Unit V: Wave Optics

Electromagnetic nature of light

The nineteenth century also saw the development of electricity and magnetism. In 1820, Oersted (often written as Ørsted) discovered that electric currents create magnetic fields and in 1926 Ampere discovered the law relating the magnetic field to the current; this law (usually referred to as Ampere's law) was later written in the form of a vector equation by Maxwell. Then, around 1830, Faraday carried out experiments which showed that

a time varying magnetic field induces an electromotive force

which is now referred to as Faraday's law; this law was also later written in the form of a vector equation by Maxwell. Around 1860, Maxwell generalized Ampere's law by stating that not only electric currents create magnetic fields, but

a time varying electric field can also create a magnetic field

–like between the plates of a capacitor when it gets charged or discharged. In his book on electricity and magnetism (see Fig. 2.10), Maxwell summed up all the laws of electricity and magnetism in the form of equations -- which are now known as Maxwell's equations; these equations are based on experimental laws. Feynman writes (Ref. 2.7):

All of electromagnetism is contained in Maxwell's equations... Untold numbers of experiments have confirmed Maxwell's equations. If we take away the scaffolding he used to build it, we find that Maxwell's edifice stands on its own.

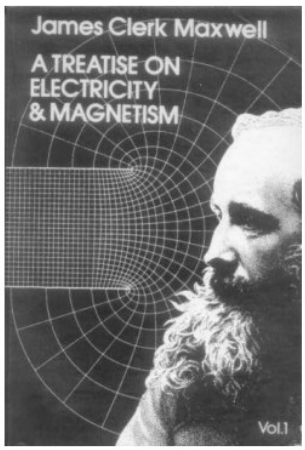

Fig. 2.10 Maxwell's book on Electricity & Magnetism in which he wrote the laws of electricity and magnetism and predicted the existence of electromagnetic waves.

In Chapter 23, we will discuss Maxwell's equations and will show that these equations have solutions

$$\left.\begin{aligned}\mathbf{E}(z,t) &= \hat{\mathbf{x}}\, E_0 \cos(kz-\omega t)\\ \mathbf{H}(z,t) &= \hat{\mathbf{y}}\, H_0 \cos(kz-\omega t)\end{aligned}\right\} \qquad (2.9)$$

where

$H_0 = \sqrt{\varepsilon_0/\mu_0}\, E_0$, and $\varepsilon_0\,(=8.8542\times10^{-12}\ \mathrm{C^2\ N^{-1}\ m^{-2}})$ and $\mu_0\,(=4\pi\times10^{-7}\ \mathrm{Ns^2/C^2})$ represent respectively the dielectric permittivity and the magnetic permeability of free space. Equation (2.9) describes propagating electromagnetic waves (see Fig. 2.11). Thus from the laws of electricity and magnetism, Maxwell predicted the existence of electromagnetic waves, and by substituting the above solutions in Maxwell's equations he showed that the velocity (in free space) would be given by

$$c = \frac{\omega}{k} = \frac{1}{\sqrt{\mu_0\,\varepsilon_0}} \approx 3\times10^8\ \mathrm{m/s} \qquad (2.10)$$

Thus Maxwell not only ***predicted*** the existence of electromagnetic waves, he also ***predicted*** that the speed of the electromagnetic waves in air should be about 3.107×10^8 m/s. He found that this value was very close to the measured value of the speed of light which according to the measurement of Fizeau in 1849 was 3.14858×10^8 m/s. The sole fact that the two values were very close to each other led Maxwell to propound (around 1865) his famous electromagnetic theory of light in his famous book [see Fig. 2.10]; according to Maxwell,

light waves are electromagnetic waves

This was one of the great unifications in physics. Max Planck had said

… (Maxwell's theory) *remains for all time one of the greatest triumphs of human intellectual endeavor*

Associated with a light wave are changing electric and magnetic fields; the changing magnetic field produces a time and space varying electric field and the changing electric field produces a time and space varying magnetic field, and this results in the propagation of the electromagnetic wave even in vacuum (see Fig. 2.11).

