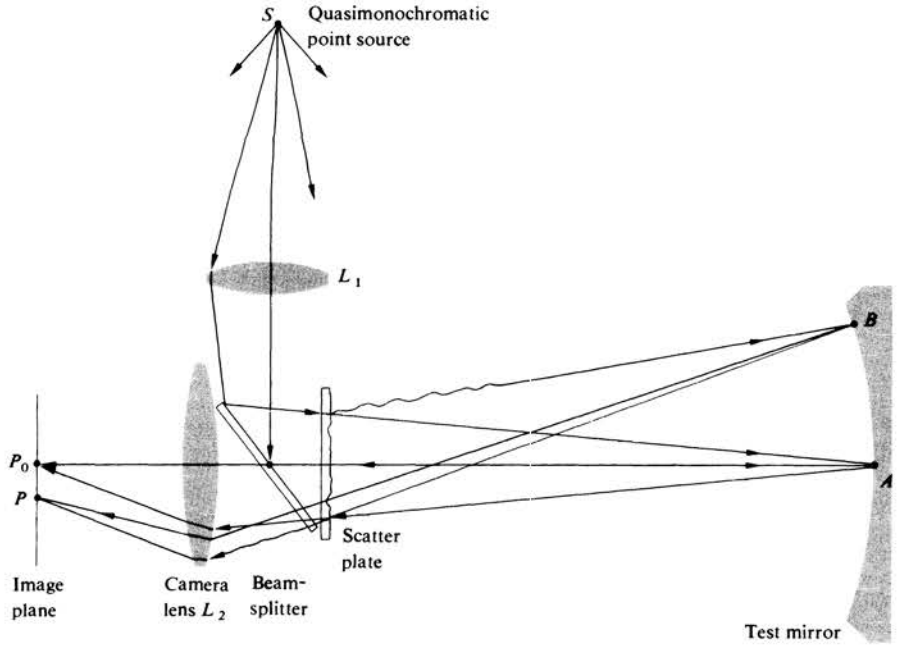


Figure 9.53 Scatter plate setup. Adapted from R. M. Scott, *Appl. Opt.* **8**, 531 (1969).

and is reflected back toward P. The resulting optical path length difference determines the interference at P. At normal incidence, the pattern is a series of concentric rings of radius*

$$\rho \approx \left[\frac{nm\lambda a^2 b^2}{d(a^2 - b^2)}\right]^{1/2}$$

Now consider a related device, which is very useful in testing optical systems. Known as a **scatter plate**, it generally consists of a slightly rough-surfaced, transparent sheet. In an arrangement such as the one shown in Fig. 9.53, it serves as an amplitude-splitting element. In this application it must have a center of symmetry; that is, each scattering site is required to have a duplicate, symmetrically located about a central point.

In the system under consideration, a point source of quasimonochromatic light S is imaged, by means of lens L_1 on the surface, at point A of the mirror being tested. A portion of the light coming from the source is scattered by the scatter plate and thereafter illuminates the entire surface of the mirror. The mirror, in turn, reflects light back to the scatter plate. This wave, as well as the light forming the image of the pinhole at point A, passes through the scatter plate again and finally reaches the image plane (either on a screen or in a camera). Fringes are formed on this latter plane. The interference process, which is manifest in the formation of these fringes, occurs because each point in the final image plane is illuminated by light arriving via two dissimilar routes, one originating at A and the other at some point B, which reflects scattered light. Indeed, as strange as they may look at first sight, well-defined fringes do result (see photo).

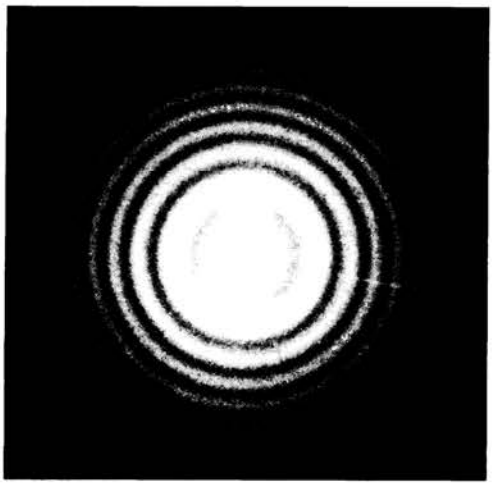
Examining the passage of light through the system in a bit more detail, consider the light initially incident on the scatter plate and assume that the wave is planar, as shown in Fig. 9.54. After it passes through the scatter plate, the incident plane wavefront $\vec{\mathbf{E}}_i$ will be distorted into a transmitted wavefront $\vec{\mathbf{E}}_T$. We envision this wave, in turn, split into a series of Fourier components consisting of plane waves, that is,

$$\vec{\mathbf{E}}_T = \vec{\mathbf{E}}_1 + \vec{\mathbf{E}}_2 + \cdots \tag{9.107}$$

Two of these constituents are shown in Fig. 9.54*a*. Now suppose we attach a specific meaning to these components; namely, $\vec{\mathbf{E}}_1$ is taken to represent the light traveling to the point A in Fig. 9.53, and $\vec{\mathbf{E}}_2$ that traveling toward B. The analysis of the

*For more of the details, see A. J. deWitte, "Interference in Scattered Light," *Am. J. Phys.* **35**, 301 (1967).

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027



Fringes in scattered light.

stages that follow could be continued in the same way. Let the portion of the wavefront returning from A be represented by the wavefront $\vec{\mathbf{E}}_A$ in Fig. 9.54*b*. The scatter plate will transform it into an irregular transmitted wave, denoted by $\vec{\mathbf{E}}_{AT}$ in the same figure. This again corresponds to a complicated configuration, but it can be split into Fourier components consisting of plane waves, as in the above case. In Fig. 9.54*b*, two of these component wavefronts have been drawn, one traveling to the left, and the other inclined at an angle θ. The latter wavefront, which is denoted by $\vec{\mathbf{E}}_{A\theta}$, is focused by lens L_2 at the point P on the screen (Fig. 9.53).

The wavefront returning from B to the scatter plate is denoted by $\vec{\mathbf{E}}_B$ in Fig. 9.54*c*. Upon traversing the scatter plate, it will be reshaped into the wave $\vec{\mathbf{E}}_{BT}$. One of the Fourier components of this wavefront, denoted by $\vec{\mathbf{E}}_{B\theta}$, is inclined at the angle θ and will therefore be focused at the same point P on the screen.

Some of the waves arriving at P will be coherent in the sense that interference occurs. To obtain the resultant irradiance I_P, first add the amplitudes of all the waves arriving at P, that is, $\vec{\mathbf{E}}_P$, and then square and time average $\vec{\mathbf{E}}_P$.

In the discussion above, only two point sources at the mirror were considered. Actually, of course, the whole surface of the mirror is illuminated by the ongoing light, and every point of it will serve as a secondary source of returning waves. All the waves will be deformed by the scatter plate, and these, in turn, can be split into plane-wave components. In each series of component waves, there will be one inclined at an angle θ, and all of these will be focused at the same point P on the screen. The resultant amplitude will then have the form

$$\vec{\mathbf{E}}_P = \vec{\mathbf{E}}_{A\theta} + \vec{\mathbf{E}}_{B\theta} + \cdots$$

The light reaching the image plane can be envisioned as made up in part of two optical fields of special interest. One of these results from light that was scattered only on its passage through the plate toward the mirror, and the other results from light that was scattered only on the way toward the image plane. The former broadly illuminates the test mirror and ultimately results in an image of it on the screen. The latter, which was initially focused to the region about A, scatters a diffuse

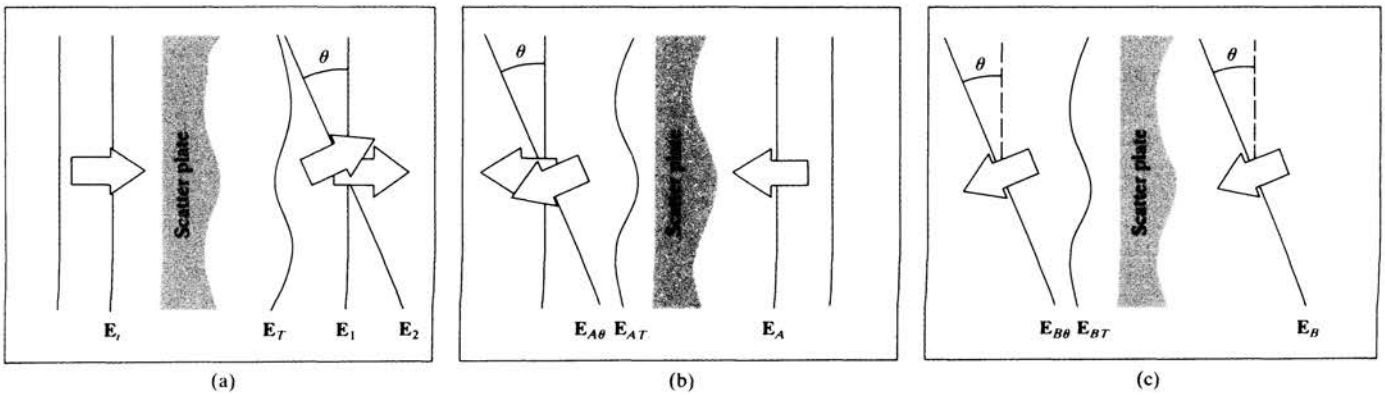


Figure 9.54 Wavefronts passing through the scatter plate.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

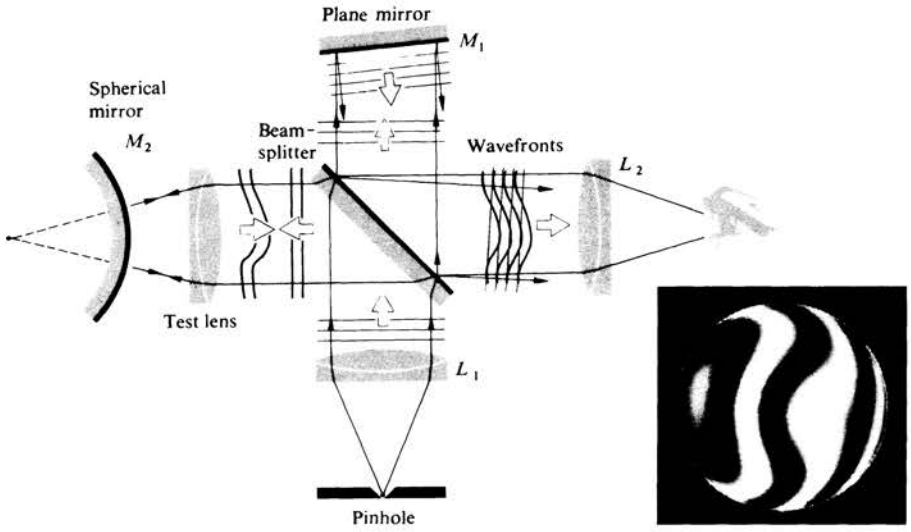


Figure 9.55 The Twyman–Green interferometer. (Photo by E.H.)

blur across the screen. The point A is chosen so that the small area in the vicinity of it is free of aberrations. In that case, the wave reflected from it serves as a reference with which to compare the wavefront corresponding to the entire mirror surface. The interference pattern will show, as a series of contour fringes, any deviations from perfection in the mirror surface.*

9.8.2 The Twyman-Green Interferometer

The Twyman–Green is essentially a variation of the Michelson Interferometer. It's an instrument of great importance in the domain of modern optical testing. Among its distinguishing physical characteristics (illustrated in Fig. 9.55) are a quasimonochromatic point source and lens L_1, to provide a source of incoming *plane waves*, and a lens L_2, which permits all the light from the aperture to enter the eye so that the entire field can be seen, that is, any portion of M_1 and M_2. A continuous laser serves as a superior source in that it provides the convenience of long path length differences and, in addition, short photographic exposure times. These tend to minimize unwanted vibration effects. Laser versions of the Twyman–Green are among the most effective testing tools in Optics. As shown in the figure, the device is set up to examine a lens. The spherical mirror M_2 has its center of curvature coincident with the focal point of the lens. If the lens being tested is free of aberrations, the emerging reflected light returning to the beamsplitter will again be a plane wave. If, however, astigmatism, coma, or spherical aberration deforms the wavefront, a fringe pattern clearly manifesting these distortions can be seen and photographed. When M_2 is replaced by a plane mirror, a number of other elements (prisms, optical flats, etc.) can be tested equally well. The optician interpreting the fringe pattern can then mark the surface for further polishing to correct high or low spots. In the fabrication of the finest optical systems, telescopes, high-altitude cameras, and so forth, the interferograms may even be scanned electronically, and the resulting data analyzed by computer. Computer-controlled plotters can then automatically produce surface contour maps or perspective "three-dimensional" drawings of the distorted wavefront generated by the element being tested. These procedures can be used throughout the fabrication process to ensure the highest quality optical instruments. Complex systems with wavefront aberrations in the fractional-wavelength range are the result of what might be called the new technology.

*For further discussion of the scatter plate, the reader might consult the rather succinct papers by J. M. Burch, *Nature* **171**, 889 (1953), and *J. Opt. Soc. Am.* **52**, 600 (1962). Reference should be made to J. Strong, *Concepts of Classical Optics*, p. 383. Also see R. M. Scott, "Scatter Plate Interferometry," *Appl. Opt.* **8**, 531 (1969), and J. B. Houston, Jr., "How to Make and Use a Scatterplate Interferometer," *Optical Spectra* (June 1970), p. 32.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

9.8.3 The Rotating Sagnac Interferometer

The Sagnac Interferometer is widely used to measure rotational speed. In particular, the ***ring laser***, which is essentially a Sagnac Interferometer containing a laser in one or more of its arms, was designed specifically for that purpose. The first ring laser gyroscope was introduced in 1963, and work is continuing on various devices of this sort (see photo). The initial experiments that gave impetus to these efforts were performed by Sagnac in 1911. At that time he rotated the entire interferometer, mirrors, source, and detector, about a perpendicular axis passing through its center (Fig. 9.56). Recall, from Section 9.4.2, that two overlapping beams traverse the interferometer, one clockwise, the other counterclockwise. The rotation effectively shortens the path taken by one beam in comparison to that of the other. In the interferometer, the result is a fringe shift proportional to the angular speed of rotation ω. In the ring laser, it is a frequency difference between the two beams that is proportional to ω.

Consider the arrangement depicted in Fig. 9.56. The corner A (and every other corner) moves with a linear speed $v = R\omega$, where R is half the diagonal of the square. Using classical reasoning, we find that the time of travel of light along AB is

$$t_{AB} = \frac{R\sqrt{2}}{c - v/\sqrt{2}}$$



An early ring laser gyro. (Photo courtesy Autonetics, a Division of Boeing North America, Inc.)

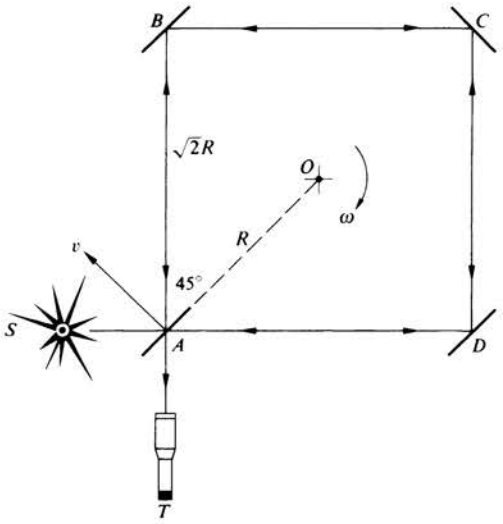


Figure 9.56 The rotating Sagnac Interferometer. Originally it was 1 m × 1 m with ω = 120 rev/min.

or

$$t_{AB} = \frac{2R}{\sqrt{2}c - \omega R}$$

The time of travel of the light from A to D is

$$t_{AD} = \frac{2R}{\sqrt{2}c + \omega R}$$

The total time for counterclockwise and clockwise travel is given respectively by

$$t_{\circlearrowleft} = \frac{8R}{\sqrt{2}c + \omega R}$$

and

$$t_{\circlearrowright} = \frac{8R}{\sqrt{2}c - \omega R}$$

For $\omega R << c$ the difference between these two intervals is

$$\Delta t = t_{\circlearrowright} - t_{\circlearrowleft}$$

or, using the Binomial Series,

$$\Delta t = \frac{8R^2\omega}{c^2}$$

This can be expressed in terms of area $A = 2R^2$ of the square formed by the beams of light as

$$\Delta t = \frac{4A\omega}{c^2}$$

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

Let the period of the monochromatic light used be $\tau = \lambda/c$; then the fractional displacement of the fringes, given by $\Delta N = \Delta t/\tau$, is

$$\Delta N = \frac{4A\omega}{c\lambda}$$

a result that has been verified experimentally. In particular, Michelson and Gale* used this method to determine the angular velocity of the Earth.

The preceding classical treatment is obviously lacking, inasmuch as it assumes speeds in excess of c, an assumption that is contrary to the dictates of Special Relativity. Furthermore, it would appear that since the system is accelerating, General Relativity would prevail. In fact, these formalisms yield the same results.

9.8.4 Radar Interferometry

In February 2000 the Space Shuttle Endeavour completed a mission to create a "three-dimensional" map of the Earth covering 119 million square kilometers. The feat was accomplished using synthetic aperture radar (SAR). In general, the larger the aperture of a viewing system, the greater the resolution (p. 471) and the more details one can see. SAR is a technique for using the motion of an airplane or spacecraft along with signal processing methods to simulate a large antenna.

Using a phased array antenna (p. 98), the Shuttle swept a radar beam back and forth perpendicular to its line of motion painting a 225-km wide swath over the Earth's surface (Fig. 9.57). Orbiting upside-down, Endeavour extended a 60-m mast with two receiving antennas at its end (Fig. 9.58). The SAR then sent out a stream of about 1700 high-powered electromagnetic pulses per second from its main antenna in the cargo bay, which was both a transmitter and receiver. Actually, the mission utilized two different radars: a C-band system operating at a wavelength of 5.6 cm that provided most of the coverage, and a higher resolution X-band 3-cm system that gave a detailed view of a narrow 50-km swath (Fig. 9.57). A radar image is made up of countless tiny uniform dots known as pixels (p. 473). The pixel is the smallest bit of information in the picture—nothing can be seen that's smaller than a single pixel. For the main C-band system, each pixel is about 12.5 m in diameter, and the smallest object that can be resolved is about 30 m across.

*Michelson and Gale, *Astrophys. J.* **61**, 140 (1925).

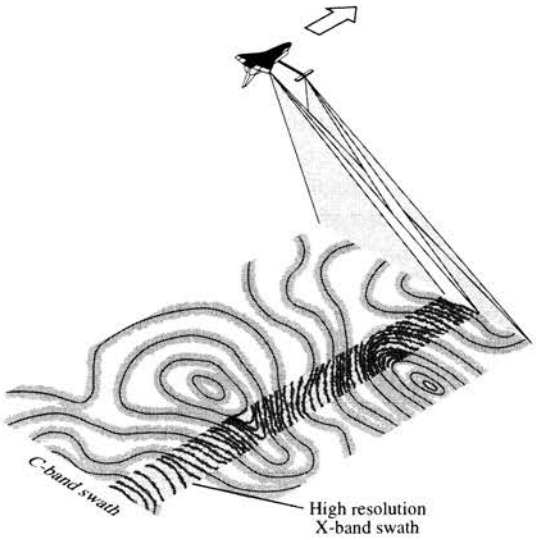


Figure 9.57 As the Shuttle orbited, its two radar systems swept out a swath across the surface of the Earth.

Ordinarily, a radar system sends out a pulse (with a pulsewidth of 10-50 μs), and then, picking up the backscattered wave, it records both the amplitude and round-trip time. That gives a rough idea of the size and location of the target. However, in order to gather data about the elevation of surface features on the Earth, the Shuttle Radar Topography Mission (SRTM) utilized interferometry—in a way that suggests

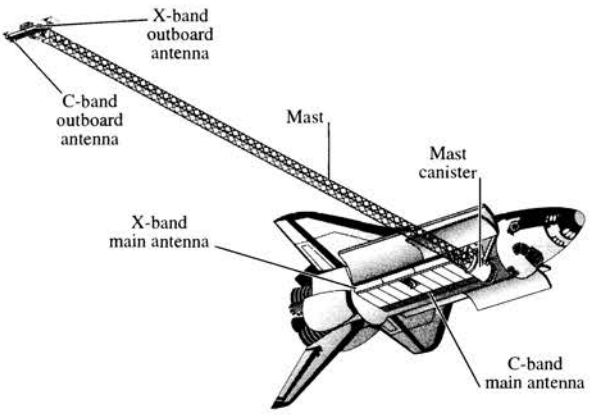


Figure 9.58 The Shuttle Endeavour carried the main C-band transmitter-receiver antenna in its cargo bay and a second receiver at the end of a 60-m long mast.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

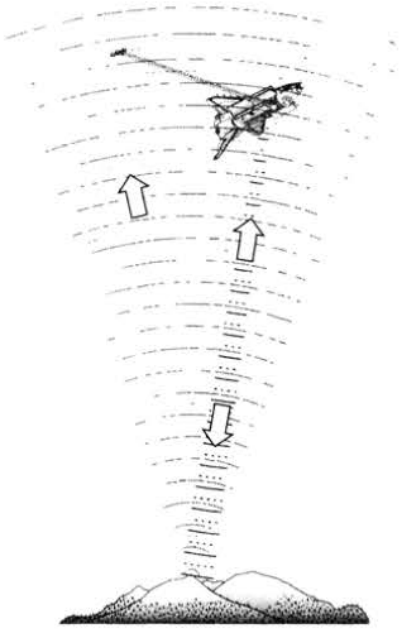


Figure 9.59 A radar pulse emitted from the Shuttle strikes the ground and reflects back. The echo is picked up by both the outboard and inboard antennas.

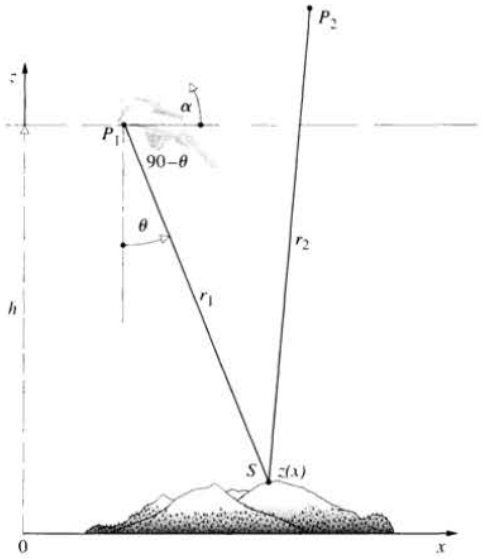
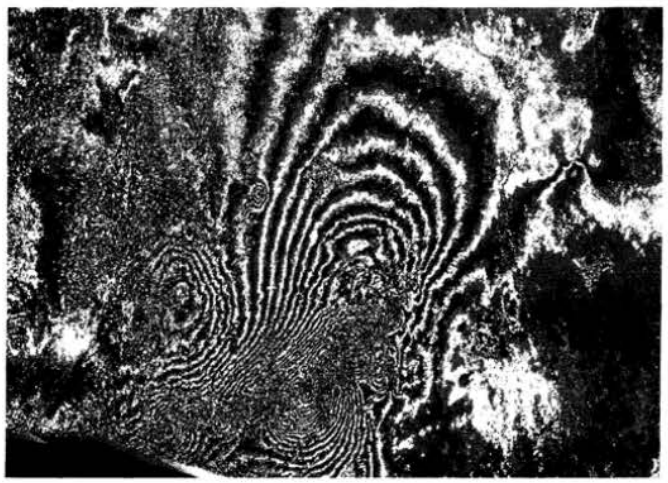


Figure 9.60 The basic geometry of the SAR interferometer. The source point S is the spot on the ground that reflects the radar pulse back to the Shuttle. Points P_1 and P_2 correspond to the two receivers, one on the mast and the other in the cargo bay of the Shuttle.

Young's Experiment run backwards (p. 393). In any event, similar interferometric techniques are of growing importance in radio and optical astronomy.

The SAR is a coherent imaging system, and it retains information about both the amplitude and phase of the radar echo during data acquisition and processing. A signal is emitted from the Shuttle (much like the flash from an ordinary camera but spectrally more controlled); it strikes the ground (Fig. 9.59) and returns to the two antennas, one in the cargo bay (P_1), the other on the boom (P_2). These are separated by a 60-m baseline a. The two radar echoes are converted into digital data which are recorded for later processing and display as an image. It's left for Problem 48 (Fig. 9.60) to show that the topography in the form of the function $z(x)$ can be expressed in terms of the altitude h, the *look angle* of the radar θ, and measured phase angle difference, or *interferometric phase* ϕ between the two signals;

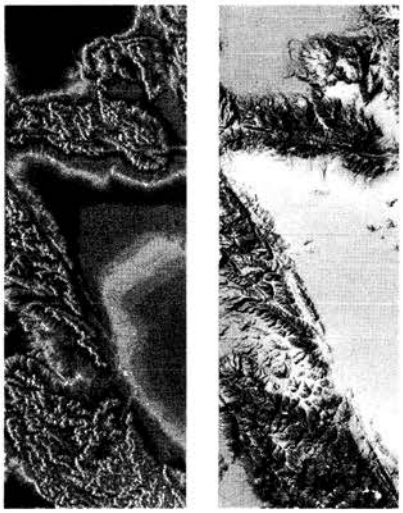
$$z(x) = h - \frac{(\lambda\phi/2\pi)^2 - a^2}{2a \sin(\alpha - \theta) - (\lambda\phi/2\pi)} \cos\theta \qquad (9.108)$$



Synthetic aperture radar was used to produce this interferogram arising from the June 1992 earthquake in Landers, California. Images taken by the ERS-1 satellite before and after the quake were combined to generate this fringe pattern which reveals the shift in the ground that took place. The picture covers an area of about 125 by 175 km. (Photo courtesy Didier Massonnet, Centre National d'Etudes Spatiales, France.)

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

An interferometer of this sort measures ϕ, the difference in phase between the signals arriving at the ends of its baseline. It does this by analytically interfering those signals using a process called *cross correlation* (p. 545). When the two separate data sets, one from each antenna, are combined on the ground the first thing produced is an interferogram or fringe map (see photo) which encodes the topography. The interferogram corresponds to a collection of "fringes of equal height," or if you will, contours of equal height. But the information needs further refining; the elevations of the contours are unknown. Based on accurate knowledge of the mast length and orientation, the height of each contour, $z(x)$, is determined, essentially via triangulation. Data collected over the oceans provides a sea-level reference for all elevations. After a considerable amount of computation, pixel by pixel, a 3-D topographical map is finally created (see photo).



These are radar images of San Andreas, California taken by the Space Shuttle Endeavour in 2000. The picture on the left (which looks a lot better in color) shows an interferogram overlaying the terrain; the picture on the right is the corresponding "three-dimensional" map that results from the analysis of all of the data.

PROBLEMS

Complete solutions to all problems—except those with an asterisk—can be found in the back of the book.

9.1 Returning to Section 9.1, let

$$\tilde{\mathbf{E}}_1(\vec{r}, t) = \tilde{\mathbf{E}}_1(\vec{r})e^{-i\omega t}$$

and

$$\tilde{\mathbf{E}}_2(\vec{r}, t) = \tilde{\mathbf{E}}_2(\vec{r})e^{-i\omega t}$$

where the wavefront shapes are not explicitly specified, and $\tilde{\mathbf{E}}_1$ and $\tilde{\mathbf{E}}_2$ are complex vectors depending on space and initial phase angle. Show that the interference term is then given by

$$I_{12} = \tfrac{1}{2}(\tilde{\mathbf{E}}_1 \cdot \tilde{\mathbf{E}}_2^* + \tilde{\mathbf{E}}_1^* \cdot \tilde{\mathbf{E}}_2) \tag{9.109}$$

You will have to evaluate terms of the form

$$\langle \tilde{\mathbf{E}}_1 \cdot \tilde{\mathbf{E}}_2 e^{-2i\omega t} \rangle_T = (\tilde{\mathbf{E}}_1 \cdot \tilde{\mathbf{E}}_2 / T) \int_t^{t+T} e^{-2i\omega t'} dt'$$

for $T >> \tau$ (take another look at Problem 3.10). Show that Eq. (9.109) leads to Eq. (9.11) for plane waves.

9.2 In Section 9.1 we considered the spatial distribution of energy for two point sources. We mentioned that for the case in which the separation $a >> \lambda$, I_{12} spatially averages to zero. Why is this true? What happens when a is much less than λ?

9.3* Return to Fig. 2.22 and prove that if two electromagnetic plane waves making an angle θ have the same amplitude, E_θ, the resulting interference pattern on the yx-plane is a cosine-squared irradiance distribution given by

$$I(y) = 4E_0^2 \cos^2\left(\frac{\pi}{\lambda} y \sin\theta\right)$$

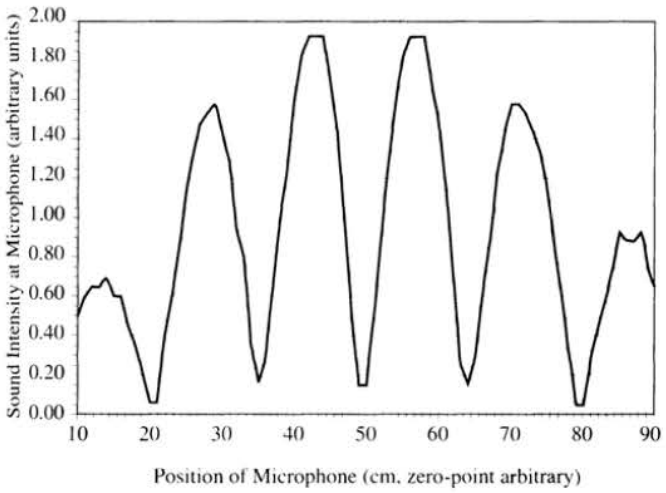
Locate the zeros of irradiance. What is the value of the fringe separation? What happens to the separation as θ increases? Compare your analysis with that leading to Eq. (9.17). [*Hint*: Begin with the wave expressions given in Section 2.7, which have the proper phases already worked out, and write them as exponentials.]

9.4 Will we get an interference pattern in Young's Experiment (Fig. 9.8) if we replace the source slit S by a single long-filament lightbulb? What would occur if we replaced the slits S_1 and S_2 by these same bulbs?

9.5* Figure P.9.5 shows an output pattern that was measured by a tiny microphone when two small piezo-loudspeakers separated by 15 cm were pointed toward the microphone at a distance of 1.5 m away. Given that the speed of sound at 20°C is 343 m/s, determine the approximate frequency at which the speakers were driven. Discuss the nature of the pattern and explain why it has a central minimum.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

Figure P.9.5 (Data courtesy of CENCO)



9.6* Two 1.0-MHz radio antennas emitting in-phase are separated by 600 m along a north–south line. A radio receiver placed 2.0 km east is equidistant from both transmitting antennas and picks up a fairly strong signal. How far north should that receiver be moved if it is again to detect a signal nearly as strong?

9.7 An expanded beam of red light from a He–Ne laser ($\lambda_0 = 632.8$ nm) is incident on a screen containing two very narrow horizontal slits separated by 0.200 mm. A fringe pattern appears on a white screen held 1.00 m away.

(a) How far (in radians and millimeters) above and below the central axis are the first zeros of irradiance?

(b) How far (in mm) from the axis is the fifth bright band?

(c) Compare these two results.

9.8* Red plane waves from a ruby laser ($\lambda_0 = 694.3$ nm) in air impinge on two parallel slits in an opaque screen. A fringe pattern forms on a distant wall, and we see the fourth bright band 1.0° above the central axis. Kindly calculate the separation between the slits.

9.9* A 3 × 5 card containing two pinholes, 0.08 mm in diameter and separated center to center by 0.10 mm, is illuminated by parallel rays of blue light from an argon ion laser ($\lambda_0 = 487.99$ nm). If the fringes on an observing screen are to be 10 mm apart, how far away should the screen be?

9.10* White light falling on two long narrow slits emerges and is observed on a distant screen. If red light ($\lambda_0 = 780$ nm) in the first-order fringe overlaps violet in the second-order fringe, what is the latter's wavelength?

9.11* Considering the double-slit experiment, derive an equation for the distance $y_{m'}$ from the central axis to the m'th irradiance *minimum*, such that the first dark bands on either side of the central maximum correspond to $m' = \pm 1$. Identify and justify all your approximations.

9.12* With regard to Young's Experiment, derive a general expression for the shift in the vertical position of the mth *maximum* as a result of placing a thin parallel sheet of glass of index n and thickness d directly over one of the slits. Identify your assumptions.

9.13* Plane waves of monochromatic light impinge at an angle θ_i on a screen containing two narrow slits separated by a distance a. Derive an equation for the angle measured from the central axis which locates the mth maximum.

9.14* Sunlight incident on a screen containing two long narrow slits 0.20 mm apart casts a pattern on a white sheet of paper 2.0 m beyond. What is the distance separating the violet ($\lambda_0 = 400$ nm) in the first-order band from the red ($\lambda_0 = 600$ nm) in the second-order band?

9.15 To examine the conditions under which the approximations of Eq. (9.23) are valid:

(a) Apply the law of cosines to triangle S_1S_2P in Fig. 9.8*c* to get

$$\frac{r_2}{r_1} = \left[1 - 2\left(\frac{a}{r_1}\right)\sin\theta + \left(\frac{a}{r_1}\right)^2\right]^{1/2}$$

(b) Expand this in a Maclaurin series yielding

$$r_2 = r_1 - a\sin\theta + \frac{a^2}{2r_1}\cos^2\theta + \cdots$$

(c) In light of Eq. (9.17), show that if $(r_1 - r_2)$ is to equal $a\sin\theta$, it is required that $r_1 >> a^2/\lambda$.

9.16 A stream of electrons, each having an energy of 0.5 eV, impinges on a pair of extremely thin slits separated by 10^{-2} mm. What is the distance between adjacent minima on a screen 20 m behind the slits? ($m_e = 9.108 \times 10^{-31}$ kg, 1 eV = 1.602×10^{-19} J.)

9.17* It is our intention to produce interference fringes by illuminating some sort of arrangement (Young's Experiment, a thin film, the Michelson Interferometer, etc.) with light at a mean wavelength of 500 nm, having a linewidth of 2.5×10^{-3} nm. At approximately what optical path length difference can you expect the fringes to vanish? [*Hint*: Think about the coherence length and revisit Problem 7.39.]

9.18* Imagine that you have an opaque screen with three horizontal very narrow parallel slits in it. The second slit is a center-to-center

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

distance a beneath the first, and the third is a distance $5a/2$ beneath the first. (a) Write a complex exponential expression in terms of δ for the amplitude of the electric field at some point P at an elevation θ on a distant screen where $\delta = ka \sin \theta$. Prove that

$$I(\theta) = \frac{I(0)}{3} + \frac{2I(0)}{9}(\cos \delta + \cos 3\delta/2 + \cos 5\delta/2)$$

Verify that at $\theta = 0$, $I(\theta) = I(0)$.

9.19* Show that a for the Fresnel biprism of Fig. 9.13 is given by $a = 2d(n - 1)\alpha$.

9.20* In the Fresnel double mirror $s = 2$ m, $\lambda_0 = 589$ nm, and the separation of the fringes was found to be 0.5 mm. What is the angle of inclination of the mirrors, if the perpendicular distance of the actual point source to the intersection of the two mirrors is 1 m?

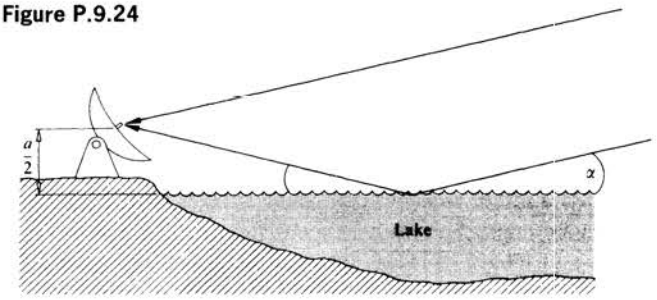
9.21* The Fresnel biprism is used to obtain fringes from a point source that is placed 2 m from the screen, and the prism is midway between the source and the screen. Let the wavelength of the light be $\lambda_0 = 500$ nm and the index of refraction of the glass be $n = 1.5$. What is the prism angle, if the separation of the fringes is 0.5 mm?

9.22 What is the general expression for the separation of the fringes of a Fresnel biprism of index n immersed in a medium having an index of refraction n'?

9.23 Using Lloyd's mirror, X-ray fringes were observed, the spacing of which was found to be 0.002 5 cm. The wavelength used was 8.33 Å. If the source–screen distance was 3 m, how high above the mirror plane was the point source of X-rays placed?

9.24 Imagine that we have an antenna at the edge of a lake picking up a signal from a distant radio star (Fig. P.9.24), which is just coming up above the horizon. Write expressions for δ and for the angular position of the star when the antenna detects its first maximum.

Figure P.9.24



9.25* If the plate in Fig. 9.17 is glass in air, show that the amplitudes of E_{1r}, E_{2r}, and E_{3r} are respectively $0.2E_{0i}$, $0.192E_{0i}$, and $0.008E_{0i}$, where E_{0i} is the incident amplitude. Make use of the Fresnel coefficients at normal incidence, assuming no absorption. You might repeat the calculation for a water film in air.

9.26 A soap film surrounded by air has an index of refraction of 1.34. If a region of the film appears bright red ($\lambda_0 = 633$ nm) in normally reflected light, what is its minimum thickness there?

9.27* A thin film of ethyl alcohol ($n = 1.36$) spread on a flat glass plate and illuminated with white light shows a color pattern in reflection. If a region of the film reflects only green light (500 nm) strongly, how thick is it?

9.28* A soap film of index 1.34 has a region where it is 550.0 nm thick. Determine the vacuum wavelengths of the radiation that is not reflected when the film is illuminated from above with sunlight.

9.29 Consider the circular pattern of Haidinger's fringes resulting from a film with a thickness of 2 mm and an index of refraction of 1.5. For monochromatic illumination of $\lambda_0 = 600$ nm, find the value of m for the central fringe ($\theta_t = 0$). Will it be bright or dark?

9.30 Illuminate a microscope slide (or even better, a thin cover-glass slide). Colored fringes can easily be seen with an ordinary fluorescent lamp (although some of the newer versions don't work well at all) serving as a broad source or a mercury street light as a point source. Describe the fringes. Now rotate the glass. Does the pattern change? Duplicate the conditions shown in Figs. 9.18 and 9.19. Try it again with a sheet of plastic food wrap stretched across the top of a cup.

9.31 Figure P.9.31 illustrates a setup used for testing lenses. Show that

$$d = x^2(R_2 - R_1)/2R_1R_2$$

when d_1 and d_2 are negligible in comparison with $2R_1$ and $2R_2$, respectively. (Recall the theorem from plane geometry that relates the products of the segments of intersecting chords.) Prove that the radius of the mth dark fringe is then

$$x_m = [R_1R_2m\lambda_f/(R_2 - R_1)]^{1/2}$$

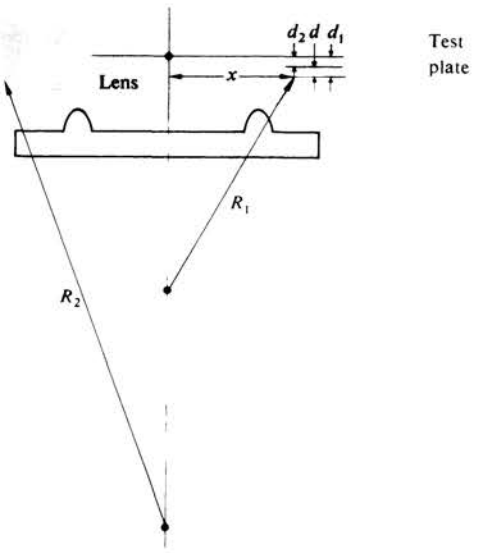
How does this relate to Eq. (9.43)?

9.32* Newton rings are observed on a film with quasimonochromatic light that has a wavelength of 500 nm. If the 20th bright ring has a radius of 1 cm, what is the radius of curvature of the lens forming one part of the interfering system?

9.33 Fringes are observed when a parallel beam of light of wavelength 500 nm is incident perpendicularly onto a wedge-shaped film with an index of refraction of 1.5. What is the angle of the wedge if the fringe separation is $\frac{1}{3}$ cm?

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

Figure P.9.31



9.34* Suppose a wedge-shaped air film is made between two sheets of glass, with a piece of paper 7.618×10^{-5} m thick used as the spacer at their very ends. If light of wavelength 500 nm comes down from directly above, determine the number of bright fringes that will be seen across the wedge.

9.35 A Michelson Interferometer is illuminated with monochromatic light. One of its mirrors is then moved 2.53×10^{-5} m, and it is observed that 92 fringe-pairs, bright and dark, pass by in the process. Determine the wavelength of the incident beam.

9.36* One of the mirrors of a Michelson Interferometer is moved, and 1000 fringe-pairs shift past the hairline in a viewing telescope during the process. If the device is illuminated with 500-nm light, how far was the mirror moved?

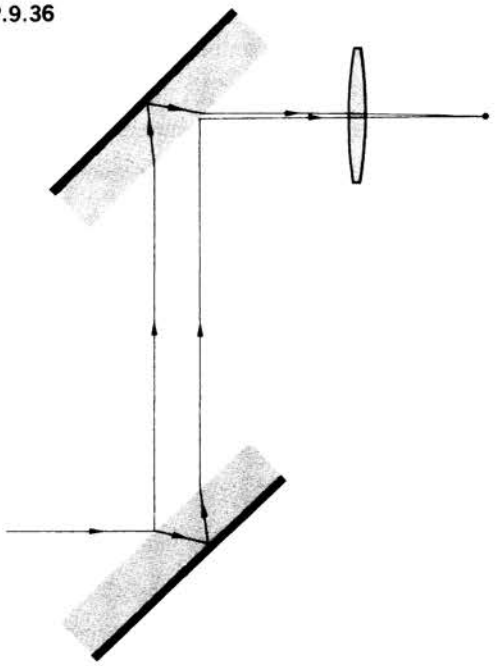
9.37* Suppose we place a chamber 10.0 cm long with flat parallel windows in one arm of a Michelson Interferometer that is being illuminated by 600-nm light. If the refractive index of air is 1.000 29 and all the air is pumped out of the cell, how many fringe-pairs will shift by in the process?

9.38* A form of the Jamin Interferometer is illustrated in Fig. P.9.38. How does it work? To what use might it be put?

9.39 Starting with Eq. (9.53) for the transmitted wave, compute the flux density, that is, Eq. (9.54).

9.40 Given that the mirrors of a Fabry–Perot Interferometer have an

Figure P.9.36



amplitude reflection coefficient of $r = 0.894\,4$, find

(a) the coefficient of finesse,

(b) the half-width,

(c) the finesse, and,

(d) the *contrast factor* defined by

$$C \equiv \frac{(I_t/I_i)_{\max}}{(I_t/I_i)_{\min}}$$

9.41 To fill in some of the details in the derivation of the smallest phase increment separating two resolvable Fabry–Perot fringes, that is,

$$(\Delta\delta) \approx 4.2/\sqrt{F} \qquad [9.73]$$

satisfy yourself that

$$[\mathscr{A}(\theta)]_{\delta = \delta_a \pm \Delta\delta/2} = [\mathscr{A}(\theta)]_{\delta = \Delta\delta/2}$$

Show that Eq. (9.72) can be rewritten as

$$2[\mathscr{A}(\theta)]_{\delta = \Delta\delta/2} = 0.81\{1 + [\mathscr{A}(\theta)]_{\delta = \Delta\delta}\}$$

When F is large γ is small, and $\sin(\Delta\delta) = \Delta\delta$. Prove that Eq. (9.73) then follows.

9.42 Consider the interference pattern of the Michelson Interferometer as arising from two beams of equal flux density. Using Eq.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

(9.17), compute the half-width. What is the separation, in δ, between adjacent maxima? What then is the finesse?

9.43* Satisfy yourself of the fact that a film of thickness $\lambda_f/4$ and index n_1 will always reduce the reflectance of the substrate on which it is deposited, as long as $n_s > n_1 > n_0$. Consider the simplest case of normal incidence and $n_0 = 1$. Show that this is equivalent to saying that the waves reflected back from the two interfaces cancel one another.

9.44 Verify that the reflectance of a substrate can be increased by coating it with a $\lambda_f/4$, high-index layer, that is, $n_1 > n_s$. Show that the reflected waves interfere constructively. The quarter-wave stack $g(HL)^m Ha$ can be thought of as a series of such structures.

9.45 Determine the refractive index and thickness of a film to be deposited on a glass surface ($n_g = 1.54$) such that no normally incident light of wavelength 540 nm is reflected.

9.46 A glass microscope lens having an index of 1.55 is to be coated with a magnesium fluoride film to increase the transmission of normally incident yellow light ($\lambda_0 = 550$ nm). What minimum thickness should be deposited on the lens?

9.47* A glass camera lens with an index of 1.55 is to be coated with a cryolite film ($n \approx 1.30$) to decrease the reflection of normally incident green light ($\lambda_0 = 500$ nm). What thickness should be deposited on the lens?

9.48* Using Fig. 9.60, which depicts the geometry of the Shuttle radar interferometer, show that

$$z(x) = h - r_1 \cos \theta$$

Then use the Law of Cosines to establish that Eq. (9.108) is correct.

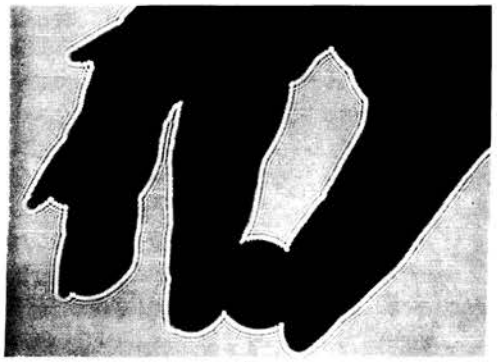
Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

10 Diffraction

10.1 Preliminary Considerations

An opaque body placed midway between a screen and a point source casts an intricate shadow made up of bright and dark regions quite unlike anything one might expect from the tenets of Geometrical Optics (see photos).* The work of Francesco Grimaldi in the 1600s was the first published detailed study of this ***deviation of light from rectilinear propagation***, something he called "*diffractio*." *The effect is a general characteristic of wave phenomena occurring whenever a portion of a wavefront, be it sound, a matter wave, or light, is obstructed in some way.* If in the course of encountering an obstacle, either transparent or opaque, a region of the wavefront is altered in amplitude or phase, diffraction will occur.† The various segments of the wavefront that propagate beyond the obstacle interfere, causing the particular energy-density distribution referred to as the diffraction pattern. There is no significant physical distinction between *interference* and *diffraction*. It has, however, become somewhat customary, if not always appropriate, to speak of interference when considering the superposition of only a few waves and diffraction when treating a large number of waves. Even so, one refers to multiple-beam interference in one context and diffraction from a grating in another.

It would be nice to treat diffraction from the perspective of the most powerful contemporary theory of light, Quantum



(a)



(b)

(*a*) The shadow of Mary's hand holding a dime, cast directly on 4 × 5 Polaroid A.S.A. 3000 film using a He–Ne beam and no lenses. (Photo by E. H.) (*b*) Fresnel diffraction of electrons by zinc oxide crystals. (After H. Boersch from *Handbuch der Physik*, edited by S.Flugge, Springer-Verlag, Heidelberg.)

*The effect is easily seen, but you need a fairly strong source. A high-intensity lamp shining through a small hole works well. If you look at the shadow pattern arising from a pencil under point-source illumination, you will see an unusual bright region bordering the edge and even a faintly illuminated band down the middle of the shadow. Take a close look at the shadow cast by your hand in direct sunlight.

†Diffraction associated with transparent obstacles is not usually considered, although if you have ever driven an automobile at night with a few rain droplets on your eyeglasses, you are no doubt quite familiar with the effect. If you have not, put a droplet of water or saliva on a glass plate, hold it close to your eye, and look directly through it at a point source. You'll see bright and dark fringes.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

Electrodynamics (QED), but that's terribly impractical; the analysis is far too complicated and wouldn't add much at that. What we *can* do is show qualitatively how QED applies to a few basic situations. For our purposes, however, the classical wave theory, which provides the simplest effective formalism, will more than suffice. Still, wherever it's appropriate, the discussion will be illuminated with insights from Fourier analysis, even though the detailed treatment of that subject is postponed to the next chapter.

The Huygens–Fresnel Principle

As an initial approach to the problem, let's reconsider Huygens's Principle (Section 4.4.2). Each point on a wavefront can be envisaged as a source of secondary spherical wavelets. The progress through space of the wavefront, or any portion thereof, can then presumably be determined. At any particular time, the shape of the wavefront is supposed to be the envelope of the secondary wavelets (Fig. 4.26). The technique, however, ignores most of each secondary wavelet, retaining only that portion common to the envelope. As a result of this inadequacy, Huygens's Principle by itself is unable to account for the details of the diffraction process. That this is indeed the case is borne out by everyday experience. Sound waves (e.g., $\nu = 500$ Hz, $\lambda \approx 68$ cm) easily "bend" around large objects like telephone poles and trees, yet these objects cast fairly distinct shadows when illuminated by light. Huygens's Principle is independent of any wavelength considerations, however, and would predict the same wavefront configurations in both situations.

The difficulty was resolved by Fresnel with his addition of the concept of interference. The corresponding **Huygens–Fresnel Principle** states that ***every unobstructed point of a wavefront, at a given instant, serves as a source of spherical secondary wavelets (with the same frequency as that of the primary wave). The amplitude of the optical field at any point beyond is the superposition of all these wavelets (considering their amplitudes and relative phases).***

Applying these ideas on the very simplest qualitative level, refer to the accompanying ripple tank photographs and the illustration in Fig. 10.1. If each unobstructed point on the incoming plane wave acts as a coherent secondary source, the maximum optical path length difference among them will be $\Lambda_{\max} = |\overline{AP} - \overline{BP}|$, corresponding to a source point at each edge of the aperture. But $\Lambda_{\max}$ is less than or equal to $\overline{AB}$, the latter being the case when P is on the screen. When $\lambda > \overline{AB}$, as

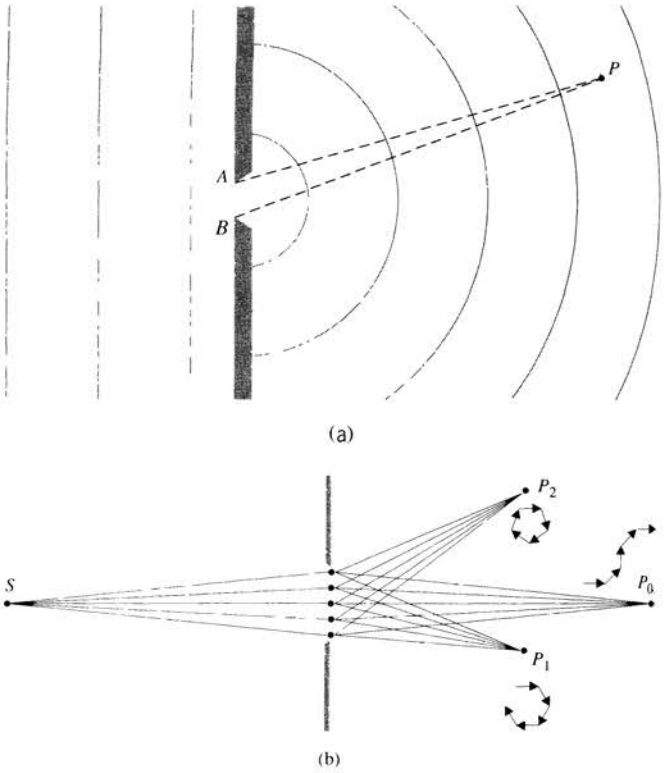


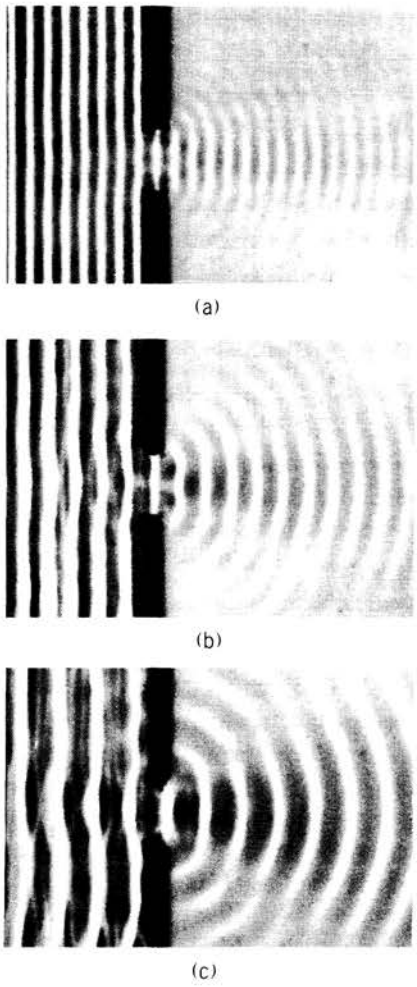
Figure 10.1 Diffraction at a small aperture. (*a*) The classical wave picture. (*b*) The view via QED and probability amplitudes.

in Fig. 10.3, it follows that $\lambda > \Lambda_{\max}$, and since the waves were initially in-phase, they all interfere constructively (to varying degrees) wherever P happens to be (see ripple tank photo *c*). Thus, ***if the wavelength is large compared to the aperture, the waves will spread out at large angles into the region beyond the obstruction***. And the smaller the aperture gets, the more nearly circular the diffracted waves become (recall the discussion of this point from a Fourier context, p. 397).

The antithetic situation occurs when $\lambda < \overline{AB}$ (as in ripple tank photo *a*). The area where $\lambda > \Lambda_{\max}$ is limited to a small region extending out directly in front of the aperture, and it is only there that all the wavelets will interfere constructively. Beyond this zone some of the wavelets can interfere destructively, and the "shadow" begins. Keep in mind that the idealized *geometric shadow* corresponds to $\lambda \to 0$.

Classically, the reason light goes where it does beyond the screen is that the multitude of wavelets emitted from the aper-

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027



Diffraction through an aperture with varying λ as seen in a ripple tank. (Photo courtesy PSSC *Physics*, D. C. Heath, Boston, 1960.)

ture "interfere"; that is, they combine (as phasors) at every point in the region, some places enhancing, some canceling, depending on the *OPL*.

Quantum mechanically (Section 4.11.1), the reason light goes where it does beyond the screen is that the multitude of probability amplitudes for photons from the aperture "interfere." That is, they combine (as phasors) at every point in the region, some places enhancing, some canceling, depending on the *OPL*. When the hole is several wavelengths wide (as in ripple tank photo *a*), the many paths to any point *P* correspond to a broad range of phasor phases. Consider all the paths to a point in the forward direction such as P_0. The straight-line route from S to P_0 corresponds to a minimum in *OPL*. Any other paths through the aperture to P_0 are somewhat longer (depending on the size of the hole) and have phasors (all of which we will take to be the same size) that are grouped around that stationary *OPL* value, much as those in Fig. 4.68. They have small mutual phase-angle differences (half +, half −) and so added tip-to-tail they turn one way, then the other, to produce a substantial resultant probability amplitude. A photon counter at P_0 will see lots of light. Off the forward direction (where the *OPL* is not stationary), the phasors each have relatively large phase-angle differences for every path and all are of the same sign. Placed tip-to-tail they spiral around, adding up to little or nothing. A detector at P_1 will record few counts, and one at P_2 fewer still.

If the aperture is now made much smaller, the number of counts at P_1 and P_2 increases, even as the number at P_0 drops off. With a narrow hole, all the paths to either P_1 or P_2 are much closer together and have nearly the same *OPL*. The phase-angle differences are therefore much smaller, the phasor spirals no longer close on themselves, and the resultant probability amplitudes, though small, are appreciable everywhere.

Qualitatively, both QED and the classical Huygens–Fresnel Principle lead to much the same general conclusion: *light diffracts and interference is at the heart of the process.*

The Huygens–Fresnel Principle has some shortcomings (which we will examine later), in addition to the fact that the whole thing at this point is rather hypothetical. Gustav Kirchhoff developed a more rigorous theory based directly on the solution of the differential wave equation. Kirchhoff, though a contemporary of Maxwell, did his work before Hertz's demonstration (and the resulting popularization) of the propagation of electromagnetic waves in 1887. Accordingly, Kirchhoff employed the older elastic-solid theory of light. His refined analysis lent credence to the assumptions of Fresnel and led to an even more precise formulation of Huygens's Principle as an exact consequence of the wave equation. Even so, the Kirchhoff theory is itself an approximation that is valid for sufficiently small wavelengths—that is, when the diffracting apertures have dimensions that are large in comparison to λ. The difficulty arises from the fact that what's required is the solution of a partial differential equation that meets the boundary conditions imposed by the obstruction. This kind of rigorous solution is obtainable only in a few special cases. Kirchhoff's theory works fairly well, even though it deals only

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

with scalar waves and is insensitive to the fact that light is a transverse vector field.*

It should be stressed that the problem of determining an exact solution for a particular diffracting configuration is among the most troublesome to be dealt with in Optics. The first such solution, utilizing the electromagnetic theory of light, was published by Arnold Johannes Wilhelm Sommerfeld (1868–1951) in 1896. Although the problem was physically somewhat unrealistic, in that it involved an infinitely thin yet opaque, perfectly conducting plane screen, the result was nonetheless extremely valuable, providing a good deal of insight into the fundamental processes involved.

Rigorous solutions of this sort do not exist even today for many of the configurations of practical interest. We will therefore, out of necessity, rely on the approximate treatments of Huygens–Fresnel and Kirchhoff. In recent times, microwave techniques have been employed to conveniently study features of the diffraction field that might otherwise be almost impossible to examine optically. The Kirchhoff theory has held up remarkably well under this kind of scrutiny.† In many cases, the simpler Huygens–Fresnel treatment will prove adequate to our needs.

10.1.1 Opaque Obstructions

Diffraction may be envisioned as arising from the interaction of electromagnetic waves with some sort of physical obstruction. We would therefore do well to reexamine briefly the processes involved; what actually takes place within the material of the opaque object?

One possible description is that a screen may be considered to be a continuum; that is, its microscopic structure may be neglected. For a nonabsorbing metal sheet (no joule heating, therefore infinite conductivity) we can write Maxwell's Equations for the metal and for the surrounding medium, and then match the two at the boundaries. Precise solutions can be obtained for very simple configurations. The reflected and diffracted waves then result from the current distribution within the sheet.

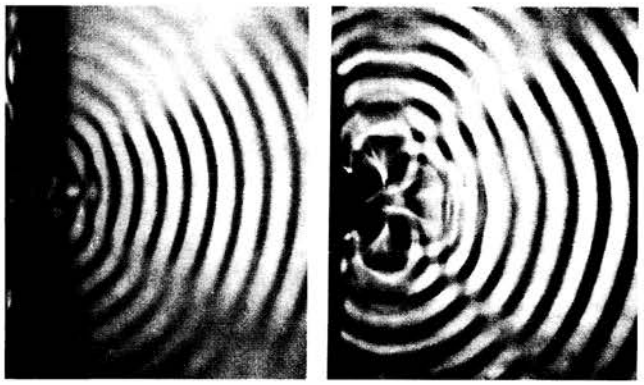
Examining the screen on a submicroscopic scale, imagine the electron cloud of each atom set into vibration by the electric field of the incident radiation. The classical model, which speaks of electron-oscillators vibrating and reemitting at the source frequency, serves quite well so that we need not be concerned with the quantum-mechanical description. The amplitude and phase of a particular oscillator within the screen are determined by the local electric field surrounding it. This in turn is a superposition of the incident field and the fields of all the other vibrating electrons. A large opaque screen with no apertures, be it made of black paper or aluminum foil, has one obvious effect: there is no optical field in the region beyond it. Electrons near the illuminated surface are driven into oscillation by the impinging light. They emit radiant energy, which is ultimately "reflected" backward, absorbed by the material, or both. In any case, the incident wave and the electron-oscillator fields superimpose in such a way as to yield zero light at any point beyond the screen. This might seem a remarkably special balance, but it actually is not. If the incident wave were not canceled completely, it would propagate deeper into the material of the screen, exciting more electrons to radiate. This in turn would further weaken the wave until it ultimately vanished (if the screen were thick enough). Even an ordinarily opaque material such as silver, in the form of a sufficiently thin sheet, is partially transparent (recall the half-silvered mirror).

Now, remove a small disk-shaped segment from the center of the screen, so that light streams through the aperture. The oscillators that uniformly cover the disk are removed along with it, so the remaining electrons within the screen are no longer affected by them. As a first and certainly approximate approach, *assume that the mutual interaction of the oscillators is essentially negligible*; that is, the electrons in the screen are completely unaffected by the removal of the electrons in the disk. The field in the region beyond the aperture will then be that which existed before the removal of the disk, namely zero, minus the contribution from the disk alone. Except for the sign, it is as if the source and screen had been taken away, leaving only the oscillators on the disk, rather than vice versa. In other words, the diffraction field can be pictured as arising exclusively from a set of fictitious noninteracting oscillators distributed uniformly over the region of the aperture. This of course, is the essence of the Huygens–Fresnel Principle.

*A vectorial formulation of the scalar Kirchhoff theory is discussed in J. D. Jackson, *Classical Electrodynamics*, p. 283. Also see Sommerfeld, *Optics*, p. 325. You might as well take a look at B. B. Baker and E. T. Copson, *The Mathematical Theory of Huygens's Principle*, as a general reference to diffraction. None of these texts is easy reading.

†C. L. Andrews, *Am. J. Phys.* **19**, 250 (1951); S. Silver, *J. Opt. Soc. Am.* **52**, 131 (1962).

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027



Ripple-tank photos. In one case, the waves are simply diffracted by a slit; in the other, a series of equally spaced point sources span the aperture and generate a similar pattern. (Photos courtesy PSSC *Physics*, D. C. Heath, Boston, 1960.)

We can expect, however, that instead of no interaction at all between electron-oscillators, there is a short-range effect, since the oscillator fields drop off with distance. In this physically more realistic view, the electrons within the vicinity of the aperture's edge are affected when the disk is removed. For larger apertures, the number of oscillators in the disk is much greater than the number along the edge. In such cases, if the point of observation is far away and in the forward direction, the Huygens–Fresnel Principle should, and does, work well (see photos). For very small apertures, or at points of observation in the vicinity of the aperture, edge effects become important, and we can anticipate difficulties. Indeed, at a point within the aperture itself, the electron-oscillators on the edge are of the greatest significance because of their proximity. Yet these electrons were certainly not unaffected by the removal of the adjacent oscillators of the disk. Thus, the deviation from the Huygens–Fresnel Principle should be appreciable.

10.1.2 Fraunhofer and Fresnel Diffraction

Imagine that we have an opaque screen, Σ, containing a single small aperture, which is being illuminated by plane waves from a very distant point source, S. The plane of observation σ is a screen parallel with, and very close to, Σ. Under these conditions an image of the aperture is projected onto the screen, which is clearly recognizable despite some slight fringing around its periphery (Fig. 10.2). If the plane of observation is

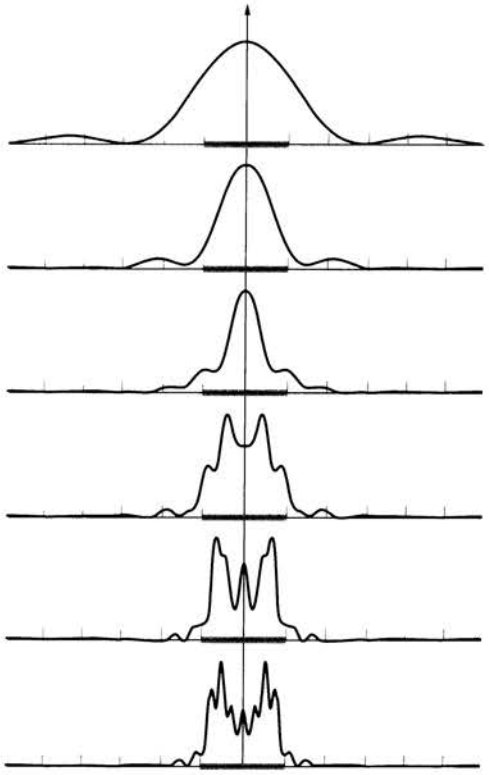


Figure 10.2 A succession of diffraction patterns at increasing distance form a single slit; Fresnel at the bottom (nearby), going toward Fraunhofer at the top (faraway). (Adapted from *Fundamentals of Waves and Oscillations* by K. U. Ingard.)

moved farther away from Σ, the image of the aperture, though still easily recognizable, becomes increasingly more structured as the fringes become more prominent. This phenomenon is known as **Fresnel** or **near-field** diffraction. If the plane of observation is moved out still farther, a continuous change in the fringes results. At a very great distance from Σ the projected pattern will have spread out considerably, bearing little or no resemblance to the actual aperture. Thereafter moving σ essentially changes only the size of the pattern and not its shape. This is **Fraunhofer** or **far-field** diffraction. If at that point we could sufficiently reduce the wavelength of the incoming radiation, the pattern would revert to the Fresnel case. If λ were decreased even more, so that it approached zero, the fringes would disappear, and the image would take on the limiting shape of the aperture, as predicted by Geometrical Optics. Returning to the original setup, if the point source

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

was now moved toward Σ, spherical waves would impinge on the aperture, and a Fresnel pattern would exist, even on a distant plane of observation.

Consider a point source S and a point of observation P, where both are very far from Σ and no lenses are present (Problem 10.1). ***As long as both the incoming and outgoing waves approach being planar (differing therefrom by a small fraction of a wavelength) over the extent of the diffracting apertures (or obstacles), Fraunhofer diffraction obtains.*** Another way to appreciate this is to realize that the *phase* of each contribution at P, due to differences in the path traversed, is crucial to the determination of the resultant field. Moreover, if the wavefronts impinging on, and emerging from, the aperture are planar, then these path differences will be describable by a linear function of the two aperture variables. ***This linearity in the aperture variables is the definitive mathematical criterion of Fraunhofer diffraction.*** On the other hand, when S or P or both are too near Σ for the curvature of the incoming and outgoing wavefronts to be negligible, Fresnel diffraction prevails.

Each point on the aperture is to be visualized as a source of Huygens wavelets, and we should be a little concerned about their relative strengths. When S is nearby, compared with the size of the aperture, a spherical wavefront will illuminate the hole. The distances from S to each point on the aperture will be different, and the strength of the incident electric field (which drops off inversely with distance) will vary from point to point over the diffracting screen. That would not be the case for incoming homogeneous plane waves. Much the same thing is true for the diffracted waves going from the aperture to P. Even if they are all emitted with the same amplitude, if P is nearby, the waves converging on it are spherical and vary in amplitude, because of the different distances from various parts of the aperture to P. Ideally, for P at infinity (whatever that means) the waves arriving there will be planar, and we need not worry about differences in field strength. That too contributes to the simplicity of the limiting Fraunhofer case.

As a practical rule-of-thumb, Fraunhofer diffraction will occur at an aperture (or obstacle) of greatest width a when

$$R > a^2/\lambda$$

where R is the smaller of the two distances from S to Σ and Σ to P (Problem 10.1). Of course, when $R = \infty$ the finite size of the aperture is of little concern. Moreover, an increase in λ clearly shifts the phenomenon toward the Fraunhofer extreme.

A practical realization of the Fraunhofer condition, where both S and P are effectively at infinity, is achieved by using an arrangement equivalent to that of Fig. 10.3. The point source S is located at F_1, the principal focus of lens L_1, and the plane of observation is the second focal plane of L_2. In the terminology

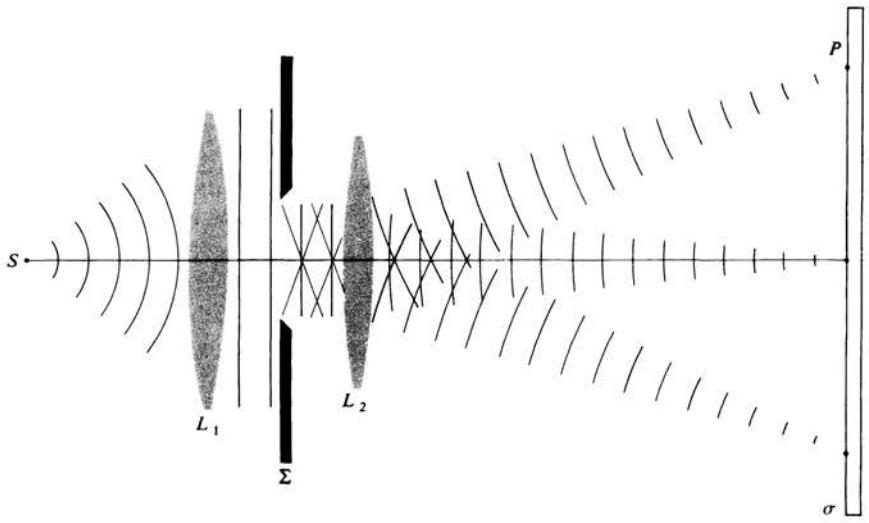


Figure 10.3 Fraunhofer diffraction using lenses so that the source and fringe pattern can both be at convenient distances from the aperture.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

of Geometrical Optics, the source plane and σ are conjugate planes.

These same ideas can be generalized to any lens system forming an image of an extended source or object (Problem 10.5).* Indeed, the image would be a Fraunhofer diffraction pattern. It is because of these important practical considerations, as well as the inherent simplicity of Fraunhofer diffraction, that we will examine it before Fresnel diffraction, even though it is a special case of the latter.

