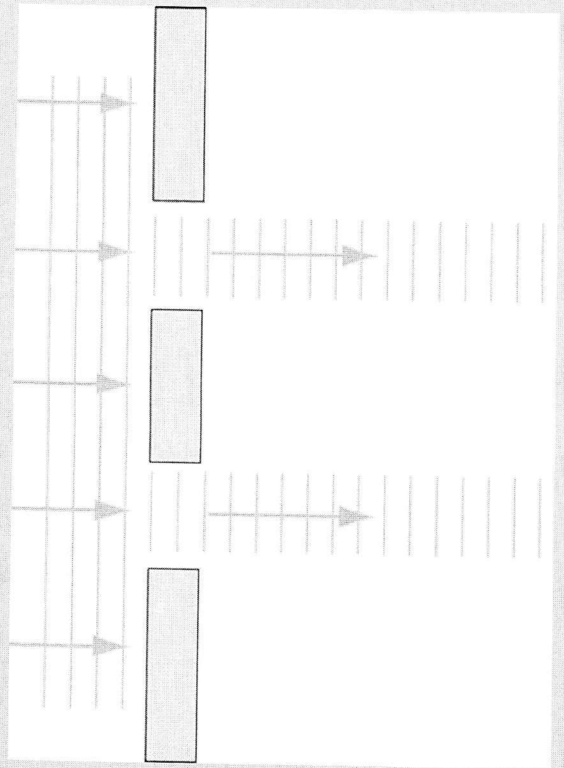
# Diffraction

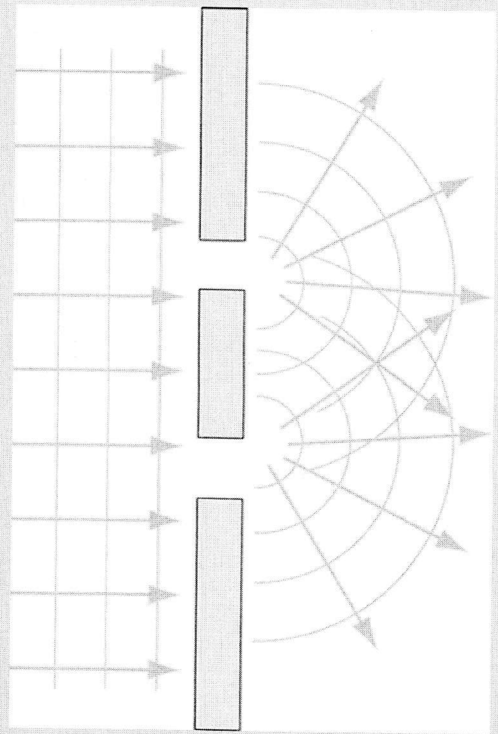
# Diffraction

**Diffraction**: The bending of waves as they pass by certain obstacles



No Diffraction

No spreading after passing though slits

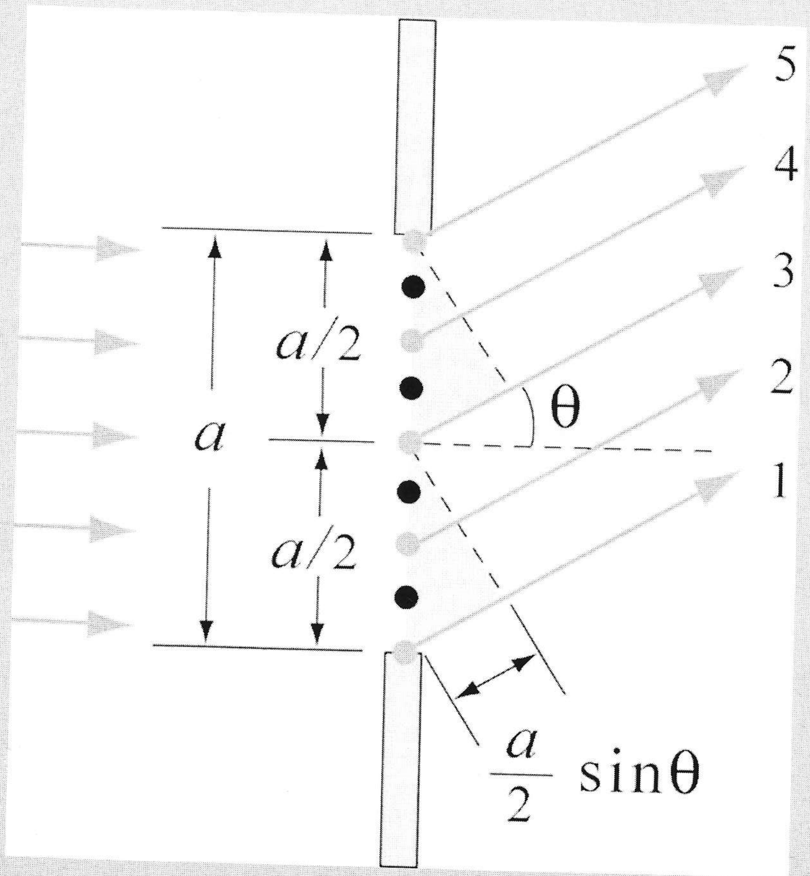


Diffraction

Spreading after passing though slits

# Single-Slit Diffraction

"Derivation" (Motivation) by Division:




Divide slit into two portions:

$$\delta = r_1 - r_3 = r_2 - r_4 = \frac{a}{2}\sin\theta$$

Destructive interference:

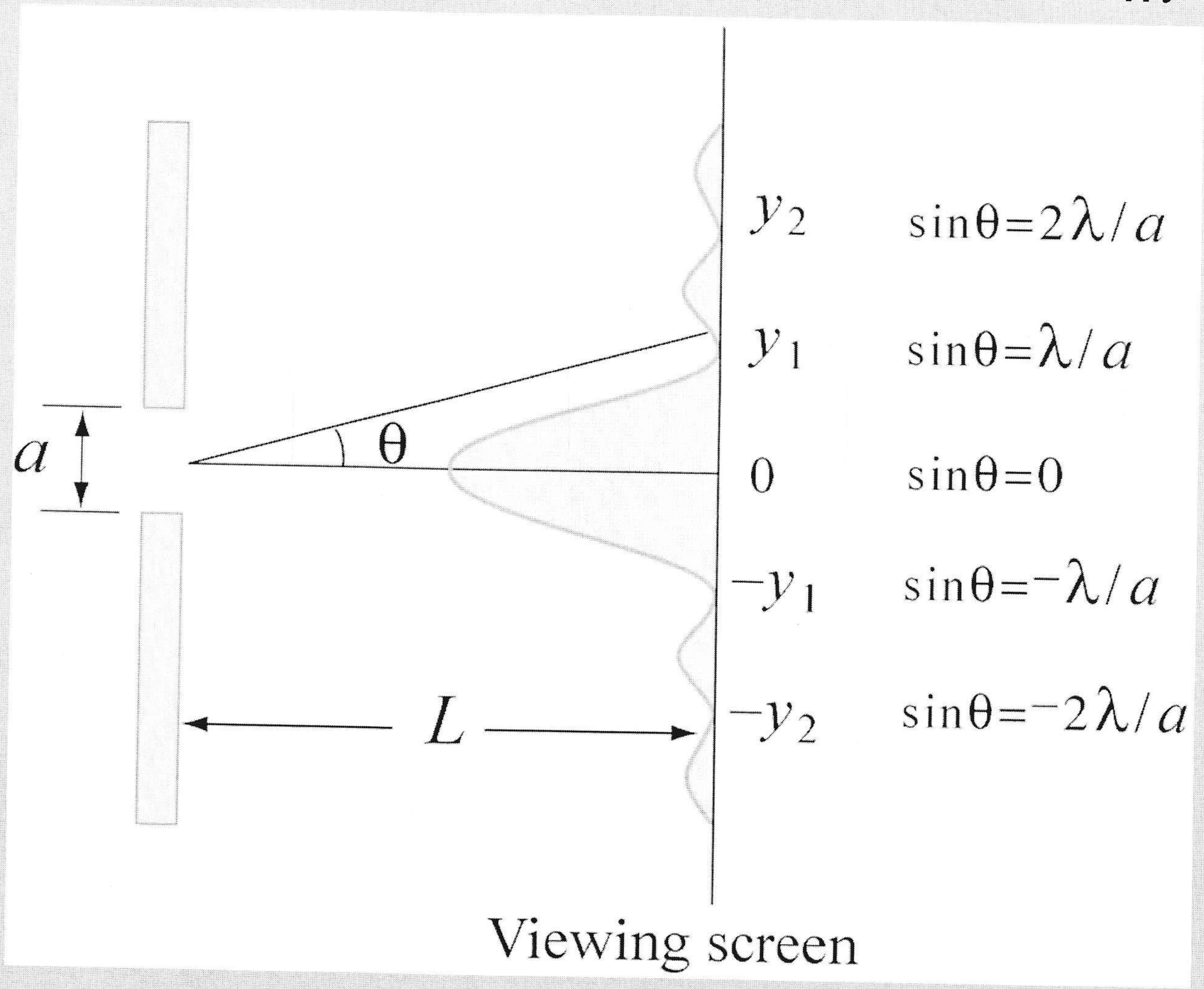
$$\delta = \frac{a}{2}\sin\theta = \left(m + \tfrac{1}{2}\right)\lambda$$

$$a\sin\theta = m\lambda \qquad m = \pm 1, \pm 2, \ldots$$

Don't get confused – this is DESTRUCTIVE!

# Intensity Distribution

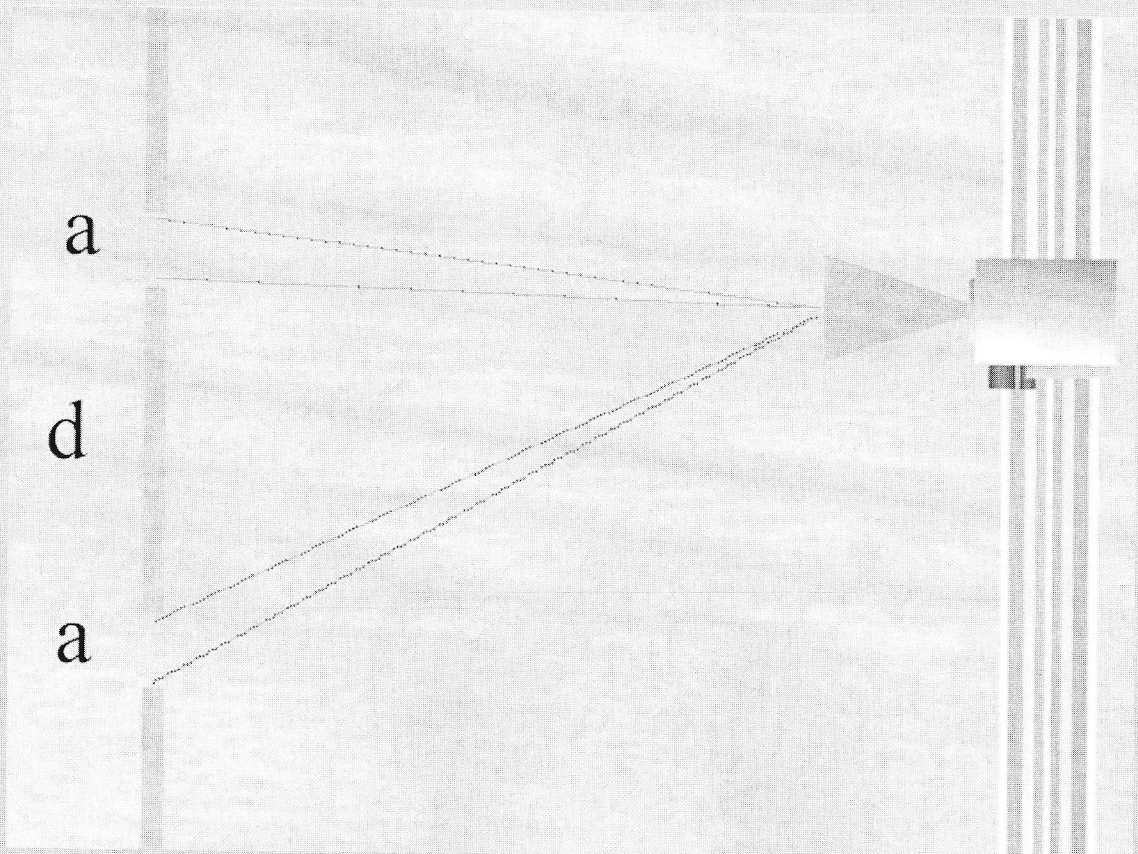
Destructive Interference: $a \sin\theta = m\lambda \qquad m = \pm 1, \pm 2, \ldots$



# Two Slits With Finite Width $a$

With more than one slit having finite width $a$, we must consider

1. Diffraction due to the individual slit
2. Interference of waves from different slits



# Two Slits With Finite width *a*

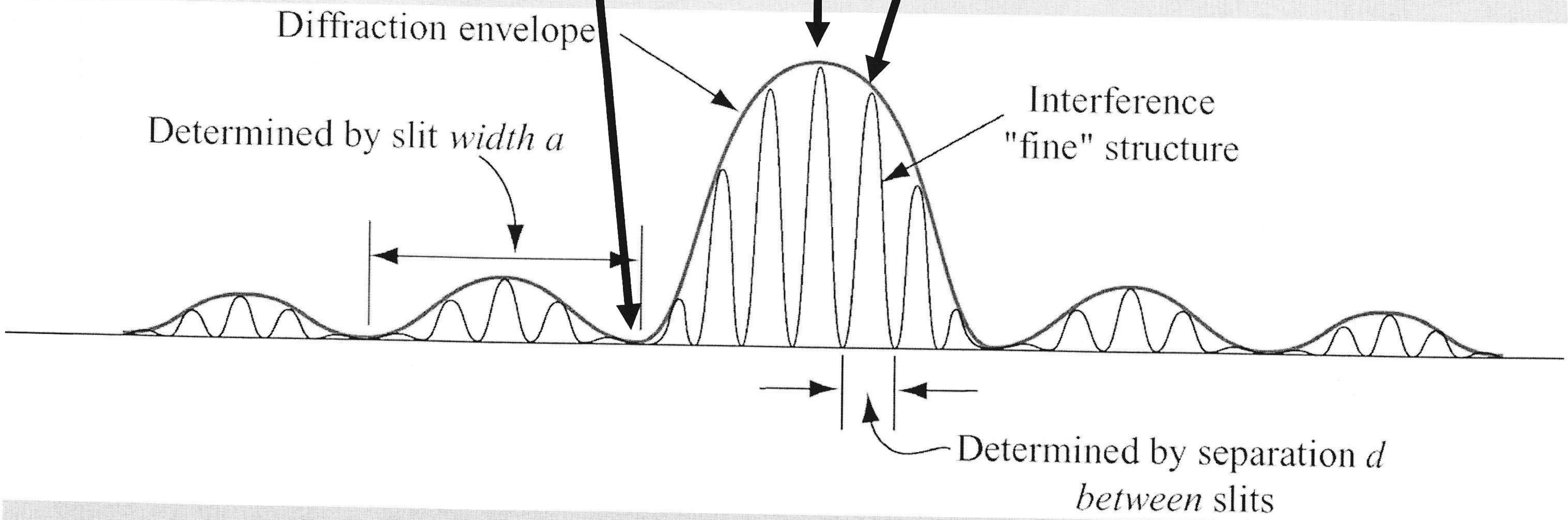
Zero Order Maximum

First Diff. Minimum

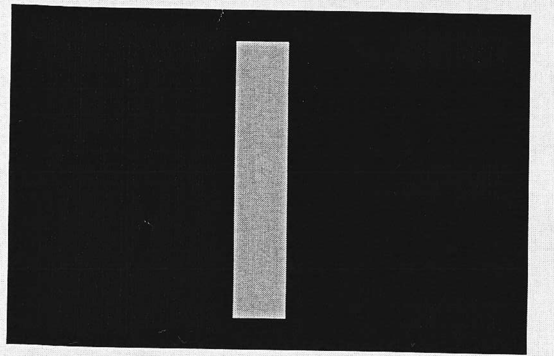
$$a \sin\theta = \lambda$$

First Order Maximum

$$d \sin\theta = \lambda$$



# Babinet's Principle



Case I: Put in a slit, get diffraction

Case II: Fill up slit, get nothing

Case III: Remove slit, get diffraction

By superposition, the E field with the slit and the E field with just the filling must be exact opposites in order to cancel:

$$E_{\text{filling}} = -E_{\text{slit}}$$

So the intensities are identical: $I_{\text{filling}} = I_{\text{slit}}$

# Diffraction

When light passes through a small aperture the wave passing through different parts can interfere ⟹ diffraction pattern on the screen placed at some distance behind it.

