternatively, you can see marks on a millimetre scale or some newsprint). Move forward or backward till these become blurred and just merge into one another. Experience tells us that 1 mm is barely resolved at 2 m. The difference is due to optical defects in the eye or the structure of retina.

You may now like to answer an SAQ.

SAQ 1
An astronaut orbiting at an height of 400 km claims that he could see the individual houses of his city as they passed beneath him. Do you believe him? If not, why?

You now know that a 40 inch telescope has a minimum angle of resolution equal to 7.2×10^{-7} rad. The minimum angle of resolution of the eye is about 3.35×10^{-4} rad. An important question that should come to our mind is: What should be the magnifying power of the telescope to take full advantage of the large diameter of the objective? The telescope must magnify about $\dfrac{3.35 \times 10^{-4}}{7.2 \times 10^{-7}}$ = 465 times. Note that any further magnification will only make the image bigger but it would not be accompanied by increase in details which are not available in the primary image. (The resolution is determined by diffraction at the objective, i.e., the magnitude of $Q_{\min}$) To get some idea about these details, you should carefully go through the following example.

Example 2
Compare the performances of two telescopes with objectives of apertures 100 cm and 200 cm. Take their focal lengths to be equal.

Solution
We know that for a telescope, the minimum angle of resolution

$$\theta_{\min} = \frac{1.22\lambda}{D}$$

For the first telescope $\theta_{\min} = \frac{1.22\lambda}{100\,cm}$, where λ is in cm. Therefore, the radius of the central diffraction disc r = $f\theta_{\min} = f\,\frac{1.22\lambda}{100\,cm}$ the area of Airy disc

$$A_1 = \pi r^2 = \pi\left(f\,\frac{1.22\lambda}{100\,cm}\right)^2$$

The area of the telescope objective which collects light is $\pi\left(\frac{100\,cm}{2}\right)^2$. This light is largely concentrated in the central maximum and gradually decreases as $\left(\frac{\sin N\gamma}{\sin\gamma}\right)^2$. If we assume that light is uniformly distributed over the disc, its brightness, i.e., light per unit area

$$I_1 \propto \pi\left(\frac{100\,cm}{2}\right)^2 \div \pi\left(f\,\frac{1.22\lambda}{100\,cm}\right)^2$$

$$= (50)^2 \times \frac{(100)^2}{f^2(1.22)^2\lambda^2}\ \text{cm}^4$$

$$= \frac{100^2 \times 100^2}{4f^2(1.22)^2\lambda^2}\ \text{cm}^4$$

For the second telescope $\theta_{\min} = \frac{1.22\lambda}{200\,cm}$. That is, the minimum angle of resolution for the second telescope is half of that for the first telescope. In other words, the R.P of 200 cm telescope is twice as large. To compare their relative performances, let us compare the brightness. As before, the area of central diffraction disc

$$A_2 = \pi\left(f\,\frac{1.22\lambda}{200\,cm}\right)^2$$

and brightness

$$I_2 \propto \frac{(200)^2(200)^2}{4f^2(1.22)^2\lambda^2}\text{ cm}^4$$
$$= 16I_1$$

In words, the area of the central diffraction disc of second telescope is four times more. And the of the image of the star will be proportional to fourth power of its area.

So we may conclude that

(i) The ability of a telescope to resolve two close stars depends on the diameter of its objective.

(ii) The intensity of the image is sixteen times since the objective collects four times more light and concentrates it over an area which is only one fourth. This means that a distant star, which is too faint to be observed by a smaller objective (of the first telescope), becomes visible by a larger telescope. That is, a bigger telescope can see farther in the sky. Therefore, **the deeper we want to penetrate the space, the greater should be the aperture of the objective of the telescope.**

You may now like to pause and ponder for a while. Then you should answer SAQ 2.

SAQ 2

We can see the stars at night but as the sun rises they gradually fade away and are not visible during the day. What measure would you suggest to enable researchers to make astronomical observations in the daytime itself?

Example 3

Calculate the dip in the resultant intensity of two $\left(\frac{\sin\beta}{\beta}\right)^2$ curves according to Rayleigh's criterion, i.e., when the maximum of one curve falls on the minimum of the other curve.

Solution

We assume that the two curves have equal intensity. These curves are symmetrical and will cross at $\beta = \pi/2$, as shown in Fig. 7.6.