10.1.3 Several Coherent Oscillators

As a simple yet logical bridge between the studies of interference and diffraction, consider the arrangement in Fig. 10.3. The illustration depicts a linear array of N coherent point oscillators (or radiating antennas), which are all identical, even to their polarization. For the moment, assume that the oscillators have no intrinsic phase difference; that is, they each have the same initial phase angle. The rays shown are all almost parallel, meeting at some very distant point P. If the spatial extent of the array is comparatively small, the separate wave amplitudes arriving at P will be essentially equal, having traveled nearly equal distances, that is,

$$E_0(r_1) = E_0(r_2) = \cdots = E_0(r_N) = E_0(r)$$

The sum of the interfering spherical wavelets yields an electric field at P, given by the real part of

$$\tilde{E} = E_0(r)e^{i(kr_1 - \omega t)} + E_0(r)e^{i(kr_2 - \omega t)} + \cdots + E_0(r)e^{i(kr_N - \omega t)} \tag{10.1}$$

It should be clear, from Section 9.1, that we need not be concerned with the vector nature of the electric field for this configuration. Now then

$$\tilde{E} = E_0(r)e^{-i\omega t}e^{ikr_1} \times [1 + e^{ik(r_2 - r_1)} + e^{ik(r_3 - r_1)} + \cdots + e^{ik(r_N - r_1)}]$$

The phase difference between adjacent sources is obtained from the expression $\delta = k_0\Lambda$, and since $\Lambda = nd \sin\theta$, in a medium of index n, $\delta = kd \sin\theta$. Making use of Fig. 10.4, it follows that $\delta = k(r_2 - r_1)$, $2\delta = k(r_3 - r_1)$, and so on. Thus the field at P may be written as

$$\tilde{E} = E_0(r)e^{-i\omega t}e^{ikr_1} \times [1 + (e^{i\delta}) + (e^{i\delta})^2 + (e^{i\delta})^3 + \cdots + (e^{i\delta})^{N-1}] \tag{10.2}$$

The bracketed geometric series has the value

$$(e^{i\delta N} - 1)/(e^{i\delta} - 1)$$

which can be rearranged into the form

$$\frac{e^{i N\delta/2}[e^{iN\delta/2} - e^{-iN\delta/2}]}{e^{i\delta/2}[e^{i\delta/2} - e^{-i\delta/2}]}$$

or equivalently

$$e^{i(N-1)\delta/2}\left(\frac{\sin N\delta/2}{\sin \delta/2}\right)$$

The field then becomes

$$\tilde{E} = E_0(r)e^{-i\omega t}\, e^{i[kr_1 + (N-1)\delta/2]}\left(\frac{\sin N\delta/2}{\sin\delta/2}\right) \tag{10.3}$$

Notice that if we define R as the distance from the center of the line of oscillators to the point P, that is,

$$R = \tfrac{1}{2}(N - 1)d \sin\theta + r_1$$

then Eq. (10.3) takes on the form

$$\tilde{E} = E_0(r)e^{i(kR - \omega t)}\left(\frac{\sin N\delta/2}{\sin \delta/2}\right) \tag{10.4}$$

Finally, then, the flux-density distribution within the diffraction pattern due to N coherent, identical, distant point sources in a linear array is proportional to $\tilde{E}\tilde{E}^*/2$ for complex E or

$$I = I_0\frac{\sin^2(N\delta/2)}{\sin^2(\delta/2)} \tag{10.5}$$

where I_0 is the flux density from any single source arriving at P. (See Problem 10.2 for a graphic derivation of the irradiance.) For $N = 0$, $I = 0$, for $N = 1$, $I = I_0$, and for $N = 2$, $I = 4I_0 \cos^2(\delta/2)$, in accord with Eq. (9.17). The functional dependence of I on θ is more apparent in the form

*A He–Ne laser can be set up to generate magnificent patterns without any auxiliary lenses, but this requires plenty of space.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

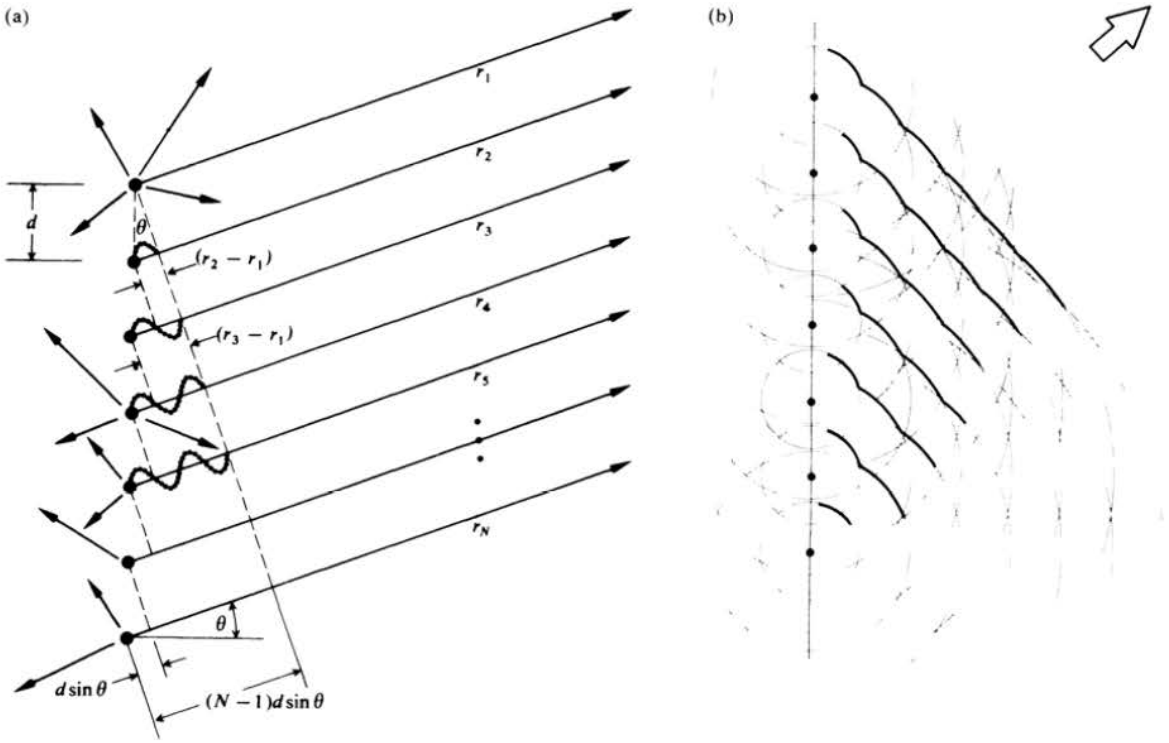


Figure 10.4 A linear array of in-phase coherent oscillators. (*a*) Note that at the angle shown $\delta = \pi$ while at $\theta = 0$, δ would be zero. (*b*) One of many sets of wavefronts emitted from a line of coherent point sources.

$$I = I_0 \frac{\sin^2[N(kd/2)\sin\theta]}{\sin^2[(kd/2)\sin\theta]} \tag{10.6}$$

The $\sin^2[N(kd/2)\sin\theta]$ term undergoes rapid fluctuations, whereas the function that modulates it, $\{\sin[(kd/2)\sin\theta]\}^{-2}$, varies relatively slowly. The combined expression gives rise to a series of sharp principal peaks separated by small subsidiary maxima. The principal maxima occur in directions θ_m, such that $\delta = 2m\pi$, where $m = 0, \pm 1, \pm 2, \ldots$. Because $\delta = kd \sin\theta$,

$$d \sin\theta_m = m\lambda \tag{10.7}$$

Since $[\sin^2 N\delta/2]/[\sin^2\delta/2] = N^2$ for $\delta = 2m\pi$ (from L'Hospital's rule), the principal maxima have values of $N^2 I_0$. This is to be expected, inasmuch as all the oscillators are in-phase at that orientation. The system will radiate a maximum in a direction perpendicular to the array ($m = 0$, $\theta_0 = 0$ and π). As θ increases, δ increases and I falls off to zero at $N\delta/2 = \pi$, its first minimum. Note that if $d < \lambda$ in Eq. (10.7), only the $m = 0$ or zero-order principal maximum exists. *If we were looking at an idealized line source of electron-oscillators separated by atomic distances, we could expect only that one principal maximum in the light field.*

An antenna array like the one in the accompanying photo can transmit radiation in the narrow beam or lobe corresponding to a principal maximum. (Parabolic dishes reflect in the forward direction, and the radiation pattern is no longer symmetrical around the common axis.) Suppose that we have a system in which we can introduce an intrinsic phase shift of ε between adjacent oscillators. In that case

$$\delta = kd \sin\theta + \varepsilon$$

The various principal maxima will occur at new angles

$$d \sin\theta_m = m\lambda - \varepsilon/k$$

Concentrating on the central maximum $m = 0$, we can vary its orientation θ_0 at will by merely adjusting the value of ε.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027



An early interferometric radio telescope at the University of Sydney, Australia ($N = 32$, $\lambda = 21$ cm, $d = 7$ m, 2 mi diameter, 700 ft. east–west base line). (Photo courtesy of W. N. Christiansen.)

The Principle of Reversibility, which states that without absorption, wave motion is reversible, leads to the same field pattern for an antenna used as either a transmitter or a receiver. The array, functioning as a radio telescope, can therefore be "pointed" by combining the output from the individual antennas with an appropriate phase shift, ε, introduced between each of them. For a given ε the output of the system corresponds to the signal impinging on the array from a specific direction in space (see the discussion of phased array radar, p.98).

The telescope in the above photograph shows the first multiple radio interferometer, designed by W. N. Christiansen and built in Australia in 1951. It consisted of 32 parabolic antennas, each 2 m in diameter, designed to function in-phase at the wavelength of the 21-cm hydrogen emission line. The antennas are arranged along an east–west base line with 7 m separating each one. This particular array utilized the Earth's rotation as the scanning mechanism.*

Examine Fig. 10.5, which depicts an idealized line source of electron-oscillators (e.g., the secondary sources of the Huygens–Fresnel Principle for a long slit whose width is much less than λ, illuminated by plane waves). Each point emits a spherical wavelet, which we write as

$$E = \left(\frac{\mathcal{E}_0}{r}\right) \sin(\omega t - kr)$$

explicitly indicating the inverse r-dependence of the amplitude. The quantity $\mathcal{E}_0$ is said to be the **source strength**. The present situation is distinct from that of Fig. 10.4, since now the sources are very weak, their number, N, is tremendously large, and the separation between them is vanishingly small. A minute but finite segment of the array Δy_i will contain $\Delta y_i(N/D)$ sources, where D is the entire length of the array. Imagine that the array is divided up into M such segments (i.e., i goes from 1 to M). The contribution to the electric-field intensity at P from the ith segment is accordingly

$$E_i = \left(\frac{\mathcal{E}_0}{r_i}\right) \sin(\omega t - kr_i)\left(\frac{N\Delta y_i}{D}\right)$$

provided that Δy_i is so small that the oscillators within it have a negligible relative phase difference (r_i = constant), and their fields simply add constructively. We can cause the array to become a continuous (coherent) line source by letting N approach infinity. This description, besides being fairly realistic on a macroscopic scale, also allows the use of the calculus for more complicated geometries. Certainly as N approaches infinity, the source strengths of the individual oscillators must diminish to nearly zero, if the total output is to be finite. We can therefore define a constant $\mathcal{E}_L$ as the ***source strength per unit length*** of the array, that is,

$$\mathcal{E}_L \equiv \frac{1}{D} \lim_{N\to\infty} (\mathcal{E}_0 N) \tag{10.8}$$

The net field at P from all M segments is

$$E = \sum_{i=1}^{M} \frac{\mathcal{E}_L}{r_i} \sin(\omega t - kr_i)\Delta y_i$$

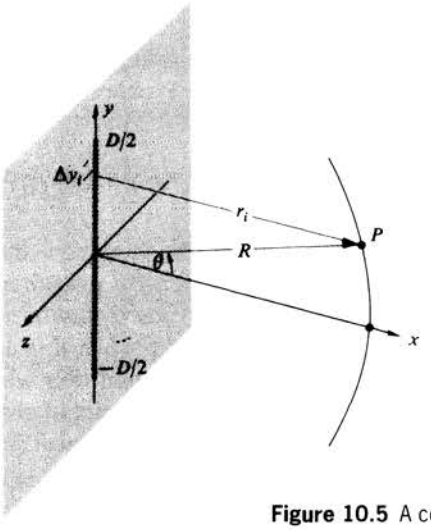


Figure 10.5 A coherent line source.

*See E. Brookner, "Phased-Array Radars," *Sci. Am.* (Feb. 1985), p. 94.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

For a continuous line source the Δy_i must become infinitesimal ($M \to \infty$), and the summation is then transformed into a definite integral

$$E = \mathcal{E}_L \int_{-D/2}^{+D/2} \frac{\sin(\omega t - kr)}{r}\, dy \qquad (10.9)$$

where $r = r(y)$. The approximation used to evaluate Eq. (10.9) must depend on the position of P with respect to the array and will therefore make the distinction between Fraunhofer and Fresnel diffraction. The coherent optical line source does not exist as a physical entity, but we will make good use of it as a mathematical device.

10.2 Fraunhofer Diffraction

10.2.1 The Single Slit

Return to Fig. 10.5, where now the point of observation is very distant from the coherent line source and $R >> D$. Under these circumstances $r(y)$ never deviates appreciably from its midpoint value R, so that the quantity $(\mathcal{E}_L/R)$ at P is essentially constant for all elements dy. It follows from Eq. (10.9) that the field at P due to the differential segment of the source dy is

$$dE = \frac{\mathcal{E}_L}{R} \sin(\omega t - kr)\, dy \qquad (10.10)$$

where $(\mathcal{E}_L/R)\, dy$ is the amplitude of the wave. Notice that the phase is much more sensitive to variations in $r(y)$ than is the amplitude, so that we will have to be more careful about introducing approximations into it. We can expand $r(y)$, in precisely the same manner as was done in Problem (9.15), to make it an explicit function of y; thus

$$r = R - y \sin\theta + (y^2/2R)\cos^2\theta + \cdots \qquad (10.11)$$

where θ is measured from the xz-plane. The third term can be ignored as long as its contribution to the phase is insignificant even when $y = \pm D/2$; that is, $(\pi D^2/4\lambda R)\cos^2\theta$ must be negligible. This will be true for all values of θ when R is adequately large. We now have the **Fraunhofer condition**, where the distance r is linear in y: the distance to the point of observation and therefore the phase can be written as a linear function of the aperture variables. Substituting into Eq. (10.10) and integrating leads to

$$E = \frac{\mathcal{E}_L}{R} \int_{-D/2}^{+D/2} \sin[\omega t - k(R - y\sin\theta)]\, dy \qquad (10.12)$$

and finally

$$E = \frac{\mathcal{E}_L D}{R} \frac{\sin[(kD/2)\sin\theta]}{(kD/2)\sin\theta} \sin(\omega t - kR) \qquad (10.13)$$

To simplify the appearance of things, let

$$\beta \equiv (kD/2)\sin\theta \qquad (10.14)$$

so that

$$E = \frac{\mathcal{E}_L D}{R} \left(\frac{\sin\beta}{\beta}\right) \sin(\omega t - kR) \qquad (10.15)$$

The quantity most readily measured is the irradiance (forgetting the constants) $I(\theta) = \langle E^2 \rangle_T$ or

$$I(\theta) = \frac{1}{2}\left(\frac{\mathcal{E}_L D}{R}\right)^2 \left(\frac{\sin\beta}{\beta}\right)^2 \qquad (10.16)$$

where $\langle \sin^2(\omega t - kR) \rangle_T = \frac{1}{2}$. When $\theta = 0$, $\sin\beta/\beta = 1$ and $I(\theta) = I(0)$, which corresponds to the *principal maximum. The irradiance resulting from an idealized coherent line source in the Fraunhofer approximation* is then

$$I(\theta) = I(0)\left(\frac{\sin\beta}{\beta}\right)^2 \qquad (10.17)$$

or, using the *sinc function* (see Table 1 of the Appendix and p. 49),

$$I(\theta) = I(0)\,\mathrm{sinc}^2\beta$$

There is symmetry about the y-axis, and this expression holds for θ measured in any plane containing that axis. Notice that since $\beta = (\pi D/\lambda)\sin\theta$, when $D >> \lambda$, the irradiance drops extremely rapidly as θ deviates from zero. This arises from the fact that β becomes very large for large values of length D (a centimeter or so when using light). The phase of the line source is equivalent, by way of Eq. (10.15), to that of a point source located at the center of the array, a distance R from P.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

Finally, a relatively long coherent line source ($D >> \lambda$) can be envisioned as a single-point emitter radiating predominantly in the forward, $\theta = 0$, direction; in other words, its emission resembles a circular wave in the xz-plane. In contrast, notice that if $\lambda >> D$, β is small, $\sin\beta \approx \beta$, and $I(\theta) \approx I(0)$. The irradiance is then constant for *all* θ, and the line source resembles a point source emitting spherical waves.

We can now turn our attention to the problem of Fraunhofer diffraction by a slit or elongated narrow rectangular hole (Fig. 10.6). An aperture of this sort might typically have a width of several hundred λ and a length of a few centimeters. The usual procedure to follow in the analysis is to divide the slit into a series of long differential strips (dz by ℓ) parallel to the y-axis, as shown in Fig. 10.7 on page 454. We immediately recognize, however, that each strip is a long coherent line source and can therefore be replaced by a point emitter on the z-axis. In effect, each such emitter radiates a circular wave in the ($y = 0$ or) xz-plane. This is certainly reasonable, since the slit is long and the merging wavefronts are practically unobstructed in the slit direction. There will thus be very little diffraction parallel to the edges of the slit. The problem has been reduced to that of finding the field in the xz-plane due to an infinite number of point sources extending across the width of the slit along the z-axis. We then need only evaluate the integral of the contribution dE from each element dz in the Fraunhofer approximation. But once again, this is equivalent to a coherent line source, so that the complete solution for the slit is, as we have seen,

$$I(\theta) = I(0)\left(\frac{\sin\beta}{\beta}\right)^2 \qquad [10.17]$$

provided that

$$\beta = (kb/2)\sin\theta \qquad (10.18)$$

and θ is measured from the xy-plane (see Problem 10.3). Note that here the line source is short, $D = b$, β is not large, and although the irradiance falls off rapidly, higher-order sub-

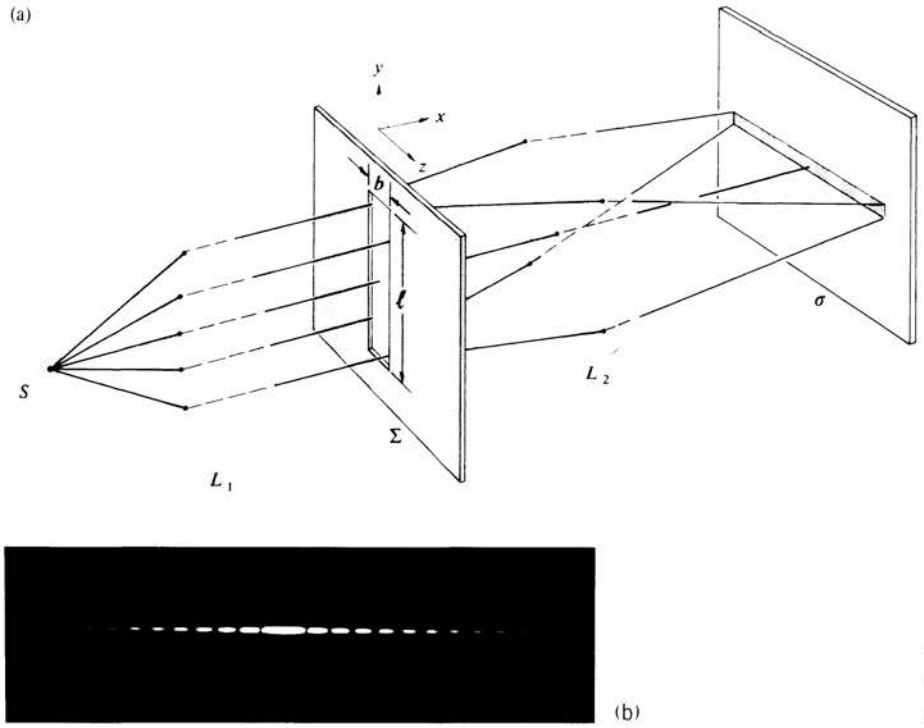


Figure 10.6 (*a*) Single-slit Fraunhofer diffraction. (*b*) Diffraction pattern of a single vertical slit under point-source illumination.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

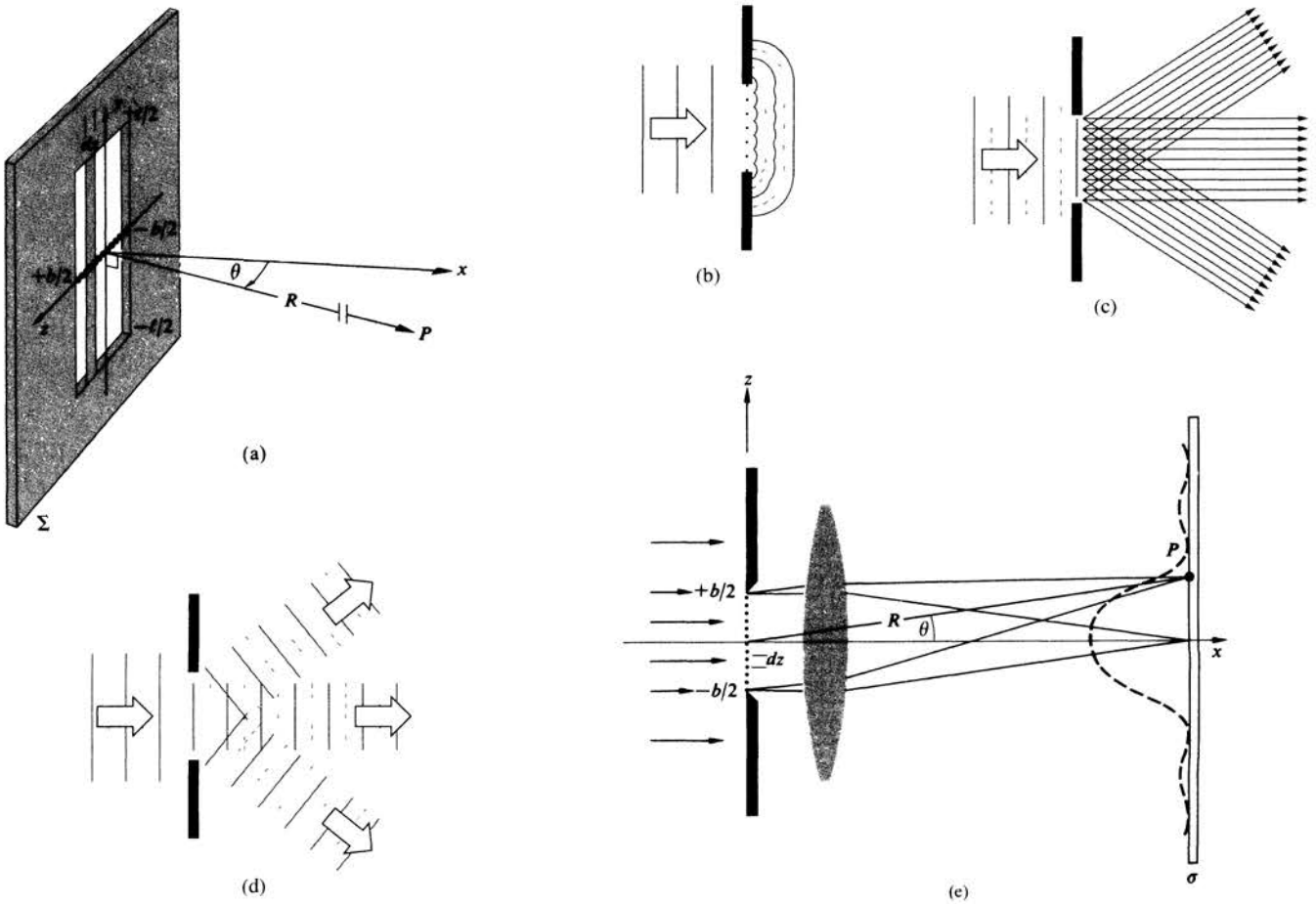


Figure 10.7 (*a*) Point P on σ is essentially infinitely far from Σ. (*b*) Huygens wavelets emitted across the aperture. (*c*) The equivalent representation in terms of rays. Each point emits rays in all directions. The parallel rays in various directions are seen. (*d*) These ray bundles correspond to plane waves, which can be thought of as the three-dimensional Fourier components. (*e*) A single slit illuminated by monochromatic plane waves.

sidiary maxima will be observable. The extrema of $I(\theta)$ occur at values of β that cause $dI/d\beta$ to be zero, that is,

$$\frac{dI}{d\beta} = I(0)\frac{2\sin\beta(\beta\cos\beta - \sin\beta)}{\beta^3} = 0 \quad (10.19)$$

The irradiance has minima, equal to zero, when $\sin\beta = 0$, whereupon

$$\beta = \pm\pi, \pm 2\pi, \pm 3\pi, \ldots \quad (10.20)$$

It also follows from Eq. (10.19) that when

$$\beta\cos\beta - \sin\beta = 0$$

$$\tan\beta = \beta \quad (10.21)$$

The solutions to this transcendental equation can be determined graphically, as shown in Fig. 10.8. The points of intersection of the curves $f_1(\beta) = \tan\beta$ with the straight line $f_2(\beta) = \beta$ are common to both and so satisfy Eq. (10.21). Only one such extremum exists between adjacent minima [Eq. (10.20)],

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

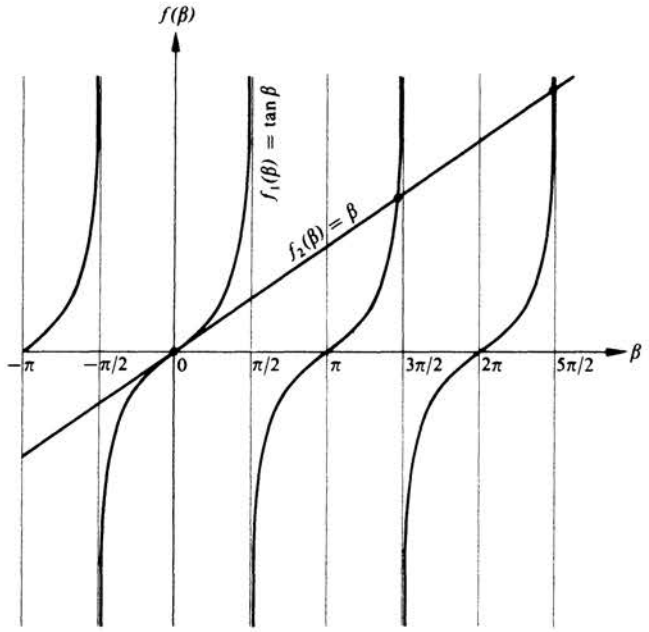


Figure 10.8 The points of intersection of the two curves are the solutions of Eq. (10.21).

so that $I(\theta)$ must have subsidiary maxima at these values of β (viz, $\pm 1.430\,3\pi$, $\pm 2.459\,0\pi$, $\pm 3.470\,7\pi$, ...).

There is a particularly easy way to appreciate what's happening here with the aid of Fig. 10.9. We envision every point in the aperture emitting rays in all directions in the xz-plane. The light that continues to propagate directly forward in Fig. 10.9*a* is the undiffracted beam, all the rays arrive on the viewing screen in-phase, and a central bright spot will be formed by them. If the screen is not actually at infinity, the rays that converge to it are not quite parallel but with it at infinity, or better still, with a lens in place, the rays are as drawn. Figure 10.9*b* shows the specific bundle of rays coming off at an angle θ_1 where the path length difference between the rays from the very top and bottom, $b \sin \theta_1$, is made equal to one wavelength. A ray from the middle of the slit will then lag $\frac{1}{2}\lambda$ behind a ray from the top and exactly cancel it. Similarly, a ray from just below center will cancel a ray from just below the top, and so on; all across the aperture ray-pairs will cancel, yielding a minimum. The irradiance has dropped from its high central maximum to the first zero on either side at $\sin \theta_1 = \pm\lambda/b$.

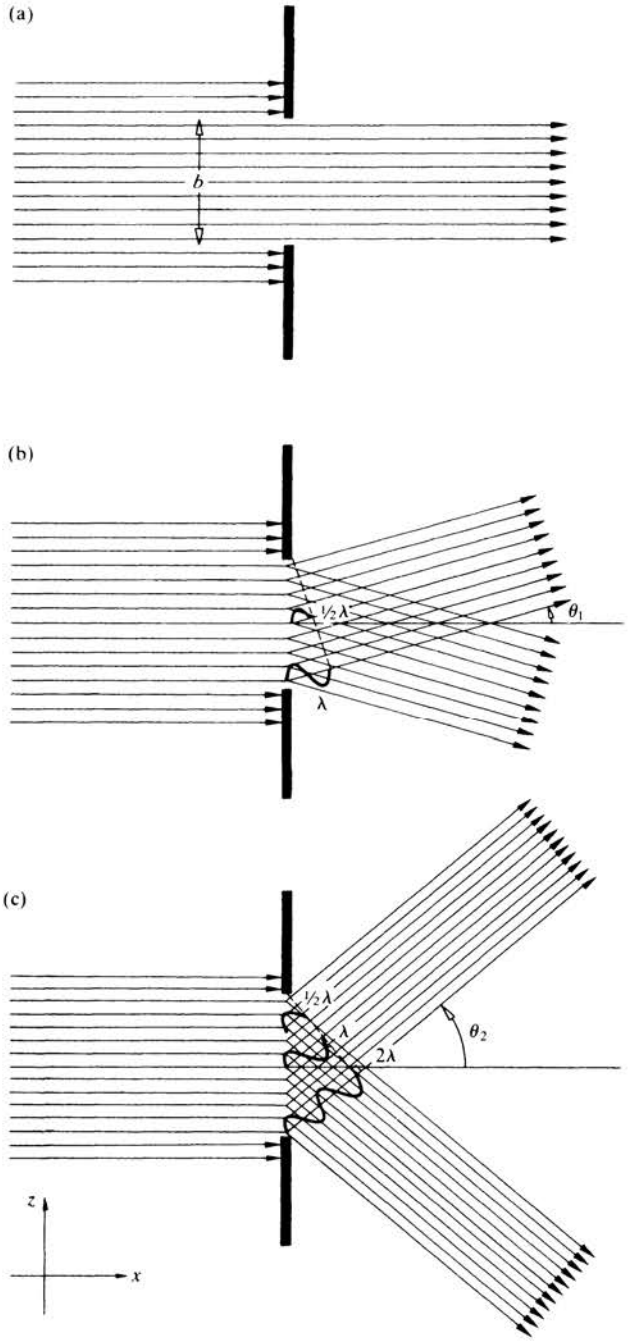


Figure 10.9 The diffraction of light in various directions. Here the aperture is a single slit, as in Fig. 10.7.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

As the angle increases further, some small fraction of the rays will again interfere constructively, and the irradiance will rise to form a subsidiary peak. A further increase in the angle produces another minimum, as shown in Fig. 10.9*c*, when $b \sin \theta_2 = 2\lambda$. Now imagine the aperture divided into quarters. Ray by ray, the top quarter will cancel the one beneath it, and the next, the third, will cancel the last quarter. Ray-pairs at the same locations in adjacent segments are $\lambda/2$ out-of-phase and destructively interfere. In general then, zeros of irradiance will occur when

$$b \sin \theta_m = m\lambda$$

where $m = \pm 1, \pm 2, \pm 3, \ldots$, which is equivalent to Eq. (10.20), since $\beta = m\pi = (kb/2) \sin \theta_m$.

We should inject a note of caution at this point: one of the frailties of the Huygens–Fresnel Principle is that it does not take proper regard of the variations in amplitude, with angle, over the surface of each secondary wavelet. We will come back to this when we consider the *obliquity factor* in Fresnel diffraction, where the effect is significant. In Fraunhofer diffraction the distance from the aperture to the plane of observation is so large that we need not be concerned about it, provided that θ remains small.

Figure 10.10 is a plot of the flux density, as expressed by Eq. (10.17). Envision some point on the curve, for example, the third subsidiary maximum at $\beta = 3.4707\pi$; since $\beta = (\pi b/\lambda) \sin \theta$, an increase in the slit width b requires a decrease in θ, if β is to be constant. Under these conditions the pattern shrinks in toward the principal maximum, as it would if λ were decreased. If the source emits white light, the higher-order maxima show a succession of colors trailing off into red with increasing θ. Each different colored light component has its minima and subsidiary maxima at angular positions characteristic of that wavelength (Problem 10.6). Indeed, only in

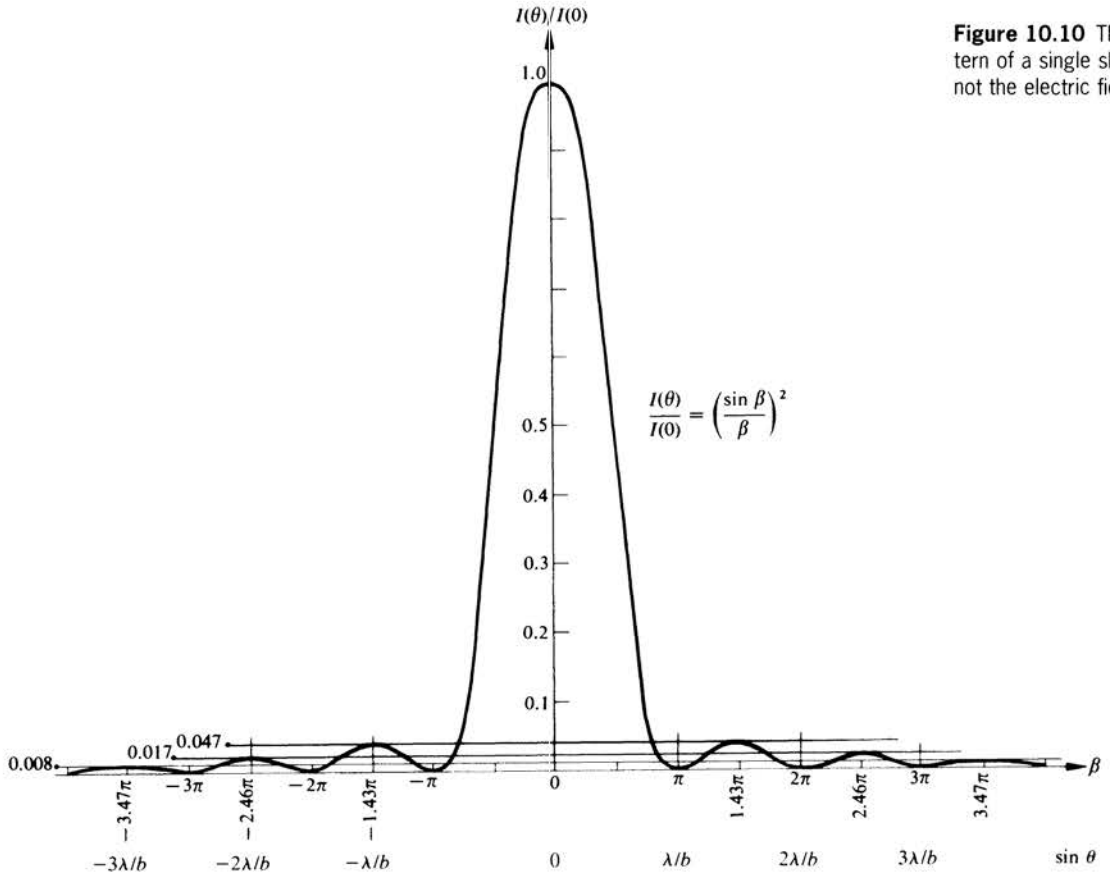


Figure 10.10 The Fraunhofer diffraction pattern of a single slit. This is the irradiance (and not the electric field) distribution

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

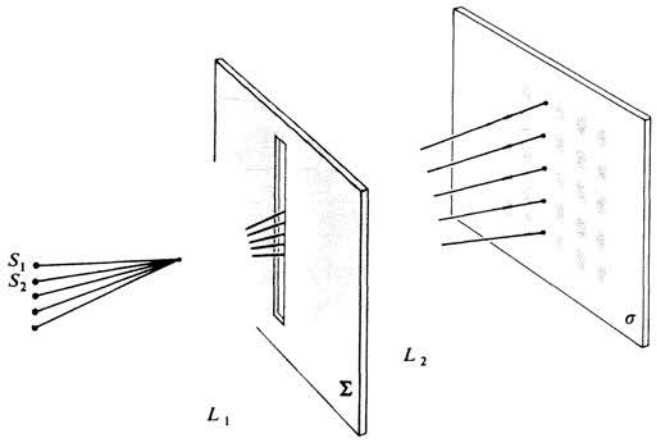


Figure 10.11 The single-slit pattern with a line source. See first photograph of Fig. 10.14.

the region about $\theta = 0$ will all the constituent colors overlap to yield white light.

The point source S in Fig. 10.6 would be imaged at the position of the center of the pattern, if the diffracting screen Σ were removed. Under this sort of illumination, the pattern produced with the slit in place is a series of dashes in the yz-plane of the screen σ, much like a spread-out image of S (Fig. 10.6*b*). An incoherent line source (in place of S) positioned parallel to the slit, in the focal plane of the collimator L_1, will broaden the pattern out into a series of bands. Any point on the line source generates an independent diffraction pattern, which is displaced, with respect to the others, along the y-direction. With no diffracting screen present, the image of the line source would be a line parallel to the original slit. With the screen in place the line is spread out, as was the point image of S (Fig. 10.11). Keep in mind that it's the small dimension of the slit that does the spreading out.

The single-slit pattern is easily observed without the use of special equipment. Any number of sources will do (e.g., a distant street light at night, a small incandescent lamp, sunlight streaming through a narrow space in a window shade); almost anything that resembles a point or line source will serve. Probably the best source for our purposes is an ordinary clear, *straight-filament* display bulb (the kind in which the filament is vertical and about 3 inches long). You can use your imagination to generate all sorts of single-slit arrangements (e.g., a comb or fork rotated to decrease the projected space between the tines, or a scratch across a layer of india ink on a microscope slide). An inexpensive vernier caliper makes a remarkably good variable slit. Hold the caliper close to your eye with the slit, a few thousandths of an inch wide, parallel to the filament of the lamp. Focus your eye beyond the slit at infinity, so that its lens serves as L_2.

10.2.2 The Double Slit

It might at first seem from Fig. 10.7 that the location of the principal maximum is always to be in line with the center of the diffracting aperture; this, however, is not generally true. The diffraction pattern is actually centered about the axis of the lens and has exactly the same shape and location, regardless of the slit's position, as long as its orientation is unchanged and the approximations are valid (Fig. 10.12). All waves traveling parallel to the lens axis converge on the second focal point of L_2; this then is the image of S and the center of the diffraction pattern. Suppose now that we have two long slits of width b and center-to-center separation a (Fig. 10.13). Each aperture, by itself, would generate the same single-slit diffraction pattern on the viewing screen σ. At any point on σ, the contributions from the two slits overlap, and even though each must be essentially equal in amplitude, they may well differ significantly in-phase. Since the same primary wave

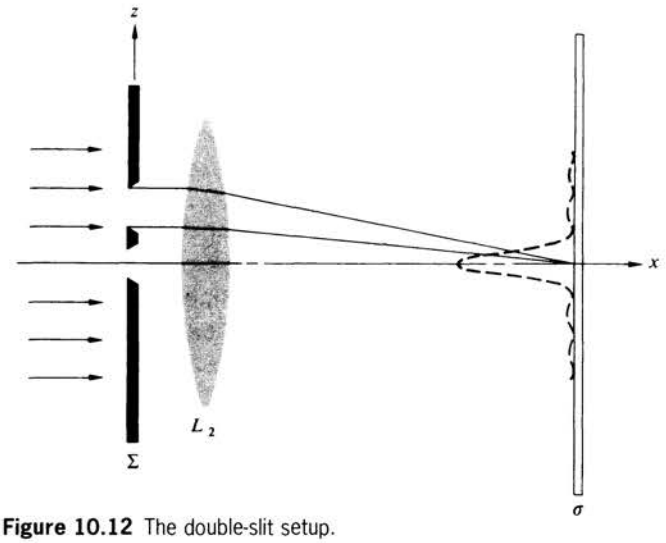


Figure 10.12 The double-slit setup.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

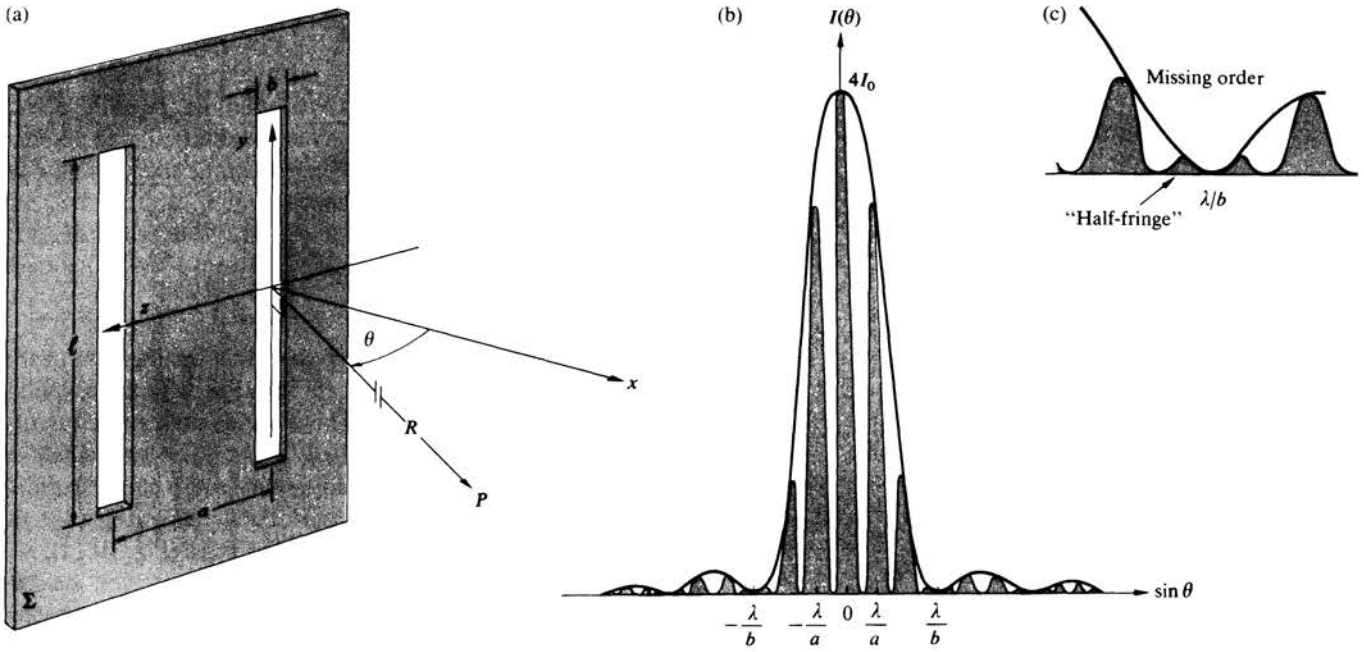


Figure 10.13 (*a*) Double-slit geometry. Point P on σ is essentially infinitely far away. (*b*) A double-slit pattern ($a = 3b$).

excites the secondary sources at each slit, the resulting wavelets will be coherent, and interference must occur. If the primary plane wave is incident on Σ at some angle θ_i (see Problem 10.3), there will be a constant relative phase difference between the secondary sources. At normal incidence, the wavelets are all emitted in-phase. The interference fringe at a particular point of observation is determined by the differences in the optical path lengths traversed by the overlapping wavelets from the two slits. As we will see, the flux-density distribution (Fig. 10.14) is the result of a rapidly varying double-slit interference system modulated by a single-slit diffraction pattern.

To obtain an expression for the optical disturbance at a point on σ, we need only slightly reformulate the single-slit analysis. Each of the two apertures is divided into differential strips (dz by ℓ), which in turn behave like an infinite number of point sources aligned along the z-axis. The total contribution to the electric field, in the Fraunhofer approximation [Eq. (10.12)], is then

$$E = C\int_{-b/2}^{b/2} F(z)\,dz + C\int_{a-b/2}^{a+b/2} F(z)\,dz \qquad (10.22)$$

where $F(z) = \sin[\omega t - k(R - z\sin\theta)]$. The constant-amplitude factor C is the secondary source strength per unit length along the z-axis (assumed to be independent of z over each aperture) divided by R, which is measured from the origin to P and is taken as constant. We will be concerned only with relative flux densities on σ, so that the actual value of C is of little interest to us now. Integration of Eq. (10.22) yields

$$E = bC\left(\frac{\sin\beta}{\beta}\right)[\sin(\omega t - kR) + \sin(\omega t - kR + 2\alpha)] \qquad (10.23)$$

with $\alpha \equiv (ka/2)\sin\theta$ and, as before, $\beta \equiv (kb/2)\sin\theta$. This is just the sum of the two fields at P, one from each slit, as given by Eq. (10.15). The distance from the first slit to P is R, giving a phase contribution of $-kR$. The distance from the second slit to P is $(R - a\sin\theta)$ or $(R - 2\alpha/k)$, yielding a phase term equal to $(-kR + 2\alpha)$, as in the second sine function. The quantity 2β is the phase difference ($k\Lambda$) between two nearly parallel rays, arriving at a point P on σ, from the edges of one of the slits. The quantity 2α is the phase difference between two waves arriving at P, one having originated at any point in

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

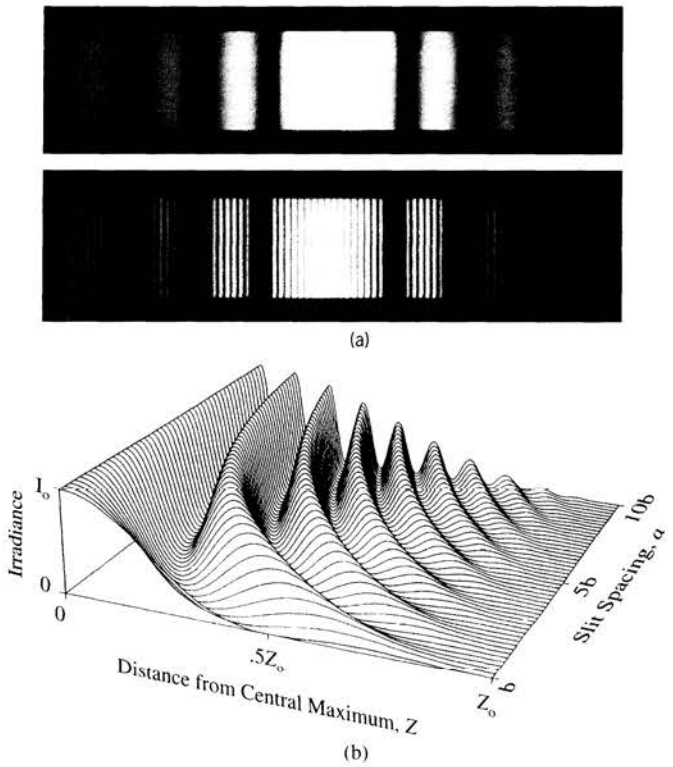


Figure 10.14 Single- and double-slit Fraunhofer patterns. (*a*) Photographs taken with monochromatic light. The faint cross-hatching arises entirely in the printing process. (Photos courtesy M. Cagnet, M. Francon, and J. C. Thrierr: *Atlas optischer Erscheinungen*, Berlin–Heidelberg–New York: Springer, 1962.) (*b*) When the slit spacing equals *b*, the two slits coalesce into one (of width 2*b*) and the single-slit pattern appears—that's the first curve closest to you. The farthest curve corresponds to the two slits separated by $a = 10b$. Notice that the two-slit patterns all have their first diffraction minimum at a distance from the central maximum of Z_0. Note how the curves gradually match Fig. 10.13*b* as the slit width *b* gets smaller in comparison to the separation *a*. (Reprinted from "Graphical Representations of Fraunhofer Interference and Diffraction" *Am. J. Phys.*, 62, 6, (1994), with permission of A. B. Bartlett, University of Colorado and B. Mechtly, Northeast Missouri State University and the American Association of Physics Teachers.)

the first slit, the other coming from the corresponding point in the second slit. Simplifying Eq. (10.23) a bit further, it becomes

$$E = 2bC\left(\frac{\sin\beta}{\beta}\right)\cos\alpha\,\sin(\omega t - kR + \alpha)$$

which when squared and averaged over a relatively long interval in time is the irradiance

$$I(\theta) = 4I_0\left(\frac{\sin^2\beta}{\beta^2}\right)\cos^2\alpha \qquad (10.24)$$

In the $\theta = 0$ direction (i.e., when $\beta = \alpha = 0$), I_0 is the flux-density contribution from either slit, and $I(0) = 4I_0$ is the total flux density. The factor of 4 comes from the fact that the amplitude of the electric field is twice what it would be at that point with one slit covered.

If in Eq. (10.24) b becomes vanishingly small ($kb << 1$), then $(\sin\beta)/\beta \approx 1$, and the equation reduces to the flux-density expression for a pair of long line sources, that is, Young's Experiment, Eq. (9.17). If on the other hand $a = 0$, the two slits coalesce into one, $\alpha = 0$, and Eq. (10.24) becomes $I(0) = 4I_0(\sin^2\beta)/\beta^2$. This is the equivalent of Eq. (10.17) for single-slit diffraction with the source strength doubled. We might then envision the total expression as being generated by a $\cos^2\alpha$ interference term modulated by a $(\sin^2\beta)/\beta^2$ diffraction term. If the slits are finite in width but very narrow, the diffraction pattern from either slit will be uniform over a broad central region, and bands resembling the idealized Young's fringes will appear within that region. At angular positions (θ-values) where

$$\beta = \pm\pi, \pm 2\pi, \pm 3\pi, \ldots$$

diffraction effects are such that no light reaches σ, and clearly none is available for interference. At points on σ where

$$\alpha = \pm\pi/2, \pm 3\pi/2, \pm 5\pi/2, \ldots$$

the various contributions to the electric field will be completely out-of-phase and will cancel, regardless of the actual amount of light made available from the diffraction process.

The irradiance distribution for a double-slit Fraunhofer pattern is illustrated in Fig. 10.13*b*. Notice that it is a combination of Figs. 9.9 and 10.6. The curve is for the particular case in which $a = 3b$ (i.e., $\alpha = 3\beta$). You can get a rough idea of what the pattern will look like, since if $a = mb$, where m is any number, there will be $2m$ bright fringes (counting "fractional fringes" as well)* within the central diffraction peak (Problem 10.10). An interference maximum and a diffraction minimum

*Notice that m need not be an integer. Moreover, if m is an integer, there will be "half-fringes," as shown in Fig. 10.13*c*.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

(zero) may correspond to the same θ-value. In that case, no light is available at that precise position to partake in the interference process, and the suppressed peak is said to be a *missing order*.

The double-slit pattern is also rather easily observed, and the seeing is well worth the effort. A straight-filament, tubular bulb is again the best line source. For slits, coat a microscope slide with India ink; if you happen to have some, a colloidal suspension of graphite in alcohol works even better (it's more opaque). Scratch a pair of slits across the dry ink with a razor blade and stand about 10 feet from the source. Hold the slits parallel to the filament and close to your eye, which, when focused at infinity, will serve as the needed lens. Interpose red or blue cellophane and observe the change in the width of the fringes. Find out what happens when you cover one and then both of the slits with a microscope slide. Move the slits slowly in the z-direction; then holding them stationary, move your eye in the z-direction. Verify that the position of the center of the pattern is indeed determined by the lens and not the aperture.

10.2.3 Diffraction by Many Slits

The procedure for obtaining the irradiance function for a monochromatic wave diffracted by many slits is essentially the same as that used when considering two slits. Here again, the limits of integration must be appropriately altered. Consider the case of N long, parallel, narrow slits, each of width b and center-to-center separation a, as illustrated in Fig. 10.15. With the origin of the coordinate system once more at the center of the first slit, the total optical disturbance at a point on the screen σ is given by

$$E = C\int_{-b/2}^{b/2} F(z)\,dz + C\int_{a-b/2}^{a+b/2} F(z)\,dz + C\int_{2a-b/2}^{2a+b/2} F(z)\,dz + \cdots + C\int_{(N-1)a-b/2}^{(N-1)a+b/2} F(z)\,dz \qquad (10.25)$$

where as before, $F(z) = \sin[\omega t - k(R - z\sin\theta)]$. This applies to the Fraunhofer condition, so that the aperture configuration must be such that all the slits are close to the origin, and the approximation [Eq. (10.11)]

$$r = R - z\sin\theta \qquad (10.26)$$

applies over the entire array. The contribution from the jth slit (where the first one is numbered zero), obtained by evaluating only that one integral in Eq. (10.25), is then

$$E_j = \frac{C}{k\sin\theta}[\sin(\omega t - kR)\sin(kz\sin\theta) - \cos(\omega t - kR)\cos(kz\sin\theta)]_{ja-b/2}^{ja+b/2}$$

provided that we require $\theta_j \approx \theta$. After some manipulation this becomes

$$E_j = bC\left(\frac{\sin\beta}{\beta}\right)\sin(\omega t - kR + 2\alpha j) \qquad (10.27)$$

recalling that $\beta = (kb/2)\sin\theta$ and $\alpha = (ka/2)\sin\theta$. Notice that this is equivalent to the expression for a line source [Eq. (10.15)] or, of course, a single slit, where in accord with Eq. (10.26) and Fig. 10.15, $R_j = R - ja\sin\theta$, so that $-kR + 2\alpha j = -kR_j$. The total optical disturbance, as given by Eq. (10.25), is simply the sum of the contributions from each of the slits; that is,

$$E = \sum_{j=0}^{N-1} E_j$$

or

$$E = \sum_{j=0}^{N-1} bC\left(\frac{\sin\beta}{\beta}\right)\sin(\omega t - kR + 2\alpha j) \qquad (10.28)$$

This in turn can be written as the imaginary part of a complex exponential:

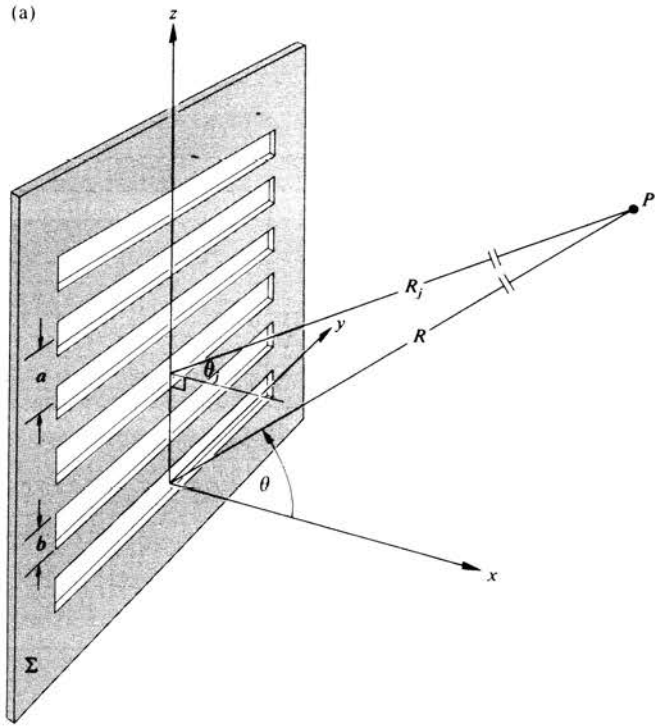
$$E = \mathrm{Im}\left[bC\left(\frac{\sin\beta}{\beta}\right)e^{i(\omega t - kR)}\sum_{j=0}^{N-1}(e^{i2\alpha})^j\right] \qquad (10.29)$$

But we have already evaluated this same geometric series in the process of simplifying Eq. (10.2). Equation (10.29) therefore reduces to the form

$$E = bC\left(\frac{\sin\beta}{\beta}\right)\left(\frac{\sin N\alpha}{\sin\alpha}\right)\sin[\omega t - kR + (N-1)\alpha] \qquad (10.30)$$

The distance from the center of the array to the point P is equal to $[R - (N-1)(a/2)\sin\theta]$, and therefore the phase of E at P corresponds to that of a wave emitted from the midpoint of the

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027



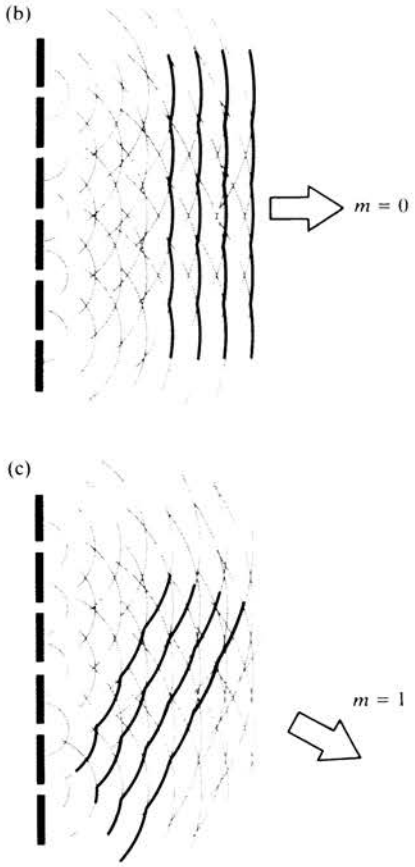


Figure 10.15 Multi-slit geometry. Again point P is on σ essentially infinitely far from Σ.

source. The flux-density distribution function is

$$I(\theta) = I_0\left(\frac{\sin\beta}{\beta}\right)^2\left(\frac{\sin N\alpha}{\sin\alpha}\right)^2 \qquad (10.31)$$

Note that I_0 is the flux density in the $\theta = 0$ direction emitted by any one of the slits and that $I(0) = N^2 I_0$. In other words, the waves arriving at P in the forward direction are all in-phase, and their fields add constructively. Each slit by itself would generate precisely the same flux-density distribution. Superimposed, the various contributions yield a multiple-wave interference system modulated by the single-slit diffraction envelope. If the width of each aperture were shrunk to zero, Eq. (10.31) would become the flux-density expression [Eq. (10.6)] for a linear coherent array of oscillators. As in that earlier treatment [Eq. (10.17)], **principal maxima** occur when $(\sin N\alpha/\sin\alpha) = N$, that is, when

$$\alpha = 0, \pm\pi, \pm 2\pi, \ldots$$

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

or equivalently, since $\alpha = (ka/2)\sin\theta$,

$$a \sin \theta_m = m\lambda \tag{10.32}$$

with $m = 0, \pm 1, \pm 2, \ldots$. This is quite general and gives rise to the same θ-locations for these maxima, regardless of the value of $N \geq 2$. Minima, of zero flux density, exist whenever $(\sin N\alpha/\sin\alpha)^2 = 0$ or when

$$\alpha = \pm\frac{\pi}{N}, \pm\frac{2\pi}{N}, \pm\frac{3\pi}{N}, \ldots, \pm\frac{(N-1)\pi}{N}, \pm\frac{(N+1)\pi}{N}, \ldots \tag{10.33}$$

Between consecutive principal maxima (i.e., over the range in α of π) there will therefore be $N - 1$ minima. And, of course, between each pair of minima there will have to be a **subsidiary maximum**. The term $(\sin N\alpha/\sin\alpha)^2$, which we can think of as embodying the interference effects, has a rapidly varying numerator and a slowly varying denominator. The subsidiary maxima are therefore located approximately at points where $\sin N\alpha$ has its greatest value, namely,

$$\alpha = \pm\frac{3\pi}{2N}, \pm\frac{5\pi}{2N}, \ldots \tag{10.34}$$

The $N - 2$ *subsidiary maxima* between consecutive principal maxima are clearly visible in Fig. 10.16. We can get some idea of the flux density at these peaks by rewriting Eq. (10.31) as

$$I(\theta) = \frac{I(0)}{N^2}\left(\frac{\sin\beta}{\beta}\right)^2\left(\frac{\sin N\alpha}{\sin\alpha}\right)^2 \tag{10.35}$$

where at the points of interest $|\sin N\alpha| = 1$. For large N, α is small and $\sin^2\alpha \approx \alpha^2$. At the first subsidiary peak $\alpha = 3\pi/2N$, in which case

$$I \approx I(0)\left(\frac{\sin\beta}{\beta}\right)^2\left(\frac{2}{3\pi}\right)^2 \tag{10.36}$$

and the flux density has dropped to about $\frac{1}{22}$ of that of the adjacent principal maximum (see Problem 10.12). Since $(\sin\beta)/\beta$ for small β varies slowly, it will not differ from 1 appreciably, close to the zeroth-order principal maximum, so that $I/I(0) \approx \frac{1}{22}$. This flux-density ratio for the next secondary peak is down to $\frac{1}{62}$, and it continues to decrease as α approaches a value halfway between the principal maxima. At that symmetry point, $\alpha \approx \pi/2$, $\sin\alpha \approx 1$, and the flux-density ratio has its lowest value, approximately $1/N^2$. Thereafter $\alpha > \pi/2$, and the flux densities of the subsidiary maxima begin to increase.

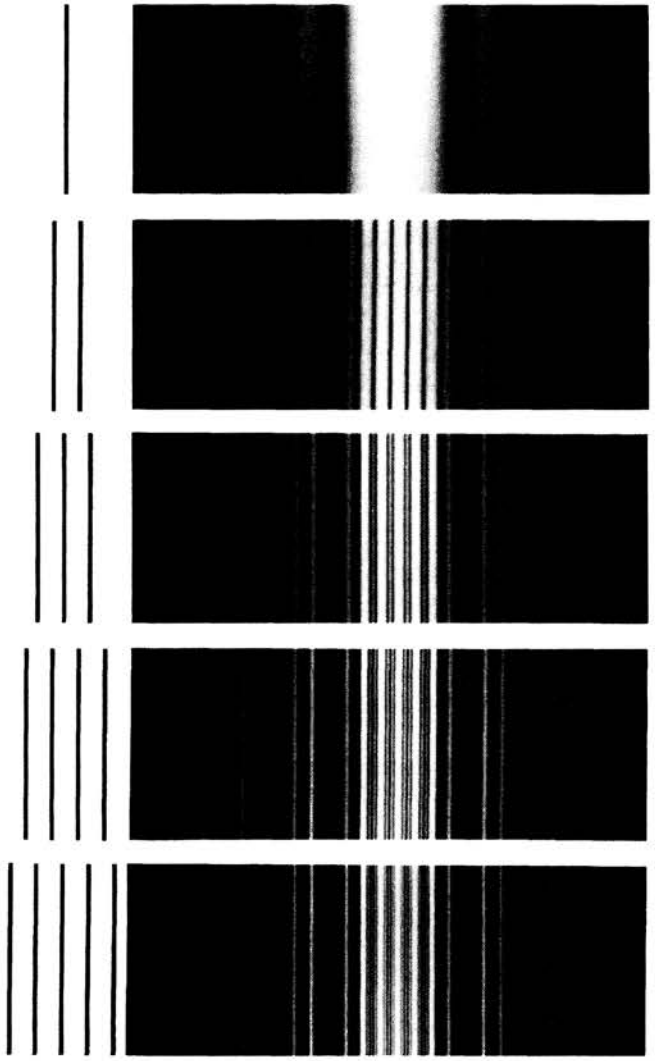


Figure 10.16 Diffraction patterns for slit systems shown at left. (Francis Weston Sears, *Optics*. Reprinted with permission of Addison Wesley Longman, Inc.)

Try duplicating Fig. 10.16 using a tubular bulb and homemade slits. You'll probably have difficulty seeing the subsidiary maxima clearly, with the effect that the only perceptible difference between the double- and multiple-slit patterns may be an apparent broadening in the dark regions between principal maxima. As in Fig. 10.16, the dark regions will become wider than the bright bands as N increases and the

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

secondary peaks fade out. If we consider each principal maximum to be bounded in width by two adjacent zeros, then each will extend over a length in θ, ($\sin\theta \approx \theta$) of approximately $2\lambda/Na$. As N increases, the principal maxima maintain their relative spacing (λ/a) while becoming increasingly narrow. Figure 10.17 shows the case of six slits, with $a = 4b$.

The multiple-slit interference term in Eq. (10.35) has the form $(\sin^2 N\alpha)/N^2 \sin^2\alpha$; thus for large N, $(N^2 \sin^2\alpha)^{-1}$ may

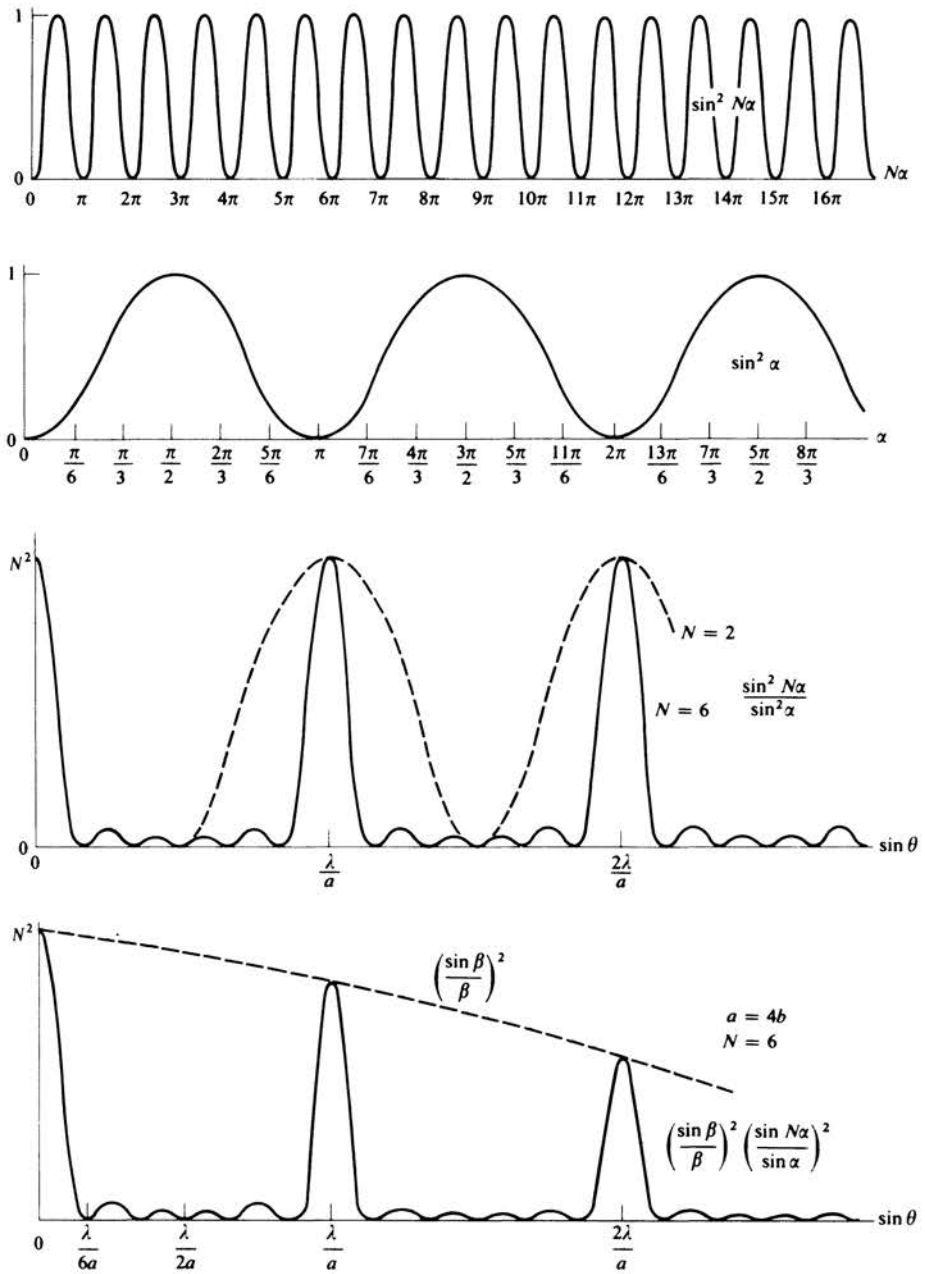


Figure 10.17 Multiple-slit pattern ($a = 4b$, $N = 6$).

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

be envisioned as the curve beneath which $\sin^2 N\alpha$ rapidly varies. Notice that for small α this interference term looks like $\text{sinc}^2 N\alpha$.

10.2.4 The Rectangular Aperture

Consider the configuration depicted in Fig. 10.18. A monochromatic plane wave propagating in the x-direction is incident on the opaque diffracting screen Σ. We wish to find the consequent (far-field) flux-density distribution in space or equivalently at some arbitrary distant point P. According to the Huygens–Fresnel Principle, a differential area dS, within the aperture, may be envisioned as being covered with coherent secondary point sources. But dS is much smaller in extent than is λ, so that all the contributions at P remain in-phase and interfere constructively. This is true regardless of θ; that is, dS emits a spherical wave (Problem 10.13). If $\mathcal{E}_A$ is the source strength per unit area, *assumed to be constant over the entire aperture*, then the optical disturbance at P due to dS is either the real or imaginary part of

$$dE = \left(\frac{\mathcal{E}_A}{r}\right) e^{i(\omega t - kr)}\, dS \tag{10.37}$$

The choice is yours and depends only on whether you like sine or cosine waves, there being no difference except for a phase shift. The distance from dS to P is

$$r = [X^2 + (Y - y)^2 + (Z - z)^2]^{1/2} \tag{10.38}$$

and as we have seen, the Fraunhofer condition occurs when this distance approaches infinity. As before, it will suffice to replace r by the distance $\overline{OP}$, that is, R, in the amplitude term, as long as the aperture is relatively small. But the approximation for r in the phase needs to be treated a bit more carefully; $k = 2\pi/\lambda$ is a large number. To that end we expand out Eq. (10.38) and, by making use of

$$R = [X^2 + Y^2 + Z^2]^{1/2} \tag{10.39}$$

obtain

$$r = R[1 + (y^2 + z^2)/R^2 - 2(Yy + Zz)/R^2]^{1/2} \tag{10.40}$$

In the far-field case R is very large in comparison to the dimensions of the aperture, and the $(y^2 + z^2)/R^2$ term is certainly negligible. Since P is very far from Σ, θ can still be kept small, even though Y and Z are fairly large, and this mitigates any concern about the directionality of the emitters (the obliquity factor). Now

$$r = R[1 - 2(Yy + Zz)/R^2]^{1/2}$$

and dropping all but the first two terms in the binomial expansion, we have

$$r = R[1 - (Yy + Zz)/R^2]$$

The total disturbance arriving at P is

$$\tilde{E} = \frac{\mathcal{E}_A e^{i(\omega t - kR)}}{R} \iint_{\text{Aperture}} e^{ik(Yy+Zz)/R}\, dS \tag{10.41}$$

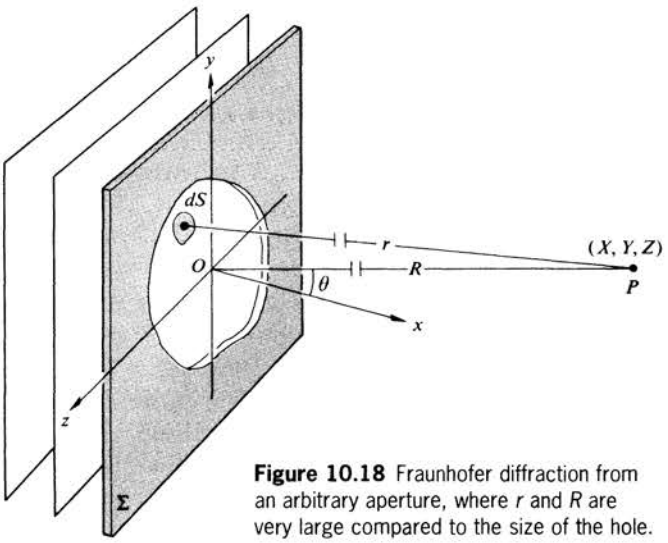


Figure 10.18 Fraunhofer diffraction from an arbitrary aperture, where r and R are very large compared to the size of the hole.

