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-

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

$$f(x) = \frac{1}{\pi}\int_0^{\infty} \left[\int_{-\infty}^{+\infty} f(x') \cos k(x' - x)\, dx' \right] dk \qquad (11.1)$$

The quantity in the square brackets is an even function of k, and therefore changing the limits on the outer integral leads to

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} f(x') \cos k(x' - x)\, dx' \right] dk \qquad (11.2)$$

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,

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} f(x') e^{ikx'}\, dx' \right] e^{-ikx}\, dk \qquad (11.3)$$

Thus we can write

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} F(k) e^{-ikx}\, dk \qquad (11.4)$$

provided that

$$F(k) = \int_{-\infty}^{+\infty} f(x) e^{ikx}\, dx \qquad (11.5)$$

*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

$$F(k) = \mathcal{F}\{f(x)\} \qquad (11.6)$$

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,

$$F(k) = A(k) + iB(k) \qquad (11.7a)$$

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*:

$$F(k) = |F(k)|e^{i\phi(k)} \qquad (11.7b)$$

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

$$f(x) = \mathcal{F}^{-1}\{F(k)\} = \mathcal{F}^{-1}\{\mathcal{F}\{f(x)\}\} \qquad (11.8)$$

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,

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} F(\omega) e^{-i\omega t} d\omega \qquad (11.9)$$

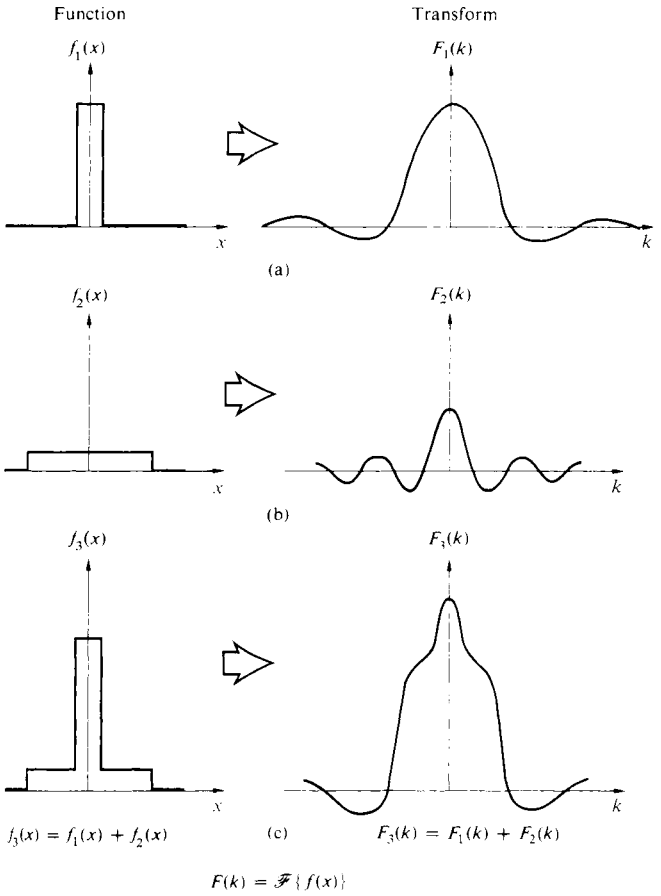


Figure 11.1 A composite function and its Fourier transform.

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 $\sigma_k =$

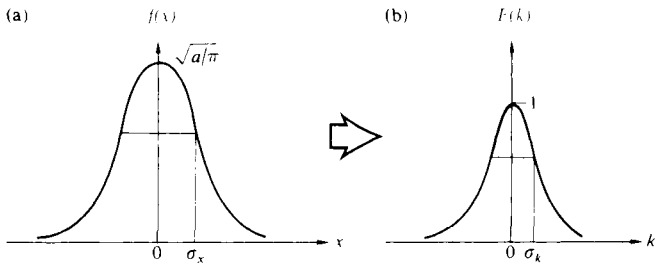


Figure 11.2 A Gaussian and its Fourier transform.

$\sqrt{2a}$ 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

$$k_\alpha = \sqrt{k_x^2 + k_y^2} \qquad (11.18)$$

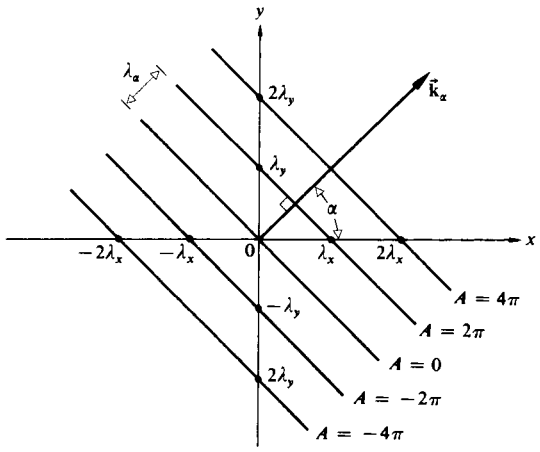


Figure 11.3 Geometry for Eq. (11.15).

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 $dr = k_\alpha^{-1}\,dw$,

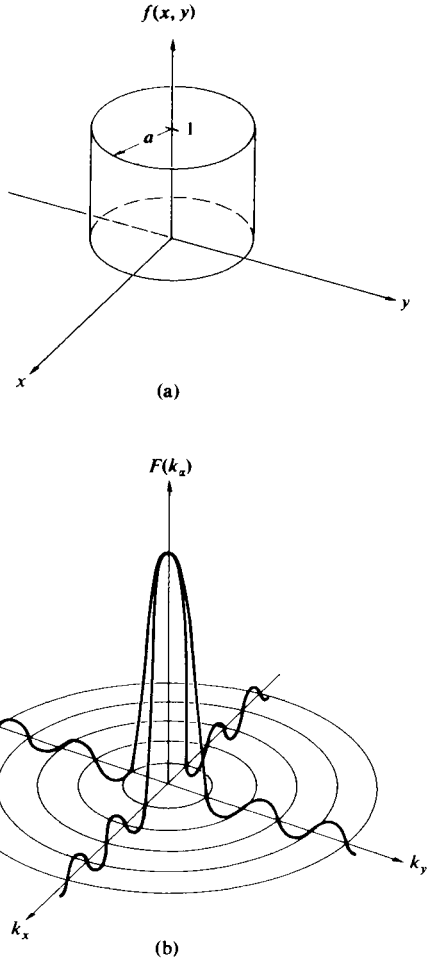


Figure 11.4 The cylinder, or top-hat, function and its transform.

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

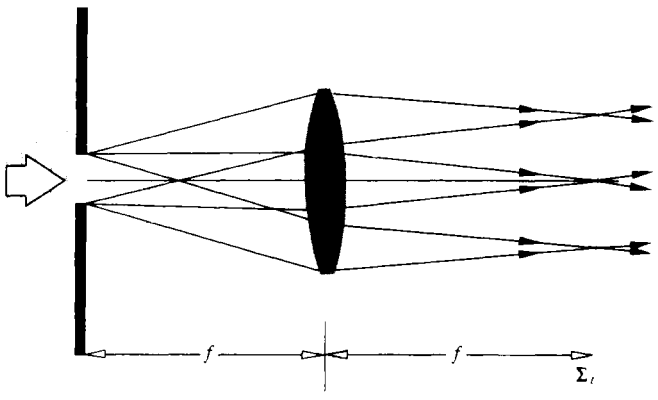


Figure 11.5 The light diffracted by a transparency at the front (or object) focal point of a lens converges to form the far-field diffraction pattern at the back (or image) focal point of the lens.

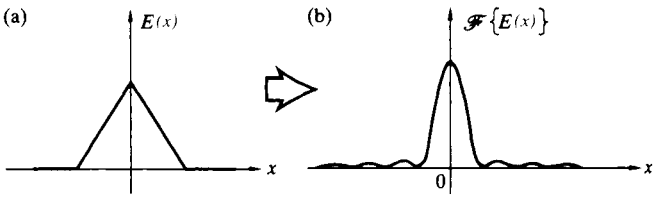


Figure 11.6 The transform of the triangle function is the sinc^2 function.

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 $\vec{E}$-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 $\vec{E}$-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

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

$$\int_{-\gamma}^{+\gamma} \delta(x)\, dx = 1$$

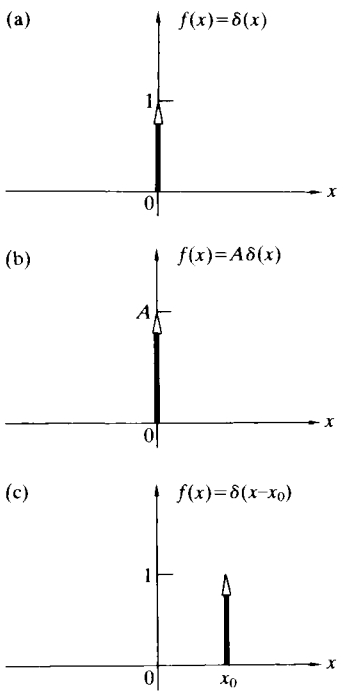


Figure 11.7 The height of the arrow representing the delta function corresponds to the area under the function.

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 $f(x' + x_0) = g(x')$,

$$\int_{-\infty}^{+\infty} \delta(x - x_0) f(x)\, dx = \int_{-\infty}^{+\infty} \delta(x') g(x')\, dx' = g(0)$$

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

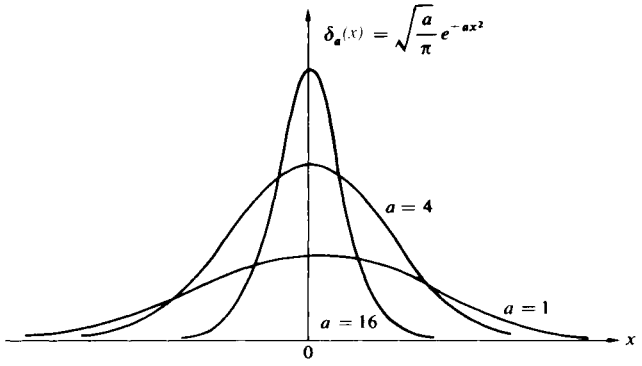


Figure 11.8 A sequence of Gaussians.

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

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 $m\pi/4$.

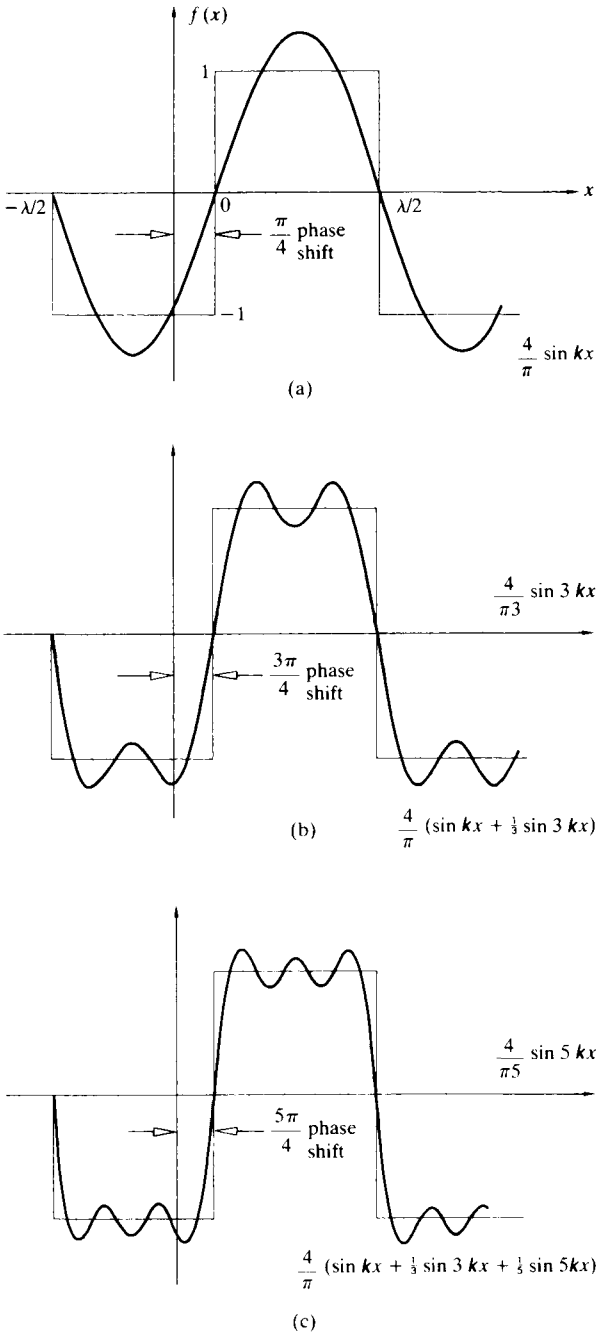


Figure 11.9 A shifted square wave showing the corresponding change in phase for each component wave.