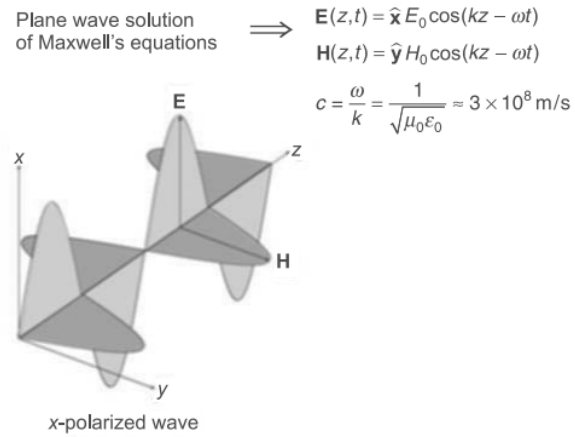

Fig. 2.11 An x-polarized electromagnetic wave propagating through free space.

In 1888, Heinrich Hertz carried out beautiful experiments which could produce and detect electromagnetic waves of frequencies smaller than those of light. These waves were produced by discharging electrically charged plates through a spark gap. The frequency of the emitted electromagnetic waves could be calculated by knowing the inductance and capacitance of the circuit. Hertz also produced standing electromagnetic waves by getting them reflected by a metal sheet. He could calculate the wavelength of the waves and knowing the frequency, he showed that the speed of the electromagnetic waves (in air) was the same as that of light: velocity of electromagnetic waves

$$= \lambda\nu \approx 3 \times 10^8 \text{ m/s} \qquad \textbf{(2.11)}$$

Using a collimated electromagnetic wave, and getting it reflected by a metal sheet he could demonstrate the laws of reflection (see Sec. 13.2). Hertz's experimental results provided dramatic confirmation of Maxwell's electromagnetic theory. In addition, there were so many other experimental results, which were quantitatively explained by using Maxwell's theory that towards the end of the nineteenth century, physicists thought that one had finally understood what light really was, i.e., light was an electromagnetic wave.

Definition and properties of wave front

A wavefront is the locus of the particles of the medium which at any instant are vibrating in the same phase

Or in simple language a wavefront is the locus of the points which are in the same phase; for example, if we drop a small stone in a calm pool of water, circular ripples spread out from the point of impact, each point on the circumference of the circle (whose center is at the point of impact) oscillates with the same amplitude and same phase and thus we have a circular wavefront. On the other hand, if we have a point source emanating waves in a uniform isotropic medium, the locus of points which have the same amplitude and are in the same phase are spheres. In this case, we have spherical wavefronts as shown in Fig. 12.1 . At large distances from the source, a small portion of the sphere can be considered as a plane and we have what is known as a **plane wave** [see Fig. 12.1].

Wavefront Types

The path followed by the particles emanating from a source determines the different types of wavefront. Wavefront are of three types.

1. Spherical Wavefront: When the point source is an isotropic medium, sending out waves in three dimensions, the wavefronts are spheres centred on the source, as shown in the figure. Such a wavefront is called a spherical wavefront.

Examples of Spherical Wavefronts

- Electromagnetic waves in a vacuum form a spherical wavefront.
- The concentric circles' formation when the stone is dropped in a water
- An army man patrolling the opposition on the radar through the camera positions the invader with the spherical wavefront symbol.

2. **Cylindrical Wavefront**: When the light source is linear, we obtain a cylindrical wavefront. In a cylindrical wavefront, all the points equidistant from the linear source lie on the surface of a cylinder, as shown in the figure.

The cylindrical wavefront appears like a cylinder. However, if we draw the wavefront from one plane, we obtain a concentric circle like a spherical wavefront.

Example of Cylindrical Wavefront

- When rays of light coming out of a lens fall on another lens, they converge at a given point. As they bend and converge at a point, it takes the form of a cylinder.

3. **Plane Wavefront**: The wavefront will appear as a plane when viewed from a considerable distance from a source of any kind. Such a wavefront is called a plane wavefront.

Moreover, the plane wavefront is obtained when the small part of the spherical or cylindrical wavefront originates from a distant source, like infinity.

Plane Wavefront Example

- Plane wavefronts are generated from a very distant source. A common example of the plane wavefront is the rays coming out of the sun.

