

FOURIER OPTICS: CONCEPTS AND APPLICATIONS

Welcome to the exciting field of photonics. Chapter 1 is a quick tour of Fourier optics, a vital foundation for the chapters that follow. The branch of optics that can be analyzed by means of the Fourier transform is known as Fourier optics. The presentation of this subject is a condensed version of several texts on the subject [1–8]. This chapter starts with expressions for plane waves and a collection of special functions that are often used in photonics. The rest of the chapter is devoted to problems that can nicely be solved by Fourier optics, including various diffraction patterns, thin lenses, optical signal processing, spatial filters, and holography. A more rigorous derivation of the diffraction equations is also added at the end.

1.1 PLANE WAVES AND SPATIAL FREQUENCY

The representation of plane waves is introduced first, followed by a discussion of spatial frequency.

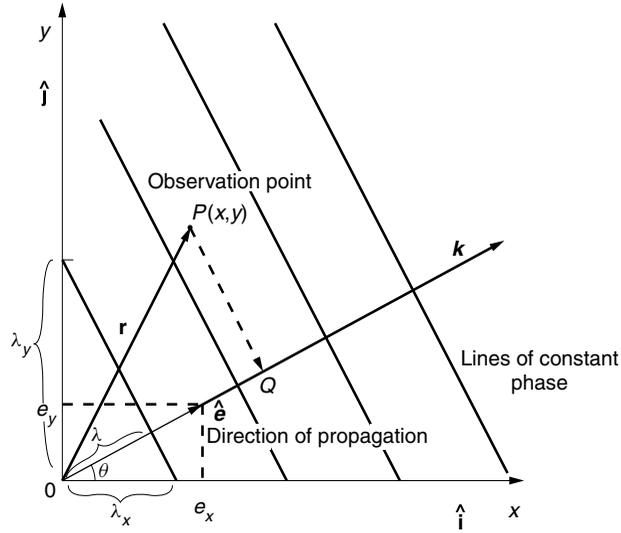
1.1.1 Plane Waves

The expression for a plane wave propagating in an arbitrary direction when observed at an arbitrary point in space will be derived. Let's first restrict ourselves to a two-dimensional (2D) vacuum medium such as shown in Fig. 1.1a.

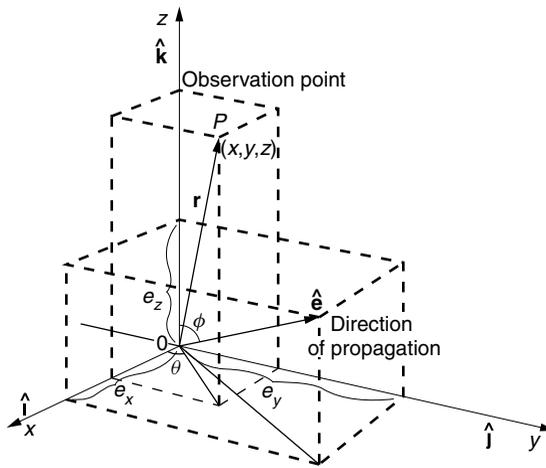
Let a plane wave observed at the origin at time t be expressed as

$$\mathbf{E}(0, 0, t) = \mathbf{E}_0(0, 0)e^{-j\omega t} \quad (1.1)$$

where the vector $\mathbf{E}_0(0, 0)$ represents the amplitude and direction of polarization, and ω represents the angular frequency. In this book the sign convention of $e^{-j\omega t}$ rather than $e^{j\omega t}$ is used. The direction of propagation associated with the sign convention is discussed in the boxed note.



(a)



(b)

Figure 1.1 Expression for a plane wave. (a) A plane wave propagating in the \hat{e} direction and observed at $P(x, y)$ in 2D space. (b) The position vector \mathbf{r} and direction \hat{e} of propagation of a plane wave in 3D space.

The direction of propagation is expressed by the unit vector

$$\hat{e} = e_x \hat{i} + e_y \hat{j} \tag{1.2}$$

where

$$e_x = \cos \theta, \quad e_y = \sin \theta \tag{1.3}$$

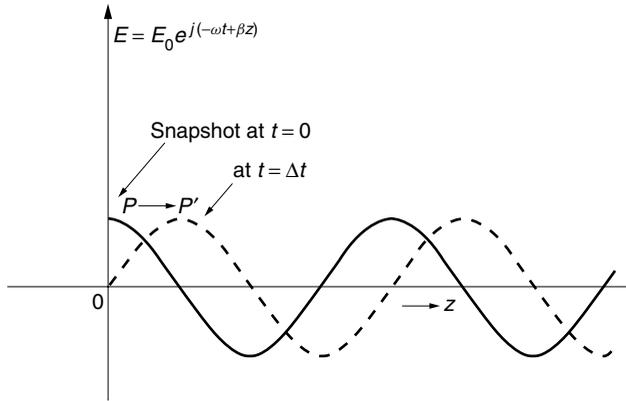
and \hat{i} and \hat{j} are unit vectors in the x and y directions, respectively.

There are two equally valid conventions for expressing the time dependence of the electric wave; they are $e^{j\omega t}$ and $e^{-j\omega t}$. Depending on the choice of convention, the wave expression is different.

Let's take the expression

$$E = \text{Re} [E_0 e^{j(-\omega t + \beta z)}] = E_0 \cos(-\omega t + \beta z)$$

as an example.



$E = E_0 e^{j(-\omega t + \beta z)}$ is a forward wave.

The peak occurs when the term within the cosine bracket equals zero. Therefore, when a snapshot of the wave is taken at $t = 0$, the location of the peak is at $z = 0$. Now, another snapshot is taken at Δt seconds later. The position Δz of the peak is again where the value of the bracketed term is zero. The peak has moved to a positive new location at $\Delta z = (\omega/\beta)\Delta t$. Thus, this equation represents a forward wave with the phase velocity

$$v_p = \omega/\beta$$

On the other hand, as time increases with

$$E = E_0 \cos(\omega t + \beta z)$$

the peak moves toward the negative z direction, $\Delta z = -(\omega/\beta)\Delta t$, and this represents the backward wave.

The rule is, whenever the signs of t and z are different, as in $E = E_0 e^{j(-\omega t + \beta z)}$ or $E = E_0 e^{j(\omega t - \beta z)}$, the waves are forward waves; and whenever the signs are the same, as in $E = E_0 e^{j(\omega t + \beta z)}$ or $E = E_0 e^{j(-\omega t - \beta z)}$, the waves are backward waves. In this book the convention of $e^{-j\omega t}$ is used, unless otherwise stated, because the forward wave $E = E_0 e^{j(-\omega t + \beta z)}$ has a positive sign on the z .

Let us now observe this wave from the observation point $P(x, y)$, which is connected to the origin by the position vector \mathbf{r} ,

$$\mathbf{r} = x\hat{\mathbf{i}} + y\hat{\mathbf{j}} \quad (1.4)$$

\overline{OQ} is the distance between the origin and the projection of point P in the direction of $\hat{\mathbf{e}}$ as shown in Fig. 1.1a. If the field travels at the speed of light, the field observed at point P is that which has left the origin \overline{OQ}/c seconds ago. \overline{OQ} can be represented by the scalar product $\hat{\mathbf{e}} \cdot \mathbf{r}$. Thus, the field $E(x, y)$ observed at P is

$$\mathbf{E}(x, y, t) = \mathbf{E}_0(0, 0)e^{-j\omega(t-\hat{\mathbf{e}}\cdot\mathbf{r}/c)} \quad (1.5)$$

Equation (1.5) is rewritten as

$$\mathbf{E}(x, y, t) = \mathbf{E}_0(0, 0)e^{-j\omega t + j\mathbf{k}\cdot\mathbf{r}} \quad (1.6)$$

where

$$\mathbf{k} = \frac{\omega}{c}\hat{\mathbf{e}} = \frac{2\pi}{\lambda}\hat{\mathbf{e}} \quad (1.7)$$

and where λ is the wavelength in vacuum. \mathbf{k} is called the *vector propagation constant*. Using Eq. (1.2) the vector propagation constant \mathbf{k} can be expanded in Cartesian coordinates as

$$\mathbf{k} = \frac{2\pi}{\lambda} \cos \theta \hat{\mathbf{i}} + \frac{2\pi}{\lambda} \sin \theta \hat{\mathbf{j}} \quad (1.8)$$

Equation (1.6) is the expression for a plane wave propagating at a speed of c in the $\hat{\mathbf{e}}$ direction observed at point P . So far, we have assumed a refractive index equal to 1. Inside a linear medium with refractive index n , the frequency does not change, but the wavelength is reduced to λ/n and the speed of propagation is slowed down to a velocity of $v = c/n$. Hence, the propagation constant becomes $n\mathbf{k}$.

1.1.2 Spatial Frequency

Next, the vector propagation constant \mathbf{k} will be rewritten in terms of spatial frequency. It is important to note that spatial frequency is different from temporal frequency. Temporal frequency f is defined as the number of wavelengths that pass through a particular point in space per unit time, whereas spatial frequency f_s is defined as the number of wavelengths in a unit of distance:

$$f_s = \frac{1}{\lambda} \quad (1.9)$$

The most popular unit of f_s is lines/mm or lines/m. In the field of spectroscopy, f_s is called the wavenumber with units of cm^{-1} .

The relationship between the spatial and temporal frequencies of a plane wave is

$$f = cf_s \quad (1.10)$$

Equation (1.10) was obtained by comparing Eq. (1.9) with the temporal frequency

$$f = \frac{c}{\lambda} \quad (1.11)$$

The spatial frequency depends on the direction in which the unit distance is taken.

Referring to Fig. 1.1a, we see that λ_x and λ_y are

$$\begin{aligned}\lambda_x &= \frac{\lambda}{\cos \theta} \\ \lambda_y &= \frac{\lambda}{\sin \theta}\end{aligned}\tag{1.12}$$

The spatial frequencies in these directions are

$$\begin{aligned}f_x &= \frac{1}{\lambda_x} = e_x f_s \\ f_y &= \frac{1}{\lambda_y} = e_y f_s \\ f_s &= \frac{1}{\lambda}\end{aligned}\tag{1.13}$$

where Eq. (1.3) was used.

Earlier, we derived the propagation vector \mathbf{k} to be

$$\mathbf{k} = 2\pi \left(\frac{\cos \theta}{\lambda} \mathbf{i} + \frac{\sin \theta}{\lambda} \mathbf{j} \right)$$

in Eq. (1.8). This can now be expressed more elegantly using Eq. (1.12) as

$$\mathbf{k} = 2\pi \left(\frac{1}{\lambda_x} \hat{\mathbf{i}} + \frac{1}{\lambda_y} \hat{\mathbf{j}} \right) = 2\pi (f_x \hat{\mathbf{i}} + f_y \hat{\mathbf{j}})$$

where, to repeat, f_x and f_y are spatial frequencies. From this, it follows that $\mathbf{k} \cdot \mathbf{r}$ is expressed as

$$\mathbf{k} \cdot \mathbf{r} = 2\pi f_x x + 2\pi f_y y\tag{1.14}$$

From Eq. (1.12), we obtain

$$\left(\frac{1}{\lambda} \right)^2 = \left(\frac{1}{\lambda_x} \right)^2 + \left(\frac{1}{\lambda_y} \right)^2$$

which can be rewritten as

$$f_s^2 = f_x^2 + f_y^2\tag{1.15}$$

Finally, the two-dimensional expression will be extended into three dimensions by adding a unit vector $\hat{\mathbf{k}}$ in the z direction. The direction of propagation $\hat{\mathbf{e}}$ becomes

$$\hat{\mathbf{e}} = e_x \hat{\mathbf{i}} + e_y \hat{\mathbf{j}} + e_z \hat{\mathbf{k}}\tag{1.16}$$

With the coordinates shown in Fig. 1.1b, the components of the unit vector $\hat{\mathbf{e}}$ are

$$\begin{aligned} e_x &= \sin \phi \cos \theta \\ e_y &= \sin \phi \sin \theta \\ e_z &= \cos \phi \end{aligned} \quad (1.17)$$

With $\phi \neq 90^\circ$ in Eq. (1.17), the expression corresponding to Eq. (1.8) is

$$\mathbf{k} = 2\pi \frac{\cos \theta}{\lambda} \sin \phi \hat{\mathbf{i}} + 2\pi \frac{\sin \theta}{\lambda} \sin \phi \hat{\mathbf{j}} + 2\pi \frac{\cos \phi}{\lambda} \hat{\mathbf{k}} \quad (1.18)$$

and

$$\begin{aligned} \lambda_x &= \frac{\lambda}{\cos \theta \sin \phi} \\ \lambda_y &= \frac{\lambda}{\sin \theta \sin \phi} \\ \lambda_z &= \frac{\lambda}{\cos \phi} \end{aligned} \quad (1.19)$$

and in terms of spatial frequencies,

$$\begin{aligned} f_x &= \frac{1}{\lambda_x} = f_s e_x \\ f_y &= \frac{1}{\lambda_y} = f_s e_y \\ f_z &= \frac{1}{\lambda_z} = f_s e_z \\ f_s &= \frac{1}{\lambda} \end{aligned} \quad (1.20)$$

The three-dimensional (3D) position vector \mathbf{r} is

$$\mathbf{r} = x\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\hat{\mathbf{k}} \quad (1.21)$$

Similar to the two-dimensional case, \mathbf{E} is expressed as

$$\mathbf{E}(x, y, z, t) = \mathbf{E}(0, 0, 0)e^{-j\omega t + j\mathbf{k} \cdot \mathbf{r}} \quad (1.22)$$

where

$$\mathbf{k} \cdot \mathbf{r} = 2\pi f_x x + 2\pi f_y y + 2\pi f_z z \quad (1.23)$$

and where

$$f_s^2 = f_x^2 + f_y^2 + f_z^2 \quad (1.24)$$

Example 1.1 For a plane wave propagating in the direction

$$\theta = 30^\circ, \quad \phi = 45^\circ$$

find the expression for the field when observed at point $P(1, 2, 3) \times 10^{-6}$ m. The free-space wavelength is $\lambda = 1.55 \mu\text{m}$. The vector \mathbf{E}_0 representing the polarization is $\hat{\mathbf{i}} + 2\hat{\mathbf{j}} - 1.86\hat{\mathbf{k}}$.

Solution From the value $\mathbf{k} \cdot \mathbf{r}$ in Eq. (1.23) and from Eqs. (1.19) and (1.20), the x component is

$$\begin{aligned} k_x x &= 2\pi f_x x \\ &= \frac{2\pi}{\lambda} e_x x \\ &= \frac{2\pi}{\lambda} \sin \phi \cos \theta \cdot x \\ &= 2.48 \end{aligned}$$

Similarly, the y and z components are

$$k_y y = 2.86$$

$$k_z z = 8.60$$

and

$$\omega = \frac{2\pi c}{\lambda} = 1.22 \times 10^{15} \text{ rad/s}$$

From Eq. (1.22), the expression for \mathbf{E} is

$$\mathbf{E}[(1, 2, 3) \times 10^{-6}, t] = (\hat{\mathbf{i}} + 2\hat{\mathbf{j}} - 1.86\hat{\mathbf{k}}) e^{j13.94 - j1.22 \times 10^{15} t} \quad \square$$

Example 1.2 A plane wave propagating in a given medium is expressed as

$$\mathbf{E}(x, y, z, t) = \mathbf{E}_0 e^{j(2x+3y+4z) \times 10^6 - j10^{15} t} \quad (1.25)$$

- Find the *unit vector* for the direction of propagation.
- What are the *values of θ and ϕ* that characterize the direction of propagation?
- Find the refractive index of the medium.
- Find the vector expression of \mathbf{E}_0 of Eq. (1.25), assuming that \mathbf{E}_0 is polarized in the $x = x_1$ plane and the amplitude is 5.0.

Solution

(a) The direction of the vector parallel to the propagation direction is $2\hat{\mathbf{i}} + 3\hat{\mathbf{j}} + 4\hat{\mathbf{k}}$. The unit vector is found by normalizing this vector. The result is

$$\hat{\mathbf{e}} = \frac{1}{\sqrt{29}} (2\hat{\mathbf{i}} + 3\hat{\mathbf{j}} + 4\hat{\mathbf{k}})$$

and therefore

$$\hat{\mathbf{e}} = 0.37\hat{\mathbf{i}} + 0.56\hat{\mathbf{j}} + 0.74\hat{\mathbf{k}}$$

(b) The direction of the unit vector in terms of θ and ϕ can be solved for by using Eq. (1.17):

$$\theta = 56.3^\circ$$

$$\phi = 42.0^\circ$$

(c) In order to determine the refractive index n , the velocity of the wave needs to be determined since $n = c/v$. The velocity is expressed by the product of the temporal frequency and the wavelength or $v = f\lambda$. The first part of the exponential term of Eq. (1.25) contains the wavelength information whereas the second term contains the temporal frequency information.

To solve for the wavelength, the spatial frequency f_s is first obtained from its components f_x , f_y , and f_z . Explicitly,

$$\mathbf{k} \cdot \mathbf{r} = 2\pi(f_x x + f_y y + f_z z) = (2x + 3y + 4z) \times 10^6$$

where

$$f_x = \frac{2 \times 10^6}{2\pi} \text{ lines/m}$$

$$f_y = \frac{3 \times 10^6}{2\pi} \text{ lines/m}$$

$$f_z = \frac{4 \times 10^6}{2\pi} \text{ lines/m}$$

and hence $f_s = 0.86 \times 10^6$ lines/m from the sum of the squares.

The inverse of the spatial frequency is the wavelength and is equal to

$$\lambda = 1.17 \times 10^{-6} \text{ m}$$

From the second part of the exponential term in Eq. (1.25) we now obtain the temporal frequency, which is

$$f = \frac{10^{15}}{2\pi} = 1.59 \times 10^{14} \text{ Hz}$$

Hence, the phase velocity v is

$$v = \lambda f = 1.86 \times 10^8 \text{ m/s}$$

Finally, from the velocity we obtain the refractive index

$$n = \frac{c}{v} = 1.61$$

(d) Let the amplitude vector be

$$\mathbf{E}_0 = (a\hat{\mathbf{i}} + b\hat{\mathbf{j}} + c\hat{\mathbf{k}})$$

With a plane wave, the direction of polarization is perpendicular to the direction of propagation. Thus, their dot product should be zero:

$$\mathbf{E}_0 \cdot \hat{\mathbf{e}} = 0$$

Since \mathbf{E}_0 is polarized in the $x = x_1$ plane, it follows that

$$a = 0$$

and

$$e_y b + e_z c = 0$$

From Eq. (1.25), and the given magnitude of 5, a pair of equations are obtained that can be solved for b and c , namely,

$$3b + 4c = 0$$

$$b^2 + c^2 = 5^2$$

Hence, the vector expression of \mathbf{E}_0 is

$$\mathbf{E}_0 = \pm(4\hat{\mathbf{j}} - 3\hat{\mathbf{k}}) \quad \square$$

1.2 FOURIER TRANSFORM AND DIFFRACTION PATTERNS IN RECTANGULAR COORDINATES

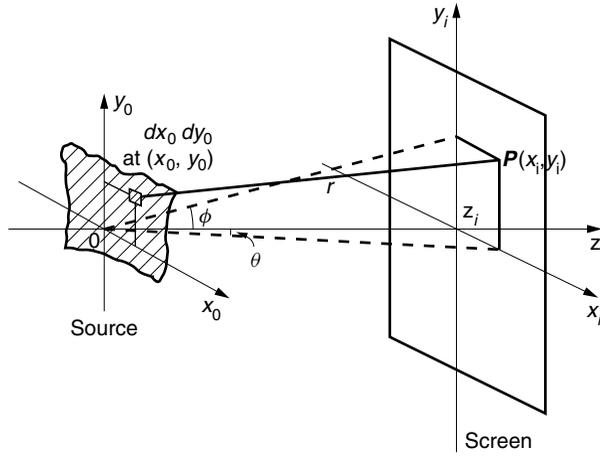
Referring to Fig. 1.2a, the field distribution from a source is observed on a screen. The field distribution on the screen is called a diffraction pattern.

We will demonstrate that the diffraction pattern can be elegantly expressed by the Fourier transform of the source. Let $E(x_i, y_i)$ represent the field at point P on the screen placed a distance z_i away from the source field $E(x_0, y_0)$. The distributed source $E(x_0, y_0)$ is considered as an ensemble of point sources. Each point source radiates a spherical wave. The field at the observation point P is comprised of contributions from an ensemble of fields radiated from all the point sources. The contribution of the point source located at (x_0, y_0) to point P at (x_i, y_i) is

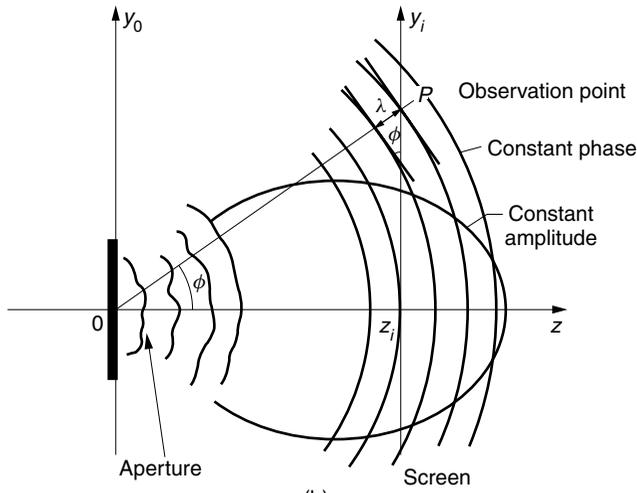
$$dE(x_i, y_i) = \frac{e^{jkr}}{r} E(x_0, y_0) dx_0 dy_0 \quad (1.26)$$

where $E(x_0, y_0)$ is the magnitude of the point source located at (x_0, y_0) and r is the distance between (x_0, y_0) and (x_i, y_i) . The distance r is expressed as

$$r = \sqrt{z_i^2 + (x_i - x_0)^2 + (y_i - y_0)^2} \quad (1.27)$$



(a)



(b)

Figure 1.2 Field distribution from a source observed on a screen. (a) Geometry. (b) Distribution of the field in the (y_i, z_i) plane.

The contribution of the spherical waves from all the point sources to $E(x_i, y_i)$ is

$$E(x_i, y_i) = K \iint \frac{e^{jkr}}{r} E(x_0, y_0) dx_0 dy_0 \tag{1.28}$$

This equation is known as the Fresnel–Kirchhoff diffraction formula. The amplitude of the diffracted field is inversely proportional to its wavelength and is expressed as

$$K = \frac{1}{j\lambda} \tag{1.29}$$

The constant K will be derived later in Appendix A of Volume I.

If the point of observation is far enough away or in the vicinity of the z axis (paraxial),

$$z_i^2 \gg (x_i - x_0)^2 + (y_i - y_0)^2 \quad (1.30)$$

then the distance r can be simplified by the binomial expansion as

$$r \doteq z_i \left(1 + \frac{(x_i - x_0)^2 + (y_i - y_0)^2}{2z_i^2} \right) \quad (1.31)$$

which can be rewritten as

$$r \cong z_i + \frac{x_i^2 + y_i^2}{2z_i} - \frac{x_i x_0 + y_i y_0}{z_i} + \frac{x_0^2 + y_0^2}{2z_i} \quad (1.32)$$

The region of z_i for which the approximate expression Eq. (1.32) is valid is called the Fresnel region or the near-field region. As the distance is further increased in the z direction, the last term in Eq. (1.32) becomes negligible for the finite size of the source. This region of z_i is called the Fraunhofer region or far-field region. In this chapter we are concerned about the far-field. In the far-field region, the approximation for r is

$$r = z_i + \frac{x_i^2 + y_i^2}{2z_i} - \frac{x_i x_0 + y_i y_0}{z_i} \quad (1.33)$$

By substituting this approximation into the exponential term of the Fresnel–Kirchhoff diffraction formula, Eq. (1.28), the field becomes

$$E(x_i, y_i) = \frac{1}{j\lambda z_i} e^{jk[z_i + (x_i^2 + y_i^2)/2z_i]} \iint_{-\infty}^{\infty} E(x_0, y_0) e^{-j2\pi(f_x x_0 + f_y y_0)} dx_0 dy_0$$

with $f_x = \frac{x_i}{\lambda z_i}$, $f_y = \frac{y_i}{\lambda z_i}$ (1.34)

We recognize that the integral is the two-dimensional Fourier transform of the field in the x, y domain into the f_x, f_y domain:

$$\mathcal{F}\{g(x, y)\} = \iint_{-\infty}^{\infty} g(x, y) e^{-j2\pi(f_x x + f_y y)} dx dy \quad (1.35)$$

Or in mathematical terms, the diffraction pattern is

$$E(x_i, y_i) = \frac{1}{j\lambda z_i} e^{jk[z_i + (x_i^2 + y_i^2)/2z_i]} \mathcal{F}\{E(x_0, y_0)\}_{f_x = x_i/\lambda z_i, f_y = y_i/\lambda z_i} \quad (1.36)$$

where \mathcal{F} denotes the Fourier transform.

In short, *the Fraunhofer diffraction pattern is the Fourier transform of the source field.*

Sometimes, the angular distribution rather than the planar distribution is desired. For this case, the azimuth angle ϕ and the elevation angle θ , measured with respect

to the center of the source field, are approximated as $\sin \theta \doteq x_i/z_i$ and $\sin \phi \doteq y_i/z_i$. Hence, the values in Eq. (1.34) are

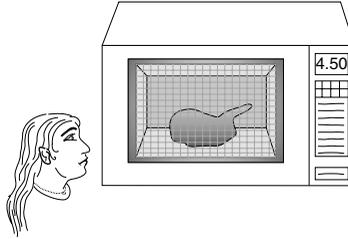
$$f_x = \frac{\sin \theta}{\lambda}, \quad f_y = \frac{\sin \phi}{\lambda} \quad (1.37)$$

The branch of optics that can be analyzed by means of the Fourier transform is categorized as Fourier optics. First, the physical meaning of f_x and f_y in Eq. (1.34) will be explored. For simplicity, only the distribution in the (y_i, z_i) plane will be considered. Figure 1.2b shows a typical phase and amplitude distribution of the field diffracted from an aperture source whose dimensions are much smaller than the distance to the screen.

In the region far from the aperture, the phase distribution is more like that of a spherical wave. With the source placed at the origin, the phase front along the y_i axis near $y_i = 0$ is always parallel to the y_i axis. In the vicinity of this point, there is no variation in the phase of the field with respect to y_i . Hence, the field has zero spatial frequency at $y_i = 0$. (The variation of the field amplitude with y_i is normally much slower than that of the phase.) As shown in Fig. 1.2b, the change in the variation of phase increases as the point of observation P moves along the y_i axis, such that, eventually, the wavelength λ_{y_i} measured along the y_i axis will approach the wavelength of free space.

Mathematically, λ_{y_i} at the observation point $P(y_i, z_i)$ is, from the geometry in Fig. 1.2b,

$$\lambda_{y_i} = \frac{\lambda}{\sin \phi}$$



Microwave oven door shields us from microwaves but not from light.

How is it that a microwave oven door shields us from microwaves but not from light waves, allowing us to see our food as it cooks? The answer is found by examining the Fresnel–Kirchhoff diffraction formula applied to the mesh of our microwave oven door. The mesh is equivalent to a series of equally spaced apertures. Hence, this causes diffraction to occur, and the Fresnel–Kirchhoff equation, Eq. (1.34), is used to calculate the diffracted field.

Let us calculate the amplitude of the microwave and light wave field observed at the center: $x_i = y_i = 0$. The integral in Eq. (1.34) is the same for both microwaves and light, but as shown by the factor K of Eq. (1.29), the amplitude of the field is inversely proportional to the wavelength. Thus, the amplitude of the microwaves is about 10,000 times smaller than that of light. Thank goodness for the presence of K .

The field located at P , therefore, has the spatial frequency of

$$f_{y_i} = \frac{\sin \phi}{\lambda} \cong \frac{y_i}{z_i \lambda}$$

If we wish to construct a spatial frequency filter to pick selectively a particular f_{y_i} spatial frequency, we can place an opaque screen in the (x_i, y_i) plane and poke a hole in it at a particular location. The location of this hole for a desired spatial frequency can be calculated by rearranging the above equation to give

$$y_i = f_{y_i} \lambda z_i$$

The usefulness of Eq. (1.36) extends throughout the electromagnetic spectrum. For example, the inverse of this calculation is used in X-ray crystallography. By knowing the X-ray diffraction pattern of a crystal, the structure of the crystal is found by its inverse Fourier transform. The same is true with radio astronomy. By probing $E(x_i, y_i)$, the radio radiation pattern of a star, the structure can be analyzed by the inverse Fourier transform in a similar manner. Yet another application is to use this relationship to find the radiation pattern of an antenna (Problem 1.4). The antenna radiation pattern for a given current distribution $I(x_0, y_0)$ can be obtained [9]. With a few simple substitutions, $E(x_0, y_0)$ is replaced by $\frac{1}{2} \eta I(x_0, y_0) \sin \theta$, where $\eta = 120\pi$ is the intrinsic impedance of free space, and $\sin \theta$ is necessary to convert E_z to E_θ , since antenna theory expresses its patterns in terms of E_θ .

Before closing this section, we will demonstrate that the Fresnel field or the *near* field can also be expressed in terms of the Fourier transform. As previously stated, the Fresnel region is valid when r is approximated with the addition of the quadratic phase factor term of the source in Eq. (1.32). By similarly substituting this value of r into the Fresnel–Kirchhoff diffraction formula, the field in the Fresnel region becomes

$$E(x_i, y_i) = \frac{1}{j\lambda z_i} e^{jk[z_i + (x_i^2 + y_i^2)/2z_i]} \mathcal{F} \left\{ \underbrace{E(x_0, y_0)}_{\text{Input}} \underbrace{e^{jk(x_0^2 + y_0^2)/2z_i}}_{\text{Part of the point spread function}} \right\}_{f_x = x_i/\lambda z_i, f_y = y_i/\lambda z_i} \quad (1.38)$$

Alternatively, the Fresnel field can be expressed elegantly as a convolution of two terms explained as follows. Recall that the approximation of r , the distance to the screen, is from Eq. (1.31)

$$r = z_i \left(1 + \frac{(x_i - x_0)^2 + (y_i - y_0)^2}{2z_i^2} \right)$$

When r is directly substituted into the Fresnel–Kirchhoff diffraction formula, Eq. (1.28), the result is

$$E(x_i, y_i) = j \frac{1}{\lambda z_i} \iint E(x_0, y_0) \{ e^{jkz_i [1 + (x_i - x_0)^2/2z_i^2 + (y_i - y_0)^2/2z_i^2]} \} dx_0 dy_0$$

The expression in the curly brackets is in the form of $f(x_i - x_0, y_i - y_0)$. From this observation, we note that the above expression takes on the form of a convolution:

$$g(x) * f(x) = \int g(\xi) f(x - \xi) d\xi$$

Thus,

$$E(x_i, y_i) = E_0(x_i, y_i) * f_{z_i}(x_i, y_i) \quad (1.39)$$

where

$$f_{z_i}(x_i, y_i) = \frac{1}{j\lambda z_i} e^{jk[z_i + (x_i^2 + y_i^2)/2z_i]} \quad (1.40)$$

The function $f_{z_i}(x_i, y_i)$ is called the point spread function (or impulse response function of free space) and is identical with the field at (x_i, y_i, z_i) when a point source is placed at the origin of the source coordinates. Whether Eq. (1.38) or (1.39) is used, the results are the same.

The Fourier transform F_{z_i} of the point spread function in Eq. (1.40) is

$$F_{z_i} = e^{jkz_i - j\pi\lambda z_i(f_x^2 + f_y^2)} \quad (1.41)$$

The derivation of Eq. (1.41) is found in the boxed note.

The Fourier transform F_{z_i} of the point spread function f_{z_i} , given by Eq. (1.41), will be derived. Let's start with the easy to remember Fourier transform [10]

$$\mathcal{F}\{e^{-\pi x^2}\} = e^{-\pi f^2}$$

The Fourier transform of this function is the original function itself.

First, the x_i component of Eq. (1.40) is rewritten as

$$e^{jkx_i^2/2z_i} = e^{[-\pi(x_i/\sqrt{j\lambda z_i})^2]}$$

The similarity theorem of the Fourier transform is

$$\mathcal{F}\{g(\alpha x)\} = \frac{1}{\alpha} G\left(\frac{f_x}{\alpha}\right) \quad (1.42)$$

where G is the Fourier transform of $g(x)$. If Eq. (1.42) is used, the Fourier transform of the x_i component is

$$\mathcal{F}\{e^{[-\pi(x_i/\sqrt{j\lambda z_i})^2]}\} = \sqrt{j\lambda z_i} e^{-j\pi\lambda z_i f_x^2}$$

For the two-dimensional case, the Fourier transform is

$$\mathcal{F}\{e^{jk(x_i^2 + y_i^2)/2z_i}\} = j\lambda z_i e^{-j\pi\lambda z_i(f_x^2 + f_y^2)} \quad (1.43)$$

Hence, the Fourier transform of the point spread function is

$$\mathcal{F}\left\{\frac{1}{j\lambda z_i} e^{jk[z_i + (x_i^2 + y_i^2)/2z_i]}\right\} = e^{jkz_i - j\pi\lambda z_i(f_x^2 + f_y^2)} \quad (1.44)$$

Example 1.3 A Fresnel diffraction pattern observed at a certain distance away from the source is identical with that obtained by a series of diffractions taking place successively from one fictitious plane after another up to the screen. This phenomenon is related to Huygens' principle. Prove this using the example of dividing the distance z_i into d_1 and d_2 .

Solution Let the source function be $E_0(x, y)$. The distance z_i to the screen is arbitrarily divided into two, d_1 and d_2 , with $z_i = d_1 + d_2$ as shown in Fig. 1.3. The diffraction pattern on a fictitious screen at distance d_1 will be used as the input for another diffraction pattern on a screen at an additional distance d_2 . This result will be compared with that obtained when a single diffraction pattern impinges directly from the input onto the screen at distance $z_i = d_1 + d_2$.

From Eq. (1.39), the diffraction pattern of $E_0(x, y)$ on the fictitious screen at d_1 is

$$E_1(x, y) = E_0(x, y) * f_{d_1}(x, y) \quad (1.45)$$

The diffraction pattern of $E_1(x, y)$ on the screen at an additional distance d_2 is

$$E_2(x_i, y_i) = E_0(x_i, y_i) * f_{d_1}(x_i, y_i) * f_{d_2}(x_i, y_i) \quad (1.46)$$

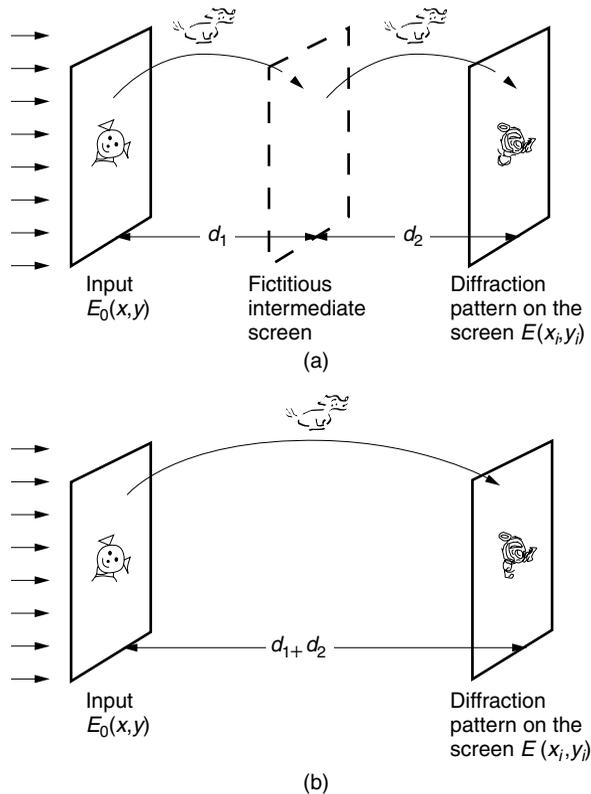


Figure 1.3 Comparison of diffraction patterns with and without a fictitious screen. (a) Two-step diffraction. (b) One-step diffraction.

The Fourier transform and inverse Fourier transform are successively performed to make use of the product rule of Fourier transforms. The result is

$$E_2(x_i, y_i) = \mathcal{F}^{-1}\{\varepsilon(f_x, f_y) \cdot F_{d_1}(f_x, f_y) \cdot F_{d_2}(f_x, f_y)\} \quad (1.47)$$

From Eq. (1.41),

$$E_2(x_i, y_i) = \mathcal{F}^{-1}\{\varepsilon(f_x, f_y)e^{jk(d_1+d_2)}e^{-j\pi\lambda(d_1+d_2)(f_x^2+f_y^2)}\} \quad (1.48)$$

where $\varepsilon(f_x, f_y)$ is the Fourier transform of $E_0(x, y)$. From Eq. (1.39), the one-step calculation of the diffraction pattern on the screen over the distance z_i is

$$E_2(x_i, y_i) = E_0(x_i, y_i) * f_{d_1+d_2}(x_i, y_i) \quad (1.49)$$

Thus, the result of the single diffraction is identical with successive diffractions. This fact conforms with Huygens' principle that a wave propagates by creating new wavefronts from the pattern of the old wavefront. \square

1.3 FOURIER TRANSFORM IN CYLINDRICAL COORDINATES

Photographic plates are rectangular, but most optical components like lenses, retarders, and apertures are cylindrically symmetric. The relationships between rectangular spatial coordinates (x, y) and spatial frequency coordinates (f_x, f_y) and cylindrical spatial coordinates (r, θ) and spatial frequency coordinates (ρ, ϕ) are

$$\begin{aligned} x &= r \cos \theta & f_x &= \rho \cos \phi \\ y &= r \sin \theta & f_y &= \rho \sin \phi \\ dx dy &= r dr d\theta & df_x df_y &= \rho d\rho d\phi \end{aligned} \quad (1.50)$$

This relationship is illustrated in Fig. 1.4. The two-dimensional Fourier transform in rectangular coordinates

$$G(f_x, f_y) = \iint_{-\infty}^{\infty} g(x, y)e^{-j2\pi(f_x x + f_y y)} dx dy \quad (1.51)$$

is converted into cylindrical coordinates using the relationships of Eq. (1.50) as

$$G(\rho, \phi) = \int_0^{\infty} \int_0^{2\pi} g(r, \theta)e^{-j2\pi\rho r \cos(\theta-\phi)} r dr d\theta \quad (1.52)$$

In order to simplify the calculation of the double integral, $g(r, \theta)$ is first separated into functions of r and θ . Since $g(r, \theta)$ is periodic with respect to θ with period 2π , it can be expanded into a Fourier series as

$$g(r, \theta) = \sum_{n=-\infty}^{\infty} a_n e^{j2\pi(n/T)\theta} \quad (1.53)$$

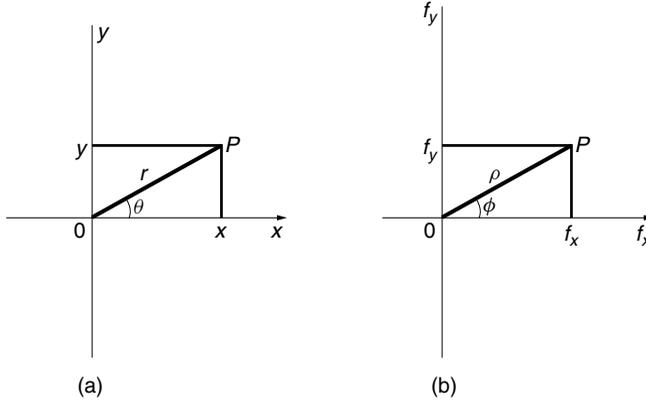


Figure 1.4 Change of coordinates from rectangular to cylindrical. (a) Space domain. (b) Spatial frequency domain.

where the coefficient a_n is

$$a_n = \frac{1}{T} \int_{-T/2}^{T/2} g(r, \theta) e^{-j2\pi(n/T)\theta} d\theta \quad (1.54)$$

Substituting 2π for the period T gives

$$g(r, \theta) = \sum_{n=-\infty}^{\infty} g_n(r) e^{jn\theta} \quad (1.55)$$

where

$$g_n(r) = \frac{1}{2\pi} \int_{-\pi}^{\pi} g(r, \theta) e^{-jn\theta} d\theta \quad (1.56)$$

Inserting Eq. (1.55) into (1.52) gives

$$G(\rho, \phi) = \sum_{n=-\infty}^{\infty} \int_0^{\infty} r dr \int_0^{2\pi} g_n(r) e^{jn\theta - j2\pi\rho r \cos(\theta - \phi)} d\theta \quad (1.57)$$

The integral with respect to θ can be expressed in terms of the Bessel function of the first kind of n th order [10] as

$$J_n(z) = \frac{1}{2\pi} \int_{\alpha}^{2\pi + \alpha} e^{j(n\beta - z \sin \beta)} d\beta \quad (1.58)$$

Noting that

$$\cos(\theta - \phi) = \sin(\theta - \phi + \pi/2) \quad (1.59)$$

and letting

$$\beta = \theta - \phi + \pi/2 \quad (1.60)$$

and inserting Eqs. (1.58), (1.59), and (1.60) into (1.52) finally gives

$$G(\rho, \phi) = \sum_{n=-\infty}^{\infty} (-j)^n e^{jn\phi} 2\pi \int_0^{\infty} r g_n(r) J_n(2\pi\rho r) dr \quad (1.61)$$

where $g_n(r)$ is given by Eq. (1.56). Conversely, the inverse Fourier transform is given by

$$g(r, \theta) = \sum_{n=-\infty}^{\infty} (j)^n e^{jn\theta} 2\pi \int_0^{\infty} \rho G_n(\rho) J_n(2\pi\rho r) d\rho \quad (1.62)$$

where

$$G_n(\rho) = \frac{1}{2\pi} \int_{-\pi}^{\pi} G(\rho, \phi) e^{-jn\phi} d\phi \quad (1.63)$$

When $n \neq 0$, the Fourier transform in cylindrical coordinates is called the Fourier–Hankel transform of the n th order. When $n = 0$, it is called the Fourier–Bessel transform and is written as $B\{g(r)\} = G(\rho)$, and $B^{-1}\{G(\rho)\} = g(r)$.

