united nations
educational, scientific
and cultural
organization

international atomic
energy agency

the
**abdus salam**
international centre for theoretical physics

**Winter College on Optics and Photonics**
**7 - 25 February 2000**

**1218-15**

# "Principles of Fourier Optics"

**W.T. RHODES**
**Georgia Institute of Technology**
**USA**

***Please note: These are preliminary notes intended for internal distribution only.***

# Preface to the Manuscript

I have taught from this and predecessor material for a number of years now, always at the introductory graduate level. Chapters 1 through 9 can be covered easily in a one-quarter course on Fourier optics, assuming that the students have had some introduction to one-dimensional Fourier theory. If desired, chapter 10 on spatial light modulators and chapter 11 on holography can be included for a fast-paced one-quarter course, perhaps with the omission of chapter 9 on coherence theory and speckle. Alternatively, Chapter 12 on coherent optical spectrum analysis, 13 on coherent spatial filtering, and 14 on incoherent spatial filtering can be covered. The entire text is well suited to two quarters or one semester of instruction.

A comparison of this text with Goodman's *Introduction to Fourier Optics* is likely to be of interest to those who have already studied or taught Fourier optics. I first learned Fourier optics from Joseph Goodman, and for many years I taught out of his classic text. I found it difficult to begin writing a book of my own that would be compared with his. Ultimately I decided in favor of the project because over the years I have developed what I think are improved ways of looking at some of the basic material of Fourier optics and, frankly, I was afraid someone else would put them into a textbook before I did! Distinctions I find important between my treatment of Fourier optics and Goodman's include a greater emphasis on the angular spectrum concept, focus on spherical wave (both expanding and converging) illumination of a transparent object as the key to understanding the Fourier transform property of a lens, a corrected and more general treatment of monochromatic image formation, a clearer distinction between coherent and incoherent imaging systems, the introduction of key concepts from coherence theory, a clearer discussion of the meaning and significance of the assumption of quasimonochromaticity, and the use of more concepts from geometrical optics. In addition, there is a somewhat different variety of applications of Fourier optics presented in the two texts. Goodman's text presents by far the more thorough treatment of the classical diffraction theory as presented in the Fresnel-Kirchhoff and Rayleigh-Sommerfeld theories. The notation used in this text is similar to that used in Goodman's, although where convenient I rely on an operator notation not used by him. The basic organization and flow of material in the first half of the text—diffraction theory, the Fourier transform property of spherical lenses, coherent and incoherent imaging, and an introduction to holography—is very similar to that in Goodman's book, and people who are familiar with his text should feel comfortable teaching from this text, despite the differences.

The text as it currently stands suffers from a lack of good bibliographical materials. References must be added before I will consider it ready to publish. More importantly, some of the applications chapters are as yet incomplete. In addition, I think the text would benefit from the inclusion of more figures. Figures are expensive from a publisher's (and, ultimately, the student's) perspective, but they can do much to increase the learning rate for the student. Finally, the text will benefit from use in its preliminary form by other teachers who can—and, I hope, will—bring to my attention the deficiencies they find in using it themselves as a primary text for their courses.

Atlanta
January, 1997

# Principles of Fourier Optics

**WILLIAM T. RHODES**
Georgia Institute of Technology

# 1. Introduction

Historically, the subjects of optical wavefield propagation and formation of optical images were introduced in physics courses, usually in the form of geometrical optics and diffraction theory. In the 1960's, however, university engineering departments began introducing their own courses in optics. The movement of optics into engineering programs was concurrent with the development of the laser, which produced greatly renewed interest in optics among both the scientific and the engineering communities. Electrical engineering departments in particular, which had grown considerably in size and importance during the earlier part of the 20th century, were eager to embrace courses that related to this significant new technology.

Of extreme importance in bringing optics courses to engineering programs was the set of familiar mathematical tools that the engineer could use in the analysis of wave propagation, image formation, and related phenomena. In particular, it was noted by scientists in the 1940's and 1950's that linear systems theory, long of use in the analysis of electrical circuits and radio signal modulation and demodulation could also be applied to optical systems.

## 1.1 THE ROLE OF LINEAR SYSTEMS IN OPTICS

The theory of linear, shift-invariant systems is well known to the upper-division electrical engineering student. Linearity implies that the amplitudes of signals in, for example, an electronic circuit can be added. Shift invariance, in the particular form of time invariance, implies that the response characteristics of the circuit do not change with time: rather, a delay in applying the input excitation produces only a corresponding delay in the production of the output response.

From a physical standpoint it is not at all surprising that analogous conditions apply to the propagation of optical wavefields and to the formation of images. Linearity in optics is found in the additivity of electromagnetic wavefields, an attribute that both electrical engineering and physics students take for granted: voltages add, Maxwell's equations imply linearity, etc. With certain restrictions (differences that exist between laser light and non-laser light are sometimes important in this regard), linearity implies that light intensities add, too: the light intensity pattern produced by two desk lamps equals the sum of the patterns produced by each lamp individually.

Shift invariance appears in its most important form in the *space invariance* of optical systems investigated in this text. Two simple examples suffice to illustrate. The first, suggested in Fig. 1-1, is that of a source of light—a flashlight—that produces a particular light intensity distribution on a distant wall. Move the source in a plane parallel to the wall and the pattern on the wall moves with it, but otherwise remains the same. The shift in source position is matched by a corresponding shift—but otherwise no change—in the resulting light pattern on the wall. The overall system, with source serving as input and pattern on the wall serving as output, exhibits the characteristics of shift invariance. We say that the system is space invariant. The other example is that of image formation. Assume that a simple imaging system produces an image of a planar object. If the object moves laterally in its plane, the image moves a corresponding distance in its plane, suggestive of shift invariance—i.e., a shift in the position of the object results in a corresponding shift in the position of the image but otherwise no change.





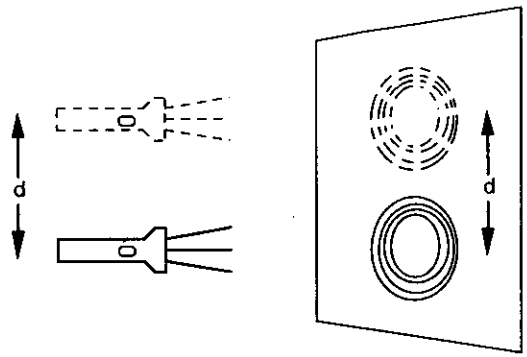



Fig. 1-1. Example of space invariance in optics. Moving the flashlight a given distance in a plane parallel to the wall moves the illumination pattern on the wall through the same distance but leaves that pattern otherwise unchanged.

The attributes of linearity and shift invariance in optical wave propagation and image formation are exploited extensively in much of this book. Among other things, these attributes imply that many optical systems can be analyzed using the tools and techniques of convolution, Fourier transforms, and transfer functions. It will be seen that these tools are powerful indeed in helping the engineer and scientist analyze, understand, and use a wide variety of optical systems.

## 1.2 THE ORGANIZATION OF THIS BOOK

This text splits naturally into two parts. The first part, comprising Chapters 3 through 9, along with the two-dimensional functions and transforms material of Chapter 2 as background, emphasizes the basic physics of optical wave propagation and image formation. The point of view is that of *analysis*: How do we analyze these phenomena? How do we understand them? Chapter 3 introduces mathematical models for optical waves and describes how wave intensities relate to wave amplitudes. Chapter 4 describes the basic phenomenon of optical wavefield propagation: in classical terms, the theory of diffraction. Chapters 5 through 8 introduce the characteristics and capabilities of lenses in optical systems: their ability to transform one wavefield into another in accord with the Fourier integral transform (Chapter 5), their ability to form images of wavefields produced by laser sources (Chapter 6) and non-laser sources (Chapter 7), and the results of their use in imaging systems with transparent objects (Chapter 8). Both space-domain (convolution) and frequency-domain (transfer function) approaches to imaging system analysis are emphasized. In Chapter 9, the distinctions that apply to light from different kinds of sources—specifically, laser light as compared with light from incandescent, fluorescent, or other non-laser sources—are elaborated upon and modeled analytically. The concept of partially coherent wavefields is explored, and the interference of such wavefields is briefly analyzed.

The remaining chapters, by way of contrast with the first nine, take the point of view of *synthesis and application*. How do we apply these concepts—the basic physics of wave propagation, image formation, etc.—to the processing of information? Chapters 10 and 11 are transitional. In Chapter 10 photographic film and spatial light modulators, which provide us with means for controlling light intensity and wave amplitude distributions, are discussed. In Chapter 11 holography is described, both as an example and as an application of basic optical wave phenomena. Chapters 12, 13, and 14 build from the analytical foundation provided in Chapters 5,

6, and 7 by describing the Fourier transforming and spatial filtering characteristics of systems constructed of lenses, but now from the perspective of application. Chapter 15 describes the construction of optical elements—computer-generated transparencies, for example—that can be used in spatial filtering systems. The remaining chapters of the book describe a variety of other important applications of Fourier optics concepts and techniques, all for processing signal information of one form or another. Chapter 16, on acousto-optic signal processing, emphasizes the processing of one-dimensional signals (e.g., radar, telecommunications) using devices that impose information on light waves through the interaction of those waves with sound waves. In Chapter 17, which introduces concepts of nonlinear optical signal processing, it is shown how linear, shift-invariant optical systems can be combined with simple nonlinear operations to perform powerful image processing operations. In Chapter 18 attention turns to more computationally-oriented signal processing, specifically to the use of optical systems to perform algebraic computations (e.g., vector-matrix multiplication). The idea of exploiting the *interconnectivity* afforded by optical beams is introduced. In Chapter 19, on shift-variant optical processing operations, it is shown how the limitations presumably inherent in a shift-invariant system can, to an extent, be overcome, allowing certain useful shift-*variant* operations to be performed. Chapters 20 and 21 introduce several historically important applications of Fourier-optics-based processors: the processing of synthetic aperture side-looking radar images (Chapter 20) and the application of Fourier optics to coded aperture imaging and tomography (Chapter 21). Finally, Chapter 22 introduces a potpourri of techniques chosen for their novelty and/or cleaverness in exploiting capabilities and principles of optics in information processing applications.

# 2. Two-Dimensional Functions and Transforms

Two-dimensional (2-D) functions play an important role in optics. In the diffraction and imaging of light wave fields, for example, the distributions of concern are usually functions of two spatial variables. Furthermore, stops, apertures, lenses, and similar elements in an optical system are normally modeled by 2-D functions.

Two-dimensional mathematical operations are also important. For example, in both wave-field propagation and imaging problems, 2-D convolution can be used to describe the relationship between light distributions in two different planes. And this fact suggests in turn that the relationship between the two distributions can also be described in terms of 2-D Fourier transforms and 2-D transfer functions.

The purpose of this chapter is to introduce a number of two-dimensional functions that can be used to model simple apertures, light intensity patterns, and other distributions of interest in imaging and optical wave propagation and to introduce 2-D convolution, correlation, and Fourier transform operations. Also introduced are the Fresnel transform and the application of the 2-D Dirac comb function to sampling and replication in two dimensions. Much of what is presented in the chapter is a straight-forward extension of corresponding one-dimensional (1-D) material.

## 2.1 TWO-DIMENSIONAL FUNCTIONS

It is convenient to have a compact notation for certain 2-D functions that are used throughout the text. Included among these are the 2-D unit rectangle and unit triangle functions, the standard 2-D Gaussian ,the 2-D sinc function, and the 2-D delta and Dirac comb (sampling) functions. These functions are illustrated in Fig. 2-1 in perspective drawings and, excepting the 2-D delta and comb functions, as grayscale images. Each of these functions is a direct extension of its 1-D counterpart, having the form f(x,y)=g(x)g(y), where g(x) is the corresponding standard 1-D function. Thus,

$$\mathrm{rect}(x,y) = \mathrm{rect}(x)\,\mathrm{rect}(y), \tag{2.1}$$

$$\mathrm{tri}(x,y) = \mathrm{tri}(x)\,\mathrm{tri}(y), \tag{2.2}$$

$$\mathrm{sinc}(x,y) = \mathrm{sinc}(x)\,\mathrm{sinc}(y), \tag{2.3}$$

$$\exp\left[-\pi\left(x^2+y^2\right)\right] = \exp\left(-\pi x^2\right)\exp\left(-\pi y^2\right), \tag{2.4}$$

$$\delta(x,y) = \delta(x)\,\delta(y), \tag{2.5}$$

$$\mathrm{comb}(x,y) = \mathrm{comb}(x)\,\mathrm{comb}(y). \tag{2.6}$$





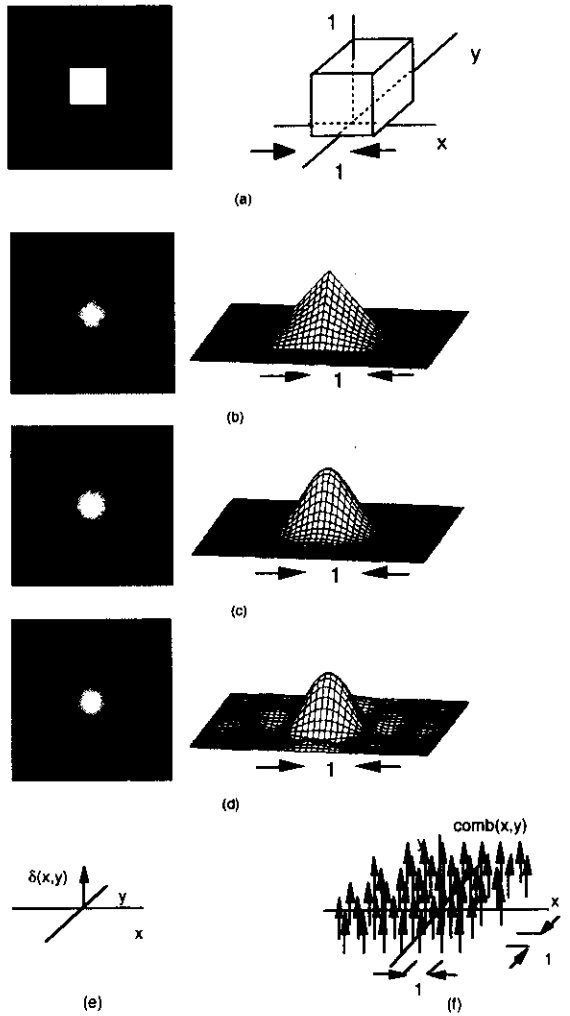



Fig. 2-1. Six standard separable 2-D functions: (a) rect(x,y), (b) tri(x,y), (c) Gaussian, (d) sinc(x,y) (with bias in the grayscale representation to accommodate negative values), (e) delta function, (f) comb(x,y) (central portion only).

The component 1-D functions rect(x), tri(x), sinc(x), and comb(x) are defined as follows:

$$\text{rect}(x) = \begin{cases} 1 & |x| < 1/2 \\ 1/2 & |x| = 1/2 \\ 0 & \text{otherwise} \end{cases}, \tag{2.7}$$

$$\text{tri}(x) = \begin{cases} 1-|x| & |x| < 1 \\ 0 & \text{otherwise} \end{cases}, \tag{2.8}$$

$$\text{sinc}(x) = \frac{\sin \pi x}{\pi x}, \tag{2.9}$$

$$\text{comb}(x) = \sum_{n=-\infty}^{\infty} \delta(x-n). \tag{2.10}$$

These 1-D functions are illustrated in Fig. 2-2. Note that sinc(x) equals zero for x=±1, ±2, ±3, etc. Also note that, since $\delta(x,y)=\delta(x)\delta(y)$, comb(x,y) can be written in either of two forms:

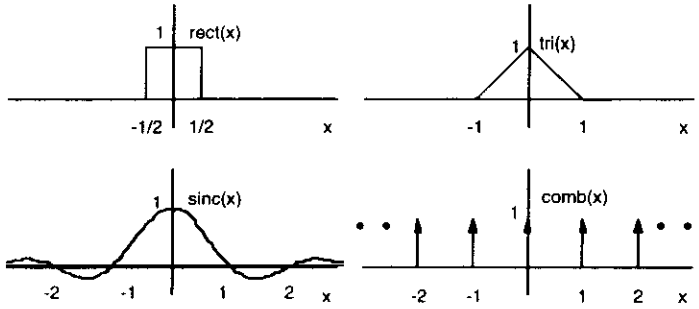



Fig. 2-2. Standard 1-D functions.

$$\text{comb}(x,y) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \delta(x-m, y-n) \tag{2.11a}$$

or

$$\text{comb}(x,y) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \delta(x-m)\delta(y-n). \tag{2.11b}$$

Although the same name may be applied to both the 1-D and the 2-D version of a particular function (e.g., *rect*), explicit use of variables removes any ambiguity as to whether the function is one- or two-dimensional. The functions rect(x,y), tri(x,y), sinc(x,y), and $\exp[-\pi(x^2+y^2)]$ are defined so as to have unit value at the origin and to have unit volume. The function $\cos[2\pi(u_o x + v_o y)]$, illustrated in Fig. 2-3 and also of great importance in the text, is not separable, but it can be written as the sum of two separable functions, $(1/2)\exp[j2\pi u_o x]\exp[j2\pi v_o y]$ and $(1/2)\exp[-j2\pi u_o x]\exp[-j2\pi v_o y]$. The function comb(x,y) has unusual scaling properties because of the behavior of the Dirac delta function under scaling: $\delta(ax) = (1/|a|)\delta(x)$. As a consequence, if it is desired to represent a regular array of unit volume impulses spaced by X and Y in the x and y directions, the function (1/|XY|)comb(x/X,y/Y) must be used, not simply comb(x/X,y/Y). This characteristic of the comb function is established in one of the homework problems. Line impulses, of the form $\delta(x)1(y)$ and $1(x)\delta(y)$, where the function 1(.) is unity for all values of its argument, are of use in representing slits in otherwise opaque masks. Their properties are explored in a homework problem.





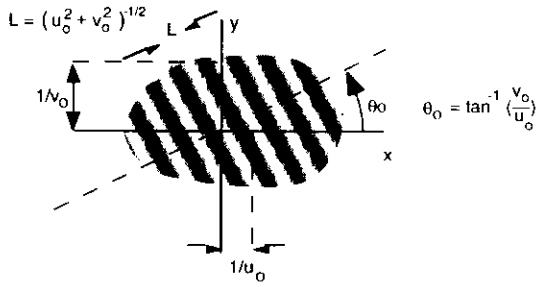



Fig. 2-3. A portion of the function $\cos[2\pi(u_o x + v_o y)]$ as seen from above the x-y plane. The bright lines represent maxima (crests) of the function, i.e., lines where the phase is an integer multiple of $2\pi$ radians. The dark lines represent the intermediate minima.

Two functions of considerable use in modeling circular openings such as apertures in an imaging system are the unit-diameter cylinder function and the unit-radius circ function, defined respectively by

$$\text{cyl}(r) = \begin{cases} 1 & r \le \frac{1}{2} \\ 0 & otherwise \end{cases}, \tag{2.12}$$

and

$$\text{circ}(r) = \begin{cases} 1 & r \le 1 \\ 0 & otherwise \end{cases}, \tag{2.13}$$

where $r$ is the radial coordinate given by

$$r = \sqrt{x^2 + y^2}. \tag{2.14}$$

These functions, which in graphical form look like cylindrical hat boxes, are illustrated in Fig. 2-4. Being functions of $r$ alone, they have circular symmetry. It is not really necessary to work with two different functions of this kind; one would suffice. However, both functions are in regular

use, and familiarity with them both is therefore desirable.[1] The cylinder function, being directly comparable in width to the 2-D rectangle function in that cyl(r) and rect(x,y) both have unit width, is preferable in cases when direct comparison of optical systems with square and circular apertures are to be made. The two functions are related by the equation $\mathrm{circ}(r) = \mathrm{cyl}(r/2)$. The 2-D Gaussian function, which can be written in the form $\exp[-\pi r^2]$, is also circularly symmetric. As noted earlier, most of the standard functions are defined so as to have unit central ordinate (value at the origin) and unit area or volume. The cylinder and circ functions are exceptions, being defined for unit diameter and radius, respectively, rather than unit volume.

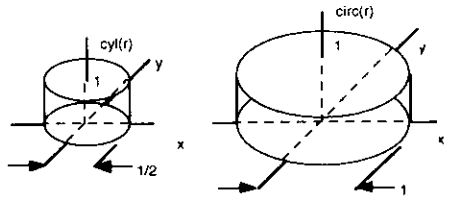



Fig. 2-4. The circularly symmetric functions cyl(r) and circ(r), where r is the radial coordinate. The cylinder function is defined to have unit diameter, whereas the circ function has unit radius.

Shifted and scaled versions of the basic functions are obtained by straightforward modification of their arguments. For example, a rectangular function of height 2 units, width in x of 2 units, and width in y of 3 units, and centered at coordinates x=4, y=5 is represented by

$$2\,\mathrm{rect}\left(\frac{x-4}{2}, \frac{y-5}{3}\right).$$

It is important to note that the arguments of all the special functions used in this text—rect(·,·), sinc(·,·), etc.—are dimensionless. Thus, in the above expression, (x-4)/2 must be dimensionless. In later chapters, variables x and y have dimensions of length, implying that all associated shift and scaling parameters must also have the dimensions of length. For example, in the function rect[(x-$x_0$)/W,(y-$y_0$)/H], if distances x and y are given in millimeters, quantities $x_0$, $y_0$, W, and H must be also.

An important operator used in this text is the rotation operator, denoted by $\mathcal{R}_{\theta_0}\{\ \}$. This operator produces a rotation through an angle $\theta_0$ of the 2-D function enclosed in the brackets. The positive direction of rotation is taken to be in the counter-clockwise direction, consistent with the convention applied to rotating phasors in engineering courses. The axis of rotation is always at the origin. Figure 2-5 shows some examples of application of the rotation operator. A simple analytical example is $\mathcal{R}_{90^\circ}\{\mathrm{rect}(x/2,y)\} = \mathrm{rect}(x,y/2)$. Note that care must be taken when rotation of a function is combined with a shift [e.g., Fig. 2-5(b)]. Formally the rotation operation can be defined by the equation

1 Bracewell, Ref. 2-1, uses the function rect(r) instead of cyl(r). The functions are essentially equal. However, the notations cyl(r) and circ(r) conjure up an image of a circularly symmetric function in a way that rect(r) may not.





$$g(x,y) = \mathcal{R}_{\theta_o}\{f(x,y)\} = f(x\cos\theta_o + y\sin\theta_o, y\cos\theta_o - x\sin\theta_o). \tag{2.15a}$$

However, except in numerical calculations, there is rarely a need to use this formula. A more compact definition can be given in polar-coordinate notation:

$$\mathcal{R}_{\theta_o}\{f_p(r,\theta)\} = f_p(r,\theta-\theta_o), \tag{2.15b}$$

where $f_p(.,.)$ denotes the function f(x,y) expressed in polar coordinates.

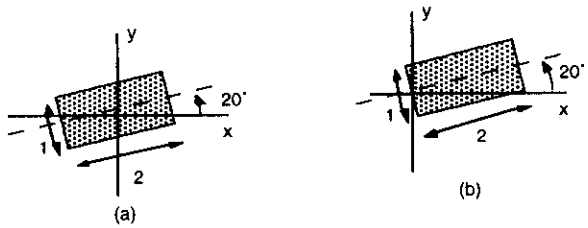



Fig. 2-5. Example of the rotation operation: (a) $\mathcal{R}_{20^\circ}\{\mathrm{rect}(\frac{x}{2},y)\}$, (b) $\mathcal{R}_{20^\circ}\{\mathrm{rect}(\frac{x-1}{2},y)\}$.

## 2.2 TWO-DIMENSIONAL CONVOLUTION AND CORRELATION

The 2-D convolution of functions f(x,y) and g(x,y) is defined by the integral expression

$$h(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\xi,\eta)g(x-\xi,y-\eta)\,d\xi\,d\eta. \tag{2.16}$$

Throughout the text a 2-D convolution operation will be denoted by a double asterisk:

$$h(x,y) = f(x,y)**g(x,y). \tag{2.17}$$

It is easily shown by substitution of variables that the convolution operation is commutative, that is,

$$f(x,y)**g(x,y) = g(x,y)**f(x,y). \tag{2.18}$$

The integral in Eq. (2.16) can be interpreted pictorially as follows. First, the function f(.,.) is sketched as a function of integration variables $\xi$ and $\eta$. The function g(.,.) is then rotated through 180° and positioned along the $\xi$ and $\eta$ axes with displacements along those axes in the amounts x and y. Thus, what was at the origin in a plot of $g(-\xi,-\eta)$ is positioned at coordinates $\xi$=x, $\eta$=y. The product of $f(\xi,\eta)$ with the rotated and shifted function $g(x-\xi,y-\eta)$ is calculated,

and the two-dimensional integral, or volume, of the product is evaluated. A particularly simple case occurs when f(x,y) and g(x,y) equal either 1 or 0, for then the product is either zero or unity and the integral equals the area of overlap of $f(\xi,\eta)$ and $g(x-\xi,y-\eta)$. An example of key steps in the convolution of two such one-zero binary functions is shown in Fig. 2-6. For the case where both f(x,y) and g(x,y) equal the unit rectangle, the result can be easily shown to be the unit triangle function: rect(x,y) ** rect(x,y) = tri(x,y).

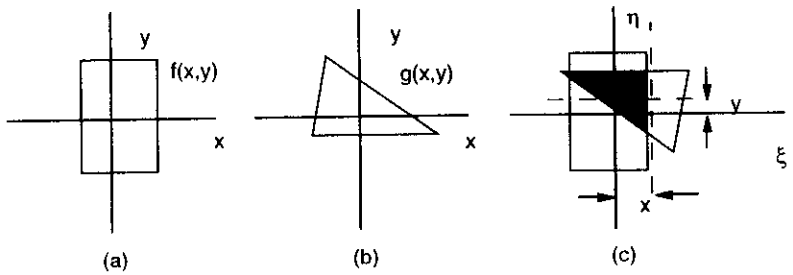



Fig. 2-6. Example of the convolution of two one-zero binary functions f(x,y) and g(x,y). The two functions are shown in (a) and (b) from above. The darker shaded region in (c) shows the area of overlap for the particular values of x and y chosen. Note that in (c) the function g(.,.) has been rotated through 180°.

An important convolution relationship is expressed by

$$f(x,y)**\delta(x-x_o,y-y_o)=f(x-x_o,y-y_o), \tag{2.19}$$

which says that a shift in the x-y plane of function f(x,y) can be represented by a convolution of f(x,y) with a shifted delta function. Mixed cartesian and polar notation can be used, as, e.g., in the expression $\mathrm{cyl}(r)**\delta(x-x_o,y-y_o)$ to represent a shifted cylinder function.

Two-dimensional correlation is similar to 2-D convolution, but there is no rotation of either of the functions, only a relative displacement, and one of the functions is conjugated. Thus, using the symbol ★★ to denote 2-D correlation,[2]

$$\begin{aligned} s(x,y) &= f(x,y)\star\star g(x,y) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\xi,\eta)g^*(\xi-x,\eta-y)\,d\xi\,d\eta \end{aligned} \tag{2.20}$$

Note that the correlation operation is not commutative. Rather, if f(x,y)★★g(x,y) = s(x,y), then g(x,y)★★f(x,y) = $s^*$(-x,-y).

---

2 There is, in fact, no universally accepted definition for the 2-D correlation integral (or the 1-D correlation integral, for that matter). Some authors include the conjugation in all cases, whereas others add it in when complex-valued functions are involved. Some shift the first function, others the second. The definition of Eq. (2.20) for correlation is attractive because of its simple relationship to convolution, as expressed in Eq. (2.23). Note that the definition in Eq. (2.23) is consistent with defining f★★g in terms of its Fourier transform: f★★g ↔ FG*.





Two useful relationships involving convolution and correlation are given by

$$[f(x,y)**g(x,y)]\star\star[f(x,y)**g(x,y)]=[f(x,y)\star\star f(x,y)]**[g(x,y)\star\star g(x,y)] \tag{2.21}$$

and

$$[f(x)g(y)]**[q(x)r(y)]=[f(x)*q(x)][g(y)*r(y)]. \tag{2.22}$$

In the latter equation, ** denotes a 2-D convolution, whereas * denotes a 1-D convolution. Equation (2.21) states, in loose terms, that the autocorrelation of a convolution can be written as the convolution of autocorrelations. Equation (2.22) says that the 2-D convolution of two separable functions can be evaluated by calculating simpler 1-D convolutions. It can also be shown that

$$\begin{aligned} f(x,y)\star\star g(x,y) &= f(x,y)**g^*(-x,-y) \\ &= f(x,y)**\mathcal{R}_{180^\circ}\{g^*(x,y)\} \end{aligned} \tag{2.23}$$

In words, the cross-correlation of f(x,y) with g(x,y) equals the convolution of f(x,y) with an inverted (rotated through 180°) and conjugated version of g(x,y).

## 2.3 THE TWO-DIMENSIONAL FOURIER TRANSFORM

The two-dimensional Fourier transform is a straight-forward extension of the one-dimensional Fourier transform (see Appendix A for a summary of basic 1-D Fourier transform relationships). In this text, the 2-D Fourier transform of function g(x,y), denoted G(u,v), is defined by

$$G(u,v)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\exp[-j2\pi(ux+vy)]\,dx\,dy. \tag{2.24}$$

The variables x and u must have reciprocal dimensions in order for their products to be dimensionless, and similarly for the variables y and v. Thus, if x and y have units of millimeters, variables u and v must have units of cycles per millimeter, or $\mathrm{mm}^{-1}$. In such a case, x and y are referred to as the *space domain* variables, and u and v as the *spatial frequency domain* variables, or simply the spatial frequency variables.

The integral expression in Eq. (2.24) is conveniently expressed using operator notation:

$$G(u,v)=\mathcal{F}\{g(x,y)\}. \tag{2.25}$$

The reciprocal relationship is given by

$$g(x,y)=\mathcal{F}^{-1}\{G(u,v)\}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(u,v)\exp[j2\pi(ux+vy)]\,du\,dv. \tag{2.26}$$

Frequently in this text the relationship between g(x,y) and G(u,v) is denoted by a two-headed arrow:

$$g(x,y)\leftrightarrow G(u,v).$$

In most cases upper-case characters will denote the Fourier transforms of functions denoted by lower-case characters; when necessary, a hat (^) will be used. Thus, F(u,v) is the Fourier transform of f(x,y), $\hat{U}(u,v)$ is the Fourier transform of U(x,y), etc.

Note that Eq. (2.26) represents the function g(x,y) in terms of a superposition of elementary components each of which has the basic form exp[j2π(ux+vy)]. Although such functions cannot be visualized directly, the corresponding real and imaginary parts cos[2π(ux+vy)] and sin[2π(ux+vy)] can be, these functions being suitably represented by surfaces with the appearance of corrugated roofs. Figure 2-3 gives some indication of how such a surface would appear from above. Only a central portion of the function is shown: cos[2π($u_o$x+$v_o$y)] extends, in fact, to infinity in all directions. The orientation of the corrugations, indicated by $\theta_o$ in the figure, is easily shown to be given by

$$\theta_o = \tan^{-1}\left(\frac{v_o}{u_o}\right). \tag{2.27}$$

The spatial period L, also shown, is given by

$$L = \frac{1}{\sqrt{u_o^2 + v_o^2}}. \tag{2.28}$$

The calculation of 2-D Fourier transforms can in many cases be difficult. Special cases where simplification occurs, such as when f(x,y) is separable in cartesian or polar coordinates, are treated in the following two sections.

### 2.3.1 Basic Relationships

A number of basic properties associated with the 2-D Fourier transform, sometimes referred to as theorems, are used throughout this text. These are listed in Table 2.1 without proof. With the exception of the rotation, skew-shift, and projection-slice theorems, which have no one-dimensional counterparts, the theorems are proved in essentially the same way the corresponding theorems for the 1-D Fourier transform are proved. The rotation theorem, formally written as

$$\mathcal{R}_{\theta_0}\{g(x,y)\} \leftrightarrow \mathcal{R}_{\theta_0}\{G(u,v)\}, \tag{2.29}$$

states that rotation through angle $\theta_0$ in the space domain corresponds to an identical rotation in the Fourier domain. Relationship 16 involving functions in polar coordinates is proved later in this chapter. The relationships $f^*(x,y) \leftrightarrow F^*(-u,-v)$ and $f^*(-x,-y) \leftrightarrow F^*(u,v)$ are easily remembered in verbal form: Conjugation in one domain corresponds to conjugation and inversion through the origin (rotation through $180^0$) in the other.

Table 2.2 lists a number of 2-D functions and their 2-D Fourier transforms. In this table and elsewhere in the text the function 1(·) equals unity, independent of the value of its argument. This function is used occasionally in the text as a placeholder to emphasize that a particular function is being considered in a 2-D framework, even though the function varies only in one direction.





## Table 2.1. Properties of the Two-Dimensional Fourier Transform

**1. Linearity**

$$\alpha g(x,y) + \beta h(x,y)\} \leftrightarrow \alpha G(u,v) + \beta H(u,v),$$

(a and b may be complex valued)

**2. Similarity (Scaling)**

$$g\left(\frac{x}{a},\frac{y}{b}\right) \leftrightarrow |ab|\, G(au,bv)$$

$$g(cx,dy) \leftrightarrow \frac{1}{|cd|} G\left(\frac{u}{c},\frac{v}{d}\right),$$

a, b, c and d real

**3. Shift**

$$g(x-x_o,y-y_o) \leftrightarrow G(u,v)\exp[-j2\pi(x_o u + y_o v)],$$

$x_o$ and $y_o$ real

**4. Combined Scaling and Shift**

$$g\left(\frac{x-x_o}{a},\frac{y-y_o}{b}\right) \leftrightarrow |ab|\, G(au,bv)\exp[-j2\pi(x_o u + y_o v)],$$

**5. Skew and Shift**

$$f(a_1x + b_1y + c_1, a_2x + b_2y + c_2)$$

$$\leftrightarrow \frac{1}{|D|}\exp[-j2\pi(x_o u + y_o v)]F\left(\frac{b_2}{D}u - \frac{a_2}{D}v, -\frac{b_1}{D}u + \frac{a_1}{D}v\right)$$

where

$$D = a_1b_2 - a_2b_1, \quad x_o = \frac{b_1c_1 - b_2c_2}{D}, \quad y_o = \frac{a_2c_1 - a_1c_2}{D}$$

**6. Modulation**

$$f(x,y)\exp[j2\pi(u_o x + v_o y)] \leftrightarrow F(u-u_o, v-v_o).$$

**7. Convolution**

$$f(x,y) ** g(x,y) \leftrightarrow F(u,v)G(u,v)$$
$$f(x,y)g(x,y) \leftrightarrow F(u,v) ** G(u,v)$$

**8. Derivative**

$$\frac{\partial}{\partial x} f(x,y) \longleftrightarrow j2\pi u F(u,v)$$

$$\frac{\partial}{\partial y} f(x,y) \longleftrightarrow j2\pi v F(u,v)$$

**9. Correlation**

$$f(x,y) \star\star g(x,y) \leftrightarrow F(u,v)G^*(u,v)$$

$$f(x,y)g^*(x,,y) \leftrightarrow F(u,v)\star\star G(u,v)$$

**10. Autocorrelation**

$$f(x,y)\star\star f(x,y) \leftrightarrow |F(u,v)|^2$$

$$|f(x,y)|^2 \leftrightarrow F(u,v)\star\star F(u,v)\ .$$

**11. Duality**

if $g(x,y) \leftrightarrow G(u,v)$, then

$G(x,y) \leftrightarrow g(-u,-v)$.

**12. Rayleigh's Theorem**

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|g(x,y)|^2\,dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|G(u,v)|^2\,dudv\,.$$

**13. Rotation**

$$\mathcal{R}_{\theta_o}\{g(x,y)\} \leftrightarrow \mathcal{R}_{\theta_o}\{G(u,v)\}.$$

**14. Separability**

$$f(x)g(y) \leftrightarrow F(u)G(v)$$

**15. Central Ordinate**

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}F(u,v)dudv = f(0,0)$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)dxdy = F(0,0)$$

**16. Circularly Symmetric Functions**

$$g_r(r) \leftrightarrow G_o(\rho)$$

where

$$G_o(\rho) = \mathcal{H}_o\{g_r(r)\} = 2\pi\int_0^{\infty} r\,g_r(r)J_o(2\pi r\rho)dr$$

$$r = \sqrt{x^2+y^2}, \quad \rho = \sqrt{u^2+v^2}$$

**17. Projection Slice Theorem**

$$f(x,y)\star\star\mathcal{R}_{\theta_o}\{\delta(x)1(y)\} \leftrightarrow F(u,v)\mathcal{R}_{\theta_o}\{1(u)\delta(v)\}.$$

**18. Symmetry Properties**

| If f(x,y) is | then F(u,v) is |
|---|---|
| real | hermitian* |
| hermitian* | real |
| even in x | even in u |
| even in y | even in v |
| odd in x | odd in u |
| odd in y | odd in v |
| imaginary | antihermitian* |
| antihermitian* | imaginary |





| | |
|---|---|
| circularly symmetric | circularly symmetric |

* A function g(x,y) is hermitian if Re{g(x,y)} is even and Im{g(x,y)} is odd. This condition implies that g*(x,y) = g(-x,-y) and that |g(x,y)| is even and arg{g(x,y)} is odd. The function g(x,y) is antihermitian if Re{g(x,y)} is odd and Im{g(x,y)} is even.