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.

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

$$\mathcal{F}\{f(x)\} = e^{ikd/2} - e^{ikd/2} = 2i\sin(kd/2) \tag{11.44}$$

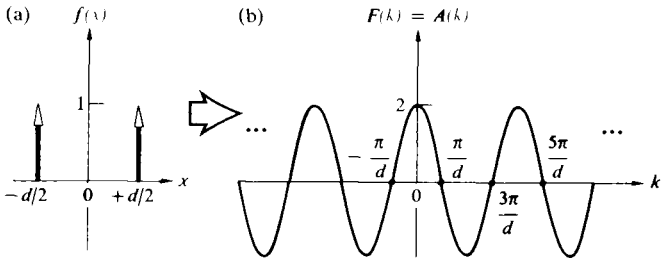


Figure 11.10 Two delta functions and their cosine-function transform.

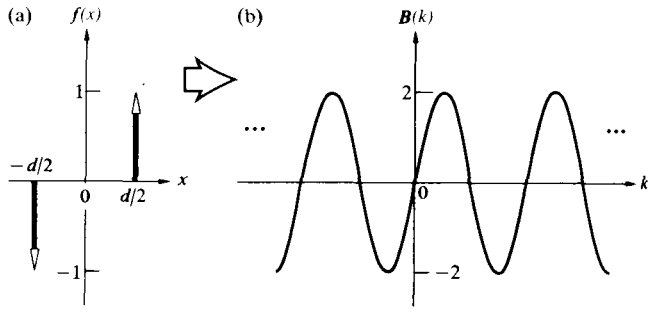


Figure 11.11 Two delta functions and their sine-function transform.

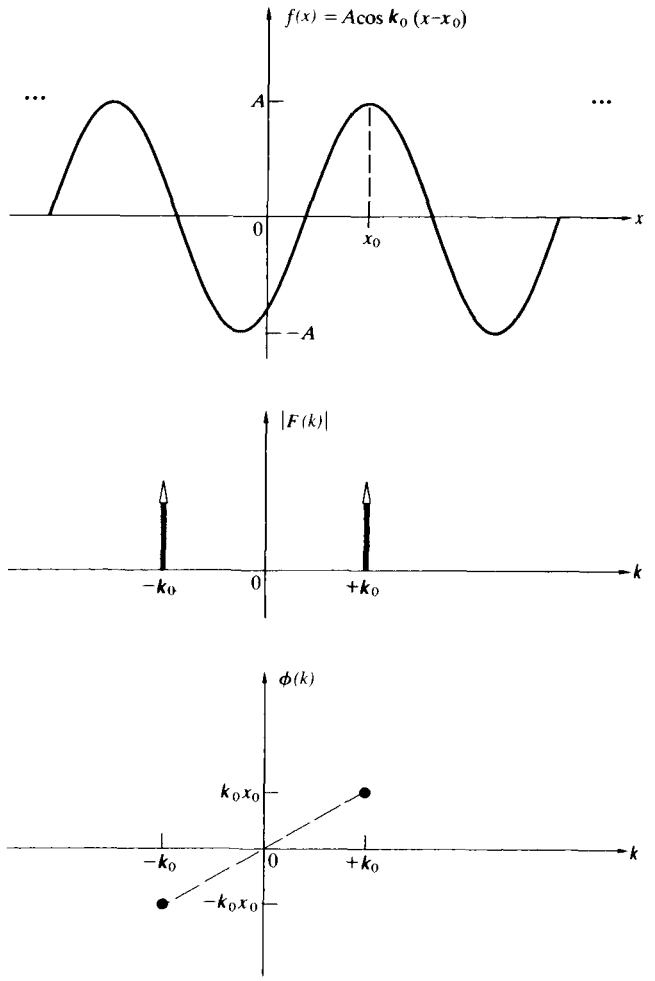


Figure 11.12 The spectra of a shifted cosine function.

The real sine transform [Eq. (11.7)] is then

$$B(k) = 2 \sin (kd/2) \tag{11.45}$$

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.

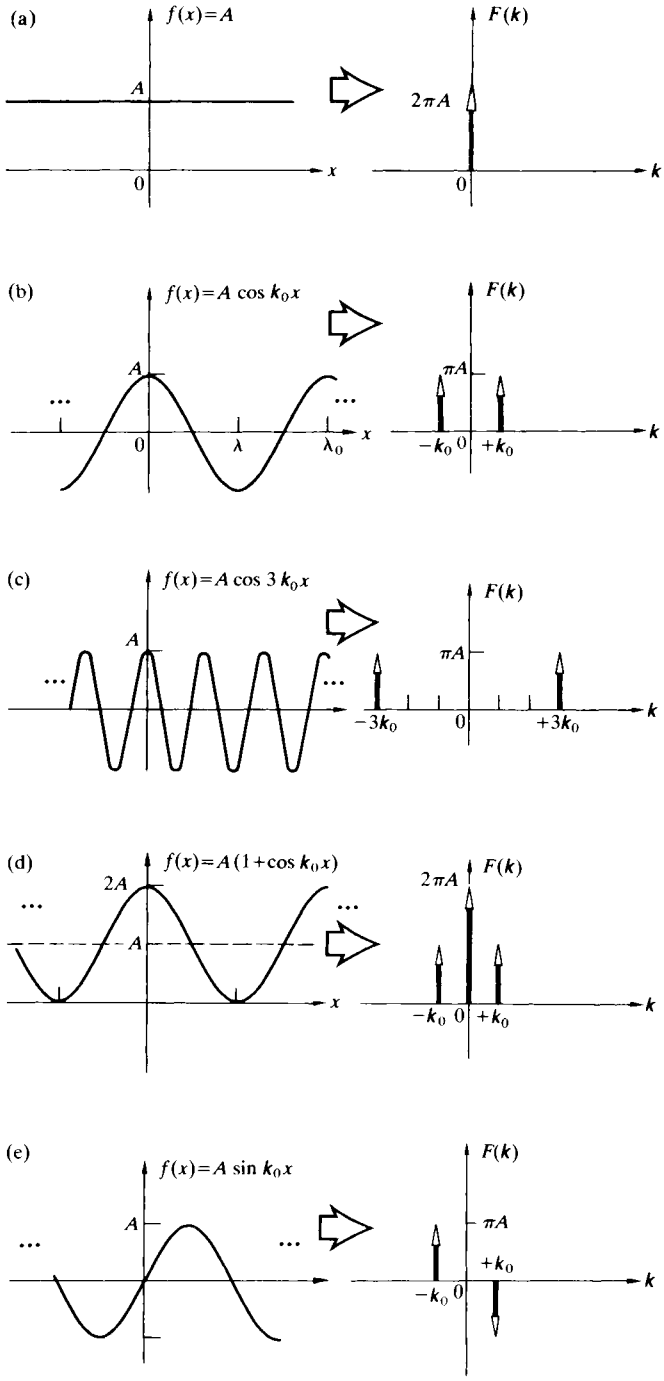


Figure 11.13 Some functions and their transforms.

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.

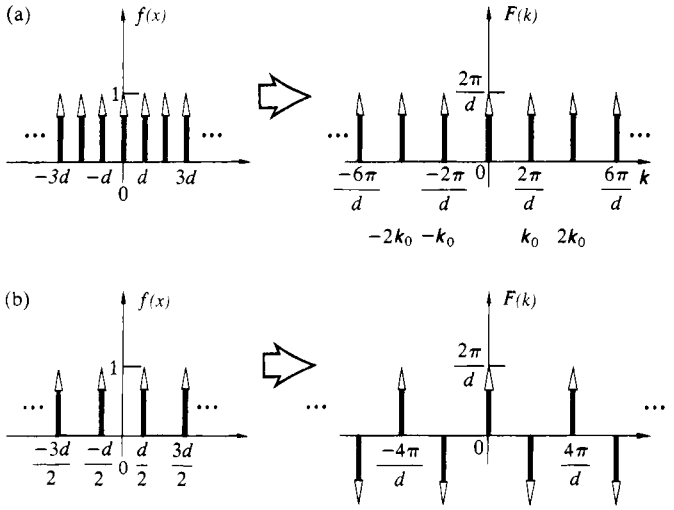


Figure 11.14 (*a*) The comb function and its transform. (*b*) A shifted comb function and its transform.

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

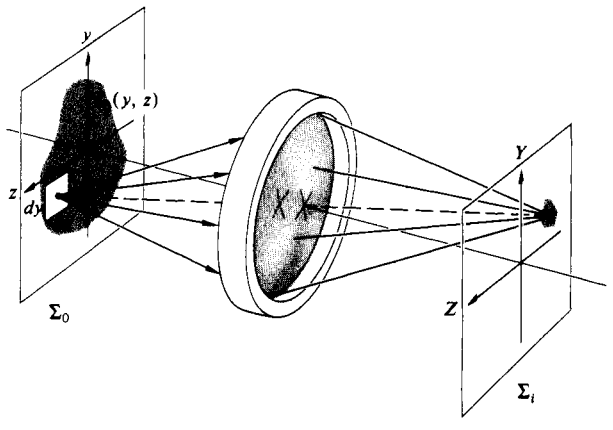


Figure 11.15 A lens system forming an image.

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.

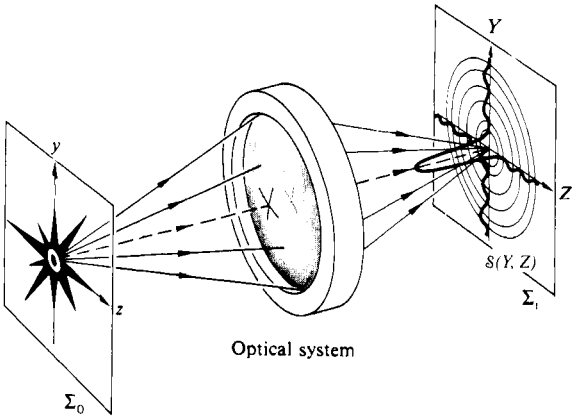


Figure 11.16 The point-spread function: the irradiance produced by the optical system with an input point source.

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.

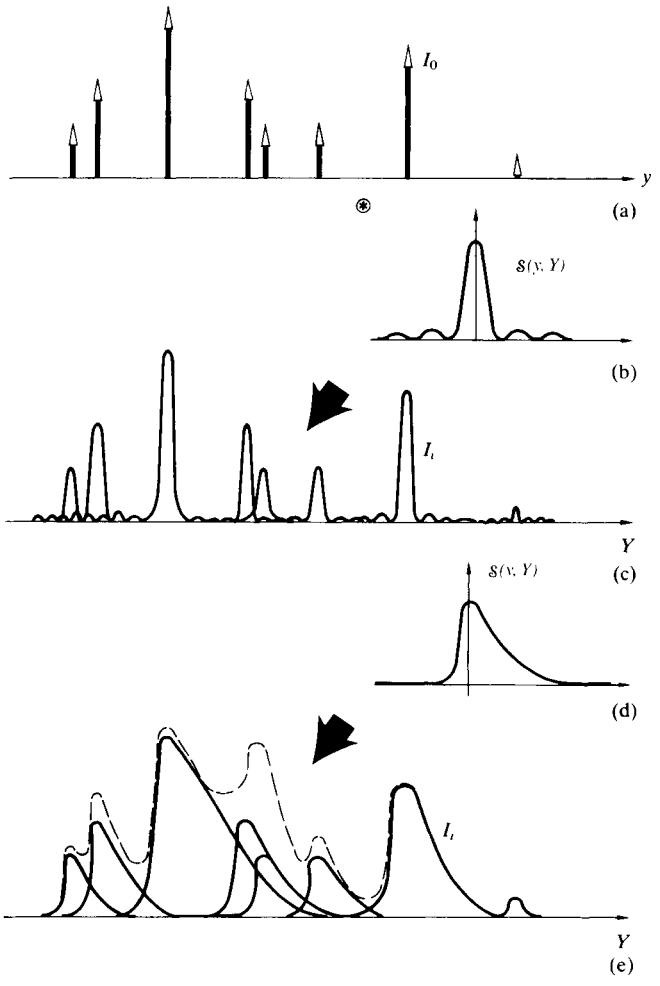


Figure 11.17 Here (*a*) is convolved first with (*b*) to produce (*c*) and then with (*d*) to produce (*e*). The resulting pattern is the sum of all the spread-out contributions as indicated by the dashed curve in (*e*).