For the special case where there is no θ dependence, such as a circular aperture, then

$$g(r, \theta) = g(r) \quad (1.64)$$

and Eq. (1.56) becomes

$$\begin{aligned} g_n(r) &= \frac{g(r)}{2\pi} \int_{-\pi}^{\pi} e^{-jn\theta} d\theta \\ &= \frac{g(r)}{2\pi} \left[\frac{e^{-jn\theta}}{-jn} \right]_{-\pi}^{\pi} \\ &= \begin{cases} g(r) & n = 0 \\ 0 & n \neq 0 \end{cases} \end{aligned} \quad (1.65)$$

Terms with nonzero n disappear from Eqs. (1.61) and (1.62), and these equations simplify to

$$G(\rho) = 2\pi \int_0^{\infty} r g(r) J_0(2\pi\rho r) dr \quad (1.66)$$

$$g(r) = 2\pi \int_0^{\infty} \rho G(\rho) J_0(2\pi\rho r) d\rho \quad (1.67)$$

which are, as mentioned above, the Fourier–Bessel transform $B\{g(r)\}$ and its inverse $B^{-1}\{G(\rho)\}$.

Example 1.4 Find the Fourier–Hankel transform of a circular aperture with a one-sixth section obstruction as shown in Fig. 1.5.

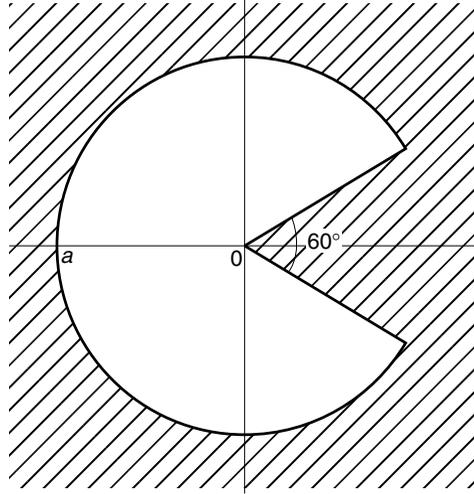


Figure 1.5 Circular aperture with obstruction.

Solution $G(\rho, \phi) = \text{whole circle} - \text{wedge portion}$. The wedge portion is expanded into a Fourier series as

$$g_n(r) = \frac{1}{2\pi} \int_{-\pi/6}^{\pi/6} g(r, \theta) e^{-jn\theta} d\theta \quad (1.68)$$

$$= \frac{1}{2\pi} \left[\frac{e^{-jn\theta}}{-jn} \right]_{-\pi/6}^{\pi/6}$$

$$g_n(r) = \begin{cases} \frac{1}{n\pi} \sin \frac{n\pi}{6} & n \neq 0 \\ \frac{1}{6} & n = 0 \end{cases} \quad (1.69)$$

$$G(\rho, \phi) = 2\pi \int_0^a r J_0(2\pi\rho r) dr$$

$$- 2 \sum_{\substack{n=-\infty \\ \text{except } n=0}}^{\infty} (-j)^n \frac{e^{jn\phi}}{n} \sin\left(\frac{n\pi}{6}\right) \int_0^a r J_n(2\pi\rho r) dr$$

$$- \frac{\pi}{3} \int_0^a r J_0(2\pi\rho r) dr \quad (1.70)$$

$$G(\rho, \phi) = \frac{5\pi a}{6\rho} J_1(2\pi\rho a) - 2 \sum_{\substack{n=-\infty \\ \text{except } n=0}}^{\infty} (-j)^n \frac{e^{jn\phi}}{n} \sin\left(\frac{n\pi}{6}\right) \int_0^a r J_n(2\pi\rho r) dr \quad (1.71)$$

□

Fourier transforms performed in either rectangular or cylindrical coordinates provide the same results, but one is usually more convenient than the other.

Now that the significance of the Fourier transform has been demonstrated, the next section describes representations of the source shape function and their Fourier transforms.

1.4 SPECIAL FUNCTIONS IN PHOTONICS AND THEIR FOURIER TRANSFORMS

1.4.1 Rectangle Function

An aperture function of unit width can be represented by the rectangle function. The rectangle function $\Pi(x)$ shown on the left in Fig. 1.6a is defined as

$$\Pi(x) = \begin{cases} 1 & |x| \leq \frac{1}{2} \\ 0 & |x| > \frac{1}{2} \end{cases} \quad (1.72)$$

The Fourier transform of the rectangle function is

$$\mathcal{F}\{\Pi(x)\} = \int_{-1/2}^{1/2} e^{-j2\pi fx} dx$$

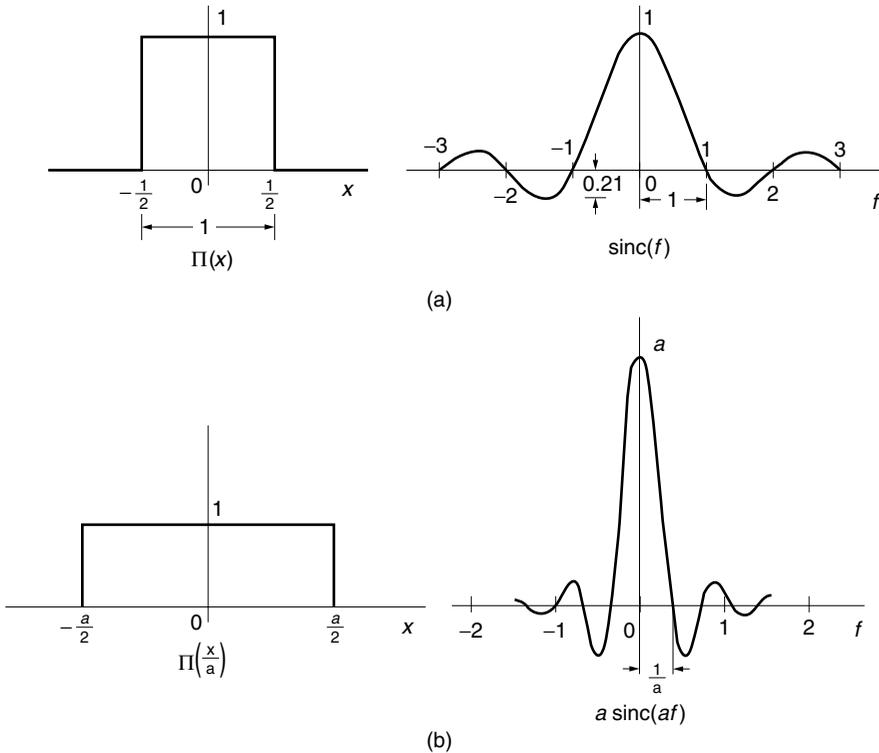


Figure 1.6 The rectangular function $\Pi(x)$ and $\Pi(x/a)$. (a) $\Pi(x)$ and its Fourier transform. (b) $\Pi(x/a)$ and its Fourier transform.

$$\begin{aligned}
&= \left[\frac{e^{-j2\pi fx}}{-j2\pi f} \right]_{-1/2}^{1/2} \\
&= \frac{\sin \pi f}{\pi f}
\end{aligned} \tag{1.73}$$

The right-hand side of Eq. (1.73) is the sinc function, defined as

$$\text{sinc}(f) = \frac{\sin \pi f}{\pi f} \tag{1.74}$$

As indicated in Fig. 1.6a, the main lobe of $\text{sinc}(f)$ has a width of 2, as measured along the f axis, and attains unit height at $f = 0$. Side lobes decrease in height as $|f|$ increases. The extrema of $\text{sinc}(f)$ occur when f is near an odd multiple of $\frac{1}{2}$, and the zeros of $\text{sinc}(f)$ are located at every integer. The height of the first side lobe is approximately 0.21 that of the main lobe. Now, let us extend the definition of Eq. (1.72) to an aperture with width a . The rectangle function for such an aperture is

$$\Pi\left(\frac{x}{a}\right) = \begin{cases} 1 & |x| \leq a/2 \\ 0 & |x| > a/2 \end{cases} \tag{1.75}$$

The Fourier transform is

$$\begin{aligned}
\mathcal{F}\left\{\Pi\left(\frac{x}{a}\right)\right\} &= \int_{-a/2}^{a/2} e^{-j2\pi fx} dx \\
&= a \text{sinc}(af)
\end{aligned} \tag{1.76}$$

Equation (1.76) certainly can be derived directly using the similarity theorem of the Fourier transform, Eq. (1.42). As indicated in Fig. 1.6b, $\text{sinc}(af)$ also has its main lobe at $f = 0$, but its height is a . The width of the main lobe is now $2/a$. Side lobes with decaying amplitudes appear with extrema near odd multiples of $1/2a$ and zeros at integral multiples of $1/a$. The ratio of the height of the first side lobe to the main lobe still remains at 0.21.

1.4.2 Triangle Function

The triangle function is defined as

$$\Lambda(x) = \begin{cases} 1 - |x| & |x| \leq 1 \\ 0 & |x| > 1 \end{cases} \tag{1.77}$$

and is shown on the left in Fig. 1.7. Unlike $\Pi(x)$, the width of the base is 2. The fact that the triangle function can be generated from the convolution of two rectangular functions makes the calculation of the Fourier transform simple:

$$\Lambda(x) = \Pi(x) * \Pi(x) \tag{1.78}$$

and

$$\mathcal{F}\{\Lambda(x)\} = \text{sinc}^2(f) \tag{1.79}$$

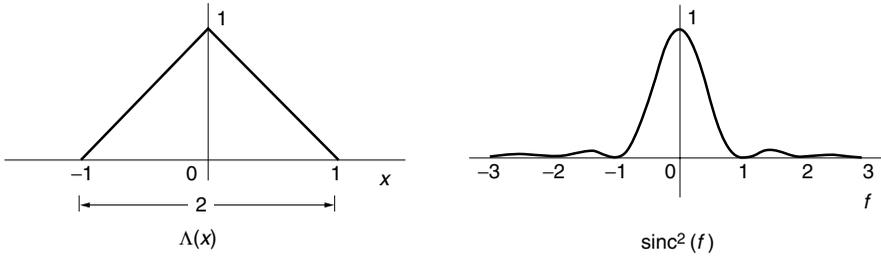


Figure 1.7 The triangle function $\Lambda(x)$ and its Fourier transform.

The triangle function with base width $2a$ can be expressed as the convolution of two rectangular functions as

$$\Lambda\left(\frac{x}{a}\right) = \frac{1}{a} \Pi\left(\frac{x}{a}\right) * \Pi\left(\frac{x}{a}\right) \quad (1.80)$$

Since the convolution of $\Pi(x/a)$ at $f = 0$ is a , the factor $1/a$ is necessary in Eq. (1.80) to make $\Lambda(x/a)$ unity at $f = 0$. The Fourier transform of Eq. (1.80) is

$$\mathcal{F}\left\{\Lambda\left(\frac{x}{a}\right)\right\} = a \operatorname{sinc}^2(af) \quad (1.81)$$

The graph of $\Lambda(x)$ with its Fourier transform is shown in Fig. 1.7. Compared to $\operatorname{sinc}(f)$, the side lobes of $\operatorname{sinc}^2(f)$ are significantly lower in height. An input aperture having a $\Lambda(x_0)$ distribution is used to reduce the side lobes in its diffraction pattern. This technique is called *apodizing* to reduce the side lobes. Apodization means a creature without legs, as, for example, eels or whales. Apodization of a lens is performed by darkening the lens toward the edge of the lens. Apodization of a radiation pattern from an antenna array is achieved by reducing the element antenna current toward the edge of the array [9].

The intensity pattern $I(x_i, y_i)$ of the diffracted field is expressed as

$$I(x_i, y_i) = E(x_i, y_i)E^*(x_i, y_i) \quad (1.82)$$

The field intensity patterns of the diffraction from a normal and an apodized slit are obtained by inserting Eqs. (1.76) and (1.79) into Eqs. (1.36) and (1.82) as

$$I_s(x_i, y_i) = \left(\frac{a}{\lambda z_i}\right)^2 \operatorname{sinc}^2\left(\frac{a}{\lambda z_i}x_i\right) \delta^2(y_i) \quad (1.83)$$

$$I_{as}(x_i, y_i) = \left(\frac{a}{\lambda z_i}\right)^2 \operatorname{sinc}^4\left(\frac{a}{\lambda z_i}x_i\right) \delta^2(y_i) \quad (1.84)$$

where δ is the delta function (see Section 1.4.5). The field intensity, which is EE^* , is not the same as the power intensity, which is $1/\eta|E|^2$ (see Section 2.3.1).

Examples of a normal and an apodized slit are shown in Fig. 1.8a, and the corresponding intensity patterns of the diffraction are compared in Fig. 1.8b. The reduction of the side lobe levels by the apodization is clearly demonstrated.

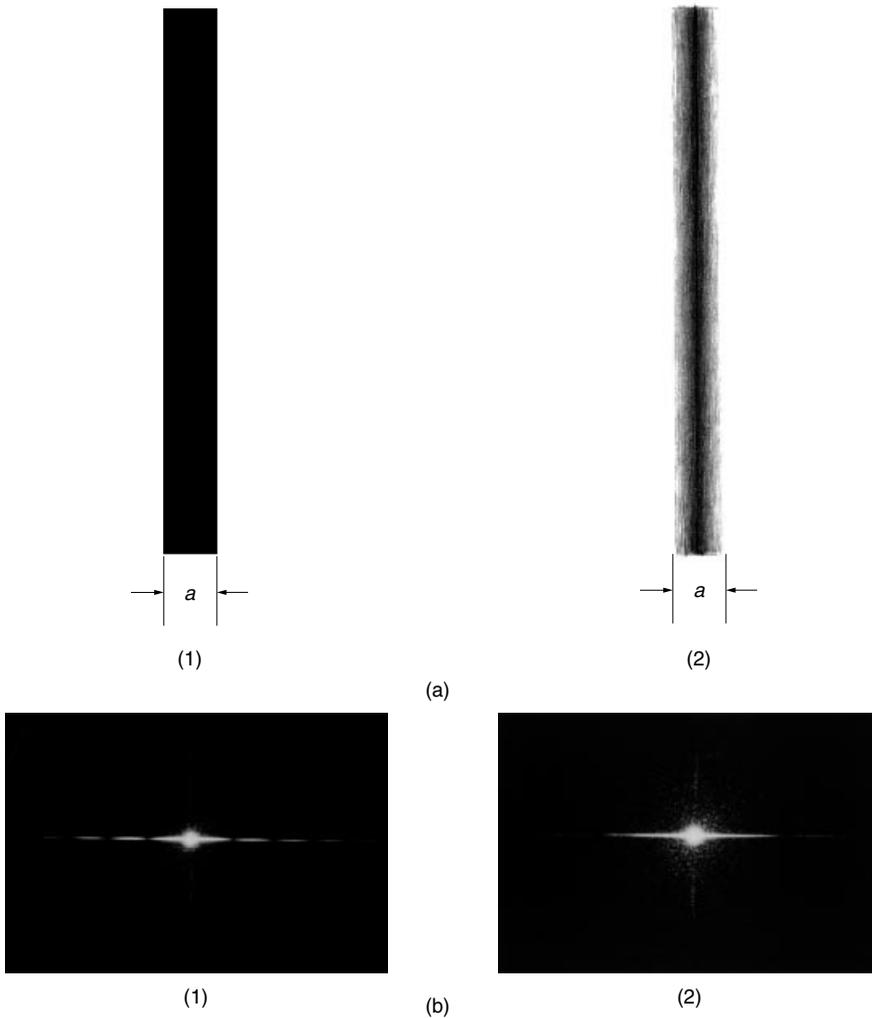


Figure 1.8. Comparison of (1) a normal slit and (2) an apodized slit. (a) Geometry of the apertures. The dark portions are the openings. (b) Diffraction patterns.

Example 1.5 Find the Fraunhofer diffraction pattern of a rectangular aperture with dimensions $a \times l$.

Solution The input function is given by

$$E(x_0, y_0) = \Pi\left(\frac{x_0}{a}\right) \Pi\left(\frac{y_0}{l}\right) \quad (1.85)$$

The field $E(x_i, y_i)$ in the $z = z_i$ plane is found from Eqs. (1.36) and (1.85):

$$E(x_i, y_i) = \frac{al}{j\lambda z_i} e^{jk[z_i + (x_i^2 + y_i^2)/2z_i]} \operatorname{sinc}\left(a \frac{x_i}{\lambda z_i}\right) \operatorname{sinc}\left(l \frac{y_i}{\lambda z_i}\right) \quad (1.86)$$

It should be noted that x_0 and y_0 are independent variables. Equation (1.86) is not the convolution of the two sinc functions but the product of the two.

The aperture and a photograph of its diffraction pattern are shown in Fig. 1.9. An important feature of the diffraction pattern is that the width of the main lobe is narrowed with widening of the aperture. To remember the concept, think of a water hose. The narrower the nozzle is pinched, the wider the water is sprayed. The smaller the structure of the source (or object), the wider the radiation pattern becomes. X-ray crystallography uses this fact very wisely to analyze molecular structure. The X-ray pattern scattered from an angstrom-sized structure is enlarged enough to be recorded by an ordinary photographic plate. \square

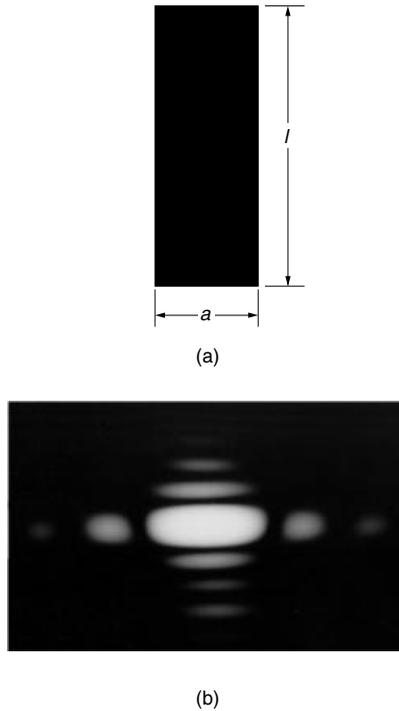


Figure 1.9 A rectangular aperture and its far-field diffraction pattern. (a) Geometry. The dark portion is the opening. (b) Diffraction pattern.



A firefighter knows that a narrower nozzle expands the beam. In the same way, a narrower source produces a larger diffraction pattern.

1.4.3 Sign and Step Functions

The sign function, $\text{sgn } x$, is positive unity for positive values of x and negative unity for negative values of x :

$$\text{sgn}(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases} \quad (1.87)$$

It is used to express a phase reversal at $x = 0$. The Fourier transform of Eq. (1.87) is

$$\mathcal{F}\{\text{sgn}(x)\} = \lim_{\alpha \rightarrow 0} \left[\int_{-\infty}^0 (-1)e^{-j2\pi f x + \alpha x} dx + \int_0^{\infty} e^{-j2\pi f x - \alpha x} dx \right] \quad (1.88)$$

The presence of α is necessary to perform the integral. After integration, α is reduced to zero. The result is

$$\mathcal{F}\{\text{sgn}(x)\} = \frac{1}{j\pi f} \quad (1.89)$$

The step function, $H(x)$, is immediately generated from the sign function:

$$H(x) = \frac{1}{2}[1 + \text{sgn}(x)] \quad (1.90)$$

The step function is used to mask one-half of a plane. Its Fourier transform is

$$\mathcal{F}\{H(x)\} = \frac{1}{2} \left(\delta(f) + \frac{1}{j\pi f} \right) \quad (1.91)$$

where δ represents the delta function and is explained in Section 1.4.5.

Figure 1.10 shows the geometry of the step function and its diffraction pattern. It is worth noting that even though the aperture does not have symmetry with respect to $x_0 = 0$, the intensity pattern of the diffraction has a symmetry with respect to the edge at $x_0 = 0$. The streak pattern is always perpendicular to the direction of the edge, and its intensity decreases monotonically with distance away from the edge.

1.4.4 Circle Function

In order to describe a circular aperture, the circle function shown in Fig. 1.11a is defined as

$$\text{circ}(r) = \begin{cases} 1 & r \leq 1 \\ 0 & r > 1 \end{cases} \quad (1.92)$$

The Fourier transform of the circle function is found from Eq. (1.66):

$$B\{\text{circ}(r)\} = 2\pi \int_0^1 r J_0(2\pi \rho r) dr \quad (1.93)$$

The Bessel function has the property

$$x^n J_{n-1}(x) = \frac{d}{dx} [x^n J_n(x)] \quad (1.94)$$

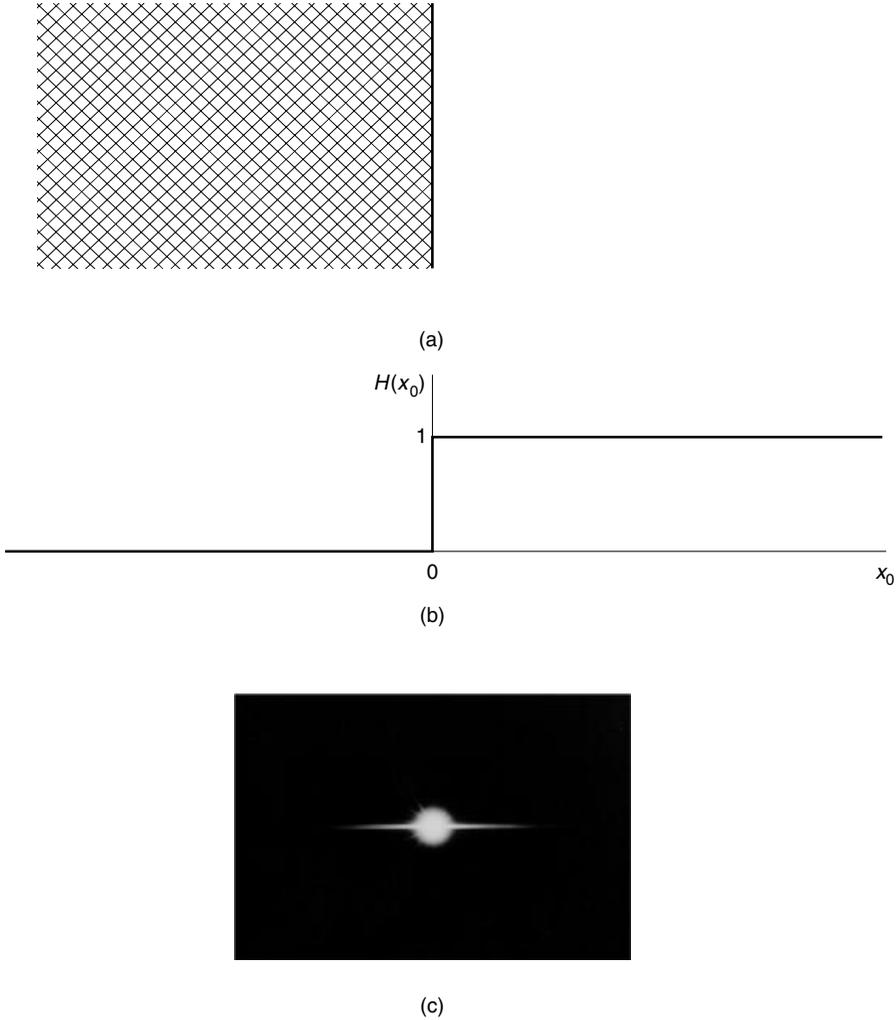


Figure 1.10 Diffraction from a semi-infinite screen. (a) Semi-infinite screen. (b) Step function. (c) Diffraction pattern.

Substituting $n = 1$ in Eq. (1.94) and integrating both sides of the equation gives

$$\int xJ_0(x) dx = xJ_1(x) \tag{1.95}$$

Using $x = 2\pi\rho r$, Eq. (1.93) becomes

$$\begin{aligned} B\{\text{circ}(r)\} &= 2\pi \frac{1}{(2\pi\rho)^2} \int_0^{2\pi\rho} xJ_0(x) dx \\ &= \frac{2\pi}{(2\pi\rho)^2} [xJ_1(x)]_0^{2\pi\rho} \\ B\{\text{circ}(r)\} &= \frac{1}{\rho} J_1(2\pi\rho) \end{aligned} \tag{1.96}$$

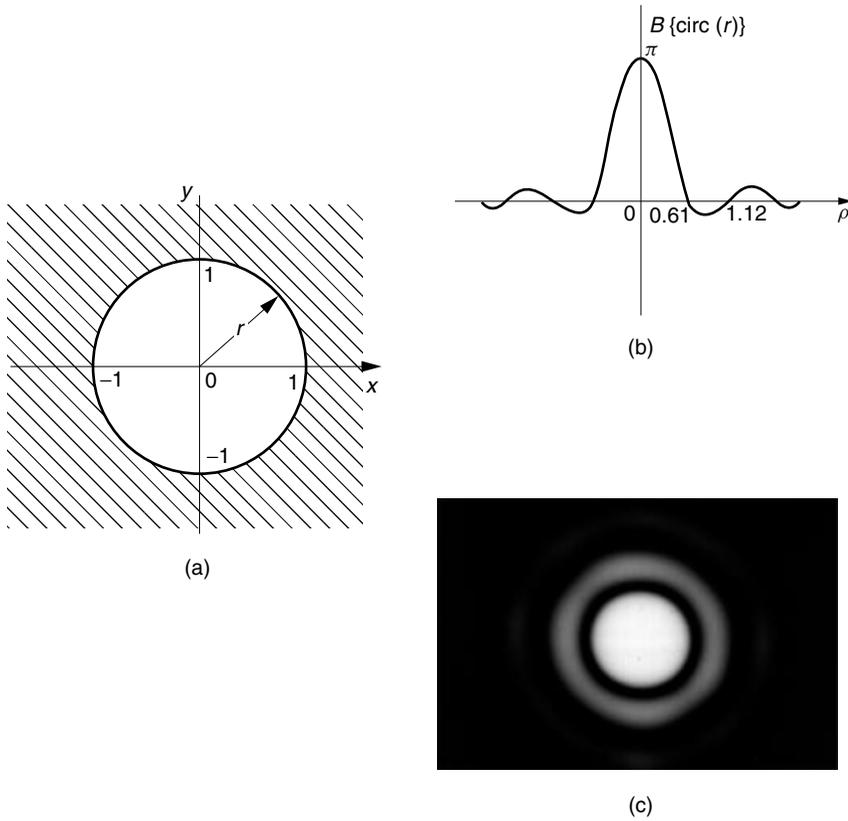


Figure 1.11 The circle function. (a) Geometry. (b) Fourier transform. (c) Diffraction pattern.

The graph of Eq. (1.96) is shown in Fig. 1.11b. The similarity theorem for the Fourier–Bessel transform does not follow Eq. (1.42); that is,

$$B\{g(ar)\} = \frac{1}{a^2} G\left(\frac{\rho}{a}\right) \quad (1.97)$$

and

$$B\left\{\text{circ}\left(\frac{r}{a}\right)\right\} = \frac{a}{\rho} J_1(2\pi\rho a) \quad (1.98)$$

The circle function and its Fourier transform are shown in Figs. 1.11a and 1.11b, respectively. The intensity pattern of the diffraction from the circular aperture with radius a is

$$I(r_i) = \left(\frac{a^2}{\lambda z_i}\right)^2 \left(\frac{J_1(2\pi a \rho)}{a \rho}\right)^2 \quad (1.99)$$

with

$$\rho = \frac{r_i}{\lambda z_i}$$

which can be rewritten as

$$I(r_i) = \left(\frac{ka^2}{z_i}\right)^2 \left(\frac{J_1(kar_i/z_i)}{kar_i/z_i}\right)^2 \tag{1.100}$$

where r_i is the radial coordinate in the plane of the diffraction pattern. This diffraction pattern was first derived by Sir George Biddell Airy and is referred to as the Airy pattern. The diffraction photograph is shown in Fig. 1.11c.

1.4.5 Delta Function

The delta function $\delta(x)$ is conveniently made to represent a point source [10]. Its amplitude is confined within a minute range of $x = \pm\varepsilon$, while it is zero outside this range, as shown in Fig. 1.12a. The amplitude grows to infinity as ε shrinks to zero, but in such a way that the area enclosed by the curve is always unity. In the limit $\varepsilon \rightarrow 0$, $\delta(0) \rightarrow \infty$, while satisfying

$$\int_{-\varepsilon}^{\varepsilon} \delta(x) dx = 1 \tag{1.101}$$

The delta function is most often used in an integral form:

$$\int_{-\infty}^{\infty} f(x)\delta(x - a) dx = f(a) \tag{1.102}$$

The integrand is shown in Fig. 1.12b. The region where the product is nonzero is only at $x = a \pm \varepsilon$. Consequently, $f(x)$ in this region can be approximated by the constant

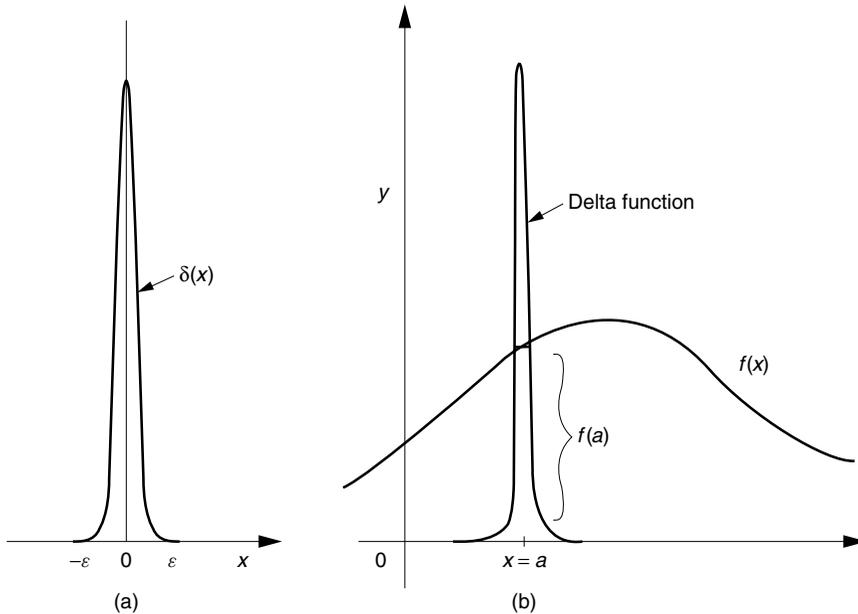


Figure 1.12 Diagram showing $\int f(x)\delta(x - a) dx = f(a)$. (a) Delta function. (b) Integral including the delta function.

value $f(a)$, and the constant $f(a)$ can be brought outside the integral. The integral of the delta function is unity and Eq. (1.102) holds. Equation (1.102) is used to sample (or sift) the value of $f(x)$ at $x = a$. This is called the sifting property of the delta function.

The Fourier transform of the delta function is

$$\mathcal{F}\{\delta(x)\} = \int_{-\infty}^{\infty} e^{-j2\pi fx} \delta(x) dx \quad (1.103)$$

Since the delta function samples the value of $e^{-j2\pi fx}$ at $x = 0$, Eq. (1.103) becomes

$$\mathcal{F}\{\delta(x)\} = 1 \quad (1.104)$$

Next, the inverse Fourier transform of $\delta(f)$ is considered. The inverse Fourier transform uses $e^{j2\pi fx}$ instead of $e^{-j2\pi fx}$ on the right-hand side of Eq. (1.103), and therefore

$$\mathcal{F}^{-1}\{\delta(f)\} = 1$$

Taking the Fourier transform of both sides gives

$$\mathcal{F}\{1\} = \int_{-\infty}^{\infty} e^{-j2\pi fx} dx = \delta(f) \quad (1.105)$$

Next, the value of $\delta(bx)$ will be expressed in terms of $\delta(x)$:

$$\int_{-\varepsilon}^{\varepsilon} f(x) \delta(bx) dx = \frac{1}{b} \int_{-b\varepsilon}^{b\varepsilon} f\left(\frac{y}{b}\right) \delta(y) dy \quad (1.106)$$

Thus, Eq. (1.106) becomes

$$\int_{-\varepsilon}^{\varepsilon} f(x) \delta(bx) dx = \frac{1}{b} f(0) \quad (1.107)$$

Equation (1.102) with $a = 0$ gives

$$\int_{-\varepsilon}^{\varepsilon} f(x) \delta(x) dx = f(0) \quad (1.108)$$

A comparison of Eqs. (1.107) and (1.108) leads to

$$\delta(bx) = \frac{\delta(x)}{|b|} \quad (1.109)$$

The absolute value is placed in the denominator of Eq. (1.109) because the result is the same for $-b$ and $+b$. Another property of the delta function is that its convolution with a function is the function itself; namely,

$$f(x) * \delta(x - a) = f(x - a) \quad (1.110)$$

because

$$\int_{-\infty}^{\infty} f(\eta)\delta(x - \eta - a) d\eta = f(x - a) \tag{1.111}$$

1.4.6 Shah Function (Impulse Train Function)

An array of equally spaced delta functions, such as shown in Fig. 1.13a, is called the shah function (or comb function), and it is denoted by $\Pi(x)$. Mathematically, the shah function is represented as

$$\Pi(x) = \sum_{n=-\infty}^{\infty} \delta(x - n) \tag{1.112}$$

Two major applications of $\Pi(x)$ are the following:

1. The generation of a sampled function. The sampled function $g_s(x)$ is an array of delta functions whose envelope is proportional to $g(x)$, such as shown in

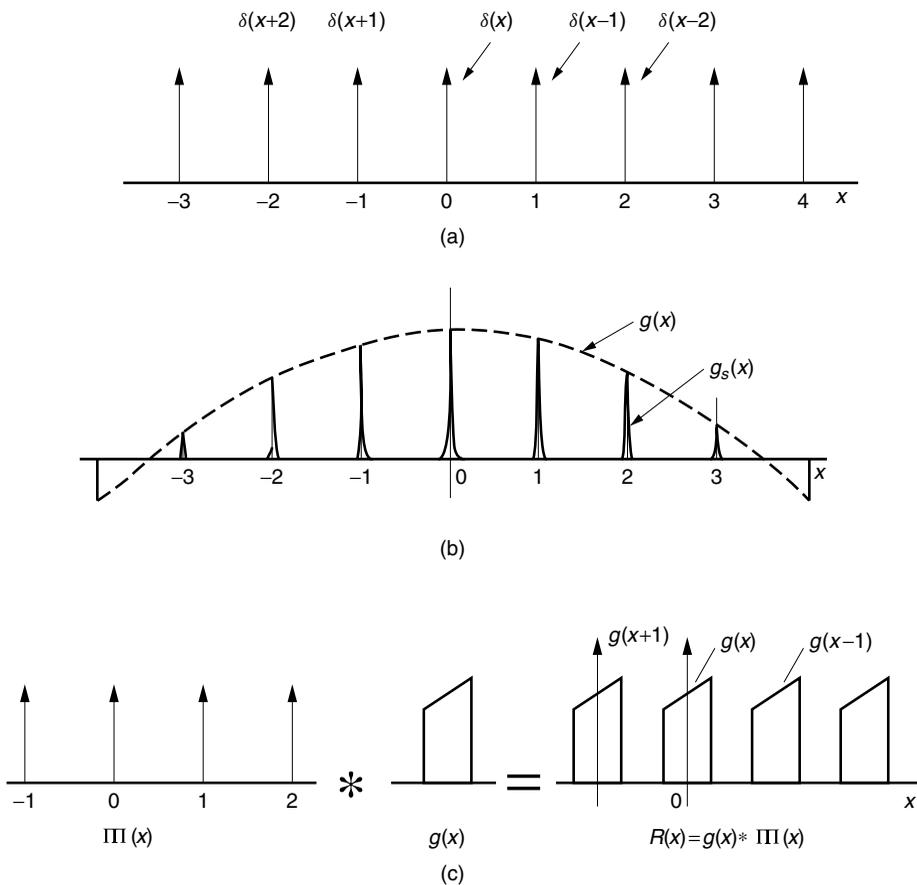


Figure 1.13 Shah function and its applications. (a) Shah function $\Pi(x)$. (b) $g(x)$ sampled by $\Pi(x)$. (c) Step and repeat function $R(x)$.

Fig. 1.13b. The sampled function can be generated by simply multiplying $g(x)$ by the shah function:

$$g_s(x) = g(x)\Pi(x) \quad (1.113)$$

A typical use for $g_s(x)$ is in optical signal processing, as, for example, processing a newspaper photograph that is made of closely sampled points.

2. The generation of a step and repeat function. The step and repeat function, such as that shown in Fig. 1.13c, can be generated by convolving $g(x)$ with the shah function:

$$R(x) = g(x) * \Pi(x) \quad (1.114)$$

From the definition of convolution, $R(x)$ is

$$\begin{aligned} R(x) &= \int_{-\infty}^{\infty} g(\tau)\Pi(x - \tau) d\tau \\ &= \int_{-\infty}^{\infty} g(\tau) \sum_{n=-\infty}^{\infty} \delta(x - \tau - n) d\tau \end{aligned} \quad (1.115)$$

Using Eq. (1.102), Eq. (1.115) becomes

$$R(x) = \sum_{n=-\infty}^{\infty} g(x - n) \quad (1.116)$$

which steps and repeats $g(x)$ at a unit interval.