**19. Coordinate Reversal and Conjugation**

| | | | | | |
|---|---|---|---|---|---|
| f(-x,y) | ↔ | F(-u,v) | f(x,-y) | ↔ | F(u,-v) |
| f(-x,-y) | ↔ | F(-u,-v) | f*(x,y) | ↔ | F*(-u,-v) |
| f*(-x,-y) | ↔ | F*(u,v) | | | |

## Table 2.2. 2-D Fourier Transform Pairs

$$G(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\exp[-j2\pi(ux+vy)]dxdy$$

$$g(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(u,v)\exp[j2\pi(ux+vy)]dudv$$

| Function | 2-D Fourier Transform |
|---|---|
| g(x,y) | G(u,v) |
| rect(x,y) | sinc(u,v) |
| sinc(x,y) | rect(u,v) |
| tri(x,y) | $\text{sinc}^2(u,v)$ |
| $\text{sinc}^2(x,y)$ | tri(u,v) |
| $\exp[-\pi(x^2+y^2)]$ | $\exp[-\pi(u^2+v^2)]$ |
| $\exp(-\pi r^2)$ | $\exp(-\pi\rho^2)$ |
| $\exp\left[-\left(\frac{r}{a}\right)^2\right]$ | $\pi a^2 \exp\left[-(\pi a\rho)^2\right]$ |
| $\exp\left[-j\pi r^2\right]$ | $-j\exp\left[j\pi\rho^2\right]$ |
| $\exp\left[-\pi\left(\frac{r^2}{a+jc}\right)\right],\quad a\geq 0$ | $(a+jc)\exp\left[-\pi(a+jc)\rho^2\right]$ |
| $\exp\left[-\pi(a+jc)r^2\right],\quad a\geq 0$ | $\frac{1}{(a+jc)}\exp\left[-\pi\left(\frac{\rho^2}{a+jc}\right)\right]$ |
| δ(x,y) | 1 |
| δ(x)1(y) | 1(u)δ(v) |
| 1(x)δ(y) | δ(u)1(v) |
| 1 | δ(u,v) |
| δ(x-a,y-b) | exp[-j2π(au+bv)] |
| sgn(x)1(y) | $-\frac{j}{2\pi u}\delta(v)$ |





| | |
|---|---|
| $\exp[j2\pi(u_o x+v_o y)]$ | $\delta(u-u_o,v-v_o)$ |
| $\cos 2\pi(u_o x+v_o y)$ | $\frac{1}{2}[\delta(u+u_o,v+v_o)+\delta(u-u_o,v-v_o)]$ |
| $\sin 2\pi(u_o x+v_o y)$ | $\frac{j}{2}[\delta(u+u_o,v+v_o)-\delta(u-u_o,v-v_o)]$ |
| comb(x,y) | comb(u,v) |
| $\frac{1}{ab}\text{comb}(\frac{x}{a},\frac{y}{b})$ | $\frac{1}{ab}[\,ab\ \text{comb}(au,bv)\,]$ (a,b>0) |
| comb(x)δ(y) | comb(u)1(v) |
| comb(x)1(y) | comb(u)δ(v) |
| cyl(r) [diameter = 1] | $\left(\frac{\pi}{4}\right)\left[2\frac{J_1(\pi\rho)}{\pi\rho}\right]=\left(\frac{\pi}{4}\right)\text{somb}(\rho)$ |
| circ(r) [radius = 1] | $\pi\left[2\frac{J_1(2\pi\rho)}{2\pi\rho}\right]=\pi\,\text{somb}(2\rho)$ |
| $\text{cyl}\left(\frac{r}{d}\right)$ | $\left(\frac{d^2\pi}{4}\right)\left[2\frac{J_1(\pi d\rho)}{\pi d\rho}\right]=\left(\frac{d^2\pi}{4}\right)\text{somb}(d\rho)$ |
| $\text{circ}\left(\frac{r}{r_o}\right)$ | $r_o^2\pi\left[2\frac{J_1(2\pi r_o\rho)}{2\pi r_o\rho}\right]=(r_o^2\pi)\text{somb}(2r_o\rho)$ |

### 2.3.2 Transforming Functions Separable in Cartesian Coordinates

As noted earlier, many of the 2-D functions of interest to us in this text are separable in Cartesian coordinates. The Fourier transforms of such functions are easily found using 1-D Fourier theory because of the following relationship: If g(x,y) is separable in Cartesian coordinates, that is, if g(x,y) is of the form

$$g(x,y) = g_x(x)g_y(y),$$

then

$$\mathcal{F}\{g(x,y)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_x(x)g_y(y)\exp[-j2\pi(ux+vy)]dxdy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_x(x)g_y(y)\exp(-j2\pi ux)\exp(-j2\pi vy)\,dx\,dy$$

$$= \int_{-\infty}^{\infty} g_x(x)\exp(-j2\pi ux)dx \int_{-\infty}^{\infty} g_y(y)\exp(-j2\pi vy)dy$$

$$= \mathcal{F}_x\{g_x(x)\}\mathcal{F}_y\{g_y(y)\},$$

where $\mathcal{F}_x$ and $\mathcal{F}_y$ denote 1-D Fourier transform operations. In compact form,

$$\mathcal{F}\{g_x(x)g_y(y)\} = G_x(u)G_y(v). \tag{2.30}$$

This relationship says that the 2-D Fourier transform of separable function $g_x(x)g_y(y)$ can be obtained by finding the 1-D Fourier transforms of $g_x(x)$ and $g_y(y)$ independently and multiplying them. Note that separability in the space-, or (x,y), domain implies separability in the spatial frequency, or (u,v), domain. As a simple example, consider the 2-D Fourier transform of the 2-D function rect(x/2,y/3):

$$\mathcal{F}\left\{\text{rect}\left(\frac{x}{2},\frac{y}{3}\right)\right\} = \mathcal{F}\left\{\text{rect}\left(\frac{x}{2}\right)\text{rect}\left(\frac{y}{3}\right)\right\} = \mathcal{F}_x\left\{\text{rect}\left(\frac{x}{2}\right)\right\}\mathcal{F}_y\left\{\text{rect}\left(\frac{y}{3}\right)\right\}$$

Recalling from 1-D Fourier theory that $\mathcal{F}_x\{\text{rect}(x)\} = \text{sinc}(u)$, we obtain

$$\mathcal{F}\left\{\text{rect}\left(\frac{x}{2},\frac{y}{3}\right)\right\} = 2\,\text{sinc}(2u)3\,\text{sinc}(3v) = 6\,\text{sinc}(2u,3v).$$

Tables of 1-D Fourier transform pairs and properties of 1-D Fourier transforms are presented in appendix A for use in connection with 1-D transform evaluation.

### 2.3.3 Transforming Functions Separable in Polar Coordinates

Various functions of interest in optics, such as the cylinder function, are best expressed in terms of polar coordinates r and θ. As a general rule, the Fourier transforms of such functions are themselves also best expressed in polar coordinates. In this section the form of a polar-coordinate Fourier transform is established and simplifications are introduced for functions that are separable in r and θ.

Consider a 2-D space-domain distribution g and its 2-D Fourier transform G, both expressed in polar coordinates: $g(r,\theta) \leftrightarrow G(\rho,\phi)$. Making the following identifications between rectangular and polar coordinates,

$$r = \sqrt{x^2 + y^2} \qquad x = r\cos\theta \tag{2.31}$$

$$\theta = \tan^{-1}\left(\frac{y}{x}\right) \qquad y = r\sin\theta \tag{2.32}$$

$$\rho = \sqrt{u^2 + v^2} \qquad u = \rho\cos\phi \tag{2.33}$$





$$\phi = \tan^{-1}\left(\frac{v}{u}\right) \qquad v = \rho\sin\phi \tag{2.34}$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dxdy = \int_0^{2\pi}\int_0^{\infty} rdrd\theta \tag{2.35}$$

and making appropriate changes in the integral expression of Eq. (2.24), we obtain the following expression for $G(\rho,\phi)$:

$$G(\rho,\phi) = \int_0^{2\pi}\left\{\int_0^{\infty} rg(r,\theta)\exp\left[-j2\pi\rho r(\cos\theta\cos\phi + \sin\theta\sin\phi)\right]dr\right\}d\theta. \tag{2.36}$$

The integral expression is in general difficult to evaluate. However, if $g(r,\theta)$ is separable in polar coordinates, that is, if

$$g(r,\theta) = g_r(r)g_\theta(\theta),$$

simplifications result. First, substituting for $g(r,\theta)$ in Eq. (2.36), rearranging terms and introducing the trigonometric identity $\cos(\alpha-\beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta$ yields

$$\mathcal{F}\{g(r,\theta)\} = \mathcal{F}\{g_r(r)g_\theta(\theta)\} = \int_0^{\infty} rg_r(r)\left\{\int_0^{2\pi} g_\theta(\theta)\exp\left[-j2\pi\rho r\cos(\theta-\phi)\right]d\theta\right\}dr. \tag{2.37}$$

Further simplification is obtained by noting that $g_\theta(\theta)$ is periodic with period $2\pi$ and can therefore be written in the form of a Fourier series:

$$g_\theta(\theta) = \sum_{m=-\infty}^{\infty} c_m \exp(jm\theta). \tag{2.38}$$

It is thus only necessary to find the polar-coordinate transform for separable functions of the form $g_r(r)\exp(jm\theta)$, m integer, and exploit linearity. The result, obtained in problems 2.11 and 2.12, is the following:

$$\mathcal{F}\{g_r(r)g_\theta(\theta)\} = \sum_{m=-\infty}^{\infty}\mathcal{H}_m\{g_r(r)\}c_m(-j)^m\exp(jm\phi), \tag{2.39}$$

where $c_m$ is the normal Fourier series coefficient,

$$c_m = \frac{1}{2\pi}\int_0^{2\pi} g_\theta(\theta)\exp(-jm\theta)\,d\theta, \tag{2.40}$$

and where $\mathcal{H}_m\{g_r(r)\}$ is the *mth-order Hankel transform* of $g_r(r)$, given by

$$\mathcal{H}_m\{g_r(r)\} = 2\pi\int_0^\infty r g_r(r) J_m(2\pi r\rho)dr, \tag{2.41}$$

$J_m$ denoting the mth order Bessel function of the first kind. Note that separability of g in $(r,\theta)$ does *not* imply separability of its Fourier transform in $(\rho,\phi)$.

Of particular interest is the special case where $g(r,\theta)$ is circularly symmetric, that is, a function of r alone:

$$g(r,\theta) = g_r(r)1(\theta). \tag{2.42}$$

In that case, only the zeroth-order Hankel transform is needed, since all Fourier series coefficients in the summation of Eq. (2.39) equal zero except for $c_0$, and the Fourier transform operation reduces to

$$\mathcal{F}\{g_r(r)1(\theta)\} = G_0(\rho)1(\phi), \tag{2.43}$$

where $G_0(\rho)$ is defined by

$$G_o(\rho) = \mathcal{H}_o\{g_r(r)\} = 2\pi\int_0^\infty r g_r(r) J_o(2\pi r\rho)dr. \tag{2.44}$$

We see that the Fourier transform of a circularly symmetric function is itself circularly symmetric and can be found by evaluating a 1-D integral, the zeroth-order Hankel transform of $g_r(r)$. This form of the 2-D Fourier transform is sometimes referred to as the *Fourier-Bessel transform*. Note that for this special case of circularly symmetric functions there is a one-to-one correspondence between the 2-D Fourier transform and the zeroth-order Hankel transform:[3]

$$\mathcal{F} \sim \mathcal{H}_o. \tag{2.45}$$

Certain consequences immediately follow from this observation. For example, since a function that is circularly symmetric is also symmetric in x and y, it follows immediately that $\mathcal{H}_0 = \mathcal{H}_0^{-1}$ (since for a symmetric function $\mathcal{F} = \mathcal{F}^{-1}$), or

$$g_r(r) = 2\pi\int_0^\infty \rho G_o(\rho) J_o(2\pi r\rho)d\rho. \tag{2.46}$$

Furthermore, a direct application of the similarity (scaling) theorem leads to the relationship

[3] There is a subtile distinction in the ways the two operators $\mathcal{F}\{.\}$ and $\mathcal{H}_m\{.\}$ are used in that the operand of $\mathcal{F}$ exists in a two-dimensional space whereas the operand of $\mathcal{H}_k$ is strictly one-dimensional: it is appropriate to write $\mathcal{F}\{g_r(r)\}$ in the form $\mathcal{F}\{g_r\sqrt{x^2+y^2}\}$; such a substitution for r in $\mathcal{H}_k\{g_r(r)\}$ is questionable at best.





$$\mathcal{H}_o\{g_r(ar)\} = \frac{1}{a^2} G_o\left(\frac{\rho}{a}\right). \tag{2.47}$$

The proportionality factor $1/a^2$ on the right side of the equation results because both x and y are scaled by a in the distribution $g_r(ar)$. In general, for circularly symmetric functions, the zeroth-order Hankel transform can be thought of simply as a mathematical route to the 2-D Fourier transform, and any characteristics of the 2-D Fourier theory must carry over.

The zeroth-order Bessel function $J_o(\cdot)$ entering into Eq. (2.44) is plotted in Fig. 2-7 in the form $J_o(\pi\xi)$. Note that $J_o(\pi\xi)$ equals unity at the origin. Its zero crossings are aperiodic, the eight shown occurring where $\xi$ equals $\pm0.765$, $\pm1.757$, $\pm2.755$, and $\pm3.753$.

The circularly symmetric function of greatest use to us is the cylinder function. Use of the zeroth-order Hankel transform leads to the following expression for its 2-D Fourier transform:

$$\mathcal{F}\{\text{cyl}(r)\} = \mathcal{H}_0\{\text{cyl}(r)\} = 2\pi\int_0^{1/2} r J_o(2\pi r\rho)dr.$$

Applying the change of variables $r' = 2\pi r\rho$ and using the identity

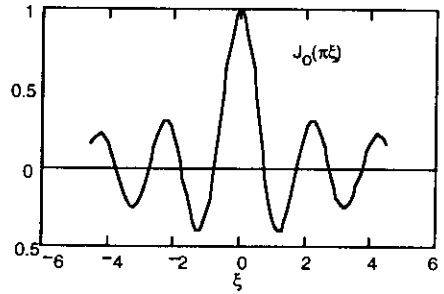



Fig. 2-7. Plot of $J_0(\pi\xi)$. The values of $\xi$ at zero crossings in the figure are $\pm0.765$, $\pm1.757$, $\pm2.755$, and $\pm3.753$.

$$\int_0^x \eta J_o(\eta)d\eta = x J_1(x), \tag{2.48}$$

we can write

$$\mathcal{H}_o\{\text{cyl}(r)\} = \frac{1}{2\pi\rho^2}\int_0^{\pi\rho} r' J_o(r')dr' = \frac{1}{2\rho} J_1(\pi\rho),$$

yielding

$$\mathcal{F}\{\text{cyl}(r)\} = \frac{\pi}{4}\left[2\frac{J_1(\pi\rho)}{\pi\rho}\right]. \tag{2.49}$$

This function, illustrated in Fig. 2-8, is circularly symmetric (as of course it should be, since it is the Fourier transform of a circularly symmetric function), consisting of a central lobe and a series of side lobes. The function in brackets, which gives the Fourier transform of the cylinder function its basic shape, is sufficiently important in Fourier optics that Gaskill (Ref. 2-5) has given it the name *sombrero function*, denoted somb(ρ):

$$\text{somb}(\rho) = 2\frac{J_1(\pi\rho)}{\pi\rho}. \tag{2.50}$$

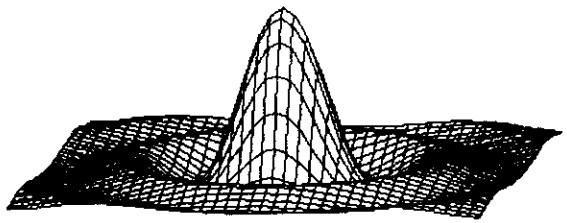


Fig. 2-8. The sombrero function *somb*(ρ), the two-dimensional Fourier transform of *cyl*(r).

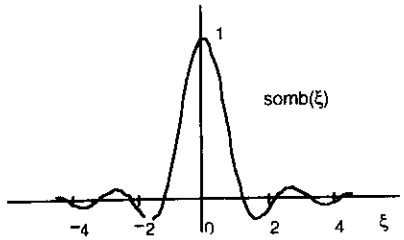



Fig. 2-9. Cross-section of the sombrero function. Zero crossings in the figure are at ξ equals ±1.220, ±2.233, ±3.238, and ±4.241.

The term *besinc* is also sometimes used in connection with this function, because it contains a Bessel function and because of its resemblance to the conventional sinc function. The sombrero function is shown in cross section in Fig. 2-9. Note that it is defined to have unit value at the origin. The values of ρ for which somb(ρ) is zero coincide with the non-periodic zeros of $J_1(\pi\rho)$. The first four occur at ρ = 1.220, 2.233, 3.238, and 4.241. An application of the similarity theorem leads to the following useful 2-D Fourier transform relationship for a cylinder function of arbitrary diameter d:





$$\mathcal{F}\left\{\text{cyl}\left(\frac{r}{d}\right)\right\} = \left(\frac{d^2\pi}{4}\right)\text{somb}(d\rho). \tag{2.51}$$

This function has its first null where dρ=1.220, or where ρ=1.220/d.

Occasionally it is convenient to use mixed polar and Cartesian notation in representing functions. For example, a pair of cylinder functions of unit diameter spaced by 4 units in the x-direction can be conveniently represented by the mixed-notation function

$$\text{cyl}(r)^{**}[\delta(x+2,y) + \delta(x-2,y)].$$

The Fourier transform of this function is also most easily expressed in mixed notation, having the form

$$\left[\frac{\pi}{4}\text{somb}(\rho)\right][2\cos(2\pi 2u)]$$

## 2.4 SAMPLING AND REPLICATION IN TWO DIMENSIONS

Sampling and replication in two dimensions is a relatively straightforward extension of 1-D sampling and replication (see, e.g., Ref. 2-1). Consider the 2-D sampling operation first. Assume that a band-limited function f(x,y) is to be sampled in two dimensions through multiplication by a sampling comb of unit-volume impulses spaced by X in the x-direction and Y in the y-direction. The result is the function

$$f_s(x,y) = f(x,y)\left[\frac{1}{XY}\text{comb}\left(\frac{x}{X},\frac{y}{Y}\right)\right]. \tag{2.52}$$

From basic definitions this function can be written in the form

$$f_s(x,y) = f(x,y)\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\delta(x-mX, y-nY)$$

$$= \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} f(mX,nY)\delta(x-mX, y-nY). \tag{2.53}$$

Fourier transforming both sides of Eq. (2.52) yields

$$F_s(u,v) = \frac{1}{XY}F(u,v)^{**}[XY\text{comb}(Xu,Yv)]$$

$$= \frac{1}{XY} F(u,v) ** \left[ \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \delta(u - \frac{m}{X}, v - \frac{n}{Y}) \right]$$

$$= \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{1}{XY} F(u - \frac{m}{X}, v - \frac{n}{Y}). \tag{2.54}$$

Equation (2.54), which shows $F_S(u,v)$ to consist of an infinite number of equally-spaced replicas of F(u,v), is valid for all functions f(x,y). In the special case when f(x,y) is band limited and the sample spacings X and Y are sufficiently small, the replicas of F(u,v) are non-overlapping. This is the case if, for example,

$$F(u,v) = 0, \quad \sqrt{u^2 + v^2} > B \tag{2.55}$$

and

$$\frac{1}{X} > 2B, \quad \frac{1}{Y} > 2B. \tag{2.56}$$

Figure 2-10 shows the appearance of the non-overlapping replicas, assuming that F(u,v) has a circular region of support of diameter 2B. Note that the replica centered at the origin is given by (1/XY)F(u,v) and that F(u,v) can therefore be obtained by multiplying $F_S(u,v)$ by (XY)rect(Xu,Yv):

$$F(u,v) = F_s(u,v)\left[ XY \,\mathrm{rect}(Xu, Yv) \right]. \tag{2.57}$$

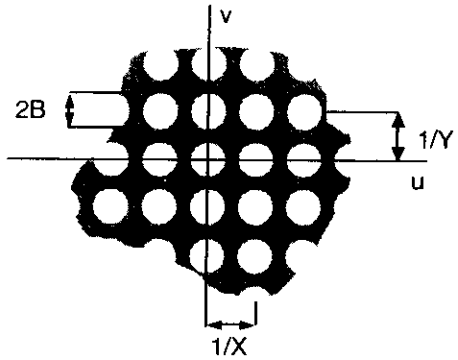



Fig. 2-10. Portion of the spectrum $F_S(u,v)$ of the sampled 2-D function $f_S(x,y)$. The white areas represent the regions of support of replicas of F(u,v).

The equivalent expressions in the space domain is the convolution





$$f(x,y) = f_s(x,y) ** \mathrm{sinc}\left( \frac{x}{X}, \frac{y}{Y} \right). \tag{2.58}$$

Replication by convolution of a function with a 2-D comb function can be used to represent functions that are periodic in two dimensions. For example, the function

$$g(x,y) = \mathrm{rect}(x, 2y) ** \frac{1}{6} \mathrm{comb}\left( \frac{x}{2}, \frac{y}{3} \right)$$

represents an infinite 2-D array of unit-amplitude rectangular pulses spaced by two units in the x-direction and by three units in the y-direction. Each pulse is one unit wide in x and one-half unit wide in y. A general periodic function with period (X,Y) is given by

$$f(x,y) = p(x,y) ** \left[ \frac{1}{XY} comb\left( \frac{x}{X}, \frac{y}{Y} \right) \right], \tag{2.59}$$

where p(x,y) is the base period function. The Fourier transform of f(x,y) is given by

$$F(u,v) = \frac{1}{XY} P(u,v)\left[ XY \,\mathrm{comb}(Xu, Yv) \right], \tag{2.60}$$

which can be written in the form

$$F(u,v) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{1}{XY} P\left( \frac{m}{X}, \frac{n}{Y} \right) \delta\left( u - \frac{m}{X}, v - \frac{n}{Y} \right). \tag{2.61}$$

The spectrum is seen to be impulsive, being zero for all frequencies except where u is an integer multiple of the fundamental period in the x-direction and v is an integer multiple of the fundamental period in the y-direction. Normally the region of support of p(x,y) will be smaller than X×Y, although it is not necessary that that condition be satisfied. A function representing an M×N array of replicas of a base function p(x,y) can be expressed, in the case where M and N are odd, in the form

$$g(x,y) = p(x,y) ** \left[ \frac{1}{XY} \mathrm{comb}\left( \frac{x}{X}, \frac{y}{Y} \right) \mathrm{rect}\left( \frac{x}{MX}, \frac{y}{NY} \right) \right]. \tag{2.62}$$

Note how the rectangle function truncates the replication array. A homework problem treats the case of even M and/or N.

## 2.5 FRESNEL TRANSFORMS*

Convolution with quadratic phase factors is important in the mathematical description of the propagation of electromagnetic wave fields. In this section, properties associated with such convolutions are presented, and a particularly important relationship, referred to in this text as the *Fresnel transform*, is introduced.[4]

[4] The Fresnel transform is named after French engineer Augustin Fresnel, who contributed significantly to the development of diffraction theory. There is, in fact, no standard definition of the Fresnel transform. The

In one dimension the Fresnel transform $\hat{f}_\alpha(x)$ of a function f(x) is defined by the convolution

$$\hat{f}_\alpha(x) = f(x) * \frac{1}{\sqrt{j\alpha}} \exp\left(j\frac{\pi}{\alpha}x^2\right). \tag{2.63}$$

Equation (2.63) can be expressed in more compact form by

$$\hat{f}_\alpha(x) = f(x) * h_\alpha(x), \tag{2.64}$$

where the convolution kernel $h_\alpha(x)$ is given by

$$h_\alpha(x) = \frac{1}{\sqrt{j\alpha}} \exp\left(j\frac{\pi}{\alpha}x^2\right). \tag{2.65}$$

The parameter $\alpha$ is restricted to be real-valued. Note that the Fresnel transform of f(x), unlike the Fourier transform, is itself a function of x, i.e., there is no change of domains. Note also that the Fresnel transform produces an entire family of functions with $\alpha$ as a parameter. In two dimensions the Fresnel transform is defined by

$$\hat{f}_\alpha(x,y) = f(x,y) ** h_\alpha(x,y), \tag{2.66}$$

where $h_\alpha(x,y)$ is given by

$$h_\alpha(x,y) = \frac{1}{j\alpha} \exp\left[j\frac{\pi}{\alpha}\left(x^2+y^2\right)\right]. \tag{2.67}$$

The functions $h_\alpha(x)$ and $h_\alpha(x,y)$ have impulsive autocorrelation properties. In particular it can be shown that

$$\begin{aligned} h_\alpha(x) \star h_\alpha(x) &= \frac{1}{\sqrt{j\alpha}} \exp\left(\frac{\pi}{\alpha}x^2\right) * \frac{1}{\sqrt{-j\alpha}} \exp\left(-j\frac{\pi}{\alpha}x^2\right) \\ &= \delta(x) \end{aligned} \tag{2.68}$$

$$\begin{aligned} h_\alpha(x,y) \star\star h_\alpha(x,y) &= \frac{1}{j\alpha} \exp\left[j\frac{\pi}{\alpha}\left(x^2+y^2\right)\right] ** \frac{1}{-j\alpha} \exp\left[-j\frac{\pi}{\alpha}\left(x^2+y^2\right)\right] \\ &= \delta(x,y) \end{aligned} \tag{2.69}$$

Using the above identities one can easily establish the following inverse transform relationships:

---

definition used here is particularly convenient in connection with wave propagation modeling, as shall be seen in Chapt. 4.





$$f(x) = \hat{f}_\alpha(x) * h_\alpha^*(x), \tag{2.70}$$

$$f(x,y) = \hat{f}_\alpha(x,y) ** h_\alpha^*(x,y). \tag{2.71}$$

The convolution relationships of Eqs. (2.64) and (2.66) can be expressed in the Fourier domain by

$$\hat{F}_\alpha(u) = F(u) H_\alpha(u) \tag{2.72}$$

$$\hat{F}_\alpha(u,v) = F(u,v) H_\alpha(u,v) \tag{2.73}$$

where $H_\alpha(u)$ and $H_\alpha(u,v)$ are the 1-D and 2-D Fourier transforms of, respectively, $h_\alpha(x)$ and $h_\alpha(x,y)$. Using the relationship

$$\begin{aligned} \mathcal{F}_x\left\{\exp(-j\pi x^2)\right\} &= \exp(j\pi u^2)\frac{1}{\sqrt{2}}(1-j) \\ &= \exp(-j\frac{\pi}{4})\exp(j\pi u^2) \end{aligned} \tag{2.74}$$

which is established in a homework problem, it is possible to show that

$$H_\alpha(u) = \mathcal{F}\left\{\frac{1}{\sqrt{j\alpha}} \exp\left(j\frac{\pi}{\alpha}x^2\right)\right\} = \exp\left(-j\pi\alpha u^2\right) \tag{2.75}$$

and

$$H_\alpha(u,v) = \mathcal{F}\left\{\frac{1}{j\alpha} \exp\left[j\frac{\pi}{\alpha}\left(x^2+y^2\right)\right]\right\} = \exp\left[-j\pi\alpha\left(u^2+v^2\right)\right]. \tag{2.76}$$

These two equations indicate that the Fresnel transform changes the phases but not the magnitudes of the frequency components of f(x) or f(x,y).

Two particularly interesting relationships involving quadratic phase factors are stated as follows:

$$\left\{\left[f(x)\exp(-j\pi x^2)\right] * \exp(j\pi x^2)\right\}\exp(-j\pi x^2) = F(x), \tag{2.77}$$

$$\left\{\left[f(x) * \exp(j\pi x^2)\right]\exp(-j\pi x^2)\right\} * \exp(j\pi x^2) = \exp\left(j\frac{\pi}{4}\right)F(x), \tag{2.78}$$

where f(.) and F(.) form a Fourier transform pair. These two equations state that the Fourier transform can be evaluated through a combination of multiplication by a quadratic phase factor and

convolution with a quadratic phase factor.[5] The proof of Eq. (2.77) is simple; Eq. (2.78) is proved with somewhat more effort. It will be seen in Chap. 5 that these relationships have counterparts in specific optical systems.

## 2.6 TWO-DIMENSIONAL LINEAR SYSTEMS

It is assumed in this text that the reader is familiar with basic concepts of linear and linear shift-invariant systems. Such concepts are introduced for one-dimensional signals in a number of undergraduate electrical engineering texts. In this section the characteristics of two-dimensional linear and linear shift-invariant systems are briefly summarized.

The input-output relationship for a 2-D linear system is characterized mathematically by a 2-D superposition integral:

$$f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(\xi,\eta)h(x,y;\xi,\eta)\,d\xi\,d\eta, \tag{2.79}$$

where g(x,y) is the system input, f(x,y) is the system output, and where h(x,y;ξ,η) is the response at coordinates (x,y) to a unit 2-D impulse applied at coordinates (ξ,η), i.e., the response to δ(x-ξ,y-η). In the special case of a *shift-invariant* system the impulse response h depends only on the coordinate differences (x-ξ) and (y-η), and the superposition integral assumes the form of a convolution:

$$f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(\xi,\eta)h(x-\xi,y-\eta)\,d\xi\,d\eta, \tag{2.80a}$$

or

$$f(x,y) = g(x,y)**h(x,y). \tag{2.80b}$$

Shift invariance implies that a shift in the input results in a corresponding shift—but no other change—in the output distribution. Thus, for a system described by Eq. (2.80), if input g(x,y) produces output f(x,y), then input $g(x-x_0,y-y_0)$ produces output $f(x-x_0,y-y_0)$. If x and y represent spatial coordinates, as they do in this text, systems that are characterized by a convolutional input-output relationship are referred to as being *space invariant*. If the system requires the more general description of Eq. (2.79), it is referred to as being *space variant*.

Space invariant systems assume a particularly simple representation in the spatial frequency domain. Taking the Fourier transform of both sides of Eq. (2.80) yields

$$F(u,v) = G(u,v)H(u,v). \tag{2.81}$$

[5] Because of the presence of the quadratic phase factors, which are also known as linear-FM functions or "chirps," these expressions are sometimes referred to as chirp-algorithm implementations of the Fourier transform operation. In discrete systems the terminology chirp-Z transform is used.





The function H(u,v) in this equation is called the *transfer function* of the system. Usually complex-valued, it specifies how the magnitude and phase of spatial frequency components of the image are modified by the system to produce the output.

## 2.7 REFERENCES


2-1 R. N. Bracewell, *The Fourier Transform and its Applications*, 2nd ed., McGraw-Hill Book Company, New York (1978).

2-2 A. Papoulis, *The Fourier Integral and its Applications*, McGraw-Hill Book Company, New York (1962).

2-3 J. W. Goodman, *Introduction to Fourier Optics*, 2nd ed., McGraw-Hill, New York (1996), Chapt. 2.

2-4 A. Papoulis, *Systems and Transforms with Applications to Optics*, McGraw-Hill Book Company, New York (1968).

2-5 J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics*, John Wiley and Sons, New York (1978).

2-6 P. M. Duffieux, *The Fourier Transform and its Applications to Optics*, Second Edition (Wiley, New York, 1993), available through University Microfilm International, Ann Arbor, Michigan.

## 2.8 PROBLEMS

2.1 The functions illustrated in Fig. P2.1 equal unity in the shaded regions, zero elsewhere. The grid spacing is unity. Represent the functions in terms of the standard 2-D functions used in this chapter and find their Fourier transforms. The rectangle functions in (a), (d), and (e) all have the same size. The rectangle in (e) is rotated from the horizontal by 20°. Represent the function in part (f) by convolving an appropriate rectangle function with a truncated comb function, as in Eq. (2.62).

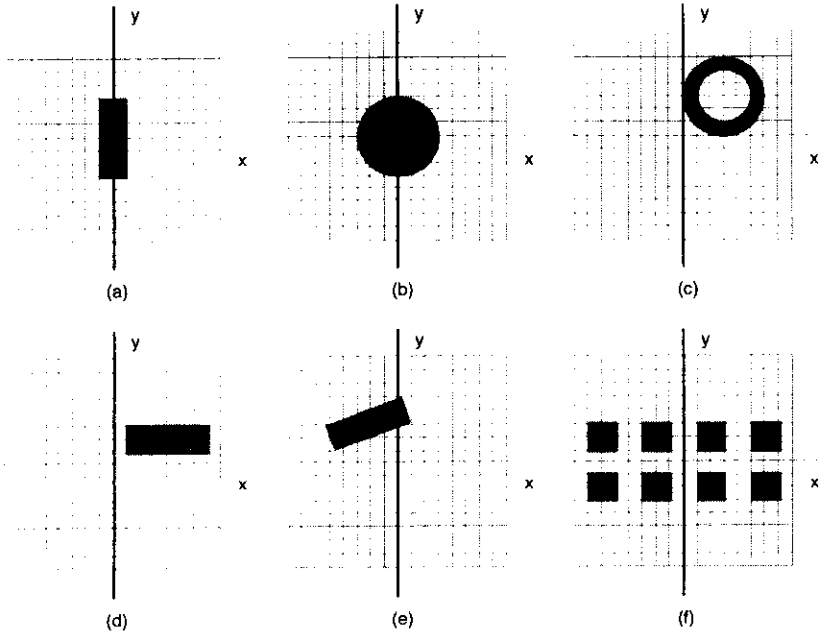



Fig. P2.1.

2.2 Using the format of Fig. P2.1, sketch the following two functions for the case $\theta = 90^{\circ}$. Note the brackets: these two functions are not the same.

(a) $\mathcal{R}_{90^{\circ}}\left\{\text{rect}\left(\frac{x}{4}, y\right) ** \delta(x-4, y)\right\}$

(b) $\mathcal{R}_{90^{\circ}}\left\{\text{rect}\left(\frac{x}{4}, y\right)\right\} ** \delta(x-4, y)$





2.3 By appealing to corresponding properties for 1-D delta functions, establish the following properties for 2-D delta function

(a) $\delta(ax,by) = \frac{1}{|ab|}\delta(x,y)$ [start with the identity $\delta(ax) = (1/|a|)\delta(x)$]

(b) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta(x,y)\,dx\,dy = 1$

(c) $\frac{1}{XY}\text{comb}\left(\frac{x}{X},\frac{y}{Y}\right) = \sum_{m=-\infty}^{\infty}\delta(x-mX)\sum_{n=-\infty}^{\infty}\delta(y-nY) = \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\delta(x-mX,y-nY)$

i.e., that the function (1/XY)comb(x/X,y/Y) represents a 2-D array of *unit volume* impulses spaced by X and Y.

2.4 Noting that $f(x,y)\delta(x - x_o,y - y_o) = f(x_o,y_o)\delta(x - x_o,y - y_o)$, find the value of the constant $A_o$ in the equation

$$\text{tri}(x,y)\delta\left(x-\frac{1}{4},y-\frac{1}{4}\right) = A_o\,\delta\left(x-\frac{1}{4},y-\frac{1}{4}\right).$$

2.5 Show by integration in polar coordinates that $(\pi r)^{-1}\delta(r)$ is an impulse at the origin having unit volume, i.e., that $(\pi r)^{-1}\delta(r)$ has the same properties as $\delta(x,y)$.

2.6 Show that the line impulse $\delta(x)1(y)$ has unit volume per unit length.

2.7 The rotation operator can be defined in polar coordinates by $\mathcal{R}_{\theta_o}\{g(r,\theta)\} = g(r,\theta - \theta_o)$. In Cartesian coordinates the definition has the form $\mathcal{R}_{\theta_o}\{f(x,y)\} = f(x\cos\theta_o + y\sin\theta_o, y\cos\theta_o - x\sin\theta_o)$. Test the validity of this latter definition by proving that $\mathcal{R}_{\theta_o}\{\delta(x-1,y)\}$ can be written in the following forms:

(a) $\delta(x-1,y)$ *for* $\theta_o = 0^{\circ}$

(b) $\delta(x,y-1)$ *for* $\theta_o = 90^{\circ}$

(c) $\delta(x+1,y)$ *for* $\theta_o = 180^{\circ}$

(d) $\delta(x,y+1)$ *for* $\theta_o = 270^{\circ}$

2.8 Using the rotation operator, find analytical representations for the two diamond-shaped one-zero functions sketched in Fig. P2.6.

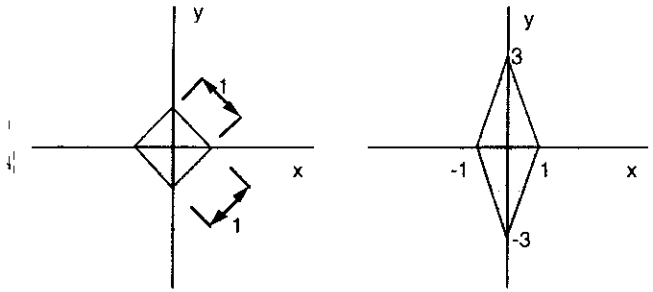



Fig. P2.6

2.9 Prove by setting up the corresponding convolution integrals that the expression $f(x,y) * g(x) * h(x)$ is equivalent to the following two expressions, where $*$ denotes 1-D convolution and where $**$ denotes 2-D convolution:

(a) $f(x,y) * *[g(x)h(y)]$

(b) $f(x,y) * *[g(x)1(y)] * *[1(x)h(y)]$.