Experiments

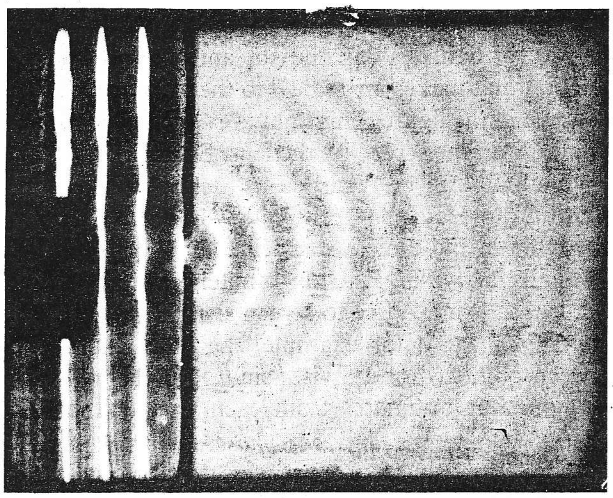


**FIGURE 26.21.** Diffraction of water waves by a single slit.

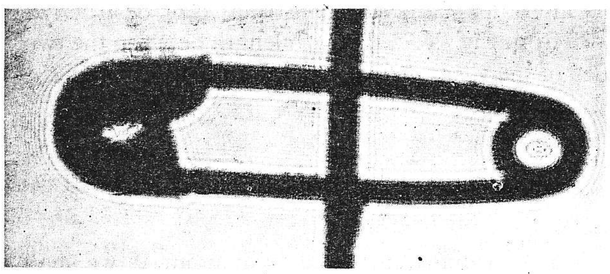


**FIGURE 26.22.** Photograph illustrating the diffraction of light by a safety pin. Courtesy of Professor T. A. Wiggins, Pennsylvania State University.

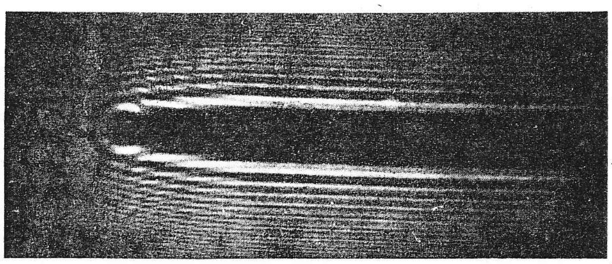
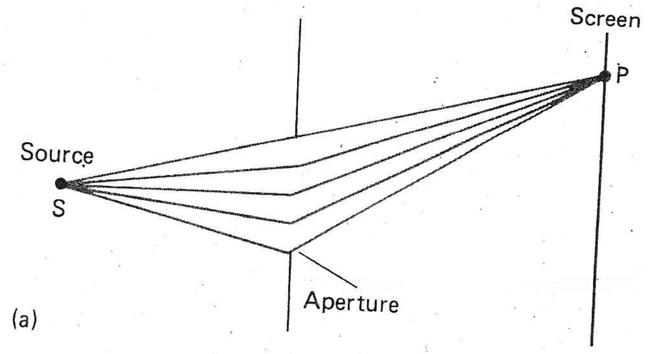


**FIGURE 26.23.** Photograph showing the diffraction of light by a needle. Courtesy of Professor T. A. Wiggins, Pennsylvania State University.

The effect is more pronounced when the wavelength $\lambda$ is comparable to the size of the apertus

For simplicity, we shall assume point monochromatic sources

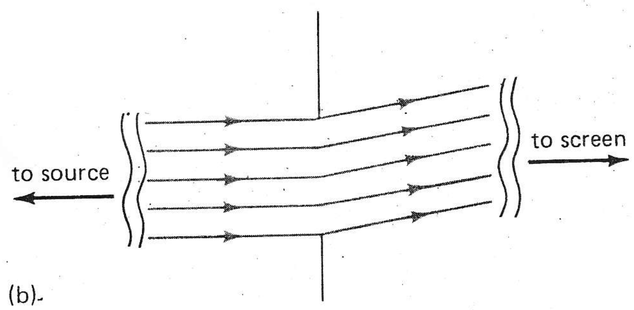
## Fresnel Diffraction




(a)

The wavefront entering and leaving the diffracting aperture are spherical rather than planar

## Fraunhofer Diffraction




The wavefront entering and leaving the apertus are planar

$\Rightarrow$ mathematical much simpler

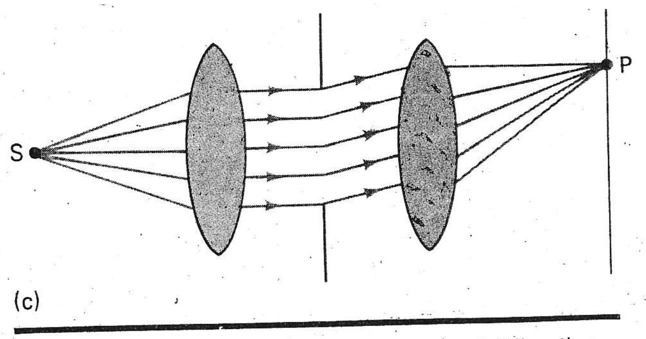
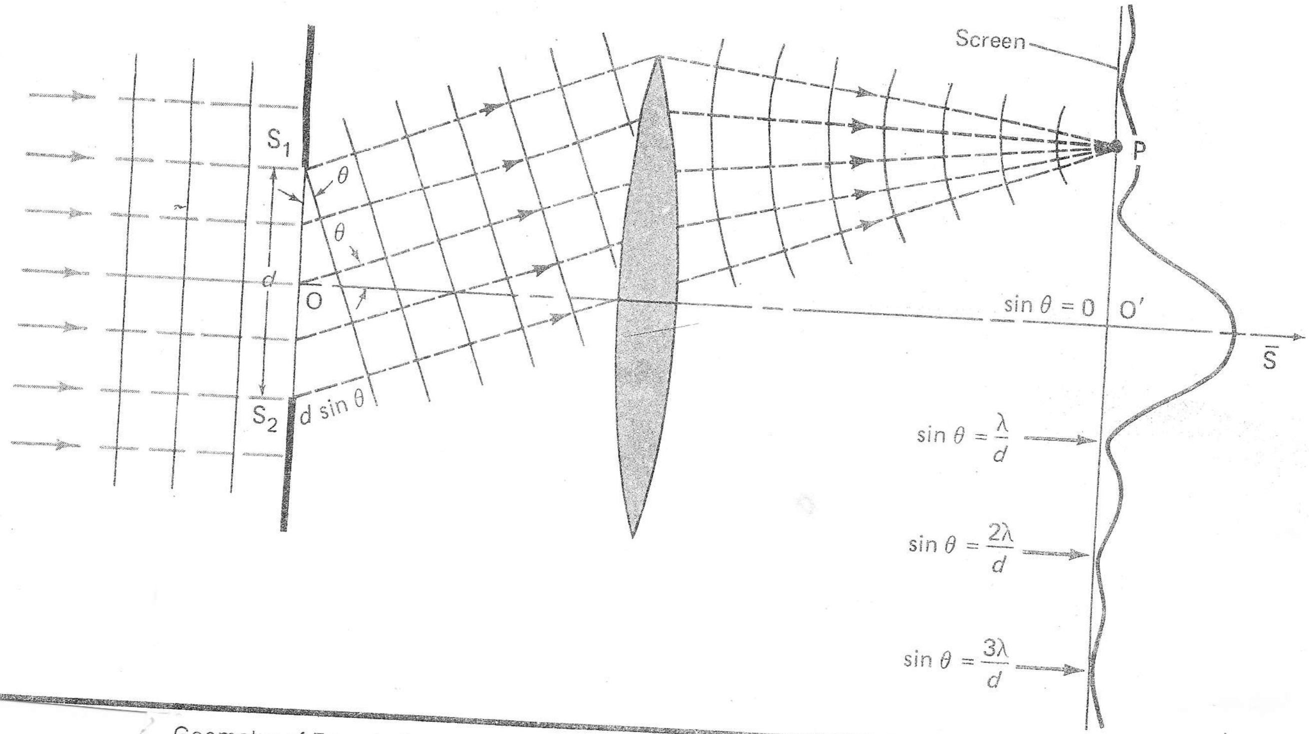



**FIGURE 26.24.** (a) General case of Fresnel diffraction, in which wavefronts at the diffracting aperture are spherical. (b) Simpler case of Fraunhofer diffraction, in which wavefronts at the diffracting aperture are planes, corresponding to source and screen at an infinite distance. (c) Arrangement of lenses used to produce the conditions of Fraunhofer diffraction with finite source and screen distances.

# Single - Slit Fraunhofer Diffraction




Geometry of Fraunhofer diffraction, as arranged in Fig. 26.24c, showing ray paths and wavefronts.

$$u_1(x,t) = A_0 \sin(kx-\omega t)$$

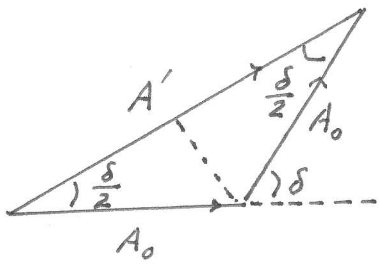
$$u_2(x,t) = A_0 \sin(kx-\omega t)$$

$$u(x,t) = A_0 [\sin(kx-\omega t) + \sin(kx-\omega t-\delta)]$$

$$= \underbrace{2A_0 \cos\frac{\delta}{2}}_{A'} \sin(kx-\omega t - \frac{\delta}{2})$$

$$A' = 2A_0 \cos\frac{\delta}{2}$$

$$I' = 4\cos^2\frac{\delta}{2}\, I_0$$




$$A'^2 = A_0^2 + A_0^2 - 2A_0^2\cos\delta$$

$$= 2A_0^2(1-\cos\delta)$$

$$= 4A_0^2 \cos\frac{\delta}{2}$$

The basic ideal of phasor

# Phasor

$$A_1 \cos(\omega t + \theta_1) + A_2 \cos(\omega t + \theta_2)$$

$$= A_3 \cos(\omega t + \theta_3)$$

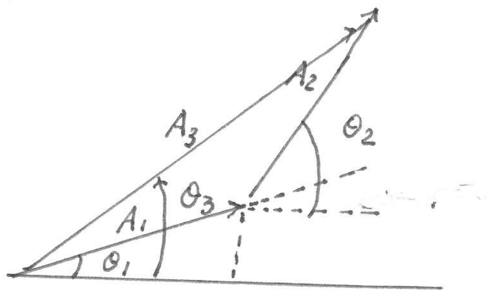
$$A_3^2 = (A_1 \cos\theta_1 + A_2 \cos\theta_2)^2 + (A_1 \sin\theta_1 + A_2 \sin\theta_2)^2$$

$$\theta_3 = \tan^{-1} \frac{A_1 \sin\theta_1 + A_2 \sin\theta_2}{A_1 \cos\theta_1 + A_2 \sin\theta_2}$$

Phasor Diagram

$\downarrow$

Graphic method $\rightarrow$ it is just a mathematical tool




Intensity $\propto A_3^2$

$\downarrow$

$A_1^2 + A_2^2 - 2A_1 A_2 \cos\delta$

$\downarrow$

obviously the result here can be generalized.

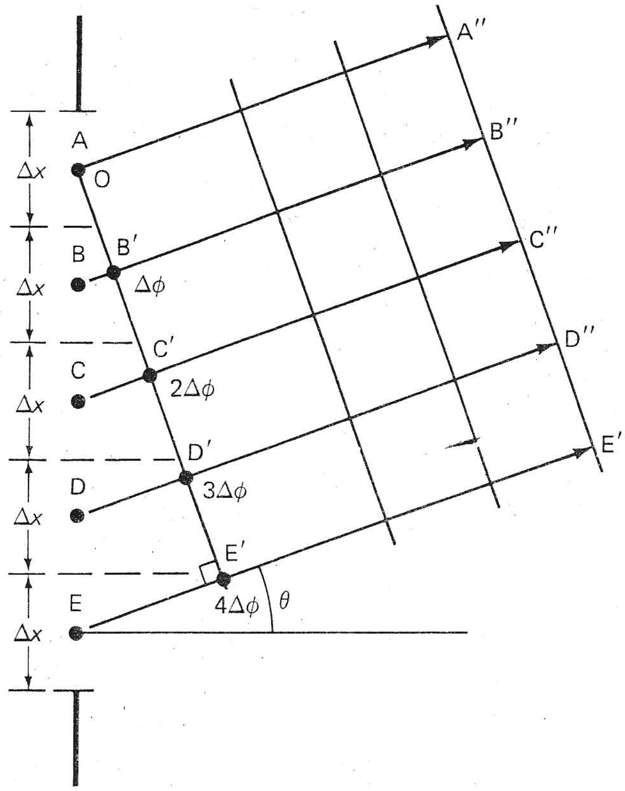



**FIGURE 26.26.** Subdivision of diffracting aperture into $N$ equal parts, each of width $\Delta x$. In the drawing, $N = 5$.

aperture up into $N$ parts, each of width $\Delta x$. We may then superpose the waves that originate from each of these sources, assigning to each of these contributions a phase angle $\phi$ that corresponds to the phase appropriate to the center of each of the $N$ intervals. Finally, we shall let $N$ become infinite while $\Delta x$ approaches zero in such a way that the product

$$N\,\Delta x = d \tag{26.6.2}$$

which represents the width of the slit, remains constant. The situation is illustrated in Fig. 26.26 for the case where $N = 5$; in this case, the points A, B, C, D, and E are regarded as point sources that contribute Huygens wavelets forming plane wavefronts that travel in the direction specified by the angle $\theta$. It is apparent that there must be a phase difference $\Delta\phi$ between each of the rays AA'', BB'', CC'', DD'', and EE'' due to the successively greater path distances BB', CC', DD', and EE' that each ray must cover.[6]

The fields produced at P by each of these contributions are harmonically varying quantities of the same frequency each of which has a different phase. We have already shown in section 9.4 that the magnitude and phase of the sum of harmonically varying quantities such as this can be obtained by regarding each of the individual quantities as a *vector* whose length represents the amplitude and whose direction represents the phase of the individual quantity. Their *vector sum* then represents the result of superposing all the individual harmonically varying components, its amplitude giving the amplitude of the resultant superposition and its phase angle giving the appropriate phase. We have found these rules for superposing sinusoidally varying quantities useful in discussing the voltages and currents in ac circuits, and they are equally useful in the present situation. The result of superposing the five amplitudes discussed in Fig. 26.26 is illustrated in Fig. 26.27a. Had we chosen to subdivide the aperture more finely into many more constitutent parts, the diagram would be as shown in