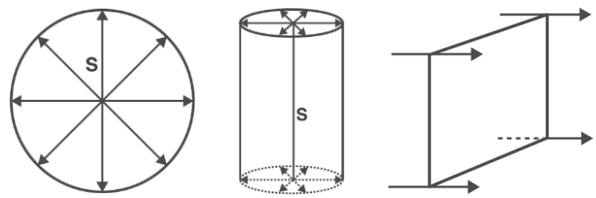

Fig. 12.1: Different types of wavefront.

Properties of Wavefront

- Wave front is defined as locus of all points having same phase at a given instant of time.
- The shape of wavefront depends on the shape of the source of disturbance.
- A wavefront is always normal to the light rays.
- A wavefront does not propagate in the backward direction.

Huygens Principle

According to Huygens' principle

Each point of a wavefront is a source of secondary disturbance and the wavelets emanating from these points spread out in all directions with the speed of the wave. The envelope of these wavelets gives the shape of the new wavefront.

In Fig. 12.2, S_1S_2 represents the shape of the wavefront (emanating from the point O) at a particular time which we denote as $t = 0$. The medium is assumed to be homogeneous and isotropic, i.e., the medium is characterized by the same property at all points and the speed of propagation of the wave is the same in all directions. Let us suppose we want to determine the

shape of the wavefront after a time interval of Δt. Then, with each point on the wavefront as center, we draw spheres of radius $v\,\Delta t$, where v is the speed of the wave in that medium. If we draw a common tangent to all these spheres, then we obtain the envelope which is again a sphere centered at O. Thus, the shape of the wavefront at a later time Δt is the sphere $S_1'S_2'$.

There is, however, one drawback with the above model, because we also obtain a backwave which is not present in practice. This backwave is shown as $S_1''S_2''$ in Fig. 12.2. In Huygens' theory, the presence of the backwave is avoided by assuming that the amplitude of the secondary wavelets is not uniform in all directions; it is maximum in the forward direction and zero in the backward direction*. The absence of the backwave is really justified through the more rigorous wave theory.

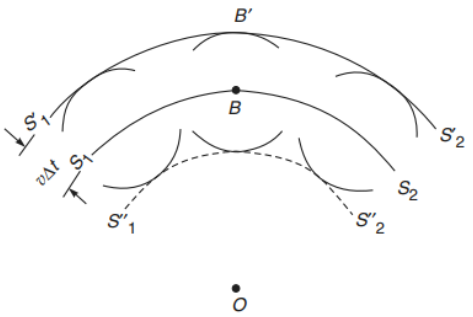

Fig. 12.2 Huygens' construction for the determination of the shape of the wavefront, given the shape of the wavefront at an earlier time. S_1S_2 is a spherical wavefront centered at O at a time, say $t = 0$. $S_1'S_2'$ corresponds to the state of the wavefront at a time Δt, which is again spherical and centered at O. The dashed curve represents the backwave.

Temporal and Spatial Coherence

Temporal Coherence

In earlier chapters on interference, we had assumed that the displacement associated with a wave remained sinusoidal for all values of time. Thus, the displacement (which we denote by E) was assumed to be given by

$$E = A\cos(kx - \omega t + \phi)$$

The above equation predicts that at any value of x, the displacement is sinusoidal for $-\infty < t < \infty$. For example, at $x = 0$ we have [see Fig. 17.1(a)].

$$E = A\cos(\omega t - \phi),\ -\infty < t < \infty \qquad \textbf{(17.1)}$$

Obviously this corresponds to an idealised situation because the radiation from an ordinary light source consists of finite size wavetrains, a typical variation of which is shown in Fig. 17.1(b). Since we will be considering only light waves, the quantity E represents the electric field associated with the light wave. Now, in Fig. 17.1(b), τ_c represents the average duration of the wavetrains, i.e., the electric field remains sinusoidal for times of the order of τ_c. Thus, at a given point, the electric fields at times t and $t + \Delta t$ will, in general, have a definite phase relationship if $\Delta t \ll \tau_c$ and will (almost) never have any phase relationship if $\Delta t \gg \tau_c$. The time duration τ_c is known as the **coherence time** of the source and the field is said to remain coherent for times $\sim\tau_c$. The length of the wavetrain, given by