2.10 Find the 2-D Fourier transforms of the following functions:

(a) $f(x,y) = \text{rect}\left(\frac{x-2}{3}, \frac{y+1}{4}\right)$

(b) $g(x,y) = \{\text{rect}(2x)\exp[i2\pi 3x]\}1(y)$

(c) $h(x,y) = \text{rect}(x,y) * *\text{tri}(2x,3y)$

(d) $q(x,y) = [\text{rect}(x)\text{tri}(y)] * *[\text{tri}(2x)\text{rect}(3y)]$

(e) $r(x,y) = [\text{rect}(x)\text{tri}(y)] * *[\text{cyl}(r)1(\theta)]$

(f) $s(x,y) = \text{rect}\left(\frac{x}{2}, y\right) ** \text{rect}\left(\frac{x}{2}, y\right)$

2.11 The function $f(x,y)$ is given by $f(x,y) = \text{rect}\left(\frac{x}{2}, y\right)[1 + \cos(2\pi 5x)]$.

(a) Find its 2-D Fourier transform.

(b) Sketch the 1-D cross-section functions f(x,0) and F(u,0).

2.12 Find and sketch the Fourier transforms of the following functions (use dotted lines to denote imaginary components):

(a) f(x,y) = exp[j2π(x+2y)].





(b) g(x,y) = cos[2π(x+2y)].

(c) h(x,y) = sin[2π(x+2y)].

$F(u,v) = \delta(u-1, v-2)$

$G(u,v) = \dfrac{\delta(u+1, v+2) + \delta(u-1, v-2)}{2}$

$H(u,v) = i\dfrac{\delta(u+1, v+2) - \delta(u-1, v-2)}{2}$

2.13 Use Fourier-domain reasoning to show that

$$[f(x)g(y)] ** [q(x)r(y)] = [f(x) * q(x)][g(y) * r(y)],$$

where $**$ denotes a 2-D convolution and $*$ denotes a 1-D convolution.

2.14 Use the result of Problem 2.13 to determine the following 2-D convolutions.

(a) $f(x,y) = \text{rect}\left(\frac{x}{2}, y\right) * *\text{rect}\left(\frac{x}{2}, y\right)$.

(b) $g(x,y) = \text{rect}\left(\frac{x}{4}, \frac{y}{4}\right) * *\text{rect}(x,y)$.

(c) $h(x,y) = \text{rect}(x+2, y) * *\text{rect}(x-2, y)$.

(d) $\text{rect}(x,y) * *\text{rect}\left(\frac{x-3}{2}, \frac{y}{2}\right)$.

2.15 Show that if f(x,y) ** g(x,y) = h(x,y), then f(ax,ay) ** g(ax,ay) = (1/|a|²)h(ax,ay):

(a) by direct substitution in Eq. (2.16).

(b) by Fourier domain reasoning.

2.16 Sketch the function $f(x,y) = \mathcal{R}_{20°}\left\{\text{rect}\left(\frac{x}{2}, y\right) * *\delta(x-2, y)\right\}$ and find its Fourier transform.

2.17 Use the *shift theorem* to find the 2-D Fourier transform of the function

$$f(x,y) = \cos[2\pi(\alpha x + \beta y) + \theta_o].$$

Compare with the result you obtain applying Euler's equation.

2.18 Use the zero-order Hankel transform to find the 2-D Fourier transform of the function

$$g(r,\theta) = \delta(r - r_o)1(\theta).$$

Note that this function can be used to represent a narrow annulus.

2.19 Starting with the equality that exists for circularly symmetric functions between the zero-order Hankel transform and the 2-D Fourier transform, show that the 2-D Fourier transform of the one-zero function f(x,y) sketched in Fig. P2.16 equals

$$F(u,v) = 2\frac{J_1\left(2\pi\sqrt{(2u^2)+v^2}\right)}{\sqrt{(2u^2)+v^2}}.$$

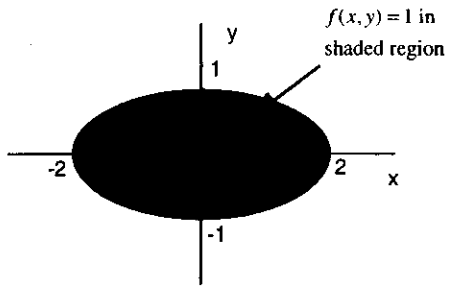



Fig. P2.16

2.20 Use Fourier-transform-domain reasoning to prove the relationship (f ** g)★★(f ** g) = (f★★f) ** (g★★g), and use this relationship to show that the autocorrelation of the function rect(x,y) ** [δ(x+2,y) + δ(x-2,y)] equals tri(x,y) ** [δ(x+4,y) + 2δ(x,y) + δ(x-4,y)]. Sketch this latter function in a perspective drawing.

2.21 Using the relationship

$$\exp(j\alpha\sin x) = \sum_{m=-\infty}^{\infty} J_m(\alpha)\exp(jmx),$$

show that if g(r,θ) is of the form

$$g(r,\theta) = g_r(r)\exp[jm\theta], \quad m \text{ integer}$$

then

$$\mathcal{F}\{g(r,\theta)\} = (-j)^m \exp[jm\phi]\,\mathcal{H}_m\{g_r(r)\},$$

where $\mathcal{H}_m\{\ \}$ denotes a Hankel transform of order m:

$$\mathcal{H}_m\{g_r(r)\} = 2\pi\int_0^{\infty} r\,g_r(r)J_m(2\pi r\rho)\,dr.$$

2.22 Show by expanding $g_\theta(\theta)$ in a Fourier series and using the results of problem 2.18 that

$$\mathcal{F}\{g_r(r)g_\theta(\theta)\} = \sum_{k=-\infty}^{\infty}(-j)^k c_k \exp(jk\phi)\mathcal{H}_k\{g_r(r)\},$$

where

$$c_k = \frac{1}{2\pi}\int_{-\infty}^{\infty} g_\theta(\theta)\exp(-jk\theta)\,d\theta.$$





2.23 Find the constant d for which the function $g(r)1(\theta) = \mathrm{cyl}(r/d)1(\theta)$ has both unit central ordinate and unit volume and find the 2-D Fourier transform of g(r)1(θ) for that case.

2.24 Sketch the following distributions as viewed from above the x-y plane. Use shading to indicate where the functions have values of zero and unity.

(a) $\mathrm{rect}(x,y) ** \left[\frac{1}{4}\mathrm{comb}\left(\frac{x}{2},\frac{y}{2}\right)\mathrm{rect}\left(\frac{x}{5},\frac{y}{9}\right)\right]$.

(b) $[\mathrm{cyl}(r) ** \mathrm{comb}(x,y)]\mathrm{rect}\left(\frac{x}{5},\frac{y}{3}\right)$.

2.25 The *x-axis projection* of f(x,y) is given by

$$p_f(x) = \int_{-\infty}^{\infty} f(x,y)\,dy.$$

(a) Show that $p_f(x)$ can be expressed in terms of a convolution of f(x,y) with line impulse δ(x)1(y) as

$$p_f(x) = [f(x,y) ** \delta(x)1(y)]\Big|_{y=0}.$$

(b) Evaluate $p_f(x)$ for the following functions:

(i) $\mathrm{rect}(x+2,y) + \mathrm{rect}\left(x-2,\frac{y-1}{2}\right)$

(ii) $\mathrm{tri}(x,y)$

(iii) $\mathrm{circ}(r)$

2.26 A generalization of the projection operation of the previous problem is of the form

$$p_{f_\theta}(x) = \left[\mathcal{R}_\theta\{f(x,y)\} ** \delta(x)1(y)\right]\Big|_{y=0}.$$

It is easy to show that $p_{f\theta}(x)$ is the same distribution that would be obtained by projecting f(x,y) onto an axis that makes an angle θ with the +x axis.

(a) Show that

$$\mathcal{F}_x\left\{p_{f_\theta}(x)\right\} = \mathcal{R}_\theta\{F(u,v)\}\Big|_{v=0}.$$

(b) Sketch $p_{f\theta}(x)$ with θ=45° for the following:

(i) $\mathrm{rect}(x,y)$

(ii) $\mathrm{rect}\left(\frac{x}{2}, y\right)$

(iii) $\mathrm{cyl}(r)\mathrm{l}(\theta)$

2.27 Show that $\alpha \exp[j\pi(\alpha x)^2] * \alpha \exp[-j\pi(\alpha x)^2] = \delta(x)$.

2.28 Use the relationship

$$\int_{-\infty}^{\infty} \cos \pi\xi^2 d\xi = \int_{-\infty}^{\infty} \sin \pi\xi^2 d\xi = 1/\sqrt{2}$$

to show that

$$\mathcal{F}_x^{-1}\left\{\exp(-j\pi u^2)\right\} = \exp(j\pi x^2)\frac{1}{\sqrt{2}}(1-j)$$

$$= \exp(-j\frac{\pi}{4})\exp(j\pi x^2)$$

and thereby establish the validity of Eq. (2.74).

2.29 Given that $\mathcal{F}_x\{\exp(j\pi x^2)\} = \exp(j\pi/4)\exp(-j\pi u^2)$, find the Fourier transform of $\cos\pi x^2$ and $\sin\pi x^2$.

2.30 Prove Eq. (2.77).

2.31 Sketch the function f(x,y) = rect(x-y,y) and show that its Fourier transform equals sinc(u,v+u).





## 2.9 INDEX - CHAPTER 2

# Appendix A. One-Dimensional Fourier Transforms and Relationships

## A.1 DEFINITIONS

**Fourier Transform and Inverse Transform**

$$F(u) = \int_{-\infty}^{\infty} f(x)e^{-j2\pi ux}\,dx \qquad f(u) = \int_{-\infty}^{\infty} F(u)e^{j2\pi ux}\,dx$$

**Convolution**

$$f(x) * g(x) = \int_{-\infty}^{\infty} f(u)g(x-u)du$$

**Correlation**

$$f(x) \star g(x) = f(x) * g^{*}(-x)$$

$$f(x) \star g(x) = \int_{-\infty}^{\infty} f(u)g^{*}(u-x)du$$

**Comb Function**

$$\sum_{m=-\infty}^{\infty} \delta(x-mT) = \frac{1}{T}\,\mathrm{comb}\left(\frac{x}{T}\right)$$

**Hermitian Functions**

Hermitian: (real,even) + (imaginary,odd)
Antihermitian: (real,odd) + (imaginary,even)
(for both, the magnitude is even, the phase odd)





## A.2 TRANSFORM RELATIONSHIPS

**Linearity**

$$f(x) + g(x) \rightarrow F(u) + G(u)$$

**Scaling**

$$f\left(\frac{x}{b}\right) \rightarrow |b|\,F(bu)$$

**Shift**

$$f(x-a) \rightarrow F(u)e^{-j2\pi au}$$

**Scale and Shift**

$$f\left(\frac{x-a}{b}\right) \rightarrow |b|F(bu)e^{-j2\pi au}$$

**Modulation**

$$f(x)e^{j2\pi ax} \rightarrow F(u-a)$$

$$f(x)\cos 2\pi ax \rightarrow \frac{1}{2}F(u+a) + \frac{1}{2}F(u-a)$$

**Modulation and Shift**

$$f(x-a)e^{j2\pi bx} \rightarrow [F(u)e^{-j2\pi au}] * \delta(u-b) = F(u-b)e^{-j2\pi a(u-b)}$$

**Modulation and Scale**

$$|b|f(b\,x)e^{-j2\pi ax} \rightarrow F\left(\frac{u+a}{b}\right)$$

**Convolution and Correlation**

$$f(x)g(x) \rightarrow F(u) * G(u)$$

$$f(x) * g(x) \rightarrow F(u)G(u)$$

$$f(x) \star g(x) \rightarrow F(u)G^{*}(u)$$

$$f(x) \star f(x) \rightarrow |F(u)|^2$$

**Derivative**

$$f'(x) \rightarrow j2\pi uF(u)$$

$$-j2\pi xf(x) \rightarrow F'(u)$$

$$f'(ax) \rightarrow \frac{1}{|a|}j2\pi\frac{u}{a}F\left(\frac{u}{a}\right)$$

**Reversal and Conjugation**

$f(-x) \rightarrow F(-u)$ (reversal → reversal)

$f^{*}(x) \rightarrow F^{*}(-u)$ (conjugation → conjugation & reversal)

$f^{*}(-x) \rightarrow F^{*}(u)$ (conjugation & reversal → conjugtion)

**Symmetry**

even → even

odd → odd

(real,even) → (real,even)

(real,odd) → (imaginary,odd)

real → Hermitian

imaginary → antihermitian

**Central Ordinate**

$$\int_{-\infty}^{\infty} f(x)dx = F(0) \quad \text{and} \quad \int_{-\infty}^{\infty} F(s)ds = f(0)$$

**Rayleigh's Theorem**

$$\int_{-\infty}^{\infty} |f(x)|^2 dx = \int_{-\infty}^{\infty} |F(s)|^2 ds$$

**Power Theorem**

$$\int_{-\infty}^{\infty} f(x)g^*(x)dx = \int_{-\infty}^{\infty} F(s)G^*(s)ds$$





## A.3 TRANSFORM PAIRS:

rect(x) → sinc(u)

tri(x) → $\text{sinc}^2(u)$

sgn(x) → $-\frac{j}{\pi u}$

step(x) → $\frac{1}{2}\delta(u) - \frac{j}{2\pi u}$

$e^{-\pi x^2} \rightarrow e^{-\pi u^2}$

$e^{-j\pi x^2} \rightarrow e^{-j\pi/4}\, e^{j\pi u^2}$

comb(x) → comb(u)

δ(x) → 1(u) = 1

δ(x-a) → $e^{-j2\pi au}$

$e^{+j2\pi ax} \rightarrow \delta(u-a)$

$\frac{j}{\pi x} \rightarrow \text{sgn}(u)$

$\exp(j2\pi u_o x) \rightarrow \delta(u-u_o)$

$\cos(2\pi ax) \rightarrow \frac{1}{2}[\delta(u+a) + \delta(u-a)]$

$\sin(2\pi ax) \rightarrow \frac{j}{2}[\delta(u+a) - \delta(u-a)]$

# 3. Optical Wave Fields and Their Representation

In this chapter complex-valued phasor representations for monochromatic waves are introduced, including those for spherical and plane waves, and the relationship between the complex wave amplitude and the optical intensity is presented. The interference of waves is discussed, and complex amplitude and intensity transmittance functions are defined. Nonmonochromatic waves are described in terms of time-varying phasors, and the concept of the coherence of wave fields is introduced.

## 3.1 MONOCHROMATIC WAVES AND COMPLEX WAVE AMPLITUDES

Light waves are normally described mathematically by the variations of their electric fields as functions of time and space. In a complete description the direction or polarization of the field oscillations must be specified, for example by means of a vector representation. However, in many cases of interest polarization can be ignored, and a much simpler *scalar*, or non-vectorial, representation is used. In this text optical wave fields are described by the scalar quantity $u$, which characterizes the instantaneous amplitude of the electric field of the optical wave as a function of position in space:[1]

$$u = u(x,y,z,t), \tag{3.1}$$

where $t$ denotes time and $(x,y,z)$ is a point in a three-dimensional Cartesian spatial coordinate system.

Of particular importance in the mathematical modeling of optical wave fields are *monochromatic* waves, which oscillate at a single frequency. For monochromatic waves, $u(x,y,z,t)$ assumes the form

$$u(x,y,z,t) = a(x,y,z)\cos\left[2\pi\nu t + \phi(x,y,z)\right], \tag{3.2}$$

where $\nu$ is the frequency of the oscillations, in hertz, $a(x,y,z)$ is the amplitude of the wave ($a > 0$) *and* $\phi(x,y,z)$ the phase. The meaning of this expression is clarified by a simple thought experiment. Assume that an extremely small dipole antenna is used as a probe to measure a monochromatic optical wave field in a room. The probe is connected to a two-dial display; as the probe is moved about the room the dials register the amplitude $a(x,y,z)$ and phase $\phi(x,y,z)$ of the wave field oscillations. (In such an experiment the phase $\phi$ must be measured relative to some reference phase, e.g., the phase of the oscillations at the origin of the coordinate system.) If the probe is held stationary, the amplitude and phase readings do not change with time, consistent with

[1] For polarized light, the scalar amplitude $u$ is proportional to the oscillatory electric field **E**; for non-polarized light, the relationship is generally more complicated.





the single-frequency nature of the monochromatic waves. Note that this experiment can only be a thought experiment, because in practice it is not possible to make such electric dipole measurements at the extremely high frequencies of optical waves—roughly $5.5\times10^{14}$ Hz for light in the middle of the visible part of the optical spectrum.

Just as for sinusoidal signals in linear systems analysis, it is convenient to represent the function of Eq. (3.2) using a complex-exponential, or phasor, notation. We obtain such a representation by writing

$$u(x,y,z,t) = \mathrm{Re}\left\{\mathbf{U}(x,y,z)\exp(-j2\pi\nu t)\right\}, \tag{3.3}$$

where

$$\mathbf{U}(x,y,z) = a(x,y,z)\exp[-j\phi(x,y,z)], \tag{3.4}$$

and where $\mathrm{Re}\{.\}$ denotes the real part.[2] So long as it is remembered that the wave fluctuations are cosinusoidal in time with frequency $\nu$, then the complex-valued function $\mathbf{U}(x,y,z)$ conveys the same information as does the real function $u(x,y,z,t)$: given $\mathbf{U}(x,y,z)$ and $\nu$ it is possible to obtain $u(x,y,z,t)$ via Eq. (3.3). The function $\mathbf{U}(x,y,z)$ is referred to as the phasor representation of the scalar wave amplitude $u(x,y,z,t)$, or the *complex amplitude*. We shall generally represent optical waves by their complex amplitudes.[3] In many cases later in the text we shall be interested in the complex amplitude distributions in planes of constant $z$. In such cases it will be convenient to write the complex amplitude either in the form $\mathbf{U}(x,y;z)$, setting $z$ off by a semicolon, or in a subscripted form, e.g., $\mathbf{U}_0(x,y)$, where the subscript denotes the plane of concern. Note that if $t$ is set equal to zero, $\mathrm{Re}\left\{\mathbf{U}(x,y,z)\exp(-j2\pi\nu t)\right\}$ has the same value as $\mathrm{Re}\left\{\mathbf{U}(x,y,z)\right\}$. Thus, we can say that the real part of complex amplitude $\mathbf{U}(x,y,z)$ gives the scalar amplitude $u(x,y,z,t)$ at time $t=0$.

The scalar amplitude $u(x,y,z,t)$ must satisfy the wave equation, and $a(x,y,z)$ and $\phi(x,y,z)$ can therefore not be completely arbitrary. In free space—i.e., in the absence of charges or currents—the scalar wave amplitude $u(x,y,z,t)$ satisfies the time-dependent scalar wave equation,

$$\nabla^2 u - \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = 0, \tag{3.5}$$

where $c$ denotes the speed of light ($3\times10^8$ m/s) and where $\nabla^2$ is the Laplacian operator:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \tag{3.6}$$

[2] Boldface is used in the text to denote phasor representations for waves and associated quantities. Note that $\mathbf{U}(x,y,z)$ could be as easily defined by writing $u(x,y,z,t) = \mathrm{Re}\{\mathbf{U}(x,y,z)\exp(j2\pi\nu t)\}$. The consequences of choosing $-j$ in the exponential of Eq. (3.3) is explored in a homework problem.

[3] Scalar amplitude $u(x,y,z,t)$ and associated complex amplitude $\mathbf{U}(x,y,z)$ are defined such that, as discussed in Section 3.3, $|\mathbf{U}(x,y,z_0)|^2$ is proportional to the power per unit area flowing through the plane $z = z_0$ in the immediate vicinity of point $(x,y)$. According to one convention for polarized light, $\mathbf{U}(x,y,z)$ is defined so as to equal the irradiance of the wave, $(1/2)\sqrt{\varepsilon/\mu}\langle\vec{E}\cdot\vec{E}\rangle$, where $\vec{E}$ is the electric field vector and where $\langle\rangle$ denotes a suitable time average. If a second convention is followed, $|\mathbf{U}|^2$ equals simply $\langle\vec{E}\cdot\vec{E}\rangle$.

Similarly, the complex amplitude $\mathbf{U}(x,y,z)$ satisfies the time-independent scalar wave equation, known as the *Helmholtz equation:*

$$\left[\nabla^2 + \left(\frac{2\pi\nu}{c}\right)^2\right]\mathbf{U}(x,y,z) = 0. \tag{3.7}$$

Because of the linearity of the wave equation, if $\mathbf{U}_1(x,y,z)$ and $\mathbf{U}_2(x,y,z)$ satisfy Eq. (3.7), then so does the weighted sum $\alpha\,\mathbf{U}_1(x,y,z) + \beta\,\mathbf{U}_2(x,y,z)$, where $\alpha$ and $\beta$ are complex constants.

## 3.2 PLANE AND SPHERICAL WAVES

Two particular wave-field functions are of special interest to us, one representing plane waves, the other representing spherical waves. Mathematical models for both types are developed in this section.

### 3.2.1 Plane Waves

Planar optical waves are observed regularly in nature—the light waves from a star, for example, are virtually planar by the time they reach the earth—and they are of great importance in the development of the theory of wave propagation presented in the next chapter. In the laboratory, planar wave fields are easily produced by expanding a laser beam with a pair of lenses, as suggested in Fig. 3-1.

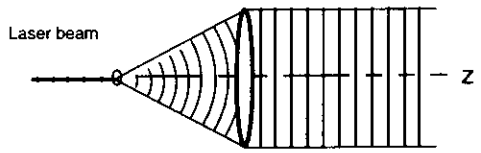



Fig. 3-1. Planar wave fields can be produced by expanding and collimating a laser beam.

If the collimated output of the beam expander is aligned to travel in the +z direction, the optical wave field in the central part of the expanded beam can be modeled by the scalar amplitude

$$u(x,y,z,t) = A\cos[2\pi\nu t - kz + \psi] \tag{3.8}$$

or by the associated complex amplitude

$$\mathbf{U}(x,y,z) = A\exp(-j\psi)\exp(jkz), \tag{3.9a}$$

which we can write as

$$\mathbf{U}(x,y,z) = \mathbf{A}\exp(jkz), \tag{3.9b}$$

where $\mathbf{A} = Ae^{-j\psi}$. Equations (3.8) and (3.9) represent a monochromatic plane wave of amplitude $A$ and phase $\psi$ traveling in the +z direction. The parameter $k$ is the so-called *wave number* of the wave. It is related to $c$, the speed of light, and to $\lambda$, the *wavelength* of the light, by the relations





$$k = \frac{2\pi}{\lambda}, \tag{3.10}$$

$$c = \lambda\nu. \tag{3.11}$$

For visible-band optical wavefields, the wavelength $\lambda$ lies in the approximate range 400 nm to 700 nm, corresponding to a range for $\nu$ of roughly $4.3\times10^{14}$ Hz to $7.5\times10^{14}$ Hz.

The motion of the wave through space is emphasized if $u(x,y,z,t)$ is written in the form

$$u(x,y,z,t) = A\cos[k(z-ct) - \psi]. \tag{3.12}$$

Recalling that $f(x\text{-}x_o)$ is the function $f(x)$ shifted a distance $x_o$ in the $+x$ direction, we see that the cosine function in Eq. (3.12) translates in the +z direction with speed $c$. A graphical representation for the scalar amplitude as a function of $z$ and at time $t$=0 is shown in Fig. 3-2. Planes of constant $z$ for which the argument of the cosine equals $n2\pi$, $n$ integer, are called the *phase fronts* of the wave.[4] The straight lines to the right of the collimating lens in Fig. 3-1 represent the phase fronts of the plane wave at some instant in time, typically chosen to be $t$=0. Since the phase angle $\psi$ cannot be measured at optical frequencies, it is often set equal to zero.

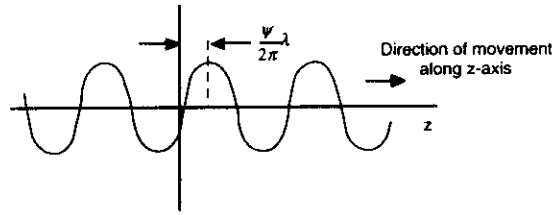



Fig. 3-2. Scalar amplitude $u(x,y,z,t) = A\cos[k(z-ct)-\psi]$ associated with plane wave propagating in +z direction, plotted for time $t$=0.

A unit-amplitude plane wave traveling in some arbitrary direction can be modeled by the real scalar amplitude

$$u(x,y,z,t) = \cos[2\pi\nu t - k(\alpha x + \beta y + \gamma z)] \tag{3.13}$$

or by the associated complex amplitude

$$\mathbf{U}(x,y,z) = \exp[jk(\alpha x + \beta y + \gamma z)]. \tag{3.14}$$

The parameters $\alpha$, $\beta$, $\gamma$ are the *direction cosines* of the wave, defined by

[4] The term wavefront is sometimes used interchangably with phase front. Usually, however, a wavefront is considered to be a surface all points on which are equal travel times from a point source. In the case of monochromatic waves, a wavefront can also be a phase front.

$$\alpha = \cos\phi_x, \tag{3.15a}$$

$$\beta = \cos\phi_y, \tag{3.15b}$$

$$\gamma = \cos\phi_z, \tag{3.15c}$$

where the angles $\phi_x$, $\phi_y$, $\phi_z$ are the angles the wave propagation direction vector makes with the +x, +y, and +z axes, respectively. The so-called *wave vector* associated with the plane wave, denoted by $\vec{k}$, is defined to have magnitude $k$ and direction given by $\alpha$, $\beta$, and $\gamma$. Thus, $\|\vec{k}\| = k$. The relation between $\vec{k}$, $\phi_x$, $\phi_y$, and $\phi_z$ is illustrated in Fig. 3-3.

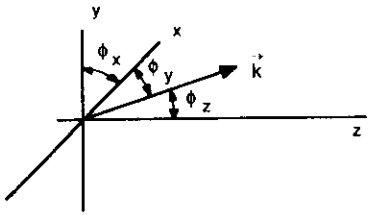



Fig. 3-3. Relationship between the wave propagation vector $\vec{k}$ and the angles $\phi_x$, $\phi_y$, and $\phi_z$ associated with direction cosines $\alpha$, $\beta$, and $\gamma$.

The direction cosines $\alpha$, $\beta$, $\gamma$ cannot be specified completely arbitrarily but are subject to the constraint

$$\alpha^2 + \beta^2 + \gamma^2 = 1 \tag{3.16}$$

So long as condition (3.16) is met, the complex amplitude of Eq. (3.14) satisfies the Helmholtz equation. Consistent with the conventional assumption of left-to-right propagation of light,[5] we consider only plane waves for which

$$\gamma > 0. \tag{3.17}$$

Of special interest to us later is a wave field whose complex amplitude in the plane $z$=0 is given by

$$\mathbf{U}(x,y,0) = \exp\left[j2\pi(u_O x + v_O y)\right]. \tag{3.18}$$

[5] The convention of left-to-right propagation of light will be followed throughout this text. Indeed, this convention is so strictly observed within the optics community that people attending optics research conferences are often temporarily confused when a diagram of an optical system shows light propagating from right to left!





Comparison with Eq. (3.14) suggests that this complex amplitude represents a unit-amplitude plane wave with direction cosines $\alpha$ and $\beta$ given by

$$\alpha = \lambda u_O, \tag{3.19}$$

$$\beta = \lambda v_O. \tag{3.20}$$

In fact, it can be shown that, since $\mathbf{U}(x,y,0)$ is governed by the Helmholtz equation, the expression on the right-hand side of Eq. (3.18) *must* represent a plane wave, evaluated in the $z$=0 plane. The remaining direction cosine, $\gamma$, from Eq. (3.16), and, through the plus sign, consistent with assumed left-to-right propagation, is given by

$$\gamma = +\sqrt{1-(\lambda u_O)^2 - (\lambda v_O)^2}. \tag{3.21}$$

Continued off the $z$=0 plane the complex amplitude of Eq. (3.18) is thus given by

$$\mathbf{U}(x,y,z) = \exp\left[j2\pi(u_O x + v_O y)\right]\exp\left(j2\pi\frac{z}{\lambda}\sqrt{1-(\lambda u_O)^2 - (\lambda v_O)^2}\right). \tag{3.22}$$

**Special Case: $\beta = 0$ and $z = 0$**

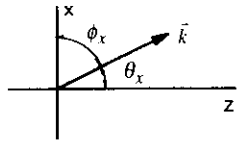
Consider the special case where $\beta = 0$ and the wave is observed in the $z$=0 plane. The complex amplitude of the wave in that plane is given by

$$\mathbf{U}(x,y,0) = \exp(jk\alpha x) = \exp\left(j2\pi\frac{\alpha}{\lambda}x\right) = \exp(j2\pi f_O x),$$

where

$$f_O = \frac{\alpha}{\lambda} = \frac{\cos\phi_x}{\lambda} = \frac{\sin\theta_x}{\lambda},$$

$\theta_x$ being the complement of the angle $\phi_x$. In this case $\vec{k}$ lies in the $x$-$z$ plane making an angle $\phi_x$ with the +$x$ axis and angle $\theta_x$ with the +$z$ axis, as illustrated in the figure.




If $\theta_x$ is small, then $\sin\theta_x \cong \theta_x$, and the spatial frequency $f_O$ is given by $f_O \cong \theta_x/\lambda$. Note that $f_O$ is the reciprocal of the spatial period of the wave field in the x direction. This period is the distance in the x-direction over which the phase of the complex wave amplitude changes by $2\pi$ radians. (Two

of the tiny dipole antennas of the earlier thought experiment, if separated in the $x$-direction by $1/f_0 = \lambda/\sin\theta_x$, would measure exactly the same sinusoidal oscillations, i.e., exactly the same amplitude and the same phase. A change in the separation of the antenna elements in the x-direction would produce a change in the measured phase.)

Figure 3-4 shows the intersections of the phase fronts of such a wave with the $x$-$z$ plane for $t$=0. The counter-clockwise tilt of the lines means that the frequency $u_o$ is positive. Note that the phase fronts intersect the $x$-axis at distances separated by $1/u_o$. As time $t$ advances, the phase fronts move upward and to the right, shifting the intersection points in the $+x$ direction.

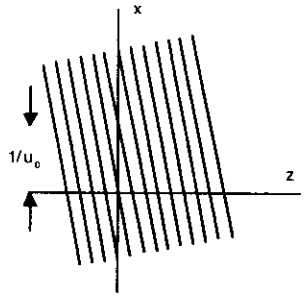



Fig. 3-4. Intersection of phase fronts with x-axis.

If $u_o$ and $v_o$ both equal zero, the complex amplitude of Eq. (3.18) assumes the particularly simple form

$$U(x,y,0) = 1,$$

which represents a unit-amplitude plane wave in the z=0 plane with $k$-vector in the $+z$ direction.

### 3.2.2 Spherical Waves

Much like planar waves, converging and diverging spherical waves can be easily produced by the focusing and subsequent expansion of a laser beam. Figure 3-5 illustrates a wave field focused to and diverging outward from a point. If the point of focus is assumed to be at the origin of the coordinate system, the complex amplitude

$$U(x,y,z) = A\frac{1}{\sqrt{x^2+y^2+z^2}}\exp\left[jk\sqrt{x^2+y^2+z^2}\right] \qquad (3.23)$$

would, upon initial inspection, appear to provide a satisfactory model for the wave, at least in the region away from the edge of the beam (where the amplitude would fall off, perhaps abruptly) and away from the focus (where the above expression suggests a non-physical infinite amplitude). In





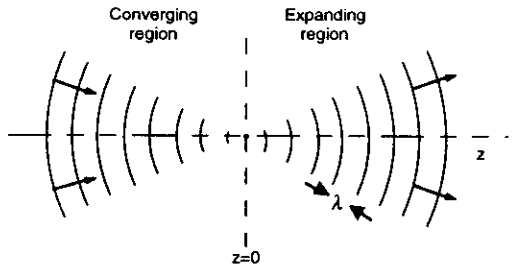



Fig. 3-5. Focused light beam showing converging and diverging spherical waves. The different offset of the wave fronts to the right and left of focus is indicative of the 90° phase shift experienced by the waves at such a focus.

particular, the phase fronts of the wave expressed in Eq. (3.23) are separated by λ, and the magnitude decreases linearly with distance from the point of focus. However, this expression is not fully suitable as a model because it fails to convey the sense of left-to-right propagation of the waves. Rather, it represents a spherical wave that is expanding outward in *all* directions from the origin, as from a hypothetical point source. Instead of the above expression, we take as our model the complex amplitude

$$U(x,y,z) = A\frac{\exp\left[jkz\sqrt{1+\left(x^2+y^2\right)/z^2}\right]}{z\sqrt{1+\left(x^2+y^2\right)/z^2}}. \qquad (3.24)$$

Note that the factor $z\sqrt{1+\left(x^2+y^2\right)/z^2}$ has the same magnitude as $\sqrt{x^2+y^2+z^2}$ but it carries with it the sign of z. Through its dependence on the sign of z, the exponential in Eq.(3.24)correctly models left-to-right advancement of the phase fronts.[6] Use of the sign-dependent quantity $z\sqrt{1+\left(x^2+y^2\right)/z^2}$ instead of $\sqrt{x^2+y^2+z^2}$ in the denominator may seem inappropriate since it leads to a 180° shift in the phase of the waves as they pass through the region of focus. In fact, however, such behavior is consistent with the wave equation, and a $\pi$-phase shift is indeed observed when converging spherical waves pass through a focus. If $A$ = 1, Eq. (3.24) represents a *unit-amplitude spherical wave*. Note that for such a wave the magnitude |U| = 1 at unit radial distance from the origin, or focus of the wave.

Although the complex amplitude of Eq. (3.24) does not, without further modification, satisfy the wave equation exactly (some indication of the necessary modifications is presented in Sect. 4.8), it nevertheless provides an excellent model of a converging/diverging spherical wave for distances away from the focus and from the edges of the light cone.

Often we are interested in spherical waves in planes of constant z that lie some distance away from the point of focus. If the region of concern is sufficiently close to the z-axis, the

[6] Note that if x and y are set equal to zero, Eq. (3.23) yields (1/z)exp(*jkz*), the factor exp(*jkz*) being consistent with the left-to-right propagating plane wave of Eq. (3.9).

complex amplitude representing the wave can be written in a mathematically more useful form. Let the distance $z$ from the focus to the plane of interest and the distance $\sqrt{x^2+y^2}$ from the $z$-axis to the observation point satisfy the condition

$$x^2 + y^2 << z^2. \tag{3.25}$$

The binomial expansion, $(1+\varepsilon)^{1/2} = 1 + \varepsilon/2 - \varepsilon^2/8 + \ldots$, allows us to write

$$z\left(1+\frac{x^2+y^2}{2z^2}\right)^{1/2} = z + \frac{x^2+y^2}{2z} - \frac{\left(x^2+y^2\right)^2}{8z^4} + \ldots \tag{3.26}$$

In using this expansion to rewrite Eq. (3.24), we retain the first two terms on the right for the exponent whereas for the denominator we retain only the first term,[7] with the result

$$\mathbf{U}(x,y,z) = A\frac{\exp(jkz)}{z}\exp\left[j\frac{k}{2z}\left(x^2+y^2\right)\right]. \tag{3.27}$$

Equation (3.27) gives the mathematical form for a converging/diverging spherical wave in the *quadratic-phase approximation*. If the wave field is observed in a specific plane of constant $z$, e.g., $z=z_o$, the factor $Ae^{jkz_o}/z_o$ can be combined into a single complex constant **A** with the following result:

$$\mathbf{U}(x,y,z_o) = \mathbf{A}\exp\left[j\frac{k}{2z_o}\left(x^2+y^2\right)\right]. \tag{3.28}$$

The condition $|\mathbf{A}| = 1$ corresponds to a unit-amplitude spherical wave. This function represents an expanding spherical wave if the plane $z=z_o$ lies to the right of the focus ($z_o>0$), a converging spherical wave if to the left ($z_o<0$). The fact that a plus sign is associated with a wave that grows with time provides a convenient memory aid. The function $\exp[j(k/6)(x^2+y^2)]$ thus represents, for a given wavelength $\lambda$ and for a given plane $z=z_o$, the complex amplitude of an expanding spherical wave of radius 3 meters, assuming that $x$ and $y$ are measured in meters. Figure 3-6 shows by means of dots the intersections of the phase fronts with the x-axis for a spherical wave expanding from a focus to the left of the origin. Note that as the wave expands, these dots move outward from the $z$-axis. The figures to the right show the real and imaginary parts of the complex amplitude, representing the scalar amplitude along the line at $t=0$ and at $t=1/2\nu$, i.e., one-half oscillatory period later.

[7] The more accurate approximation is used in the exponent because there the distance is multiplied by $k=2\pi/\lambda$, an exceedingly large number at optical wavelengths.





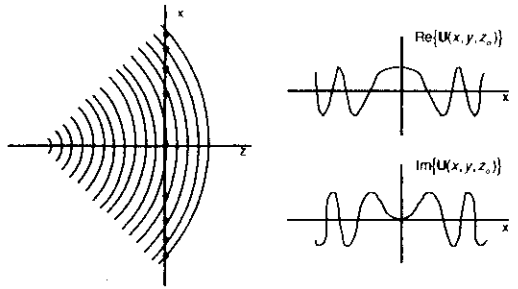



Fig. 3-6. The quadratic phase oscillations shown graphically.