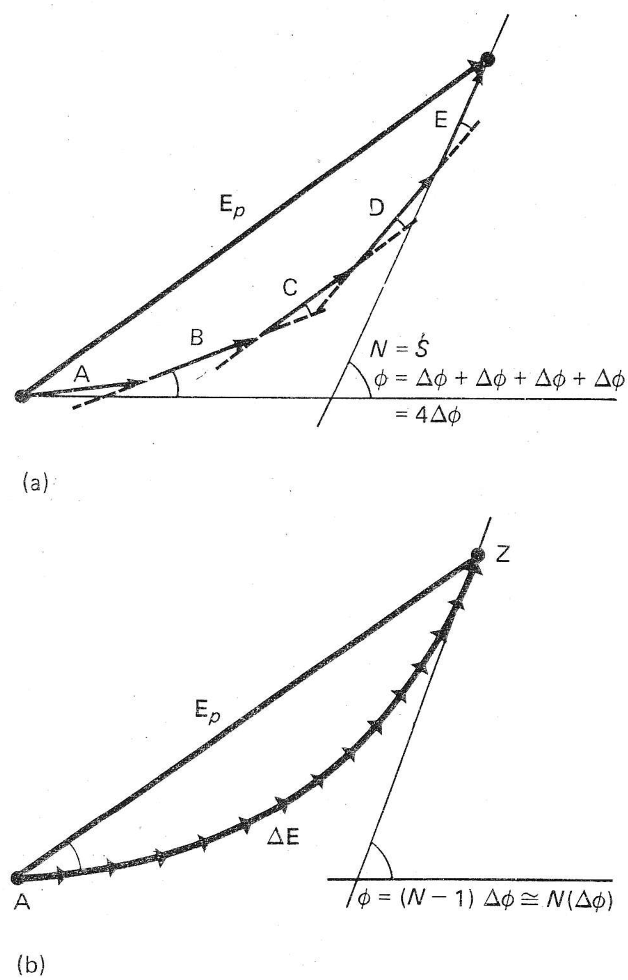



**FIGURE 26.27.** Superposition of harmonically varying fields at the screen using the laws of vector addition (a) for $N = 5$ and (b) for more minute subdivision, corresponding to a large value of $N$.

[6] It is important to note that according to Fermat's principle, the distances $A''P$, $B''P$, $C''P$, $D''P$, and $E''P$ are traversed *in the same time* by all the rays, even though the lens happens to be there. There is, therefore, no difference in the effective optical paths between those points even though a lens is interposed to focus the light. The path differences $BB'$, . . . $EE'$ are, therefore, the only ones that need be discussed.

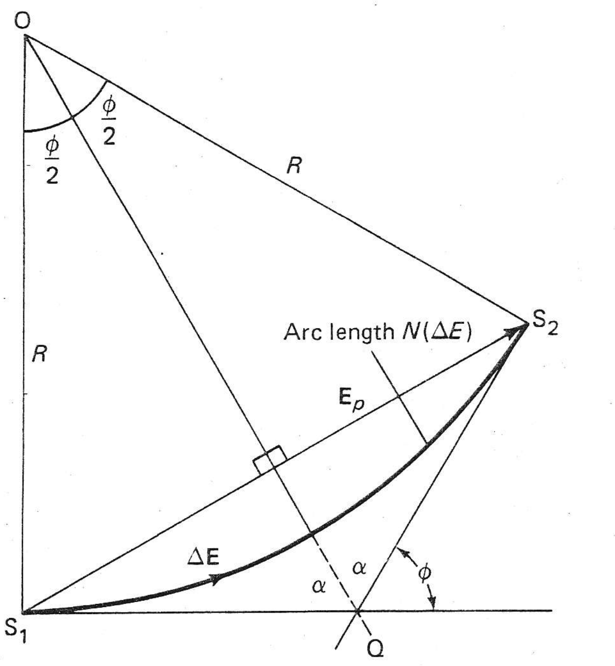



**FIGURE 26.28.** Limiting case of infinitesimal subdivision, in which the **ΔE** vectors describe a circular arc.

Fig. 26.27b. It is evident that as the number of subdivisions $N$ increases without limit, the constituent vectors will describe the arc of a *circle*, since a circle is the curve whose direction changes at a uniform rate as you traverse its arc.[7] From Fig. 26.27, it is also apparent that the total phase difference $\phi$ given by (26.6.1) will be $\phi = (N - 1)\,\Delta\phi$. But in the limit where the number of subdivisions $N$ becomes infinitely large and the phase difference $\Delta\phi$ between adjacent subdivisions infinitesimally small, this may just as well be written

$$\phi = N\,\Delta\phi \tag{26.6.3}$$

Suppose now that we denote the magnitude of the electric field arising from each of the subdivisions $\Delta E$. It is evident, then, from Fig. 26.27b that in the limit of large $N$, the arc length of the circular segment AZ will approach $N\,\Delta E$. At the same time, the vector **AZ**, representing the resultant of all the individual vectors **ΔE**, will represent the electric field of the light wave reaching P in amplitude and phase. The total phase difference $\phi$ of Eq. (26.6.1) is represented by the direction of the last ΔE vector, whose head rests on Z. In the limit, this is the direction of the *tangent* to the circular arc at Z. The limiting situation, therefore, is as represented in Fig. 26.28. In that diagram, it is apparent that the angles $\phi$ and $\alpha$ are related by

[7] In other words, you can drive a car around a perfectly circular track without touching the steering wheel once you have it properly aimed. This is clearly the situation illustrated by the vector diagrams in Fig. 26.27.

$$\phi + 2\alpha = 180^\circ$$

$$\alpha = 90 - \frac{\phi}{2} \tag{26.6.4}$$

Since $OS_1Q$ and $OS_2Q$ are both right triangles, this means that the vertex angles $S_1OQ$ and $S_2OQ$ are both equal to $\phi/2$, as illustrated.

Since the length of the circular arc connecting $S_1$ and $S_2$ is $N\,\Delta E$, we may write

$$R\phi = N\,\Delta E \tag{26.6.5}$$

Also, from elementary trigonometry, the half-length of the vector $\mathbf{E}_p$ that represents the amplitude of the resultant wave at the point P on the screen is

$$\frac{E_p}{2} = R \sin\frac{\phi}{2} \tag{26.6.6}$$

Expressing the radius $R$ in terms of $\phi$ and $\Delta E$ by (26.6.5), this can be written

$$E_p = N\,\Delta E\,\frac{\sin(\phi/2)}{\phi/2} \tag{26.6.7}$$

Now, as the number of parts $N$ into which the aperture is subdivided increases, it is evident that the strength $\Delta E$ associated with each of them must decrease proportionally. This means that $\Delta E$ is inversely proportional to $N$, or that the quantity $N\,\Delta E$ remains *constant* as $N$ is varied. If we denote this constant quantity, which ultimately governs the total intensity of the observed diffraction pattern, as $E_0$, (26.6.7) takes the form

$$E_p = E_0\,\frac{\sin(\phi/2)}{\phi/2} \tag{26.6.8}$$

In this equation, the quantity $E_p$ refers to the amplitude of a harmonically varying electric field at the point P. The time variation of this field can be expressed by incorporating a factor $\cos(\omega t + \delta)$ on the right side of the equation. Since the *intensity* $\overline{S}$ is proportional to the time average of the *square* of the amplitude, it can be represented as

$$\overline{S} = S_0\left(\frac{\sin(\phi/2)}{\phi/2}\right)^2 \tag{26.6.9}$$

In this expression, the quantity $S_0$ represents the time average of $E_0{}^2\cos^2(\omega t + \delta)$, which turns out to be $\frac{1}{2}E_0{}^2$, since the time average of $\cos^2(\omega t + \delta)$ is $\frac{1}{2}$. The trigonometric functions that appear in (26.6.8) and (26.6.9) above are plotted in Fig. 26.29. The angle $\phi$ represents the total difference in phase for light arriving at P between waves starting at the upper and lower ends of the slit, $S_1$ and $S_2$, in Fig. 26.25. This phase difference may be expressed in terms of the slit width $d$, the wavelength $\lambda$, and the angle $\theta$ (in terms of which the location of point P can be

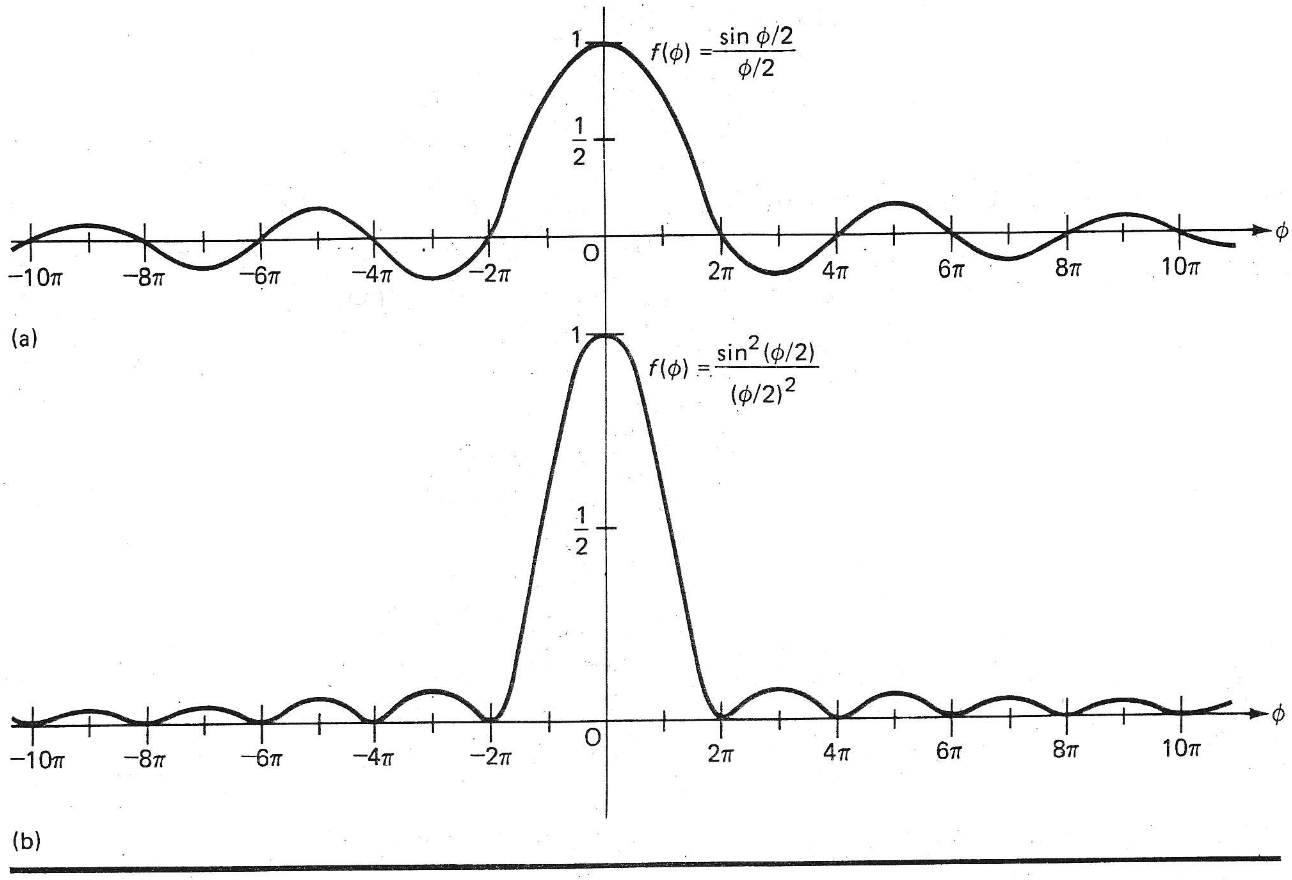



**FIGURE 26.29.** Plot of (a) the field amplitude and (b) the light intensity at the screen, as a function of the variable $\phi = (2\pi d/\lambda) \sin \theta$, for the case of single-slit Fraunhofer diffraction.

expressed) by (26.6.1).[8] Since $\phi$ increases as the angular displacement $\theta$ of point P from the center of the screen increases, there will be a pattern of intensity maxima and minima on the screen, whose locations correspond to the maxima and minima of the function $\sin^2(\phi/2)/(\phi/2)^2$, as plotted in Fig. 26.29b. This pattern has a number of important and rather unusual features which we shall now examine in detail.

First, from Fig. 26.29b, it is easy to see that the maximum intensity occurs at $\phi = 0$, corresponding to $\theta = 0$ from (26.6.1). At this point, which is at the center of the screen, directly opposite the slit, the intensity is $S_0$, since the limiting value of the function $\sin(\phi/2)/(\phi/2)$ is unity as $\phi$ approaches zero. As the angle $\theta$ increases, $\phi$ becomes larger also, and when $\sin \theta = \lambda/d$, $\phi$ attains the value $2\pi$ radians, and the intensity as given by (26.6.9) is zero. Under these conditions, the phase difference between $S_1$ and $S_2$ in Figure 26.25 is 360°. The phase difference between $S_1$ and O is, therefore just half this value or 180°. Light entering the slit at $S_1$ and arriving at P now *interferes destructively* with light entering the slit at O and arriving at P. In fact, light entering *anywhere* above O will interfere destructively with light entering at a point a half-slit width lower along $S_1S_2$, since the phase difference between any two such points is 180°. In this way, we can easily see that when $\sin \theta = \lambda/d$, light entering the upper half of the slit interferes destructively with light coming through the lower half. For this value of $\theta$, therefore, the light intensity on the screen is zero.

If the angle $\theta$ is increased further, the interference of light entering various parts of the slit is no longer totally destructive, and there is, therefore, a resultant intensity given by (26.6.9). When $\sin \theta = 2\lambda/d$, however, the total phase difference $\phi$ between $S_1$ and $S_2$ is $4\pi$, and, under these circumstances, total destructive interference again occurs. Now, if we divide the slit into four quarters, we find that all the light entering the first and third quarters interferes destructively, as does all the light entering the second and fourth quarters. As $\theta$ increases still further, a series of alternating intensity maxima and minima are observed,

[8] In this discussion, it is assumed for convenience that the silt width $d$ is considerably larger than the wavelength, so that the quantity $2\pi d/\lambda$ is quite a large number. This condition need not always be satisfied, of course. We shall see later what happens under those circumstances.

corresponding, respectively, to

$$\phi = 2\pi, 4\pi, 6\pi, 8\pi, \ldots$$
or
$$\sin\theta = \frac{\lambda}{d}, \frac{2\lambda}{d}, \frac{3\lambda}{d}, \frac{4\lambda}{d}, \ldots \tag{26.6.10}$$

for the intensity minima, and, approximately,

$$\phi = 3\pi, 5\pi, 7\pi, 9\pi, \ldots$$
or
$$\sin\theta = \frac{3\lambda}{2d}, \frac{5\lambda}{2d}, \frac{7\lambda}{2d}, \frac{9\lambda}{2d}, \ldots \tag{26.6.11}$$

for the maxima.