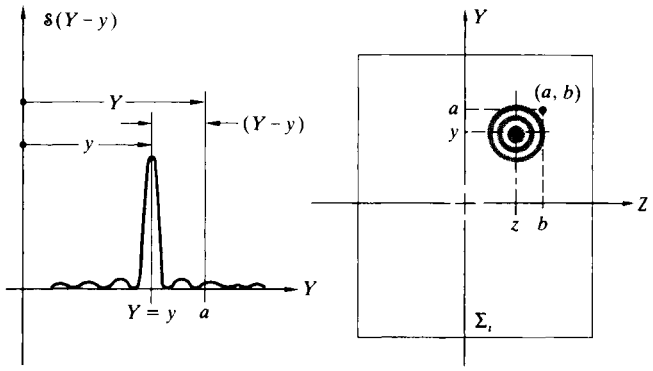


Figure 11.18 The point-spread function.

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-

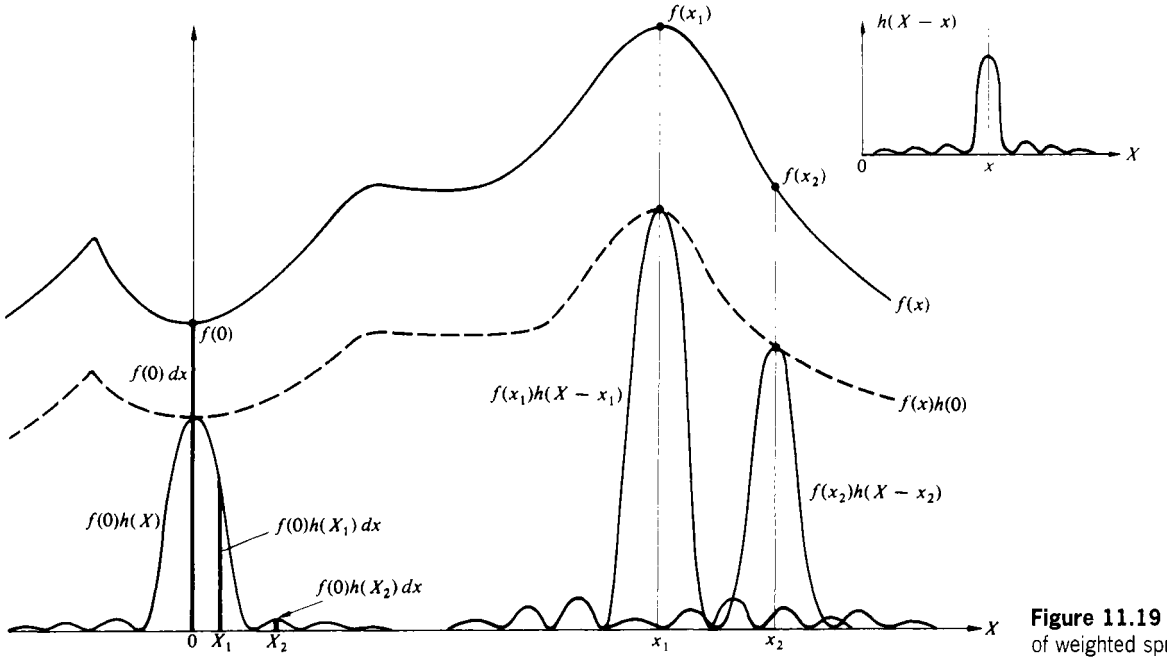


Figure 11.19 The overlapping of weighted spread functions.

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 $g(X_I)$, 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 $dg(X_I) = f(x)h(X_I - x)\,dx$. 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

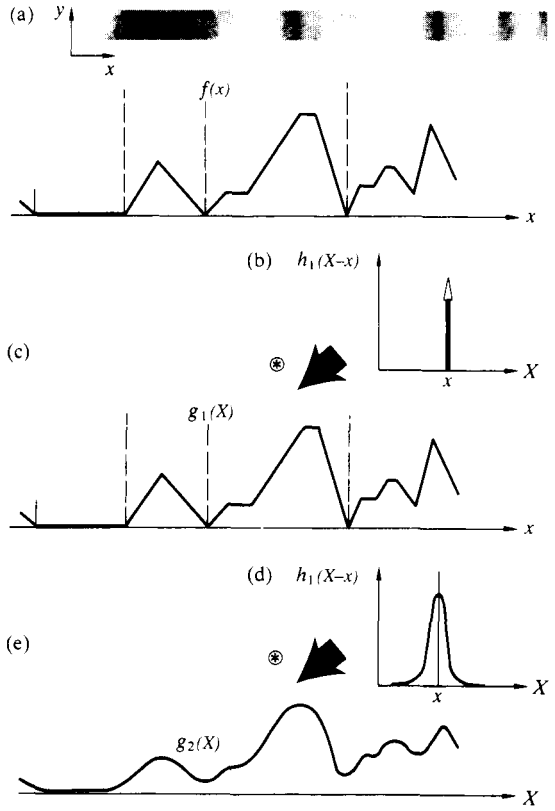


Figure 11.20 The irradiance distribution is converted to a function $f(x)$ shown in (*a*). This is convolved with a δ-function (*b*) to yield a duplicate of $f(x)$. By contrast, convolving $f(x)$ with the spread function h_2 in (*d*) yields a smoothed-out curve represented by $g_2(x)$ in (*e*).

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*

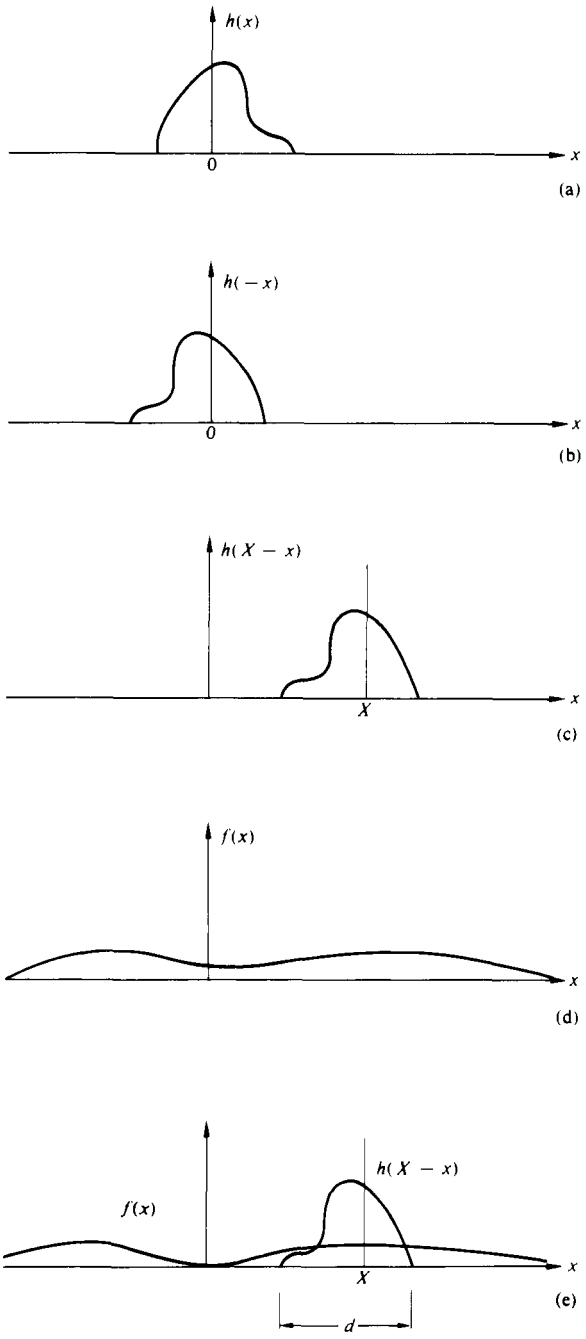


Figure 11.21 The geometry of the convolution process in the object coordinates.

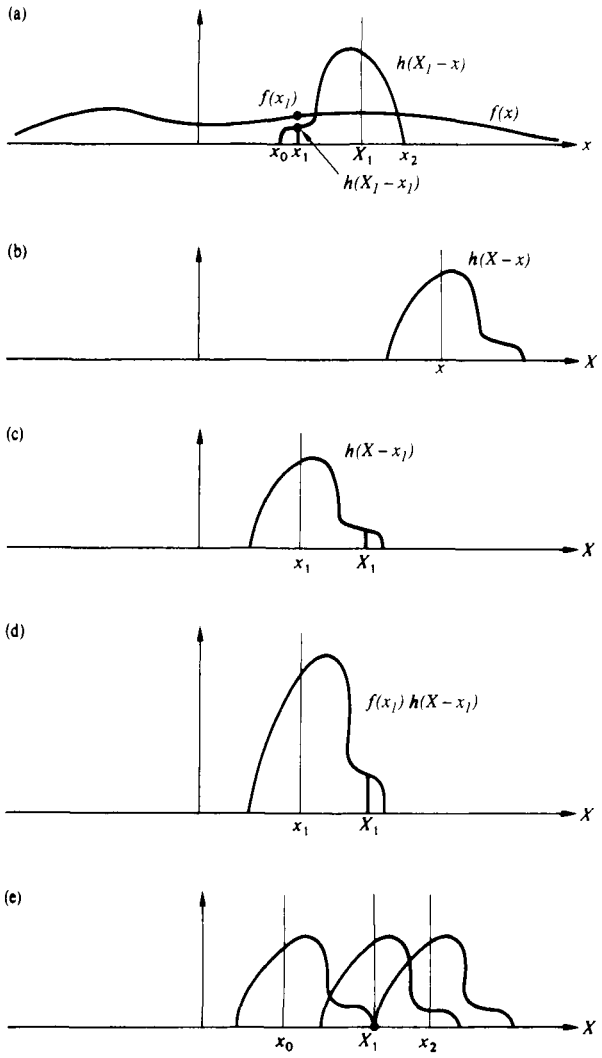


Figure 11.22 The geometry of the convolution process in the image coordinates.

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

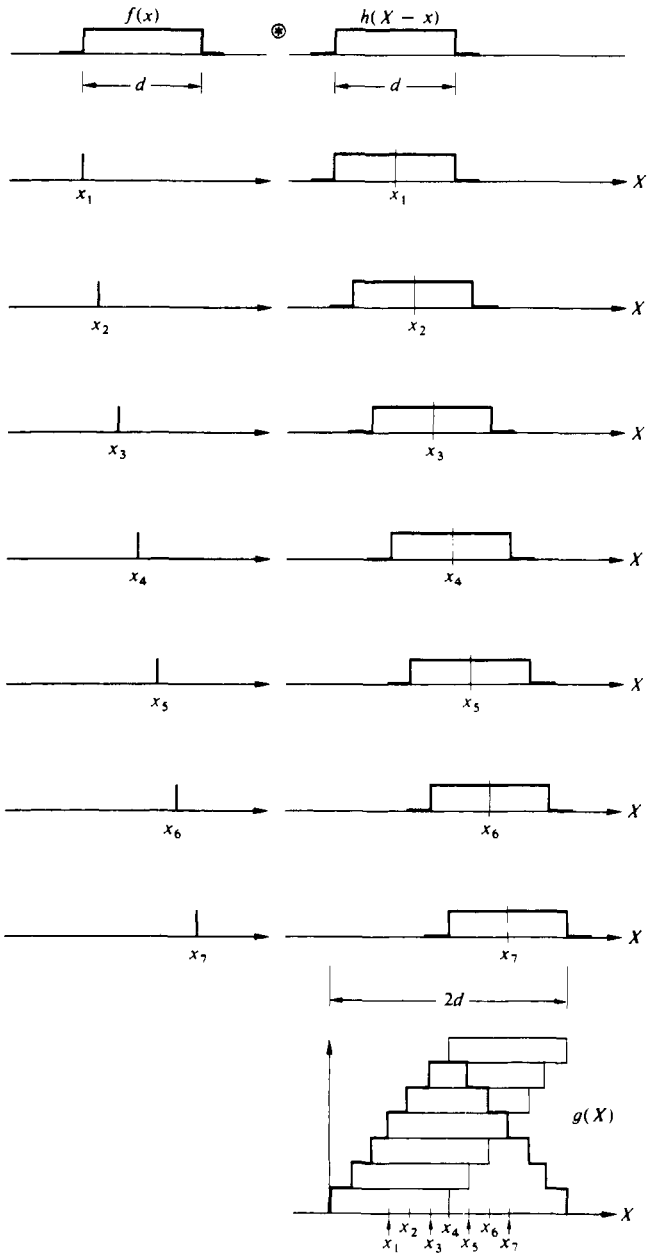


Figure 11.23 Convolution of two square pulses. The fact that we represented $f(x)$ by a finite number of delta functions (viz., 7) accounts for the steps in $g(X)$.