The shah function that steps and repeats at an interval other than the unit interval warrants special attention. The step and repeat function at an interval of a is expressed as

$$g(x) * \Pi\left(\frac{x}{a}\right) = \int_{-\infty}^{\infty} g(\tau) \sum_{-\infty}^{\infty} \delta\left(\frac{x - \tau - an}{a}\right) d\tau \quad (1.117)$$

From Eq. (1.109), Eq. (1.117) becomes

$$g(x) * \Pi\left(\frac{x}{a}\right) = a \sum_{n=-\infty}^{\infty} g(x - an) \quad (1.118)$$

Thus, the step and repeat function at an interval of a is

$$R(x) = \frac{1}{a}g(x) * \Pi\left(\frac{x}{a}\right) \quad (1.119)$$

The existence of the factor $1/a$ should be noted.

Next, the Fourier transform of the shah function will be derived. Because the shah function is a periodic function, it can be expanded into a Fourier series with a period

of unity as

$$\Pi(x) = \sum_{n=-\infty}^{\infty} a_n e^{j2\pi(n/T)x} \quad (1.120)$$

where

$$a_n = \frac{1}{T} \int_{-T/2}^{T/2} \delta(x) e^{-j2\pi(n/T)x} dx = 1$$

and where T is the period of the delta functions and is unity. Thus, Eq. (1.120) becomes

$$\Pi(x) = \sum_{n=-\infty}^{\infty} e^{j2\pi nx} \quad (1.121)$$

Whenever x is an integer, $\Pi(x)$ becomes infinite, so that Eq. (1.121) constitutes an array of delta functions spaced by unity. Equation (1.121) is an alternate expression for the shah function. Thus, using Eqs. (1.35), (1.105), and (1.121), the Fourier transform of the shah function becomes

$$\mathcal{F}\{\Pi(x)\} = \sum_{n=-\infty}^{\infty} \delta(f - n) \equiv \Pi(f) \quad (1.122a)$$

Similarly,

$$\mathcal{F}\{\Pi(x/a)\} = a\Pi(af) \quad (1.122b)$$

$\Pi(x)$ is a very special function in that the Fourier transform is the same as the function itself.

1.4.7 Diffraction from an Infinite Array of Similar Apertures with Regular Spacing

Making use of the shah function, the diffraction pattern will be calculated for a one-dimensional array of slits such as shown in Fig. 1.14a(1). The slits are identical rectangle functions with width a , and the slits are equally spaced with period b , thereby forming a step and repeat function. The transmittance of this step and repeat function is expressed using Eqs. (1.75) and (1.119) as

$$E(x_0, y_0) = \frac{1}{b} \Pi\left(\frac{x_0}{a}\right) * \Pi\left(\frac{x_0}{b}\right) \quad (1.123)$$

The diffraction pattern is given by Eq. (1.34) as

$$E(x_i, y_i) = a \operatorname{sinc}\left(\frac{a}{\lambda z_i} x_i\right) \cdot \Pi\left(\frac{b}{\lambda z_i} x_i\right) \delta\left(\frac{y_i}{\lambda z_i}\right) \quad (1.124)$$

where the quadratic phase factor will be suppressed in this section.

A photograph of the diffraction pattern is shown in Fig. 1.14b(1). The pattern consists of an array of bright spikes with spacing $\lambda z_i/b$. For this particular array, the ratio between the slit width a and the period b is $b/a = 5$. This means that the fifth spike overlaps with the first null of $\operatorname{sinc}(af_x)$ and the intensity of the fifth spike is faint.

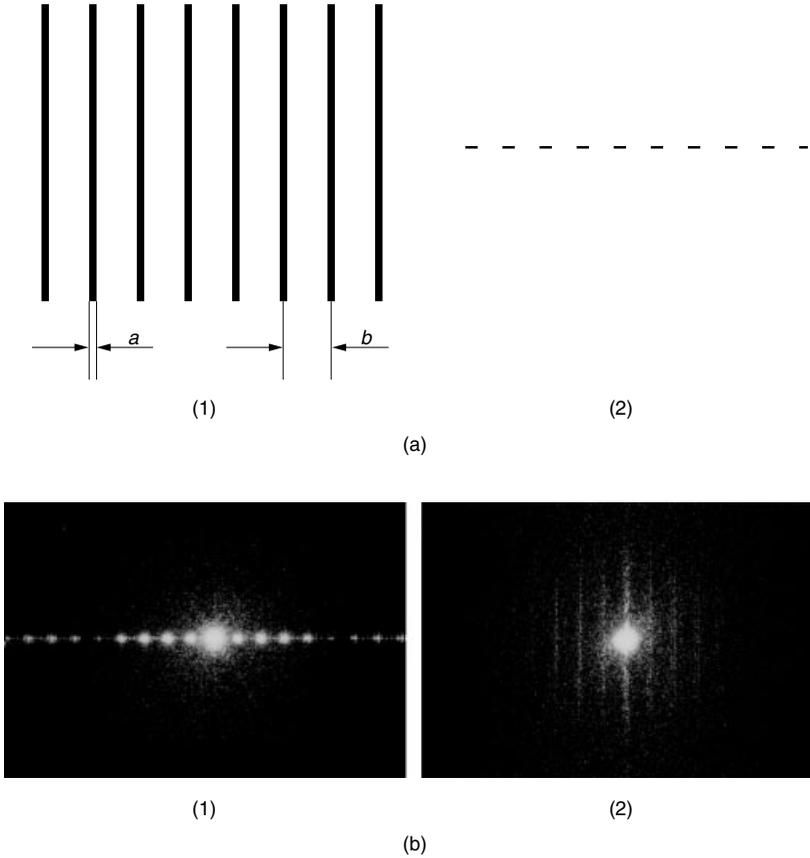


Figure 1.14 Comparison of (1) an array of long slits and (2) an array of short slits. (a) Geometry of the apertures. The dark portions are the openings. (b) Diffraction patterns.

Figure 1.14b(2) was included in order to demonstrate how the dimension in f_y of the diffraction pattern stretches as the height of the slits y_0 is shortened, while keeping other parameters unchanged.

Next, a two-dimensional array will be formed out of rectangular apertures with dimensions $a \times l$, with period b in the x direction and m in the y direction, as shown in Fig. 1.15a(2).

The aperture is represented by

$$E(x_0, y_0) = \frac{1}{bm} \left[\Pi \left(\frac{x_0}{a} \right) * \Pi \left(\frac{x_0}{b} \right) \right] \left[\Pi \left(\frac{y_0}{l} \right) * \Pi \left(\frac{y_0}{m} \right) \right] \quad (1.125)$$

The diffraction pattern is given by

$$E(x_i, y_i) = \frac{al}{j\lambda z_i} \operatorname{sinc} \left(\frac{a}{\lambda z_i} x_i \right) \operatorname{sinc} \left(\frac{l}{\lambda z_i} y_i \right) \Pi \left(\frac{b}{\lambda z_i} x_i \right) \Pi \left(\frac{m}{\lambda z_i} y_i \right) \quad (1.126)$$

The photograph of the diffraction pattern in Fig. 1.15b(2) shows a grid of bright spikes. The spacing of the spikes is $\lambda z_i/b$ in the x_i direction and $\lambda z_i/m$ in the y_i direction. Note that the brightness of the spikes is not uniform and the overall distribution of the bright spikes is similar to that of the single rectangular aperture shown in Fig. 1.15b(1).

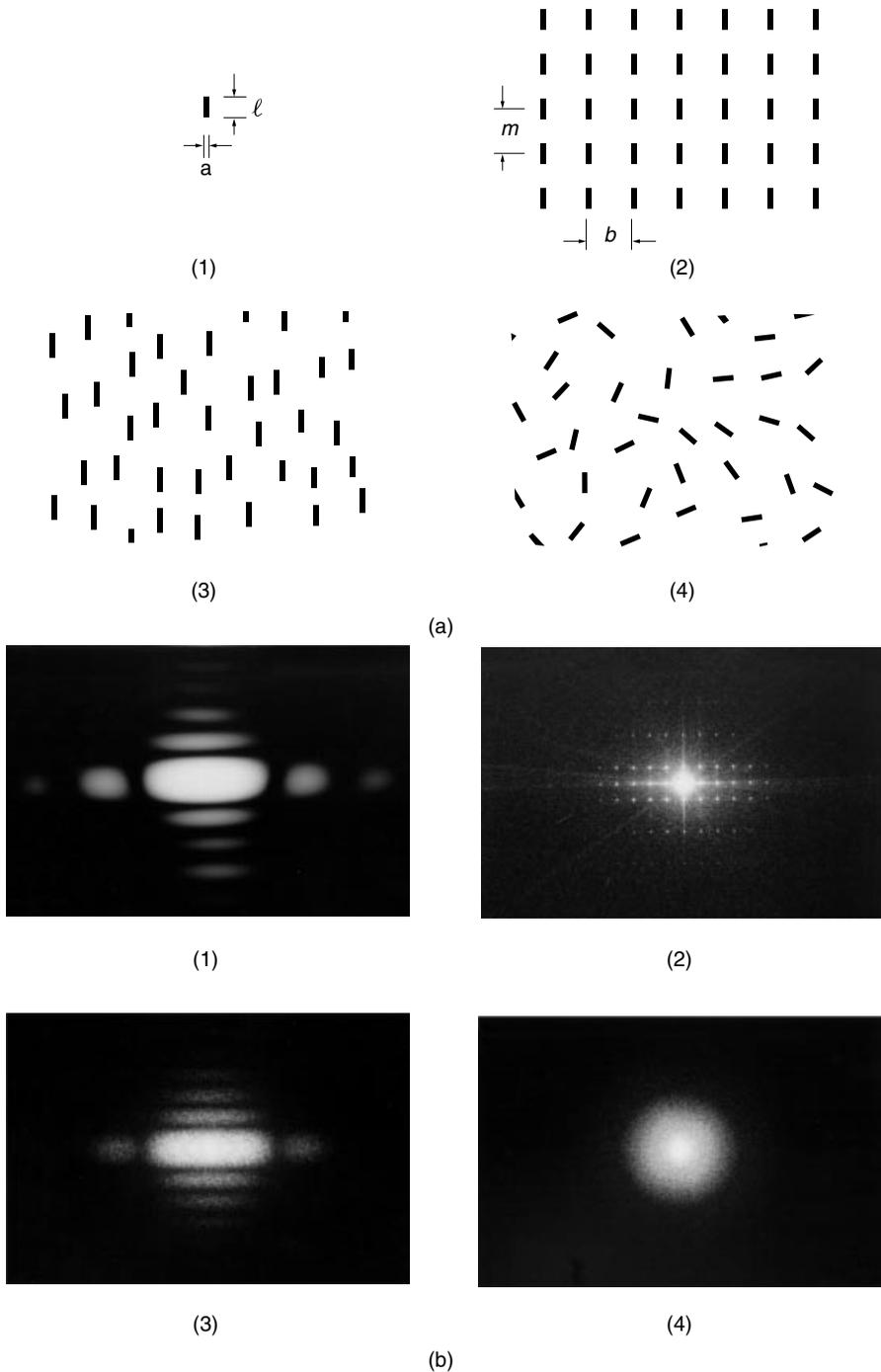


Figure 1.15 Comparison of rectangular apertures in various configurations: (1) single rectangular aperture, (2) rectangular apertures in a grid pattern, (3) rectangular apertures with random position in the vertical orientation, and (4) randomly arranged rectangular apertures. (a) Geometry. (b) Diffraction patterns.

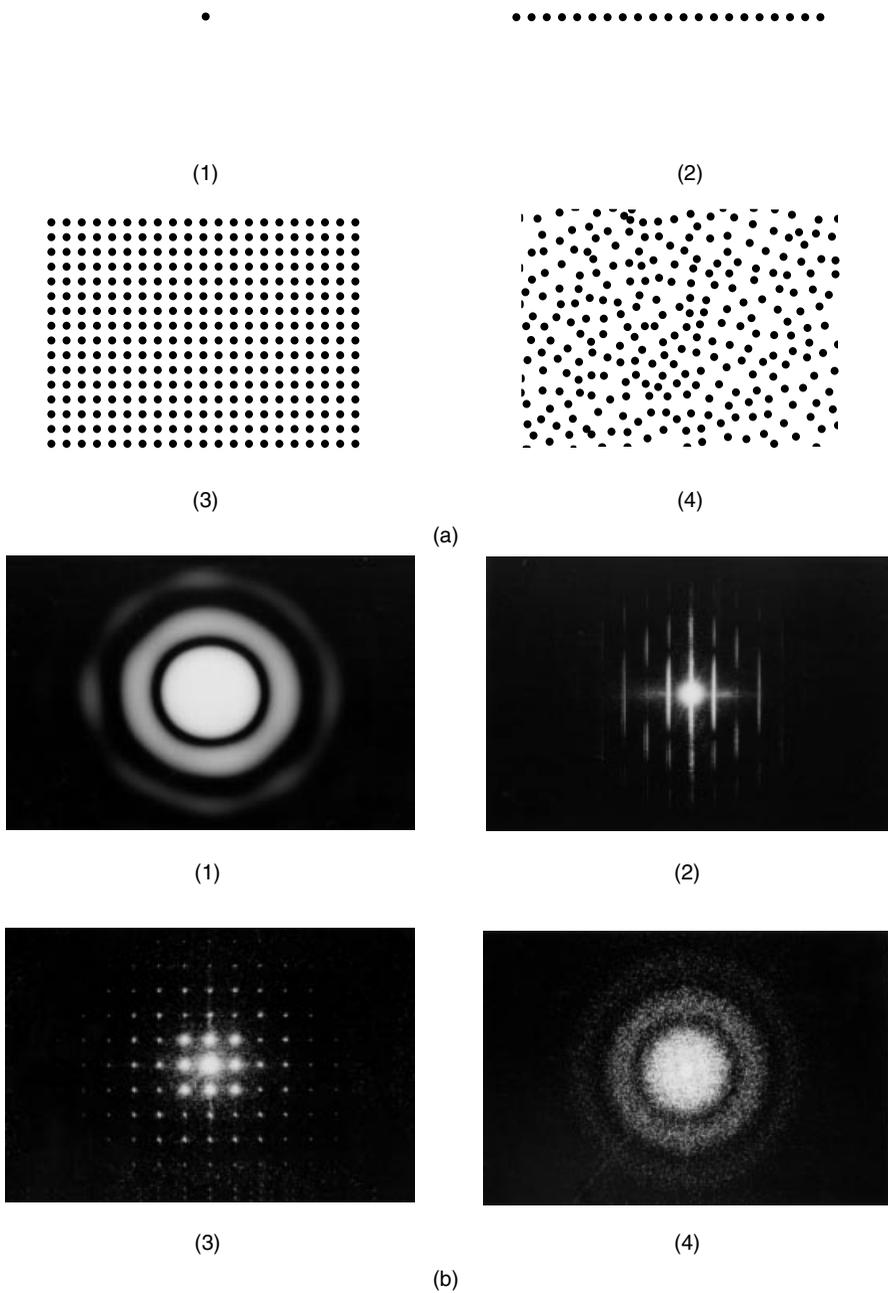


Figure 1.16 Circular apertures in various configurations: (1) single circular aperture, (2) circular apertures in a linear array, (3) circular apertures in a grid pattern, and (4) randomly arranged circular apertures. (a) Geometry. (b) Diffraction patterns.

1.4.8 Diffraction from an Infinite Array of Similar Apertures with Irregular Spacing

In the configuration shown in Fig. 1.15a(3), the elements of the array are randomly arranged, but without any rotation of the individual elements. For simplicity, consider a one-dimensional array with random spacing. If each rectangular aperture is translated in the x_0 direction by a random distance b_j from the origin, then the aperture distribution is represented by

$$E(x_0, y_0) = \sum_{j=0}^{N-1} \Pi\left(\frac{x_0 - b_j}{a}\right) \Pi\left(\frac{y_0}{l}\right) \quad (1.127)$$

Using the shift theorem of the Fourier transform

$$\mathcal{F}\{g(x - a)\} = e^{-j2\pi f a} G(f) \quad (1.128)$$

and

$$I(x_i, y_i) = E(x_i, y_i) \times E^*(x_i, y_i)$$

the intensity distribution of the diffraction pattern is

$$\begin{aligned} I(x_i, y_i) &= E(x_i, y_i)(1 + e^{-j2\pi b_1 f_x} + e^{-j2\pi b_2 f_x} + \dots) \\ &\quad \times E^*(x_i, y_i)(1 + e^{j2\pi b_1 f_x} + e^{j2\pi b_2 f_x} + \dots) \\ &= |E(x_i, y_i)|^2 \left(N + 2 \sum_{j=1}^{N-1} \sum_{k=1}^{N-1} \cos 2f_x(b_j - b_k) \right) \end{aligned} \quad (1.129)$$

where $|E(x_i, y_i)|^2$ is the diffraction pattern of a single rectangular aperture.

Since b_j and b_k are random, the second term is a superposition of cosine functions of a random period, which means that the resultant is zero and the diffraction pattern becomes $N|E(x_i, y_i)|^2$. The intensity distribution of the diffraction is the same as that of the single rectangular aperture but with N times the intensity, as shown in Fig. 1.15b(3).

Finally, the rectangular apertures are randomized with respect to rotation as well as translation. The diffraction photograph in Fig. 1.15b(4) looks like one that would have been obtained by rotating the single rectangular aperture diffraction pattern in Fig. 1.15b(1) about its center.

Figure 1.16a and 1.16b show what happens when the rectangular apertures in Fig. 1.15a are replaced by circular apertures.

The diffraction pattern of the single circular aperture in Fig. 1.16a(1) is shown in Fig. 1.16b(1). When this circular aperture is arranged in a one-dimensional array, as shown in Fig. 1.16a(2), the diffraction pattern is made up of an array of vertical lines (shah function) as shown in Fig. 1.16b(2). The brightness of the lines is not uniform, but the overall brightness distribution resembles the diffraction pattern of the single circular aperture shown in Fig. 1.16a(1).

Next, the circular apertures are arranged in grid form with period b in both the x_0 and y_0 directions as indicated in Fig. 1.16a(3). A photograph of the diffraction pattern is shown in Fig. 1.16b(3). The grid pattern of the diffraction looks like the pattern that would be obtained by the product of the pattern of an array in the x_0 direction and the pattern of an array in the y_0 direction.

As in the case of the rectangular element apertures shown in Fig. 1.15b(2), the overall pattern of the brightness of the spikes in Fig. 1.16b(3) has a similar distribution to the diffraction pattern of a circular element aperture.

In Fig. 1.16a(4) the circular apertures are arranged in a random manner. As with the rectangular aperture, the photograph of the diffraction pattern in Fig. 1.16b(4) resembles that of a single circular aperture but with N times the intensity.

The speckle patterns are due to the finite number of element apertures and decrease with an increase in the number of elements.

1.4.9 Diffraction from a Finite Array

So far the dimensions of the array have been assumed to be infinite. In this section, the effect of the finiteness of the array will be explained. An example of a two-dimensional finite size rectangular aperture array is shown in Fig. 1.17a. The element apertures have a size of $a \times l$ and are spaced b and m apart in the x_0 and y_0 directions, respectively.

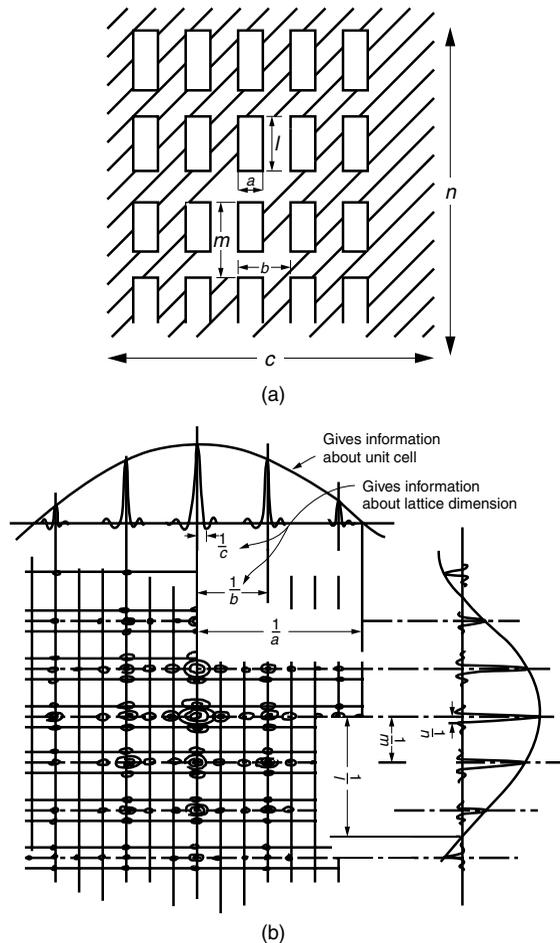


Figure 1.17 Diffraction from an array of finite size. (a) Two-dimensional array of windows with external dimensions $c \times n$. (b) Diffraction pattern of (a).

The extent of the array is limited to $c \times n$ in the x_0 and y_0 directions. The expression for the transmittance is

$$E(x_0, y_0) = \frac{1}{bm} \left[\Pi\left(\frac{x_0}{a}\right) * \text{III}\left(\frac{x_0}{b}\right) \right] \Pi\left(\frac{x_0}{c}\right) \\ \times \left[\Pi\left(\frac{y_0}{l}\right) * \text{III}\left(\frac{y_0}{m}\right) \right] \Pi\left(\frac{y_0}{n}\right) \quad (1.130)$$

The step and repeat functions in the square brackets are truncated by the rectangle functions $\Pi(x_0/c)$ and $\Pi(y_0/n)$.

The Fourier transform of the transmittance pattern is obtained from Eqs. (1.76) and (1.122b):

$$\mathcal{F}\{E_0(x_0, y_0)\} = acln [\text{sinc}(af_x)\text{III}(bf_x)] * \text{sinc}(cf_x) \\ \times [\text{sinc}(lf_y)\text{III}(mf_y)] * \text{sinc}(nf_y) \quad (1.131)$$

The order of the dimensions in the transmittance pattern is

$$a < b < c$$

where a is the width of the window, b is the spacing between windows, and c is the overall dimension. In the diffraction pattern in Fig. 1.17b, the order of the dimensions is inverted and reversed; namely,

$$\frac{1}{c} < \frac{1}{b} < \frac{1}{a} \quad (1.132)$$

where $1/c$ is the size of an individual spike, $1/b$ is the spacing between spots, and $1/a$ is the overall size of the diffraction pattern. Thus, the external size of the array controls the size of the individual spike in the diffraction pattern.

Figure 1.18 illustrates how the external shape of the array controls the shape of the individual spikes in the diffraction pattern. In Fig. 1.18a, the element apertures are circular, and the grids are bordered by four different boundaries — circular, rectangular, rectangular tilted at 45° , and triangular. The corresponding diffraction patterns are shown in Fig. 1.18b.

From Eq. (1.131), the pattern of an individual spike is proportional to $[\text{sinc}(cf_x)\text{sinc}(nf_y)]^2$. As the shape of the outer boundary is rotated, the shape of the individual spike rotates as shown in Figs. 1.18b(2) and 1.18b(3).

Finally, Fig. 1.18b(4) demonstrates that when the outer boundary is an equilateral triangle, the shape of each spike becomes the diffraction pattern of a single triangular aperture.

For large c , Eq. (1.131) can be approximated as

$$\mathcal{F}\{E_0(x_0, y_0)\} = acln \text{sinc}(af_x)[\text{III}(bf_x) * \text{sinc}(cf_x)] \\ \times \text{sinc}(lf_y)[\text{III}(mf_y) * \text{sinc}(nf_y)] \quad (1.133)$$

Now, Eq. (1.133) can be separated into two factors of different natures. The factors outside the square brackets are determined solely by the shape of each cell and are called the *element pattern*. The factors inside the square brackets are determined solely

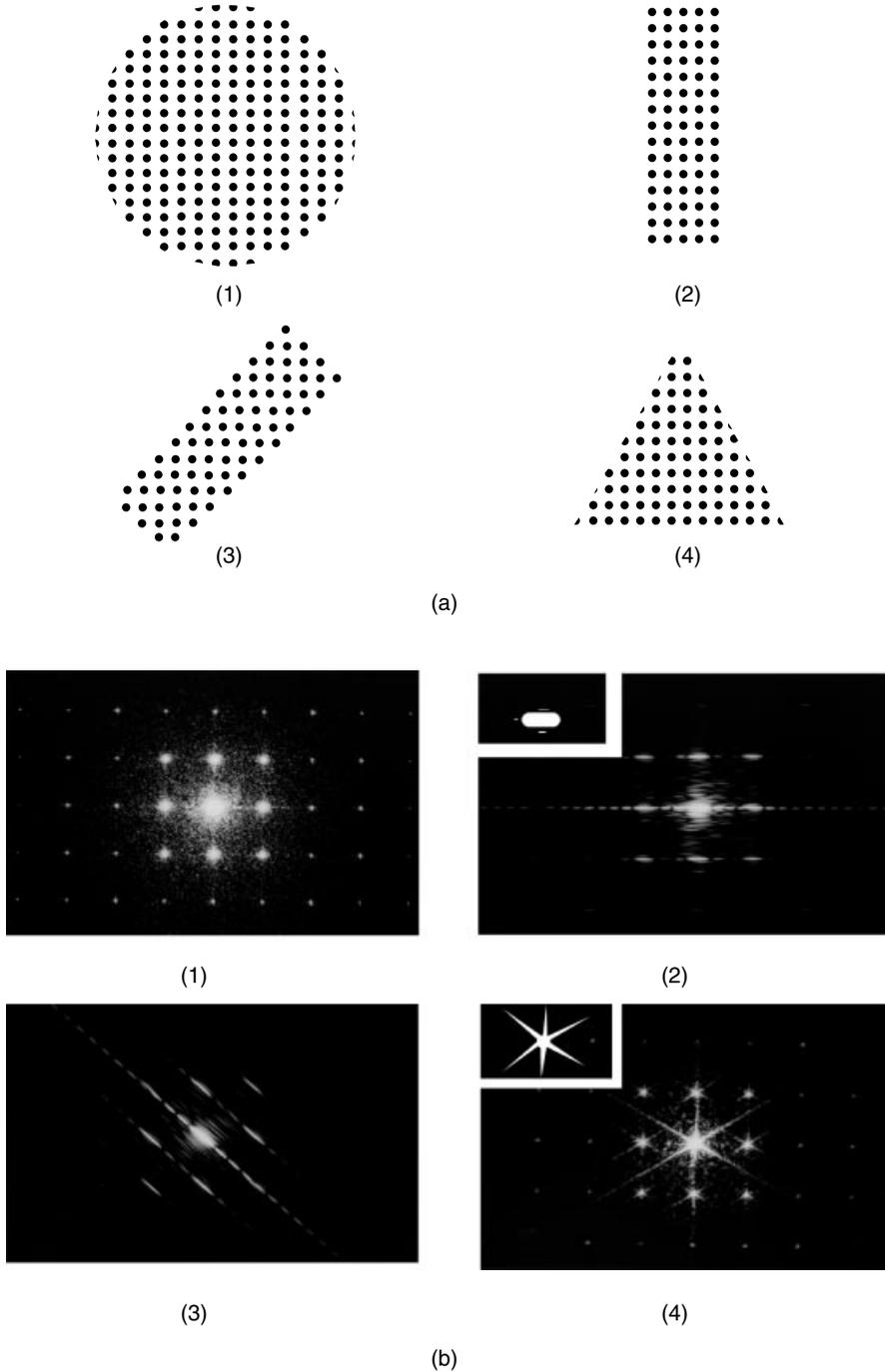


Figure 1.18 The effect of external border shapes on a grid of circular apertures: (1) circular border, (2) rectangular border, (3) tilted rectangular border, and (4) triangular border. (a) Geometry. (b) Diffraction patterns. Each inset is the diffraction pattern of a single aperture with the same shape as the border.

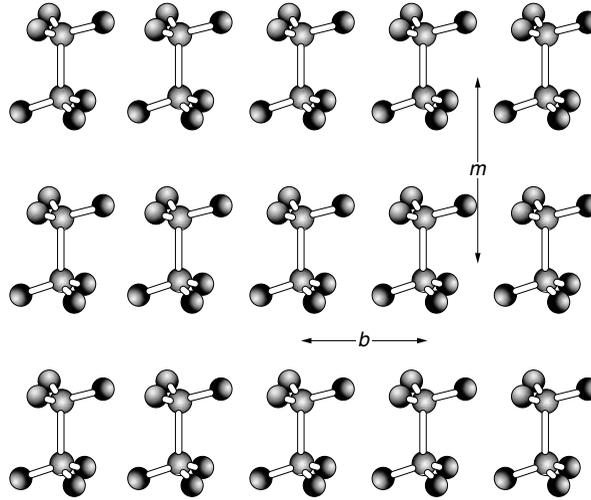


Figure 1.19 Two-dimensional crystal array.

by the spacing and overall dimensions of the lattice, are independent of the shape of each cell, and are called the *array pattern*. The diffraction pattern is the product of these two patterns. The array pattern, which is periodic, provides the lattice dimension. The element pattern, which is the envelope, provides the shape of the cell.

$$(\text{Pattern}) = (\text{Element pattern}) \times (\text{array pattern}) \quad (1.134)$$

The two-dimensional crystal pattern such as shown in Fig. 1.19 can be obtained by simply replacing $\text{sinc}(af_x) \text{sinc}(lf_y)$ by $G(f_x, f_y) = \mathcal{F}\{g(x_0, y_0)\}$. The amplitude envelope function $G(f_x, f_y)$ contains information about the structure of the unit cell. The structure of the unit cell $g(x_0, y_0)$ can be derived by taking the inverse Fourier transform of the envelope function. This is precisely the principle of X-ray crystallography.

In X-ray crystallography the darkness of each spot of the diffraction pattern in an X-ray photograph is measured by a microdensitometer to obtain the envelope of the intensity distribution $|G(f_x, f_y)|^2$ of the diffraction pattern. This method, however, measures the intensity pattern, and not the amplitude pattern of $G(f_x, f_y)$. The inverse Fourier transform of the intensity pattern gives

$$\mathcal{F}^{-1}|G(f_x, f_y)|^2 = g(x_0, y_0) * g(-x_0, -y_0)$$

which is called the Harker pattern. From the Harker pattern, $g(x_0, y_0)$ is resolved using additional information derived from the physical chemistry of the molecule.

1.5 THE CONVEX LENS AND ITS FUNCTIONS

Geometrical optics is most often applied to find the location and size of the image formed by a lens. Geometrical optics, however, fails to provide information about the wavelength and polarization dependences of the field distribution, and the image resolution for a given lens size. On the other hand, Fourier optics describes the wave nature of optics, corrects these failures, and hence will be explained here.

1.5.1 Phase Distribution After a Plano-Convex Lens

Figure 1.20 shows a thin plano-convex lens made by slicing a small section of a sphere by the plane $A-A'$. The lens medium is glass and the surrounding medium is air. The phase distribution across the tangent plane $B-B'$ will be calculated. Plane $B-B'$ is parallel to plane $A-A'$. The incident light is a plane wave whose propagation direction is normal to the plane $A-A'$. The incident wave is represented as an array of parallel rays, as shown on the left side of Fig. 1.20. The ray passing through the fat lens center suffers the longest phase delay, while that passing through the thin rim undergoes the shortest phase delay. The exact distribution $\phi(x, a)$ will be calculated in the coordinates whose origin coincides with the center of the sphere. A light ray at an arbitrary height x goes through both glass and air to reach plane $B-B'$ from plane $A-A'$. The total phase delay $\phi(x, a)$ is

$$\phi(x, a) = k[n(\sqrt{a^2 - x^2} - b) + a - \sqrt{a^2 - x^2}] \tag{1.135}$$

where the y component is suppressed. With $(x/a)^2 \ll 1$ and using the binomial expansion, $\phi(x, a)$ is approximated as

$$\phi(x, a) = \phi_0 - k \frac{x^2}{2f_0}$$

where

$$\phi_0 = kn(a - b) \tag{1.136}$$

and

$$f_0 = \frac{a}{n - 1} \tag{1.137}$$

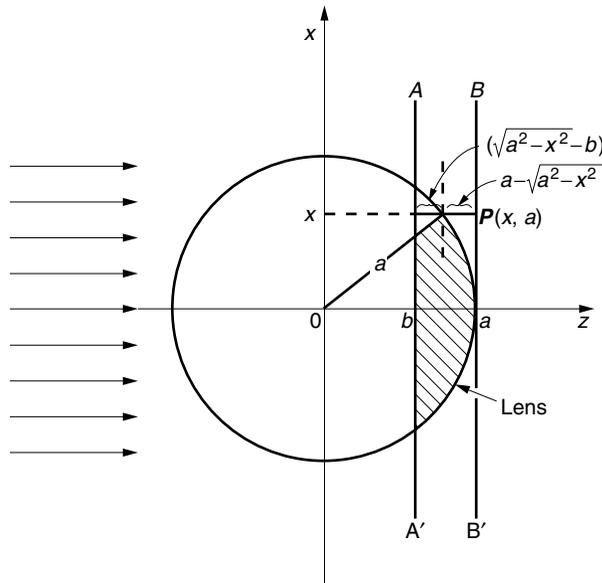


Figure 1.20 Geometry of a thin lens.

Thus, the emergent field from the plane $B-B'$ is

$$E(x) = Ae^{-jkx^2/2f_0} \tag{1.138}$$

The phase ϕ_0 is a constant that causes the same phase delay everywhere. It has no physical significance and is suppressed. In the two-dimensional case of a spherical lens, the emergent field $E(x, y)$ is

$$E(x, y) = Ae^{-jk(x^2+y^2)/2f_0} \tag{1.139}$$

where f_0 is the focal length of the lens, as will be seen shortly. The lens generates a quadratic phase distribution with a negative sign. This negative sign plays an important role in the convex lens.

1.5.2 Collimating Property of a Convex Lens

The collimating property of a lens is one of the simplest Fourier optics examples and is a good starting point for more detailed lens analysis. A delta function source is placed at F in front of a convex lens at a distance f_0 as shown in Fig. 1.21. The field incident on the lens is

$$E(x, y) = \frac{1}{j\lambda f_0} e^{jk[f_0+(x^2+y^2)/2f_0]} \tag{1.140}$$

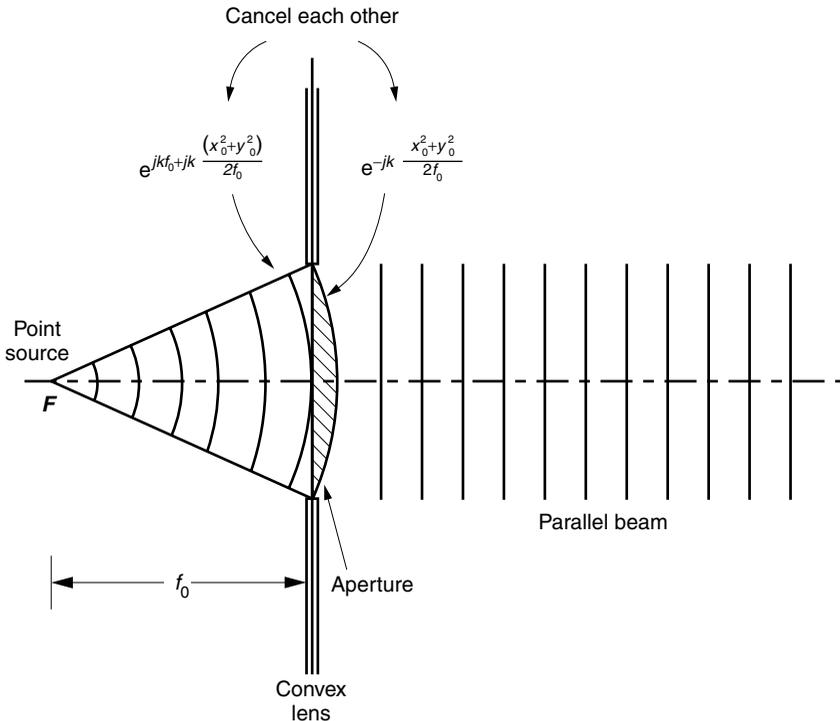


Figure 1.21 Interpretation of the generation of a parallel beam by Fourier optics.

where the coordinates (x, y) are in the plane of the lens. When the field passes through the lens it experiences a phase delay, which has previously been shown in Eq. (1.139) to be $\Delta\phi = -k[(x^2 + y^2)/2f_0]$. Hence, the resultant Fresnel diffraction of this field at a distant $z = z_i$ plane is obtained from the Fresnel–Kirchhoff diffraction formula (Eq. (1.38))

$$E(x_i, y_i) = \frac{1}{j\lambda z_i} e^{jk[z_i + (x_i^2 + y_i^2)/2z_i]} \times \mathcal{F} \left\{ \underbrace{\frac{1}{j\lambda f_0} e^{jk[f_0 + (x^2 + y^2)/2f_0]}}_{\text{Input}} \underbrace{e^{-jk(x^2 + y^2)/2f_0}}_{\text{Lens}} \underbrace{e^{jk(x^2 + y^2)/2z_i}}_{\text{Part of the point spread function}} \right\}_{f_x = x_i/\lambda z_i, f_y = y_i/\lambda z_i} \quad (1.141)$$

The divergence of the input field factor is partially cancelled by the lens factor. The rearrangement of Eq. (1.141) results in

$$E(x_i, y_i) = \frac{1}{j\lambda f_0} e^{jkf_0 + jk[(x_i^2 + y_i^2)/2z_i]} \mathcal{F} \left\{ \frac{1}{j\lambda z_i} e^{jk[z_i + (x^2 + y^2)/2z_i]} \right\}_{f_x = x_i/\lambda z_i, f_y = y_i/\lambda z_i} \quad (1.142)$$

The quantity in the curly brackets is the point spread function and its Fourier transform is obtained from Eq. (1.41):

$$E(x_i, y_i) = \frac{1}{j\lambda f_0} e^{jk(f_0 + z_i) + jk[(x_i^2 + y_i^2)/2z_i]} [e^{-j\pi\lambda z_i(f_x^2 + f_y^2)}]_{f_x = x_i/\lambda z_i, f_y = y_i/\lambda z_i} \quad (1.143)$$

Due to cancellation we finally obtain a function independent of x_i and y_i :

$$E(x_i, y_i) = \frac{1}{j\lambda f_0} e^{jk(f_0 + z_i)} \quad (1.144)$$

Hence, the field is a parallel beam and is a plane wave with constant amplitude $1/\lambda f_0$. The amplitude stays constant with distance z_i as shown in Fig. 1.21. This proves that f_0 is indeed the focal length of the lens.

1.5.3 Imaging Property of a Convex Lens

The Gaussian lens formula is derived by geometrical optics as

$$\frac{1}{d_1} + \frac{1}{d_2} = \frac{1}{f_0} \quad (1.145)$$

where f_0 is the focal length of the convex lens, d_1 is the distance from the lens to the object, and d_2 is the distance from the lens to the image. The Gaussian lens formula will be obtained using Fourier optics. The imaging condition is that the light emanating from a point on the object converges back to a point in the image plane.

