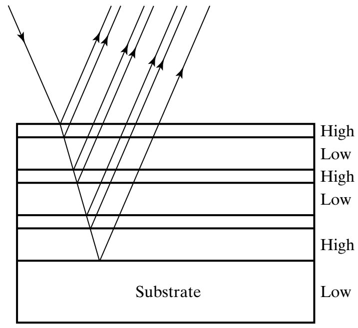
Third Edition

Introduction to Optics

FRANK L. PEDROTTI, S.J.

LENO M. PEDROTTI

LENO S. PEDROTTI



7 Interference of Light

INTRODUCTION

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, due precisely to their superposition, these effects are usually said to be due to the interference of light. When conditions of enhancement, or *constructive interference*, and diminution, or *destructive interference*, alternate in a spatial display, the interference is said to produce a pattern of *fringes*, as in the double-slit interference pattern. The same conditions may lead to the enhancement of one visible wavelength interval or color at the expense of the others, in which case interference colors are produced, as in oil slicks and soap films. The simplest explanation of these phenomena can be undertaken successfully by treating light as a wave motion. In this and following chapters, several such applications, considered under the general heading of interference, are presented.

1 TWO-BEAM INTERFERENCE

We consider first the interference of two *plane waves* of the same frequency, represented by $\vec{\mathbf{E}}_1$ and $\vec{\mathbf{E}}_2$. We may express the two electric fields at a point P where the fields are combined as

$$\vec{\mathbf{E}}_1 = \vec{\mathbf{E}}_{01} \cos(ks_1 - \omega t + \phi_1) \tag{1}$$

$$\vec{\mathbf{E}}_2 = \vec{\mathbf{E}}_{02} \cos(ks_2 - \omega t + \phi_2) \tag{2}$$

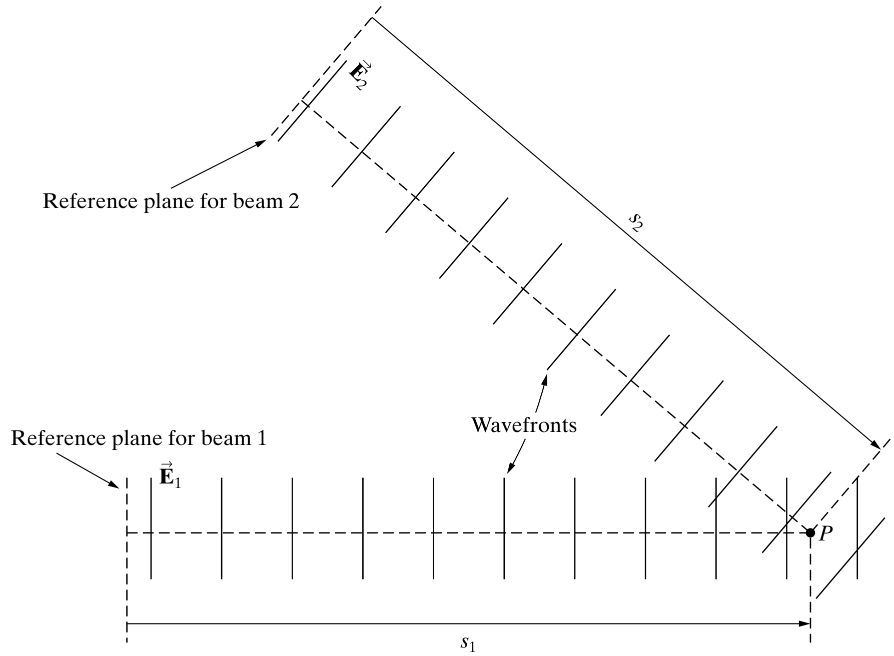


Figure 1 Two-beam interference.

In these relations $k = 2\pi/\lambda$, and s_1 and s_2 can be taken to be the distances traveled by each beam along its respective path from its source to the observation point P. (See Figure 1.) Then ϕ_1 and ϕ_2 represent the phases of these waves at their respective sources at time $t = 0$. These waves combine to produce a disturbance at point P, whose electric field $\vec{\mathbf{E}}_p$ is given by the principle of superposition,

$$\vec{\mathbf{E}}_p = \vec{\mathbf{E}}_1 + \vec{\mathbf{E}}_2$$

It should be noted that $\vec{\mathbf{E}}_1$ and $\vec{\mathbf{E}}_2$ are rapidly varying functions with optical frequencies of order 10^{14} to 10^{15} Hz for visible light. Thus both $\vec{\mathbf{E}}_1$ and $\vec{\mathbf{E}}_2$ average to zero over very short time intervals. Measurement of the waves by their effect on the eye or some other light detector depends on the energy of the light beam. The radiant power density, or *irradiance*, E_e (W/m^2), measures the time average of the square of the wave amplitude. In practice, the time average is carried out by a detector. The averaging time for the eye is on the order of 1/30 of a second; other detectors have averaging times as short as a nanosecond. In general, the averaging time of physical detectors greatly exceeds an optical period ($10^{-14} - 10^{-15}$ s).

Unfortunately, the standard symbol for irradiance, except for the subscript, is the same as that for the electric field. To avoid confusion, we use here the symbol I for irradiance, so that

$$I = \varepsilon_0 c \langle \vec{\mathbf{E}} \cdot \vec{\mathbf{E}} \rangle \tag{3}$$

Thus, the resulting irradiance at P is given by

$$\begin{aligned} I &= \varepsilon_0 c \langle \vec{\mathbf{E}}_p^2 \rangle = \varepsilon_0 c \langle \vec{\mathbf{E}}_p \cdot \vec{\mathbf{E}}_p \rangle \\ &= \varepsilon_0 c \langle (\vec{\mathbf{E}}_1 + \vec{\mathbf{E}}_2) \cdot (\vec{\mathbf{E}}_1 + \vec{\mathbf{E}}_2) \rangle \end{aligned}$$

or

$$I = \varepsilon_0 c \langle \vec{\mathbf{E}}_1 \cdot \vec{\mathbf{E}}_1 + \vec{\mathbf{E}}_2 \cdot \vec{\mathbf{E}}_2 + 2\vec{\mathbf{E}}_1 \cdot \vec{\mathbf{E}}_2 \rangle \tag{4}$$

In Eq. (4), the first two terms correspond to the irradiances of the individual waves, I_1 and I_2. The last term depends on an interaction of the waves and is called the *interference term*, I_{12}. We may then write

$$I = I_1 + I_2 + I_{12} \tag{5}$$

If light behaved without interference, like classical particles, we would then expect $I = I_1 + I_2$. The presence of the third term I_{12} is indicative of the wave nature of light, which can produce enhancement or diminution of the irradiance through interference. Notice that when $\vec{\mathbf{E}}_1$ and $\vec{\mathbf{E}}_2$ are orthogonal, so that their dot product vanishes, no interference results. When the electric fields are parallel, on the other hand, the interference term makes its maximum contribution. Two beams of unpolarized light produce interference because each can be resolved into orthogonal components of $\vec{\mathbf{E}}$ that can then be paired off with similar components of the other beam. Each component produces an interference term with $\vec{\mathbf{E}}_1 \| \vec{\mathbf{E}}_2$ ($\vec{\mathbf{E}}_1$ parallel to $\vec{\mathbf{E}}_2$).

Consider the interference term,

$$I_{12} = 2\varepsilon_0 c\langle \vec{\mathbf{E}}_1 \cdot \vec{\mathbf{E}}_2 \rangle \tag{6}$$

where $\vec{\mathbf{E}}_1$ and $\vec{\mathbf{E}}_2$ are given by Eqs. (1) and (2). Their dot product,

$$\vec{\mathbf{E}}_1 \cdot \vec{\mathbf{E}}_2 = \vec{\mathbf{E}}_{01} \cdot \vec{\mathbf{E}}_{02} \cos(ks_1 - \omega t + \phi_1)\cos(ks_2 - \omega t + \phi_2)$$

can be simplified in an instructive manner using a trigonometric identity. To this end, let us define

$$\alpha \equiv ks_1 + \phi_1 \quad \text{and} \quad \beta \equiv ks_2 + \phi_2$$

so that

$$2\vec{\mathbf{E}}_1 \cdot \vec{\mathbf{E}}_2 = 2\vec{\mathbf{E}}_{01} \cdot \vec{\mathbf{E}}_{02} \cos(\alpha - \omega t)\cos(\beta - \omega t)$$

The identity $2\cos(A)\cos(B) = \cos(A + B) + \cos(B - A)$ helps us cast the time average of $2\vec{\mathbf{E}}_1 \cdot \vec{\mathbf{E}}_2$ as

$$2\langle \vec{\mathbf{E}}_1 \cdot \vec{\mathbf{E}}_2 \rangle = \vec{\mathbf{E}}_{01} \cdot \vec{\mathbf{E}}_{02}[\langle \cos(\alpha + \beta - 2\omega t\rangle + \langle \cos(\beta - \alpha)\rangle]$$

The first time average in this relation is taken over a rapidly oscillating cosine function and so is zero. Thus,

$$\begin{aligned} 2\langle \vec{\mathbf{E}}_1 \cdot \vec{\mathbf{E}}_2 \rangle &= \vec{\mathbf{E}}_{01} \cdot \vec{\mathbf{E}}_{02}\langle \cos(\beta - \alpha)\rangle = \vec{\mathbf{E}}_{01} \cdot \vec{\mathbf{E}}_{02}\langle \cos(k(s_2 - s_1) + \phi_2 - \phi_1)\rangle \\ &\equiv \vec{\mathbf{E}}_{01} \cdot \vec{\mathbf{E}}_{02}\langle \cos\delta\rangle \end{aligned} \tag{7}$$

where we have defined the phase difference between $\vec{\mathbf{E}}_2$ and $\vec{\mathbf{E}}_1$ as

$$\delta = k(s_2 - s_1) + \phi_2 - \phi_1 \tag{8}$$

For purely monochromatic fields, δ is time-independent, in which case $\langle \cos\delta\rangle = \cos\delta$. However, as we will discuss, for real fields, which are not perfectly monochromatic, care must be taken in treating this time average. Combining Eqs. (6) and (7),

$$I_{12} = \varepsilon_0 c\vec{\mathbf{E}}_{01} \cdot \vec{\mathbf{E}}_{02}\langle \cos\delta\rangle \tag{9}$$

The irradiance terms I_1 and I_2 of Eq. (5) can be shown to yield

$$I_1 = \varepsilon_0 c\langle \vec{\mathbf{E}}_1 \cdot \vec{\mathbf{E}}_1\rangle = \varepsilon_0 c E_{01}^2\langle \cos^2(\alpha - \omega t)\rangle = \frac{1}{2}\varepsilon_0 c E_{01}^2 \tag{10}$$

and

$$I_2 = \varepsilon_0 c\langle \vec{\mathbf{E}}_2 \cdot \vec{\mathbf{E}}_2\rangle = \varepsilon_0 c E_{02}^2\langle \cos^2(\beta - \omega t)\rangle = \frac{1}{2}\varepsilon_0 c E_{02}^2 \tag{11}$$

In Eqs. (10) and (11) we used the fact that the time average of the square of a rapidly oscillating sinusoidal function is 1/2. In Eq. (9) when $\mathrm{E}_{01}\|\mathrm{E}_{02}$, their dot product is identical with the product of their magnitudes E_{01} and E_{02}. These may be expressed in terms of I_1 and I_2 by the use of Eqs. (10) and (11), and when combined with Eq. (9) results in

$$I_{12} = 2\sqrt{I_1 I_2}\langle \cos\delta\rangle \tag{12}$$

so that we may write, finally,

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\langle \cos\delta\rangle \tag{13}$$

Notice that once we have made the assumption that the $\vec{\mathbf{E}}$ fields are parallel, the treatment becomes much the same as the scalar theory.

Interference of Mutually Incoherent Fields

In practice, for electric fields $\vec{\mathbf{E}}_1$ and $\vec{\mathbf{E}}_2$ originating from different sources, the time average in Eq. (13) is zero. This occurs because no source is perfectly monochromatic. To model real sources, Eqs. (1) and (2) must be modified to account for departures from monochromaticity. One way to do this is to allow the phases ϕ_1 and ϕ_2 to be functions of time. For laser sources, these phases would typically be random functions of time that vary on a time scale much longer than an optical period but still shorter than typical detector averaging times. The interference term I_{12}, in this case, takes the form,

$$2\sqrt{I_1 I_2}\langle \cos(k(s_2 - s_1) + \phi_2(t) - \phi_1(t))\rangle$$

As stated, for real detectors and for all but those laser sources with state-of-the-art frequency stability, the time average in the preceding relation will be zero. In such a case we say that the sources are *mutually incoherent* and the detected irradiance will be

$$I = I_1 + I_2 \qquad \text{Mutually incoherent beams}$$

It is often said, therefore, that light beams from independent sources, even if both sources are the same kind of laser, do not interfere with each other. In fact, these fields do interfere but the interference term averages to zero over the averaging times of most real detectors.

Interference of Mutually Coherent Beams

If light from the same laser source is split and then recombined at a detector, the time average in Eq. (13) need not be zero. This occurs because the departures from monochromaticity of each beam, while still present, will be correlated since both beams come from the same source. In this case, the phase difference $\phi_2(t) - \phi_1(t)$ will be strictly zero if the beams travel paths of *equal duration* before being recombined at the detector. In such a case, δ is a constant and the interference term takes the form,

$$2\sqrt{I_1 I_2}\langle\cos(k(s_2 - s_1) + \phi_1(t) - \phi_1(t))\rangle = 2\sqrt{I_1 I_2}\cos(k(s_2 - s_1))$$
$$= 2\sqrt{I_1 I_2}\cos\delta$$

Even if the electric fields travel paths that differ in duration by a time δt, the phase difference resulting from the departure from monochromaticity, $\phi_1(t) - \phi_1(t + \delta t)$, will still be nearly zero so long as δt is less than the so-called *coherence* time, τ_0, of the source. Qualitatively, the coherence time of the source is the time interval over which departures from monochromaticity are small. You will learn the coherence time of a source is inversely proportional to the range of frequencies, $\Delta\nu$, of the components that make up the electric field. That is,

$$\tau_0 = \frac{1}{\Delta\nu}$$

Associated with the coherence time of a source is a coherence length, $l_t = c\tau_0$, which is the distance that the electric field travels in a coherence time. For a white light source the coherence length is about 1 μm; laser sources have coherence lengths that range from tens of centimeters to tens of kilometers. Throughout the rest of this chapter, we will presume that the difference in the lengths of paths traveled by beams originating from the same source is considerably less than the coherence length of the source. In such a case, the electric fields are said to be *mutually* coherent and the irradiance of the combined fields will have the form

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos\delta \qquad \text{Mutually coherent beams} \qquad (14)$$

where δ is the total phase difference at the point of recombination of the beam. As we have noted, if the beams originate from the same source, this phase difference accumulates as a result of a difference in path lengths traveled by the respective beams. In many cases of interest, other factors can lead to a phase difference between the beams as well. Important mechanisms of this sort include differing phase shifts due to reflection from beam splitters and differing indices of refraction in the separate paths taken by the two beams. Depending on whether $\cos\delta > 0$ or $\cos\delta < 0$ in Eq. (14), the interference term either augments or diminishes the sum of the individual irradiances I_1 and I_2, leading to constructive or destructive interference, respectively. Since the relative distances traveled by the two beams will, in general, differ for different observation points in the region of overlap, the phase difference δ will also differ for different observation points. Typically, $\cos\delta$ will take on alternating maximum and minimum values, and *interference fringes*, spatially separated, will occur in the observation plane.

To be more specific, when $\cos\delta = +1$, constructive interference yields the maximum irradiance

$$I_{\text{max}} = I_1 + I_2 + 2\sqrt{I_1 I_2} \qquad (15)$$

This condition occurs whenever the phase difference $\delta = 2m\pi$, where m is any integer or zero. On the other hand, when $\cos\delta = -1$, destructive interference yields the minimum, or background, irradiance

$$I_{\text{min}} = I_1 + I_2 - 2\sqrt{I_1 I_2} \qquad (16)$$

a condition that occurs whenever $\delta = (2m + 1)\pi$. A plot of irradiance I versus phase δ, in Figure 2a, exhibits periodic fringes. Destructive interference is complete, that is, cancellation is complete, when $I_1 = I_2 = I_0$. Then, Eqs. (15) and (16) give

$$I_{\text{max}} = 4I_0 \quad \text{and} \quad I_{\text{min}} = 0$$

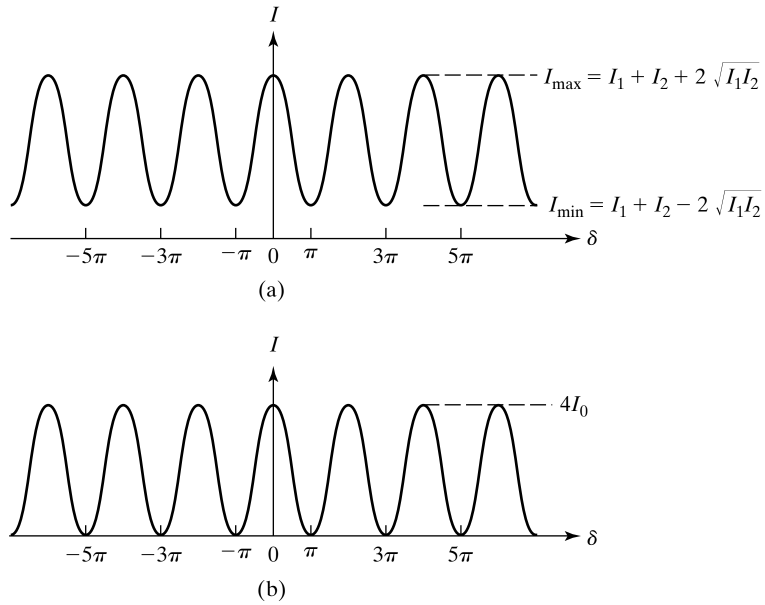


Figure 2 Irradiance of interference fringes as a function of phase difference δ. Visibility is enhanced in (b), where the background irradiance $I_{min} = 0$ when $I_1 = I_2$.

Resulting fringes, shown in Figure 2b, now exhibit better contrast. A measure of *fringe contrast*, called *visibility*, with values between 0 and 1, is given by the quantity

$$\text{visibility} = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \tag{17}$$

In the experimental utilization of fringe patterns, it is therefore usually desirable to ensure that the interfering beams have the same amplitudes.

Another useful form of Eq. (14), for the case of interfering beams of equal amplitude so that $I_1 = I_2 = I_0$, is found by writing

$$I = I_0 + I_0 + 2\sqrt{I_0^2}\cos\delta = 2I_0(1 + \cos\delta)$$

and then making use of the trigonometric identity

$$1 + \cos\delta \equiv 2\cos^2\left(\frac{\delta}{2}\right)$$

The irradiance for two equal interfering beams is then

$$I = 4I_0\cos^2\left(\frac{\delta}{2}\right) \tag{18}$$

Notice that energy is not conserved at each point of the superposition, that is, $I \neq 2I_0$, but that over at least one spatial period of the fringe pattern $I_{av} = 2I_0$. This situation is typical of interference and diffraction phenomena: If the power density falls below the average at some points, it rises above the average at other points in such a way that the total pattern satisfies the principle of energy conservation.

Example 1

Consider two interfering beams with parallel electric fields that are superposed. Take the electric fields of the individual beams to be

$$E_1 = 2\cos(ks_1 - \omega t) \qquad (\text{kV/m})$$

$$E_2 = 5\cos(ks_2 - \omega t) \qquad (\text{kV/m})$$

Let us determine the irradiance contributed by each beam acting alone and that due to their mutual interference at a point where their path difference is such that $k(s_2 - s_1) = \pi/12$. We have

$$I_1 = \tfrac{1}{2}\varepsilon_0 c E_{01}^2 = \tfrac{1}{2}\varepsilon_0 c(2000)^2 = 5309\ \text{W/m}^2$$
$$I_2 = \tfrac{1}{2}\varepsilon_0 c E_{02}^2 = \tfrac{1}{2}\varepsilon_0 c(5000)^2 = 33{,}180\ \text{W/m}^2$$
$$I_{12} = 2\sqrt{I_1 I_2}\cos\delta = 2\sqrt{(5309 \times 33180)}\cos(\pi/12) = 25{,}640\ \text{W/m}^2$$

To find the visibility near this point of recombination, we must calculate

$$I_{\max} = I_1 + I_2 + 2\sqrt{I_1 I_2} = 5309 + 33180 + 2\sqrt{(5309 \times 33180)}$$
$$= 65{,}034\ \text{W/m}^2$$
$$I_{\min} = I_1 + I_2 - 2\sqrt{I_1 I_2} = 5309 + 33180 - 2\sqrt{(5309 \times 33180)}$$
$$= 11{,}945\ \text{W/m}^2$$

The visibility is then given by Eq. (17), or

$$\text{visibility} = \frac{65{,}034 - 11{,}945}{65{,}034 + 11{,}945} = 0.690$$

If the amplitudes of the two waves were equal, then $I_{\max} = 4I_0$, $I_{\min} = 0$, and the visibility would be 1.

In the analysis leading to the irradiance that results from the superposition of two mutually coherent beams, Eq. (14), we assumed that the individual beams were plane waves described by Eqs. (1) and (2). In fact, the analysis holds for any sort of harmonic wave (e.g., spherical, cylindrical, or Gaussian). However, for these types of waves, the amplitudes E_{01} and E_{02} (and so the irradiances I_1 and I_2) depend on the distance from the source to the observation point.