One of the more unexpected features of diffraction can be exhibited by calculating the angle $\theta_m$ between the $m$th minimum of intensity and the center of the pattern. From (26.6.10), this is

$$\sin\theta_m = \frac{m\lambda}{d} \tag{26.6.12}$$

In this expression, the slit width $d$ is in the denominator. Therefore, when the slit is very wide, $\theta_m$ will be an extremely small angle, and the diffraction pattern will occupy only a very narrow region near the center of the screen at O′ in Fig. 26.25. Under such circumstances, the intensity minima and maxima may be so closely spaced that they are not readily observed, and the only light that is easily seen is the large central maximum at $\theta = 0$, as illustrated in Fig. 26.30a. The pattern then looks very much like the geometric shadow image of the slit, since there is a lot of light hitting the screen at $\theta = 0$ and very little anywhere else. This result is not unexpected and certainly not very exciting.

But now, suppose we make the slit width $d$ smaller and smaller. Then according to (26.6.12), the angle $\theta_m$ will become increasingly *larger*, the diffraction pattern spreading out to occupy an appreciable area of the screen, as shown at (b) in Fig. 26.30. Under these conditions, the intensity maxima and minima are rather easily observed, although the pattern is much less bright because there is now less total energy coming through the slit and it is spread over a larger part of the screen.

Finally, if $d$ becomes less than $m\lambda$, $m\lambda/d$ will exceed unity, and there will be no real value of $\theta_m$ for which Eq. (26.6.12) is satisfied. The $m$th minimum then disappears, leaving only a pattern in which there are $m - 1$ minima between the center of the pattern and the edge of the screen. As the slit becomes narrower still, the pattern spreads out more and more and successive minima disappear from the screen until, finally, when $d$ becomes smaller than $\lambda$, the first minimum vanishes also. Then, only a small part of the pattern near the central maximum of intensity remains, but it is spread out all over the screen, from $\theta = -90°$ to $\theta = +90°$, as shown in Fig. 26.30c. Ultimately, when the slit width is very much less than the wavelength, only a small region in the neighborhood of the central maximum fills the whole screen, which is then practically uniformly, though very dimly, illuminated. You will recall, that when we studied double-slit interference in section 26.3, we assumed that each of our two slits was very narrow in comparison with the light wavelength. The reason for that assumption is now understandable, since if that condition is not satisfied, the intensity variations on the screen will result not only from interference between the two slits but also from the *diffraction* of light passing through each of them.

This curious effect, in which the diffraction pattern spreads out to occupy an increasingly large area on the screen as the slit width is reduced, is

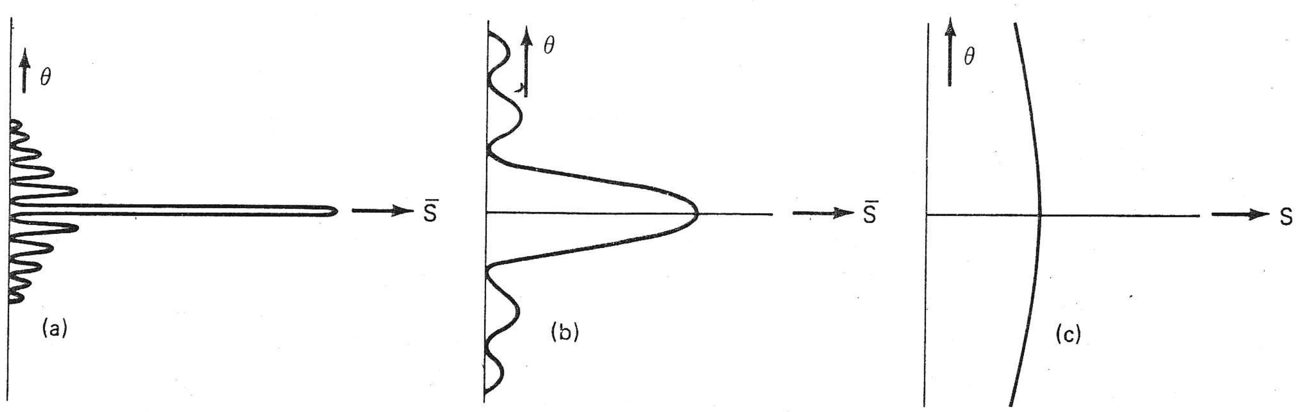



**FIGURE 26.30.** Successive stages in the appearance of the single-slit Fraunhofer diffraction pattern as the slit width is decreased. Note that as the slit becomes narrower, the pattern becomes wider.

characteristic of diffraction in general. It occurs not only for light diffracted by a slit but also for light diffracted by an aperture of any shape or for light diffracted by an opaque obstacle rather than an aperture. It is this effect that eventually frustrates our attempts to improve the sharpness of fuzzy pinhole images by making the pinhole size smaller and smaller.

In concluding the discussion of the diffraction of light by a single slit, it is appropriate to examine the difference between this situation and the double-slit interference phenomena discussed in section 26.3 in connection with Young's experiment. In the former case, we were concerned with the *interference* of light coming from *two separate coherent sources*. These sources were required to be so small that the diffraction effects arising from light emitted by different parts of each source would not be observable. In the present case, only a *single* slit source is involved, but its dimensions are such that light coming through different parts of the slit may interfere differently at each point on the screen. In both cases a series of alternating maxima and minima of light intensity is observed. In the case of double slit interference, however, the maxima all have essentially the same intensity, while for single-slit diffraction, the intensity of the maxima fall off rapidly as the distance from the central maximum increases, as illustrated in Fig. 26.29b. Also, for double-slit interference, the spacing of the fringes is determined by the distance between the two slit sources, while in the case of diffraction, the distance between the striations is determined by the width of the single slit source. It is important to remember that although diffraction is a phenomenon that arises from interference, it is not synonymous with interference. Diffraction of light is the interference of light rays that originate from different parts of an aperture or that come from different locations in the neighborhood of an opaque obstacle. It can, therefore, be regarded as a special case of the more general phenomenon of interference.

EXAMPLE 26.6.1

A Fraunhofer diffraction pattern is observed using light of wavelength 5500 Å. If the slit width is $2.5 \times 10^{-4}$ cm, find the angles for which maximum and minimum intensity occurs. Find the ratio of the intensities of the fourth and first maxima beyond the central maximum.

The zero-order or central maximum occurs for $\theta = 0$. The other maxima are found, according to (26.6.11), for

$$\sin\theta = \frac{3\lambda}{2d}, \frac{5\lambda}{2d}, \frac{7\lambda}{2d}, \ldots$$

corresponding to $\phi = 3\pi, 5\pi, 7\pi, \ldots$. From this, we find

$$\sin\theta = \frac{(3, 5, 7, \ldots)(5.5 \times 10^{-5})}{(2)(2.5 \times 10^{-4})} = (3, 5, 7, \ldots)(0.110)$$

The possible values of $\theta$ corresponding to *maxima* are, therefore,

$$\sin\theta_1 = (3)(0.110) = 0.330 \qquad \theta_1 = 19.27^\circ$$
$$\sin\theta_2 = (5)(0.110) = 0.550 \qquad \theta_2 = 33.37^\circ$$
$$\sin\theta_3 = (7)(0.110) = 0.770 \qquad \theta_3 = 50.35^\circ$$
$$\sin\theta_4 = (9)(0.110) = 0.990 \qquad \theta_4 = 81.89^\circ$$

For minima, (26.6.10) tells us that we must have

$$\sin\theta' = \frac{\lambda}{d}, \frac{2\lambda}{d}, \frac{3\lambda}{d}, \ldots$$

corresponding to $\phi = 2\pi, 4\pi, 6\pi, \ldots$. The possible values of $\theta$ for minimum intensity are, therefore,

$$\sin\theta' = \frac{(1, 2, 3, \ldots)(5.5 \times 10^{-5})}{(2.5 \times 10^{-4})} = (1, 2, 3, \ldots)(0.220)$$

or

$$\sin\theta_1' = (1)(0.220) = 0.220 \qquad \theta_1' = 12.71^\circ$$
$$\sin\theta_2' = (2)(0.220) = 0.440 \qquad \theta_2' = 26.10^\circ$$
$$\sin\theta_3' = (3)(0.220) = 0.660 \qquad \theta_3' = 41.30^\circ$$
$$\sin\theta_4' = (4)(0.220) = 0.880 \qquad \theta_4' = 61.64^\circ$$

For the first maximum, corresponding to $\theta = \theta_1 = 19.27^\circ$, the total phase difference $\phi$ is $3\pi$; for the fourth maximum, at $\theta = \theta_4 = 81.89^\circ$, it is $9\pi$. From Eq. (26.6.9), the intensities at these two points must be

$$\bar{S}_1 = S_0\left[\frac{\sin(3\pi/2)}{(3\pi/2)}\right]^2 = (2/3\pi)^2 S_0$$

and

$$\bar{S}_4 = S_0\left[\frac{\sin(9\pi/2)}{(9\pi/2)}\right]^2 = (2/9\pi)^2 S_0$$

The intensity ratio for the two maxima is, therefore,

$$\frac{\bar{S}_4}{\bar{S}_1} = \left(\frac{2}{9\pi}\frac{3\pi}{2}\right)^2 = \frac{1}{9}$$

EXAMPLE 26.6.2

The sodium $D_1$ and sodium $D_2$ spectral lines have wavelengths of approximately 5896 and 5890 Å. A sodium lamp sends incident plane waves onto a slit of width $2 \times 10^{-4}$ cm. A screen is located 3 meters from the slit. Find the spacing between the first maxima of the two sodium lines, as measured on the screen.

Using Eq. (26.6.11) for each of the two wavelengths, we find

$$\sin\theta_1 = \frac{(3)(5.896 \times 10^{-5})}{(2)(2 \times 10^{-4})} = 0.44220$$

$$\sin \theta_1' = \frac{(3)(5.890 \times 10^{-5})}{(2)(2 \times 10^{-4})} = 0.44175$$

From which

$\theta_1 = 0.45805$ rad or $26.2443°$

$\theta_1' = 0.45755$ rad or $26.2156°$

The vertical distances $h_1$ and $h_1'$ from the central maximum are approximately given by

$h_1 \cong 3 \tan \theta_1 = 1.4791$ m $= 147.91$ cm

$h_1' \cong 3 \tan \theta_1' = 1.4772$ m $= 147.72$ cm

Therefore,

$\Delta h = h_1 - h_1' = 0.19$ cm

The Na lines, which are in the yellow portion of the visible spectrum, were first resolved by Josef Fraunhofer.

## 26.7 Multiple-Slit Diffraction

We have now studied interference due to two very narrow slits and also diffraction due to a single slit. In this section, we shall study the interference pattern established by two slits each of which produces a diffraction pattern. We shall also extend the discussion to a multiple-slit system known as a diffraction grating. The double-slit pattern will be derived explicitly, but the case of a grating will be discussed only in a qualitative way.

In Fig. 26.31, a plane wave with wavelength $\lambda$ is incident on a two-slit system. The slits have widths $a$ and are separated by a distance $d$. To determine the electric vector at a point P on a distant screen, we assume that every point on the wavefront acts as a source of secondary waves, and we superpose all of these using graphical methods. Let $\mathbf{E}_0$ be the electric vector which would be present at $\theta = 0$ if one of the slits were covered. In Fig. 26.28, we showed how to superpose graphically the field vectors due to a single slit. For a double slit, we simply continue the addition process. However, referring to Fig. 26.31, we must realize that the wave sent out from point A differs in phase from that at point B by the amount $(2\pi/\lambda)(AB)\sin\theta$, where the distance $AB$ is $d - a$. Referring now to Fig. 26.32, we first obtain the resultant electric field $\mathbf{E}_l$ from the lower slit by vector addition, as in Fig. 26.28. This vector makes an angle of $\phi/2$ with the horizontal. To obtain the second vector $\mathbf{E}_u$, we again begin the process of adding infinitesimal vectors, but the first small vector (emitted at A) now has a phase of

$$\delta = \frac{2\pi}{\lambda}(d - a)\sin\theta \qquad (26.7.1)$$

with respect to the small vector from B. The length of $\mathbf{E}_u$ is the same as that of $\mathbf{E}_l$, but according to the figure, the *angle* between the two vectors is seen to be $\phi + \delta$. Let $\beta = \phi/2$ and $\gamma = (\phi + \delta)/2$. Then, since the vectors $\mathbf{E}_l$ and $\mathbf{E}_u$ both have the same magnitude $E_l$, from Fig. 26.32b it is evident that

$$E_p = 2E_l \cos\left(\frac{\phi}{2} + \frac{\delta}{2}\right) \qquad (26.7.2)$$

But since $E_l$ can be expressed in terms of $E_0$ by (26.6.8), just as in the preceding section, this can be written as

$$E_p = 2E_0 \frac{\sin(\phi/2)}{\phi/2} \cos\left(\frac{\phi}{2} + \frac{\delta}{2}\right)$$

or, more simply,

$$E_p = 2E_0 \frac{\sin\beta}{\beta} \cos\gamma \qquad (26.7.3)$$

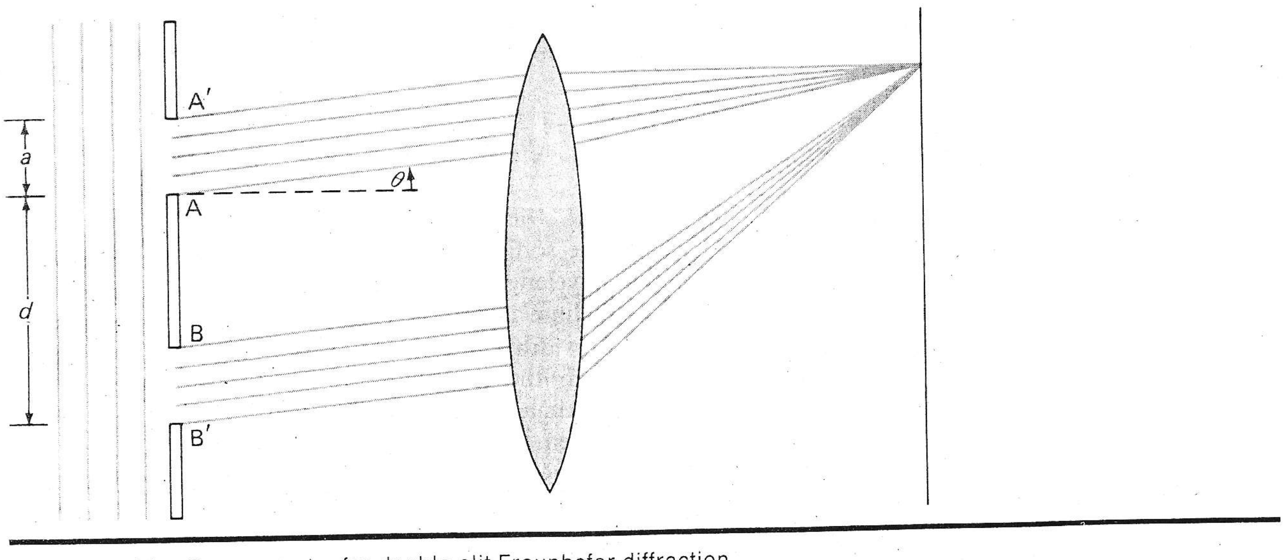



**FIGURE 26.31.** Ray geometry for double-slit Fraunhofer diffraction.

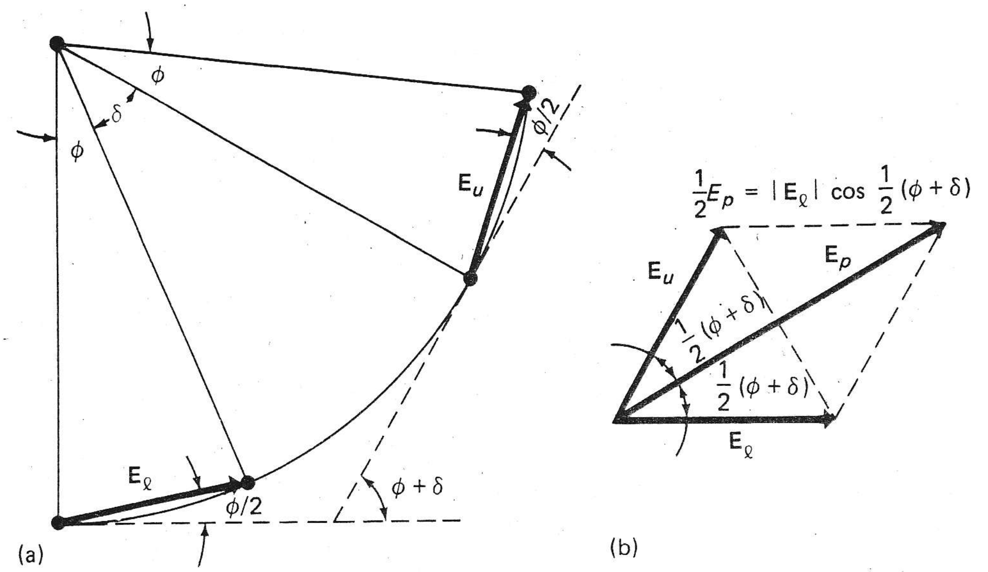



**FIGURE 26.32.** Phase diagram for double-slit Fraunhofer diffraction.

where, from (26.6.1),

$$\beta = \phi/2 = (\pi/\lambda)a \sin \theta \tag{26.7.4}$$

$$\gamma = \frac{\phi}{2} + \frac{\delta}{2} = (\pi/\lambda)d \sin \theta \tag{26.7.5}$$