At the point of intersection, both curves have equal intensity:

$$I = \left(\frac{\sin\frac{\pi}{2}}{\pi/2}\right)^2 = \frac{4}{\pi^2} = 0.4053$$

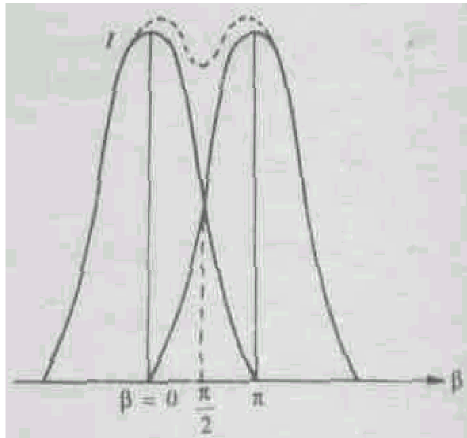


Fig. 7.6: Resolution of two single slit patterns: Rayleigh's criterion

At this point the resultant intensity will be equal to the sum of the two intensities and therefore equal to 0.8106. This means that according to Rayleigh's criterion, the resultant intensity will show a dip of about 20%. And this dip is easily visible to even unaided human eye. If these two curves are brought closer, the dip will gradually decrease and it becomes difficult to resolve the images. Moreover, if these intensities were unequal, the dip will not be 20%.

In the above example we have taken the intensity of both the curves to be equal. This essentially means that in Rayleigh criterion we take both the stars to be equally luminous. Another important point to note is that the curves are of finite angular (or lateral) width. In the case of grating (or prism), two spectrum lines, though assumed to be of equal intensity, are very sharp. Now the question arises: Can we use the same criterion even for a grating? From your second level physics laboratory you may recall the answer to this question: we do use the same criterion. Is the dip 20% or so even in this case? To discover, answer to this question, you should answer the following SAQ.

SAQ 3

What is the dip in the resultant intensity of two $\left(\dfrac{\sin N\gamma}{\sin\gamma}\right)^2$ curves according to the Rayleigh criterion?

A more realistic criterion for resolving power has been proposed by Sparrow. We know that at the Rayleigh limit there is a central dip or saddle point between adjacent peaks. As the distance between two point sources is less than the Rayleigh limit, the central dip will grow shallower and may ultimately disappear (Fig. 7.7) The angular separation corresponding to that configuration is said to be Sparrow's limit. Note that the resultant maximum has a broad flat top; there is no change in slope. However, we will not discuss it any further.

Another useful image forming device is the microscope. Let us now learn to calculate its resolving power.

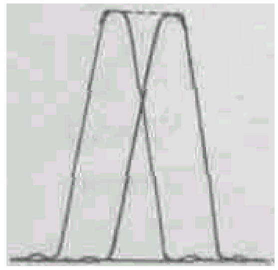


Fig. 7.7 Sparrow's resolution criterion

7.3.2 Microscope

We know that an astronomical telescope is used to view far off objects the exact distances of which are usually unknown. However, we were chiefly interested in their smallest permissible angular separation at the objective. In the case of an optical microscope, the objects being examined are very close to the objective and subtend *a* large angle. For this reason by resolving power of a microscope we mean the smallest distance, rather than the minimum angular separation, between two point objects (*O* and *O'*) when their fringed images (I and I') are just resolved.

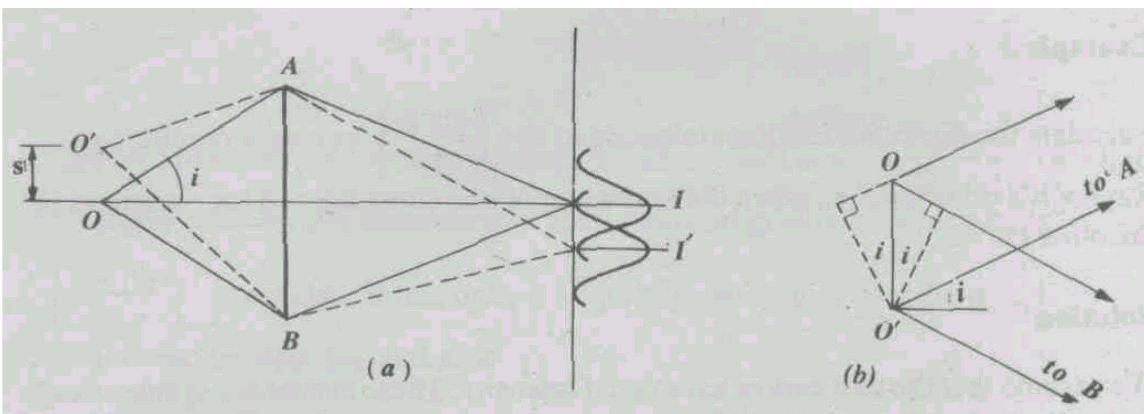


Fig. 7.8: The optical microscope, (a) Airy pattern images of two objects O and *O'* separated through a distance s (b) Ray diagram for computation of path difference *O'B* -*O'A*