2 YOUNG'S DOUBLE-SLIT EXPERIMENT

The decisive experiment performed by Thomas Young in 1802 is shown schematically in Figure 3. Monochromatic light is first allowed to pass through a single small hole in order to approximate a single point source S. The light spreads out in spherical waves from the source S according to Huygens' principle and is allowed to fall on a plane with two closely spaced holes, S_1 and S_2. In a modern version of this experiment, a laser is typically used to illuminate the two holes. In either case, the holes become two coherent sources of light, whose interference can be observed on a screen some distance away. If the two holes are equal in size, light waves emanating from the holes have comparable amplitudes, and the irradiance at any point of superposition is given by Eq. (18). Referring to Figure 3, we will now develop an expression for the irradiance at observation points such as P on a screen that is a distance L from the plane containing the two holes S_1 and S_2. The phase difference δ between the two waves arriving at the observation point P must be determined to calculate the resultant irradiance there. Clearly, if $S_2P - S_1P = s_2 - s_1 = m\lambda$, the waves will arrive in phase, and maximum irradiance or brightness results. If $s_2 - s_1 = (m + \frac{1}{2})\lambda$, the requisite condition for destructive interference or darkness is met. Practically speaking, the hole separation a is much smaller than the screen distance L, allowing a simple expression for the path distance, $s_2 - s_1$. Using P as a center, let an arc S_1Q be drawn of radius s_1 so that it intersects the line S_2P at Q. Then $s_2 - s_1$ is equal to the segment Δ, as shown. The first approximation is to regard arc S_1Q as a straight-line segment that

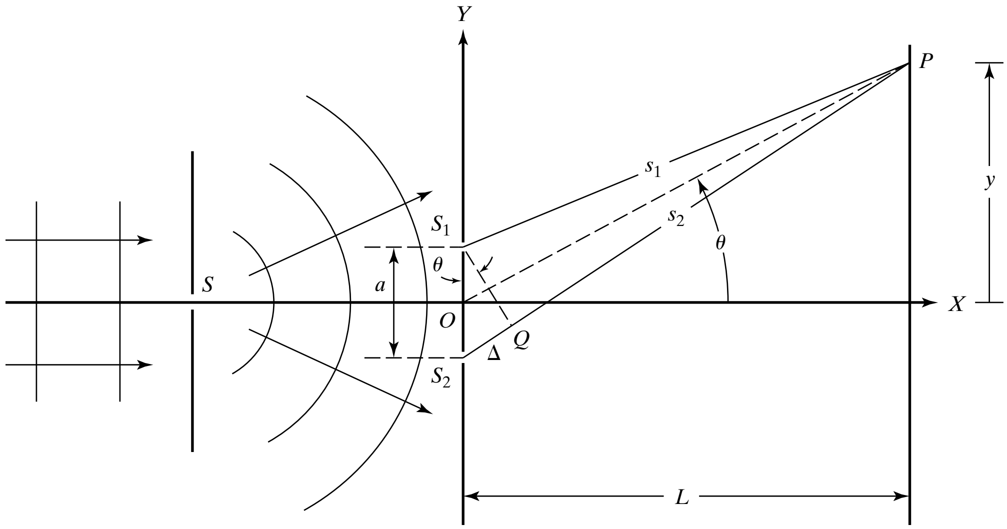


Figure 3 Schematic for Young's double-slit experiment. The holes S_1 and S_2 are usually slits, with the long dimensions extending into the page. The hole at S is not necessary if the source is a spatially coherent laser.

forms one leg of the right triangle S_1S_2Q. If θ is the angle between the line segments S_1S_2 and S_1Q, then $\Delta = a \sin\theta$. The second approximation identifies the angle θ with the angle between the optical axis OX and the line drawn from the midpoint O between holes to the point P at the screen. Observe that the corresponding sides of the two angles θ are related such that $OX \perp S_1S_2$, and OP is almost exactly perpendicular to S_1Q.

The condition for *constructive interference* at a point P on the screen is, then, to a very good approximation

$$s_2 - s_1 = \Delta = m\lambda \cong a \sin\theta \tag{19}$$

whereas for *destructive interference*,

$$\Delta = \left(m + \tfrac{1}{2}\right)\lambda \cong a \sin\theta \tag{20}$$

where m is zero or of integral value. Typically, at observation points of interest, the electric field *amplitudes* of the beams originating from the two slits are nearly the same so that the irradiance on the screen, at a point determined by the angle θ, is found using Eq. (18) and the relationship between path difference Δ and phase difference δ,

$$\delta = k(s_2 - s_1) = \frac{2\pi}{\lambda}\Delta$$

The result is

$$I = 4I_0 \cos^2\left(\frac{\pi\Delta}{\lambda}\right) = 4I_0 \cos^2\left(\frac{\pi a \sin\theta}{\lambda}\right)$$

For points P near the optical axis, where $y \ll L$, we may approximate further: $\sin\theta \cong \tan\theta \cong y/L$, so that

$$I = 4I_0 \cos^2\left(\frac{\pi a y}{\lambda L}\right) \tag{21}$$

By allowing the cosine function in Eq. (21) to become alternately ± 1 and 0, the conditions expressed by Eqs. (19) and (20) for constructive and destructive interference are reproduced.

Arguing now from Eq. (19) and the small angle relation $\sin\theta \cong \tan\theta \cong y/L$, we find the bright fringe positions to be given by

$$y_m = \frac{m\lambda L}{a}, \qquad m = 0, \pm 1, \pm 2, \ldots \tag{22}$$

Consequently, there is a constant separation between irradiance maxima, corresponding to successive values of m, given by

$$\Delta y = y_{m+1} - y_m = \frac{\lambda L}{a} \tag{23}$$

with minima situated midway between the maxima. Thus, fringe separation is proportional both to wavelength and screen distance and inversely proportional to the hole spacing. Reducing the hole spacing expands the fringe pattern formed by each color. Measurement of the fringe separation provides a means of determining the wavelength of the light. The single hole, used to secure a degree of spatial coherence, may be eliminated if laser light, both highly monochromatic and spatially coherent, is used to illuminate the double slit. In the observational arrangement just described, fringes are observed on a screen placed perpendicular to the optical axis at some distance from the aperture, as indicated in Figure 4. Fringe maxima coincide with integral orders of m, and fringe minima fall halfway between adjacent maxima.

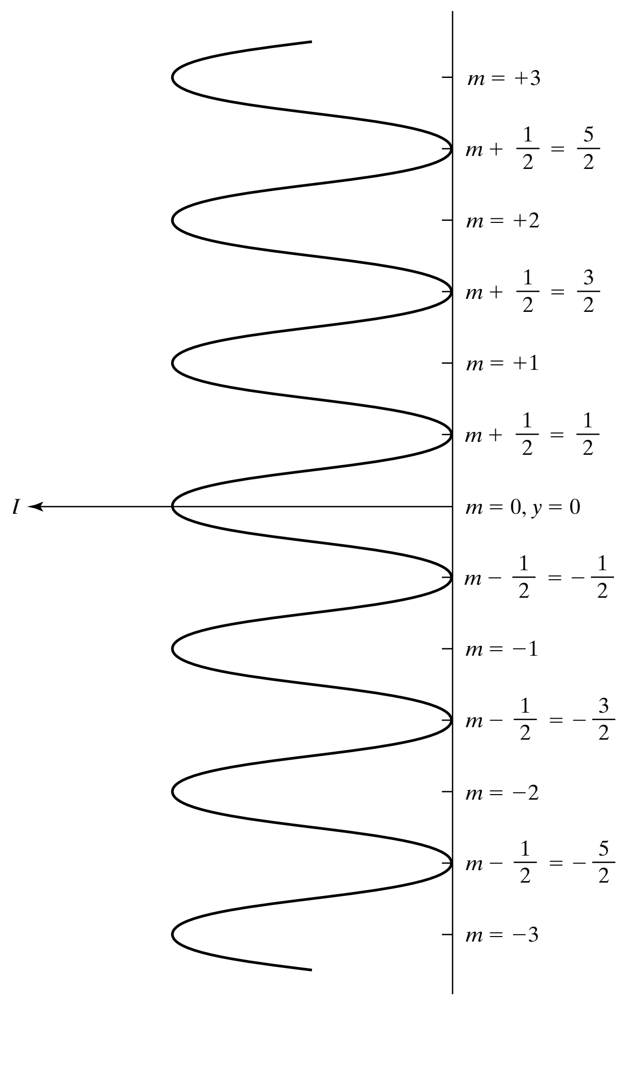


Figure 4 Irradiance versus distance from the optical axis for a double-slit fringe pattern. The *order* of the interference pattern is indicated by m, with integral values of m determining positions of fringe maxima.

Example 2

Laser light passes through two identical and parallel slits, 0.2 mm apart. Interference fringes are seen on a screen 1 m away. Interference maxima are separated by 3.29 mm. What is the wavelength of the light? How does the irradiance at the screen vary, if the contribution of one slit alone is I_0?

Solution

From Eq. (23),

$$\lambda = a\Delta y/L = (0.0002\text{ m})(3.29 \times 10^{-3}\text{ m})/(1\text{ m})$$
$$= 6.58 \times 10^{-7}\text{ m} = 658\text{ nm}$$

According to Eq. (21), $I = 4I_0 \cos^2[\pi a y/\lambda L]$. In this case,

$$I = 4I_0 \cos^2[\pi(0.0002)y/(658 \times 10^{-9})(1\text{m})] = 4I_0 \cos^2[(955/\text{m})y]$$

An alternative way to view the formation of bright (B) positions of constructive interference and dark (D) positions of destructive interference is shown in Figure 5. The crests and valleys of spherical waves from S_1 and S_2 are shown approaching the screen. Along directions marked B, wave crests (or wave valleys) from both slits coincide, producing maximum irradiance. Along directions marked D, on the other hand, the waves are seen to be out of step by half a wavelength, and destructive interference results.

Obviously, fringes should be present in all the space surrounding the holes, where light from the holes is allowed to interfere, though the irradiance is greatest in the forward direction. If we imagine two coherent point sources of light radiating in all directions, then the condition given by Eq. (19) for bright fringes,

$$s_2 - s_1 = m\lambda \tag{24}$$

defines a family of bright fringe surfaces in the space surrounding the holes. To visualize this set of surfaces, we may take advantage of the inherent symmetry in the arrangement. In Figure 6, the intersection of several bright fringe surfaces with a plane that includes the two sources is shown, each surface corresponding

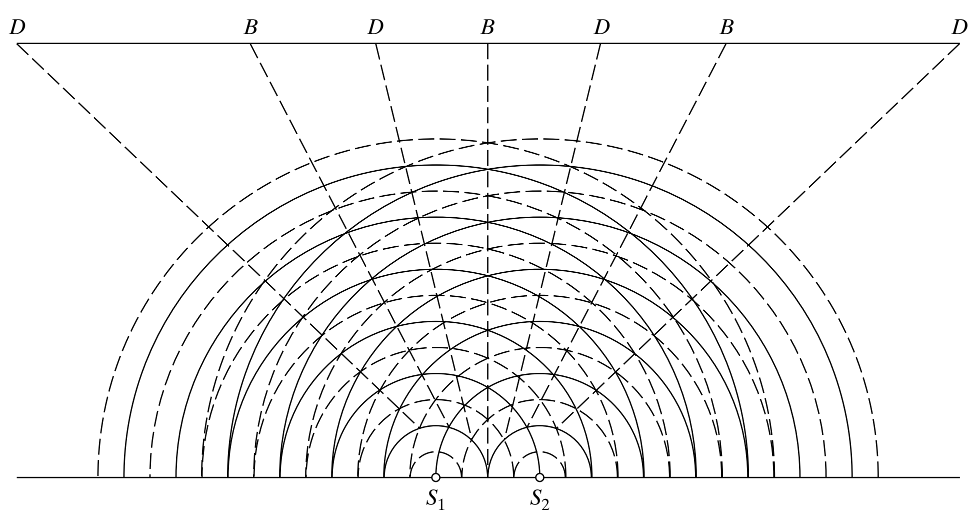


Figure 5 Alternating bright and dark interference fringes are produced by light from two coherent sources. Along directions where crests (solid circles) from S_1 intersect crests from S_2, brightness (B) results. Along directions where crests meet valleys (dashed circles), darkness (D) results.

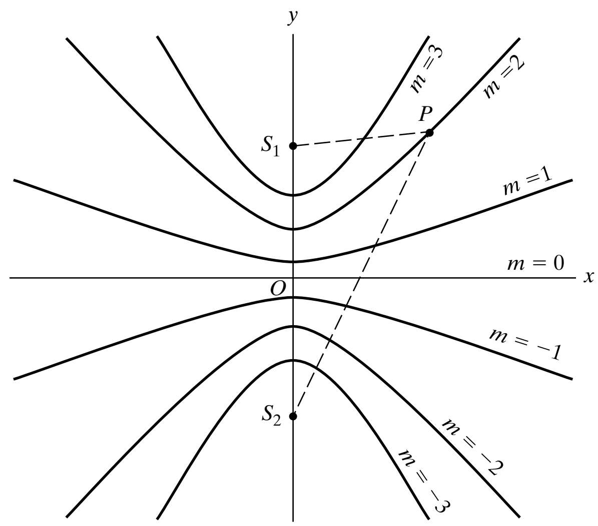


Figure 6 Bright fringe surfaces for two coherent point sources. The distances from S_1 and S_2 to any point P on a bright fringe surface differ by an integral number of wavelengths. The surfaces are generated by rotating the pattern about the y-axis.

to an integral value of order m. The surfaces are hyperbolic, since Eq. (24) is precisely the condition for a family of hyperbolic curves with parameter m. Inasmuch as the y-axis is an axis of symmetry, the corresponding bright fringe surfaces are generated by rotating the entire pattern about the y-axis. One should then be able to visualize the intercept of these surfaces with the plane of an observational screen placed anywhere in the vicinity. In particular, a screen placed perpendicular to the OX axis, as in Figure 3, intercepts hyperbolic arcs that appear as straight-line fringes near the axis, whereas a screen placed perpendicular to the OY axis shows concentric circular fringes centered on the axis. Because the fringe system extends throughout the space surrounding the two sources, the fringes are said to be *nonlocalized*.

The holes S, S_1, and S_2 of Figure 3 are usually replaced by parallel, narrow slits (oriented with their long sides perpendicular to the page in Figure 3) to illuminate more fully the interference pattern. The effect of the array of point sources along the slits, each set producing its own fringe system as just described, is simply to elongate the pattern parallel to the fringes, without changing their geometrical relationships. This is true even when two points along a source slit are not mutually coherent.

3 DOUBLE-SLIT INTERFERENCE WITH VIRTUAL SOURCES

Interference fringes may sometimes appear in arrangements when only one light source is present. It is possible, through reflection or refraction, to produce virtual images that, acting together or with the actual source, behave as two coherent sources that can produce an interference pattern. Figures 7 to 9 illustrate three such examples. These examples are not only of some historic importance; they also serve to impress us with the variety of ways unexpected fringe patterns may appear in optical experiments, especially when the extremely coherent light of a laser is being used.

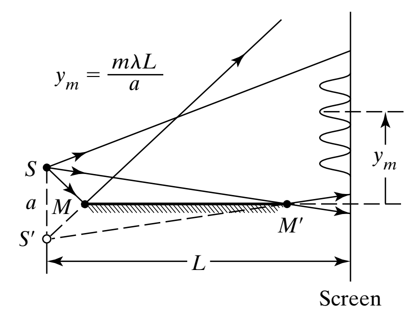


Figure 7 Interference with Lloyd's mirror. Coherent sources are the point source S and its virtual image, S'.

Lloyd's Mirror

In Figure 7, interference fringes are produced due to the superposition of light at the screen that originates at the actual source S and, by reflection, also originates effectively from its virtual source S' below the surface of the plane mirror MM'. Where the direct and reflected beams strike the screen, fringes will appear. The position of bright fringes is given by Eq. (22), where a is

PROBLEMS

1 Two mutually coherent beams having parallel electric fields are described by

$$E_1 = 3\cos\left(ks_1 - \omega t + \frac{\pi}{5}\right)$$
$$E_2 = 4\cos\left(ks_2 - \omega t + \frac{\pi}{6}\right)$$

with amplitudes in kV/m. The beams interfere at a point P where the phase difference due to path is $\pi/3$ (the first beam having the longer path). At the point of superposition, calculate (a) the irradiances I_1 and I_2 of the individual beams; (b) the irradiance I_{12} due to their interference; (c) the net irradiance; (d) the fringe visibility.

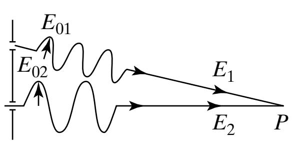


Figure 23 Problem 1.

2 Two harmonic light waves with amplitudes of 1.6 and 2.8 interfere at some point P on a screen. What visibility results there if (a) their electric field vectors are parallel and (b) if they are perpendicular?

3 The ratio of the amplitudes of two beams forming an interference fringe pattern is 2/1. What is the visibility? What ratio of amplitudes produces a visibility of 0.5?

4 **a.** Show that if one beam of a two-beam interference setup has an irradiance of N times that of the other beam, the fringe visibility is given by

$$V = \frac{2\sqrt{N}}{N + 1}$$

b. Determine the beam irradiance ratios for visibilities of 0.96, 0.9, 0.8, and 0.5.

5 A mercury source of light is positioned behind a glass filter, which allows transmission of the 546.1-nm green light from the source. The light is allowed to pass through a narrow, horizontal slit positioned 1 mm above a flat mirror surface. Describe both qualitatively and quantitatively what appears on a screen 1 m away from the slit.

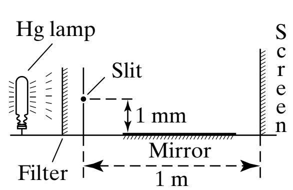


Figure 24 Problem 5.

6 Two slits are illuminated by light that consists of two wavelengths. One wavelength is known to be 436 nm. On a screen, the fourth minimum of the 436-nm light coincides with the third maximum of the other light. What is the wavelength of the other light?

7 In a Young's experiment, narrow double slits 0.2 mm apart diffract monochromatic light onto a screen 1.5 m away. The distance between the fifth minima on either side of the zeroth-order maximum is measured to be 34.73 mm. Determine the wavelength of the light.

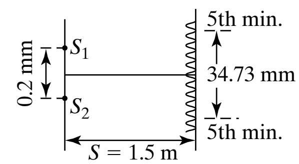


Figure 25 Problem 7.

8 A quasi-monochromatic beam of light illuminates Young's double-slit setup, generating a fringe pattern having a 5.6-mm separation between consecutive dark bands. The distance between the plane containing the apertures and the plane of observation is 7 m, and the two slits are separated by 1.0 mm. Sketch the experimental arrangement. Why is an initial single slit necessary? What is the wavelength of the light?

9 In an interference experiment of the Young type, the distance between slits is 0.5 mm, and the wavelength of the light is 600 nm.

a. If it is desired to have a fringe spacing of 1 mm at the screen, what is the proper screen distance?
b. If a thin plate of glass ($n = 1.50$) of thickness 100 microns is placed over one of the slits, what is the lateral fringe displacement at the screen?
c. What path difference corresponds to a shift in the fringe pattern from a peak maximum to the (same) peak half-maximum?

10 White light (400 to 700 nm) is used to illuminate a double slit with a spacing of 1.25 mm. An interference pattern falls on a screen 1.5 m away. A pinhole in the screen allows some light to enter a spectrograph of high resolution. If the pinhole in the screen is 3 mm from the central white fringe, where would one expect dark lines to show up in the spectrum of the pinhole source?

11 Sodium light (589.3 nm) from a narrow slit illuminates a Fresnel biprism made of glass of index 1.50. The biprism is twice as far from a screen on which fringes are observed as it is from the slit. The fringes are observed to be separated by 0.03 cm. What is the biprism angle α?

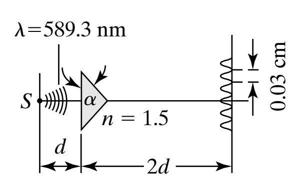
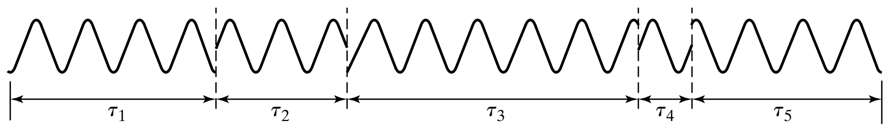


Figure 26 Problem 11.

12 The small angle θ between two plane, adjacent reflecting surfaces is determined by examining the interference fringes produced in a Fresnel mirror experiment. A source slit is parallel to the intersection between the mirrors and 50 cm



9 *Coherence*

INTRODUCTION

The term *coherence* is used to describe the correlation between phases of monochromatic radiations. Beams with random phase relationships are, generally speaking, incoherent beams, whereas beams with a constant phase relationship are coherent beams. The requirement of coherence between interfering beams of light, if they are to produce observable fringe patterns, should be familiar to you, as should the relationship between coherence and the net irradiance of interfering beams. In the superposition of in-phase coherent beams, individual amplitudes add together, whereas in the superposition of incoherent beams, individual irradiances add together. In this chapter, we examine the property of coherence in greater detail, distinguishing between *longitudinal coherence*, which is related to the spectral purity of the source, and *lateral* or *spatial coherence*, which is related to the size of the source. We also describe a quantitative measure of *partial coherence*, the condition under which most experimental measurements of interference take place. We begin our treatment with a brief description of Fourier analysis, which we will need in this chapter.

1 FOURIER ANALYSIS

When a number of harmonic waves of the same frequency are added together, even though they differ in amplitude and phase, the result is again a harmonic wave of the given frequency. If the superposed waves differ in frequency as well, the result is periodic but anharmonic and may assume an arbitrary

5 SPATIAL COHERENCE

In speaking of temporal coherence, we have been considering the correlation in phase between temporally distinct points of the radiation field of a source along its line of propagation. For this reason, temporal coherence is also called *longitudinal coherence*. The degree of coherence can be observed by examining the interference fringe contrast in an amplitude-splitting instrument, such as the Michelson interferometer. As we have seen, temporal coherence is a measure of the average length of the constituent harmonic waves, which depends on the radiation properties of the source. In contrast, we now turn our attention to what is referred to as *spatial*, or *lateral, coherence*, the correlation in phase between spatially distinct points of the radiation field. This type of coherence is important when using a wavefront-splitting device, such as the double slit. The quality of the interference pattern in the double-slit experiment depends on the degree of coherence between distinct regions of the wavefield at the two slits.

To sharpen our understanding of the coherence of a wavefield radiating from a source, consider the situation depicted in Figure 11. Light from a source S passes through a double slit and is also sampled by a Michelson interferometer located nearby. Spatial coherence between wavefront points A and B at the slits is insured as long as the source S is a true point source. In that case, all rays emanating from S are associated with a single set of spherical waves that have the same phase on any given wavefront. Are clear distinguishable fringes then formed on a screen near point P_1? The answer, of course, depends on whether the light from S, traveling along the two distinct paths SAP_1 and SBP_1, is *temporally* as well as spatially coherent. The matter of temporal coherence requires a comparison between the path difference $\Delta = SAP_1 - SBP_1$ and the coherence length of the radiation. This is equivalent to a comparison of coherence along any radial direction of light propagation from the source at two wavefronts separated by the same path difference. It is this property of temporal coherence that is measured by the Michelson interferometer. If the path difference Δ is much less than the coherence length $(\Delta \ll l_t)$, clean interference fringes are formed at P_1; if the path difference is equal to or greater than the coherence length $(\Delta \geq l_t)$, interference fringes are poorly defined or absent altogether. In practice, of course, S is always an extended source, so that rays reach A and B from many points of the source. In ordinary (nonlaser) sources, light emitted by different points of a source, well over a wavelength in separation, is not correlated in phase and so lacks coherence. Thus, the spatial coherence of light at

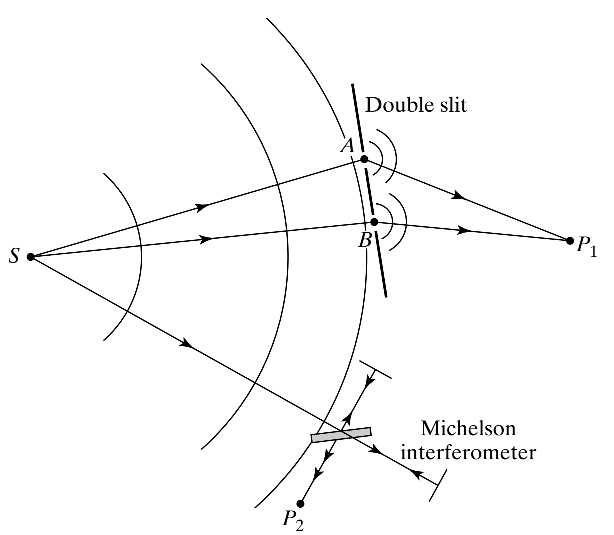


Figure 11 Wavefront and amplitude division of radiation from source S, illustrating the practical requirements of spatial and temporal coherence.