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).

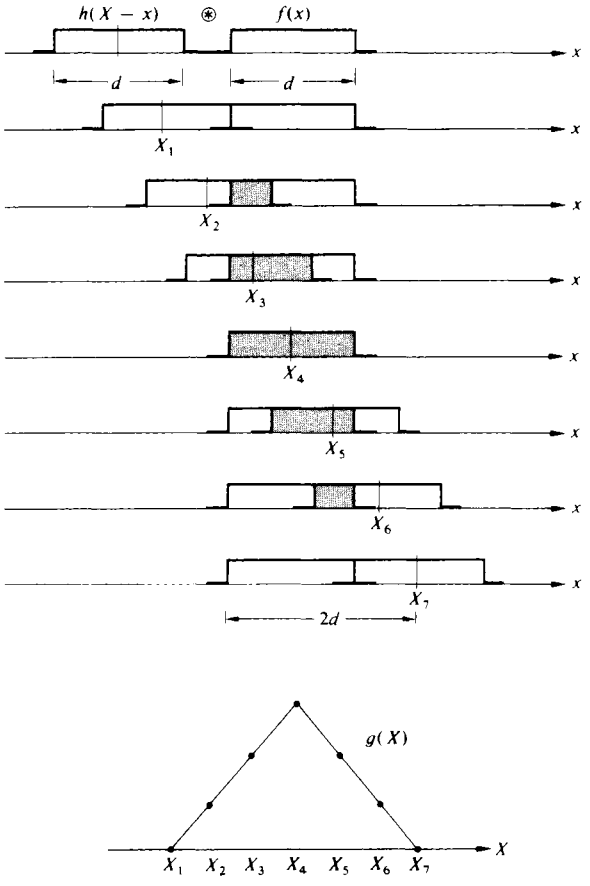


Figure 11.24 Convolution of two square pulses.

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,

$$\mathcal{F}\{g\} = [d \text{ sinc } (kd/2)]^2 \tag{11.55}$$

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.

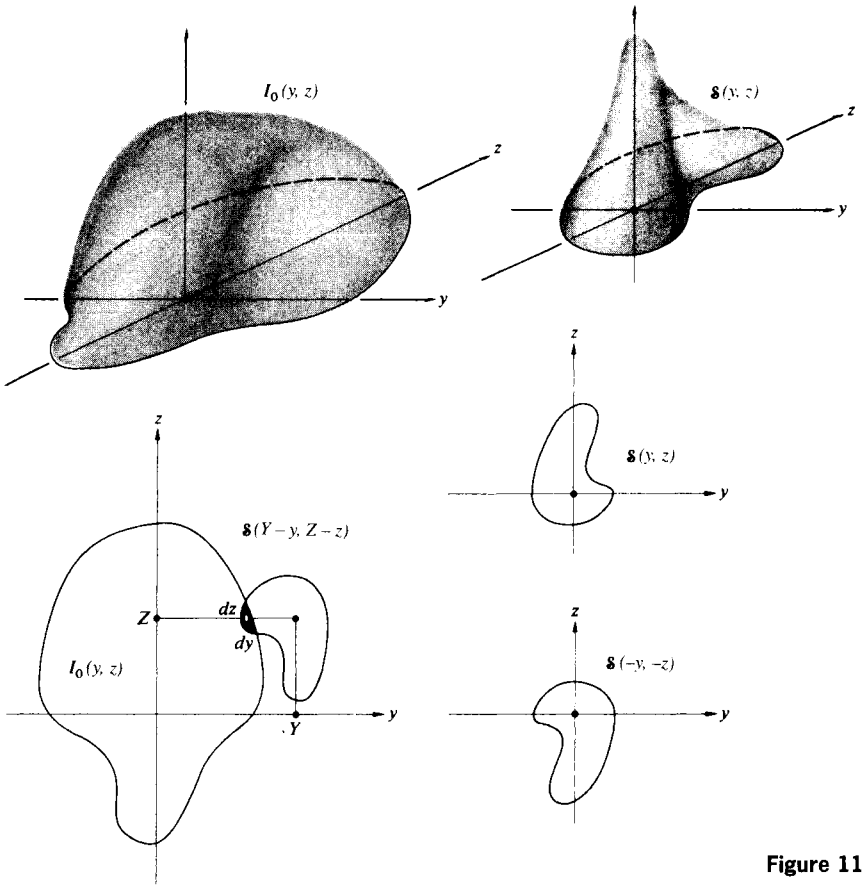


Figure 11.25 Convolution in two dimensions.

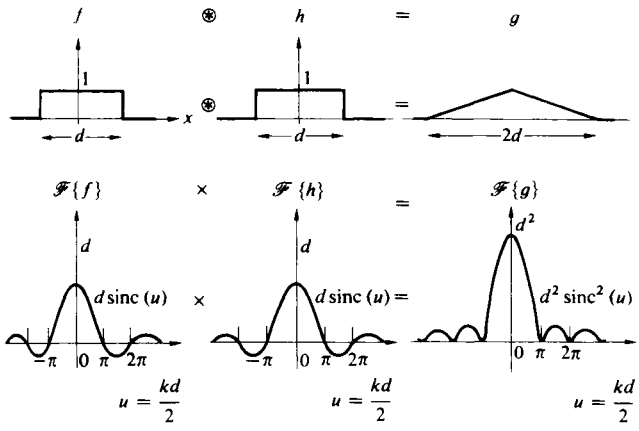


Figure 11.26 An illustration of the convolution theorem.

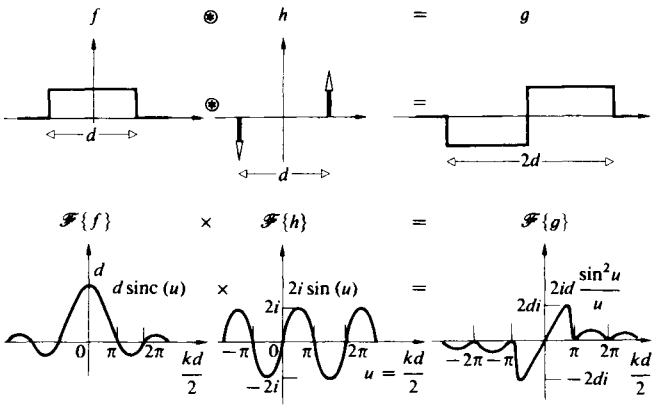


Figure 11.27 An illustration of the convolution theorem.

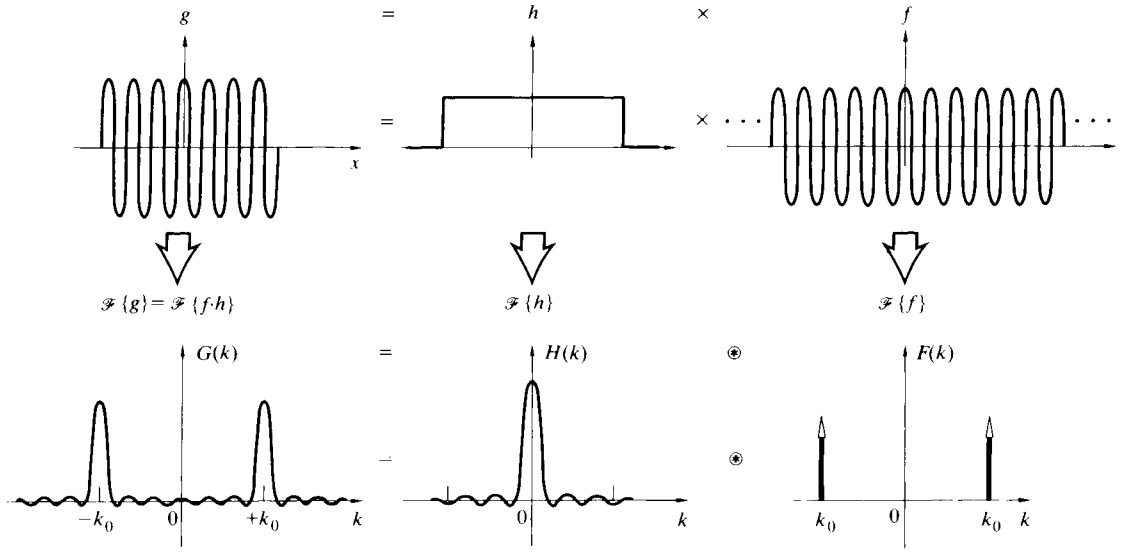


Figure 11.28
An example of the frequency convolution theorem.

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

$$F(k') = \int_{-\infty}^{+\infty} f(x)e^{ik'x}\,dx = F(k - k_0) \tag{11.58}$$

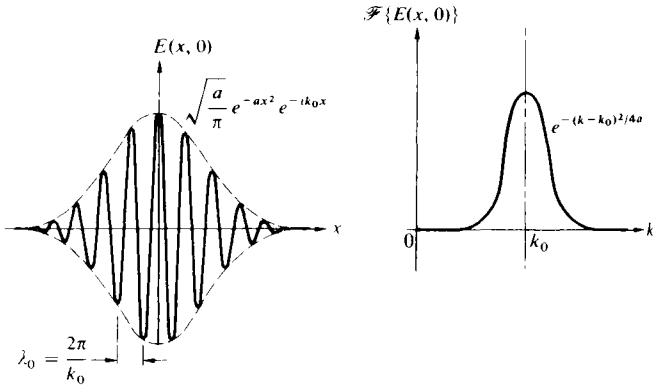


Figure 11.29 A Gaussian wave packet and its transform.

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!

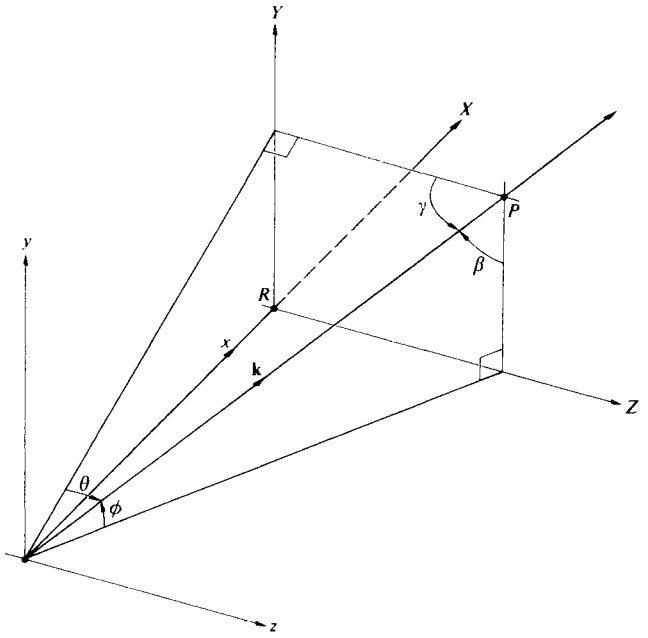


Figure 11.30 A bit of geometry.