$$L = c\,\tau_c \qquad \textbf{(17.2)}$$

duration of the wavetrains, i.e., the electric field remains sinusoidal for times of the order of τ_c. Thus, at a given point, the electric fields at times t and $t + \Delta t$ will, in general, have a definite phase relationship if $\Delta t \ll \tau_c$ and will (almost) never have any phase relationship if $\Delta t \gg \tau_c$. The time duration τ_c is known as the **coherence time** of the source and the field is said to remain coherent for times $\sim\tau_c$. The length of the wavetrain, given by

$$L = c\,\tau_c \qquad \textbf{(17.2)}$$

(where c is the speed of light in free space) is referred to as **coherence length**. For example, for the neon line (λ = 6328 Å), $\tau_c \sim 10^{-10}$ sec and for the red cadmium line (λ = 6438 Å), $\tau_c \sim 10^{-9}$ sec; the corresponding coherence lengths are 3 cm and 30 cm, respectively. The finite value of the coherence time τ_c could be due to many factors; for example, if a radiating atom undergoes collision with another atom, then the wavetrain undergoes an abrupt phase shift of the type shown in Fig. 17.1(b). The finite coherence time

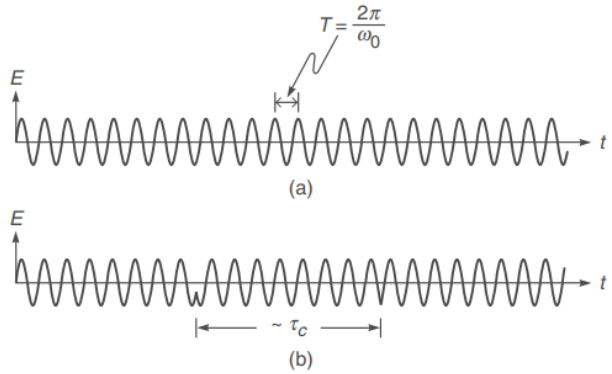

Fig. 17.1 (a) For a perfectly monochromatic beam, the displacement remains sinusoidal for $-\infty < t < +\infty$. (b) For an actual source, a definite phase relationship exists for times of the order of τ_c, which is known as the *temporal coherence of the beam*. For $\nu \sim 5 \times 10^{14}$ Hz and $\tau_c \sim 10^{-10}$ sec, one has about 50,000 oscillations in the time τ_c.

could also be on account of the random motion of atoms or due to the fact that an atom has a finite lifetime in the energy level from which it drops to the lower energy level while radiating.*

In order to understand the concept of coherence time (or of coherence length) we consider Young's double hole experiment as shown in Fig. 17.2; the interference pattern produced by this experimental arrangement was discussed in considerable detail in Sec. 14.4. Now, the interference pattern observed around the point P at time t is due to the superposition of waves emanating from S_1 and S_2 at times $t - r_1/c$ and $t - r_2/c$, respectively, where r_1 and r_2 are the distances S_1P and S_2P, respectively. Obviously, if

$$\frac{r_2 - r_1}{c} \ll \tau_c$$

then the waves arriving at P from S_1 and S_2 will have a definite phase relationship and an interference pattern of good contrast will be obtained. On the other hand, if the path difference $(r_2 - r_1)$ is large enough such that

$$\frac{r_2 - r_1}{c} \gg \tau_c$$

then the waves arriving at P from S_1 and S_2 will have no fixed phase relationship and no interference pattern will be observed. Thus the central fringe (for which $r_1 = r_2$) will, in general, have a good contrast and as we move towards higher order fringes the contrast of the fringes will gradually become poorer. This point is discussed in greater detail in Sec. 17.7.