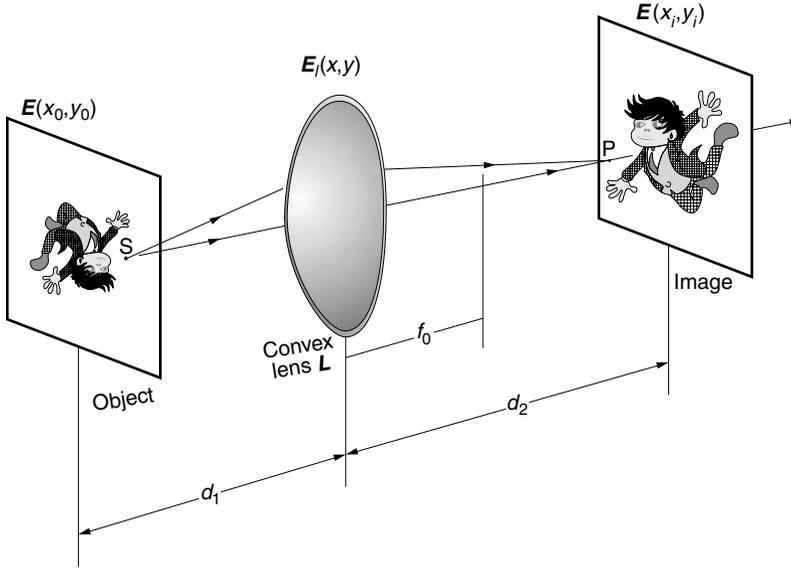


Figure 1.22 Imaging condition of a convex lens.

For simplicity, the object is a delta function source S located on the z axis, as shown in Fig. 1.22. The input field $E_l(x, y)$ to the lens in x, y coordinates is

$$E_l(x, y) = \frac{1}{j\lambda d_1} e^{jk[d_1+(x^2+y^2)/2d_1]} \quad (1.146)$$

The Fresnel diffraction from the lens to the image plane at $z_i = d_2$ is, from Eq. (1.38),

$$E(x_i, y_i) = -\frac{e^{jk[d_2+(x_i^2+y_i^2)/2d_2]}}{\lambda^2 d_1 d_2} \mathcal{F} \left\{ \underbrace{e^{jk[d_1+(x^2+y^2)/2d_1]}}_{\text{Input}} \underbrace{e^{-jk(x^2+y^2)/2f_0}}_{\text{Lens}} \underbrace{e^{jk(x^2+y^2)/2d_2}}_{\text{Part of the point spread function}} \right\}_{f_x=x_i/\lambda d_2, f_y=y_i/\lambda d_2} \quad (1.147)$$

If the observation is made at the particular value of d_2 that satisfies the Gaussian lens formula, Eq. (1.145), then the value inside the curly brackets becomes unity and

$$E(x_i, y_i) = -\frac{1}{\lambda^2 d_1 d_2} e^{jk[d_1+d_2+(x_i^2+y_i^2)/2d_2]} \delta\left(\frac{x_i}{\lambda d_2}\right) \delta\left(\frac{y_i}{\lambda d_2}\right) \quad (1.148)$$

The delta function image P is recovered if the imaging condition is satisfied. Thus, Fourier optics has proved Eq. (1.145).

$E(x_i, y_i)$ will be rewritten in a form more conducive to extracting the physical meaning. Using Eq. (1.109), the last two factors of Eq. (1.148) can be rewritten as

$$\delta\left(\frac{x_i}{\lambda d_2}\right) \delta\left(\frac{y_i}{\lambda d_2}\right) = (\lambda d_1)^2 \delta\left(\frac{d_1}{d_2} x_i\right) \delta\left(\frac{d_1}{d_2} y_i\right) \quad (1.149)$$

Let m be the magnification factor:

$$m = \frac{d_2}{d_1} \quad (1.150)$$

The expression for the image finally becomes

$$E(x_i, y_i) = -\frac{1}{m} e^{jk[d_1+d_2+(x_i^2+y_i^2)/2d_2]} \delta\left(\frac{x_i}{m}\right) \delta\left(\frac{y_i}{m}\right) \quad (1.151)$$

Since the object is a delta function source $\delta(x_0)\delta(y_0)$, its image is also expected to be a delta function and is $\delta(x_i/m)\delta(y_i/m)$. The width of $\delta(x_i/m)$ is wider than $\delta(x_0)$ by m times, and hence the image is magnified by m times. As far as the light intensity is concerned, the amplitude of the light is diluted by $1/m$ times.

The advantage of the Fourier optics approach is that it can be applied to situations where geometric optics is inadequate, such as dealing with the finite size of the lens. For example, the case of the finite-sized square aperture of dimensions $a \times a$ is represented as

$$A(x, y) = \Pi\left(\frac{x}{a}\right) \Pi\left(\frac{y}{a}\right) \quad (1.152)$$

This has the effect of altering the input by this factor. With this insertion into Eq. (1.146), the finite lens equivalent of Eq. (1.148) is

$$\begin{aligned} E(x_i, y_i) = & -\frac{1}{\lambda^2 d_1 d_2} e^{jk[d_1+d_2+(x_i^2+y_i^2)/2d_2]} \\ & \times \underbrace{[a^2 \operatorname{sinc}(af_x) \operatorname{sinc}(af_y)]}_{\text{Due to a finite square aperture}} * \underbrace{\delta(f_x)\delta(f_y)}_{\text{Image}} \Big|_{f_x=x_i/\lambda d_2, f_y=y_i/\lambda d_2} \end{aligned} \quad (1.153)$$

The final result is

$$E(x_i, y_i) = \frac{1}{\lambda^2 d_1 d_2} e^{jk[d_1+d_2+(x_i^2+y_i^2)/2d_2]} a^2 \operatorname{sinc}\left(a\frac{x_i}{\lambda d_2}\right) \operatorname{sinc}\left(a\frac{y_i}{\lambda d_2}\right) \quad (1.154)$$

Thus, the image of the delta function object is no longer a delta function but a sinc function whose main lobe size is $(2\lambda d_2/a) \times (2\lambda d_2/a)$. The larger the aperture $a \times a$ is, the smaller the width of the lobe becomes. However, as long as the dimension of the aperture is finite, the lobe is also finite. Hence, the image cannot be the same as the original delta function, even when the lens is designed perfectly and is aberration free. If the only limitation on resolution is due to the finiteness of the aperture causing the diffraction, then the imaging system is said to be *diffraction limited*.

1.5.4 Fourier Transformable Property of a Convex Lens

Besides the properties of converging, diverging, collimating, and forming images, the convex lens also has a Fourier transformable property [1,6,8]. It takes place at the back focal plane of the lens. For instance, when a parallel beam (constant with respect to x_0 and y_0) is incident onto a convex lens, the distribution of light that results on the focal plane behind the lens is a point or delta function. The delta function is the Fourier transform of a constant. In this way, the convex lens can be considered to possess a *Fourier transformable* property.

Now consider a source (an input function or input mask) placed on the back surface of a convex lens, as shown in Fig. 1.23. The screen is located in the near field and hence Fresnel's near-field diffraction equation, Eq. (1.38), is used. The field on the screen is the Fourier transform of the field just behind the lens multiplied by the point spread function.

$$E(x_i, y_i) = \frac{1}{j\lambda z_i} e^{jk[z_i + (x_i^2 + y_i^2)/2z_i]}$$

$$\mathcal{F}\{E(x_0, y_0)e^{-jk(x_0^2 + y_0^2)/2f_0} \cdot e^{jk(x_0^2 + y_0^2)/2z_i}\}_{f_x=x_i/\lambda z_i, f_y=y_i/\lambda z_i} \quad (1.155)$$

If the observation is made at a distance $z_i = f_0$, the point spread phase factor is cancelled by the quadratic phase factor of the convex lens and Eq. (1.155)

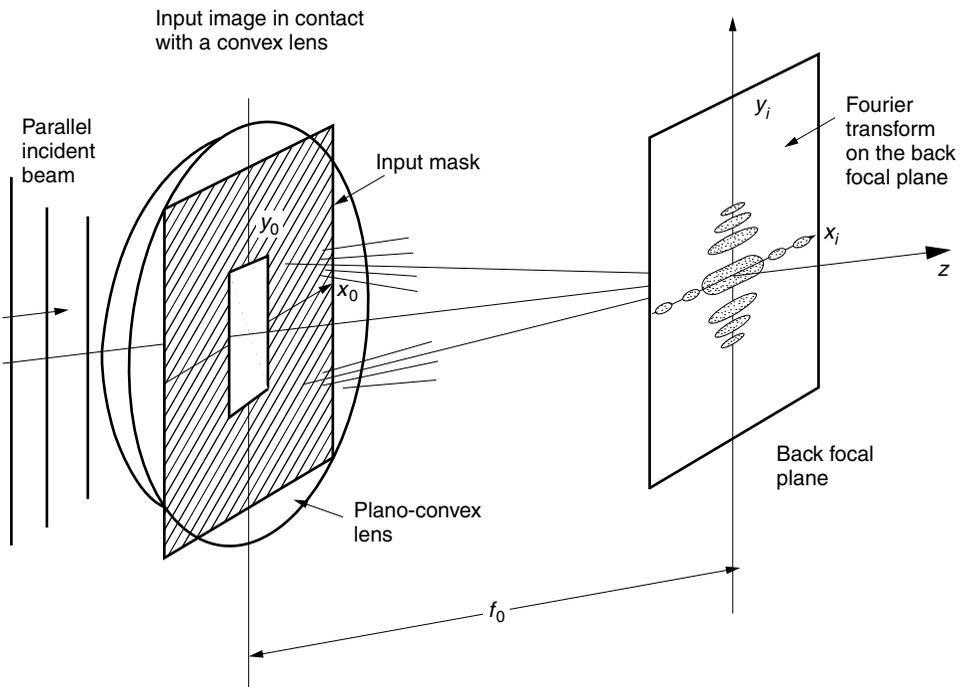


Figure 1.23 Fourier transform when the input image is pressed against a plano-convex lens.

becomes

$$E(x_i, y_i) = \frac{1}{j\lambda f_0} e^{jk[f_0 + (x_i^2 + y_i^2)/2f_0]} \varepsilon\left(\frac{x_i}{\lambda f_0}, \frac{y_i}{\lambda f_0}\right) \tag{1.156}$$

where

$$\varepsilon(f_x, f_y) = \mathcal{F}\{E(x_0, y_0)\}$$

Thus, the output $E(x_i, y_i)$ is the Fourier transform of the input function. It should be noted, however, that the output is not exactly the Fourier transform but is altered by a quadratic phase factor $\exp[jk(x_i^2 + y_i^2)/2f_0]$. Furthermore, the size of the Fourier transform pattern $\varepsilon(x_i/\lambda f_0, y_i/\lambda f_0)$ depends on the focal length of the lens.

Next, a different problem is investigated. Consider the case when the input function is placed at the focal plane in front of the lens, as arranged in Fig. 1.24. The field incident upon the lens is in the near field. Hence, the Fresnel diffraction pattern of the input function, in the form of convolution with the point spread function as given by Eq. (1.39), is used.

$$E_l(x, y) = E_0(x, y) * f_{f_0}(x, y) \tag{1.157}$$

Using the previous discovery that the field in the back focal plane of a convex lens is merely the Fourier transform of the input function (Eq. (1.156)), the distribution on

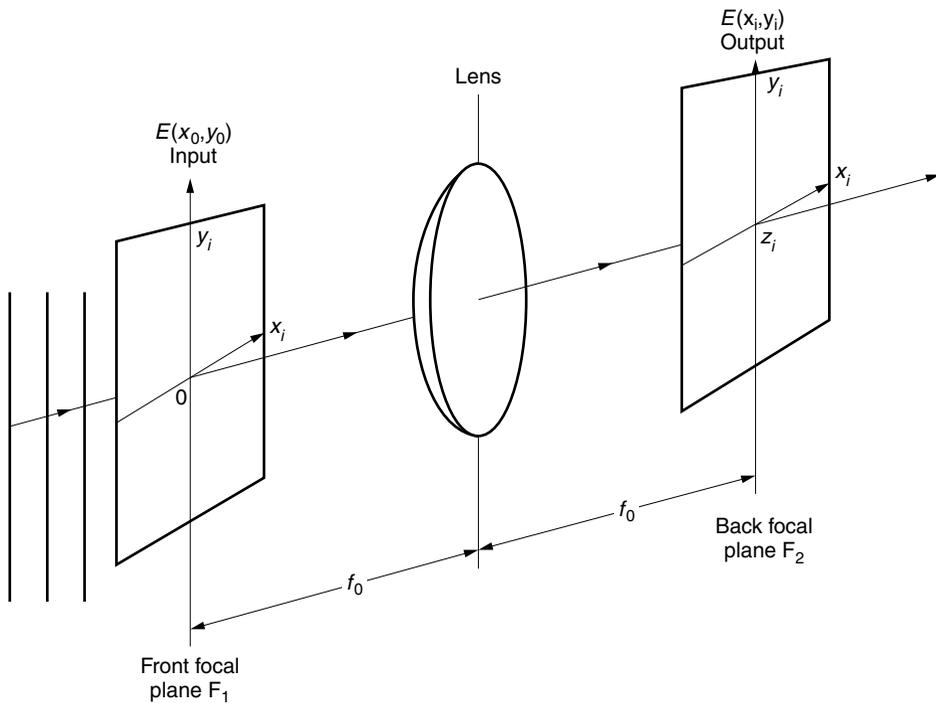


Figure 1.24 Fourier transform by a convex lens when the input is at the front focal plane F_1 and the Fourier transform is on the back focal plane F_2 .

the screen in the back focal plane is given by

$$E(x_i, y_i) = \frac{1}{j\lambda f_0} e^{jk[f_0 + (x_i^2 + y_i^2)/2f_0]} \mathcal{F}\{E_0(x, y) * f_{f_0}(x, y)\}_{f_x = x_i/\lambda f_0, f_y = y_i/\lambda f_0} \quad (1.158)$$

Using the product rule of Fourier transforms and the Fourier transform of the point spread function from Eq. (1.41), the field is equal to

$$E(x_i, y_i) = \frac{1}{j\lambda f_0} e^{jk[f_0 + (x_i^2 + y_i^2)/2f_0]} [\varepsilon(f_x, f_y) \cdot e^{jkf_0 - j\pi\lambda f_0(f_x^2 + f_y^2)}]_{f_x = x_i/\lambda f_0, f_y = y_i/\lambda f_0} \quad (1.159)$$

$$E(x_i, y_i) = \frac{e^{j2kf_0}}{j\lambda f_0} \varepsilon\left(\frac{x_i}{\lambda f_0}, \frac{y_i}{\lambda f_0}\right) \quad (1.160)$$

This final result can be compared to the previous findings for the lens system of Fig. 1.23. This formula is very similar except for the absence of the quadratic phase factor.

Thus, the Fourier transform without the quadratic phase factor is obtained when the input is placed in the front focal plane and observed on the back focal plane.

Lastly, consider what happens when the input is placed in an arbitrary converging beam, as shown in Fig. 1.25. In this case, the Fourier transform is obtained in a plane containing the point of convergence. It does not matter whether the converging beam has been made by a single lens or a composite of lenses, as long as a converging spherical wave is the incident input beam.

Before going into detail, the difference in the expressions for diverging and converging rays needs to be understood. Figure 1.26 shows how two phase fronts evolve as time elapses from $t = 1$ to $t = 3$. Both beams are propagating in the positive

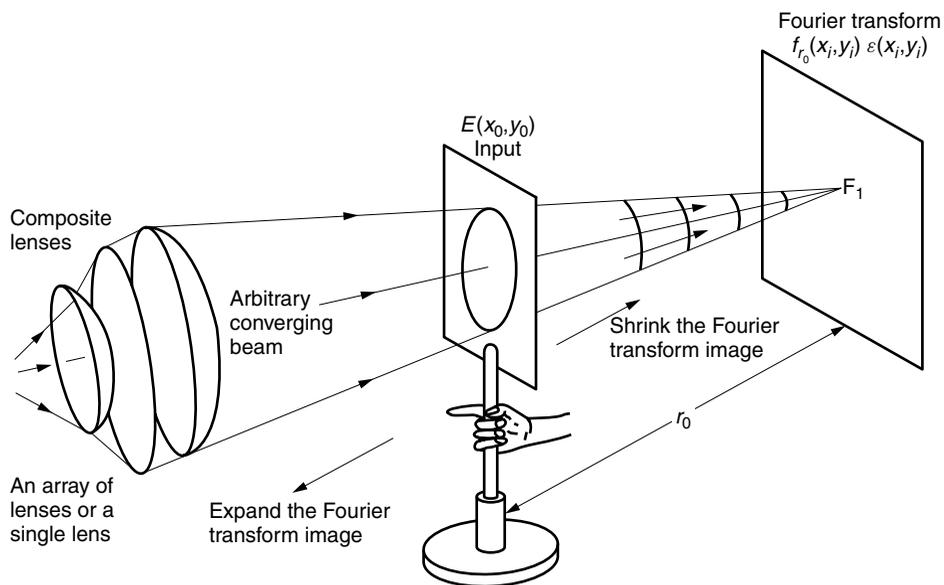


Figure 1.25 Fourier transform by a converging beam.

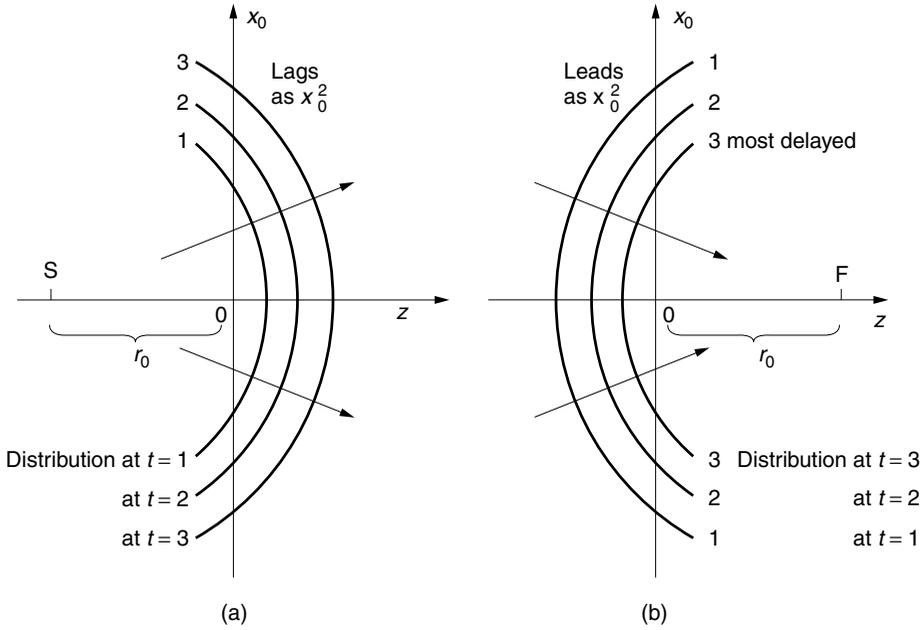


Figure 1.26 Diverging and converging spherical waves.

z direction and have a quadratic phase distribution in the transverse (x_i, y_i) plane. However, the signs of the phase are opposite. The difference in signs can be explained as follows. In Fig. 1.26a, if the phase is observed outward from the origin (increasing $x_0^2 + y_0^2$), the later or the more delayed phase is observed. Therefore, the factor is positive and is expressed as

$$E(x_0, y_0) = E_0 e^{jk(x_0^2+y_0^2)/2r_0} \tag{1.161}$$

where r_0 is the distance between the point source and the screen.

On the other hand, for the converging beam in Fig. 1.26b, the earlier or more leading phase is observed as the point of observation moves away from the center of the lens; thus, the phase factor has to be

$$E(x_0, y_0) = E_0 e^{-jk(x_0^2+y_0^2)/2r_0} \tag{1.162}$$

Now, let us return to the diffraction pattern of the input and consider the case when the input is inserted into a converging beam. From the Fresnel diffraction formula of Eq. (1.38) and the phase factor, the field observed at a distance r_0 from the input is

$$E(x_i, y_i) = \frac{1}{j\lambda z_i} e^{jk[r_0+(x_i^2+y_i^2)/2r_0]} \mathcal{F}\{e^{-jk(x_0^2+y_0^2)/2r_0} E(x_0, y_0) e^{jk(x_0^2+y_0^2)/2r_0}\}_{f_x=x_i/\lambda r_0, f_y=y_i/\lambda r_0} \tag{1.163}$$

The phase of the converging beam and the point spread function cancel each other, and the final expression becomes

$$E(x_i, y_i) = \frac{e^{jk[r_0+(x_i^2+y_i^2)/2r_0]}}{j\lambda r_0} \varepsilon\left(\frac{x_i}{\lambda r_0}, \frac{y_i}{\lambda r_0}\right) \quad (1.164)$$

The Fourier transform is obtained, but with the inclusion of the quadratic phase factor once again. This time, however, the size of the Fourier transform diffraction pattern can be controlled without changing the focal length of the lens. The size of the image is enlarged as r_0 is increased, until the input is obstructed by the lens.

1.5.5 How Can a Convex Lens Perform the Fourier Transform?

Figure 1.27 gives a pictorial explanation of how a convex lens performs the Fourier transform of the input image. The input is placed in the front focal plane of the convex lens, and the output is observed in the back focal plane. The observation point $P(x_i)$ will always be at $x_i = x_i$ in the back focal plane.

In case (1), a delta function source is placed at the origin of the input plane. The emergent light from the lens is a parallel beam. The phase at $P(x_i)$ is the same as that at the origin. In case (2), the light source is moved to $x_0 = x_1$. The parallel beam will be tilted and propagates slightly downward. The phase at $P(x_i)$ is leading that at the origin by $-jk(\sin\theta_i)x_i$ radians or approximately $-j(2\pi/\lambda)(x_1/f_0)x_i$ radians. The field at $P(x_i)$ is

$$E(x_i) = E_1 \exp\left(-j\frac{2\pi}{\lambda} \frac{x_i}{f_0} x_1\right)$$

In case (3), one more source is added at $x_0 = x_2$, and the sum of the contributions of the two sources is

$$E(x_i) = E_1 \exp\left(-j\frac{2\pi}{\lambda} \frac{x_i}{f_0} x_1\right) + E_2 \exp\left(-j\frac{2\pi}{\lambda} \frac{x_i}{f_0} x_2\right)$$

In case (4), a distributed source $E(x_0)$ is placed in the input plane. The contribution of the distributed source is expressed by an integral over the source plane.

$$E(x_i) = \int E(x_0) \exp\left(-j\frac{2\pi}{\lambda} \frac{x_i}{f_0} x_0\right) dx_0$$

Thus, the lens performs the Fourier transform of the input image and the output is

$$E(x_i) = \mathcal{F}\{E(x_0)\}_{f_x=x_i/\lambda f_0}$$

1.5.6 Invariance of the Location of the Input Pattern to the Fourier Transform

Even when the input pattern is moved sideways, the location of the Fourier transform pattern stays in close proximity to the back focal point of the lens. Figure 1.28 shows how the ray paths change as the input object (letter envelope) is moved vertically in the x - y plane and horizontally along the z axis. From the figure, we see that the

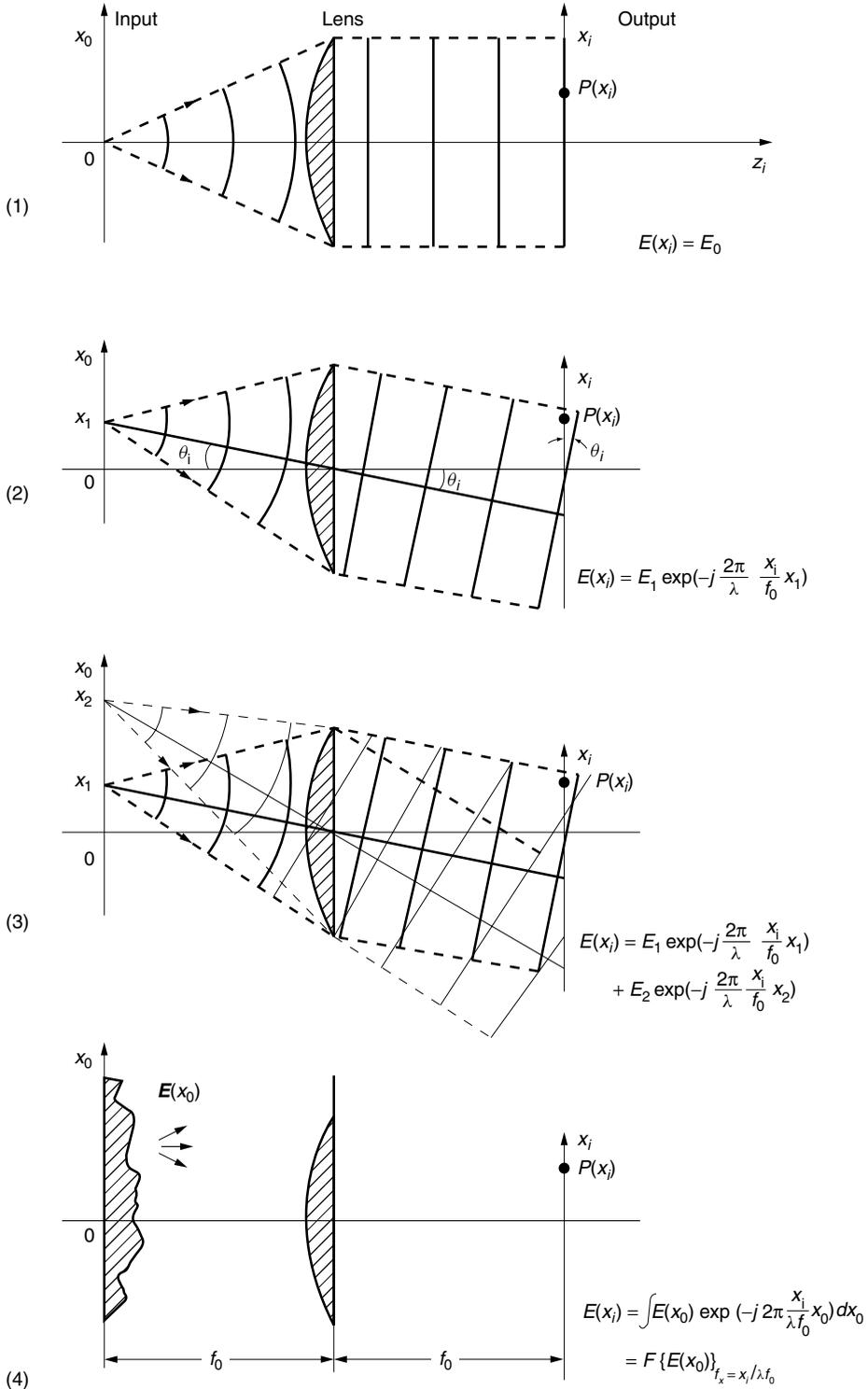


Figure 1.27 How a convex lens performs the Fourier transform.

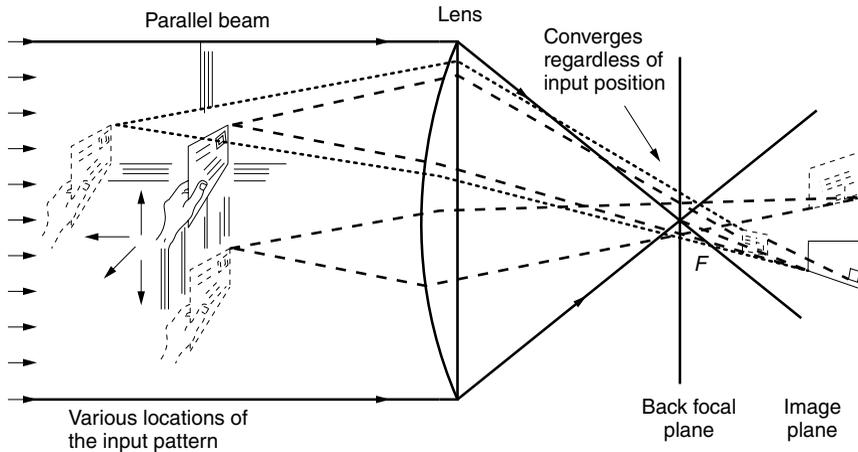


Figure 1.28 Demonstrating that the Fourier transform is confined to the vicinity of the back focal point F regardless of the location of the input image.

incident parallel beam converges to the back focal point and the diffraction pattern of the input remains around the focal point regardless of the location of the input. When the envelope is lowered, the scattered beam starts from a lower location but the scattered field first converges to the center of the back focal plane and then forms an image at a higher location. In fact, the Fourier transform pattern is usually confined within a small range. For instance, with $\lambda = 0.63 \mu\text{m}$, the focal length $f = 10 \text{ cm}$, and with a maximum spatial frequency of $f_x = 100 \text{ lines/mm}$, the Fourier transform appears within $x_i = \pm f_x \lambda f = \pm 6.3 \text{ mm}$ at the back focal point. This is one of the most useful features of the Fourier transform nature of a lens. In practical applications, such as mass processing of addresses on letters, this means that stringent positioning requirements of the input image are not necessary.

The reason for this phenomenon mathematically is that the Fourier transform of the shifted input $g(x - a)$ is $e^{-j2\pi f a} G(f)$. The shift causes the phase shift $e^{-j2\pi f a}$. However, the human eye cannot recognize phase shifts and thus the pattern does not show any change when the input is shifted. However, this is only true with a translational shift and not true with a rotational shift.

1.6 SPATIAL FREQUENCY APPROACHES IN FOURIER OPTICS

Another way of calculating the diffraction pattern will be introduced. The given input pattern is first decomposed into its constituent spatial frequencies by Fourier transforming the input, and then the propagated field of each spatial frequency component is calculated. The desired diffraction pattern of the given input pattern is the sum of the propagated patterns of the spatial frequency components [1, 11].

1.6.1 Solution of the Wave Equation by Means of the Fourier Transform

With the geometry shown in Fig. 1.2, the field $\mathbf{E}(x_i, y_i, z, t)$ at $z = z_i$ will be calculated from the spectrum of the input field at $z = 0$. Assuming a sinusoidal time dependence,

the field is expressed as

$$\mathbf{E}(x, y, z, t) = \mathbf{E}(x, y, z)e^{-j\omega t} \quad (1.165)$$

Equation (1.165) has to satisfy the wave equation

$$\nabla^2 \mathbf{E} + k^2 \mathbf{E} = 0 \quad (1.166)$$

If Cartesian coordinates are used to express Eq (1.166), the general solutions for E_x , E_y , and E_z are all identical. What makes E_x , E_y , and E_z different is that the boundary conditions depend on the components. Let the solution of Eq. (1.166) be denoted as $E(x, y, z)$. This approach is called the *scalar wave approach*. The scalar wave approach is much simpler than the vector wave approach but is less accurate because it assumes that the same boundary conditions are applicable to both normal and tangential components. Accepting this trade of simplicity for accuracy, the scalar equivalent of Eq. (1.165) is

$$E(x, y, z, t) = E(x, y, z)e^{-j\omega t} \quad (1.167)$$

Thus, the scalar wave equation becomes

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + \frac{\partial^2 E}{\partial z^2} + k^2 E = 0 \quad (1.168)$$

Now, the wave equation (1.168) can be solved using the Fourier transform method [11]. The Fourier transform of E with respect to x in the f_x, f_y domain is

$$\mathcal{F}_x\{E\} = \varepsilon(f_x, y, z) \quad (1.169)$$

The derivative rule of the Fourier transform is

$$\mathcal{F}\left\{\frac{\partial E}{\partial x}\right\} = j2\pi f_x \varepsilon(f_x, y, z) \quad (1.170)$$

The Fourier transform of Eq. (1.168) with respect to x gives

$$(j2\pi f_x)^2 \varepsilon(f_x, y, z) + \frac{\partial^2}{\partial y^2} \varepsilon(f_x, y, z) + \frac{\partial^2}{\partial z^2} \varepsilon(f_x, y, z) + k^2 \varepsilon(f_x, y, z) = 0 \quad (1.171)$$

Similarly, the Fourier transform of the above equation with respect to y gives

$$(j2\pi f_x)^2 \varepsilon(f_x, f_y, z) + (j2\pi f_y)^2 \varepsilon(f_x, f_y, z) + \frac{\partial^2}{\partial z^2} \varepsilon(f_x, f_y, z) + k^2 \varepsilon(f_x, f_y, z) = 0 \quad (1.172)$$

Hence, the Fourier transform of the scalar wave equation, Eq. (1.168), with respect to both x and y is

$$\left[\frac{\partial^2}{\partial z^2} + (2\pi f_z)^2 \right] \varepsilon(f_x, f_y, z) = 0 \quad (1.173)$$

where

$$f_z = \sqrt{f_s^2 - f_x^2 - f_y^2} \quad (1.174)$$

f_s is the spatial frequency in the direction of propagation and $f_s = 1/\lambda$ as defined by Eq. (1.20). Equation (1.173) is a second order partial differential equation with respect to z whose solution is

$$\varepsilon(f_x, f_y, z) = Ae^{j2\pi f_z z} + Be^{-j2\pi f_z z} \quad (1.175)$$

The first term is a forward wave while the second term is a backward wave. The values of A and B are to be found from the boundary conditions of either ε or its derivative. We will restrict ourselves to the simple case where only the forward wave is present, and $B = 0$. By setting $z = 0$ in Eq. (1.175), A is found to be $\varepsilon(f_x, f_y, 0)$. The solution becomes

$$\varepsilon(f_x, f_y, z) = \varepsilon(f_x, f_y, 0)e^{j2\pi f_z z} \quad (1.176)$$

Finally, we solve for the field $E(x_i, y_i, z_i)$ by taking the inverse Fourier transform of Eq. (1.176). This gives

$$E(x_i, y_i, z_i) = \iint_{-\infty}^{\infty} \varepsilon(f_x, f_y, 0)e^{j2\pi\sqrt{f_s^2 - f_x^2 - f_y^2}z_i} e^{j2\pi f_x x_i + j2\pi f_y y_i} df_x df_y \quad (1.177)$$

where

$$\varepsilon(f_x, f_y, 0) = \iint_{-\infty}^{\infty} E(x_0, y_0, 0)e^{-j2\pi f_x x_0 - j2\pi f_y y_0} dx_0 dy_0 \quad (1.178)$$

The combination of these equations is called the *Rayleigh–Sommerfeld diffraction formula* and provides the field at $z = z_i$ from the Fourier transform of the input field at $z = 0$.

Even though Eqs. (1.177) and (1.178) are simple expressions, it is difficult to obtain a highly accurate representation of $E(x_0, y_0, 0)$ that expresses the field for an aperture of finite size. For example, the field of an aperture illuminated from the back by a plane wave is almost, but not exactly, uniform across the aperture. The aperture contains contributions from the waves scattered by the edges, as well as multiple scatterings between the facing edges. As a result, it is quite difficult to obtain the exact expression of the input field.

Example 1.6 Demonstrate that the Rayleigh–Sommerfeld and Fresnel–Kirchhoff diffraction formulas provide the same answer using the example of the diffraction pattern of a pinhole. The pinhole is located at the origin of the input plane and is illuminated by a plane wave of unit amplitude from behind as shown in Fig. 1.29a.

Solution

1. Solution by the Rayleigh–Sommerfeld formula.

In this method, the Fourier transform ε of the input field is first obtained and then ε is allowed to propagate to the output screen where the inverse Fourier transform is performed to obtain the final result.

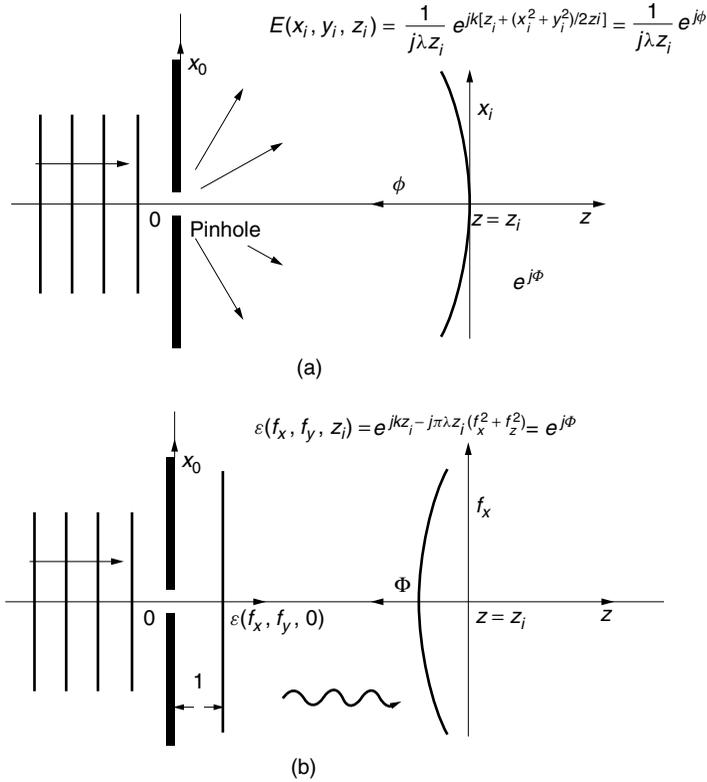


Figure 1.29 Field from a pinhole in space and spatial frequency domains. (a) Space domain. (b) Spatial frequency domain.

The pinhole is represented by

$$E(x_0, y_0) = \delta(x_0)\delta(y_0) \quad (1.179)$$

Inserting Eq. (1.179) into (1.178) gives

$$\varepsilon(f_x, f_y, 0) = 1 \quad (1.180)$$

Since there is no backward wave in the region between the input and output planes, $B = 0$ in Eq. (1.175). Let $\varepsilon(f_x, f_y, 0)$ propagate to the output screen using Eq. (1.176):

$$\varepsilon(f_x, f_y, z_i) = e^{j2\pi f_z z_i} \quad (1.181)$$

If we apply the para-axial approximation, which means that the propagation directions of the component waves are almost along the z axis, then f_s satisfies

$$f_s^2 \gg f_x^2 + f_y^2 \quad (1.182)$$

Recalling that

$$f_s = \frac{1}{\lambda}$$

the binomial expansion of Eq. (1.174) becomes

$$f_z = \frac{1}{\lambda} - \frac{\lambda}{2}(f_x^2 + f_y^2) \quad (1.183)$$

Inserting Eq. (1.183) into (1.181) gives

$$\varepsilon(f_x, f_y, 0)e^{j2\pi f_z z_i} = e^{jkz - j\pi\lambda z_i(f_x^2 + f_y^2)} \quad (1.184)$$

Using the Fourier transform relationship of Eq. (1.44), Eq. (1.184) is inverse Fourier transformed to obtain the final result as

$$E(x_i, y_i, z_i) = \frac{1}{j\lambda z_i} e^{jk[z_i + (x_i^2 + y_i^2)/2z_i]} \quad (1.185)$$

2. Solution by the Fresnel–Kirchhoff integral.

Inserting Eq. (1.179) into (1.38) gives

$$\begin{aligned} E(x_i, y_i, z_i) &= \frac{1}{j\lambda z_i} e^{jk[z_i + (x_i^2 + y_i^2)/2z_i]} \iint_{-\infty}^{\infty} \delta(x_0)\delta(y_0) \\ &\quad \times e^{jk(x_0^2 + y_0^2)/2z_i} \cdot e^{-j2\pi(f_x x_0 + f_y y_0)} dx_0 dy_0 \end{aligned} \quad (1.186)$$

The integral in Eq. (1.186) is unity and Eq. (1.186) is the same as Eq. (1.185). \square

Example 1.7 As shown in Fig. 1.30, light is incident from an optically dense medium with refractive index n into free space. The interface between the medium and free space is in the plane of $z = 0$. The propagation direction of the incident wave is in the x – z plane and the incident angle to the interface is θ_i . Find the field at point (x_i, y_i, z_i) in free space and its propagation direction.