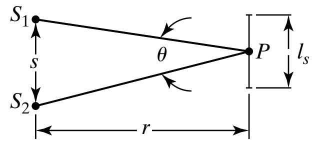


Figure 12 Lateral region of coherence l_s, due to two independent point sources.

the slits A and B depends on how closely the source S resembles a point source of light, either in extension or in its actual coherence properties.

We show in the next section that if two source points S_1 and S_2, as in Figure 12, are separated by a distance s and if light of wavelength λ from these sources is observed at a distance r away, there will be a region of high spatial coherence of dimension l_s, given by

$$l_s < \frac{\lambda}{\theta} \tag{35}$$

where θ is the angle subtended by the point sources at the observation point P. Accepting this result for the moment and combining it with the temporal or longitudinal coherence length l_t, we conclude that there exists at any point in the radiation field of a real light source a region of space in which the light is coherent. This region has lateral dimensions of l_s and longitudinal dimensions of l_t relative to the source and thus occupies a volume of roughly $l_s^2 l_t$ around the point P. It is from this volume that any interferometer must accept radiation if it is to produce observable interference fringes.

6 SPATIAL COHERENCE WIDTH

Consider now the spatial coherence at points P_1 and P_2 in the radiation field of a quasi-monochromatic extended source, simply represented by two mutually incoherent emitting points A and B at the edges of the source (Figure 13). We may think of P_1 and P_2 as two slits that propagate light to a screen, where interference fringes may be viewed. Each point source, acting alone, then produces a set of double-slit interference fringes on the screen. When both sources act together, however, the fringe systems overlap. If the fringe systems overlap with their maxima and minima falling together, the resulting fringe pattern is highly visible, and the radiation from the two incoherent sources is considered highly coherent! When the fringe systems are relatively displaced, however, so that the maxima of one fall on the minima of the other, the composite pattern is not visible and the radiation is considered incoherent. Suppose that source B is at the position of source A, or that the distance s in Figure 13 is zero. The fringe systems at the screen then coincide and correspond to the fringes of a single point source. A maximum in the interference pattern occurs at P if P lies on the perpendicular bisector of the two slits. In this condition,

$$BP_2 - BP_1 = AP_2 - AP_1 = 0$$

If source B is moved below A, the fringe systems separate until, at a certain distance s, where

$$BP_2 - BP_1 = \Delta = \frac{\lambda}{2}$$

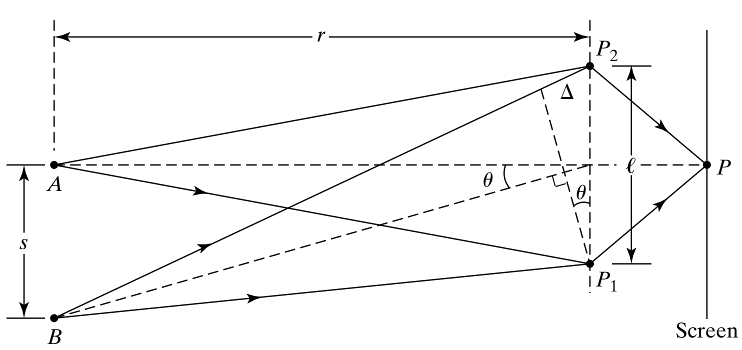


Figure 13 Light from each of two point sources A and B reach points P_1 and P_2 in the radiation field and are allowed to interfere at the screen. In practice, $s \ll \ell$ and angles θ are approximately equal.

the maximum in the fringe system at P due to source B is replaced by a minimum, and the composite fringe pattern disappears.

If the angle θ represents the angular separation of the sources from the plane of the slits, then from the diagram, $\Delta \cong \ell\theta$, where ℓ is the distance between slits, and $\theta \cong s/r$, where r is the distance to the sources. It follows that

$$\Delta = \frac{\lambda}{2} = \frac{s\ell}{r} \quad \text{or} \quad s = \frac{r\lambda}{2\ell} \tag{36}$$

When the distance AB is considered instead to be a continuous array of point sources, the individual fringe systems do not give complete cancellation until the spatial extent AB of the source reaches twice the value of s in Eq. (36). If extreme points are separated by an amount $s < r\lambda/\ell$, then fringe definition is assured. Regarding this result as describing instead the maximum slit separation ℓ, given a source dimension s, we have for the *spatial coherence width* ℓ_s,

$$\ell_s < \frac{r\lambda}{s} \cong \frac{\lambda}{\theta} \tag{37}$$

As ℓ_s is restricted to smaller fractions of this value, the fringe contrast is correspondingly improved.

According to this argument, moving the source B even farther should bring the fringe system into coincidence again, so that the degree of coherence $|\gamma|$ between P_1 and P_2 is a periodic function. In a more complete mathematical argument, the extended source is represented by a continuous array of elemental emitting areas rather than by two point sources. Results show that outside the coherence width given by Eq. (37), the fringe visibility, while oscillatory, is negligible. According to a general theorem, known as the *Van Cittert-Zernike theorem*[1], a plot of the degree of coherence versus spatial separation ℓ of points P_1 and P_2 is the same as a plot of the diffraction pattern due to an aperture of the same size and shape as the extended source.

The significance of Eq. (37) is apparent in the case of Young's double-slit experiment, where an extended source is used together with a single slit to render the light striking the double slit reasonably coherent, as in Figure 14. We may now use Eq. (37) to determine how small the single slit must be to ensure coherence and the production of fringes at the screen. The two slits S_1 and S_2 must fall within the lateral coherence width l_s due to the primary slit of width s.

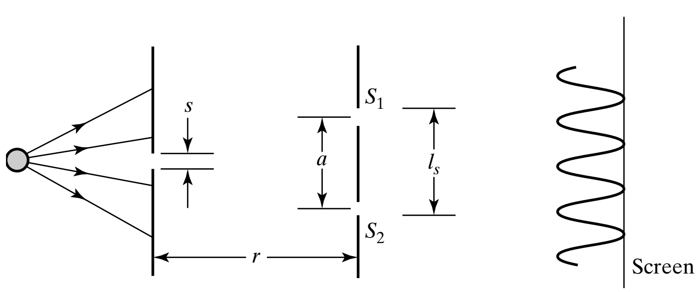


Figure 14 Young's double-slit setup. Slits S_1 and S_2 must fall within the lateral coherence width l_s due to the single-slit source.

[1]Born, M. and E. Wolf. *Principals of Optics*, 5th ed., (New York: Pergamon Press, 1975.)

Example 2

Let the source-to-slit distance be 20 cm, the slit separation 0.1 mm, and the wavelength 546 nm. Determine the maximum width of the primary or single slit.

Solution

Using Eq. (37),

$$s < \frac{r\lambda}{l_s} = \frac{(0.2)(546 \times 10^{-9})}{1 \times 10^{-4}} = 1.1 \text{ mm}$$

Now suppose that the source slit in the example is made exactly 1.1 mm in width and that the separation between slits S_1 and S_2 is adjustable. When the slits are very close together ($a \ll l_s$), they fall within a high coherence region and the fringes in the interference pattern appear sharply defined. As the slits are moved farther apart, the degree of coherence $|\gamma|$ decreases and the fringe contrast begins to degrade. When the slit separation a reaches a value of 0.1 mm, $|\gamma| = 0$ and the fringes disappear. Evidently an experimental determination of this slit separation could be used to deduce the size s of the extended source. This technique was employed by Michelson to measure the angular diameter of stars. Stars are so distant that imaging techniques are unable to resolve their diameters. If a star is regarded as an extended, incoherent source with light emanating from a continuous array of points extending across a diameter s of the star (see Figure 15b), then the spatial coherent width l_s in Eq. (38) becomes

$$l_s < \frac{1.22\lambda}{\theta} \tag{38}$$

Here the factor 1.22 arises from the circular shape of the source, as it does in the Fraunhofer diffraction of a circular aperture. Since the angular diameter θ of a star is extremely small, l_s will be correspondingly large. The movable slits were therefore arranged as in Figure 15a, using mirrors that direct widely separated portions of the radiation wavefront into a double-slit-telescope instrument. The spacing of the interference fringes depends on the double-slit separation a, whereas their visibility depends on the separation l_s. As l_s is increased, the fringes disappear when equality in Eq. (38) is satisfied.

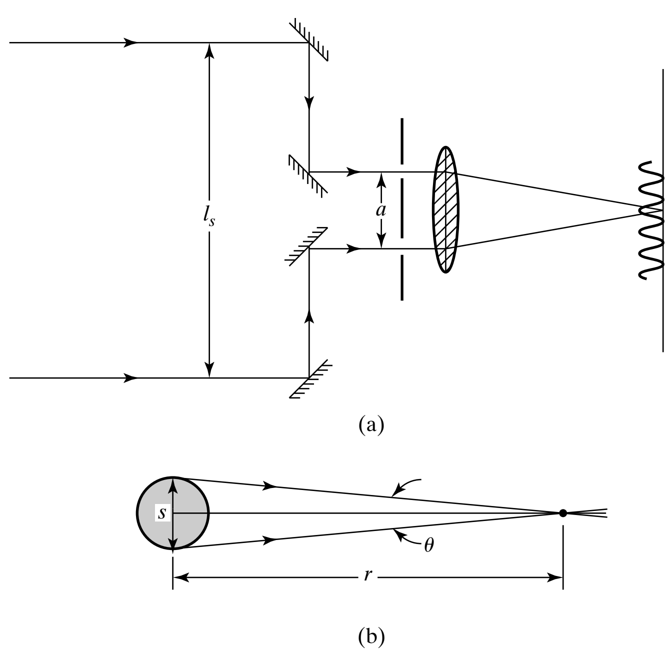


Figure 15 Michelson stellar interferometer (a) used to determine a stellar diameter (b).

Example 3

When Michelson used this technique on the star Betelgeuse in the constellation Orion, he found a first minimum in the fringes at $l_s = 308$ cm. Using an average wavelength of 570 nm, what is the angular diameter of the star?

Solution

Taking Eq. (38) as an equality,

$$\theta = \frac{1.22\lambda}{l_s} = \frac{1.22(570 \times 10^{-9})}{3.08} = 2.26 \times 10^{-7} \text{ rad}$$

Since Orion is known to be about 1×10^{15} mi away, the stellar diameter is $s = r\theta = 2.26 \times 10^8$ mi, or about 260 solar diameters.

PROBLEMS

1 Determine the Fourier series for the function of spatial period L given by

$$f(x) = \begin{cases} -1, & \dfrac{-L}{2} < x < 0 \\ +1, & 0 < x < \dfrac{+L}{2} \end{cases}$$

2 A half-wave rectifier removes the negative half-cycles of a sinusoidal waveform, given by $E = E_0 \cos \omega t$. Find the Fourier series of the resulting wave.

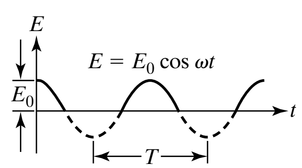


Figure 16 Problem 2.

3 Find the Fourier transform of the Gaussian function given by

$$f(t) = he^{-t^2/2\sigma^2}$$

where h is the height and σ the "width." (*Hint*: Remember how to complete a square? You will also need the definite integral

$$\int_{-\infty}^{+\infty} e^{-x^2}\, dx = \sqrt{\pi}$$

in your calculations.) Does the transform, interpreted as the frequency spectrum, show the proper relationship to the original "pulse" width?

4 Using the Fourier transform, determine the power spectrum of a single square pulse of amplitude A and duration τ_0. Sketch the power spectrum, locating its zeros, and show that the frequency bandwidth for the pulse is inversely proportional to its duration.

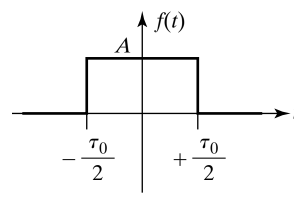


Figure 17 Problem 4.

5 Two light filters are used to transmit yellow light centered around a wavelength of 590 nm. One filter has a "broad" transmission width of 100 nm, whereas the other has a "narrow" pass band of 10 nm. Which filter would be better to use for an interference experiment? Compare the coherence lengths of the light from each.

6 A continuous He-Ne laser beam (632.8 nm) is "chopped," using a spinning aperture, into 1-μs pulses. Compute the resultant line width $\Delta\lambda$, bandwidth $\Delta\nu$, and coherence length.

7 The angular diameter of the sun viewed from the earth is approximately 0.5 degree. Determine the spatial coherence length for "good" coherence, neglecting any variations in brightness across the surface. Let us consider, somewhat arbitrarily, that "good" coherence will exist over an area that is 10% of the maximum area of coherence.

8 Michelson found that the cadmium red line (643.8 nm) was one of the most ideal monochromatic sources available, allowing fringes to be discerned up to a path difference of 30 cm in a beam-splitting interference experiment, such as with a Michelson interferometer. Calculate (a) the wavelength spread of the line and (b) the coherence time of the source.

9 A narrow band-pass filter transmits wavelengths in the range 5000 ± 0.5 Å. If this filter is placed in front of a source of white light, what is the coherence length of the transmitted light?

10 Let a collimated beam of white light fall on one refracting face of a prism and let the light emerging from the second face be focused by a lens onto a screen. Suppose that the linear dispersion at the screen is 20 Å/mm. By introducing a narrow "exit slit" in the screen, one has a type of monochromator that provides a nearly monochromatic beam of light. Sketch the setup. For an exit slit of 0.02 cm, what is the coherence time and coherence length of the light of mean wavelength 5000 Å?

11 A pinhole of diameter 0.5 mm is used in front of a sodium lamp (5890 Å) as a source in a Young interference experiment. The distance from pinhole to slits is 1 m. What is the maximum slit space insuring interference fringes that are just visible?

12 Determine the linewidth in angstroms and hertz for laser light whose coherence length is 10 km. The mean wavelength is 6328 Å.

13 **a.** A monochromator is used to obtain quasi-monochromatic light from a tungsten lamp. The linear dispersion of the instrument is 20 Å/mm and an exit slit of 200 μm is used. What is the coherence time and length of the light from the monochromator when set to give light of mean wavelength 500 nm?

b. This light is used to form fringes in an interference experiment in which the light is first amplitude-split into two equal parts and then brought together again. If the optical path difference between the two paths is 0.400 mm, calculate the magnitude of the normalized correlation function and the visibility of the resulting fringes.

c. If the maximum irradiance produced by the fringes is 100 on an arbitrary scale, what is the difference between maximum irradiance and background irradiance on this scale?

14 Determine the length and base area of the cylindrical volume within which light received from the sun is coherent. For this purpose, let us assume "good" spatial coherence occurs within a length that is 25% of the maximum value given by Eq. (38). The sun subtends an angle of 0.5° at the earth's surface. The mean value of the visible spectrum may be taken at 550 nm. Express the coherence volume also in terms of number of wavelengths across cylindrical length and diameter.

15 **a.** Show that the fringe visibility may be expressed by

$$V = \frac{2\sqrt{I_1 I_2}|\gamma(\tau)|}{(I_1 + I_2)}$$

b. What irradiance ratio of the interfering beams reduces the fringe visibility by 10% of that for equal-amplitude beams?

16 Show that the visibility of double-slit fringes in the mth order is given by

$$V = 1 - \left(m\frac{\Delta\lambda}{\lambda}\right)$$

where λ is the average wavelength of the light and $\Delta\lambda$ is its linewidth.

17 A filtered mercury lamp produces green light at 546.1 nm with a linewidth of 0.05 nm. The light illuminates a double slit of spacing 0.1 mm. Determine the visibility of the fringes on a screen 1 m away, in the vicinity of the fringe of order $m = 20$. (See problem 16.) If the discharge lamp is replaced with a white light source and a filter of bandwidth 10 nm at 546 nm, how does the visibility change?

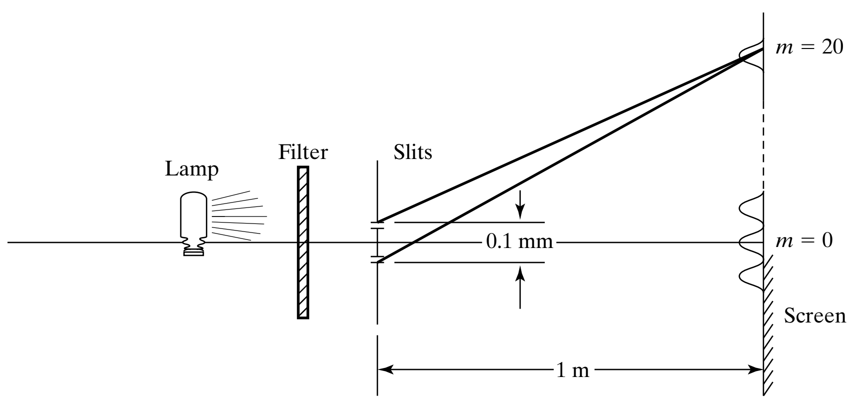
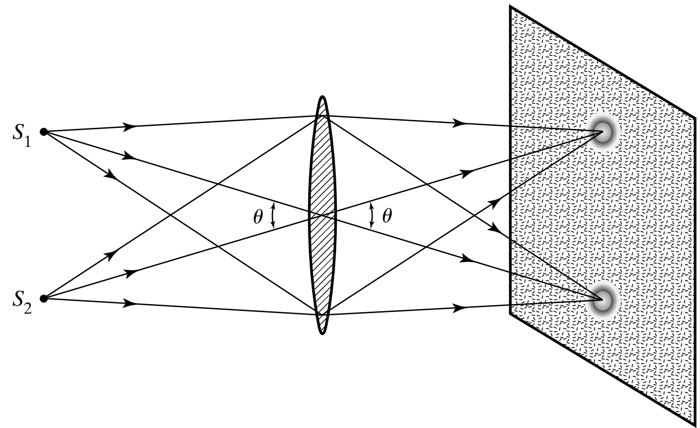


Figure 18 Problem 17.

18 A Michelson interferometer forms fringes with cadmium red light of 643.847 nm and linewidth of 0.0013 nm. What is the visibility of the fringes when one mirror is moved 1 cm from the position of zero path difference between arms? How does this change when the distance moved is 5 cm? At what distance does the visibility go to zero?

19 **a.** Repeat problem 18 when the light is the green mercury line of 546.1 nm with a linewidth of 0.025 nm.

b. How far can the mirror be moved from zero path difference so that fringe visibility is at least 0.85?



11 Fraunhofer Diffraction

INTRODUCTION

The wave character of light has been invoked to explain a number of phenomena, classified as "interference effects". In each case, two or more individual coherent beams of light, originating from a single source and separated by amplitude or wavefront division, were brought together again to interfere. Fundamentally, the same effect is involved in the *diffraction* of light. In its simplest description, diffraction is any deviation from geometrical optics that results from the obstruction of a wavefront of light. For example, an opaque screen with a round hole represents such an obstruction. On a viewing screen placed beyond the hole, the circle of light may show complex edge effects. This type of obstruction is typical in many optical instruments that utilize only the portion of a wavefront passing through a round lens. Any obstruction, however, shows detailed structure in its own shadow that is quite unexpected on the basis of geometrical optics.

Diffraction effects are a consequence of the wave character of light. Even if the obstacle is not opaque but causes local variations in the amplitude or phase of the wavefront of the transmitted light, such effects are observed. Tiny bubbles or imperfections in a glass lens, for example, produce undesirable diffraction patterns when transmitting laser light. Because the edges of optical images are blurred by diffraction, the phenomenon leads to a fundamental limitation in instrument resolution. More often, though, the sharpness of optical images is more seriously degraded by optical aberrations due to the imaging components themselves. *Diffraction-limited* optics is good optics indeed.

The double slit studied previously constitutes an obstruction to a wavefront in which light is blocked everywhere except at the two apertures. Recall that the irradiance of the resulting fringe pattern was calculated by treating the two openings as point sources, or long slits whose widths could be treated as points. A more complete analysis of this experiment must take into account the finite size of the slits. When this is done, the problem is treated as a diffraction problem. The results show that the interference pattern determined earlier is modified in a way that accounts for the actual details of the observed fringes.

Adequate agreement with experimental observations is possible through an application of the *Huygens-Fresnel principle*. According to Huygens, every point of a given wavefront of light can be considered a source of secondary spherical wavelets. To this, Fresnel added the assumption that the actual field at any point beyond the wavefront is a superposition of all these wavelets, taking into account both their amplitudes and phases. Thus, in calculating the diffraction pattern of the double slit at some point on a screen, one considers every point of the wavefront emerging from each slit as a source of wavelets whose superposition produces the resultant field. This procedure then takes into account a continuous array of sources across both slits, rather than two isolated point sources, as in the interference calculation. Diffraction is often distinguished from interference on this basis: In diffraction phenomena, the interfering beams originate from a continuous distribution of sources; in interference phenomena, the interfering beams originate from a discrete number of sources. This is not, however, a fundamental *physical* distinction.

A further classification of diffraction effects arises from the mathematical approximations possible when calculating the resultant fields. If both the source of light and observation screen are *effectively* far enough from the diffraction aperture so that wavefronts arriving at the aperture and observation screen may be considered plane, we speak of *Fraunhofer*, or *far-field, diffraction*, the type treated in this chapter. When this is not the case and the curvature of the wavefront must be taken into account, we speak of *Fresnel*, or *near-field, diffraction*. In the far-field approximation, as the viewing screen is moved relative to the aperture, the *size* of the diffraction pattern scales uniformly, but the *shape* of the diffraction pattern does not change. In the near-field approximation, the situation is more complicated. Both the shape and size of the diffraction pattern depend on the distance between the aperture and the screen. As the screen is moved away from the aperture, the image of the aperture passes through the forms predicted in turn by geometrical optics, near-field diffraction, and far-field diffraction.