Consider the specific configuration shown in Fig. 10.19. Equation (10.41) can now be written as

$$\tilde{E} = \frac{\mathcal{E}_A e^{i(\omega t - kR)}}{R} \int_{-b/2}^{b/2} e^{ikYy/R}\, dy \int_{-a/2}^{a/2} e^{ikZz/R}\, dz$$

where $dS = dy\, dz$. With $\beta' \equiv kbY/2R$ and $\alpha' \equiv kaZ/2R$,

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

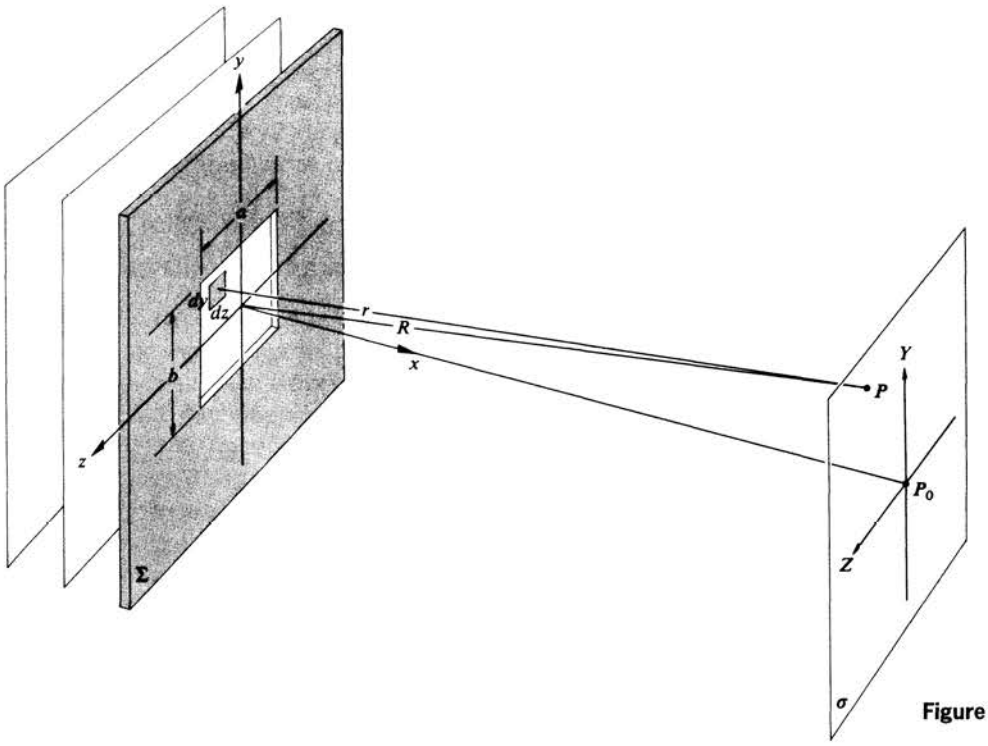


Figure 10.19 A rectangular aperture.

we have

$$\int_{-b/2}^{+b/2} e^{ikYy/R}\,dy = b\left(\frac{e^{i\beta'} - e^{-i\beta'}}{2i\beta'}\right) = b\left(\frac{\sin\beta'}{\beta'}\right)$$

and similarly

$$\int_{-a/2}^{+a/2} e^{ikZz/R}\,dz = a\left(\frac{e^{i\alpha'} - e^{-i\alpha'}}{2i\alpha'}\right) = a\left(\frac{\sin\alpha'}{a'}\right)$$

so that

$$\tilde{E} = \frac{A\,\mathcal{E}_A e^{i(\omega t - kR)}}{R}\left(\frac{\sin\alpha'}{\alpha'}\right)\left(\frac{\sin\beta'}{\beta'}\right) \qquad (10.42)$$

where A is the area of the aperture. Since $I = \langle(\mathrm{Re}\ \tilde{E})^2\rangle_T$,

$$I(Y, Z) = I(0)\left(\frac{\sin\alpha'}{\alpha'}\right)^2\left(\frac{\sin\beta'}{\beta'}\right)^2 \qquad (10.43)$$

where $I(0)$ is the irradiance at P_0; that is, at $Y = 0, Z = 0$ (see photos on page 466). At values of Y and Z such that $\alpha' = 0$ or $\beta' = 0$, $I(Y, Z)$ assumes the familiar shape of Fig. 10.10. When β' or α' are nonzero integer multiples of π or equivalently when Y and Z are nonzero integer multiples of $\lambda R/b$ and $\lambda R/a$, respectively, $I(Y, Z) = 0$, and we have a rectangular grid of nodal lines, as indicated in Fig. 10.20. Notice that the pattern in the Y-, Z-directions varies *inversely* with the y-, z-aperture dimensions. A horizontal, rectangular opening will produce a pattern with a vertical rectangle at its center.

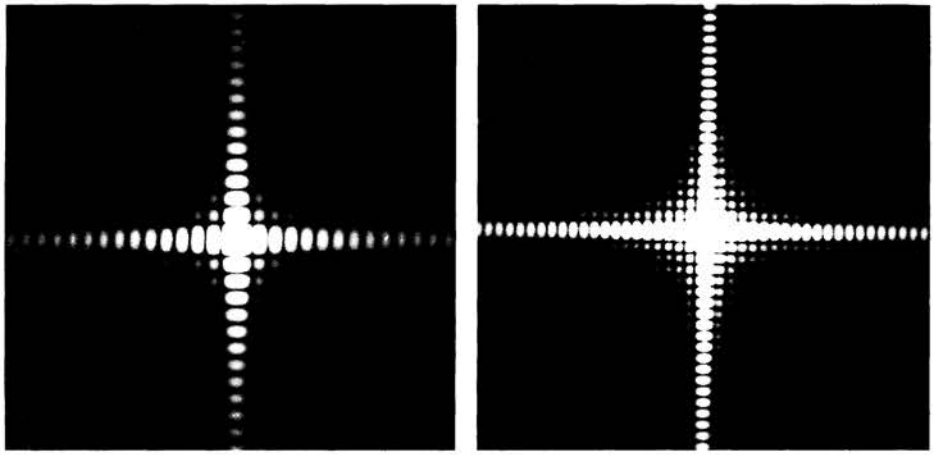
Along the β'-axis, $\alpha' = 0$ and the subsidiary maxima are located approximately halfway between zeros, that is, at $\beta'_m = \pm 3\pi/2, \pm 5\pi/2, \pm 7\pi/2, \ldots$. At each subsidiary maximum $\sin\beta'_m = 1$, and of course along the β'-axis, since $\alpha' = 0$, $(\sin\alpha')/\alpha' = 1$, so that the relative irradiances are approximated simply by

$$\frac{I}{I(0)} = \frac{1}{\beta'^2_m} \qquad (10.44)$$

Similarly, along the α'-axis

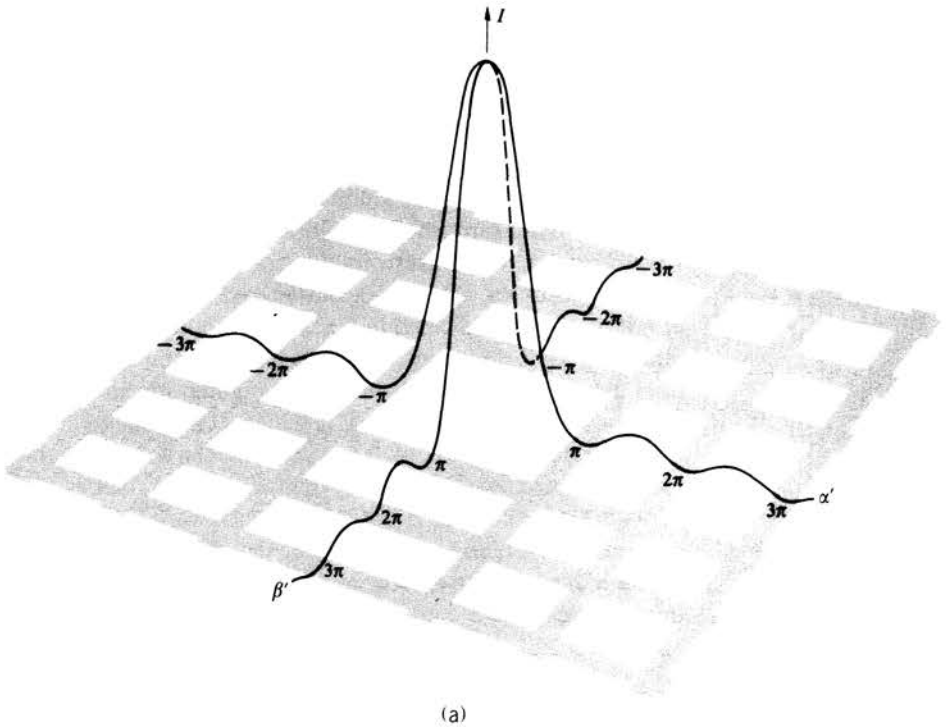
$$\frac{I}{I(0)} = \frac{1}{\alpha'^2_m} \qquad (10.45)$$

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

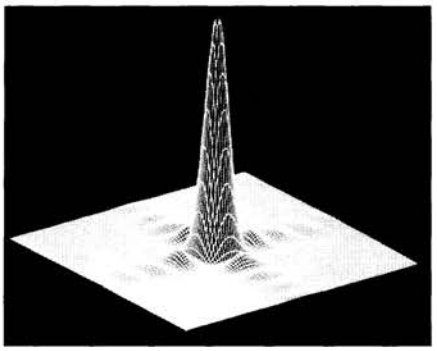


(a) (b)

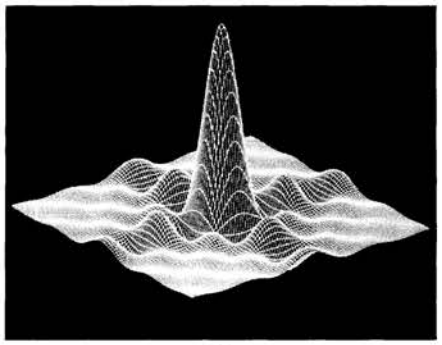
(a) Fraunhofer pattern of a square aperture. (b) The same pattern further exposed to bring out some of the faint terms. (Photos by E. H.)



(a)



(b)



(c)

Figure 10.20 (*a*) The irradiance distribution for a square aperture. (*b*) The irradiance produced by Fraunhofer diffraction at a square aperture. (*c*) The electric-field distribution produced by Fraunhofer diffraction via a square aperture. (Photos courtesy R. G. Wilson, Illinois Wesleyan University.)

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

The flux-density ratio* drops off rather rapidly from 1 to $\frac{1}{22}$ to $\frac{1}{62}$ to $\frac{1}{122}$, and so on. Even so, the off-axis secondary peaks are still smaller; for example, the four corner peaks (whose coordinates correspond to appropriate combinations of $\beta' = \pm 3\pi/2$ and $\alpha' = \pm 3\pi/2$) nearest to the central maximum each have relative irradiances of $(\frac{1}{22})^2$.

10.2.5 The Circular Aperture

Fraunhofer diffraction at a circular aperture is an effect of great practical significance in the study of optical instrumentation. Envision a typical arrangement: plane waves impinging on a screen Σ containing a circular aperture and the consequent far-field diffraction pattern spread across a distant observing screen σ. By using a focusing lens L_2, we can bring σ in close to the aperture without changing the pattern. Now, if L_2 is positioned within and exactly fills the diffracting opening in Σ, the form of the pattern is essentially unaltered. The lightwave reaching Σ is cropped, so that only a circular segment propagates through L_2 to form an image in the focal plane. This is obviously the same process that takes place in an eye, telescope, microscope, or camera lens. The image of a distant point source, as formed by a perfectly aberration-free converging lens, is never a point but rather some sort of diffraction pattern. We are essentially collecting only a fraction of the incident wavefront and therefore cannot hope to form a perfect image. As shown in the last section, the expression for the optical disturbance at P, arising from an arbitrary aperture in the far-field case, is

$$\tilde{E} = \frac{\mathcal{E}_A e^{i(\omega t - kR)}}{R} \iint_{\text{Aperture}} e^{ik(Y\mathcal{Y} + Z\mathcal{Z})/R}\, dS \qquad [10.41]$$

For a circular opening, symmetry would suggest introducing spherical coordinates in both the plane of the aperture and the plane of observation, as shown in Fig. 10.21 on page 468. Therefore, let

$$z = \rho \cos\phi \qquad y = \rho \sin\phi$$
$$Z = q \cos\Phi \qquad Y = q \sin\Phi$$

The differential element of area is now

$$dS = \rho\, d\rho\, d\phi$$

Substituting these expressions into Eq. (10.41), it becomes

$$\tilde{E} = \frac{\mathcal{E}_A e^{i(\omega t - kR)}}{R} \int_{\rho=0}^{a} \int_{\phi=0}^{2\pi} e^{i(kpq/R)\cos(\phi - \Phi)} \rho\, d\rho\, d\phi \qquad (10.46)$$

Because of the complete axial symmetry, the solution must be independent of Φ. We might just as well solve Eq. (10.46) with $\Phi = 0$ as with any other value, thereby simplifying things slightly.

The portion of the double integral associated with the variable ϕ,

$$\int_0^{2\pi} e^{i(kpq/R)\cos\phi}\, d\phi$$

is one that arises quite frequently in the mathematics of physics. It is a unique function in that it cannot be reduced to any of the more common forms, such as the various hyperbolic, exponential, or trigonometric functions, and indeed with the exception of these, it is perhaps the most often encountered. The quantity

$$J_0(u) = \frac{1}{2\pi} \int_0^{2\pi} e^{iu\cos v}\, dv \qquad (10.47)$$

is known as the *Bessel function* (of the first kind) of order zero. More generally,

$$J_m(u) = \frac{i^{-m}}{2\pi} \int_0^{2\pi} e^{i(mv + u\cos v)}\, dv \qquad (10.48)$$

represents the Bessel function of order m. Numerical values of $J_0(u)$ and $J_1(u)$ are tabulated for a large range of u in most mathematical handbooks. Just like sine and cosine, the Bessel functions have series expansions and are certainly no more

*These particular photographs were taken during an undergraduate laboratory session. A 1.5-mW He–Ne laser was used as a plane-wave source. The apparatus was set up in a long darkened room, and the pattern was cast directly on 4 × 5 Polaroid (ASA 3000) film. The film was located about 30 feet from a small aperture, so that no focusing lens was needed. The shutter, placed directly in front of the laser, was a student-contrived cardboard guillotine arrangement, and therefore no exposure times are available. Any camera shutter (a single-lens reflex with the lens removed and the back open) will serve, but the cardboard one was more fun.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

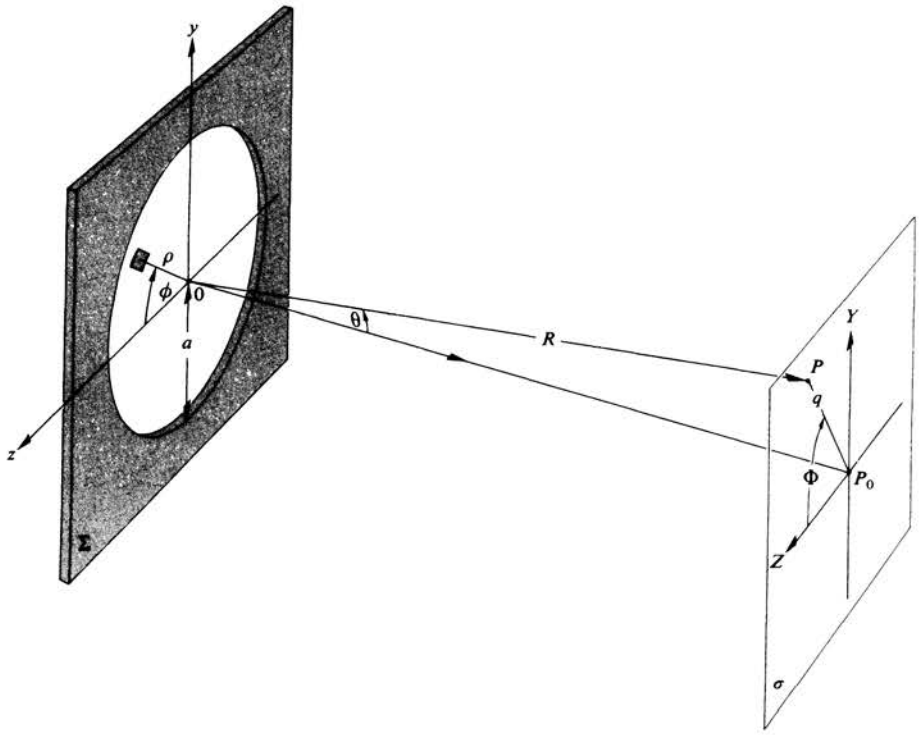


Figure 10.21 Circular aperture geometry.

esoteric than these familiar childhood acquaintances. As seen in Fig. 10.22, $J_0(u)$ and $J_1(u)$ are slowly decreasing oscillatory functions that do nothing particularly dramatic.

Equation (10.46) can be rewritten as

$$\tilde{E} = \frac{\mathcal{E}_A e^{i(\omega t - kR)}}{R} 2\pi \int_0^a J_0(k\rho q/R)\rho \, d\rho \qquad (10.49)$$

Another general property of Bessel functions, referred to as a recurrence relation, is

$$\frac{d}{du}[u^m J_m(u)] = u^m J_{m-1}(u)$$

When $m = 1$, this clearly leads to

$$\int_0^u u' J_0(u') \, du' = u J_1(u) \qquad (10.50)$$

with u' just serving as a dummy variable. If we now return to the integral in Eq. (10.49) and change the variable such that $w = k\rho q/R$, then $d\rho = (R/kq)\, dw$ and

$$\int_{\rho=0}^{\rho=a} J_0(k\rho q/R)\rho \, d\rho = (R/kq)^2 \int_{w=0}^{w=kaq/R} J_0(w) w \, dw$$

Making use of Eq. (10.50), we get

$$\tilde{E}(t) = \frac{\mathcal{E}_A e^{i(\omega t - kR)}}{R} 2\pi a^2 (R/kaq) J_1(kaq/R) \qquad (10.51)$$

The irradiance at point P is $\langle(\text{Re}\ \tilde{E})^2\rangle$ or $\frac{1}{2}\tilde{E}\tilde{E}^*$, that is,

$$I = \frac{2\mathcal{E}_A^2 A^2}{R^2} \left[\frac{J_1(kaq/R)}{kaq/R}\right]^2 \qquad (10.52)$$

where A is the area of the circular opening. To find the irradiance at the center of the pattern (i.e., at P_0), set $q = 0$. It follows from the above recurrence relation ($m = 1$) that

$$J_0(u) = \frac{d}{du} J_1(u) + \frac{J_1(u)}{u} \qquad (10.53)$$

From Eq. (10.47) we see that $J_0(0) = 1$, and from Eq. (10.48), $J_1(0) = 0$. The ratio of $J_1(u)/u$ as u approaches zero has the same limit (L'Hospital's rule) as the ratio of the separate derivatives of its numerator and denominator, namely, $dJ_1(u)/du$ over 1. But this means that the right-hand side of Eq. (10.53) is twice that limiting value, so that $J_1(u)/u = \frac{1}{2}$ at

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

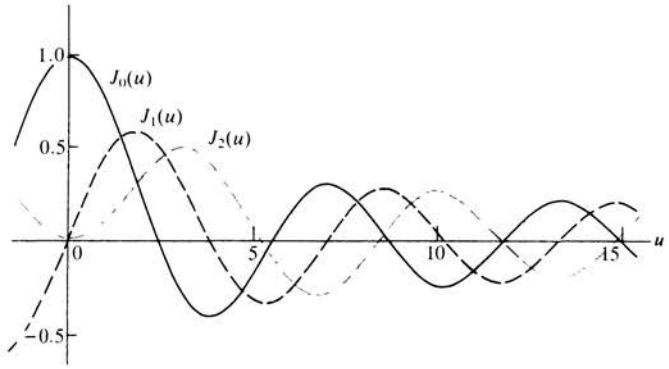


Figure 10.22 Bessel functions.

$u = 0$. The irradiance at P_0 is therefore

$$I(0) = \frac{\mathcal{E}_A^2 A^2}{2R^2} \tag{10.54}$$

which is the same result obtained for the rectangular opening [Eq. (10.43)]. If R is assumed to be essentially constant over the pattern, we can write

$$I = I(0)\left[\frac{2J_1(kaq/R)}{kaq/R}\right]^2 \tag{10.55}$$

Since $\sin\theta = q/R$, the irradiance can be written as a function of θ,

$$I(\theta) = I(0)\left[\frac{2J_1(ka\sin\theta)}{ka\sin\theta}\right]^2 \tag{10.56}$$

and as such is plotted in Fig. 10.23. Because of the axial symmetry, the towering central maximum corresponds to a high-irradiance circular spot known as the **Airy disk**. It was Sir George Biddell Airy (1801–1892), Astronomer Royal of England, who first derived Eq. (10.56). The central disk is surrounded by a dark ring that corresponds to the first zero of the function $J_1(u)$. From Table 10.1 on page 470 $J_1(u) = 0$ when $u = 3.83$, that is, $kaq/R = 3.83$. The radius q_1 drawn to the center of this first dark ring can be thought of as the extent of the Airy disk. It is given by

$$q_1 = 1.22\frac{R\lambda}{2a} \tag{10.57}$$

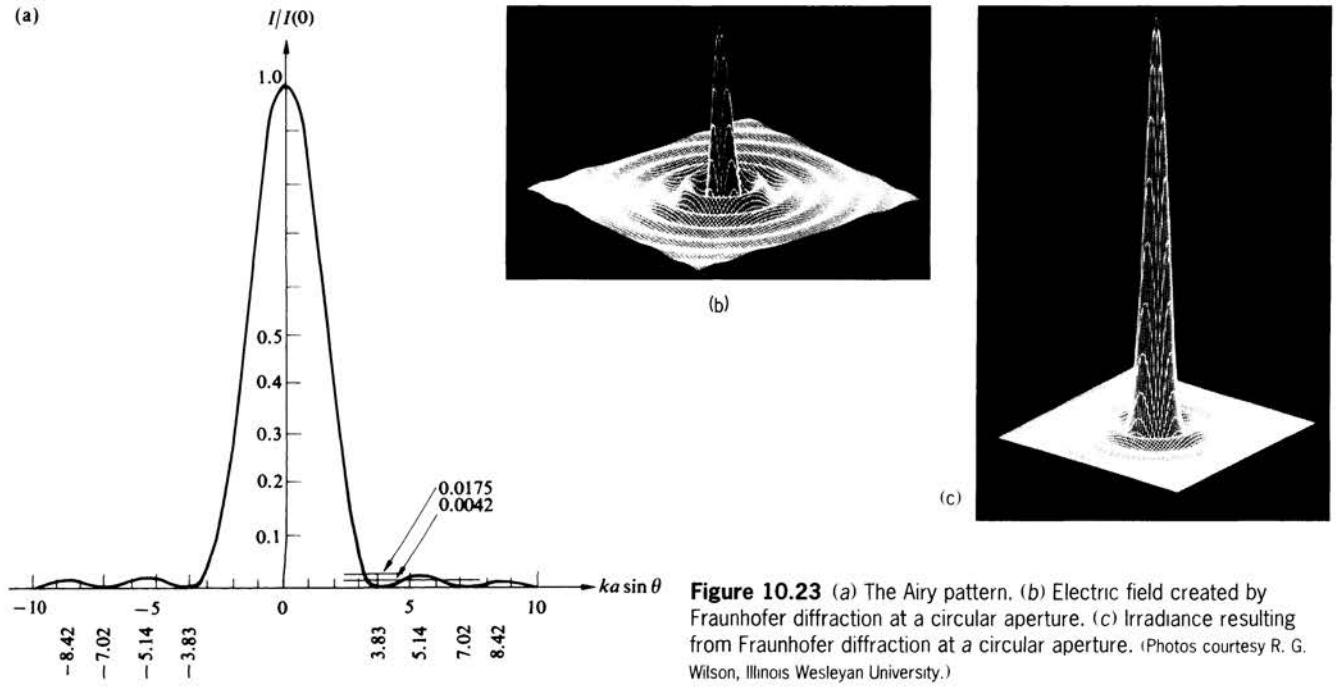
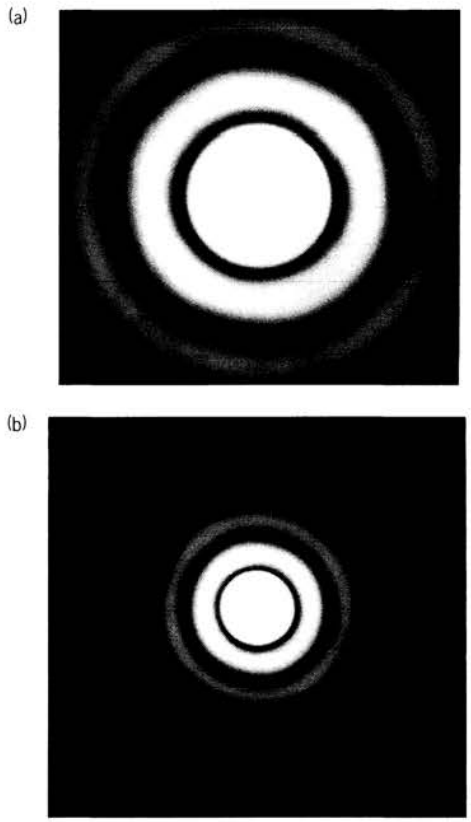


Figure 10.23 (*a*) The Airy pattern. (*b*) Electric field created by Fraunhofer diffraction at a circular aperture. (*c*) Irradiance resulting from Fraunhofer diffraction at *a* circular aperture. (Photos courtesy R. G. Wilson, Illinois Wesleyan University.)

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027



Airy rings using (a) a 0.5-mm hole diameter and (b) a 1.0-mm hole diameter. (Photo by E. H.)

For a lens focused on the screen σ, the focal length $f \approx R$, so

$$q_1 \approx 1.22 \frac{f\lambda}{D} \tag{10.58}$$

where D is the aperture diameter, in other words, $D = 2a$. (The *diameter* of the Airy disk in the visible spectrum is *very roughly* equal to the $f/\#$ of the lens in millionths of a meter.) As shown in the accompanying photos, q_1 varies inversely with the hole's diameter. As D approaches λ, the Airy disk can be very large indeed, and the circular aperture begins to resemble a point source of spherical waves.

The higher-order zeros occur at values of kaq/R equal to 7.02, 10.17, and so forth. The secondary maxima are located where u satisfies the condition

$$\frac{d}{du}\left[\frac{J_1(u)}{u}\right] = 0$$

TABLE 10.1 Bessel Functions*

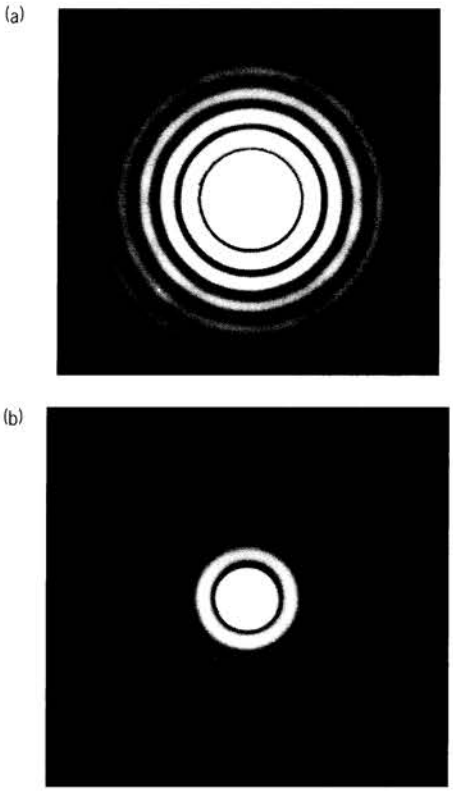
x	$J_1(x)$*	x	$J_1(x)$	x	$J_1(x)$
0.0	0.0000	3.0	0.3391	6.0	−0.2767
0.1	0.0499	3.1	0.3009	6.1	−0.2559
0.2	0.0995	3.2	0.2613	6.2	−0.2329
0.3	0.1483	3.3	0.2207	6.3	−0.2081
0.4	0.1960	3.4	0.1792	6.4	−0.1816
0.5	0.2423	3.5	0.1374	6.5	−0.1538
0.6	0.2867	3.6	0.0955	6.6	−0.1250
0.7	0.3290	3.7	0.0538	6.7	−0.0953
0.8	0.3688	3.8	0.0128	6.8	−0.0652
0.9	0.4059	3.9	−0.0272	6.9	−0.0349
1.0	0.4401	4.0	−0.0660	7.0	−0.0047
1.1	0.4709	4.1	−0.1033	7.1	0.0252
1.2	0.4983	4.2	−0.1386	7.2	0.0543
1.3	0.5220	4.3	−0.1719	7.3	0.0826
1.4	0.5419	4.4	−0.2028	7.4	0.1096
1.5	0.5579	4.5	−0.2311	7.5	0.1352
1.6	0.5699	4.6	−0.2566	7.6	0.1592
1.7	0.5778	4.7	−0.2791	7.7	0.1813
1.8	0.5815	4.8	−0.2985	7.8	0.2014
1.9	0.5812	4.9	−0.3147	7.9	0.2192
2.0	0.5767	5.0	−0.3276	8.0	0.2346
2.1	0.5683	5.1	−0.3371	8.1	0.2476
2.2	0.5560	5.2	−0.3432	8.2	0.2580
2.3	0.5399	5.3	−0.3460	8.3	0.2657
2.4	0.5202	5.4	−0.3453	8.4	0.2708
2.5	0.4971	5.5	−0.3414	8.5	0.2731
2.6	0.4708	5.6	−0.3343	8.6	0.2728
2.7	0.4416	5.7	−0.3241	8.7	0.2697
2.8	0.4097	5.8	−0.3110	8.8	0.2641
2.9	0.3754	5.9	−0.2951	8.9	0.2559

*$J_1(x) = 0$ for $x = 0, 3.832, 7.016, 10.173, 13.324, \ldots$
Adapted from E. Kreyszig, *Advanced Engineering Mathematics*, reprinted by permission of John Wiley & Sons, Inc.

which is equivalent to $J_2(u) = 0$. From the tables then, these secondary peaks occur when kaq/R equals 5.14, 8.42, 11.6, and so on, whereupon $I/I(0)$ drops from 1 to 0.0175, 0.0042, and 0.0016, respectively (Problem 10.22).

Circular apertures are preferable to rectangular ones, as far as lens shapes go, since the circle's irradiance curve is broader around the central peak and drops off more rapidly thereafter. Exactly what fraction of the total light energy incident on σ is confined to within the various maxima is a question of

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027



(a) Airy rings—long exposure (1.5-mm hole diameter). (b) Central Airy disk—short exposure with the same aperture. (Photos by E. H.)

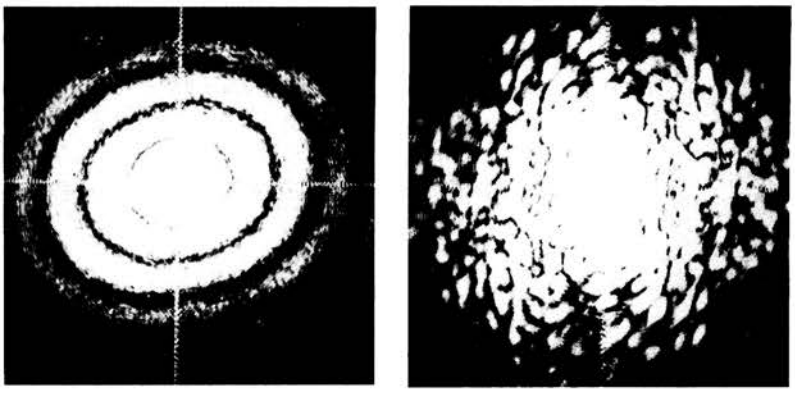
interest, but one somewhat too involved to solve here.* On integrating the irradiance over a particular region of the pattern, one finds that 84% of the light arrives within the Airy disk and 91% within the bounds of the second dark ring.

10.2.6 Resolution of Imaging Systems

Imagine that we have some sort of lens system that forms an image of an extended object. If the object is self-luminous, it is likely that we can regard it as made up of an array of incoherent sources. On the other hand, an object seen in reflected light will surely display some phase correlation between its various scattering points. When the point sources are in fact incoherent, the lens system will form an image of the object that consists of a distribution of partially overlapping, yet independent, Airy patterns. In the finest lenses, which have negligible aberrations, the spreading out of each image point due to diffraction represents the ultimate limit on image quality.

Suppose that we simplify matters somewhat and examine only two equal-irradiance, incoherent, distant point sources. For example, consider two stars seen through the objective lens of a telescope, where the entrance pupil corresponds to the diffracting aperture. In the previous section we saw that the radius of the Airy disk was given by $q_1 = 1.22f\lambda/D$. If $\Delta\theta$ is the corresponding angular measure, then $\Delta\theta = 1.22\lambda/D$, inasmuch as $q_1/f = \sin\Delta\theta \approx \Delta\theta$. The Airy disk for each star will be spread out over an angular half-width $\Delta\theta$ about its geometric image point, as shown in Fig. 10.24. If the angular separation of the stars is $\Delta\varphi$ and if $\Delta\varphi >> \Delta\theta$, the images will be

*See Born and Wolf, Principles of Optics, p. 398, or the very fine elementary text by Towne, Wave Phenomena, p.464.



Diffraction is being studied as a possible means of rapid automatic analysis of Pap tests for cancer. (a) The Fraunhofer diffraction pattern of a normal cervical cell. (b) The diffraction pattern of a malignant cervical cell is very different. (Photos courtesy Benjamin J. Pernick.)

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

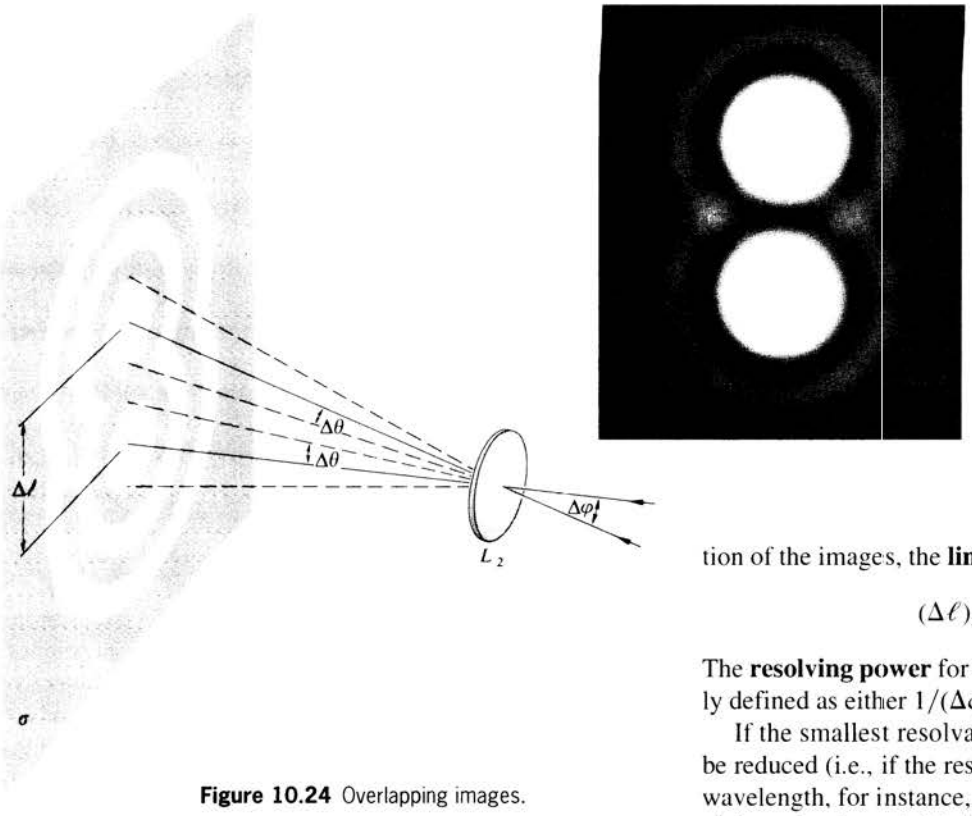


Figure 10.24 Overlapping images.

distinct and easily resolved. As the stars approach each other, their respective images come together, overlap, and commingle into a single blend of fringes. If Lord Rayleigh's criterion is applied, the stars are said to be *just resolved* when the center of one Airy disk falls on the first minimum of the Airy pattern of the other star. (We can certainly do a bit better than this, but Rayleigh's criterion, however arbitrary, has the virtue of being particularly uncomplicated.*) The *minimum resolvable angular separation* or *angular limit of resolution* is

$$(\Delta\varphi)_{\min} = \Delta\theta = 1.22\lambda/D \tag{10.59}$$

as depicted in Fig. 10.25. If $\Delta\ell$ is the center-to-center separation of the images, the **limit of resolution** is

$$(\Delta\ell)_{\min} = 1.22f\lambda/D \tag{10.60}$$

The **resolving power** for an image-forming system is generally defined as either $1/(\Delta\varphi)_{\min}$ or $1/(\Delta\ell)_{\min}$.

If the smallest resolvable separation between images is to be reduced (i.e., if the resolving power is to be increased), the wavelength, for instance, might be made smaller. Using ultraviolet rather than visible light in microscopy allows for the perception of finer detail. The electron microscope utilizes equivalent wavelengths of about 10^{-4} to 10^{-5} that of light. This makes it possible to examine objects that would otherwise be completely obscured by diffraction effects in the visible spectrum. On the other hand, the resolving power of a telescope can be increased by increasing the diameter of the objective lens or mirror. Besides collecting more of the incident radiation, this will also result in a smaller Airy disk and therefore a sharper, brighter image. The Mount Palomar 200-in telescope has a mirror 5 m in diameter (neglecting the obstruction of a small region at its center). At 550 nm it has an angular limit of resolution of 2.7×10^{-2} s of arc. In contrast, the Jodrell Bank radio telescope, with a 250-ft diameter, operates at a rather long, 21-cm wavelength. It therefore has a limit of resolution of only about 700 s of arc. The human eye has a pupil diameter that of course varies. Taking it, under bright conditions, to be about 2 mm, with $\lambda = 550$ nm, $(\Delta\varphi)_{\min}$ turns out to be roughly 1 min of arc. With a focal length of about 20 mm, $(\Delta\ell)_{\min}$ on the retina is 6700 nm. This is roughly twice the mean spacing between receptors. The human eye should there-

*In Rayleigh's own words: "This rule is convenient on account of its simplicity and it is sufficiently accurate in view of the necessary uncertainty as to what exactly is meant by resolution." See Section 9.6.1. for further discussion.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

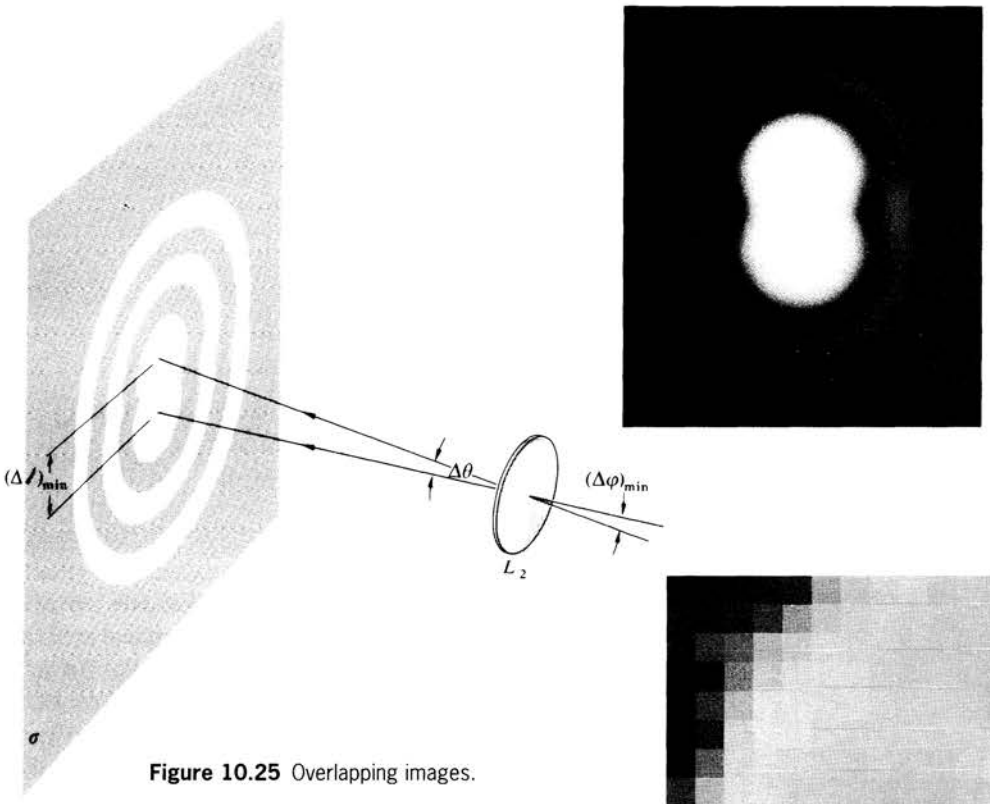
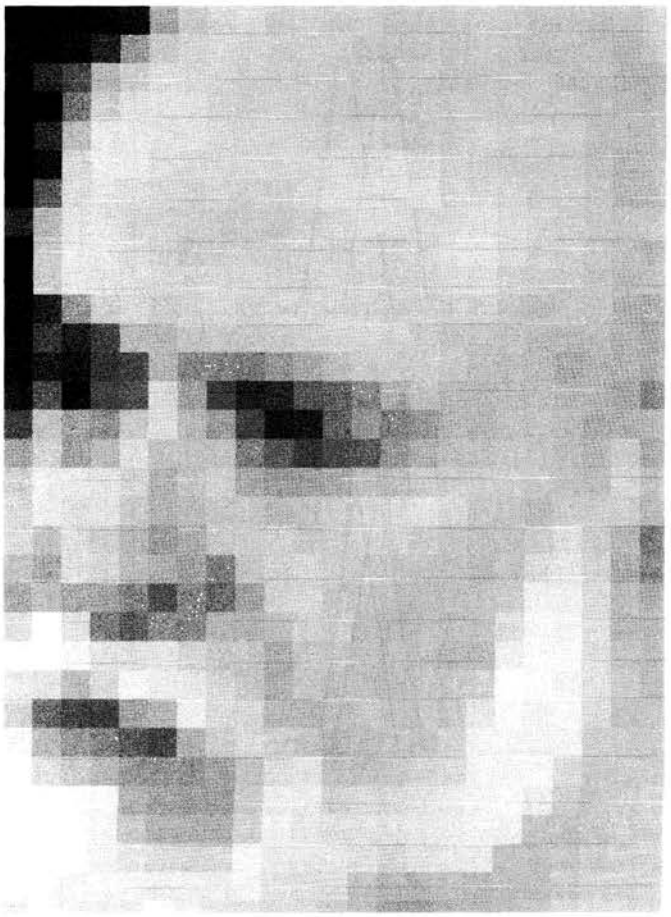


Figure 10.25 Overlapping images.

fore be able to resolve two points, an inch apart, at a distance of some 100 yards. You will probably not be able to do quite that well; one part in one thousand is more likely.

A more appropriate criterion for resolving power has been proposed by C. Sparrow. Recall that at the Rayleigh limit there is a central minimum or saddle point between adjacent peaks. A further decrease in the distance between the two point sources will cause the central dip to grow shallower and ultimately disappear. The angular separation corresponding to that configuration is Sparrow's limit. The resultant maximum has a broad flat top. In other words, at



This picture is made up of only about 750 pixels. The subject is hard to see when the page is close to you because you can resolve the individual pixel squares. That's especially true when you hold the picture close to your face. To perceive it more clearly, decrease your ability to resolve each separate pixel: decrease D, the aperture of your eye, (squint) or decrease the angular separation of the edges of each pixel (hold the picture farther away). If you do either, you should be able to make out the image of your humble author. (Photo by E.H.)

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

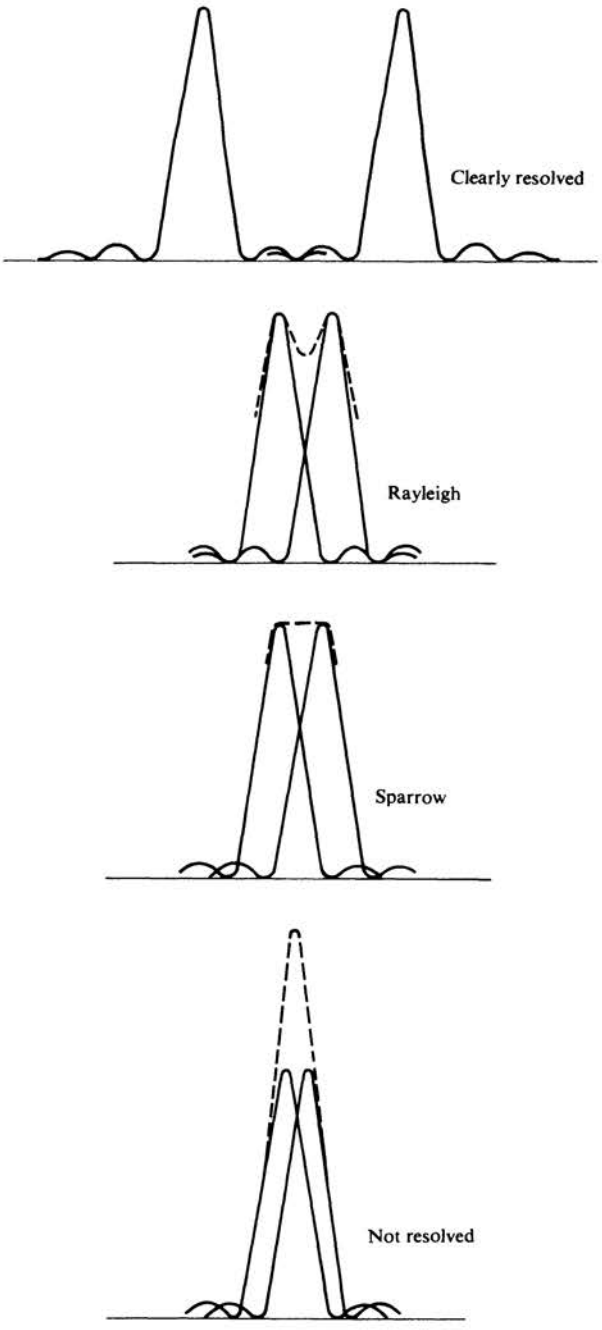


Figure 10.26 The Rayleigh and Sparrow criteria for overlapping point images.

the origin, which is the center of the peak, the second derivative of the irradiance function is zero; there is no change in slope (Fig. 10.26).

Unlike the Rayleigh rule, which rather tacitly assumes incoherence, the Sparrow condition can readily be generalized to coherent sources. In addition, astronomical studies of equal-brightness stars have shown that Sparrow's criterion is by far the more realistic.

10.2.7 The Zeroth-Order Bessel Beam

When light emerges from a narrow circular aperture, the diffracted beam has a central Airy disc that increases with distance via Eq. (10.57)—the beam spreads out as it propagates. Even though they resemble parallel ray bundles, laserbeams also diverge. The simplest and one of the most common laser-beam configurations is the TEM_{00} mode Gaussian beam (p. 594). If D_0 is the *waist diameter* (i.e., the diameter where the beam is narrowest), it will double its cross-sectional area after propagating a distance z_R. This is known as the *Rayleigh range*, where $z_R = \pi D_0^2/4\lambda$. Indeed, all real beams, no matter how well collimated they are, diverge.

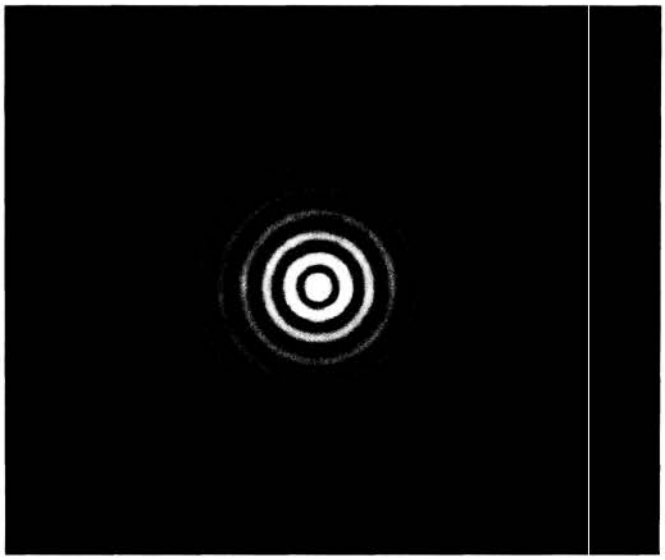
Nonetheless, there is a class of solutions to the differential wave equation for free space that are "nondiffracting." The simplest of these nonspreading beam solutions corresponds to a monochromatic wave propagating in the z-direction with an electric field proportional to the zeroth-order Bessel function J_0:

$$\tilde{E}(r, \theta, z, t) \propto J_0(k_\perp r) e^{i(k_\parallel z - \omega t)}$$

Here $\tilde{E}(r, \theta, z, t)$ is expressed in cylindrical coordinates (p. 31), $k_\parallel = k \cos\phi$, $k_\perp = k \sin\phi$, and the angle ϕ is fixed between 0 and 90°. Note that when $\phi = 0$, $\sin\phi = 0$, $J_0(0) = 1$ and the solution is a plane wave. Ideal plane waves don't spread out as they travel. But then again they're not localized in a narrow beam, nor do they actually exist.

We'll see in Chapter 11, when we study Fourier transforms, that in general a complex waveform like $\tilde{E}(r, \theta, z, t)$ can be represented as an infinite sum of plane waves taken over a continuum of k values. In particular, $\tilde{E}(r, \theta, z, t)$ may be considered to be a superposition of an infinite number of plane waves, all with propagation vectors (or if you like, wave vectors) lying along a cone whose half-angle, measured from its central axis, the z-axis, is ϕ. This is the defining characteristic of the **Bessel** or J_0**-beam**.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027



A Bessel beam. (Photo courtesy Ryan P. MacDonald.)

Because the irradiance is proportional to $\tilde{E}\tilde{E}^*$, all dependence on z vanishes; $I(r, \theta, t) \propto J_0^2(k_\perp r)$ and the irradiance is the same in every plane perpendicular to the z-axis. This means that the transverse irradiance pattern does not spread out as the wave advances. That pattern consists of a narrow central region (of diameter $2.405/k_\perp$) surrounded by concentric rings (see photo above). Each ring carries roughly the same energy as does the central peak, which is only about 5% of the initial energy of the beam.

In reality, one cannot create perfect plane waves from which to fabricate an ideal J_0-beam. A plane wave has infinite spatial extent and is therefore an unattainable idealization. So, at best, we can only build a wave that approximates a J_0-beam over a finite region of space; several methods have been used to accomplish just that.

Figure 10.27*a* shows an elegant scheme for generating a quasi-J_0-beam. A narrow ($\approx$10 μm) circular slit, or annulus, a few millimeters in diameter (a), is illuminated by monochromatic plane waves of wavelength λ. Every point in the aperture acts like a point source of spherical waves. The annulus is located in the front focal plane of a lens of radius R. Each spherical wavelet leaves the lens as a plane wave propagating at an angle ϕ, such that

$$\phi = \tan^{-1}\left(\tfrac{1}{2}a/f\right)$$

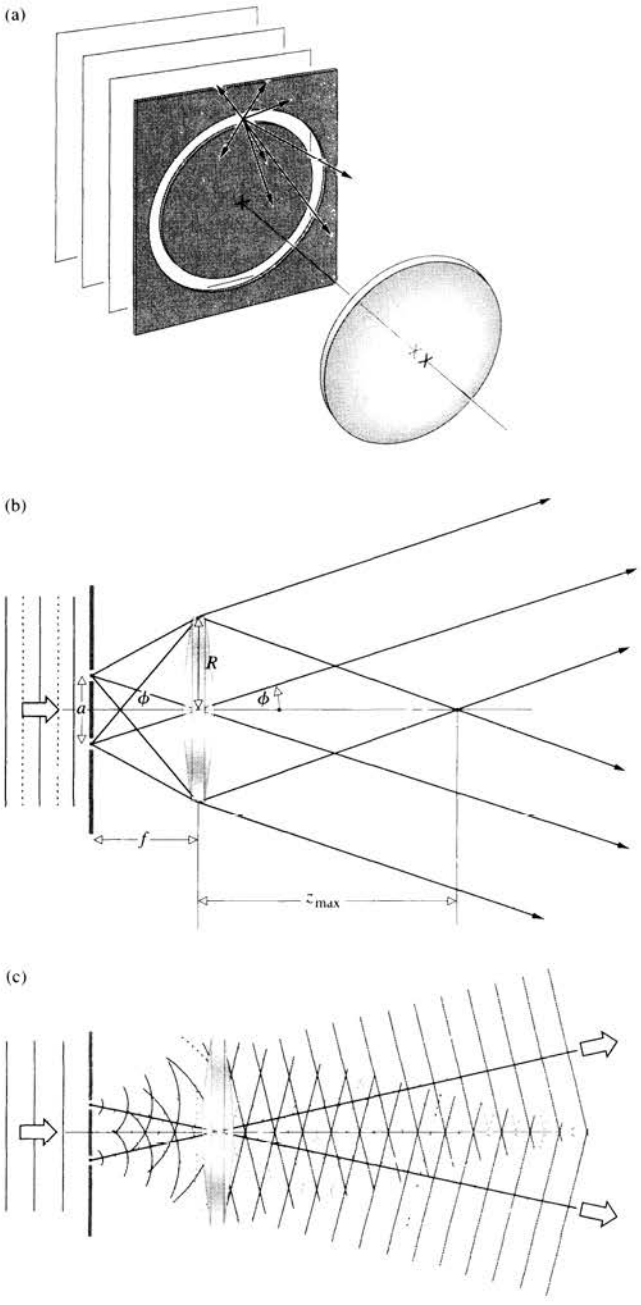


Figure 10.27 An arrangement for producing a Bessel beam using a circular slit. (a) A ring-shaped opening is illuminated by plane waves. (b) The aperture is placed in the front focal plane of the lens so that parallel rays leave the lens. (c) The plane waves, which all have propagation vectors residing on a cone, overlap out to a distance of z_{max}.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

The region of overlap of the plane waves in Fig.10.27*b* extends out to a distance z_{max} where $\tan\phi = R/z_{max} = \frac{1}{2}a/f$ and so

$$z_{max} = \frac{2Rf}{a}$$

This is the propagation length or range of the Bessel beam. If a is kept small and R large, it can be substantially greater than the Rayleigh range for a Gaussian beam of comparable diameter.*

10.2.8 The Diffraction Grating

A repetitive array of diffracting elements, either apertures or obstacles, that has the effect of producing periodic alterations in the phase, amplitude, or both of an emergent wave is said to be a **diffraction grating**. One of the simplest such arrangements is the multiple-slit configuration of Section 10.2.3. It seems to have been invented by the American astronomer David Rittenhouse in about 1785. Some years later Joseph von Fraunhofer independently rediscovered the principle and went on to make a number of important contributions to both the theory and technology of gratings. The earliest devices were indeed multiple-slit assemblies, usually consisting of a grid of fine wire or thread wound about and extending between two parallel screws, which served as spacers. A wavefront, in passing through such a system, is confronted by alternate opaque and transparent regions, so that it undergoes a modulation in *amplitude*. Accordingly, a multiple-slit configuration is said to be a ***transmission amplitude grating***. Another, more common form of transmission grating is made by ruling or scratching parallel notches into the surface of a flat, clear glass plate (Fig. 10.28*a*). Each of the scratches serves as a source of scattered light, and together they form a regular array of parallel line sources. When the grating is totally transparent, so that there is negligible amplitude modulation, the regular variations in the optical thickness across the grating yield a modulation in-

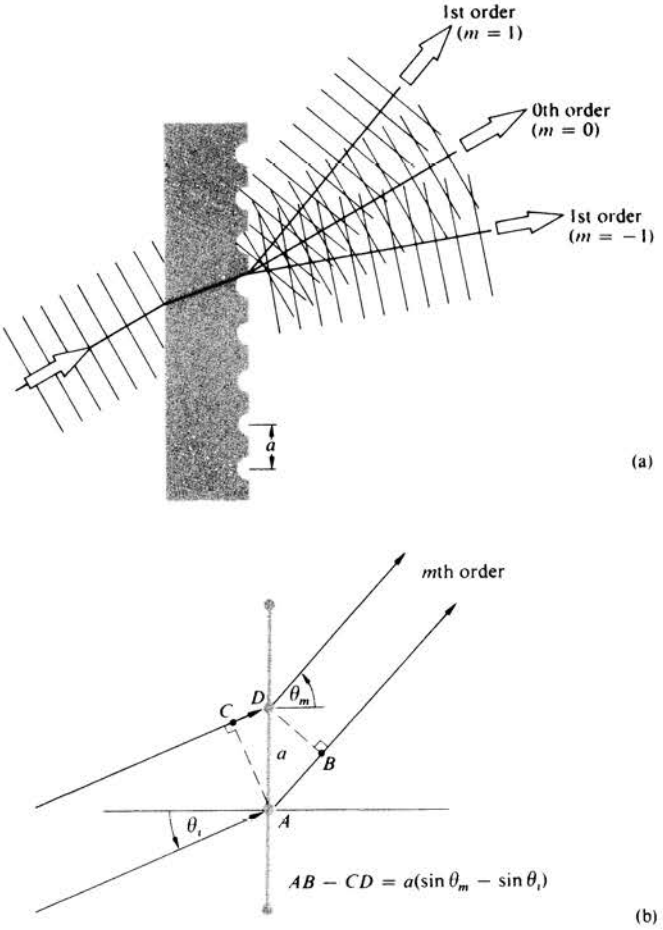


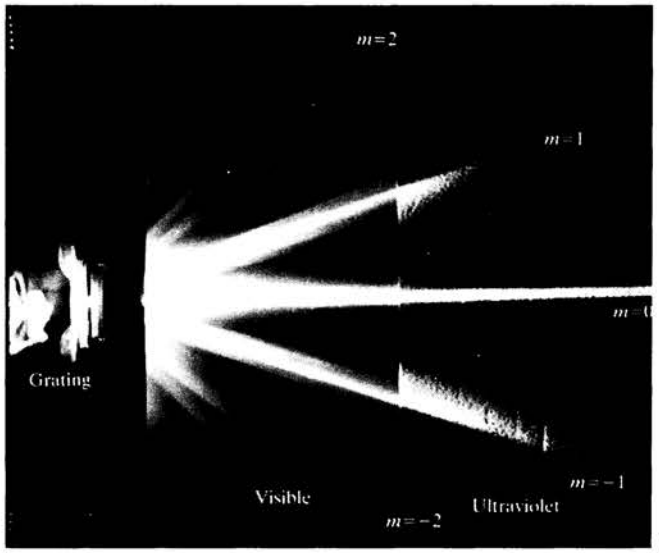
Figure 10.28 A transmission grating.

phase, and we have what is known as a ***transmission phase grating*** (see photo). In the Huygens–Fresnel representation you can envision the wavelets as radiated with different phases over the grating surface. An emerging wavefront therefore contains periodic variations in its shape rather than its amplitude. This in turn is equivalent to an angular distribution of constituent plane waves.

On reflection from this kind of grating, light scattered by the various periodic surface features will arrive at some point P with a definite phase relationship. The consequent interference pattern generated after reflection is quite similar to that arising from transmission. Gratings designed specifically to

*Lord Rayleigh, "On the passage of electric waves through tubes, or the vibrations of dielectric cylinders," *Phil. Mag.*, S. 5, **43**, No. 261, 125 (Feb. 1897). J. Durnin, "Exact solutions for nondiffracting beams. I. The scalar theory," *J. Opt. Soc. Am. A* **4**, 651 (1987). J. Durnin, J. J. Miceli, Jr., and J. H. Eberly, "Diffraction-Free Beams," *Phys. Rev. Lett.* **58**, 1499 (1987). C. A. McQueen, J. Arit, and K. Dholakia, "An experiment to study a 'nondiffracting' light beam," *Am. J. Phys.* **67**, 912 (1999).

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027



(a)



(b)



(c)

Light passing through a grating. (*a*) The region on the left is the visible spectrum; that on the right, the ultraviolet. (Photo courtesy Klinger Educational Prod. Corp., College Point, N.Y.) (*b*) Head-on views of the $m = 0$ and $m = \pm 1$ diffracted beams arising when light from a He–Ne laser passed through a 530 lines/mm grating. In the upper version, the grating was in air ($\lambda = 632.8$ nm). In the lower version, the grating was immersed in water. From the measured value of θ_1 the grating equation yielded $\lambda_w = 471$ nm and therefore $n_w = 1.34$. (Photo courtesy A. F. Leung, The Chinese University of Hong Kong.)

function in this fashion are known as ***reflection phase gratings*** (Fig. 10.29). Gratings of this sort have traditionally been ruled in thin films of aluminum that have been evaporated onto optically flat glass blanks. The aluminum, being fairly soft, results in less wear on the diamond ruling tool and is also a better reflector in the ultraviolet region.

The manufacture of ruled gratings is extremely difficult, and relatively few are made. In actuality, most gratings are exceedingly good plastic castings or *replicas* of fine, master ruled gratings. Today, large numbers of gratings are made holographically (p.623).

If you were to look perpendicularly through a transmission grating at a distant parallel line source, your eye would serve as a focusing lens for the diffraction pattern. Recall the analysis of Section 10.2.3 and the expression

$$a \sin \theta_m = m\lambda \qquad [10.32]$$

which is known as the **grating equation** for normal incidence. The values of m specify the *order* of the various principal maxima. For a source having a broad continuous spectrum, such as a tungsten filament, the $m = 0$, or zeroth-order, image corresponds to the undeflected, $\theta_0 = 0$, white-light view of the source (see photo). The grating equation is dependent on λ, and so for any value of $m \neq 0$ the various colored images of the source corresponding to slightly different angles (θ_m) spread out into a continuous spectrum. The regions occupied

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

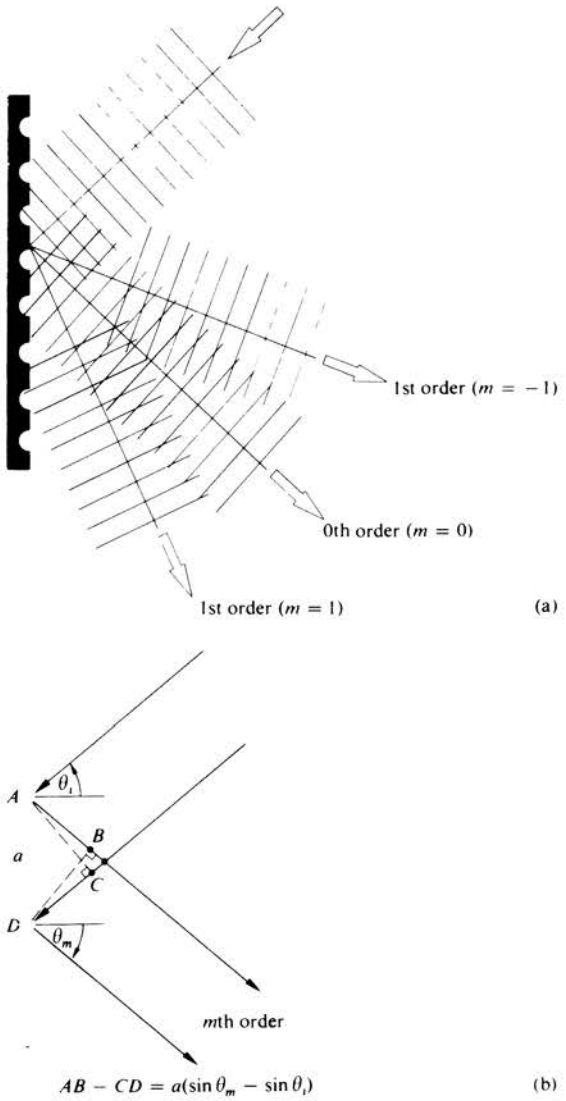


Figure 10.29 A reflection grating.

by the faint subsidiary maxima will show up as bands seemingly devoid of any light. The first-order spectrum $m = \pm 1$ appears on either side of $\theta = 0$ and is followed, along with alternate intervals of darkness, by the higher-order spectra, $m = \pm 2, \pm 3, \ldots$. Notice that the smaller a becomes in Eq. (10.32), the fewer will be the number of visible orders.

It should be no surprise that the grating equation is in fact Eq. (9.29), which describes the location of the maxima in Young's double-slit setup. The interference maxima, all located at the same angles, are now simply sharper (just as the multiple-beam operation of the Fabry–Perot etalon made its fringes sharper). In the double-slit case when the point of observation is somewhat off the exact center of an irradiance maximum, the two waves, one from each slit, will still be more or less in-phase, and the irradiance, though reduced, will still be appreciable. Thus the bright regions are fairly broad. By contrast, with multiple-beam systems, although all the waves interfere constructively at the centers of the maxima, even a small displacement will cause certain ones to arrive out-of-phase by $\frac{1}{2}\lambda$ with respect to others. For example, suppose P is slightly off from θ_1 so that $a \sin\theta = 1.010\lambda$ instead of 1.000λ. Each of the waves from successive slits will arrive at P shifted by 0.01λ with respect to the previous one. Then 50 slits down from the first, the path length will have shifted by $\frac{1}{2}\lambda$, and the light from slit 1 and slit 51 will essentially cancel. The same would be true for slit-pairs 2 and 52, 3 and 53, and so forth. The result is a rapid falloff in irradiance beyond the centers of the maxima.

Consider next the somewhat more general situation of oblique incidence, as depicted in Figs. 10.28 and 10.29. The grating equation, for both transmission and reflection, becomes

$$a(\sin\theta_m - \sin\theta_i) = m\lambda \tag{10.61}$$

This expression applies equally well, regardless of the refractive index of the transmission grating itself (Problem 10.37). One of the main disadvantages of the devices examined thus far, and in fact the reason for their obsolescence, is that they spread the available light energy out over a number of low-irradiance spectral orders. For a grating like that shown in Fig. 10.29, most of the incident light undergoes *specular reflection*, as if from a plane mirror. It follows from the grating equation that $\theta_m = \theta_i$ corresponds to the zeroth order, $m = 0$. All of this light is essentially wasted, at least for spectroscopic purposes, since the constituent wavelengths overlap.

In an article in the *Encyclopedia Britannica* of 1888, Lord Rayleigh suggested that it was at least theoretically possible to shift energy out of the useless zeroth order into one of the higher-order spectra. So motivated, Robert Williams Wood (1868–1955) succeeded in 1910 in ruling grooves with a controlled shape, as shown in Fig. 10.30. Most modern gratings are of this shaped or **blazed** variety. The angular positions of the nonzero orders, θ_m-values, are determined by a, λ, and, of more immediate interest, θ_i. But θ_i and θ_m are measured from the normal to the grating plane and not with respect to the indi-

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

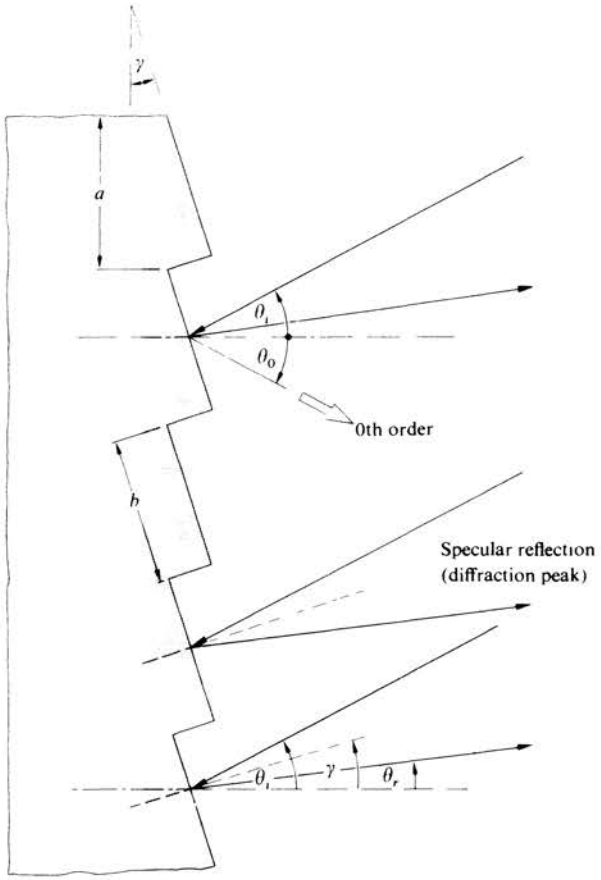


Figure 10.30 Section of a blazed reflection phase grating.

vidual groove surfaces. On the other hand, the location of the peak in the single-facet diffraction pattern corresponds to *specular reflection* off that face, for each groove. It is governed by the *blaze angle* γ and can be varied independently of θ_m. This is somewhat analogous to the antenna array of Section 10.1.3, where we were able to control the spatial position of the interference pattern [Eq. (10.6)] by adjusting the relative phase shift between sources without actually changing their orientations.

Consider the situation depicted in Fig. 10.31 when the incident wave is normal to the plane of a blazed reflection grating; that is, $\theta_i = 0$, so for $m = 0$, $\theta_0 = 0$. For *specular reflection* $\theta_i - \theta_r = 2\gamma$ (Fig. 10.30), most of the diffracted radiation is concentrated about $\theta_r = -2\gamma$. (θ_r is negative because the incident and reflected rays are on the same side of the grating normal.) This will correspond to a particular nonzero order, on one side of the central image, when $\theta_m = -2\gamma$; in other words, $a \sin(-2\gamma) = m\lambda$ for the desired λ and m.

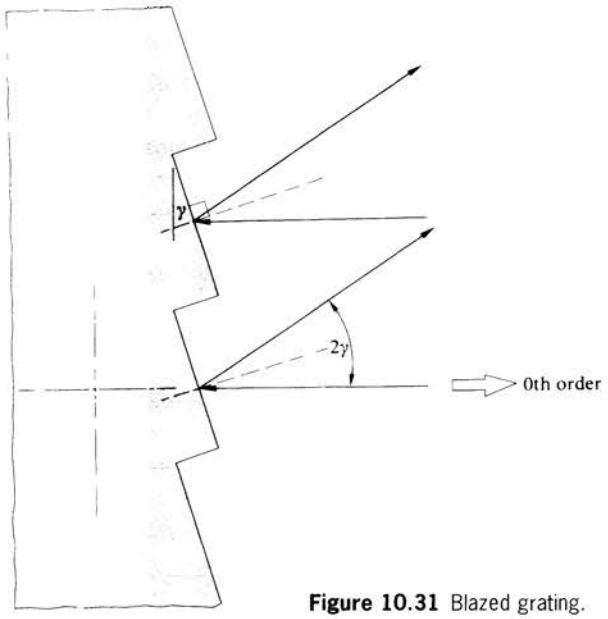


Figure 10.31 Blazed grating.