Solution The wavelength λ_{x_0} in the x_0 direction at the interface is

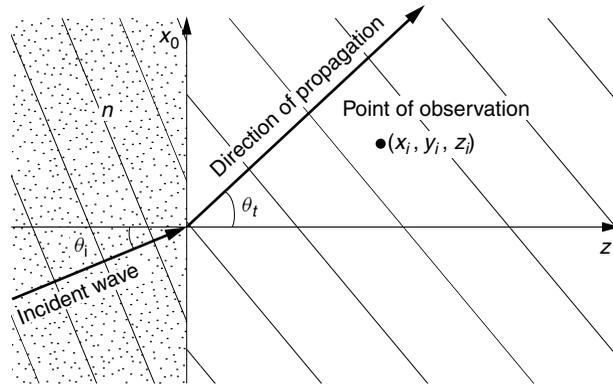
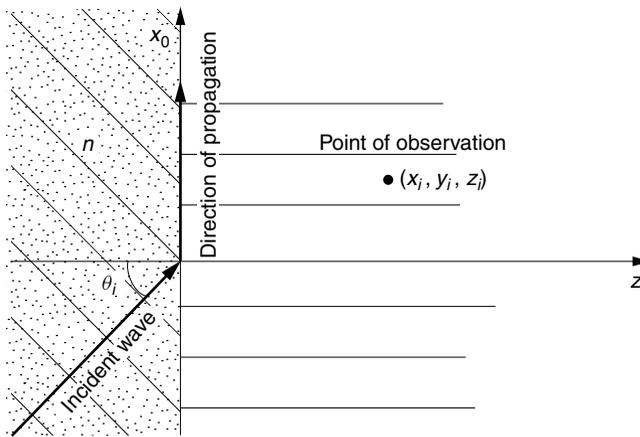
$$\lambda_{x_0} = \frac{\lambda}{n \sin \theta_i} \quad (1.187)$$

The corresponding spatial frequency from Eq. (1.13) is

$$f_{x_0} = n f_s \sin \theta_i \quad (1.188)$$

Note that B in Eq. (1.175) is zero because there is no wave propagating in the negative z direction in the free-space region. Thus, the field $E(x_0, y_0, 0)$ in the $z = 0$ plane is

$$E(x_0, y_0, 0) = E_0 e^{j2\pi n f_s x_0 \sin \theta_i} \quad (1.189)$$


 (a) $n \sin \theta_i < 1$

 (b) $n \sin \theta_i > 1$
Figure 1.30 Transmission from an optically dense medium into a less dense medium is calculated.

From Eq. (1.189), Eq. (1.178) becomes

$$\varepsilon(f_x, f_y, 0) = E_0 \delta(f_x - n f_s \sin \theta_i) \delta(f_y) \quad (1.190)$$

Thus, the input field contains only one spatial frequency component, $n f_s \sin \theta_i$. Next, Eq. (1.177) will be calculated:

$$E(x_i, y_i, z_i) = \iint E_0 \delta(f_x - n f_s \sin \theta_i) \delta(f_y) e^{j2\pi \sqrt{f_s^2 - f_x^2 - f_y^2} z_i} e^{j2\pi (f_x x_i + f_y y_i)} df_x df_y \quad (1.191)$$

Applying Eq. (1.102) to (1.191) gives

$$E(x_i, y_i, z_i) = E_0 \exp[j2\pi \sqrt{f_s^2 - n^2 f_s^2 \sin^2 \theta_i} z_i + j2\pi n f_s (\sin \theta_i) x_i] \quad (1.192)$$

From Snell's law at the boundary,

$$n \sin \theta_i = \sin \theta_t \quad (1.193)$$

the final result is

$$E(x_i, y_i, z_i) = E_0 \exp[j2\pi f_s(\cos \theta_t)z_i + j2\pi f_s(\sin \theta_t)x_i] \quad (1.194)$$

Equation (1.194) is a plane wave whose propagation unit vector is

$$\hat{\mathbf{e}} = \sin \theta_t \hat{\mathbf{i}} + \cos \theta_t \hat{\mathbf{k}} \quad (1.195)$$

The propagation direction with respect to the normal of the input plane is

$$\tan^{-1} \frac{e_x}{e_z} = \theta_t \quad (1.196)$$

Next, the case of the evanescent wave will be explained. Equation (1.192) can be rewritten as

$$E(x_i, y_i, z_i) = E_0 \exp[-2\pi f_s \sqrt{n^2 \sin^2 \theta_i - 1} z_i + j2\pi n f_s(\sin \theta_i)x_i] \quad (1.197)$$

When either θ_i or n is large and the condition $n \sin \theta_i > 1$ is satisfied, the amplitude of the wave decays exponentially in the z direction. This is an example of an evanescent wave. For the evanescent wave, the phase varies in the x_i direction whereas the amplitude of the wave does not, as shown by Eq. (1.197) and Fig. 1.30b. The evanescent wave is a very important subject and will be treated in more detail in the next chapter. \square

1.6.2 Rayleigh–Sommerfeld Integral

This section is devoted to explaining the conceptual differences between the Rayleigh–Sommerfeld and the Fresnel–Kirchhoff diffraction formulas. Mathematically, the two integrals are equivalent, and Appendix A of Volume I presents a general proof of this. Conceptually, there are differences between the two approaches. Understanding these differences is important so that the most appropriate choice is made in solving a given problem.

The Fresnel–Kirchhoff diffraction formula, Eq. (1.28), is repeated here for convenience.

$$E(x_i, y_i, z_i) = \frac{1}{j\lambda} \iint \underbrace{\frac{e^{j2\pi f_s r}}{r}}_{\text{Spherical wave}} \underbrace{E(x_0, y_0)}_{\text{Amplitude of the spherical wave}} dx_0 dy_0 \quad (1.198)$$

$$\text{with } r = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2 + z_i^2}$$

As mentioned in Section 1.2, $e^{j2\pi f_s r}/r$ represents a spherical wave emanating from the point (x_0, y_0) with $E(x_0, y_0) dx_0 dy_0$ as its amplitude. The representation of the source function as a collection of spherical wave point sources is indicated

by $S_1, S_2, S_3, \dots, S_n$ in Fig. 1.31a. In order to find the field observed at P , the contributions of all spherical wave sources $S_1, S_2, S_3, \dots, S_n$ are integrated over the entire shape of the source.

Next, let us examine the Rayleigh–Sommerfeld diffraction formula. Equation (1.177) is rewritten below:

$$E(x_i, y_i, z_i) = \iint \underbrace{e^{j2\pi f_x x + j2\pi f_y y + j2\pi f_z z}}_{\substack{\text{Plane wave propagating} \\ \text{in the } (f_x \mathbf{i}, f_y \mathbf{j}, f_z \mathbf{k}) \\ \text{direction}}} \underbrace{\varepsilon(f_x, f_y, 0)}_{\text{Amplitude}} df_x df_y \quad (1.199)$$

$$\text{with } f_z = \sqrt{f_s^2 - f_x^2 - f_y^2}$$

The first factor of the integral in Eq. (1.199) is a plane wave component propagating in the direction

$$(f_x \mathbf{i}, f_y \mathbf{j}, f_z \mathbf{k}) \quad (1.200)$$

The second factor is the amplitude of the plane wave. The integral with respect to f_x and f_y means the integration of the plane wave contributions from all propagating directions. Since $f_s = 1/\lambda$ is given, once f_x and f_y are specified, the value of f_z is accordingly set from the lower equation of Eq. (1.199). There are few noteworthy points with regard to this integral. As shown in Fig. 1.31b, the integral includes not only the plane wave B whose wave normal is aimed at P but also all other wave normals denoted by A, C, \dots . However, a wave normal is not the same thing as a light beam. Any line drawn parallel to the propagation direction of a given wave can be the wave normal of that same wave. For example, the wave represented by B' in Fig. 1.31b is identical with the wave represented by B .

It is also important to include the contributions of plane waves propagating in the negative x_0 direction, such as shown by A in Fig. 1.31b.

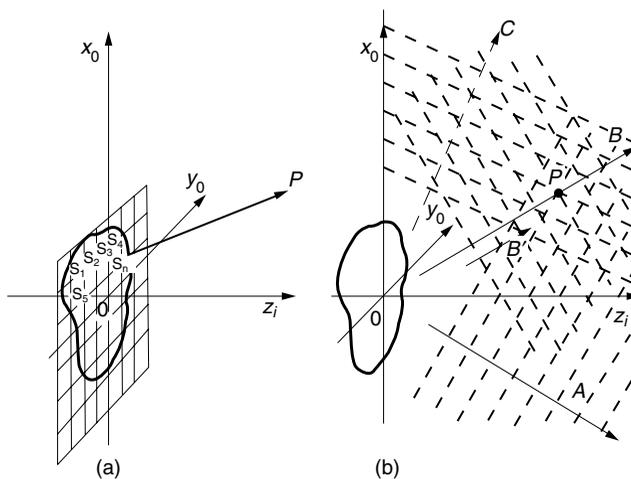


Figure 1.31 Conceptual comparison of the diffraction formulas. (a) Fresnel–Kirchhoff. (b) Rayleigh–Sommerfeld.

According to Eq. (1.199), another significant point to remember is that f_z becomes an imaginary number for

$$f_x^2 + f_y^2 > f_s^2 \tag{1.201}$$

When f_z is an imaginary number, the integrand becomes an evanescent wave whose magnitude decays exponentially with z . The contribution of the evanescent wave is insignificant unless the point of observation is close to the source and z is small.

1.6.3 Identifying the Spatial Frequency Components

The spatial frequency components shown in Fig. 1.31b can be identified by using a convex lens as illustrated in Fig. 1.32. Only two spatial frequency components are considered. Both of them are incident from the left of the convex lens. Because both waves are plane waves, they are focused on the back focal plane. The component with $f_x = f_y = 0$ is incident normal to the lens and is focused at $(0, 0, f)$, where f is the focal length of the convex lens. The component with $f_x \neq 0, f_y = 0$ is incident to the lens with an incident angle

$$\theta = \tan^{-1}(f_x/f_z) \tag{1.202}$$

and is focused on the back focal plane at $(x_i, 0, f)$, where

$$x_i = f \tan \theta \doteq f \lambda f_x \tag{1.203}$$

and where the approximation of $f_z = f_s$ was made. The location of the focused light rises along the x_i axis as f_x is increased. The higher the location, the higher the spatial frequency of the converging light. This means that by placing a mask of a predetermined

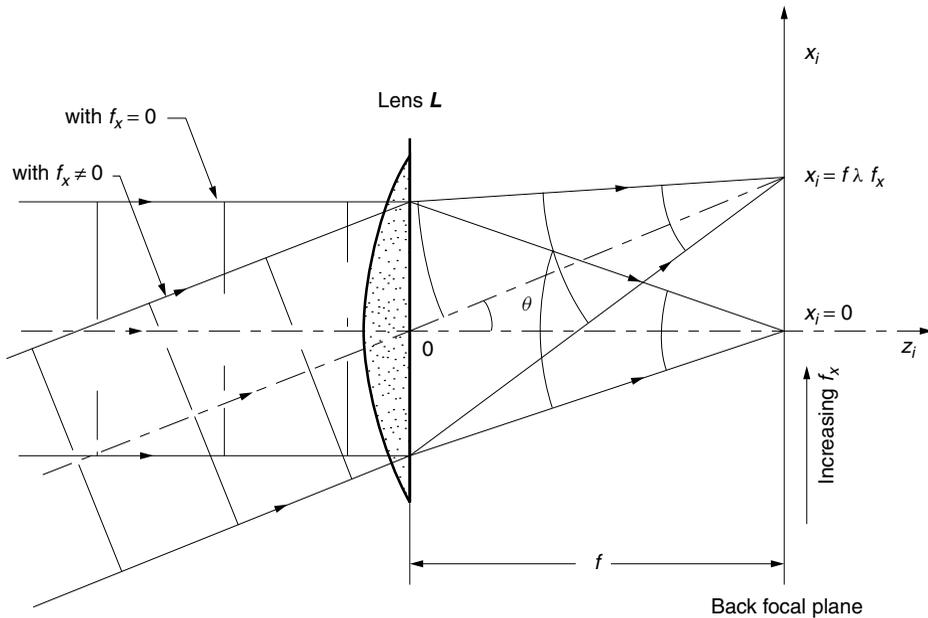


Figure 1.32 Identifying the spatial frequency components by means of a convex lens.

transmittance distribution in the back focal plane of the convex lens, the waves of a particular spatial frequency component can be either selectively transmitted through or blocked by this mask. The following sections elaborate on this technique. (See Section 1.2 for the case without a lens.)

1.7 SPATIAL FILTERS

This section is devoted to a discussion of optical signal processing based on manipulation of spatial frequency components.

1.7.1 Image Processing Filters

Figure 1.33 shows different arrangements for image processing using various types of spatial frequency filters [1,8]. For all the cases, there is a point source that is collimated

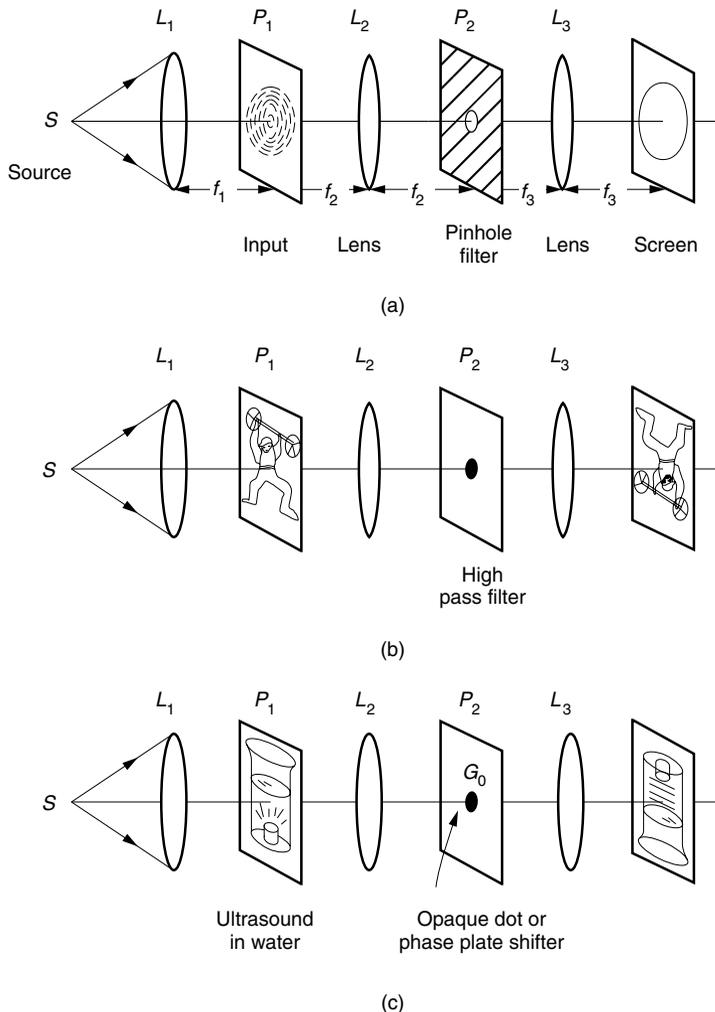


Figure 1.33 Various types of spatial filters. (a) Low-pass filter. (b) High-pass filter. (c) Schlieren camera. (d) Spatial derivative operation. (e) Step and repeat operation.

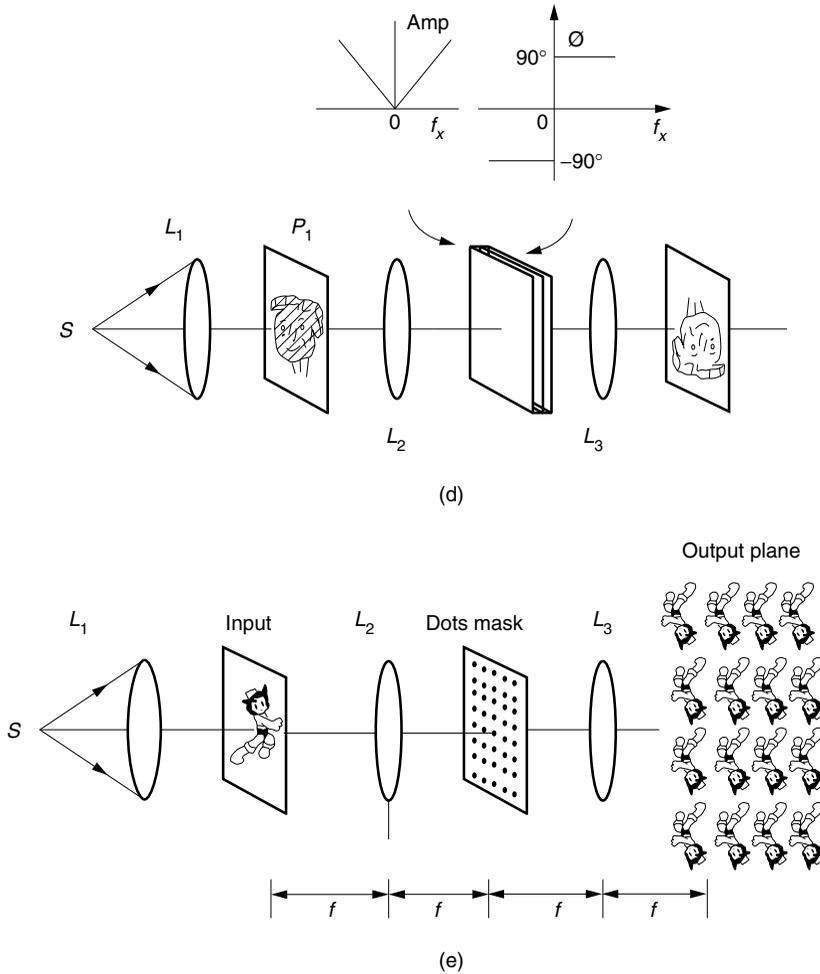


Figure 1.33 (Continued)

by lens L_1 . The collimated beam illuminates a transparency overlay forming the input image. Lens L_2 Fourier transforms the transparency pattern in its back focal plane. The spatial filter is placed in the back focal plane of L_2 and modifies the pattern. The modified pattern is Fourier transformed again by means of lens L_3 , and the processed image is finally projected onto the screen.

Figure 1.33a shows a low-pass spatial filter. In the plane of the spatial filter, the spatial frequency is zero on the z axis and increases linearly with distance from the axis. The higher frequency components, which are diffracted to the area away from the axis, are blocked by the mask. A typical low-pass filter consists of a pinhole and a lens combination, commonly marketed as a “spatial filter.”* This filter filters out all the spatial frequency components except the zero spatial frequency, resulting in a pure parallel beam.

* Keep in mind that the true meaning of a spatial filter is broader in scope than the marketplace meaning of spatial filter.

Figure 1.33b shows an example of image processing with a high-pass spatial filter. Here, the lower spatial frequency components are suppressed. In general, sharp edges and fine lines in the image generate higher spatial frequency components, which are then accentuated in the image. Thus, fine point or edge enhancement is achieved by this type of filter.

Figure 1.33c shows an example of a Schlieren camera. This filter is used to view transparent objects. It is normally difficult to view or photograph objects that are transparent although they may have varying indices of refraction within them. Examples include such objects as microbes, turbulent air, or ultrasound patterns in a liquid. The difficulty is that the images of these objects have only phase variations, which neither our eyes nor an ordinary camera can detect.

The Schlieren camera, however, creates an interference pattern between the image and a constant phase reference wave. The constant phase reference wave is generated by placing an opaque dot at the back focal point of lens L_2 on P_2 . The generation of the reference wave is explained as follows. The zero spatial frequency is blocked by the opaque dot. Blocking the zero spatial frequency component means there is zero field at the location of the opaque dot, and a zero field is equivalent to the sum of two waves of equal amplitude and opposite phase. Thus, the field at the location of the opaque dot can be expressed as the sum of the original zero spatial frequency wave, and a wave of equal amplitude but opposite phase (reference wave). This reference wave can be thought of as a fictitious point source located at the front focal point of lens L_3 and projecting a constant phase reference field onto the screen, while the input wave is Fourier transformed by L_3 , forming an inverted image on the same screen. The superposition of image and reference waves on the screen creates an interference pattern. The contours of the constructive interference are the brightest and those of the destructive interference are the darkest. The phase variation of the object is converted into an intensity variation by the Schlieren camera.

A quarter-wave phase plate can also be used instead of an opaque dot (Problem 1.10). The effects of a Schlieren camera can also be replicated by placing a knife edge to block the entire lower half domain of the Fourier transform.

Figure 1.34 shows a photograph of a fetal mouse taken by a Schlieren camera. Practically no image is formed if an ordinary microscope is used.

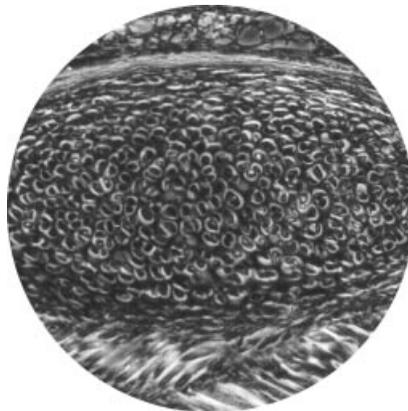


Figure 1.34 Schlieren photograph of a fetal mouse. (Courtesy of Olympus Optical Co., Ltd.)

Figure 1.33d shows an example of the spatial derivative operation. The Fourier transform of the derivative is

$$\mathcal{F}\{g'(x_0)\} = j2\pi f_x \cdot \mathcal{F}\{g\} \quad (1.204)$$

The filter placed at the back focal plane of the Fourier transform lens L_2 has transmittance characteristics of $\tau(f_x) = f_x$ in the spatial frequency domain. The filter, however, has to have this characteristic from $f_x = -\infty$ to $f_x = \infty$ in order for $\tau(f_x) = f_x$ to be realized. The positive f_x region can easily be realized, but the negative f_x region is much more difficult to realize. The phase of the negative f_x region can be reversed by a thin layer of a substance with a higher refractive index. This layer causes a π -radian phase delay. After another Fourier transform through lens L_3 , the image of $g'(x_0)$ is obtained on the screen. One of the practical applications of such an operation is in edge enhancement or outlining the input image for easy identification of the shape of the object.

Figure 1.33e shows how to generate the repeated image of the input. A mask of a grid of pinholes with periods a and b in the x and y directions is placed in the back focal plane of lens L_2 . The transmittance of such a mask is

$$\frac{1}{ab} \text{III} \left(\frac{x}{a} \right) \text{III} \left(\frac{y}{b} \right) \quad (1.205)$$

Thus, the field distribution after passing through the pinhole mask would be

$$\frac{1}{ab} \frac{e^{jkf}}{j\lambda f} G \left(\frac{x}{\lambda f}, \frac{y}{\lambda f} \right) \text{III} \left(\frac{x}{a} \right) \text{III} \left(\frac{y}{b} \right) \quad (1.206)$$

where $G(f_x, f_y)$ is the Fourier transform of the input $g(x, y)$ function. The effect to the field then, after passing through lens L_3 , is an additional Fourier transform producing the field

$$E(x_i, y_i) = -e^{j2kf} g(-x_i, -y_i) * \left\{ \text{III} \left(\frac{ax_i}{\lambda f} \right) \text{III} \left(\frac{by_i}{\lambda f} \right) \right\} \quad (1.207)$$

on the back focal plane of lens L_3 . From the above relationship, we see that the resultant image is a grid of repeated input images.

1.7.2 Optical Correlators

Optical correlators determine whether a particular image exists within a given picture. The two most popular types of correlators are the Vander Lugt correlator [1,12,13] and the joint transform correlator [14].

1.7.2.1 Vander Lugt Correlator

The Vander Lugt correlator (VLC) was first proposed by Vander Lugt in 1964. The VLC will be explained in two stages: a brief description of the principle in this section, followed by more detailed mathematical expressions in the next section.

Figure 1.35a shows a schematic of the VLC correlator. The input image $h(x_1, y_1)$ is interrogated to determine if it has the same shape as a given reference image $g(x_1, y_1)$.

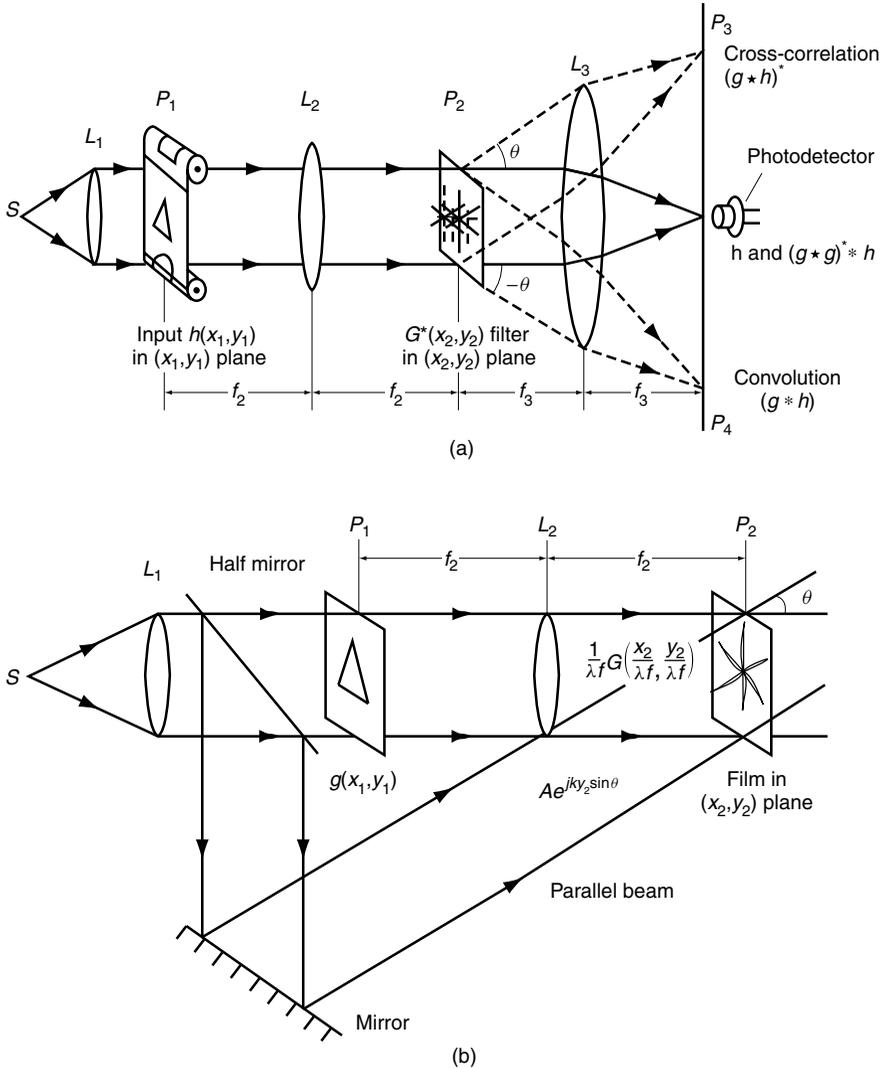


Figure 1.35 Vander Lugt correlator (VLC). (a) Layout. (b) Fabrication of a VLC filter.

First, with the input $h(x_1, y_1) = g(x_1, y_1)$ in the front focal plane of lens L_2 , its Fourier transform

$$G(f_x, f_y) = |G(f_x, f_y)|e^{j\phi(f_x, f_y)} \quad (1.208)$$

is projected onto the P_2 plane in the back focal plane of lens L_2 . The encoded filter, whose transmission is the complex conjugate of Eq. (1.208),

$$G^*(f_x, f_y) = |G(f_x, f_y)|e^{-j\phi(f_x, f_y)} \quad (1.209)$$

is inserted in the P_2 plane. The light transmitted through the filter is the product of Eqs. (1.208) and (1.209) and

$$E(f_x, f_y) = |G(f_x, f_y)|^2 \quad (1.210)$$

This function $E(f_x, f_y)$ has no spatial variation in phase, and the wavefront of the incident light on lens L_3 is parallel to the lens surface. The incident light converges to the back focal point of lens L_3 as indicated by the solid line where a photodetector is located to measure the light intensity.

If, however, the input is $h(x_1, y_1) \neq g(x_1, y_1)$, then the light projected through the filter is

$$H(f_x, f_y) = |H(f_x, f_y)|e^{j\phi'(f_x, f_y)} \quad (1.211)$$

and the transmitted light through the filter on P_2 becomes

$$E'(f_x, f_y) = |H(f_x, f_y)||G(f_x, f_y)|e^{j\phi'(f_x, f_y) - j\phi(f_x, f_y)} \quad (1.212)$$

The phase distribution is not uniform and not all the light converges to the back focal point of lens L_3 . Moreover, the patterns of $|H(f_x, f_y)|$ and $|G(f_x, f_y)|$ may not overlap and the total light power reaching lens L_3 will be less than the case of $h(x_1, y_1) = g(x_1, y_1)$. Hence, the input that best matches the reference image $g(x_1, y_1)$ gives the largest light intensity at the output.

1.7.2.2 Detailed Analysis of the Vander Lugt Correlator

It is not always a simple matter to fabricate a filter with the prescribed complex transmission coefficient. One way, which is similar to fabricating a hologram, is illustrated in Fig. 1.35b. From Eq. (1.160), when the reference image $g(x_1, y_1)$ is put in the front focal plane of lens L_2 , its Fourier transform,

$$E(x_2, y_2) = \frac{1}{j\lambda f} G\left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f}\right) \quad (1.213)$$

is projected onto the film in the back focal plane. The factor e^{j2kf_0} is suppressed. A parallel beam at angle θ to the normal of the film is added at the same time.

The transmittance $t(x_2, y_2)$ of the exposed and then developed film is

$$t(x_2, y_2) = t_0 - \beta \left| A e^{jk_y y_2 \sin \theta} + \frac{1}{j\lambda f} G\left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f}\right) \right|^2 \quad (1.214)$$

where t_0 and β specify the photographic characteristics of the film. Equation (1.214) is rewritten as

$$\begin{aligned} t(x_2, y_2) = & A^2 + \left(\frac{1}{\lambda f}\right)^2 \left| G\left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f}\right) \right|^2 \\ & + \frac{jA}{\lambda f} G^*\left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f}\right) e^{jk_y y_2 \sin \theta} \\ & + \frac{A}{j\lambda f} G\left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f}\right) e^{-jk_y y_2 \sin \theta} \end{aligned} \quad (1.215)$$

where t_0 and $-\beta$ were suppressed. The third term of Eq. (1.215) is the most important term. The desired quantity for the filter is $G^*(x_2/\lambda f, y_2/\lambda f)$ but this term has an extra factor of $e^{jk_y y_2 \sin \theta}$. This factor, however, does not harm the operation but just shifts the location of the correlation peak by θ degrees from the center.

Now, the input image $h(x_1, y_1)$ to be interrogated is installed in the front focal plane of lens L_2 in Fig. 1.35a and the filter is installed in the back focal plane of the same lens. The light transmitted through the filter is Fourier transformed by lens L_3 and its field $E(x_3, y_3)$ in the P_3 plane is given by

$$E(x_3, y_3) = \frac{1}{j\lambda f} \mathcal{F} \left\{ \frac{1}{j\lambda f} t(x_2, y_2) H \left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f} \right) \right\}_{f_x=x_3/\lambda f, f_y=y_3/\lambda f} \quad (1.216)$$

Equation (1.216) has the same number of terms as Eq. (1.215), and we can designate these terms as $E_1(x_3, y_3)$, $E_2(x_3, y_3)$, $E_3(x_3, y_3)$, and $E_4(x_3, y_3)$.

The first term is

$$E_1(x_3, y_3) = \frac{1}{j\lambda f} \mathcal{F} \left\{ \frac{A^2}{j\lambda f} H \left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f} \right) \right\}_{f_x=x_3/\lambda f, f_y=y_3/\lambda f} \quad (1.217)$$

Recall that lens L_3 performs the Fourier transform but not the inverse Fourier transform, and

$$E_1(x_3, y_3) = -A^2 h(-x_3, -y_3) \quad (1.218)$$

The inverted image of the input picture is seen around the origin of the P_3 plane.

The contribution of the second term is

$$E_2(x_3, y_3) = \frac{1}{j\lambda f} \mathcal{F} \left\{ \frac{1}{j(\lambda f)^3} \left| G \left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f} \right) \right|^2 H \left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f} \right) \right\}_{f_x=x_3/\lambda f, f_y=y_3/\lambda f} \quad (1.219)$$

Changing variables as

$$\frac{x_2}{\lambda f} = \xi, \quad \frac{y_2}{\lambda f} = \eta$$

Eq. (1.219) becomes

$$\begin{aligned} E_2(x_3, y_3) &= -\frac{1}{(\lambda f)^2} \iint_{-\infty}^{\infty} G(\xi, \eta) G^*(\xi, \eta) H(\xi, \eta) \{e^{-j2\pi\lambda f(f_x\xi + f_y\eta)}\} d\xi d\eta_{f_x=x_3/\lambda f, f_y=y_3/\lambda f} \\ & \quad (1.220) \end{aligned}$$

$$\begin{aligned} E_2(x_3, y_3) &= -\frac{1}{(\lambda f)^2} g(-\lambda f f_x, -\lambda f f_y) * g^*(\lambda f f_x, \lambda f f_y) * h(-\lambda f f_x, -\lambda f f_y)_{f_x=x_3/\lambda f, f_y=y_3/\lambda f} \end{aligned}$$

Inserting f_x and f_y into the equation and using Rule (5) from the boxed note, $E_2(x_3, y_3)$ is expressed as

$$E_2(x_3, y_3) = -\frac{1}{(\lambda f)^2} [g(x_3, y_3) \star g(x_3, y_3)]^* * h(-x_3, -y_3) \quad (1.221)$$

$E_2(x_3, y_3)$ is the autocorrelation of $g(x_3, y_3)$ convolved with $h(-x_3, -y_3)$. It is spread around the origin and is considered background noise (zero order noise).

The contribution of the third term is the most important one.

$$E_3(x_3, y_3) = \frac{1}{j\lambda f} \mathcal{F} \left\{ \frac{A}{(\lambda f)^2} G^* \left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f} \right) H \left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f} \right) e^{jk y_2 \sin \theta} \right\}_{f_x=x_3/\lambda f, f_y=y_3/\lambda f} \quad (1.222)$$

Changing variables as

$$\frac{x_2}{\lambda f} = \xi, \quad \frac{y_2}{\lambda f} = \eta$$

Eq. (1.222) becomes

$$E_3(x_3, y_3) = \frac{A}{j\lambda f} \iint_{-\infty}^{\infty} G^*(\xi, \eta) H(\xi, \eta) \{ e^{j2\pi(\sin \theta/\lambda)\lambda f \eta} e^{-j2\pi\lambda f (f_x \xi + f_y \eta)} d\xi d\eta \}_{f_x=x_3/\lambda f, f_y=y_3/\lambda f} \quad (1.223)$$

Equation (1.223) can be expressed as the convolution of three factors. Using the rules in the boxed note, Eq. (1.223) is written as

$$E_3(x_3, y_3) = \frac{A}{j\lambda f} g^*(\lambda f f_x, \lambda f f_y) * h(-\lambda f f_x, -\lambda f f_y) * \delta \left[\lambda f f_x, \lambda f \left(f_y - \frac{\sin \theta}{\lambda} \right) \right]_{f_x=x_3/\lambda f, f_y=y_3/\lambda f} \quad (1.224)$$

Inserting f_x and f_y into the equation gives

$$E_3(x_3, y_3) = \frac{A}{j\lambda f} g^*(x_3, y_3) * h(-x_3, -y_3) * \delta(x_3, y_3 - f \sin \theta) \quad (1.225)$$

Using Rule (5) in the boxed note, Eq. (1.225) is expressed as

$$E_3(x_3, y_3) = \frac{A}{j\lambda f} [g(x_3, y_3) \star h(x_3, y_3)]^* * \delta(x_3, y_3 - f \sin \theta) \quad (1.226)$$

Equation (1.226) is the expression of the cross-correlation between g and h . When $g(x_1, y_1)$ and $h(x_1, y_1)$ match, the peak value rises at

$$(0, f \sin \theta) \quad (1.227)$$

in the P_3 plane as indicated by the dotted line.

Finally, the contribution of the fourth term of the Eq. (1.216) is considered.

$$E_4(x_3, y_3) = \frac{1}{j\lambda f} \mathcal{F} \left\{ \frac{-A}{(\lambda f)^2} G \left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f} \right) H \left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f} \right) e^{-jk y_2 \sin \theta} \right\}_{f_x=x_3/\lambda f, f_y=y_3/\lambda f} \quad (1.228)$$

$E_4(x_3, y_3)$ is quite similar to Eq. (1.221). The differences are the minus sign in the front, the absence of the complex conjugate sign on G , and the minus sign in the exponent

(a) The following are relationships for the Fourier transforms of Fourier transforms:

$$\begin{aligned}
 (1) \quad \mathcal{F}\{G(f)\} &= \int_{-\infty}^{\infty} G(f)e^{-j2\pi fx} df \\
 &= \int_{-\infty}^{\infty} G(f)e^{j2\pi f(-x)} df \\
 &= g(-x)
 \end{aligned}$$

$$\begin{aligned}
 (2) \quad \mathcal{F}\{G^*(f)\} &= \int_{-\infty}^{\infty} G^*(f)e^{-j2\pi fx} df \\
 &= \left[\int_{-\infty}^{\infty} G(f)e^{j2\pi fx} df \right]^* \\
 &= g^*(x)
 \end{aligned}$$

$$\begin{aligned}
 (3) \quad \mathcal{F}\{|G(f)|^2\} &= \mathcal{F}\{G(f) \cdot G^*(f)\} \\
 &= g(-x) * g^*(x)
 \end{aligned}$$

(b) The correlation symbol is \star and the correlation operation is defined as

$$\begin{aligned}
 (4) \quad g(x) \star h(x) &\triangleq \int g(\xi)h^*(\xi - x) d\xi \\
 &= \int g(\xi + x)h^*(\xi) d\xi
 \end{aligned}$$

The operation $g \star h$ is called the cross-correlation, and the operation $g \star g$ is called the autocorrelation. Thus, the relationship between the convolution and the cross-correlation becomes

$$\begin{aligned}
 (5) \quad g(x) * h^*(-x) &= \int g(\xi)h^*(\xi - x) d\xi \\
 &= g(x) \star h(x)
 \end{aligned}$$

and from Rules (3) and (5)

$$(6) \quad \mathcal{F}\{|G(f)|^2\} = [g(x) \star g(x)]^*$$

of the last factor. Thus, the contribution of the fourth term is obtained directly from Eq. (1.226) as

$$E_4(x_3, y_3) = \frac{jA}{\lambda f} g(-x_3, -y_3) * h(-x_3, -y_3) * \delta(x_3, y_3 + f \sin \theta) \quad (1.229)$$

The convolution of g and h appears around

$$(0, -f \sin \theta) \quad (1.230)$$

in the P_3 plane as indicated by the dotted line.

In summary, the cross-correlation term $E_3(x_3, y_3)$ peaks up when $h = g$, indicating a match. The peak appears at $(0, f \sin \theta)$ in the P_3 plane. The value of θ should be chosen large enough to ensure that the $E_3(x_3, y_3)$ peak is well separated from the fields $E_1(x_3, y_3)$ and $E_2(x_3, y_3)$ that are spread around the origin.