The two approximations for $z\sqrt{1+\left(x^2+y^2\right)/z^2}$ used in obtaining Eq. (3.27) are known as the *Fresnel approximations*, named after the French scientist-engineer Augistin Fresnel who, in 1818, first introduced them in his mathematical treatment of Huygens' principle. The resulting complex amplitude of Eq. (3.27) is thus often referred to as the *Fresnel approximation to a spherical wave*. Note that in replacing $z\sqrt{1+\left(x^2+y^2\right)/z^2}$ with $z+(x^2+y^2)/2z$, we have in essence replaced spherical phase fronts with parabolic phase fronts, as suggested in Fig. 3-7. So long as the distance $\sqrt{x^2+y^2}$ is sufficiently small compared to $|z|$, the difference between the original spherical wave and its parabolic approximation can be neglected. This condition is, in fact, often not met in many laboratory experiments. For example, a phase error exceeding $2\pi/10$ radians (corresponding to a tenth of a wavelength) is incurred in using the quadratic phase approximation for a spherical wave of wavelength 633 nm measured little more than 5.5 mm off axis a distance 100 mm from the focus. However, despite the apparent limitation of the Fresnel approximations, they are often used in calculations and, for reasons discussed later in the text, generally with excellent results.

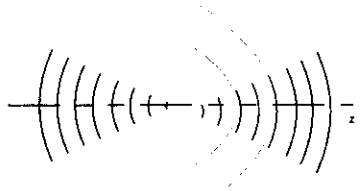



Fig. 3-7. In the Fresnel regime, spherical phase fronts (thick lines) are approximated by parabolic phase fronts (thin lines).

## 3.3 WAVE INTENSITY AND INTERFERENCE

The scalar wave amplitude $u(x,y,z,t)$ is proportional to the instantaneous electric field associated with the optical wave disturbance. When an optical wave field is probed, either by our eyes or some other detector, the response is not to the instantaneous wave amplitude itself but rather to the power conveyed by the wave. The amount of energy flowing per unit time through a suitably small area $\Delta x\Delta y$ in the vicinity of point $(x,y)$ in a plane of constant $z$ is proportional to $\langle u^2(x,y,z,t)\rangle \Delta x\Delta y$, where the angle brackets denote a time average. Accordingly, the *optical intensity* of the wave field, $I(x,y,z)$, is defined by

$$I(x,y,z) = 2\langle u^2(x,y,z,t)\rangle. \tag{3.29}$$

Instruments for measuring optical power are usually calibrated in terms of the light wave *irradiance*, which has the units of power per unit area. The irradiance and the optical intensity of a wave are proportional. The factor of 2 is introduced in the definition as a convenience, as will be seen in connection with Eq. (3.30) following. The duration of the time average is determined by the reciprocal of the temporal frequency bandwidth of the wave field or by a suitable characteristic time of the experiment performed, whichever is shorter.[8] For light from a filtered low-pressure gas-discharge lamp the reciprocal bandwidth is perhaps a tenth of a microsecond; for light from a gas laser it might be several milliseconds or longer. For ideal monochromatic waves, the time average of Eq. (3.29) is taken to be infinite, in which case $I(x,y,z)$ can easily be shown to have the form

$$\begin{aligned} I(x,y,z) &= \mathbf{U}(x,y,z)\mathbf{U}^*(x,y,z) \\ &= |\mathbf{U}(x,y,z)|^2. \end{aligned} \tag{3.30}$$

Note that the optical intensity of a plane wave is uniform throughout space. Sometimes we are interested in the total power carried by a light wave as it passes through a plane. The actual power flow, in watts, is given by the spatial integral of the incident irradiance distribution across the plane. Proportional to that quantity is what will be referred to in this text as the *optical power*. Thus, the optical power flowing through the plane $z=z_o$ is defined by

$$P = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(x,y,z_o)\,dx\,dy \tag{3.31}$$

Of particular importance is the optical intensity produced by two or more plane waves or spherical waves incident on the same plane. Consider first the sum of two monochromatic unit-amplitude plane waves, given by

$$\mathbf{U}_{tot}(x,y,z) = \exp[jk(\alpha_1 x + \beta_1 y + \gamma_1 z)] + \exp[jk(\alpha_2 x + \beta_2 y + \gamma_2 z)]. \tag{3.32}$$

The associated optical intensity, for convenience evaluated in the z=0 plane, is given by

$$I(x,y,0) = |\mathbf{U}_{tot}(x,y,0)|^2 = \mathbf{U}_{tot}(x,y,0)\mathbf{U}^*_{tot}(x,y,0)$$

[8] The issue of what constitutes the proper interval for the time average can, in fact, be a complex one. Special considerations apply, for example, in the case where a laser source is modulated to convey signal information. Specific homework problems in the text are designed to illustrate some of the complexities.





$$= \{\exp[jk(\alpha_1 x + \beta_1 y)] + \exp[jk(\alpha_2 x + \beta_2 y)]\}\{\exp[-jk(\alpha_1 x + \beta_1 y)] + \exp[-jk(\alpha_2 x + \beta_2 y)]\}$$

$$= 2 + 2\,\mathrm{Re}\{\exp\{jk[(\alpha_1 - \alpha_2)x + (\beta_1 - \beta_2)y]\}\}$$

$$= 2\left\{1 + \cos\left[2\pi\left(\frac{\Delta\alpha}{\lambda}x + \frac{\Delta\beta}{\lambda}y\right)\right]\right\}, \tag{3.33}$$

where

$$\Delta\alpha = \alpha_1 - \alpha_2 = \cos\phi_{x_1} - \cos\phi_{x_2}, \tag{3.34a}$$

$$\Delta\beta = \beta_1 - \beta_2 = \cos\phi_{y_1} - \cos\phi_{y_2}. \tag{3.34b}$$

The sinusoidal variations in this optical intensity distribution have spatial period $\lambda/\Delta\alpha$ in the x-direction and $\lambda/\Delta\beta$ in the y-direction. This distribution is often referred to as an *interference fringe pattern* or a set of *interference fringes*, corresponding to the interference—alternatingly constructive (adding) and destructive (subtracting)—of the two plane wave amplitudes. A photograph of such a sinusoidal fringe pattern is reproduced in Fig. 3-4 with the spatial periods in the x- and y-directions indicated. The contrast of the fringe pattern of Eq. (3.33) is 100%, owing to the equal amplitudes of the interfering plane waves. More generally, for plane waves of different amplitudes and phases the optical intensity pattern in the z=0 plane has the form

$$I(x,y,0) = I_o\left\{1 + \mu\cos\left[2\pi\left(\frac{\Delta\alpha}{\lambda}x + \frac{\Delta\beta}{\lambda}y\right) + \psi\right]\right\}, \tag{3.35}$$

where $\mu$ is a measure of the fringe contrast (generally referred to as the *fringe visibility*) and $\psi$ gives the spatial phase of the fringes (see homework problem).

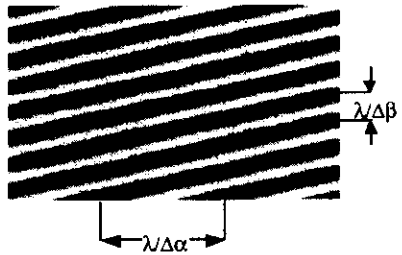



Fig. 3-4. Photograph of sinusoidal fringe pattern produced by the interference of two plane waves. The spatial periods in the x- and y-directions are indicated.

As a second example of a wave interference calculation, consider the optical intensity in the $z = z_o$ plane produced by the addition of two monochromatic unit-amplitude spherical waves expanding from points in the z=0 plane separated by a distance $S$. The geometry is shown in Fig.

3-5. For simplicity both points are assumed to lie on the x-axis at distances ±S/2 from the origin. Assuming $z_o$ to be sufficiently large to allow the quadratic phase approximation, we can generalized from Eq. (3.28) to write

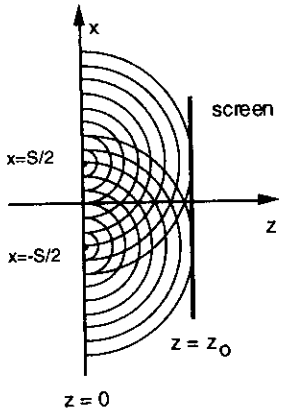



Fig. 3-5. Interference geometry for two expanding spherical waves.

$$\begin{aligned}\mathbf{U}_{tot}(x,y,z_o) &= A\exp\left\{j\frac{k}{2z_o}\left[\left(x+\frac{S}{2}\right)^2+y^2\right]\right\}+A\exp\left\{j\frac{k}{2z_o}\left[\left(x-\frac{S}{2}\right)^2+y^2\right]\right\}\\ &= A\exp\left\{j\frac{k}{2z_o}\left[x^2+y^2+\frac{S^2}{4}+Sx\right]\right\}+A\exp\left\{j\frac{k}{2z_o}\left[x^2+y^2+\frac{S^2}{4}-Sx\right]\right\}\end{aligned} \tag{3.36}$$

The corresponding interference pattern is described by the optical intensity





$$\begin{aligned}I(x,y,z_o) &= A^2\left\{\exp\left[j\frac{k}{2z_o}\left(x^2+\frac{S^2}{4}+y^2+Sx\right)\right]+\exp\left[j\frac{k}{2z_o}\left(x^2+\frac{S^2}{4}+y^2-Sx\right)\right]\right\}\\ &\times\left\{\exp\left[-j\frac{k}{2z_o}\left(x^2+\frac{S^2}{4}+y^2+Sx\right)\right]+\exp\left[-j\frac{k}{2z_o}\left(x^2+\frac{S^2}{4}+y^2-Sx\right)\right]\right\}\\ &= A^2\left[2+\exp\left(j\frac{k}{2z_o}2Sx\right)+\exp\left(-j\frac{k}{2z_o}2Sx\right)\right]\\ &= 2A^2\left[1+\cos 2\pi\left(\frac{S}{\lambda z_o}\right)x\right]\end{aligned} \tag{3.37}$$

We see that, subject to the conditions and approximations imposed in the analysis, the interference of two spherical waves also produces a sinusoidal interference fringe pattern. The spatial frequency of the sinusoidal fringes is proportional to the separation of the source points S and inversely proportional to the wavelength of the light and to the distance from the plane of the two source points to the (parallel) plane of observation. The spatial frequency of the sinusoidal interference pattern depends only on the separation S of the two source points in the plane containing them, not on their absolute positions. Thus, if one source point were at the origin and the other at coordinates (S,0) in the z=0 plane, the resulting interference pattern would have the same form as that given in Eq. (3.37), to within a shift in phase of the cosine function. A different result is obtained if the source points do not lie in a plane parallel to the observation plane or if the quadratic phase approximation is not used.

## 3.4 OPTICAL TRANSMITTANCE FUNCTIONS

When an optical wave passes through an aperture or photographic transparency, it is in general modified both in magnitude and in phase. The modification can be represented mathematically, as is now discussed.

Consider a monochromatic wave incident on a photo-transparency or other thin transparent object, as illustrated in Fig. 3-6. For convenience, assume the transparency to be in the z=0 plane. The incident complex amplitude is given by the function U(x,y,0-), which describes the wave field just to the left of the z=0 plane. For compactness in notation we denote this distribution by $\mathbf{U}_{inc}(x,y)$. As the wave passes through the object, it is attenuated by varying amounts in different places, and it may also be retarded to a greater or lesser degree, depending on whether the object has variations in thickness or refractive index. The resultant transmitted wave U(x,y,0+) just to the right of the transparency is denoted by $\mathbf{U}_{trans}(x,y)$. The ratio of $\mathbf{U}_{trans}(x,y)$ to $\mathbf{U}_{inc}(x,y)$ is, by definition, the *complex amplitude transmittance* of the object, written $\mathbf{t}(x,y)$:

$$\mathbf{t}(x,y)=\frac{\mathbf{U}_{trans}(x,y)}{\mathbf{U}_{inc}(x,y)}. \tag{3.38}$$

Note that the magnitude of $\mathbf{t}(x,y)$ must satisfy the condition $0 \le |\mathbf{t}(x,y)| \le 1$, since the wave amplitude can only be attenuated, not amplified. The complex transmittance function $\mathbf{t}(x,y)$ allows us to represent mathematically many commonly used apertures and mask functions—e.g., a circular aperture of diameter d can be represented by $\mathrm{cyl}(r/d)$—as well as more general thin transmissive objects such as photographic film or an object on a microscope slide. It is assumed

that the object is sufficiently thin that its transmittance does not depend on the angle of incidence of the illumination. (Viewed from a ray-optics perspective, a "thin" object satisfies the condition that a ray entering the object at given point leaves the object at the same point.) Note that if a mask in the $z$=0 plane with complex transmittance $\mathbf{t}(x,y)$ is illuminated by a normally incident unit amplitude plane wave, the incident wave field is represented by

$$\mathbf{U}_{inc}(x,y) = 1, \tag{3.39}$$

and the transmitted wave distribution is

$$\mathbf{U}_{trans}(x,y) = \mathbf{t}(x,y). \tag{3.40}$$

A transparency that changes only the phase of the transmitted wave (for example, by refractive index variations within the transparency) is represented by a transmittance function of the form $\mathbf{t}(x,y) = \exp[j\theta(x,y)]$. Consistent with our scalar treatment of light waves, we require that the transparency not affect the polarization of the light passing through it. This requirement generally means that the transparency may not be birefringent and that the finest-scale structure in the transparency must be somewhat larger than the wavelength of the light.

If two masks or transparencies are placed in close contact, their transmittance functions multiply. Thus, a pair of transparencies with complex amplitude transmittance functions $\mathbf{t}_1(x,y)$ and $\mathbf{t}_2(x,y)$ would, if placed one immediately behind the other, be represented by

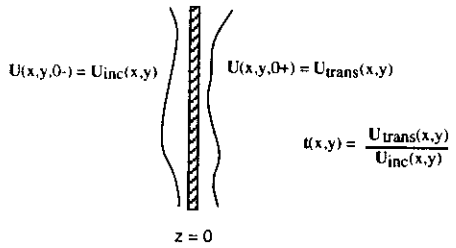



Fig. 3-6. Incident and transmitted waves at a thin transparency.

$$\mathbf{t}(x,y) = \mathbf{t}_1(x,y)\mathbf{t}_2(x,y). \tag{3.41}$$

Often it is the *optical intensity transmittance* of a thin transmissive object that is of interest. The intensity transmittance, denoted $\tau(x,y)$, is defined as the ratio of the optical wave intensity transmitted by the object at point $(x,y)$ to the optical intensity of the wave incident at that point. It is easily shown that intensity transmittance and complex amplitude transmittance are related by

$$\tau(x,y) = |\mathbf{t}(x,y)|^2. \tag{3.42}$$





## 3.5 NONMONOCHROMATIC WAVES AND SPATIAL INCOHERENCE

In reality, monochromatic waves do not exist. Although frequency-stabilized lasers can produce waves with extremely small fractional bandwidths, the magnitude and phase of these waves nonetheless vary with time, with bandwidths exceeding hundreds of kilohertz for all but the most frequency-stable lasers.

In order to accommodate nonmonochromatic waves, the concept of the *time-varying complex amplitude* is introduced. Consider the nonmonochromatic scalar wave amplitude

$$u(x,y,z,t) = A(x,y,z,t)\cos\left[2\pi\bar{\nu}t + \phi(x,y,z,t)\right]. \tag{3.43}$$

Whereas for the monochromatic case both amplitude $A$ and phase $\phi$ are constant with time, in the nonmonochromatic case they vary. The frequency $\bar{\nu}$ is the center, or average, frequency of the electromagnetic field oscillations. The temporal frequency bandwidth of the wave, $\Delta\nu$, depends on the linewidth of the source, $\Delta\lambda$. For the best frequency-stabilized lasers, the bandwidth of $u(x,y,z,t)$ can be as small as several kilohertz, corresponding to a fractional bandwidth of roughly $10^{-12}$. For a white-light source, the bandwidth is almost 300 terahertz (one terahertz equals $10^{12}$ hertz), corresponding to a fractional bandwidth of roughly 0.5. For the case when the fractional bandwidth $\Delta\nu / \bar{\nu}$ satisfies the condition

$$\frac{\Delta\nu}{\bar{\nu}} << 1, \tag{3.44a}$$

we say that the wave field is *narrowband*. It is easily shown that this condition is equivalent to the condition

$$\frac{\Delta\lambda}{\bar{\lambda}} << 1, \tag{3.44b}$$

where $\bar{\lambda}$ denotes the average wavelength of the waves.

Like monochromatic waves, nonmonochromatic waves can be represented by complex amplitudes. Specifically, $u(x,y,z,t)$ can be written in the form

$$u(x,y,z,t) = \mathrm{Re}\left\{\mathbf{U}(x,y,z,t)\exp(-j2\pi\bar{\nu}t)\right\}, \tag{3.45}$$

where $\mathbf{U}(x,y,z,t)$, given by

$$\mathbf{U}(x,y,z,t) = A(x,y,z,t)\exp\left[-j\phi(x,y,z,t)\right], \tag{3.46}$$

is the *time-varying phasor* associated with the real scalar amplitude disturbance. For thermal light sources, such as tungsten filaments and gas discharge tubes, the variations in the magnitude $A(x,y,z,t)$ as a function of time can be quite large, and the phase $\phi(x,y,z,t)$ can make excursions that are extremely large compared to $2\pi$ radians over a time comparable to or greater than $1/\Delta\nu$, the reciprocal bandwidth of the source. Even for sources such as gas lasers, the amplitude $A$ can

easily vary by ten percent or more over time, and temporal variations in $\phi$ are still large compared to $2\pi$ radians.

The optical intensity associated with a nonmonochromatic wave is given, as before, by Eq. (3.29). In terms of the time-varying phasor representation, $I(x,y,z)$ assumes the form

$$I(x,y,z) = \langle \mathbf{U}(x,y,z,t)\mathbf{U}^*(x,y,z,t) \rangle = \langle |\mathbf{U}(x,y,z,t)|^2 \rangle, \tag{3.47}$$

where again the time average is evaluated over a suitable interval, typically at least several times the reciprocal bandwidth of the source, $1/\Delta\nu$. Although $I(x,y,z)$ is not written as a function of $t$, that fact is not meant to imply that the optical intensity cannot in fact vary with time. However, the temporal variations that we are normally aware of through our eyes or some other detector—the modulation of a laser beam by a shutter, for example—are on a time scale much larger than $1/\Delta\nu$. Because such variations are usually of no concern to us in this text, the optical intensity will in most cases be expressed as an explicit function of spatial coordinates only.

Of special interest are nonmonochromatic waves for which the time-varying phasor amplitudes at all points in a given region of space maintain a fixed relationship, that is, waves for which the ratio of the complex amplitudes at two different spatial locations in that region does not vary with time but only as a function of the two spatial positions.[9] In such a case, The complex amplitude can be written in the separable form

$$\mathbf{U}(x,y,z,t) = \mathbf{U}_s(x,y,z)\mathbf{B}(t), \tag{3.48}$$

where the time-independent spatial distribution $\mathbf{U}_s(x,y,z)$ gives the phasor amplitude relative to some reference amplitude. Such waves are said to be *spatially coherent* in the region of concern. Often the function $\mathbf{B}(t)$ is given by

$$\mathbf{B}(t) = \frac{\mathbf{U}(0,0,0,t)}{\sqrt{\langle |\mathbf{U}(0,0,0,t)|^2 \rangle}}, \tag{3.49}$$

in which case the region over which the wave is spatially coherent is in the vicinity of the origin and the reference amplitude is chosen as that at the origin. Note that $\mathbf{B}(t)$ is normalized so as to be dimensionless, and $\mathbf{U}_s(x,y,z)$ thus has the units of $\mathbf{U}(x,y,z)$ itself. If $\mathbf{B}(t)$ is normalized as in Eq. (3.49), the optical intensity of the wave is given by $I(x,y,z) = |\mathbf{U}_s(x,y,z)|^2$, just as though the wave were monochromatic. In later chapters we shall be concerned with waves that are spatially coherent in a plane of constant z. In such cases we can suppress reference to z altogether and write simply

$$\mathbf{U}(x,y,t) = \mathbf{U}_s(x,y)\mathbf{B}(t). \tag{3.50}$$

The interference of nonmonochromatic waves is more complicated to analyze than is that of monochromatic waves. Consider first the optical intensity associated with two overlapping

[9] Such conditions occur, for example, when light from a point-like narrowband source is observed in a volume for which all optical pathlengths from the source differ by amounts small compared to $c/\Delta\nu$.





nonmonochromatic waves $\mathbf{U}_1$ and $\mathbf{U}_2$. The total complex amplitude in a plane of constant z is given by

$$\mathbf{U}_{tot}(x,y,t) = \mathbf{U}_1(x,y,t) + \mathbf{U}_2(x,y,t), \tag{3.51}$$

the associated optical intensity being

$$I_{tot}(x,y) = \langle |\mathbf{U}_1(x,y,t) + \mathbf{U}_2(x,y,t)|^2 \rangle = \langle |\mathbf{U}_1(x,y,t)|^2 \rangle + \langle |\mathbf{U}_2(x,y,t)|^2 \rangle + \langle \mathbf{U}_1(x,y,t)\mathbf{U}_2^*(x,y,t) \rangle + \langle \mathbf{U}_1^*(x,y,t)\mathbf{U}_2(x,y,t) \rangle. \tag{3.52}$$

The first two terms in this expression equal the optical intensities that would be observed if, respectively, only wave field $\mathbf{U}_1$ or $\mathbf{U}_2$ were present. The term $\langle \mathbf{U}_1(x,y,t)\mathbf{U}_2^*(x,y,t) \rangle$ and its conjugate correspond to the temporal correlation of $\mathbf{U}_1(x,y,t)$ with $\mathbf{U}_2(x,y,t)$ at point $(x,y)$. If we denote the first two terms by $I_1(x,y)$ and $I_2(x,y)$, then $I_{tot}(x,y)$ can be written in the form

$$I_{tot}(x,y) = I_1(x,y) + I_2(x,y) + 2\,\mathrm{Re}\{\langle \mathbf{U}_1(x,y,t)\mathbf{U}_2^*(x,y,t) \rangle\}. \tag{3.53}$$

The value of the third term in Eq. (3.53), which represents the interference of the two waves, depends intimately on the specific amplitude and phase functions associated with $\mathbf{U}_1$ and $\mathbf{U}_2$. The term could be positive, corresponding to the case of constructive interference, or negative, corresponding to destructive interference.

Of great importance is the case when the two wave distributions $\mathbf{U}_1(x,y,t)$ and $\mathbf{U}_2(x,y,t)$ are produced by *independent* sources—e.g., different lasers, atoms in different regions of a gas discharge lamp, or atoms in different regions of a radiating tungsten filament—for in that case the cross-correlation of the two wave fields evaluates to zero, with the simple result

$$I_{tot}(x,y) = I_1(x,y) + I_2(x,y). \tag{3.54}$$

Under such circumstances, we say that the two wave fields $\mathbf{U}_1$ and $\mathbf{U}_2$ are *mutually incoherent*, or that they add *incoherently*. For mutually incoherent wave fields, it is seen that there is no interference of the waves and that the wave fields can be added directly on an intensity, or power, basis.[10] Similar conditions hold if the wave fields from more than two independent nonmonochromatic sources are present: all cross-product terms in expressions like that of Eq. (3.52) have zero time averages, and the total optical intensity can be determined simply by adding the wave intensities associated with the individual contributing wave fields rather than by going through the intermediate step of calculating the total complex amplitude.

## REFERENCES

3-1 J. W. Goodman, *Introduction to Fourier Optics*, 2nd ed., McGraw-Hill, New York (1996).

[10] It can be argued that there is interference that remains relatively stable over an interval that is small compared to the reciprocal of the wave-field bandwidth. However, for non-laser sources this interval is so short that such interference cannot be observed by available detectors, and the issue of whether there is or is not short-term interference is thus of no practical importance.

3-2 J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics*, John Wiley and Sons, New York (1978).

## PROBLEMS

3.1 A monochromatic wave in the $z$=0 plane is represented by the complex amplitude $\mathbf{U}(x,y) = \mathrm{rect}(x,y)\exp(j2\pi 4x)$.

(a) Assuming the temporal frequency of the wave to be $\nu$, find the corresponding real (scalar) distribution in that plane, $u(x,y,0,t)$.

(b) Describe the nature of this wave in words.

3.2 In this and many optics texts for engineers $\cos\omega t$ is represented, consistent with Eq. (3.3), by $\mathrm{Re}\{\exp(-j\omega t)\}$, a phasor in the complex plane that rotates clockwise with time. However, some authors follow the opposite convention, representing $\cos\omega t$ by $\mathrm{Re}\{\exp(j\omega t)\}$, a counter-clockwise rotating phasor. Assuming that this alternative convention is followed, find the complex amplitudes for (a) a unit amplitude plane wave with direction cosines $\alpha$, $\beta$, $\gamma$, and (b) a spherical wave, traveling nominally left-to-right, that converges to and diverges from the origin. Compare your results with the expressions in Eqs. (3.14) and (3.24).

3.3 Assuming that a monochromatic scalar wave amplitude $u(x,y,z,t)$ satisfies the real scalar wave equation, show by using Eq. (3.3) in Eq. (3.5) that the corresponding complex amplitude $\mathbf{U}(x,y,z)$ satisfies the Helmholtz equation, Eq. (3.7).

3.4 Show that complex amplitude $\exp[jk(\alpha x+\beta y+\gamma z)]$ satisfies the Helmholtz equation if $\alpha^2+\beta^2+\gamma^2 = 1$ and $k = 2\pi\nu/c$.

3.5 Find the phasor representation for a monochromatic plane wave whose $k$-vector lies in the $x$-$z$ plane and makes an angle of $30^\circ$ with the $+x$ axis.

3.6 Assume that the k-vector of a monochromatic plane wave makes a $20^\circ$ angle with the $+x$ axis and an $85^\circ$ angle with the $+y$ axis.

(a) What angle does it make with the $+z$ axis?

(b) Assuming $\lambda = 500$ nm, find the spatial frequency of this wave distribution in the x and y directions. Express your answer in cycles per millimeter ($\mathrm{mm}^{-1}$).

3.7 Assuming the $k$-vector of a plane wave lies in the $x$-$z$ plane and that it makes an angle $\theta_o$ with the $+z$ axis that is small, show that $\alpha = \sin\theta_o = \theta_o$, $\beta = 0$.

3.8 Consider a unit-amplitude plane wave with k-vector in the $x$-$z$ plane making an angle $\theta_x = 1^\circ$ with the $+z$ axis.

(a) Find the corresponding complex amplitude $\mathbf{U}(x,y,z)$.

(b) Assuming that $\lambda = 633$ nm and that $y$ and $z$ are held constant, by how much must $x$ change if the phase of $\mathbf{U}(x,y,z)$ is to change by $2\pi$ radians?





(c) Assume again that $\lambda = 633$ nm and that now $x$ and $y$ are held constant. By how much must $z$ change for the phase of $\mathbf{U}(x,y,z)$ to change by $2\pi$ radians?

3.9 A unit amplitude plane wave propagates with the $k$-vector in the $x$-$z$ plane and making a *small* angle $\theta_x$ with the $+z$ axis. Show that in the $z = 0$ plane the wave can be represented by complex amplitude $\mathbf{U}(x,y,0) = \exp(j2\pi f_o x)$, where $f_o \approx \theta_x/\lambda$.

3.10 The lines in Fig. P3.11 represent the plane wave with complex amplitude $\mathbf{U}(x,y,z) = \exp[jk(\alpha x+\gamma z)]$. Provide a corresponding sketch for the wave $\mathbf{U}(x,y,z) = \exp(j\pi/2)\exp[jk(\alpha x+\gamma z)]$.

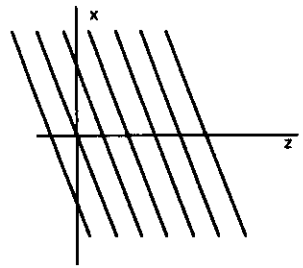



Fig. P3.1

3.11 Assuming that $(z-z_s)^2$ is sufficiently large compared to $[(x-x_s)^2 + (y-y_s)^2]$, show that a spherical wave expanding from the point $(x_s,y_s,z_s)$, as observed at point $(x,y,z)$, can be written in the approximate (quadratic-phase) form

$$\mathbf{U}(x,y,z) \approx \frac{A}{|z-z_s|}\exp[jk(z-z_s)]\exp\left\{j\frac{k}{2(z-z_s)}\left[(x-x_s)^2+(y-y_s)^2\right]\right\}.$$

3.12 A spherical wave diverging from a point source at coordinates $(0,0,-Z)$ is incident on the $z$=0 plane. Find the quadratic-phase approximation for $\mathbf{U}(x,y,0)$ assuming that $|\mathbf{U}(x,y,0)| = 1$.

3.13 Show that the equation $\mathbf{U}(x,y;z) = \mathbf{A}\exp\left[j1.67\pi(x^2+y^2)\right]$ gives the complex amplitude, in the quadratic-phase approximation, for a spherical wave of wavelength $\lambda = 600$ nm expanding from the origin and observed in a plane a distance 1 meter in the $+z$ direction, assuming $x$ and $y$ to be expressed in millimeters.

3.14 Figure P3.15 shows a spherical wave converging to a point $(x_o,0,z_o)$. Find the quadratic phase approximation for the complex amplitude of that wave as observed in the $z$=0 plane. Assume that $|\mathbf{U}(x,y,0)| = 1$.

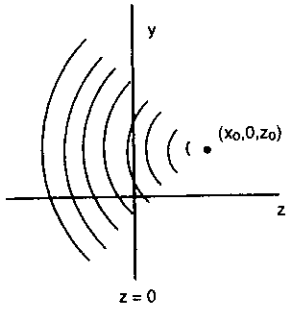



Fig. P3.15

3.15 Establish the validity of Eq. (3.27) by direct substitution in Eq. (3.26).

3.16 Find the optical intensity distributions in the $z$=0 plane associated with the following complex amplitudes.

(a) $U(x,y,0) = \exp(j2\pi f_o x)$

(b) $U(x,y,0) = \exp(j2\pi f_o x) + \exp(-j2\pi f_o x)$

(c) $U(x,y,0) = 1 + \exp(j2\pi f_o x)$

(d) $U(x,y,0) = 1 + \exp(-j2\pi f_o x)$

(e) $U(x,y,0) = 1 + \exp(j\theta)\exp(j2\pi f_o x)$

3.17 Two unit-amplitude plane waves are incident on the $z$=0 plane. Their $k$-vectors, both in the $x$-$z$ plane, make angles $\pm\theta_o$ to the $+z$ axis.

(a) Show that the corresponding optical intensity distribution in that plane is given by

$$I(x,y,0) = 2 + 2\cos\left[2\pi\left(2\frac{\sin\theta_o}{\lambda}\right)x\right].$$

(b) Repeat the problem, but assume that one of the waves has its $k$-vector along the $+z$ axis.

3.18 Show that the interference in the $z$=0 plane of the two plane waves $A_1\exp\{j[k(\alpha_1 x + \beta_1 y) + \theta_1]\}$ and $A_2\exp\{j[k(\alpha_2 x + \beta_2 y) + \theta_2]\}$ produces an optical intensity distribution of the form given in Eq. (3.35). Find the values for $I_o$, $\mu$, and $\psi$. Assume that $A_1$ and $A_2$ are real.

3.19 Incident on a screen are a normally incident unit-amplitude plane wave and a diverging unit-amplitude spherical wave, both monochromatic and at the same wavelength $\lambda$, as shown in Fig. P3.19. Find the optical intensity of the corresponding interference pattern assuming that the source of the expanding spherical wave is at the origin of the coordinate system, the





observation screen is in the plane $z$=$z_o$, and that the observation region is sufficiently small compared to $z_o$ that the spherical wave can be prepresented in the quadratic-phase (Fresnel) approximation.

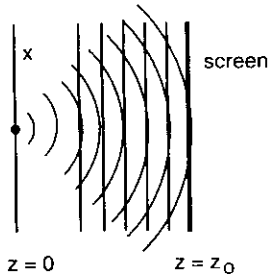



Fig. P3.19. Geometry for producing interference pattern.

3.20 The wave produced by a typical gas laser has a circularly symmetric Gaussian amplitude distribution. Show that a wave field with complex amplitude distribution

$$U(x,y;0) = A\exp\left(-\frac{x^2+y^2}{r_o^2}\right) = A\exp\left[-\left(\frac{r}{r_o}\right)^2\right]$$

in the z=0 plane carries unit optical power across that plane if the amplitude A equals $(1/r_o)\sqrt{2/\pi}$.

3.21 A unit amplitude plane wave is normally incident on a transparency with complex amplitude transmittance $t(x,y) = \exp[-ja(x^2+y^2)]$. Show that the resulting transmitted wave corresponds to the quadratic-phase approximation to a spherical wave converging to a point on the $z$-axis a distance $z_o$ from the transparency and find $z_o$. (In chapter 5 it is shown that a simple thin lens has a complex amplitude transmittance function of this form.)

3.22 A glass prism takes an incident beam and redirects it as shown in Fig. P3.22. Find a simple expression for the complex amplitude transmittance for such a prism, assuming that the prism has ignorable thickness and that it tilts a normally-incident beam clockwise through an angle $\theta_o$.

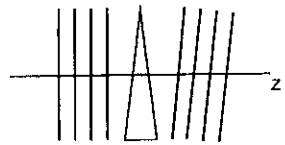



Fig. P3.22

3.23 A thin transparency has optical intensity transmittance suggested by Fig. P3.23: In the middle strip $\tau(x,y)$ equals 1; for both side strips it equals 1/2. The strips are one unit wide in the $x$-direction and separated by unit-wide gaps. In the $y$-direction they are four units long. Assume that the opaque part of the transparency extends to infinity.

(a) Find an analytical expression for the optical intensity transmittance of the transparency.

(b) Find an analytical expression for the complex amplitude transmittance of the transparency assuming the left and right strips introduce no phase shift whereas the strip in the middle shifts the phase of the incident wave by 180°.

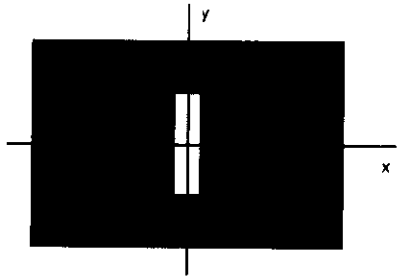



Fig. P3.23

3.24 Show that condition (3.44), $\Delta\nu/\nu \ll 1$, is equivalent to the condition $\Delta\lambda/\lambda \ll 1$.





INDEX - CHAPTER 3

# 4. Propagation of Waves: Diffraction Theory

In this chapter the mathematics describing the propagation of light waves in free space is presented and the diffraction of light by apertures and planar transparencies is explored. Traditionally, the relationships obtained are derived through one or another formulation of the Huygens-Fresnel diffraction principle (Ref. 4-1). The approach taken in this text is based instead on the so-called angular spectrum of plane waves, which allows the description of an arbitrary wave amplitude distribution in terms of a 2-D Fourier integral and its propagation in terms of the propagation of elementary plane wave components. A spatial-frequency-domain description of wave propagation is first derived. Then, with the introduction of simplifying approximations, wave propagation is considered in what are referred to as the Fresnel and Fraunhofer regimes. Examples of diffraction in both regimes are presented, including the singularly important case of converging spherical wave illumination of an aperture and the resulting Fourier transform relationship between the aperture function and the complex amplitude in the plane of convergence. Conditions allowing extension of the theory to non-monochromatic waves are then given, and an exact expression for monochromatic wave propagation is presented.

## 4.1 THE ANGULAR SPECTRUM AND THE PROPAGATION TRANSFER FUNCTION

Assume that a source of monochromatic light produces a complex wave amplitude distribution $\mathbf{U}(x,y,0)$ in the $z$=0 plane of the system of Fig. 4-1. This wave propagates, nominally left-to-right in the figure, producing the wave amplitude $\mathbf{U}(x,y,z)$ in a parallel plane a distance $z$ from the first. How can $\mathbf{U}(x,y,z)$ be determined given $\mathbf{U}(x,y,0)$?

As suggested by Fig. 4-2, this problem can be viewed as a 2-D systems problem with $\mathbf{U}(x,y,0)$ being the input to the system, $\mathbf{U}(x,y,z)$ the output, and $z$ a parameter of the system. Consistent with this point of view, the two wave amplitudes are written as $\mathbf{U}(x,y;0)$ and $\mathbf{U}(x,y;z)$, the semicolon emphasizing that $z$ is considered to be a fixed but arbitrary parameter of the system rather than an independent variable of the wave field distribution.

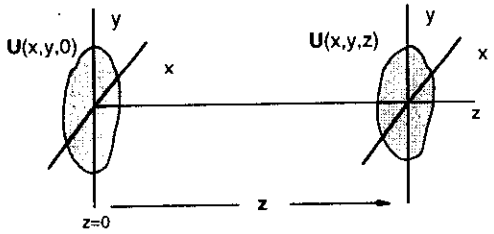



Fig. 4-1. Coordinate system for wave propagation problem.





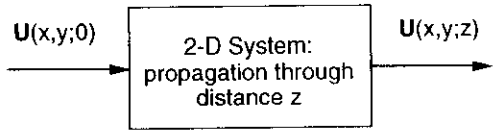



Fig. 4-2. 2-D system representation for wave propagation.