Each image consists of a central Airy disc (surrounded by a system of rings which are very faint and not considered.) According to the Rayleigh criterion, the first maximum of I should be at the same position where the first minimum of I' lies. The angular separation between the two discs on the limit of resolution $\theta_{\min} = \dfrac{1.22\lambda}{D}$. When two images are just resolved, the wave from *O'* diffracted to I' has zero intensity (first dark ring) and the path difference *O'B* - *O'A* = 1.22 λ (Fig. 7.8 (a)). We show an enlarged part in Fig. 7.8 (b) from which we see that *O'B* is longer than *OB* by $s \sin i$, and *O'A* is shorter by the same amount. Here the point *O* subtends an angle 2 *i* at the objective of the microscope. Thus the path difference of the extreme rays from *O'* to the objective is $2s\sin i$. Upon equating this to 1.22 λ we find that the minimum separation between two points in an object that can be resolved by a microscope is given by

$$2s\sin i = 1.22$$

or

$$s = \frac{1.22\lambda}{2\sin i} = \frac{0.61\lambda}{\sin i} s =$$

In high power microscopes, the space between the object and objective is filled with oil of refractive index μ. For an oil-immersed objective, the above expression becomes

$$s = \frac{0.61\lambda}{\mu \sin i} \tag{7.3}$$

You may now like to answer an SAQ.

SAQ4
In the above discussion we assumed that the two point objects were self-luminous. Suppose two objects are illuminated by the same source. Will Eq. (7.3) still hold?

Abbe investigated this problem of image formation in detail and found that the resolving power depends on the mode of illumination of the object. In the above treatment both O and O' were treated as self-luminous objects and thus the light given out by these had no constant phase relationship. For all practical modes of illumination, the resolving power may be taken simply as

$$R.\ P. = \frac{0.61\lambda}{\mu \sin i}$$

The term $\mu \sin i$ is termed the **numerical aperture** (NA) of the microscope objective. The maximum value of i is 90°. This gives the microscopic limit on *R.P* approximately as $\frac{\lambda}{2\mu}$. This shows that smaller the *NA*, greater will be the *R.P*. In practice, good objectives have $N.\ A \approx 1$ so that the smallest distance that can be resolved by a microscope is of the order of the wavelength of light used. Obviously, with light of shorter wavelength, say ultraviolet rather than visible light, microscopy allows for perception of finer details. (We may have to take the photographs and then examine the images.)

In your school physics curriculum you have learnt that electrons exhibit diffraction effects. The de Broglie wavelength of an electron is given by

$$\lambda\ (\text{Å}) = \frac{12.3}{\sqrt{V}} \tag{7.4}$$

For electrons accelerated to 100 kV, the wavelength is

$$\lambda\ (\text{Å}) = \frac{12.3}{\sqrt{10^5}} = 0.039 \times 10^{-10}\ \text{m} \tag{11.5}$$

This wavelength is 10^5 times smaller than that for visible light. The resolving power of an **electron microscope** will therefore be very high. This makes it possible to examine objects that would otherwise be completely obscured by diffraction effects in the visible spectrum. In this connection we may mention tremendous utility of electron microscope in the study of minute objects like viruses, microbes and finer details of crystal structures. It is better than even ultraviolet microscope for high-resolution applications.

7.3.3 Diffraction Grating

You are familiar with a sodium lamp. It gives out two close spectral lines, the so-called D_1 and D_2 lines with wavelengths λ_1 = 5890 Å and λ_2 = 5896 Å. For such lines, the resultant peak may become somewhat ambiguous. The problem we now wish to consider is: What is the smallest difference $\Delta\lambda$, that a diffraction grating can resolve? The resolving power of a grating is defined as

$$R.P. = \frac{\lambda}{(\Delta\lambda)_{\min}} \tag{7.6}$$

where $(\Delta\lambda)_{\min}$ is the least resolvable wavelength difference or limit of resolution and λ is the mean wavelength. It is sometimes also called chromatic resolving power.

The de Broglie wavelength of an electron is given by

$$\lambda = \frac{h}{m_e v}$$

where h is Planck's constant, m_e is electronic mass and v is electron speed. When an electron beam is accelerated through a potential difference V, we can write

$$v = \sqrt{\frac{2v_e}{m_e}}$$

On combining these relations we find that

$$\lambda = \frac{h}{\sqrt{2m_e e}}\frac{1}{\sqrt{V}}$$

Substituting the values

$h = 6.6\times10^{-34}\ Js$, $m_e = 9.11\times10^{-31}$ kg and $e = 1.6\times10^{-19}\,C$, you will find that

$$\lambda\ (\text{Å}) = \frac{12.3}{\sqrt{V}}$$

where $(\Delta\lambda)_{\min}$ is the least resolvable wavelength difference or the limit of resolution and λ is the mean wavelength. It is sometimes also called the chromatic resolving power.