It should be stated at the outset that the Huygens-Fresnel principle we shall employ to calculate diffraction patterns is itself an approximation. When no light penetrates an opaque screen, it means that the interaction of the incident radiation with the electronic oscillators, set into motion within the screen, is such as to produce zero net field beyond the screen. This balance is not maintained at the edge of an aperture in the screen, where the distribution of oscillators is interrupted. The Huygens-Fresnel principle does not include the contribution to the diffraction field of the electronic oscillators in the screen material at the edge of the aperture. Such edge effects are important, however, only when the observation point is very near the aperture itself.

1 DIFFRACTION FROM A SINGLE SLIT

We first calculate the Fraunhofer diffraction pattern from a *single slit*, a rectangular aperture characterized by a length much larger than its width. For Fraunhofer diffraction, the wavefronts of light reaching the slit must be essentially plane. In practice, this is easily accomplished by placing a source in the focal plane of a positive lens or by simply using a laser beam with a small divergence angle as the source. Similarly, we consider the observation screen to be effectively at infinity by using a lens on the exit side of the slit, as shown

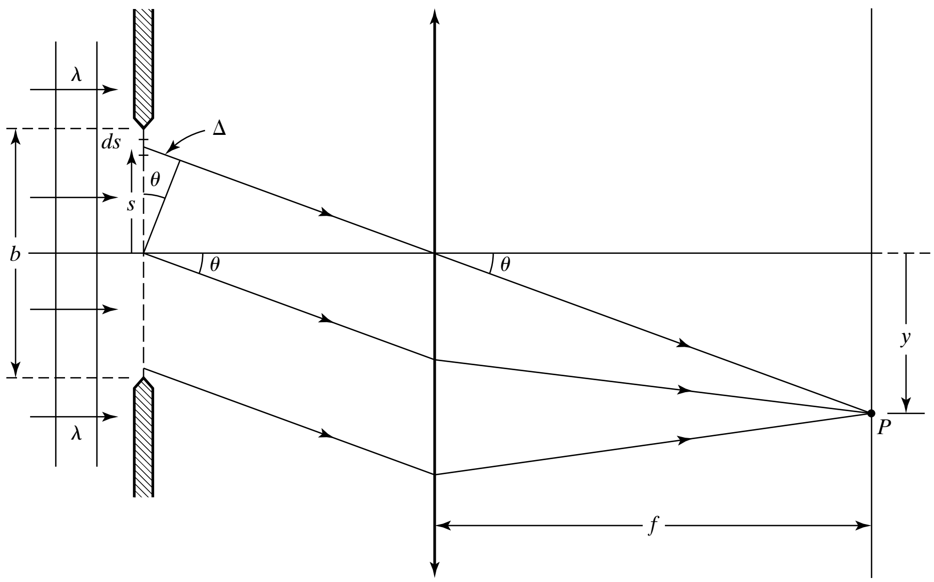


Figure 1 Construction for determining irradiance on a screen due to Fraunhofer diffraction by a single slit.

in Figure 1. Then the light reaching any point such as P on the screen is due to parallel rays of light from different portions of the wavefront at the slit (dashed line). According to the Huygens-Fresnel principle, we can consider spherical wavelets to be emanating from each point of the wavefront as it reaches the plane of the slit and then calculate the resultant field at P by adding the waves according to the principle of superposition. As shown in Figure 1, the waves do not arrive at P in phase. A ray from the center of the slit, for example, has an optical-path length that is an amount Δ shorter than one leaving from a point a vertical distance s above the optical axis.

The plane portion of the wavefront in the slit opening represents a continuous array of Huygens' wavelet sources. We consider each interval of length ds as a source and calculate the result of all such sources by integrating over the entire slit width b. Each interval contributes a spherical wavelet at P whose magnitude is directly proportional to the infinitesimal length ds. Thus,

$$dE_p = \left(\frac{E_L\,ds}{r}\right)e^{i(kr-\omega t)} \tag{1}$$

where r is the optical-path length from the interval ds to the point P. The amplitude $(E_L\,ds/r)$ has a $1/r$ dependence because spherical waves decrease in irradiance with distance, in accordance with the inverse square law. That is, for spherical waves the irradiance (which is proportional to the square of the electric field amplitude) is proportional to $1/r^2$ and so the electric field amplitude of a spherical wave is proportional to $1/r$. The proportionality constant E_L, here taken to be constant, determines the strength of the electric field contribution coming from each slit interval ds. Let us set $r = r_0$ for the wave from the center of the slit (at $s = 0$). Then, for any other wave originating at height s, taking the difference in phase into account, the differential field at P is

$$dE_p = \left(\frac{E_L\,ds}{r_0 + \Delta}\right)e^{i[k(r_0+\Delta)-\omega t]} = \left(\frac{E_L\,ds}{r_0 + \Delta}\right)e^{i(kr_0-\omega t)}e^{ik\Delta} \tag{2}$$

Note that the quantity $r_0 + \Delta$ appears both in the amplitude factor and in the phase factor. The path difference Δ is much smaller than r_0 and so (to lowest order) can be ignored in the amplitude factor. However, this path difference

Δ cannot be ignored in the phase factor. To understand why this is so, note that $k\Delta = (2\pi/\lambda)\,\Delta$. So as Δ varies by one wavelength, the phase $k\Delta$ varies over an entire cycle of range 2π. Figure 1 shows that $\Delta = s \sin\theta$. With these modifications, Eq. 2 can be rewritten as

$$dE_P = \left(\frac{E_L ds}{r_0}\right) e^{i(kr_0 - \omega t)} e^{iks\sin\theta} \tag{3}$$

The total electric field at the point P is found by integrating over the width of the slit. That is,

$$E_P = \int_{\text{slit}} dE_p = \frac{E_L}{r_0} e^{i(kr_0-\omega t)} \int_{-b/2}^{b/2} e^{iks\sin\theta}\, ds \tag{4}$$

Integration gives

$$E_P = \frac{E_L}{r_0} e^{i(kr_0-\omega t)} \left(\frac{e^{iks\sin\theta}}{ik\sin\theta}\right)_{-b/2}^{b/2} \tag{5}$$

Inserting the limits of integration into Eq. (5),

$$E_P = \frac{E_L}{r_0} e^{i(kr_0-\omega t)} \frac{e^{(ikb\sin\theta)/2} - e^{-(ikb\sin\theta)/2}}{ik\sin\theta} \tag{6}$$

The phases of the exponential terms suggest we make a convenient substitution,

$$\beta \equiv \tfrac{1}{2} kb \sin\theta \tag{7}$$

Then,

$$E_P = \frac{E_L}{r_0} e^{i(kr_0-\omega t)} \frac{b(e^{i\beta} - e^{-i\beta})}{2i\beta} = \frac{E_L}{r_0} e^{i(kr_0-\omega t)} \frac{b(2i\sin\beta)}{2i\beta} \tag{8}$$

where we have applied Euler's equation to obtain the last equality. Simplifying, we find

$$E_P = \frac{E_L b}{r_0} \frac{\sin\beta}{\beta} e^{i(kr_0-\omega t)} \tag{9}$$

Thus, the amplitude of the resultant field at P, given by Eq. (9), includes the *sinc* function $(\sin\beta)/\beta$, where β varies with θ and thus with the observation point P on the screen. We may give physical significance to β by interpreting it as a *phase difference*. Since a phase difference is given in general by $k\Delta$, Eq. (7) indicates a path difference associated with β of $\Delta = (b/2)\sin\theta$, shown in Figure 1. Thus $|\beta|$ represents the magnitude of the phase difference, at point P, between waves from the center and either endpoint of the slit, where $|s| = b/2$. In the analysis leading to Eq. (9), we assumed that the field strength E_L associated with each slit interval ds was a constant. If the field strength is not uniform across the slit, then the Fraunhofer diffraction pattern is the Fourier transform of the function that describes the field strength at various points within the aperture.

The irradiance I at P is proportional to the square of the resultant field amplitude there. The amplitude of the electric field given in Eq. (9) is

$$E_0 = \frac{E_L b}{r_0} \frac{\sin\beta}{\beta}$$

Thus, we find the irradiance I to be

$$I = \left(\frac{\varepsilon_0 c}{2}\right)E_0^2 = \frac{\varepsilon_0 c}{2}\left(\frac{E_L b}{r_0}\right)^2 \frac{\sin^2 \beta}{\beta^2}$$

or

$$I = I_0\left(\frac{\sin^2 \beta}{\beta^2}\right) \equiv I_0 \operatorname{sinc}^2(\beta) \tag{10}$$

where I_0 includes all constant factors. Equations (9) and (10) now permit us to plot the variation of irradiance with vertical displacement y from the symmetry axis at the screen. The sinc function has the property that it approaches 1 as its argument approaches 0:

$$\lim_{\beta \to 0} \operatorname{sinc}(\beta) = \lim_{\beta \to 0}\left(\frac{\sin \beta}{\beta}\right) = 1 \tag{11}$$

Otherwise, the zeros of $\operatorname{sinc}(\beta)$ occur when $\sin \beta = 0$, that is, when

$$\beta = \tfrac{1}{2}(kb \sin \theta) = m\pi \qquad m = \pm 1, \pm 2, \ldots$$

Equation (11) shows that the value $m = 0$ should not be included in this condition. The irradiance is plotted as a function of β in Figure 2. Setting $k = 2\pi/\lambda$, the condition for zeros of the sinc function (and so of the irradiance) is

$$m\lambda = b \sin \theta \qquad m = \pm 1, \pm 2, \ldots \tag{12}$$

Referring to Figure 1, note that the distance y from the center of the screen to a point on the screen P located by the angle θ is given approximately by $y \cong f \sin \theta$, where we have made the small angle approximation $\sin \theta \cong \tan \theta$. On the screen, therefore, in accordance with Eqs. (11) and (12), the irradiance is a maximum at $\theta = 0$ ($y = 0$) and drops to zero at values y_m such that

$$y_m \cong \frac{m\lambda f}{b} \tag{13}$$

The irradiance pattern is symmetrical about $y = 0$.

The secondary maxima of the single-slit diffraction pattern do not quite fall at the midpoints between zeros, even though this condition is more nearly

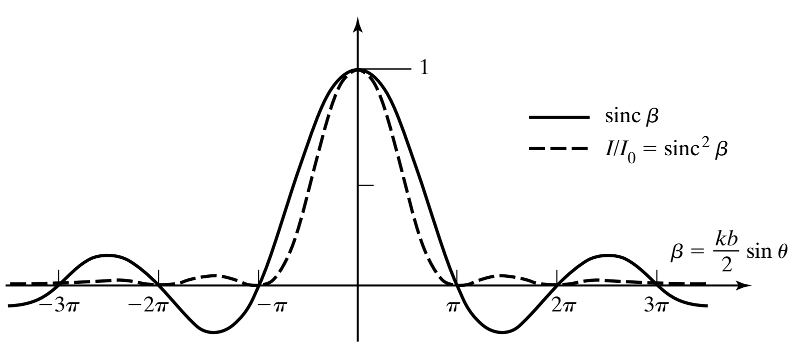


Figure 2 Sinc function (solid line) plotted as a function of β. The normalized irradiance function I/I_0 (dashed line) for single-slit Fraunhofer diffraction is the square of $\operatorname{sinc}(\beta)$.

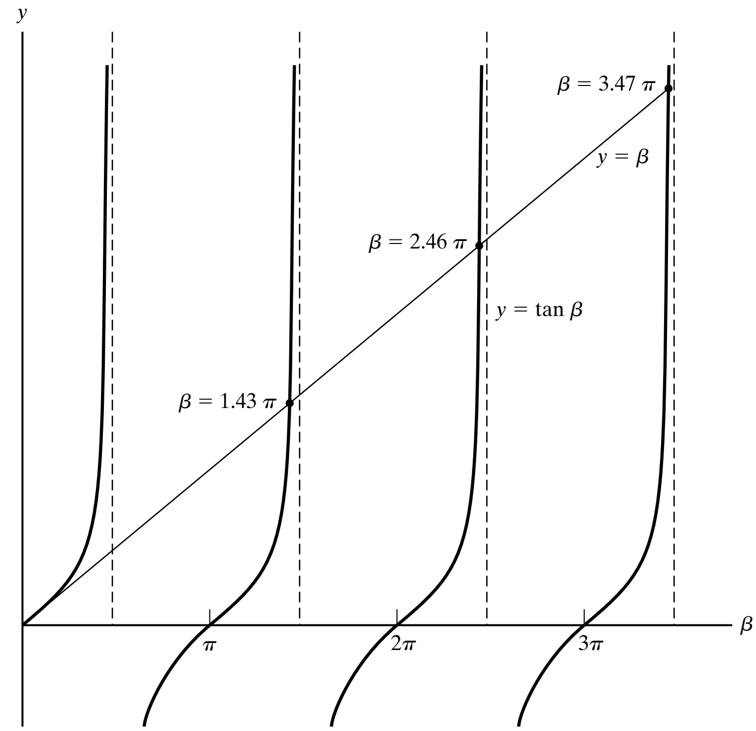


Figure 3 Intersections of the curves $y = \beta$ and $y = \tan\beta$ determine the angles β at which the sinc function is a maximum.

approached as β increases. The maxima coincide with maxima of the sinc function, which occur at points satisfying

$$\frac{d}{d\beta}\left(\frac{\sin\beta}{\beta}\right) = \frac{\beta\cos\beta - \sin\beta}{\beta^2} = 0$$

or $\beta = \tan\beta$. An angle equals its tangent at intersections of the curves $y = \beta$ and $y = \tan\beta$, both plotted in Figure 3. Intersections, excluding $\beta = 0$, occur at 1.43π (rather than 1.5π), 2.46π (rather than 2.5π), 3.47π (rather than 3.5π), and so on. The plot clearly shows that intersection points approach the vertical lines defining midpoints more closely as β increases. Thus, in the irradiance plot of Figure 2, secondary maxima are skewed slightly away from the midpoints toward the central peak. Most of the energy of the diffraction pattern falls under the central maximum, which is much larger than the adjoining maximum on either side.

Example 1

What is the ratio of irradiances at the central peak maximum to the first of the secondary maxima?

Solution

The ratio to be calculated is

$$\frac{I_{\beta=0}}{I_{\beta=1.43\pi}} = \frac{(\sin^2\beta/\beta^2)_{\beta=0}}{(\sin^2\beta/\beta^2)_{\beta=1.43\pi}} = \frac{1}{(\sin^2\beta/\beta^2)_{\beta=1.43\pi}}$$

$$= \left(\frac{\beta^2}{\sin^2\beta}\right)_{1.43\pi} = \frac{20.18}{0.952} = 21.2$$

Thus the maximum irradiance of the nearest secondary peak is only 4.7% that of the central peak.

The central maximum represents essentially the image of the slit on a distant screen. We observe that the edges of the image are not sharp but reveal a series of maxima and minima that tail off into the shadow surrounding the image. These effects are typical of the blurring of images due to diffraction and will be seen again in other cases of diffraction to be considered. The angular width of the central maximum is defined as the angle $\Delta\theta$ between the first minima on either side. Using Eq. (12) with $m = \pm 1$ and approximating $\sin\theta$ by θ, we get

$$\Delta\theta = \frac{2\lambda}{b} \tag{14}$$

From Eq. (14), it follows that the central maximum will spread as the slit width is narrowed. Since the length of the slit is very large compared to its width, the diffraction pattern due to points of the wavefront along the length of the slit has a very small angular width and is not prominent on the screen. Of course, the dimensions of the diffraction pattern also depend on the wavelength, as indicated in Eq. (14).

2 BEAM SPREADING

According to Eq. (14), the angular spread $\Delta\theta$ of the central maximum in the far field is independent of distance between aperture and screen. The linear dimensions of the diffraction pattern thus increase uniformly with distance L, as shown in Figure 4, such that the width W of the central maximum is given by

$$W = L\,\Delta\theta = \frac{2L\lambda}{b} \tag{15}$$

We may describe the content of Eq. (15) as a linear spread of a beam of light, originally constricted to a width b. Indeed, the means by which the beam is originally narrowed is not relevant to the nature of the diffraction pattern that occurs. If one dispenses with the slit in Figure 4 and merely assumes an original beam of constant irradiance across a finite width b, all our results follow in the same way. After collimation, a "parallel" beam of light spreads just as if it emerged from a single opening.

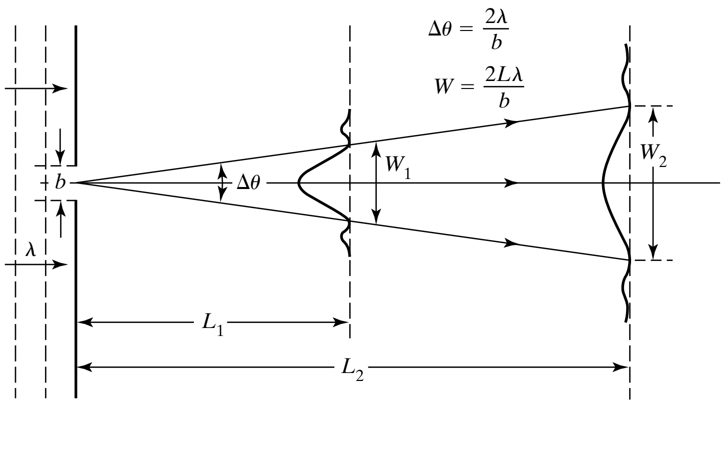


Figure 4 Spread of the central maximum in the far-field diffraction pattern of a single slit.

Example 2

Imagine a parallel beam of 546-nm light of width $b = 0.5$ mm propagating a distance of 10 m across the laboratory. Estimate the final width W of the beam due to diffraction spreading.

Solution

Using Eq. (15),

$$W = \frac{2L\lambda}{b} = \frac{2(10\text{ m})(546 \times 10^{-9}\text{ m})}{0.5 \times 10^{-3}\text{ m}} = 0.0218\text{ m} = 21.8\text{ mm}$$

Even highly collimated laser beams are subject to beam spreading as they propagate, due to diffraction. It is a fundamental consequence of the wave nature of light that beams of finite transverse extent must spread as they propagate.

The beam spreading described by Eq. (14) is valid for a rectangular aperture of width much less than its length. As we show in the next section, the spreading due to diffraction from a circular aperture follows a form similar to Eq. (14) but with the replacement of the width b of the slit by the diameter D of the circular aperture and with the replacement of the wavelength λ by the factor 1.22λ. Furthermore, one must keep in mind that this treatment assumes a plane wavefront of uniform irradiance.[1] The spreading described by Eq. (15) has been deduced on the basis of Fraunhofer, or far-field, diffraction, which means here that L must remain reasonably large. If L is taken small enough, for example, the equation predicts a beam width less than b, contrary to assumption. Evidently L must be larger than some minimum value, $L_{\min}$, which gives a beam width $W = b$, that is,

$$L_{\min} = \frac{b^2}{2\lambda}$$

We may conclude that we are in the *far field* when

$$L \gg \frac{b^2}{\lambda}$$

A more general approach leads to the commonly stated criterion for *far-field diffraction* in the form[2]

$$L \gg \frac{\text{area of aperture}}{\lambda} \tag{16}$$

3 RECTANGULAR AND CIRCULAR APERTURES

We have been describing diffraction from a slit having a width b much smaller than its length a, as illustrated in Figure 5a. When both dimensions of

[1]A laser beam usually does not have constant irradiance across its diameter. In its fundamental mode, the transverse profile is a Gaussian function. Still, its spread formula is essentially that of Eq. (14) with the beam diameter replacing b and the constant factor of 2 replaced by $4/\pi \cong 1.27$. In comparing formulas for divergence angles, care must be taken to distinguish between the *full angular spread* illustrated in Figure 4 and the *half-angle spread*.

[2]Many practitioners in the field of high-energy lasers use the far-field criterion, $L > 100$ (area of aperture)$/\lambda$.

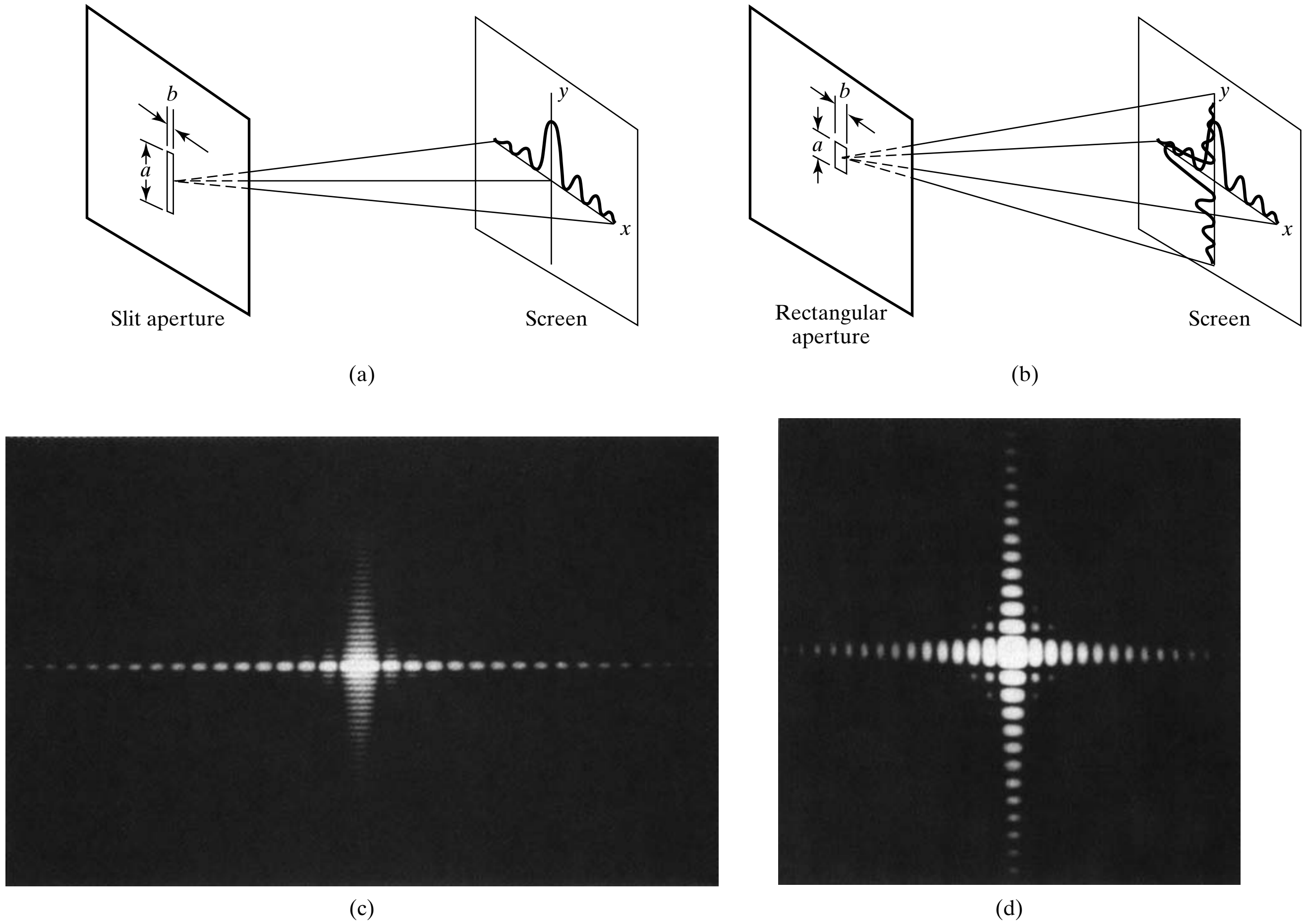


Figure 5 (a) Single-slit diffraction. Only the small dimension b of a long, narrow slit causes appreciable spreading of the light along the x-direction on the screen. (b) Rectangular aperture diffraction. Both dimensions of the rectangular aperture are small and a two-dimensional diffraction pattern is discernible on the screen. (c) Photograph of the diffraction image of a rectangular aperture with $b < a$, as in the representation of Figure 5a. (d) Photograph of the diffraction image of a rectangular aperture with $b = a$, as in the representation of Figure 5b. (Both photos are from M. Cagnet, M. Francon, and J. C. Thrierr, *Atlas of Optical Phenomenon*, Plate 17, Berlin: Springer-Verlag, 1962.)

the slit are comparable and small, each produces appreciable spreading, as illustrated in Figure 5b. For the aperture dimension a, we write analogously, for the irradiance, as in Eq. (10),

$$I = I_0\left(\frac{\sin \alpha}{\alpha}\right)^2 \quad \text{where } \alpha \equiv \left(\frac{k}{2}\right)a \sin \theta \tag{17}$$

The two-dimensional pattern now gives zero irradiance for points x, y satisfied by either

$$y_m = \frac{m\lambda f}{b} \quad \text{or} \quad x_n = \frac{n\lambda f}{a}$$

where both m and n represent nonzero integral values. The irradiance over the screen turns out to be just a product of the irradiance functions in each dimension, or

$$I = I_0(\text{sinc}^2\, \beta)(\text{sinc}^2\, \alpha) \tag{18}$$

In calculating this result, the single integration over one dimension of the slit is replaced by a double integration over both dimensions of the aperture. Photographs of single-aperture diffraction patterns for rectangular and square apertures are shown in Figure 5c and d.