The factor $(\sin \beta)/\beta$ characterizes single-slit diffraction, while $\cos \gamma$ occurs in double slit interference. In the present case, both factors are present. Therefore, the relative intensity, which we may write as

$$\bar{S} = 4S_0\left(\frac{\sin \beta}{\beta}\right)^2 \cos^2 \gamma \tag{26.7.6}$$

contains variations typical of both. Figure 26.33 illustrates the resulting pattern for the case in which $d = 4a$, which implies that $\gamma = 4\beta$. The $\cos^2$ function has a maximum whenever $\gamma = 0, \pi, 2\pi, \ldots$. These occur whenever

$$d \sin \theta = m\lambda \qquad m = 0, 1, 2, 3, 4, \ldots \tag{26.7.7}$$

At these angles, we have so-called *principal maxima* in the intensity pattern. The integer $m$ denotes the order of the pattern, and $m = 0, 1, 2, 3$, etc., corresponds to the zeroth, first, second, third maxima, etc. For double-slit interference, which involves extremely narrow slits, for which $(\sin^2 \beta)/\beta^2 = 1$, these maxima would all be of uniform intensity. However, when the slit width is nonzero, additional changes in intensity can occur due to the $(\sin^2 \beta)/\beta^2$ factor. For example, whenever, as in (26.6.10), $\phi/2 = \beta = \pi, 2\pi, 3\pi, \ldots,$

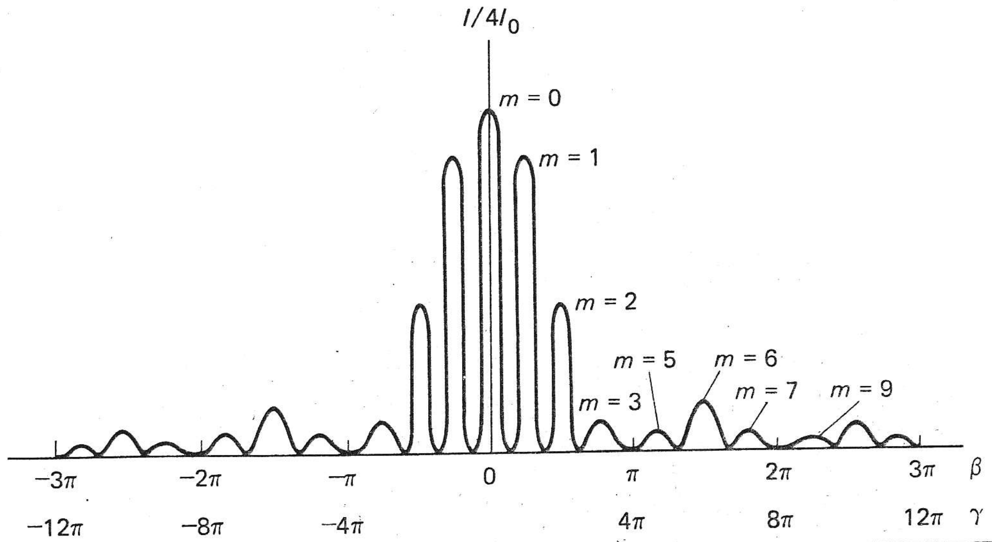



**FIGURE 26.33.** Double-slit Fraunhofer diffraction intensity pattern for the case where $d = 4a$, corresponding to $\gamma = 4\beta$.

we will have an intensity minimum, since $\sin\beta$ then vanishes. From (26.7.4), it is seen that this occurs for

$$a \sin\theta = n\lambda \qquad n = 1, 2, 3, \ldots \tag{26.7.8}$$

The pattern may be described by saying that the $(\sin^2\beta)/\beta^2$ factor *modulates* the $\cos^2\gamma$ variations. Thus, we see in Fig. 26.33 that the various interference maxima are contained within an envelope corresponding to the diffraction factor $(\sin^2\beta)/\beta^2$

EXAMPLE 26.7.1

A double slit is illuminated by light for which $\lambda =$ 4500 Å. If the two slits are each 9000 Å in width and are separated by a distance of 27,000 Å, find all angles $\theta$ at which the intensity is zero. Find also the angles at which an intensity maximum occurs.

Apart from the central maximum, the intensity will be zero when *either* $\sin\beta$ or $\cos\gamma$ vanishes. Since $a = 2\lambda$ and $d = 6\lambda$, this will occur for

$$\beta = (\pi a/\lambda)\sin\theta = 2\pi\sin\theta = \pi, 2\pi, 3\pi, \ldots$$

and for

$$\gamma = (\pi d/\lambda)\sin\theta = 6\pi\sin\theta = \pi/2, 3\pi/2, 5\pi/2, \ldots$$

Thus, there are minima when

$$\sin\theta = 0.5 \text{ or } 1.0 \qquad \theta = 30^\circ \text{ and } 90^\circ$$

and

$$\sin\theta = \tfrac{1}{12}, \tfrac{3}{12}, \tfrac{5}{12}, \tfrac{7}{12}, \tfrac{9}{12}, \tfrac{11}{12}$$
$$\theta = 4.78^\circ, 14.48^\circ, 24.62^\circ, 35.69^\circ, 48.59^\circ, 66.44^\circ$$

The maxima will occur whenever $\cos^2\gamma = 1$ *unless* $\sin\beta$ vanishes at one of these maxima. If this happens, we say that a maximum has been *suppressed*. Since $\cos^2\gamma = 1$ implies that

$$\gamma = (\pi d/\lambda)\sin\theta = 6\pi\sin\theta = 0, \pi, 2\pi, \ldots$$

we have

$$\sin\theta = 0, \tfrac{1}{6}, \tfrac{2}{6}, \tfrac{3}{6}, \tfrac{4}{6}, \tfrac{5}{6}, 1$$
$$\theta = 0^\circ, 9.59^\circ, 19.47^\circ, 30^\circ, 41.81^\circ, 56.44^\circ, 90^\circ$$

We see that the maxima at $\theta = 30^\circ$ and $90^\circ$ are indeed suppressed since there are minima of $\sin\beta$ at these angles. More generally, we have

$$d\sin\theta = m\lambda \qquad m = 0, 1, 2, \ldots$$

giving the $m$th maximum. But

$$a\sin\theta = n\lambda \qquad n = 1, 2, \ldots$$

gives a minimum of zero. Thus, whenever

$$\left(\frac{d}{a}\right)n = m$$

a maximum of the two slit interference pattern is eliminated. In the present case, for which $d/a = 3$, the third and sixth orders are missing.

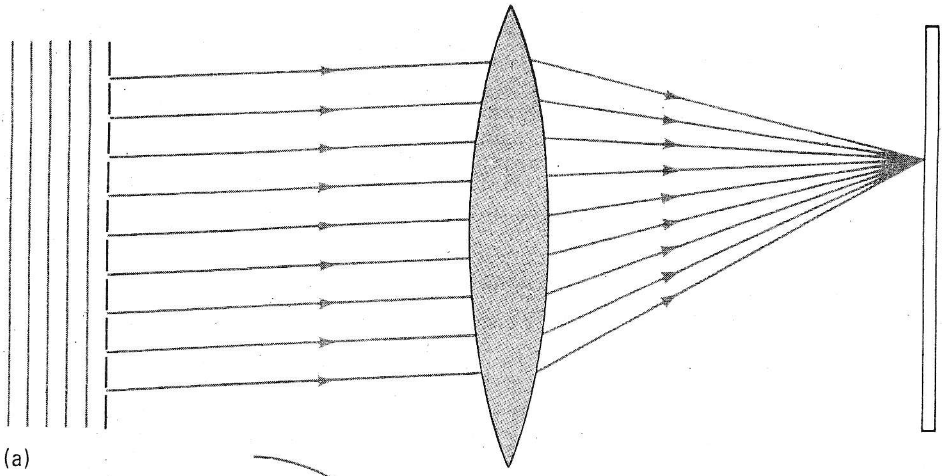
Let us now study in a qualitative way the interference pattern produced by a *multiple-slit* system, also known as a *diffraction grating.* Figure 26.34a is a schematic representation of a grating. We assume that there are $N$ narrow slits, each of width $d$ and separated by a distance $a$. A typical grating might have 6000 or more slits per centimeter. It is no small accomplishment to construct such a grating. Fraunhofer made some early gratings by winding fine wire very closely on two fine screws. He was also able to rule gratings on glass, although this could then not be done with great accuracy. Sixty years after Fraunhofer's pioneering work, the American physicist Henry Rowland constructed a "ruling engine" for cutting gratings on an aluminum coating deposited on glass. A diamond cutting edge was used to produce very straight grooves. In this way, successful and practical diffraction gratings were first obtained.

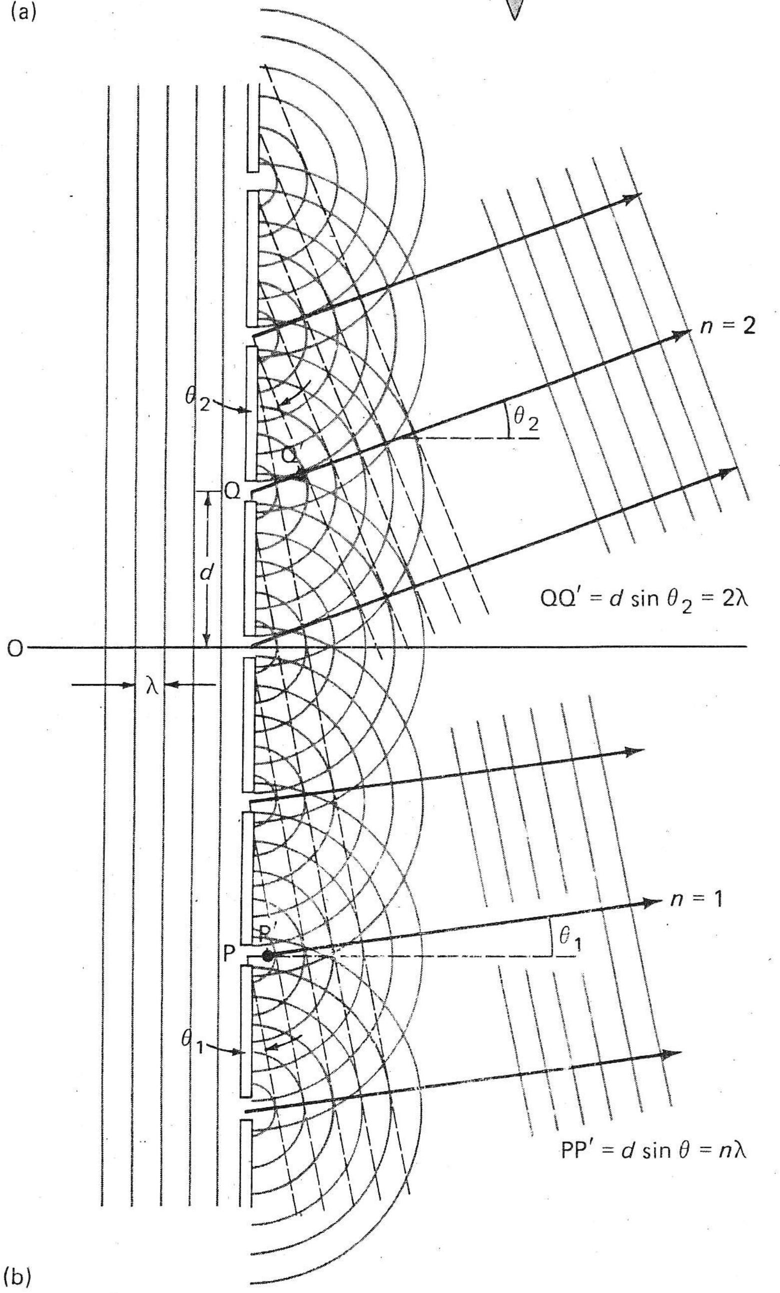
To understand the physical principles that govern the behavior of a diffraction grating, let us first consider the simple system shown in Fig. 26.34b. In this system, there are a very large number of very narrow regularly spaced slits. The distance between neighboring slits, as shown in the diagram, is $d$. The slit width in this particular example is assumed to be *very much less than the light wavelength,* so that diffraction effects arising from the passage of light through apertures of finite width are unimportant. We are, therefore, concerned in this specific case with what might be more accurately referred to as an *interference* grating. The grating is illuminated with normally incident plane monochromatic light waves of wavelength $\lambda$. We may then regard each slit as a point source of cylindrical Huygens wavelets whose wavefronts are illustrated in the diagram.

Let us now consider how outgoing plane wavefronts might be formed from the system of Huygens wavelets on the far side of the grating. It is, of course, possible to draw a set of tangent planes parallel to the incoming waves and to the grating which represent plane waves that propagate horizontally to the right beyond the grating. These wavefronts do actually exist but contribute only to a continuation of the incident wave. It is possible also to draw a set of wavefronts that are tangent to Huygens wavelets from neighboring slits whose phase *differs* by an entire wavelength, as shown by the dashed lines in the lower part of the figure. These wavefronts propagate at an angle $\theta_1$ to the horizontal. Since the path difference $PP'$ between succeeding wavefronts is $d\sin\theta_1$ and since this must be an entire wavelength, we may determine the direction of propagation, $\theta_1$, from the condition

$$d\sin\theta_1 = \lambda \tag{26.7.9}$$

In much the same way, a series of wavefronts propagating at a larger angle $\theta_2$ to the horizontal may be constructed by drawing tangents connecting wavelets from neighboring slits that are out of phase by two


(a)



(b)

**FIGURE 26.34.** (a) Ray diagram for Fraunhofer diffraction by a diffraction grating. (b) Geometry of rays and wavefronts for diffraction by a grating.

wavelengths. These wavefronts are shown as the set of dashed lines in the upper part of Fig. 26.34b. Since the path difference $QQ'$ between successive wavelets is now twice the wavelength, and since $QQ' = d \sin \theta_2$, we may write for this set of outgoing waves

$$d \sin \theta_2 = 2\lambda \tag{26.7.10}$$

We can in the same manner find outgoing waves in other directions, $\theta_3, \theta_4, \theta_5, \ldots$, corresponding to larger integral path differences between neighboring slits. Their propagation directions are clearly given by

$$\boxed{d \sin \theta_n = n\lambda} \tag{26.7.11}$$

where the integer $n$ identifies the so-called *order* of the outgoing diffracted wave.