We know that the grating forms a principal maximum corresponding to wavelength λ at the diffraction angle θ. Similarly, the principal maxima at corresponding to $\lambda + \Delta\lambda$ will be at $\theta + \Delta\theta$. At first thought you may argue that the two colours will be separated and always appear to be resolved since the two angles are different. This could be so if the principal maxima, i.e. the spectrum lines in the experimental arrangement, were truly sharp like an ideal geometric line. But we know that the principal maximum has a finite angular width. Therefore, the question is: How close can these be brought so that they are seen distinct? Obviously, sharper the lines, the closer these can be brought and still be seen as two.

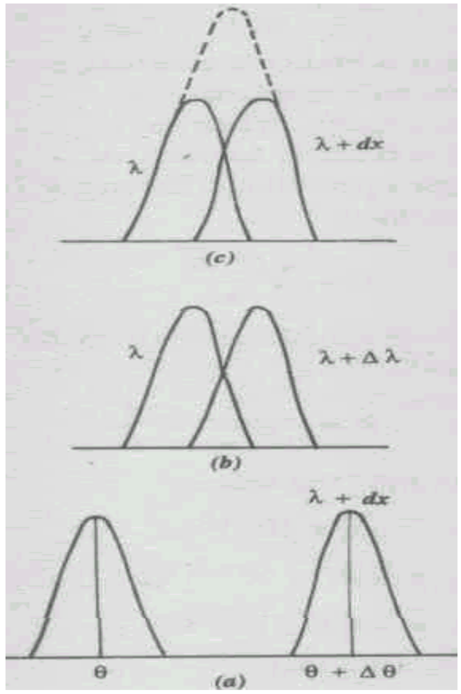


Fig. 7.9 Resolution of two spectral lines

This question was also carefully examined by Rayieigh. In Fig. 7.9 (a) we show plots of two widely separated principal maxima. In Fig. 7.9 (b) we have brought these closer so that the principal maximum of $\lambda + \Delta\lambda$, is situated at the position where the minimum of λ falls. The dotted line defines resultant intensity, which shows a dip. You will recall that according to the Rayleigh criterion, this is the closest that we can bring these curves and still regard them as separate. If we bring them still closer as in Fig. 7.9 (c), the resultant intensity (shown by the dotted line) signifies a single enhanced principal maxima.

According to the Rayleigh criterion, the condition for resolution of two spectral lines by a diffraction grating is obtained by noting that for the common diffraction angle θ, the following two equations should be satisfied simultaneously:

$$d \sin\theta = n(\lambda + \Delta\lambda)$$

for principal maxima of $\lambda + \Delta\lambda$ and

$$d \sin\theta = n\lambda + \frac{\lambda}{N}$$

for first minimum adjacent to the principal maximum for wavelength λ. On simplifying these we get

$$\frac{\lambda}{\Delta\lambda} = nN \tag{7.7}$$

We note that in a given order *n*, the *R.P* is proportional to the total number of slits. Does this mean that *R.P* increases indefinitely with *N*? It is not so. Think why? Does it have some connection with the width of the grating? You will also note that the resolving power is independent of grating constant. It means that the resolving powers of two gratings having equal number of lines but different grating constants will be equal.

To enable you to grasp these concepts and appreciate the numerical values, we now give some more solved examples.

Example 4

For D_1 and D_2 sodium lines, λ_{D_1} = 5890 Å and λ_{D_2} = 5896 Å. Calculate the minimum number of lines in a grating which will resolve the doublet in the first order.

Solution

Let us take the average wavelength as 5893 Å. From Eq. (7.6) we find that the resolving power is

$$\frac{\lambda}{\Delta\lambda} = \frac{5893\times10^{-8}}{6\times10^{-8}} = 982.2$$

Therefore, we must have a grating with more than 983 lines to resolve sodium doublet in first order. A grating of 1000 lines will serve the purpose.

Example 5

Suppose that to observe sodium doublet we use a grating having $d = 10^{-3}$ cm and a lens of focal length 2 m. Let us calculate the linear separation of the two lines in the 1st and 2nd order.

Solution

We know that

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

For the D_1 line

$$\sin\theta_1 = \frac{5890\times10^{-8}\,cm}{10^{-3}\,cm} = 5890\times10^{-5}$$

or

$$\theta_1 = 5890\times10^{-5}\ \text{rad}$$

Similarly for the D_2 line

$$\sin\theta_2 = \frac{5896\times10^{-8}\,cm}{10^{-3}\,cm} = 5896\times10^{-5}$$

or

$$\theta_2 = 5896 \times 10^{-5} \text{ rad}$$

With a lens of focal length 200 cm, we find that linear separation between D_1 and D_2 lines is

$$l = f\Delta\theta$$

$$= (200 \text{ cm}) \times (6 \times 10^{-5} \text{ rad})$$

$$= 12 \times 10^{-3} \text{ cm} = 0.12 \text{ mm}$$

This shows that 6 Å are separated by 0.12 mm in 1st order. Alternatively we may say that linear separation is nearly 50 Å per millimetre in the first order. You can readily check that in the second order this linear separation will be 25 Å per millimetre.