It will be shown in the following analysis that the relationship between $\mathbf{U}(x,y;z)$ and $\mathbf{U}(x,y;0)$ has the form of a convolution,

$$\mathbf{U}(x,y;z) = \mathbf{U}(x,y;0)**\mathbf{h}_z(x,y), \tag{4.1}$$

where $\mathbf{h}_z(x,y)$ is the *system impulse response* appropriate to propagation of the wave through a plane-to-plane distance $z$. Such a relationship tells us that the propagation operation is both linear and shift-invariant.[1]

Application of the convolution theorem leads to an alternative input-output relationship, obtained by Fourier transforming both sides of Eq. (4.1):

$$\hat{\mathbf{U}}(u,v;z) = \hat{\mathbf{U}}(u,v;0)\mathbf{H}_z(u,v), \tag{4.2}$$

where $\hat{\mathbf{U}}(u,v;z)$ and $\hat{\mathbf{U}}(u,v;0)$ are the 2-D Fourier transforms of $\mathbf{U}(x,y;z)$ and $\mathbf{U}(x,y;0)$ with respect to $x$ and $y$,

$$\hat{\mathbf{U}}(u,v;z) = \mathcal{F}\{\mathbf{U}(x,y;z)\}, \tag{4.3}$$

$$\hat{\mathbf{U}}(u,v;0) = \mathcal{F}\{\mathbf{U}(x,y;0)\}, \tag{4.4}$$

and where $\mathbf{H}_z(u,v)$, the *wave propagation transfer function*, is the Fourier transform of $\mathbf{h}_z(x,y)$:

$$\mathbf{H}_z(u,v) = \mathcal{F}\{\mathbf{h}_z(x,y)\}. \tag{4.5}$$

There are various methods for determining $\mathbf{h}_z(x,y)$ and $\mathbf{H}_z(u,v)$. The method used in this text exploits the equivalence of form between a 2-D Fourier component and the complex amplitude of a plane wave in a plane of constant $z$. To begin with, $\mathbf{U}(x,y;0)$ is written as an inverse Fourier transform:

$$\mathbf{U}(x,y;0) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \hat{\mathbf{U}}(u,v;0)\exp\{j2\pi(ux+vy)\}du\,dv. \tag{4.6}$$

Each Fourier component in this integral has the same form as does a unit amplitude plane wave as measured in the $z$=0 plane, as can be seen by setting $z$ equal to zero in Eq. (3.14):

[1] As discussed in Chapt. 1, the notion that the wave propagation system is linear and shift-invariant is reasonable from an intuitive viewpoint. The free-space superposition property of electromagnetic wave fields is of course well known from electromagnetic theory, and the shift invariance of the system is made evident through simple geometrical considerations. For example, the wave field surrounding a point-source emitter moves with that emitter. (To avoid complexities associated with the finite propagation speed of the wave, we must assume that the source motion is suitably slow.) If the input distribution is displaced by vector distance $(\Delta x,\Delta y,\Delta z)$, the entire response wave field is displaced through the same distance.

$$\mathbf{U}_{pw}(x,y,z)\big|_{z=0} = \exp\big[jk(\alpha x+\beta y)\big] = \exp\left[j2\pi\left(\frac{\alpha}{\lambda}x+\frac{\beta}{\lambda}y\right)\right] \tag{4.7}$$

With this similarity in mind, the identities

$$u=\frac{\alpha}{\lambda},\quad v=\frac{\beta}{\lambda} \tag{4.8}$$

are introduced and Eq. (4.6) is written in the form

$$\mathbf{U}(x,y;0)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\hat{\mathbf{U}}\left(\frac{\alpha}{\lambda},\frac{\beta}{\lambda};0\right)\exp\left[j2\pi\left(\frac{\alpha}{\lambda}x+\frac{\beta}{\lambda}y\right)\right]d\frac{\alpha}{\lambda}d\frac{\beta}{\lambda}. \tag{4.9}$$

Whereas Eq. (4.6) describes $\mathbf{U}(x,y;0)$ in terms of a superposition of 2-D Fourier components, Eq. (4.9) describes the input wave field in terms of a superposition of plane waves, all evaluated in the $z$=0 plane, and all traveling in different directions. The component wave with direction cosines $\alpha$ and $\beta$ has complex amplitude $\hat{\mathbf{U}}(\alpha/\lambda,\beta/\lambda;0)\,\mathrm{d}(\alpha/\lambda)\,\mathrm{d}(\beta/\lambda)$. The function $\hat{\mathbf{U}}(\alpha/\lambda,\beta/\lambda;0)$, which describes the relative amplitudes and phases of the different component waves, is known as the *angular spectrum of plane waves* of $\mathbf{U}(x,y;0)$ (angular because of its direction cosine dependence, spectrum because of its Fourier transform aspect), sometimes denoted $\mathbf{A}_0(\alpha/\lambda,\beta/\lambda)$.

Equation (4.9) is specific to the $z$=0 plane. However, the analytical form of each plane wave component making up the right-hand side of that equation is known for all values of $x$, $y$, and $z$. Specifically, from Eq. (3.22), we know that a plane wave having the form given in Eq. (4.7) in the $z$=0 plane has the form

$$\mathbf{U}_{pw}(x,y,z)=\exp\left[j2\pi\left(\frac{\alpha}{\lambda}x+\frac{\beta}{\lambda}y\right)\right]\exp\left(j2\pi\frac{z}{\lambda}\sqrt{1-\alpha^2-\beta^2}\right) \tag{4.10}$$

in a plane of arbitrary $z$, where we have used the relationship, from Eq. (3.16),[2]

$$\gamma=\sqrt{1-\alpha^2-\beta^2}. \tag{4.11}$$

Since all the component plane waves in the integral of Eq. (4.9) must continue throughout the coordinate system in accord with Eq. (4.10), the complex amplitude in a plane of arbitrary $z$, $\mathbf{U}(x,y;z)$, must have the form

$$\mathbf{U}(x,y;z)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\hat{\mathbf{U}}\left(\frac{\alpha}{\lambda},\frac{\beta}{\lambda};0\right)\exp\left[j2\pi\left(\frac{\alpha}{\lambda}x+\frac{\beta}{\lambda}y\right)\right]\exp\left[j2\pi\frac{z}{\lambda}\sqrt{1-\alpha^2-\beta^2}\right]d\frac{\alpha}{\lambda}d\frac{\beta}{\lambda}. \tag{4.12}$$

Substituting back for $\alpha/\lambda$ and $\beta/\lambda$ yields

$$\mathbf{U}(x,y;z)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{\hat{\mathbf{U}}(u,v;0)\exp\left[j2\pi\frac{z}{\lambda}\sqrt{1-(\lambda u)^2-(\lambda v)^2}\right]\right\}\exp\big[j2\pi(ux+vy)\big]du\,dv, \tag{4.13a}$$

which can be written in the form

[2] The choice of a plus sign in front of the radical is consistent with a presumed nominal left-to-right propagation of the wave field.





$$\mathbf{U}(x,y;z)=\mathcal{F}^{-1}\left\{\hat{\mathbf{U}}(u,v;0)\exp\left[j2\pi\frac{z}{\lambda}\sqrt{1-(\lambda u)^2-(\lambda v)^2}\right]\right\}. \tag{4.13b}$$

Comparison of Eq. (4.13b) with Eq. (4.2) confirms the statement made earlier, that the system of Fig. 1 is linear and shift-invariant, and yields the following equation for the propagation transfer function:

$$\mathbf{H}_z(u,v)=\exp\left[j2\pi\frac{z}{\lambda}\sqrt{1-(\lambda u)^2-(\lambda v)^2}\right]. \tag{4.14a}$$

This equation warrants discussion, for it is of fundamental importance in describing the propagation of optical wave fields. First, note that $\mathbf{H}_z(u,v)$ has circular symmetry and can therefore be written in the form

$$\mathbf{H}_z(\rho)=\exp\left[j2\pi\frac{z}{\lambda}\sqrt{1-(\lambda\rho)^2}\right]. \tag{4.14b}$$

Second, note that the character of the transfer function changes dramatically depending on whether $\rho=\sqrt{u^2+v^2}$ is greater or smaller than $1/\lambda$. So long as the condition

$$\rho=\sqrt{u^2+v^2}<\frac{1}{\lambda} \tag{4.15}$$

is satisfied, the quantity $\sqrt{1-(\lambda\rho)^2}$ is real-valued and $\mathbf{H}_z(u,v)$ is a pure phase function: in propagating a plane-to-plane distance $z$, Fourier components of $\mathbf{U}(x,y;0)$ that satisfy (4.15) undergo a shift in phase but are unchanged in magnitude. The amount of the phase shift depends on $\lambda$, $z$, $u$, and $v$. This aspect of $\mathbf{H}_z(u,v)$ is consistent with the behavior of plane waves making up the angular spectrum superposition of Eq. (4.12): plane waves propagating between two parallel planes separated a distance $z$ undergo a shift in phase—by an amount that depends on $z$, $\lambda$, $\alpha$, and $\beta$—but experience no other change.

If, on the other hand, the condition

$$\rho=\sqrt{u^2+v^2}>\frac{1}{\lambda} \tag{4.16}$$

is satisfied—i.e., if transverse spatial frequencies of the complex amplitude in the $z$=0 plane are sufficiently high—the square root in Eq. (4.14) is imaginary and $\mathbf{H}_z(u,v)$ assumes the form

$$\mathbf{H}_z(u,v)=\exp\big[-\mu(u,v)z\big], \tag{4.17}$$

where $\mu(u,v)$, given by

$$\mu(u,v)=\frac{2\pi}{\lambda}\sqrt{(\lambda u)^2+(\lambda v)^2-1}=2\pi\sqrt{\rho^2-1/\lambda^2}, \tag{4.18}$$

is real-valued. There is no phase shift, and, indeed, there is no propagation of that part of $\mathbf{U}(x,y;0)$ corresponding to these higher spatial frequency components! The pure attenuation of Fourier components satisfying (4.16) has its physical interpretation in so-called *inhomogeneous*, or

*evanescent* wave fields. Such electromagnetic fields oscillate at the optical frequency $\nu$, they have finite amplitudes, but they do not propagate: they simply decrease in amplitude as $z$ increases. It should be noted that a wave field containing evanescent components cannot be produced by the superposition of plane waves, since no plane-wave component contains spatial frequencies exceeding $1/\lambda$ in any direction. Instead, it is necessary that a propagating wave interact with some material object—a high-spatial-frequency grating, a finely detailed photo-transparency, an opaque mask with tiny pinholes in it, or some other such structure. Only then will an evanescent field be produced. Evanescent wave fields are directly analogous to the non-propagating electromagnetic fields in a microwave waveguide that is operated below its cutoff frequency. Because at optical frequencies $\lambda$ is so small, evanescent optical wave fields become negligible in amplitude for even very small values of $z$ and so are generally ignored. Note that the boundary between the propagating and non-propagating cases, where $\rho = \sqrt{u^2 + v^2} = 1/\lambda$, corresponds to the condition $\alpha^2 + \beta^2 = 1$. This condition implies, through Eq. (3.16), that $\gamma = \cos\phi_z = 0$, which is the condition for plane waves traveling in directions perpendicular to the $z$-axis.[3]

Since evanescent fields die out so rapidly with distance $z$, it is not unreasonable to ignore them totally, and to a good approximation $\mathbf{H}_z(u,v)$ can be written in the form

$$\mathbf{H}_z(u,v) = \begin{cases} \exp\left[ j2\pi \frac{z}{\lambda} \sqrt{1-(\lambda u)^2 - (\lambda v)^2} \right] & \text{for } u^2 + v^2 < \frac{1}{\lambda^2} \\ 0 & \text{otherwise} \end{cases} \tag{4.19a}$$

or, alternatively,

$$\mathbf{H}_z(\rho) = \exp\left[ j2\pi z \left( \sqrt{1/\lambda^2 - \rho^2} \right) \right] \mathrm{circ}\left( \frac{\rho}{1/\lambda} \right). \tag{4.19b}$$

## 4.2 PROPAGATION IN THE FRESNEL REGIME — FRESNEL DIFFRACTION

Equation (4.2) describes wave propagation through plane-to-plane distance $z$ in terms of a multiplication of the Fourier transform of the input wave amplitude by a transfer function, $\mathbf{H}_z(u,v)$, given by Eq. (4.19). Corresponding to that transfer function is the associated spatial impulse response $\mathbf{h}_z(x,y)$. The impulse response $\mathbf{h}_z(x,y)$ can be found by a rather complicated direct inverse Fourier transformation of $\mathbf{H}_z(u,v)$, as shown in Sec. 4.8. However, the task is greatly simplified if the square root in Eq. (4.14) is removed through application of a binomial

[3] The alert reader may detect what appears to be a logical flaw in the derivation of Eq. (4.13), for that equation is described as representing a superposition of plane waves, whereas if $\rho > 1/\lambda$ the wave field components are *not* plane waves but rather evanescent waves. Resolution of this apparent flaw in reasoning lies in recognizing that the complex amplitude of Eq. (4.10) satisfies the Helmholtz equation *regardless* of whether $\sqrt{1 - \alpha^2 - \beta^2}$ is real or imaginary. If the square root is real, Eq. (4.10) represents a propagating plane wave, and $\alpha$, $\beta$, and $\gamma$ can be interpreted as the cosines of the angles $\phi_x$, $\phi_y$, and $\phi_z$ that the k-vector makes with the $x$, $y$, and $z$ axes, as discussed in Sec. 3.2. If, on the other hand, the square root is imaginary, Eq. (4.10) represents a non-propagating wave field component, and the interpretation of $\alpha$, $\beta$, and $\gamma$ as cosines of real angles no longer applies. Equation (4.13) is nevertheless valid.





approximation and $\mathbf{H}_z(x,y)$ is written in the form of a quadratic phase factor. Let the quantity $(u^2 + v^2)$ satisfy the condition

$$u^2 + v^2 << \left( \frac{1}{\lambda} \right)^2. \tag{4.20}$$

The binomial approximation can then be used to write

$$\left[ 1 - (\lambda u)^2 - (\lambda v)^2 \right]^{1/2} \approx 1 - \frac{1}{2}(\lambda u)^2 - \frac{1}{2}(\lambda v)^2 \tag{4.21}$$

Substitution of the above approximation in Eq. (4.19) yields the following approximate form, denoted by $\tilde{\mathbf{H}}_z(u,v)$, for the propagation transfer function:

$$\tilde{\mathbf{H}}_z(u,v) = \exp\left( j2\pi \frac{z}{\lambda} \right) \exp\left[ -j\pi\lambda z \left( u^2 + v^2 \right) \right], \tag{4.22a}$$

or

$$\tilde{\mathbf{H}}_z(\rho) = \exp\left( j2\pi \frac{z}{\lambda} \right) \exp\left[ -j\pi\lambda z \rho^2 \right]. \tag{4.22b}$$

This transfer function, being now separable in u and v, can be easily Fourier transformed to yield the corresponding convolution kernel $\tilde{\mathbf{h}}_z(x,y)$. Key to the operation is the Fourier transform relationship, established in problem 2.28,

$$\mathcal{F}_x^{-1}\left\{ \exp(-j\pi u^2) \right\} = \exp(-j\frac{\pi}{4}) \exp(j\pi x^2) \tag{4.23}$$

Application of the similarity theorem, or scaling relationship, yields

$$\mathcal{F}_x^{-1}\left\{ \exp\left( -j\pi\lambda z u^2 \right) \right\} = \frac{1}{\sqrt{\lambda z}} \exp\left( -j\frac{\pi}{4} \right) \exp\left( j\frac{\pi}{\lambda z} x^2 \right), \tag{4.24}$$

with the result (setting $2\pi / \lambda = k$)

$$\tilde{\mathbf{h}}_z(x,y) = \frac{\exp(jkz)}{j\lambda z} \exp\left[ j\frac{k}{2z}\left( x^2 + y^2 \right) \right], \tag{4.25a}$$

or, alternatively,

$$\tilde{\mathbf{h}}_z(r) = \frac{\exp(jkz)}{j\lambda z} \exp\left[ j\frac{k}{2z} r^2 \right]. \tag{4.25b}$$

Although derived here by other methods and using different approximations, this result is exactly the same as is obtained in what is known as the Fresnel regime formulation of diffraction theory presented in many optics texts. We shall therefore refer to $\tilde{\mathbf{h}}_z(x,y)$ as the *Fresnel regime impulse response*. The transfer function of Eq. (4.22), accordingly, is referred to as the *Fresnel regime transfer function*. It should be emphasized that the Fresnel regime expressions have been obtained through application of the approximation given in (4.21). The legitimacy of this step is discussed in Sect. 4.8.

Having found $\tilde{\mathbf{h}}_z(x,y)$, we can now write the convolution of Eq. (4.1) in the form

$$\mathbf{U}(x,y;z) = \mathbf{U}(x,y;0) ** \left\{ \frac{\exp(jkz)}{j\lambda z} \exp\left[ j\frac{k}{2z}\left(x^2+y^2\right)\right]\right\} \tag{4.26a}$$

or

$$\mathbf{U}(x,y;z) = \frac{\exp(jkz)}{j\lambda z} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{U}(\xi,\eta;0) \exp\left\{ j\frac{k}{2z}\left[(x-\xi)^2+(y-\eta)^2\right]\right\} d\xi\, d\eta \tag{4.26b}$$

Comparison of Eq. (3.27) with Eq. (4.25a) shows that $\tilde{\mathbf{h}}_z(x,y)$ has the same analytical form as a *quadratic-phase approximation to a spherical wave* expanding from the origin, the only difference between the two expressions being the factor $j\lambda$ that appears in the denominator of $\tilde{\mathbf{h}}_z(x,y)$ but not in the spherical wave representation. Consistent with this observation, the wave amplitude $\mathbf{U}(x,y;z)$ of Eq. (4.26) can be thought of as resulting from the superposition of a large number of spherical waves (in a quadratic phase approximation), with each spherical wave originating at a hypothetical point source with coordinates $(\xi,\eta)$ in the $z$=0 plane. The point source at that location has amplitude proportional to $\mathbf{U}(\xi,\eta;0)$. This point of view is consistent with the *Huygens-Fresnel principle*, which forms the basis for classical diffraction theory. The quadratic phase functions entering into the superposition integral represent what are often referred to as Huygens-Fresnel "wavelets."

Equation (4.26) assumes that the input wave distribution is in the $z$=0 plane. In fact, there is nothing special about the plane $z$=0, and Eq. (4.26) and other propagation formulas can be rewritten to express propagation between parallel planes of arbitrary $z$. As an example, the free-space propagation of a monochromatic light wave distribution from plane $z=z_1$ to parallel plane $z=z_2$ is expressed by the convolution

$$\mathbf{U}(x,y;z_2) = \mathbf{U}(x,y;z_1) ** \frac{\exp(jkZ)}{j\lambda Z} \exp\left[ j\frac{k}{\lambda Z}\left(x^2+y^2\right)\right], \tag{4.27}$$

where $Z$ is given by

$$Z = z_2 - z_1. \tag{4.28}$$

The propagation distance $Z$ can be either positive or negative, depending on whether the plane $z=z_2$ lies to the right or the left of the plane $z=z_1$; in either case, Eq. (4.27) shows the relationship between the complex amplitude distributions in the two planes. Note that the absence of restrictions on the sign of $Z$ allows $\mathbf{U}(x,y;z_1)$ to be calculated from $\mathbf{U}(x,y;z_2)$ using the inverse propagation formula

$$\mathbf{U}(x,y;z_1) = \mathbf{U}(x,y;z_2) ** \frac{\exp(-jkZ)}{-j\lambda Z} \exp\left[ -j\frac{k}{\lambda Z}\left(x^2+y^2\right)\right], \tag{4.29}$$

the distance $Z$ again being given by Eq. (4.28). Inverse propagation is governed by the same convolution operation as forward propagation, with only the sign of the propagation distance being changed: $Z$ is positive for forward propagation, negative for inverse propagation, again assuming light propagates nominally in the $+z$ direction. The correctness of Eq. (4.29) is easily proved if the right-hand side of the equation is convolved with the forward propagation kernel $[\exp(jkZ)/j\lambda Z]\exp[j(k/\lambda Z)(x^2+y^2)]$: properties of quadratic-phase functions discussed in Sect. 2.5 can be used to show that $\mathbf{U}(x,y;z_2)$ results, as required.





Equation (4.26) expresses wave propagation in terms of a convolution integral. An alternate form results if the quantity $[(x-\xi)^2+(y-\eta)^2]$ in Eq. (4.26b) is replaced by $[x^2 - 2x\xi + \xi^2 + y^2 - 2y\eta + \eta^2]$, with the result, after some rearrangement of terms,

$$\mathbf{U}(x,y;z) = \frac{\exp(jkz)}{j\lambda z} \exp\left[ j\frac{k}{2z}\left(x^2+y^2\right)\right] \times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{U}(\xi,\eta;0) \exp\left[ j\frac{k}{2z}\left(\xi^2+\eta^2\right)\right] \exp\left[ -j\frac{2\pi}{\lambda z}(x\xi+y\eta)\right] d\xi\, d\eta \tag{4.30}$$

Noting that the last term of the integrand has the form of a 2-D Fourier transform kernel ($2\pi/\lambda$ has been deliberately substituted for $k$ to emphasize this form), we can write Eq. (4.30) in the convenient operator form

$$\mathbf{U}(x,y;z) = \frac{\exp(jkz)}{j\lambda z} \exp\left[ j\frac{k}{2z}(x^2+y^2)\right] \mathcal{F}\left\{ \mathbf{U}(x,y;0) \exp\left[ j\frac{k}{2z}(x^2+y^2)\right]\right\}\Big|_{u=x/\lambda z \atop v=y/\lambda z} \tag{4.31}$$

This form indicates that one method of finding the output wave amplitude $\mathbf{U}(x,y;z)$ is to perform the following steps:

(1) Multiply the input wave amplitude $\mathbf{U}(x,y;0)$ by quadratic phase factor $\exp[j(k/2z)(x^2+y^2)]$.

(2) Fourier transform the resulting product.

(3) Replace the Fourier variables $u$ and $v$ by $x/\lambda z$ and $y/\lambda z$, respectively.

(4) Multiply the result by the factor $(1/j\lambda z)\exp(jkz)\exp[j(k/2z)(x^2+y^2)]$.

Either Eq. (4.26) or Eq. (4.30), which express $\mathbf{U}(x,y;z)$ in terms of what is referred to as a *Fresnel diffraction integral*, can be used to calculate $\mathbf{U}(x,y;z)$. The convolution form of Eq. (4.26) emphasizes the linear, shift-invariant aspect of the propagation phenomenon. The Fourier transform form of Eq. (4.30), on the other hand, makes that equation attractive for numerical computation of $\mathbf{U}(x,y;z)$, using, e.g., a fast Fourier transform algorithm on a computer. In either case, the corresponding intensity pattern is obtained by taking the squared modulus of the calculated complex wave amplitude.

The results of a numerical computation, shown in Fig. 4-3, illustrate the effects of propagation on a wave having the form $\mathbf{U}(x,y;0) = \mathrm{rect}(x/w, y/w)$ in the $z$=0 plane, obtained by illuminating a square aperture with a normally incident monochromatic plane wave. The figure shows the x-axis cross-section of the optical intensity of the propagated wave, i.e., $I(x,0;z)$. Figure 4-4 shows photographs of the actual optical intensities associated with the cross-sectional plots.

To be supplied

Fig. 4-3. Propagation of wave amplitude $U(x,y;0) = \text{rect}(x/w,y/w)$ in the Fresnel regime. Shown are x-axis cross sections of the optical intensity of the propagated wave. Note that the scale changes for larger values of $z$.

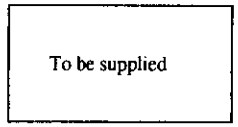



Fig. 4-4. Photographs of light intensity distributions corresponding to the propagated wave amplitudes of Fig. 4-3

The following observations are made in connection with these figures:

1. For small values of z, the propagated wave amplitude still has a highly rectangular appearance. The principle effect of propagation through small distances is to introduce small oscillations, or ringing, near the edges of the distribution.
2. As the wave propagates farther, this ringing affects more and more of the distribution; in addition, the intensity pattern begins to spread significantly.
3. With even greater propagation distances the spreading becomes quite significant, until for sufficiently large values of $z$ the width of the observed intensity distribution scales in proportion to $z$.
4. Finally, for $z$ sufficiently large, the shape of the optical intensity distribution ceases to change significantly as $z$ increases further. For such values of $z$, what is observed is sometimes referred to as the "far field" intensity distribution associated with the input wave amplitude. For reasons given below it is also referred to as the *Fraunhofer intensity distribution.*

In the next section it is shown that this far field, or Fraunhofer, intensity distribution has the functional form of the squared modulus of the Fourier transform of the input wave amplitude function. In the case illustrated in Figs. 4-3 and 4-4 this distribution would have the form of a 2-D $sinc^2$ function. Observations 1 through 4, although made in connection with the wave intensity distributions of these figures, apply generally to wave fields produced by illuminating an aperture by a plane wave, whether the aperture is square, circular, or otherwise: ringing is first observed near edges, then the distribution begins to spread in width, and finally the pattern stabilizes to its far field form. Only the distances over which these transitions are observed differ from case to case, for they depend strongly on the actual structure of the input wave amplitude distribution.

Thus far in this chapter the terminology of diffraction has been largely avoided. Rather, the emphasis has been on wave propagation, Eqs. (4.26) through (4.31) being integral manifestations of the physics governing the free-space propagation of electromagnetic waves. In practice, however, patterns like the ones shown in Fig. 4-4 are usually called *diffraction patterns*, and the phenomenon governing their appearance is usually referred to as *diffraction*. In producing the distributions of Fig. 4-4 a plane wave was made to illuminate a square aperture, or opening, in an opaque mask. That aperture is often referred to as the *diffracting aperture*, and the resulting wave is the *diffracted* wave.

Diffraction occurs whenever light waves interact with obstructing objects: masks with holes in them, gratings, photo transparencies, fine hairs, etc. Rather than traveling in straight lines, as the ray optics theories of Newton's time and earlier required, light appears to "bend" in





the vicinity of such obstructions. The term diffraction, coming from the Latin word meaning to break, was used to describe this perceived anomaly. There was of course no violation of the laws of physics, only of what were at that time thought to be the laws of physics. The motion of light through either vacuum or substance is properly described in terms of waves, a ray representation of that propagation being only an approximation, an approximation that is more and more valid for smaller and smaller wavelengths, since diffraction effects scale with wavelength.

The term diffraction is usually used to describe the behavior of an optical wave that is transmitted through an aperture in an opaque screen and that travels some distance in free space. The precise physics of the wave's interaction with the screen may be difficult to specify, depending on whether the screen is conductive, dielectric, etc., but the subsequent propagation of the wave is described by the free-space wave propagation integrals in their various forms. The usual assumption, in the case when the aperture is at least several times larger than a wavelength, is that the incident wave is unchanged within the aperture and is reduced to zero amplitude by the opaque part of the screen. Thus, by this assumption, a plane wave of amplitude $A$ normally incident on an opaque screen with a rectangular hole cut in it produces a transmitted wave of the form $U(x,y) = A\,\text{rect}(x/w_1,y/w_2)$. Strictly speaking, a complex amplitude like $A\text{rect}(x/w_1,y/w_2)$ violates the assumptions [e.g., condition (4.20)] leading to the Fresnel-regime propagation formulas, since the Fourier spectrum of such a distribution can contain significant energy out to high spatial frequencies. In practice, however, if $w_x$ and $w_y$ are at least several times $\lambda$, the errors incurred are usually negligible.

## 4.3 PROPAGATION IN THE FRAUNHOFER REGIME — FRAUNHOFER DIFFRACTION

Despite the relative simplicity of Eq. (4.30), it is impossible to evaluate $U(x,y;z)$ in closed form for many wave amplitude distributions of interest because of the presence of the quadratic phase factor inside the integral. If this factor can be eliminated, then a simple Fourier transform relationship exists between $U(x,y;0)$ and $U(x,y;z)$. As will be presented in the next chapter, one way of eliminating this factor is by imposing a lens in the path of the propagating wave.

It is not, however, necessary to actually eliminate the quadratic phase factor; it is only necessary that its phase remain close enough to zero that the factor can be ignored. Assume that the input wave has zero or essentially zero amplitude in the $z$=0 plane beyond a radial distance $r_{max}$ from the origin and that the plane-to-plane propagation distance $z$ is sufficiently large that the condition

$$\frac{k}{2z} r_{\max}^2 << 1, \tag{4.32a}$$

or, equivalently,

$$z >> \frac{\pi}{\lambda}\left(x^2 + y^2\right)_{\max} \tag{4.32b}$$

is satisfied. Under this condition, known as the *Fraunhofer condition*, the operand of the Fourier transform operator in Eq. (4.30) becomes negligible before the quadratic phase factor changes significantly from unity, and $U(x,y;z)$ can be written in the form

$$U(x,y;z) = \frac{\exp(jkz)}{j\lambda z}\exp\left[j\frac{k}{2z}\left(x^2+y^2\right)\right]\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U(\xi,\eta;0)\exp\left[-j\frac{2\pi}{\lambda z}(x\xi+y\eta)\right]d\xi d\eta, \tag{4.33a}$$

or, alternatively,

$$\mathbf{U}(x,y;z) = \frac{\exp(jkz)}{j\lambda z}\exp\left[j\frac{k}{2z}(x^2+y^2)\right]\mathcal{F}\{\mathbf{U}(x,y;0)\}\Big|_{u=x/\lambda z,\; v=y/\lambda z}. \tag{4.33b}$$

The formula presented in Eq. (4.33a) is known as the *Fraunhofer approximation to the wave propagation integral*, or, simply, the *Fraunhofer diffraction integral*. Note that the quadratic phase factor in front of the Fourier transform integral cannot be removed in the way the one inside the integral was, since the condition expressed in Eq. (4.32) refers to the input plane, whereas the quadratic phase factor remaining in Eq. (4.33b) refers to the observation plane, upon which no condition has been imposed. Of course, if observation is restricted to a region satisfying the condition $\left(x^2+y^2\right) << \lambda z/\pi$, then this latter quadratic phase factor can be approximated by unity too. The quantity of ultimate concern is frequently the associated optical intensity, given by [using Eq. (3.30)]

$$I(x,y;z) = |\mathbf{U}(x,y;z)|^2 = \left(\frac{1}{\lambda z}\right)^2\left|\hat{\mathbf{U}}\left(\frac{x}{\lambda z},\frac{y}{\lambda z};0\right)\right|^2. \tag{4.34}$$

The remaining quadratic phase factor of course disappears in the calculation of this latter quantity. Equations (4.33) and (4.34) describe *wave propagation in the Fraunhofer regime*. If the distance $z$ is sufficiently large that condition (4.32) is satisfied, we say, as noted in the previous section, that $\mathbf{U}(x,y;z)$ is the far field or Fraunhofer amplitude distribution associated with $\mathbf{U}(x,y;0)$. It is evident from Eq. (4.34) that the far field optical intensity distribution does not change basic functional form; rather, it remains proportional to $|\hat{\mathbf{U}}|^2$, simply growing larger as $z$ increases.

Note that in the Fraunhofer regime, wave propagation is no longer described by a convolution integral: at a mathematical level the shift invariance property has been lost. However, it must be kept in mind that Eq. (4.33) represents an approximation—valid only under restricted conditions—to the Fresnel regime integral, the latter being valid under much broader conditions. The shift invariance of the wave propagation "system" is always present in a physical sense. If, for example, the input wave field distribution is shifted transversely by some fraction of its width, the corresponding far field distribution moves through the same distance. However, because of the size of the Fraunhofer pattern, the motion of the far field distribution goes virtually unnoticed.

It should be noted that the Fraunhofer integral expression in Eq. (4.33a) can be interpreted as representing a superposition of plane waves. At the observation distance $z$ the radius of curvature of a Huygens-Fresnel spherical wavelet expanding from point $(\xi,\eta)$ is sufficiently large that the wavelet is adequately represented by a plane wave of basic form $\exp[-j(2\pi/\lambda z)(x\xi + y\eta)]$.

The Fraunhofer condition, (4.32), is quite severe at optical wavelengths because the wave number $k$ is so large. For example, at a wavelength of 500 nm and with a circular aperture of diameter 2 cm (corresponding to a value of 1 cm for $r_{max}$), the plane-to-plane propagation distance $z$ must satisfy the condition

$$z >> 628 \text{ meters}.$$

## 4.4 CALCULATIONS OF FRAUNHOFER DIFFRACTION PATTERNS





In this section Fraunhofer-regime wave amplitudes and intensitites are calculated for several simple but important input wave amplitude distributions. The input distributions can be obtained experimentally by illuminating a suitable aperture or photo transparency with a normally incident plane wave.

### 4.4.1 Rectangular Aperture

Consider first a wave field produced by illumination of a rectangular aperture by a unit amplitude, normally incident, monochromatic plane wave. The complex amplitude transmittance of the aperture is given by $\mathbf{t}_0(x,y) = \text{rect}(x/w_x, y/w_y)$, and $\mathbf{U}(x,y;0)$ is thus given by

$$\mathbf{U}(x,y;0) = \text{rect}\left(\frac{x}{w_x},\frac{y}{w_y}\right). \tag{4.35}$$

In this expression, $w_x$ and $w_y$ are the width and height, respectively, of the aperture. To calculate the Fraunhofer pattern we first calculate the Fourier transform of the transmitted wave amplitude: $\mathcal{F}\{\mathbf{U}(x,y;0)\} = w_x w_y \text{sinc}(w_x u, w_y v)$. Substitution in Eq. (4.30b) yields

$$\mathbf{U}(x,y;z) = \frac{\exp[jkz]}{j\lambda z}\exp\left[j\frac{k}{2z}\left(x^2+y^2\right)\right]w_x w_y \text{sinc}\left(w_x\frac{x}{\lambda z}, w_y\frac{y}{\lambda z}\right). \tag{4.36}$$

The corresponding optical intensity pattern, $I(x,y;z) = |\mathbf{U}(x,y;z)|^2$, is given by

$$I(x,y;z) = \left(\frac{w_x w_y}{\lambda z}\right)^2 \text{sinc}^2\left(\frac{x}{\lambda z/w_x}, \frac{y}{\lambda z/w_y}\right). \tag{4.37}$$

The basic shape of the Fraunhofer pattern is seen to remain unchanged as $z$ increases; only the scale of the pattern changes, in proportion to $z$. The distance $z$ must, of course, be sufficiently large to satisfy the Fraunhofer condition. The factor $(w_x w_y/\lambda z)^2$ is consistent with energy conservation, reducing the intensity of the pattern as it spreads out with increasing $z$. Figure 4-5 shows the x-axis cross section of this Fraunhofer intensity pattern. The distance from the peak of the central lobe of the pattern to the first zero in the x-direction is $\lambda z/w_x$, giving a full width (distance between nulls) for the main lobe of

$$\Delta_x = 2\frac{\lambda z}{w_x}. \tag{4.38}$$

Note that increasing either $\lambda$ or $z$ broadens the pattern, whereas increasing the aperture width $w_x$ narrows the pattern. Figure 4-6 shows a photograph of this diffraction pattern produced by a rectangular aperture with aspect ratio $w_x/w_y = 2$.

Sometimes it is the angular width of the central lobe (the angular separation between the first nulls on either side of the lobe) of this Fraunhofer pattern, as viewed from the aperture plane, that is of interest. This angular width, designated $\Delta_\theta$, is given for the x-direction by

$$\Delta_\theta = \tan^{-1}\left(\frac{\Delta_x}{z}\right) = \tan^{-1}\left(2\frac{\lambda}{w_x}\right) \cong 2\frac{\lambda}{w_x}, \tag{4.39}$$

where the approximation holds for the small angles implicit in Fresnel- and Fraunhofer-regime calculations.

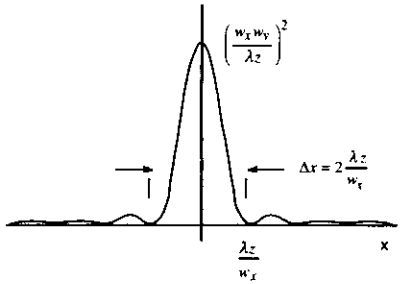



Fig. 4-5. X-axis cross section of the Fraunhofer pattern associated with a rectangular aperture of width $w_x$ in the x-direction.

Fig. 4-6. Photograph of Fraunhofer pattern produced by a rectangular aperture with aspect ratio $w_x/w_y = 2$.