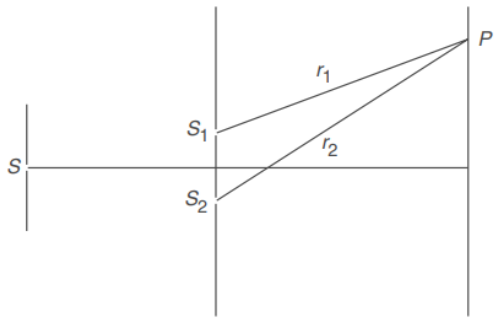

Fig. 17.2 Young's double-hole experiment. The interference pattern observed around the point P at time t is due to the superposition of waves emanating from S_1 and S_2 at times $t - \frac{r_1}{c}$ and $t - \frac{r_2}{c}$, respectively; thus inter-ference fringes of good contrast will be observed at P if $(r_2 - r_1)/c << \tau_c$.

We next consider the Michelson interferometer experiment (see Sec. 15.10). A light beam falls on a beam splitter G (which is usually a partially silvered plate) and the waves reflected from the mirrors M_1 and M_2 interfere (see Fig. 17.3). Let M_2' represent the image of the mirror M_2 (formed by the plate G) as seen by the eye. If the distance M_1M_2' is denoted by d, then the beam which gets reflected by mirror M_2 travels an additional path equal to $2d$. Thus, the beam reflected from M_1 interferes with the beam reflected by M_2 which had originated $2d/c$ seconds earlier.

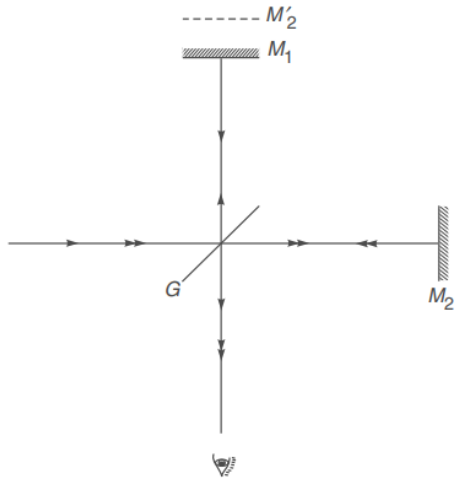

Fig. 17.3 The Michelson interferometer arrangement. G represents the beam splitter. M_2' represents the image of M_2 as formed by G.

If the distance d is such that

$$\frac{2d}{c} << \tau_c$$

then a definite phase relationship exists between the two beams and well-defined interference fringes are observed. On the other hand, if

$$\frac{2d}{c} >> \tau_c$$

then, in general, there is no definite phase relationship between the two beams and no interference pattern is observed. It may be mentioned that there is no definite distance at which the interference pattern disappears; as the distance increases, the contrast of the fringes becomes gradually poorer and eventually the fringe system disappears. For the neon line ($\lambda = 6328$ Å), the disappearance occurs when the path difference is about a few centimetres giving $\tau_c \sim 10^{-10}$ s. On the other hand, for the red cadmium line ($\lambda = 6438$ Å), the coherence length is of the order of 30 cm giving $\tau_c \sim 10^{-9}$ sec.

The coherence time for a laser beam is usually much large in comparison to ordinary light sources. Indeed, for helium–neon laser, coherence times as large as 50 milliseconds have been obtained Ref. 17.9]; this would imply a coherence length of 15,000 km! Commercially available helium–neon lasers have $\tau_c \sim 50$ nsec implying coherence lengths of about 15 m. Thus using such a laser beam, high contrast interference fringes can be obtained even for a path difference of a few metres.

Spatial Coherence

In order to demonstrate the large coherence length of the laser beam we consider an experimental arrangement shown in Fig. 17.4. A parallel beam of light is incident normally on a pair of circular holes. The Fraunhofer diffraction pattern is observed on the focal plane of a convex lens. We first use a helium neon laser beam, the resulting interference pattern is shown in Fig. 17.5(a) which is simply the product of the Airy pattern and the interference pattern produced by two point sources.* We next introduce a $\frac{1}{2}$ mm thick glass plate in front of one of the circular holes; there is almost no change in the interference pattern as can be seen from Fig. 17.5(b). Clearly, the extra path introduced by the plate [$= (\mu - 1)\, t$, see Sec. 14.10] is very small in comparison to the coherence length associated with the laser beam. If we repeat the experiment with a collimated mercury arc beam, we would find that with the introduction of the glass plate the interference pattern disappears (Fig. 17.6). This implies that the extra path length introduced by the glass plate is so large that there is no definite phase relationship between the waves arriving on the screen from the two circular apertures.