7.4 IMPROVING RESOLUTION

You now know that with the help of a telescope, we can view a faint star, resolve two close stars and measure the angle subtended by the double star at the objective of the telescope. However, it is worth noting that based on Fraunhofer diffraction image of a star, we cannot measure its angular diameter. To overcome this limitation, Fizeau suggested a slight modification in that we should use a two slit adjustable aperture (with provision for lateral adjustment), in front of the objective of the telescope. As a result, the plane wavefront falling on the double slit is diffracted and collected by the objective. The Fraunhofer diffraction pattern of the double slit is formed in the back focal plane of the objective. The measurements to determine angular diameter are made from the observations on these interference fringes.

Refer to Fig. 7.10. Two slit apertures S_1 and S_2 are at a distance d apart. The telescope is first pointed towards the double stars, which act as two point sources O and O'. The two point sources are separated by an angle θ in a direction at right angles to the lengths of the slits. Such objects emit white light and because of intensity considerations, the observations have to be made with white light fringes. It is therefore customary to assume an effective value of the wavelength emitted by the source. This depends upon the distribution of intensity of the light and the colour response of the eye. The interference patterns due to O and O' have the same fringe spacing since this spacing depends upon separation between slit apertures and the focal length of the objective. Moreover, these fringe patterns are shifted with respect to each other by an angle θ. Therefore, as shown in the figure the central maximum of the pattern due to O is at P and that due to O' is at P'. If O and O' are two incoherent sources, the combined pattern is formed by summing the intensities of these two patterns at each point. Assuming that both O and O' have equal brightness, we can plot two $\cos^2\gamma$ curves on the same scale and shift them suitably to obtain the resultant curve.

We can show graphically that if this shift is a small fraction of the angular separation θ, the resultant intensity distribution resembles a $\cos^2\gamma$ curve. However, the intensity does not fall to zero at the minimum. The net result is a fringe pattern (shown in Fig. 7.10(b)). By successive adjustments a stage can come when the maximum of one pattern, say due to O, coincides with the minimum of O'. Then we $\frac{\theta_1}{2} = \frac{\lambda}{2d}$. And the paths from the two sources differ by $\frac{\lambda}{2}$. We can show graphically that the resultant curve 2 shows a uniform intensity and the fringes have disappeared. If we displace the two curves further, the fringes reappear and become sharp when the fringes are displaced by a whole fringe width, i.e. $\theta = \theta_1$. They disappear again when $\theta = \frac{3\theta_1}{2}$ or $\frac{5\theta_1}{2}$. Therefore, with two point sources subtending an angle θ at the double slit, the condition for the disappearance of fringes is

$$\theta = \frac{\lambda}{2d}, \frac{3\lambda}{2d}, \frac{5\lambda}{2d}$$

The intensity of the double slit pattern is given by

$$I = 4R^2 \frac{\sin^2 \beta}{\beta^2} \cos^2 \gamma$$

where $\beta = \frac{\pi a \sin\theta}{\lambda}$ and $\gamma = \frac{\pi d \sin\theta}{\lambda}$

in which *a* is the slit width and *d* is the slit separation. The positions of the maxima are given by

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

where $n = 0, 1,2,3, ...$ When θ is small, the successive maxima occur at

$$\theta = 0, \frac{\lambda}{d}, \frac{2\lambda}{d}, \frac{3\lambda}{d}, \ldots$$

so that the angular separation between successive maxima is given by $\theta_1 = \frac{\lambda}{d}$.

Further, if *a* is
small, the interference pattern will be essentially a $\cos^2\gamma$ curve near the centre.

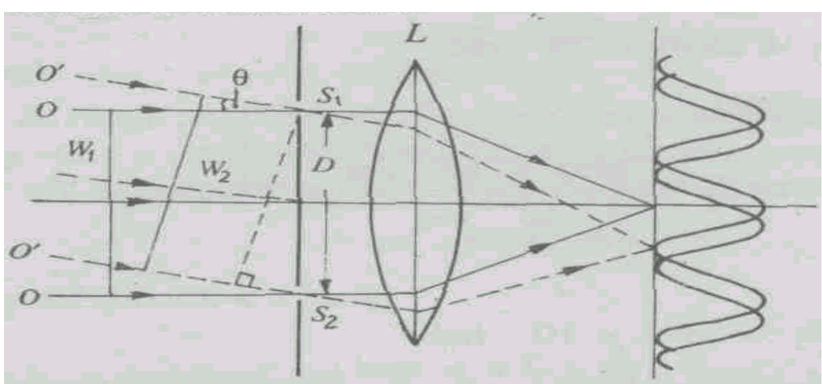


Fig. 7.10 Principle of measurement of angular diameter of stellar objects by interferometry