When the aperture is circular, the integration is over the entire area of the aperture since both vertical and horizontal dimensions of the aperture are comparable. Equation (4), which describes the total electric field at point P of Figure 1 due to single-slit diffraction, can be modified to describe diffraction from a circular aperture. The required modification involves the replacement of the incremental *electric field amplitude* $E_L\, ds/r_0$ by $E_A\, dA/r_0$ and the conversion of the integral over the slit width to an integral over the aperture area. Here, E_A is a constant factor (with "units" of electric field per unit length) that determines the strength of the electric field in the aperture and dA is the elemental area of the aperture. The electric field at P (as in Figure 1) due to diffraction through a circular aperture can then be written as

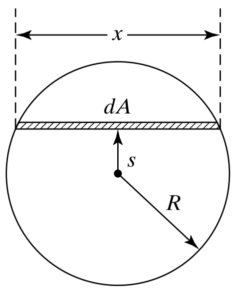


Figure 6 Geometry used in the integration over a circular aperture.

$$E_p = \frac{E_A}{r_0} e^{i(kr_0 - \omega t)} \iint_{\text{Area}} e^{isk \sin\theta}\, dA$$

We take a rectangular strip of area $dA = x\, ds$ as the elemental area of integration, shown in Figure 6. Using the equation of a circle, we calculate the length x at height s to be given by

$$x = 2\sqrt{R^2 - s^2}$$

where R is the aperture radius. The preceding integral can then be rewritten, leading to

$$E_P = \frac{2E_A}{r_0} e^{i(kr_0 - \omega t)} \int_{-R}^{R} e^{isk \sin\theta} \sqrt{R^2 - s^2}\, ds$$

The integral takes the form of a standard definite integral upon making the substitutions $v = s/R$ and $\gamma = kR \sin\theta$:

$$E_P = \frac{2E_A R^2}{r_0} e^{i(kr_0 - \omega t)} \int_{-1}^{+1} e^{i\gamma v} \sqrt{1 - v^2}\, dv$$

The integral has the value

$$\int_{-1}^{+1} e^{i\gamma v} \sqrt{1 - v^2}\, dv = \frac{\pi J_1(\gamma)}{\gamma}$$

where $J_1(\gamma)$ is the first-order *Bessel function of the first kind*, expressible by the infinite series

$$J_1(\gamma) = \frac{\gamma}{2} - \frac{(\gamma/2)^3}{1^2 \cdot 2} + \frac{(\gamma/2)^5}{1^2 \cdot 2^2 \cdot 3} - \cdots$$

As can be verified from this series expansion, the ratio $J_1(\gamma)/\gamma$ has the limit $\frac{1}{2}$ as $\gamma \rightarrow 0$. Thus, the circular aperture requires, instead of the sine function for the single slit, the Bessel function J_1, which oscillates somewhat like the sine function, as shown in the plot of Figure 7. One important difference is that the amplitude of the oscillation of the Bessel function decreases as its argument departs from zero.

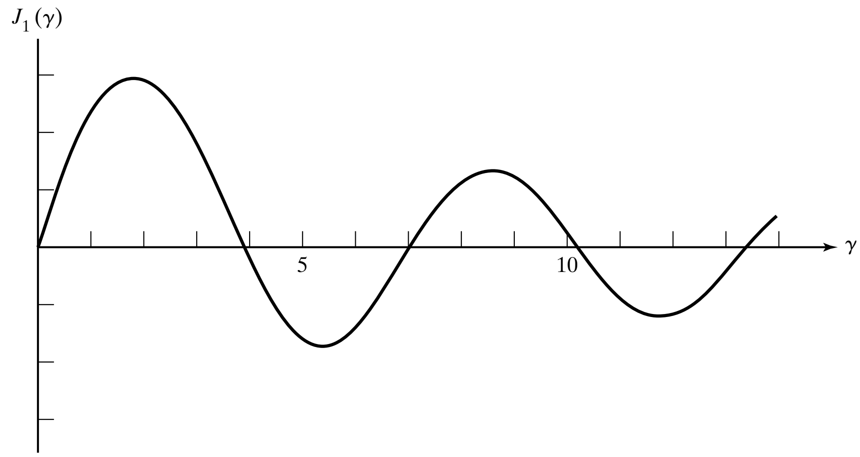


Figure 7 A plot of the Bessel function $J_1(\gamma)$ vs. γ. The first few zeroes of the Bessel function occur at $\gamma = 0$, $\gamma = 3.832$, $\gamma = 7.016$, $\gamma = 10.173$, and $\gamma = 13.324$.

The irradiance for a circular aperture of diameter D can now be written as

$$I = I_0\left(\frac{2J_1(\gamma)}{\gamma}\right)^2, \quad \text{where } \gamma \equiv \tfrac{1}{2}kD\sin\theta \tag{19}$$

where I_0 is the irradiance at $\gamma \to 0$ or at $\theta = 0$. The equations should be compared with those of Eq. (17) to appreciate the analogous role played by the Bessel function. Like $(\sin x)/x$, the function $J_1(x)/x$ approaches a maximum as x approaches zero, so that the irradiance is greatest at the center of the pattern ($\theta = 0$). (In fact, $J_1(x)/x$ tends to $1/2$ as x tends to zero, so the irradiance tends to I_0 as γ tends to zero.) The pattern is symmetrical about the optical axis through the center of the circular aperture and has its first zero when $\gamma = 3.832$, as indicated in Figure 8a and b. Thus, the irradiance first falls to zero when

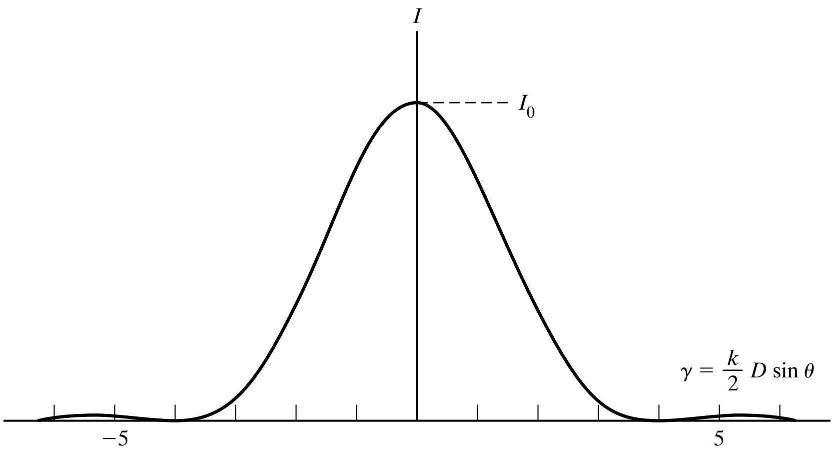
$$\gamma = \left(\frac{k}{2}\right)D\sin\theta = 3.832 \quad \text{or} \quad \text{when } D\sin\theta = 1.22\lambda \tag{20}$$

The irradiance pattern of Eq. (19) is plotted in Figure 8a. The first few zeroes, and maxima of the normalized irradiance $I/I_0 = (2J_1(\gamma)/\gamma)^2$ are listed in Figure 8b. The pattern is similar to that of Figure 2 for a slit, except that the pattern for a circular aperture has rotational symmetry about the optical axis. A photograph is shown in Figure 8c. The central maximum is a circle of light, the diffracted "image" of the circular aperture, and is called the *Airy disc*. Equation (20) should be compared with the analogous equation for the narrow rectangular slit, $m\lambda = b\sin\theta$. We see that $m = 1$ for the first minimum in the slit pattern is replaced by the number 1.22 in the case of the circular aperture. Successive minima are determined in a similar way from other zeros of the Bessel function, as indicated in the table in Figure 8b.

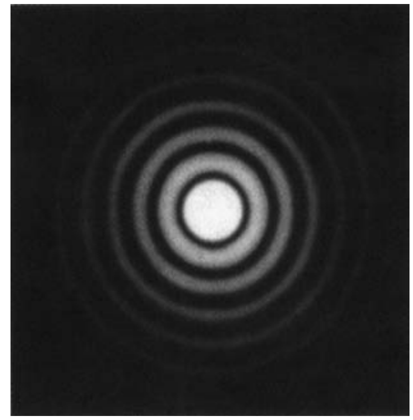
Note that the far-field angular *radius* (i.e., the angular half-width) of the Airy disc, according to Eq. (20), is very nearly

$$\Delta\theta_{1/2} = \frac{1.22\lambda}{D} \tag{21}$$

In Example 3, the beam spread from a circular aperture is compared with that from a single slit.



(a)


(c)

	γ	$I/I_0 = (2J_1(\gamma)/\gamma)^2$
1st Maximum	0	1
1st Zero	3.832	0
2nd Maximum	5.136	0.0175
2nd Zero	7.016	0
3rd Maximum	8.417	0.00416
3rd Zero	10.173	0
4th Maximum	11.620	0.00160
4th Zero	13.324	0

(b)

Figure 8 Circular aperture diffraction pattern. (a) Irradiance $I = I_0(2J_1(\gamma)/\gamma)^2$ of the diffraction pattern of a circular aperture. By far the largest amount of light energy is diffracted into the central maximum. (b) The first few zeroes and maxima of the normalized irradiance $I/I_0 = (2J_1(\gamma)/\gamma)^2$. (c) Diffraction image of a circular aperture. The circle of light at the center corresponds to the zeroth order of diffraction and is known as the Airy disc. (From M. Cagnet, M. Francon, and J. C. Thrierr, *Atlas of Optical Phenomenon*, Plate 16, Berlin: Springer-Verlag, 1962.)

Example 3

Find the diameter of the Airy disc at the center of the diffraction pattern formed on a wall at a distance $L = 10$ m from a uniformly illuminated circular aperture of diameter $D = 0.5$ mm. Assume that the illuminating light has wavelength of $\lambda = 546$ nm. Compare the beam spread to that from the slit of width $b = 0.5$ mm of Example 2.

Solution

The angular radius of the Airy disc is found using Eq. 21,

$$\Delta\theta_{1/2} = \frac{1.22\lambda}{D} = \frac{1.22(546 \times 10^{-9}\text{ m})}{5 \times 10^{-4}\text{ m}} = 1.33 \times 10^{-3}\text{ rad}$$

The radius r_d of the Airy disc is then found using an argument similar to that used in Figure 4 for single-slit diffraction,

$$r_\text{d} = L\,\Delta\theta_{1/2} = (10\text{ m})(1.33 \times 10^{-3}) = 0.013\text{ m} = 13\text{ mm}$$

The diameter D_d of the Airy disc is, then,

$$D_d = 2r_d = 26 \text{ mm}$$

The beam spread is comparable to, but slightly more than, that from the single slit of Example 2, where W was found to be near 22 mm.

4 RESOLUTION

In forming the Fraunhofer diffraction pattern of a single slit, as in Figure 1, we notice that the distance between slit and lens is not crucial to the details of the pattern. The lens merely intercepts a larger solid angle of light when the distance is small. If this distance is allowed to go to zero, aperture and lens coincide, as in the objective of a telescope. Thus, the image formed by a telescope with a round objective is subject to the diffraction effects described by Eq. (19) for a circular aperture. The sharpness of the image of a distant point object—a star, for example—is, then, limited by diffraction. The image occupies essentially the region of the Airy disc. An eyepiece viewing the primary image and providing further magnification merely enlarges the details of the diffraction pattern formed by the lens. The limit of resolution is already set in the primary image. The inevitable blur that diffraction produces in the image restricts the resolution of the instrument, that is, its ability to provide distinct images for distinct object points, either physically close together (as in a microscope) or separated by a small angle at the lens (as in a telescope). Figure 9a illustrates the diffraction of two point objects S_1 and S_2 formed

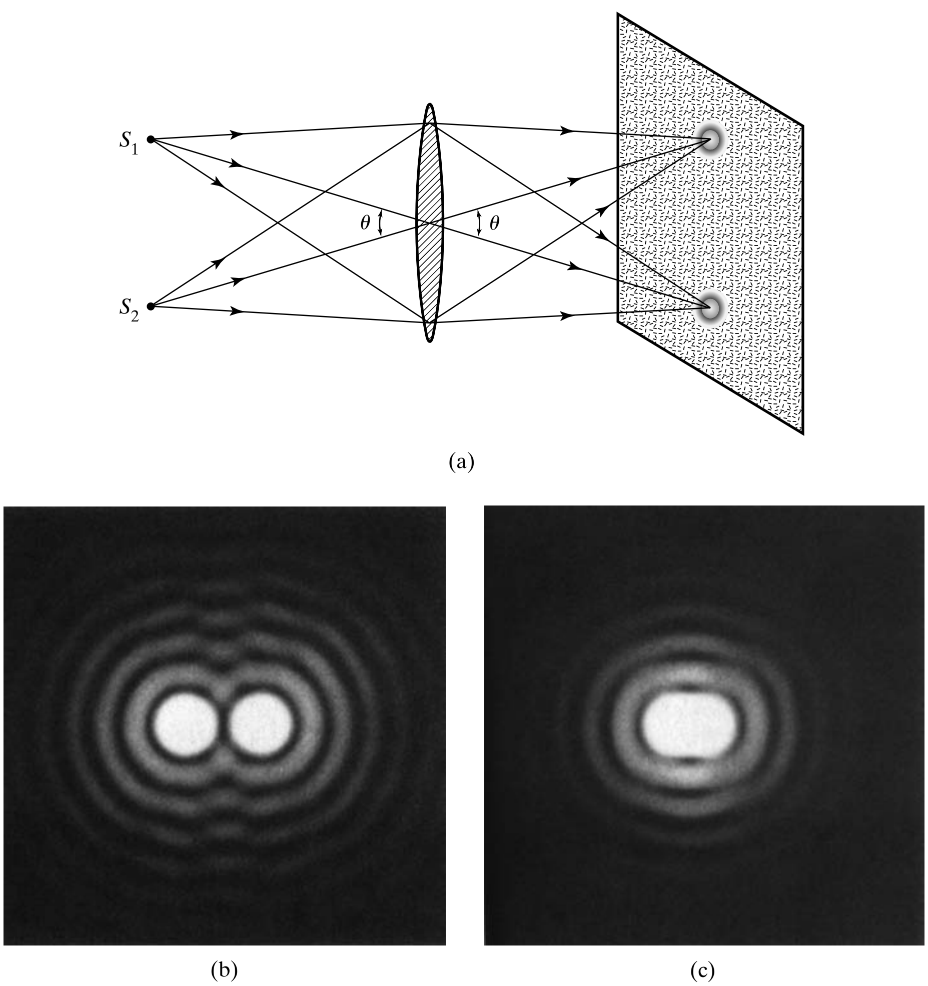


Figure 9 (a) Diffraction-limited images of two point objects formed by a lens. As long as the Airy discs are well separated, the images are well resolved. (b) Separated images of two incoherent point sources. In this diffraction pattern, the two images are well resolved. (c) Image of a pair of incoherent point sources at the limit of resolution. (Reproduced by permission from "Atlas of Optical Phenomena", 1962, Michael Cagnet, Maurice Franco and Jean Claude Thrierr; Plate 12. Copyright © Springer-Verlag GmbH & Co KG. With Kind Permission of Springer Science and Business Media.)

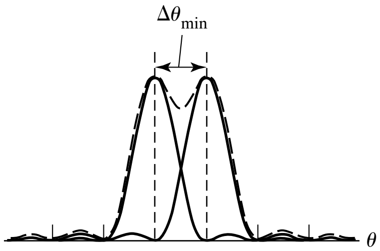


Figure 10 Rayleigh's criterion for just-resolvable diffraction patterns. The dashed curve is the observed sum of independent diffraction peaks.

by a single lens. The point objects and the centers of their Airy discs are both separated by the angle θ. If the angle is large enough, two distinct images will be clearly seen, as shown in the photograph of Figure 9b. Imagine now that the objects S_1 and S_2 are brought closer together. When their image patterns begin to overlap substantially, it becomes more difficult to discern the patterns as distinct, that is, to resolve them as belonging to distinct object points. A photograph of the two images at the limit of resolution is shown in Figure 9c.

Rayleigh's criterion for just-resolvable images—a somewhat arbitrary but useful criterion—requires that the angular separation of the centers of the image patterns be no less than the angular radius of the Airy disc, as in Figure 10. In this condition, the maximum of one pattern falls directly over the first minimum of the other. Thus, for the *limit of resolution*, we have, using Eq. (21),

$$(\Delta\theta)_{min} = \frac{1.22\lambda}{D} \tag{22}$$

where D is now the diameter of the lens. In accordance with this result, the minimum resolvable angular separation of two object points may be reduced (the resolution improved) by increasing the lens diameter and decreasing the wavelength.

We consider several applications of Eq. (22), beginning with the following example.

Example 4

Suppose that each lens on a pair of binoculars has a diameter of 35 mm. How far apart must two stars be before they are theoretically resolvable by either of the lenses in the binoculars?

Solution

According to Eq. (22),

$$(\Delta\theta)_{min} = \frac{1.22(550 \times 10^{-9})}{35 \times 10^{-3}} = 1.92 \times 10^{-5} \text{ rad}$$

or about 4″ of arc, using an average wavelength for visible light. If the stars are near the center of our galaxy, a distance, d, of around 30,000 light-years, then their actual separation s is approximately

$$s = d\,\Delta\theta_{min} = (30{,}000)(1.92 \times 10^{-5}) = 0.58 \text{ light-years}$$

To get some appreciation for this distance, consider that the planet Pluto at the edge of our solar system is only about 5.5 light-*hours* distant. If the stars are being detected by their long-wavelength radio waves—the lenses being replaced by dish antennas—the resolution must, by Eq. (22), be much less.

If the lens is the objective of a microscope, as indicated in Figure 11, the problem of resolving nearby objects is basically the same. Making only rough estimates, we shall ignore the fact that the wavefronts striking the lens from nearby object points A and B are not plane, as required in far-field

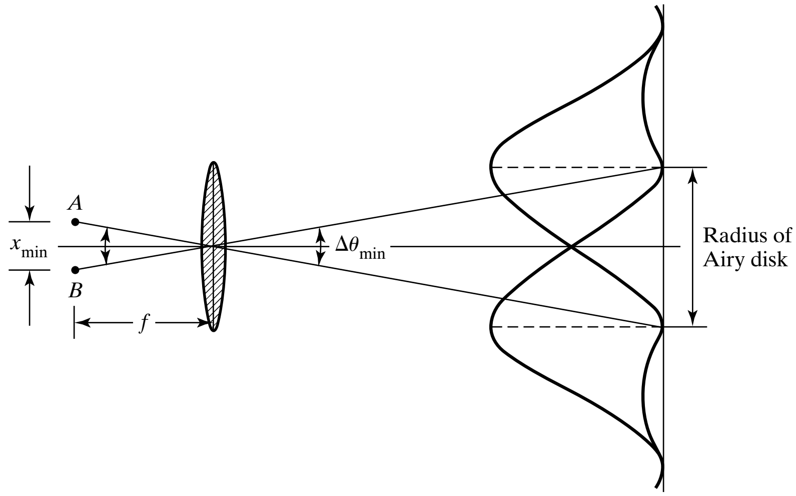


Figure 11 Minimum angular resolution of a microscope.

diffraction equations. The minimum separation, x_{min}, of two just-resolved objects near the focal plane of the lens of diameter D is then given by

$$x_{min} = f\,\Delta\theta_{min} = f\left(\frac{1.22\lambda}{D}\right)$$

The ratio *D/f* is the *numerical aperture*, with a typical value of 1.2 for a good oil-immersion objective. Thus,

$$x_{min} \cong \lambda$$

The resolution of a microscope is roughly equal to the wavelength of light used, a fact that explains the advantage of ultraviolet, X-ray, and electron microscopes in high-resolution applications. Know that some techniques used in *near-field microscopy* allow one to surpass the diffraction-limited resolution just discussed.

The limits of resolution due to diffraction also affect the human eye, which may be approximated by a circular aperture (pupil), a lens, and a screen (retina), as in Figure 12. Night vision, which takes place with large, adapted pupils of around 8 mm, is capable of higher resolution than daylight vision. Unfortunately, there is not enough light to take advantage of the situation! On a bright day the pupil diameter may be 2 mm. Under these conditions, Eq. (22) gives $(\Delta\theta)_{min} = 33.6 \times 10^{-5}$ rad, for an average wavelength of 550 nm. Experimentally, one finds that a separation of 1 mm at a distance of about 2 m is just barely resolvable, giving $(\Delta\theta)_{min} = 50 \times 10^{-5}$ rad, about 1.5 times the theoretical limit. One's own resolution (*visual acuity*) can easily be tested by viewing two lines drawn 1 mm apart at increasing distances until they can no longer be seen as distinct. It is interesting to note that the theoretical resolution just determined for a 2-mm-diameter pupil is consistent with the value of 1′ of arc (29×10^{-5} rad) used by Snellen to characterize normal visual acuity.