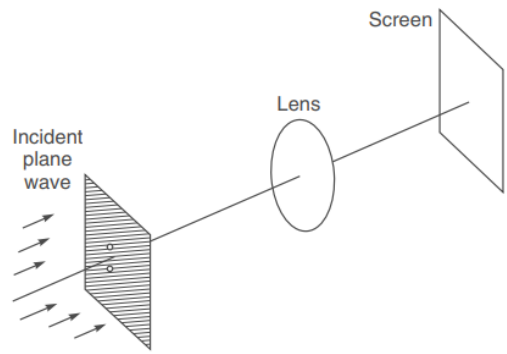

Fig. 17.4 A parallel beam of light is incident normally on a pair of circular holes and the Fraunhofer diffraction pattern is observed on the focal plane of a convex lens.

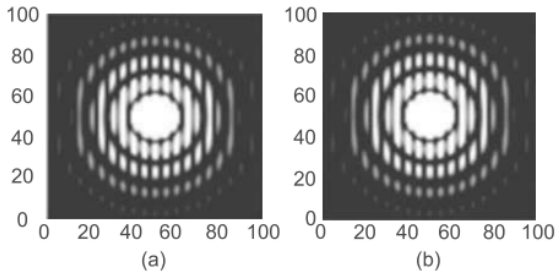

Fig. 17.5 (a) The interference pattern produced for the arrangement shown in Fig. 17.4 using a helium-neon laser beam. (b) The interference pattern produced by the same arrangement with 1 mm thick glass plate in front of one of the holes. (The above figures are computer-generated; the experimentally obtained photographs are very similar—see Ref. 17.16.)

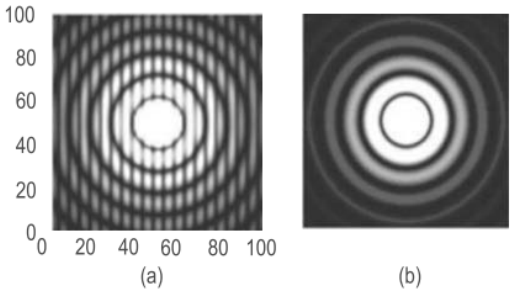

Fig. 17.6 (a) The interference pattern produced for the arrangement shown in Fig. 17.4 using a collimated mercury arc. (b) The interference pattern is washed out when 0.5 mm thick glass plate is introduced in front of one of the holes. (The above figures are computer-generated; the experimentally obtained photographs are very similar—see Ref. 17.16.)

Till now we have considered the coherence of two fields arriving at a particular point in space from a point source through two different optical paths. In this section, we will discuss the coherence properties of the field associated with the finite dimension of the source.

We consider the Young's double-hole experiment with the point source S being equidistant from S_1 and S_2 [see Fig. 17.7(a)]. We assume S to be nearly monochromatic so that it produces interference fringes of good contrast on the screen PP'. The point O on the screen is such that $S_1O = S_2O$. Clearly, the point source S will produce an intensity maximum around the point O. We next consider another similar source S' at a distance l from S. We assume that the waves from S and S' have no definite phase relationship. Thus the interference pattern observed on the screen PP' will be a superposition of the intensity distributions of the interference patterns formed due to S and S' (see Sec. 17.5). If the separation l is slowly increased from zero, the contrast of the fringes on the screen PP' becomes poorer because of the fact that the interference pattern produced by S' is slightly shifted from that produced by S. Clearly, if

$$S'S_2 - S'S_1 = \frac{\lambda}{2} \tag{17.11}$$

the minima of the interference pattern produced by S will fall on the maxima of the interference pattern produced by S' and no fringe pattern will be observed. It can be easily seen that

$$S'S_2 = \left[a^2 + \left(\frac{d}{2} + l\right)^2\right]^{1/2} \simeq a + \frac{1}{2a}\left(\frac{d}{2} + l\right)^2$$