To measure the angular separation of a double star, the double slit is mounted in front of the objective of the telescope which points towards the double star. (We should remember that the line joining the stars should be perpendicular to the length of the slits.) We expect interference pattern due to the double slit. If on adjusting the separation between the slits, the interference fringes can be made to disappear, we can infer that the star is a double star. The first disappearance should take place when the angular separation is $\frac{\lambda}{2d}$. Let us compare this with the expression for the resolving power of a telescope $(\theta = 1.22\lambda / a)$, where *a* is the diameter of the objective). If the double slits are *a* apart and the first disappearance occurs for $d = a$, the angle θ between the double stars is $\theta = \frac{\lambda}{2d} = \frac{\lambda}{2a}$. This angle is effectively half of the *R.P* of the telescope. It explains the genesis of the statement: **The *R.P* of a telescope may be doubled by placing a double slit in front of it.** You must however note that with a double slit, we can only infer the presence of a double star (from the disappearance of the fringes); we neither get the images of the stars nor resolve them. Indeed, even before the disappearance of the fringes, a blurring of fringes starts. This angle is only a small fraction of θ_1. You may have realised that this method enables us to measure the **angular diameter** of the disc of the star and Michelson successfully used it in 1920.

Angular Diameter of a Star

For measuring the angular diameter of the disc of a star we should first know the condition for the disappearance of fringes for a double slit placed in front of a telescope. In contrast to two point sources, the disc of a star consists of a series of points extending from one end O_1 to another end O_2 In Fig. 7.10, we see that when O_l and the central point O satisfy the condition for disappearance of fringes, the point just next to O_l will have a similar point next to O and so on. Thus all the points between O_1 and O will have corresponding points lying between O and O_2 satisfying the condition for disappearance of fringes. Since the angle between O_l and O for the first disappearance of fringes is $\frac{\lambda}{2d}$, the angle between (O_1, and O_2 (which is for the total disc) equals $\frac{\lambda}{d}$. Thus the angular disc θ of the star, computed from the first disappearance of fringes, is given by $\theta = \frac{\lambda}{d}$. For successive disappearance θ is given by $\theta = \frac{2\lambda}{d}, \frac{3\lambda}{d}, \ldots$ If the source is a circular disc, the condition for the first disappearance is $\theta = 1.22\frac{\lambda}{d}$. This method was successfully used to measure angular diameters of planetary satellites. But attempts to apply it for single stars failed because of their small angular diameters. Even with the largest slit separation possible with the available telescopes, the fringes remained distinct; no disappearance was achieved. To overcome this difficulty, Michelson devised the stellar interferometer in 1890. We will discuss it now.

7.4.1 Michelson Stellar Interferometer

The principle of Michelson's Stellar Interferometer is illustrated in the Fig. 7.11. The slit apertures S_1 and S_2 in front of the telescope are fixed. Light reaches them after reflection from a symmetrical system of mirrors M_1, M_2, M_3 and M_4 mounted on a rigid girder in front of the telescope. The inner mirrors M_3 and M_4 are fixed but the outer mirrors M_1 and M_2 can be separated out symmetrically in a direction perpendicular to the lengths of the slit apertures. Therefore light from one edge of the star (shown as solid line) reaches the point *P* in the focal plane via the paths $OM_1\ M_3\ S_1\ P$ and $OM_2\ M_4\ S_2\ P$. This will form interference fringes with the angular separation equal to $\frac{\lambda}{d}$.