These outgoing "diffracted" beams strike a distant screen at various distances above the axis $OO'$ of the grating. They can, in fact, be focused into sharp lines of light by a converging lens placed between grating and screen, as shown in Fig. 26.25. By measuring these distances, or by measuring the angles $\theta_1, \theta_2, \theta_3, \ldots$, it is possible to measure the wavelength $\lambda$ of the incoming light very accurately, provided that the distance $d$ between slits is precisely known. Also, since light of different wavelength will be diffracted at different angles, a grating acts to produce a *spatial separation between light of different colors*, in somewhat the same way as a dispersive prism.

In this example, we have assumed that the slit width is very much smaller than the light wavelength so that the effects that we observe are really those of interference rather than diffraction. In actual gratings, this assumption is hardly ever justified; and for this reason, though the outgoing beams propagate in precisely the same directions, as defined by (26.7.11), their intensities may be strongly affected by effects of diffraction associated with slits of finite width.

To derive the intensity pattern due to a grating with finite slits of width $a$, we would follow the previous procedure of superposing infinitesimal electric fields to find the field due to each slit. Then we would add vectorially the fields due to each slit, taking into account the phase differences between each of these fields. For a double slit, we found that the intensity could be expressed by (26.8.3). This can be rewritten as

$$\bar{S} = 4S_0\left(\frac{\sin \beta}{\beta}\right)^2 \cos^2 \gamma = S_0\left(\frac{\sin \beta}{\beta}\right)^2\left(\frac{\sin 2\gamma}{\sin \gamma}\right)^2 \tag{26.7.12}$$

since

$$\sin 2\gamma = 2 \sin \gamma \cos \gamma$$

The generalization of this expression to the case of $N$ slits is a complex mathematical task which we shall not attempt here. The result, however, is simple enough and is obtained simply by replacing $\sin 2\gamma$ by $\sin N\gamma$ in Eq. (26.7.12). Thus, for a diffraction grating having $N$ slits, the intensity is given by

$$\bar{S} = S_0\left(\frac{\sin \beta}{\beta}\right)^2\left(\frac{\sin N\gamma}{\sin \gamma}\right)^2 \tag{26.7.13}$$

where $S_0$ is the intensity at $\theta = 0$ and where $\beta$ and $\gamma$ are defined by (26.7.4) and (26.7.5), respectively. The factor $(\sin^2 \beta)/\beta^2$ is due to single-slit *diffraction*, while $(\sin N\gamma/\sin \gamma)^2$ is present due to the *interference* of $N$ slits.

Let us now discuss the principal features of (26.7.13) to establish the dependence of the intensity on the angle $\theta$. We shall first examine the interference factor. Whenever $\sin N\gamma$ vanishes, it appears, at first sight, that the intensity $I$ should be zero. There is an exception to this, however; for if $\sin \gamma$ also vanishes, we have an indeterminate ratio. In fact, it turns out that when both of these factors vanish, $(\sin N\gamma/\sin \gamma)^2$ attains a *maximum* rather than a minimum. To see this, we note that whenever $\gamma = n\pi$, $\sin \gamma$, and $\sin N\gamma$ will both vanish. Let us, therefore, take the limit of $\sin N\gamma/\sin \gamma$ as $\gamma \to n\pi$. We can do this by using l'Hôpital's rule from calculus to obtain

$$\lim_{\gamma \to n\pi} \frac{\sin N\gamma}{\sin \gamma} = \lim_{\gamma \to n\pi} \frac{\dfrac{d}{d\gamma}\sin N\gamma}{\dfrac{d}{d\gamma}\sin \gamma} = \lim_{\gamma \to n\pi} \frac{N \cos N\gamma}{\cos \gamma} = \pm N \tag{26.7.14}$$

Thus, whenever $\gamma$ is an integral multiple of $\pi$, we have a so-called *principal maximum* of the pattern, and a factor of $N^2$ coming from the $N$-slit interference factor is present. It might happen, of course, that $\sin \beta$ vanishes or is small at such a principal maximum. In this event, the principal maximum will be suppressed due to the diffraction term.

Whenever $N\gamma$ is an integral multiple of $\pi$ but $\gamma$ *is not* a multiple of $\pi$, the intensity will vanish since $\sin N\gamma$ is zero and $\sin \gamma$ is nonzero. For angles which satisfy this condition, a *minimum* intensity is observed. For large $N$, this implies a very large number of subsidiary maxima and minima between any two principal maxima because of the rapid variation of the function $\sin N\gamma$ in Eq. (26.7.13). These subsidiary maxima, in gratings with many slits, are of very low intensity in comparison to the principal maxima and ordinarily are not even visible.

The effect of all this is that when the number of slits becomes very large, the principal maxima become very intense and the subsidiary maxima very weak. Since the principal maxima occur when $\gamma = n\pi$ and since $\gamma$ is defined by (26.7.5), the principal maxima occur when

$$n\pi = (\pi d/\lambda) \sin \theta$$

$$\xi_1 = A_1 \sin(\omega t - kr_1)$$

$$\xi_2 = A_2 \sin(\omega t - kr_2)$$

$$\xi = \xi_1 + \xi_2$$

$$= A_1 \sin(\omega t - kr_1) + A_2 \sin(\omega t - kr_2)$$

$$\xi^2 = A_1^2 \sin^2(\omega t - kr_1) + 2A_1A_2 \sin(\omega t - kr_1)\sin(\omega t - kr_2)$$
$$+ A_2^2 \sin^2(\omega t - kr_2)$$

$$\underbrace{\sin(\omega t - kr_1)}_{A} \sin(\underbrace{\omega t - kr_2}_{B})$$

$$= \tfrac{1}{2}[\cos(A - B) - \cos(A + B)]$$

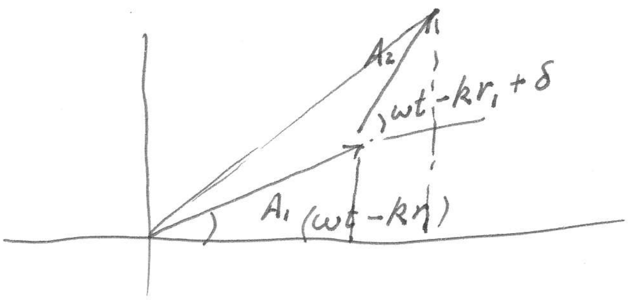
$$= \tfrac{1}{2}[\cos[\underbrace{\omega t - kr_1 - \omega t + kr_2}_{\delta}] - \cos(2\omega t - kr_1 - kr_2]$$

Average over time $T \to$ period

$$\xi^2 = \tfrac{1}{2}A_1^2 + \tfrac{1}{2}2A_1A_2 \cos\delta + \tfrac{1}{2}A_2^2$$

$$A^2 = \tfrac{1}{2}[A_1^2 + A_2^2 + \tfrac{1}{2}A_1A_2\cos\delta]$$

$$A_1 \sin(\omega t - kr_1) + A_2 \sin(\omega t - kr_2)$$



through the primary slit affects the phase of the light emitted by the two secondary slits in exactly the same way and at exactly the same time.

If the two secondary slits are illuminated by two independent sources, even though they are "monochromatic" in the sense described above, there is no phase relationship between the wavefronts emitted by the two secondary slits that persists for any appreciable length of time. There is, accordingly, no stable interference pattern, and the screen appears to be more or less evenly illuminated.

### 3. SLITS ARE NO LONGER NARROW IN COMPARISON WITH THE LIGHT WAVELENGTH

If the slit width is not narrow, there will be not only interference of light waves from the two separate slits, but also interference of light coming from *different portions of a single slit*. When this happens, the interference pattern becomes much more complicated, and will be discussed in a subsequent section under the subject of *diffraction*.

### 4. THE SECONDARY SLITS ARE CLOSE TO THE SCREEN ON WHICH THE PATTERN IS OBSERVED

The interference pattern is still present, but the geometry is now such that the spacing of the fringes must be determined from Eqs. (26.3.14) and (26.3.15) rather than from the approximations which follow from the condition $d \ll D$ used earlier. Also, the approximations that the two fields $\mathbf{E}_1$ and $\mathbf{E}_2$ from the two separate slits are the same in both magnitude and direction may no longer be justified.

In all the discussion above, the entire analysis is based on the underlying notion that light must be described as a wave. We have, of course, made use of the harmonically oscillating electric field as the underlying entity used to determine the intensity of light at any point. But when Young carried out his experiments, he did not know that such a connection existed between light and electricity and magnetism. Indeed, Maxwell had not yet been born! It is very much to Young's credit that he was able to advance the bold hypothesis that light is a wave phenomenon and to assert that the effect of *interference* provided the proof.

In spite of Young's remarkable success, most of his contemporaries publicly ridiculed his ideas, for English physicists were biased in favor of the corpuscular views advanced by Newton a century earlier. Even though Newton himself had suggested that the true nature of light must ultimately be decided by experiment, the critics of Young were apparently not convinced by his accomplishments. This seems strange

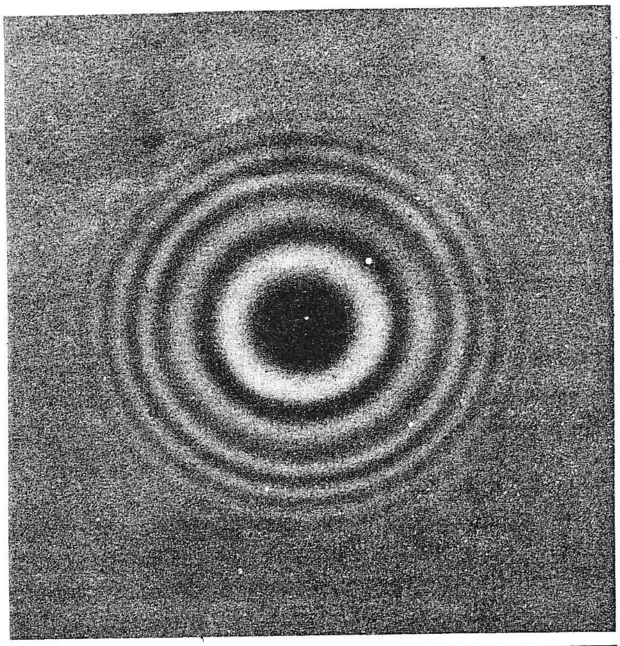


FIGURE 26.12. Photograph of Newton's rings. Photo courtesy of Professor T. A. Wiggins, Pennsylvania State University.

now, because if we assume that light consists of a stream of particles that travel in straight lines along the direction of the light rays of geometrical optics, in the experimental arrangement illustrated in Fig. 26.4 there could be no interference pattern at all, and the entire screen would be dark. In 1815, however, fourteen years after Thomas Young discovered the interference of light, the French engineer Fresnel explained diffraction patterns on the basis of a wave theory of light by using Huygens's principle. He also made some theoretical predictions which were experimentally confirmed by the French physicist Dominique Arago. Arago became a convert to the ideas of Young and Fresnel, and the wave theory of light became generally accepted soon thereafter.

## 26.4 Interference in Thin Films

The subject of thin-film optics dates back to the work of Isaac Newton, who observed fringes produced by the thin air film between a convex lens and a plane surface. These *Newton rings*, shown in Fig. 26.12, provide an early example of thin-film interference. Of course, Newton was unaware of the wave character of light, but Thomas Young was able to explain these rings on the basis of his ideas about wave interference.[3] There are many naturally occurring examples of *thin-*

[3] In a strict sense, the observation of Newton's rings could be held to justify the point of view that it was Newton rather than Young who really discovered the interference of light.

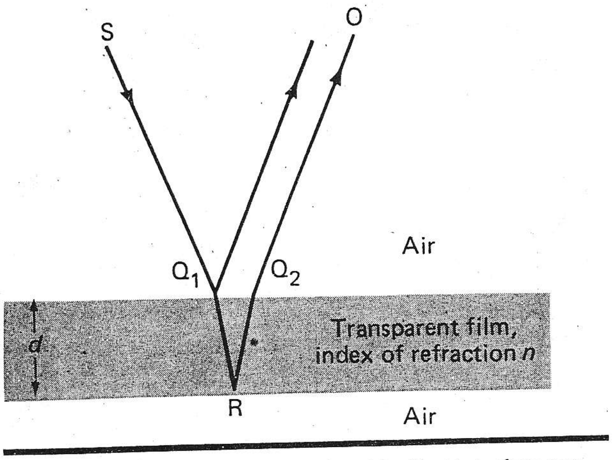



**FIGURE 26.13.** Ray diagram for thin-film interference.

*film interference* of this type. All of us have, at some time or other, seen the beautiful colors reflected from an oil film on water or from a soap bubble, or the colored reflections from the "coated" lenses of cameras or binoculars. These are also illustrations of thin-film interference.

Let us consider an example in which a source of light with a dominant wavelength $\lambda$ in air illuminates a reflecting surface upon which a thin transparent film, with index of refraction $n$, has been deposited. We shall further assume that the incident light rays are *almost normal* to the surface, as are the reflected rays arriving at an observer located at O, as illustrated in Fig. 26.13. Upon reflection, such rays are split into two beams; one is reflected directly from the surface of the film at $Q_1$, and one enters the film and is reflected from the underlying substrate, emerging at $Q_2$. It is the interference between two such light beams that leads to the observed thin-film optical effects. If the light has a wavelength $\lambda$ in air, the wavelength in the film will be $\lambda' = \lambda/n$. To determine whether or not the two light rays interfere constructively or destructively at O, we must consider the phase difference between the two rays arriving there. There are two factors that can lead to phase differences. First, there is a path difference $Q_1RQ_2$ between the two rays. Assuming that the incident and outgoing rays are normal to the surface, the phase change associated with this path difference is

$$\frac{2\pi}{\lambda'} Q_1RQ_2 = \frac{2\pi}{\lambda/n} Q_1RQ_2 \tag{26.4.1}$$

Also, there are possible phase changes that take place upon reflection at the reflecting surfaces, as discussed in Chapter 24. If light is incident upon a reflecting surface from a medium of refractive index $n_1$ and if the refractive index of the material beyond the reflecting surface is $n_2$, then, if $n_2 > n_1$, the reflected wave changes phase by $\pi$, while, if $n_1 > n_2$, no phase change occurs. The above statement is summarized in the short verse[4] by F. K. du Pré, which goes:

Low to High
Phase Change $\pi$
High to Low
Phase Change? No.

In this illustration, we are led to conclude that there is a phase change $\pi$ upon reflection from the upper surface but none for the reflection from the lower surface.