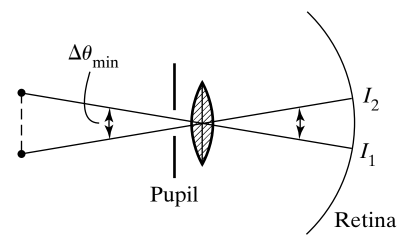


Figure 12 Diffraction by the eye with pupil as aperture limits the resolution of objects subtending angle $\Delta\theta_{min}$.

5 DOUBLE-SLIT DIFFRACTION

The diffraction pattern of a plane wavefront that is obstructed everywhere except at two narrow slits is calculated in the same manner as for the single slit. The mathematical argument departs from that for the single slit with Eq. (4). Here, the limits of integration covering the apertures of the two slits become those indicated in Figure 13.

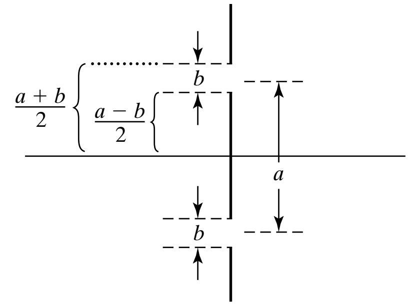


Figure 13 Specification of slit width and separation for double-slit diffraction.

We find

$$E_P = \frac{E_L}{r_0} e^{i(kr_0 - \omega t)} \int_{-(1/2)(a+b)}^{-(1/2)(a-b)} e^{isk \sin \theta}\, ds + \frac{E_L}{r_0} e^{i(kr_0 - \omega t)} \int_{(1/2)(a-b)}^{(1/2)(a+b)} e^{isk \sin \theta}\, ds \tag{23}$$

Integration and substitution of the limits leads to

$$E_P = \frac{E_L}{r_0} e^{i(kr_0 - \omega t)} \frac{1}{ik \sin \theta} [e^{(1/2)ik(-a+b) \sin \theta} - e^{(1/2)ik(-a-b) \sin \theta} + e^{(1/2)ik(a+b) \sin \theta} - e^{(1/2)ik(a-b) \sin \theta}]$$

Reintroducing the substitution of Eq. (7), involving the slit width b,

$$\beta \equiv \tfrac{1}{2} kb \sin \theta \tag{24}$$

and a similar one involving the slit separation a,

$$\alpha \equiv \tfrac{1}{2} ka \sin \theta \tag{25}$$

our equation is written more compactly as

$$E_P = \frac{E_L}{r_0} e^{i(kr_0 - \omega t)} \frac{b}{2i\beta} [e^{i\alpha}(e^{i\beta} - e^{-i\beta}) + e^{-i\alpha}(e^{i\beta} - e^{-i\beta})]$$

Employing Euler's equation,

$$E_P = \frac{E_L}{r_0} e^{i(kr_0 - \omega t)} \frac{b}{2i\beta} (2i \sin \beta)(2 \cos \alpha)$$

Finally,

$$E_P = \frac{E_L}{r_0} e^{i(kr_0 - \omega t)} \frac{2b \sin \beta}{\beta} \cos \alpha \tag{26}$$

The amplitude of this electric field is

$$E_0 = \frac{E_L}{r_0} \frac{2b \sin \beta}{\beta} \cos \alpha$$

so that the irradiance at point P in the double-slit diffraction pattern is

$$I = \left(\frac{\varepsilon_0 c}{2}\right) E_0^2 = \left(\frac{\varepsilon_0 c}{2}\right) \left(\frac{2E_L b}{r_0}\right)^2 \left(\frac{\sin \beta}{\beta}\right)^2 \cos^2 \alpha$$

or

$$I = 4I_0 \left(\frac{\sin \beta}{\beta}\right)^2 \cos^2 \alpha \tag{27}$$

where

$$I_0 = \left(\frac{\varepsilon_0 c}{2}\right) \left(\frac{E_L b}{r_0}\right)^2$$

as defined in Eq. (10) for the single slit. Since the maximum value of Eq. (27) is $4I_0$, we see that the double slit provides four times the maximum irradiance in the pattern center as compared with the single slit. This is exactly what should be expected where the two beams are in phase and amplitudes add.

On closer inspection of Eq. (27), we find that the irradiance is just a product of the irradiances found for double-slit interference and single-slit diffraction. The factor $[(\sin\beta)/\beta]^2$ is that of Eq. (10) for single-slit diffraction. The $\cos^2\alpha$ factor, when α is written out as in Eq. (25), is

$$\cos^2\alpha = \cos^2\left[\frac{ka(\sin\theta)}{2}\right] = \cos^2\left[\frac{\pi a(\sin\theta)}{\lambda}\right]$$

The sinc and cosine factors of Eq. (27) are plotted in Figure 14a for the case $a = 6b$ or $\alpha = 6\beta$. Because $a > b$, the $\cos^2\alpha$ factor varies more rapidly than the $(\sin^2\beta)/\beta^2$ factor. The product of the sine and cosine factors may be considered a modulation of the interference fringe pattern by a single-slit diffraction envelope, as shown in Figure 14b. The diffraction envelope has a minimum

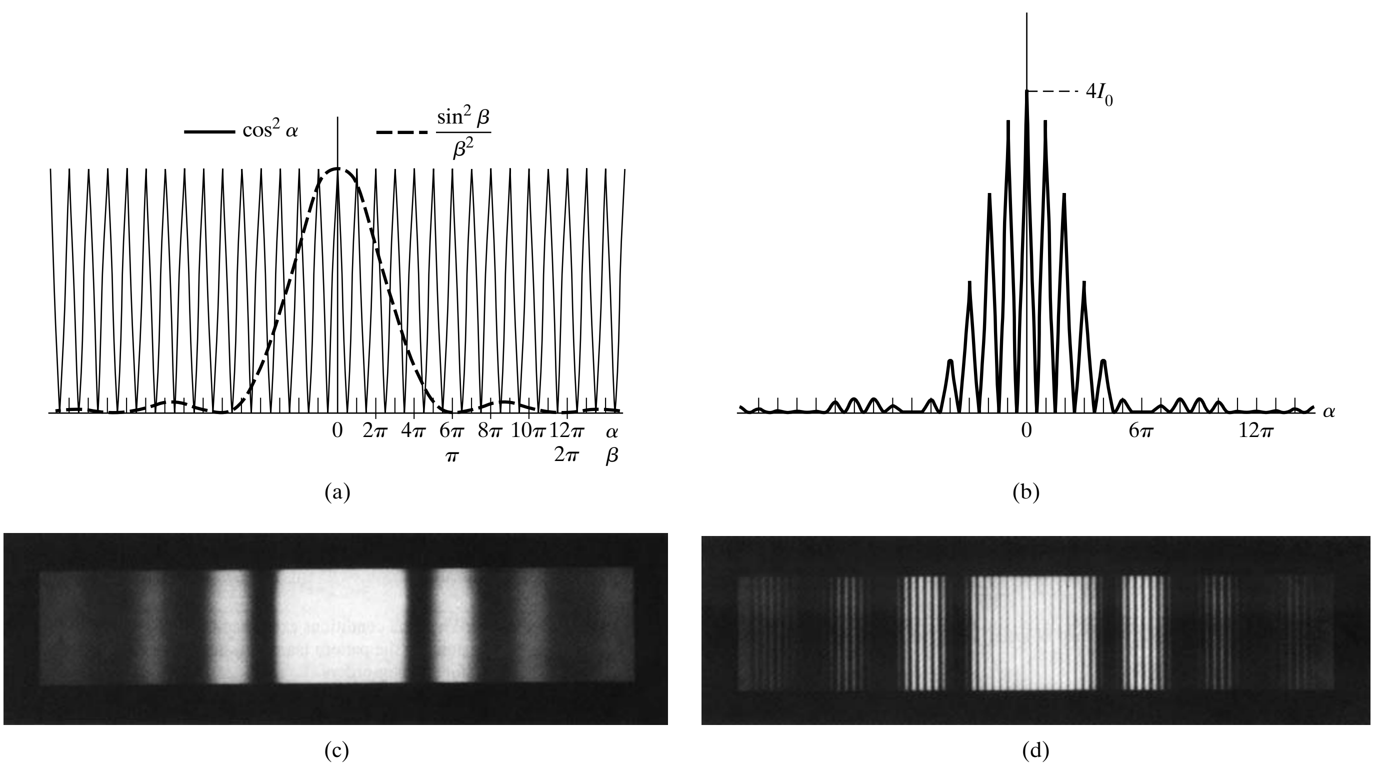


Figure 14 (a) Interference (solid line) and diffraction (dashed line) functions plotted for double-slit Fraunhofer diffraction when the slit separation is six times the slit width ($a = 6b$). (b) Irradiance for the double slit of (a). The curve represents the product of the interference and diffraction factors. (c) Diffraction pattern due to a single slit. (d) Diffraction pattern due to a double-slit aperture, with each slit of width b like the one that produced (c), but with *a/b* unspecified. (Both photos are from M. Cagnet, M. Francon, and J. C. Thrierr, *Atlas of Optical Phenomenon*, Plate 18, Berlin: Springer-Verlag, 1962.)

when $\beta = m\pi$, with $m = \pm1, \pm2, \ldots$, as shown. In terms of the spatial angle θ, this condition is

$$\text{diffraction minima: } m\lambda = b \sin\theta \tag{28}$$

as in Eq. (12). When these minima happen to coincide with interference fringe maxima, the fringe is missing from the pattern. Interference maxima occur for $\alpha = p\pi$, with $p = 0, \pm1, \pm2, \ldots$, or when

$$\text{interference maxima: } p\lambda = a \sin\theta \tag{29}$$

When the conditions expressed by Eqs. (28) and (29) are satisfied at the same point in the pattern (same θ), dividing one equation by the other gives the condition for missing orders.

$$\text{condition for missing orders: } a = \left(\frac{p}{m}\right)b \tag{30}$$

or

$$\alpha = \left(\frac{p}{m}\right)\beta$$

Thus, when the slit separation is an integral multiple of the slit width, the condition for missing order is met exactly. For example, when $a = 2b$, then $p = 2m = \pm2, \pm4, \pm6, \ldots$ gives the missing orders of interference. For the case plotted in Figure 14a and b, $a = 6b$, and the missing orders are those for which $p = \pm6, \pm12$, and so on. Figure 14c and d contains photographs of a single-slit pattern and a double-slit pattern with the same slit width. (What is the ratio of *a/b* in this case? Would a ratio of $a/b = 9$ fit the pattern shown?) Evidently, when $a = Nb$ and N is large, the first missing order at $p = \pm N$ is far from the center of the pattern. To produce a simple Young's interference pattern for two slits, one accordingly makes $a \gg b$ so that N is large. A large number of fringes then fall under the central maximum of the diffraction envelope. As a trivial but satisfying case, observe that when $a = b$, Eq. (30) requires that all orders (except $p = 0$) are missing. These dimensions cannot be satisfied, however, unless the two slits have merged into one and are unable to produce interference fringes. When $a = b$, the resulting pattern is, of course, that of a single slit.

6 DIFFRACTION FROM MANY SLITS

For an aperture of multiple slits (a *grating*), the integrals of Eq. (23), together with Figure 13, are extended by integrating over N slits. The individual slits are identified by the index j in the following expression for the resultant amplitude:

$$E_P = \frac{E_L}{r_0} e^{i(kr_0 - \omega t)} \sum_{j=1}^{N/2} \left\{ \int_{[-(2j-1)a-b]/2}^{[-(2j-1)a+b]/2} e^{isk \sin\theta}\, ds + \int_{[(2j-1)a-b]/2}^{[(2j-1)a+b]/2} e^{isk \sin\theta}\, ds \right\} \tag{31}$$

As j increases, pairs of slits symmetrically placed below (first integral) and above (second integral) the origin are included in the integration. When $j = 1$, for example, Eq. (31) reduces to the double-slit case, Eq. (23). When $j = 2$, the next two slits are included, whose edges are located at

$\frac{1}{2}(-3a - b)$ and $\frac{1}{2}(-3a + b)$ below the origin and $\frac{1}{2}(3a - b)$ and $\frac{1}{2}(3a + b)$ above the origin.[3] When $j = N/2$, all slits are accounted for.

Let us first concentrate on the integrals contained within the curly brackets, which we shall refer to as K, temporarily. After integration and substitution of limits, we get

$$K = \frac{1}{ik \sin\theta}\{e^{-ik \sin\theta[(2j-1)a-b]/2} - e^{-ik \sin\theta[(2j-1)a+b]/2}\}$$
$$+ \frac{1}{ik \sin\theta}\{e^{ik \sin\theta[(2j-1)a+b]/2} - e^{ik \sin\theta[(2j-1)a-b]/2}\}$$

Using Eqs. (24) and (25) again for α and β,

$$K = \frac{b}{2i\beta}[e^{-i(2j-1)\alpha}(e^{i\beta} - e^{-i\beta}) + e^{i(2j-1)\alpha}(e^{i\beta} - e^{-i\beta})]$$

With the help of Euler's equation, this can be written as

$$K = \frac{b}{2i\beta}(2i \sin\beta)\{2\cos[(2j-1)\alpha]\}$$

or

$$K = 2b\frac{\sin\beta}{\beta} \operatorname{Re}[e^{i(2j-1)\alpha}]$$

where we have expressed the cosine as the real part of the corresponding exponential. Returning to Eq. (31), we need next the sum S:

$$S = 2b\frac{\sin\beta}{\beta} \operatorname{Re}\sum_{j=1}^{N/2} e^{i(2j-1)\alpha}$$

Expanding the sum, we find

$$S = 2b\frac{\sin\beta}{\beta} \operatorname{Re}[e^{i\alpha} + e^{i3\alpha} + e^{i5\alpha} + \cdots + e^{i(N-1)\alpha}]$$

The series in brackets is a geometric series whose first term a and ratio r can be used to find its sum, given by

$$a\left(\frac{r^n - 1}{r - 1}\right) = e^{i\alpha}\left[\frac{(e^{2i\alpha})^{N/2} - 1}{e^{2i\alpha} - 1}\right] = \frac{e^{iN\alpha} - 1}{e^{i\alpha} - e^{-i\alpha}}$$

Using Euler's equation, this can be recast into the form

$$\frac{(\cos N\alpha - 1) + i \sin N\alpha}{2i \sin\alpha} = \frac{i(\cos N\alpha - 1) - \sin N\alpha}{-2\sin\alpha}$$

whose real part is $(\sin N\alpha)/(2\sin\alpha)$. Then,

$$S = b\frac{\sin\beta}{\beta}\frac{\sin N\alpha}{\sin\alpha}$$

[3]This expression is adapted to N even. For N large, one need not be concerned about the parity of N. For N small, however, N odd can be handled by taking the origin at the center of the central slit. This approach is left to the problems.

and

$$E_P = \frac{E_L}{r_0} e^{i(kr_0 - \omega t)} \left\{ \frac{b \sin \beta}{\beta} \frac{\sin N\alpha}{\sin \alpha} \right\}$$

As before, the irradiance is proportional to the square of the field amplitude,

$$I = I_0 \underbrace{\left(\frac{\sin \beta}{\beta} \right)^2}_{\text{diffraction}} \underbrace{\left(\frac{\sin N\alpha}{\sin \alpha} \right)^2}_{\text{interference}} \tag{32}$$

where I_0 includes all the constants, the first set of brackets encloses the diffraction factor, and the second set of brackets encloses the interference factor.

Although derived here for an even number N of slits, the result expressed by Eq. (32) is valid also for N odd (see problem 21). When $N = 1$ and $N = 2$, Eq. (32) reduces to the results obtained previously for single- and double-slit diffraction, respectively. By now we are familiar with the factor in β representing the diffraction envelope of the resultant irradiance. Let us examine the factor $(\sin N\alpha/\sin \alpha)^2$, which evidently describes interference between slits. When $\alpha = 0$ or some multiple of π, the expression reduces to an indeterminate form. We can show, in fact, that for such values, the expression is a maximum. Employing L'Hôpital's rule for any $m = 0, \pm 1, \pm 2, \ldots,$

$$\lim_{\alpha \to m\pi} \frac{\sin N\alpha}{\sin \alpha} = \lim_{\alpha \to m\pi} \frac{N \cos N\alpha}{\cos \alpha} = \pm N$$

Thus, the interference factor in Eq. (32) describes a series of sharp irradiance peaks (*principal maxima*). The irradiance at a principal maximum is proportional to N^2 and the principal maxima are centered at values for which $\alpha = 0, \pm\pi, \pm 2\pi, \pm 3\pi$, and so on. For the case $N = 8$, four such peaks, at $\alpha = 0, \pi, 2\pi$, and 3π are shown in Figure 15a. In between successive peaks there are shown $N - 2 = 6$ *secondary* peaks. The diffraction factor in Eq. (32) is plotted as the dotted line in Figure 15a, and the full irradiance which is proportional to the product of the diffraction and interference factors is plotted in Figure 15b. Note that the resulting irradiance in Figure 15b reflects the presence of the limiting diffraction envelope.

Let us now develop a more explicit understanding of the formation of the *secondary* peaks. The interference factor $(\sin(N\alpha)/\sin \alpha)^2$ goes to zero when the function in its numerator $(\sin(N\alpha))$ goes to zero but the function in its denominator $(\sin \alpha)$ does not. The numerator is identically zero under the condition $\alpha = p\pi/N$, where p takes on integer values. For the 8-slit case $(N = 8)$ and p from 0 to $N = 8$, the numerator goes to zero for the sequence of values $\alpha = 0, \pi/8, 2\pi/8, 3\pi/8, 4\pi/8, 5\pi/8, 6\pi/8, 7\pi/8$, and $8\pi/8$. Note that $\alpha = 0$ when $p = 0$ and $\alpha = \pi$ when $p = N = 8$. These values, $\alpha = 0$ and $\alpha = \pi$, correspond to the first two principal maxima in Figure 15. For $N = 8$, the function $\sin(N\alpha)$ in the numerator of the interference factor goes to zero for each of the seven intermediate terms in the sequence ($\alpha = \pi/8$ to $\alpha = 7\pi/8$), but the function in the denominator $\sin \alpha$ does *not* go to zero for the these seven intermediate values. Thus, for the case at hand, there are $N - 1 = 7$ zeroes, and as a consequence $N - 2 = 6$ secondary maxima, between the principal maxima. For the case of arbitrary N, there will be $N - 1$ zeroes and $N - 2$ secondary peaks between principal maxima. We have looked in detail at the behavior as p ranges from 0 to N. This pattern simply repeats for p from N to $2N$ and so on, thereby accounting for all of the principal and secondary peaks. The situation described by Eq. (32) and

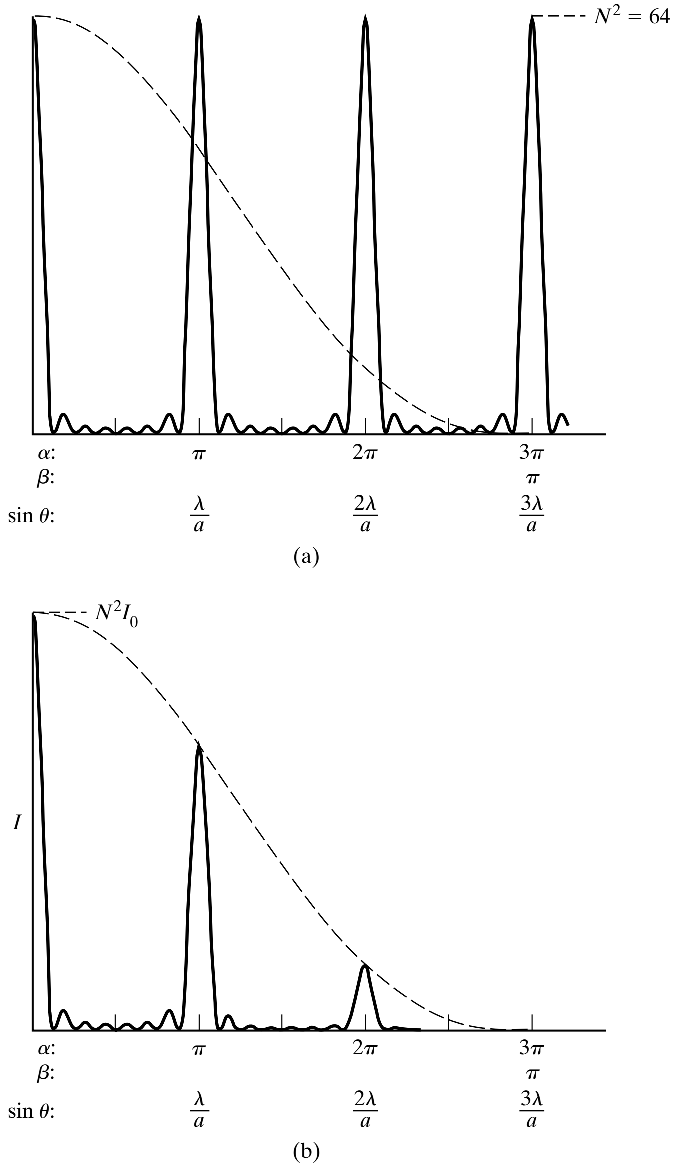


Figure 15 (a) Interference factor $\sin^2(N\alpha)/\sin^2(\alpha)$ (solid line) and diffraction factor $\sin^2\beta/\beta^2$ (dashed line) plotted for multiple-slit Fraunhofer diffraction when $N = 8$ and $a = 3b$. The interference factor peaks at $N^2 = 8^2 = 64$. The diffraction factor has a maximum value of 1 for $\beta = 0$. (b) Irradiance function $I = I_0 \dfrac{\sin^2\beta}{\beta^2}\dfrac{\sin^2(N\alpha)}{\sin^2\alpha}$ for the multiple slit of (a). The irradiance at the peak of the central principal maximum (at $\alpha = 0$) is $I = N^2 I_0$. Subsequent principal maxima are less bright since they are limited by the diffraction envelope, $\sin^2\beta/\beta^2$ (dashed line).

presented graphically in Figure 15 is precisely described by the following set of equations and conditions:

$$\text{for } \alpha = \frac{p\pi}{N}, \qquad p = 0, \pm1, \pm2, \ldots \pm N \ldots \pm 2N \ldots$$
$$\text{principal maxima occur for } p = 0, \pm N, \pm 2N, \ldots \tag{33}$$
$$\text{secondary minima occur for } p = \text{ all other integer values}$$

A practical device that makes use of multiple-slit diffraction is the *diffraction grating*. For large N, its principal maxima are bright, distinct, and spatially well separated. According to Eq. (33) the principal maxima occur for $p/N = m = 0, \pm1, \pm2, \ldots$. Thus the condition for the principal maxima is simply

$$\alpha = m\pi \qquad m = 0, \pm1, \pm2 \ldots$$

Recall from Eq. (25) that $\alpha = (1/2)ka\sin\theta = \pi a \sin\theta/\lambda$, so that the condition for the existence of a principal maximum can be recast as

$$m\lambda = a\sin\theta \tag{34}$$

Equation (34) is sometimes called the *diffraction grating equation* and m is identified as the *order* of the diffraction.