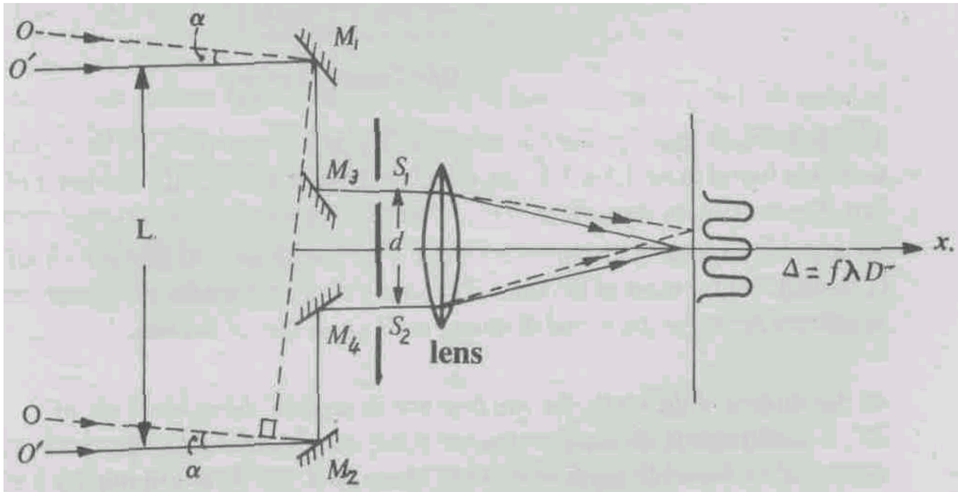


Fig. 7.11: Schematics of Michelson Stellar Interferometer

The other edge of the star sends light along the dotted lines and produces a similar system displaced slightly with the central fringe at *P'*. You now know that when two extreme fringe systems are displaced by a whole fringe width, the resultant intensity pattern will show uniform intensity and the fringes will disappear. The angular diameter of the star $\alpha = 1.22\frac{\lambda}{D}$, where *D* is the separation of outer mirrors M_1 and M_2. You can easily convince yourself by noting that the optical paths $M_1M_3S_1$ and $M_2M_4S_2$ have been maintained equal so that the optical path difference for light from the two edges of the star is the same at S_1 and S_2 as at M_1 and M_2. If the path difference at M_1 and M_2 is one whole wavelength, the path difference at S_1 and S_2 is also one whole λ and fringe shift is equal to one fringe width. This leads to disappearance of fringes. As shown in the diagram, the dotted lines inclined at an angle α will have a path difference of λ when $\alpha = \frac{\lambda}{D}$. In this arrangement the smallest angular diameter that can be measured is determined by the separation of the outer mirrors M_1 and M_2 rather than the diameter of the objective of the telescope. Therefore, the stellar interferometer magnifies the effective resolving power of the telescope in the ratio $\frac{D}{d}$. We may emphasize that for a circular star disc, the fringes will disappear when $\alpha = 1.22\frac{\lambda}{D}$. This implies that the outer mirrors have to be moved out somewhat.

The interferometer was mounted on the large reflecting telescope (diameter 100 inch) of the Mount Wilson observatory, which was used because of its mechanical strength. The first star whose diameter was measured by this method was Betelegeuse (α-Orion) whose fringes disappeared when the separation between M_1 and M_2 was equal to 121 inches. Assuming λ = 5700 Å, we find that

$$\alpha = \frac{1.22\lambda}{D} = \frac{1.22 \times 5700 \times 10^{-8}\, cm}{121 \times 2.54\, cm}$$

$$= 22.7 \times 10^{-8} \text{ rad}$$

$$= 0.047 \text{ seconds of arc}$$

The distance of Betelegeuse was measured by parallax method. Its linear diameter was then found to be 4.1×10^8 km, which is about 300 times the diameter of the sun. The maximum separation of the outer mirrors was 6.1 m, so that the smallest measurable angular diameter with λ = 5500 Å was about 0.02 seconds of arc. This is insufficient for most of the stars. The smallest star for which measurements were made was Arcturus. Its actual diameter is 27 times that of the sun.

At the surface of the earth, the sun disc has an angular diameter of about 32' ~ 0.018 rad. If we imagine the sun to be at a distance of the nearest star, its disc would subtend an angle only 0.007 seconds of arc. This will require a mirror separation of 20 m for disappearance of fringes. It is difficult to achieve this since we require a rigid mechanical connection between mirrors and the eyepiece.

Let us now summarise what you have learnt in this unit.

7.5 SUMMARY

- Diffraction constrains an optical device in the formation of a sharp point-like image of a point source.

- Rayleigh criterion for the resolution of two images demands that the first minimum of diffraction pattern of one object and the central maximum of the diffraction pattern of the other should fall at the same position.

- The minimum resolvable angular separation or angular limit of resolution of two close objects by a telescope is given by

  $$\theta_{\min} = \frac{1.22\lambda}{D}$$

  where λ is the wavelength and D is diameter of the objective of the telescope.

- The resolving power of a telescope is the inverse of the angular limit of resolution. The deeper we want to penetrate the space, the greater should be the aperture of the objective of telescope.

- The resolving power of a microscope is defined as the smallest **distance** between two point objects when their fringed images are just resolved:

  $$R.P = \frac{0.61\lambda}{\mu \sin i} = \frac{0.61\lambda}{N.A.}$$

where i is the angle of incidence, sin I is known as numerical aperture and is approximately equal to 1 for good objective.

- The resolving power of a diffraction grating is defined as

$$R.P = \frac{\lambda}{(\Delta\lambda)_{\min}} = nN$$

where $\Delta\lambda$ is the least resolvable wavelength difference, n is the order of spectrum and N is the total number of slits.