Now, whenever the *total phase difference* between two waves is given by $\delta = 0, 2\pi, 4\pi, 6\pi, \ldots$, the waves will reinforce to produce a maximum intensity. On the other hand, if the difference is $\delta = \pi, 3\pi, 5\pi, \ldots$, the waves will interfere destructively to produce minimum intensity. In the present case, assuming normal incidence, the ray $SQ_1RQ_2O$ undergoes a phase change $4\pi nd/\lambda$ in passing through the film as a result of the optical path difference, while the ray $SQ_1O$ experiences a phase change of $\pi$ as a consequence of reflection from the optically denser medium. The *total phase difference* between waves is, therefore, $(4\pi nd/\lambda) - \pi$, and hence the conditions for maxima and minima are, respectively,

*Maxima*

$$\delta = \frac{4\pi nd}{\lambda} - \pi = 0, 2\pi, 4\pi, \ldots$$

*Minima* (26.4.2)

$$\delta = \frac{4\pi nd}{\lambda} - \pi = -\pi, \pi, 3\pi, 5\pi, \ldots$$

or

*Maxima*

$$d = \frac{\lambda}{4n}, \frac{3\lambda}{4n}, \frac{5\lambda}{4n}, \ldots$$

*Minima* (26.4.3)

$$d = 0, \frac{\lambda}{2n}, \frac{2\lambda}{2n}, \frac{3\lambda}{2n}, \ldots$$

We see, therefore, that for any given wavelength there are many possible film thicknesses which could give rise to maximum or minimum reflected intensities. Since energy must be conserved, we are forced to conclude that under circumstances leading to a *minimum* in the reflected light, there will be a *maximum* in the transmitted light.

Now suppose that instead of using incident monochromatic light we use white light, which consists of all optical frequencies. Then, for a given film thickness, some wavelengths may interfere constructively on reflection while others may interfere destructively. As a consequence, the reflected light

[4] F. K. du Pré, *Appl. Optics* **10**, 2345 (1371).

**FIGURE 26.14.** Interference fringes from a soap film, produced using a monochromatic source. Photo by the authors.

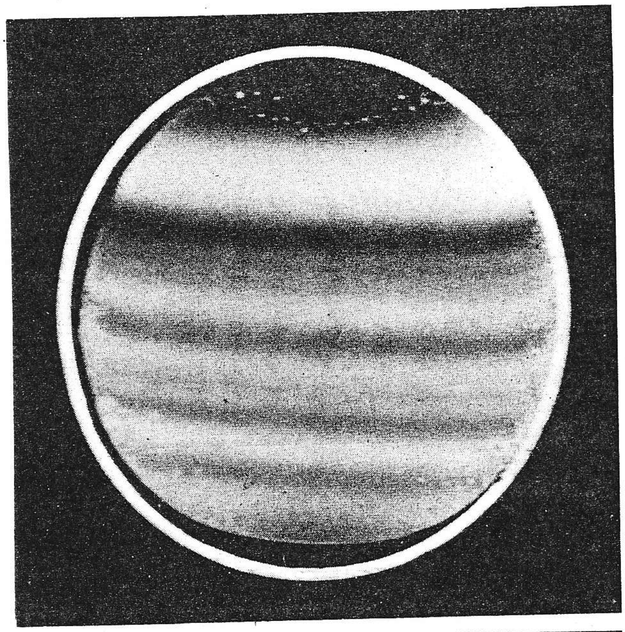


**FIGURE 26.15.** White light interference fringes from a soap film. Photo courtesy of Professor T. A. Wiggins, Pennsylvania State University.

corresponding to wavelengths for which constructive interference occurs will be quite intense, while light of wavelength corresponding to destructive interference will be essentially absent. For this reason, the reflected light appears to be colored, the precise shade depending upon the film thickness and the index of refraction.

Figures 26.14 and 26.15 show photographs of a soap film suspended on a wire ring. The soap film has a variable thickness since it is held in a vertical position and, therefore, drains downward. As a consequence, certain portions of the film satisfy the condition for constructive interference while other portions satisfy the destructive interference criteria. In Fig. 26.14, the soapy film is illuminated with monochromatic light, and a pattern of light and dark bands shows the effect of interference. If white light is used, as in Fig. 26.15, we expect to see constructive interference of various frequencies occurring at different locations. In this way, the soap film separates different spectral hues spatially and exhibits bands of *interference colors.*

In the example just above, we considered the case of a thin film suspended in air, in which case it is bounded both above and below by a less highly refractive medium. It is also important, however, to examine the case illustrated in Fig. 26.16, where the film is supported upon a more highly refractive solid transparent substrate, such as optical glass. Such thin films have been found to be very useful in a number of important practical situations. They can drastically reduce the reflectivity of the underlying transparent medium and can, under certain conditions, even reduce it to zero. They can also make it highly reflecting, if that is what is wanted. Let us now study some of the problems involved in making "coated" lenses and other optical components by applying thin films to their surfaces so as to reduce their reflectivity as much as possible.

As we learned in Chapter 24, the reflection coefficient of a typical glass–air interface for light at

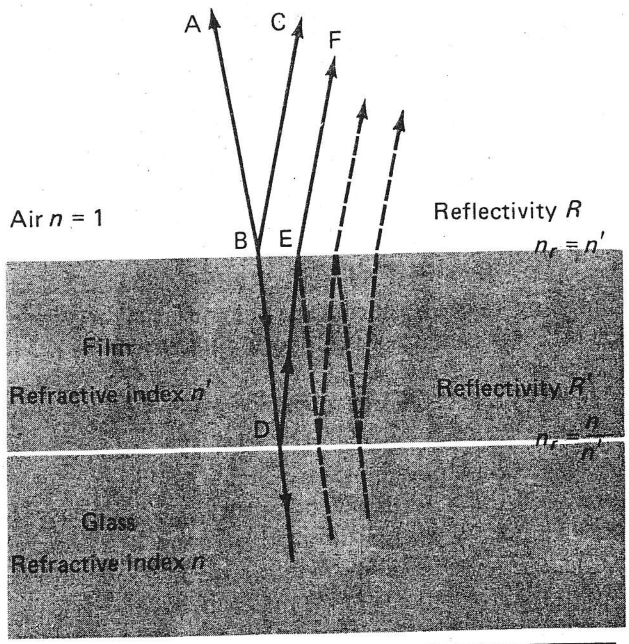



**FIGURE 26.16.** Ray diagram for light reflected from a coated lens.

normal incidence is about 5 percent. In optical systems such as complex photographic lenses, there may be 10 or even 20 such interfaces. Clearly, if we had to put up with a 5 percent reflection loss at each lens surface, we would end up with a lens that would transmit only a small percentage of the light incident on its front surface. For a lens having 10 glass–air surfaces, we would expect a transmission of $0.95^{10} = 0.5987$, or 59.9 percent, since only 95 percent of the light survives reflection at each surface. For a zoom lens having 20 such interfaces, the figure would be only 35.8 percent. It is desirable to reduce this reflection loss by optical coating, not only because of the light loss itself but also because much of the reflected light ends up on the film, or in the eyepiece, in the form of "ghost images." The ghost images reduce image contrast and sometimes produce annoying visible spots and haloes when the lens is aimed in the direction of bright point sources of light. It is, therefore, very desirable to reduce the reflectivity of the glass–air surfaces to the lowest possible value, or to eliminate it completely where that can be done. These objectives are usually realized in practice by applying a thin refracting layer to the glass in such a way that the reflected light from its top and bottom surfaces interferes destructively, as shown in Fig. 26.16.

Let us assume that the film has an index of refraction which is greater than that of air but less than that of glass. Under these conditions, the light rays reflected at the upper and lower surfaces both experience a phase change of $\pi$ radians, since the reflection occurs from an optically denser underlying medium. Thus, the two light rays ABDEF and ABC will have a phase difference which results only from their path difference. For approximately *normal incidence*, the phase difference will, therefore, be $4\pi nd/\lambda$. If this difference is $\pi$, the two waves arrive out of phase. This will happen if the film thickness is chosen to be $\lambda/4n$, in other words, if it is one fourth of the wavelength in the medium.

Now, if the *intensities* of the beams reflected at the upper and lower surfaces are not equal, their destructive interference will be partial rather than complete, and the total reflected intensity will be lowered but not entirely eliminated. If we can somehow arrange to have them exactly equal in intensity, however, there will be complete destructive cancellation and no reflected intensity at all. In Example 24.4.4, we saw that the reflectivity of a glass surface, defined as the ratio of reflected to incident intensity, is given for normal incidence by Eq. (24.4.54). In the present example, there are *two* interfaces whose reflectivity must, according to these results, be

$$R = \left(\frac{n_r - 1}{n_r + 1}\right)^2 \quad \text{and} \quad R' = \left(\frac{n_r' - 1}{n_r' + 1}\right)^2 \tag{26.4.4}$$

for the upper and lower surfaces, respectively. For most transparent substances, these reflectivities are relatively small, and only a few percent of the incident light is reflected at each interface. Under these circumstances, the intensities of the incident ray AB and the transmitted ray BD are practically equal. In such a situation, if we make the reflectivities $R$ and $R'$ the same, the reflected rays BC and DEF that interfere destructively will have practically the same intensities, and there will be no significant reflected light at all. But $R$ and $R'$ will be equal only if the relative indices $n_r = n'$ and $n_r' = n/n'$ are the same, and this, in turn, requires that[5]

$$n' = \frac{n}{n'}$$

or

$$n' = \sqrt{n} \tag{26.4.5}$$

We, must, therefore seek a coating film whose index of refraction is approximately the same as the square root of the index of the glass. For optical glass of refractive index 1.75, we would, therefore, look for a film of refractive index 1.32. It is not easy, in practice, to find materials that have the required refractive index along with all the other desired characteristics of transparency, durability, and ease of deposition. Magnesium fluoride, $MgF_2$, whose refractive index is 1.38, comes reasonably close to satisfying all the requirements mentioned above. A quarter-wave film of $MgF_2$ will reduce the reflectivity of most optical glasses in the visible light range to under 1 percent. If the coating is designed to minimize reflection of green light to which the eye is most sensitive ($\lambda \sim 5500$ Å), there will be, of necessity, some residual reflectance for the red and violet regions of the spectrum, since for these colors the film thickness is no longer very close to $\lambda/4n$. This small residual reflection of red and violet light gives rise to the familiar purplish hue of light reflected from coated lenses. The technology of optical coating is now very highly developed. By coating glass with a number of films whose thicknesses and refractive indices are appropriately chosen, it is possible to reduce the reflectivity of optical glass to very small values over the whole visible spectrum.

In the processes of reflection and refraction, there is *no loss* of light energy at all. Therefore, all

[5] If a significant fraction of the incident light is reflected, of course, as it would be if $n'$ and $n$ were very large, then the transmitted beam's intensity would be much reduced; and simply making the reflectivities $R$ and $R'$ the same no longer suffices, even approximately, to produce complete destructive interference. Under these circumstances, the multiple reflection of beams within the film, shown by dotted lines in Fig. 26.16, also becomes important. Fortunately, most of the optical glasses and coating films used in practice have reflectivities of less than 10 percent.

light that is not reflected from a transparent substance must be transmitted. In the present example, therefore, the process of coating a transparent substance with a film a quarter-wavelength thick not only reduces the reflectivity of the substance but actually *increases* the amount of light that is transmitted. Although the physical details of the processes that act to accomplish this increase in transmission are not easily understood, there is no question that it must take place. In the examples mentioned previously, of a lens with 10 glass–air surfaces and a zoom lens with 20 such surfaces, the application of a quarter-wave film that reduces the reflectivity to 1 percent at each interface increases the fraction of light transmitted to $0.99^{10} = 0.9043$ and $0.99^{20} = 0.8179$, respectively. Comparison of these figures with the ones given previously for uncoated lenses reveals that there is a significant increase in the transmission properties of the lenses. For the zoom lens, the fraction of light transmitted is more than doubled by the coating process. Indeed, the manufacture of zoom lenses, which are often comprised of 15 or 20 separate elements, would be quite out of the question without optical coating.

EXAMPLE 26.4.1

A film 10,000 Å thick is used to coat a certain type of glass. At what wavelengths in the visible spectrum will the reflected light interfere destructively? The index of refraction of the film is assumed to be 1.40, which is less than that of the glass.

In the present example, the condition of a minimum in reflected intensity is that the path difference between the two reflected light rays is an odd multiple of one half the wavelength in the medium. Compensating phase changes occur upon reflection from the air–film and film–glass interfaces. Since the index of the film is not specified to be equal to the square root of the index of the glass, those light wavelengths for which the reflected light is a minimum will be reduced in intensity but not necessarily extinguished. Since the path difference is twice the film thickness $d$ and since this must equal an odd number of half-wavelengths in the medium, we may write

$$2d = \frac{1}{2}\frac{\lambda}{n'}, \frac{3}{2}\frac{\lambda}{n'}, \frac{5}{2}\frac{\lambda}{n'}, \ldots$$

or

$$\lambda = 4n'd, \frac{4n'd}{3}, \frac{4n'd}{5}, \frac{4n'd}{7}, \ldots$$

Substituting the numerical values given above for $n'$ and $d$, we find

$$\lambda = 56{,}000\ \text{Å}, 18{,}667\ \text{Å}, 11{,}200\ \text{Å}, 8{,}000\ \text{Å}, 6{,}222\ \text{Å}$$
$$5{,}091\ \text{Å}, 4{,}308\ \text{Å}, 3{,}733\ \text{Å}, \ldots$$

Of these wavelengths, only 6222 Å, 5091 Å, and 4308 Å are in the visible spectrum, and, therefore, these optical wavelengths are those for which minimum reflectivity is observed. Other wavelengths are more or less strongly reflected.

EXAMPLE 26.4.2

A soap film is formed on a wire frame as shown in Fig. 26.17. The film drains downward due to gravity and is, therefore, thicker at the bottom than at the very top, where it might be only a few molecular layers thick. At a particular instant of time, the film has a thickness of 50 Å near the top, 1500 Å somewhere in the middle, and 4000 Å near the bottom. When the film is viewed with reflected light, what colors would be eliminated at the top, middle, and bottom? Assume the film has an index of refraction of 1.35.

Let us find out which wavelengths will be absent from the reflected light due to destructive interference. From (26.4.3), we should expect intensity minima when $d = \lambda/2n, 2\lambda/2n, 3\lambda/2n, \ldots$, from which

$$\lambda = 2nd, \frac{2nd}{2}, \frac{2nd}{3}, \ldots$$

Now, near the top of the film, where the thickness is only 50 Å, the phase difference between rays reflected from the two surfaces of the film is nearly $\pi$ for *all* visible wavelengths. The phase difference resulting from the optical path difference is negligible in this case since the film is very thin. The top of the film then will reflect very little light of any color and will, therefore, appear quite dark.

Near the middle of the film, the thickness is 1500 Å, and, therefore, destructive interference can occur at wavelength $\lambda$ for which

$$\lambda = 2nd = (1500)(2)(1.35) = 4050\ \text{Å}$$

This is in the violet part of the visible spectrum, and, therefore, violet is not reflected. The other wavelengths for which destructive interference occurs are not in the visible spectrum. Constructive interference occurs,

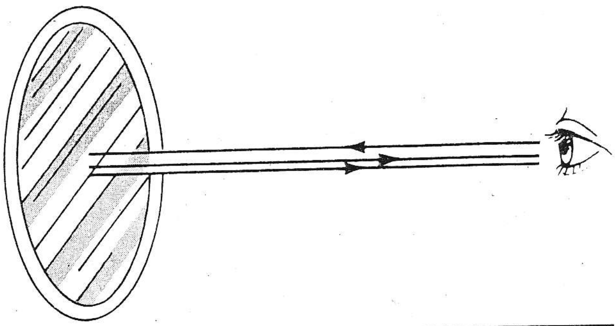


**FIGURE 26.17.** Interfering beams reflected from the surfaces of a soap film.

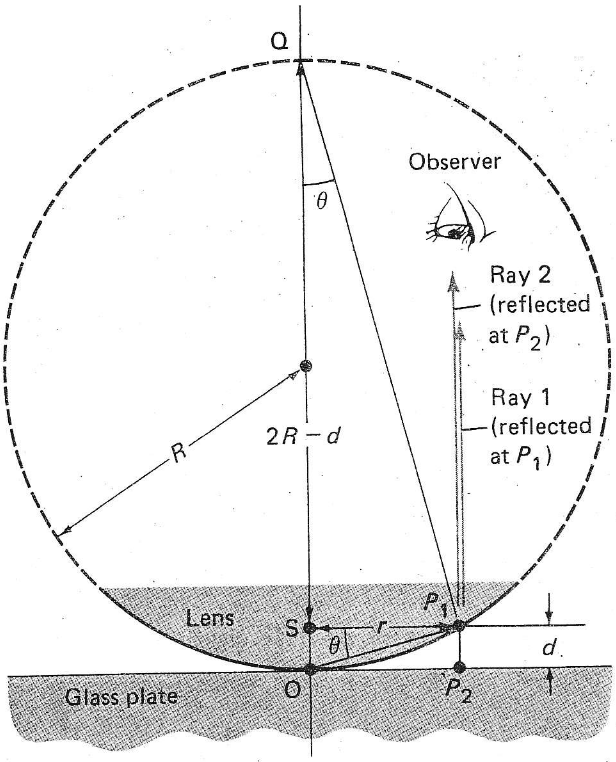



**FIGURE 26.18.** Interference geometry for Newton's rings.

according to (26.4.3), for wavelengths 8100 Å, 2700 Å, and smaller values, none of which is in the visible region.