Now as the number N of slits increases, the brightness of the principal maxima increase as N^2. This increase in irradiance at the peaks of the principal maxima must be accompanied by an overall decrease in irradiance between the peaks of the principal maxima. Thus gratings with more slits direct a greater fraction of the energy emerging from the slits towards the positions of the peaks of the principal maxima than do gratings with fewer slits. Gratings with more slits produce brighter and narrower principal maxima.

Returning to Eq. (34), some insight is gained by examining Figure 16, which shows representative slits of a grating illuminated by plane wavefronts of monochromatic light. Wavelets emerging from each slit arrive in phase at angular deviation θ from the axis if every path difference like AB $(= a \sin \theta)$ equals an integral number m of wavelengths. When $AB = m\lambda$, the grating Eq. (34) follows immediately. When all waves arrive in phase, the resulting phasor diagram is formed by adding N phasors all in the same "direction," giving a maximum resultant. At such points, the principal maxima of Figure 15 are produced. Secondary maxima result because a uniform phase difference between waves from adjoining slits causes the phase diagram to curl up, with a smaller resultant. At each of the minima, the phasor diagram forms a closed figure, so that cancellation is complete. The phase difference between waves from adjoining slits and in the direction of θ can be found from Figure 15a by recalling that the angle α represents half the phase difference between successive slits. Thus, the first principal maximum from the center, at $\alpha = \pi$, occurs when the phase difference between successive waves is precisely 2π.

Photographs of diffraction fringes produced by 2, 3, 4, and 5 slits are shown in Figure 17. An examination of the four photographs shows that the principal maxima become narrower and secondary maxima begin to appear as the number of slits increases. For example, notice that the $N - 2 = 3$ secondary maxima appear between the principal maxima for the case $N = 5$. The diffraction grating—for N very large—is discussed further in some detail in the next chapter.

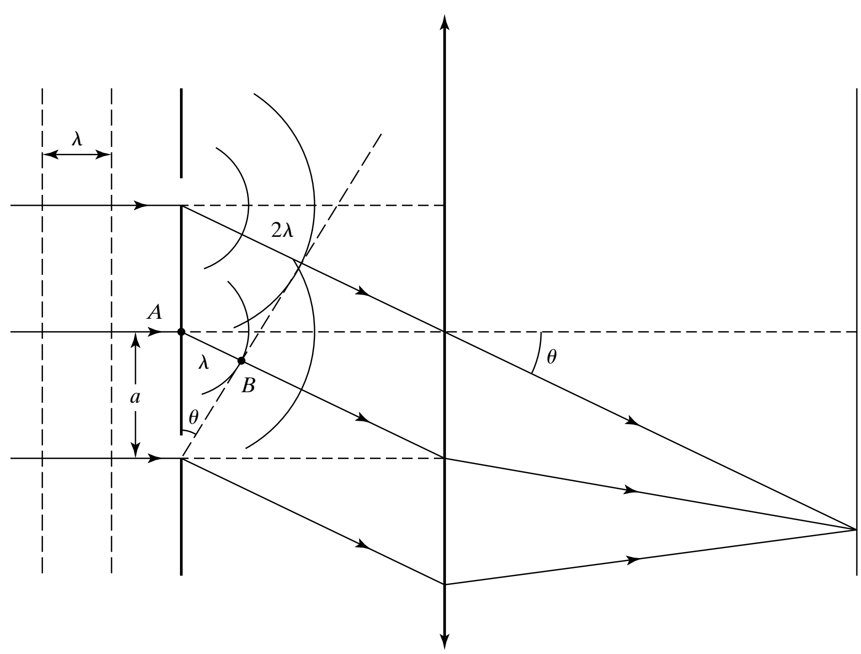
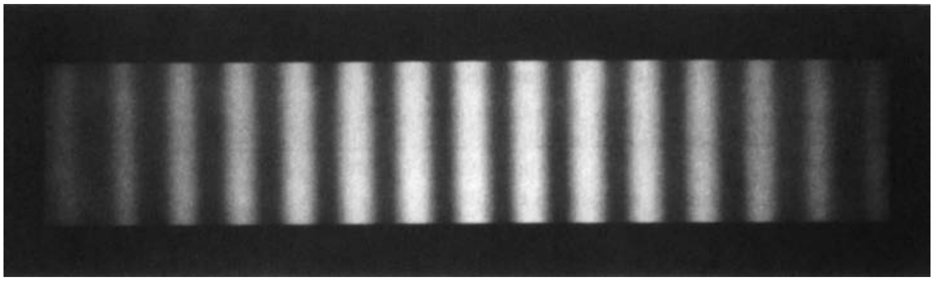
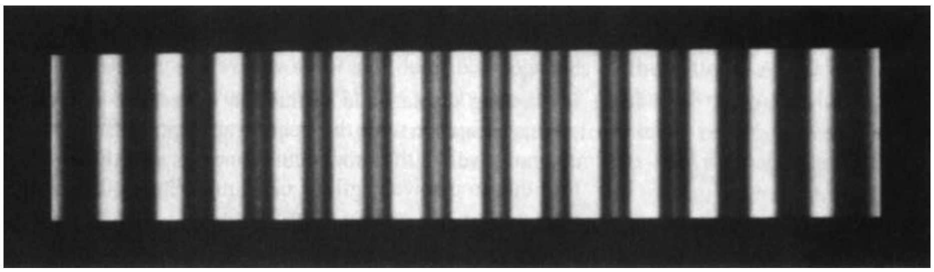


Figure 16 Representative grating slits illuminated by collimated monochromatic light. Formation of the first-order diffraction maximum is shown.

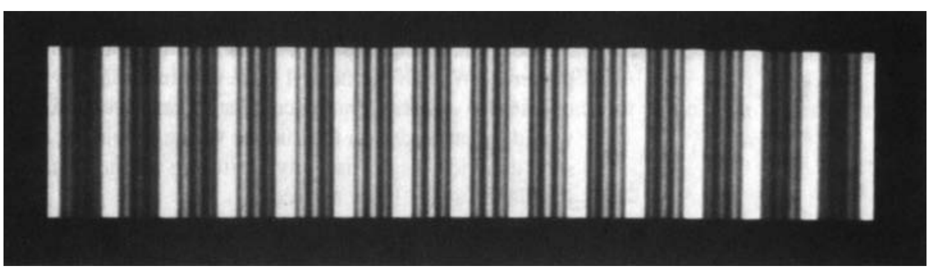


(a) $N = 2$



(b) $N = 3$

(c) $N = 4$



(d) $N = 5$

Figure 17 Diffraction fringes produced in turn by two, three, four, and five slits. (From M. Cagnet, M. Francon, and J. C. Thrierr, *Atlas of Optical Phenomenon*, Plate 19, Berlin: Springer-Verlag, 1962.)

PROBLEMS

1 A collimated beam of mercury green light at 546.1 nm is normally incident on a slit 0.015 cm wide. A lens of focal length 60 cm is placed behind the slit. A diffraction pattern is formed on a screen placed in the focal plane of the lens. Determine the distance between (a) the central maximum and first minimum and (b) the first and second minima.

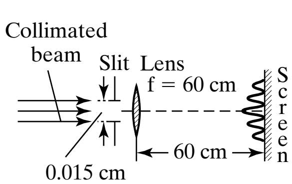


Figure 18 Problem 1.

2 Call the irradiance at the center of the central Fraunhofer diffraction maximum of a single slit I_0 and the irradiance at some other point in the pattern I. Obtain the ratio I/I_0 for a point on the screen that is 3/4 of a wavelength farther from one edge of the slit than the other.

3 The width of a rectangular slit is measured in the laboratory by means of its diffraction pattern at a distance of 2 m from the slit. When illuminated normally with a parallel beam of laser light (632.8 nm), the distance between the third minima on either side of the principal maximum is measured. An average of several tries gives 5.625 cm.

a. Assuming Fraunhofer diffraction, what is the slit width?
b. Is the assumption of far-field diffraction justified in this case? What is the ratio L/L_{min}?

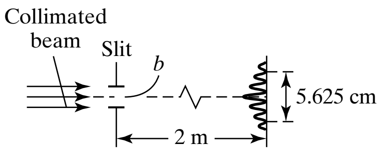


Figure 19 Problem 3.

4 In viewing the far-field diffraction pattern of a single slit illuminated by a discrete-spectrum source with the help of absorption filters, one finds that the fifth minimum of one wavelength component coincides exactly with the fourth minimum of the pattern due to a wavelength of 620 nm. What is the other wavelength?

5 Calculate the rectangular slit width that will produce a central maximum in its far-field diffraction pattern having an angular breadth of 30°, 45°, 90°, and 180°. Assume a wavelength of 550 nm.

6 Consider the far-field diffraction pattern of a single slit of width 2.125 μm when illuminated normally by a collimated beam of 550-nm light. Determine (a) the angular radius of its central peak and (b) the ratio I/I_0 at points making an angle of $\theta = 5°, 10°, 15°$, and $22.5°$ with the axis.

7 **a.** Find the values of β for which the fourth and fifth secondary maxima of the single-slit diffraction pattern occur. (See the discussion surrounding Figure 3.)
b. Find the ratio of the irradiance of the maxima of part (a) to the irradiance at the central maximum of the single-slit diffraction pattern.

8 Compare the relative irradiances of the first two secondary maxima of a circular diffraction pattern to those of a single-slit diffraction pattern.

9 The Lick Observatory has one of the largest refracting telescopes, with an aperture diameter of 36 in. and a focal length of 56 ft. Determine the radii of the first and second bright rings surrounding the Airy disc in the diffraction pattern formed by a star on the focal plane of the objective. See Figure 8b.

10 A telescope objective is 12 cm in diameter and has a focal length of 150 cm. Light of mean wavelength 550 nm from a distant star enters the scope as a nearly collimated beam. Compute the radius of the central disk of light forming the image of the star on the focal plane of the lens.

11 Suppose that a CO_2 gas laser emits a diffraction-limited beam at wavelength 10.6 μm, power 2 kW, and diameter 1 mm. Assume that, by multimoding, the laser beam has an essentially uniform irradiance over its cross section. Approximately how large a spot would be produced on the surface of the moon, a distance of 376,000 km away from such a device, neglecting any scattering by the earth's atmosphere? What will be the irradiance at the lunar surface?

12 Assume that a 2-mm-diameter laser beam (632.8 nm) is diffraction limited and has a constant irradiance over its cross section. On the basis of spreading due to diffraction alone, how far must it travel to double its diameter?

13 Two headlights on an automobile are 45 in. apart. How far away will the lights appear to be if they are just resolvable to a person whose nocturnal pupils are just 5 mm in diameter? Assume an average wavelength of 550 nm.

14 Assume that the pupil diameter of a normal eye typically can vary from 2 to 7 mm in response to ambient light variations.
a. What is the corresponding range of distances over which such an eye can detect the separation of objects 1 mm apart?
b. Experiment to find the range of distances over which you can detect the separation of lines placed 1 mm. apart. Use the results of your experiment to estimate the diameter range of your own pupils.

15 A double-slit diffraction pattern is formed using mercury green light at 546.1 nm. Each slit has a width of 0.100 mm. The pattern reveals that the fourth-order interference maxima are missing from the pattern.
a. What is the slit separation?
b. What is the irradiance of the first three orders of interference fringes, relative to the zeroth-order maximum?

16 **a.** Show that the number of bright fringes seen under the central diffraction peak in a Fraunhofer double-slit pattern is given by $2(a/b) - 1$, where a/b is the ratio of slit separation to slit width.
b. If 13 bright fringes are seen in the central diffraction peak when the slit width is 0.30 mm, determine the slit separation.

17 **a.** Show that in a double-slit Fraunhofer diffraction pattern, the ratio of widths of the central diffraction peak to the central interference fringe is $2(a/b)$, where a/b is the ratio of slit separation to slit width. Notice that the result is independent of wavelength.
b. Determine the peak-to-fringe ratio, in particular when $a = 10b$.

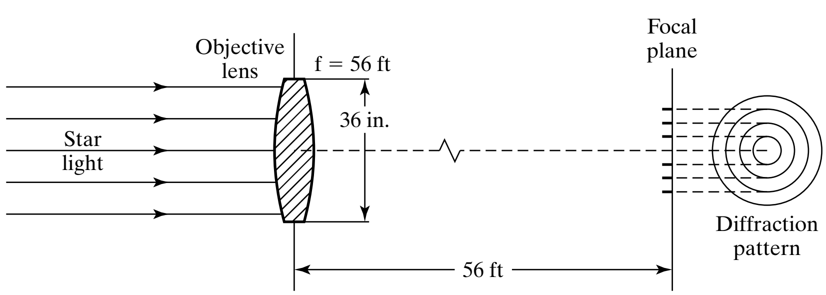


Figure 20 Problem 9.

18 Calculate by integration the irradiance of the diffraction pattern produced by a three-slit aperture, where the slit separation a is three times the slit width b. Make a careful sketch of I versus $\sin\theta$ and describe properties of the pattern. Also show that your results are consistent with the general result for N slits, given by Eq. (32).

19 Make a rough sketch for the irradiance pattern from seven equally spaced slits having a separation-to-width ratio of 4. Label points on the x-axis with corresponding values of α and β.

20 A 10-slit aperture, with slit spacing five times the slit width of 1×10^{-4} cm, is used to produce a Fraunhofer diffraction pattern with light of 435.8 nm. Determine the irradiance of the principal interference maxima of orders 1, 2, 3, 4, and 5 relative to the central fringe of zeroth order.

21 Show that one can arrive at Eq. (32) by taking the origin of coordinates at the midpoint of the central slit in an array where N is odd.

22 A rectangular aperture of dimensions 0.100 mm along the x-axis and 0.200 mm along the y-axis is illuminated by coherent light of wavelength 546 nm. A 1-m focal length lens intercepts the light diffracted by the aperture and projects the diffraction pattern on a screen in its focal plane. See Figure 21.

a. What is the distribution of irradiance on the screen near the pattern center as a function of x and y (in mm) and I_0, the irradiance at the pattern center?
b. How far from the pattern center are the first minima along the x and y directions?
c. What fraction of the I_0 irradiance occurs at 1 mm from the pattern center along the x- and y-directions?
d. What is the irradiance at the point $(x = 2,\ y = 3)$ mm?

23 What is the angular half-width (from central maximum to first minimum) of a diffracted beam for a slit width of (a) λ; (b) 5λ; (c) 10λ?

24 A property of the Bessel function $J_1(x)$ is that, for large x, a closed form exists, given by

$$J_1(x) = \frac{\sin x - \cos x}{\sqrt{\pi x}}$$

Find the angular separation of diffraction minima far from the axis of a circular aperture.

25 We have shown that the secondary maxima in a single-slit diffraction pattern do not fall exactly halfway between minima, but are quite close. Assuming they are halfway:

a. Show that the irradiance of the mth secondary peak is given approximately by

$$I_m \cong I_0 \frac{1}{\left[\left(m + \frac{1}{2}\right)\pi\right]^2}$$

b. Calculate the percent error involved in this approximation for the first three secondary maxima.

26 Three antennas broadcast in phase at a wavelength of 1 km. The antennas are separated by a distance of $\frac{2}{3}$ km and each antenna radiates equally in all horizontal directions. Because of interference, a broadcast "beam" is limited by interference minima. How many well-defined beams are broadcast and what are their angular half-widths?

27 A collimated light beam is incident normally on three very narrow, identical slits. At the center of the pattern projected on a screen, the irradiance is $I_{\max}$.

a. If the irradiance I_P at some point P on the screen is zero, what is the phase difference between light arriving at P from neighboring slits?
b. If the phase difference between light waves arriving at P from neighboring slits is π, determine the ratio $I_P/I_{\max}$.
c. What is $I_P/I_{\max}$ at the first principal maximum?
d. If the average irradiance on the entire screen is I_{av}, what is the ratio I_P/I_{av} at the central maximum?

28 Draw phasor diagrams illustrating the principal maxima and zero irradiance points for a four-slit aperture.

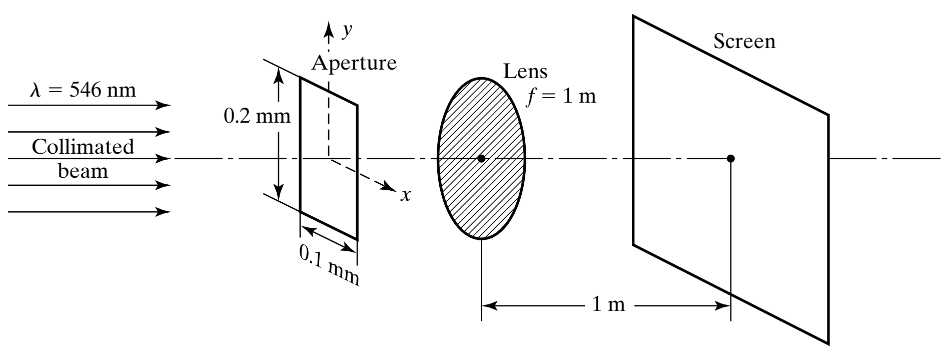
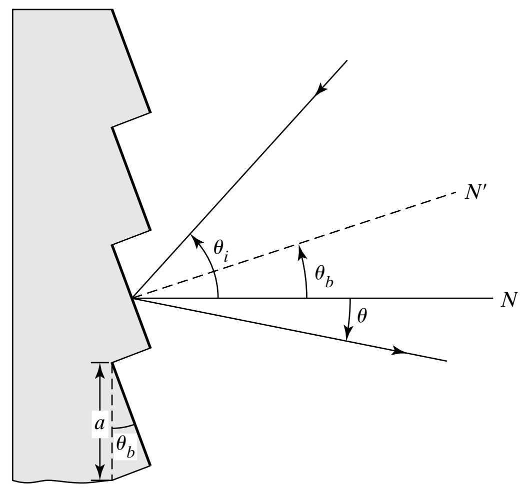


Figure 21 Problem 22.



12 The Diffraction Grating

INTRODUCTION

In this chapter we give a formal treatment of diffraction due to a large number of slits or apertures. The diffraction grating equation is first generalized to handle light beams incident on the grating at an arbitrary angle. Performance parameters of practical interest are then developed in discussions of the *spectral range, dispersion, resolution*, and *blaze* of a grating. A brief discussion of interference gratings and several conventional types of grating spectrographs ends the chapter.

1 THE GRATING EQUATION

A periodic, multiple-slit device designed to take advantage of the sensitivity of its diffraction pattern to the wavelength of the incident light is called a *diffraction grating*. A grating equation may be generalized for the case when the incident plane wavefronts of light make an angle θ_i with the plane of the grating, as in Figure 1. The net path difference for waves from successive slits is then

$$\Delta = \Delta_1 + \Delta_2 = a \sin \theta_i + a \sin \theta_m \tag{1}$$

The two sine terms in the path difference may add or subtract, depending on the direction θ_m of the diffracted light. To make Eq. (1) correct for all angles of diffraction, we need to adopt a sign convention for the angles. When the incident and diffracted rays are on the same side of the grating normal, as they are in Figure 1, θ_m is considered positive. When the diffracted rays are on the side of the grating normal opposite to that of the incident rays, θ_m is

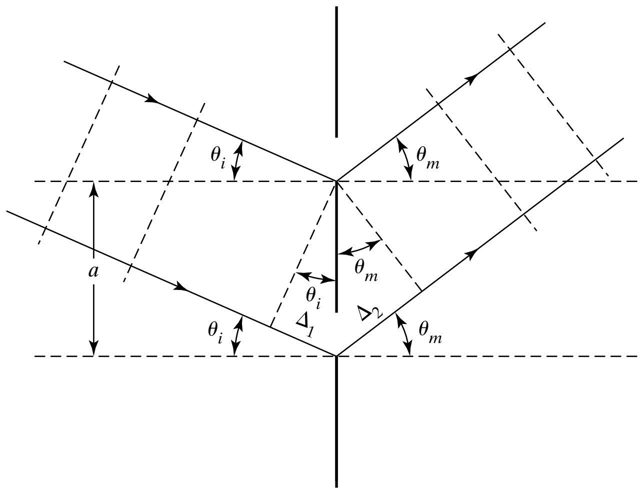


Figure 1 Neighboring grating slits illuminated by light incident at angle θ_i with the grating normal. For light diffracted in the direction θ_m, the net path difference from the two slits is $\Delta_1 + \Delta_2$.

considered negative. In the latter case, the net path difference for waves from successive slits is the difference $\Delta_1 - \Delta_2$, as would be evident in a modified sketch of Figure 1. In either case, when $\Delta = m\lambda$, all diffracted waves are in phase and the grating equation becomes

$$a(\sin\theta_i + \sin\theta_m) = m\lambda, \qquad m = 0, \pm1, \pm2, \ldots \tag{2}$$

When it is not necessary to distinguish between angles, the subscript on the angle of diffraction, θ_m, is often dropped. For each value of m, monochromatic radiation of wavelength λ is enhanced by the diffractive properties of the grating. By Eq. (2), the zeroth order of interference, $m = 0$, occurs at $\theta_m = -\theta_i$, the direction of the incident light, for all λ. Thus, light of all wavelengths appears in the central or zeroth-order peak of the diffraction pattern. Higher orders—both plus and minus—produce spectral *lines* appearing on either side of the zeroth order. For a fixed direction of incidence given by θ_i, the direction θ_m of each principal maximum varies with wavelength. For orders $m \neq 0$, therefore, the grating separates different wavelengths of light present in the incident beam, a feature that accounts for its usefulness in wavelength measurement and spectral analysis. As a dispersing element, the grating is superior to a prism in several ways. Figure 2a illustrates the formation of the spectral orders of diffraction for monochromatic light. Figure 2b shows the angular spread of the continuous spectrum of visible light for a particular grating. Note that second and third orders in this case partially overlap. Before wavelengths of spectral lines appearing in a region of overlap can be assigned, the actual order of the line must first be ascertained so that the appropriate value of m can be used in Eq. (2). Unlike the prism, a grating produces greater deviation from the zeroth-order point for longer wavelengths. Thus, when the spectrum is not a simple one, the overlap ambiguity is often resolved experimentally by using a filter that removes, say, the shorter wavelengths from the incident light. In this way, the spectral range of the incident light is limited by filtering until overlap is removed and each line can be correctly identified. At other times it may be advisable to limit the wavelength range accepted by the grating by first using an instrument of lower dispersion.