### 4.4.2 Circular Aperture

If a circular aperture of diameter $w$ is illuminated in the z=0 plane by a unit amplitude, normally incident, monochromatic plane wave, the transmitted wave amplitude is given by

$$\mathbf{U}(r;0) = \mathrm{cyl}\left(\frac{r}{w}\right). \tag{4.40}$$

Assuming that the Fraunhofer approximation is valid, the observed wave amplitude a distance $z$ away is given by Eq. (4.30b) expressed in polar-coordinate form:

$$\mathbf{U}(r;z) = \frac{\exp(jkz)}{j\lambda z}\exp\left(j\frac{k}{2z}r^2\right)\mathcal{F}\{\mathbf{U}(r;0)\}\Big|_{\rho=r/\lambda z}. \tag{4.41}$$

But





$$\mathcal{F}\left\{\mathrm{cyl}\left(\frac{r}{w}\right)\right\} = \mathcal{H}_0\left\{\mathrm{cyl}\left(\frac{r}{w}\right)\right\} = \frac{\pi w^2}{4}\mathrm{somb}(w\rho),$$

and thus

$$\begin{aligned}\mathbf{U}(r;z) &= \frac{\exp(jkz)}{j\lambda z}\exp\left(j\frac{k}{2z}r^2\right)\left[\left(\frac{\pi w^2}{4}\right)\mathrm{somb}\left(\frac{wr}{\lambda z}\right)\right] \\ &= \frac{\exp(jkz)}{j\lambda z}\exp\left(j\frac{k}{2z}r^2\right)\left[\left(\frac{\pi w^2}{2}\right)\frac{J_1(\pi wr/\lambda z)}{\pi wr/\lambda z}\right].\end{aligned} \tag{4.42}$$

The corresponding optical intensity distribution, obtained by evaluating $|\mathbf{U}(r;z)|^2$, has the form

$$\begin{aligned}I(r;z) &= \left(\frac{\pi w^2}{4\lambda z}\right)^2 \mathrm{somb}^2\left(\frac{wr}{\lambda z}\right) \\ &= \left(\frac{\pi w^2}{4\lambda z}\right)^2\left[2\frac{J_1(\pi wr/\lambda z)}{\pi wr/\lambda z}\right]^2\end{aligned} \tag{4.43}$$

This intensity distribution is plotted in Fig. 4.7. The distribution has nulls at radial distances given by

$$\frac{wr}{\lambda z} = 1.220,\ 2.233,\ 3.238,\ldots \tag{4.44}$$

The central lobe thus has a full width between first nulls given by

$$\Delta = 2.44\frac{\lambda z}{w}, \tag{4.45}$$

22% larger than the central lobe of the diffraction pattern from a square aperture of equal width. The angular subtense of the central lobe is given by

$$\Delta_\theta = \tan^{-1}\left(\frac{\Delta}{z}\right) = \tan^{-1}\left(2.44\frac{\lambda}{w}\right) \approx 2.44\frac{\lambda}{w} \tag{4.46}$$

The optical intensity distribution of Eq. (4.43) is often referred to as the *Airy distribution*, after astronomer G. B. Airy, who first determined its mathematical form. Figure 4-8 shows the same pattern as a grayscale distribution. Contrast has been reduced to show more clearly the ring structure surrounding the central lobe.

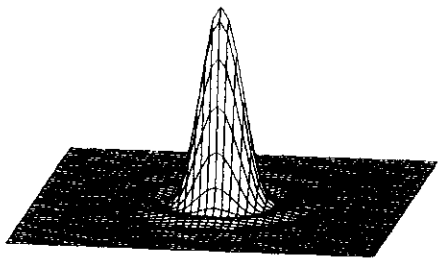


Fig. 4-7. Surface plot of the Fraunhofer intensity pattern for a circular aperture: the Airy distribution. The radial distance from the peak to the first null equals 1.22λz/w, where w is the diameter of the aperture.

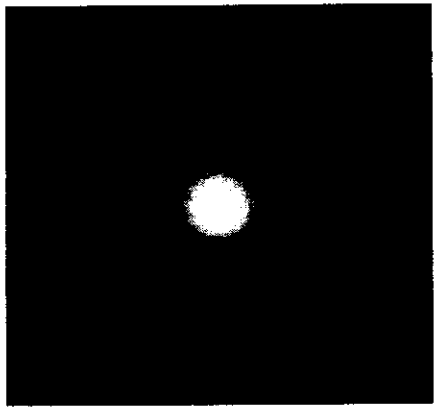


Fig. 4-8. Photograph of Fraunhofer pattern from circular aperture.

### 4.4.3 Sinusoidal Amplitude Grating

Gratings play an important role in optics, most notably in grating-type spectrometers. A grating structure of particular importance is the sinusoidal amplitude grating (usually referred to simply as a *sine grating*), represented by complex amplitude transmittance

$$\mathbf{t}(x,y)=\left[\frac{1}{2}+\frac{m}{2}\cos(2\pi f_o x)\right]\mathrm{rect}\left(\frac{x}{w},\frac{y}{w}\right), \tag{4.47}$$

where the parameter $m$ is real-valued and satisfies the condition $|m| \le 1$. This function is plotted in cross-section in Fig. 4-9. Assume such an object to be illuminated by a normally incident, unit amplitude, monochromatic plane wave. To calculate the resulting Fraunhofer pattern it is





necessary to evaluate the Fourier transform of $\mathbf{t}(x,y)$. This is done by invoking the convolution theorem and noting that

$$\mathcal{F}\left\{\frac{1}{2}+\frac{m}{2}\cos(2\pi f_o x)\right\}=\frac{1}{2}\delta(u,v)+\frac{m}{4}\left[\delta(u+f_o,v)+\delta(u-f_o,v)\right], \tag{4.48}$$

and

$$\mathcal{F}\{\mathrm{rect}(\tfrac{x}{w},\tfrac{y}{w})\}=w^2\,\mathrm{sinc}(wu,wv). \tag{4.49}$$

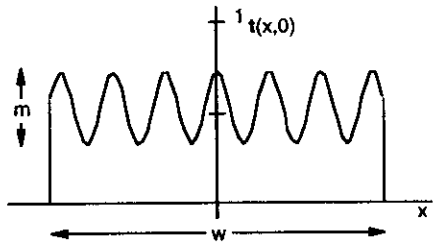



Fig. 4-9. Sinusoidal amplitude grating transmittance function.

Thus

$$\begin{aligned}\mathcal{F}\{\mathbf{U}(x,y;0)\}&=\mathcal{F}\{t(x,y)\}\\&=\frac{w^2}{2}\mathrm{sinc}(wu,wv)\\&\quad+\frac{w^2}{4}m\,\mathrm{sinc}\left[w(u+f_o),wv\right]+\frac{w^2}{4}m\,\mathrm{sinc}\left[w(u-f_o),wv\right]\end{aligned}. \tag{4.50}$$

Substituting in Eq. (4.30b) yields

$$\begin{aligned}\mathbf{U}(x,y;z)&=\frac{w^2}{2}\frac{\exp(jkz)}{j\lambda z}\exp\left[j\frac{k}{2z}(x^2+y^2)\right]\\&\quad\times\left\{\mathrm{sinc}(\frac{x}{\lambda z/w},\frac{y}{\lambda z/w})\right.\\&\quad\left.+\frac{m}{2}\mathrm{sinc}\left[\frac{(x+\lambda z f_o)}{\lambda z/w},\frac{y}{\lambda z/w}\right]+\frac{m}{2}\mathrm{sinc}\left[\frac{(x-\lambda z f_o)}{\lambda z/w},\frac{y}{\lambda z/w}\right]\right\}\end{aligned}. \tag{4.51}$$

The corresponding optical intensity has the form

$$I(x,y;z)=\left(\frac{w^2}{2\lambda z}\right)^2\left\{\operatorname{sinc}^2\left(\frac{x}{\lambda z/w},\frac{y}{\lambda z/w}\right)+\frac{m^2}{4}\operatorname{sinc}^2\left[\frac{(x+f_o\lambda z)}{\lambda z/w},\frac{y}{\lambda z/w}\right]\right.$$
$$\left.+\frac{m^2}{4}\operatorname{sinc}^2\left[\frac{(x-f_o\lambda z)}{\lambda z/w},\frac{y}{\lambda z/w}\right]+\text{cross-product terms}\right\} \tag{4.52}$$

The cross-product terms are products of sinc functions separated from each other along the $x$-axis by $f_o\lambda z$ and $2f_o\lambda z$. If the condition

$$f_o >> \frac{1}{w} \tag{4.53}$$

is satisfied, the sinc functions are effectively non-overlapping and their cross-products, which are insignificant in amplitude compared to their self-products, can be neglected. The $x$-axis cross section of this Fraunhofer pattern is shown in Fig. 4-10. In diffraction terminology, the central structure of the pattern is called the *zero-order diffraction component*, while the two side patterns are called the *first-order diffraction* components. Note that the higher the spatial frequency $f_o$ of the grating the farther from the zero-order component the two first-order components appear. As viewed from the plane of the grating itself, the two first-order components have an angular separation from the +z axis in the amount

$$\theta = \pm\tan^{-1}(f_o\lambda), \tag{4.54}$$

which is given approximately by $\pm f_o\lambda$ for small angles.

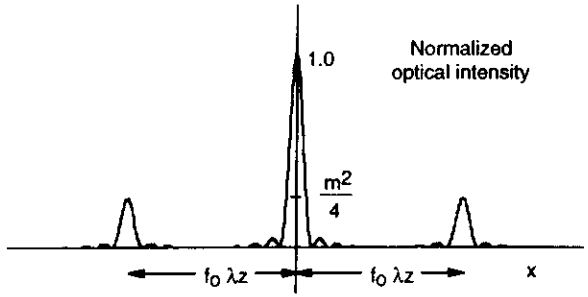



Fig. 4-10. The x-axis cross section of the Fraunhofer pattern from a sine grating.

### 4.4.4 Sinusoidal Phase Grating

Another grating of interest is the *sinusoidal phase grating*, represented by complex wave amplitude transmittance

$$t(x,y)=\exp\left[j\frac{m}{2}\sin(2\pi f_o x)\right]\operatorname{rect}\left(\frac{x}{w},\frac{y}{w}\right). \tag{4.55}$$





In this expression the parameter $m$ is the peak-to-peak excursion of the optical phase shift of the transmitted wave. Such a grating can be manufactured by embossing a sheet of transparent plastic in such a way that the thickness varies sinusoidally as a function of $x$. Alternatively, the refractive index of a transparent sheet might somehow be made to vary spatially.

In calculating the corresponding Fraunhofer pattern produced when such a grating is illuminated by a normally incident plane wave, we use the identity

$$\exp\left[j\frac{m}{2}\sin(2\pi f_o x)\right]=\sum_{n=-\infty}^{\infty}J_n\left(\frac{m}{2}\right)\exp(j2\pi n f_o x) \tag{4.56}$$

(note that the right-hand side of this equation is the Fourier series expansion for the periodic phase function) to obtain

$$\mathcal{F}\{U(x,y;0)\}=\sum_{n=-\infty}^{\infty}w^2J_n\left(\frac{m}{2}\right)\operatorname{sinc}[w(u-nf_o),wv]. \tag{4.57}$$

Substituting in Eq. (4.34) for the Fraunhofer intensity pattern yields

$$I(x,y;z)=\left(\frac{w^2}{\lambda z}\right)^2\sum_{n=-\infty}^{\infty}J_n^2\left(\frac{m}{2}\right)\operatorname{sinc}^2\left[\frac{w}{\lambda z}(x-nf_o\lambda z),\frac{w}{\lambda z}y\right], \tag{4.58}$$

where, assuming $f_0 >> 1/w$, cross product terms have been ignored. Figure 4-11 shows the $x$-axis cross section of this Fraunhofer pattern for the case m = 6.8. Light is diffracted by the grating into a number of diffraction components, with the optical intensity of the $n$th component being proportional to $\left[wJ_n(m/2)/kz\right]^2$. Note that where $m/2$ is a root of $J_o$, the zero-order component vanishes.

To be supplied

Fig. 4-11. X-axis cross section of the Fraunhofer pattern produced by a sinusoidal phase grating.

Another grating of special interest is the square-wave grating, or Ronchi ruling. The diffraction pattern produced by this structure is investigated in a homework problem.

## 4.5 CALCULATIONS OF FRESNEL DIFFRACTION PATTERNS

Although evaluation of the integrals in Eqs. (4.26) and (4.30) is in general quite difficult, there are several cases where it can be done with relative ease. One distribution for which the Fresnel propagation integral can be calculated in closed form is the Gaussian wave amplitude distribution. A Gaussian wave amplitude is of great practical importance, for it is produced by many lasers operating in their lowest order transverse mode (see, e.g., Ref. 4-8).

### 4.5.1 Gaussian Beam Propagation

Figure 4-12 illustrates a typical Gaussian laser beam converging to and expanding from a focus. At the *beam waist*, denoted by $z_o$, the beam diameter is a minimum. There, the optical wave field can be represented by complex amplitude

$$\mathbf{U}(x,y;z_o) = \mathbf{A}\exp\left(-\frac{x^2+y^2}{r_o^2}\right). \tag{4.59}$$

Note that $r_o$, the *beam radius at the waist*, gives the distance, measured from the center of the beam, at which the wave amplitude is reduced by a factor 1/e from its value at beam center. In specification sheets for gas lasers, the *beam diameter* is the diameter at which the laser beam irradiance is down by a factor $1/e^2$ from its value on axis. It is easily shown that the beam diameter at the waist, according to this definition, equals $2r_o$. Assuming the waist to lie in the $z$=0 plane, the beam amplitude can be written in the form (with $x^2+y^2 = r^2$)

$$\mathbf{U}(r;0) = A\exp\left[-\left(\frac{r}{r_o}\right)^2\right]. \tag{4.60}$$

According to Eq. (4.31), re-expressed in polar coordinates, propagation of this Gaussian wave through distance z produces the complex amplitude

$$\mathbf{U}(r;z) = \frac{\exp(jkz)}{j\lambda z}\exp\left(j\frac{k}{2z}r^2\right)\mathcal{F}\left\{\exp\left(-\frac{r^2}{r_o^2}\right)\exp\left(j\frac{k}{2z}r^2\right)\right\}\Big|_{\rho=r/\lambda z}. \tag{4.61}$$

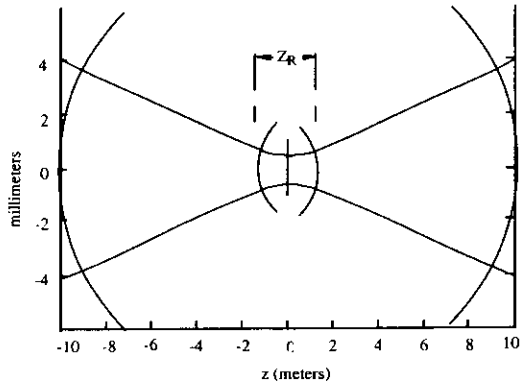



Fig. 4-12. Gaussian beam propagation. The bow-tie pattern (note the different scales for the horizontal and vertical axes) shows the $1/e^2$ intensity beam diameter for a $\lambda$=633 nm laser beam with a waist diameter of 1 mm. The circular arcs show the wavefront curvature at $z$=10 m, at the Rayleigh distance $z_R$ ($z$=1241 mm), and at the waist (where the wavefront is planar).

If the operand of the Fourier transform operator is written in the form





$$\exp\left[-\pi\left(\frac{1}{\pi r_o^2}+j\frac{-k}{2\pi z}\right)r^2\right],$$

the 2-D Fourier transform relationship ($a$ and $c$ real, a>0; see Ch. 2)

$$\exp\left[-\pi(a+jc)r^2\right] \leftrightarrow \frac{1}{a+jc}\exp\left[-\pi\frac{\rho^2}{a+jc}\right] \tag{4.62}$$

can be used in evaluating $\mathbf{U}(r;z)$. The final result, obtained after some manipulation, is

$$\mathbf{U}(r;z) = \exp(jkz)\exp[-j\psi(z)]\frac{1}{\eta(z)}\exp\left[j\frac{k}{2R(z)}r^2\right]\exp\left[-\left(\frac{r}{\eta(z)}\right)^2\right], \tag{4.63}$$

where

$$\eta(z) = r_o\sqrt{1+\left(\frac{z}{z_R}\right)^2}, \tag{4.64a}$$

$$R(z) = z\left[1+\left(\frac{z_R}{z}\right)^2\right], \tag{4.64b}$$

$$\Psi(z) = \tan^{-1}\left(\frac{z}{z_R}\right). \tag{4.64c}$$

The parameter $z_R$, the so-called *Rayleigh range* of the beam, is given by

$$z_R = \pi r_o^2/\lambda. \tag{4.65}$$

Note first that $\mathbf{U}(r;z)$ has a Gaussian magnitude for all values of z. The 1/e-amplitude radius $r_1(z)$ equals $r_o$ at the waist and approaches the value $\lambda z/\pi r_o$ for $z$ large compared to the Rayleigh range. Second, note that the phase is quadratic with $r$, the same as for a spherical wave in quadratic-phase form. The wave $\mathbf{U}(r;z)$ has, in fact, the same curvature as an expanding spherical wave with radius of curvature equal to $R(z)$ and is generally referred to as a *Gaussian spherical wave*. For large z, the radius of curvature $R(z)$ approximately equals z, and as $z\to 0$ and $z\to\infty$ the radius becomes infinite. It is easily shown that the radius of curvature attains its minimum value when $z = z_R$, i.e., at the Rayleigh range, where $R(z)$ equals $2Z_R$. At the Rayleigh range the 1/e-amplitude radius $r_1$ equals $r_o\sqrt{2}$. Additional information on Gaussian beams is contained in Ref. 4-9.

### 4.5.2 Converging Spherical Wave Illumination of an Aperture

We consider as a second example of Fresnel diffraction the singularly important case of an aperture illuminated by a converging monochromatic spherical wave. Assume, as illustrated in Fig. 4-13, that the spherical wave is converging toward the origin and that the aperture, characterized by complex transmittance $\mathbf{t}_o(x,y)$, is in a plane an adjustable distance $d$ in front of

the $z$=0 plane. The amplitude of the illuminating wave is obtained from Eq. (3.27) with $z$ set equal to $-d$, yielding

$$\mathbf{U}(x,y;-d+) = A\frac{\exp(-jkd)}{d}\exp\left[-j\frac{k}{2d}\left(x^2+y^2\right)\right]\mathbf{t}_o(x,y) \tag{4.66}$$

for the wave immediately to the right of the aperture. The amplitude of the propagated wave in a plane of arbitrary $z$ ($z$>0) is given by Eq. (4.27) with $z_2$ set equal to $z$, $z_1$ set equal to $-d$, and [from Eq. (4.28)] $Z$=$z$+$d$. Evaluating this expression for arbitrary $z$ is no easier than calculating the general Fresnel diffraction integral. However, for the special case $z$=0, the propagation distance Z equals $d$ and major simplifications result. Thus, from Eq. (4.27) and its expansion into the form given in Eq., (4.31),

$$\begin{aligned}\mathbf{U}(x,y;0) &= \mathbf{U}(x,y;-d+) ** \frac{\exp(jkd)}{j\lambda d}\exp\left[j\frac{k}{2d}\left(x^2+y^2\right)\right] \\ &= \frac{\exp(jkd)}{j\lambda d}\exp\left[j\frac{k}{2d}\left(x^2+y^2\right)\right]\mathcal{F}\left\{\mathbf{U}(x,y;-d+)\exp\left[j\frac{k}{2d}\left(x^2+y^2\right)\right]\right\}\Big|_{u=x/\lambda d \atop v=y/\lambda d}\end{aligned} \tag{4.67}$$

Substitution for $\mathbf{U}(x,y;-d+)$ and cancellation of common phase factors quickly leads to the result

$$\mathbf{U}(x,y;0) = \frac{A}{j\lambda d^2}\exp\left[j\frac{k}{2d}(x^2+y^2)\right]\mathbf{T}_o\left(\frac{x}{\lambda d},\frac{y}{\lambda d}\right), \tag{4.68}$$

where $\mathbf{T}_o(u,v)$ is the Fourier transform of $\mathbf{t}_o(x,y)$.

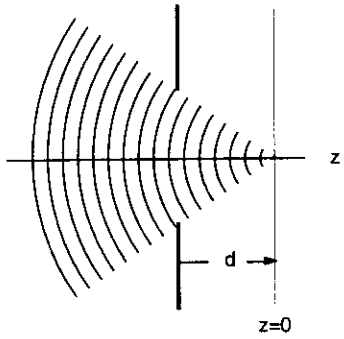



Fig. 4-13. Converging spherical wave illumination of an aperture. The complex amplitude in the plane of convergence is proportional to a scaled version of the Fourier transform of the aperture function, i.e., to the amplitude of the aperture Fraunhofer pattern.

The corresponding optical intensity is given by

$$I(x,y;0) = \left(\frac{1}{\lambda d^2}\right)^2\left|\mathbf{T}_o\left(\frac{x}{\lambda d},\frac{y}{\lambda d}\right)\right|^2, \tag{4.69}$$





which is seen to be proportional to the Fraunhofer intensity pattern associated with the aperture, scaled by the factor $\lambda d$.

As an example, consider the case of a square aperture of width $w$. Then $\mathbf{t}_o(x,y)$ = rect($x/w$,$y/w$), $\mathbf{T}_o(u,v) = w^2\mathrm{sinc}(wu,wv)$, and the complex amplitude in the $z$=0 plane is

$$\mathbf{U}(x,y;0) = \frac{A}{j\lambda}\exp\left[j\frac{k}{2d}\left(x^2+y^2\right)\right]\left(\frac{w}{d}\right)^2\mathrm{sinc}\left(\frac{x}{\lambda d/w},\frac{y}{\lambda d/w}\right). \tag{4.70}$$

Note that the magnitude of this wave depends on the aperture width $w$ and the convergence distance $d$ only through their ratio $w/d$. This condition implies that if $w$ and $d$ are increased proportionately—e.g., if both are doubled in size—the magnitude, and therefore the intensity, of the wave distribution in the plane of convergence remains the same. Only the quadratic phase factor in Eq. (4.70) changes. This conclusion is not unique to the case of a square aperture but is, in fact, quite general.

Equation (4.68) states that converging spherical wave illumination of an aperture produces, centered on the nominal point of convergence, a complex amplitude distribution proportional to the Fourier transform of the aperture transmittance function, scaled by the wavelength times the convergence distance. This result is one of the most important reached in the development of Fourier optics. It plays a central role in connection with the theory of image formation by spherical lenses, and it also leads to many of the important signal processing applications of optical systems. Although illustrated for the case of a wave converging to a point on the $z$-axis, it can be shown to hold also for the case where the point of convergence lies off-axis (see homework problem).

## 4.6 NON-MONOCHROMATIC WAVE PROPAGATION: THE QUASI-MONOCHROMATIC CONDITION

Thus far in the discussion of wave propagation it has been assumed that the wave is monochromatic. However, as noted in Chap. 3, truly monochromatic waves do not exist in nature. The important question thus arises, what must be done to modify the theory of wave propagation to properly take into account the finite spectral bandwidths of real-world wave fields? The answer depends strongly on how large the spectral bandwidth is. For narrowband wave amplitudes, the modification is much simpler than for the case of broadband wave amplitudes. The interested reader is referred to Refs. 4-7 and 4-8 for a discussion of the general theory. Fortunately, in many cases of practical interest the spectral bandwidth is sufficiently small that the theory for monochromatic wave fields provides accurate results even for time-varying phasors, and the various equations in this chapter can be used directly if $\mathbf{U}(x,y,z)$ is replaced by $\mathbf{U}(x,y,z,t)$, $\lambda$ by $\bar{\lambda}$, $k$ by $\bar{k}$, and $\nu$ by $\bar{\nu}$, where the bars denote the average values. In order for such a procedure to be allowable, the following two conditions on $\Delta\nu$ are imposed:

$$\frac{\Delta\nu}{\bar{\nu}} = \frac{\Delta\lambda}{\bar{\lambda}} << 1 \tag{4.71}$$

and

$$\Delta\nu << \frac{c}{d}, \tag{4.72}$$

where $d$ is the longest optical path involved in the analysis. The first of these conditions is simply the narrowband condition of Eq. (3.44). It is necessary if significant scale changes with

wavelength are to be avoided. The second, or *small path condition*, is a sufficient condition that assures that the phase shifts represented by the transfer function of Eq. (4.19) are accurate over the entire observation region for the entire range of values of $\lambda$. The two conditions taken together constitute the *quasimonochromatic condition*, and light satisfying them is referred to as being *quasimonochromatic*. As noted, light satisfying the quasimonochromatic condition can be treated in propagation and diffraction calculations as though it were truly monochromatic. Light from a laboratory laser often satifies the quasimonochromatic condition.

## 4.7 ACCURACY OF FRESNEL-REGIME CALCULATIONS AND THE EXACT FORM OF $\mathbf{h}_z(x,y)$*

The Fresnel regime wave propagation impulse response of Eq. (4.25), $\tilde{\mathbf{h}}_z(x,y)$, was obtained by imposing condition (4.20) and approximating the square root in the exponent of $\mathbf{H}_z(u,v)$ by the first two terms of its binomial expansion. Condition (4.20) can be interpreted in either of two ways: $\mathbf{U}(x,y;0)$ contains only spatial frequency components that are well below the frequency $1/\lambda$ in any direction, or, equivalently, the plane waves that make up $\mathbf{U}(x,y;0)$ through the angular spectrum superposition integral of Eq. (4.9) all have k-vectors that make suitably small angles with the positive $z$-axis.

The actual error in the phase of $\tilde{\mathbf{H}}_z(u,v)$ that is introduced through this approximation increases in proportion to $z$, and simple calculations suggest that the Fresnel regime approximation for $\mathbf{H}_z(u,v)$ should be used only under highly restrictive conditions. For example, if it is assumed that $z$ = 1.0 m and $\lambda$ = 500 nm, and if the magnitude of the phase error is to be no greater than $2\pi/10$, then $\rho=\sqrt{u^2+v^2}$ may take on values no greater than about 50/mm . The corresponding maximum angle the $k$-vector may make with the $+z$ axis is less than 1.5°.

However, in assessing the validity of the Fresnel-regime approximation, it is extremely important that its effect on the propagated wave field $\mathbf{U}(x,y;z)$ itself be considered rather than simply its effect on the Fourier transform $\hat{\mathbf{U}}(u,v;z)$. Stated differently, it is necessary to consider not just the relationship between $\mathbf{H}_z(u,v)$ and $\tilde{\mathbf{H}}_z(u,v)$ but—and much more importantly—also that between $\mathbf{U}(x,y;0)**\mathbf{h}_z(x,y)$ and $\mathbf{U}(x,y;0)**\tilde{\mathbf{h}}_z(x,y)$. One way of assessing this relationship is numerically, using, for example, the discrete Fourier transform and knowledge of $\mathbf{H}_z(u,v)$ and $\tilde{\mathbf{H}}_z(u,v)$ to calculate corresponding values for $\mathbf{U}(x,y;z)$. Alternatively, the results obtained numerically or, when possible, analytically using $\tilde{\mathbf{h}}_z(x,y)$ can be compared with experimental results obtained in the laboratory. In either case, a wide variety of experiments have shown that the use of the Fresnel kernel $\tilde{\mathbf{h}}_z(x,y)$ will yield remarkably good results except when small angle approximations are seriously violated, i.e., when angles involved in the problem exceed ±20° or so. The reasons are varied and complex and are not elaborated on here. However, relevant discussions are to be found in Ref. 4-1 as well as in various journal articles (see, e.g., Ref. 4-12).





It is also possible to compare $\tilde{\mathbf{h}}_z(x,y)$ with $\mathbf{h}_z(x,y)$ directly, for, as noted in Sect. 4.2, the exact propagation kernel can be determined by direct evaluation of $\mathcal{F}^{-1}\{\mathbf{H}_z(u,v)\}$ without the use of approximations, as is now shown.

We start with Eq. (4.14b) for the exact propagation transfer function, restated here:

$$\mathbf{H}_z(\rho) = \exp\left[j2\pi\frac{z}{\lambda}\sqrt{1-(\lambda\rho)^2}\right]. \tag{4.73}$$

Since this function has circular symmetry, its inverse Fourier transform can be evaluated using the zeroth-order inverse Hankel transform. Thus,

$$\mathbf{h}_z(r) = \mathcal{H}_0^{-1}\{\mathbf{H}_z(\rho)\} = 2\pi\int_0^\infty \rho\exp\left[j2\pi\frac{z}{\lambda}\sqrt{1-(\lambda\rho)^2}\right]J_0(2\pi r\rho)d\rho. \tag{4.74}$$

Stark (Ref. 4-10) points out that this integral can be cast in standard form and evaluated through the use of integral tables. The result, after modification to allow for negative values of $z$, is

$$\mathbf{h}_z(x,y) = \frac{\exp\left(jkz\sqrt{1+(x^2+y^2)/z^2}\right)}{j\lambda z\sqrt{1+(x^2+y^2)/z^2}}\frac{1}{\sqrt{1+(x^2+y^2)/z^2}}\left(1-\frac{1}{jkz\sqrt{1+(x^2+y^2)/z^2}}\right). \tag{4.75}$$

In classical diffraction theory this function equals the kernel of the general *Rayleigh-Sommerfeld diffraction formula*, obtained through the application of Green's theorem and a particular choice of the Green's function (see, e.g., Ref. 4-1). No approximations have been introduced in obtaining Eq. (4.75), which gives the exact expression for the propagation kernel.

In almost all cases of interest the condition $z>>\lambda$ is satisfied, the second term in the parentheses on the right-hand side of Eq. (4.75) is therefore quite small compared to the first, and $\mathbf{h}_z(x,y)$ is well approximated by

$$\mathbf{h}_z(x,y) \approx \frac{1}{j\lambda}\frac{1}{\sqrt{1+(x^2+y^2)/z^2}}\left[\frac{\exp\left(jkz\sqrt{1+(x^2+y^2)/z^2}\right)}{z\sqrt{1+(x^2+y^2)/z^2}}\right]. \tag{4.76}$$

Comparison of this equation with Eq. (3.24) shows the function in brackets to be the complex amplitude of our model for a unit amplitude spherical wave converging to and expanding from the origin [which, recall, is the location of the "impulse" giving rise to the impulse response $\mathbf{h}_z(x,y)$]. The spherical wave function is shifted 90° in phase by the factor $1/j$, reduced in amplitude by the factor $1/\lambda$, and multiplied by $1/\sqrt{1+(x^2+y^2)/z^2}$. This latter factor is the modulus of the so-called *obliquity factor*, which equals the cosine of the angle that a line from the origin—i.e., from the source point—to the observation point (x,y,z) makes with the positive $z$-axis. If the region of interest in the output plane is reasonably small compared to $z$, the obliquity factor is well approximated by unity and $\mathbf{h}_z(x,y)$ becomes

$$\mathbf{h}_z(x,y) \approx \frac{1}{j\lambda}\frac{\exp\left(jkz\sqrt{1+(x^2+y^2)/z^2}\right)}{z\sqrt{1+(x^2+y^2)/z^2}}. \tag{4.77}$$

It is now only a short step to the Fresnel regime approximation of Eq. (4.25). The remaining approximations are the same as those made in going from Eq. (3.24) to Eq. (3.27) in the representation of an expanding spherical wave: $z\sqrt{1+(x^2+y^2)/z^2}$ is replaced in the denominator by $z$ and in the exponent by $[z + (x^2+y^2)/2z]$, with the familiar result[4]

$$\mathbf{h}_z(x,y) \approx \frac{\exp(jkz)}{j\lambda z}\exp\left[j\frac{k}{2z}\left(x^2+y^2\right)\right], \tag{4.78}$$

the same expression as given in Eq. (4.25) for $\tilde{\mathbf{h}}_z(x,y)$.

It is interesting to note that the nature of the approximations used in this section differ from those imposed in Sect. 4.2, yet with the same result for the propagation kernel. In Sect. 4.2 a condition is imposed on the *input* wave: $\mathbf{U}(x,y;0)$ may not contain high spatial frequency components. In this section, on the other hand, third and higher-order terms in the binomial approximation to $\sqrt{x^2+y^2+z^2}$ are ignored, thereby imposing a condition on *the region of observation* in the output plane. As noted earlier, however, it is not so much the specific nature of the approximations on **H** or **h** that is important as is the effect these approximations have on the propagation integral, and experience suggests that this effect is quite often not significant.

## 4.8 CONCLUDING REMARKS

Confusion easily arises between the terms *interference* and *diffraction*. The term interference, as introduced in Chap. 3, is commonly used only when a small number of waves are involved or when the optical system is clearly an "interferometer." However, the distinction between *interference pattern* and *diffraction pattern* is often vague. For example, the Fraunhofer diffraction pattern from a two-pinhole mask is commonly described as a Young's interference experiment pattern. However, when the pinholes become large or are replaced by a random array of pinholes or by the 2-D transmittance function of a general 35 mm photo transparency, the resulting pattern is more often than not referred to as a diffraction pattern. No advice is given here, only an acknowledgment of the mix of terminology.

What is possibly the most useful equation in this text is the following:

$$\rho_o = \frac{\sin\theta_o}{\lambda}, \tag{4.79}$$

where $\rho_o$ is the radial spatial frequency associated with a plane wave whose k-vector makes an angle $\theta_o$ with the +z axis. For the case when the k-vector lies in the x-z plane, Eq. (4.79) assumes the form

$$u_o = \frac{\sin\theta_x}{\lambda}. \tag{4.80}$$

[4] These two approximations are often referred to collectively as the *Fresnel approximations*.





where $\theta_x$ is the angle the k-vector makes with the +z axis. Using this equation one can solve a remarkably large number of wave propagation problems without much effort. Keep it in mind. Examples of its use will appear in later chapters. Note also that $\sin\theta_o \approx \tan\theta_o \approx \theta_o$ for small values of $\theta_o$, simplifying results even more for many practical cases.

## REFERENCES


4-1 J. W. Goodman, *Introduction to Fourier Optics*, 2nd ed., McGraw-Hill, New York, 1996; Chaps. 3 and 4.

4-2 J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics*, Wiley, New York, 1978, Chap. 10.

4-3 M. Born and E. Wolf, *Principles of Optics*, 6th ed., Pergamon, New York, 1980;.Chap. 8.

4-4 H. Stark, "Theory and Measurement of the Optical Fourier Transform," in H. Stark, ed., *Applications of Optical Fourier Transforms*, Academic, New York, 1982, pp. 2-7.

4-5 B. B. Baker and E. T. Copson, *The Mathematical Theory of Huygens' Principle*, 2d edition, Clarendon, Oxford, 1949.

4-6 V. Ronchi, *The Nature of Light*, Harvard University Press, Cambridge, 1970.

4-7 J. W. Goodman, *Statistical Optics*, Wiley, New York, 1985, Ch.. 3.

4-8 A. S. Marathay, *Elements of Optical Coherence Theory*, Wiley, New York, 1982.

4-9 A. E. Siegman, *Lasers*, University Science Books, Mill Valley, California, 1986.

4-10 D. C. O'Shea, *Elements of Modern Optical Design*, Wiley, New York, 1985.

4-11 H. Stark, ed., *Applications of Optical Fourier Transforms*, Academic Press, New York, 1982, Ch. 1.

4-12 W. Southwell, "Validity of the Fresnel approximation in the near field," *J. Opt. Soc. Am.* vol. 71, pp. 7-14 (1981).


## PROBLEMS

4.1 Assume that a square aperture of width w is illuminated by a normally incident, unit amplitude, monochromatic plane wave.

(a) Find the angular spectrum $\hat{\mathbf{U}}(\alpha/\lambda,\beta/\lambda;0+)$ of the transmitted wave field immediately following the aperture and sketch its $\alpha/\lambda$ cross section, $\hat{\mathbf{U}}(\alpha/\lambda,0;0+)$.

(b) The angular spectrum of part (a) has periodic zeros along the $\alpha/\lambda$ axis. The direction corresponding to the first null from the origin can be expressed in terms of the angle $\theta$ that the $k$-vector makes with the $+z$ axis. Find that angle, assuming that $\lambda$ = 633 nm and $w$ = 2 μm.

4.2 A monochromatic wave is modeled by complex wave amplitude

$$\mathbf{U}(x,y;0) = 1 + \cos(2\pi f_o x).$$

(a) Find the angular spectrum $\hat{\mathbf{U}}(\alpha/\lambda,\beta/\lambda;0)$ associated with this wave.

(b) Assuming $\lambda$ = 600 nm, for what spatial frequency $f_o$, in $mm^{-1}$, does the cosine term describe two plane waves with propagation directions at $\pm 45^\circ$ to the $+z$ axis?

(c) Repeat part (b) for $\pm 90^\circ$.

(d) Assuming the propagation directions of part (b), find the plane-to-plane distance $z$ through which the wave must propagate to result in a complex wave amplitude of the form

$$U(x,y;z) = 1 + \exp(j\theta)\cos(2\pi f_o x),$$

where $\theta$ is some constant that depends on the specific value of $z$.

(e) The wave distribution of part (b) is incident on a square aperture 2 mm in width. Sketch the $\alpha/\lambda$ cross section of the angular spectrum of the resulting transmitted wave field, assuming $f_o = 50\ \text{mm}^{-1}$.

(f) If $f_o > 1/\lambda$, the cosinusoidal wave field is evanescent and dies out with distance. For what distance $z$ is it reduced in amplitude by a factor of 10 if $f_o$ is 1% larger than $1/\lambda$ [i.e., $f_o = 1.01(1/\lambda)$]?

4.3 Figure P4.13 illustrates the propagation of a plane wave whose $k$-vector lies in the $x$-$z$ plane. Assuming that the phase of the plane wave is zero in the $z$=0 plane, use simple trigonometry to show that its phase in the plane $z$=$z_o$ is consistent with Eq. (4.10).

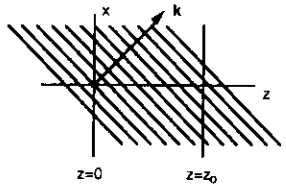



Fig. P4.13

4.4 An optical wave with complex amplitude $\mathbf{U}(x,y;0)$ has angular spectrum

$$\hat{\mathbf{U}}\left(\frac{\alpha}{\lambda},\frac{\beta}{\lambda};0\right) = \delta(\rho-\rho_o),$$

where

$$\rho = \sqrt{\left(\frac{\alpha}{\lambda}\right)^2 + \left(\frac{\beta}{\lambda}\right)^2}.$$

(a) Find $\mathbf{U}(x,y;0)$.