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),

$$E(k_Y, k_Z) = \mathcal{F}\{\mathcal{A}(y, z)\}$$

$$= \int_{y=-b/2}^{+b/2} \int_{z=-a/2}^{+a/2} \mathcal{A}_0 e^{i(k_Y y + k_Z z)} dy\, dz$$

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-

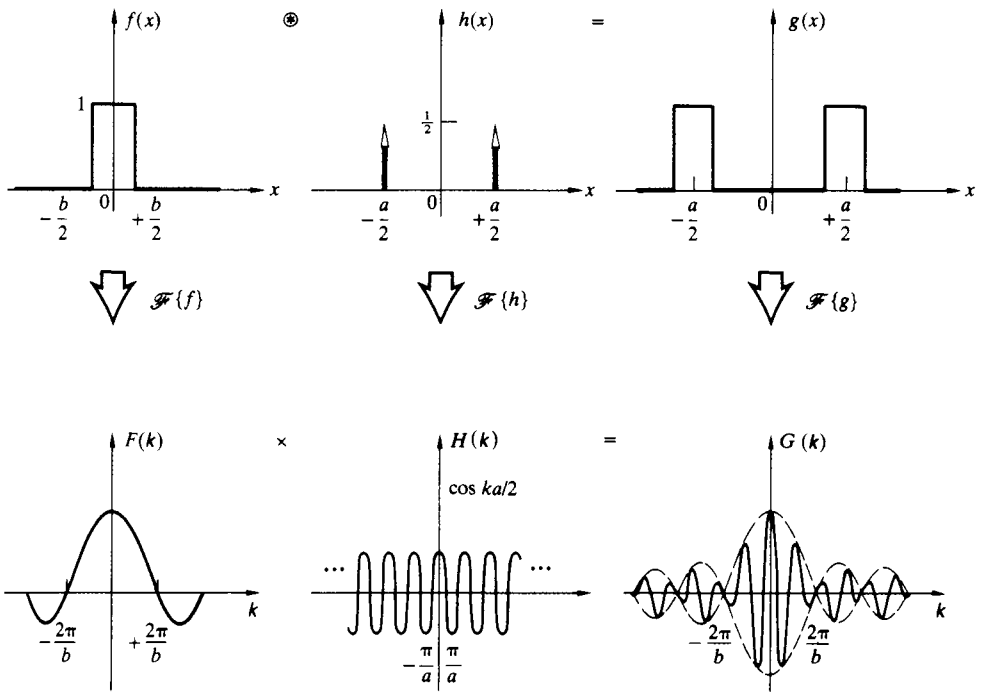


Figure 11.31 An illustration of the convolution theorem.

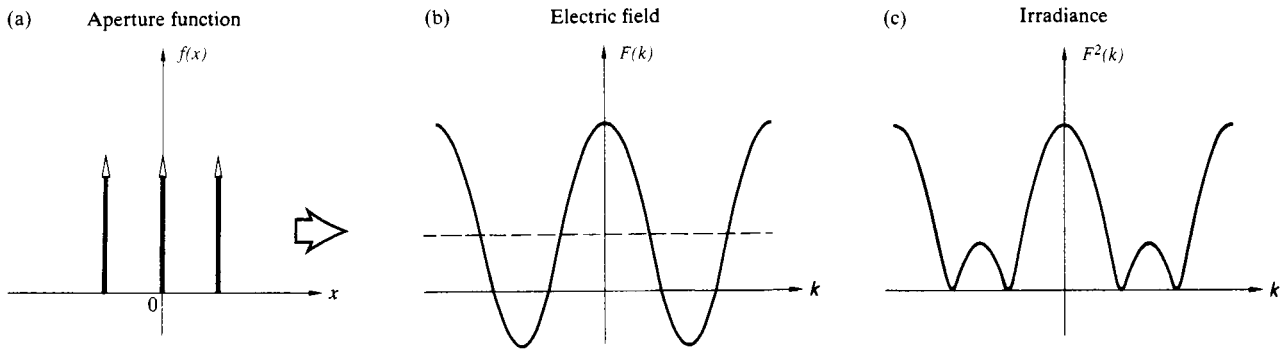


Figure 11.32 The Fourier transform of three equal δ-functions representing three slits.

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*.

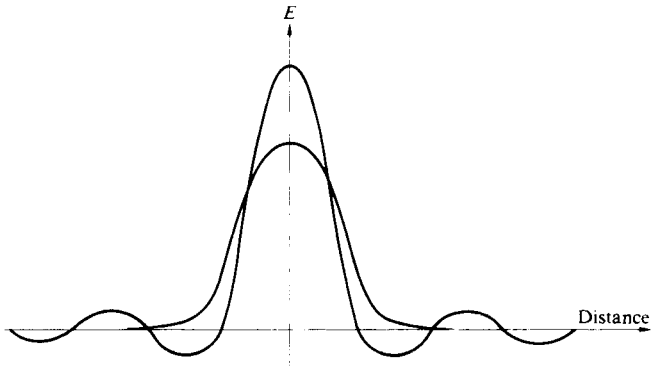


Figure 11.33 An Airy pattern compared with a Gaussian.

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).***

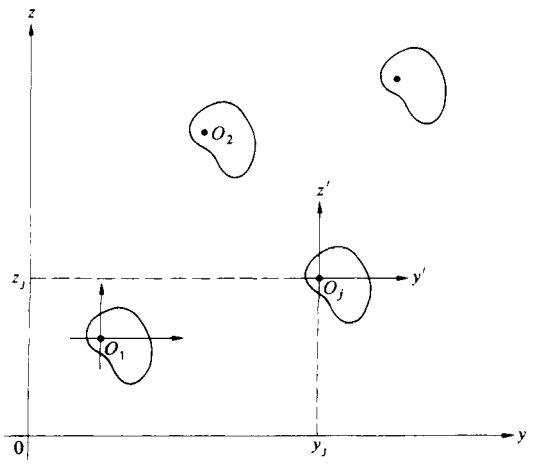


Figure 11.34 Multiple-aperture configuration.

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

$$\mathscr{A}(y, z) = \mathscr{A}_1(y', z') \circledast A_\delta \tag{11.75}$$

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,

$$\mathscr{A}_1(z') = \begin{cases} \mathscr{A}_{10} & \text{when } |z'| \leq b/2 \\ 0 & \text{when } |z'| > b/2 \end{cases}$$

and so

$$\mathscr{F}\{\mathscr{A}_1(z')\} = \mathscr{A}_{10} b \text{ sinc } k_Z b/2$$

With the slits located at $z = \pm a/2$,

$$A_\delta = \delta(z - a/2) + \delta(z + a/2)$$

and from Eq. (11.43)

$$\mathscr{F}\{A_\delta\} = 2\cos k_Z a/2$$

Thus

$$E(k_Z) = 2\mathscr{A}_{10} b \text{ sinc}\left(\frac{k_Z b}{2}\right)\cos\left(\frac{k_Z a}{2}\right)$$

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 $F(\omega) = \mathscr{F}\{f(t)\}$ 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 $|f(t)|^2 = f(t)f^*(t) = f(t)\cdot[\mathscr{F}^{-1}\{\mathscr{F}(\omega)\}]^*$,

$$\int_{-\infty}^{+\infty} |f(t)|^2\,dt = \int_{-\infty}^{+\infty} f(t)\left[\frac{1}{2\pi}\int_{-\infty}^{+\infty} F^*(\omega)e^{+i\omega t}\,d\omega\right]dt$$

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

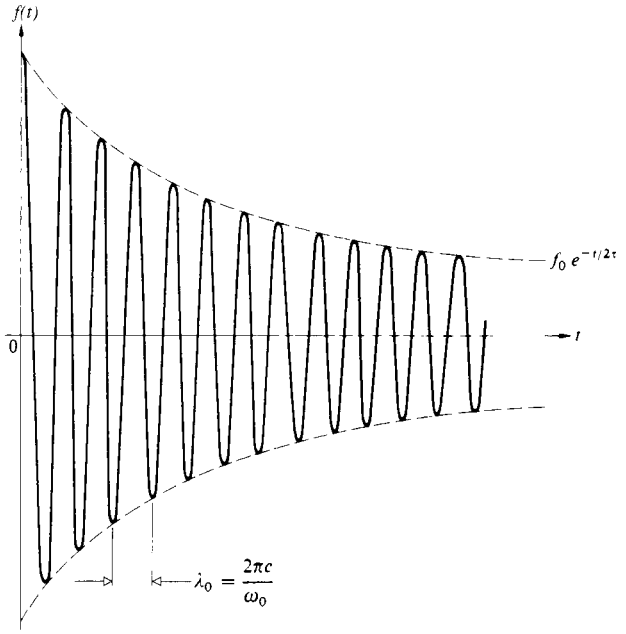


Figure 11.35 A damped harmonic wave.

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.

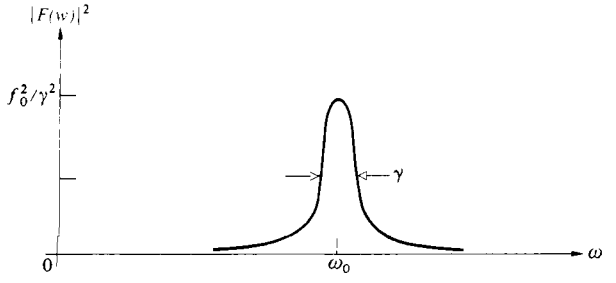


Figure 11.36 The resonance or Lorentz profile.

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

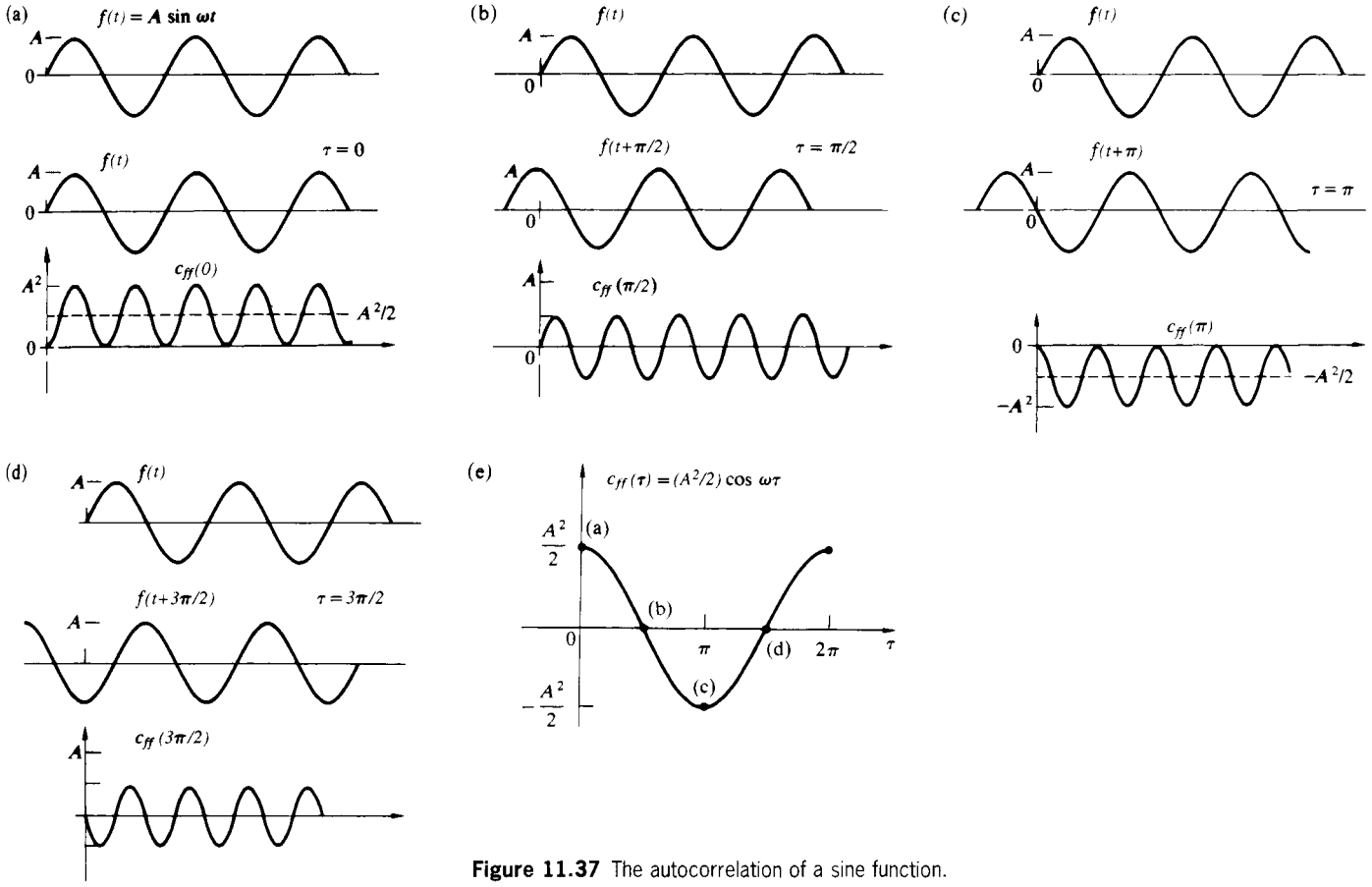


Figure 11.37 The autocorrelation of a sine function.