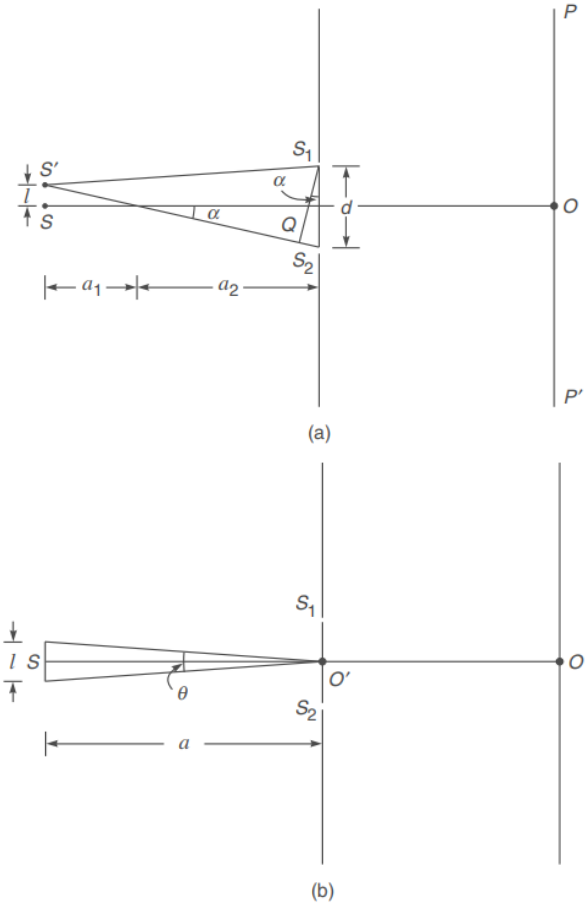

Fig. 17.7 (a) Young's double-hole interference experiment with two independent point sources S and S'. (b) The same experiment with an extended source.

and

$$S'S_1 = \left[a^2 + \left(\frac{d}{2} - l\right)^2\right]^{1/2} \simeq a + \frac{1}{2a}\left(\frac{d}{2} - l\right)^2$$

where

$$a = a_1 + a_2$$

and we have assumed $a >> d, l$. Thus,

$$S'S_2 - S'S_1 \simeq \frac{ld}{a}$$

Thus for the fringes to disappear, we must have

$$\frac{\lambda}{2} = S'S_2 - S'S_1 \simeq \frac{ld}{a}$$

or

$$l \simeq \frac{\lambda a}{2d}$$

Now, if we have an extended incoherent source whose linear dimension is $\sim \lambda a/d$ then for every point on the source, there is a point at a distance of $\lambda a/2d$ which produces fringes which are shifted by half a fringe width. Therefore, the interference pattern will not be observed. Thus for an extended incoherent source, interference fringes of good contrast will be observed only when

$$l << \frac{\lambda a}{d} \tag{17.12}$$

Now, if θ is the angle subtended by the source at the slits [see Fig. 17.7(b)] then $\theta \simeq l/a$ and the above condition for obtaining fringes of good contrast takes the form

$$d << \frac{\lambda}{\theta} \tag{17.13}$$

On the other hand, if

$$d \sim \frac{\lambda}{\theta} \tag{17.14}$$

the fringes will be of very poor contrast. Indeed, a more rigorous diffraction theory tells us that the interference pattern disappears when*

$$d = 1.22\frac{\lambda}{\theta}, 2.25\frac{\lambda}{\theta}, 3.24\frac{\lambda}{\theta}, \ldots \tag{17.15}$$

Thus as the separation of the pinholes is increased from zero, the interference fringes disappear when $d = 1.22\lambda/\theta$; if d is further increased the fringes reappear with relatively poor contrast and they are washed out again when $d = 2.25\lambda/\theta$, and so on. The distance

$$l_w = \lambda/\theta \quad \text{(Lateral coherence width)} \tag{17.16}$$

gives the distance over which the beam may be assumed to be spatially coherent and is referred to as the *lateral coherence width*.

References:

1. Optics, Ajoy Ghatak, 2017, Tata McGraw Hill