Grating Spectroscopy

Quantum Mechanics, which evolved in the early 1920s, had its initial thrust in the area of atomic physics. Predictions were made concerning the detailed structure of the hydrogen atom as manifested by its emitted radiation, and spectroscopy provided the vital proving ground. The need for larger and better gratings became apparent. Grating spectrometers, used over the range from soft X-rays to the far infrared, have enjoyed continued interest. In the hands of the astrophysicist or rocket-borne, they yield information concerning the very origins of the Universe, information as varied as the temperature of a star, the rotation of a galaxy, and the red shift in the spectrum of a quasar. In the mid-1900s George R. Harrison and George W. Stroke remarkably improved the quality of high-resolution gratings. They used a ruling engine* whose operation was controlled by an interferometrically guided servomechanism.

*For more details about these marvelous machines, see A. R. Ingalls, *Sci. Amer.* **186**, 45 (1952), or the article by E. W. Palmer and J. F. Verrill, *Contemp. Phys.* **9**, 257 (1968).

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

Let us now examine in some detail a few of the major features of the grating spectrum. Assume an infinitesimally narrow incoherent source. The effective width of an emergent spectral line may be defined as the angular distance between the zeros on either side of a principal maximum; in other words, $\Delta\alpha = 2\pi/N$, which follows from Eq. (10.33). At oblique incidence we can redefine α as $(ka/2)$ (sin θ − sin θ_i), and so a small change in α is given by

$$\Delta\alpha = (ka/2)\cos\theta\,(\Delta\theta) = 2\pi/N \qquad (10.62)$$

where the angle-of-incidence is constant, that is, $\Delta\theta_i = 0$. Thus, even when the incident light is monochromatic,

$$\Delta\theta = 2\lambda/(Na\cos\theta_m) \qquad (10.63)$$

is the *angular width of a line*, due to *instrumental broadening*. Interestingly enough, the angular linewidth varies inversely with the width of the grating itself, Na. Another important quantity is the difference in angular position corresponding to a difference in wavelength. The **angular dispersion**, as in the case of a prism, is defined as

$$\mathcal{D} \equiv d\theta/d\lambda \qquad (10.64)$$

Differentiating the grating equation yields

$$\mathcal{D} = m/(a\cos\theta_m) \qquad (10.65)$$

This means that the angular separation between two different frequency lines will increase as the order increases.

Blazed plane gratings with nearly rectangular grooves are most often mounted so that the incident propagation vector is almost normal to either one of the groove faces. This is the condition of *autocollimation*, in which θ_i and θ_m are on the same side of the normal and $\gamma \approx \theta_i \approx -\theta_m$ (see Fig. 10.32), whereupon

$$\mathcal{D}_{\text{auto}} = (2\tan\theta_i)/\lambda \qquad (10.66)$$

which is independent of a.

When the wavelength difference between two lines is small enough so that they overlap, the resultant peak becomes somewhat ambiguous. The chromatic resolving power $\mathcal{R}$ of a spectrometer is defined as

$$\mathcal{R} \equiv \lambda/(\Delta\lambda)_{\min} \qquad [9.76]$$

where $(\Delta\lambda)_{\min}$ is the least resolvable wavelength difference, or **limit of resolution**, and λ is the mean wavelength. Lord Rayleigh's criterion for the resolution of two fringes with equal flux density requires that the principal maximum of one coincide with the first minimum of the other. (Compare this with the equivalent statement used in Section 9.6.1.) As shown in Fig. 10.26, at the limit of resolution the angular separation is half the linewidth, or from Eq. (10.63)

$$(\Delta\theta)_{\min} = \lambda/(Na\cos\theta_m)$$

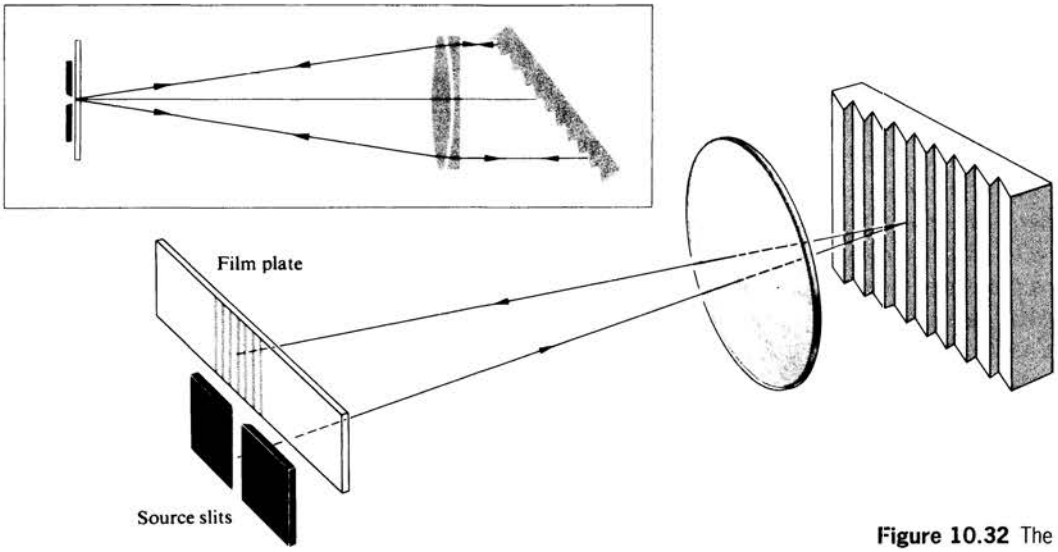


Figure 10.32 The Littrow autocollimation mounting.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

Applying the expression for the dispersion, we get

$$(\Delta\theta)_{\min} = (\Delta\lambda)_{\min}\, m/(a \cos \theta_m)$$

The combination of these two equations provides us with $\mathcal{R}$, that is,

$$\lambda/(\Delta\lambda)_{\min} = mN \tag{10.67}$$

or

$$\mathcal{R} = \frac{Na(\sin \theta_m - \sin \theta_i)}{\lambda} \tag{10.68}$$

The resolving power is a function of the grating width Na, the angle-of-incidence, and λ. A grating 6 inches wide and containing 15,000 lines per inch will have a total of 9×10^4 lines and a resolving power, in the second order, of 1.8×10^5. In the vicinity of 540 nm, the grating could resolve a wavelength difference of 0.003 nm. Notice that the resolving power cannot exceed $2Na/\lambda$, which occurs when $\theta_i = -\theta_m = 90°$. The largest values of $\mathcal{R}$ are obtained when the grating is used in autocollimation, whereupon

$$\mathcal{R}_{\text{auto}} = \frac{2Na \sin \theta_i}{\lambda} \tag{10.69}$$

and again θ_i and θ_m are on the same side of the normal. For one of Harrison's 260-mm-wide blazed gratings at about 75° in a Littrow mount, with $\lambda = 500$ nm, the resolving power just exceeds 10^6.

We now need to consider the problem of overlapping orders. The grating equation makes it quite clear that a line of 600 nm in the first order will have precisely the same position, θ_m, as a 300-nm line in the second order or a 200-nm line when $m = 3$. If two lines of wavelength λ and $(\lambda + \Delta\lambda)$ in successive orders $(m + 1)$ and m just coincide, then

$$a(\sin \theta_m - \sin \theta_i) = (m + 1)\lambda = m(\lambda + \Delta\lambda)$$

The precise wavelength difference is known as the **free spectral range**,

$$(\Delta\lambda)_{\text{fsr}} = \lambda/m \tag{10.70}$$

as it was for the Fabry–Perot Interferometer. In comparison with that device, whose resolving power was

$$\mathcal{R} = \mathcal{F}m \tag{9.76}$$

we might take N to be the finesse of a diffraction grating (Problem 10.38).

A high-resolution grating blazed for the first order, so as to have the greatest free spectral range, will require a high groove density (up to about 1200 lines per millimeter) in order to maintain $\mathcal{R}$. Equation (10.68) shows that $\mathcal{R}$ can be kept constant by ruling fewer lines with increasing spacing, such that the grating width Na is constant. But this requires an increase in m and a subsequent decrease in free spectral range, characterized by overlapping orders. If this time N is held constant while a alone is made larger, $\mathcal{R}$ increases as does m, so that $(\Delta\lambda)_{\text{fsr}}$ again decreases. The angular width of a line is reduced (i.e., the spectral lines become sharper), the coarser the grating is, but the dispersion in a given order diminishes, with the effect that the lines in that spectrum approach each other.

Thus far we have considered a particular type of periodic array, namely, the *line grating*. A good deal more information is available in the literature* concerning their shapes, mountings, uses, and so forth.

A few unlikely household items can be used as crude gratings. The grooved surface of a phonograph record works nicely near grazing incidence and CDs are lovely reflection gratings. Surprisingly enough, with $\theta_i \approx 90°$ an ordinary fine-toothed comb (or a stick of staples) will separate out the constituent wavelengths of white light. This occurs in exactly the same fashion as it would with a more orthodox reflection grating. In a letter to a friend dated May 12, 1673, James Gregory pointed out that sunlight passing through a feather would produce a colored pattern, and he asked that his observations be conveyed to Mr. Newton. If you've got one, a flight feather makes a nice transmission grating.

Two- and Three-Dimensional Gratings

Suppose that the diffracting screen Σ contains a large number, N, of identical diffracting objects (apertures or obstacles). These are to be envisioned as distributed over the surface of Σ in a completely random manner. We also require that each and every one be similarly oriented. Imagine the diffracting screen to be illuminated by plane waves that are focused by a perfect lens L_2, after emerging from Σ (see Fig. 10.12). The individ-

*See F. Kneubühl, "Diffraction Grating Spectroscopy," *Appl. Opt.* **8**, 505 (1969); R. S. Longhurst, *Geometrical and Physical Optics*; and the extensive article by G. W. Stroke in the *Encyclopedia of Physics*, Vol. 29, edited by S. Flügge, p. 426.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

ual apertures generate identical Fraunhofer diffraction patterns, all of which overlap on the image plane σ. If there is no regular periodicity in the location of the apertures, we cannot anticipate anything but a random distribution in the relative phases of the waves arriving at an arbitrary point P on σ. We have to be rather careful, however, because there is one exception, which occurs when P is on the central axis, that is, $P = P_0$. All rays, from all apertures, parallel to the central axis will traverse equal optical path lengths before reaching P_0. They will therefore arrive in-phase and interfere constructively.

Now consider a group of arbitrarily directed parallel rays (not in the direction of the central axis), each one emitted from a different aperture. These will be focused at some point on σ, such that each corresponding wave will have an equal probability of arriving with any phase between 0 and 2π. What must be determined is the resultant field arising from the superposition of N equal-amplitude phasors all having random relative phases. The solution to this problem requires an elaborate analysis in terms of probability theory, which is a little too far afield to do here.* The important point is that the sum of a number of phasors taken at random angles is not simply zero, as might be thought. The general analysis begins, for statistical reasons, by assuming that there are a large number of individual aperture screens, each containing N random diffracting apertures and each illuminated, in turn, by a monochromatic wave. We shouldn't be surprised if there is some difference, however small, between the diffraction patterns of two different random distributions of, say, $N = 100$ holes. After all, they are different, and the smaller N is, the more obvious that becomes. Indeed, we can expect their similarities to show up statistically on considering a large number of such masks—ergo the general approach.

If the many individual resulting irradiance distributions are all averaged for a particular *off-axis* point on σ, it will be found that the average irradiance (I_{av}) there equals N times the irradiance (I_0) due to a single aperture: $I_{av} = NI_0$. Still, the irradiance at any point arising from any one aperture screen can differ from this average value by a fairly large amount, regardless of how great N is. These point-to-point fluctuations about the average manifest themselves in each particular pattern as a granularity that tends to show a radial fiberlike structure. If this fine-grained mottling is averaged over a small region of the pattern, which nonetheless contains many fluctuations, it will average out to NI_0.

Of course, in any real experiment the situation will not quite match the ideal—there is no such thing as monochromatic light or a truly random array of (nonoverlapping) diffracting objects. Nonetheless, with a screen containing N "random" apertures illuminated by quasimonochromatic, nearly plane-wave illumination, we can anticipate seeing a mottled flux-density distribution closely resembling that of an individual aperture but N times as strong. Moreover, a bright spot will exist on-axis at its center, which will have a flux density of N^2 times that of a single aperture. If, for example, the screen contains N rectangular holes (Fig. 10.33*a*), the resultant pattern (Fig. 10.33*b*) will resemble the images on page 466. Similarly, the array of circular holes depicted in Fig. 10.33*c* will produce the diffraction rings of Fig. 10.33*d*.

As the number of apertures increases, the central spot will tend to become so bright as to obscure the rest of the pattern. Note as well that the above considerations apply when all the apertures are illuminated completely coherently. In actuality, the diffracted flux-density distribution will be determined by the degree of coherence (see Chapter 12). The pattern will run the gamut from no interference with completely incoherent light to the case discussed above for completely coherent illumination (Problem 10.40).

The same kind of effects arise from what we might call a two-dimensional *phase grating*. For example, the halo or corona often seen about the Sun or Moon results from diffraction by random droplets of water vapor (i.e., cloud particles). If you would like to duplicate the effect, fog up a microscope slide with your breath, or rub a very thin film of talcum powder on it and then fog it up. Look at a white-light point source. You should see a pattern of clear, concentric, colored rings [Eq. (10.56)] surrounding a white central disk. If you just see a white blur, you don't have a distribution of roughly equal-sized droplets; have another try at the talcum. Strikingly beautiful patterns approximating concentric ring systems can be seen through an ordinary *mesh* nylon stocking. If you are fortunate enough to have mercury-vapor street lights, you'll have no trouble seeing all their constituent visible spectral frequencies. (If not, block out most of a fluorescent lamp, leaving something resembling a small source.) Notice the increased symmetry as you increase the number of layers of nylon. Inci-

*For a statistical treatment, consult J. M. Stone, *Radiation and Optics*, p. 146, and Sommerfeld, *Optics*, p. 194. Also take a look at "Diffraction Plates for Classroom Demonstrations," by R. B. Hoover, *Am. J. Phys.* 37, 871 (1969), and T. A. Wiggins, "Hole Gratings for Optics Experiments," *Am. J. Phys.* **53**, 227 (1985).

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

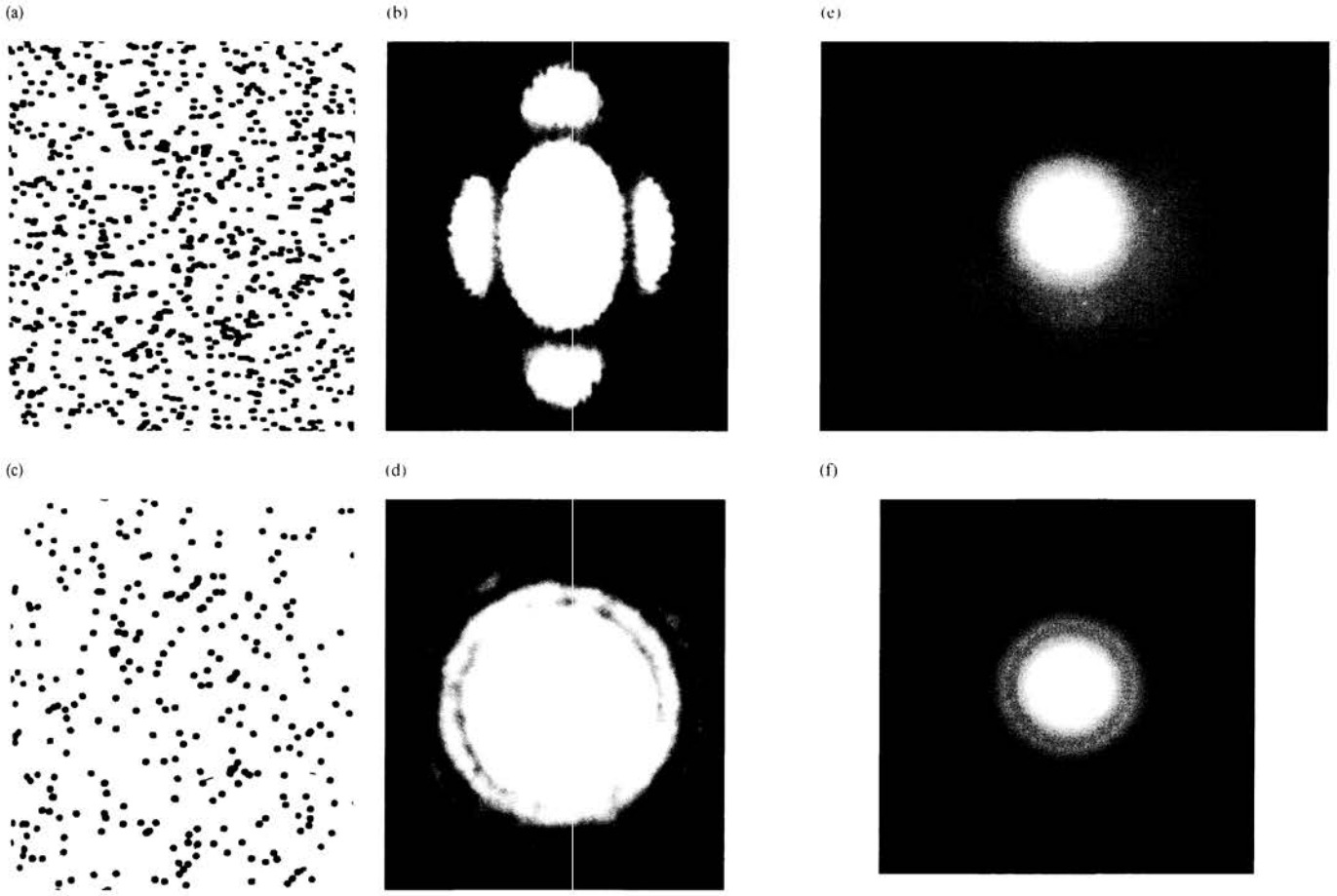


Figure 10.33 (*a*) A random array of rectangular apertures. (*b*) The resulting white-light Fraunhofer pattern. (*c*) A random array of circular apertures. (*d*) The resulting white-light Fraunhofer pattern. (Photos courtesy The Ealing Corporation and Richard B. Hoover.) (*e*) A candle flame viewed through a fogged piece of glass. The spectral colors are visible as concentric rings. (Photo by E. H.) (*f*) A similar colored ring system created by viewing a white-light point source through a glass plate covered with transparent spherical lycopodium spores. (Photo by E. H.)

dentally, this is precisely the way Rittenhouse, the inventor of the grating, became interested in the problem—only he used a silk handkerchief.

Consider the case of a *regular* two-dimensional array of diffracting elements (Fig. 10.34) under normally incident plane-wave illumination. Each small element behaves as a coherent source. And because of the regular periodicity of the lattice of emitters, each emergent wave bears a fixed-phase relation to the others. There will now be certain directions in which constructive interference prevails. Obviously, these occur when the distances from each diffracting element to P are such that the waves are nearly in-phase at arrival. The phenomenon can be observed by looking at a point source through a piece of *square woven*, thin cloth (such as nylon curtain material; see Fig. 10.34*e*) or the fine metal mesh of a tea strainer. The diffracted image is effectively the superposition of two grating patterns at right angles. Examine the center of the pattern carefully to see its gridlike structure.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

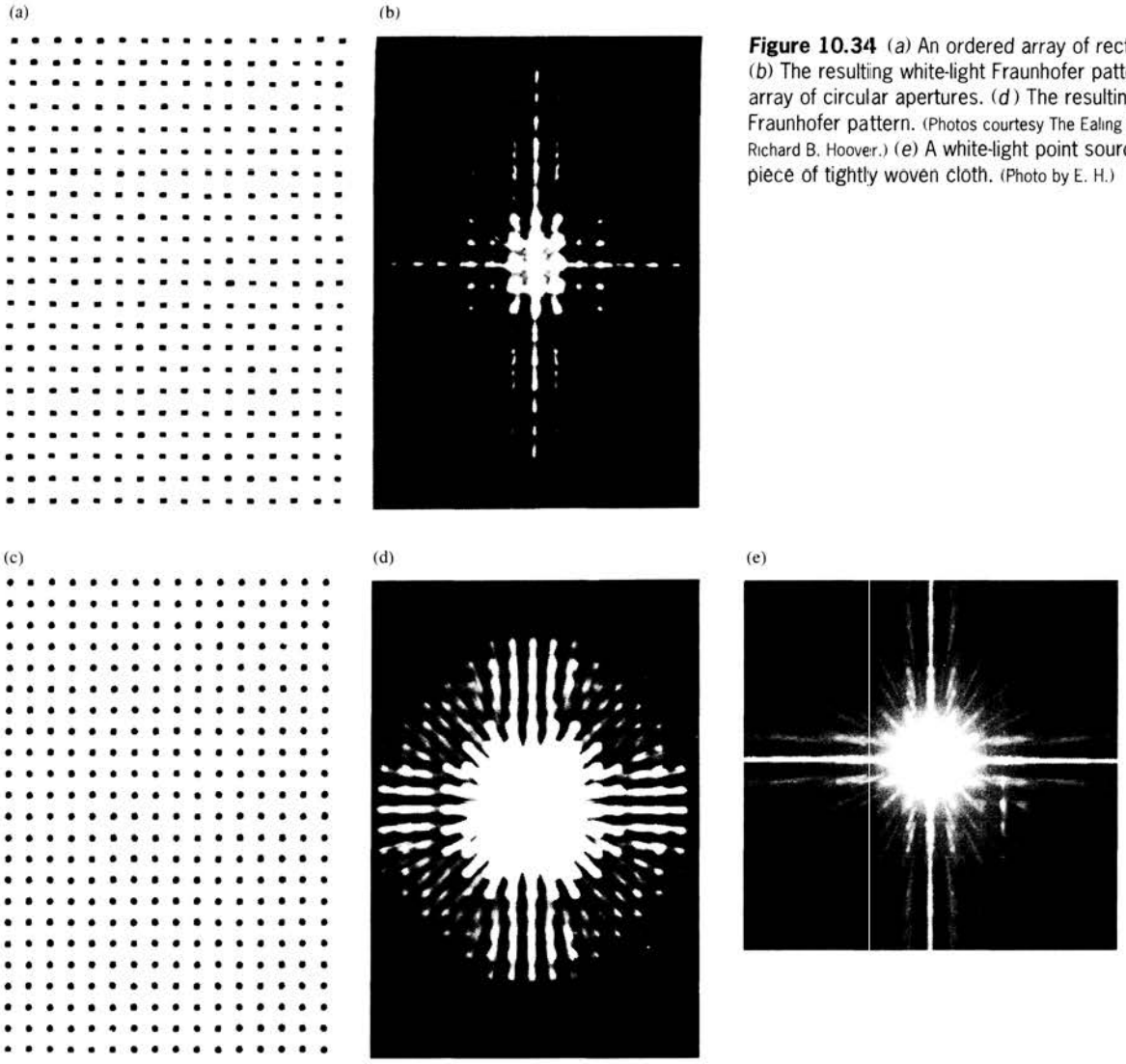


Figure 10.34 (*a*) An ordered array of rectangular apertures. (*b*) The resulting white-light Fraunhofer pattern. (*c*) An ordered array of circular apertures. (*d*) The resulting white-light Fraunhofer pattern. (Photos courtesy The Ealing Corporation and Richard B. Hoover.) (*e*) A white-light point source seen through a piece of tightly woven cloth. (Photo by E. H.)

As for the possibility of a *three-dimensional grating*, there seems to be no particular conceptual difficulty. A regular spatial array of scattering centers would certainly yield interference maxima in preferred directions. In 1912 Max von Laue (1879–1960) conceived the ingenious idea of using the regularly spaced atoms within a crystal as a three-dimensional grating. It is apparent from the grating equation [Eq. (10.61)] that if λ is much greater than the grating spacing, only the zeroth order ($m = 0$) is possible. This is equivalent to $\theta_0 = \theta_r$, that is, specular reflection. Since the spacing between atoms in a crystal is generally several angstroms (1 Å $= 10^{-1}$ nm), light can be diffracted only in the zeroth order.

Von Laue's solution to the problem was to probe the lattice, not with light but with X-rays whose wavelengths were comparable to the interatomic distances (Fig. 10.35a). A narrow beam of white radiation (the broad continuous frequency range emitted by an X-ray tube) was directed onto a thin single crystal. The film plate (Fig. 10.35b) revealed a Fraunhofer pattern consisting of an array of precisely located spots. These sites of constructive interference occurred whenever the angle

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

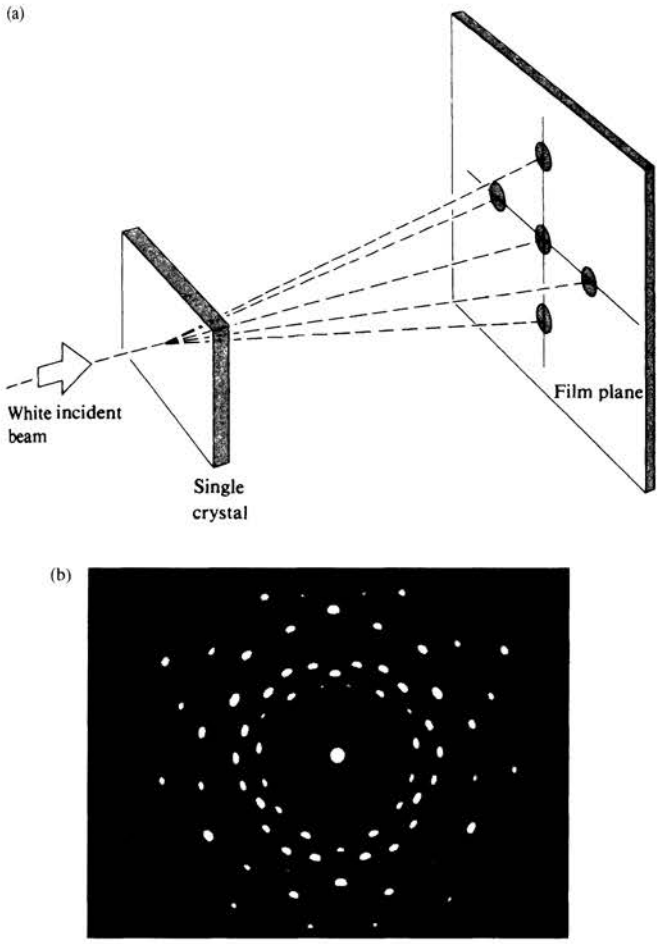


Figure 10.35 (*a*) Transmission Laue pattern. (*b*) X-ray diffraction pattern for quartz (SiO_2).

between the beam and a set of atomic planes within the crystal obeyed Bragg's Law:

$$2d \sin \theta = m\lambda \tag{10.71}$$

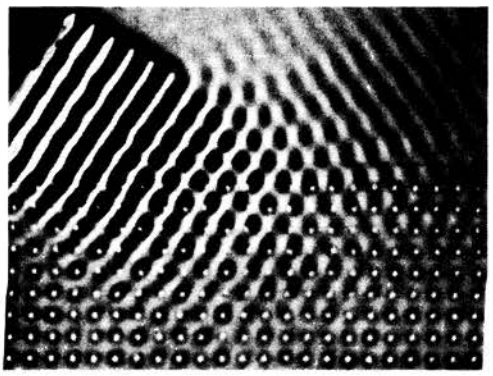
Notice that in X-ray work θ is traditionally measured from the plane and not the normal to it. Each set of planes diffracts a particular wavelength into a particular direction. The accompanying photo rather strikingly shows the analogous behavior in a ripple tank.

Instead of reducing λ to the X-ray range, we could have scaled everything up by a factor of about a billion and made a lattice of metal balls as a grating for microwaves.

10.3 Fresnel Diffraction

10.3.1 The Free Propagation of a Spherical Wave

In the Fraunhofer configuration, the diffracting system was relatively small, and the point of observation was very distant. Under these circumstances a few potentially problematic features of the Huygens–Fresnel Principle could be completely passed over without concern. But we are now going to deal with the near-field region, which extends right up to the diffracting element itself, and any such approximations would be inappropriate. We therefore return to the Huygens–Fresnel Principle in order to reexamine it more closely. At any instant, every point on the primary wavefront is envisioned as a continuous emitter of spherical secondary wavelets. But if each wavelet radiated uniformly in all directions, in addition to generating an ongoing wave, there would also be a reverse wave traveling back toward the source. No such wave is found experimentally, so we must somehow modify the radiation pattern of the secondary emitters. We now introduce the function $K(\theta)$, known as the **obliquity** or **inclination factor**, in order to describe the directionality of the secondary emissions.



Water waves in a ripple tank reflecting off an array of pegs acting as point scatterers. (Photo courtesy PSSC *Physics*, D. C. Heath, Boston, 1960.)

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

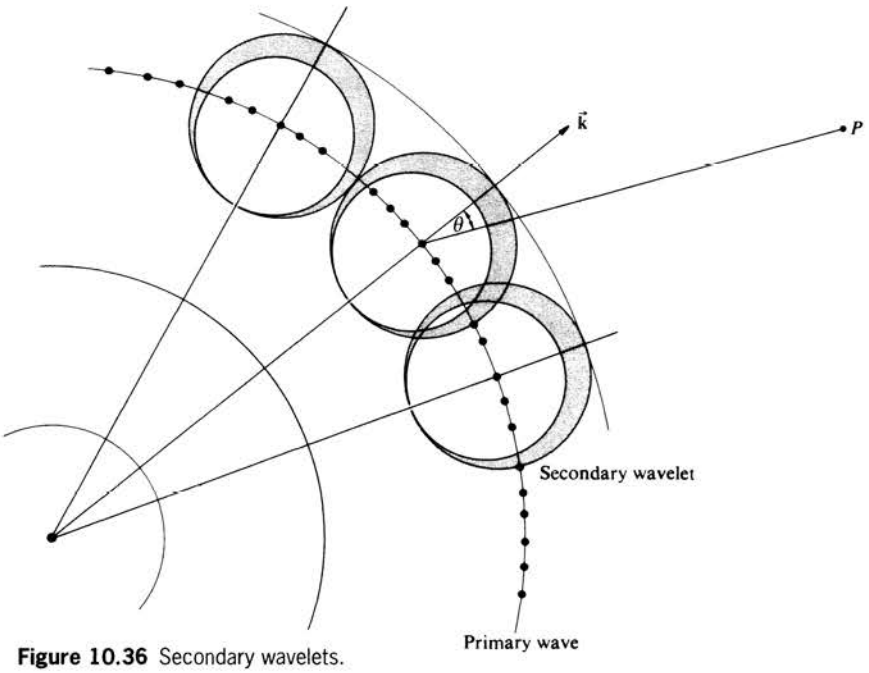


Figure 10.36 Secondary wavelets.

Fresnel recognized the need to introduce a quantity of this kind, but he did little more than conjecture about its form.* It remained for the more analytic Kirchhoff formulation to provide an actual expression for $K(\theta)$, which, as we will see in Section 10.4, turns out to be

$$K(\theta) = \tfrac{1}{2}(1 + \cos\theta) \qquad (10.72)$$

As shown in Fig. 10.36, θ is the angle made with the normal to the primary wavefront, $\vec{\mathbf{k}}$. This has its maximum value, $K(0) = 1$, in the forward direction and also dispenses with the back wave, since $K(\pi) = 0$.

Let us now examine the free propagation of a spherical monochromatic wave emitted from a *point source S*. If the Huygens–Fresnel Principle is correct, we should be able to add up the secondary wavelets arriving at a point P and thus obtain the unobstructed primary wave. In the process we will gain some insights, recognize a few shortcomings, and develop a very useful technique. Consider the construction shown in Fig. 10.37. The spherical surface corresponds to the primary wavefront at some arbitrary time t' after it has been emitted from S at $t = 0$. The disturbance, having a radius ρ, can be represented by any one of the mathematical expressions describing a harmonic spherical wave, for example,

$$E = \frac{\mathcal{E}_0}{\rho}\cos(\omega t' - k\rho) \qquad (10.73)$$

As illustrated, we have divided the wavefront into a number of annular regions. The boundaries of the various regions correspond to the intersections of the wavefront with a series of spheres centered at P of radius $r_0 + \lambda/2$, $r_0 + \lambda$, $r_0 + 3\lambda/2$, and so forth. These are the **Fresnel** or **half-period zones**. Notice that, for a secondary point source in one zone, there will be a point source in the adjacent zone that is further from P by an amount $\lambda/2$. Since each zone, though small, is finite in extent, we define a ring-shaped differential area element dS, as indicated in Fig. 10.38. All the point sources within dS are coherent, and *we assume that each radiates in-phase with the primary wave* [Eq. (10.73)]. The secondary wavelets travel a distance r to reach P, at a time t, all arriving there with the same phase, $\omega t - k(\rho + r)$. The amplitude of the primary wave at a distance ρ from S is $\mathcal{E}_0/\rho$. We assume, accordingly, that the source strength per unit area $\mathcal{E}_A$ of the secondary emitters on dS is proportional to $\mathcal{E}_0/\rho$ by way of a constant Q, that is, $\mathcal{E}_A = Q\mathcal{E}_0/\rho$. The contribution of the optical disturbance at P from the secondary sources on dS is, therefore,

$$dE = K\frac{\mathcal{E}_A}{r}\cos[\omega t - k(\rho + r)]\,dS \qquad (10.74)$$

The obliquity factor must vary slowly and may be assumed to be constant over a single Fresnel zone. To get dS as a function

*It is interesting to read Fresnel's own words on the matter, keeping in mind that he was talking about light as an elastic vibration of the aether.

> Since the impulse communicated to every part of the primitive wave was directed along the normal, the motion which each tends to impress upon the aether ought to be more intense in this direction than in any other; and the rays which would emanate from it, if acting alone, would be less and less intense as they deviated more and more from this direction.
>
> The investigation of the law according to which their intensity varies about each center of disturbance is doubtless a very difficult matter;...

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

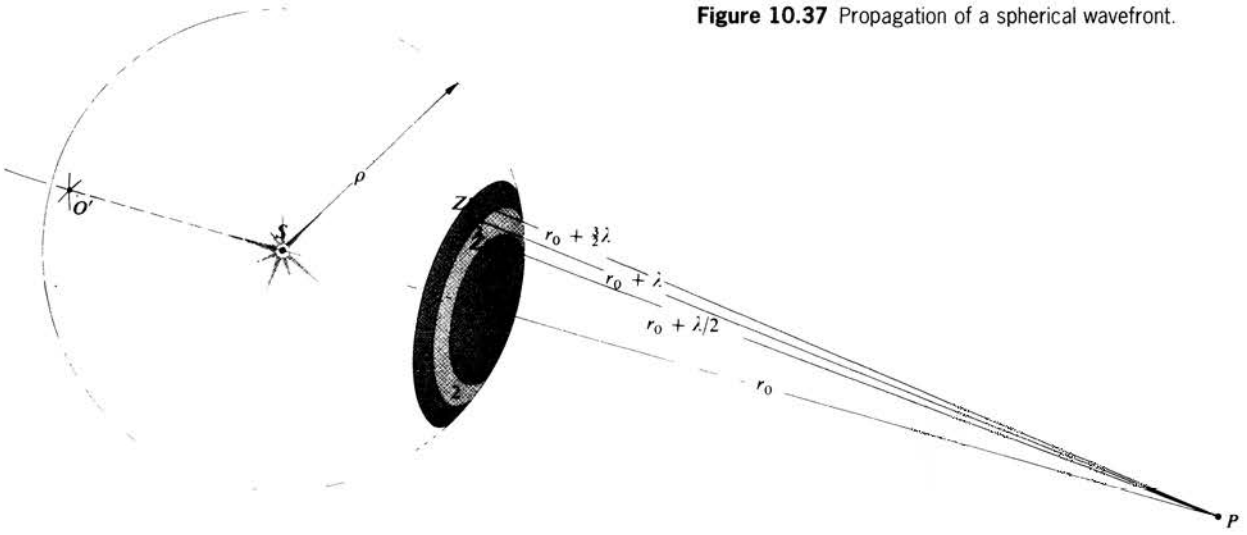


Figure 10.37 Propagation of a spherical wavefront.

of r, begin with

$$dS = \rho\, d\varphi\, 2\pi(\rho \sin \varphi)$$

Applying the law of cosines, we get

$$r^2 = \rho^2 + (\rho + r_0)^2 - 2\rho(\rho + r_0)\cos \varphi$$

Upon differentiation, this yields

$$2r\, dr = 2\rho(\rho + r_0) \sin \varphi\, d\varphi$$

with ρ and r_0 held constant. Making use of the value of $d\varphi$, we find that the area of the element is therefore

$$dS = 2\pi \frac{\rho}{(\rho + r_0)} r\, dr \tag{10.75}$$

The disturbance arriving at P from the lth zone is

$$E_l = K_l 2\pi \frac{\mathcal{E}_A \rho}{(\rho + r_0)} \int_{r_{l-1}}^{r_l} \cos[\omega t - k(\rho + r)]\, dr$$

Hence

$$E_l = \frac{-K_l \mathcal{E}_A \rho \lambda}{(\rho + r_0)} [\sin(\omega t - k\rho - kr)]_{r=r_{l-1}}^{r=r_l}$$

Upon the introduction of $r_{l-1} = r_0 + (l-1)\lambda/2$ and $r_l = r_0 + l\lambda/2$, the expression reduces (Problem 10.42) to

$$E_l = (-1)^{l+1} \frac{2K_l \mathcal{E}_A \rho \lambda}{(\rho + r_0)} \sin[\omega t - k(\rho + r_0)] \tag{10.76}$$

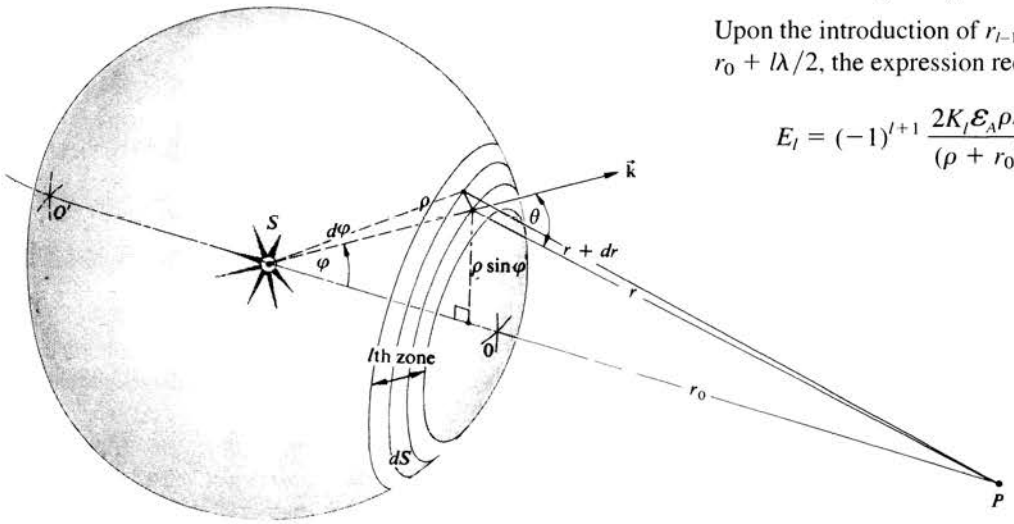


Figure 10.38 Propagation of a spherical wavefront.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

Observe that the amplitude of E_l alternates between positive and negative values, depending on whether l is odd or even. This means that the contributions from adjacent zones are out-of-phase and tend to cancel. It is here that the obliquity factor makes a crucial difference. As l increases, θ increases and K decreases, so that successive contributions do not in fact completely cancel each other. It is interesting to note that E_l/K_l is independent of any position variables. Although the areas of each zone are almost equal, they do increase slightly as l increases, which means an increased number of emitters. But the average distance from each zone to P also increases, such that E_l/K_l remains constant (see Problem 10.43).

The sum of the optical disturbances from all m zones at P is

$$E = E_1 + E_2 + E_3 + \cdots + E_m$$

and since these alternate in sign, we can write

$$E = |E_1| - |E_2| + |E_3| - \cdots \pm |E_m| \tag{10.77}$$

If m is odd, the series can be reformulated in two ways, either as

$$E = \frac{|E_1|}{2} + \left(\frac{|E_1|}{2} - |E_2| + \frac{|E_3|}{2}\right) + \left(\frac{|E_3|}{2} - |E_4| + \frac{|E_5|}{2}\right) + \cdots + \left(\frac{|E_{m-2}|}{2} - |E_{m-1}| + \frac{|E_m|}{2}\right) + \frac{|E_m|}{2} \tag{10.78}$$

or as

$$E = |E_1| - \frac{|E_2|}{2} - \left(\frac{|E_2|}{2} - |E_3| + \frac{|E_4|}{2}\right) - \left(\frac{|E_4|}{2} - |E_5| + \frac{|E_6|}{2}\right) + \cdots + \left(\frac{|E_{m-3}|}{2} - |E_{m-2}| + \frac{|E_{m-1}|}{2}\right) - \frac{|E_{m-1}|}{2} + |E_m| \tag{10.79}$$

There are now two possibilities: either $|E_l|$ is greater than the arithmetic mean of its two neighbors $|E_{l-1}|$ and $|E_{l+1}|$, or it is less than that mean. This is really a question concerning the rate of change of $K(\theta)$. When

$$|E_l| > (|E_{l-1}| + |E_{l+1}|)/2$$

each bracketed term is negative. It follows from Eq. (10.78) that

$$E < \frac{|E_1|}{2} + \frac{|E_m|}{2} \tag{10.80}$$

and from Eq. (10.79) that

$$E > |E_1| - \frac{|E_2|}{2} - \frac{|E_{m-1}|}{2} + |E_m| \tag{10.81}$$

Since the obliquity factor goes from 1 to 0 over a great many zones, we can neglect any variation between adjacent zones, that is, $|E_1| \approx |E_2|$ and $|E_{m-1}| \approx |E_m|$. Expression (10.81), to the same degree of approximation, becomes

$$E > \frac{|E_1|}{2} + \frac{|E_m|}{2} \tag{10.82}$$

We conclude from Eqs. (10.80) and (10.82) that

$$E \approx \frac{|E_1|}{2} + \frac{|E_m|}{2} \tag{10.83}$$

This same result is obtained when

$$|E_l| < (|E_{l-1}| + |E_{l+1}|)/2$$

If the last term, $|E_m|$, in the series of Eq. (10.77) corresponds to an even m, the same procedure (Problem 10.44) leads to

$$E \approx \frac{|E_1|}{2} - \frac{|E_m|}{2} \tag{10.84}$$

Fresnel conjectured that the obliquity factor was such that the last contributing zone occurred at $\theta = 90°$, that is,

$$K(\theta) = 0 \quad \text{for } \pi/2 \leq |\theta| \leq \pi$$

In that case Eqs. (10.83) and (10.84) both reduce to

$$E \approx \frac{|E_1|}{2} \tag{10.85}$$

when $|E_m|$ goes to zero, because $K_m(\pi/2) = 0$. Alternatively, using Kirchhoff's correct obliquity factor, we divide the *entire* spherical wave into zones with the last or mth zone surrounding O'. Now θ approaches π, $K_m(\pi) = 0$, $|E_m| = 0$, and once again $E \approx |E_1|/2$. *The optical disturbance generated by the entire unobstructed wavefront is approximately equal to one-half the contribution from the first zone.*

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

If the primary wave were simply to propagate from S to P in a time t, it would have the form

$$E = \frac{\mathcal{E}_0}{(\rho + r_0)} \cos [\omega t - k(\rho + r_0)] \qquad (10.86)$$

Yet the disturbance synthesized from secondary wavelets, Eqs. (10.76) and (10.85), is

$$E = \frac{K_1 \mathcal{E}_A \rho \lambda}{(\rho + r_0)} \sin [\omega t - k(\rho + r_0)] \qquad (10.87)$$

These two equations must, however, be exactly equivalent, and we interpret the constants in Eq. (10.87) to make them so. Note that there is some latitude in how we do this. We prefer to have the obliquity factor equal to 1 in the forward direction, that is, $K_1 = 1$ (rather than $1/\lambda$), from which it follows that Q must be equal to $1/\lambda$. In that case, $\mathcal{E}_A \rho \lambda = \mathcal{E}_0$, which is fine dimensionally. Keep in mind that $\mathcal{E}_A$ is the secondary-wavelet source strength per unit area over the primary wavefront of radius ρ, and $\mathcal{E}_0/\rho$ is the amplitude of that primary wave $E_0(\rho)$. Thus $\mathcal{E}_A = E_0(\rho)/\lambda$. There is one other problem, and that's the $\pi/2$ phase difference between Eqs. (10.86) and (10.87). This can be accounted for if we are willing to assume that the secondary sources radiate one-quarter of a wavelength out-of-phase with the primary wave (see Section 4.2.3).

We have found it necessary to modify the initial statement of the Huygens–Fresnel Principle, but this should not distract us from our rather pragmatic reasons for using it, which are twofold. First, the Huygens–Fresnel theory can be shown to be an approximation of the Kirchhoff formulation and as such is no longer merely a contrivance. Second, it yields, in a fairly simple way, many predictions that are in fine agreement with experimental observations. Don't forget that it worked quite well in the Fraunhofer approximation.

10.3.2 The Vibration Curve

We now develop a graphic method for qualitatively analyzing a number of diffraction problems that arise predominantly from circularly symmetric configurations.

Imagine that the first, or polar, Fresnel zone in Fig. 10.37 is divided into N subzones by the intersection of spheres, centered on P, of radii

$$r_0 + \lambda/2N,\ r_0 + \lambda/N,\ r_0 + 3\lambda/2N, \ldots,\ r_0 + \lambda/2$$

Each subzone contributes to the disturbance at P, the resultant of which is of course just E_1. Since the phase difference across the entire zone, from O to its edge, is π rad (corresponding to $\lambda/2$), each subzone is shifted by π/N rad. Figure 10.39 depicts the vector addition of the subzone phasors, where, for convenience, $N = 10$. The chain of phasors deviates very slightly from the circle, because the obliquity factor shrinks each successive amplitude. When the number of subzones is increased to infinity (i.e., $N \to \infty$), the polygon of vectors blends into a segment of a smooth spiral called a **vibration curve**. For each additional Fresnel zone, the vibration curve swings through *one half-turn* and a phase of π as it spirals inward. As shown in Fig. 10.40, the points $O_s, Z_{s1}, Z_{s2}, Z_{s3}, \ldots, O'_s$ on the spiral correspond to points $O, Z_1, Z_2, Z_3, \ldots, O'$, respectively, on the wavefront in Fig. 10.37. Each point $Z_1, Z_2, \ldots, Z_m$ lies on the periphery of a zone, so each point $Z_{s1}, Z_{s2}, \ldots, Z_{sm}$ is separated by a half-turn. We will see later, in Eq. (10.91), that the radius of each zone is proportional to the square root of its numerical designation, m. The radius of the hundredth zone will be only 10 times that of the first zone. Initially, therefore, the angle θ increases rapidly; thereafter it gradually slows down as m becomes larger. Accordingly, $K(\theta)$ decreases rapidly only for the first few zones. The result is that as the spiral circulates around with increasing m, it becomes tighter and tighter, deviating from a circle by a smaller amount for each revolution.

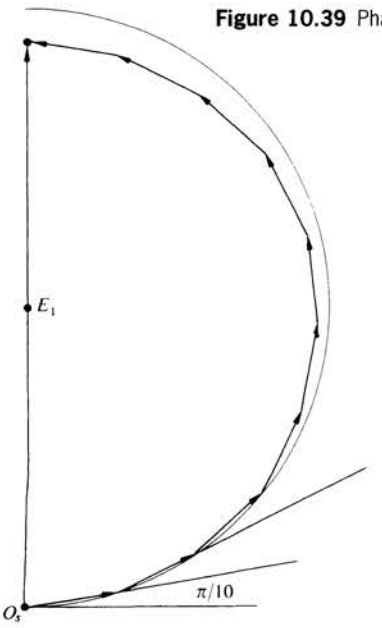


Figure 10.39 Phasor addition.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

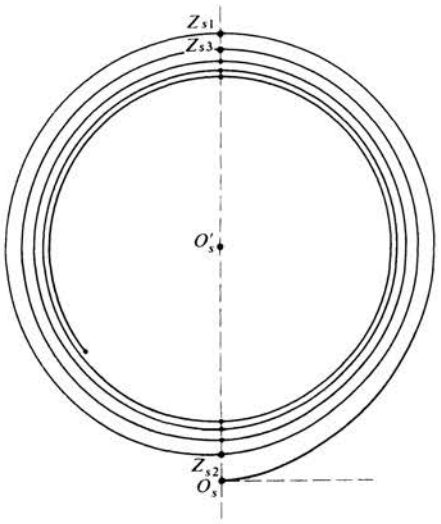


Figure 10.40 Overlapping point images.

Keep in mind that the spiral is made up of an infinite number of phasors, each shifted by a small phase angle. The relative phase between any two disturbances at P, coming from two points on the wavefront, say O and A, can be depicted as shown in Fig. 10.41. The angle made by the tangents to the vibration curve, at points O_s and A_s, is β, and this is the desired phase difference. If the point A is considered to lie on the boundary of a cap-shaped region of the wavefront, the resultant at P from the whole region is $\overrightarrow{O_s A_s}$ at an angle δ.

The total disturbance arriving at P from an unimpeded wave is the sum of the contributions from all the zones between O and O'. The length of the vector from O_s to O'_s is therefore precisely that amplitude. Note that as expected, the amplitude $O_s O'_s$ is just about one-half the contribution from the first zone, $O_s Z_{s1}$. Observe that $\overrightarrow{O_s O'_s}$ has a phase of 90° with respect to the wave arriving at P from O. A wavelet emitted at O in-phase with the primary excitation gets to P still in-phase with the primary wave. This means that $\overrightarrow{O_s O'_s}$ is 90° out-of-phase with the unobstructed primary wave. This, as we have seen, is one of the shortcomings of the Fresnel formulation.

10.3.3 Circular Apertures

Spherical Waves

Fresnel's procedure, applied to a point source, can be used as a semiquantitative method to study diffraction at a circular aperture. Envision a monochromatic spherical wave impinging on a screen containing a small hole, as illustrated in Fig. 10.42. We first record the irradiance arriving at a very small sensor placed at point P on the symmetry axis. Our intention is to move the sensor around in space and so get a point-by-point map of the irradiance of the region beyond Σ.

Let us assume that the sensor at P "sees" an integral number of zones, m, filling the aperture. In actuality, the sensor

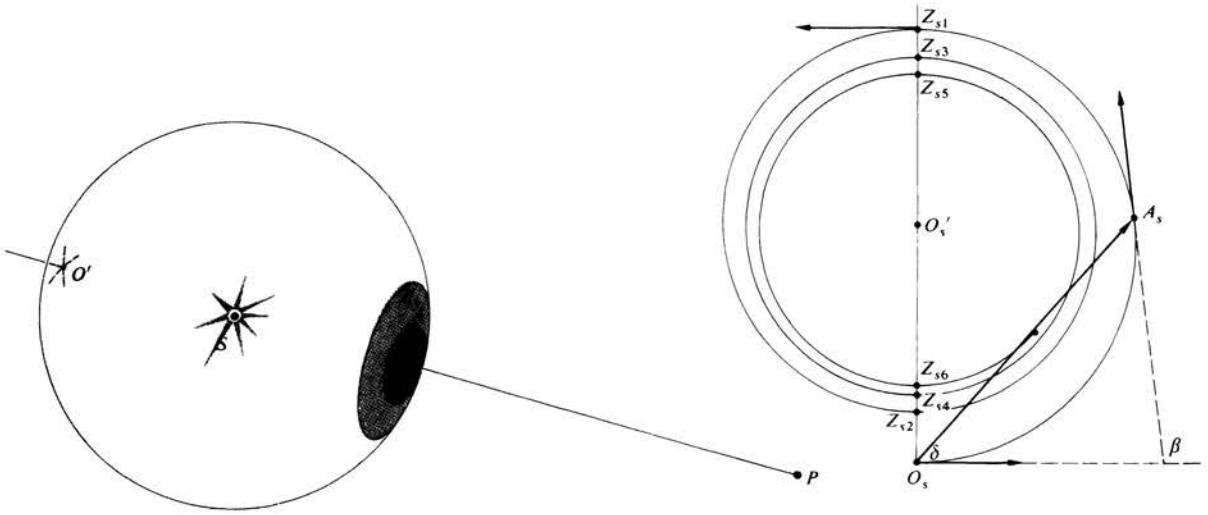


Figure 10.41 Wavefront and corresponding vibration curve.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

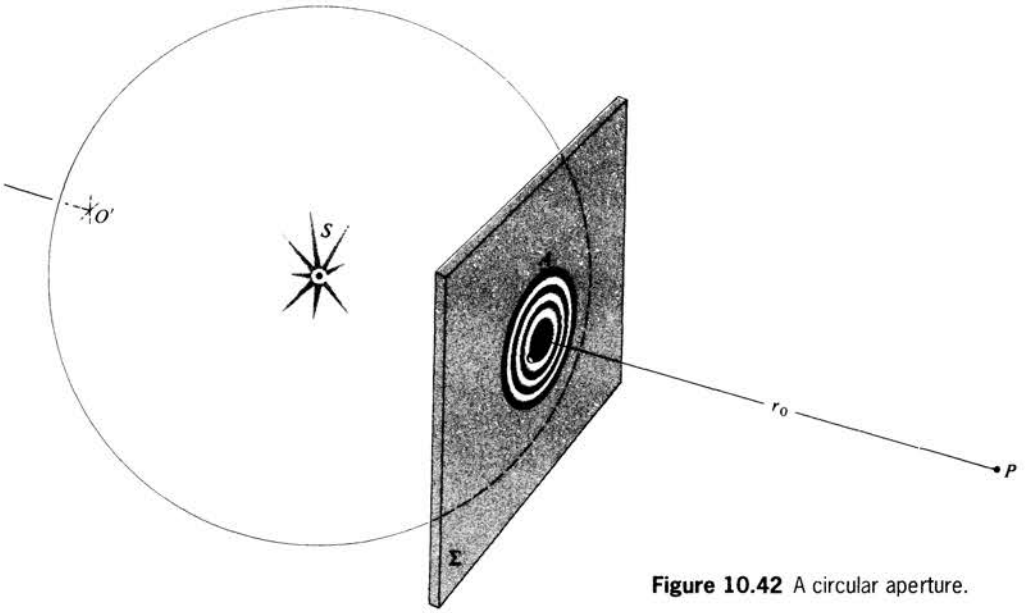


Figure 10.42 A circular aperture.

merely records the irradiance at P, the zones having no reality. If m is even, then since $K_m \neq 0$,

$$E = (|E_1| - |E_2|) + (|E_3| - |E_4|) + \cdots + (|E_{m-1}| - |E_m|)$$

Because each adjacent contribution is nearly equal,

$$E \approx 0$$

and $I \approx 0$. If, on the other hand, m is odd,

$$E = |E_1| - (|E_2| - |E_3|) - (|E_4| - |E_5|) - \cdots - (|E_{m-1}| - |E_m|)$$

and

$$E \approx |E_1|$$

which is roughly twice the amplitude of the unobstructed wave. This is truly an amazing result. By inserting a screen in the path of the wave, thereby blocking out most of the wavefront, we have increased the irradiance at P by a factor of four. Conservation of energy clearly demands that there be other points where the irradiance has decreased. Because of the complete symmetry of the setup, we can expect a circular ring pattern. If m is not an integer (i.e., a fraction of a zone appears in the aperture), the irradiance at P is somewhere between zero and its maximum value.

You might see this all a bit more clearly if you imagine that the aperture is expanding smoothly from an initial value of nearly zero. The amplitude at P can be determined from the vibration curve, where A is any point on the edge of the hole. The phasor magnitude O_sA_s is the desired amplitude of the optical field. Return to Fig. 10.41; as the hole increases, A_s moves counterclockwise around the spiral toward Z_{s1} and a maximum. Allowing the second zone in reduces O_sA_s to O_sZ_{s2}, which is nearly zero, and P becomes a dark spot. As the aperture increases, O_sA_s oscillates in length from nearly zero to a number of successive maxima, which themselves gradually decrease. Finally, when the hole is fairly large, the wave is essentially unobstructed, A_s approaches O_s', and further changes in O_sA_s are imperceptible.

To map the rest of the pattern, we now move the sensor along any line perpendicular to the axis, as shown in Fig. 10.43. At P we assume that two complete zones fill the aperture and $E \approx 0$. At P_1 the second zone has been partially obscured and the third begins to show; E is no longer zero. At P_2 a good fraction of the second zone is hidden, whereas the third is even more evident. Since the contributions from the first and third zones are in-phase, the sensor, placed anywhere on the dotted circle passing through P_2, records a bright spot. As it moves radially outward and portions of successive zones are uncovered, the sensor detects a series of relative maxima and minima. The photo on page 493 shows the diffraction patterns for a number of holes ranging in diameter from 1 mm to

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

4 mm as they appear on a screen 1 m away. Starting from the top left and moving right, the first four holes are so small that only a fraction of the first zone is uncovered. The sixth hole uncovers the first and second zones and is therefore black at its center. The ninth hole uncovers the first three zones and is once again bright at its center. Notice that even slightly beyond the geometric shadow at P_3, in Fig. 10.43, the first zone is partially uncovered. Each of the last few contributing segments is only a small fraction of its respective zone and as such is negligible. The sum of all the amplitudes of the fractional zones, though small, is therefore still finite. Farther into the geometric shadow, however, the entire first zone is obscured, the last terms are again negligible, and this time the series does indeed go to zero and darkness.

We can gain a better appreciation of the actual size of the things we are dealing with by computing the number of zones in a given aperture. The area of each zone (from Problem 10.43) is given by

$$A = \frac{\rho}{(\rho + r_0)} \pi r_0 \lambda \tag{10.88}$$

If the aperture has a radius R, a good approximation of the number of zones within it is simply

$$\frac{\pi R^2}{A} = \frac{(\rho + r_0)R^2}{\rho r_0 \lambda} \tag{10.89}$$

For example, with a point source 1 m behind the aperture ($\rho \approx 1$ m), a plane of observation 1 m in front of it ($r_0 = 1$ m), and $\lambda = 500$ nm, there are four zones when $R = 1$ mm, and 400 zones when $R = 1$ cm. When both ρ and r_0 are increased to the point where only a small fraction of a zone appears in the aperture, Fraunhofer diffraction occurs. This is essentially a restatement of the Fraunhofer condition of Section 10.1.2; see Problem 10.1 as well.

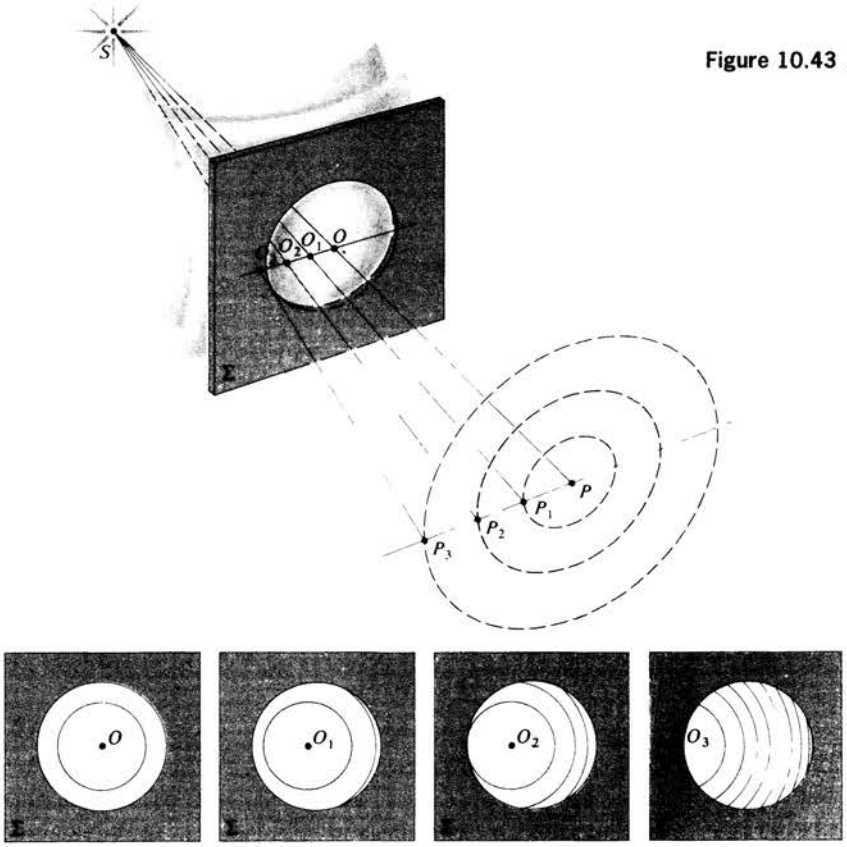
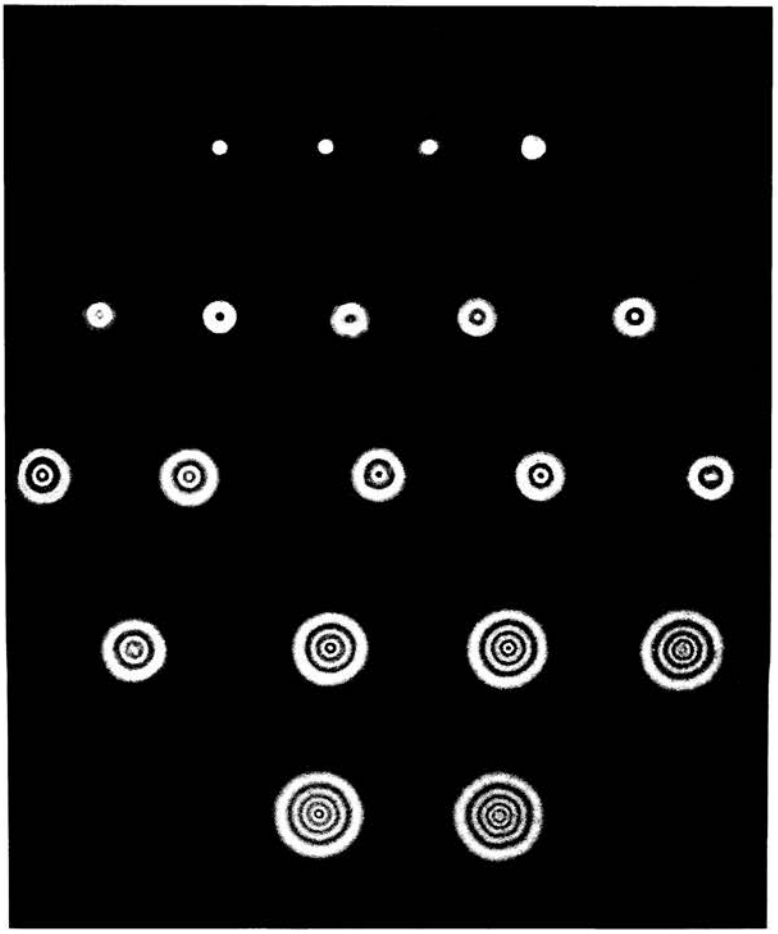


Figure 10.43 Zones in a circular aperture.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027



Diffraction patterns for circular apertures of increasing size. (Francis Weston Sears, *Optics*, ©1949, Addison-Wesley Reading, MA.

It follows from Eq. (10.89) that the number of zones filling the aperture depends on the distance r_0 from P to O. As P moves in either direction along the central axis, the number of uncovered zones, whether increasing or decreasing, oscillates between odd and even integers. As a result, the irradiance goes through a series of maxima and minima. Clearly, this does not occur in the Fraunhofer configuration, where by definition, more than one zone cannot appear in the aperture.*

Plane Waves

Suppose now that the point source has been moved so far from the diffracting screen that the incoming light can be regarded as a plane wave ($\rho \to \infty$). Referring to Fig. 10.44, we derive an expression for the radius of the mth zone, R_m. Since $r_m = r_0 + m\lambda/2$,

$$R_m^2 = (r_0 + m\lambda/2)^2 - r_0^2$$

and so

$$R_m^2 = mr_0\lambda + m^2\lambda^2/4 \tag{10.90}$$

Under most circumstances, the second term in Eq. (10.90) is negligible as long as m is not extremely large; consequently

$$R_m^2 = mr_0\lambda \tag{10.91}$$

and *the radii are proportional to the square roots of integers.* Using a collimated He–Ne laser ($\lambda_0 = 632.8$ nm), the radius of

*D. S. Burch, "Fresnel Diffraction by a Circular Aperture," *Am. J. Phys.* **53**, 255 (1985).

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

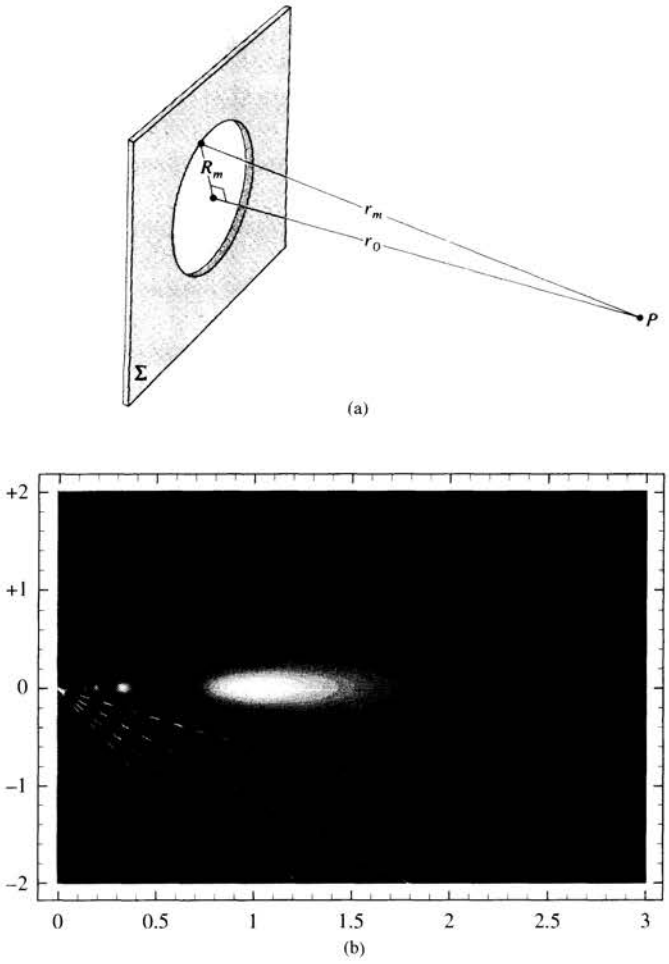


Figure 10.44 (*a*) Plane waves incident on a circular hole. (*b*) A cross section of the three-dimensional irradiance distribution. The horizontal axis is scaled in units of R^2/λ and the vertical axis in R, where R is the radius of the hole. Thus the aperture extends from +1 to −1. At a distance of $R^2/\lambda = r_0$ one Fresnel zone fills the aperture and the irradiance has a maximum. Beyond that $I(r)$ falls off monotonically until it reaches the far-field regime. The first four zeros of the Fraunhofer irradiance distribution lie on the dashed lines. (Photo courtesy, G. W. Forbes, The Institute of Optics, University of Rochester.)

the first zone is 1 mm when viewed from a distance of 1.58 m. Under these particular conditions Eq. (10.91) is applicable as long as $m << 10^7$, in which case $R_m = \sqrt{m}$ in millimeters. Figure 10.43 requires a slight modification in that now the lines $\overline{O_1P_1}$, $\overline{O_2P_2}$, and $\overline{O_3P_3}$ are perpendiculars dropped from the points of observation to Σ.

10.3.4 Circular Obstacles

In 1818 Fresnel entered a competition sponsored by the French Academy. His paper on the theory of diffraction ultimately won first prize and the title *Mémoire Courronné*, but not until it had provided the basis for a rather interesting story. The judging committee consisted of Pierre Laplace, Jean B. Biot, Siméon D. Poisson, Dominique F. Arago, and Joseph L. Gay-Lussac—a formidable group indeed. Poisson, who was an ardent critic of the wave description of light, deduced a remarkable and seemingly untenable conclusion from Fresnel's theory. He showed that a bright spot would be visible at the center of the shadow of a circular opaque obstacle, a result that he felt proved the absurdity of Fresnel's treatment. We can come to the same conclusion by considering the following, somewhat oversimplified argument. Recall that an unobstructed wave yields a disturbance [Eq. (10.85)] given by $E \approx |E_1|/2$. If some sort of obstacle precisely covers the first Fresnel zone, so that its contribution of $|E_1|$ is subtracted out, then $E \approx -|E_1|/2$. It is therefore possible that at some point P on the axis, the irradiance will be unaltered by the insertion of that obstruction. This surprising prediction, fashioned by Poisson as the death blow to the wave theory, was almost immediately verified experimentally by Arago; the spot actually existed. Amusingly enough, Poisson's spot, as it is now called, had been observed many years earlier (1723) by Maraldi, but this work had long gone unnoticed.*

We now examine the problem a bit more closely, since it's quite evident from the accompanying photo that there is a good deal of structure in the actual shadow pattern. If the opaque obstacle, be it a disk or sphere, obscures the first ℓ zones, then

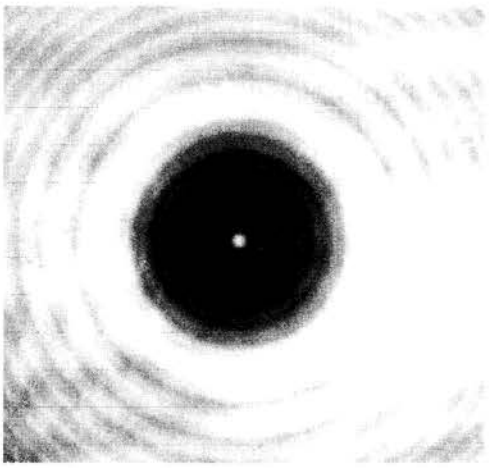
$$E = |E_{\ell+1}| - |E_{\ell+2}| + \cdots + |E_m|$$

(where, as before, there is no absolute significance to the signs other than that alternate terms must subtract). Unlike the analysis for the circular aperture, E_m now approaches zero, because $K_m \to 0$. The series must be evaluated in the same manner as that of the unobstructed wave [Eqs. (10.78) and (10.79)]. Repeating that procedure yields

$$E \approx \frac{|E_{\ell+1}|}{2} \tag{10.92}$$

*See J. E. Harvey and J. L. Forgham, "The Spot of Arago: New Relevance for an Old Phenomenon," *Am. J. Phys.* **52**, 243 (1984).