7.6 TERMINAL QUESTIONS

1. A diffraction limited laser beam (λ = 6300 Å) of diameter 5 mm is directed at the earth from a space laboratory orbiting at an altitude of 500 km. How large an area would the central beam illuminate?

2. The resolving power of a prism is given by

$$\frac{\lambda}{d\lambda} = t\frac{d\mu}{d\lambda}$$

where t is the length of the base of the prism, μ is the refractive index of the material of prism for wavelength λ . A prism is made of dense flint glass for which refractive indices for λ = 6560 Å and 4860 Å are 1.743 and 1.773 respectively. Calculate the length of the base of the prism.

7.7 SOLUTIONS AND ANSWERS

SAQs

1. The minimum angle of resolution of eye

$$\theta = \frac{1.22\lambda}{D} = \frac{1.22\times(5.5\times10^{-5}\,cm)}{0.2\,cm} = 3.36\times10^{-4}\,\text{rad}$$

The lateral width for resolution

$$l = r\theta = (4\times10^{5}\,m)\times(3.36\times10^{-6}\,\text{rad}) = 1.34\text{ m}$$

Since it is must less than the width of individual houses, it is not wise to believe the astronaut.

2. As we increase the aperture of the telescope, the light collected by it from a star gradually increases and gets concentrated in the image (the diffraction disc). Ultimately a stage will come when the image of the star becomes brighter than the background and is visible (This is because the intensity of the image of a star is proportional to the fourth power while the background sky light increases as the square of the area of the aperture.) This means that you can see stars during the day by using a telescope of sufficient aperture!

3. The maximum is at $Nn\pi$ and minimum at $(Nn + 1)\pi$. The two curves are symmetrical and if they are of equal intensity, they will cross at $N\gamma = Nn\pi + \frac{\pi}{2}$. Therefore, if you evaluate the function $\left(\frac{\sin N\gamma}{\sin \gamma}\right)^2$ at $N\gamma = Nn\pi$ and $N\gamma = Nn\pi + \frac{\pi}{2}$, i.e., $\gamma = n\pi$ and $\gamma = n\pi + \frac{\pi}{2N}$, you will find that

$$\left(\frac{\sin Nn\pi}{\sin n\pi}\right)^2 = N^2$$

and

$$\left[\frac{\sin\left(Nn\pi + \frac{\pi}{2}\right)}{\sin\left(n\pi + \frac{\pi}{2N}\right)}\right] = \frac{1}{\sin^2\left(\frac{\pi}{2N}\right)} = \frac{1}{\left(\frac{\pi}{2N}\right)^2}$$

$$= \frac{4N^2}{\pi^2}$$

Hence the required ratio is $\frac{4}{\pi^2} = 0.4053$

Therefore the resultant intensity will show a dip of about 20% as in the case of a telescope.

4. The waves given out by each self-luminous object bear no constant phase relationship so that the intensities can be added up. The objects viewed with microscopes are illuminated by the same source and there will be some phase relationship between the waves emanating from these. Strictly speaking the intensities will not be additive. But Abbe found that Eq. (7.3) gives the correct order for the limit of resolution.

TQs

1. We know that angular spread of light beam is given by

$$\theta = \frac{1.22\lambda}{D} = \frac{1.22 \times 6300 \times 10^{-8}\,cm}{0.5\,cm}$$

$$= 1.54 \times 10^{-4} \text{ rad}$$

Since the diameter of light patch

$$x = 2r\theta$$

the area of the earth illuminated by the beam focussed from the space laboratory at an altitude of 500 km is

$$A = \frac{\pi x^2}{4} = \pi r^2 \theta^2$$

$$= \frac{22}{7} \times (25 \times 10^{10} m^2) \times (1.54 \times 10^{-4})^2$$

$$= 10934 \text{ m}^2 = 0.01 \text{ km}^2$$

2. $d\mu = 1.773 - 1.743 = 0.03$

$$d\lambda = 6560 - 4860 = 1700 \text{ Å} = 1700 \times 10^{-8} \text{ cm}$$

Note that spectral spread is very wide whereas $d\lambda$ should be a small change. Assuming that μ changes linearly between these two colours, we have

$$\frac{d\mu}{d\lambda} = -\frac{0.03}{1700 \times 10^{-8} cm} = -\frac{3}{17} \times 10^4 \text{ cm}^{-1}$$

The negative sign signifies inverse value of relationship between λ and μ. The prism is made of dense flint glass and to just resolve D_1 and D_2 lines find that

$$R.P = \frac{5893}{6} = 982$$

so that

$$\frac{\lambda}{d\lambda} = 982 = 1765\, t$$

and

$$t = \frac{982}{1765\, cm^{-1}} = 0.556 \text{ cm} \approx 0.6 \text{ cm}$$