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

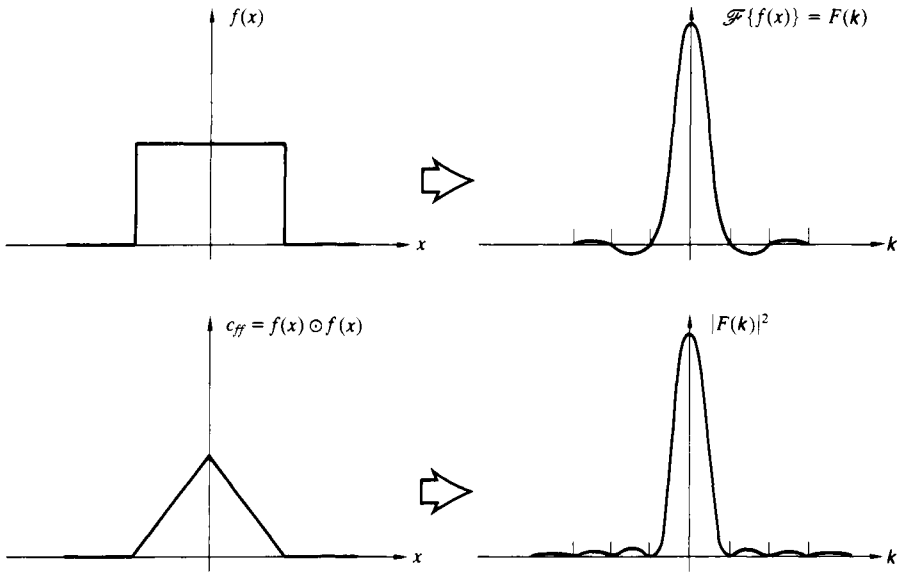


Figure 11.38 The square of the Fourier transform of the rectangular pulse $f(x)$ (i.e., $|F(k)|^2$) equals the Fourier transform of the autocorrelation of $f(x)$.

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

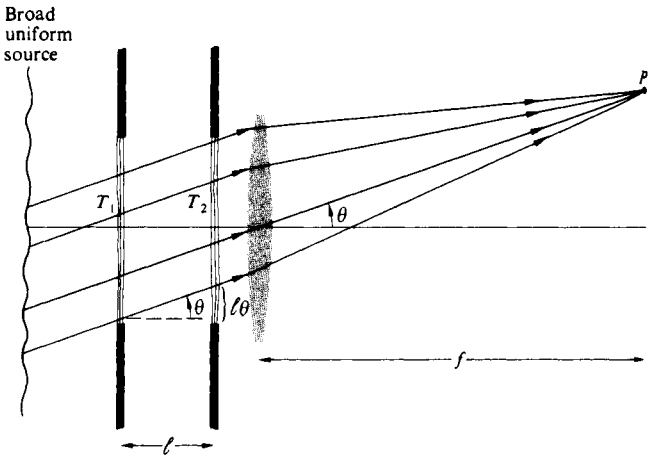


Figure 11.39 Optical correlation of two functions.

*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).

(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

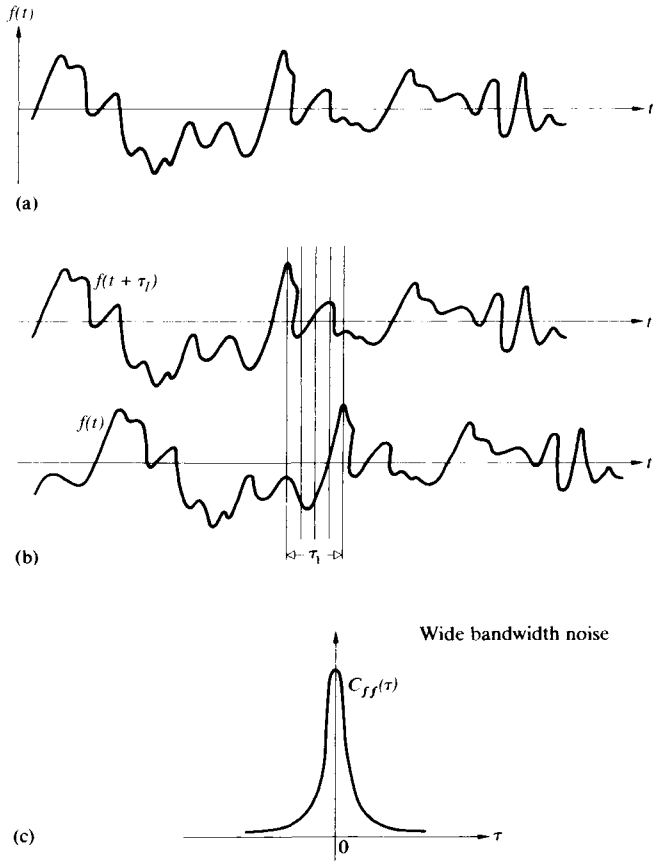


Figure 11.40 A signal $f(t)$ and its autocorrelation.

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

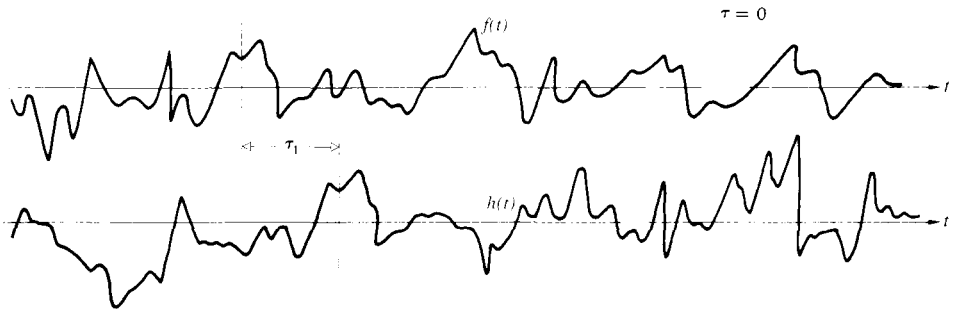


Figure 11.41 The cross-correlation of $f(t)$ and $h(t)$.

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

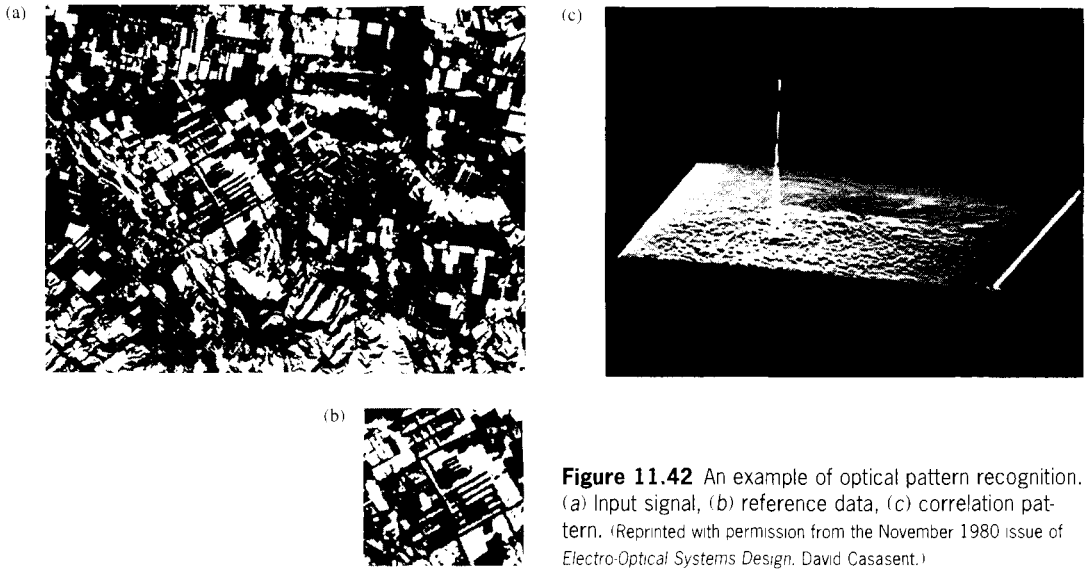


Figure 11.42 An example of optical pattern recognition. (*a*) Input signal, (*b*) reference data, (*c*) correlation pattern. (Reprinted with permission from the November 1980 issue of *Electro-Optical Systems Design*. David Casasent.)

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*).

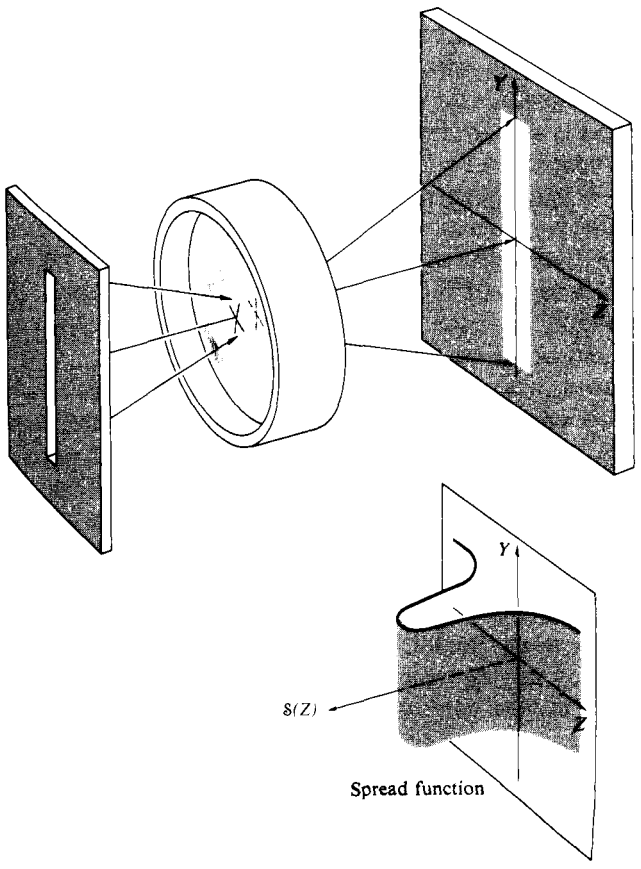


Figure 11.43 The line-spread function.

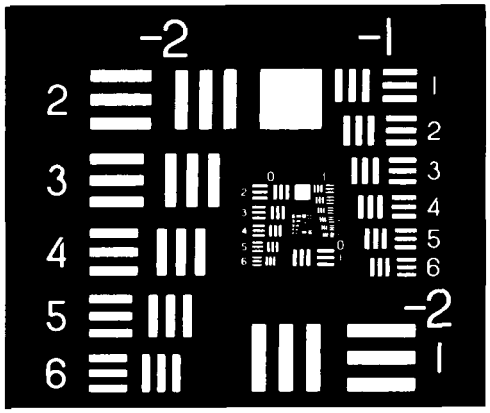


Figure 11.44 A bar target resolution chart.

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

$$\text{Modulation} \equiv \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \tag{11.89}$$

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

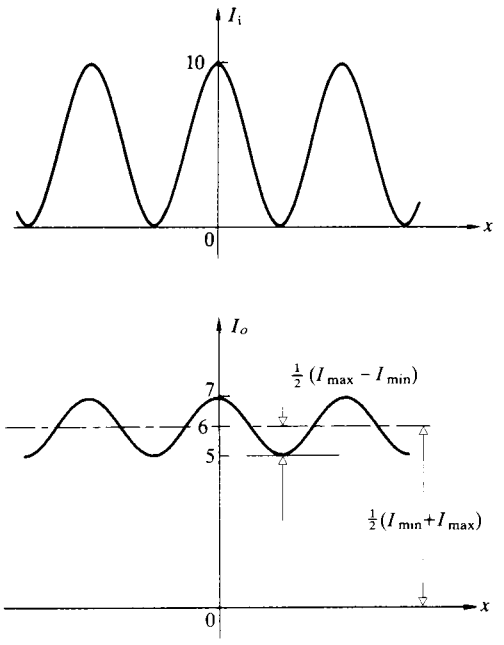


Figure 11.45 The irradiance into and out of a system.