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027



Shadow of a 1/8-inch diameter ball bearing. The bearing was glued to an ordinary microscope slide and illuminated with a He–Ne laserbeam. There are some faint extraneous nonconcentric fringes arising from both the microscope slide and a lens in the beam. (Photo by E. H.)

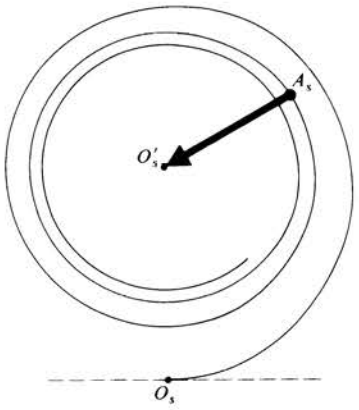


Figure 10.45 The vibration curve applied to a circular obstruction.

and the irradiance on the central axis is generally only slightly less than that of the unobstructed wave. *There is a bright spot everywhere along the central axis except immediately behind the circular obstacle.* The wavelets propagating beyond the disk's circumference meet in-phase on the central axis. Notice that as P moves close to the disk, θ increases, $K_{\ell+1} \to 0$, and the irradiance gradually falls off to zero. If the disk is large, the $(\ell + 1)$th zone is very narrow, and any irregularities in the obstacle's surface may seriously obscure that zone. For Poisson's spot to be readily observable, the obstacle must be smooth and circular.

If A is a point on the periphery of the disk or sphere, A_s is the corresponding point on the vibration curve (Fig. 10.45). As the disk increases for a fixed P, A_s spirals in counterclockwise toward O'_s, and the amplitude $A_sO'_s$ gradually decreases. The same thing happens as P moves toward a disk of constant size.

Off the axis, the zones covered in Fig. 10.43 for the circular aperture will now be exposed and vice versa. Accordingly, a whole series of concentric bright and dark rings will surround the central spot.

The opaque disk images S at P and would similarly form a crude image of every point in an extended source. R. W. Pohl has shown that a small disk can therefore be used as a crude positive lens.

The diffraction pattern can be seen with little difficulty, but you need a telescope or binoculars. Glue a small ball bearing ($\approx\frac{1}{8}$ or $\frac{1}{4}$ inch in diameter) to a microscope slide, which then serves as a handle. Place the bearing a few meters beyond the point source and observe it from 3 or 4 meters away. Position it so that it is directly in front of and completely obscuring the source. You will need the telescope to magnify the image, since r_0 is so large. If you can hold the telescope steady, the ring system should be quite clear.

10.3.5 The Fresnel Zone Plate

In our previous considerations, we utilized the fact that successive Fresnel zones tended to nullify each other. This suggests that we will observe a tremendous increase in irradiance at P, if we remove either all the even or all the odd zones. A screen that alters the light, either in amplitude or phase, coming from every other half-period zone is called a **zone plate**.*

Suppose that we construct a zone plate that passes only the first 20 odd zones and obstructs the even zones,

$$E = E_1 + E_3 + E_5 + \cdots + E_{39}$$

and each of these terms is approximately equal. For an unobstructed wavefront, the disturbance at P would be $E_1/2$,

*Lord Rayleigh seems to have invented the zone plate, as witnessed by this entry of April 11, 1871, in his notebook: "The experiment of blocking out the odd Huygens zones so as to increase the light at centre succeeded very well."

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

whereas with the zone plate in place, $E \approx 20E_1$. The irradiance has been increased by a factor of 1600. The same result would obviously be true if the even zones were passed instead.

To calculate the radii of the zones shown in Fig. 10.46, refer to Fig. 10.47. The outer edge of the mth zone is marked by the point A_m. By definition, a wave that travels the path S–A_m–P must arrive out-of-phase by $m\lambda/2$ with a wave that traverses the path S–O–P, that is,

$$(\rho_m + r_m) - (\rho_0 + r_0) = m\lambda/2 \qquad (10.93)$$

Clearly, $\rho_m = (R_m^2 + \rho_0^2)^{1/2}$ and $r_m = (R_m^2 + r_0^2)^{1/2}$. Expand both these expressions using the binomial series. Since R_m is

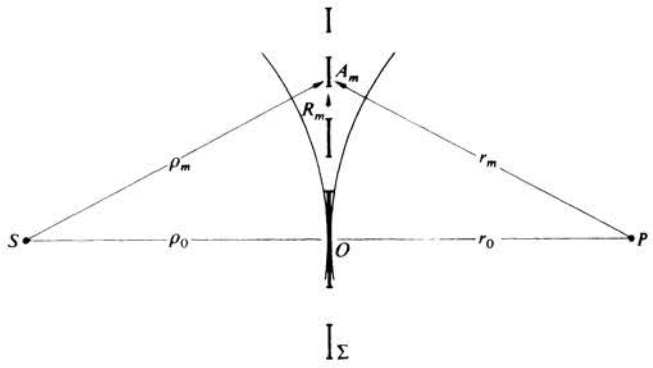


Figure 10.47 Zone-plate geometry.

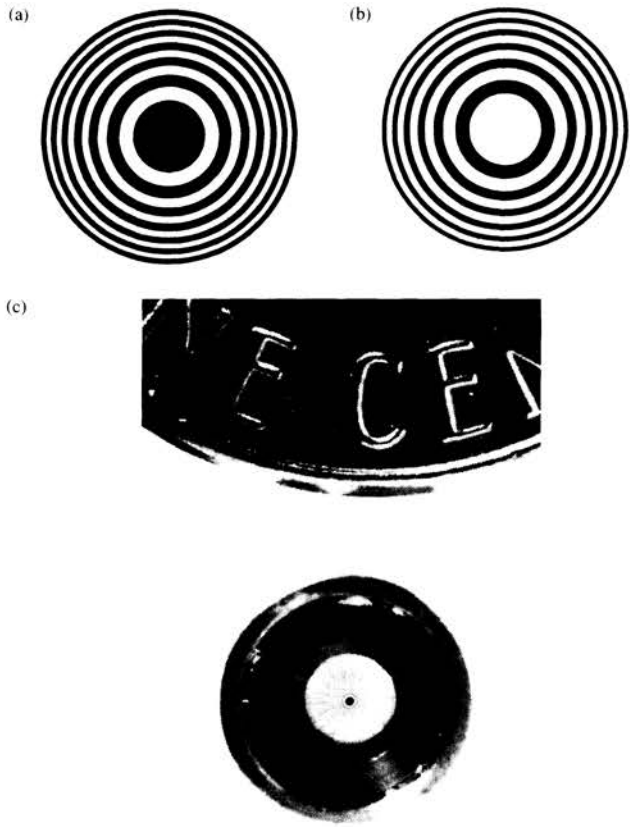


Figure 10.46 (*a*) and (*b*) zone plates. (*c*) A zone plate used to image alpha particles coming from a target 1 cm in front, on photographic film 5 cm behind. The plate is 2.5 mm in diameter and contains 100 zones, the narrowest of which is 5.3 μm wide. (Photo courtesy of the University of California, Lawrence Livermore National Laboratory and the Department of Energy.

comparatively small, retaining only the first two terms yields

$$\rho_m = \rho_0 + \frac{R_m^2}{2\rho_0} \quad \text{and} \quad r_m = r_0 + \frac{R_m^2}{2r_0}$$

Finally, substituting into Eq. (10.93), we obtain

$$\left(\frac{1}{\rho_0} + \frac{1}{r_0}\right) = \frac{m\lambda}{R_m^2} \qquad (10.94)$$

Under plane-wave illumination ($\rho_0 \to \infty$), and Eq. (10.94) reduces to

$$R_m^2 = mr_0\lambda \qquad [10.91]$$

which is an approximation of the exact expression stated by Eq. (10.90). Equation (10.94) has a form identical to that of the thin-lens equation, which is not merely a coincidence, since S is actually imaged in converging diffracted light at P. Accordingly, the *primary focal length* is said to be

$$f_1 = \frac{R_m^2}{m\lambda} \qquad (10.95)$$

(Note that the zone plate will show extensive chromatic aberration.) The points S and P are said to be conjugate foci. With a collimated incident beam (Fig. 10.48), the image distance is the primary or *first-order* focal length, which in turn corresponds to a principal maximum in the irradiance distribution. In addition to this real image, there is also a virtual image formed of diverging light a distance f_1 in front of Σ. At a distance of f_1 from Σ, each ring on the plate is filled by exactly one half-period zone on the wavefront. If we move a sensor along the S–P axis toward Σ, it registers a series of very small irradiance maxima and minima until it arrives at a point $f_1/3$

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

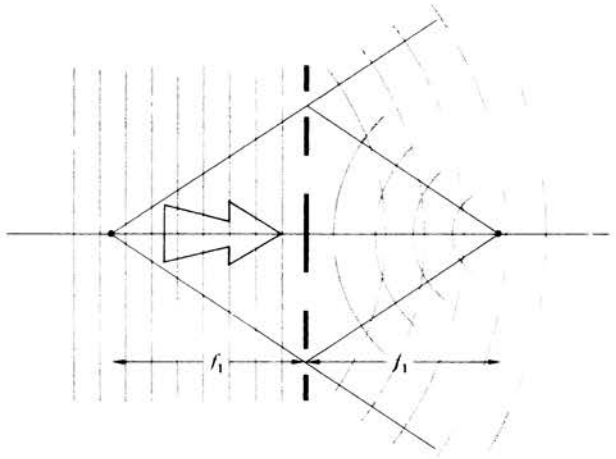


Figure 10.48 Zone-plate foci.

from Σ. At that *third-order focal point*, there is a pronounced irradiance peak. Additional focal points will exist at $f_1/5$, $f_1/7$, and so forth, unlike a lens but even more unlike a simple opaque disk.

Following a suggestion by Lord Rayleigh, R. W. Wood constructed a *phase-reversal zone plate*. Instead of blocking out every other zone, he increased the thickness of alternate zones, thereby retarding their phase by π. Since the entire plate is transparent, the amplitude should double, and the irradiance should increase by a factor of four. In actuality, the device does not work quite that well because the phase is not really constant over each zone. Ideally, the retardation should be made to vary gradually over a zone, jumping back by π at the start of the next zone.*

The usual way to make an optical zone plate is to draw a large-scale version and then photographically reduce it. Plates with hundreds of zones can be made by photographing a Newton's ring pattern, in collimated quasimonochromatic light. Rings of aluminum foil on cardboard work very well for microwaves.

*See Ditchburn, *Light*, 2nd ed., p. 232; M. Sussman, "Elementary Diffraction Theory of Zone Plates," *Am. J. Phys.* **28**, 394 (1960); Ora E. Myers, Jr., "Studies of Transmission Zone Plates," *Am. J. Phys.* **19**, 359 (1951); and J. Higbie, "Fresnel Zone Plate: Anomalous Foci," *Am. J. Phys.* **44**, 929 (1976).

Zone plates can be made of metal with a self-supporting spoked structure, so that the transparent regions are devoid of any material. These will function as lenses in the range from ultraviolet to soft X-rays, where ordinary glass is opaque.

10.3.6 Fresnel Integrals and the Rectangular Aperture

We now consider a class of problems within the domain of Fresnel diffraction, which no longer have the circular symmetry of the previously studied configurations. Consider Fig. 10.49 where dS is an area element situated at some arbitrary point A whose coordinates are (y, z). The location of the origin O is determined by a perpendicular drawn to Σ from the position of the monochromatic point source. The contribution to the optical disturbance at P from the secondary sources on dS has the form given by Eq. (10.74). Making use of what we learned from the freely propagating wave ($\mathcal{E}_A\rho\lambda = \mathcal{E}_0$), we can rewrite that equation as

$$dE_p = \frac{K(\theta)\mathcal{E}_0}{\rho r\lambda}\cos\left[k(\rho + r) - \omega t\right] dS \qquad (10.96)$$

The sign of the phase has changed from that of Eq. (10.74) and

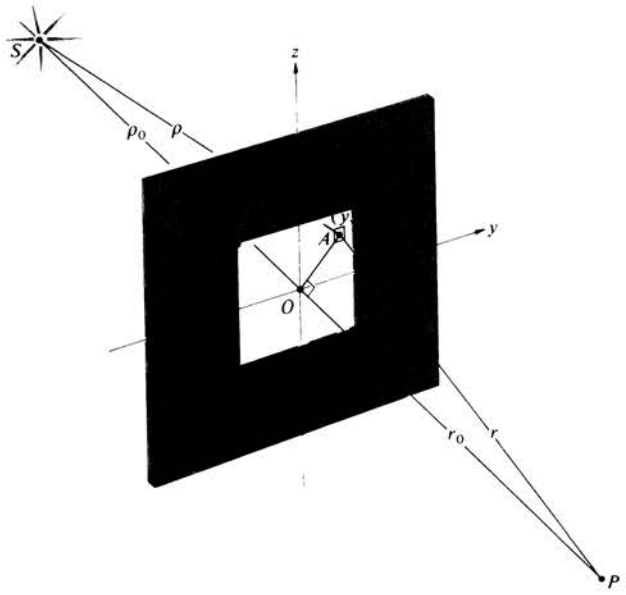


Figure 10.49 Fresnel diffraction at a rectangular aperture.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

is written in this way to conform with traditional treatment. *In the case where the dimensions of the aperture are small* in comparison to ρ_0 and r_0, we can set $K(\theta) = 1$ and let $1/\rho r$ equal $1/\rho_0 r_0$ in the amplitude coefficient. Being more careful about approximations introduced into the phase, apply the Pythagorean theorem to triangles *SOA* and *POA* to get

$$\rho = (\rho_0^2 + y^2 + z^2)^{1/2}$$

and

$$r = (r_0^2 + y^2 + z^2)^{1/2}$$

Expand these using the binomial series and form

$$\rho + r \approx \rho_0 + r_0 + (y^2 + z^2)\frac{\rho_0 + r_0}{2\rho_0 r_0} \tag{10.97}$$

Observe that this is a more sensitive approximation than that used in the Fraunhofer analysis [Eq. (10.40)], where the terms quadratic and higher in the aperture variables were neglected. The disturbance at P in the complex representation is

$$\tilde{E}_p = \frac{\mathcal{E}_0 e^{-i\omega t}}{\rho_0 r_0 \lambda} \int_{y_1}^{y_2} \int_{z_1}^{z_2} e^{ik(\rho + r)}\, dy\, dz \tag{10.98}$$

Following the usual form of derivation, we introduce the dimensionless variables u and v defined by

$$u \equiv y\left[\frac{2(\rho_0 + r_0)}{\lambda \rho_0 r_0}\right]^{1/2} \qquad v \equiv z\left[\frac{2(\rho_0 + r_0)}{\lambda \rho_0 r_0}\right]^{1/2} \tag{10.99}$$

Substituting Eq. (10.97) into Eq. (10.98) and utilizing the new variables, we arrive at

$$\tilde{E}_p = \frac{\mathcal{E}_0}{2(\rho_0 + r_0)} e^{i[k(\rho_0 + r_0) - \omega t]} \int_{u_1}^{u_2} e^{i\pi u^2/2}\, du \int_{v_1}^{v_2} e^{i\pi v^2/2}\, dv \tag{10.100}$$

The term in front of the integral represents the unobstructed disturbance at P divided by 2; let us call it $\tilde{E}_u/2$. The integral itself can be evaluated using two functions, $\mathscr{C}(w)$ and $\mathscr{S}(w)$, where w represents either u or v. These quantities, which are known as the **Fresnel integrals**, are defined by

$$\mathscr{C}(w) \equiv \int_0^w \cos(\pi w'^2/2)\, dw'$$

$$\mathscr{S}(w) \equiv \int_0^w \sin(\pi w'^2/2)\, dw' \tag{10.101}$$

Both functions have been extensively studied, and their numerical values are well tabulated. Their interest to us at this point derives from the fact that

$$\int_0^w e^{i\pi w'^2/2}\, dw' = \mathscr{C}(w) + i\mathscr{S}(w)$$

and this, in turn, has the form of the integrals in Eq. (10.100). The disturbance at P is then

$$\tilde{E}_p = \frac{\tilde{E}_u}{2}[\mathscr{C}(u) + i\mathscr{S}(u)]_{u_1}^{u_2}[\mathscr{C}(v) + i\mathscr{S}(v)]_{v_1}^{v_2} \tag{10.102}$$

which can be evaluated using the tabulated values of $\mathscr{C}(u_1)$, $\mathscr{C}(u_2)$, $\mathscr{S}(u_1)$, and so on. The mathematics becomes rather involved if we compute the disturbance at all points of the plane of observation, leaving the position of the aperture fixed. Instead we will fix the *S–O–P* line and imagine that we move the aperture through small displacements in the Σ-plane. This has the effect of translating the origin O with respect to the fixed aperture, thereby scanning the pattern over the point P. Each new position of O corresponds to a new set of relative boundary locations y_1, y_2, z_1, and z_2. These in turn mean new values of u_1, u_2, v_1, and v_2, which, when substituted into Eq. (10.102), yield a new $\tilde{E}_p$. The error encountered in such a procedure is negligible, as long as the aperture is displaced by distances that are small compared with ρ_0. This approach is therefore even more appropriate to incident plane waves. In that case, if E_0 is the amplitude of the incoming plane wave at Σ, Eq. (10.96) becomes simply

$$dE_p = \frac{E_0 K(\theta)}{r\lambda} \cos(kr - \omega t)\, dS$$

where, as before, $\mathcal{E}_A = E_0/\lambda$. This time, with

$$u = y\left(\frac{2}{\lambda r_0}\right)^{1/2} \qquad v = z\left(\frac{2}{\lambda r_0}\right)^{1/2} \tag{10.103}$$

where we have divided the numerator and denominator in Eq. (10.99) by ρ_0 and then let it go to infinity, $\tilde{E}_p$ takes the same form as Eq. (10.102), where $\tilde{E}_u$ is again the unobstructed disturbance. The irradiance at P is $\tilde{E}_p \tilde{E}_p^*/2$; hence

$$I_p = \frac{I_0}{4}\{[\mathscr{C}(u_2) - \mathscr{C}(u_1)]^2 + [\mathscr{S}(u_2) - \mathscr{S}(u_1)]^2\} \times \{[\mathscr{C}(v_2) - \mathscr{C}(v_1)]^2 + [\mathscr{S}(v_2) - \mathscr{S}(v_1)]^2\} \tag{10.104}$$

where I_0 is the unobstructed irradiance at P.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

As a simple example, envision a square hole 2 mm on each side under plane-wave illumination at 500 nm. If P is 4 m away and directly opposite point O at the center of the aperture, $u_2 = 1.0$, $u_1 = -1.0$, $v_2 = 1.0$, and $v_1 = -1.0$. The Fresnel integrals are both odd functions, that is,

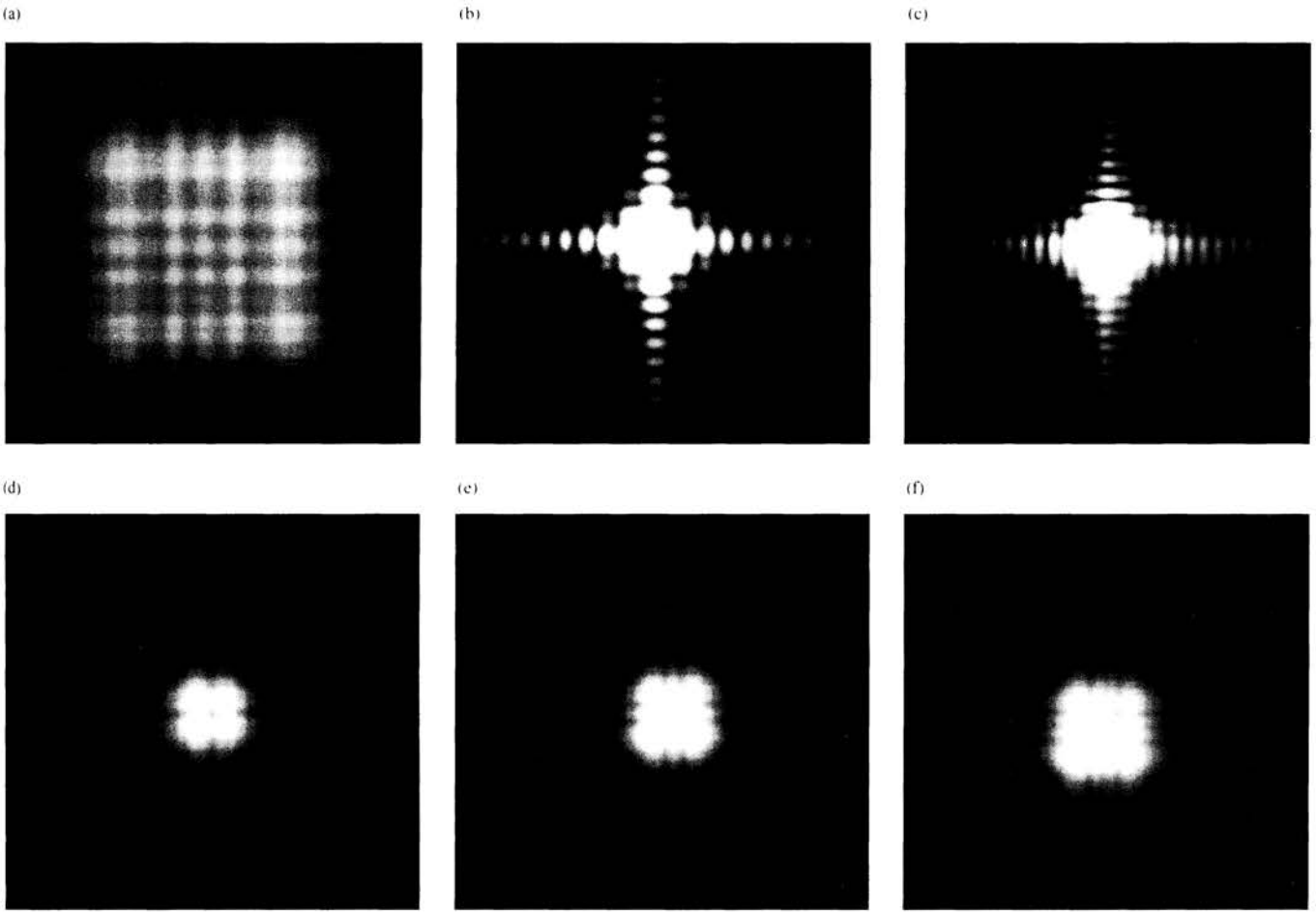
$$\mathscr{C}(w) = -\mathscr{C}(-w) \quad \text{and} \quad \mathscr{S}(w) = -\mathscr{S}(-w)$$

Consequently

$$I_p = \frac{I_0}{4}\{[2\mathscr{C}(1)]^2 + [2\mathscr{S}(1)]^2\}^2$$

and a numerical value is easily obtained. To find the irradiance somewhere else in the pattern, for example, 0.1 mm to the left of center, move the aperture relative to the OP-line accordingly, whereupon $u_2 = 1.1$, $u_1 = -0.9$, $v_2 = 1.0$, and $v_1 = -1.0$. The resultant I_p will also be equal to that found at 0.1 mm to the right of center. Indeed, because the aperture is square, the same value obtains 0.1 mm directly above and below center as well (see photo).

We can approach the limiting case of free propagation by allowing the aperture dimensions to increase indefinitely. Making use of the fact that $\mathscr{C}(\infty) = \mathscr{S}(\infty) = \frac{1}{2}$ and $\mathscr{C}(-\infty)$



(a) A typical Fresnel pattern for a square aperture. (b)–(f) A series of Fresnel patterns for increasing square apertures under identical conditions. Note that as the hole gets larger, the pattern changes from a spread-out Fraunhofer-like distribution to a far more localized structure. (Photos by E. H.)

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

numerical values are taken from Table 10.2. If $d\ell$ is an element of arc length measured along the curve, then

$$d\ell^2 = d\mathscr{C}^2 + d\mathscr{S}^2$$

From the definitions (10.101),

$$d\ell^2 = (\cos^2 \pi w^2/2 + \sin^2 \pi w^2/2)dw^2$$

and

$$d\ell = dw$$

Values of w correspond to the arc length and are marked off along the spiral in Fig. 10.50. As w approaches $\pm\infty$, the curve spirals into its limiting values at $\tilde{B}^+ = \frac{1}{2} + i\frac{1}{2}$ and $\tilde{B}^- = -\frac{1}{2} - i\frac{1}{2}$. The slope of the spiral is

$$\frac{d\mathscr{S}}{d\mathscr{C}} = \frac{\sin \pi w^2/2}{\cos \pi w^2/2} = \tan \frac{\pi w^2}{2} \qquad (10.105)$$

and so the angle between the tangent to the spiral at any point and the $\mathscr{C}$-axis is $\beta = \pi w^2/2$.

The Cornu spiral can be used either as a convenient tool for quantitative determinations or as an aid to gaining a qualitative picture of a diffraction pattern (which was also the case with the vibration curve). As an example of its quantitative uses, reconsider the problem of a 2-mm-square hole, dealt with in the previous section ($\lambda = 500$ nm, $r_0 = 4$ m, and plane-wave illumination). We wish to find the irradiance at P directly opposite the aperture's center, where in this case $u_1 = -1.0$ and $u_2 = 1.0$. The variable u is measured along the arc; that is, w is replaced by u on the spiral. Place two points on the spiral at distances from O_s equal to u_1 and u_2. (These are symmetrical with respect to O_s, because P is now opposite the aperture's center.) Label the two points $\tilde{B}_1(u)$ and $\tilde{B}_2(u)$, respectively, as in Fig. 10.51. The phasor $\tilde{\mathbf{B}}_{12}(u)$ drawn from $\tilde{B}_1(u)$ to $\tilde{B}_2(u)$ is just the complex number $\tilde{B}_2(u) - \tilde{B}_1(u)$,

$$\tilde{\mathbf{B}}_{12}(u) = [\mathscr{C}(u) + i\mathscr{S}(u)]_{u_1}^{u_2}$$

and is the first term in the expression [Eq. (10.102)] for $\tilde{E}_p$. Similarly for $v_1 = -1.0$ and $v_2 = 1.0$, $\tilde{B}_2(v) - \tilde{B}_1(v)$ is

$$\tilde{\mathbf{B}}_{12}(v) = [\mathscr{C}(v) + i\mathscr{S}(v)]_{v_1}^{v_2}$$

which is the latter portion of $\tilde{E}_p$. The magnitudes of these two complex numbers are just the lengths of the appropriate $\tilde{\mathbf{B}}_{12}$-phasors, which can be read off the curve with a ruler, using either axis as a scale. The irradiance is then simply

$$I_p = \frac{I_0}{4} |\tilde{\mathbf{B}}_{12}(u)|^2 |\tilde{\mathbf{B}}_{12}(v)|^2 \qquad (10.106)$$

TABLE 10.2 Fresnel Integrals

w	$\mathscr{C}(w)$	$\mathscr{S}(w)$	w	$\mathscr{C}(w)$	$\mathscr{S}(w)$
0.00	0.0000	0.0000	4.50	0.5261	0.4342
0.10	0.1000	0.0005	4.60	0.5673	0.5162
0.20	0.1999	0.0042	4.70	0.4914	0.5672
0.30	0.2994	0.0141	4.80	0.4338	0.4968
0.40	0.3975	0.0334	4.90	0.5002	0.4350
0.50	0.4923	0.0647	5.00	0.5637	0.4992
0.60	0.5811	0.1105	5.05	0.5450	0.5442
0.70	0.6597	0.1721	5.10	0.4998	0.5624
0.80	0.7230	0.2493	5.15	0.4553	0.5427
0.90	0.7648	0.3398	5.20	0.4389	0.4969
1.00	0.7799	0.4383	5.25	0.4610	0.4536
1.10	0.7638	0.5365	5.30	0.5078	0.4405
1.20	0.7154	0.6234	5.35	0.5490	0.4662
1.30	0.6386	0.6863	5.40	0.5573	0.5140
1.40	0.5431	0.7135	5.45	0.5269	0.5519
1.50	0.4453	0.6975	5.50	0.4784	0.5537
1.60	0.3655	0.6389	5.55	0.4456	0.5181
1.70	0.3238	0.5492	5.60	0.4517	0.4700
1.80	0.3336	0.4508	5.65	0.4926	0.4441
1.90	0.3944	0.3734	5.70	0.5385	0.4595
2.00	0.4882	0.3434	5.75	0.5551	0.5049
2.10	0.5815	0.3743	5.80	0.5298	0.5461
2.20	0.6363	0.4557	5.85	0.4819	0.5513
2.30	0.6266	0.5531	5.90	0.4486	0.5163
2.40	0.5550	0.6197	5.95	0.4566	0.4688
2.50	0.4574	0.6192	6.00	0.4995	0.4470
2.60	0.3890	0.5500	6.05	0.5424	0.4689
2.70	0.3925	0.4529	6.10	0.5495	0.5165
2.80	0.4675	0.3915	6.15	0.5146	0.5496
2.90	0.5624	0.4101	6.20	0.4676	0.5398
3.00	0.6058	0.4963	6.25	0.4493	0.4954
3.10	0.5616	0.5818	6.30	0.4760	0.4555
3.20	0.4664	0.5933	6.35	0.5240	0.4560
3.30	0.4058	0.5192	6.40	0.5496	0.4965
3.40	0.4385	0.4296	6.45	0.5292	0.5398
3.50	0.5326	0.4152	6.50	0.4816	0.5454
3.60	0.5880	0.4923	6.55	0.4520	0.5078
3.70	0.5420	0.5750	6.60	0.4690	0.4631
3.80	0.4481	0.5656	6.65	0.5161	0.4549
3.90	0.4223	0.4752	6.70	0.5467	0.4915
4.00	0.4984	0.4204	6.75	0.5302	0.5362
4.10	0.5738	0.4758	6.80	0.4831	0.5436
4.20	0.5418	0.5633	6.85	0.4539	0.5060
4.30	0.4494	0.5540	6.90	0.4732	0.4624
4.40	0.4383	0.4622	6.95	0.5207	0.4591

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

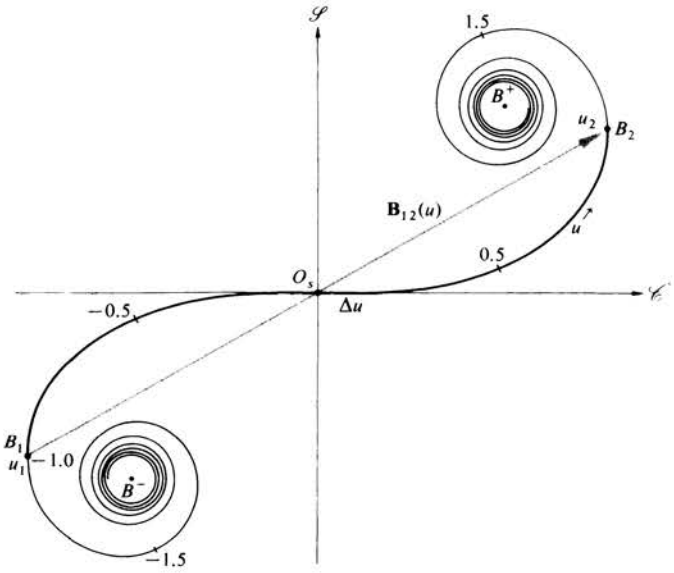


Figure 10.51 Cornu spiral.

and the problem is solved. Notice that the arc lengths along the spiral (i.e., $\Delta u = u_2 - u_1$ and $\Delta v = v_2 - v_1$) are proportional to the aperture's overall dimensions in the y- and z-direction, respectively. *The arc lengths are therefore constant, regardless of the position of P in the plane of observation.* On the other hand, the phasors $\tilde{\mathbf{B}}_{12}(u)$ and $\tilde{\mathbf{B}}_{12}(v)$, which span the arc lengths, are not constant, and they do depend on the location of P.

Maintaining the position of P opposite the center of the diffracting hole, now suppose that the aperture size is adjustable. As the square hole is gradually opened, Δv and Δu increase accordingly. The endpoints $\tilde{B}_1$ and $\tilde{B}_2$ of either of these arc lengths spiral around counterclockwise toward their limiting values of $\tilde{B}^-$ and $\tilde{B}^+$, respectively. The phasors $\tilde{\mathbf{B}}_{12}(u)$ and $\tilde{\mathbf{B}}_{12}(v)$, which are identical in this instance because of the symmetry, pass through a series of extrema. The central spot in the pattern therefore gradually shifts from relative brightness to darkness and back. All the while, the entire irradiance distribution varies continually from one beautifully intricate display to the next (see photo p. 499). For any particular aperture size, the off-center diffraction pattern can be computed by repositioning P. It is helpful to visualize the arc length as a piece of string, whose measure is equal to either Δv or Δu. Imagine it lying on the spiral, with O_s initially at its midpoint. As P is moved, for example, to the left along the y-axis (Fig. 10.49), y_1 and therefore u_1 both become less negative, and y_2 and u_2 increase positively. The result is that our Δu-string slides up the spiral. As the distance between the endpoints of the Δu-string changes, $|\tilde{\mathbf{B}}_{12}(u)|$ changes, and the irradiance [Eq. (10.106)] varies accordingly. When P is at the left edge of the geometric shadow, $y_1 = u_1 = 0$. As the point of observation moves into the geometric shadow, u_1 increases *positively*, and the Δu-string is now entirely on the upper half of the Cornu spiral. As u_1 and u_2 continue to increase, the string winds ever more tightly about the $\tilde{B}^+$-limit. Its ends, $\tilde{B}_1$ and $\tilde{B}_2$, become closer to each other, with the result that $|\tilde{\mathbf{B}}_{12}(u)|$ becomes quite small, and I_p decreases within the geometric shadow region. (We will come back to this point in more detail in the next section.) The same process applies when we scan in the z-direction; Δv is constant and $\tilde{\mathbf{B}}_{12}(v)$ varies.

If the aperture is completely opened out, revealing an unobstructed wave, $u_1 = v_1 = -\infty$, which means that $\tilde{B}_1(u) = \tilde{B}_1(v) = \tilde{B}^-$ and $\tilde{B}_2(u) = \tilde{B}_2(v) = \tilde{B}^+$. The $\tilde{B}^-\tilde{B}^+$-line makes a 45° angle with the $\mathscr{C}$-axis and has a length equal to $\sqrt{2}$. Consequently, the phasors $\tilde{\mathbf{B}}_{12}(u)$ and $\tilde{\mathbf{B}}_{12}(v)$ each have magnitude $\sqrt{2}$ and phase $\pi/4$, that is, $\tilde{\mathbf{B}}_{12}(u) = \sqrt{2}\exp(i\pi/4)$ and $\tilde{\mathbf{B}}_{12}(v) = \sqrt{2}\exp(i\pi/4)$. It follows from Eq. (10.102) that

$$\tilde{E}_p = \tilde{E}_u e^{i\pi/2} \tag{10.107}$$

and as in Section 10.3.1, we have the unobstructed amplitude, except for a $\pi/2$ phase discrepancy.* Finally, using (10.106), $I_p = I_0$.

We can construct a more palpable picture of what the Cornu spiral represents by considering Fig. 10.52, which depicts a cylindrical wavefront propagating from a coherent line source. The present procedure is exactly the same as that used in deriving the vibration curve, and the reader is referred back to Section 10.3.2 for a more leisurely discussion. Suffice it to say that the wavefront is divided into *half-period strip zones* by its intersection with a family of cylinders having a common axis and radii of $r_0 + \lambda/2$, $r_0 + \lambda$, $r_0 + 3\lambda/2$, and so on. *The contributions from these strip zones are proportional to their areas, which decrease rapidly.* This is in contrast to the circular zones, whose radii increase, thereby keeping the areas

*The phase discrepancy will be resolved by the Kirchhoff theory in Section 10.4.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

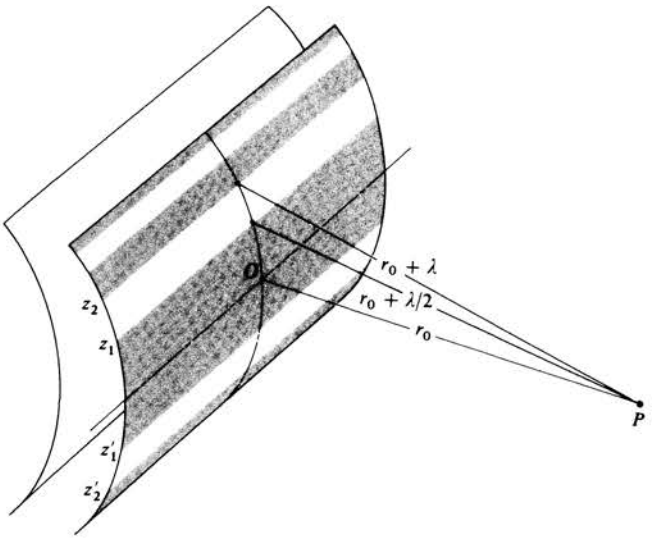


Figure 10.52 Cylindrical wavefront zones.

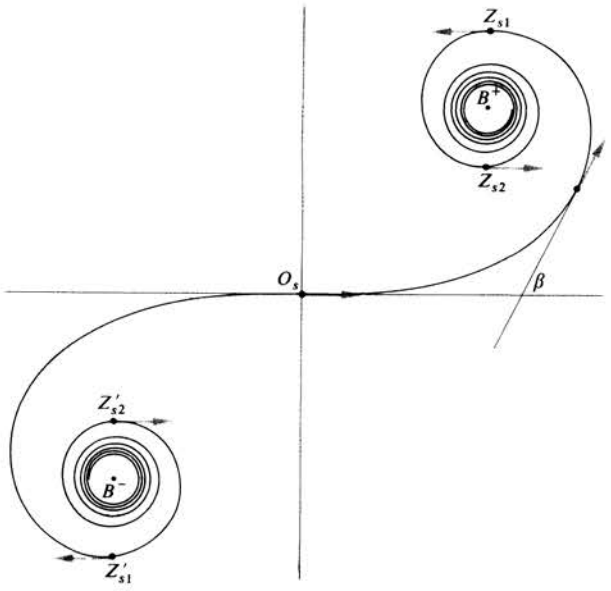


Figure 10.53 Cornu spiral related to the cylindrical wavefront.

nearly constant. Each strip zone is similarly divided into N subzones, which have a relative phase difference of π/N. The vector sum of all the amplitude contributions from zones above the center line is a spiraling polygon. If N goes to ∞ and the contributions generated by the strip zones below the center line are included, the polygon smooths out into a continuous Cornu spiral. This is not surprising, since the coherent line source generates an infinite number of overlapping point-source patterns.

Figure 10.53 shows a number of unit tangent vectors at various positions along the spiral. The vector at O_s corresponds to the contribution from the central axis passing through O on the wavefront. The points associated with the boundaries of each strip zone can be located on the spiral, since at those positions the relative phase, β, is either an even or odd multiple of π. For example, the point Z_{s1} on the spiral (Fig. 10.53), which is related to z_1 (Fig. 10.52) on the wavefront, is by definition 180° out-of-phase with O_s. Therefore Z_{s1} must be located at the top of the spiral, where $w = \sqrt{2}$ inasmuch as there $\beta = \pi w^2/2 = \pi$.

It will be helpful as we go along in the treatment to visualize the blocking out of these strip zones when analyzing the effects of obstructions. Obviously, one could even make an appropriate zone plate, which would accomplish this to some advantage, and such devices are in use.

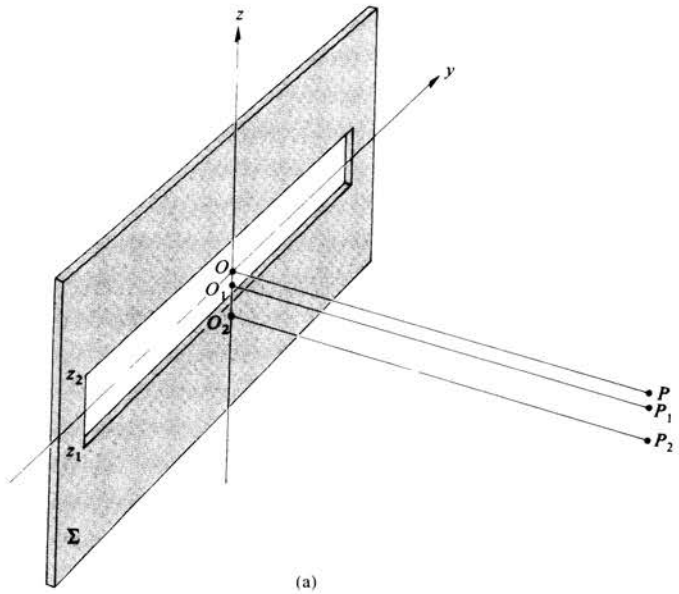
10.3.8 Fresnel Diffraction by a Slit

We can treat Fresnel diffraction at a long slit as an extension of the rectangular-aperture problem. We need only elongate the rectangle by allowing y_1 and y_2 to move very far from O, as shown in Fig. 10.54. As the point of observation moves along the y-axis, as long as the vertical boundaries at either end of the slit are still essentially at infinity, $u_2 \approx \infty$, $u_1 \approx -\infty$, and $\tilde{\mathbf{B}}_{12}(u) \approx \sqrt{2}e^{i\pi/4}$. From Eq. (10.106), for either point-source or plane-wave illumination,

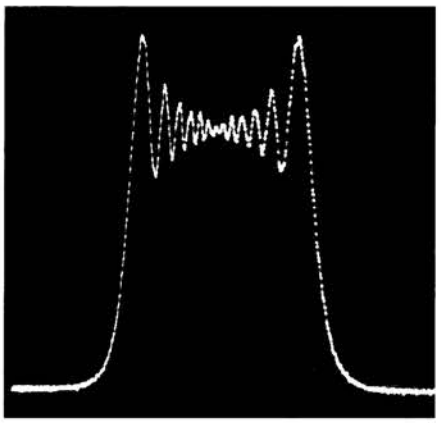
$$I_p = \frac{I_0}{2}|\tilde{\mathbf{B}}_{12}(v)|^2 \tag{10.108}$$

and the pattern is independent of y. The values of z_1 and z_2, which fix the slit width, determine the important parameter $\Delta v = v_2 - v_1$, which in turn governs $\tilde{\mathbf{B}}_{12}(v)$. Imagine once again that we have a string of length Δv lying along the spiral. At P, opposite point O, the aperture is symmetrical, and the string is centered on O_s (Fig. 10.55). The chord $|\tilde{\mathbf{B}}_{12}(v)|$ need only be measured and substituted into Eq. (10.108) to find I_p. At point P_1, z_1 and therefore v_1 are smaller negative numbers, whereas z_2 and v_2 have increased positively. The

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027



(a)



(b)

Figure 10.54 (*a*) Single-slit geometry. (*b*) A typical near-field irradiance distribution fairly close to a wide slit. The aperture was illuminated by a He–Ne laser and the pattern detected via a photodiode. Here the horizontal is parallel to the *z*-axis in the diagram. (Photo courtesy W. Klein, I. Physikalisches Institut, Germany.)

arc length Δv (the string) moves up the spiral (Fig. 10.55), and the chord decreases. As the point of observation moves down into the geometric shadow, the string winds about B^+, and the chord goes through a series of relative extrema. If Δv is very small, our imaginary piece of string is small, and the chord $|\tilde{\mathbf{B}}_{12}(v)|$ decreases appreciably only when the radius of curvature of the spiral itself is small. This occurs in the vicinity of $\tilde{B}^+$ or $\tilde{B}^-$, that is, far out into the geometric shadow. There will therefore be light well beyond the edges of the aperture, as long as the aperture is relatively small. Note too that with small Δv there will be a broad central maximum. In fact, if Δv is much less than 1, $r_0\lambda$ is much greater than the aperture width, and the Fraunhofer condition prevails. This transition of Eq. (10.108) into the form of Eq. (10.17) is more plausible when we realize that for large w the Fresnel integrals have trigonometric representations (see Problem 10.46).

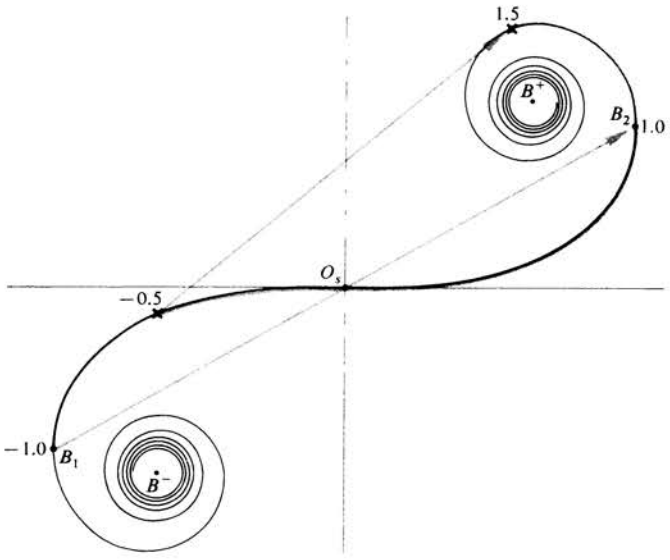


Figure 10.55 Cornu spiral for the slit.

As the slit widens, Δv becomes larger, for a fixed r_0, until a configuration like that in Fig. 10.56 exists for a point opposite the slit's center. If the point of observation is moved vertically either up or down, Δv slides either down or up the spiral. Yet the chord increases in both cases, so that the center of the diffraction pattern must be a relative minimum. Fringes now appear within the geometric image of the slit, unlike the Fraunhofer pattern.

Figure 10.57 shows two curves of $|\tilde{\mathbf{B}}_{12}(w)|^2$ plotted against $(w_1 + w_2)/2$, which is the center point of the arc length Δw. (Recall that the symbol w stands for either u or v.) A family of such curves running the range in Δw from about 1 to 10 would

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

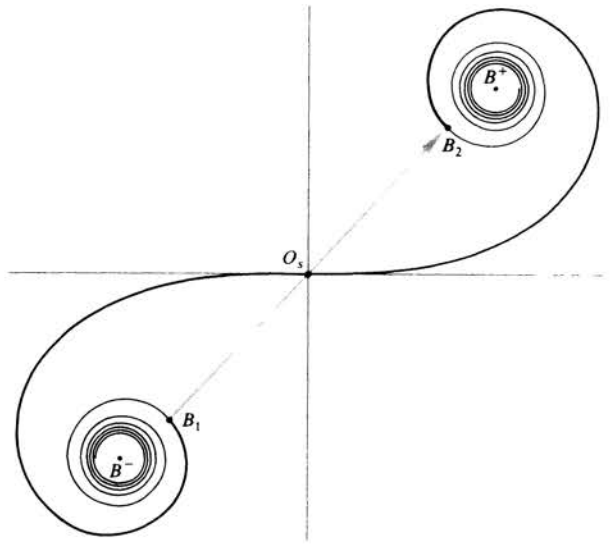


Figure 10.56 An irradiance minimum in the slit pattern.

cover the region of interest. The curves are computed by first choosing a particular Δw and then reading the appropriate $|\tilde{\mathbf{B}}_{12}(w)|$ values off the Cornu spiral as Δw slides along it. For a long slit

$$I_p = \frac{I_0}{2}|\tilde{\mathbf{B}}_{12}(v)|^2 \qquad [10.108]$$

and since Δz is the slit width that corresponds to Δv, each curve in Fig. 10.57 *is proportional to the irradiance distribution* for a given slit. For example, Fig. 10.57*a* can be read as $|\tilde{\mathbf{B}}_{12}(v)|^2$ versus $(v_1 + v_2)/2$ for $\Delta v = 2.5$. The abscissa relates to $(z_1 + z_2)/2$, that is, the displacement of the point of observation from the center of the slit. In Fig. 10.57*b* $\Delta w = 3.5$, which means that a slit having a $\Delta v = 3.5$ clearly has fringes appearing within the geometric image as expected (Problem 10.45). The curves could, of course, be plotted in terms of values of Δz or Δy explicitly, but that would unnecessarily limit them to one set of configuration parameters ρ_0, r_0, and λ.

As the slit is widened still further, Δv approaches and then surpasses 10. An increasing number of fringes appear within the geometric image, and the pattern no longer extends appreciably beyond that image.

The same kind of reasoning applies equally well to the analysis of the rectangular aperture, where use can also be made of the curves in Fig. 10.57.

To observe Fresnel slit diffraction, form a long narrow space between two fingers held at arm's length. Make a similar parallel slit close to your eye, using your other hand. With a *bright* source, such as the daytime sky or a large lamp, illuminating the far slit, observe it through the nearby aperture. After inserting the near slit, the far slit will appear to widen, and rows of fringes will be evident.

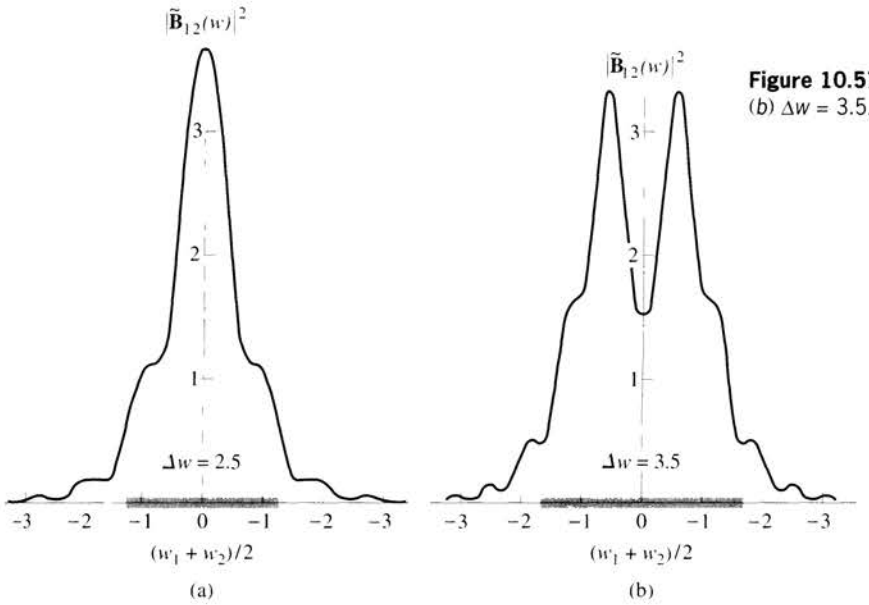


Figure 10.57 $|\tilde{\mathbf{B}}_{12}(w)|^2$ versus $(w_1 + w_2)/2$ for (*a*) $\Delta w = 2.5$ and (*b*) $\Delta w = 3.5$.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

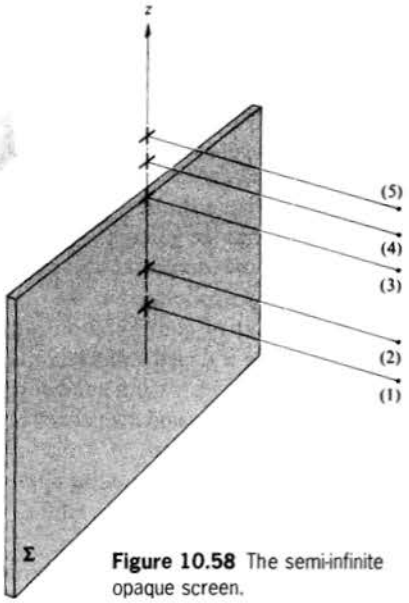


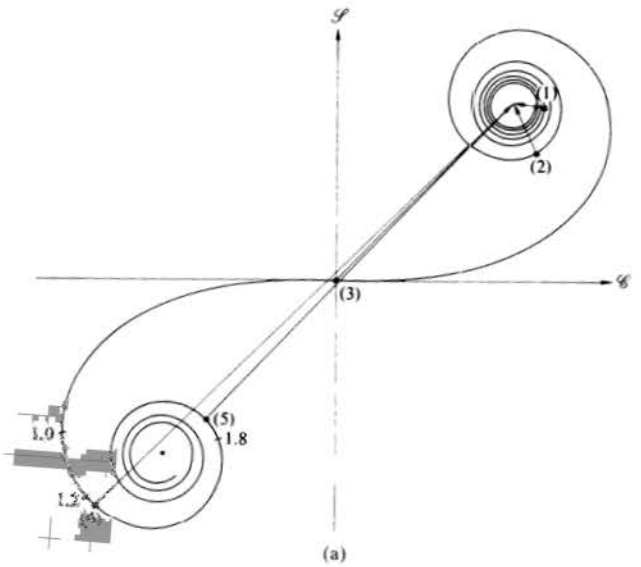
Figure 10.58 The semi-infinite opaque screen.

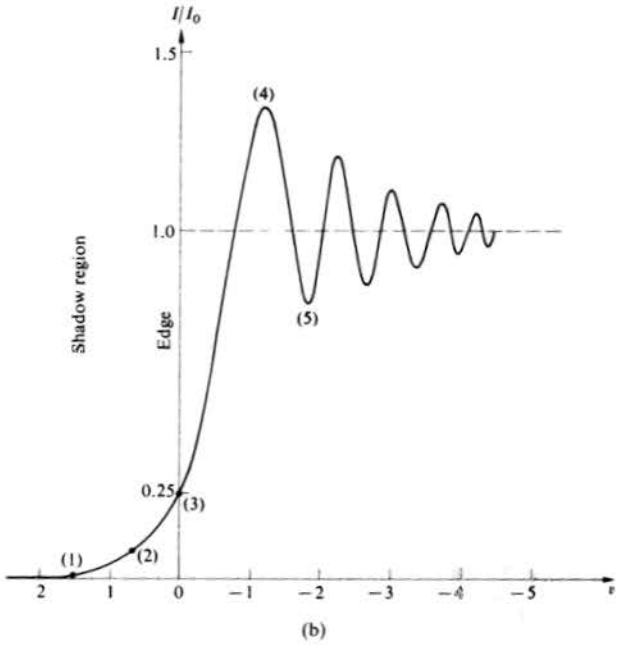
10.3.9 The Semi-Infinite Opaque Screen

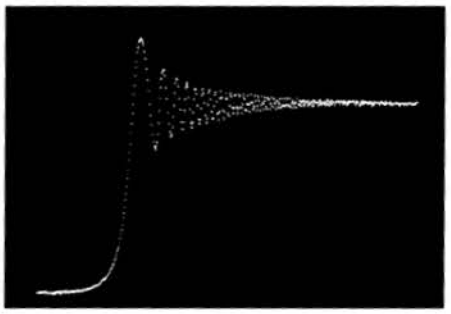
We now form a semi-infinite planar opaque screen by removing the upper half of Σ in Fig. 10.54. This is done simply enough, by letting $z_2 = y_1 = y_2 = \infty$. Remembering the original approximations, we limit the geometry so that the point of observation is close to the screen's edge. Since $v_2 = u_2 = \infty$ and $u_1 = -\infty$, Eq. (10.104) or (10.108) leads to

$$I_p = \frac{I_0}{2}\{[\tfrac{1}{2} - \mathscr{C}(v_1)]^2 + [\tfrac{1}{2} - \mathscr{S}(v_1)]^2\} \quad (10.109)$$

When the point P is directly opposite the edge, $v_1 = 0$, $\mathscr{C}(0) = \mathscr{S}(0) = 0$, and $I_p = I_0/4$. This was to be expected, since half the wavefront is obstructed, the amplitude of the disturbance is halved, and the irradiance drops to one quarter. This occurs at point (3) in Figs. 10.58 and 10.59. Moving into the geometric shadow region to point (2) and then on to (1) and still further, the successive chords clearly decrease monotonically (Problem 10.46). No irradiance oscillations exist



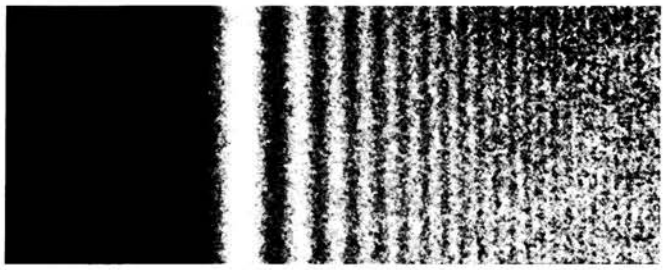




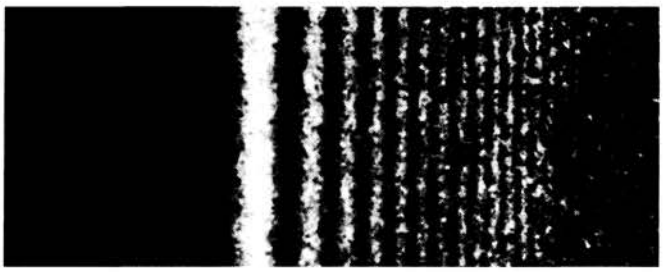
(c)

Figure 10.59 (*a*) The Cornu spiral for a semi-infinite screen. (*b*) The corresponding calculated irradiance distribution. (*c*) The same irradiance pattern under He–Ne laser illumination measured with a photodiode. (Photo courtesy W. Klein, I. Physikalisches Institut, Koln, Germany.)

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027



(a)



(b)

(a) The fringe pattern for a half-screen formed with light. (Francis Weston Sears, *Optics*, ©1949, Addison-Wesley Reading, MA.) (b) Fresnel electron diffraction at a half plane (MgO crystal)–electrons behave like photons. (Photo from *Handbuch der Physik*, edited by S. Flugge, Springer-Verlag, Heidelberg.)

within that region; the irradiance merely drops off rapidly. At any point above (3) the screen's edge will be below it; in other words, $z_1 < 0$ and $v_1 < 0$. At about $v_1 = -1.2$ the chord reaches a maximum, and the irradiance is a maximum. Thereafter, I_p oscillates about I_0, gradually diminishing in magnitude. With sensitive electronic techniques, many hundreds of these fringes can be observed.*

It is evident that the diffraction pattern in the accompanying photo would appear in the vicinity of the edges of a wide *slit* (Δv greater than about 10) as a limiting case. The irradiance distribution suggested by geometrical optics is obtained only when λ goes to zero. Indeed, as λ decreases, the fringes move closer to the edge and become increasingly fine in extent.

The straight-edge pattern can be observed using any kind of slit, held up in front of a broad lamp at arm's length, as a source. Introduce an opaque obstruction (e.g., a blackened microscope slide or a razor blade) very near your eye. As the edge of the obstruction passes in front of the source slit parallel to it, a series of fringes will appear.

*J. D. Barnett and F. S. Harris, Jr., *J. Opt. Soc. Amer.* **52**, 637 (1962).

10.3.10 Diffraction by a Narrow Obstacle

Refer back to the description of the single narrow slit; consider the complementary case in which the slit is opaque, and the screen transparent. Let's envision, for example, a vertical opaque wire. At a point directly opposite the wire's center, there will be two separate contributing regions extending from y_1 to $-\infty$ and from y_2 to $+\infty$. On the Cornu spiral, these correspond to two arc lengths from u_1 to B^- and from u_2 to B^+. The amplitude of the disturbance at a point P on the plane of observation is the magnitude of the *vector* sum of the two phasors $\overrightarrow{B^-u_1}$ and $\overrightarrow{u_2B^+}$, illustrated in Fig. 10.60. As with the opaque disk, the symmetry is such that there will always be an illuminated region along the central axis. This can be seen from the spiral, since when P is on the central axis, $\overrightarrow{B^-u_1} = \overrightarrow{u_2B^+}$ and their sum can never be zero. The arc length Δu represents the obscured region of the spiral, which increases as the diameter of the wire increases. For thick wires, u_1 approaches B^-, u_2 approaches B^+, the phasors decrease in length, and the irradiance on the shadow's axis drops off. This is evident from the accompanying photos, which show the patterns actually cast by a thin piece of lead from a mechanical pencil and by a rod with a $\frac{1}{8}$-inch diameter. Imagine that we have a small irradiance sensor at point P on the plane of observation (or the film plate). As P moves off the central axis to the right, y_1 and u_1 increase negatively, whereas y_2 and u_2, which are positive, decrease. The opaque region, Δu, slides down the spiral. When the sensor is at the right edge of the geometric shadow $y_2 = 0$, $u_2 = 0$; in other words, u_2 is at O_s. Notice that if the wire is thin, that is,

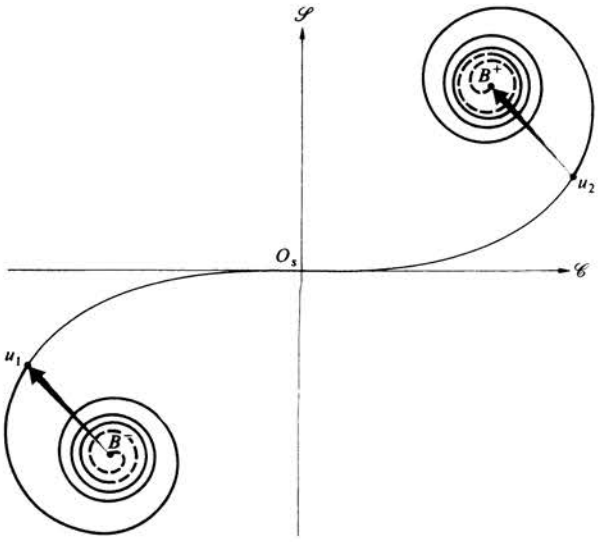
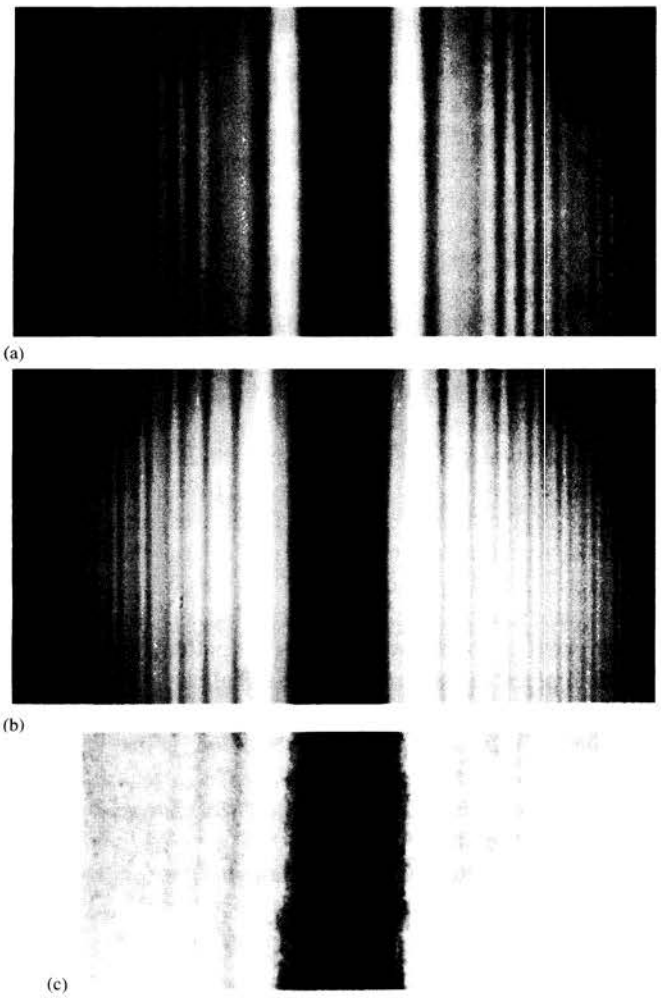


Figure 10.60 The Cornu spiral as applied to a narrow obstacle.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027



(*a*) The shadow pattern cast by the lead from a mechanical pencil. (*b*) The pattern cast by a 1/8-inch diameter rod. (Photos by E. H.) (c) Matter-wave diffraction. Fresnel electron-diffraction pattern of a 2-μm diameter metallized quartz filament. [Photo from O. E. Klemperer, *Electron Physics*, Butterworths, London (1972), pgs 188–191.]

if Δu is small, the sensor will record a gradual decrease in irradiance as u_2 approaches O_s. On the other hand, if the wire is thick, Δu is large and u_1 and u_2 are large. As Δu slides down the spiral, the two phasors revolve through a number of complete rotations, going in- and out-of-phase in the process. The resulting additional extrema appearing within the geometric shadow are evident in the middle photo. In fact, the separation between internal fringes varies inversely with the width of the rod, just as if the pattern arose from the interference of two waves (Young's Experiment) reflected at the rod's edges.

10.3.11 Babinet's Principle

Two diffracting screens are said to be *complementary* when the transparent regions on one exactly correspond to the opaque regions on the other and vice versa. When two such screens are overlapped, the combination is obviously completely opaque. Now then, let E_1 or E_2 be the scalar optical disturbance arriving at P when either complementary screen Σ_1 or Σ_2, respectively, is in place. The total contribution from each aperture is determined by integrating over the area bounded by that aperture. If both *apertures* are present at once, there are no opaque regions at all; the limits of integration go to infinity, and we have the unobstructed disturbance E_0, whereupon

$$E_1 + E_2 = E_0 \tag{10.110}$$

which is the statement of **Babinet's Principle**. Take a close look at Figs. 10.56 and 10.60, which depict the Cornu spiral configurations for a transparent slit and a narrow opaque obstacle. If the two arrangements are made complementary, Fig. 10.61 illustrates Babinet's Principle quite clearly. The phasor arising from a narrow obstacle ($\overrightarrow{B^-B_1}+\overrightarrow{B_2B^+}$) added to that from a slit $\overrightarrow{B_2B_1}$ yields the unobstructed phase $\overrightarrow{B^-B^+}$.

The principle implies that when $E_0 = 0$, $E_1 = -E_2$; in other words, these disturbances are precisely equal in magnitude and 180° out-of-phase. One would therefore observe exactly the same irradiance distribution with either Σ_1 or Σ_2 in place, an interesting result indeed. It is evident, however, that the principle cannot be exactly true, since for an unobstructed wave from a point source, there are no zero-amplitude points

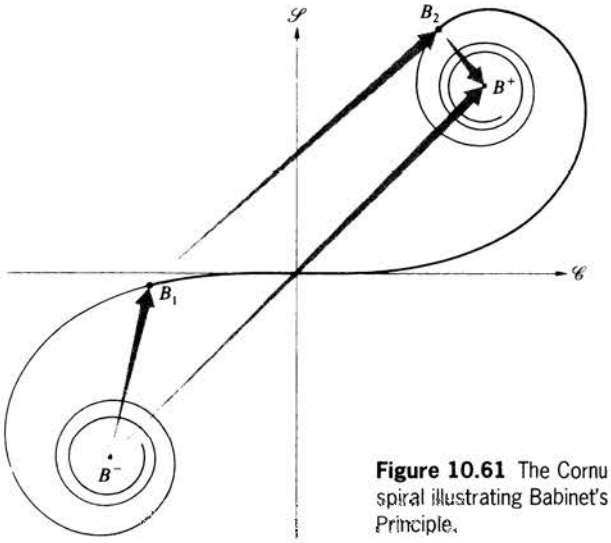


Figure 10.61 The Cornu spiral illustrating Babinet's Principle.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

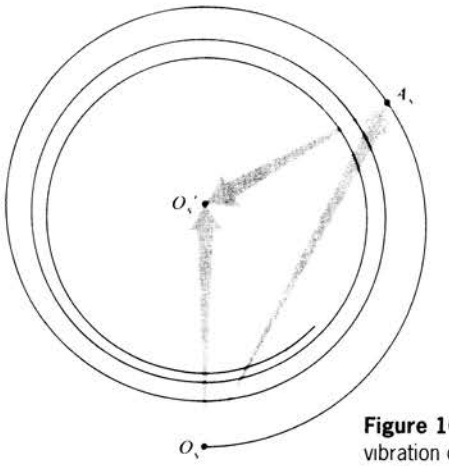


Figure 10.62 The vibration curve illustrating Babinet's Principle.

(i.e., $E_0 \neq 0$ everywhere). Yet if the source is imaged at P_0 by perfect lenses, as in Fig. 10.6 (with neither Σ_1 nor Σ_2 present), there will be a large, essentially zero-amplitude region beyond the immediate vicinity of P_0 (beyond the Airy disk) in which $E_1 + E_2 = E_0 = 0$. It is therefore only for the case of Fraunhofer diffraction that complementary screens will generate equivalent irradiance distributions, that is, $E_1 = -E_2$ (excluding point P_0). Nonetheless, Eq. (10.110) is valid in Fresnel diffraction, even though the irradiances obey no simple relationship. This is exemplified by the slit and narrow obstacle of Fig. 10.61. For a circular hole and disk examine Fig. 10.62. Equation (10.110) is again clearly applicable, even though the diffraction patterns are certainly not equivalent.

The real beauty of Babinet's Principle is most evident when applied to Fraunhofer diffraction, as shown in Fig. 10.63, where the patterns from complementary screens are almost identical.

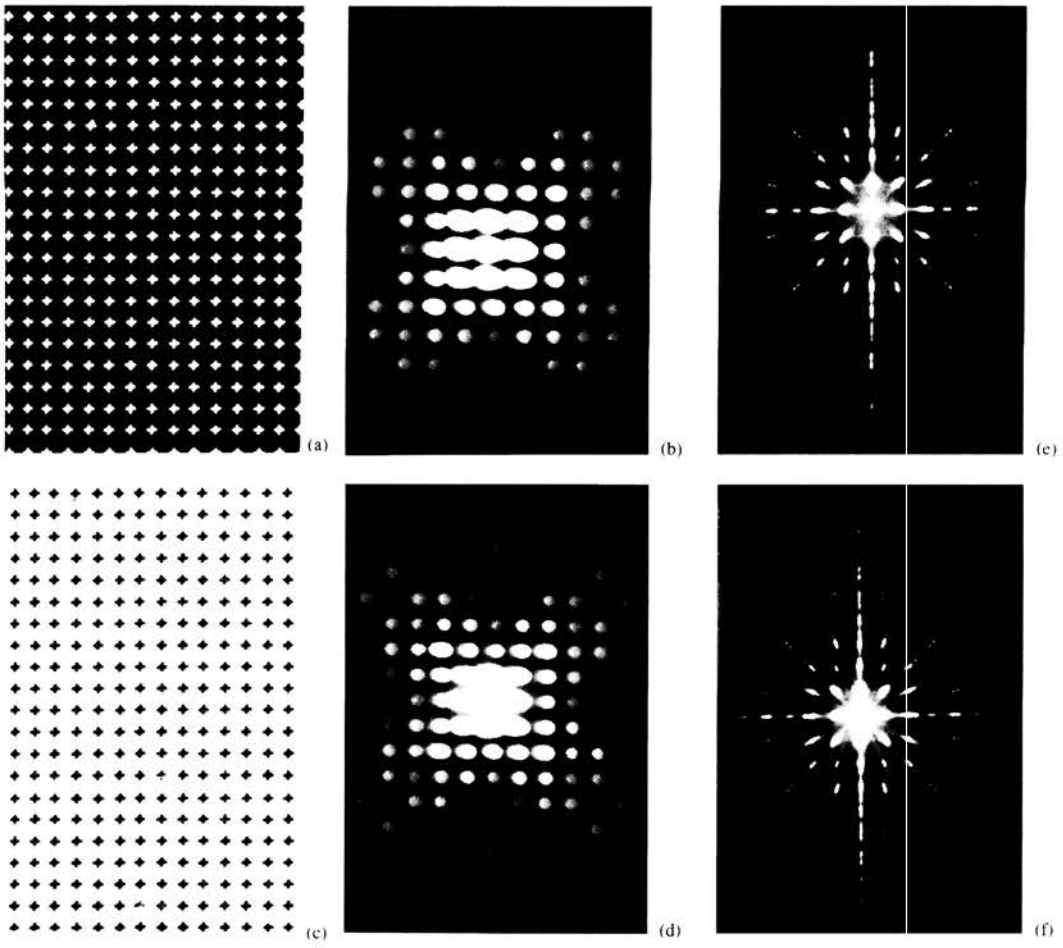


Figure 10.63 (*a*)–(*d*) White-light diffraction patterns for regular arrays of apertures and complementary obstacles in the form of rounded plus signs. (*e*) and (*f*) Diffraction patterns for a regular array of rectangular apertures and obstacles, respectively. (Photos courtesy The Ealing Corporation and Richard B. Hoover.)