2 FREE SPECTRAL RANGE OF A GRATING

For diffraction gratings, the nonoverlapping wavelength range in a particular order is called the *free spectral range*, λ_{fsr}. Overlapping occurs because in the grating equation, the product $a \sin\theta$ may be equal to several possible combinations of $m\lambda$ for the light actually incident and processed by the optical system.

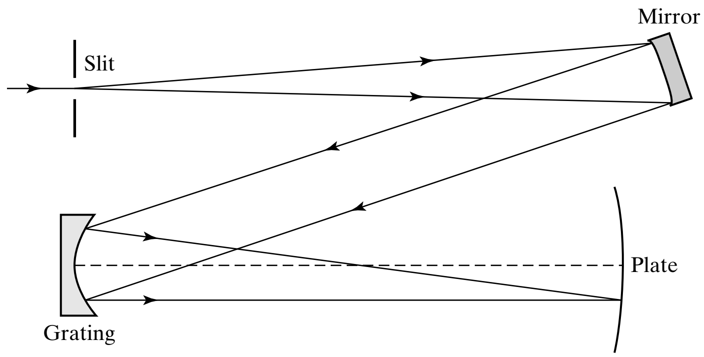


Figure 12 Wadsworth mount for a concave grating.

space occupied by this spectrograph can be quite large. The first three orders of diffraction are most commonly used. Typical angles of incidence may vary within the range 30° to 45°, and angles of diffraction may vary between 25° on the opposite side of the grating normal to 85° on the same side of the normal as the slit. Thus, much of the Rowland circle is useful for recording various portions of the spectrum. In Figure 11, the first-order spectrum spread (200 to 1200 nm) around the Rowland circle is shown for $\theta_i = 38°$ and a grating of 1200 grooves/mm. Spectral lines formed in this way may suffer rather severely from astigmatism. The *Wadsworth spectrograph* (Figure 12) uses a concave mirror, a concave grating, and a plate holder. The plate is mounted normal to the grating. The primary mirror collimates the light incident on the grating. This arrangement eliminates astigmatism and spherical aberration and dispenses with the need for the Rowland circle. Spectra are observed over a range making small angles to the grating normal, perhaps 10° to either side. To record different regions of the spectrum, the grating can be rotated and higher orders can be used. This version of a grating spectrograph allows more compact construction than does the Paschen-Runge design.

The ability of diffraction gratings to direct light of different wavelengths in different directions finds use in several other applications. For example, a Littrow grating can be used as a wavelength-selective mirror to ensure that only one of several laser lines experiences low loss in a laser cavity. Diffraction gratings are also sometimes used in wavelength-division multiplexing and demultiplexing systems in order to combine different-wavelength signals prior to launching them into an optical fiber and then to separate these signals once they have exited the fiber.

PROBLEMS

1 What is the angular separation in second order between light of wavelengths 400 nm and 600 nm when diffracted by a grating of 5000 grooves/cm?

2 **a.** Describe the dispersion in the red wavelength region around 650 nm (both in °/nm and in nm/mm) for a transmission grating 6 cm wide, containing 3500 grooves/cm, when it is focused in the third-order spectrum on a screen by a lens of focal length 150 cm.

b. Find the resolving power of the grating under these conditions.

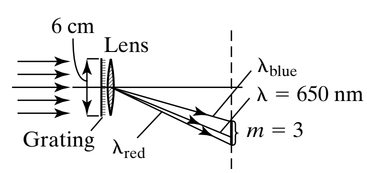


Figure 13 Problem 2.

3 **a.** What is the angular separation between the second-order principal maximum and the neighboring minimum on either side for the Fraunhofer pattern of a 24-groove grating having a groove separation of 10^{-3} cm and illuminated by light of 600 nm?

b. What slightly longer (or slightly shorter) wavelength would have its second-order maximum on top of the

13. (a) $M_e = 10^4$ W/m², $I_e(\theta = 0) = 7.96$ W/sr, $L_e =$ 3180 W/m²-sr (b) 1.56×10^{-4} W (c) 35.9 W/m²
15. 5.7 cm
16. $f = 53.3$ cm, 13.33 cm, 1.86 cm
17. 5.3 to 7.0 ft
18. 1.3×10^5 W/cm²
19. (a) 0.90 cm (b) 5.45 cm, 3 ×
20. (a) 27.8 mm (b) $f/3.1$, $f/5.4$, $f/9.4$
    (c) 16.0, 9.26, 5.35 mm (d) 0.03, 0.09, 0.27 s
22. (a) 2.8 cm (b) 10 ×
23. (a) 320 × (b) 0.516 cm
24. (a) 46.7 × (b) 8.68 cm
25. 5 cm
26. 14.9 cm
27. (a) 7 × (b) 2 cm (c) 5 mm (d) 2.3 cm (e) 337 ft
28. (b) 7.50 ×; 8.70 ×
29. 1.05 cm
30. (a) 8 cm, 3 × (b) 7.38 cm, 2.6 ×
31. 1.25 cm farther from the objective
32. (a) 12.5 × (b) 15 × (c) 0.13 cm, 3 mm, (d) 3.8°
34. −2.5 ft, −180 ×

Chapter 4

1. $y = ae^{-b(x+10t)^2}$
2. (a) $y = \frac{4\text{ m}^3}{(x+(2.5\text{ m/s})t)^2+2\text{ m}^2}$
3. (a) (1) and (2) qualify because they satisfy the wave equation; more simply, if $w = z + vt$, they are functions of w: $y = A\sin^2(4\pi w)$ and $y = Aw^2$.
   (b) (i) $v = 1$ m/s in $-z$-direction; (ii) $v = 1$ m/s in $+x$-direction
4. 10 m/s in $+x$-direction
5. (a) $\psi = 2\sin[2\pi(z/5\text{ m} + t/3\text{ s})]$
   (b) $\psi = 2\sin(2\pi/5)(z/\text{m} + \frac{5}{3}t/\text{s})$
   (c) $\psi = 2\exp[(2\pi i(z/5\text{ m} + t/3\text{ s})]$
6. (a) $y = (5\text{ m})\sin(\pi x/25\text{ m})$
   (b) $y = (5\text{ m})\sin[(\pi/25)(x/\text{m} + 8)]$
7. (a) 0.01 cm (b) 1000 Hz (c) 628.3 cm⁻¹ (d) 6283 s⁻¹
   (e) 1 ms (f) 10 cm/s (g) 10 cm
8. (a) +1 in y-direction (b) $-C/B$ in x-direction
   (c) C in z-direction
9. $y = 15\sin(kx + \pi/3)$
10. (b) $\pi/2, \pi/3, 0, -\pi/2, 0.6\pi$ (c) Subtract $\pi/2$ from each.
11. (a) $A\sin(2\pi/\lambda)(z - vt)$ (b) $A\sin(2\pi/\lambda)(\sqrt{2}x \pm vt)$
    (c) $A\sin(2\pi/\lambda)\left[\left(\sqrt{3}/3\right)(x + y + z) \pm vt\right]$
15. $E = 870$ V/m, $B = 2.90 \times 10^{-6}$ T
16. (a) 5×10^{-7} T (b) 19.88 W/m²
17. (a) 1.01×10^3 V/m, 3.37×10^{-6} T
    (b) 4.76×10^{21}/m²-s
    (c) $E = 1010\sin 2\pi(1.43 \times 10^6 r + 4.28 \times 10^{14} t)$, r in m, t in s
18. (a) 8.75×10^{-3} W/m², 2.57 V/m (b) 2×10^{13} W/m², 1.23×10^8 V/m, 0.410 T
21. $v = 0.168c$
22. $v = -0.917c$
23. $2\Delta\lambda = 0.12$ Å

Chapter 5

1. (a) The waves move in opposite directions along the x-axis, E_1 to the right, E_2 to the left, with equal speeds of $\frac{4}{3}$ m/s. (b) $t = \frac{3}{4}$ s (c) $x = 1$ m
2. (b) $E_R = 8.53\cos(0.20\pi - \omega t)$
3. $E_R = 6.08\cos(0.36\pi - 2\pi t/\text{s})$
4. $y = 11.6\sin(\omega t + 0.402\pi)$
5. $E = 0.695\cos(0.349 - \pi t/\text{s})$
6. (a) 2 V/m (v) 0.2 V/m
7. $\psi(t) = (2.48\text{ cm})\cos(2.51 - (20/\text{s})t)$
8. (a) $v_g = v_p[1 - (\omega/n)(dn/d\omega)]$ (b) $v_g < v_p$
9. $c/1.56$
10. $v_p = c/1.5$; $v_g = c/1.73$
12. $v_g = A =$ constant
14. $2(v/c)v_0$
15. 14 cm; 1.57 cm; 0.785 cm; 0 cm/s; T seconds
16. (a) 1.5 cm; 25 Hz; 20 cm; 5 m/s; opposite directions
    (b) 10 cm (c) −3 cm; 0 cm/s; 7.40×10^4 cm/s²
18. 40

Chapter 6

1. (a) 122 nm, 103 nm, 97.3 nm; ultraviolet
   (b) 656 nm, 486 nm, 434 nm; visible.
2. (a) no (b) less than 91.2 nm (c) less than 365 nm
3. (b) 2.4×10^{-21} J = 0.015 eV (c) 0.55
4. (a) 8.6×10^{-20} J = 0.54 eV (b) 5×10^{-10}
6. e^{-76}
10. (a) 0.4830 μm (b) 0.0756 W
11. 6266 K; 462.5 nm
12. 6105 K
15. 0.45 nm
16. 10^{-5} s; 3000 m (b) 5×10^{-10} s; 15 cm
17. 6
18. (a) half angle spread: 0.4 mrad (b) 80 cm.
19. 3.6 mm
20. (a) 0.81 μm; 0.75 μm; 0.585 μm; 0.525 μm
    (b) 76%; 70%; 55%; 49%
21. (a) 31.5 W (b) 1.26%

Chapter 7

1. (a) 11,950 and 21,240 W/m² (b) 12,960 W/m²
   (c) 33,200 W/m² (d) 0.95
2. 0.86, 0
3. 0.8; 3.73/1
4. (b) 1.78, 2.55, 4.00, 13.9
5. Lloyd's mirror interference fringes are produced, aligned parallel to the slit, and separated by 0.273 mm. The irradiance of the pattern is given by $I = 4I_0\sin^2(115y)$, with y measured in cm from the mirror surface.
6. 509 nm
7. 514.5 nm
8. To acquire coherent beams; 800 nm

9. (a) 83.3 cm (b) 83.3 fringes (c) 150 nm
10. 556 nm, 455 nm
11. 20.3′
12. 6′5″
13. 35′40″
14. 9.09×10^{-5} cm; orders 4 and 3, respectively
15. 498 nm
16. 1.33; 103 nm
17. (a) 2.78% (b) 89.3 nm (c) 1%
18. Soap film becomes wedge-shaped under gravity; the angle of the wedge is 1′14″
19. 15
20. 1.16×10^{-3} cm
21. 1.09 mm; 184
23. 3 m
25. 603.5 nm; 2.39 mm; 2.87×10^{-4} cm
26. 928 nm
27. (a) 980 V/m (b) 30° (c) $r' = 0.28; tt' = 0.9216$
(d) 274, 253, 19.8 V/m; 7.8%, 6.7%, 0.041%
(e) 903, 70.8 V/m; 85%, 0.52% (f) 258 nm

Chapter 8

1. 436 nm
2. One mirror makes a wedge angle of 0.0172° with the image of the other, reflected through the beam splitter. Fizeau fringes result.
3. 23.75 μm
4. (a) 80,000; (b)79,994
5. (a) $n = 1 + N\lambda/2L$ (b) 153
6. (a) 11.2° (b) 45.9°
7. 79.1 nm or $\lambda/8$
8. (a) 48,260 (b) 0.01013 cm
9. (a) 3.996×10^6 (b) 3.16×10^6
(c) 0.318 mm (d) 6.29 Å (e) 0.002 Å
10. (a) 329, 670 (b) 361 (c) 9.8×10^6
11. 2.18 cm
12. 0.161 mm
13. (a) 360° (b)180° (c) 2
14. 1; 0.47
15. (a) $R = \frac{4r^2 \sin^2 \delta/2}{(1-r^2)^2 + 4r^2 \sin^2 \delta/2}$
16. 16
17. For lossless mirrors with $R_{1,2} \equiv |r_{1,2}|^2$,
$T = \frac{(1-R_1)(1-R_2)}{(1-\sqrt{R_1R_2})^2 + 4\sqrt{R_1R_2}\sin^2\delta/2}; R = 1 - T$
18. (a) 70 (b) 1.5 GHz (c) 21 MHz (d) 2.2×10^7 (e) 8 ns
20. $9.99 \times 10^5; 1570; 3.14 \times 10^8; 8.4 \times 10^{-8}; 3.14 \times 10^8;$
1.6×10^{-6} nm; 250 nm; 0.16 nm; 3 GHz; 1.91 MHz
21. (a) 0.5 μm (b) 2×10^{-8} μm (c) -2.48×10^{-6} μm
22. (a) 10.6 GHz (b) 0.83 GHz; 0.62 GHz

Chapter 9

1. $f(x) = (4/\pi)\left(\sin kx + \frac{1}{3}\sin 3kx + \frac{1}{5}\sin 5kx + \cdots\right)$
2. $f(t) = \frac{E_0}{\pi} + \frac{E_0}{2}\cos\omega t + \frac{2E_0}{3\pi}\cos 2\omega t - \frac{2E_0}{15\pi}\cos 4\omega t + \cdots$
3. $g(\omega) = \frac{\sigma h}{\sqrt{2\pi}} e^{(-\sigma^2\omega^2)/2}$
If the width of the first is σ, the width of the second is $1/\sigma$. Thus the spectrum broadens as the original Gaussian narrows, and vice versa.
4. $|g(\omega)|^2 = \left(A^2\tau_0^2/4\pi^2\right)(\sin u/u)^2$, where $u = \omega\tau_0/2$
5. The narrow-band filter has a coherence length better by one order of magnitude: 3.48×10^{-5} m
6. 1.3 fm; 10^6 Hz; 300 m
7. 0.0243 mm
8. (a) 0.00138 nm (b) 1 ns
9. 2.5 mm
10. 0.0625 cm; 2.08×10^{-12} s
11. 0.144 cm
12. 4×10^{-7} Å; 3×10^4 Hz
13. (a) 2.08×10^{-12} s, 0.0625 cm (b) 0.36, 0.36 (c) 53
14. 1.01×10^{-4} cm, 2.90×10^{-6} cm²; 1.8, 35
15. (b) 2.55
17. 0.998, 0.63
18. 0.937, 0.686, 15.95 cm
19. (a) 0, 0, 0.596 cm (b) 0.895 mm

Chapter 10

1. 672
2. (a) 32 million (b) 0.67 million
3. (b) 1284
4. (a) 68.1° (b) 0.567 (c) 34.5°
5. (a) 0.64 (b) 79.5° (c) 6624, 3281
6. (c) 432 μm; 429 μm; 10.07 m
7. 159
8. 10.2 μm
9. 12 and 120, counting both polarizations
10. −70 db/km
11. 0.080 mW
12. 3.33 km; 10 km
13. 0.136 db/km
14. (b) −1.25 db, −6.02 db, −10 db, −20 db
15. (a) 1.0069 km; 1 km (b) 4.900 μs; 4.867 μs
16. 431 ns; 2.32 MHz
17. 77.2 ns
18. 14.6 ns/km
20. 457 ps; 1/146
21. 25 MHz
22. (a) 4 ns (b) 0.4 ns
23. 48.9 ns
24. (b) 3.9 ps/km; 4.3 ps/km
25. (a) 50.5 ns; 1.075 ns; 0.075 ns (b) 50.5 ns
27. (a) 100 GHz (b) 4 THz
28. (a) No ΔL satisfies Eqs. (21) and (22) with *exactly* integer m, but many approximate solutions exist. One such solution is $n\Delta L = 950.01\lambda_1 = 950.500\lambda_2$
(b) Ouput 2. (c) 0.94

Chapter 11

1. (a) 0.218 cm (b) 0.218 cm
2. 0.090
3. (a) 0.135 mm (b) 139
4. 496 nm

5. 2.125, 1.44, 0.778, and 0.55 μm
6. (a) 15° (b) 0.678, 0.166, 0, 0.0461
7. (a) 4.477π; 5.482π (b) 1/199; 1/298
8. Single slit: 0.047, 0.017; Circular Aperture: 0.018, 0.0042
9. 1.68×10^{-3} cm; 2.75×10^{-3} cm
10. 8.4×10^{-4} cm
11. 9725 km in diameter; 2.69×10^{-11} W/m^2
12. 5.2 m
13. 5.3 miles
14. 3 to 10.4
15. (a) 0.400 mm (b) 0.8106, 0.4053, 0.09006
16. (b) 2.10 mm
17. (b) 20
20. 0.875, 0.573, 0.255, 0.0547, 0
22. (a) $I = I_0 \frac{\sin^2(1.151y)\sin^2(0.575x)}{0.438x^2y^2}$
    (b) 5.46 mm along x; 2.73 mm along y
    (c) 0.895 along x; 0.629 along y (d) 0.005
23. (a) 90° (b) 11.5° (c) 5.7°
24. $\Delta\theta = (\lambda/D)(1/\cos\theta)$
25. (b) 4.7%, 1.8%, 0.84% for $m = 1, 2, 3$, respectively
26. $m = 0$; $\theta_{1/2} = 30°$
27. (a) 120° (b) $I_p = \left(\frac{1}{9}\right) I_{\max}$ (c) $I_p = I_{\max}$ (d) $I_p = 3I_{\text{av}}$

Chapter 12

1. 13°18′
2. (a) 0.0823°/nm; 0.464 nm/mm (b) 63,000
3. (a) 8.66′ (b) 612.5 nm (or 587.5 nm) (c) 48; 48
4. 987; 494
5. (a) 700 nm, 360 nm (b) 57.1°, 25.6° (c) 350 nm and 175 nm for crown glass; 180 nm and 90 nm for quartz
6. 120,000; 0.069 Å
7. (a) third order (b) any width smaller than light beam
8. (a) 21.8 cm, in each case (b) 9, in each case
   (c) 21.8 cm, 4.37 cm, and 0.0029 cm, respectively
9. (a) 8750 grooves/cm (b) 18.89°
   (c) 37.77° (d) 7.88 nm/deg
10. (a) 7000 (b) 0.018 mm
11. (a) −5.7° to +11.5° (b) 100,000 (c) 10 Å/mm (d) 1 m
12. about 5000 grooves/cm
13. (a) 1.16 μm (b) 18.4 Å/mm
14. (a) 11.5° (b) 11.8°
15. 3550 grooves/mm; reduces it
16. (a) 3647 (b) 1200 grooves/mm (c) 3.04 mm
17. (a) 557 to 318 (b) 960 (c) 388,800; 0.014 Å
    (d) 0.41°/nm (e) 5.5 Å

Chapter 13

1. near, near, far
2. maxima: 409, 136, 81.8 cm;
   minima: 204.5, 102, 68 cm
3. (a) 1.88 and 3.26 mm (b) 2.66 and 3.76 mm
4. (a) 0.0346 cm (b) 833 (c) 20 cm, 6.67 cm, 4 cm
6. (a) 0.02 cm (b) 2500
7. (a) 4 × (b) very nearly zero (c) 5; 6
8. 0.0012%
9. (a) 1/100 (b) 50.31 cm
10. 1.05, 1.48, 1.82 mm
12. 1.97 mm radius; zero
13. 14.8 cm
14. (a) $0.829I_u$ (b) $0.213I_u$
15. (a) $0.0119I_u$ (b) $1.23I_u$
16. (a) $0.018I_u$ (b) $0.223I_u$
17. $1.19I_u$; $0.861I_u$
18. $0.55I_u$
19. 21%
20. (b) 0.145 mm (c) $0.655I_u$
21. 19 μm

Chapter 14

2. (a) $\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}$: linearly polarized at −45°
   (b) $\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}$: linearly polarized at +45°
   (c) $\frac{1}{\sqrt{2}}\begin{bmatrix}1\\\frac{1}{\sqrt{2}}(1-i)\end{bmatrix}$: right-elliptically polarized at +45°
   (d) $\frac{1}{\sqrt{2}}\begin{bmatrix}1\\i\end{bmatrix}$: left-circularly polarized
3. (a) linearly polarized along x-direction, traveling in $+z$-direction with amplitude of $2E_0$
   (b) linearly polarized at 53.1° relative to the x-axis, traveling in the $+z$-direction with amplitude of $5E_0$
   (c) right-circularly polarized, traveling in $-z$-direction with amplitude of $5E_0$
4. 75°
5. right-circularly polarized light
6. (a) $\vec{\mathbf{E}} = E_0\left(\sqrt{3}\hat{\mathbf{y}} + \hat{\mathbf{z}}\right)e^{i(kx-\omega t)}$
   (b) $\vec{\mathbf{E}} = E_0(2\hat{\mathbf{z}} - i\hat{\mathbf{x}})e^{i(ky-\omega t)}$
   (c) $\vec{\mathbf{E}} = \hat{\mathbf{z}}E_0 \exp\left\{i\left[(x+y)k/\sqrt{2} - \omega t\right]\right\}$
7. (a) $C = 0, m\pi$ (b) $B = 0, \left(m + \frac{1}{2}\right)\pi$
   (c) $B = 0, A = \pm C, \left(m + \frac{1}{2}\right)\pi$
9. (a) linearly polarized, $\alpha = 18.4°, A = \sqrt{10}$ (b) right-circularly polarized, $A = 1$ (c) right-elliptically polarized; semimajor axis = 5 along y-axis, semiminor axis = 4 along x-axis (d) linearly polarized, horizontal, $A = 5$ (e) left-circularly polarized, $A = 2$ (f) linearly polarized, $\alpha = 56.3°, A = \sqrt{13}$ (g) left-elliptically polarized, $\varepsilon = 53.1°, \alpha = -7°, E_{0x} = 2, E_{0y} = 10$
10. right-elliptically polarized, symmetrical with x- and y-axes, $E_{0x}/E_{0y} = \sqrt{3}$
13. right-circularly polarized light
14. no light emerges
15. (a) right-elliptically polarized, major axis along x-axis
    (b) vertically linearly polarized
16. (a) linearly polarized at ±45° (b) elliptically polarized
17. $\begin{bmatrix}1 & -i\\ i & 1\end{bmatrix}$
20. (a) Elliptical polarization with inclination angle $\alpha = -25.097°$: $\frac{1}{\sqrt{13}}\begin{bmatrix}2\\\frac{3}{2} + \frac{i3\sqrt{3}}{2}\end{bmatrix}$ (b) Elliptical polarization