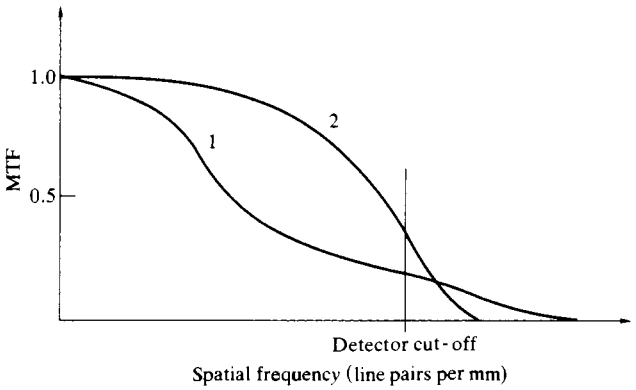


Figure 11.46 Modulation versus spatial frequency for two lenses.

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

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

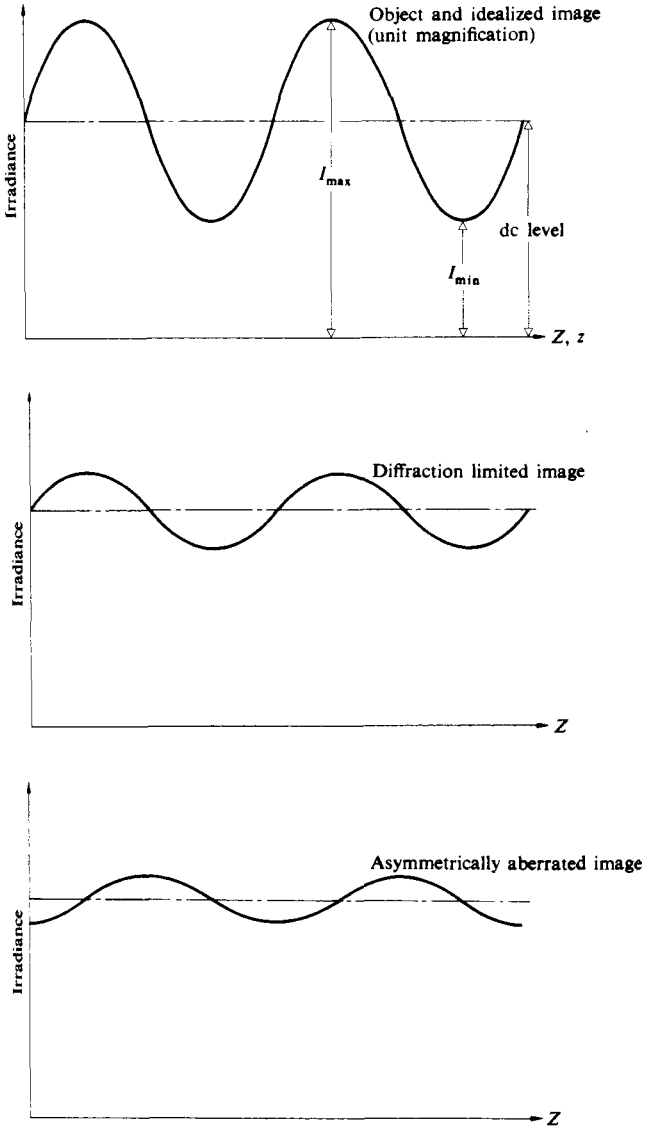


Figure 11.47 Harmonic input and resulting output.

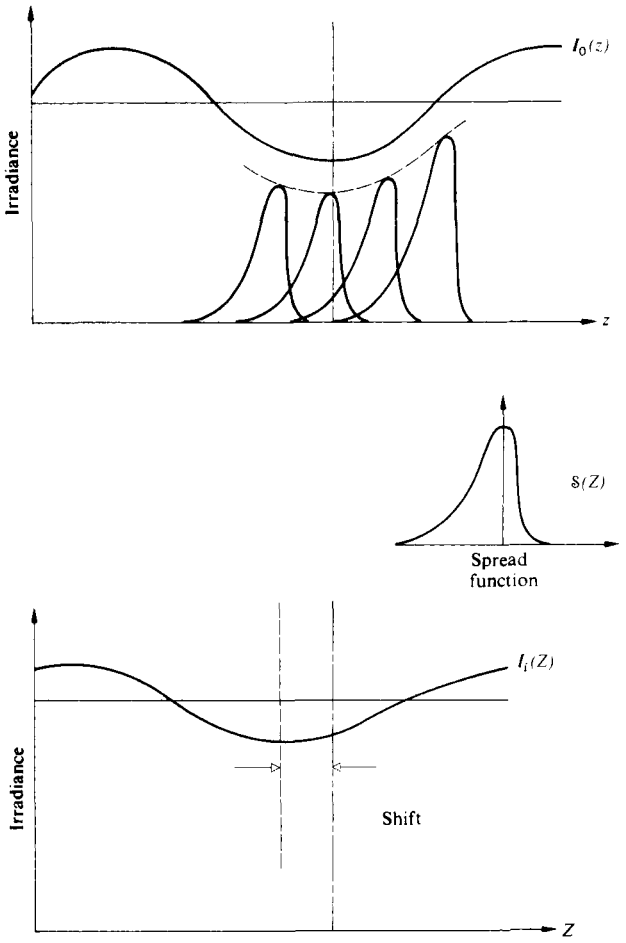


Figure 11.48 Harmonic input and output with an asymmetric spread function.

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

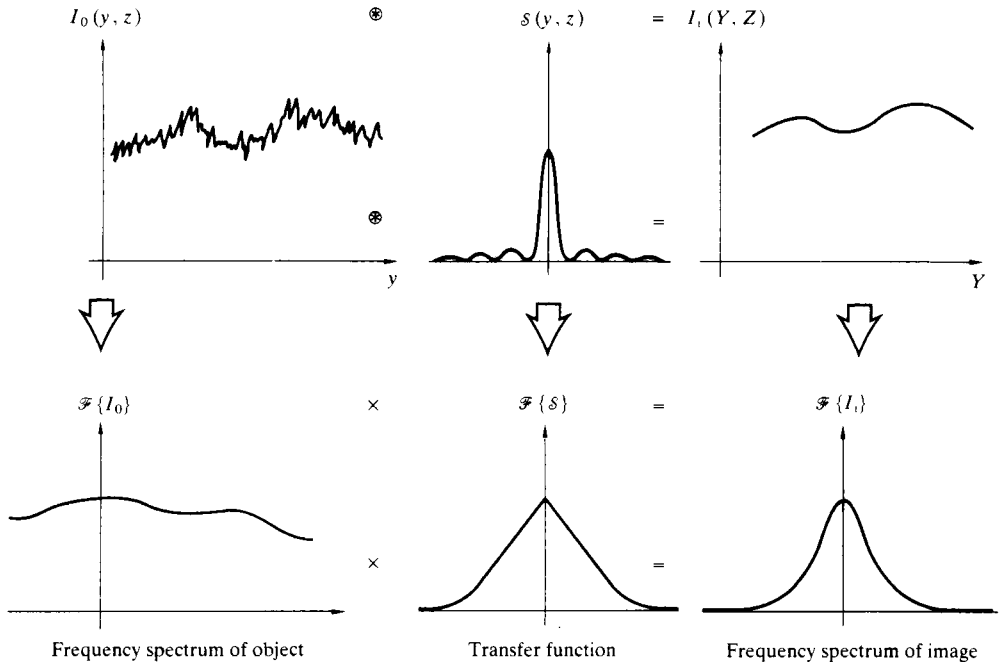


Figure 11.49 The relationships between the object and image spectra by way of the OTF, and the object and image irradiances by way of the point-spread function—all in incoherent illumination.

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,

$$\mathcal{F}\{I_i(Y,Z)\} = \mathcal{F}\{I_0(y,z)\} \cdot \mathcal{F}\{s(y,z)\} \tag{11.91}$$

*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).

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.

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

$$\mathcal{S}_n(z) = \frac{\mathcal{S}(z)}{\int_{-\infty}^{+\infty} \mathcal{S}(z)\,dz} \tag{11.100}$$

while the **normalized OTF** is

$$T(k_Z) \equiv \frac{\mathcal{F}\{\mathcal{S}(z)\}}{\int_{-\infty}^{+\infty} \mathcal{S}(z)\,dz} = \mathcal{F}\{\mathcal{S}_n(z)\} \tag{11.101}$$

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 $= M(k_Z)$.

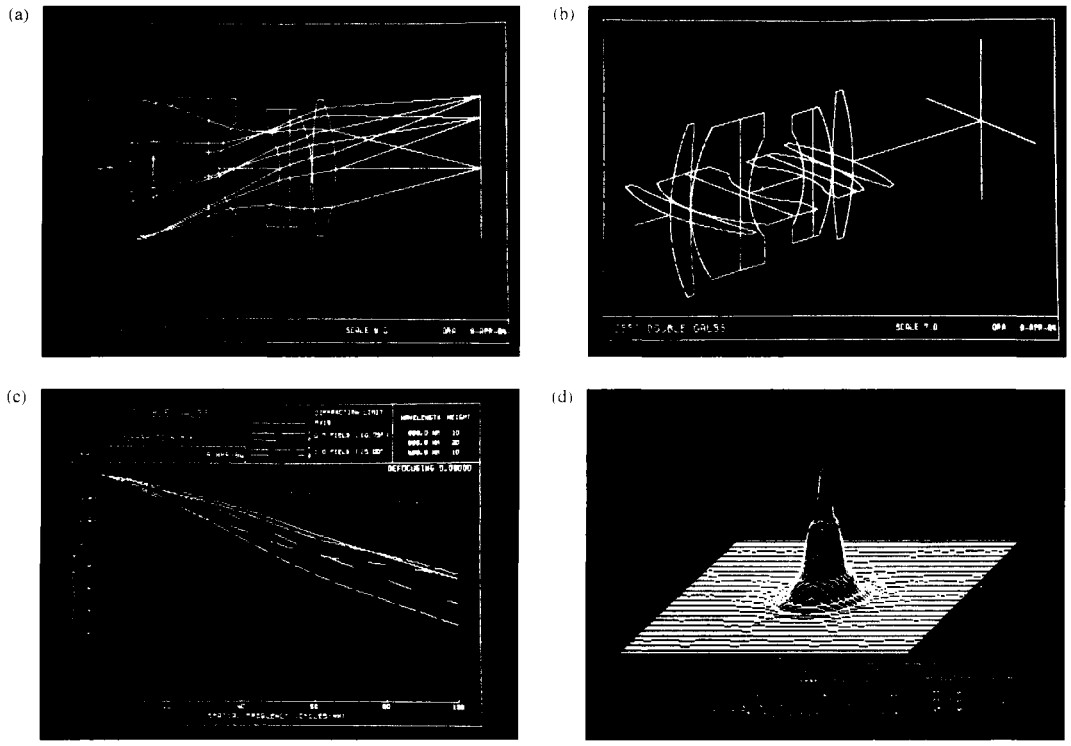


Figure 11.50 An example of the kind of lens design information available via computer techniques. (Photos courtesy Optical Research Associates.)

This discussion is only an introductory one designed more as a strong foundation than a complete structure. There are many other insights to be explored, such as the relationship between the autocorrelation of the pupil function and the OTF, and from there, the means of computing and measuring transfer functions (Fig. 11.50)—but for this the reader is directed to the literature.*

*See the series of articles "The Evolution of the Transfer Function," by F. Abbott, beginning in March 1970 in *Optical Spectra*; the articles "Physical Optics Notebook," by G. B. Parrent, Jr., and B. J. Thompson, beginning in December 1964, in the *S.P.I.E. Journal*, Vol. 3; or "Image Structure and Transfer," by K. Sayanagi, 1967, available from the Institute of Optics, University of Rochester. A number of books are worth consulting for practical emphasis, e.g., *Modern Optics*, by E. Brown; *Modern Optical Engineering*, by W. Smith; and *Applied Optics*, by L. Levi. In all of these, be careful of the sign convention in the transforms.