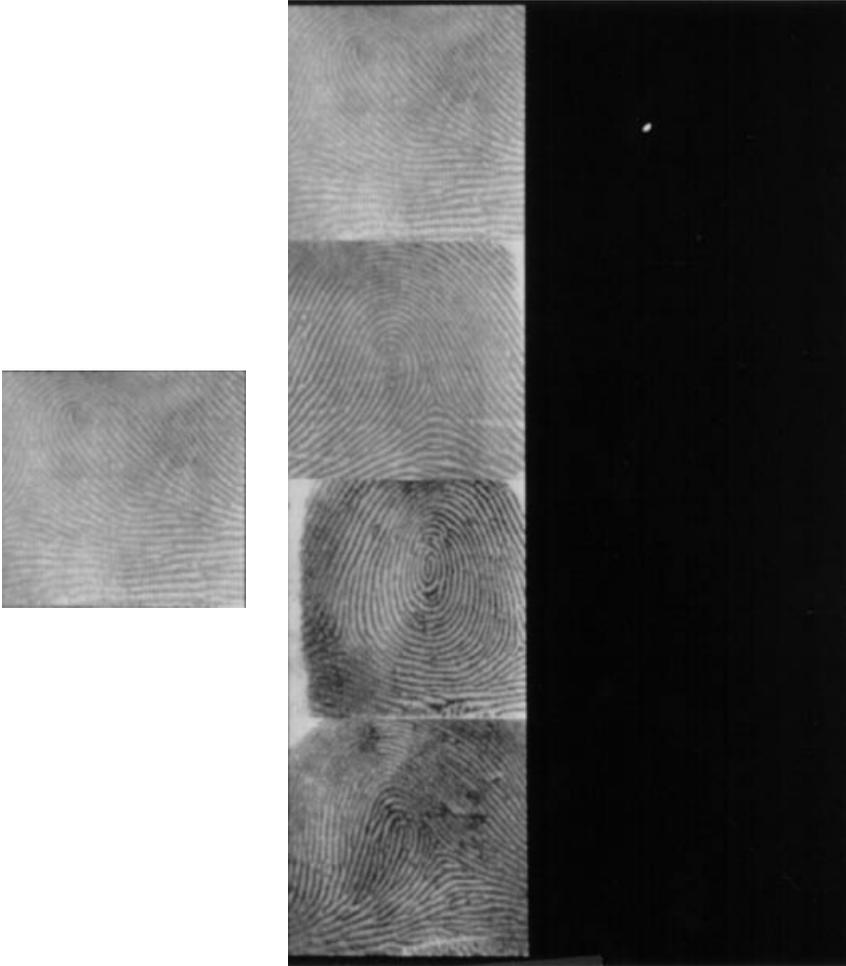


Figure 1.36 Fingerprint interrogator. Correlation of fingerprints. *Left:* Reference fingerprint. *Middle:* Sample fingerprints. *Right:* Correlation peak. Only the sample fingerprint that matches the reference fingerprint generates a correlation peak. (Courtesy of A. Bergeron, J. Gauvin, and INO.)

Figure 1.36 shows the results when the VLC is applied in indentifying a specific fingerprint from multiple samples. The similarity to the encoded fingerprint is indicated by the brightness of the cross-correlation of $E_3(x_3, y_3)$.

1.7.2.3 Joint Transform Correlator

The joint transform correlator (JTC) was first proposed by Weaver and Goodman in 1966 [13,14,15]. A schematic diagram of the joint transform correlator is shown in Fig. 1.37. The difference between the VLC and JTC is the arrangement of the input. In a sense, the former is arranged in series and the latter in parallel. With the VLC, the Fourier transform H is projected onto the prefabricated G^* filter to generate HG^* , while with the JTC, the input images are put side by side in plane P_1 to generate G^*H . The JTC is based on the principle of the lateral shift invariance of the Fourier transform mentioned in Section 1.5.6.

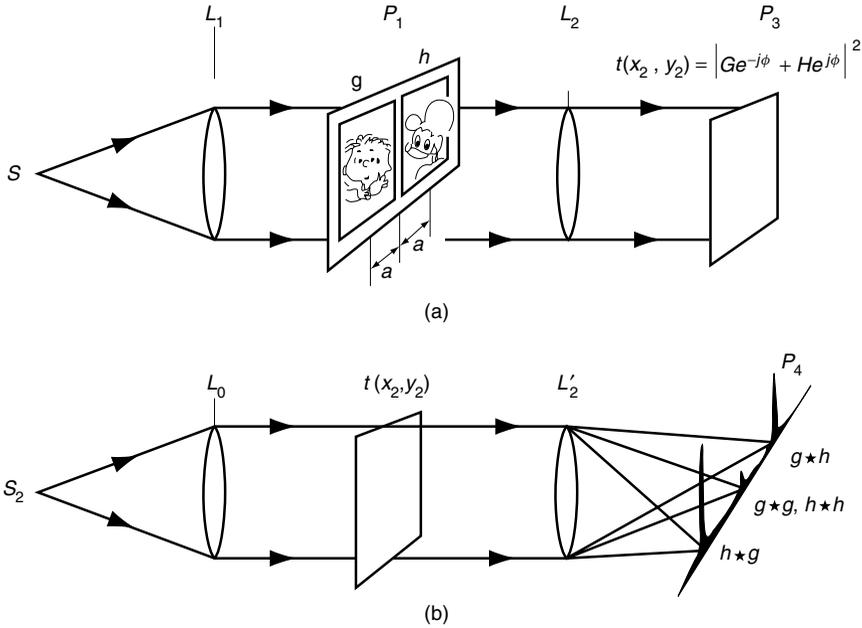


Figure 1.37 Joint transform correlator (JTC). (a) Step 1: Fabrication of the film $t(x_2, y_2)$. (b) Step 2: Display of correlation.

The JTC procedure is a two-step operation. The first step is shown in Fig. 1.37a and is the operation of the square detection of the Fourier transforms of the input images. The second step in Fig. 1.37b is the illumination of the detected Fourier transforms to display the cross-correlation.

In Fig. 1.37a, the reference image is placed a distance a to the left of the center of the P_1 plane, and the input image is placed a distance a to the right of the center. Before considering the general case of an arbitrary input image h , we can gain some useful insight into the JTC by examining the simple case of $g = h$. When $g = h$ in Fig. 1.37a, the input consists of two identical g images located side by side. The Fourier transform of such a pair, which appears in the output plane P_3 , is

$$\begin{aligned} \mathcal{F}\{g(x - a) + g(x + a)\} &= G(f)(e^{-j2\pi af} + e^{j2\pi af}) \\ &= 2G(f) \cos 2\pi af \end{aligned}$$

Thus, a sinusoidal striation appears in the recorded film pattern. It is this sinusoidal striation that plays a key role in the JTC.

The developed film is now inserted in the front focal plane of lens L'_2 in Fig. 1.37b for interrogation. The Fourier transform of the sinusoidal striation generates two peaks in the output plane P_4 (see Problem 1.3), indicating $g = h$.

If, however, $g \neq h$, the sinusoidal striation is absent from the film, and no peaks appear in the output plane P_4 , indicating $g \neq h$. A more precise explanation follows.

The total input to the system is

$$g(x, -a, y_1) + h(x_1 + a, y_1) \quad (1.231)$$

The transmittance of the exposed and then developed film is

$$t(x_2, y_2) = \frac{1}{(\lambda f)^2} \left| G \left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f} \right) e^{-j2\pi(a/\lambda f)x_2} + H \left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f} \right) e^{j2\pi(a/\lambda f)x_2} \right|^2 \quad (1.232)$$

where the transmittance of the film was assumed to be linearly proportional to the square of the incident field. Equation (1.232) is expanded as

$$\begin{aligned} t(x_2, y_2) = & \left(\frac{1}{\lambda f} \right)^2 \left[\left| G \left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f} \right) \right|^2 + \left| H \left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f} \right) \right|^2 \right. \\ & + G^* \left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f} \right) H \left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f} \right) e^{j4\pi(a/\lambda f)x_2} \\ & \left. + H^* \left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f} \right) G \left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f} \right) e^{-j4\pi(a/\lambda f)x_2} \right] \quad (1.233) \end{aligned}$$

Thus, the square detection generates a G^*H term.

The second step is the generation of the correlation peaks. As shown in Fig. 1.37b, the film is placed in the front focal plane of lens L_2 and is illuminated by a parallel beam with amplitude A . The field in the back focal plane of L_2 is

$$E(x_3, y_3) = \frac{A}{j\lambda f} \mathcal{F}\{t(x_2, y_2)\}_{f_x=x_3/\lambda f, f_y=y_3/\lambda f} \quad (1.234)$$

Equation (1.234) can be calculated directly by comparing Eq. (1.233) with the VLC results. The first two terms of Eq. (1.233) are compared to Eq. (1.219), and the last two terms with Eq. (1.222). The final result is

$$\begin{aligned} E(x_3, y_3) = & \frac{A}{j\lambda f} [g(x_3, y_3) \star g(x_3, y_3) \\ & + h(x_3, y_3) \star h(x_3, y_3) \\ & + g(x_3, y_3) \star h(x_3, y_3) * \delta(x_3 - 2a, y_3) \\ & + h(x_3, y_3) \star g(x_3, y_3) * \delta(x_3 + 2a, y_3)]^* \quad (1.235) \end{aligned}$$

The last two terms of Eq. (1.235) are the cross-correlation terms appearing at $(2a, 0)$ and $(-2a, 0)$. If g and h are pure real, the two peaks are of identical shape.

When $g = h$, both curves are not only identical but their intensities peak up, indicating a match.

The complex conjugate signs appearing in Eqs. (1.221), (1.226), and (1.235) disappear if the (x_3, y_3) coordinates are further transformed as $x_3 \rightarrow -x_3$, $y_3 \rightarrow -y_3$, namely, rotating the coordinates of (x_3, y_3) by 180° in its plane.

1.7.2.4 Comparison Between VLC and JTC

VLC and JTC are compared as follows [16,17]:

1. While VLC needs a prefabricated reference filter, JTC does not.
2. After the reference filter has been made, the VLC can interrogate the input in one step. If the same reference image is used, countless interrogations can be made without changing the filter. The JTC requires a two-step operation. For

every interrogation, a Fourier transform filter has to be made, which then has to be illuminated for the correlation peak.

3. VLC demands a very precise lateral alignment of the filter. Even though the lateral location of the input image h in the P_1 plane is arbitrary in Fig. 1.35a, its Fourier transform H always appears at the same location in the P_2 plane. The location of the filter G^* has to match precisely with this location of H . In fact, it has to match within microns. In short, the location of h is arbitrary but that of G^* has to be very precise. JTC does not demand this precision.
4. The required diameter of the JTC lens is twice that of the VLC lens.
5. A higher signal-to-noise ratio (S/N) is obtainable with VLC because a larger separation from the zero order terms is possible by increasing θ , whereas with JTC, the separation from the zero order term is limited by a and hence by the size of the input lens.

1.7.3 Rotation and Scaling

Even though the JTC is impervious to lateral misalignment, its sensitivity is significantly reduced if the input and reference images are rotated with respect to one another or there is a difference in their sizes.

First, let us consider a rotation countermeasure [18]. This countermeasure is explained by way of example in Fig. 1.38.

The points on the rectangle are replotted in polar coordinates with $\rho(\theta)$ as the vertical axis and θ as the horizontal axis. To obtain the polar graph, the transformation

$$\begin{aligned}\rho(\theta) &= \sqrt{x^2 + y^2} \\ \theta &= \tan^{-1}\left(\frac{y}{x}\right)\end{aligned}\tag{1.236}$$

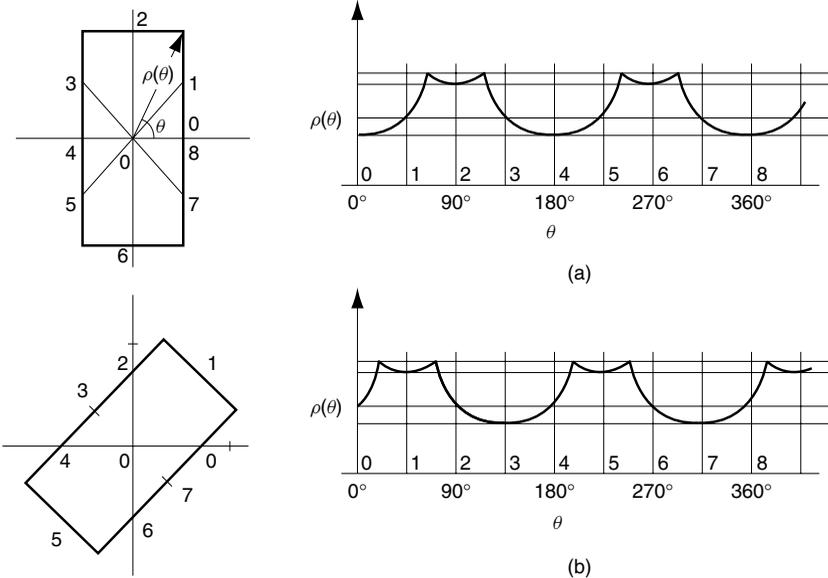


Figure 1.38 Polar coordinate representation of a rectangle. (a) Before rotation of the input. (b) After rotation.

was performed. Figure 1.38a shows the polar graph before the rotation of the rectangle and Fig. 1.38b, after the rotation.

It is clear that the shape of the polar graphs are the same regardless of the rotation of the input image. The only difference is a shift of the curve in the horizontal direction. If the polar coordinate image is used as the input, the property of the lateral shift invariance of the Fourier transform can be utilized.

The shape of the transformed result, however, depends on the choice of the center of rotation. This polar graph method should be used in the Fourier transform domain rather than in the input image domain because the location of the Fourier transformed image is independent of the location of the input in the input plane.

Next, let's look at a countermeasure for differences in scale [19]. Figure 1.39 shows three similar triangles whose heights are in the ratio of 1:5:10. Figure 1.40 shows these same triangles replotted on a logarithmic graph. The triangular shape is distorted but all three are congruent. The locations of the logarithmic images are shifted to the right according to the size of the input image. The logarithmic images can be used as the input for the VLC or JTC processor when the scales of the inputs are different.

Logarithmic scaling for the VLC will be explained using mathematical expressions. The transformation from the (x, y) plane to the (ξ, η) log-log plane is

$$\xi = \log x, \quad \eta = \log y \quad (1.237)$$

The reference image $g(x, y)$ in the xy plane is transformed to $m(\xi, \eta)$ in the log-log plane.

$$g(x, y) \Rightarrow m(\xi, \eta) \quad (1.238)$$

The image $m(\xi, \eta)$ is used as the input to the VLC in the P_1 plane in Fig. 1.35b in order to fabricate the filter. In the P_2 plane, this quantity is Fourier transformed to $M(u)$. For simplicity, the one-dimensional, rather than the two-dimensional, Fourier transform will be used.

$$M(u) = \int_0^{\infty} m(\xi) e^{-j2\pi u \xi} d\xi \quad (1.239)$$

Taking the derivative of Eq. (1.237) gives

$$\frac{d\xi}{dx} = \frac{1}{x}$$

With the help of this derivative, $M(u)$ is expressed in terms of x as

$$M(u) = \int_0^{\infty} f(x) e^{-j2\pi u \log x} \frac{dx}{x} \quad (1.240)$$

Let's rewrite the exponential by putting

$$Y = e^{-j2\pi u \log x} \quad (1.241)$$

Taking the log of both sides of Eq. (1.241) gives

$$\log Y = -j2\pi u \log x \log e$$

$$\log Y = \log x^{-j2\pi u \log e}$$

and hence,

$$Y = x^{-j2\pi u \log e} \quad (1.242)$$

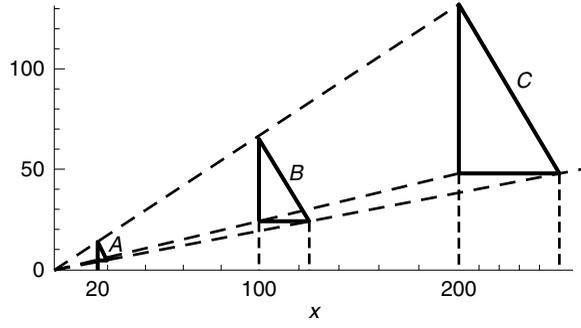


Figure 1.39 Three similar triangles with ratio 1:5:10.

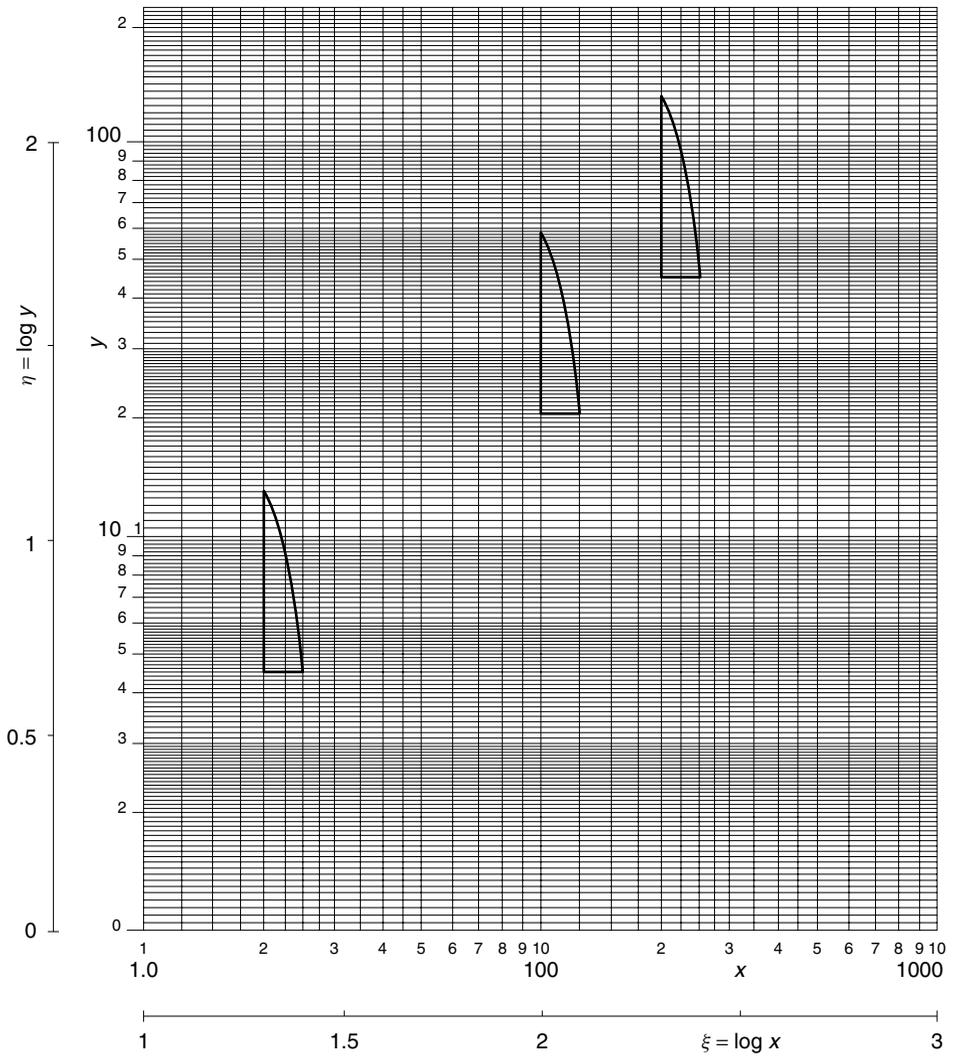


Figure 1.40 The three similar triangles in Fig. 1.39 are replotted on log-log graph paper.

Inserting Eq. (1.242) back into (1.240) gives

$$M(u) = \int_0^{\infty} f(x)x^{-j2\pi u \log_{10} e^{-1}} dx \quad (1.243)$$

If the same is repeated using a \ln - \ln graph instead of \log - \log graph, Eq. (1.243) becomes

$$M(u) = \int_0^{\infty} f(x)x^{-j2\pi u-1} dx \quad (1.244)$$

where $f(x)$ is defined in $0 \leq x < \infty$. This integral is known as the Mellin transform; namely, the Fourier transform of the logarithm of the input function is the Mellin transform of the input function. The VLC filter $M^*(u)$ is made out of $M(u)$ according to the method described in Section 1.7.2.2.

Next, the input image h to be interrogated has to be \ln - \ln transformed in a similar manner. Let us say the input image h is a times the reference image g , namely,

$$h(x, y) = g\left(\frac{x}{a}, \frac{y}{a}\right) \quad (1.245)$$

The \ln - \ln transform of Eq. (1.245), $m_a(\xi, \eta)$, is put into the P_1 plane in Fig. 1.35a. The x component of the output in the P_2 plane is

$$M_a(u) = \mathcal{M}\left\{g\left(\frac{x}{a}\right)\right\} \quad (1.246)$$

where $\mathcal{M}\{ \}$ represents the operation of the Mellin transform.

From Eq. (1.244), $M_a(u)$ is

$$M_a(u) = \int_0^{\infty} g\left(\frac{x}{a}\right)x^{-j2\pi u-1} dx \quad (1.247)$$

Putting $x/a = X$ gives

$$M_a(u) = a^{-j2\pi u} \int_0^{\infty} f(X)X^{-j2\pi u-1} dX \quad (1.248)$$

Note the similarity between Eqs. (1.244) and (1.248). In order to rewrite the first factor of Eq. (1.248) as a power of e , let

$$\begin{aligned} Y &= a^{-j2\pi u} \\ \ln Y &= -j2\pi u \ln a \\ Y &= e^{-j2\pi u \ln a} \end{aligned} \quad (1.249)$$

Putting Eq. (1.249) into (1.248) and comparing with Eq. (1.244) gives the final result:

$$M_a(u) = e^{-j2\pi u \ln a} M(u) \quad (1.250)$$

The enlargement of the input generates an additional phase shift.

$M_a(u)$ is the pattern projected to P_2 in Fig. 1.35a, where the Fourier transform $M^*(u)$ of the reference image has already been placed. Thus, the input to lens L_3 in Fig. 1.35a is

$$M^*(u)M_a(u) = e^{-j2\pi u \ln a} |M(u)|^2 \quad (1.251)$$

where λf was assumed unity for simplicity.

The output from lens L_3 is now proportional to

$$[m(x_3) \star m(x_3)]^* \delta(x_3 + \ln a) \quad (1.252)$$

Thus, the magnitude of the correlation peak is always

$$[m(x_3) \star m(x_3)]^*$$

regardless of the enlargement factor a . Only the location of the peak shifts in accordance with the enlargement a .

1.7.4 Real-Time Correlation

Correlators have many practical applications [13,15]. They are used as robotic eyes in automatic assembly lines, and as security devices for checking biometric indicators such as fingerprints, facial images, voice, and DNA. For applications such as these, correlators with real-time response are crucial. The bottleneck for the real-time operation of either the VLC or JTC is the recording of G and H by means of photographic film, which acts as a square detector to produce G^*H . The photographic film can be replaced either by a photorefractive (PR) crystal whose index of refraction is changed by light intensity (see Section 5.6) or simply by using a CCD camera with the camera lens removed.

Figure 1.41 shows an example of the real-time operation of the JTC using a photorefractive crystal. The pattern of $|G + H|^2$ is generated by laser light S_1 from the left. This pattern is recorded by the photorefractive crystal in the P_2 plane. The recorded pattern is then read by laser light S_2 from the right. S_1 and S_2 have different wavelengths. The read image is projected into a CCD camera whose camera lens has been removed. As far as the input method is concerned in this particular example, the input scene h has been taken by another CCD camera whose camera lens is intact. The input scene is displayed on the spatial light modulator (SLM) located in the input plane P_1 . The SLM is a liquid crystal display panel, which is described in Section 5.10.4.4. One of the advantages of the SLM is its optically flat display surface. The surface

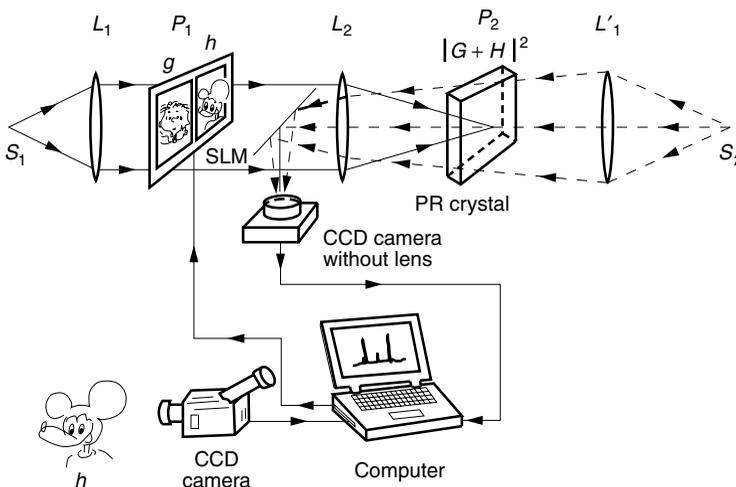


Figure 1.41 Real-time JTC using a photorefractive (PR) crystal.

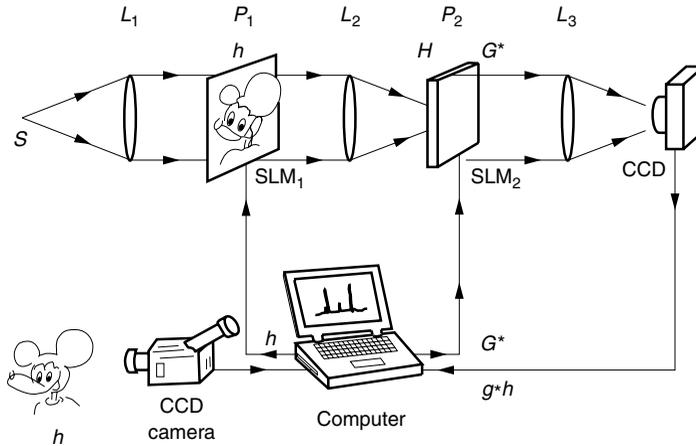


Figure 1.42 Real-time VLC using optics and a computer.

condition of the images is an important factor in the performance of optical signal processors that use coherent light.

Optical signal processing provides high speed and large information capacity while computers provide versatility and reliability. The examples shown in Figs. 1.41 and 1.42 make use of both.

Now let's look at the real-time VLC example shown in Fig. 1.42. The input image, taken by the CCD camera whose lens is intact, is fed into the computer, and is displayed on SLM₁ in the input plane P₁. The input image is then optically Fourier transformed and projected over the computer-generated G* displayed on SLM₂. The operation of $\mathcal{F}\{G^*H\}$ is again performed optically and is projected onto the CCD camera whose lens has been removed. The signal from the lensless CCD camera is fed back to the computer. The computer processes the results to arrive at a final decision.

Figure 1.43 shows a similar arrangement but with the JTC. The process of obtaining G*H is quite similar to the previous case. The laser light S, which has been branched off to the bottom of the figure, optically performs the operation of $\mathcal{F}\{G^*H\}$ to provide g * h on the CCD camera whose lens has been removed.

Figure 1.44 shows a system that relies more heavily on the computer. The scene captured by the lensed CCD camera is displayed on the SLM. The optical system Fourier transforms (g + h) to give |G + H|² on the lensless CCD camera. The computer takes over the rest of the processing including the operation of $\mathcal{F}\{G^*H\}$ as well as the decision on the result of the interrogation.

1.7.5 Cryptograph

Another special application of the spatial filter is the cryptograph [20,21] (meaning encoding) of an image for security purposes. Optical signal processing needs to provide fast and reliable identification of people and verification of their signatures on a document.

First, the method of encryption will be described referring to Fig. 1.45a. Let the input image to be encrypted be E₀(x₀, y₀). The Fourier transform of E₀(x₀, y₀) is projected onto the back focal plane of L₂. The key card, onto which a white sequence noise pattern n(x, y) is imprinted, is also placed in the back focal plane of L₂. The

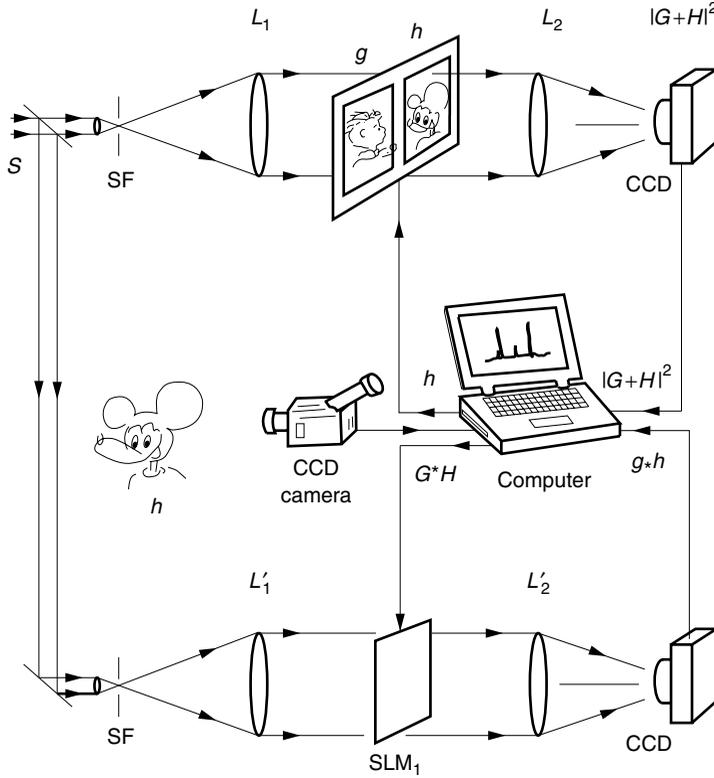


Figure 1.43 Real-time JTC using optics and a computer.

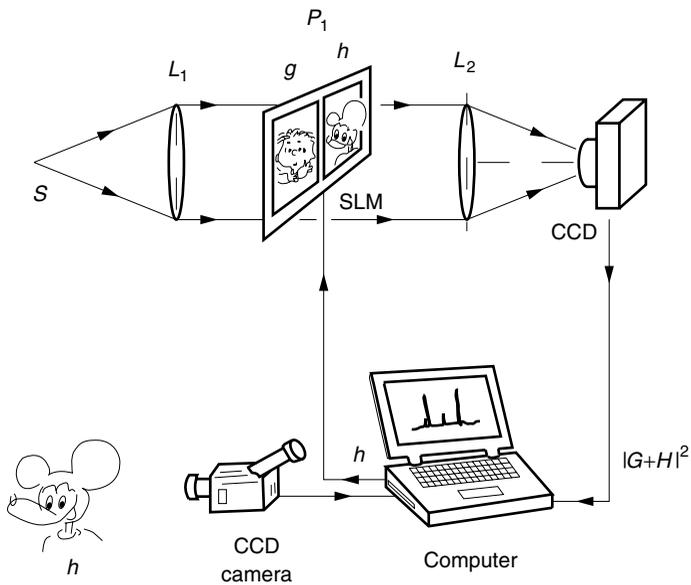


Figure 1.44 Real-time JTC that heavily relies on the use of a computer.

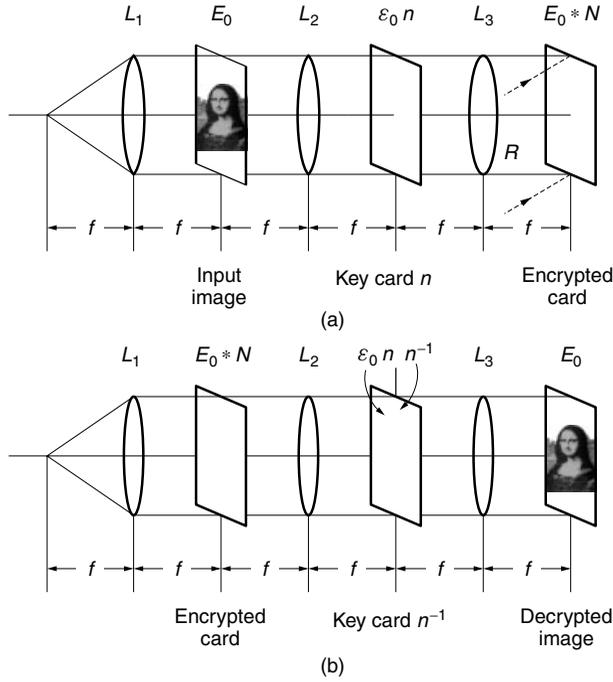


Figure 1.45 Cryptographic spatial filters. (a) Encryption. (b) Decryption.

input to the second Fourier transform lens L_3 becomes

$$\mathcal{F}\{E_0(x, y)\}n(x, y) \tag{1.253}$$

In the back focal plane of L_3 , the image given by Eq. (1.253) is further Fourier transformed to

$$E(x_i, y_i) = E_0(-x_i, -y_i) * N(x_i, y_i) \tag{1.254}$$

where $N(x_i, y_i)$ is the Fourier transform of the white noise sequence. The output image represented by Eq. (1.254) does not resemble the original input image $E_0(x_0, y_0)$. The original input image can be recovered only when the key card is used. In order to photographically record $E(x_i, y_i)$ including the phase onto a card, a reference plane wave $R(x_i, y_i)$ (not shown in the figure) is added at the time of recording. Thus, the encrypted card has the intensity pattern $I(x_i, y_i)$:

$$\begin{aligned} I(x_i, y_i) &= |R(x_i, y_i) + E_0(-x_i, -y_i) * N(x_i, y_i)|^2 \\ &= |R(x_i, y_i)|^2 + |E_0(-x_i, -y_i) * N(x_i, y_i)|^2 \\ &\quad + R^*(x_i, y_i)[E_0(-x_i, -y_i) * N(x_i, y_i)] \\ &\quad + R(x_i, y_i)[E_0(-x_i, -y_i) * N(x_i, y_i)]^* \end{aligned} \tag{1.255}$$

Next, the method of decrypting the original image is explained referring to Fig. 1.45b. The encrypted pattern $I(x_i, y_i)$ is placed in the front focal plane of lens L_2 , and all four terms in Eq. (1.255) are Fourier transformed in the back focal plane.

Only the Fourier transform $E_3(x, y)$ of the third term in Eq. (1.255) is of prime concern,

$$E_3(x, y) = R_0 \mathcal{F}\{E_0(-x, -y)\} \cdot n(-x, -y) \quad (1.256)$$

where for simplicity of expression, the direction of incidence of the plane reference wave was assumed normal to the card surface in order to make $R(x_i, y_i)$ a constant R_0 across the surface of the card. If a key card imprinted with the pattern of $n^{-1}(-x, -y)$ is placed in the back focal plane of L_2 , the second factor in Eq. (1.256) is canceled. The input to the Fourier transform lens L_3 becomes $R_0 \mathcal{F}\{E_0(-x, -y)\}$ and in the back focal plane of L_3 , the original input image is recovered. Recovery of the original image is only possible for a key card imprinted with $n^{-1}(x, y)$.

A method for fabricating a key card that can be used for both encrypting and decrypting is described as follows. A pseudorandom pattern of 0s and 1s is written onto a half-wavelength thick optical film on a substrate. The light passing through the “0” location experiences no phase shift while that passing through the “1” experiences a π -radian phase shift. The transmission pattern of the key card is then

$$n(x, y) = e^{j\pi b(x, y)} \quad (1.257)$$

where $b(x, y)$ is the pseudorandom pattern. Note that such a pattern satisfies $n^{-1}(x, y) = n(x, y)$ and can be used for both encrypting and decrypting. The pattern $n^{-1}(-x, -y)$ can be obtained by rotating the card by 180° in the plane of the card.

1.8 HOLOGRAPHY

Holography was invented by Dennis Gabor in 1948 when he was trying to improve the quality of electron microscope images. The word *holo* in Greek means complete and *gram* means recording, so that a “hologram” is a complete recording of the wave scattered from an object. Holography uses both phase and amplitude distributions of the scattered light to record the image of the object [1,6,8].

Both conventional photography and holography utilize light-sensitive film as the recording medium. In both cases, the film records light intensity. In conventional photography, the camera’s lens generates an image of an object in the film plane, and the film records the image’s intensity pattern. In a hologram, the film is directly exposed to the light scattered by an object. By itself, recording the scattered wave’s intensity is not sufficient to make a hologram, as phase information would be lost. This shortcoming is overcome by illuminating the holographic film with a reference wave, as well as the scattered object wave. The holographic film records the fringe pattern that results from the interference of the reference wave and the scattered object wave. Fringe contours with high intensity indicate the scattered object wave and the reference wave are in phase; likewise, low intensity contours indicate the waves are out of phase. Note that the phase information has been converted into an intensity pattern. The holographic film thus exposed and developed is the hologram.

In order to see the image from the hologram, a laser beam illuminates the hologram. The laser beam is diffracted by the recorded fringe patterns on the hologram. The diffracted light recreates the image of the original object. An observer looking through the hologram toward the laser will see a view just as if the real, original object were present behind the hologram.

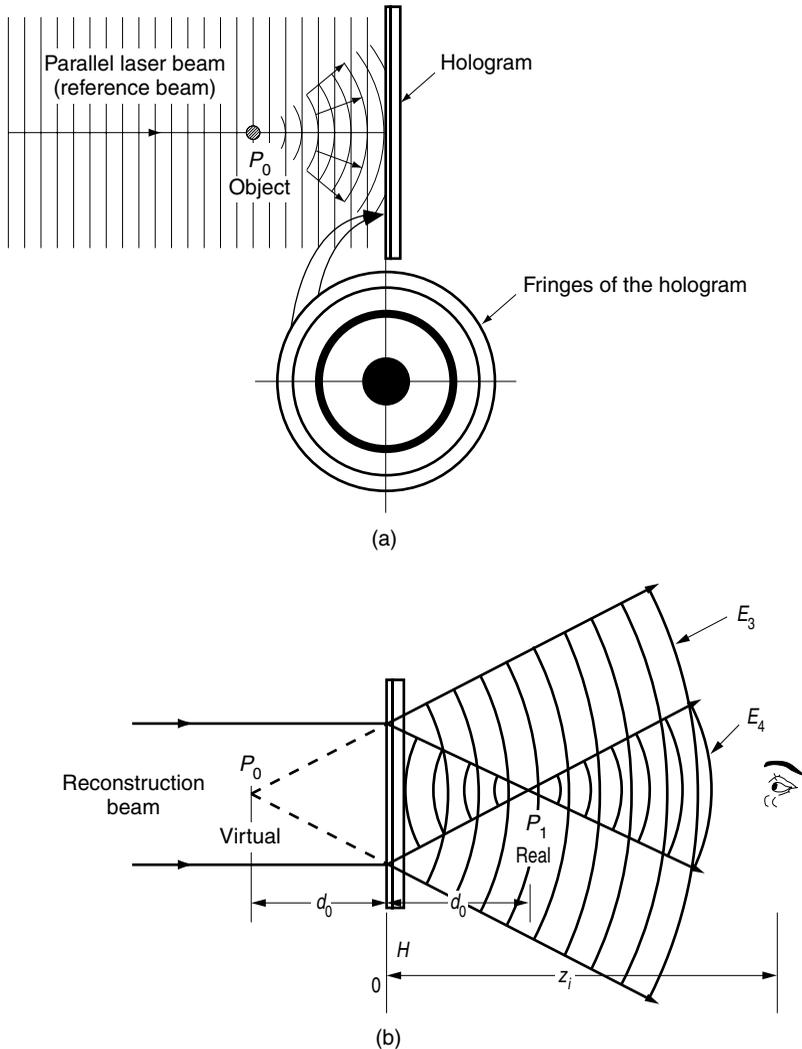


Figure 1.46 Illustration of the Gabor-type hologram. (a) Fabrication of the hologram (point object). (b) Reconstruction of the image.

1.8.1 Gabor-Type Hologram

Figure 1.46a shows an arrangement for fabricating a Gabor-type hologram. A parallel laser beam illuminates both the object and the photographic film. A special feature of the Gabor-type hologram is that only one beam illuminates both the object and the film. In contrast, the Leith–Upatnieks type hologram uses two beams: one for illuminating the object and the other for illuminating the film.