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

10.4 Kirchhoff's Scalar Diffraction Theory

We have described a number of diffracting configurations, quite satisfactorily, within the context of the relatively simple Huygens–Fresnel theory. Yet the whole imagery of surfaces covered with fictitious point sources, which was the basis of that analysis, was merely postulated rather than derived from fundamental principles. The Kirchhoff treatment shows that these results are actually derivable from the *scalar* differential wave equation.

The discussion to follow is rather formal and involved. Portions of it have therefore been relegated to an appendix, where we can indulge in succinctness and risk sacrificing readability for rigor.

In the past, when dealing with a distribution of monochromatic point sources, we computed the resultant optical disturbance at point P (i.e., E_p) by carrying out a superposition of the individual waves. There is, however, a completely different approach, which is founded in potential theory. Here one is concerned not with the sources themselves but rather with the scalar optical disturbance and its derivatives over an arbitrary closed surface surrounding P. We assume that a Fourier analysis can separate the constituent frequencies, so that we need only deal with one such frequency at a time. The monochromatic optical disturbance E is a solution of the differential wave equation

$$\nabla^2 E = \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2} \tag{10.111}$$

Without specifying the precise spatial nature of the wave, we can write it as

$$\tilde{E} = \tilde{\mathscr{E}} e^{-ikct} \tag{10.112}$$

Here $\tilde{\mathscr{E}}$ represents the complex space part of the disturbance. Substituting this into the wave equation, we obtain

$$\nabla^2 \tilde{\mathscr{E}} + k^2 \tilde{\mathscr{E}} = 0 \tag{10.113}$$

This is known as the *Helmholtz Equation* and is solved, with the aid of Green's Theorem, in Appendix 2. The optical disturbance existing at a point P, expressed in terms of the optical disturbance and its gradient evaluated on an arbitrary closed surface S, enclosing P, is

$$\tilde{\mathscr{E}}_p = \frac{1}{4\pi}\left[\oiint_S \frac{e^{ikr}}{r}\nabla\tilde{\mathscr{E}}\cdot d\vec{\mathbf{S}} - \oiint_S \tilde{\mathscr{E}}\nabla\left(\frac{e^{ikr}}{r}\right)\cdot d\vec{\mathbf{S}}\right] \tag{10.114}$$

Known as the *Kirchhoff Integral Theorem*, Eq. (10.114) relates to the geometric configuration illustrated in Fig. 10.64.

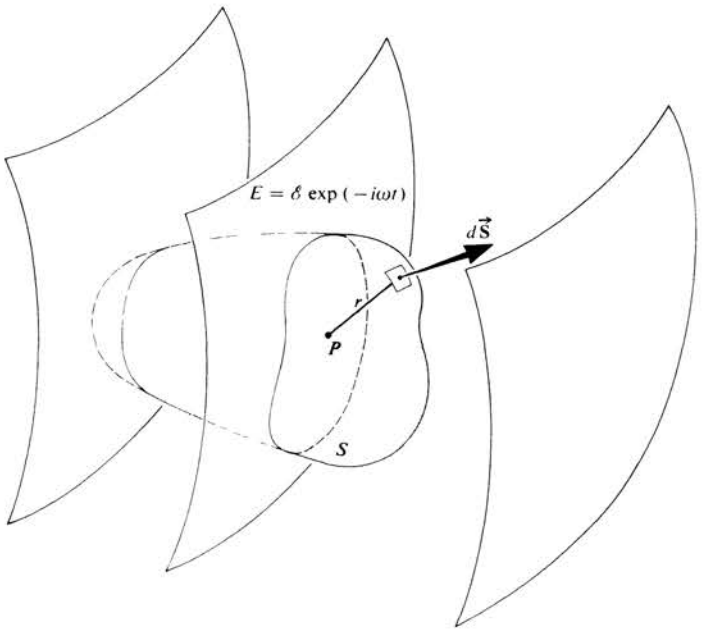


Figure 10.64 An arbitrary closed surface S enclosing point P.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

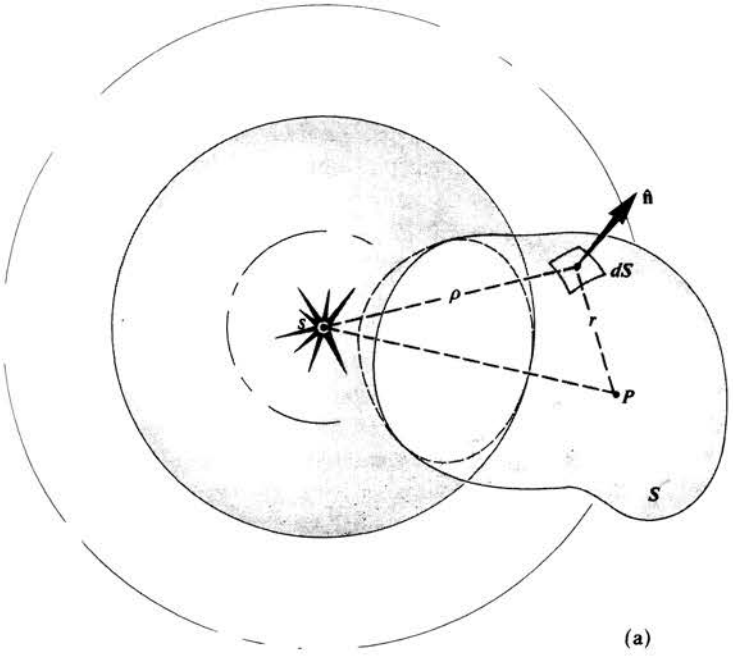
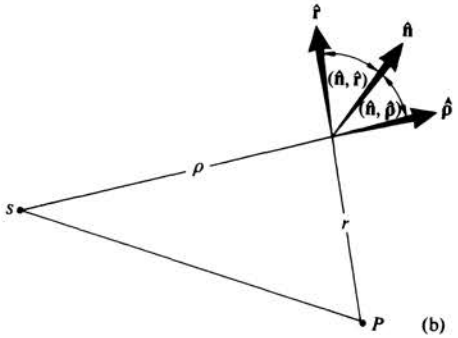


Figure 10.65 A spherical wave emitted from point *s*.



We now apply the theorem to the specific instance of an unobstructed spherical wave originating at a point source s, as shown in Fig. 10.65. The disturbance has the form

$$\tilde{E}(\rho, t) = \frac{\mathcal{E}_0}{\rho} e^{i(k\rho - \omega t)} \tag{10.115}$$

in which case

$$\tilde{\mathscr{E}}(\rho) = \frac{\mathcal{E}_0}{\rho} e^{ik\rho} \tag{10.116}$$

If we substitute this into Eq. (10.114), it becomes

$$\tilde{\mathscr{E}}_\rho = \frac{1}{4\pi}\left[\iint_S \frac{e^{ikr}}{r} \frac{\partial}{\partial \rho}\left(\frac{\mathcal{E}_0}{\rho} e^{ik\rho}\right) \cos(\hat{\mathbf{n}}, \hat{\boldsymbol{\rho}})\, dS - \iint_S \frac{\mathcal{E}_0}{\rho} e^{ik\rho} \frac{\partial}{\partial r}\left(\frac{e^{ikr}}{r}\right) \cos(\hat{\mathbf{n}}, \hat{\mathbf{r}})\, dS \right]$$

where $d\vec{\mathbf{S}} = \hat{\mathbf{n}}\, dS$, $\hat{\mathbf{n}}$, $\hat{\mathbf{r}}$, and $\hat{\boldsymbol{\rho}}$ are unit vectors,

$$\nabla\left(\frac{e^{ikr}}{r}\right) = \hat{\mathbf{r}} \frac{\partial}{\partial r}\left(\frac{e^{ikr}}{r}\right)$$

and

$$\nabla \tilde{\mathscr{E}}(\rho) = \hat{\boldsymbol{\rho}} \partial \tilde{\mathscr{E}} / \partial \rho$$

The differentiations under the integral signs are

$$\frac{\partial}{\partial \rho}\left(\frac{e^{ik\rho}}{\rho}\right) = e^{ik\rho}\left(\frac{ik}{\rho} - \frac{1}{\rho^2}\right)$$

and

$$\frac{\partial}{\partial r}\left(\frac{e^{ikr}}{r}\right) = e^{ikr}\left(\frac{ik}{r} - \frac{1}{r^2}\right)$$

When $\rho >> \lambda$ and $r >> \lambda$, the $1/\rho^2$ and $1/r^2$ terms can be neglected. This approximation is fine in the optical spectrum but certainly need not be true for microwaves. Proceeding, we write

$$\tilde{\mathscr{E}}_p = -\frac{\mathcal{E}_0 i}{\lambda} \iint_S \frac{e^{ik(\rho + r)}}{\rho r}\left[\frac{\cos(\hat{\mathbf{n}}, \hat{\mathbf{r}}) - \cos(\hat{\mathbf{n}}, \hat{\boldsymbol{\rho}})}{2}\right] dS \tag{10.117}$$

This is the ***Fresnel–Kirchhoff diffraction formula***.

Take a long look at Eq. (10.96), which represents the disturbance at P arising from an element dS in the Huygens–Fresnel theory, and compare it with Eq. (10.117). In Eq. (10.117) the angular dependence is contained in the single term $\frac{1}{2}[\cos(\hat{\mathbf{n}}, \hat{\mathbf{r}}) - \cos(\hat{\mathbf{n}}, \hat{\boldsymbol{\rho}})]$, which we shall call the **obliquity factor** $K(\theta)$, showing it to be equivalent to Eq. (10.72) later on. Notice as well that k can be replaced by $-k$ every-

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

where, since we certainly could have chosen the phase of Eq. (10.115) to have been $(\omega t - k\rho)$. With Eq. (10.112) in mind, multiply both sides of Eq. (10.117) by $\exp(-i\omega t)$; the differential element is then

$$dE_p = \frac{K(\theta)\mathcal{E}_0}{\rho r \lambda} \cos[k(\rho + r) - \omega t - \pi/2]\, dS \qquad (10.118)$$

This is the contribution to E_p arising from an element of surface area dS a distance r from P. The $\pi/2$ term in the phase results from the fact that $-i = \exp(-i\pi/2)$. The Kirchhoff formulation therefore leads to the same total result, with the exception that it includes the correct $\pi/2$ phase shift, which is lacking in the Huygens–Fresnel treatment [Eq. (10.96)].

We have yet to ensure that the surface S can be made to correspond to the unobstructed portion of the wavefront, as it does in the Huygens–Fresnel theory. For the case of a freely propagating spherical wave emanating from the point source s, we construct the doubly connected region shown in Fig. 10.66. The surface S_2 completely surrounds the small spherical surface S_1. At $\rho = 0$ the disturbance $E(\rho, t)$ has a singularity and is therefore properly excluded from the volume V between S_1 and S_2. The integral must now include both surfaces S_1 and S_2. But we can have S_2 increase outward indefinitely by requiring its radius to go to infinity. In that case, the contribution to the surface integral vanishes. (This is true whatever the form of the incoming disturbance, as long as it drops off at least as rapidly as a spherical wave.) The remaining surface S_1 is a sphere centered at the point source. Since, over S_1, $\hat{\mathbf{n}}$ and $\hat{\boldsymbol{\rho}}$ are antiparallel, it is evident form Fig. 10.65*b* that the angles $(\hat{\mathbf{n}}, \hat{\mathbf{r}})$ and $(\hat{\mathbf{n}}, \hat{\boldsymbol{\rho}})$ are θ and 180°, respectively. The obliquity factor then becomes

$$K(\theta) = \frac{\cos\theta + 1}{2}$$

which is Eq. (10.72). Clearly, since the surface of integration S_1 is centered at s, it does indeed correspond to the spherical wavefront at some instant. ***The Huygens–Fresnel Principle is therefore directly traceable to the scalar differential wave equation.***

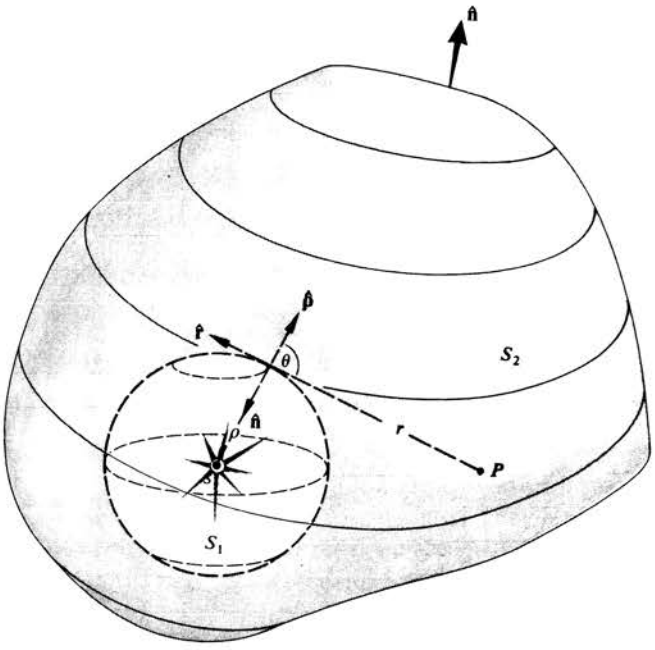


Figure 10.66 A doubly connected region surrounding point s.

We shall not pursue the Kirchhoff formulation any further, other than to point out briefly how it is applied to diffracting screens. The single closed surface of integration surrounding the point of observation P is generally taken to be the entire screen Σ capped by an infinite hemisphere. There are then three distinct areas with which to be concerned. The contribution to the integral from the region of the infinite hemisphere is zero. Moreover, it is assumed that there is no disturbance immediately behind the opaque screen, so that this second region contributes nothing. The disturbance at P is therefore determined solely by the contributions arising from the aperture, and one need only integrate Eq. (10.117) over that area.

The fine results obtained by using the Huygens–Fresnel Principle are now justified theoretically, the main limitations being that $\rho >> \lambda$ and $r >> \lambda$.

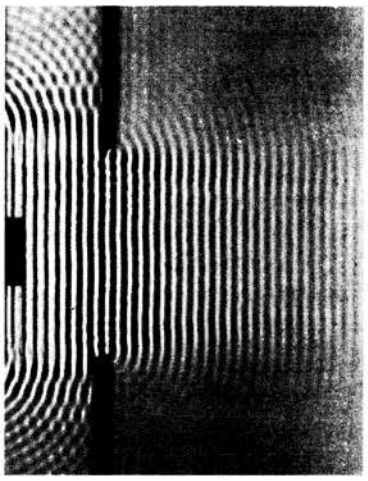
10.5 Boundary Diffraction Waves

In Section 10.1.1 we said that the diffracted wave could be envisioned as arising from a fictitious distribution of secondary emitters spread across the unobstructed portion of the wavefront, namely, the Huygens–Fresnel Principle. There is, however, another, completely different, and rather appealing possibility. Suppose that an incoming wave sets the electrons on the rear of the diffracting screen Σ into oscillation, and these in turn radiate. We anticipate a twofold effect. First, all the oscillators that are remote from the edge of the aperture radiate back toward the source in such a fashion as to cancel

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

the incoming wave at all points, except within the projection of the aperture itself. In other words, if this were the only contributing mechanism, a perfect geometrical image of the aperture would appear on the plane of observation. There is, however, an additional contribution arising from those oscillators in the vicinity of the aperture's edge. A portion of the energy radiated by these secondary sources propagates in the forward direction. The superposition of this scattered wave (known as the *boundary diffraction wave*) and the unobstructed portion of the primary wave (known as the *geometrical wave*) yield the diffraction pattern. A rather cogent reason for contemplating such a scheme becomes apparent when one examines the following arrangement. Tear a small hole ($\approx\frac{1}{2}$ cm in diameter) of arbitrary shape in a piece of paper, and holding it at arm's length, view an ordinary lightbulb some meters distant. Even with your eye in the shadow region, the edges of the aperture will be brightly illuminated. The accompanying ripple-tank photograph also illustrates the process. Notice how each edge of the slit seems to serve as a center for a circular disturbance, which then propagates beyond the aperture. There are no electron-oscillators here, which implies that these ideas have a certain generality, being applicable to elastic waves as well.

The formulation of diffraction in terms of the interference of a scattered edge wave and a geometrical wave is perhaps more physically appealing than the fictitious emitters of the Huygens–Fresnel Principle. It is not, however, a new concept. Indeed, it was first propounded by the ubiquitous Thomas Young even before Fresnel's celebrated memoir on diffraction. But in time Fresnel's brilliant successes unfortunately convinced Young to reject his own ideas, and he finally did so in a letter to Fresnel in 1818. Strengthened by Kirchhoff's work, the Fresnel conception of diffraction became generally accepted and has persisted (right up to Section 10.4). The resurrection of Young's theory began in 1888. At that time, Gian Antonio Maggi proved that Kirchhoff's analysis, for a point source at least, was equivalent to two contributing terms. One of these was a geometrical wave, but the other, unhappily, was an integral, which allowed no clear physical interpretation at the time.



Ripple-tank waves passing through a slit. (Photo courtesy PSSC *Physics*, D. C. Heath, Boston, 1960.)

In his doctoral thesis (1893), Eugen Maey showed that an edge wave could indeed be extracted from a modified Kirchhoff formulation for a semi-infinite half-plane. Arnold Sommerfeld's rigorous solution of the half-plane problem (see Section 10.1) showed that a cylindrical wave actually does proceed from the screen's edge. It propagates into both the geometrical shadow region and the illuminated region. In the latter, the boundary diffraction wave combines with the geometrical wave, in complete accord with Young's theory. In 1917 Adalbert (Wojciech) Rubinowicz was able to prove that Kirchhoff's formula for a plane or spherical wave can be appropriately decomposed into the two desired waves, thereby revealing the basic correctness of Young's ideas. He also later established that the boundary diffraction wave, to a first approximation, was generated by reflection of the primary wave from the aperture's edge. In 1923 Friedrich Kottler pointed out the equivalence of the solutions of Maggi and Rubinowicz, and one now speaks of the Young-Maggi-Rubinowicz theory. Most recently, Kenro Miyamoto and Emil Wolf (1962) have extended the boundary diffraction theory to the case of arbitrary incident waves.*

A very useful contemporary approach to the problem has been devised by Joseph B. Keller. He has developed a geometric theory of diffraction that is closely related to Young's edge wave picture. Along with the usual rays of geometrical optics, Keller hypothesizes the existence of diffracted rays. Rules governing these diffracted rays, which are analogous to the Laws of Reflection and Refraction, are employed to determine the resultant fields.

*A fairly complete bibliography can be found in the article by A. Rubinowicz in *Progress in Optics*, Vol. 4, p. 199.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

PROBLEMS

Complete solutions to all problems—except those with an asterisk—can be found in the back of the book.

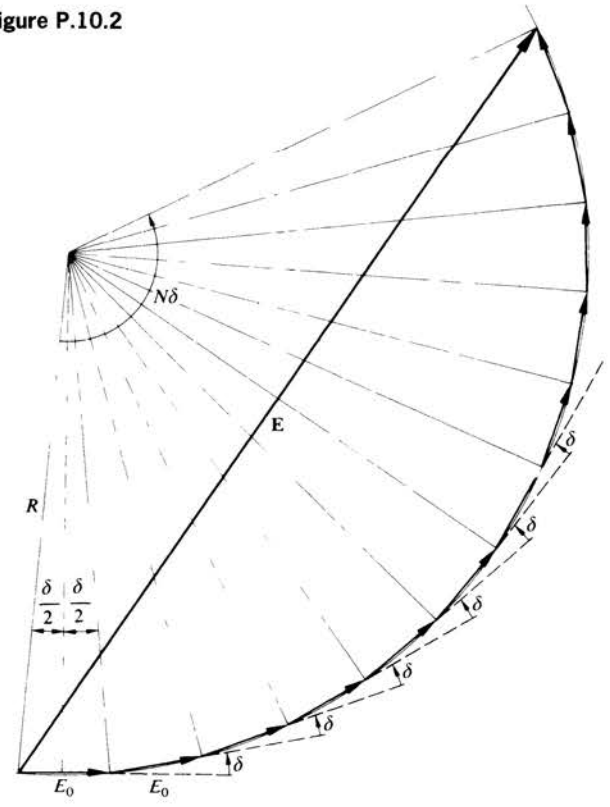
10.1 A point source S is a perpendicular distance R away from the center of a circular hole of radius a in an opaque screen. If the distance to the periphery is $(R + \ell)$, show that Fraunhofer diffraction will occur on a very distant screen when

$$\lambda R >> a^2/2$$

What is the smallest satisfactory value of R if the hole has a radius of 1 mm, $\ell \leq \lambda/10$, and $\lambda = 500$ nm?

10.2 Using Fig. P.10.2, derive the irradiance equation for N coherent oscillators, Eq. (10.5).

Figure P.10.2



10.3* In Section 10.1.3 we talked about introducing an intrinsic phase shift ε between oscillators in a linear array. With this in mind, show that Eq. (10.18) becomes

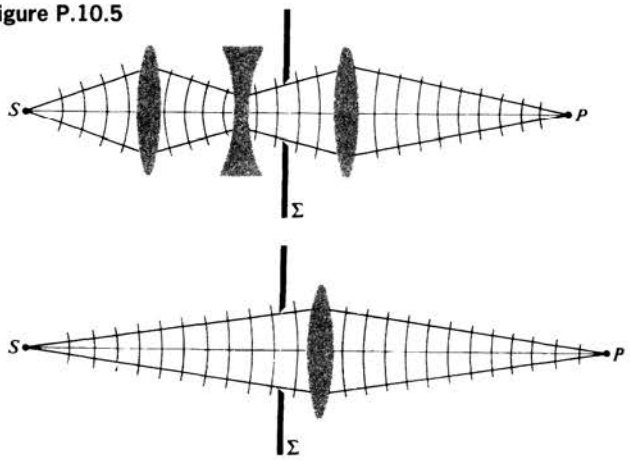
$$\beta = (kb/2)(\sin\theta - \sin\theta_i)$$

when the incident plane wave makes an angle θ_i with the plane of the slit.

10.4 Referring back to the multiple antenna system on p.451, compute the angular separation between successive lobes or principal maxima and the width of the central maximum.

10.5 Examine the setup of Fig. 10.3 in order to determine what is happening in the image space of the lenses; in other words, locate the exit pupil and relate it to the diffraction process. Show that the configurations in Fig. P.10.5 are equivalent to those of Fig. 10.3 and will therefore result in Fraunhofer diffraction. Design at least one more such arrangement.

Figure P.10.5



10.6 The angular distance between the center and the first minimum of a single-slit Fraunhofer diffraction pattern is called the *half-angular breadth*; write an expression for it. Find the corresponding *half-linear width* (a) when no focusing lens is present and the slit-viewing screen distance is L, and (b) when a lens of focal length f_2 is very close to the aperture. Notice that the half-linear width is also the distance between the successive minima.

10.7* A single slit in an opaque screen 0.10 mm wide is illuminated (in air) by plane waves from a krypton ion laser ($\lambda_0 = 461.9$ nm). If the observing screen is 1.0 m away, determine whether or not the resulting diffraction pattern will be of the far-field variety and then compute the angular width of the central maximum.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

10.8* A narrow single slit (in air) in an opaque screen is illuminated by infrared from a He–Ne laser at 1152.2 nm, and it is found that the center of the tenth dark band in the Fraunhofer pattern lies at an angle of 6.2° off the central axis. Please determine the width of the slit. At what angle will the tenth minimum appear if the entire arrangement is immersed in water ($n_w = 1.33$) rather than air ($n_a = 1.000\,29$)?

10.9 A collimated beam of microwaves impinges on a metal screen that contains a long horizontal slit that is 20 cm wide. A detector moving parallel to the screen in the far-field region locates the first minimum of irradiance at an angle of 36.87° above the central axis. Determine the wavelength of the radiation.

10.10 Show that for a double-slit Fraunhofer pattern, if $a = mb$, the number of bright fringes (or parts thereof) within the central diffraction maximum will be equal to $2m$.

10.11* Two long slits 0.10 mm wide, separated by 0.20 mm, in an opaque screen are illuminated by light with a wavelength of 500 nm. If the plane of observation is 2.5 m away, will the pattern correspond to Fraunhofer or Fresnel diffraction? How many Young's fringes will be seen within the central bright band?

10.12 What is the relative irradiance of the subsidiary maxima in a three-slit Fraunhofer diffraction pattern? Draw a graph of the irradiance distribution, when $a = 2b$, for two and then three slits.

10.13* Starting with the irradiance expression for a finite slit, shrink the slit down to a minuscule area element and show that it emits equally in all directions.

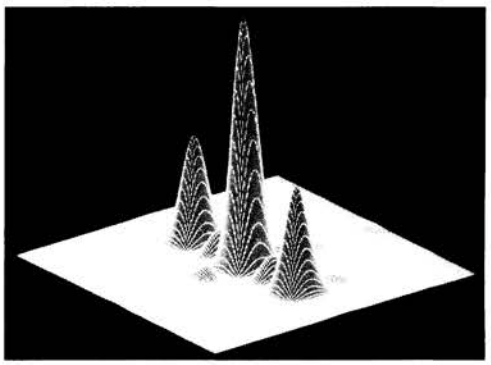
10.14* Show that Fraunhofer diffraction patterns have a center of symmetry [i.e., $I(Y, Z) = I(-Y, -Z)$], regardless of the configuration of the aperture, as long as there are no phase variations in the field over the region of the hole. Begin with Eq. (10.41). We'll see later (Chapter 11) that this restriction is equivalent to saying that the aperture function is real.

10.15 With the results of Problem 10.14 in mind, discuss the symmetries that would be evident in the Fraunhofer diffraction pattern of an aperture that is itself symmetrical about a line (assuming normally incident quasimonochromatic plane waves).

10.16 From symmetry considerations, create a rough sketch of the Fraunhofer diffraction patterns of an equilateral triangular aperture and an aperture in the form of a plus sign.

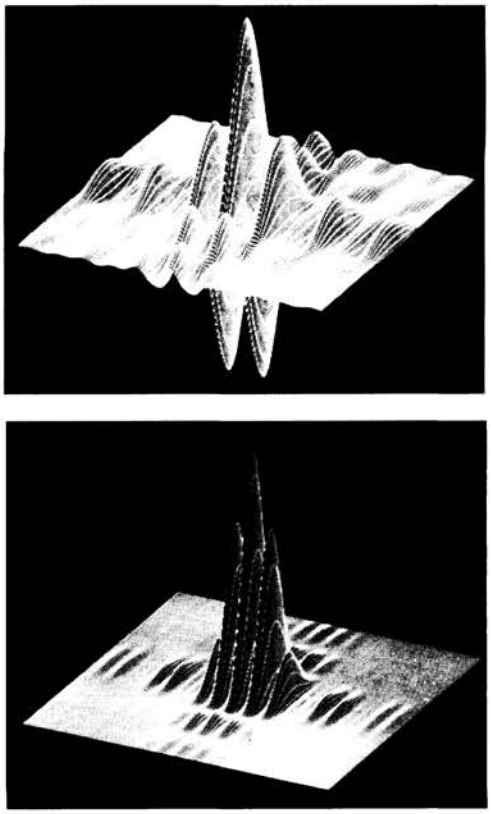
10.17 Figure P.10.17 is the irradiance distribution in the far field for a configuration of elongated rectangular apertures. Describe the arrangement of holes that would give rise to such a pattern and give your reasoning in detail.

Figure P.10.17 (Photo courtesy R. G. Wilson, Illinois Wesleyan University.)



10.18 In Fig. P.10.18*a* and *b* are the electric field and irradiance distributions, respectively, in the far field for a configuration of elongated rectangular apertures. Describe the arrangement of holes that would give rise to such patterns and discuss your reasoning.

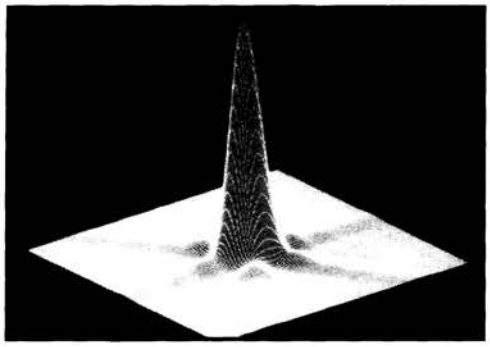
Figure P.10.18 (Photo courtesy R. G. Wilson, Illinois Wesleyan University.)



Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

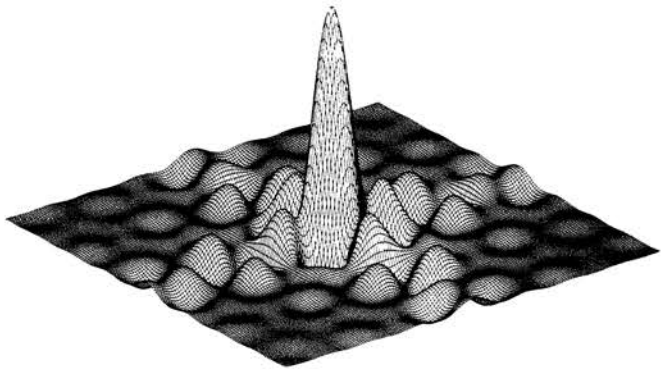
10.19 Figure P.10.19 is a computer-generated Fraunhofer irradiance distribution. Describe the aperture that would give rise to such a pattern and give your reasoning in detail.

Figure P.10.19 (Photo courtesy R. G. Wilson, Illinois Wesleyan University.)



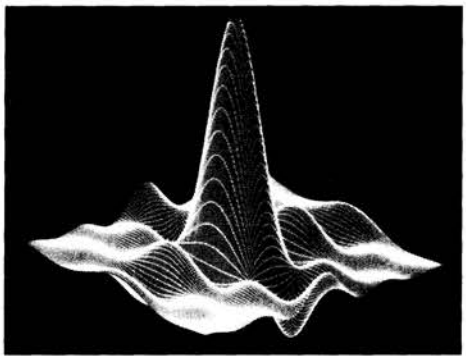
10.20 Figure P.10.20 is the electric-field distribution in the far field for a hole of some sort in an opaque screen. Describe the aperture that would give rise to such a pattern and give your reasoning in detail.

Figure P.10.20 (Photo courtesy R. G. Wilson, Illinois Wesleyan University.)



10.21 In light of the five previous questions, identify Fig. P.10.21, explaining what it is and what aperture gave rise to it.

Figure P.10.21 (Photo courtesy R. G. Wilson, Illinois Wesleyan University.)



10.22* Verify that the peak irradiance I_1 of the first "ring" in the Airy pattern for far-field diffraction at a circular aperture is such that $I_1/I(0) = 0.0175$. You might want to use the fact that

$$J_1(u) = \frac{u}{2}\left[1 - \frac{1}{1!2!}(\tfrac{1}{2}u)^2 + \frac{1}{2!3!}(\tfrac{1}{2}u)^4 - \frac{1}{3!4!}(\tfrac{1}{2}u)^6 + \cdots\right]$$

10.23 No lens can focus light down to a perfect point because there will always be some diffraction. Estimate the size of the minimum spot of light that can be expected at the focus of a lens. Discuss the relationship among the focal length, the lens diameter, and the spot size. Take the *f-number* of the lens to be roughly 0.8 or 0.9, which is just about what you can expect for a fast lens.

10.24 Figure P.10.24 shows several aperture configurations. Roughly sketch the Fraunhofer patterns for each. Note that the circular regions should generate Airy-like ring systems centered at the origin.

Figure P.10.24



10.25* Suppose that we have a laser emitting a diffraction-limited beam ($\lambda_0 = 632.84$ nm) with a 2-mm diameter. How big a light spot would be produced on the surface of the Moon a distance of 376×10^3 km away from such a device? Neglect any effects of the Earth's atmosphere.

10.26* If you peered through a 0.75-mm hole at an eye chart, you would probably notice a decrease in visual acuity. Compute the angular limit of resolution, assuming that it's determined only by diffraction; take $\lambda_0 = 550$ nm. Compare your results with the value of 1.7×10^{-4} rad, which corresponds to a 4.0-mm pupil.

10.27 The neoimpressionist painter Georges Seurat was a member of the pointillist school. His paintings consist of an enormous number of closely spaced small dots ($\approx \frac{1}{10}$ inch) of pure pigment. The illusion of color mixing is produced only in the eye of the observer. How far from such a painting should one stand in order to achieve the desired blending of color?

10.28* The Mount Palomar telescope has an objective mirror with a 508-cm diameter. Determine its angular limit of resolution at a wavelength of 550 nm, in radians, degrees, and seconds of arc. How far apart must two objects be on the surface of the Moon if they are to be resolvable by the Palomar telescope? The Earth–Moon distance is 3.844×10^8 m; take $\lambda_0 = 550$ nm. How far apart must two objects be on the Moon if they are to be distinguished by the eye? Assume a pupil diameter of 4.00 mm.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

10.29* A transmission grating whose lines are separated by 3.0×10^{-6} m is illuminated by a narrow beam of red light ($\lambda_0 = 694.3$ nm) from a ruby laser. Spots of diffracted light, on both sides of the undeflected beam, appear on a screen 2.0 m away. How far from the central axis is each of the two nearest spots?

10.30* A diffraction grating with slits 0.60×10^{-3} cm apart is illuminated by light with a wavelength of 500 nm. At what angle will the third-order maximum appear?

10.31* A diffraction grating produces a second-order spectrum of yellow light ($\lambda_0 = 550$ nm) at 25°. Determine the spacing between the lines on the grating.

10.32 White light falls normally on a transmission grating that contains 1000 lines per centimeter. At what angle will red light ($\lambda_0 = 650$ nm) emerge in the first-order spectrum?

10.33* Light from a laboratory sodium lamp has two strong yellow components at 589.5923 nm and 588.9953 nm. How far apart in the first-order spectrum will these two lines be on a screen 1.00 m from a grating having 10 000 lines per centimeter?

10.34* Sunlight impinges on a transmission grating that is formed with 5000 lines per centimeter. Does the third-order spectrum overlap the second-order spectrum? Take red to be 780 nm and violet to be 390 nm.

10.35 Light having a frequency of 4.0×10^{14} Hz is incident on a grating formed with 10 000 lines per centimeter. What is the highest-order spectrum that can be seen with this device? Explain.

10.36* Suppose that a grating spectrometer while in vacuum on Earth sends 500-nm light off at an angle of 20.0° in the first-order spectrum. By comparison, after landing on the planet Mongo, the same light is diffracted through 18.0°. Determine the index of refraction of the Mongoian atmosphere.

10.37 Prove that the equation

$$a(\sin \theta_m - \sin \theta_i) = m\lambda \qquad [10.61]$$

when applied to a transmission grating, is independent of the refractive index.

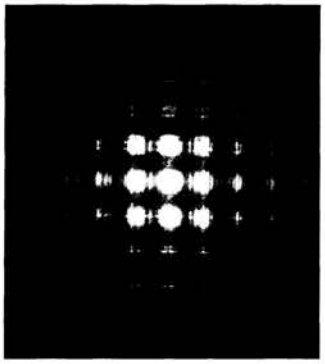
10.38 A high-resolution grating 260 mm wide, with 300 lines per millimeter, at about 75° in autocollimation has a resolving power of just about 10^6 for $\lambda = 500$ nm. Find its free spectral range. How do these values of $\mathcal{R}$ and $(\Delta\lambda)_{\text{fsr}}$ compare with those of a Fabry–Perot etalon having a 1-cm air gap and a finesse of 25?

10.39 What is the total number of lines a grating must have in order just to separate the sodium doublet ($\lambda_1 = 5895.9$ Å, $\lambda_2 = 5890.0$ Å) in the third order?

10.40* Imagine an opaque screen containing 30 randomly located circular holes. The light source is such that every aperture is coherently illuminated by its own plane wave. Each wave in turn is completely incoherent with respect to all the others. Describe the resulting far-field diffraction pattern.

10.41 Imagine that you are looking through a piece of square woven cloth at a point source ($\lambda_0 = 600$ nm) 20 m away. If you see a square arrangement of bright spots located about the point source (Fig. P.10.41), each separated by an apparent nearest-neighbor distance of 12 cm, how close together are the strands of cloth?

Figure P.10.41 (Photo by E. H.)



10.42* Perform the necessary mathematical operations needed to arrive at Eq. (10.76).

10.43 Referring to Fig. 10.38, integrate the expression $dS = 2\pi\rho^2 \sin\varphi\, d\varphi$ over the lth zone to get the area of that zone,

$$A_l = \frac{\lambda \pi \rho}{\rho + r_0}\left[r_0 + \frac{(2l-1)\lambda}{4}\right]$$

Show that the mean distance to the lth zone is

$$r_l = r_0 + \frac{(2l-1)\lambda}{4}$$

so that the ratio A_l/r_l is constant.

10.44* Derive Eq. (10.84).

10.45 Use the Cornu spiral to make a rough sketch of $|\tilde{\mathbf{B}}_{12}(w)|^2$ versus $(w_1 + w_2)/2$ for $\Delta w = 5.5$. Compare your results with those of Fig. 10.57.

10.46 The Fresnel integrals have the asymptotic forms (corresponding to large values of w) given by

$$\mathscr{C}(w) \approx \tfrac{1}{2} + \left(\frac{1}{\pi w}\right)\sin\left(\frac{\pi w^2}{2}\right)$$

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

$$\mathcal{S}(w) \approx \tfrac{1}{2} - \left(\frac{1}{\pi w}\right) \cos\left(\frac{\pi w^2}{2}\right)$$

Using this fact, show that the irradiance in the shadow of a semi-infinite opaque screen decreases in proportion to the inverse square of the distance to the edge, as z_1 and therefore v_1 become large.

10.47 What would you expect to see on the plane of observation if the half-plane Σ in Fig. 10.58 were semi-transparent?

10.48 Plane waves from a collimated He–Ne laserbeam ($\lambda_0 = 632.8$ nm) impinge on a steel rod with a 2.5-mm diameter. Draw a rough *graphic* representation of the diffraction pattern that would be seen on a screen 3.16 m from the rod.

10.49 Make a rough sketch of the irradiance function for a Fresnel diffraction pattern arising from a double slit. What would the Cornu spiral picture look like at point P_0?

10.50* Make a rough sketch of a possible Fresnel diffraction pattern arising from each of the indicated apertures (Fig. P.10.50).

Figure P.10.50



10.51* Suppose the slit in Fig. 10.54 is made very wide. What will the Fresnel diffraction pattern look like?

10.52* Collimated light from a krypton ion laser at 568.19 nm impinges normally on a circular aperture. When viewed axially from a distance of 1.00 m, the hole uncovers the first half-period Fresnel zone. Determine its diameter.

10.53* Plane waves impinge perpendicularly on a screen with a small circular hole in it. It is found that when viewed from some axial point P the hole uncovers $\frac{1}{2}$ of the first half-period zone. What is the irradiance at P in terms of the irradiance there when the screen is removed?

10.54* A collimated beam from a ruby laser (694.3 nm) having an irradiance of 10 W/m^2 is incident perpendicularly on an opaque screen containing a square hole 5.0 mm on a side. Compute the irradiance at a point on the central axis 250 cm from the aperture.

10.55* A long narrow slit 0.10 mm wide is illuminated by light of wavelength 500 nm coming from a point source 0.90 m away. Determine the irradiance at a point 2.0 m beyond the screen when the slit is centered on, and perpendicular to, the line from the source to the point of observation. Write your answer in terms of the unobstructed irradiance.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

11 Fourier Optics

11.1 Introduction

In what is to follow we will extend the discussion of Fourier methods introduced in Chapter 7. It is our intent to provide a strong basic introduction to the subject rather than a complete treatment. Besides its real mathematical power, Fourier analysis leads to a marvelous way of treating optical processes in terms of spatial frequencies.* It is always exciting to discover a new bag of analytic toys, but it's perhaps even more valuable to unfold yet another way of thinking about a broad range of physical problems—we shall do both.†

The primary motivation here is to develop an understanding of the way optical systems process light to form images. In the end we want to know all about the amplitudes and phases of the lightwaves reaching the image plane. Fourier methods are especially suited to that task, so we first extend the treatment of Fourier transforms begun earlier. Several transforms are particularly useful in the analysis, and these will be considered first. Among them is the delta function, which will subsequently be used to represent a point source of light. How an optical system responds to an object comprising a large number of delta-function point sources will be considered in Section 11.3.1. The relationship between Fourier analysis and Fraunhofer diffraction is explored throughout the discussion, but is given special attention in Section 11.3.3. The chapter ends with a return to the problem of image evaluation, this time from a different, though related, perspective: the object is treated not as a collection of point sources but as a scatterer of plane waves.

*See Section 13.2 for a further nonmathematical discussion.

†As general references for this chapter, see R. C. Jennison, *Fourier Transforms and Convolutions for the Experimentalist;* N. F. Barber, *Experimental Correlograms and Fourier Transforms;* A. Papoulis, *Systems and Transforms with Applications in Optics;* J. W. Goodman, *Introduction to Fourier Optics;* J Gaskill, *Linear Systems, Fourier Transforms, and Optics;* R. G. Wilson, *Fourier Series and Optical Transform Techniques in Contemporary Optics;* and the excellent series of booklets *Images and Information,* by B. W. Jones et al.

11.2 Fourier Transforms

11.2.1 One-Dimensional Transforms

It was seen in Section 7.4 that a one-dimensional function of some space variable $f(x)$ could be expressed as a linear combination of an infinite number of harmonic contributions:

$$f(x) = \frac{1}{\pi}\left[\int_0^\infty A(k)\cos kx\,dk + \int_0^\infty B(k)\sin kx\,dk\right] \quad [7.56]$$

The weighting factors that determine the significance of the various angular spatial frequency (k) contributions, that is, $A(k)$ and $B(k)$, are the *Fourier cosine and sine transforms of* $f(x)$ given by

$$A(k) = \int_{-\infty}^{+\infty} f(x')\cos kx'\,dx'$$

and

$$B(k) = \int_{-\infty}^{+\infty} f(x')\sin kx'\,dx' \quad [7.57]$$

respectively. Here the quantity x' is a dummy variable over which the integration is carried out, so that neither $A(k)$ nor $B(k)$ is an explicit function of x', and the choice of symbol used to denote it is irrelevant. The sine and cosine transforms can be consolidated into a single complex exponential expres-

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

sion as follows: substituting [Eq. 7.57] into [Eq. 7.56], we obtain

$$f(x) = \frac{1}{\pi}\int_0^{\infty} \cos kx \int_{-\infty}^{+\infty} f(x') \cos kx'\, dx'\, dk + \frac{1}{\pi}\int_0^{\infty} \sin kx \int_{-\infty}^{+\infty} f(x') \sin kx'\, dx'\, dk$$

But since $\cos k(x' - x) = \cos kx \cos kx' + \sin kx \sin kx'$, this can be rewritten as

$$f(x) = \frac{1}{\pi}\int_0^{\infty} \left[\int_{-\infty}^{+\infty} f(x') \cos k(x'-x)\, dx'\right] dk \quad (11.1)$$

The quantity in the square brackets is an even function of k, and therefore changing the limits on the outer integral leads to

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} f(x') \cos k(x'-x)\, dx'\right] dk \quad (11.2)$$

Inasmuch as we are looking for an exponential representation, Euler's theorem comes to mind. Consequently, observe that

$$\frac{i}{2\pi}\int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} f(x') \sin k(x'-x)\, dx'\right] dk = 0$$

because the factor in brackets is an odd function of k. Adding these last two expressions yields the complex* form of the Fourier integral,

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} f(x') e^{ikx'}\, dx'\right] e^{-ikx}\, dk \quad (11.3)$$

Thus we can write

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} F(k) e^{-ikx}\, dk \quad (11.4)$$

provided that

$$F(k) = \int_{-\infty}^{+\infty} f(x) e^{ikx}\, dx \quad (11.5)$$

*To keep the notation in standard form, and when there's no loss of clarity, we omit the tilde symbol that would otherwise indicate a complex quantity.

having set $x' = x$ in Eq. (11.5). The function $F(k)$ is the **Fourier transform** of $f(x)$, which is symbolically denoted by

$$F(k) = \mathcal{F}\{f(x)\} \quad (11.6)$$

Actually, several equivalent, slightly different ways of defining the transform appear in the literature. For example, the signs in the exponentials could be interchanged, or the factor of $1/2\pi$ could be split symmetrically between $f(x)$ and $F(k)$; each would then have a coefficient of $1/\sqrt{2\pi}$. Note that $A(k)$ is the real part of $F(k)$, while $B(k)$ is its imaginary part, that is,

$$F(k) = A(k) + iB(k) \quad (11.7a)$$

As was seen in Section 2.4, a complex quantity like this can also be written in terms of a real-valued amplitude, $|F(k)|$, the *amplitude spectrum*, and a real-valued phase, $\phi(k)$, the *phase spectrum*:

$$F(k) = |F(k)|e^{i\phi(k)} \quad (11.7b)$$

and sometimes this form can be quite useful [see Eq. (11.96)].

Just as $F(k)$ is the transform of $f(x)$, $f(x)$ itself is said to be the **inverse Fourier transform** of $F(k)$, or symbolically

$$f(x) = \mathcal{F}^{-1}\{F(k)\} = \mathcal{F}^{-1}\{\mathcal{F}\{f(x)\}\} \quad (11.8)$$

and $f(x)$ and $F(k)$ are frequently referred to as a Fourier-transform pair. It's possible to construct the transform and its inverse in an even more symmetrical form in terms of the spatial frequency $\kappa = 1/\lambda = k/2\pi$. Still, in whatever way it's expressed, the transform will not be precisely the same as the inverse transform because of the minus sign in the exponential. As a result (Problem 11.10), in the present formulation,

$$\mathcal{F}\{F(k)\} = 2\pi f(-x) \quad \text{while} \quad \mathcal{F}^{-1}\{F(k)\} = f(x)$$

This is most often inconsequential, especially for even functions where $f(x) = f(-x)$, so we can expect a good deal of parity between functions and their transforms.

Obviously, if f were a function of time rather than space, we would merely have to replace x by t and then k, the angular spatial frequency, by ω, the angular temporal frequency, in order to get the appropriate transform pair in the time domain, that is,

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} F(\omega) e^{-i\omega t} d\omega \quad (11.9)$$

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

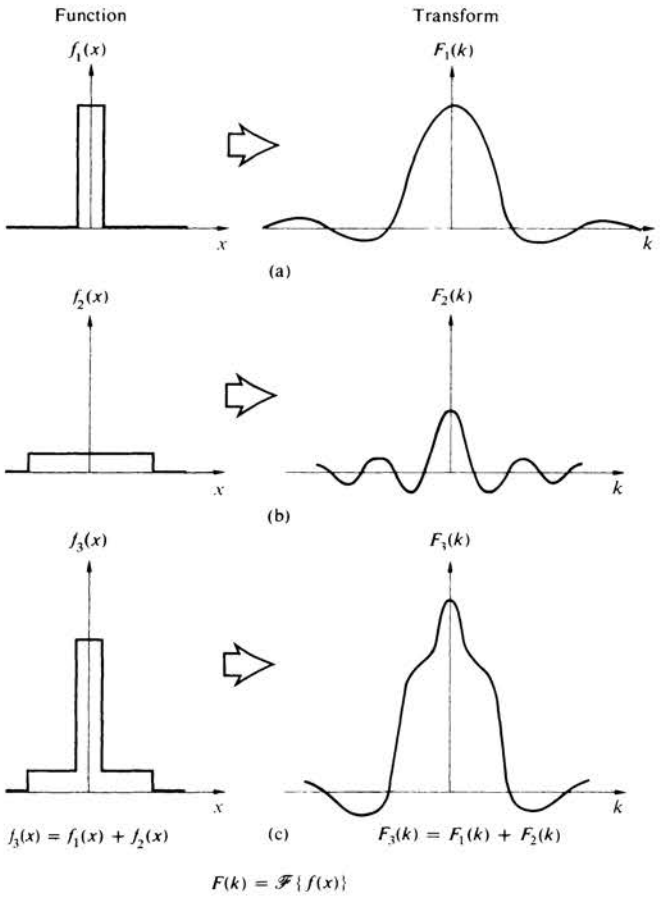


Figure 11.1 A composite function and its Fourier transform.

and
$$F(\omega) = \int_{-\infty}^{+\infty} f(t)e^{i\omega t}\,dt \qquad (11.10)$$

It should be mentioned that if we write $f(x)$ as a sum of functions, its transform [Eq. (11.5)] will apparently be the sum of the transforms of the individual component functions. This can sometimes be quite a convenient way of establishing the transforms of complicated functions that can be constructed from well-known constituents. Figure 11.1 makes this procedure fairly self-evident.

Transform of the Gaussian Function

As an example of the method, let's examine the Gaussian probability function,

$$f(x) = Ce^{-ax^2} \qquad (11.11)$$

where $C = \sqrt{a/\pi}$ and a is a constant. If you like, you can imagine this to be the profile of a pulse at $t = 0$. The familiar bell-shaped curve (Fig. 11.2*a*) is quite frequently encountered in Optics. It will be germane to a diversity of considerations, such as the wave packet representation of individual photons, the cross-sectional irradiance distribution of a laserbeam in the TEM_{00} mode, and the statistical treatment of thermal light in coherence theory. Its Fourier transform, $\mathcal{F}\{f(x)\}$, is obtained by evaluating

$$F(k) = \int_{-\infty}^{+\infty} (Ce^{-ax^2})e^{ikx}\,dx$$

On completing the square, the exponent, $-ax^2 + ikx$, becomes $-(x\sqrt{a} - ik/2\sqrt{a})^2 - k^2/4a$, and letting $x\sqrt{a} - ik/2\sqrt{a} = \beta$ yields

$$F(k) = \frac{C}{\sqrt{a}}\,e^{-k^2/4a}\int_{-\infty}^{+\infty} e^{-\beta^2}\,d\beta$$

The definite integral can be found in tables and equals $\sqrt{\pi}$; hence

$$F(k) = e^{-k^2/4a} \qquad (11.12)$$

which is again a Gaussian function (Fig. 11.2*b*), this time with k as the variable. The standard deviation is defined as the range of the variable (x or k) over which the function drops by a factor of $e^{-1/2} = 0.607$ of its maximum value. Thus the standard deviations for the two curves are $\sigma_x = 1/\sqrt{2a}$ and $\sigma_k =$

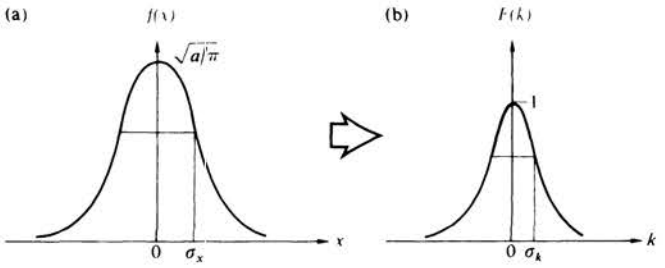


Figure 11.2 A Gaussian and its Fourier transform.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

$\sqrt{2a}$ and $\sigma_x \sigma_k = 1$. As a increases, $f(x)$ becomes narrower while, in contrast, $F(k)$ broadens. In other words, the shorter the pulse length, the broader the spatial frequency bandwidth.

11.2.2 Two-Dimensional Transforms

Thus far the discussion has been limited to one-dimensional functions, but Optics generally involves two-dimensional signals: for example, the field across an aperture or the flux-density distribution over an image plane. The Fourier-transform pair can readily be generalized to two dimensions, whereupon

$$f(x,y) = \frac{1}{(2\pi)^2} \int_{-\infty}^{+\infty}\int F(k_x,k_y) e^{-i(k_x x + k_y y)}\, dk_x\, dk_y \qquad (11.13)$$

and

$$F(k_x,k_y) = \int_{-\infty}^{+\infty}\int f(x,y) e^{i(k_x x + k_y y)}\, dx\, dy \qquad (11.14)$$

The quantities k_x and k_y are the angular spatial frequencies along the two axes. Suppose we were looking at the image of a tiled floor made up alternately of black and white squares aligned with their edges parallel to the x- and y-directions. If the floor were infinite in extent, the mathematical distribution of reflected light could be regarded in terms of a two-dimensional Fourier series. With each tile having a length ℓ, the spatial period along either axis would be 2ℓ, and the associated fundamental angular spatial frequencies would equal π/ℓ. These and their harmonics would certainly be needed to construct a function describing the scene. If the pattern was finite in extent, the function would no longer be truly periodic, and the Fourier integral would have to replace the series. In effect, Eq. (11.13) says that $f(x, y)$ can be constructed out of a linear combination of elementary functions having the form $\exp[-i(k_x x + k_y y)]$, each appropriately weighted in amplitude and phase by a complex factor $F(k_x, k_y)$. The transform simply tells you how much of and with what phase each elementary component must be added to the recipe. In three dimensions, the elementary functions appear as $\exp[-i(k_x x + k_y y + k_z z)]$ or $\exp(-i\vec{\mathbf{k}}\cdot\vec{\mathbf{r}})$, which correspond to planar surfaces. Furthermore, if f is a wave function, that is, some sort of three-dimensional wave $f(\vec{r}, t)$, these elementary contributions become plane waves that look like $\exp[-i(\vec{\mathbf{k}}\cdot\vec{\mathbf{r}} - \omega t)]$. In other words, *the disturbance can be synthesized out of a linear combination of plane waves having various propagation numbers and moving in various directions.* Similarly, in two dimensions the elementary functions are "oriented" in different directions as well. That is to say, for a given set of values of k_x and k_y, the exponent or phase of the elementary functions will be constant along lines

$$k_x x + k_y y = \text{constant} = A$$

or

$$y = -\frac{k_x}{k_y} x + \frac{A}{k_y} \qquad (11.15)$$

The situation is analogous to one in which a set of planes normal to and intersecting the xy-plane does so along the lines given by Eq. (11.15) for differing values of A. A vector perpendicular to the set of lines, call it $\vec{\mathbf{k}}_\alpha$, would have components k_x and k_y. Figure 11.3 shows several of these lines (for a given k_x and k_y), where $A = 0, \pm 2\pi, \pm 4\pi \ldots$. The slopes are all equal to $-k_x/k_y$ or $-\lambda_y/\lambda_x$ while the y-intercepts equal $A/k_y = A\lambda_y/2\pi$. The orientation of the constant phase lines is

$$\alpha = \tan^{-1}\frac{k_y}{k_x} = \tan^{-1}\frac{\lambda_x}{\lambda_y} \qquad (11.16)$$

The wavelength, or spatial period λ_α, measured along $\vec{\mathbf{k}}_\alpha$, is obtained from the similar triangles in the diagram, where $\lambda_\alpha/\lambda_y = \lambda_x/\sqrt{\lambda_x^2 + \lambda_y^2}$ and

$$\lambda_\alpha = \frac{1}{\sqrt{\lambda_x^{-2} + \lambda_y^{-2}}} \qquad (11.17)$$

The angular spatial frequency k_α, being $2\pi/\lambda_\alpha$, is then

$$k_\alpha = \sqrt{k_x^2 + k_y^2} \qquad (11.18)$$

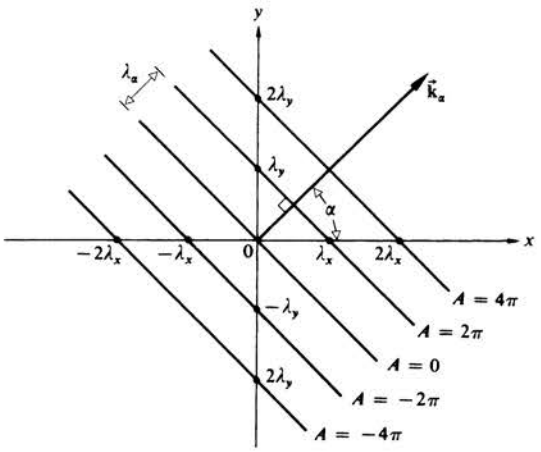


Figure 11.3 Geometry for Eq. (11.15).

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

as expected. All of this just means that in order to construct a two-dimensional function, harmonic terms in addition to those of spatial frequency k_x and k_y will generally have to be included as well, and these are oriented in directions other than along the x- and y-axes.

Return for a moment to Fig. 10.7, which shows an aperture, with the diffracted wave leaving it represented by several different conceptions. One of these ways to envision the complicated emerging wavefront is as a superposition of plane waves coming off in a whole range of directions. These are the Fourier-transform components, which emerge in specific directions with specific values of angular spatial frequency—the zero spatial frequency term corresponding to the undeviated axial wave, the higher spatial frequency terms coming off at increasingly great angles from the central axis. These Fourier components make up the diffracted field as it emerges from the aperture.

Transform of the Cylinder Function

The cylinder function

$$f(x,y) = \begin{cases} 1 & \sqrt{x^2+y^2} \le a \\ 0 & \sqrt{x^2+y^2} > a \end{cases} \tag{11.19}$$

(Fig. 11.4*a*) provides an important practical example of the application of Fourier methods to two dimensions. The mathematics will not be particularly simple, but the relevance of the calculation to the theory of diffraction by circular apertures and lenses amply justifies the effort. The evident circular symmetry suggests polar coordinates, and so let

$$\begin{aligned} k_x &= k_\alpha \cos\alpha \\ k_y &= k_\alpha \sin\alpha \\ x &= r\cos\theta \\ y &= r\sin\theta \end{aligned} \tag{11.20}$$

in which case $dx\,dy = r\,dr\,d\theta$. The transform, $\mathcal{F}\{f(x)\}$, then reads

$$F(k_\alpha,\alpha) = \int_{r=0}^{a}\left[\int_{\theta=0}^{2\pi} e^{ik_\alpha r\cos(\theta-\alpha)}\,d\theta\right] r\,dr \tag{11.21}$$

Inasmuch as $f(x, y)$ is circularly symmetric, its transform must be symmetrical as well. This implies that $F(k_\alpha, \alpha)$ is independent of α. The integral can therefore be simplified by letting α equal some constant value, which we choose to be zero, whereupon

$$F(k_\alpha) = \int_0^a\left[\int_0^{2\pi} e^{ik_\alpha r\cos\theta}\,d\theta\right] r\,dr \tag{11.22}$$

It follows from Eq. (10.47) that

$$F(k_\alpha) = 2\pi\int_0^a J_0(k_\alpha r) r\,dr \tag{11.23}$$

the $J_0(k_\alpha r)$ being a Bessel function of order zero. Introducing a change of variable, namely, $k_\alpha r = w$, we have $dr = k_\alpha^{-1}\,dw$,

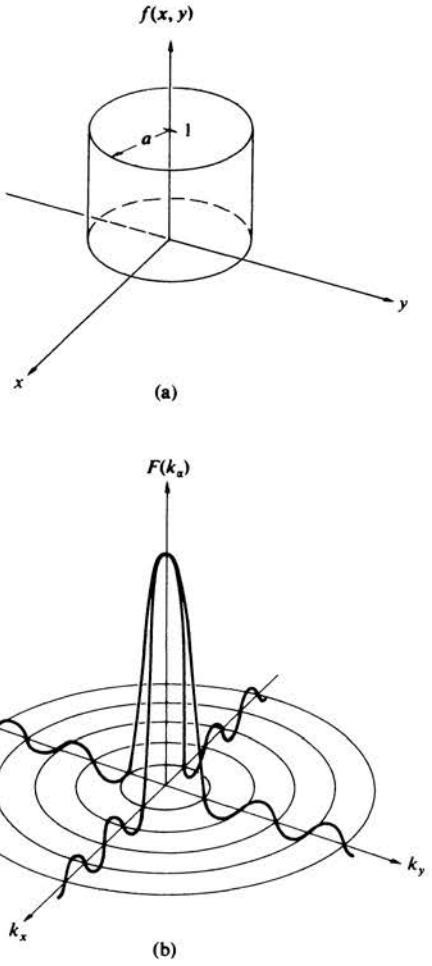


Figure 11.4 The cylinder, or top-hat, function and its transform.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

and the integral becomes

$$\frac{1}{k_\alpha^2}\int_{w=0}^{k_\alpha a} J_0(w)w\,dw \tag{11.24}$$

Using Eq. (10.50), the transform takes the form of a first-order Bessel function (see Fig. 10.22), that is,

$$F(k_\alpha) = \frac{2\pi}{k_\alpha^2}\,k_\alpha a J_1(k_\alpha a)$$

or

$$F(k_\alpha) = 2\pi a^2\left[\frac{J_1(k_\alpha a)}{k_\alpha a}\right] \tag{11.25}$$

The similarity between this expression (Fig. 11.4*b*) and the formula for the electric field in the Fraunhofer diffraction pattern of a circular aperture [Eq. (10.51)] is, of course, not accidental.

The Lens as a Fourier Transformer

Figure 11.5 shows a transparency, located in the front focal plane of a converging lens, being illuminated by parallel light. This object, in turn, scatters plane waves, which are collected by the lens, and parallel bundles of rays are brought to convergence at its back focal plane. If a screen were placed there, at Σ_t, the so-called **transform plane**, we would see the far-field diffraction pattern of the object spread across it. (This is essentially the configuration of Fig. 10.7*e*.) In other words, the electric-field distribution across the object mask, which is known

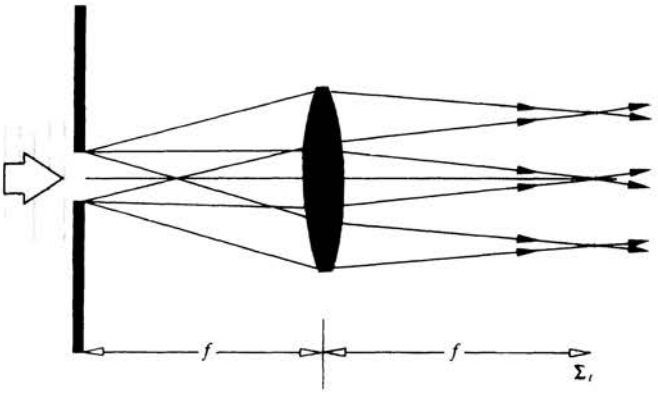


Figure 11.5 The light diffracted by a transparency at the front (or object) focal point of a lens converges to form the far-field diffraction pattern at the back (or image) focal point of the lens.

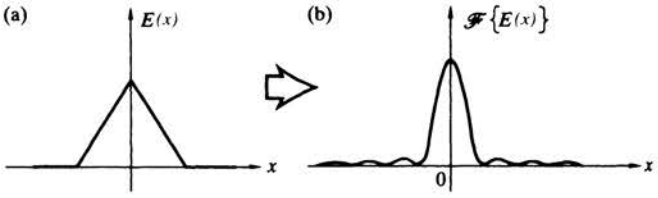


Figure 11.6 The transform of the triangle function is the sinc² function.

as the **aperture function**, is transformed by the lens into the far-field diffraction pattern. Although this assertion is true enough for most purposes, it's not exactly true. After all, the lens doesn't actually form its image on a plane.

Remarkably, that Fraunhofer $\vec{\mathbf{E}}$-field pattern corresponds to the exact Fourier transform of the aperture function—a fact we shall confirm more rigorously in Section 11.3.3. Here the object is in the front focal plane, and all the various diffracted waves maintain their phase relationships traveling essentially equal optical path lengths to the transform plane. That doesn't quite happen when the object is displaced from the front focal plane. Then there will be a phase deviation, but that is actually of little consequence, since we are generally interested in the irradiance where the phase information is averaged out and the phase distortion is unobservable.

Thus if an otherwise opaque object mask contains a single circular hole, the $\vec{\mathbf{E}}$-field across it will resemble the top hat of Fig. 11.4*a*, and the diffracted field, the Fourier transform, will be distributed in space as a Bessel function, looking very much like Fig. 11.4*b*. Similarly, if the object transparency varies in density only along one axis, such that its amplitude transmission profile is triangular (Fig. 11.6*a*), then the amplitude of the electric field in the diffraction pattern will correspond to Fig. 11.6*b*—the Fourier transform of the triangle function is the sinc-squared function.

11.2.3 The Dirac Delta Function

Many physical phenomena occur over very short durations in time with great intensity, and one is frequently concerned with the consequent response of some system to such stimuli. For example: How will a mechanical device, like a billiard ball, respond to being slammed with a hammer? Or how will a particular circuit behave if the input is a short burst of current? In much the same way, we can envision some stimulus that is a

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

sharp pulse in the space, rather than the time, domain. A bright minute source of light embedded in a dark background is essentially a highly localized, two-dimensional, spatial pulse—a spike of irradiance. A convenient idealized mathematical representation of this sort of sharply peaked stimulus is the **Dirac delta function** $\delta(x)$. This is a quantity that is zero everywhere except at the origin, where it goes to infinity in a manner so as to encompass a *unit area*, that is,

$$\delta(x) = \begin{cases} 0 & x \neq 0 \\ \infty & x = 0 \end{cases} \tag{11.26}$$

and

$$\int_{-\infty}^{+\infty} \delta(x)\, dx = 1 \tag{11.27}$$

This is not really a function in the traditional mathematical sense. In fact, because it is so singular in nature, it remained the focus of considerable controversy long after it was reintroduced and brought into prominence by P.A.M. Dirac in 1930. Yet physicists, pragmatic as they sometimes are, found it so highly useful that it soon became an established tool, despite what seemed a lack of rigorous justification. The precise mathematical theory of the delta function evolved roughly 20 years later, in the early 1950s, principally at the hands of Laurent Schwartz.

Perhaps the most basic operation to which $\delta(x)$ can be applied is the evaluation of the integral

$$\int_{-\infty}^{+\infty} \delta(x) f(x)\, dx$$

Here the expression $f(x)$ corresponds to any continuous function. Over a tiny interval running from $x = -\gamma$ to $+\gamma$ centered about the origin, $f(x) \approx f(0) \approx$ constant, since the function is continuous at $x = 0$. From $x = -\infty$ to $x = -\gamma$ and from $x = +\gamma$ to $x = +\infty$, the integral is zero, simply because the δ-function is zero there. Thus the integral equals

$$f(0) \int_{-\gamma}^{+\gamma} \delta(x)\, dx$$

Because $\delta(x) = 0$ for all x other than 0, the interval can be vanishingly small, that is, $\gamma \to 0$, and still

$$\int_{-\gamma}^{+\gamma} \delta(x)\, dx = 1$$

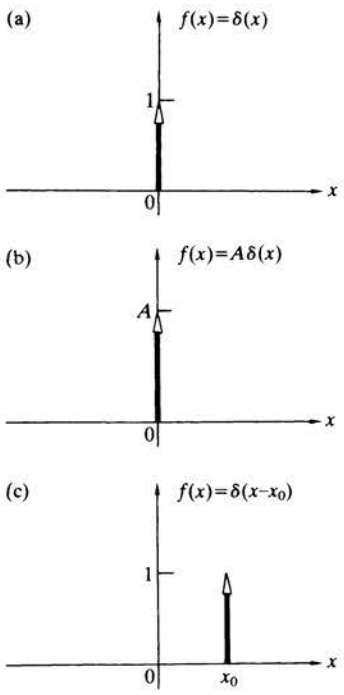


Figure 11.7 The height of the arrow representing the delta function corresponds to the area under the function.

from Eq. (11.27). Hence we have the exact result that

$$\int_{-\infty}^{+\infty} \delta(x) f(x)\, dx = f(0) \tag{11.28}$$

This is often spoken of as the **sifting property** of the δ-function because it manages to extract only the one value of $f(x)$ taken at $x = 0$ from all its possible values. Similarly, with a shift of origin of an amount x_0,

$$\delta(x - x_0) = \begin{cases} 0 & x \neq x_0 \\ \infty & x = x_0 \end{cases} \tag{11.29}$$

and the spike resides at $x = x_0$ rather than $x = 0$, as shown in Fig. 11.7. The corresponding sifting property can be appreciated by letting $x - x_0 = x'$, then with $f(x' + x_0) = g(x')$,

$$\int_{-\infty}^{+\infty} \delta(x - x_0) f(x)\, dx = \int_{-\infty}^{+\infty} \delta(x') g(x')\, dx' = g(0)$$

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

and since $g(0) = f(x_0)$,

$$\int_{-\infty}^{+\infty} \delta(x - x_0) f(x)\, dx = f(x_0) \tag{11.30}$$

Formally, rather than worrying about a precise definition of $\delta(x)$ for each value of x, it would be more fruitful to continue along the lines of defining the effect of $\delta(x)$ on some other function $f(x)$. Accordingly, Eq. (11.28) is really the definition of an entire operation that assigns a number $f(0)$ to the function $f(x)$. Incidentally, an operation that performs this service is called a *functional*.