PROBLEMS

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

11.1 Determine the Fourier transform of the function

$$E(x) = \begin{cases} E_0 \sin k_p x & |x| < L \\ 0 & |x| > L \end{cases}$$

Make a sketch of $\mathcal{F}\{E(x)\}$. Discuss its relationship to Fig. 11.11.

11.2* Determine the Fourier transform of

$$f(x) = \begin{cases} \sin^2 k_p x & |x| < L \\ 0 & |x| > L \end{cases}$$

Make a sketch of it.

11.3 Determine the Fourier transform of

$$f(t) = \begin{cases} \cos^2 \omega_p t & |t| < T \\ 0 & |t| > T \end{cases}$$

Make a sketch of $F(\omega)$, then sketch its limiting form as $T \to \pm\infty$.

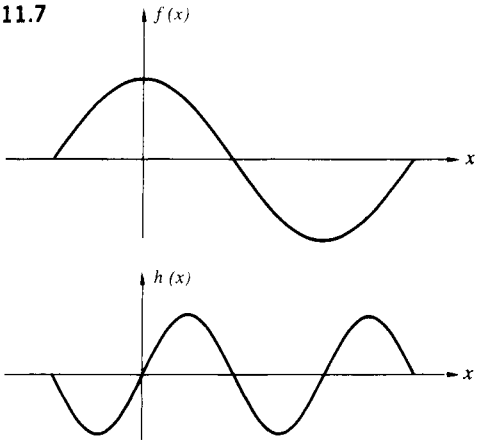
11.4* Show that $\mathcal{F}\{1\} = 2\pi\delta(k)$.

11.5* Determine the Fourier transform of the function $f(x) = A\cos k_0 x$.

11.6 Given that $\mathcal{F}\{f(x)\} = F(k)$ and $\mathcal{F}\{h(x)\} = H(k)$, if a and b are constants, determine $\mathcal{F}\{af(x) + bh(x)\}$.

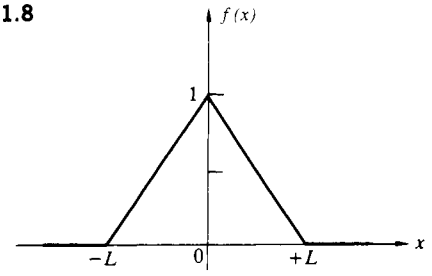
11.7* Figure P.11.7 shows two periodic functions, $f(x)$ and $h(x)$, which are to be added to produce $g(x)$. Sketch $g(x)$; then draw diagrams of the real and imaginary frequency spectra, as well as the amplitude spectra for each of the three functions.

Figure P.11.7



11.8 Compute the Fourier transform of the triangular pulse shown in Fig. P.11.8. Make a sketch of your answer, labeling all the pertinent values on the curve.

Figure P.11.8



11.9* Given that $\mathcal{F}\{f(x)\} = F(k)$, introduce a constant scaling factor $1/a$ and determine the Fourier transform of $f(x/a)$. Show that the transform of $f(-x)$ is $F(-k)$.

11.10* Show that the Fourier transform of the transform, $\mathcal{F}\{f(x)\}$, equals $2\pi f(-x)$, and that this is not the inverse transform of the transform, which equals $f(x)$. This problem was suggested by Mr. D. Chapman while a student at the University of Ottawa.

11.11* The rectangular function is often defined as

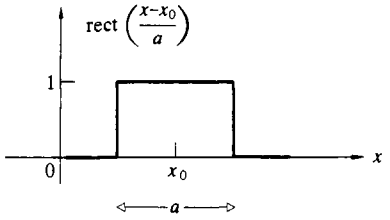
$$\text{rect}\left|\frac{x - x_0}{a}\right| = \begin{cases} 0, & |(x - x_0)/a| > \frac{1}{2} \\ \frac{1}{2}, & |(x - x_0)/a| = \frac{1}{2} \\ 1, & |(x - x_0)/a| < \frac{1}{2} \end{cases}$$

where it is set equal to $\frac{1}{2}$ at the discontinuities (Fig. P.11.11). Determine the Fourier transform of

$$f(x) = \text{rect}\left|\frac{x - x_0}{a}\right|$$

Notice that this is just a rectangular pulse, like that in Fig. 11.1*b*, shifted a distance x_0 from the origin.

Figure P.11.11



11.12* With the last two problems in mind, show that $\mathcal{F}\{(1/2\pi)\,\text{sinc}\,(\frac{1}{2}x)\} = \text{rect}(k)$, starting with the knowledge that $\mathcal{F}\{\text{rect}(x)\} = \text{sinc}\,(\frac{1}{2}k)$, in other words, Eq. (7.58) with $L = a$, where $a = 1$.

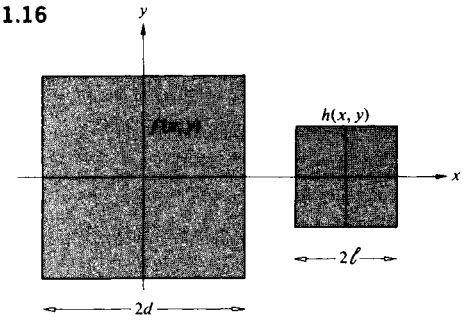
11.13* Utilizing Eq. (11.38), show that $\mathcal{F}^{-1}\{\mathcal{F}\{f(x)\}\} = f(x)$.

11.14* Given $\mathcal{F}\{f(x)\}$, show that $\mathcal{F}\{f(x - x_0)\}$ differs from it only by a linear phase factor.

11.15 Prove that $f \circledast h = h \circledast f$ directly. Now do it using the convolution theorem.

11.16* Suppose we have two functions, $f(x, y)$ and $h(x, y)$, where both have a value of 1 over a square region in the xy-plane and are zero everywhere else (Fig. P.11.16). If $g(X, Y)$ is their convolution, make a plot of $g(X, 0)$.

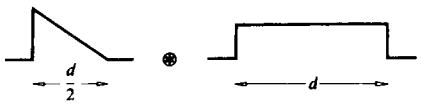
Figure P.11.16



11.17 Referring to the previous problem, justify the fact that the convolution is zero for $|X| \geq d + \ell$ when h is viewed as a spread functio00n.

11.18* Use the method illustrated in Fig. 11.23 to convolve the two functions depicted in Fig. P.11.18.

Figure P.11.18



11.19 Given that $f(x) \circledast h(x) = g(X)$, show that after shifting one of the functions an amount x_0, we get $f(x - x_0) \circledast h(x) = g(X - x_0)$.

11.20* Prove analytically that the convolution of any function $f(x)$ with a delta function, $\delta(x)$, generates the original function $f(X)$. You might make use of the fact that $\delta(x)$ is even.

11.21 Prove that $\delta(x - x_0) \circledast f(x) = f(X - x_0)$ and discuss the meaning of this result. Make a sketch of two appropriate functions and convolve them. Be sure to use an asymmetrical $f(x)$.

11.22* Show that $\mathcal{F}\{f(x) \cos k_0 x\} = [F(k - k_0) + F(k + k_0)]/2$ and that $\mathcal{F}\{f(x) \sin k_0 x\} = [F(k - k_0) - F(k + k_0)]/2i$.

11.23* Figure P.11.23 shows two functions. Convolve them graphically and draw a plot of the result.

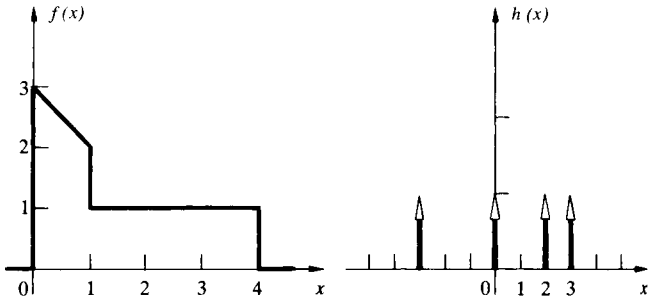
11.24 Given the function

$$f(x) = \text{rect}\left|\frac{x - a}{a}\right| + \text{rect}\left|\frac{x + a}{a}\right|$$

determine its Fourier transform. (See Problem 11.11.)

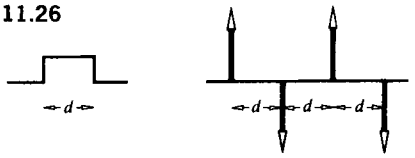
11.25 Given the function $f(x) = \delta(x + 3) + \delta(x - 2) + \delta(x - 5)$, convolve it with the arbitrary function $h(x)$.

Figure P.11.23



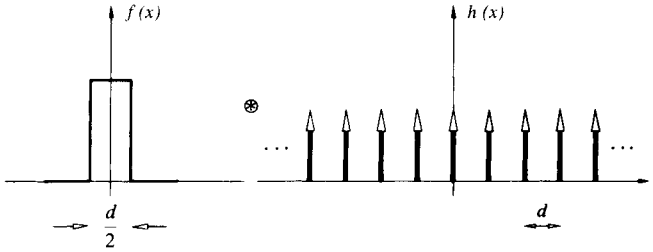
11.26* Make a sketch of the function arising from the convolution of the two functions depicted in Fig. P.11.26.

Figure P.11.26



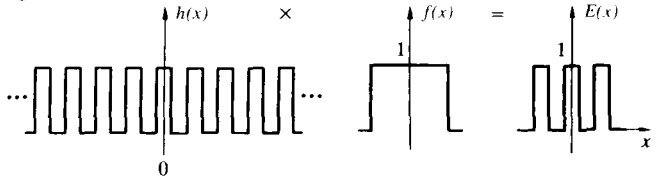
11.27* Figure P.11.27 depicts a *rect* function (as defined above) and a periodic *comb* function. Convolve the two to get $g(x)$. Now sketch the transform of each of these functions against spatial frequency $k/2\pi = 1/\lambda$. Check your results with the convolution theorem. Label all the relevant points on the horizontal axes in terms of d—like the zeros of the transform of $f(x)$.

Figure P.11.27



11.28 Figure P.11.28 shows, in one dimension, the electric field across an illuminated aperture consisting of several opaque bars forming a grating. Considering it to be created by taking the product of a periodic rectangular wave $h(x)$ and a unit rectangular function $f(x)$, sketch the resulting electric field in the Fraunhofer region.

Figure P.11.28



11.29 Show (for normally incident plane waves) that if an aperture has a center of symmetry (i.e., if the aperture function is even), then the diffracted field in the Fraunhofer case also possesses a center of symmetry.

11.30 Suppose a given aperture produces a Fraunhofer field pattern $E(Y, Z)$. Show that if the aperture's dimensions are altered such that the aperture function goes from $\mathcal{A}(y, z)$ to $\mathcal{A}(\alpha y, \beta z)$, the newly diffracted field will be given by

$$E'(Y, Z) = \frac{1}{\alpha\beta} E\left(\frac{Y}{\alpha}, \frac{Z}{\beta}\right)$$

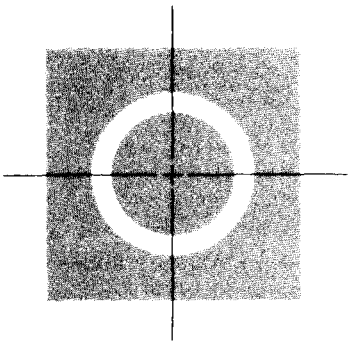
11.31 Show that when $f(t) = A \sin(\omega t + \varepsilon)$, $C_{ff}(\tau) = (A^2/2)\cos \omega t$, which confirms the loss of phase information in the autocorrelation.

11.32 Suppose we have a single slit along the y-direction of width b where the aperture function is constant across it at a value of $\mathcal{A}_0$. What is the diffracted field if we now apodize the slit with a cosine function amplitude mask? In other words, we cause the aperture function to go from $\mathcal{A}_0$ at the center to 0 at $\pm b/2$ via a cosinusoidal dropoff.

11.33* Show, from the integral definitions, that $f(x) \odot g(x) = f(x) \circledast g(-x)$.

11.34* Figure P.11.34 shows a transparent ring on an otherwise opaque mask. Make a rough sketch of its autocorrelation function, taking l to be the center-to-center separation against which you plot that function.

Figure P.11.34



11.35* Consider the function in Fig. 11.35 as a cosine carrier multiplied by an exponential envelope. Use the frequency convolution theorem to evaluate its Fourier transform.