(b) Using the wave propagation transfer function $\mathbf{H}_z(u,v)$ given in Eq. (4.19), find the complex wave amplitude $\mathbf{U}(x,y;z)$ obtained by propagation of the above wave field through a distance $z$.

4.5 Prove the following statements:

(a) The condition $\sqrt{u^2+v^2} << 1/\lambda$ corresponds to the case of plane wave components with k-vectors that make small angles to the +z axis.

(b) The condition $\rho = \sqrt{u^2+v^2} = 1/\lambda$ corresponds to the case of plane wave components propagating in directions perpendicular to the $z$-axis.





4.6 A Ronchi ruling—a grating with a square-wave amplitude transmission profile (a Ronchi ruling is manufactured by scribing grooves in a piece of glass and filling the grooves with black paint or other opaque substance)—can be represented by the transmittance function

$$\mathbf{t}(x,y) = \left[\text{rect}\left(\frac{x}{d}\right) * \frac{1}{\Lambda}\text{comb}\left(\frac{x}{\Lambda}\right)\right]\text{rect}\left(\frac{x}{w},\frac{y}{h}\right), \qquad d < \Lambda,$$

where $\Lambda$ is the grating period and $d$ the grove width.

(a) Sketch the $x$-axis cross section of this transmittance function.

(b) Assuming such a grating to be illuminated by a unit amplitude, normally incident, monochromatic plane wave, find the resulting Fraunhofer intensity pattern a distance $z$ away. Sketch its $x$-axis cross section, labeling all critical distances. Assume that $w >> \Lambda$.

(c) Describe what happens in the special case when $d = \Lambda/2$.

4.7 An aperture consisting of two rectangular slits of width $X$ and height $Y$ and separated by a center-to-center spacing $D$ is illuminated by a unit-amplitude, normally incident plane wave at wavelength $\lambda$.

(a) Find the resulting Fraunhofer intensity pattern.

(b) Sketch $x$- and $y$-axis cross sections of this pattern, assuming $\lambda$ = 500 nm, $D$ = 5 mm, $X$ = 0.5 mm, $Y$ = 20 mm, and $z$ = 200 m.

4.8 A normally incident unit-amplitude, monochromatic plane wave of wavelength $\lambda$ illuminates an annular aperture (essentially a circular slit in a mask) with outer diameter $D$ and inner diameter $d$.

(a) Find the resulting Fraunhofer intensity pattern.

(b) Show that as the difference between $D$ and $d$ becomes small, the Fraunhofer pattern assumes the form

$$I(r;z) = \left(\frac{\pi A d}{\lambda z}\right)^2 J_o^2\left(\frac{\pi d r}{z}\right),$$

and find the constant $A$. (Hint: Represent the aperture by a delta function of the proper volume.)

4.9 Show by integrating $I(x,y;0)$ and $I(x,y;z)$ over the two respective $x$-$y$ planes that the factor $(1/\lambda z)^2$ in Eq. (4.34) is consistent with energy conservation.

4.10 An object with complex wave amplitude transmittance

$$t_o(x,y) = \left[\frac{1}{2} + \frac{1}{4}\cos(2\pi f_o x) + \frac{1}{4}\cos(2\pi 2 f_o x)\right]\text{rect}\left(\frac{x}{w},\frac{y}{w}\right)$$

is illuminated by a normally incident monochromatic plane wave.

(a) Find the angular spectrum for the transmitted wave and sketch its $\alpha/\lambda$ cross section, assuming that $f_o >> 1/w$. Be sure to label critical distances.

(b) In the vicinity of the object, assuming $w >> \lambda$, the transmitted wave field corresponds approximately to five beams of light of rectangular cross-section. What is the angle $\theta$ separating the nominal propagation directions of the beams, assuming $\lambda = 500$ nm, $f_o = 50$ mm$^{-1}$?

(c) For $f_o$ is sufficiently large the higher frequency components of the wave field amplitude become evanescent. For what approximate range of frequencies $f_o$ are only three beams produced, assuming the same wavelength given in part (b)? State any assumptions you make.

4.11 A mask containing two small pinholes at $x = \pm D$ is illuminated by a normally incident, unit-amplitude, monochromatic plane wave of wavelength $\lambda$. Find the intensity of the resulting wave a distance $z$ away, assuming propagation in the Fresnel regime, and sketch its $x$-axis cross-section. (Hint: Model the amplitude transmittance of the mask by a pair of delta functions.)

4.12 An opaque mask contains a circular opening of diameter $D$ obstructed by an off-axis disk of diameter $D/2$, as shown in the Fig. P4.12.

(a) Write an expression for the complex wave amplitude transmittance of this mask.

(b) Assuming the mask to be illuminated by a normally incident unit-amplitude plane wave at wavelength $\lambda$, find the Fraunhofer intensity pattern resulting in a plane a distance $z$ away.

(c) Explain how your answer to part (b) changes if the illuminating plane wave is obliquely incident on the mask, its $k$-vector lying in the $x$-$z$ plane and making an angle $\theta_x$ with the $+z$ axis.

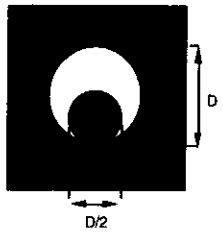



Fig. P4.12

4.13 In the Fresnel regime, propagation from the $z$=0 plane to a plane of arbitrary $z$ is expressed by the convolution formula $\mathbf{U}(x,y;z) = \mathbf{U}(x,y;0) ** \tilde{\mathbf{h}}_z(x,y)$, where $\tilde{\mathbf{h}}_z(x,y)$ is the Fresnel regime propagation kernel.





(a) Show, by proving that $\tilde{\mathbf{h}}_z(x,y) * \tilde{\mathbf{h}}_z^*(x,y) = \delta(x,y)$, that $\tilde{\mathbf{h}}_z^*(x,y)$ is the inverse propagation kernel in the Fresnel regime.

(b) Use the result from part (a) to prove that $\mathbf{U}(x,y;0)$ can be obtained from $\mathbf{U}(x,y;z)$ by means of the formula $\mathbf{U}(x,y;0) = \mathbf{U}(x,y;z) * \tilde{\mathbf{h}}_z^*(x,y)$.

(c) Show that, in the Fresnel regime, inverse propagation is represented in the spatial frequency domain by the multiplication with the *inverse propagation transfer function* $\tilde{\mathbf{H}}_z^*(u,v)$.

(c) Prove that, more generally, inverse propagation can be represented exactly in the spatial frequency domain by multiplication with $\mathbf{H}_z^*(u,v)$, where $\mathbf{H}_z(u,v)$ is the propagation transfer function given in Eq. (4.14) and, therefore, that inverse propagation in the space domain is represented exactly by convolution with the convolution kernel $\mathbf{h}_z(x,y)$ given in Eq. (4.75).

4.14 Show that condition (4.72) is consistent with the condition $d << \ell_c$, where $\ell_c$, the *coherence length* of the light, equals the distance the light travels in a time equal to the reciprocal bandwidth $1/\Delta\nu$.

4.15 As illustrated in Fig. P4.16, a spherical wave converging to a point at coordinates $(x_o,y_o,z_o)$ illuminates an aperture in the $z$=0 plane. Show that, as in the example of Sect. 4.5.2, the optical intensity resulting in the $z$=0 plane consists of the Fraunhofer diffraction pattern of the aperture, centered on the point of convergence.

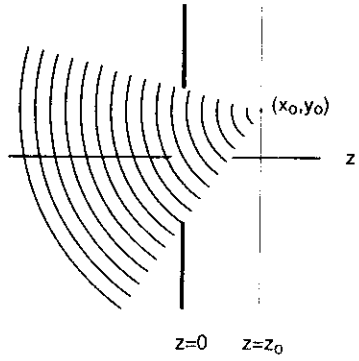



Fig. P4.16.

**INDEX - CHAPTER 4**

# 5. Lenses: Models and Properties

Lenses are among the most important elements in optical systems, for they can perform the amazing feat of forming images of light distributions. In addition, they have the remarkable property of being able to "calculate" the spatial Fourier transform of an input distribution, either a complex amplitude or a complex amplitude transmittance function. In this chapter we explore the properties associated with lenses and develop mathematical models for them and for the operations they can perform in simple configurations. We begin with an introduction to the most basic properties and models for lenses, from both a ray optics and a wave optics perspective. We then analyze the ability of a lens to form images of complex amplitude and optical intensity distributions and investigate in detail the Fourier transforming property . We conclude with a brief look at the effects of non-monochromatic light, methods for modeling "thick" lenses, and certain practical aspects of lenses and their use.

## 5.1 THIN SPHERICAL LENSES: BASIC PROPERTIES AND MODELS

Only a small set of properties and associated rules is needed for one to be able to understand the behavior of lenses in a variety of systems. Basic ray-optics properties are discussed first, followed by a wave-optics-based model for a lens illuminated by monochromatic light.

### 5.1.1 Ray-Optics Properties: Collimating, Focusing, and Imaging

A simple single-element spherical lens is illustrated in Fig. 5-1. It consists of a piece of glass (it could be plastic) with two spherical surfaces of radii $R_1$ and $R_2$. The centers of curvature of the surfaces lie along the *optical axis* of the lens, taken in the remainder of the text as being coincident with the z-axis.

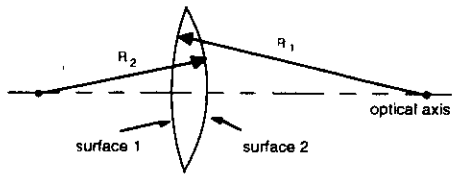



Fig. 5-1. Simple spherical lens.

As illustrated in Fig. 5-2(a), if a such lens is illuminated by a plane wave or, from a ray optics viewpoint, by a collimated beam of light, the result is a focusing of the light to what approximates a point in a plane a distance $f$ beyond the lens. The distance $f$ is called the *focal length* of the lens, and the plane is the *back focal plane*. The point where the optical axis intersects the back focal plane is refereed to as the *back focal point*. The location of the point of focus is determined by passing a ray through the center of the lens at the incidence angle and noting





where that ray intersects the focal plane, as shown in the figure.[1] Similarly, a point of light in a plane a focal distance $f$ in front of such a lens—in the *front focal plane*—produces a collimated beam of light that travels in the direction established by a line from the source point through the center of the lens. This configuration is illustrated in Fig. 5-2(b). If the point source lies at the *front focal point* of the lens—at the intersection of the optical axis with the front focal plane—the resulting collimated beam travels along the z-axis.

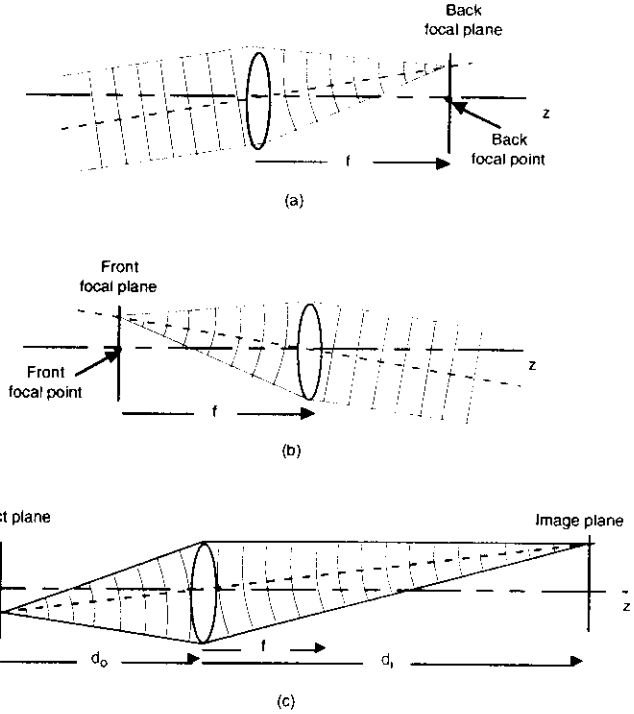



Fig. 5-2. Effect of a positive lens on light rays and waves: (a) focusing, (b) collimation, (c) imaging.

Figure 5-2(c) illustrates the imaging of light from an object point to a corresponding image point. From a wave-optics viewpoint the effect of the lens is to convert diverging spherical waves into converging spherical waves. In this text we are concerned almost exclusively with the imaging

[1] The precise determination of the focal distance $f$ is discussed in Sect. 5.4. The notion of a focal plane is reasonable only if the angle of incidence the light waves make with the lens normal is relatively small. If the angle of incidence is too large, off-axis aberrations become noticeable, and the light is focused to a blurred rather than to a compact point. Furthermore, the point of best focus may be closer to or farther away from the lens than the nominal back focal plane. Note that a large angle of incidence is inconsistent with condition (4.20) assumed in deriving the Fresnel-regime propagation formulas.

of objects that are confined to planes of constant $z$. If paraxial conditions (see box) are satisfied, the image of such an object is itself planar, lying in a second plane of constant $z$. The distances between the planes containing the object, image, and lens are governed by the *Gaussian imaging formula,*[2] or *lens law*,

$$\frac{1}{d_o}+\frac{1}{d_i}=\frac{1}{f}, \tag{5.1}$$

where $d_o$ is the signed distance (see box) from the object plane to the lens plane and where $d_i$ is the signed distance from the lens plane to the image plane. Because of the tightly coupled relationship between these two distances, the two planes are said to be *conjugate* to one another, as are object and image points themselves. Figure 5-2(c) illustrates for the case where $f$, $d_o$, and $d_i$ are all positive. The *transverse magnification* of the image, defined by

$$M=-\frac{d_i}{d_o}, \tag{5.2}$$

is negative for the single-lens system, agreeing with the inversion, or 180° rotation, of the image relative to the object. Thus, for this system, if the object point is at coordinates $(x_o,y_o)$ in the object plane, the conjugate image point is at coordinates $(-d_ix_o/d_o,-d_iy_o/d_o)$ or $(Mx_o,My_o)$ in the image plane.

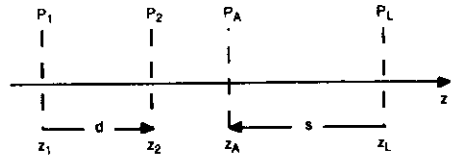
**Paraxial Optics**

Except when the effects of diffraction are to be considered, systems of lenses are analyzed by means of ray-trace methods: lines representing rays from points in object space are traced through a diagram of the optical system to corresponding points in image space. Ray-trace analysis is greatly simplified if small-angle approximations can be used for trigonometric functions. Such approximations are valid if all rays of concern make only small angles with the optical axis and change angle at lens elements only by small amounts. Such conditions will generally be met if the lens elements share a common axis and if all rays of interest remain sufficiently close to the optical axis—i.e., if all rays are *paraxial*. If these paraxial conditions are not satisfied, more exact methods of analysis must be used. Throughout this text paraxial conditions will be assumed to be satisfied unless otherwise stated.

**Signed Distances**

Using signed distances in connection with imaging systems is a little like giving a traveler compass-point directions for driving from Philadelphia to Chicago. If he already has a sense of where Chicago lies, it is adequate to tell him that it is about 700 miles distant. But if he doesn't know the territory, it may help him a great deal to know that his destination lies toward the west!

With reference to the figure, signed distances along the z-axis can be signified by notation of the form $d=\overline{P_1P_1}$. This equation should be read "$d$ is the distance along the $z$ axis from plane $P_1$ to plane $P_2$." If $P_2$ lies to the right of $P_1$, then $d$ is positive and, as in the figure, is denoted by an arrow pointing to the right—i.e., in the $+z$ direction. The signed distance $s$ in the figure is negative—a condition conveyed by the arrow's pointing toward the left—indicating that $s=\overline{P_LP_A}$, i.e., the distance from plane $P_L$ to plane $P_A$. Since $P_A$ lies to the left of plane $P_L$, this distance is in the -$z$ direction, and $s$ is therefore negative. The signed distances $d$ and $s$ can be defined in terms of the $z$-coordinates of the planes of concern. Thus, $d=z_2-z_1>0$, and $s=z_A-z_L<0$. In Figs. 5-2, 5-3, and 5-4, the arrows associated with focal distance $f$ point to the right, signifying that $f$ is positive for these lenses. In Fig. 5-5, $f$ is negative, a fact conveyed by the leftward-pointing arrow.




In the analysis of the single-lens imaging systems in this chapter, object distance $d_o$ is defined as the signed distance from the object plane to the lens plane: $d_o=\overline{P_oP_L}=z_L-z_o$. This quantity will always be positive if the object plane lies to the left of the lens (as it must if we demand that light travel left to right). Similarly, image distance $d_i$ is given by $d_i=\overline{P_LP_i}=z_i-z_L$. This distance is positive if the image lies to the right of the lens. It can, however, be negative if $z_i-z_L<0$, i.e., if the image—virtual in this case—lies to the left of the lens [Fig. 5-3(c)].

Signed distances are denoted in figures in the text through the use of single-headed arrows. If only the magnitude of a distance is of concern (a rare case), a two-sided arrow is used.

Other examples of imaging with a single positive-focal-length lens are illustrated in Fig. 5-3. These examples illustrate basic techniques in ray-trace methods. Note specifically that in each case three rays are traced from the object point to its conjugate image point:

(a) A ray parallel to the optical axis. This ray, after passage through the lens, passes through the back focal point of the lens.

(b) A ray that passes undeviated through the center of the lens.

(c) A ray that passes through, or whose extension passes through [Fig. 5-3(c)], the front focal plane of the lens, emerging parallel to the optical axis on the right side of the lens.

The intersection of these rays denotes the location of the conjugate image point. Clearly only two of the three rays are actually needed to locate this intersection. Note that in Fig. 5-3(c), where the object lies between the front focal plane and the lens, the image is *virtual*: the rays of light diverging outward to the right of the lens only *appear* to be coming from the image point; the light rays do not in fact pass through the image point as they do for the real images points in Figs. 5-3(a) and (b).

Two more definitions are of occasional us to us. With reference to Fig. 5-4, if a lens of diameter D and focal length $f$ is used to bring collimated light to a focus, the lens is said to operate

[2] Named after German mathematician Carl Friedrich Gauss, whose work on imaging in the geometrical optics regime led to this formula.

with an *f-number* (designated in this text by $F^{\#}$; in many photographic texts the symbol f/# is used) given by

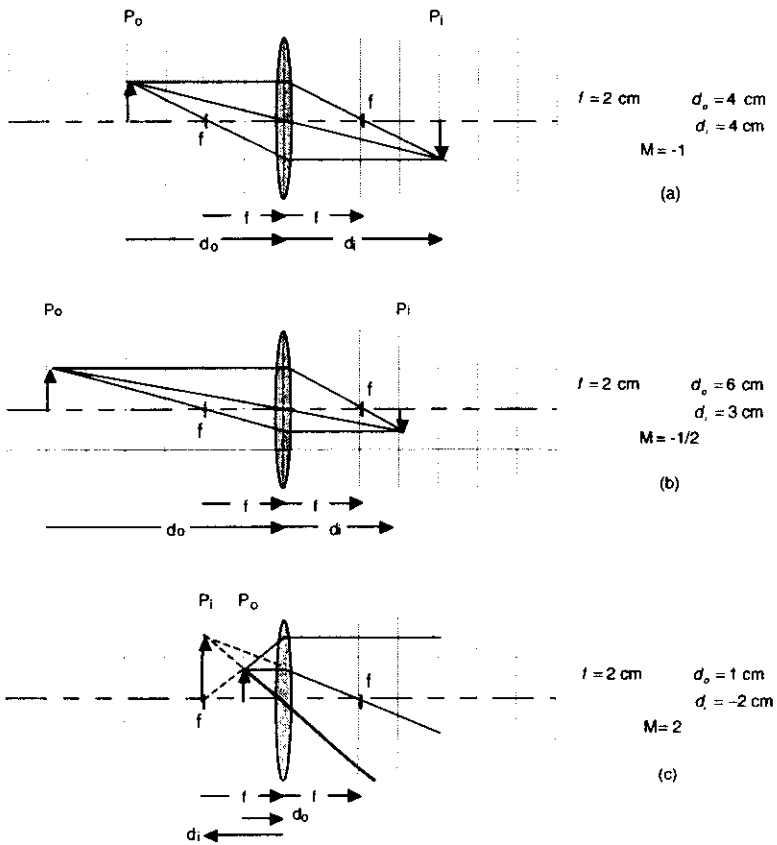



Fig. 5-3. Examples of imaging with a single thin lens of positive focal length. The gridwork is in centimeters.

$$F^{\#} = \frac{f}{D}. \tag{5.3}$$

Cameras with adjustable apertures can operate at different f-numbers. Thus, a camera with a 50 mm focal-length lens and an aperture adjusted to a diameter of 6.25 mm is said to operate with an f-number of 8.0 (often written f/8). Note that the larger the aperture is, the smaller the f-number,





and thus smaller f-numbers correspond to greater amounts of light reaching the focal plane. If light is brought to a focus in a plane other than the back focal plane of the lens, it is still possible to speak of an *effective* f-number, $F^{\#}_{eff}$, for the light cone. Thus, assuming the lens in Fig. 5-2(c) has diameter D, we can say that the cone of light brought to a focus at the image point has an effective f-number given by $F^{\#}_{eff} = d_i / D$. In some cases, an effective f-number may be applied to the cone of light on the object side of the lens.

Another frequently-used number characterizing a lens is its *numerical aperture*, designated NA and given by the sine of the half-angle $\alpha$ shown in Fig. 5-4:

$$\text{NA} = \sin\alpha. \tag{5.4}$$

Since $\alpha$ can never exceed 90°, the maximum possible value for NA is unity.[3] It is easily shown that if $\alpha$ is small the numerical aperture of the lens is approximately equal to $1/(2F^{\#})$. Larger NAs correspond to greater amounts of light reaching the focal plane. Later we shall see that, because of the wave nature of light and corresponding diffraction effects, the higher the NA the better the resolution provided by a lens used for imaging.

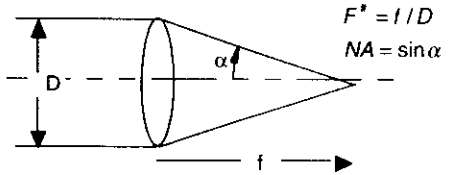



Fig. 5-4. Parameters relating to the f-number and numerical aperture of a lens.

In most cases our concern is with *positive* lenses, where $f$ is greater than zero. Figures 5-2 and 5-3 illustrate the action of positive lenses. Occasionally, negative lenses, where $f$ is less than zero, are of interest to us. A negative lens, illustrated in Fig. 5-5, takes an incident plane wave and converts it into a spherical wave that appears to be expanding from a point a focal distance in front of the lens, i.e., from a *virtual* source point. Negative lenses are thinner in the middle than at the edges, whereas positive lenses are thicker in the middle.

[3] This statement is not true in all cases in that it assumes that the lens is operating in air. The actual definition of the numerical aperture includes a multiplicative factor equal to the refractive index $n$ of the medium in which the light is being focused: $NA = n\sin\alpha$. Thus, if the focusing is in oil having a refractive index of 1.5, an NA of 1.5 is theoretically possible (though in practice an NA exceeding 1.3 is rarely achieved).

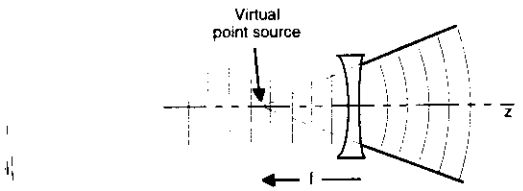



Fig. 5-5. A negative lens, when illuminated by a plane wave, produces an expanding spherical wave. The focal length $f$ for such a lens, being negative, is represented by an arrow pointing in the -z direction.

### 5.1.2 Wave-Optics Properties: Complex Amplitude Transmittance Function

We now consider the wave optics properties of lenses. One way to develop a mathematical model for the effect of a spherical lens is to calculate the retardation of a monochromatic light wave introduced by the lens as the wave passes through it. Such an approach is taken by Goodman (Ref. 5-1, Sect. 5.1.1), and the interested reader is referred to his analysis. An alternative approach is simply to note that normally incident plane wave illumination of the lens produces what is essentially a segment of a converging spherical wave. Assuming that the incident plane wave has unit magnitude, i.e., $U_{inc}(x,y) = 1$, then immediately after the lens the transmitted wave must still have unit magnitude. Its phase, however, must be that of a spherical wave converging to a point a distance $f$ away. If the quadratic-phase approximation of Sect. 3.2.2 is employed, the resulting complex amplitude is, from Eq. (3.27), $U_{trans}(x,y) = \exp[j(k/2z_o)(x^2+y^2)]$, with $z_0$ set equal to $-f$. The resulting quadratic-phase-approximation expression for the *complex amplitude transmittance of a thin spherical lens of focal length f* is thus given by

$$t_{lens}(x,y) = \exp\left[-j\frac{k}{2f}\left(x^2+y^2\right)\right]. \tag{5.5}$$

The action of the lens is to modify the phase of the incident wave but not its magnitude. Note that the quadratic phase approximation to the spherical wave factor is consistent with wave propagation in the Fresnel regime.

The form of Eq. (5.5) implies that the lens has infinite extent. The finite extent of an actual lens can be modeled through inclusion of an appropriate wave transmittance function representing a confining aperture. This transmittance function is generally referred to as the *pupil function* of the lens, and is denoted by $\mathbf{p}(x,y)$. Thus,

$$t_{lens}(x,y) = \mathbf{p}(x,y)\exp\left[-j\frac{k}{2f}\left(x^2+y^2\right)\right], \tag{5.6}$$

where $\mathbf{p}(x,y)$ equals unity within the lens aperture and zero otherwise. In most cases $\mathbf{p}(x,y)$ will be a circ or cylinder function of appropriate diameter, or perhaps a rectangle function.

An assumption implicit in describing a lens by Eqs. (5.5) and (5.6) is that the lens is "thin." According to this assumption, a ray entering the lens at transverse coordinates $(x_o,y_o)$,





although in fact bent by both front and back surfaces of the lens, exits at effectively the same coordinates $(x_o,y_o)$. Equivalently, the action performed by the lens can be thought of as being confined to a plane: the physical thickness of the lens is ignored. The assumption that lenses are thin will generally be made in this text in the modeling of optical systems. Figure 5-6(a) illustrates the behavior of a thin positive lens: all bending of the rays is modeled as occurring in a single plane. Figure 5-6(b) shows a symbolic representation for a thin positive lens that is often used in diagrammatic representations of optical systems. As illustrated in Fig. 5-6(c), real lenses have finite thickness, and rays are bent at both front and back (and, possibly, internal) surfaces. However, as discussed in Section 5.4, thick, and even compound or multiple-component lenses, can be modeled by the thin-lens expression.

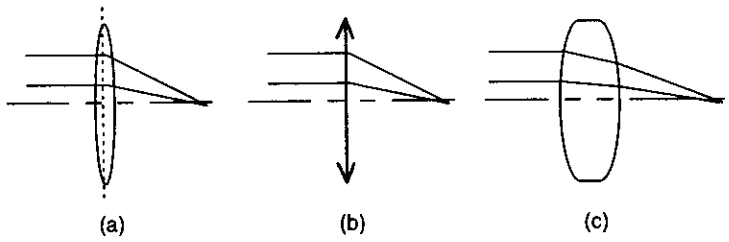



Fig. 5-6. Concept of the "thin" lens. Rays passing through a thin lens can be assumed to be bent at a single surface, as in (a). A thin positive lens is sometimes modeled diagrammatically as illustrated in (b). With a thick lens, illustrated in (c), rays must be modeled as being bent at both front and back (and, possibly, internal) surfaces.

## 5.2 SINGLE-LENS IMAGING AND THE EFFECT OF DIFFRACTION

Having developed some understanding of the basic geometrical-optics properties of thin spherical lenses, we now work to obtain mathematical models for their imaging property in the wave optics regime. We consider the imaging of complex amplitudes and wave intensity distributions, both coherent and incoherent. The system illustrated in Fig. 5-7 serves as the basis for our analyses. The lens is assumed to have finite diameter aperture modeled by pupil function $\mathbf{p}(x,y)$; object and image distances $d_o$ and $d_i$ are assumed to satisfy the lens law, Eq. (5.1).

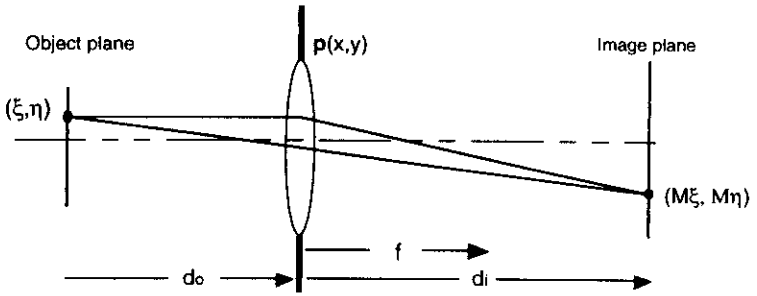



Fig. 5-7. Single-lens imaging system.

### 5.2.1 Imaging Complex Amplitudes

Consistent with the superposition property of light waves, we express the imaging operation for complex amplitude distributions in terms of a general superposition integral:

$$\mathbf{U}_{im}(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{U}_{obj}(\xi,\eta)h(x,y;\xi,\eta)\,d\xi\,d\eta, \tag{5.7}$$

where $\mathbf{U}_{im}(x,y)$ and $\mathbf{U}_{obj}(x,y)$ denote[4] the image- and object-plane complex amplitude distributions and where $h(x,y;\xi,\eta)$ is the response at image-plane coordinates $(x,y)$ to a unit-impulse point source at object-plane coordinates $(\xi,\eta)$, sometimes referred to as the *complex amplitude point spread function*

According to the rules of geometrical optics, a point with coordinates $(\xi,\eta)$ in the object plane is imaged, in the ideal case, to a point with coordinates $(M\xi,M\eta)$ in the image plane, where $M$ is given by Eq. (5.2). Assuming that all light from the object is brought to the image plane with perfect resolution and geometrical fidelity, an object intensity distribution $I_{obj}(x,y)$ produces image distribution $(1/M^2)I_{obj}(x/M,y/M)$, which we define as the *geometrical optics image intensity distribution*, $I_g(x,y)$:

$$I_g(x,y) = \frac{1}{M^2} I_{obj}\left(\frac{x}{M},\frac{y}{M}\right), \tag{5.8}$$

the factor $1/M^2$ being consistent with conservation of energy principles. For a monochromatic object wave field $\mathbf{U}_{obj}(x,y)$, the corresponding *geometrical optics image amplitude distribution*, denoted $\mathbf{U}_g(x,y)$, is defined by

[4] The more compact subscript notation is now used since our concern is with complex amplitudes in specific planes.





$$\mathbf{U}_g(x,y) = \frac{1}{M}\mathbf{U}_{obj}\left(\frac{x}{M},\frac{y}{M}\right). \tag{5.9}$$

Noting that $\mathbf{U}_g(x,y)$ can be obtained by evaluating the superposition integral

$$\mathbf{U}_g(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{U}_{obj}(\xi,\eta)\frac{1}{M}\delta\left(\xi-\frac{x}{M},\eta-\frac{y}{M}\right)d\xi\,d\eta, \tag{5.10}$$

we define the impulse response of the geometrical optics imaging system by

$$h_g(x,y;\xi,\eta) = \frac{1}{M}\delta\left(\xi-\frac{x}{M},\eta-\frac{y}{M}\right). \tag{5.11}$$

The definitions just presented represent an idealization of imaging system operation, and $h_g(x,y;\xi,\eta)$, as given in in Eq. (5.11), although useful to us in later formulations, is often an unrealistic approximation to the actual impulse response of an imaging system. To begin with, it ignores the effects of distortion and aberrations, generally present to some degree. In addition, and of great importance to us in this text, it ignores the effects of diffraction by the lens aperture and does not take the phase of the light waves into account.

To calculate the *diffraction-limited*[5] impulse response for imaging in the wave optics regime with the system of Fig. 5-7, we assume that the object amplitude distribution is given by a unit impulse, $\delta(x\text{-}\xi,y\text{-}\eta)$. The resulting Fresnel-regime complex amplitude incident on the lens is then, from Eq. (4.25a)

$$\mathbf{U}_l(x,y) = \frac{1}{\lambda d_o}\exp\left\{j\frac{k}{2d_o}\left[(x-\xi)^2+(y-\eta)^2\right]\right\}, \tag{5.12}$$

where an unimportant constant phase factor $(1/j)\exp(jkd_o)$ has been discarded.[6] Multiplication by the lens transmittance function, Eq. (5.6), yields the transmitted wave amplitude

$$\mathbf{U}_l'(x,y) = \mathbf{U}_l(x,y)\mathbf{p}(x,y)\exp\left[-j\frac{k}{2f}\left(x^2+y^2\right)\right], \tag{5.13}$$

and Fresnel-regime propagation through the distance $d_i$ [Eq. (4.30)] then yields (again discarding unimportant constant phase factors) the wave-optics impulse response for single-lens imaging, denoted $h_w(x,y;\xi,\eta)$:

[5] By diffraction-limited it is meant that the performance of the imaging lens is limited strictly by diffraction effects and not by aberrations. Diffraction-limited performance is the best that one can achieve with an imaging lens of a given aperture.

[6] In the expressions for Fresnel- and Fraunhofer-regime wave propagation, the factor $(1/j)\exp[jkz]$ is retained because the behavior of the wave distribution with changes in $z$ is of interest. In the current case, however, the distance $d_o$ is fixed, the corresponding constant phase factor $(1/j)\exp(jkd_o)$ has no practical significance, and it is consequently dropped for the sake of simplicity and compactness in the resulting expression. In the remainder of the text, constant phase factors will be dropped consistently.

$$h_w(x,y;\xi,\eta) = \frac{1}{\lambda d_i}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{U}_i(\mu,\zeta)\exp\left\{j\frac{k}{2d_i}\left[(x-\mu)^2+(y-\zeta)^2\right]\right\}d\mu\,d\zeta. \qquad (5.14)$$

Substituting from Eqs. (5.13) and (5.12) and invoking the lens law, Eq. (5.1), yields, after some effort, the result

$$h_w(x,y;\xi,\eta) = \exp\left[j\frac{k}{2d_o}\left(\xi^2+\eta^2\right)\right]\exp\left[j\frac{k}{2d_i}\left(x^2+y^2\right)\right] \times\frac{1}{\lambda^2 d_o d_i}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathbf{p}(\mu,\zeta)\exp\left\{-jk\left[\left(\frac{\xi}{d_o}+\frac{x}{d_i}\right)\mu+\left(\frac{\eta}{d_o}+\frac{y}{d_i}\right)\zeta\right]\right\}d\mu\,d\zeta \qquad (5.15)$$

The integral expression has the basic form of a Fourier transform integral, allowing $h_w(x,y;\xi,\eta)$ to be written in the more compact form

$$h_w(x,y;\xi,\eta) = \exp\left[j\frac{k}{2d_o}\left(\xi^2+\eta^2\right)\right]\frac{1}{\lambda^2 d_o d_i}\mathbf{P}\left(\frac{\xi}{\lambda d_o}+\frac{x}{\lambda_{d_i}},\frac{\eta}{\lambda d_o}+\frac{y}{\lambda d_i}\right)\exp\left[j\frac{k}{2d_i}\left(x^2+y^2\right)\right], \qquad (5.16)$$

where $\mathbf{P}(u,v)$ is the Fourier transform of the pupil function: $\mathbf{P}(u,v) = \mathcal{F}\{\mathbf{p}(x,y)\}$.

The above expression for $h_w(x,y;\xi,\eta)$ is the principal result of this calculation. Although it can be written in a variety of forms, $h_w(x,y;\xi,\eta)$ cannot be written as a function of the differences of coordinates, i.e., as a function of $(x-\xi)$, $(y-\eta)$, alone. We thus conclude that the single-lens imaging system of Fig. 5-7 is space-variant and must be represented by the general superposition integral of Eq. (5.7) rather than by a simpler convolution integral. In part the space-variance comes from the inversion and magnification of the image, although in the next chapter we shall see that that aspect of its shift-variance is relatively unimportant. Of much more fundamental importance is the phase factor involving object-space coordinates, $\exp[j(k/2d_o)(\xi^2+\eta^2)]$.

How this factor can introduce significant space-variance in the imaging operation is seen by way of a specific example. Assume that the imaging system is characterized by pupil function

$$\mathbf{p}(x,y) = \mathrm{rect}(x/w, y/w) \qquad (5.17)$$

and the object by

$$\mathbf{U}_{obj}(x,y) = \delta\left(x+x_o+\frac{1}{2}\frac{\lambda d}{w},y\right)+\delta\left(x+x_o-\frac{1}{2}\frac{\lambda d}{w},y\right), \qquad (5.18)$$

corresponding to a pair of point sources at mean location $x_o$ and separated by $\lambda d/w$. It is not difficult to show that optical intensity in the image plane is given by





$$I_{im}(x,y) = \left(\frac{w}{\lambda d}\right)^4\left|\mathrm{sinc}\left(\frac{x+x_o+\lambda d/2w}{\lambda d/w},\frac{y}{\lambda d/w}\right) +\exp\left[-j2\pi\frac{1}{w}x_o\right]\mathrm{sinc}\left(\frac{x+x_o-\lambda d/2w}{\lambda d/w},\frac{y}{\lambda d/w}\right)\right|^2, \qquad (5.19)$$

Consider this expression for two specific values of $x_o$. For $x_o=0$, $I_{im}(x,y)$ evaluates to

$$I_{im}(x,y)\big|_{x_o=0} = \left(\frac{w}{\lambda d}\right)^4\left|\mathrm{sinc}\left(\frac{x+\lambda d/2w}{\lambda d/w},\frac{y}{\lambda d/w}\right)+\mathrm{sinc}\left(\frac{x-\lambda d/2w}{\lambda d/w},\frac{y}{\lambda d/w}\right)\right|^2, \qquad (5.20)$$

whereas for $x_o=w/2$ the result is

$$I_{im}(x,y)\big|_{x_o=w/2} = \left(\frac{w}{\lambda d}\right)^4\left|\mathrm{sinc}\left(\frac{x+\lambda d/2w}{\lambda d/w},\frac{y}{\lambda d/w}\right)-\mathrm{sinc}\left(\frac{x-\lambda d/2w}{\lambda d/w},\frac{y}{\lambda d/w}\right)\right|^2. \qquad (5.21)$$

The sum of the two sinc functions within the squared-modulus bars has changed to a difference. The $x$-axis cross-sections of these two intensity distributions are shown in Fig. 5-8. The effect of moving the two points off axis, and by an amount equal to only half the lens aperture width, is seen to be quite significant. If the imaging operation were shift-invariant, a change in $x_0$ would have no effect on the shape of the image: it would only shift its position.