In the region where the thickness is 4000 Å, destructive interference occurs at

$$\lambda = (4000)(2)(1.35) = 10{,}800 \text{ Å}$$

which is not in the visible range, but also at 10,800/2 = 5400 Å, which is in the green part of the spectrum. None of the other wavelengths at which the reflected waves are canceled is visible. At 4000 Å thickness, maxima can occur at 21,600/1, 21,600/3, 21,600/5, . . . Å. Only one of these, the wavelength 4400 Å, is in the visible spectrum. Therefore, light of this wavelength, which is violet, will be strongly visible on reflection. In general, the exact colors that are seen on reflection depend on the composition of the incoming light, the amount of each wavelength that is reflected, and also the sensitivity of the human eye.

EXAMPLE 26.4.3
NEWTON'S RINGS

An air film is formed between a convex lens and a plane reflecting surface, as shown in Fig. 26.18. The lens is assumed to have a radius of curvature $R$. Interference rings are observed when light of wavelength $\lambda$ is viewed at normal incidence. Find an expression which can be used to determine the location of the circular fringes.

A photograph of these Newton rings is shown in Figure 26.12. To analyze this problem, consider the two light rays designated in Fig. 26.18 as 1 and 2. Since the thickness $d$ of the air film is very small, the circular and plane surfaces are almost parallel, and we may therefore assume that the two reflected rays are also parallel. Ray 2 undergoes a phase change of $\pi$ radians at $P_2$, but ray 1 has no phase change at $P_1$. The phase difference between the rays is therefore $(2\pi/\lambda)(2d) + \pi$, and therefore the conditions for minima in the reflected light are

$$\frac{4\pi d}{\lambda} + \pi = \pi, 3\pi, 5\pi, \ldots$$

or

$$d = 0, \frac{\lambda}{2}, \frac{2\lambda}{2}, \frac{3\lambda}{2}, \ldots$$

Let us now write this condition as

$$d = N\frac{\lambda}{2} \tag{26.4.6}$$

where $N = 0, 1, 2, 3, \ldots$. Now, referring to Fig. 26.18, the properties of the similar triangles $OSP_1$ and $P_1SQ$ lead us to conclude that $OS/SP_1 = SP_1/SQ$, hence that

$$\frac{d}{r} = \frac{r}{2R - d}$$

or

$$r^2 = 2Rd - d^2$$

But since $d \ll R$,

$$r^2 \cong 2Rd = 2RN\frac{\lambda}{2}$$

$$r = \sqrt{NR\lambda} \tag{26.4.7}$$

We see from this that the radii of the destructive interference fringes are proportional to the square root of the fringe order $N$. A formula such as this can be used to find the wavelength of the light if the radius $R$ can be found by measuring the diameter of the rings. Alternatively, if $\lambda$ is known, it may be used to determine $R$.

If the fringes are not circular, this would indicate that the lens surface is not spherical or perhaps that the plane surface is not perfectly flat. Interference fringes are often used to test whether a given optical surface has been ground to the desired curvature. For example, if we are certain that the plane surface is completely flat, then any departure from circular Newton rings would indicate that the lens was not spherical.

In Fig. 26.19, a method for testing whether a surface has been ground optically flat is suggested. If description of the effects that are observed is much simpler than in the more general instance of Fresnel diffraction. Of course, it is inconvenient from an experimental point of view to have both source and observing screen infinitely far from the diffracting system. This is easily remedied, however, by using lenses to convert the diverging rays from a point source at a finite distance into a parallel beam and to focus parallel outgoing rays onto an observing screen at a finite distance, as shown in Fig. 26.24c. In this arrangement, the source must be placed at the focal point of the first lens and the screen at the focal point of the second. The rays that enter the diffracting aperture have no way of "knowing" that they did not originate from an infinitely distant point source. In the same way, a set of parallel rays that leave the aperture headed in a given direction do not know that they are not headed for an infinitely distant screen. This scheme enables us to observe the simpler Fraunhofer diffraction effects very easily. Though Fresnel diffraction is important in many situations, we shall confine ourselves in this book to a discussion of the much easier case of Fraunhofer diffraction.

## 26.6 Single-Slit Fraunhofer Diffraction

The simplest case of Fraunhofer diffraction occurs when a single, narrow slit of width $d$ is illuminated normally by plane monochromatic light waves, as illustrated in Fig. 26.25. We shall try to find the intensity of light leaving the slit and traveling in a direction that makes an angle $\theta$ to the incoming light, as shown in the figure. This light is eventually focused at point P on the screen by the lens. Each point along the line $S_1S_2$ can be regarded as a source of waves, according to Huygens's principle, whose fronts may propagate in the direction specified by $\theta$ and which eventually arrive at P. Unfortunately, however, the path length from source to screen is *different for each point* along $S_1S_2$; for example, the path difference between points $S_1$ and $S_2$ is $d \sin \theta$. For this reason, the light that originates from each part of the slit arrives at P with a *different phase*. It is necessary, therefore, to account properly for these phase relationships in superposing the contribution from each source point. Another way of expressing the essential facts is to observe that light entering different parts of the slits interferes as it arrives at the screen and that this interference may be constructive or destructive. Since from Fig. 26.25 the path length between the two sides of the slit is $d \sin \theta$, the total phase difference $\phi$ between these points must be

$$\boxed{\phi = \frac{2\pi}{\lambda} d \sin \theta} \qquad (26.6.1)$$

To calculate the intensity at point P, we shall find the electric field amplitude $E_p$ associated with the light that arrives there. To do this, we shall split the

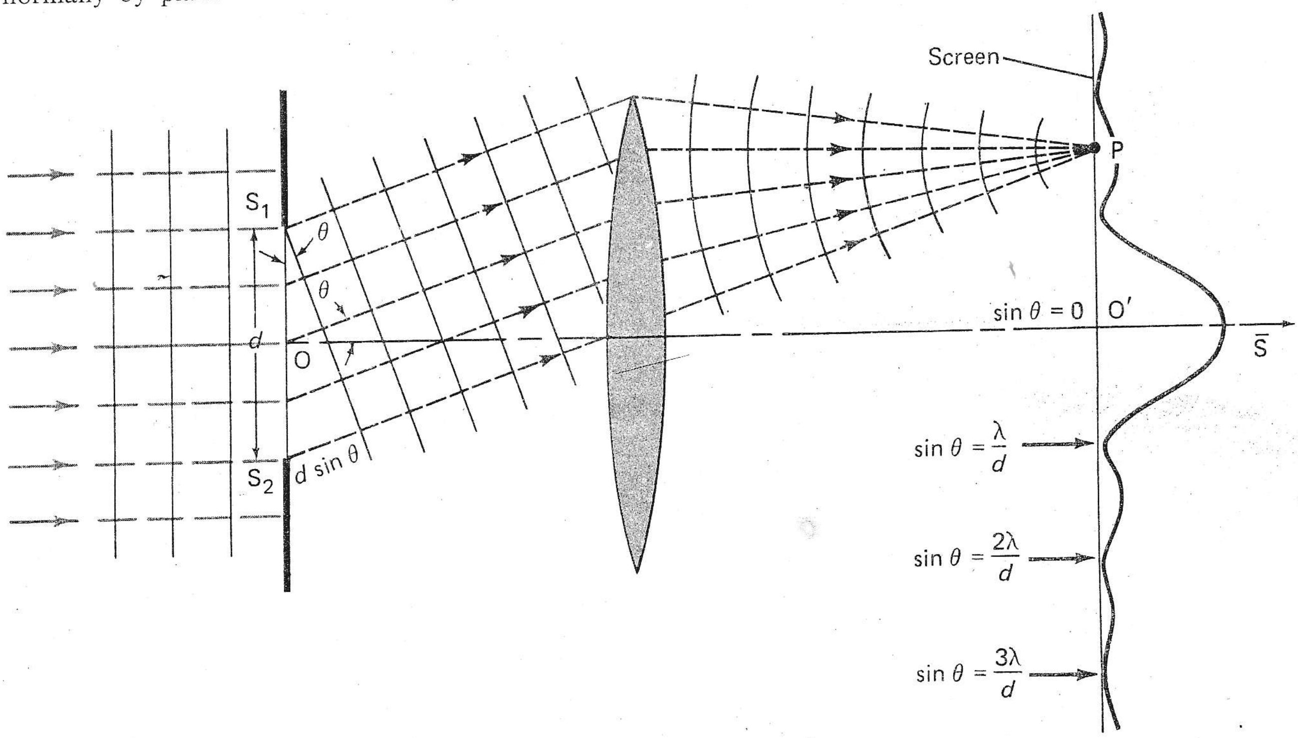



**FIGURE 26.25.** Geometry of Fraunhofer diffraction, as arranged in Fig. 26.24c, showing ray paths and wavefronts.

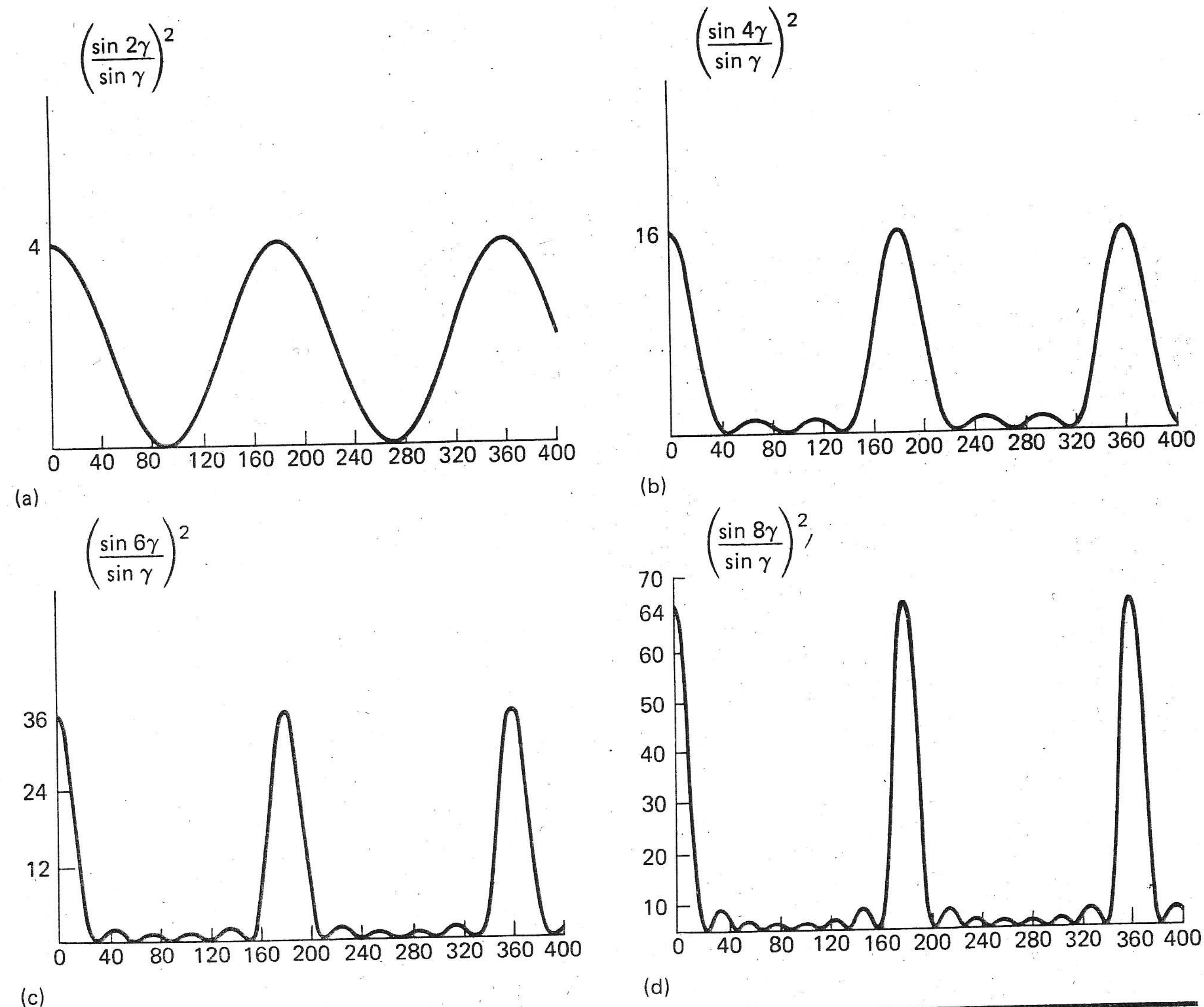



**FIGURE 26.35.** Plots of interference function $(\sin N\gamma/\sin \gamma)^2$ for multislit array with (a) $N = 2$, (b) $N = 4$, (c) $N = 6$, and (d) $N = 8$.

or

$$\boxed{n\lambda = d \sin \theta} \tag{26.7.15}$$

This is identical to (26.7.11), which defines the beams diffracted by a grating having an infinite number of very narrow slits. The principal maxima, therefore, are formed as in the earlier case, except that now their intensity is modulated by the diffraction factor $(\sin^2 \beta)/\beta^2$ in (26.7.13). In Fig. 26.35, the function $(\sin N\gamma/\sin \gamma)^2$ is plotted for several values of $N$ to demonstrate the predominance of the principal maxima and the weakness of the subsidiary maxima that develops as $N$ increases. The same effect is shown in Fig. 26.36, in which the diffraction pattern arising from arrays of various numbers of slits has been photographed. It should be noted that the effect of the diffraction factor $(\sin^2 \beta)/\beta^2$ is to cause an overall decrease in the intensity of the pattern with increasing values of the angle $\theta$. Also, if this factor should happen to be zero at a value of $\theta$ that coincides with one of the principal interference maxima defined by (26.7.15), that particular principal maximum will be suppressed and there will be a "missing order of interference" in the resulting pattern. These effects are illustrated in Fig. 26.37.

The plane diffraction grating we have been discussing has been, and still is, an extremely valuable tool for studying the spectrum of light radiated by atoms and molecules. It has also been used extensively in astronomy to identify the composition of planets and stars, and has been very useful in a wide variety of other laboratory situations.

We have so far discussed the pattern produced by a single wavelength. If more than one wavelength is present in the incident light, each wavelength will be diffracted through different sets of angles as defined