It is possible to construct a number of sequences of pulses, each member of which has an ever-decreasing width and a concomitantly increasing height, such that any one pulse encompasses a unit area. A sequence of square pulses of height a/L and width L/a for which $a = 1, 2, 3, \ldots$ would fit the bill; so would a sequence of Gaussians [Eq. (11.11)],

$$\delta_a(x) = \sqrt{\frac{a}{\pi}}\, e^{-ax^2} \tag{11.31}$$

as in Fig. 11.8, or a sequence of sinc functions

$$\delta_a(x) = \frac{a}{\pi} \operatorname{sinc}(ax) \tag{11.32}$$

Such strongly peaked functions that approach the sifting property, that is, for which

$$\lim_{a \to \infty} \int_{-\infty}^{+\infty} \delta_a(x) f(x)\, dx = f(0) \tag{11.33}$$

are known as *delta sequences*. It is often useful, but not actually rigorously correct, to imagine $\delta(x)$ as the convergence

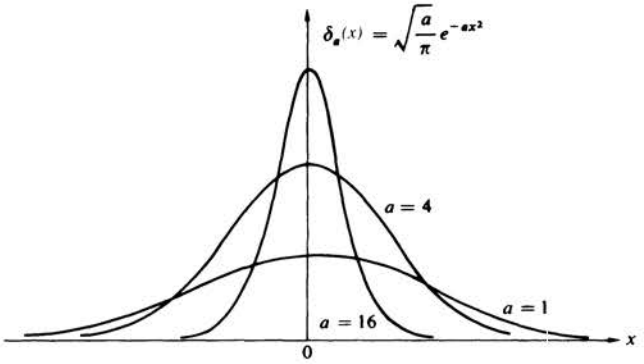


Figure 11.8 A sequence of Gaussians.

limit of such sequences as $a \to \infty$. The extension of these ideas into two dimensions is provided by the definition

$$\delta(x,y) = \begin{cases} \infty & x = y = 0 \\ 0 & \text{otherwise} \end{cases} \tag{11.34}$$

and

$$\int_{-\infty}^{+\infty}\int \delta(x,y)\, dx\, dy = 1 \tag{11.35}$$

and the sifting property becomes

$$\int_{-\infty}^{+\infty}\int f(x,y)\delta(x - x_0)\delta(y - y_0)\, dx\, dy = f(x_0,y_0) \tag{11.36}$$

Another representation of the δ-function follows from Eq. (11.3), the Fourier integral, which can be restated as

$$f(x) = \int_{-\infty}^{+\infty} \left[\frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-ik(x-x')} dk \right] f(x')\, dx'$$

and hence

$$f(x) = \int_{-\infty}^{+\infty} \delta(x - x') f(x')\, dx' \tag{11.37}$$

provided that

$$\delta(x - x') = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-ik(x-x')} dk \tag{11.38}$$

Equation (11.37) is identical to Eq. (11.30), since by definition from Eq. (11.29) $\delta(x - x') = \delta(x' - x)$. The (divergent) integral of Eq. (11.38) is zero everywhere except at $x = x'$. Evidently, with $x' = 0$, $\delta(x) = \delta(-x)$ and

$$\delta(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-ikx} dk = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{ikx} dk \tag{11.39}$$

This implies, via Eq. (11.4), that the delta function can be thought of as the inverse Fourier transform of unity, that is, $\delta(x) = \mathcal{F}^{-1}\{1\}$ and so $\mathcal{F}\{\delta(x)\} = 1$. We can imagine a square pulse becoming narrower and taller as its transform, in turn, grows broader, until finally the pulse is infinitesimal in width, and its transform is infinite in extent, in other words, a constant.

Displacements and Phase Shifts

If the δ-spike is shifted off $x = 0$ to, say, $x = x_0$, its transform will change phase but not amplitude—that remains equal to

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

one. To see this, evaluate

$$\mathcal{F}\{\delta(x - x_0)\} = \int_{-\infty}^{+\infty} \delta(x - x_0) e^{ikx} \, dx$$

From the sifting property [Eq. (11.30)] the expression becomes

$$\mathcal{F}\{\delta(x - x_0)\} = e^{ikx_0} \tag{11.40}$$

What we see is that only the phase is affected, the amplitude being one as it was when $x_0 = 0$. This whole process can be appreciated somewhat more intuitively if we switch to the time domain and think of an infinitesimally narrow pulse (such as a spark) occurring at $t = 0$. This results in the generation of an infinite range of frequency components, which are all initially in-phase at the instant of creation ($t = 0$). On the other hand, suppose the pulse occurs at a time t_0. Again every frequency is produced, but in this situation the harmonic components are all in phase at $t = t_0$. Consequently, if we extrapolate back, the phase of each constituent at $t = 0$ will now have to be different, depending on the particular frequency. Besides, we know that all these components superimpose to yield zero everywhere except at t_0, so that a frequency-dependent phase shift is quite reasonable. This phase shift is evident in Eq. (11.40) for the space domain. Note that it does vary with the angular spatial frequency k.

All of this is quite general in its applicability, and we observe that ***the Fourier transform of a function that is displaced in space (or time) is the transform of the undisplaced function multiplied by an exponential that is linear in phase*** (Problem 11.14). This property of the transform will be of special interest presently, when we consider the image of several point sources that are separated but otherwise identical. The process can be appreciated diagramatically with the help of Figs. 11.9 and 7.29. To shift the square wave by $\pi/4$ to the right, the fundamental must be shifted $\frac{1}{8}$-wavelength (or, say, 1.0 mm), and every component must then be displaced an equal distance (i.e., 1.0 mm). Thus each component must be shifted in phase by an amount specific to it that produces a 1.0-mm displacement. Here each is displaced, in turn, by a phase of $m\pi/4$.

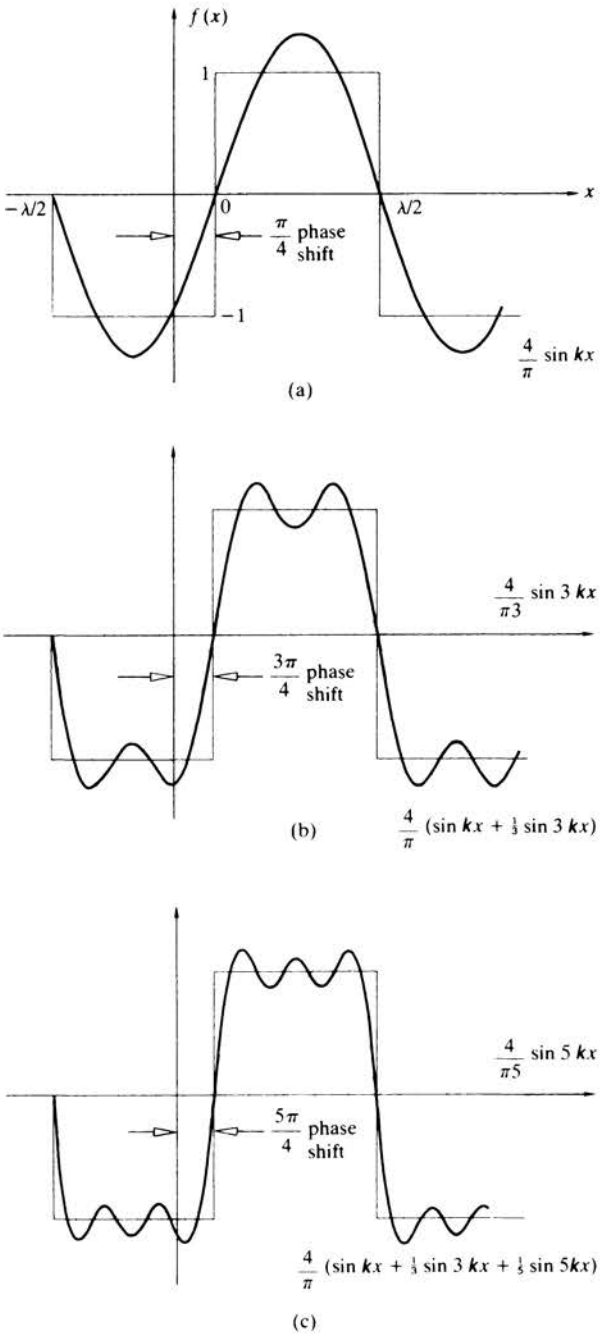


Figure 11.9 A shifted square wave showing the corresponding change in phase for each component wave.

Sines and Cosines

We saw earlier (Fig. 11.1) that if the function at hand can be written as a sum of individual functions, its transform is simply the sum of the transforms of the component functions.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

Suppose we have a string of delta functions spread out uniformly like the teeth on a comb,

$$f(x) = \sum_j \delta(x - x_j) \tag{11.41}$$

When the number of terms is infinite, this periodic function is often called *comb*(x). In any event, the transform will simply be a sum of terms, such as that of Eq. (11.40):

$$\mathcal{F}\{f(x)\} = \sum_j e^{ikx_j} \tag{11.42}$$

In particular, if there are two δ-functions, one at $x_0 = d/2$ and the other at $x_0 = -d/2$,

$$f(x) = \delta[x - (+d/2)] + \delta[x - (-d/2)]$$

and $$\mathcal{F}\{f(x)\} = e^{ikd/2} + e^{-ikd/2}$$

which is just

$$\mathcal{F}\{f(x)\} = 2\cos(kd/2) \tag{11.43}$$

as in Fig. 11.10. Thus the transform of the sum of these two symmetrical δ-functions is a cosine function and vice versa. The composite is a real even function, and $F(k) = \mathcal{F}\{f(x)\}$ will also be real and even. This should be reminiscent of Young's Experiment with infinitesimally narrow slits—we'll come back to it later. If the phase of one of the δ-functions is shifted, as in Fig. 11.11, the composite function is asymmetrical, it's odd,

$$f(x) = \delta[x - (+d/2)] - \delta[x - (-d/2)]$$

and $$\mathcal{F}\{f(x)\} = e^{ikd/2} - e^{ikd/2} = 2i\sin(kd/2) \tag{11.44}$$

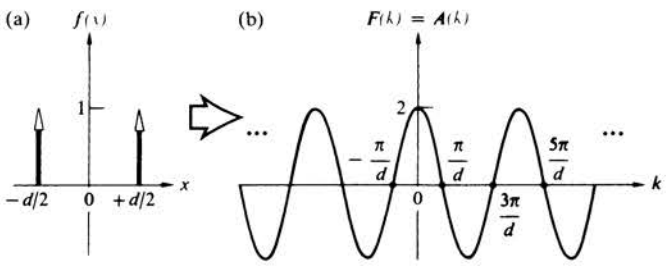


Figure 11.10 Two delta functions and their cosine-function transform.

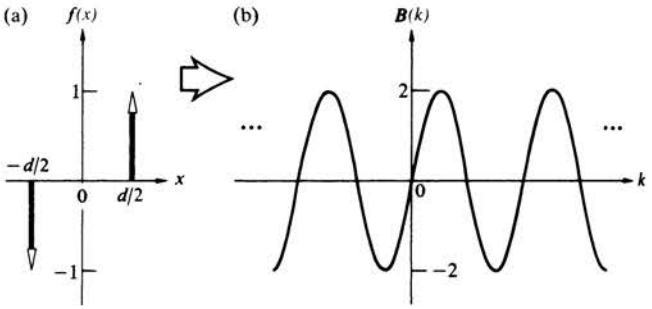


Figure 11.11 Two delta functions and their sine-function transform.

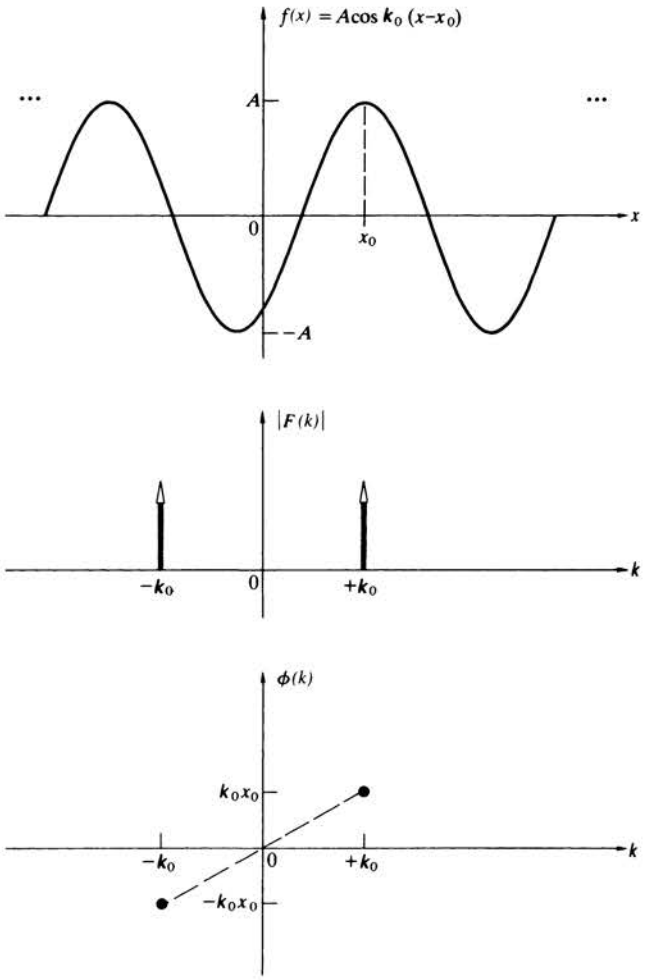


Figure 11.12 The spectra of a shifted cosine function.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

The real sine transform [Eq. (11.7)] is then

$$B(k) = 2 \sin (kd/2) \qquad (11.45)$$

and it too is an odd function.

This raises an interesting point. Recall that there are two alternative ways to consider the complex transform: either as the sum of a real and an imaginary part, from Eq. (11.7a), or as the product of an amplitude and a phase term, from Eq. (11.7b). It happens that the cosine and sine are rather special functions; the former is associated with a purely real contribution, and the latter is associated with a purely imaginary one. Most functions, even harmonic ones, will usually be a blend of real and imaginary parts. For example, once a cosine is displaced a little, the new function, which is typically neither odd nor even, has both a real and an imaginary part. Moreover, it can be expressed as a cosinusoidal amplitude spectrum, which is appropriately phase-shifted (Fig. 11.12). Notice that when the cosine is shifted $\frac{1}{4}\lambda$ into a sine, the relative phase difference between the two component delta functions is again π rad.

Figure 11.13 displays in summary form a number of transforms, mostly of harmonic functions. Observe how the functions and transforms in (*a*) and (*b*) combine to produce the function and its transform in (*d*). As a rule, each member of the pair of δ-pulses in the frequency spectrum of a harmonic function is located along the k-axis at a distance from the origin equal to the fundamental angular spatial frequency of $f(x)$. Since any well-behaved periodic function can be expressed as a Fourier series, it can also be represented as an array of pairs of delta functions, each weighted appropriately and each a distance from the k-origin equal to the angular spatial frequency of the particular harmonic contribution—*the frequency spectrum of any periodic function will be discrete*. One of the most remarkable of the periodic functions is *comb*(x): as shown in Fig. 11.14, its transform is also a comb function.

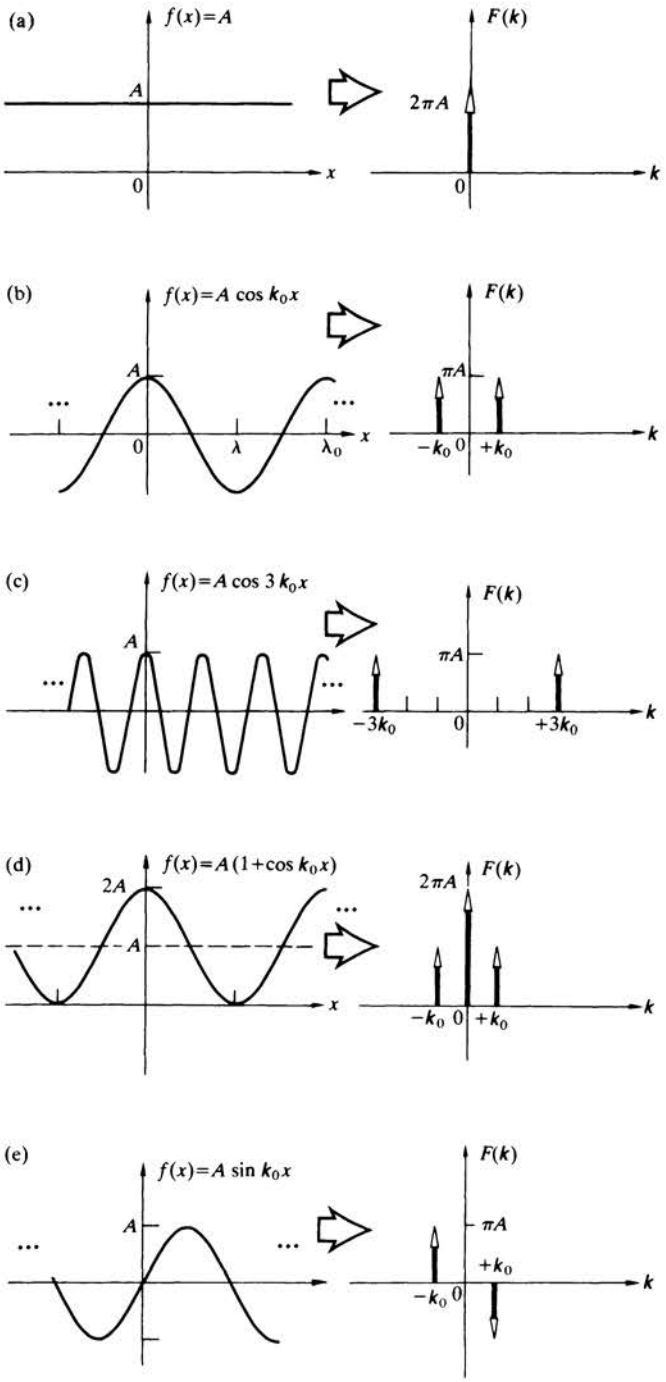


Figure 11.13 Some functions and their transforms.

11.3 Optical Applications

11.3.1 Linear Systems

Fourier techniques provide a particularly elegant framework from which to evolve a description of the formation of images. And for the most part, this will be the direction in which we shall be moving, although some side excursions are unavoidable in order to develop the needed mathematics.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

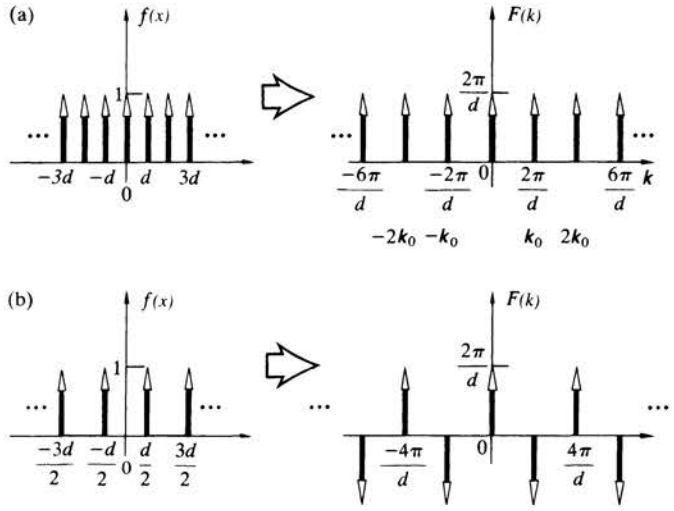


Figure 11.14 (*a*) The comb function and its transform. (*b*) A shifted comb function and its transform.

A key point in the analysis is the concept of a **linear system**, which in turn is defined in terms of its input–output relations. Suppose then that an input signal $f(y, z)$ passing through some optical system results in an output $g(Y, Z)$. The system is linear if:

1. multiplying $f(y, z)$ by a constant a produces an output $ag(Y, Z)$.
2. when the input is a weighted sum of two (or more) functions, $af_1(y, z) + bf_2(y, z)$, the output will similarly have the form $ag_1(Y, Z) + bg_2(Y, Z)$, where $f_1(y, z)$ and $f_2(y, z)$ generate $g_1(Y, Z)$ and $g_2(Y, Z)$ respectively.

Furthermore, a linear system will be *space invariant* if it possesses the property of *stationarity*; that is, in effect, changing the position of the input merely changes the location of the output without altering its functional form. The idea behind much of this is that the output produced by an optical system can be treated as a linear superposition of the outputs arising from each of the individual points on the object. In fact, if we symbolically represent the operation of the linear system as $\mathcal{L}\{\}$, the input and output can be written as

$$g(Y,Z) = \mathcal{L}\{f(y,z)\} \tag{11.46}$$

Using the sifting property of the δ-function [Eq. (11.36)], this becomes

$$g(Y, Z) = \mathcal{L}\left\{\int_{-\infty}^{+\infty}\int f(y', z')\delta(y' - y)\delta(z' - z)\, dy'\, dz'\right\}$$

The integral expresses $f(y, z)$ as a linear combination of elementary delta functions, each weighted by a number $f(y', z')$. It follows from the second linearity condition that the system operator can equivalently act on each of the elementary functions; thus

$$g(Y, Z) = \int_{-\infty}^{+\infty}\int f(y', z')\mathcal{L}\{\delta(y' - y)\delta(z' - z)\}\, dy'\, dz' \tag{11.47}$$

The quantity $\mathcal{L}\{\delta(y' - y)\delta(z' - z)\}$ is the response of the system [Eq. (11.46)] to a delta function located at the point (y', z') in the input space—it's called the **impulse response**. Apparently, if the impulse response of a system is known, the output can be determined directly from the input by means of Eq. (11.47). If the elementary sources are coherent, the input and output signals will have to be electric fields; if incoherent, they'll be flux densities.

Consider the self-luminous and, therefore, incoherent source depicted in Fig. 11.15. We can imagine that each point on the object plane, Σ_0, emits light that is processed by the optical system. It emerges to form a spot on the focal or image plane, Σ_i. In addition, *assume that the magnification between object and image planes is one*. The image will be life-sized

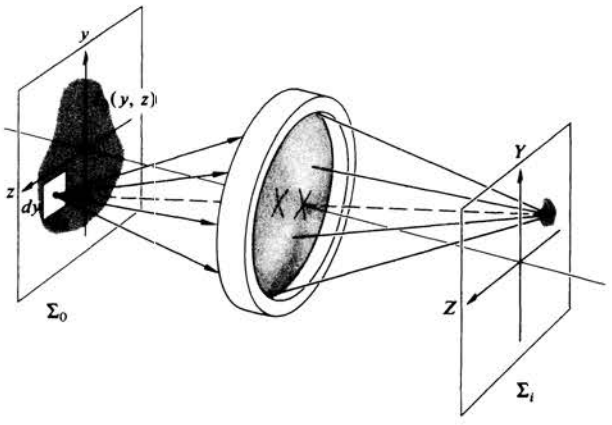


Figure 11.15 A lens system forming an image.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

and erect, which makes it a little easier to deal with for the time being. Notice that if the magnification (M_T) was greater than one, the image would be larger than the object. Consequently, all of its structural details would be larger and broader, so the spatial frequencies of the harmonic contributions that go into synthesizing the image would be lower than those of the object. For example, an object that is a transparency of a sinusoidally varying black and white linear pattern (a sinusoidal amplitude grating) would be imaged having a greater space between maxima and therefore a lower spatial frequency. Besides that, the image irradiance would be decreased by M_T^2, because the image area would be increased by a factor of M_T^2.

If $I_0(y, z)$ is the irradiance distribution on the object plane, an element $dy\,dz$ located at (y, z) will emit a radiant flux of $I_0(y, z)\,dy\,dz$. Because of diffraction (and the possible presence of aberrations), this light is smeared out into some sort of blur spot over a finite area on the image plane rather than focused to a point. The spread of radiant flux is described mathematically by the function $\mathcal{S}(y, z; Y, Z)$, such that the flux density arriving at the image point from $dy\,dz$ is

$$dI_i(Y, Z) = \mathcal{S}(y, z; Y, Z) I_0(y, z)\,dy\,dz \qquad (11.48)$$

This is the patch of light in the image plane at (Y, Z), and $\mathcal{S}(y, z; Y, Z)$ is known as the **point-spread function**. In other words, when the irradiance $I_0(y, z)$ over the source element $dy\,dz$ is 1 W/m^2, $\mathcal{S}(y, z; Y, Z)\,dy\,dz$ is the profile of the resulting irradiance distribution in the image plane. Because of the incoherence of the source, the flux-density contributions from each of its elements are additive, so

$$I_i(Y, Z) = \int_{-\infty}^{+\infty}\int I_0(y, z)\mathcal{S}(y, z; Y, Z)\,dy\,dz \qquad (11.49)$$

In a "perfect," diffraction-limited optical system having no aberrations, $\mathcal{S}(y, z; Y, Z)$ would correspond in shape to the diffraction figure of a point source at (y, z). Evidently, if we set the input equal to a δ-pulse centered at (y_0, z_0), then $I_0(y, z) = A\delta(y - y_0)\delta(z - z_0)$. Here the constant A of magnitude one carries the needed units (i.e., irradiance times area). Thus

$$I_i(Y, Z) = A\int_{-\infty}^{+\infty}\int \delta(y - y_0)\delta(z - z_0)\mathcal{S}(y, z; Y, Z)\,dy\,dz$$

and so from the sifting property,

$$I_i(Y, Z) = A\mathcal{S}(y_0, z_0; Y, Z)$$

The point-spread function has a functional form identical to that of the image generated by a δ-pulse input. It's the impulse response of the system [compare Eqs. (11.47) and (11.49)], whether optically perfect or not. In a well-corrected system $\mathcal{S}$, apart from a multiplicative constant, is the Airy irradiance distribution function [Eq. (10.56)] centered on the Gaussian image point (Fig. 11.16).

If the system is space invariant, a point-source input can be moved about over the object plane without any effect other than changing the location of its image. Equivalently, one can say that the spread function is the same for any point (y, z). In practice, however, the spread function will vary, but even so, the image plane can be divided into small regions, over each of which $\mathcal{S}$ doesn't change appreciably. Thus if the object, and therefore its image, is small enough, the system can be taken to be space invariant. We can imagine a spread function sitting at every Gaussian image point on Σ_i, each multiplied by a different weighting factor $I_0(y, z)$ but all of the same general shape independent of (y, z). Since the magnification was set at one, the coordinates of any object and conjugate image point have the same magnitude.

If we were dealing with coherent light, we would have to consider how the system acted upon an input that was again a δ-pulse, but this time one representing the field amplitude.

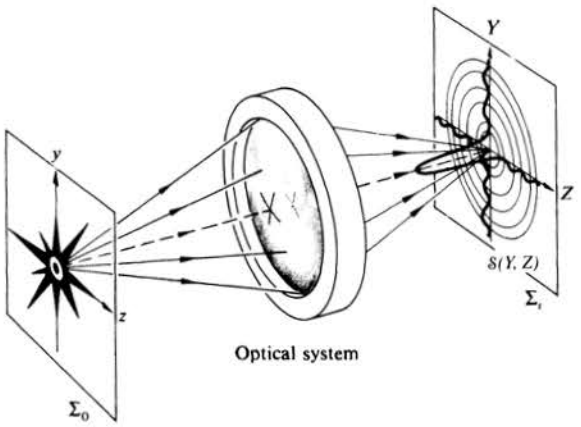


Figure 11.16 The point-spread function: the irradiance produced by the optical system with an input point source.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

Once more the resulting image would be described by a spread function, although it would be an *amplitude* spread function. For a diffraction-limited circular aperture, the amplitude spread function looks like Fig. 10.23*b*. And finally, we would have to be concerned about the interference that would take place on the image plane as the coherent fields interacted. By contrast, with incoherent object points the process occurring on the image plane is simply the summation of overlapping irradiances, as depicted in one dimension in Fig. 11.17. Each source point, with its own strength, corresponds to an appropriately scaled δ-pulse, and in the image plane each of these is smeared out, via the spread function. The sum of all the overlapping contributions is the image irradiance.

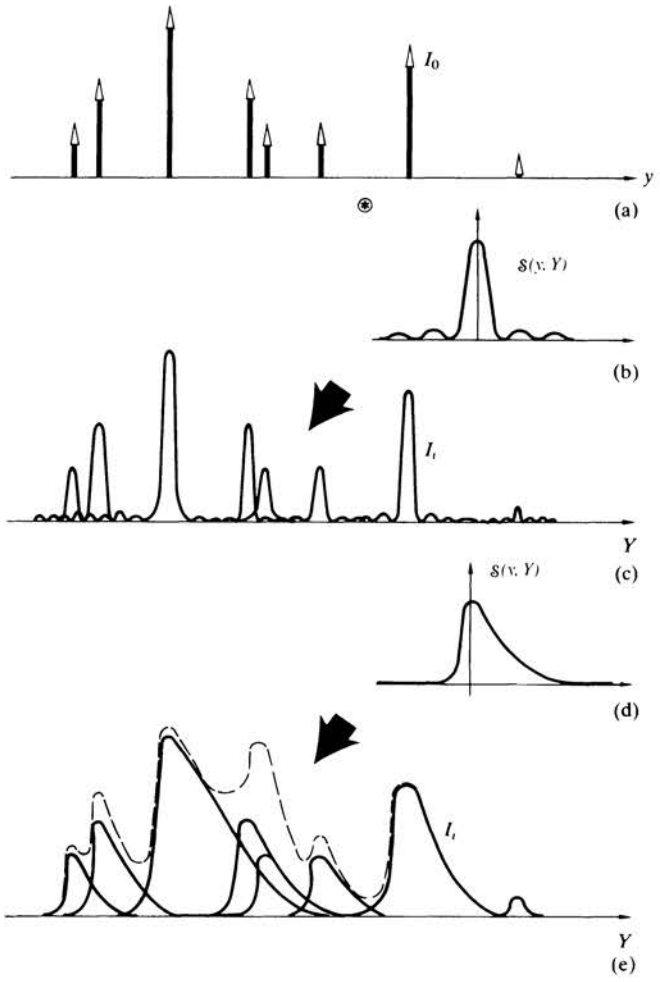


Figure 11.17 Here (*a*) is convolved first with (*b*) to produce (*c*) and then with (*d*) to produce (*e*). The resulting pattern is the sum of all the spread-out contributions as indicated by the dashed curve in (*e*).

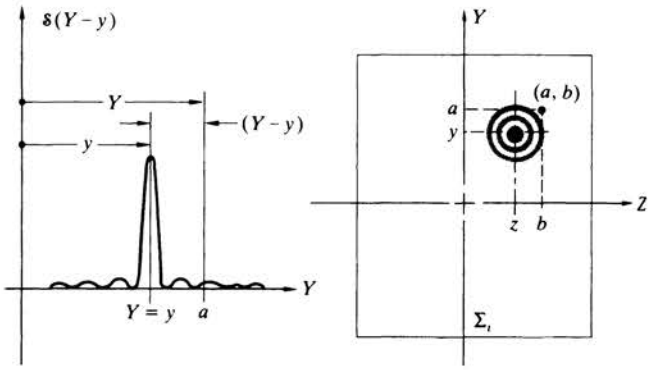


Figure 11.18 The point-spread function.

What kind of dependence on the image and object space variables will $\mathcal{S}(y, z; Y, Z)$ have? The spread function can only depend on (y, z) as far as the location of its center is concerned. Thus the value of $\mathcal{S}(y, z; Y, Z)$ anywhere on Σ_i merely depends on the displacement at that location from the particular Gaussian image point ($Y = y$, $Z = z$) on which $\mathcal{S}$ is centered (Fig. 11.18). In other words,

$$\mathcal{S}(y, z; Y, Z) = \mathcal{S}(Y - y, Z - z) \tag{11.50}$$

When the object point is on the central axis ($y = 0, z = 0$), the Gaussian image point is as well, and the spread function is then just $\mathcal{S}(Y, Z)$, as depicted in Fig. 11.16. Under the circumstances of space invariance and incoherence,

$$I_i(Y, Z) = \int_{-\infty}^{+\infty}\int I_0(y, z)\mathcal{S}(Y - y, Z - z)\, dy\, dz \tag{11.51}$$

11.3.2 The Convolution Integral

Figure 11.17 shows a one-dimensional representation of the distribution of point-source δ-functions that make up the object. The corresponding image is essentially obtained by "dealing out" an appropriately weighted point-spread function to the location of each image point on Σ_i and then adding up all the contributions at each point along Y. This dealing out of one function to every point of (and weighted by) another function is a process known as **convolution**, and we say that one func-

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

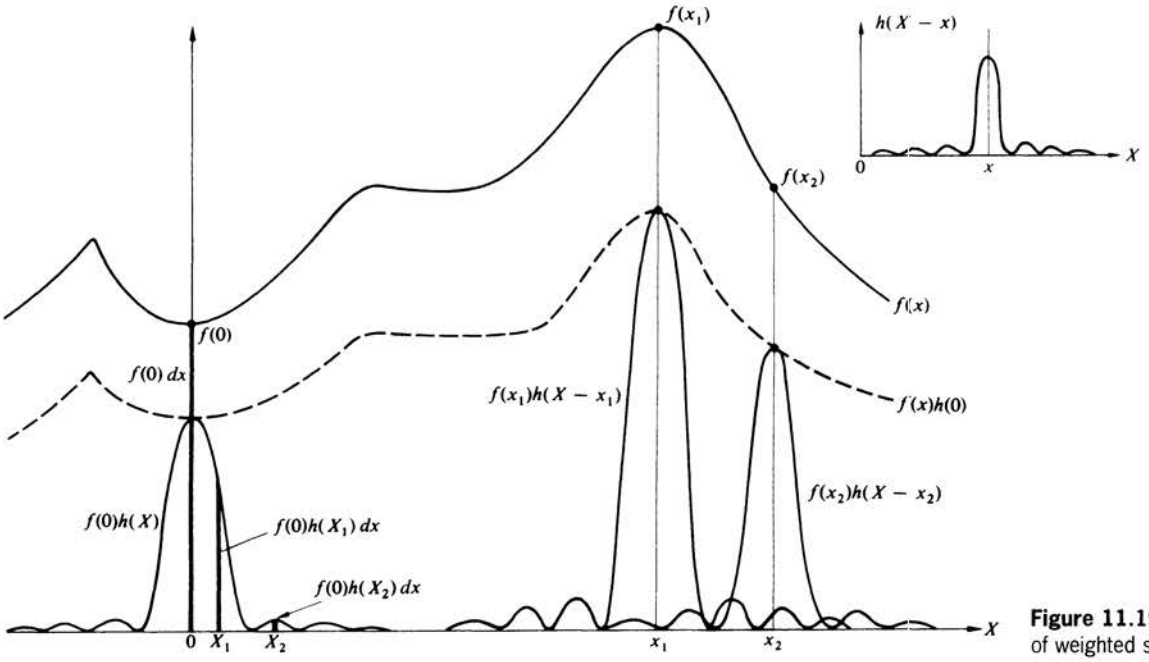


Figure 11.19 The overlapping of weighted spread functions.

tion $I_0(y)$, is convolved with another, $\mathcal{S}(y, Y)$, or vice versa.

This procedure can be carried out in two dimensions as well, and that's essentially what is being done by Eq. (11.51), the so-called ***convolution integral***. The corresponding one-dimensional expression describing the convolution of two functions $f(x)$ and $h(x)$,

$$g(X) = \int_{-\infty}^{+\infty} f(x)h(X - x)\,dx \qquad (11.52)$$

is easier to appreciate. In Fig. 11.17 one of the two functions was a group of δ-pulses, and the convolution operation was particularly easy to visualize. Still, we can imagine any function to be composed of a "densely packed" continuum of δ-pulses and treat it in much the same fashion. Let us now examine in some detail exactly how the integral of Eq. (11.52) mathematically manages to perform the convolution. The essential features of the process are illustrated in Fig. 11.19. The resulting signal $g(X_1)$, at some point X_1 in the output space, is a linear superposition of all the individual overlapping contributions that exist at X_1. In other words, each source element dx yields a signal of a particular strength $f(x)\,dx$, which is then smeared out by the system into a region centered about the Gaussian image point ($X = x$). The output at X_1 is then $dg(X_1) = f(x)h(X_1 - x)\,dx$. The integral sums up all of these contributions from each source element. Of course the elements more remote from a given point on Σ_i contribute less because the spread function generally drops off with displacement. Thus we can imagine $f(x)$ to be a one-dimensional irradiance distribution, such as a series of vertical bands, as in Fig. 11.20. If the one-dimensional **line-spread function**, $h(X - x)$, is that of Fig. 11.20*d*, the resulting image will simply be a somewhat blurred version of the input (Fig. 11.20*e*).

Let's now examine the convolution a bit more as a mathematical entity. Actually it's a rather subtle beast, performing a process that might certainly not be obvious at first glance, so let's approach it from a slightly different viewpoint. Accordingly, we will have two ways of thinking about the convolution integral, and we shall show that they are equivalent.

Suppose $h(x)$ looks like the asymmetrical function in Fig. 11.21*a*. Then $h(-x)$ appears in Fig. 11.21*b*, and its shifted form $h(X - x)$ is shown in (*c*). The convolution of $f(x)$ [depicted in (*d*)] and $h(x)$ is $g(X)$, as given by Eq. (11.52). This is often written more concisely as $f(x) \circledast h(x)$. The integral simply says that the area under the product function $f(x)h(X - x)$ for all x is $g(X)$. Evidently, the product is nonzero only over the range d wherein $h(X - x)$ is nonzero, that is, where the two curves overlap (Fig. 11.21*e*). At a particular point X_1 in the

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

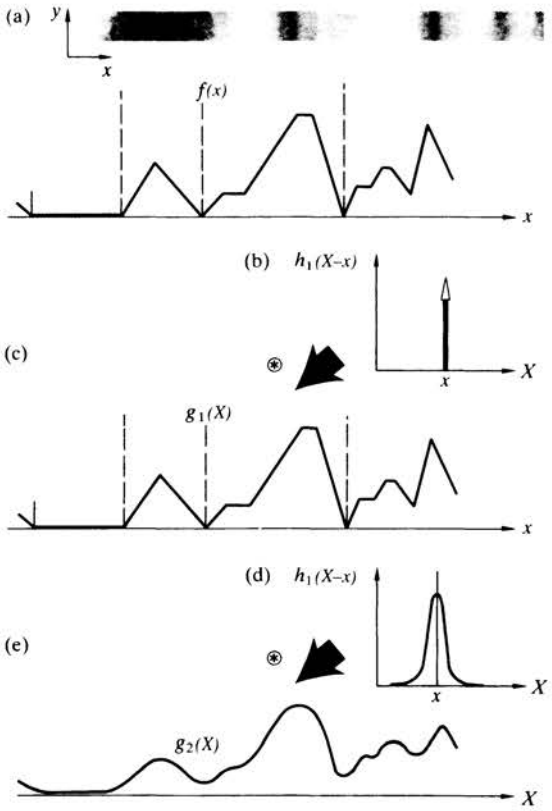


Figure 11.20 The irradiance distribution is converted to a function $f(x)$ shown in (*a*). This is convolved with a δ-function (*b*) to yield a duplicate of $f(x)$. By contrast, convolving $f(x)$ with the spread function h_2 in (*d*) yields a smoothed-out curve represented by $g_2(x)$ in (*e*).

output space, the area under the product $f(x)h(X_1 - x)$ is $g(X_1)$. This fairly direct interpretation can be related back to the physically more pleasing view of the integral in terms of overlapping point contributions, as depicted previously in Fig. 11.19. Remember that there we said that each source element was smeared out in a blur spot on the image plane having the shape of the spread function. Now suppose we take the direct approach and wish to compute the product area of Fig. 11.21*e* at X_1, that is, $g(X_1)$. A differential element dx centered on any point in the region of overlap (Fig. 11.22*a*), say x_1, will contribute an amount $f(x_1)h(X_1 - x_1)\,dx$ to the area. This same differential element will make an identical contribution when viewed in the overlapping spread-function scheme. To see this, examine (*b*) and (*c*) in Fig. 11.22, which are *now drawn*

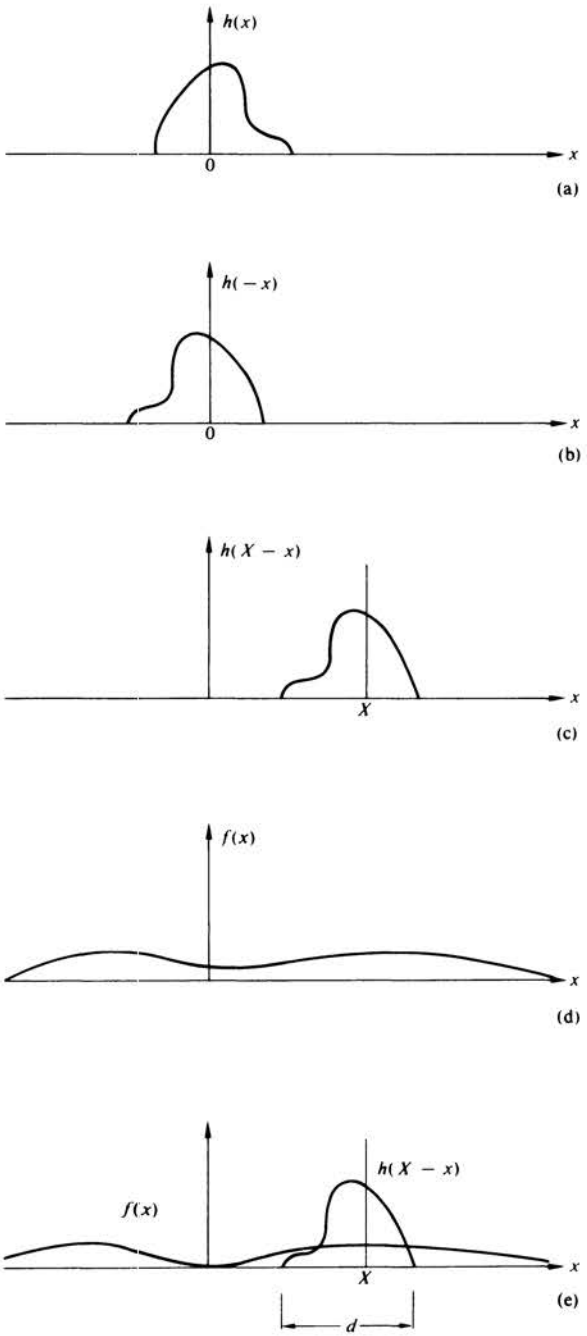


Figure 11.21 The geometry of the convolution process in the object coordinates.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

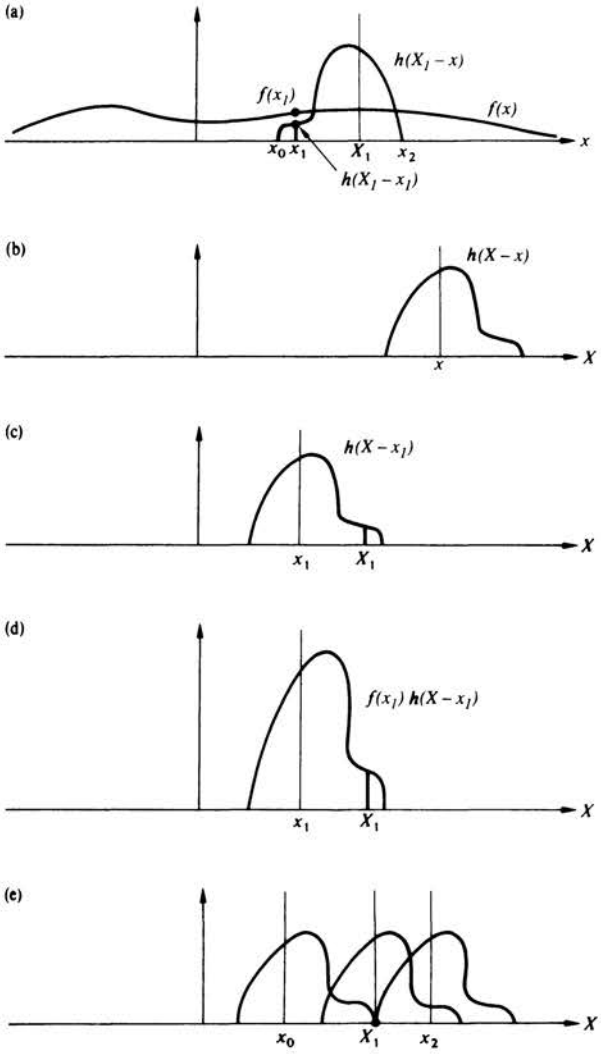


Figure 11.22 The geometry of the convolution process in the image coordinates.

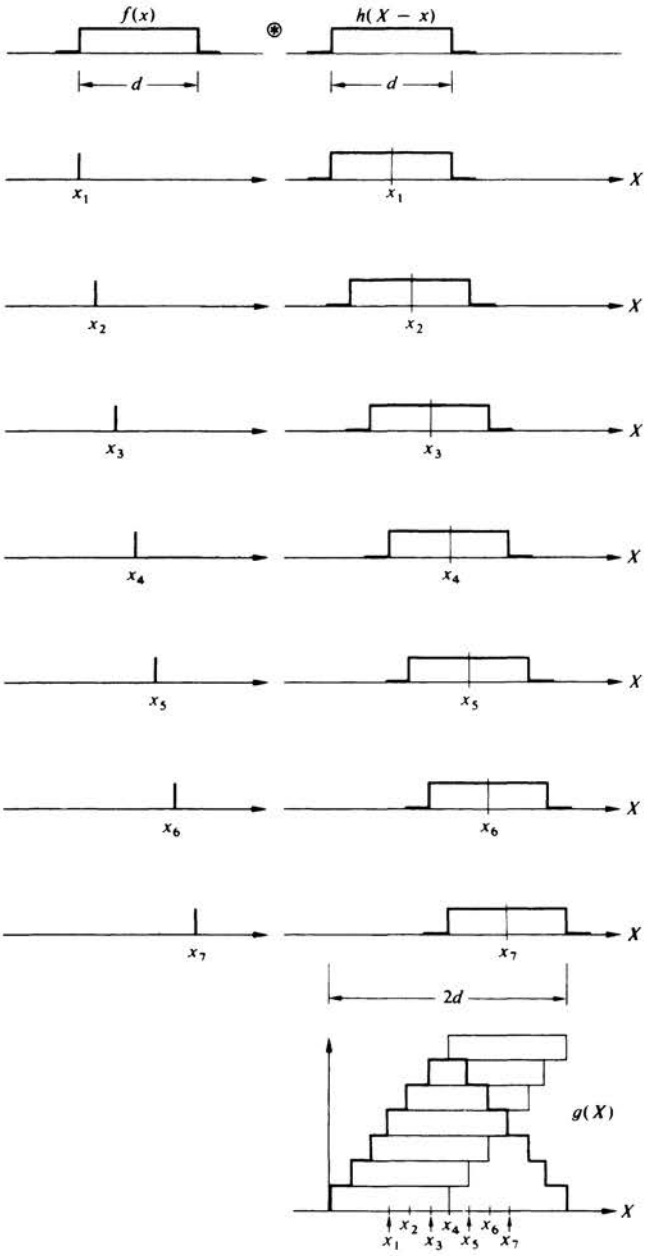


Figure 11.23 Convolution of two square pulses. The fact that we represented $f(x)$ by a finite number of delta functions (viz., 7) accounts for the steps in $g(X)$.

in the output space. The latter shows the spread function "centered" at $X = x_1$. A source element dx, in this case located on the object at x_1, generates a smeared-out signal proportional to $f(x_1)h(X - x_1)$, as in (d), where $f(x_1)$ is just a number. The piece of this signal that exists at X_1 is $f(x_1)h(X_1 - x_1)\,dx$, which indeed is identical to the contribution made by dx at x_1 in (a). Similarly, each differential element of the product area (at any $x = x'$) in Fig. 11.22*a* has its counterpart in a curve like

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

that of (d) but "centered" on a new point $(X = x')$. Points beyond $x = x_2$ make no contribution because they are not in the overlap region of (a) and, equivalently, because they are too far from X_1 for the smear to reach it, as shown in (e).

If the functions being convolved are simple enough, $g(X)$ can be determined roughly without any calculations at all. The convolution of two identical square pulses is illustrated, from both of the viewpoints discussed above, in Figs. 11.23 and 11.24. In Fig. 11.23 each impulse constituting $f(x)$ is spread out into a square pulse and summed. In Fig. 11.24 the overlapping area, as h varies, is plotted against X. In both instances the result is a triangular pulse. Incidentally, observe that $(f \circledast h) = (h \circledast f)$, as can be seen by a change of variable $(x' = X - x)$ in Eq. (11.52), being careful with the limits (see Problem 11.15).

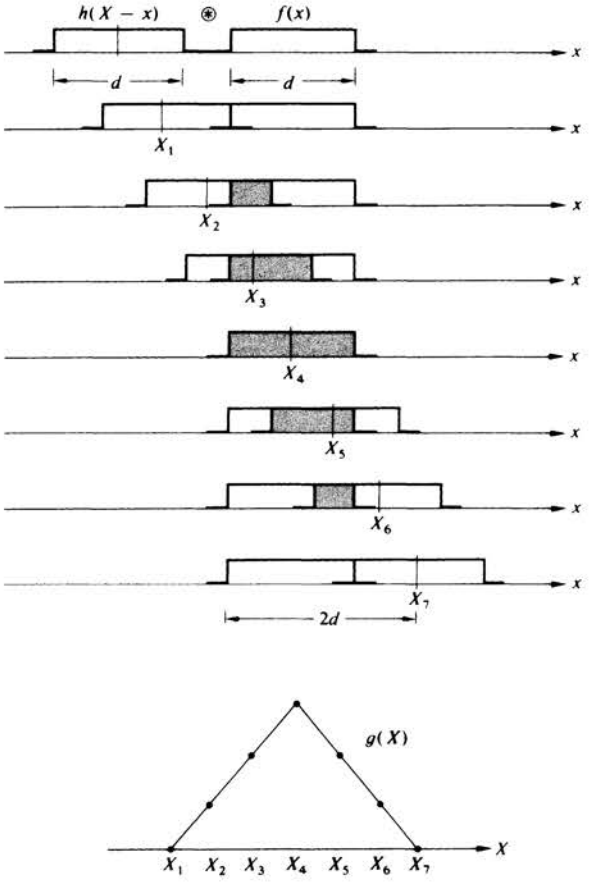


Figure 11.24 Convolution of two square pulses.

Figure 11.25 illustrates the convolution of two functions $I_0(y, z)$ and $\mathcal{S}(y, z)$ in two dimensions, as given by Eq. (11.51). Here the volume under the product curve $I_0(y, z)\mathcal{S}(Y - y, Z - z)$, that is, the region of overlap, equals $I_i(Y, Z)$ at (Y, Z); see Problem 11.16.

The Convolution Theorem

Suppose we have two functions $f(x)$ and $h(x)$ with Fourier transforms $\mathcal{F}\{f(x)\} = F(k)$ and $\mathcal{F}\{h(x)\} = H(k)$, respectively. The **convolution theorem** states that if $g = f \circledast h$,

$$\mathcal{F}\{g\} = \mathcal{F}\{f \circledast h\} = \mathcal{F}\{f\} \cdot \mathcal{F}\{h\} \tag{11.53}$$

or

$$G(k) = F(k)H(k) \tag{11.54}$$

where $\mathcal{F}\{g\} = G(k)$. The proof is quite straightforward:

$$\mathcal{F}\{f \circledast h\} = \int_{-\infty}^{+\infty} g(X)e^{ikX}\,dX$$

$$= \int_{-\infty}^{+\infty} e^{ikX}\left[\int_{-\infty}^{+\infty} f(x)h(X - x)\,dx\right]dX$$

Thus

$$G(k) = \int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty} h(X - x)e^{ikX}dX\right]f(x)\,dx$$

If we put $w = X - x$ in the inner integral, then $dX = dw$ and

$$G(k) = \int_{-\infty}^{+\infty} f(x)e^{ikx}dx \int_{-\infty}^{+\infty} h(w)e^{ikw}dw$$

Hence

$$G(k) = F(k)H(k)$$

which verifies the theorem. As an example of its application, refer to Fig. 11.26. Since the convolution of two identical square pulses $(f \circledast h)$ is a triangular pulse (g), the product of their transforms (Fig. 7.17) must be the transform of g, namely,

$$\mathcal{F}\{g\} = [d \operatorname{sinc}(kd/2)]^2 \tag{11.55}$$

As an additional example, convolve a square pulse with the two δ-functions of Fig. 11.11. The transform of the resulting double pulse (Fig. 11.27) is again the product of the individual transforms.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

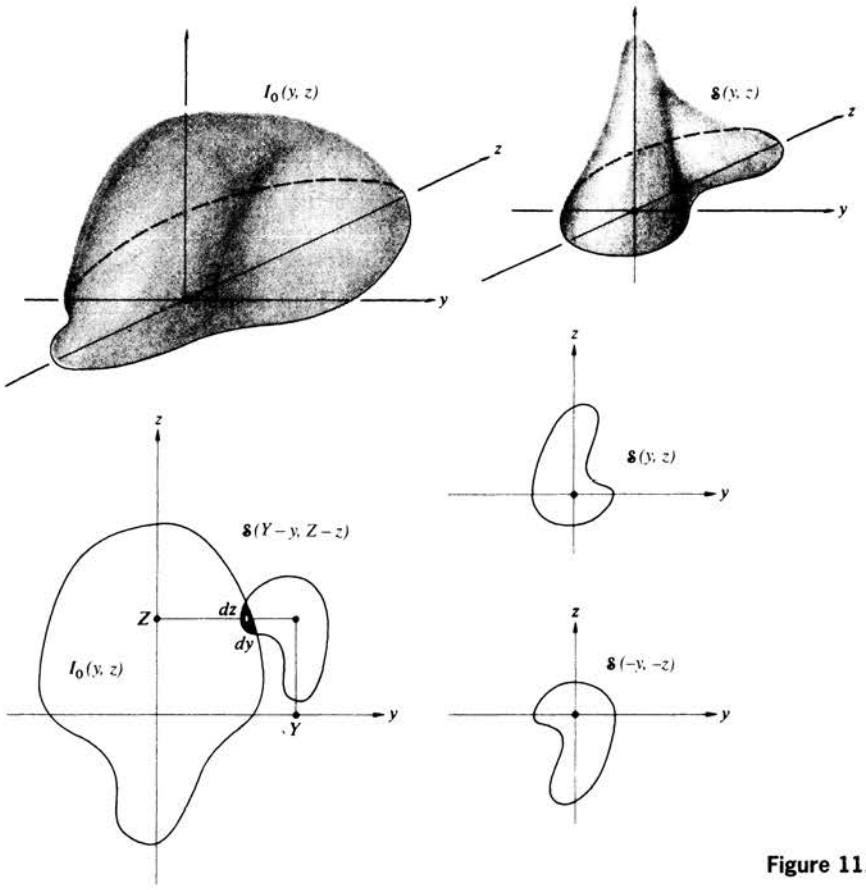


Figure 11.25 Convolution in two dimensions.

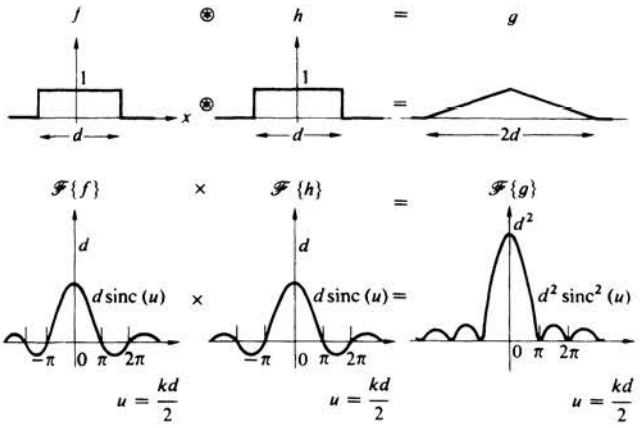


Figure 11.26 An illustration of the convolution theorem.

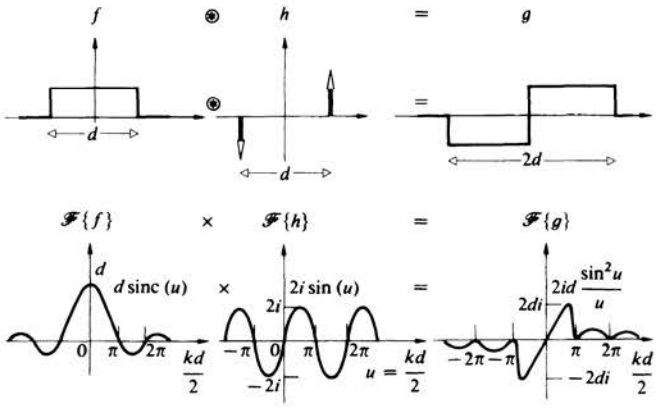


Figure 11.27 An illustration of the convolution theorem.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

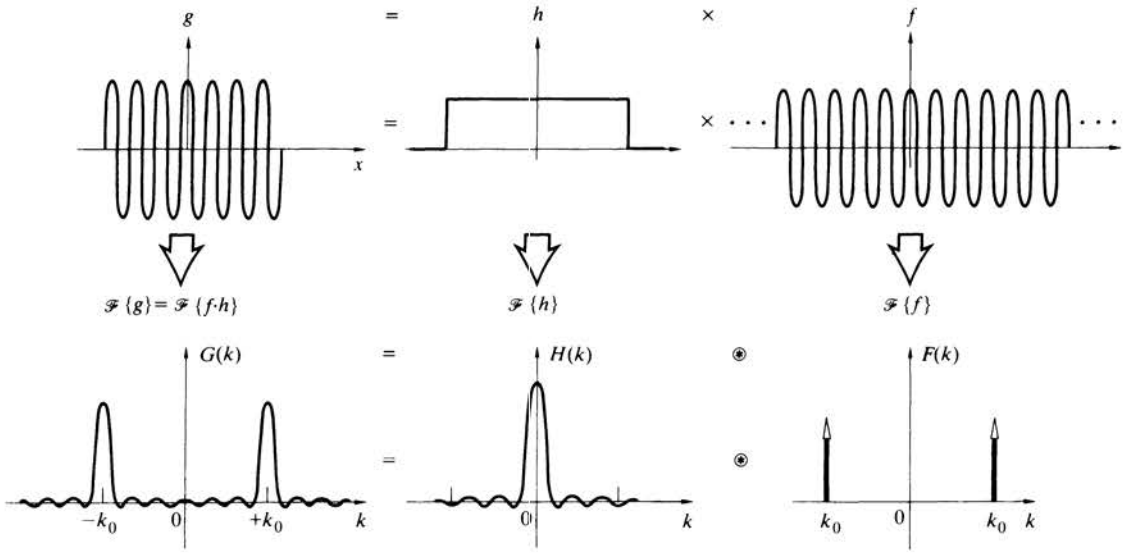


Figure 11.28
An example of the frequency convolution theorem.

The *k*-space counterpart of Eq. (11.53), namely, the *frequency convolution theorem*, is given by

$$\mathcal{F}\{f \cdot h\} = \frac{1}{2\pi}\mathcal{F}\{f\} \circledast \mathcal{F}\{h\} \tag{11.56}$$

That is, the transform of the product is the convolution of the transforms.

Figure 11.28 makes the point rather nicely. Here an infinitely long cosine, $f(x)$, is multiplied by a rectangular pulse, $h(x)$, which truncates it into a short oscillatory wavetrain, $g(x)$. The transform of $f(x)$ is a pair of delta functions, the transform of the rectangular pulse is a sinc function, and the convolution of the two is the transform of $g(x)$. Compare this result with that of Eq. (7.60).

Transform of the Gaussian Wave Packet

As a further example of the usefulness of the convolution theorem, let's evaluate the Fourier transform of a pulse of light in the configuration of the wave packet of Fig. 11.29. Taking a rather general approach, notice that since a one-dimensional harmonic wave has the form

$$\tilde{E}(x, t) = E_0 e^{-i(k_0 x - \omega t)}$$

one need only modulate the amplitude to get a pulse of the desired structure. Assuming the wave's profile to be independent of time, we can write it as

$$\tilde{E}(x, 0) = f(x)e^{-ik_0 x}$$

Now, to determine $\mathcal{F}\{f(x)e^{-ik_0 x}\}$ evaluate

$$\int_{-\infty}^{+\infty} f(x)e^{-ik_0 x}e^{ikx}\,dx \tag{11.57}$$

Letting $k' = k - k_0$, we get

$$F(k') = \int_{-\infty}^{+\infty} f(x)e^{ik'x}\,dx = F(k - k_0) \tag{11.58}$$

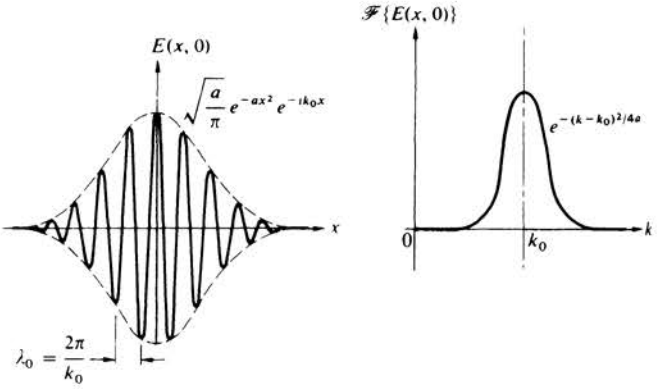


Figure 11.29 A Gaussian wave packet and its transform.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

In other words, if $F(k) = \mathcal{F}\{f(x)\}$, then $F(k - k_0) = \mathcal{F}\{f(x)e^{-ik_0x}\}$. For the specific case of a Gaussian envelope [Eq. (11.11)], as in the figure, $f(x) = \sqrt{a/\pi}\, e^{-ax^2}$, that is,

$$\tilde{E}(x, 0) = \sqrt{a/\pi}\, e^{-ax^2} e^{-ik_0x} \tag{11.59}$$

From the foregoing discussion and Eq. (11.12), it follows that

$$\mathcal{F}\{\tilde{E}(x, 0)\} = e^{-(k-k_0)^2/4a} \tag{11.60}$$

In quite a different way, the transform can be determined from Eq. (11.56). The expression $\tilde{E}(x, 0)$ is now viewed as the product of the two functions $f(x) = \sqrt{a/\pi} \exp(-ax^2)$ and $h(x) = \exp(-ik_0x)$. One way to evaluate $\mathcal{F}\{h\}$ is to set $f(x) = 1$ in Eq. (11.57). This yields the transform of 1 with k replaced by $k - k_0$. Since $\mathcal{F}\{1\} = 2\pi\delta(k)$ (see Problem 11.4), we have $\mathcal{F}\{e^{-ik_0x}\} = 2\pi\delta(k - k_0)$. Thus $\mathcal{F}\{\tilde{E}(x, 0)\}$ is $1/2\pi$ times the convolution of $2\pi\delta(k - k_0)$, with the Gaussian $e^{-k^2/4a}$ centered on zero. The result* is once again a Gaussian centered on k_0, namely, $e^{-(k-k_0)^2/4a}$.

11.3.3 Fourier Methods in Diffraction Theory

Fraunhofer Diffraction

Fourier-transform theory provides a particularly beautiful insight into the mechanism of Fraunhofer diffraction. Let's go back to Eq. (10.41), rewritten as

$$E(Y, Z) = \frac{\mathcal{E}_A e^{i(\omega t - kR)}}{R} \iint_{\text{Aperture}} e^{ik(Yy+Zz)/R}\, dy\, dz \tag{11.61}$$

This formula refers to Fig. 10.18, which depicts an arbitrary diffracting aperture in the yz-plane upon which is incident a monochromatic plane wave. The quantity R is the distance from the center of the aperture to the output point where the field is $E(Y, Z)$. The source strength per unit area of the aperture is denoted by $\mathcal{E}_A$. We are talking about electric fields that are of course time-varying; the term $\exp i(\omega t - kR)$ just relates the phase of the net disturbance at the point (Y, Z) to that at the center of the aperture. The $1/R$ corresponds to the dropoff of field amplitude with distance from the aperture. The phase term in front of the integral is of little present concern, since we are interested in the relative amplitude distribution of the field, and it doesn't much matter what the resultant phase is at any particular output point. Thus if we limit ourselves to a small region of output space over which R is essentially constant, everything in front of the integral, with the exception of $\mathcal{E}_A$, can be lumped into a single constant.

The $\mathcal{E}_A$ has thus far been assumed to be invariant over the aperture, but that certainly need not be the case. Indeed, if the aperture were filled with a bumpy piece of dirty glass, the field emanating from each area element $dy\, dz$ could differ in both amplitude and phase. There would be nonuniform absorption, as well as a position-dependent optical path length through the glass, which would certainly affect the diffracted field distribution. The variations in $\mathcal{E}_A$, as well as the multiplicative constant, can be combined into a single complex quantity

$$\mathcal{A}(y, z) = \mathcal{A}_0(y, z)e^{i\phi(y,z)} \tag{11.62}$$

which we call the **aperture function**. The amplitude of the field over the aperture is described by $\mathcal{A}_0(y, z)$, while the point-to-point phase variation is represented by $\exp[i\phi(y, z)]$. Accordingly, $\mathcal{A}(y, z)\, dy\, dz$ is proportional to the diffracted field emanating from the differential source element $dy\, dz$. Consolidating this much, we can reformulate Eq. (11.61) more generally as

$$E(Y, Z) = \int_{-\infty}^{+\infty}\int \mathcal{A}(y, z)e^{ik(Yy+Zz)/R}\, dy\, dz \tag{11.63}$$

The limits on the integral can be extended to $\pm\infty$ because the aperture function is nonzero only over the region of the aperture.

It might be helpful to envision $dE(Y, Z)$ at a given point P as if it were a plane wave propagating in the direction of $\vec{\mathbf{k}}$ as in Fig. 11.30 and having an amplitude determined by $\mathcal{A}(y, z)\, dy\, dz$. To underscore the similarity between Eq. (11.63) and Eq. (11.14), let's define the *spatial frequencies* k_Y

*We should actually have used the real part of $\exp(-ik_0x)$ to start with in this derivation, since the transform of the complex exponential is different from the transform of $\cos k_0x$ and taking the real part afterward is insufficient. This is the same sort of difficulty one always encounters when forming products of complex exponentials. The final answer [Eq. (11.60)] should, in fact, contain an additional $\exp[-(k + k_0)^2/4a]$ term, as well as a multiplicative constant of $\frac{1}{2}$. This second term is usually negligible in comparison, however. Even so, had we used $\exp(+ik_0x)$ to start with [Eq. (11.59)], only the negligible term would have resulted! Using the complex exponential to represent the sine or cosine in this fashion is *rigorously incorrect*, albeit pragmatically common practice. As a short-cut, it should be indulged in only with the greatest caution!

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

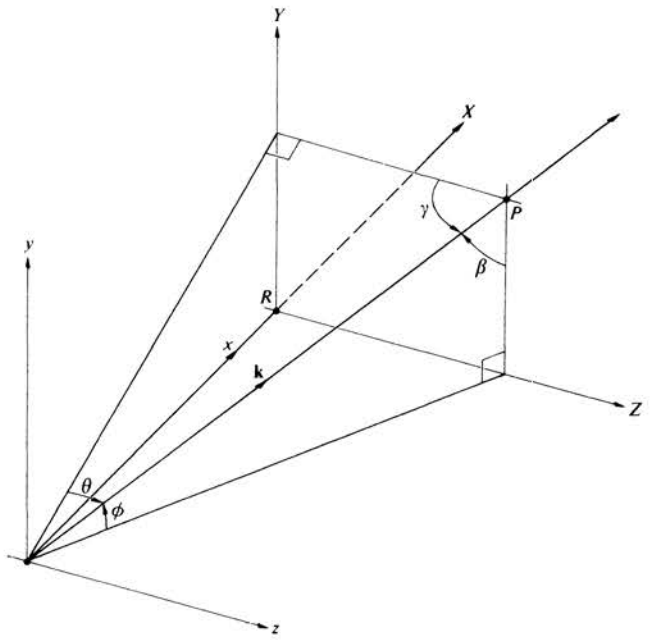


Figure 11.30 A bit of geometry.

and k_Z *as*

$$k_Y \equiv kY/R = k \sin \phi = k \cos \beta \tag{11.64}$$

and

$$k_Z \equiv kZ/R = k \sin \theta = k \cos \gamma \tag{11.65}$$

For each point on the image plane, there is a corresponding spatial frequency. The diffracted field can now be written as

$$E(k_Y, k_Z) = \int_{-\infty}^{+\infty}\int \mathcal{A}(y,z)e^{i(k_y y + k_z z)}\, dy\, dz \tag{11.66}$$

and we've arrived at the key point: ***the field distribution in the Fraunhofer diffraction pattern is the Fourier transform of the field distribution across the aperture (i.e., the aperture function)***. Symbolically, this is written as

$$E(k_Y, k_Z) = \mathcal{F}\{\mathcal{A}(y,z)\} \tag{11.67}$$

The field distribution in the image plane is the spatial-frequency spectrum of the aperture function. The inverse transform is then

$$\mathcal{A}(y,z) = \frac{1}{(2\pi)^2}\int_{-\infty}^{+\infty}\int E(k_Y, k_Z)e^{-i(k_y y + k_z z)}\, dk_Y\, dk_Z \tag{11.68}$$

that is,

$$\mathcal{A}(y,z) = \mathcal{F}^{-1}\{E(k_Y, k_Z)\} \tag{11.69}$$

As we have seen time and again, the more localized the signal, the more spread out is its transform—the same is true in two dimensions. The smaller the diffracting aperture, the larger the angular spread of the diffracted beam or, equivalently, the larger the spatial frequency bandwidth.

There is a minor issue that should be mentioned here. If we actually try to observe a Fraunhofer pattern on a distant screen (without a lens), what we get will only be an approximation; the true Fraunhofer pattern is formed in parallel light that doesn't converge at any finite distance. That doesn't generally cause any grief because what we do observe is the irradiance, and that is indistinguishable from the ideal distribution at great distances. Still, at any distant, but finite, location the diffracted electric-field distribution will differ in phase very slightly from the Fourier transform of the aperture function. Since we cannot even measure the electric field, the problem is not likely to be a practical one and we shall henceforth simply overlook it.

The Single Slit

As an illustration of the method, consider the long slit in the y-direction of Fig. 10.11, illuminated by a plane wave. Assuming that there are no phase or amplitude variations across the aperture, $\mathcal{A}(y, z)$ has the form of a square pulse (Fig. 7.23):

$$\mathcal{A}(y,z) = \begin{cases} \mathcal{A}_0 & \text{when } |z| \le b/2 \\ 0 & \text{when } |z| > b/2 \end{cases}$$

where $\mathcal{A}_0$ is no longer a function of y and z. If we take it as a one-dimensional problem,

$$E(k_Z) = \mathcal{F}\{\mathcal{A}(z)\} = \mathcal{A}_0\int_{z=-b/2}^{+b/2} e^{ik_z z}\, dz$$

$$E(k_Z) = \mathcal{A}_0 b \operatorname{sinc} k_Z b/2$$

With $k_Z = k \sin \theta$, this is precisely the form derived in Section 10.2.1. The far-field diffraction pattern of a rectangular aperture (Section 10.2.4) is the two-dimensional counterpart of the slit. With $\mathcal{A}(y, z)$ again equal to $\mathcal{A}_0$ over the aperture (Fig. 10.19),

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

$$E(k_Y, k_Z) = \mathcal{F}\{\mathcal{A}(y, z)\}$$

$$= \int_{y=-b/2}^{+b/2} \int_{z=-a/2}^{+a/2} \mathcal{A}_0 e^{i(k_Y y + k_Z z)}\, dy\, dz$$

hence,

$$E(k_Y, k_Z) = \mathcal{A}_0 ba \operatorname{sinc} \frac{bkY}{2R} \operatorname{sinc} \frac{akZ}{2R}$$

just as in Eq. (10.42), where ba is the area of the hole.