Let's take a closer look at the geometry of the Gabor-type hologram in Fig. 1.46a. For simplicity, the object at P_0 is a point object. The reference beam R is incident along the z axis, which is perpendicular to the film. The object beam is the wave scattered from the object. The film is simultaneously illuminated by both the reference beam and the object beam. The film records the interference pattern between the reference

and object beams. For the recording of the interference pattern to be successful, it is necessary to have a mechanically stable setup, a spectrally pure source, and a high-resolution film. After being exposed, the film is developed, and the result is a Gabor-type hologram.

The human eye is not able to decipher the interference pattern of the hologram directly. Instead, the image has to be recreated by a process called reconstruction.

Let's assume that a laser was used to provide the spectrally pure source for creating the hologram, and that a similar laser with the same wavelength is used in the reconstruction process. Figure 1.46b shows the arrangement for reconstructing the image from the hologram. The hologram is illuminated from the back by the reconstruction laser beam. The light scattered from the interference pattern on the hologram forms the image of the original point object.

A more quantitative description is in order. Let's say the film is in the (x, y) plane at a distance d_0 from the object. The field scattered by the point object is a spherical wave centered at the object. The scattered field observed on the film is

$$O(x, y) = \frac{A}{j\lambda} \frac{e^{jkr}}{r} \quad (1.258)$$

where

$$r = \sqrt{d_0^2 + x^2 + y^2}$$

If d_0 is much larger than the size of the photographic film, Eq. (1.258) can be approximated as in Eq. (1.31)

$$O(x, y) = \frac{A}{j\lambda d_0} e^{jk[d_0 + (x^2 + y^2)/2d_0]} \quad (1.259)$$

The reference beam is a plane wave propagating in the z direction and is expressed by

$$R = R_0 e^{jkz} \quad (1.260)$$

where R_0 is a real number.

The photographic film is now exposed to the interference pattern of the object and reference beams. The developed film looks like a Fresnel zone plate and is composed of a set of concentric rings as shown below the hologram in Fig. 1.46a. The expression $t(x, y)$ for the transmission coefficient of the film is

$$\begin{aligned} t(x, y) &= t_0 - \beta |R_0 + O(x, y)|^2 \\ &= t_0 - \beta [R_0^2 + |O(x, y)|^2 + R_0 O(x, y) + R_0 O^*(x, y)] \end{aligned} \quad (1.261)$$

We are now ready to reconstruct the image from the hologram by illuminating it with the reconstruction beam. The Fresnel diffraction pattern forms the reconstructed image. In order to use the Fresnel diffraction formula Eq. (1.38), we need to find the integrand $E(x, y)$, which is the light distribution that has just passed through the hologram. For simplicity, the reconstruction beam is assumed to be the same as the reference beam used at the time of the hologram's fabrication. From Eqs. (1.260) and (1.261), the expression for $E(x, y)$ is

$$E(x, y) = R_0^2 + |O(x, y)|^2 R_0 + R_0^2 O(x, y) + R_0^2 O^*(x, y) \quad (1.262)$$

where t_0 and β were suppressed. The hologram is placed at $z = 0$ and the diffraction pattern is observed in the (x_i, y_i) plane at $z = z_i$.

The first two terms of Eq. (1.262) do not have much of a spatial variation, which means that this portion of the reconstruction beam passes straight through along the z axis with some attenuation. The Fresnel diffraction pattern $E_3(x_i, y_i)$ associated with the third term of Eq. (1.262) is from Eqs. (1.38), (1.259), and (1.260):

$$E_3(x_i, y_i) = AR_0^2 \frac{e^{jk[z_i+d_0+(x_i^2+y_i^2)/2z_i]}}{j\lambda z_i} \frac{1}{j\lambda d_0} \mathcal{F}\{e^{jk(x^2+y^2)/2d_0} e^{jk(x^2+y^2)/2z_i}\}_{f_x=x_i/\lambda z_i, f_y=y_i/\lambda z_i} \quad (1.263)$$

Equation (1.263) can be rewritten as

$$E_3(x_i, y_i) = AR_0^2 \frac{e^{jk[z_i+d_0+(x_i^2+y_i^2)/2z_i]}}{j\lambda z_i} \frac{1}{j\lambda d_0} \mathcal{F}\{e^{jk(x^2+y^2)/2D}\}_{f_x=x_i/\lambda z_i, f_y=y_i/\lambda z_i} \quad (1.264)$$

where

$$\frac{1}{D} = \frac{1}{z_i} + \frac{1}{d_0}$$

Using the Fourier transform relationship Eq. (1.44) gives

$$E_3(x_i, y_i) = AR_0^2 \frac{e^{jk[z_i+d_0+(x_i^2+y_i^2)/2z_i]}}{j\lambda z_i j\lambda d_0} \cdot j\lambda D \{e^{-j\pi\lambda D(f_x^2+f_y^2)}\}_{f_x=x_i/\lambda z_i, f_y=y_i/\lambda z_i} \quad (1.265)$$

$$E_3(x_i, y_i) = AR_0^2 \frac{e^{jk(z_i+d_0)}}{j\lambda(z_i+d_0)} \cdot e^{jk(x_i^2+y_i^2)/2(z_i+d_0)} \quad (1.266)$$

Equation (1.266) is the expression for a diverging spherical wave that would be established if a point source were located at a distance of $z_i + d_0$ from the observer. Referring to Fig. 1.46b, this location is P_0 and is exactly where the point object had been placed. No light rays, however, converge to this point so that the image is a virtual image of the point object.

Next, the diffraction pattern due to the fourth term is obtained in a similar manner and the field observed is

$$E_4(x_i, y_i) = AR_0^2 \frac{e^{jk(z_i-d_0)}}{j\lambda(z_i-d_0)} e^{jk(x_i^2+y_i^2)/2(z_i-d_0)} \quad (1.267)$$

Equation (1.267) is the expression of a spherical wave that would be established if a point source were located at a distance of $z_i - d_0$ from the observer. Referring to Fig. 1.46b, this location is P_1 and is symmetric to P_0 with respect to the hologram. The spherical wave is convergent first in the region $z_i < d_0$ and actually converges at $z_i = d_0$; then it is divergent again in the region of $z_i > d_0$. If a sheet of paper is placed at $z_i = d_0$, a bright spot is observed. This bright spot is the real image of the point object.

The object was assumed to be a point object, but the analysis can be extended to a more complex object. The complex object can be considered as an ensemble

of individual points; thus, the virtual image of the original object is observed at the original object location. This image is called the orthoscopic image.

The real image is in a plane that is symmetric to the object with respect to the hologram. This image is called the pseudoscopic image of the hologram and is explained further in Section 1.8.3. The observer can view this image either by inserting a piece of paper or by positioning his/her eyes at a distance $z_i > d_0$.

A major disadvantage of the Gabor-type hologram is that the wavefronts overlap. With the geometry shown in Fig. 1.46b, the observer sees three wavefronts at the same time; one from the undiffracted portion of the reconstructing beam that passes straight through the hologram, one from the extension of the virtual image P_0 , and one from the real image P_1 . When the observer tries to focus his/her eyes on P_0 , the reconstruction beam and the blurred image of P_1 are in the background and the quality of the reconstructed image of the Gabor-type hologram is not high.

1.8.2 Off-Axis Hologram

The disadvantage of the Gabor-type hologram was overcome by Leith and Upatnieks in 1962. They proposed a scheme for slanting the direction of incidence of the reference beam in order to spatially separate the locations of the reconstructed images. The quality of the images were substantially improved.

Figure 1.47a shows an off-axis reference hologram. The reference beam is incident on the hologram at an angle θ with respect to the normal to the hologram and is expressed by

$$R = R_0 e^{jkx \sin \theta + jkz \cos \theta} \quad (1.268)$$

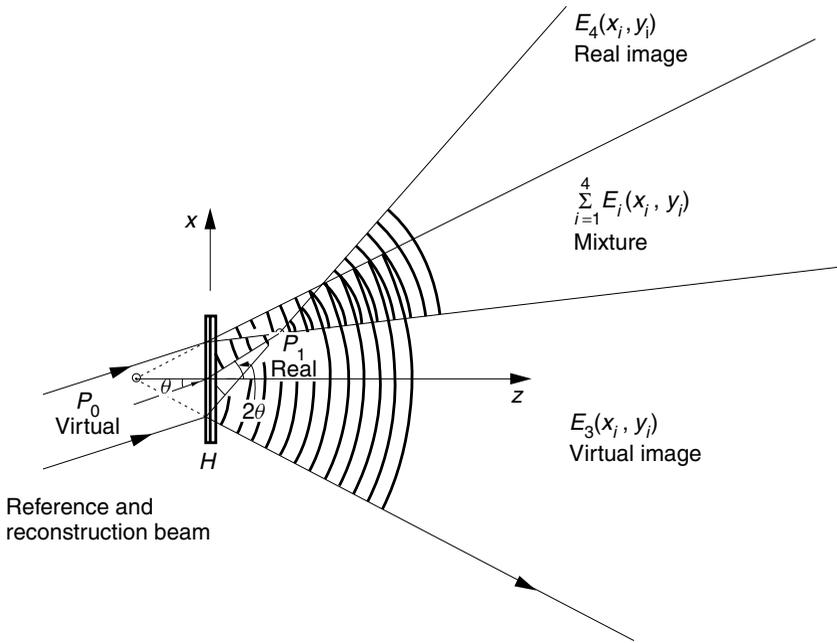
where R_0 is a real number, again. The same light beam will be used as the reconstruction beam. The change to the slanted reference beam from the straight-on reference beam needs only a minor modification to the previous results obtained with the Gabor-type hologram in Section 1.8.1. The transmission coefficient of the hologram at $z = 0$ is

$$\begin{aligned} t(x, y) &= t_0 - \beta |R_0 e^{jkx \sin \theta} + O(x, y)|^2 \\ &= t_0 - \beta [R_0^2 + |O(x, y)|^2 + R_0 e^{-jkx \sin \theta} O(x, y) + R_0 e^{jkx \sin \theta} O^*(x, y)] \end{aligned} \quad (1.269)$$

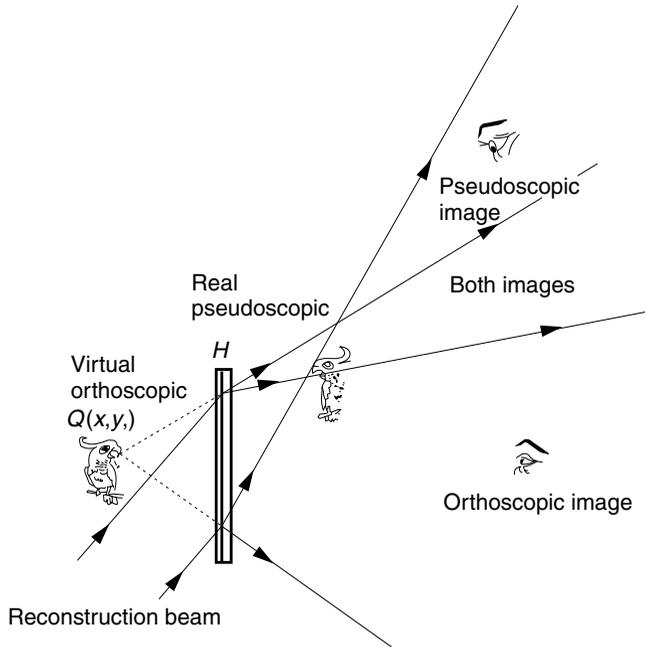
In the process of reconstructing the holographic image, the hologram is illuminated with a reconstruction beam that is the same as Eq. (1.268). The input pattern $E(x, y)$ to the Fresnel diffraction formula is obtained by the multiplication of $t(x, y)$ and R .

$$E(x, y) = R_0^3 e^{j2kx \sin \theta} + |O(x, y)|^2 R_0 e^{jkx \sin \theta} + R_0^2 O(x, y) + R_0^2 e^{j2kx \sin \theta} O^*(x, y) \quad (1.270)$$

The amplitude of the first two terms has practically no spatial variation; these terms are the parallel beams in the θ direction. The third term is identical to the earlier result in Eq. (1.263). This means that tilting the reference beam does not affect the location of the virtual image (as long as the reference beam is used as a reconstruction beam). The virtual image occupies the same location as the original object, which in Fig. 1.47a is along the z axis. The fourth term can be obtained the same way that Eq. (1.267)



(a)



(b)

Figure 1.47 Off-axis reference beam hologram. (a) Geometry. (b) Regions for observing pseudoscopic and orthoscopic images.

was obtained:

$$E_4(x_i, y_i) = AR_0^2 \frac{e^{jk[z_i - d_0 + (x_i^2 + y_i^2)/2z_i]}}{j\lambda z_i} \frac{1}{j\lambda d_0} \mathcal{F}\{e^{jk(x^2 + y^2)/2D'} \cdot e^{j2kx \sin \theta}\}_{f_x = x_i/\lambda z_i, f_y = y_i/\lambda z_i} \quad (1.271)$$

where

$$\frac{1}{D'} = \frac{1}{z_i} - \frac{1}{d_0} \quad (1.272)$$

The Fourier transform of the second factor in the braces is

$$\mathcal{F}\{e^{j2kx \sin \theta}\} = \delta\left(f_x - \frac{2 \sin \theta}{\lambda}\right) \quad (1.273)$$

Using the convolution relationship and then the delta function property, Eqs. (1.110) and (1.44), the final expression for Eq. (1.271) is

$$E_4(x_i, y_i) = \frac{jAR_0^2}{\lambda(z_i - d_0)} \times \exp\left[jk\left(z_i - d_0 + \frac{(x_i - 2d_0 \sin \theta)^2 + y_i^2 + 4d_0(z_i - d_0) \sin^2 \theta}{2(z_i - d_0)}\right)\right] \quad (1.274)$$

Thus, the real image appears at

$$(x_i, y_i, z_i) = (2d_0 \sin \theta, 0, d_0) \quad (1.275)$$

with some aberration for large θ .

For small θ , Fig. 1.47a summarizes the positions of the images. These images are the virtual image at $\theta = 0$, the undiffracted beam at θ , and the real image at 2θ .

Except for the overlap region, the images are clearly separated and can be observed without interference.

1.8.3 Pseudoscopic Image

When the observer views the real image, his/her eyes will see a peculiar image. Let's suppose the object is a bird, as shown in Fig. 1.47b. The object bird was facing the observer when the hologram was made, but the reconstructed real image of the bird is facing away from the observer. The hologram can record only the side of the object that is facing the hologram. The light scattered from the other side never reaches the hologram and cannot be recorded. Because of this, the observer looking at the real reconstructed image sees the tip of the beak further away than the bird's eyes. The observer has the sensation of looking straight through the back surface and seeing the front surface image from the inside out (like looking at the inside of a mask). Such an inside-out image is called a pseudoscopic image.

1.8.4 Volume Hologram

Important applications such as the high-density recording of images and white light or color holograms are based on the volume hologram. As the thickness of the emulsion

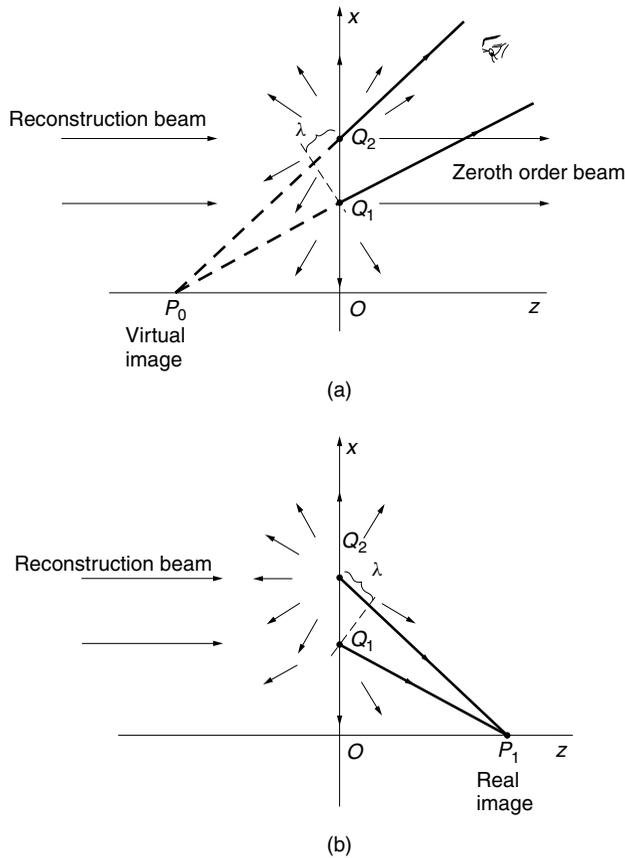


Figure 1.48 Illustration of the reconstruction of the holographic image from a thin emulsion film. (Cross section of the hologram). (a) Reconstruction of the virtual image. (b) Reconstruction of the real image.

of the photographic film is increased, properties that do not exist with thin emulsion holograms begin to surface. Before going into the case of the thick emulsion, the case of the thin emulsion will be reviewed.

The geometry considered is exactly the same as already studied in Fig. 1.46, but Fig. 1.48 is intended to show the finer details of the fringe patterns on the hologram. Points Q_1, Q_2, \dots (actually rings) form an array of scattering centers. The reconstruction of the image can be treated approximately as the problem of obtaining the scattering pattern from a periodic array of point scatterers. As mentioned earlier in Section 1.4.9, the field pattern can be separated into the element pattern, which is the radiation pattern of the individual element, and the array pattern, which is the pattern determined by the spacing between the elements and the overall dimension of the array. The scattering pattern is the product of these two patterns as indicated by Eq. (1.134).

When the emulsion of the film is thin, the size of the individual Q 's are comparable to the light wavelength, and each Q scatters the light in all directions. Hence, the element pattern is omnidirectional when the emulsion is thin.

With the geometry shown in Fig. 1.48a, the path difference between $\overline{P_0Q_1}$ and $\overline{P_0Q_2}$ is exactly one wavelength and the rays scattered from Q_1 and Q_2 enhance each other in the direction of $\overline{P_0Q_2}$. The observer sees this peak of light as the virtual image.

Another peak is observed in the direction of propagation of the reconstruction beam because the light scattered from Q_1 and Q_2 are also in phase in this direction. This peak is associated with the undiffracted or zeroth order beam and is treated as noise.

Furthermore, with the geometry shown in Fig. 1.48b, the path difference between $\overline{Q_1P_1}$ and $\overline{Q_2P_1}$ is exactly one wavelength and the array pattern makes another peak that is associated with the real image. Thus, the array pattern of a thin emulsion hologram has three peaks; namely, in the directions of the virtual image, the zeroth order noise, and the real image. The sharpness of the image is determined by the product of the element and array patterns. In the case of a thin emulsion, the element pattern is omnidirectional and the sharpness of the image is predominantly determined by the array pattern.

Next, the thick emulsion case will be explained. The scattering points Q_1, Q_2, \dots grow into a set of mirror platelets made out of silver grains as shown in Fig. 1.49a. The mirror platelets are formed by the standing wave pattern created by the interference between the reference and object beams. They are oriented in the plane of the bisector of the angle between the reference and object beams. If the reconstruction beam is the same as the reference beam, the orientation of each mirror platelet is such that the reflected beam is directed toward the extension of the object beams $\overline{P_0Q_1}$ and $\overline{P_0Q_2}$, thus forming the virtual image.

Because of the reflective nature of the platelet mirror surfaces, the element pattern has only one peak, which corresponds to the virtual image at P_0 . Light is not reflected toward the real image P_1 , and *no pseudoscopic image is observed from the volume hologram*.

As far as the array pattern is concerned, the spacing between the mirror platelets is the same as for the thin emulsion, and the array pattern is also directed toward the extension of the object beam. The diffraction pattern is the product of the element and array patterns, both of which are sharply directed in the same direction. Thus, a very sharp image is reconstructed from a thick emulsion hologram. This type of hologram is often called a volume hologram.

Bragg reflection is the basis for X-ray analysis of atomic layers. Bragg reflection takes place when the directivity established by the path difference between the rays diffracted from adjacent atomic layers matches up with the direction of the specular reflection from the surface of the atomic layer. With the geometry shown in Fig. 1.49c, the Bragg reflection is in the direction given by

$$2d \sin \theta = m\lambda \quad (1.276)$$

where θ is traditionally taken with respect to the surface rather than the normal. In other words, Bragg reflection takes place when the array pattern of the atomic layer matches the element pattern of the atomic layer. So, we can say that the reconstruction of the image from the volume hologram is based on Bragg diffraction.

A volume hologram is capable of storing a large number of images. If the images are each recorded using a different angle of incidence for the reference beam, then the images can selectively be reconstructed by adjusting the angle of the reconstruction beam. The reason we do not see all the images reconstructed simultaneously is

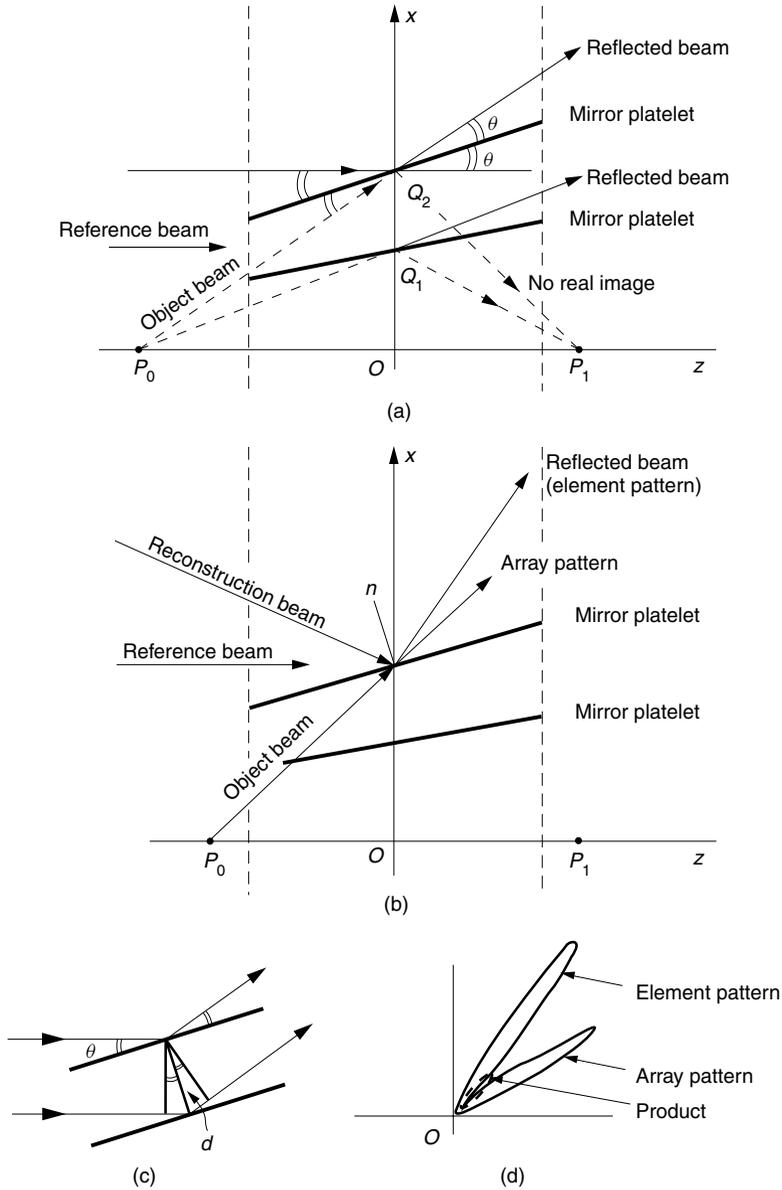


Figure 1.49 The reconstruction of the holographic image from a thick emulsion film. (Cross section of the hologram). (a) Reference and reconstruction beams are identical. (b) Reference and reconstruction beams are not identical. (c) Bragg reflection. (d) Polar diagram of the pattern.

explained as follows. Suppose the reference and reconstruction beams for a given image are not identical, as shown in Fig. 1.49b. With the change in the direction of the reconstruction beam, both the element and the array patterns shift, but they shift differently. Their peaks no longer match, as indicated by the polar diagram in Fig. 1.49c, and the sharp peak disappears from the product of the polar diagram. The image can be reconstructed only when the reconstruction beam is identical to the reference beam. This fact is used for the high-density recording of the volume

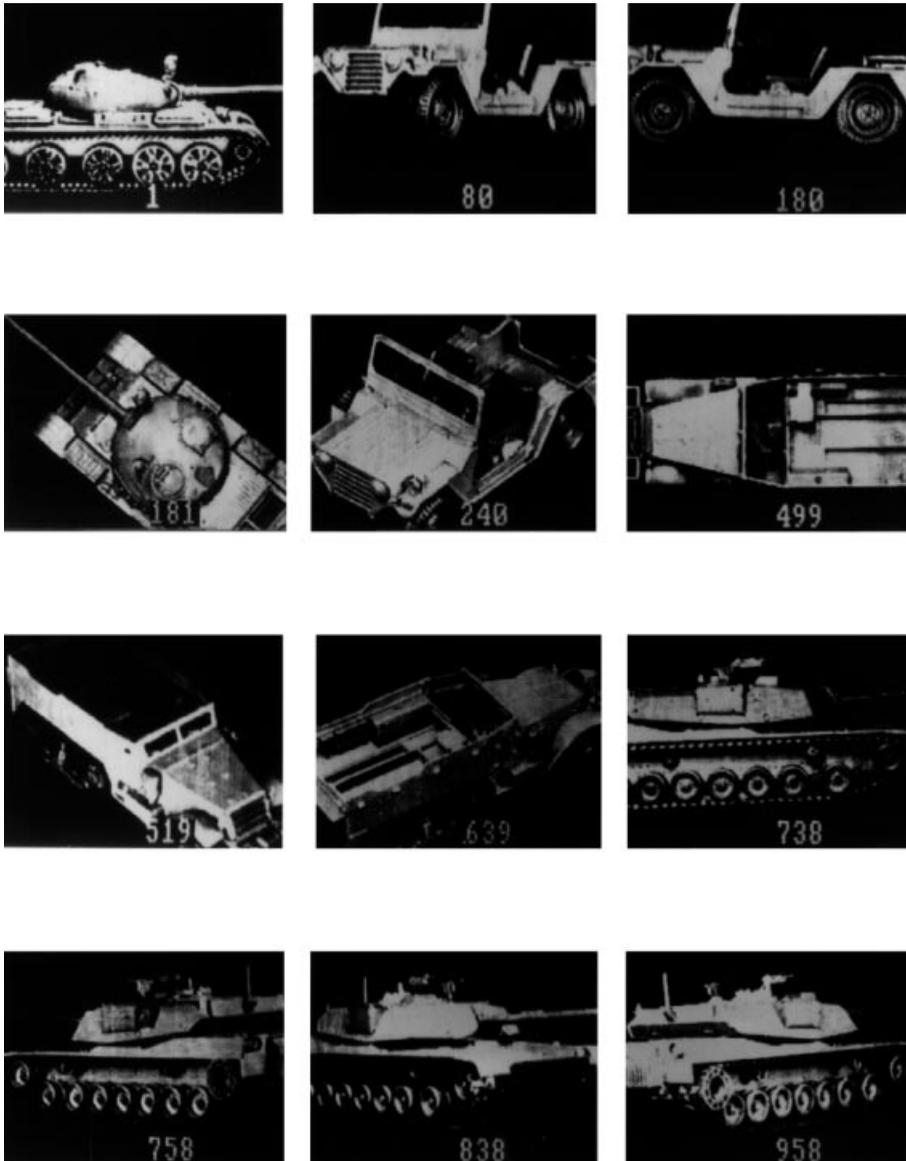


Figure 1.50. Sample reconstructions from 1000 holograms recorded in a lithium niobate crystal. (Courtesy of F. H. Mok [22]).

hologram. Figure 1.50 shows a sample of reconstructions from 1000 holograms recorded in a $2 \times 1.5 \times 1\text{-cm}^3$ photorefractive crystal of lithium niobate, Fe:LiNbO_3 . The direction of the reference beam was stepped at less than 0.01° [22,23].

The principle of the thick emulsion hologram is also applied to the white light hologram. The direction of the peak of the array pattern moves with the wavelength of the reconstruction beam but that of the element pattern more or less remains unchanged. Hence, when white light is used for reconstructing the holographic image, the image is reconstructed only by the light spectrum that creates a match between the peaks of the



Figure 1.51 White light hologram. (Courtesy of Dainippon Printing Co.)

element and array patterns. Thus, the reconstruction process is wavelength selective, and the reconstructed image is of a single color. The color is the same as that of the reference beam used to fabricate the hologram. This kind of white light hologram is called either a Lippmann or Denisyuk hologram. An example is shown in Fig. 1.51.

The wavelength selectivity of the thick emulsion hologram is also the basis of the color hologram. As just mentioned, the thick emulsion hologram reconstructs the image in the color used to fabricate the hologram. To produce a color hologram, the thick emulsion hologram is triple exposed by three kinds of lasers whose wavelengths correspond to basic colors such as red, green, and blue. When the hologram is illuminated with white light, images of the three different colors are reconstructed, and a color image is produced.

1.8.5 Applications of Holography

Holography has gained recognition both as an art form and as a measurement tool. A few of the numerous applications are described next [8,24].

1.8.5.1 Three-Dimensional Displays

When a hologram is illuminated, it causes the wavefront that was originally formed by the object to be reconstructed. It is the lens of the observer's eye that forms an image on the retina from the reconstructed wavefront. Depending on the viewing angle of the observer's left and right eyes relative to the hologram, the left and right eyes intercept different portions of the reconstructed wavefront. Figure 1.52a is a photograph of the reconstructed image taken by a camera set at the position of the left eye of the observer and Fig. 1.52b is a photograph of the same image taken with the camera set at the position of the right eye.

Notice the movement of the chess piece in the foreground with respect to the others: Fig. 1.52a is what the left eye sees and Fig. 1.52b is what the right eye sees. This difference of scenes produces the perception of viewing the original object as if it were present in front of the observer in three dimensions. This is called *binocular parallax*. Even with one eye, the side to side movement of the observer's face provides the observer with a differential movement (the foreground appears to move faster than

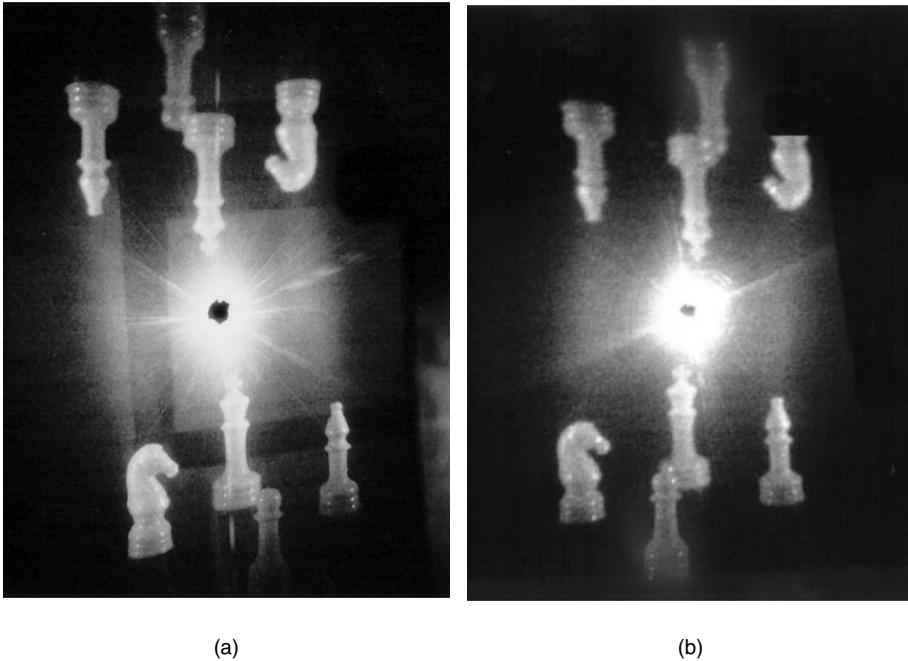


Figure 1.52 The reconstructed image of a hologram varies with the viewing angle. The upside down image is the pseudoscopic image. (a) When viewed from the left. (b) When viewed from the right.

the background), and as the observer's face moves hidden portions of the reconstructed image become visible. These movement effects are usually perceived and contribute to the observer's sensation of seeing the image in three dimensions. This is called *movement parallax*. The tendency of the observer's eyes to focus on the object, which is called *accommodation*, also contributes to the three-dimensional perception.

1.8.5.2 Microfiche Recording

Since the hologram does not use a lens in the fabrication process, the reconstructed image has neither limitations on the depth of focus nor aberrations due to the lens.

The scattered wave from the object is spread and recorded over the entire hologram, so that even if a portion of the hologram is missing or damaged, the same image could still be reconstructed. The damage to the hologram results in an overall degradation in quality, but no specific portion of the image is lost. Consequently, the hologram has a high tolerance to mishandling and is ideal for applications such as high-density microfiche recordings.

1.8.5.3 Measurement of Displacement

An ordinary photograph records only the intensity pattern, but the image reconstructed from a hologram has both phase and amplitude information. If the light waves of the reconstructed image are overlaid with those of the original object, an interference pattern is generated. The interference fringes indicate minute distortions of the object between its current state and its prior state when the hologram was recorded. This technique is called *interference holography*. An interference hologram can be fabricated by exposing a hologram twice to the same object, with and without the deformation. The resultant fringe patterns are seen in the reconstructed image.

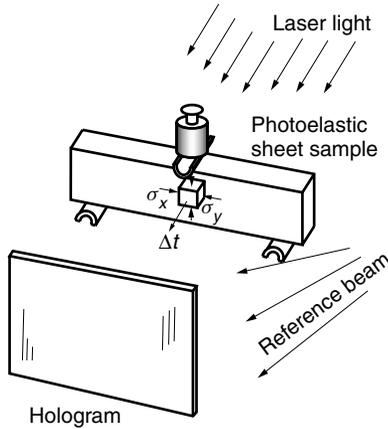


Figure 1.53 Interferometric hologram of a photoelastic sheet under load.

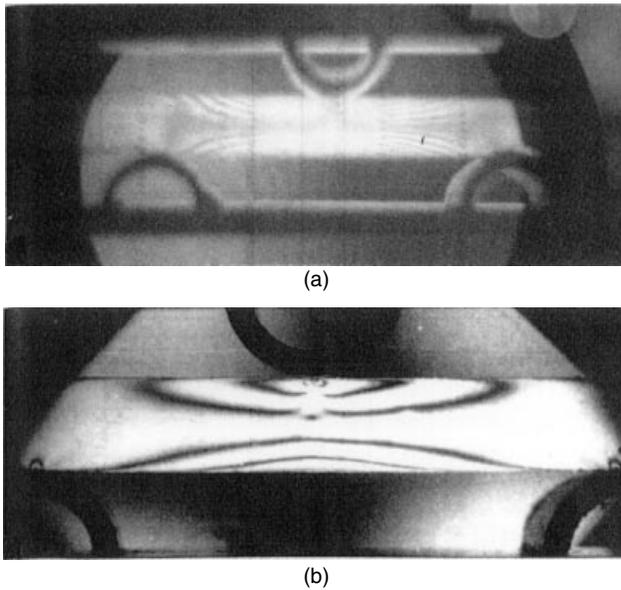


Figure 1.54 A photoelastic sheet under three-point loading. (a) Fringes of the interferometric hologram, which represent $\sigma_x + \sigma_y$. (b) Photoelastic fringes, which represent $\sigma_x - \sigma_y$. (Courtesy of X. Yan, T. Ohsawa, [25] and T. Ozaki.)

An example of combining the interferometric hologram with photoelasticity is presented in Figs. 1.53 and 1.54 [25]. With the geometry shown in Fig. 1.53, an interferometric hologram is made to visualize the pattern of strain established in a photoelastic sheet (e.g., PlexiglasTM). When a photoelastic sheet of uniform thickness is loaded, the thickness becomes nonuniform. When the direction of the illuminating laser beam is normal to the sheet surface, the phase change $\Delta\phi$ associated with the increase in the optical path length is

$$\Delta\phi = kn\Delta t \tag{1.277}$$

where Δt is the increase in thickness. The hologram is doubly exposed. The first exposure is made without compression and the second, under compression. Dark fringes will appear in the reconstructed image where

$$\Delta\phi = (2n + 1)\pi \quad (1.278)$$

due to the destructive interference between the two reconstructed images. This pattern is called the *isopachic fringe pattern*, meaning the locus of points of constant thickness of the sheet [26]. Both of the principal stresses σ_x and σ_y can be attributed to the change in thickness of the sheet. Δt is proportional to the sum of $\sigma_x + \sigma_y$. From the isopachic fringe alone, the contributions of σ_x and σ_y cannot be known separately.

Next, we will describe the procedure for observing the birefringence pattern of the same sample with the same loading. In order to do so, a polariscope (Section 6.7) is built around the sample. A polarizer is placed between the sample and the illuminating laser light, and the hologram is replaced by an analyzer whose transmission axis is perpendicular to that of the polarizer.

When the photoelastic sheet is stressed, the sheet becomes birefringent. That is, if the direction of compression is in the x direction, the index of refraction seen by the light polarized in the x direction is no longer the same as that of the y direction, and the degree of birefringence is proportional to the difference $\sigma_x - \sigma_y$ between the principal stresses. The degree of birefringence can be observed using the polariscope. With the isotropic (uncompressed) sample inserted in between the polarizer and the analyzer, no light gets through. With the birefringent (compressed) sample inserted in between the polarizer and analyzer, light is observed emerging from the analyzer in accordance with the degree of birefringence caused by the stress. Thus, the polariscope displays the pattern of the stress by means of the birefringence in the photoelastic sheet. By combining $\sigma_x + \sigma_y$ from the *isopachic fringes* with $\sigma_x - \sigma_y$ from the *photoelastic fringes*, it is possible to obtain σ_x and σ_y separately.

Figure 1.54a is the isopachic fringe pattern observed by means of the interferometric hologram. Figure 1.54b is the photoelastic fringe pattern of the same sample under the same compression as in Figure 1.54a observed by means of the polariscope.

1.8.5.4 Measurement of Vibration

Another application of holography is in recording vibration patterns. The hologram is able to record fringe patterns on the order of microns. While the holographic film is being exposed, both the object and the holographic plate must be very steady, otherwise the fringe patterns cannot be recorded. This fact can be used conversely to record the vibration pattern of an object since the vibrating portions of the object would appear dark in the reconstructed image. Figure 1.55 shows an example of mapping vibration patterns in a loudspeaker [27].

1.8.5.5 Nonoptical Holographies

So far, light has been assumed to be the wave source for generating holograms. In fact, other types of waves are capable of forming an interference pattern that can be used for generating holograms. For instance, microwave [8] and acoustic wave [28] holograms are possible. These hologram images can then be viewed optically by photographically reducing the dimensions of the hologram to the ratio of their relative wavelengths. By doing so, one can visualize how microwaves radiate from antennas (radiation patterns) or reflect from objects.