Young's Experiment: The Double Slit

In our first treatment of Young's Experiment (Section 9.3), we took the slits to be infinitesimally wide. The aperture function was then two symmetrical δ-pulses, and the corresponding idealized field amplitude in the diffraction pattern was the Fourier transform, namely, a cosine function. Squared, this yields the familiar cosine-squared irradiance distribution of Fig. 9.9. More realistically, each aperture actually has some finite shape, and the real diffraction pattern will never be quite so simple. Figure 11.31 shows the case in which the holes are actual slits. The aperture function, $g(x)$, is obtained by convolving the δ-function spikes, $h(x)$, that locate each slit with the rectangular pulse, $f(x)$, that corresponds to the particular opening. From the convolution theorem, the product of the transforms is the modulated cosine amplitude function representing the diffracted field as it appears on the image plane. Squaring that would produce the anticipated double-slit irradiance distribution shown in Fig. 10.14. The one-dimensional transform curves are plotted against k, but that's equivalent to plotting against image-space variables by means of Eq. (11.64). (The same reasoning applied to circular apertures yields the fringe pattern of Fig. 12.2.)

Three Slits

Looking at Fig. 11.13*d*, it should be clear that the transform of the array of three δ-functions in the diagram will generate a cosine that is raised by an amount proportional to the zero-frequency term, that is, the δ-function at the origin. When that delta function has twice the amplitude of the other two, the cosine is totally positive. Now suppose we have three ideally narrow parallel slits uniformly illuminated. The aperture func-

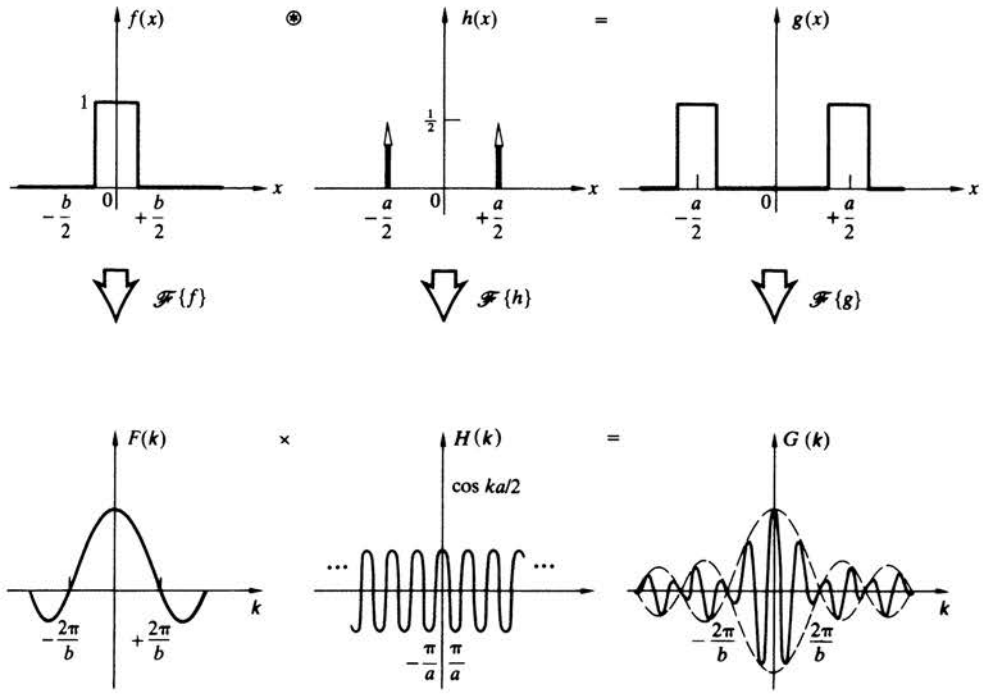


Figure 11.31 An illustration of the convolution theorem.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

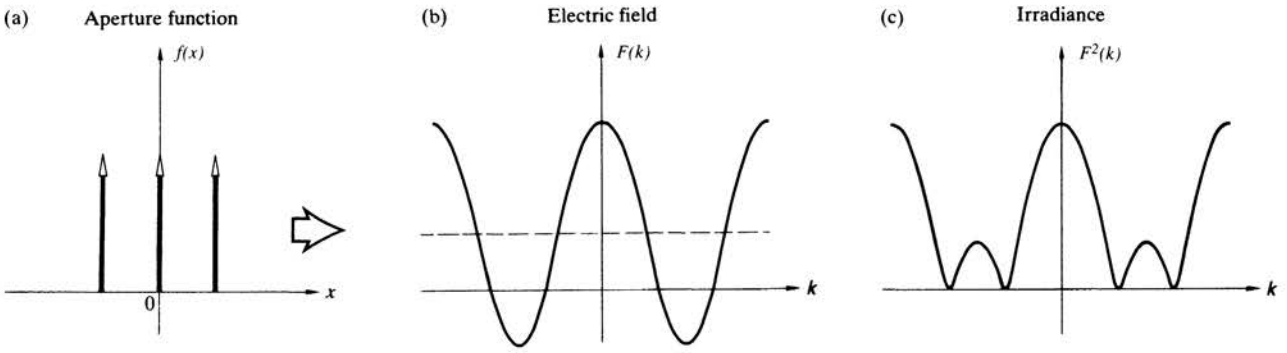


Figure 11.32 The Fourier transform of three equal δ-functions representing three slits.

tion corresponds to Fig. 11.32*a*, where the central δ-function is half its previous size. Accordingly, the cosine transform will drop one quarter of the way down, as indicated in Fig. 11.32*b*. This corresponds to the diffracted electric-field amplitude, and its square, Fig. 11.32*c*, is the three-slit irradiance pattern.

Apodization

The term **apodization** derives from the Greek α, to take away, and $\pi o \delta o \sigma$, meaning foot. It refers to the process of suppressing the secondary maxima (side lobes) or feet of a diffraction pattern. In the case of a circular pupil (Section 10.2.5), the diffraction pattern is a central spot surrounded by concentric rings. The first ring has a flux density of 1.75% that of the central peak—it's small, but it can be troublesome. About 16% of the light incident on the image plane is distributed in the ring system. The presence of these side lobes can diminish the resolving power of an optical system to a point where apodization is called for, as is often the case in astronomy and spectroscopy. For example, the star Sirius, which appears as the brightest star in the sky (it's in the constellation *Canis Major*—the big dog), is actually one of a binary system. It's accompanied by a faint white dwarf as they both orbit about their mutual center of mass. Because of the tremendous difference in brightness (10^4 to 1), the image of the faint companion, as viewed with a telescope, is generally completely obscured by the side lobes of the diffraction pattern of the main star.

Apodization can be accomplished in several ways, for example, by altering the shape of the aperture or its transmission characteristics.* We already know from Eq. (11.66) that the diffracted field distribution is the transform of $\mathcal{A}(y, z)$. Thus we could effect a change in the side lobes by altering $\mathcal{A}_0(y, z)$ or $\phi(y, z)$. Perhaps the simplest approach is the one in which only $\mathcal{A}_0(y, z)$ is manipulated. This can be accomplished physically by covering the aperture with a suitably coated flat glass plate (or coating the objective lens itself). Suppose that the coating becomes increasingly opaque as it goes radially out from the center (in the yz-plane) toward the edges of a circular pupil. The transmitted field will correspondingly decrease off-axis until it is made to become negligible at the periphery of the aperture. In particular, imagine that this dropoff in amplitude follows a Gaussian curve. Then $\mathcal{A}_0(y, z)$ is a Gaussian function, as is its transform $E(Y, Z)$, and consequently the ring system vanishes. Even though the central peak is broadened, the side lobes are indeed suppressed (Fig. 11.33).

Another rather heuristic but appealing way to look at the process is to realize that the higher spatial frequency contributions go into sharpening up the details of the function being synthesized. As we saw earlier in one dimension (Fig. 7.29), the high frequencies serve to fill in the corners of the square pulse. In the same way, since $\mathcal{A}(y, z) = \mathcal{F}^{-1}\{E(k_Y, k_Z)\}$, sharp edges on the aperture necessitate the presence of appreciable contributions of high spatial frequency in the diffracted field. It follows that making $\mathcal{A}_0(y, z)$ fall off gradually will reduce

*For an extensive treatment of the subject, see P. Jacquinot and B. Roizen-Dossier, "Apodization," in Vol. III of *Progress in Optics*.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

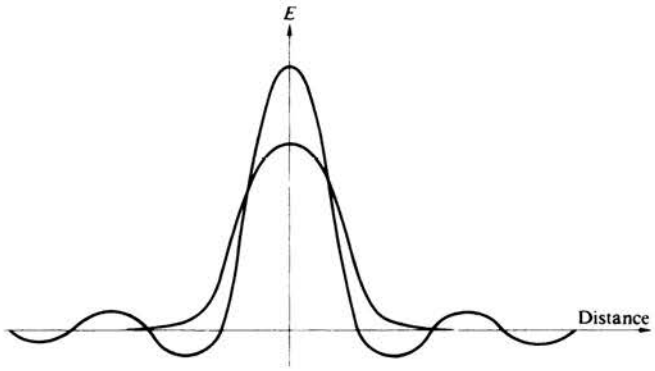


Figure 11.33 An Airy pattern compared with a Gaussian.

these high frequencies, which in turn is manifest in a suppression of the side lobes.

Apodization is one aspect of the more encompassing technique of *spatial filtering*, which is discussed in an extensive yet nonmathematical treatment in Chapter 13.

The Array Theorem

Generalizing some of our previous ideas to two dimensions, imagine that we have a screen containing N identical holes, as in Fig. 11.34. In each aperture, at the same relative position, we locate a point O_1, O_2,..., O_N at (y_1, z_1), (y_2, z_2),..., (y_N, z_N), respectively. Each of these, in turn, fixes the origin of a local coordinate system (y', z'). Thus a point (y', z') in the local frame of the jth aperture has coordinates $(y_j + y', z_j + z')$ in the (y, z)-system. Under coherent monochromatic illumination, the resulting Fraunhofer diffraction field $E(Y, Z)$ at some point P on the image plane will be a superposition of the individual fields at P arising from each separate aperture; in other words,

$$E(Y,Z) = \sum_{j=1}^{N} \int_{-\infty}^{+\infty}\int \mathcal{A}_I(y',z')e^{ik[Y(y_j+y')+Z(z_j+z')]/R}dy'\,dz' \tag{11.70}$$

or

$$E(Y,Z) = \int_{-\infty}^{+\infty}\int \mathcal{A}_I(y',z')e^{ik(Yy'+Zz')/R}dy'\,dz' \times \sum_{j=1}^{N} e^{ik(Yy_j+Zz_j)/R} \tag{11.71}$$

where $\mathcal{A}_I(y', z')$ is the individual aperture function applicable to each hole. This can be recast, using Eqs. (11.64) and (11.65), as

$$E(k_Y, k_Z) = \int_{-\infty}^{+\infty}\int \mathcal{A}_I(y',z')e^{i(k_Yy'+k_Zz')}\,dy'\,dz' \times \sum_{j=1}^{N} e^{i(k_Yy_j)}e^{i(k_Zz_j)} \tag{11.72}$$

Notice that the integral is the Fourier transform of the individual aperture function, while the sum is the transform [Eq. (11.42)] of an array of delta functions

$$A_\delta = \sum_j \delta(y - y_j)\delta(z - z_j) \tag{11.73}$$

Inasmuch as $E(k_Y, k_Z)$ itself is the transform $\mathcal{F}\{\mathcal{A}(y, z)\}$ of the total aperture function for the entire array, we have

$$\mathcal{F}\{\mathcal{A}(y,z)\} = \mathcal{F}\{\mathcal{A}_I(y',z')\}\cdot\mathcal{F}\{A_\delta\} \tag{11.74}$$

This equation is a statement of the **array theorem**, which says that ***the field distribution in the Fraunhofer diffraction pattern of an array of similarly oriented identical apertures equals the Fourier transform of an individual aperture function (i.e., its diffracted field distribution) multiplied by the pattern that would result from a set of point sources arrayed in the same configuration (which is the transform of A_δ).***

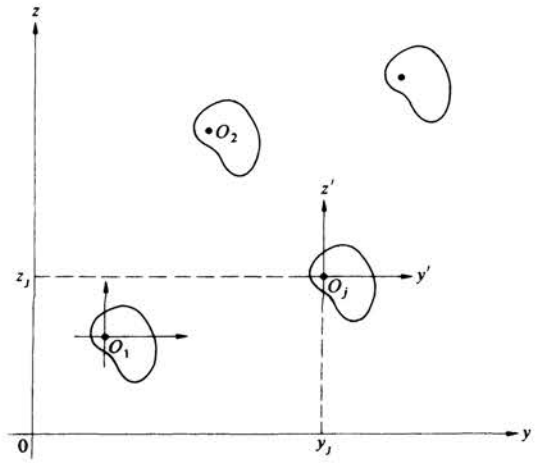


Figure 11.34 Multiple-aperture configuration.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

This can be seen from a slightly different point of view. The total aperture function may be formed by convolving the individual aperture function with an appropriate array of delta functions, each sitting at one of the coordinate origins (y_1, z_1), (y_2, z_2), and so on. Hence

$$\mathcal{A}(y, z) = \mathcal{A}_1(y', z') \circledast A_\delta \tag{11.75}$$

whereupon the array theorem follows directly from the convolution theorem [Eq. (11.53)].

As a simple example, imagine that we again have Young's Experiment with two slits along the y-direction, of width b and separation a. The individual aperture function for each slit is a step function,

$$\mathcal{A}_1(z') = \begin{cases} \mathcal{A}_{10} & \text{when } |z'| \le b/2 \\ 0 & \text{when } |z'| > b/2 \end{cases}$$

and so

$$\mathcal{F}\{\mathcal{A}_1(z')\} = \mathcal{A}_{10} b \text{ sinc } k_Z b/2$$

With the slits located at $z = \pm a/2$,

$$A_\delta = \delta(z - a/2) + \delta(z + a/2)$$

and from Eq. (11.43)

$$\mathcal{F}\{A_\delta\} = 2\cos k_Z a/2$$

Thus

$$E(k_Z) = 2\mathcal{A}_{10} b \text{ sinc}\left(\frac{k_Z b}{2}\right)\cos\left(\frac{k_Z a}{2}\right)$$

which is the same conclusion arrived at earlier (Fig. 11.31). The irradiance pattern is a set of *cosine-squared* interference fringes modulated by a *sinc-squared* diffraction envelope.

11.3.4 Spectra and Correlation

Parseval's Formula

Suppose that $f(x)$ is a pulse of finite extent, and $F(k)$ is its Fourier transform [Eq. (11.5)]. Thinking back to Section 7.8, we recognize the function $F(k)$ as the amplitude of the spatial frequency spectrum of $f(x)$. And $F(k)\,dk$ then connotes the amplitude of the contributions to the pulse within the frequency range from k to $k + dk$. Hence it seems that $|F(k)|$ serves as a spectral amplitude density, and its square, $|F(k)|^2$, should be proportional to the energy per unit spatial frequency interval. Similarly, in the time domain, if $f(t)$ is a radiated electric field, $|f(t)|^2$ is proportional to the radiant flux or power, and the total emitted energy is proportional to $\int_0^\infty |f(t)|^2 dt$. With $F(\omega) = \mathcal{F}\{f(t)\}$ it appears that $|F(\omega)|^2$ must be a measure of the radiated energy per unit temporal frequency interval. To be a bit more precise, let's evaluate $\int_{-\infty}^{+\infty} |f(t)|^2 dt$ in terms of the appropriate Fourier transforms. Inasmuch as $|f(t)|^2 = f(t)f^*(t) = f(t)\cdot[\mathcal{F}^{-1}\{\mathcal{F}(\omega)\}]^*$,

$$\int_{-\infty}^{+\infty} |f(t)|^2\,dt = \int_{-\infty}^{+\infty} f(t)\left[\frac{1}{2\pi}\int_{-\infty}^{+\infty} F^*(\omega)e^{+i\omega t}\,d\omega\right]dt$$

Interchanging the order of integration, we obtain

$$\int_{-\infty}^{+\infty} |f(t)|^2\,dt = \frac{1}{2\pi}\int_{-\infty}^{+\infty} F^*(\omega)\left[\int_{-\infty}^{+\infty} f(t)e^{i\omega t}\,dt\right]d\omega$$

and so

$$\int_{-\infty}^{+\infty} |f(t)|^2 dt = \frac{1}{2\pi}\int_{-\infty}^{+\infty} |F(\omega)|^2\,d\omega \tag{11.76}$$

where $|F(\omega)|^2 = F^*(\omega)F(\omega)$. This is *Parseval's formula*. As expected, the total energy is proportional to the area under the $|F(\omega)|^2$ curve, and consequently $|F(\omega)|^2$ is sometimes called the **power spectrum** or *spectral energy distribution*. The corresponding formula for the space domain is

$$\int_{-\infty}^{+\infty} |f(x)|^2 dx = \frac{1}{2\pi}\int_{-\infty}^{+\infty} |F(k)|^2\,dk \tag{11.77}$$

The Lorentzian Profile

As an indication of the manner in which these ideas are applied in practice, consider the damped harmonic wave $f(t)$ at $x = 0$ depicted in Fig. 11.35. Here

$$f(t) = \begin{cases} 0 & \text{from } t = -\infty \text{ to } t = 0 \\ f_0 e^{-t/2\tau}\cos\omega_0 t & \text{from } t = 0 \text{ to } t = +\infty \end{cases}$$

The negative exponential dependence arises, quite generally, whenever the rate-of-change of a quantity depends on its instantaneous value. In this case, we might suppose that the power radiated by an atom varies as $(e^{-t/\tau})^{1/2}$. In any event, τ

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

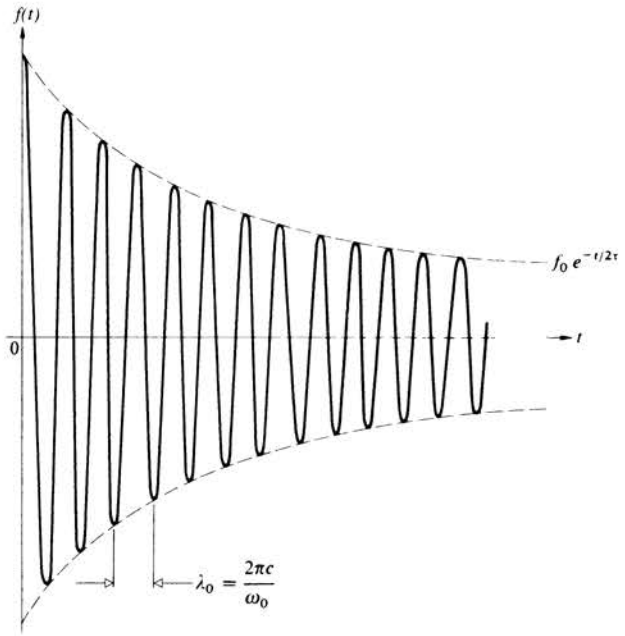


Figure 11.35 A damped harmonic wave.

is known as the time constant of the oscillation, and $\tau^{-1} = \gamma$ is the damping constant. The transform of $f(t)$ is

$$F(\omega) = \int_0^\infty (f_0 e^{-t/2\tau} \cos \omega_0 t) e^{i\omega t}\, dt \qquad (11.78)$$

The evaluation of this integral is explored in the problems. One finds on performing the calculation that

$$F(\omega) = \frac{f_0}{2}\left[\frac{1}{2\tau} - i(\omega + \omega_0)\right]^{-1} + \frac{f_0}{2}\left[\frac{1}{2\tau} - i(\omega - \omega_0)\right]^{-1}$$

When $f(t)$ is the radiated field of an atom, τ denotes the *lifetime* of the excited state (from around 1.0 ns to 10 ns). Now if we form the power spectrum $F(\omega)F^*(\omega)$, it will be composed of two peaks centered on $\pm\,\omega_0$ and thus separated by $2\omega_0$. At optical frequencies where $\omega_0 >> \gamma$, these will be both narrow and widely spaced, with essentially no overlap. The shape of these peaks is determined by the transform of the modulation envelope in Fig. 11.35, that is, a negative exponential. The location of the peaks is fixed by the frequency of the modulated cosine wave, and the fact that there are two such peaks is a reflection of the spectrum of the cosine in this symmetrical frequency representation (Section 7.8). To determine the observable spectrum from $F(\omega)F^*(\omega)$, we need only consider the positive frequency term, namely,

$$|F(\omega)|^2 = \frac{f_0^2}{\gamma^2}\,\frac{\gamma^2/4}{(\omega - \omega_0)^2 + \gamma^2/4} \qquad (11.79)$$

This has a maximum value of f_0^2/γ^2 at $\omega = \omega_0$, as shown in Fig. 11.36. At the half-power points $(\omega - \omega_0) = \pm\gamma/2$, $|F(\omega)|^2 = f_0^2/2\gamma^2$, which is half its maximum value. The width of the spectral line between these points is equal to γ.

The curve given by Eq. (11.79) is known as the ***resonance*** or ***Lorentz profile***. The frequency bandwidth arising from the finite duration of the excited state is called the ***natural linewidth***.

If the radiating atom suffers a collision, it can lose energy and thereby further shorten the duration of emission. The frequency bandwidth increases in the process, which is known as *Lorentz broadening*. Here again, the spectrum is found to have a Lorentz profile. Furthermore, because of the random thermal motion of the atoms in a gas, the frequency bandwidth will be increased via the Doppler effect. *Doppler broadening*, as it is called, results in a Gaussian spectrum (Section 7.10). The Gaussian drops more slowly in the immediate vicinity of ω_0 and then more quickly away from it than does the Lorentzian profile. These effects can be combined mathematically to yield a single spectrum by convolving the Gaussian and Lorentzian functions. In a low-pressure gaseous discharge, the Gaussian profile is by far the wider and generally predominates.

Autocorrelation and Cross-Correlation

Let's now go back to the derivation of Parseval's formula and follow it through again, this time with a slight modification.

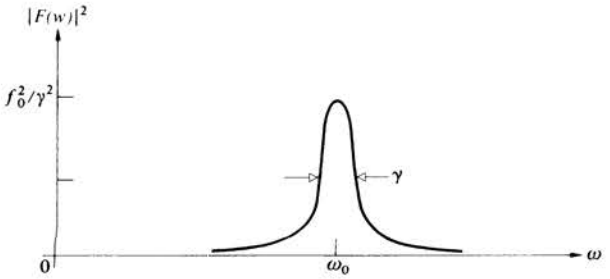


Figure 11.36 The resonance or Lorentz profile.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

We wish to evaluate $\int_{-\infty}^{+\infty} f(t+\tau)f^*(t)\,dt$, using much the same approach as before. Thus, if $F(\omega) = \mathcal{F}\{f(t)\}$,

$$\int_{-\infty}^{+\infty} f(t+\tau)f^*(t)\,dt = \int_{-\infty}^{+\infty} f(t+\tau) \times \left[\frac{1}{2\pi}\int_{-\infty}^{+\infty} F^*(\omega)e^{+i\omega t}\,d\omega\right] dt \tag{11.80}$$

Changing the order of integration, we obtain

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty} F^*(\omega)\left[\int_{-\infty}^{+\infty} f(t+\tau)e^{i\omega t}\,dt\right] d\omega = \frac{1}{2\pi}\int_{-\infty}^{+\infty} F^*(\omega)\mathcal{F}\{f(t+\tau)\}d\omega$$

To evaluate the transform within the last integral, notice that

$$f(t+\tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} F(\omega)e^{-i\omega(t+\tau)}\,d\omega$$

by a change of variable in Eq. (11.9). Hence,

$$f(t+\tau) = \mathcal{F}^{-1}\{F(\omega)e^{-i\omega\tau}\}$$

so as discussed earlier, $\mathcal{F}\{f(t+\tau)\} = F(\omega)e^{-i\omega\tau}$, Eq. (11.80) becomes

$$\int_{-\infty}^{+\infty} f(t+\tau)f^*(t)\,dt = \frac{1}{2\pi}\int_{-\infty}^{+\infty} F^*(\omega)F(\omega)e^{-i\omega\tau}\,d\omega \tag{11.81}$$

and both sides are functions of the parameter τ. The left-hand side of this formula is said to be the **autocorrelation** of $f(t)$, denoted by

$$c_{ff}(\tau) \equiv \int_{-\infty}^{+\infty} f(t+\tau)f^*(t)\,dt \tag{11.82}$$

which is often written symbolically as $f(t) \odot f^*(t)$. If we take the transform of both sides, Eq. (11.81) then becomes

$$\mathcal{F}\{c_{ff}(\tau)\} = |F(\omega)|^2 \tag{11.83}$$

This is a form of the ***Wiener–Khintchine theorem***. It allows for determination of the spectrum by way of the autocorrelation of the generating function. The definition of $c_{ff}(\tau)$ applies when the function has finite energy. When it doesn't, things will have to be changed slightly. The integral can also be restated as

$$c_{ff}(\tau) = \int_{-\infty}^{+\infty} f(t)f^*(t-\tau)\,dt \tag{11.84}$$

by a simple change of variable ($t+\tau$ to t). Similarly, the **cross-correlation** of the functions $f(t)$ and $h(t)$ is defined as

$$c_{fh}(\tau) = \int_{-\infty}^{+\infty} f^*(t)h(t+\tau)\,dt \tag{11.85}$$

Correlation analysis is essentially a means for comparing two signals in order to determine the degree of similarity between them. In autocorrelation the original function is displaced in time by an amount τ, the product of the displaced and undisplaced versions is formed, and the area under that product (corresponding to the degree of overlap) is computed by means of the integral. The autocorrelation function, $c_{ff}(\tau)$, provides the result that will be obtained in such a process for all values of τ. One reason for doing such a thing, for example, is to extract a signal from a background of random noise.

To see how the business works step by step, let's take the autocorrelation of a simple function, such as $A\sin(\omega t+\varepsilon)$, shown in Fig. 11.37. In each part of the diagram the function is shifted by a value of τ, the product $f(t)\cdot f(t+\tau)$ is formed, and then the area under that product function is computed and plotted in part (e). Notice that the process is indifferent to the value of ε. The final result is $c_{ff}(\tau) = \frac{1}{2}A^2\cos\omega\tau$, where this function unfolds through one cycle as τ goes through 2π, so it has the same frequency as $f(t)$. Accordingly, if we had a process for generating the autocorrelation, we could reconstruct from that both the original amplitude A and the angular frequency ω.

Assuming the functions to be real, we can rewrite $c_{fh}(\tau)$ as

$$c_{fh}(\tau) = \int_{-\infty}^{+\infty} f(t)h(t+\tau)\,dt \tag{11.86}$$

which is obviously similar to the expression for the convolution of $f(t)$ and $h(t)$. Equation (11.86) is written symbolically as $c_{fh}(\tau) = f(t) \odot h(t)$. Indeed, if either $f(t)$ or $h(t)$ is even, then $f(t) \circledast h(t) = f(t) \odot h(t)$, as we shall see by example presently. Recall that the convolution flips one of the functions over and then sums up the overlap area (Fig. 11.21), that is, the area under the product curve. In contrast, the correlation sums up the overlap without flipping the function, and thus if the function is even, $f(t) = f(-t)$, it isn't changed by being flipped (or folded about the symmetry axis), and the two integrands are

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

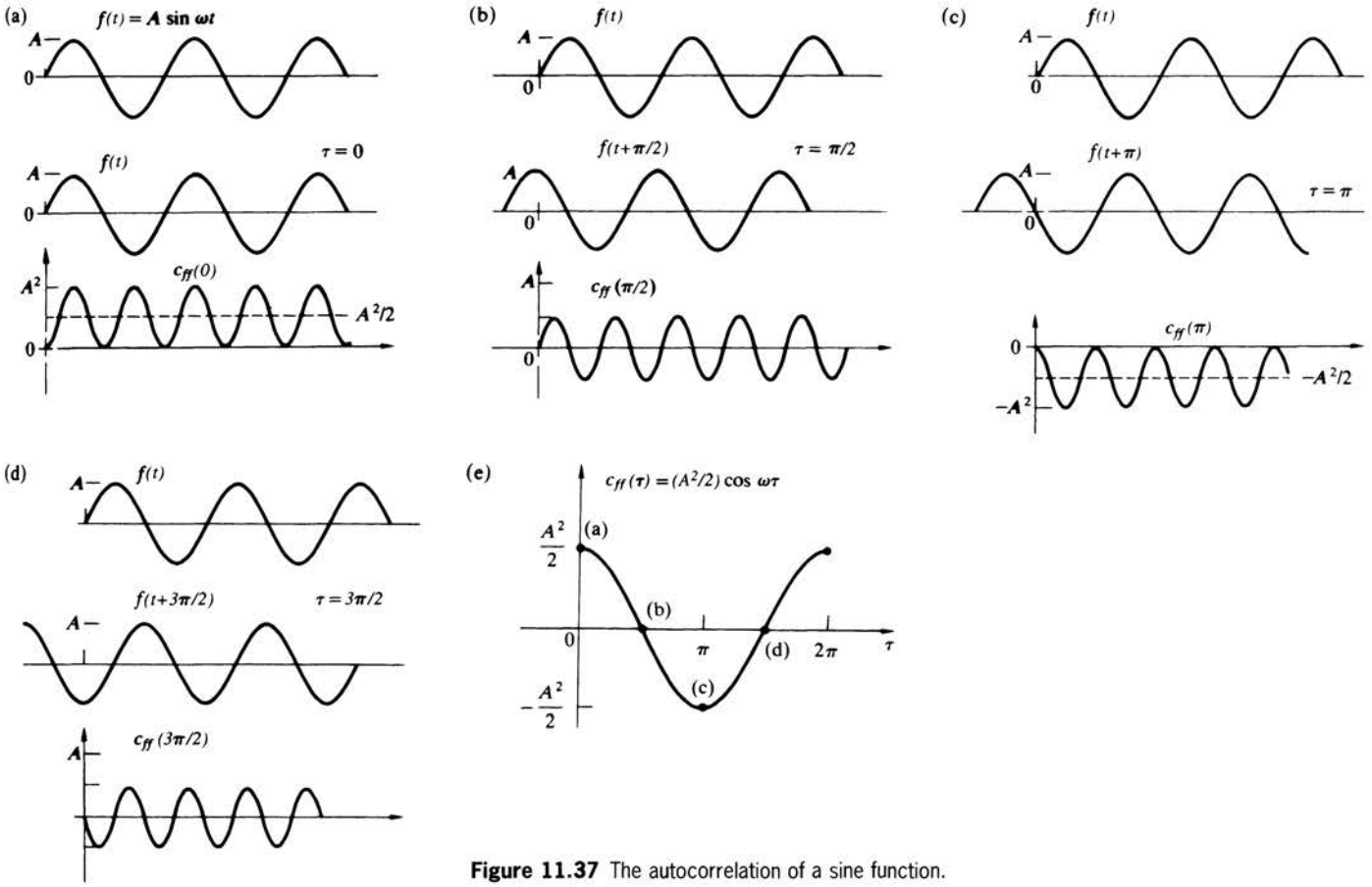


Figure 11.37 The autocorrelation of a sine function.

identical. For this to obtain, either function must be even, since $f(t) \circledast h(t) = h(t) \circledast f(t)$. The autocorrelation of a square pulse is therefore equal to the convolution of the pulse with itself, which yields a triangular signal, as in Fig. 11.24. This same conclusion follows from Eq. (11.83) and Fig. 11.26. The transform of a square pulse is a sinc function, so that the power spectrum varies as $\text{sinc}^2 u$. The inverse transform of $|F(\omega)|^2$, that is, $\mathcal{F}^{-1}\{\text{sinc}^2 u\}$, is $c_{ff}(\tau)$, which as we have seen, is again a triangular pulse (Fig. 11.38).

It is clearly possible for a function to have infinite energy [Eq. (11.76)] over an integration ranging from $-\infty$ to $+\infty$ and yet still have a finite *average power*

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} |f(t)|^2 \, dt$$

Accordingly, we will define a correlation that is divided by the integration interval:

$$C_{fh}(\tau) \equiv \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} f(t)h(t+\tau) \, dt \qquad (11.87)$$

For example, if $f(t) = A$ (i.e., a constant), its autocorrelation

$$C_{ff}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} (A)(A) \, dt = A^2$$

and the power spectrum, which is the transform of the autocorrelation, becomes

$$\mathcal{F}\{C_{ff}(\tau)\} = A^2 2\pi\delta(\omega)$$

a single impulse at the origin ($\omega = 0$), which is sometimes

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

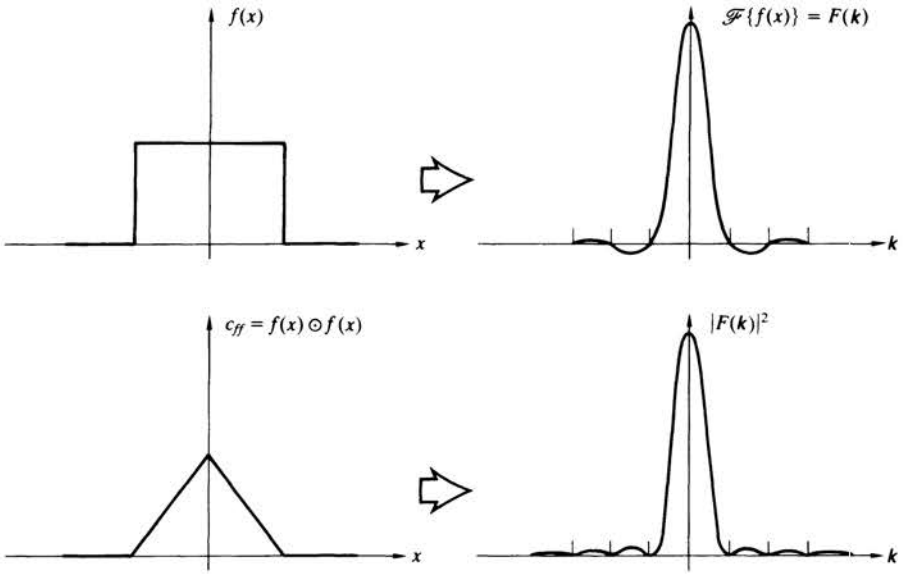


Figure 11.38 The square of the Fourier transform of the rectangular pulse $f(x)$ (i.e., $|F(k)|^2$) equals the Fourier transform of the autocorrelation of $f(x)$.

referred to as a *dc*-term. Notice that $C_{fh}(\tau)$ can be thought of as the time average of a product of two functions, one of which is shifted by an interval τ. In the next chapter, expressions of the form $\langle f^*(t)h(t+\tau)\rangle$ arise as coherence functions relating electric fields. They are also quite useful in the analysis of noise problems, for example, film grain noise.

We can obviously reconstruct a function from its transform, but once the transform is squared, as in Eq. (11.83), we lose information about the signs of the frequency contributions, that is, their relative phases. In the same way, the autocorrelation of a function contains no phase information and is not unique. To see this more clearly, imagine we have a number of harmonic functions of different amplitude and frequency. If their relative phases are altered, the resultant function changes, as does its transform, but in all cases the amount of energy available at any frequency must be constant. Thus, whatever the form of the resultant profile, its autocorrelation is unaltered. It is left as a problem to show analytically that when $f(t) = A\sin(\omega t + \varepsilon)$, $C_{ff}(\tau) = (A^2/2)\cos\omega\tau$, which confirms the loss of phase information.

Figure 11.39 shows a means of optically correlating two two-dimensional spatial functions. Each of these signals is represented as a point-by-point variation in the irradiance transmission property of a photographic transparency (T_1 and T_2). For relatively simple signals opaque screens with appropriate apertures could serve instead of transparencies (e.g., for square pulses).* The irradiance at any point P on the image is due to a focused bundle of parallel rays that has traversed both transparencies. The coordinates of P, $(\theta f, \varphi f)$, are fixed by the orientation of the ray bundle, that is, the angles θ and φ. If the transparencies are identical, a ray passing through any point

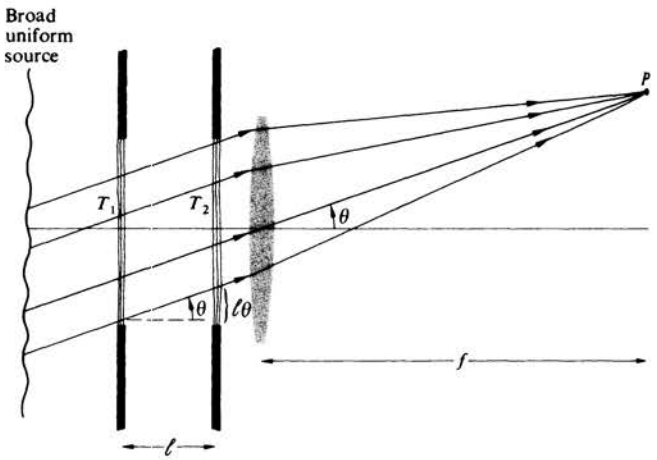


Figure 11.39 Optical correlation of two functions.

*See L. S. G. Kovasznay and A. Arman, *Rev. Sci. Instr.* **28**, 793 (1958), and D. McLachlan, Jr., *J. Opt. Soc. Am.* **52**, 454 (1962).

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

(x, y) on the first film with a transmittance $g(x, y)$ will pass through a corresponding point $(x + X, y + Y)$ on the second film where the transmittance is $g(x + X, y + Y)$. The shifts in coordinate are given by $X = \ell\theta$ and $Y = \ell\varphi$, where ℓ is the separation between the transparencies. The irradiance at P is therefore proportional to the autocorrelation of $g(x, y)$, that is,

$$c_{ff}(X, Y) = \int_{-\infty}^{+\infty}\int g(x, y)g(x + X, y + Z)\, dx\, dy \tag{11.88}$$

and the entire flux-density pattern is called a *correlogram*. If the transparencies are different, the image is of course representative of the cross-correlation of the functions. Similarly, if one of the transparencies is rotated by 180° with respect to the other, the convolution can be obtained (see Fig. 11.25).

Before moving on, let's make sure that we actually do have a good physical feeling for the operation performed by the correlation functions. Accordingly, suppose we have a random noise-like signal (e.g., a fluctuating irradiance at a point in space or a time-varying voltage or electric field), as in Fig. 11.40*a*. The autocorrelation of $f(t)$ in effect compares the function with its value at some other time, $f(t + \tau)$. For example, with $\tau = 0$ the integral runs along the signal in time, summing up and averaging the product of $f(t)$ and $f(t + \tau)$; in this case it's simply $f^2(t)$. Since at each value of t, $f^2(t)$ is positive, $C_{ff}(0)$ will be a comparatively large number. On the other hand, when the noise is compared with itself shifted by an amount $+\tau_1$, $C_{ff}(\tau_1)$ will be somewhat reduced. There will be points in time where $f(t)f(t + \tau_1)$ is positive and other points where it will be negative, so that the value of the integral drops off (Fig. 11.40*b*). In other words, by shifting the signal with respect to itself, we have reduced the point-by-point similarity that previously ($\tau = 0$) occurred at any instant. As this shift τ increases, what little correlation existed quickly vanishes, as depicted in Fig. 11.40*c*. We can assume from the fact that the autocorrelation and the power spectrum form a Fourier transform pair [Eq. (11.83)] that the broader the frequency bandwidth of the noise, the narrower the autocorrelation. Thus for wide-bandwidth noise even a slight shift markedly reduces any similarity between $f(t)$ and $f(t + \tau)$. Furthermore, if the signal comprises a random distribution of rectangular pulses, we can see intuitively that the similarity we spoke of earlier persists for a time commensurate with the width of the pulses. The wider (in time) the pulses are, the more slowly the correlation decreases as τ increases. But this is equivalent to saying that reducing the signal bandwidth

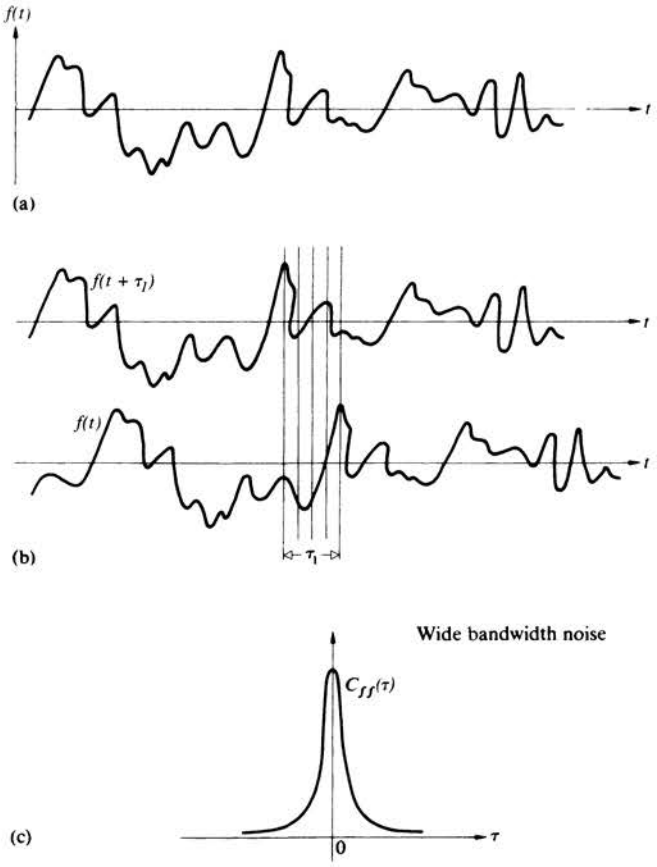


Figure 11.40 A signal $f(t)$ and its autocorrelation.

broadens $C_{ff}(\tau)$. All of this is in keeping with our previous observation that the autocorrelation tosses out any phase information, which in this case would correspond to the locations in time of the random pulses. Clearly, $C_{ff}(\tau)$ shouldn't be affected by the position of the pulses along t.

In very much the same way, the cross-correlation is a measure of the similarity between two different waveforms, $f(t)$ and $h(t)$, as a function of the relative time shift τ. Unlike the autocorrelation, there is now nothing special about $\tau = 0$. Once again, for each value of τ we average the product $f(t)h(t + \tau)$ to get $C_{fh}(\tau)$ via Eq. (11.87). For the functions shown in Fig. 11.41, $C_{fh}(\tau)$ would have a positive peak at $\tau = \tau_1$.

Since the 1960s a great deal of effort has gone into the development of optical processors that can rapidly analyze

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

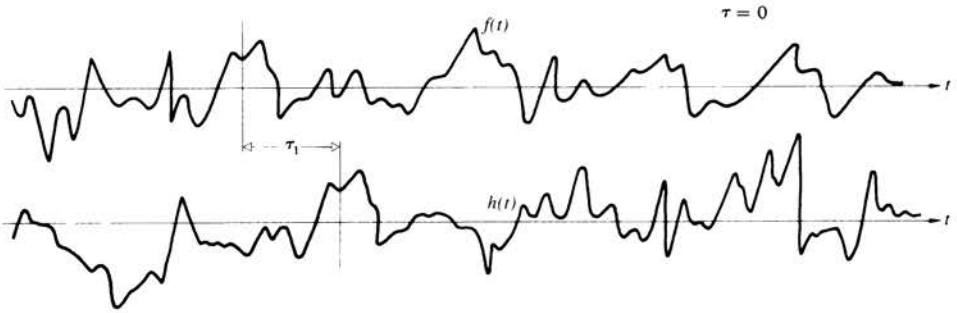


Figure 11.41 The cross-correlation of $f(t)$ and $h(t)$.

pictorial data. The potential uses range from comparing fingerprints to scanning documents for words or phrases; from screening aerial reconnaissance pictures to creating terrain-following guidance systems for missiles. An example of this kind of *optical pattern recognition*, accomplished using correlation techniques, is shown in Fig. 11.42. The input signal $f(x, y)$ depicted in photograph (*a*) is a broad view of some region that is to be searched for a particular group of structures [photograph (*b*)] isolated as the reference signal $h(x, y)$. Of course, that small frame is easy enough to scan directly by eye, so to make things more realistic, imagine the input to be a few hundred feet of reconnaissance film. The result of optically correlating these two signals is displayed in photograph (*c*), where we immediately see, from the correlation peak (i.e., the spike of light), that indeed the desired group of structures is in the input picture, and moreover its location is marked by the peak.

11.3.5 Transfer Functions

An Introduction to the Concepts

Until recent times, the traditional means of determining the quality of an optical element or system of elements was to evaluate its limit of resolution. The greater the resolution, the better the system was presumed to be. In the spirit of this approach, one might train an optical system on a resolution target consisting, for instance, of a series of alternating light

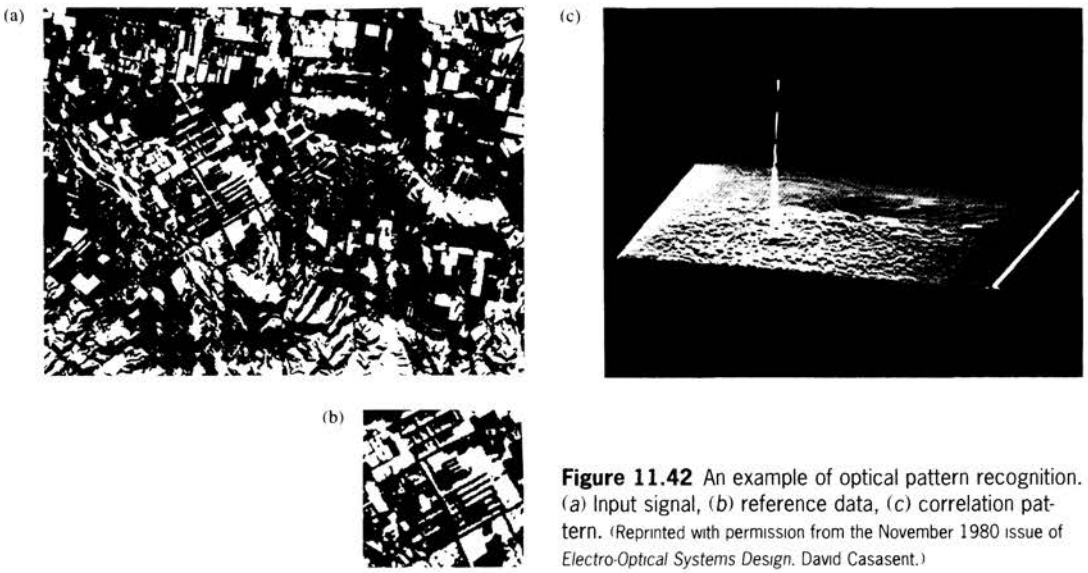


Figure 11.42 An example of optical pattern recognition. (*a*) Input signal, (*b*) reference data, (*c*) correlation pattern. (Reprinted with permission from the November 1980 issue of *Electro-Optical Systems Design*. David Casasent.)

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

and dark parallel rectangular bars. We have already seen that an object point is imaged as a smear of light described by the point-spread function $\mathcal{S}(Y, Z)$, as in Fig. 11.18. Under incoherent illumination, these elementary flux-density patterns overlap and add linearly to create the final image. The one-dimensional counterpart is the *line-spread function* $\mathcal{S}(Z)$, which corresponds to the flux-density distribution across the image of a geometrical line source having infinitesimal width (Fig. 11.43). Because even an ideally perfect system is limited by diffraction effects, the image of a resolution target (Fig. 11.44) will be somewhat blurred (see Fig. 11.20). Thus, as the width of the bars on the target is made narrower, a limit will be reached where the fine-line structure (akin to a *Ronchi ruling*) will no longer be discernible—this then is the resolution limit of the system. We can think of it as a spatial frequency cutoff where each bright and dark bar pair constitutes one cycle on the object (a common measure of which is *line pairs per mm*).

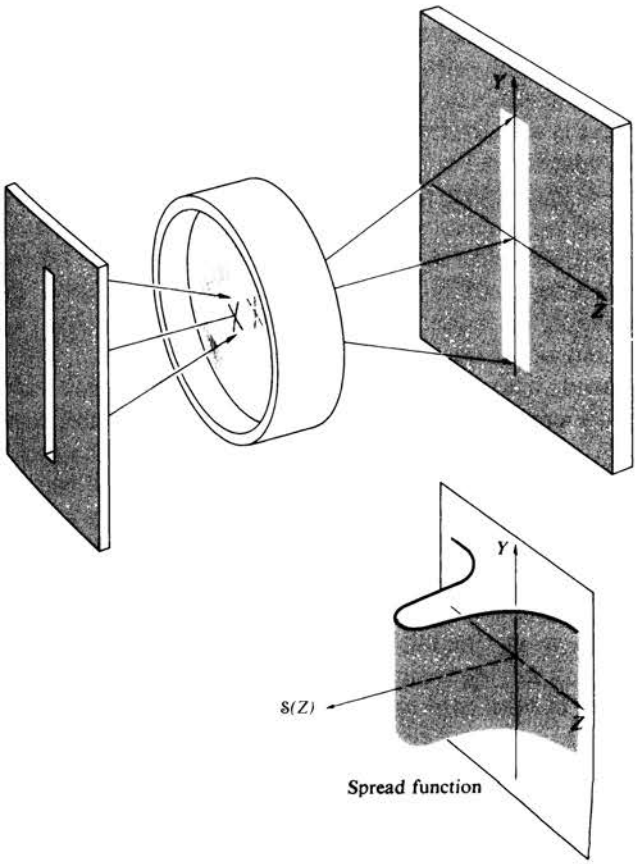


Figure 11.43 The line-spread function.

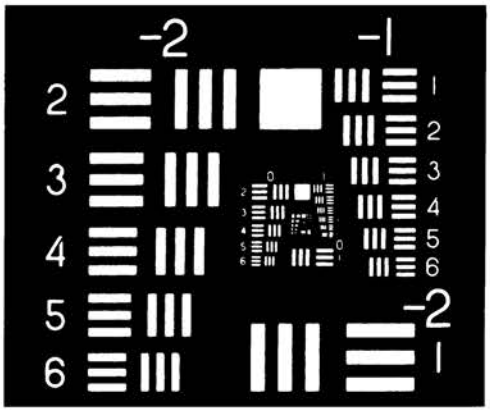


Figure 11.44 A bar target resolution chart.

An obvious analogy which underscores the shortcomings of this approach would be to evaluate a high-fidelity sound system simply on the basis of its upper-frequency cutoff. The limitations of this scheme became quite apparent with the introduction of detectors such as the plumbicon, image orthicon, and vidicon. These tubes have a relatively coarse scanning raster, which fixes the resolution limit of the lens-tube system at a fairly low spatial frequency. Accordingly, it would seem reasonable to design the optics preceding such detectors so that it provided the most contrast over this limited frequency range. It would clearly be unnecessary and perhaps, as we shall see, even detrimental to select a mating lens system merely because of its own high limit of resolution. Evidently, it would be more helpful to have some figure of merit applicable to the entire operating frequency range.

We have already represented the object as a collection of point sources, each of which is imaged as a point-spread function by the optical system, and that patch of light is then convolved into the image. Now we approach the problem of image analysis from a different, though related, perspective. Consider the object to be the source of an input lightwave, which itself is made up of plane waves. These travel off in specific directions corresponding, via Eqs. (11.64) and (11.65), to particular values of spatial frequency. How does the system modify the amplitude and phase of each plane wave as it transfers it from object to image?

A highly useful parameter in evaluating the performance of a system is the **contrast** or **modulation**, defined by

$$\text{Modulation} \equiv \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \tag{11.89}$$

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

As a simple example, suppose the input is a cosinusoidal irradiance distribution arising from an incoherently illuminated transparency (Fig. 11.45). Here the output is also a cosine, but one that's somewhat altered. The modulation, which corresponds to the amount the function varies about its mean value divided by that mean value, is a measure of how readily the fluctuations will be discernible against the dc background. For the input the modulation is a maximum of 1.0, but the output modulation is only 0.17. This is only the response of our hypothetical system to essentially one spatial frequency input—it would be nice to know what it does at all such frequencies. Moreover, here the input modulation was 1.0, and the comparison with the output was easy. In general it will not be 1.0, and so we define *the ratio of the image modulation to the object modulation at all spatial frequencies* as the **modulation transfer function**, or MTF.

Figure 11.46 is a plot of the MTF for two hypothetical lenses. Both start off with a zero-frequency (dc) value of 1.0, and both cross the zero axis somewhere where they can no longer resolve the data at that *cutoff frequency*. Had they both been diffraction-limited lenses, that cutoff would have depended only on diffraction and, hence, on the size of the aperture. In any event, suppose one of these is to be coupled to a detector

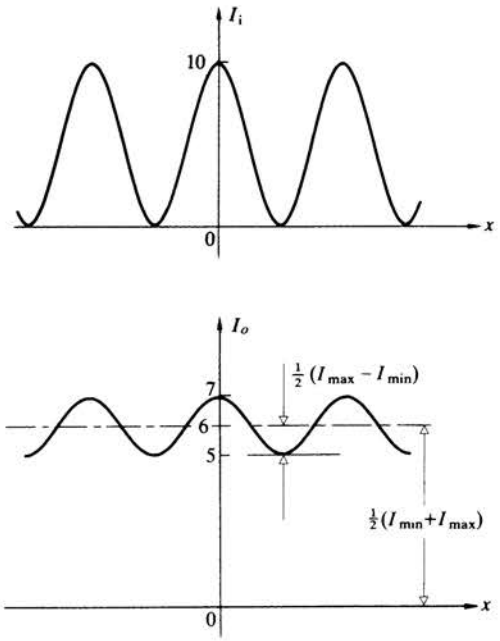


Figure 11.45 The irradiance into and out of a system.

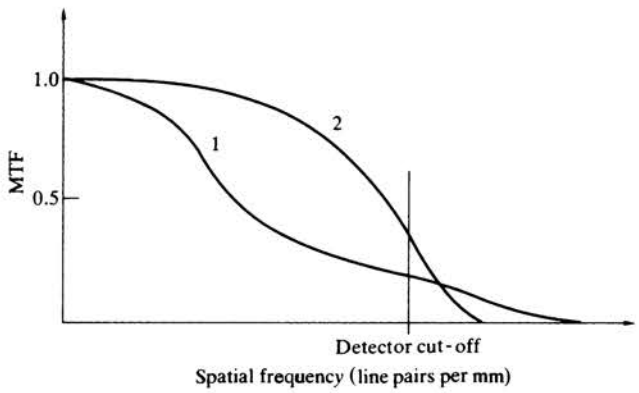


Figure 11.46 Modulation versus spatial frequency for two lenses.

whose cutoff frequency is indicated in the diagram. Despite the fact that lens-1 has a higher limit of resolution, lens-2 would certainly provide better performance when coupled to the particular detector.

It should be pointed out that a square bar target provides an input signal that is a series of square pulses, and the contrast in image is actually a superposition of contrast variations due to the constituent Fourier components. Indeed, one of the key points in what is to follow is that *optical elements functioning as linear operators transform a sinusoidal input into an undistorted sinusoidal output.* Despite this, the input and output irradiance distributions as a rule will not be identical. For example, the system's magnification affects the spatial frequency of the output (henceforth, the magnification will be taken as one). Diffraction and aberrations reduce the sinusoid's amplitude (contrast). Finally, asymmetrical aberrations (e.g., coma) and poor centering of elements produce a shift in the position of the output sinusoid corresponding to the introduction of a phase shift. This latter point, which was considered in Fig. 11.12, can be appreciated using a diagram like that of Fig. 11.47.

If the spread function is symmetrical, the image irradiance will be an unshifted sinusoid, whereas an asymmetrical spread function will apparently push the output over a bit, as in Fig. 11.48. In either case, *regardless of the form of the spread function, the image is harmonic if the object is harmonic.* Consequently, if we envision an object as being composed of Fourier components, the manner in which these individual harmonic components are transformed by the optical system into the corresponding harmonic constituents of the image is the quintessential feature of the process. The function that performs

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

this service is known as the **optical transfer function**, or OTF. It is a spatial frequency-dependent complex quantity whose modulus is the *modulation transfer function* (MTF) and whose phase, naturally enough, is the **phase transfer function** (PTF). The former is a measure of the reduction in contrast from object to image over the spectrum. The latter represents the commensurate relative phase shift. Phase shifts in centered optical systems occur only off-axis, and often the PTF is of less interest than the MTF. Even so, each application of the transfer function must be studied carefully; there are situations wherein the PTF plays a crucial role. In point of fact, the MTF has become a widely used means of specifying the performance of all sorts of elements and systems, from lenses, magnetic tape, and films to telescopes, the atmosphere, and the eye, to mention but a few. Moreover, it has the advantage that if the MTFs for the individual independent components in a system are known, the total MTF is often simply their product. This is inapplicable to the cascading of lenses, since the aberrations in one lens can compensate for those of another lens in tandem with it, and they are therefore not independent. Thus if we photograph an object having a modulation of 0.3 at 30

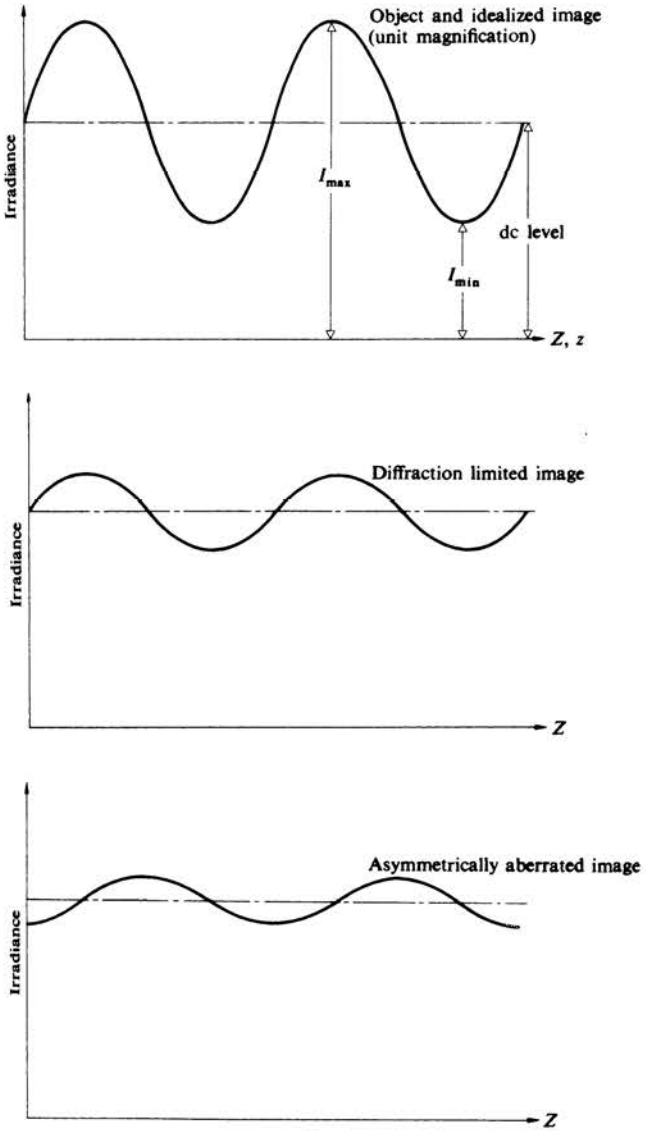


Figure 11.47 Harmonic input and resulting output.

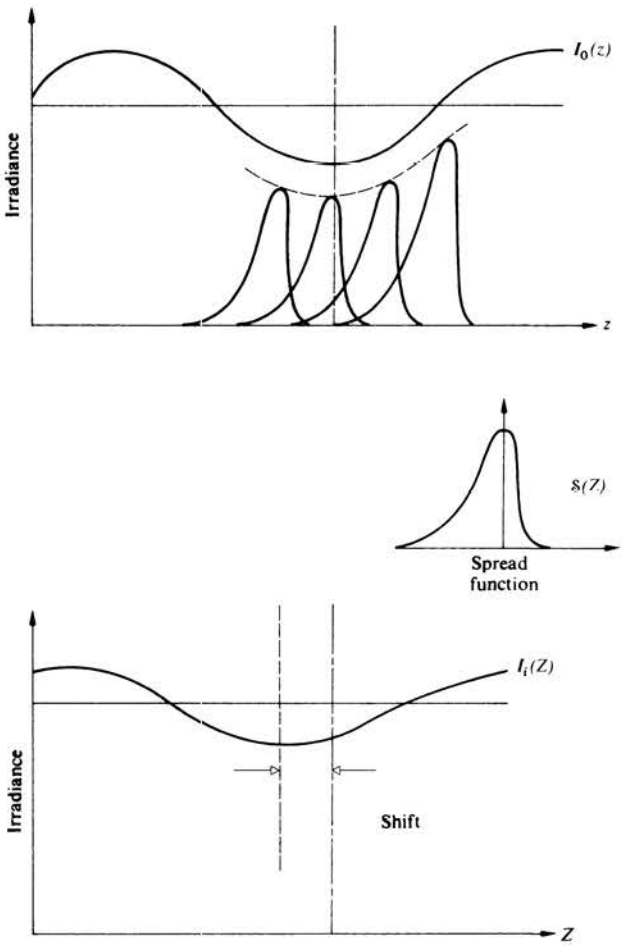


Figure 11.48 Harmonic input and output with an asymmetric spread function.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

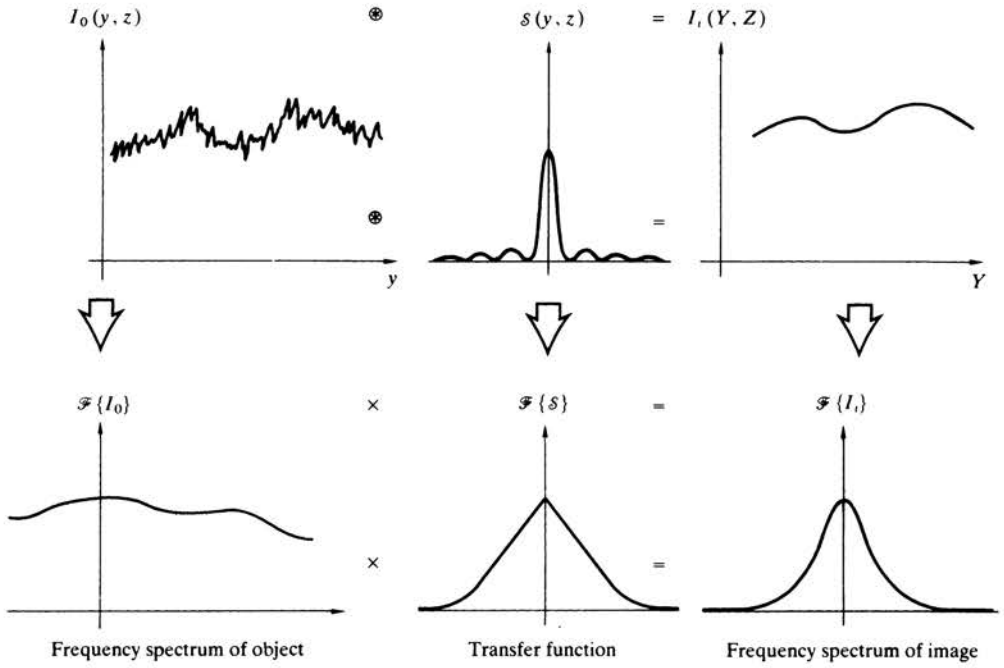


Figure 11.49 The relationships between the object and image spectra by way of the OTF, and the object and image irradiances by way of the point-spread function—all in incoherent illumination.

cycles per mm, using a camera whose lens at the appropriate setting has an MTF of 0.5 at 30 cycles/mm and a film* such as Tri-X with an MTF of 0.4 at 30 cycles/mm, the image modulation will be $0.3 \times 0.5 \times 0.4 = 0.06$.

A More Formal Discussion

We saw in Eq. (11.51) that the image (under the conditions of space invariance and incoherence) could be expressed as the convolution of the object irradiance and the point-spread function, in other words,

$$I_i(Y,Z) = I_0(y,z) \circledast s(y,z) \tag{11.90}$$

The corresponding statement in the spatial frequency domain is obtained by a Fourier transform, namely,

$$\mathcal{F}\{I_i(Y,Z)\} = \mathcal{F}\{I_0(y,z)\} \cdot \mathcal{F}\{s(y,z)\} \tag{11.91}$$

where use was made of the convolution theorem [Eq. (11.53)]. This says that *the frequency spectrum of the image irradiance distribution equals the product of the frequency spectrum of the object irradiance distribution and the transform of the spread function* (Fig. 11.49). Thus, it is multiplication by $\mathcal{F}\{s(y,z)\}$ that produces the alteration in the frequency spectrum of the object, converting it into that of the image spectrum. In other words, it is $\mathcal{F}\{s(y,z)\}$ that, in effect, transfers the object spectrum into the image spectrum. This is just the service performed by the OTF, and indeed we shall define the **unnormalized OTF** as

$$\mathcal{T}(k_Y, k_Z) \equiv \mathcal{F}\{s(y,z)\} \tag{11.92}$$

The modulus of $\mathcal{T}(k_Y, k_Z)$ will effect a change in the amplitudes of the various frequency components of the object spectrum, while its phase will, of course, appropriately alter the phase of these components to yield $\mathcal{F}\{I_i(Y,Z)\}$. Bear in mind that in the right-hand side of Eq. (11.90) the only quantity dependent on the actual optical system is $s(y,z)$, so it's not surprising that the spread function is the spatial counterpart of the OTF.

Let's now verify the statement made earlier that a harmonic input transforms into a somewhat altered harmonic output.

*Incidentally, the whole idea of treating film as a noise-free linear system is somewhat suspect. For further reading see J. B. De Velis and G. B. Parrent, Jr., "Transfer Function for Cascaded Optical Systems," *J. Opt. Soc. Am.* **57**, 1486 (1967).

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027

To that end, suppose

$$I_0(z) = 1 + a\cos(k_Z z + \varepsilon) \tag{11.93}$$

where for simplicity's sake, we'll again use a one-dimensional distribution. The 1 is a dc bias, which makes sure the irradiance doesn't take on any unphysical negative values. Insofar as $f \circledast h = h \circledast f$, it will be more convenient here to use

$$I_i(Z) = \mathcal{S}(Z) \circledast I_0(z)$$

and so

$$I_i(Z) = \int_{-\infty}^{+\infty} \{1 + a\cos[k_Z(Z - z) + \varepsilon]\}\mathcal{S}(z)\, dz$$

Expanding out the cosine, we obtain

$$I_i(Z) = \int_{-\infty}^{+\infty} \mathcal{S}(z)dz + a\cos(k_Z Z + \varepsilon)\int_{-\infty}^{+\infty} \cos k_Z z\, \mathcal{S}(z)\, dz$$

$$+ a\sin(k_Z Z + \varepsilon)\int_{-\infty}^{+\infty} \sin k_Z z\, \mathcal{S}(z)\, dz$$

Referring back to Eq. (7.57), we recognize the second and third integrals as the Fourier cosine and sine transforms of $\mathcal{S}(z)$, respectively, that is to say, $\mathcal{F}_c\{\mathcal{S}(z)\}$ and $\mathcal{F}_s\{\mathcal{S}(z)\}$. Hence

$$I_i(z) = \int_{-\infty}^{+\infty} \mathcal{S}(z)dz + \mathcal{F}_c\{\mathcal{S}(z)\}a\cos(k_Z Z + \varepsilon)$$

$$+ \mathcal{F}_s\{\mathcal{S}(z)\}a\sin(k_Z Z + \varepsilon) \tag{11.94}$$

Recall that the complex transform we've become so used to working with was defined such that

$$\mathcal{F}\{f(z)\} = \mathcal{F}_c\{f(z)\} + i\mathcal{F}_s\{f(z)\} \tag{11.95}$$

or

$$F(k_Z) = A(k_Z) + iB(k_Z) \tag{11.7}$$

In addition,

$$\mathcal{F}\{f(z)\} = |F(k_Z)|e^{i\varphi(k_Z)} = |F(k_Z)|[\cos\varphi + i\sin\varphi]$$

where

$$|F(k_Z)| = [A^2(k_Z) + B^2(k_Z)]^{1/2} \tag{11.96}$$

and

$$\varphi(k) = \tan^{-1}\frac{B(k_Z)}{A(k_Z)} \tag{11.97}$$

In precisely the same way, we apply this to the OTF, writing it as

$$\mathcal{F}\{\mathcal{S}(z)\} \equiv \mathcal{T}(k_Z) = \mathcal{M}(k_Z)e^{i\Phi(k_Z)} \tag{11.98}$$

where $\mathcal{M}(k_Z)$ and $\Phi(k_Z)$ are the unnormalized MTF and the PTF, respectively. It is left as a problem to show that Eq. (11.94) can be recast as

$$I_i(Z) = \int_{-\infty}^{+\infty} \mathcal{S}(z)\, dz + a\mathcal{M}(k_Z)\cos[k_Z Z + \varepsilon - \Phi(k_Z)] \tag{11.99}$$

Notice that this is a function of the same form as the input signal [Eq. (11.93)], $I_0(z)$, which is just what we set out to determine. If the line-spread function is symmetrical (i.e., even), $\mathcal{F}_s\{\mathcal{S}(z)\} = 0$, $\mathcal{M}(k_Z) = \mathcal{F}_c\{\mathcal{S}(z)\}$, and $\Phi(k_Z) = 0$; there is no phase shift, as was pointed out in the previous section. For an asymmetric (odd) spread function, $\mathcal{F}_s\{\mathcal{S}(z)\}$ is nonzero, as is the PTF.

It has now become customary practice to define a set of *normalized transfer functions* by dividing $\mathcal{T}(k_Z)$ by its zero spatial frequency value, that is, $\mathcal{T}(0) = \int_{-\infty}^{+\infty} \mathcal{S}(z)\, dz$. The normalized spread function becomes

$$\mathcal{S}_n(z) = \frac{\mathcal{S}(z)}{\displaystyle\int_{-\infty}^{+\infty} \mathcal{S}(z)\, dz} \tag{11.100}$$

while the **normalized OTF** is

$$T(k_Z) \equiv \frac{\mathcal{F}\{\mathcal{S}(z)\}}{\displaystyle\int_{-\infty}^{+\infty} \mathcal{S}(z)\, dz} = \mathcal{F}\{\mathcal{S}_n(z)\} \tag{11.101}$$

or in two dimensions

$$T(k_Y, k_Z) = M(k_Y, k_Z)e^{i\Phi(k_Y, k_Z)} \tag{11.102}$$

where $M(k_Y, k_Z) \equiv \mathcal{M}(k_Y, k_Z)/\mathcal{T}(0, 0)$. Therefore $I_i(Z)$ in Eq. (11.99) would then be proportional to

$$1 + aM(k_Z)\cos[k_Z Z + \varepsilon - \Phi(k_Z)]$$

The image modulation [Eq. (11.89)] becomes $aM(k_Z)$, the object modulation [Eq. (11.93)] is a, and the ratio is, as expected, the normalized MTF = $M(k_Z)$.

Giesecke+Devrient GmbH,
Giesecke+Devrient Currency Technology GmbH,
and Papierfabrik Louisenthal GmbH. - Ex. 1027