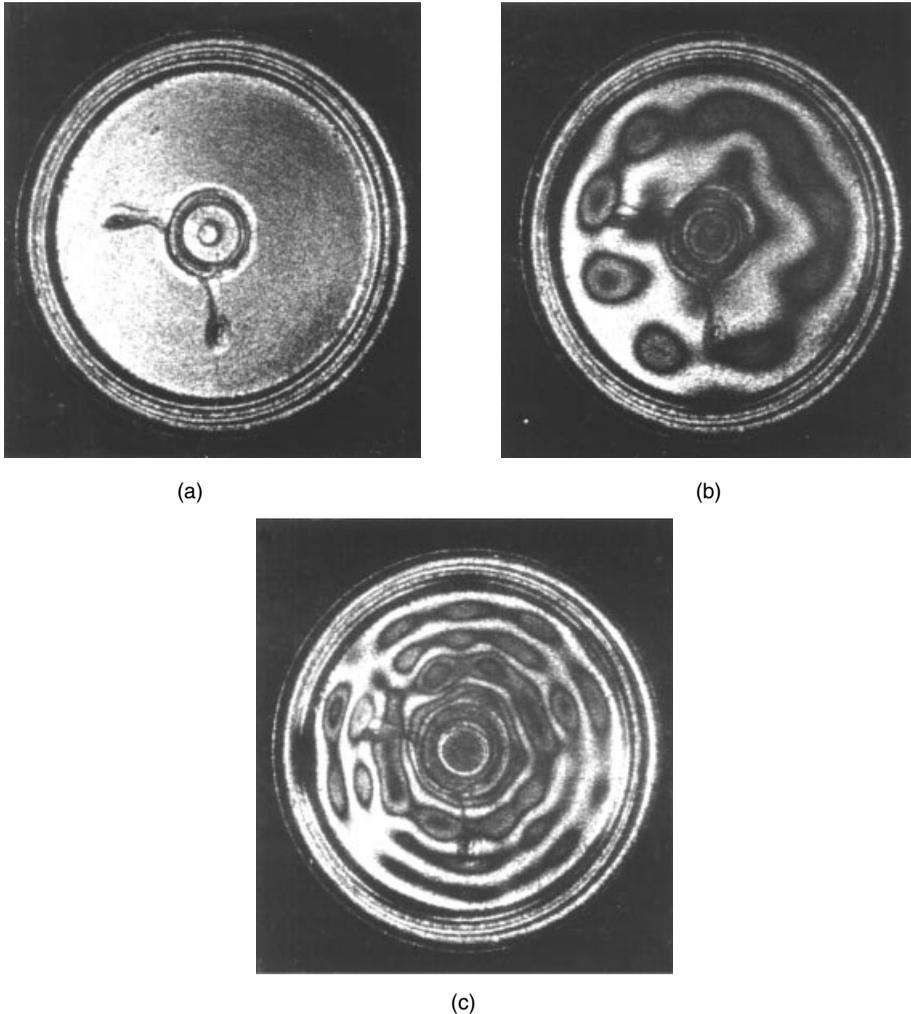


Figure 1.55 Images of a vibrating loudspeaker. (Courtesy of A. Marrakchi, J.-P. Huignard, and J. P. Herriau [27].)

Figure 1.56 shows an example of a microwave hologram used to examine the radiation pattern from an antenna [29]. Figure 1.56a shows a photograph of the object, which is a radiating monopole driven at the end of a microwave waveguide and Fig. 1.56b shows the reconstructed image of the plane of the monopole. The hologram reveals that the tip and the driving point are the major sources of radiation, and the function of the antenna wire is simply that of a feed wire for bringing current to the tip of the antenna. Figure 1.56c and 1.56d are the reconstructed images as the point of observation moves away from the antenna.

Figure 1.57a shows the geometry of a side scan sonar. The ship tows the sonar transducer whose beam is swept perpendicular to the direction of the ship's motion. The composite image is constructed as the ship cruises. The transducer is towed instead of being kept aboard the ship in order to decouple the motion of the ship. Figure 1.57b is the recorded image of a sunken ship resting on the sea bottom at a depth of 25 m.

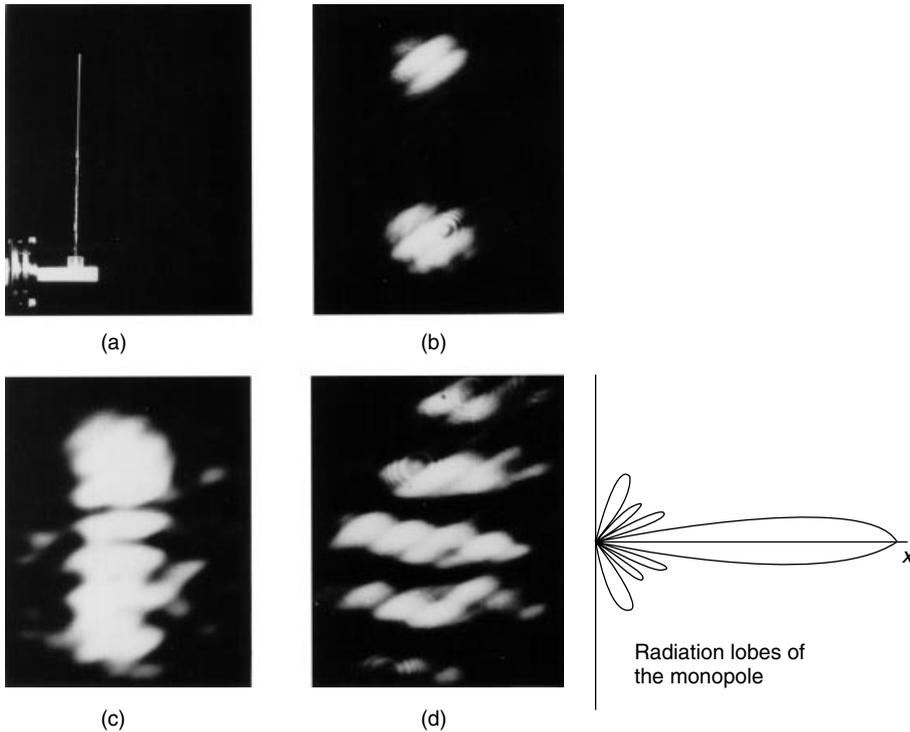


Figure 1.56 Visualization of a radiation pattern from a monopole antenna by means of a microwave hologram. (a) Photograph of the monopole used as the object. (b) At the antenna plane. (c) At a plane in the Fresnel region. (d) At the Fraunhofer region [29].

Figure 1.57c is the image of a bicycle sunken 4.2 m deep in the ocean [30]. The sound wave signal was processed by synthetic aperture [8] holography and even the spokes of the bicycle wheels can almost be resolved.

Another advantage of nonoptical holography is that an object imbedded in an optically opaque medium can be visualized.

A further advantage of the nonoptical hologram over “shadowgrams” such as an X ray is the freedom of choice of the hologram’s location with regard to the source and object. With X rays, the film always has to be behind the object, which is not true in holograms. A particularly useful example of this advantage is when nonoptical holograms are used in geological surveys, where placing a film deep beneath the surface is neither efficient nor desirable.

1.8.5.6 Computer-Generated Holograms

Holograms need not be fabricated using actual electromagnetic waves. They can be computer generated from values based on theory [31,32]. For instance, holograms can also be generated using the expected field pattern for a given object that can be calculated using the Fresnel diffraction formula of Eq. (1.38) or (1.39). The interference of the object and reference beams can then be calculated and drawn by computer. This image is then photographically reduced so that the holographic image can be reconstructed by light. Figure 1.58 shows an example of a computer-generated hologram.

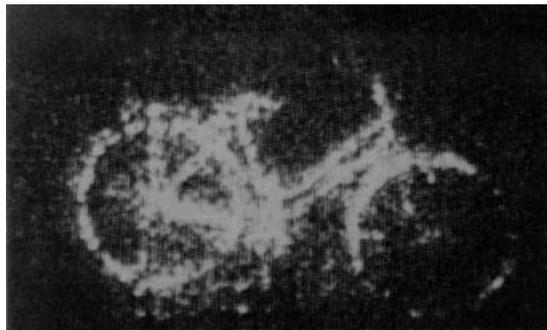
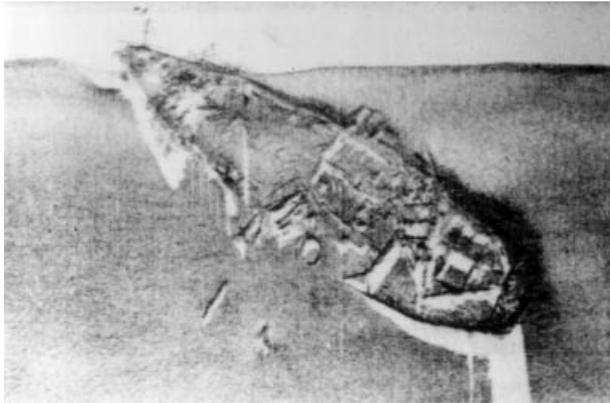
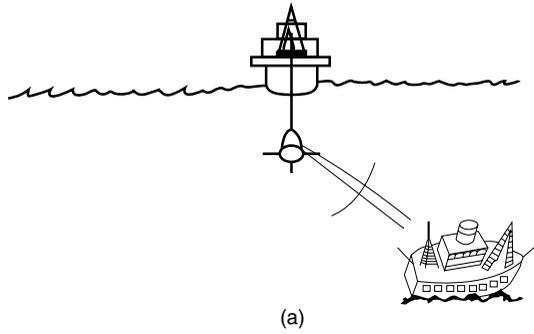


Figure 1.57 High-resolution underwater acoustic images. (a) Geometry of the side scan sonar. (b) *Vineyard Sound Lightship* on the sea bottom off Nantucket, Massachusetts, recorded with a HYDROSCAN. (Courtesy of Klein Associates, Inc.) (c) Synthetic aperture holographic acoustic image at 4.2 m. (Courtesy of K. Mano and K. Nitadori [30].)

Figure 1.59 shows another kind of computer-generated hologram. The image of the hologram is reconstructed by light propagating in the plane of the hologram [33]. As shown in Fig. 1.59a, the reconstruction beam is fed through an optical fiber pigtail. Figure 1.59b is a photograph of a scanning electron microscope image of the computer-generated hologram. The pattern was generated using electron-beam

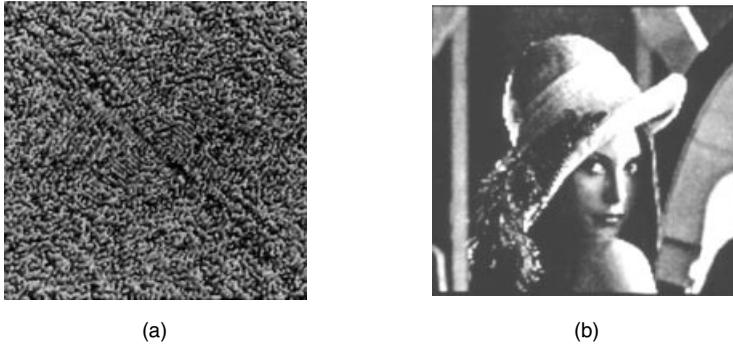


Figure 1.58. Computer-generated holography. (a) Hologram. (b) Reconstructed image. (Courtesy of D. Asselin, A. Bergeron, and INO)

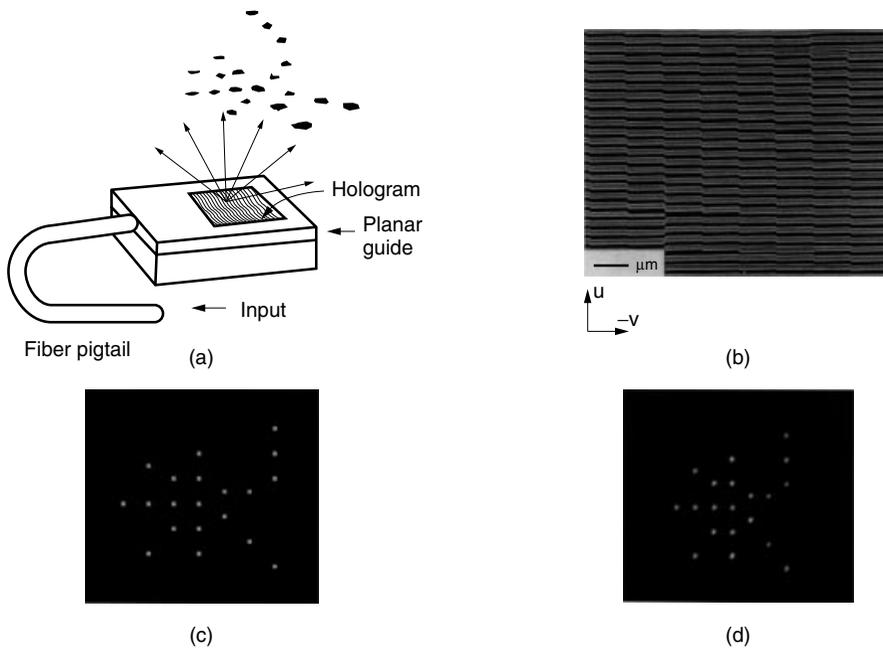


Figure 1.59. Off-plane computer-generated waveguide hologram (OP-CGWH). (a) Geometry. (b) Photograph of scanning electron microscope image of the hologram. (c) Designed pattern. (d) Obtained pattern. (After M. Li et al. [33].)

lithography. Figure 1.59c gives the desired pattern of the reconstructed image. Figure 1.59d shows the reconstructed image from the fabricated computer hologram. A hologram such as this can be used as an optical interconnect, where the output light from an optical fiber has to be connected to multiple terminals through free-space propagation.

1.8.5.7 Holographic Video Display

Figure 1.60 shows a schematic of a holographic video display system [34]. The system is intended to animate computer-generated holograms. The key component of the system is the surface acoustooptic modulator (AOM) on which three of the line fringe patterns are written for three primary color holograms.

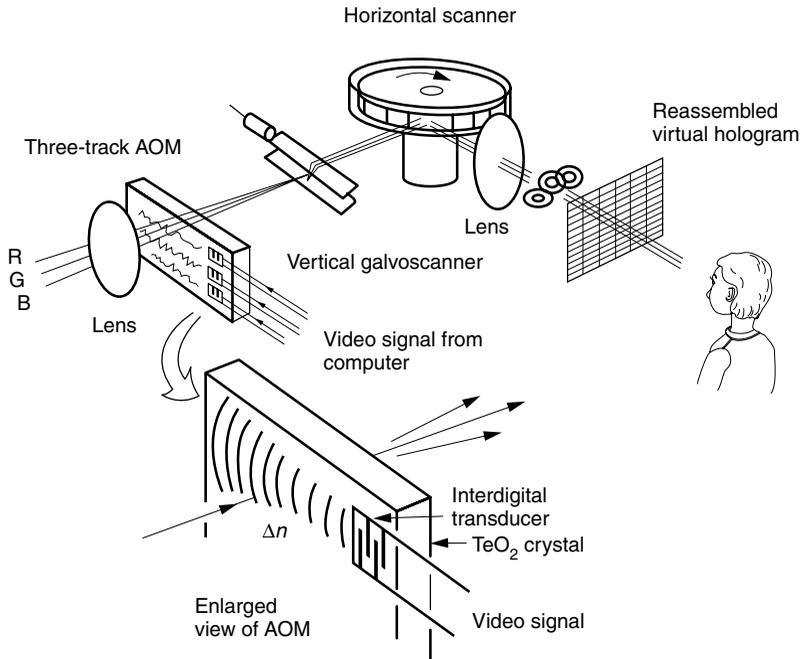


Figure 1.60 Schematic of a color holographic video display. (After P. St.-Hilaire et al. [33].)

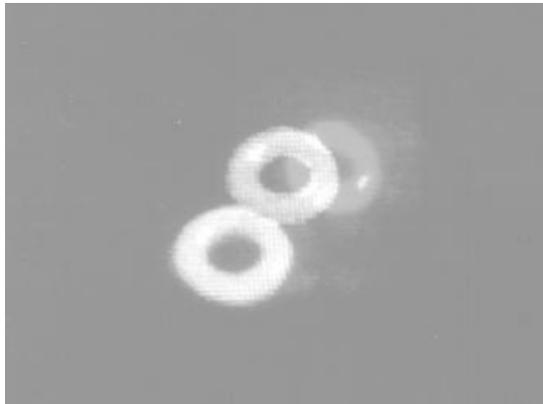


Figure 1.61. A frame of the animated color computer-generated hologram of colorful donuts. (Courtesy of P. St.-Hilaire et al. [33].)

The fringes of the holograms are scanned as a composite of line holograms. The diffracted images from the line holograms are reassembled into a complete image of a computer-generated hologram using horizontal as well as vertical scanners. The AOM is made of a TeO_2 crystal. A pair of interdigital electrodes are deposited on the crystal surface. The interdigital electrodes launch a surface acoustic wave (SAW) due to the piezoelectric effect in accordance with the video signal from the computer. The surface wave spatially modulates the index of refraction due to the acoustooptic effect and writes a one-line hologram on the crystal as it propagates along the crystal (see Example 5.6).

When laser light is illuminated perpendicular to the surface, a diffracted image moving at the speed of the surface acoustic wave is generated.

The horizontal scanner is an 18-sided polygonal mirror. It scans in the opposite direction to the movement of the diffracted image to immobilize the image. The horizontal scanner also multiplexes the image of the crystal, creating a virtual crystal that is exactly the same length as one line of the computer-generated hologram.

The galvanometric scanner shifts the horizontal one-line fringe vertically after each horizontal scan completing the reassembly of the computer-generated hologram.

Figure 1.61 shows one of the frames of an animated color computer-generated hologram of colorful donuts. Three primary color lasers were used simultaneously. These are a HeNe laser at $\lambda = 633 \text{ nm}$ for red, a frequency-doubled YAG laser at $\lambda = 532 \text{ nm}$ for green, and a HeCd laser at $\lambda = 442 \text{ nm}$ for blue.

PROBLEMS

1.1 For a plane wave that is propagating (Fig. P1.1) in the direction

$$\theta = 45^\circ \quad \phi = 45^\circ$$

the light field observed at $P(2, 3, 4) \times 10^{-6} \text{ m}$ is expressed as

$$E = E_0 e^{j67.32 - j2.44 \times 10^{15} t}$$

- (a) Find the wavelength of light in the medium.
- (b) Find the index of refraction of the medium.

1.2 The spatial frequencies of an incident wave (Fig. P1.2) were measured along the x and y axes as

$$f_x = 0.6 \text{ lines}/\mu\text{m}, \quad f_y = 0.8 \text{ lines}/\mu\text{m}$$

The wavelength of the received light is $\lambda = 0.84 \mu\text{m}$.

- (a) What is the direction of propagation of the incident wave?
- (b) What would be the spatial frequency if the measurement were made along a line in the direction $\hat{\mathbf{I}} = 3\hat{\mathbf{i}} + 4\hat{\mathbf{j}}$?

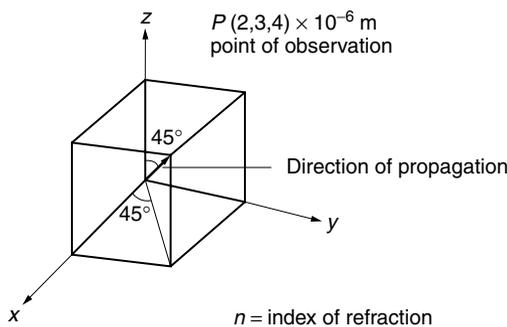


Figure P1.1 Geometry of the propagation.

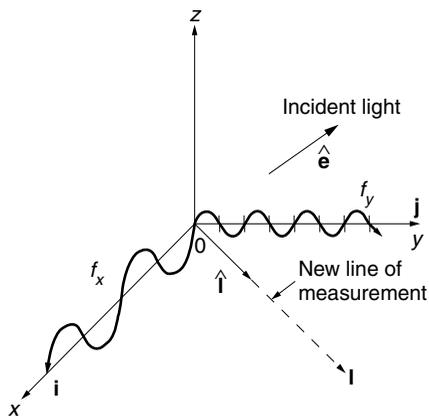


Figure P1.2 Geometry of the incident wave.

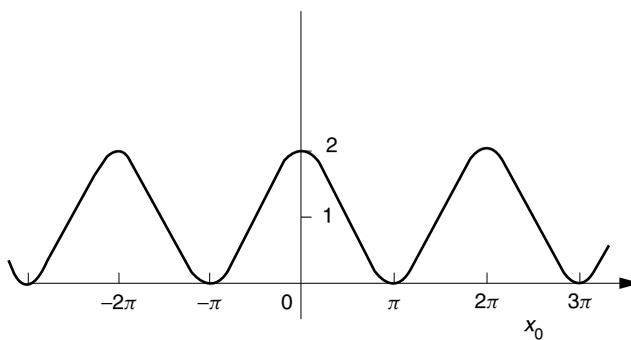


Figure P1.3 Sinusoidal distribution $1 + \cos 2\pi f_{x_0} x_0$.

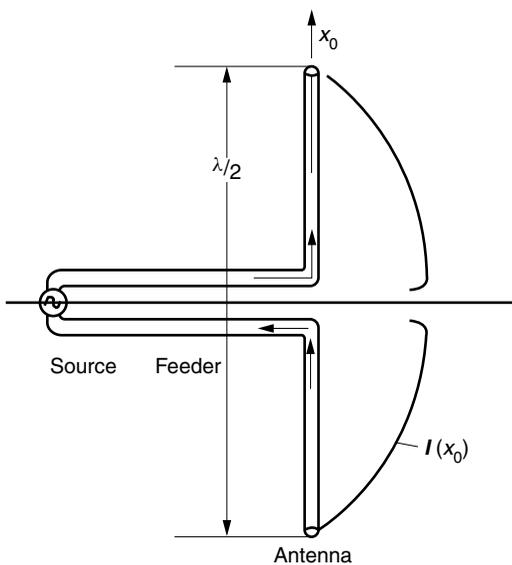


Figure P1.4 Half-wave dipole antenna and its current distribution.

- 1.3** Find the diffraction pattern from a film whose transmittance is modulated sinusoidally as shown in Fig. P1.3:

$$t(x_0, y_0) = 1 + \cos 2\pi f_{x_0} x_0$$

Assume that the size of the film is infinitely large.

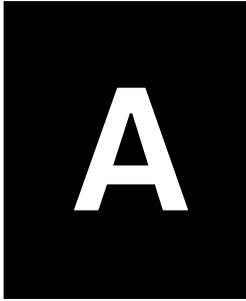
- 1.4** Derive the radiation pattern of a half-wave dipole using Fourier optics. Assume that the current distribution along the antenna is sinusoidal (Fig. P1.4).
- 1.5** Match each of the letter apertures in Fig. P1.5a with its correct diffraction pattern in Fig. P1.5b.
- 1.6** Figure P1.6a shows apertures of three of the fingerprints that appeared in Fig. 1.36, and an aperture in the shape of a cartoon character. Figure P1.6b shows the photographs of the diffraction patterns. Match the apertures with their respective diffraction patterns.
- 1.7** The finest possible light spot is to be obtained using a finite size convex lens. The aperture of the lens is square $\Pi(x/a)\Pi(y/a)$ and the focal length is f_0 . The incident light is a parallel beam.
- (a) What is the smallest spot size obtainable?
- (b) What is the required size of a diffraction-limited lens that can resolve $1 \mu\text{m}$. The wavelength of the light is $\lambda = 0.555 \mu\text{m}$ and $f_0 = 50 \text{ mm}$. Assume a rectangular lens.
- 1.8** Explain the principle of operation of a pinhole camera using Fourier optics (Fig. P1.8). Assume the shape of the pinhole is a square.
- 1.9** An input function $g(x_0, y_0)$ is placed in the front focal plane of lens L_1 with focal length f_1 (Fig. P1.9). A second lens L_2 with focal length f_2 is placed behind lens L_1 at a distance $f_1 + f_2$ from L_1 . Find the expression for the field at the back focal plane of lens L_2 .
- 1.10** The optical system shown in Fig. 1.33c is a Schlieren camera made by placing an opaque dot or $\pi/2$ -radian phase plate at the back focal point G_0 of lens L_2 . Let us say that the object is transparent but with a small variation of the index of refraction so that the input function

$$g(x_0, y_0) = e^{j\phi(x_0, y_0)}$$

can be approximated as

$$g(x_0, y_0) = 1 + j\phi(x_0, y_0)$$

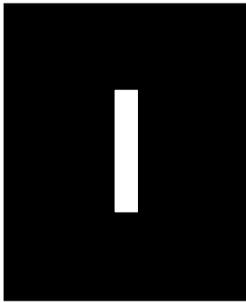
- (a) Find the expression for the output intensity distribution $I_a(x_i, y_i)$ with the opaque dot placed at the back focal point G_0 of lens L_2 .
- (b) Find the output intensity distribution $I_b(x_i, y_i)$ with the $\pi/2$ -radian phase plate at the same location as the opaque dot in part (a).
- (c) Compare the results of parts (a) and (b).



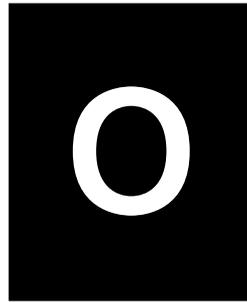
(A)



(B)

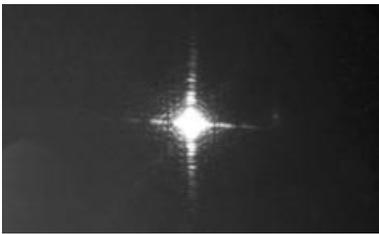


(C)

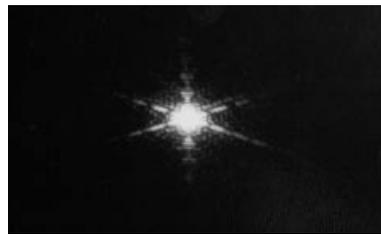


(D)

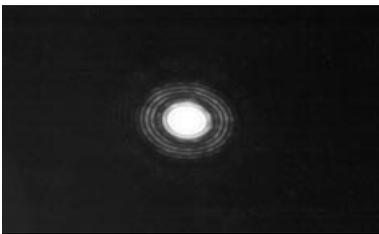
(a)



(1)



(2)



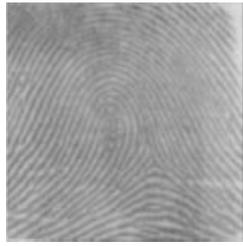
(3)



(4)

(b)

Figure P1.5. Match the aperture with its diffraction pattern. (a) Apertures of letters. (b) Photographs of diffraction patterns.



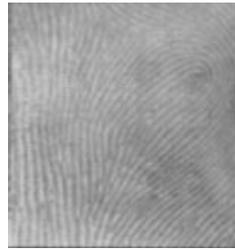
(A)



(B)



(C)

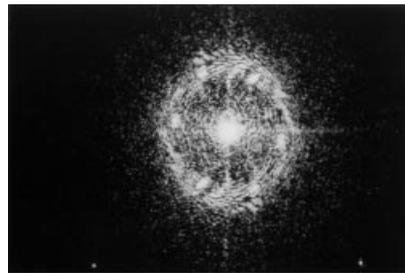


(D)

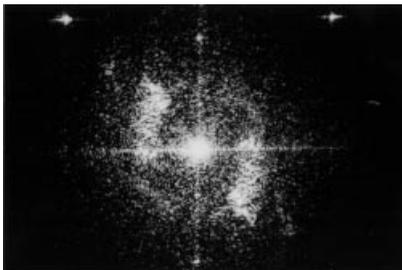
(a)



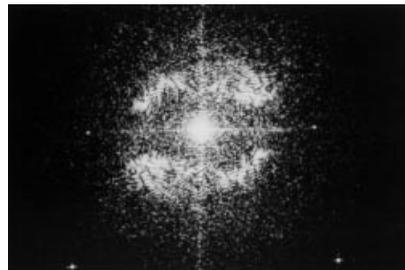
(1)



(2)



(3)



(4)

(b)

Figure P1.6. Match the aperture with its diffraction pattern. (a) Apertures of fingerprints and a cartoon character. (b) Photographs of diffraction patterns.

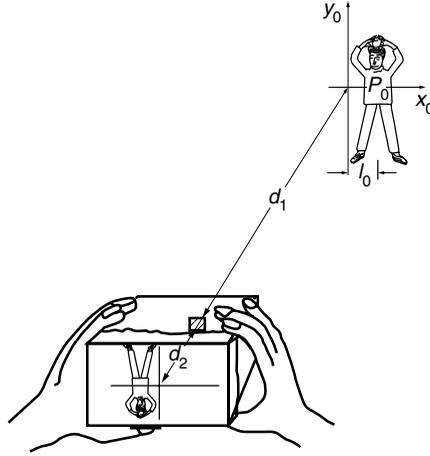


Figure P1.8 Principle of the pinhole camera by Fourier optics.

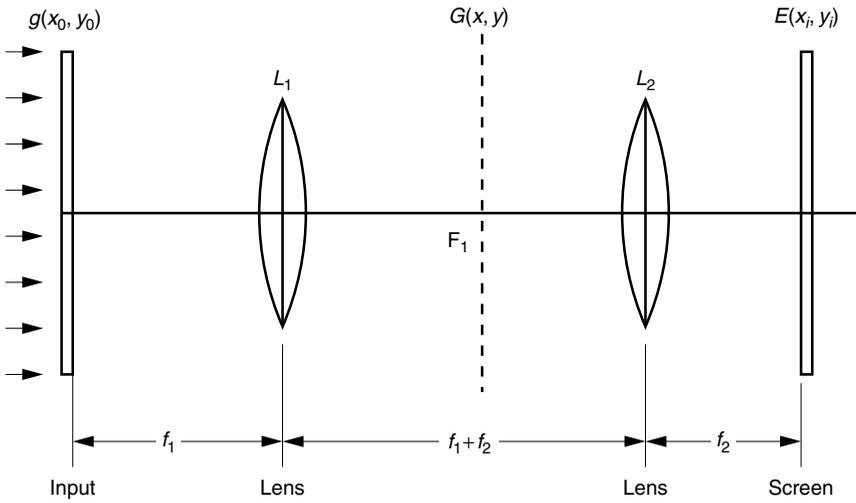


Figure P1.9 Two Fourier transform lenses in series.

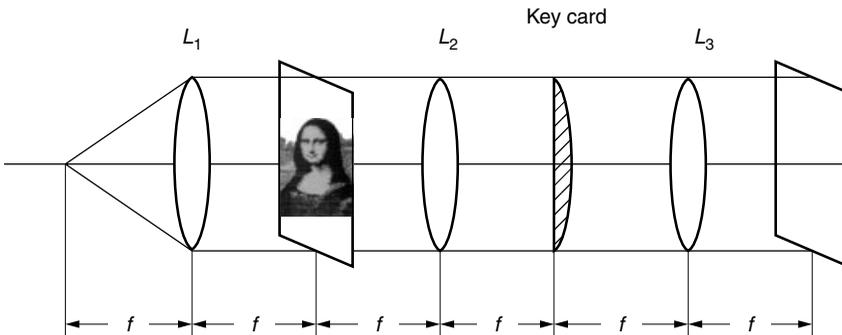


Figure P1.11 A convex lens with focal length f_0 is used as the key card for encryption.

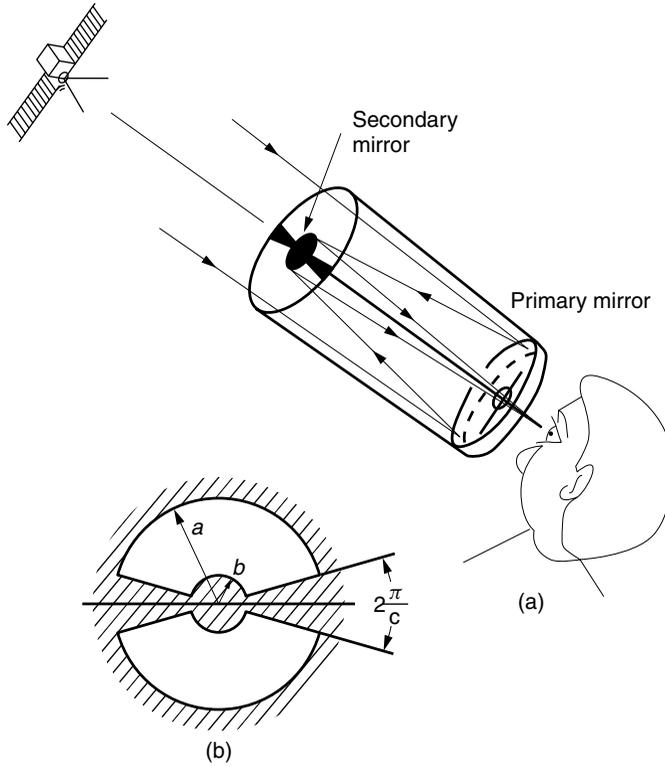


Figure P1.12 Cassegrain reflecting telescope with obstruction. (a) Telescope. (b) Obstruction.

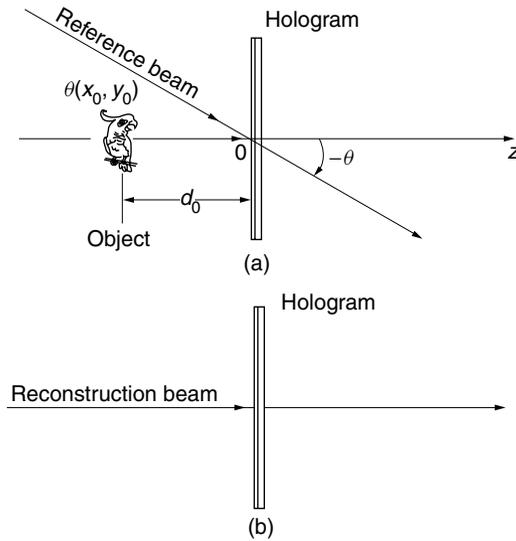


Figure P1.13 Locations of the reconstructed images. (a) Fabrication of hologram. (b) Reconstructing the images.

- 1.11** With the encryption scheme shown in Fig. P1.11, if a convex lens of focal length f_0 is used as the key card for the encryption, what is the key card for the decryption?
- 1.12** A Cassegrain telescope has a primary concave mirror and a secondary convex mirror, as shown in Fig. P1.12a. The supports holding the secondary mirror present an obstruction and act like a mask, as shown in Fig. P1.12b. Find the diffraction pattern of the masked aperture shown in Fig. P1.12b.
- 1.13** A hologram was fabricated with the geometry shown in Fig. P1.13. Assume the thin emulsion case. The object was placed on the z axis, which is perpendicular to the hologram. A plane wave reference beam was incident at angle θ to the normal of the hologram. Find the locations of the reconstructed images when the reconstruction beam is not identical with the reference beam and is incident along the z axis.

REFERENCES

1. J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, New York, 1968.
2. M. V. Klein and T. E. Furtak, *Optics*, 2nd ed., Wiley, New York, 1986.
3. E. Hecht, *Optics*, 3rd ed., Addison-Wesley, Reading, MA, 1998.
4. F. T. S. Yu and X. Yang, *Introduction to Optical Engineering*, Cambridge University Press, Cambridge, 1997.
5. R. D. Guenther, *Modern Optics*, Wiley, New York, 1990.
6. B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*, Wiley, New York, 1991.
7. G. R. Fowles, *Introduction to Modern Optics*, 2nd ed., Dover Publications, New York, 1989.
8. K. Iizuka, *Engineering Optics*, 2nd ed., Springer Verlag, Berlin, 1986.
9. C. A. Balanis, *Antenna Theory Analysis and Design*, 2nd ed., Wiley, New York, 1997.
10. C. R. Wylie and L. C. Barrett, *Advanced Engineering Mathematics*, 6th ed., McGraw-Hill, 1995.
11. G. A. Massey, "Microscopy and pattern generation with scanned evanescent waves," *Appl. Opt.* **23**(5), 658–660 (1984).
12. A. Vander Lugt, "Practical considerations for the use of spatial carrier-frequency filters," *Appl. Opt.* **5**(11), 1760–1765 (1966).
13. F. T. S. Yu and Suganda Jutamulia, *Optical Pattern Recognition*, Cambridge University Press, Cambridge, 1998.
14. C. S. Weaver and J. W. Goodman, "A technique for optically convolving two functions," *Appl. Opt.* **5**(7), 1248–1249 (1966).
15. G. Lu, Z. Zhang, S. Wu, and F. T. S. Yu, "Implementation of a non-zero-order joint-transform correlator by use of phase-shifting techniques," *Appl. Opt.* **36**(2), 470–483 (1997).
16. P. Purwosumarto and F. T. S. Yu, "Robustness of joint transform correlator versus Vander Lugt correlator," *Opt. Eng.* **36**(10), 2775–2780 (1997).
17. F. T. S. Yu, Q. W. Song, Y. S. Cheng, and D. A. Gregory, "Comparison of detection efficiencies for Vander Lugt and joint transform correlators," *Appl. Opt.* **29**(2), 225–232 (1990).
18. L. Leclerc, Y. Sheng, and H. H. Arsenault, "Circular harmonic covariance filters for rotation invariant object recognition and discrimination," *Opt. Commun.* **85**, 299–305 (1991).

19. D. Casasent and D. Psaltis, "Position, rotation, and scale invariant optical correlation," *Appl. Opt.* **15**(7), 1795–1799 (1976).
20. B. Javidi, "Optical information processing for encryption and security systems," *Opt. Photonics News*, 29–33 (Mar 1997).
21. B. Javidi and A. Sergent, "Fully phase encoded key and biometrics for security verification," *Opt. Eng.* **36**(3), 935–942 (1997).
22. F. H. Mok, "Angle-multiplexed storage of 5000 holograms in lithium niobate," *Opt. Lett.* **18**(11), 915–917 (1993).
23. G. Barbastathis and D. J. Brady, "Multidimensional tomographic imaging using volume holography," *Proc. IEEE* **87**(12), 2098–2120 (1999).
24. P. Hariharan, *Optical Holography, Principles, Techniques and Applications*, 2nd ed., Cambridge University Press, Cambridge, 1996.
25. X. Yan and T. Ohsawa, "Measurement of the internal local stress distribution of composite materials by means of laser imaging methods," *Composites* **25**(6), 443–450 (1994).
26. J. W. Dally and W. F. Riley, *Experimental Stress Analysis*, 3rd ed., McGraw-Hill, New York, 1991.
27. A. Marrakchi, J.-P. Huignard, and J. P. Herriau, "Application of phase conjugation in $\text{Bi}_{12}\text{SiO}_{20}$ crystals to mode pattern visualization of diffuse vibrating structures," *Opt. Commun.* **34**(1), 15–18 (1980).
28. B. P. Hildebrand and B. B. Brenden, *An Introduction to Acoustic Holography*, Plenum Press, New York, 1972.
29. K. Iizuka and L. G. Gregoris, "Application of microwave holography in the study of the field from a radiating source," *Applied Physics Letters* **17**(12), 509–512 (1970).
30. K. Mano and K. Nitadori, "An experimental underwater viewing system using acoustical holography," *1977 IEEE Ultrasonics Symposium Proceedings*, Phoenix, Arizona, Oct. 26–28, IEEE Cat. #77CH/264-1SU, pp. 272–277, 1977.
31. A. Bergeron, H. H. Arsenault, J. Gauvin, and D. J. Gingras, "Computer-generated holograms improved by a global iterative coding," *Opt. Eng.* **32**(9), 2216–2226 (1993).
32. H. Yoshikawa and H. Taniguchi, "Computer generated rainbow hologram," *Opt. Rev.* **6**(2), 118–123 (1999).
33. M. Li, J. Bengtsson, M. Hagberg, A. Larsson, and T. Suhara, "Off-plane computer-generated waveguide hologram," *IEEE J. Selected Topics Quantum Electro.* **2**(2), 226–235 (1996).
34. P. St.-Hilaire, S. A. Benton, M. Lucente, and P. M. Hubel, "Color images with the MIT holographic video display," *SPIE Proceedings*, Vol. 1667, *Practical Holography VI*, San Jose, pp. 73–84, 1992.