Equation (5.16) for $h_w(x,y;\xi,\eta)$ takes diffraction at the finite lens aperture into account. If the lens is sufficiently large compared to the object and the object spatial frequency bandwidth is sufficiently small (see homework problem), essentially *all* of the light from the object passes through to the image plane, and the effect of the lens aperture can be neglected. To determine the corresponding form for $h_w(x,y;\xi,\eta)$, let $\mathbf{p}(x,y) = 1$, in which case $\mathbf{P}(u,v) = \delta(u,v)$. Substituting for $\mathbf{P}(.,.)$ in Eq. (5.16), letting $-d_i/d_o = M$, and applying the scaling relationship for impulses (Section 2.1) results in the equation

$$h_\infty(x,y;\xi,\eta) = \exp\left[j\frac{k}{2d_o}\left(\xi^2+\eta^2\right)\right]\frac{1}{M}\delta\left(\xi-\frac{x}{M},\eta-\frac{y}{M}\right)\exp\left[j\frac{k}{2d_i}\left(x^2+y^2\right)\right], \qquad (5.22)$$

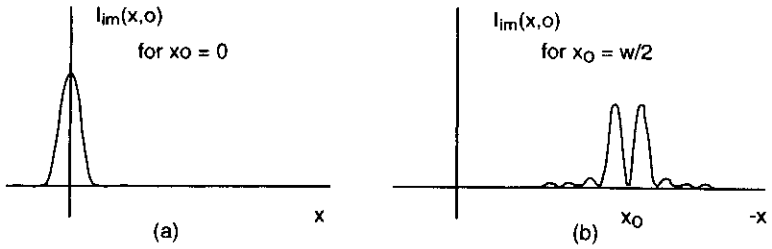



Fig. 5-8. Effect of space variance on image of two closely-spaced monochromatic point sources: (a) points centered on-axis; (b) points centered off axis a distance equal to half the lens aperture width.

where the subscript ∞ is suggestive of the infinite-aperture assumption. The sifting property of the delta function allows the first phase factor to be rewritten as $\exp\left[j\left(k/2M^2d_o\right)\left(x^2+y^2\right)\right]$. Combination of this factor with the remaining quadratic phase factor and application once again of the lens law yields

$$h_\infty(x,y;\xi,\eta)=\frac{1}{M}\delta\left(\xi-\frac{x}{M},\eta-\frac{y}{M}\right)\exp\left[j\frac{k}{2d'}\left(x^2+y^2\right)\right], \tag{5.23}$$

where the distance $d'$ is given by

$$d'=(d_i-f). \tag{5.24}$$

The wave-optics impulse response associated with an infinite-aperture lens is thus seen to equal the geometrical-optics impulse response of Eq. (5.9) times a quadratic phase factor.

It should be emphasized that Eq. (5.23), if it is to be useful, requires not only that diffraction at the aperture be negligible but also that distortion and the effects of aberrations introduced by the lens can be ignored. It can provide reasonable results only when paraxial conditions hold and when essentially all of the light from the object passes through the lens.

## 5.3 FOURIER TRANSFORMING PROPERTY

We now turn attention to the second important property of a spherical lens, its Fourier transform property. There are two distinct cases to consider. In one, the input to the Fourier transform "operation" is a complex amplitude $\mathbf{U}_o(x,y)$ propagating toward the lens. In the other, the input is considered to be a thin object, characterized by complex amplitude transmittance $\mathbf{t}_o(x,y)$, that is inserted in a plane before or behind a lens illuminated by light originating at a point source.

### 5.3.1 Complex Amplitude Distribution as Input

We consider first the Fourier transform of complex amplitude distributions. The transform operation is related to the collimating and, particularly, the focusing property of a spherical lens. Recall that a collimated beam of light is focused to a point in the back focal plane of a lens. In wave terminology, a plane wave is focused to a point; in mathematical terminology a linear phase factor is mapped into a delta function, one essential feature of the Fourier transform. Conversely, light





from a point in the front focal plane of a lens is collimated: a point source is mapped to a plane wave, or, in mathematical terms, a delta function is mapped to a linear phase factor. Thus, we could argue, from front focal plane to back focal plane, a lens performs a Fourier transform operation: linear phase factors map into impulses, impulses into linear phase factors. To investigate this relationship more fully, we consider the system of Fig. 5-9. To allow for maximum generality, we assume a complex amplitude distribution $\mathbf{U}_o(x,y)$ in a plane an arbitrary distance $d$ in front of a spherical lens of focal lens $f$ and derive the complex amplitude in the back focal plane. In anticipation that this distribution will bear a Fourier transform relationship to $\mathbf{U}_o(x,y)$, we denote the back focal plane distribution $\mathbf{U}_{FT}(x,y)$.

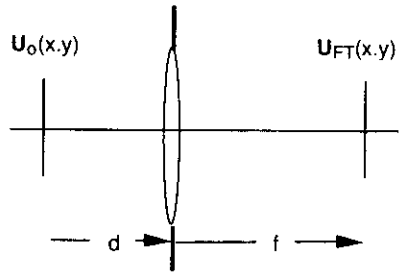



Fig. 5-9. System for Fourier transformation of complex amplitude distribution. Complex amplitude $\mathbf{U}_{FT}(x,y)$ in the back focal plane of the lens is proportional to the scaled Fourier transform of the complex amplitude $\mathbf{U}_0(x,y)$.

We begin by denoting the complex amplitude incident on the lens by $\mathbf{U}_l(x,y)$ and that immediately after the lens by $\mathbf{U}_l'(x,y)$. The relationship between these two distributions is given by

$$\mathbf{U}_l'(x,y)=\mathbf{U}_l(x,y)\exp\left[j\frac{k}{2f}\left(x^2+y^2\right)\right], \tag{5.25}$$

where by neglecting a pupil function we have temporarily assumed that the lens is infinite in extent. Repeating the analysis of Section 4.5.2, it is easy to show that the complex amplitude in the back focal plane, $\mathbf{U}_{FT}(x,y)$, is given by

$$\mathbf{U}_{FT}(x,y)=\frac{\exp(jkf)}{j\lambda f}\exp\left[j\frac{k}{2f}(x^2+y^2)\right]\hat{\mathbf{U}}_l\left(\frac{x}{\lambda z_o},\frac{y}{\lambda z_o}\right). \tag{5.26}$$

Furthermore, through an application of Eqs. (4.2) and (4.22), we know that

$$\hat{\mathbf{U}}_l(u,v)=\hat{\mathbf{U}}_o(u,v)\exp\left(j2\pi\frac{f}{\lambda}\right)\exp\left[-j\pi\lambda f\left(u^2+v^2\right)\right]. \tag{5.27}$$

Combining Eqs. (5.26) and (5.27) and discarding unimportant constant phase factors yields the result

$$\mathbf{U}_{FT}(x,y) = \exp\left\{ j\frac{k}{2f}\left[\left(1-\frac{d}{f}\right)(x^2+y^2)\right]\right\}\frac{1}{\lambda f}\hat{\mathbf{U}}_o\left(\frac{x}{\lambda f},\frac{y}{\lambda f}\right). \tag{5.28}$$

We thus reach the extremely important conclusion that the complex amplitude in the back focal plane of a positive lens is proportional to a scaled version of the 2-D Fourier transform of a complex amplitude in front of the lens. It is common to speak of this Fourier transform distribution as lying in *the Fourier transform plane*, or, alternatively, in the Fraunhofer plane, since $\mathbf{U}_{FT}(x,y)$ has the basic form of a Fraunhofer amplitude distribution with scaling parameter $f$.

In the majority of cases it is the optical intensity of the wave field in the Fourier transform plane that is of interest. This distribution is given by

$$I_{FT}(x,y) = \left(\frac{1}{\lambda f}\right)^2\left|\hat{\mathbf{U}}_o\left(\frac{x}{\lambda f},\frac{y}{\lambda f}\right)\right|^2. \tag{5.29}$$

If it is desired to remove the quadratic phase factor of Eq. (5.28) from the Fourier transform plane amplitude distribution, either of two methods can be used. The more complicated method involves the insertion of a second lens in the system, just before the Fourier transform plane, as illustrated in Fig. 5-10. This lens, which must have a focal length $f_2$ obtained by solving

$$\frac{1}{f_2} = \frac{1}{f}\left(1-\frac{d}{f}\right), \tag{5.30}$$

is characterized by a complex amplitude transmittance that just cancels the quadratic phase factor in Eq. (5.28). The result is a Fourier plane distribution given by

$$\mathbf{U}_{FT}(x,y) = \frac{1}{\lambda f}\hat{\mathbf{U}}_o\left(\frac{x}{\lambda f},\frac{y}{\lambda f}\right). \tag{5.31}$$

It can be shown that the focal length $f_2$ given by Eq. (5.30) is such that light from an on-axis point source a distance $d$ in front of the Fourier transform lens results in a plane wave in the Fourier transform plane.

In the second, much easier method, the distance $d$ is made to equal the focal distance $f$, in which case the quadratic phase factor of Eq. (5.28) evaluates to unity. The result, once again, is the Fourier plane distribution given by Eq. (5.31).





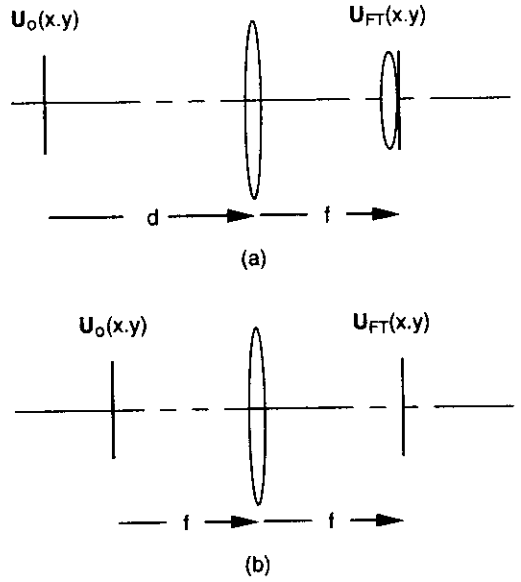



Fig. 5-10. The quadratic phase factor accompanying the Fourier transform amplitude distribution can be removed in two ways: (a) by inserting a second lens, of appropriate focal length, in the back focal plane of the Fourier transform lens; (b) by setting the distance $d$ equal to the focal length $f$.

For an example of an optical Fourier transform operation, assume a complex amplitude $\mathbf{U}_o(x,y)$ given by

$$\mathbf{U}_o(x,y) = \mathrm{rect}\left(\frac{x}{w},\frac{y}{w}\right)\exp(jk\alpha x). \tag{5.32}$$

a distance $d$ in front of the lens. This function represents a truncated plane wave traveling with $k$-vector in the $x$-$z$ plane making an angle $\cos^{-1}\alpha$ with the $+x$ axis or, alternatively, an angle $\sin^{-1}\alpha$ with the $+z$ axis. Consistent with Fresnel-regime conditions, that angle is assumed to be small: $\alpha \ll 1$. Noting again that $k=2\pi/\lambda$, we can immediate write the Fourier transform of $\mathbf{U}_o(x,y)$:

$$\begin{aligned}\hat{\mathbf{U}}_o(u,v) &= w^2\,\mathrm{sinc}(wu,wv) ** \delta\left(u-\frac{\alpha}{\lambda},v\right)\\ &= w^2\,\mathrm{sinc}\left(\frac{u-\alpha/\lambda}{1/w},\frac{v}{1/w}\right)\end{aligned} \tag{5.33}$$

with the result, upon substitution in Eqs. (5.28) and (5.29),

$$U_{FT}(x,y) = \exp\left\{ j\frac{k}{2f}\left[\left(1-\frac{d}{f}\right)(x^2+y^2)\right]\right\}\frac{w^2}{\lambda f}\,\mathrm{sinc}\left(\frac{x-\alpha f}{\lambda f/w}, \frac{y}{\lambda f/w}\right) \tag{5.34}$$

and

$$I_{FT}(x,y) = \left(\frac{w^2}{\lambda f}\right)^2 \mathrm{sinc}^2\left(\frac{x-\alpha f}{\lambda f/w}, \frac{y}{\lambda f/w}\right). \tag{5.35}$$

The Fourier plane distribution is thus seen to consist of a sinc-squared-distributed spot of light positioned off-axis in the $x$-direction an amount proportional to $\alpha$ and to $f$.

The effect of the finite lens aperture can in principle be calculated by including a pupil function $\mathbf{p}(x,y)$ in the lens factor of Eq. (5.25). However, the result of such a formulation is extremely complicated and lends little insight into actual limitations imposed by the aperture. Instead, a method based on a combination of ray and wave optics concepts is used. With reference to Fig. 5-11(a), consider light from the input distribution, $\mathbf{U}_o(x,y)$, that is focused to a point on axis in the back focal plane. The amplitude of the light at that point conveys information about the amplitude of the zero spatial frequency component of $\mathbf{U}_o(x,y)$. Now from a geometrical optics standpoint, we know that only rays of light parallel to the optical axis are focused to that point. Figure 5-11(a) shows the region of the input distribution, denoted by the brace, from which that light can come. It is evident from the figure that if the lens is smaller than the input distribution, some of the input amplitude distribution—that portion sufficiently far off-axis—is not included in the Fourier transform "calculation" for the zero-spatial-frequency component: In order for $\mathbf{U}_{FT}(0,0)$ to properly represent the zero-frequency Fourier component of $\mathbf{U}_o(x,y)$, it is at a minimum necessary that the lens be larger in diameter than the object.

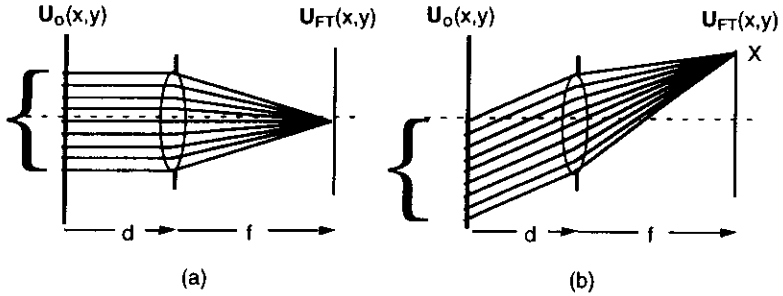



Fig. 5-11. Effect of finite-aperture lens in Fourier transform operation. Only a limited-spatial-extent portion of the input wave distribution, indicated by the single brace, can contribute to the light focused to a point in the back focal plane: (a) on-axis focus corresponding to the zero spatial-frequency component; (b) off axis focus corresponding to a frequency X/λf.

More generally, consider light focused to a point a distance $X$ off axis in the Fourier plane, as shown in Fig. 5-11(b). Only rays traveling at angle $\theta=\tan^{-1}(X/f)$ to the $z$-axis are focused to that point. It is clear from the figure that for sufficiently large values of $X$ there will be regions of the input—again, those lying outside the region designated by the brace—that cannot contribute to the light beam focused to that point. With an aperture around the lens, some portions of a centered





input distribution cannot be "seen" from points far off axis in the back focal plane. Indeed, for small on-axis objects and sufficiently large values of $X$, the entire input distribution may be obscured from view by the aperture. We must conclude that the Fourier analysis performed by the apertured lens is "frequency variant," in the sense that only restricted portions of the input may contribute to different portions of the Fourier plane spectrum. The basic effect—that a smaller and smaller fraction of the input is "seen" from points in the focal plane that are farther and farther from the axis—is referred to as *vignetting* by the lens aperture.[7]

Assuming that the lens has diameter $D_L$ and that the input distribution has diameter $D_O$, it is easily shown (see homework problem) that the maximum spatial frequency—in any direction for a circular aperture—for which the focal plane distribution accurately represents the Fourier spectrum of the input is given by

$$\rho_1 = \frac{D_L - D_O}{2\lambda d}. \tag{5.36}$$

For spatial frequencies below $\rho_1$ there is essentially no vignetting. The spatial frequency above which *no* light reaches the Fourier plane, even though the input may have non-zero Fourier components at higher frequencies, is given by

$$\rho_2 = \frac{D_L + D_O}{2\lambda d}. \tag{5.37}$$

For spatial frequencies above $\rho_2$ vignetting is essentially complete: virtually no light reaches the Fourier transform plane. Between these two frequencies—a range corresponding to partial vignetting—only parts of the input distribution contribute to the Fourier plane distribution. It is clear that if a lens is to be used to produce an accurate Fourier spectrum or Fraunhofer pattern of an input complex amplitude, it is necessary that the lens be larger than the input and that the distance $d$ separating input and lens be sufficiently small. The optimum configuration has the input distribution $\mathbf{U}_o(x,y)$ right up against the Fourier transform lens, i.e., with $d$ equal to zero.

The relationship between the angular spectrum representation of a wave field and the Fourier transform property of the lens should be emphasized. In particular, the angular spectrum analysis tells us that each Fourier component of the input wave field corresponds to a plane wave traveling in a particular direction. The higher the frequency of the component, the greater the angle the propagation vector makes with the +z axis. But the effect of a lens is to take a plane wave and focus it to a point in the back focal plane, the distance from the axis being determined by the tilt angle of the wave. It is this action by the lens on the plane wave components of the input wave field that produces the Fourier transform distribution in the back focal plane of the lens.

### 5.3.2 Transparent Object as Input

In the previous section the notion of a lens performing a Fourier transform operation on an input complex amplitude $\mathbf{U}_o(x,y)$ was introduced. In this section we consider the Fourier transform property of a lens in connection with thin transparent objects illuminated by spherical waves. The

[7] The effects of vignetting are sometimes observed in photographs made with inexpensive cameras: the corners of the image are noticeably darker than is the center. Vignetting was common with cameras built early in the history of photography; it is sometimes deliberately introduced in photographic printing to yield an "old-time" appearance to the photographic print.

input to the transform operation is no longer considered to be a complex amplitude but, rather, a complex amplitude transmittance function $t_o(x,y)$.

Two canonical cases, illustrated in Fig. 5-12, are considered. In the first case, illustrated in Fig. 5-12(a), a thin object transparency with complex amplitude transmittance $t_o(x,y)$ is placed in the path of a converging spherical wave produced by the illumination of a positive lens with monochromatic light coming from a point source S. In the second case, illustrated in Fig. 5-12(b), the object lies in front of the lens and is illuminated by a diverging spherical wave originating at the point source. It will be shown that in both cases a complex amplitude proportional to the Fourier transform of the object transmittance function appears in the plane of convergence of the spherical wave, centered on the image S′ of the source point. These two cases lie at the heart of the Fourier transformation of object transparencies using spherical lenses and provide general principals applicable to many imaging systems and to optical signal processing systems. A special case, examined separately, is that where the point source is at infinity. In that case, the object is illuminated by a normally incident plane wave, and, as might be expected from the analysis of the previous section, the Fourier transform distribution appears in the back focal plane of the lens. A second special case, as shall be seen, occurs when the object transparency lies in the front focal plane of the lens.

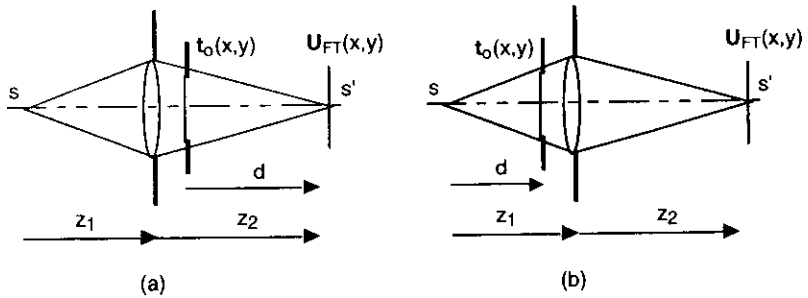



Fig. 5-12. Fourier transforming property of a spherical lens with illumination from a point source: (a) object in converging beam produced by lens; (b) object in expanding beam illuminating lens. In both cases a scaled version of the Fourier transform of the object transmittance function $t_o(x,y)$ appears centered on the image S′ of the point source S.

### Analysis: Object in Converging Beam

For the case illustrated in Fig. 5-12(a), a monochromatic on-axis point source $S$ is imaged by a positive lens of focal length $f$ to point $S'$ in the conjugate image plane. A thin planar object with complex amplitude transmittance $t_o(x,y)$ is positioned in the converging beam a distance $d$ in front of the point of convergence. The object is assumed to be sufficiently small in diameter that it is completely illuminated by the converging beam. The wave incident on the object transparency is modeled by complex amplitude $\exp[-j(k/2d)(x^2+y^2)]$, representative in the quadratic phase approximation of a spherical wave, of unit amplitude, converging to the point $S'$.[8] For convenience the magnitude factor $A$ is set equal to unity. Recalling Eq. (4.67) and the analysis of Sect. 4.5.2,

[8] For this model to be valid, the wave incident on the object transparency must be essentially free from diffraction effects introduced by the lens aperture. This condition will not be achieved if the distance from the lens to the object is too large, a point discussed later.





we can immediately write an expression for the complex amplitude, again denoted $U_{FT}(x,y)$, in the plane to which the illuminating wave is converging:

$$\mathbf{U}_{FT}(x,y) = \exp\left[j\frac{k}{2d}(x^2+y^2)\right]\left[\frac{1}{\lambda d}\mathbf{T}_o\left(\frac{x}{\lambda d},\frac{y}{\lambda d}\right)\right], \tag{5.38}$$

where $\mathbf{T}_o(.,.)$ denotes the Fourier transform of $\mathbf{t}_o(.,.)$ and where an unimportant constant phase factor has been discarded. The notation $U_{FT}(x,y)$ is again used to suggest the basic Fourier transform nature of this distribution. The corresponding optical intensity is given by

$$I_{FT}(x,y) = |\mathbf{U}_{FT}(x,y)|^2 = \left(\frac{1}{\lambda d}\right)^2\left|\mathbf{T}_o\left(\frac{x}{\lambda d},\frac{y}{\lambda d}\right)\right|^2, \tag{5.39}$$

which is recognized as the Fraunhofer pattern associated with the object transparency, with scale factor $\lambda d$.

As an example, consider the case where the object is a Ronchi ruling. Such an object, illustrated in Fig. 5-13, is modeled by complex amplitude transmittance

$$\mathbf{t}_o(x,y) = \left\{\left[\mathrm{rect}\left(\frac{x}{\delta}\right) * \frac{1}{\Lambda}\mathrm{comb}\left(\frac{1}{\Lambda}\right)\right]1(y)\right\}\mathrm{rect}\left(\frac{x}{w_x},\frac{y}{w_y}\right), \qquad \delta < \Lambda. \tag{5.40}$$

The Fourier transform of this function is easily shown to be given by

$$\mathbf{T}_o(u,v) = \left(\frac{w_x w_y \delta}{\Lambda}\right)\sum_{-\infty}^{\infty}\mathrm{sinc}\left(\frac{n}{\Lambda/\delta}\right)\mathrm{sinc}\left(\frac{u-n/\Lambda}{1/w_x},\frac{v}{1/w_y}\right). \tag{5.41}$$

Inserting this expression in Eq. (5.38) and assuming that diffraction order overlap is sufficiently small that cross-product terms can be neglected, the following Fourier plane optical intensity is obtained:

$$I_{FT}(x,y) = \left(\frac{w_x w_y \delta}{\lambda d \Lambda}\right)^2\sum_{n=-\infty}^{\infty}\mathrm{sinc}^2\left(\frac{n\delta}{\Lambda}\right)\mathrm{sinc}^2\left(\frac{x-n\lambda d/\Lambda}{\lambda d/w_x},\frac{y}{\lambda d/w_y}\right). \tag{5.42}$$

This intensity pattern, illustrated in cross-section in Fig. 5-14, is seen to consist of a multitude of diffraction orders separated by a distance inversely proportional to the fundamental period of the ruling.

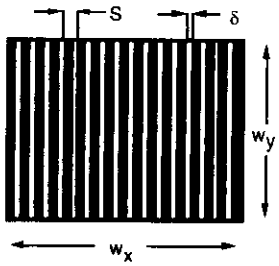



Fig. 5-13. Ronchi ruling.

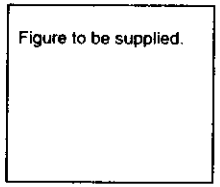



Fig. 5-14. Fourier-plane optical intensity associated with Ronchi ruling.

Unlike the case considered in the previous section, there is no position for the object transparency for which the quadratic phase factor in Eq. (5.38) disappears. However, if desired, that factor can be removed by placing a second positive lens of focal length $d$ in the Fourier transform plane. The complex wave amplitude $\mathbf{U}'_{FT}(x,y)$ immediately following the second lens is then given by

$$\mathbf{U}'_{FT}(x,y) = \mathbf{U}_{FT}(x,y)\exp\left[-j\frac{k}{2d}(x^2+y^2)\right] = \frac{1}{\lambda d}\mathbf{T}_o\left(\frac{x}{\lambda d},\frac{y}{\lambda d}\right) \tag{5.43}$$

It is assumed in the above analysis that the wave incident on the object has the form of a converging spherical wave. However, this assumption is not valid if the effects of diffraction introduced by the lens aperture itself are too large. From the discussion of Sect. 4.5.2 it should be clear that a wave with complex amplitude $\exp[-j(k/2d)r^2]\mathrm{cyl}(r/w)$ in the plane of the lens undergoes a continuous transition to a wave having the form proportional to $\mathrm{somb}(wr/\lambda d)$ in the plane of convergence: only in the plane of the lens does the wave actually have the assumed form, and if the object is positioned too close to the point of convergence, the result obtained is far from the Fourier transform relationship developed above. However, it can be demonstrated both experimentally and numerically that results well-approximated by Eqs. (5.38), (5.39), and (5.43) are obtained if the diameter of the object is sufficiently small compared to that of the illuminating beam. With reference to Fig. 5-15, restricting the object to lie within an illumination cone whose apex lies roughly half the distance to $S'$ should generally yield good results.





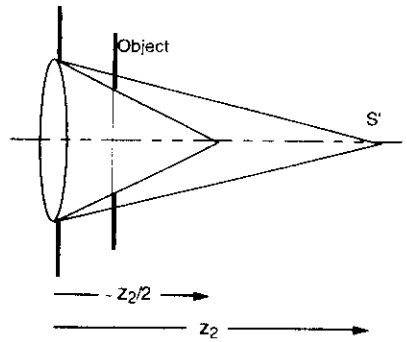



Fig. 5-15. Usually the object will be smaller than the illumination cone provided by the lens. In order to prevent undesired effects of diffraction by the lens aperture, the object should be fully illuminated by a cone of light (shaded) that is significantly smaller than the actual cone.

### Analysis: Object in Diverging Beam

In the second canonical case, illustrated in Fig. 5-12(b), the object is inserted a distance $d$ from the point source, in the path of the beam expanding toward the lens. Again, the Fourier transform of the object transmittance function appears centered on the image point $S'$ conjugate to the source point $S$. In this case, however, the proof is more difficult. The approach taken here is to establish first that there is a *virtual Fourier transform distribution centered on the source point S itself.* It is the real image of this virtual Fourier transform that then appears centered on $S'$.

The wave amplitude distribution illuminating the object again has the form of a spherical wave of radius $d$, but this time expanding: $\mathbf{U}_{\mathrm{inc}}(x,y) = \exp[j(k/2d)(x^2+y^2)]$, to within a constant proportionality factor assumed to be unity. The complex amplitude immediately following the object is given by

$$\mathbf{U}_{obj}(x,y) = \exp\left[j\frac{k}{2d}(x^2+y^2)\right]\mathbf{t}_o(x,y). \tag{5.44}$$

The form of the virtual Fourier transform centered on point $S$ is determined by means of a trick: we ask ourselves the question, What wave amplitude distribution in the plane of the source would, after propagating through distance $d$, produce the *same* distribution in the plane of the object as that given by Eq. (5.44)? To determine this effective source-plane distribution, denoted $\mathbf{U}_{eff}(x,y)$, we exploit the technique of inverse propagation, introduced in Sect. 4.2, by convolving $\mathbf{U}_{obj}(x,y)$ with the inverse propagation kernel appropriate for plane-to-plane distance $d$, $(-1/j\lambda d)\exp(-jkd) \times\exp[-j(k/2d)(x^2+y^2)]$. The analysis proceeds almost exactly as that of Sect. 4.5.2 but with the distance $z_o$ replaced by $-d$. The result, obtained after discarding unnecessary constant phase factors, is the complex amplitude

$$\mathbf{U}_{eff}(x,y) = \exp\left[-j\frac{k}{2d}(x^2+y^2)\right]\left[\frac{1}{\lambda d}\mathbf{T}_o\left(\frac{-x}{\lambda d},\frac{-y}{\lambda d}\right)\right]. \tag{5.45}$$

This distribution is nearly the same as that of the distribution of Eq. (5.38): only the orientation of the Fourier transform has been changed—by a 180° rotation—and the diverging spherical wave factor has been replaced by a converging spherical wave factor. It should be emphasized that this Fourier transform distribution is *virtual* in the sense that, as in the case of a virtual image, the wave amplitude distribution $\mathbf{U}_{eff}(x,y)$ does not actually exist in the plane of the source point: it only appears to be there—for example, to an observer looking back through the object toward the source point.[9]

The complex amplitude distribution $\mathbf{U}_{FT}(x,y)$ in the plane containing $S'$ is determined by calculating the image in that plane of the virtual Fourier transform distribution $\mathbf{U}_{eff}(x,y)$. This calculation is made by assuming the lens to be infinite in extent and using the imaging relationship given by Eq. (5.10) with impulse response $h_\infty(x,y;\xi,\eta)$ given by Eq. (5.23):

$$\mathbf{U}_{FT}(x,y)=\exp\left[j\frac{k}{2d'}(x^2+y^2)\right]\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathbf{U}_{eff}(\xi,\eta)\frac{1}{M}\delta\left(\xi+\frac{x}{M},\eta+\frac{y}{M}\right)d\xi\,d\eta, \tag{5.46}$$

where

$$d'=z_2-f \tag{5.47}$$

and

$$M=\frac{z_2}{z_1}. \tag{5.48}$$

Evaluating the integral yields

$$\begin{aligned}\mathbf{U}_{FT}(x,y)&=\exp\left[j\frac{k}{2d'}(x^2+y^2)\right]\frac{1}{M}\mathbf{U}_{eff}\left(\frac{-x}{M},\frac{-y}{M}\right)\\&=\exp\left[j\frac{k}{2d'}(x^2+y^2)\right]\exp\left[-j\frac{k}{2dM^2}(x^2+y^2)\right]\frac{1}{\lambda Md}\mathbf{T}_o\left(\frac{x}{\lambda Md},\frac{y}{\lambda Md}\right).\end{aligned} \tag{5.49}$$

The two quadratic phase factors can be combined to yield the final result

$$\mathbf{U}_{FT}(x,y)=\exp\left[-j\frac{k}{2d''}(x^2+y^2)\right]\left[\frac{1}{\lambda Md}\mathbf{T}_o\left(\frac{x}{\lambda Md},\frac{y}{\lambda Md}\right)\right], \tag{5.50}$$

where $d''$ satisfies the equation

$$\frac{1}{z_1-d}+\frac{1}{z_2+d''}=\frac{1}{f}. \tag{5.51}$$

[9] A simple experiment serves to illustrate. Place a grating or other diffracting object transparency in front of your eye and look through it toward a bright point of light some distance away: a spot on a wall illuminated by a laser beam, for example, or a distant street lamp at night. Apparently centered on the point of light will be the Fraunhofer pattern associated with the object transparency. For the case of the street-lamp illumination, the Fraunhofer pattern will be polychromatic, being made up of the λ-scaled distributions associated with the different wavelengths of the source.





The corresponding optical intensity is

$$I_{FT}(x,y)=\left(\frac{1}{\lambda Md}\right)^2\left|\mathbf{T}_o\left(\frac{x}{\lambda Md},\frac{y}{\lambda Md}\right)\right|^2. \tag{5.52}$$

Not too surprisingly, the main result of the imaging operation is a magnification of the Fourier transform distribution and an inversion through the origin. The factor $1/\lambda Md$ in Eqs. (5.50) and (5.52) is consistent with energy conservation for light traveling the (lossless) path from the object transparency to the Fourier transform plane. The distance $d''$ can be either positive or negative, depending respectively on whether the object lies before or behind the front focal plane of the lens.

Two special cases are of interest to us. In the first, the object transparency is positioned exactly one focal distance in front of the lens. In this case, the image of the object appears at infinity, $d''$ is infinite, and the quadratic phase factor of Eq. (5.50) simplifies to unity. It is easily shown that in this case $Md=f$, and $\mathbf{U}_{FT}(x,y)$ thus assumes the particularly simple form

$$\mathbf{U}_{FT}(x,y)=\frac{1}{\lambda f}\mathbf{T}_o\left(\frac{x}{\lambda f},\frac{y}{\lambda f}\right). \tag{5.53}$$

In words, if the object transparency is in the front focal plane of the lens, the Fourier transform distribution produced in the Fourier transform plane is governed by a scale factor that depends on $\lambda$ and $f$ alone and is free from the quadratic phase factor that is characteristic of all other cases. This condition is true regardless of the distance $z_1$ to the source point (and, therefore, of the distance $z_2$ to the Fourier transform plane). The form of Eq. (5.53) is of course reminiscent of Eq. (5.25) for the Fourier plane distribution assuming complex amplitude $\mathbf{U}_o(x,y)$ in the front focal plane of the lens. However, it must be emphasized that the physical conditions are quite different, and that in the case of the transparent object treated in this section, the Fourier transform is in general not in the back focal plane of the lens but, rather, in whatever plane is conjugate to the source point.

In the second special case of interest the point source $S$ is removed to infinity. The wave illuminating the object is therefore a plane wave, and by the lens law the Fourier transform distribution moves to the back focal plane of the lens. This special configuration is frequently analyzed in contemporary texts on optics—so much so, in fact, that there is a widespread misconception that the only place that an optical Fourier transform can be found is in the back focal plane of the lens.

To determine the form of the Fourier plane distribution in this case, we start with Eq. (5.50) and note that as $z_1$ grows large, both $z_2$ and the factor $Md$ approach $f$ in value, with the result

$$\mathbf{U}_{FT}(x,y)=\exp\left[-j\frac{k}{2d'}(x^2+y^2)\right]\left[\frac{1}{\lambda f}\mathbf{T}_o\left(\frac{x}{\lambda f},\frac{y}{\lambda f}\right)\right], \tag{5.54}$$

where $d'$ is again the distance from the back focal plane to the plane that is conjugate to the object, as illustrated in Fig. 5-16. Note that in this figure $d$ is now used to denote the distance between object and lens rather than between source and object. The distance $d'$ is easily expressed in terms of this object-lens distance $d$, resulting in the alternative expression

$$\mathbf{U}_{FT}(x,y) = \exp\left[j\frac{k}{2f}\left(1-\frac{d}{f}\right)(x^2+y^2)\right]\left[\frac{1}{\lambda f}\mathbf{T}_o\left(\frac{x}{\lambda f},\frac{y}{\lambda f}\right)\right]. \qquad (5.55)$$

It is particularly clear from this latter equation that if the object is placed in the front focal plane of the lens, making $d=f$, the quadratic phase disappears, and the simpler quadratic-phase-free Fourier transform relationship of Eq. (5.53) results: $\mathbf{U}_{FT}(x,y) = (1/\lambda f)\mathbf{T}_o(x/\lambda f, y/\lambda f)$.

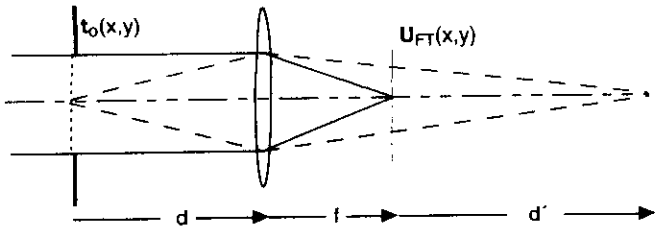



Fig. 5-16. Configuration with plane wave illumination of object. The solid lines show how a plane wave is mapped into a point of light, whereas the dotted lines show how a point of light is mapped into a spherical wave.

In calculating $\mathbf{U}_{FT}(x,y)$ for the case of Fig. 5-12(b) it has been tacitly assumed that the lens is infinite in diameter. However, as for the case in Section 5.3.1 of the complex amplitude $\mathbf{U}_o(x,y)$ being transformed by a lens, the effect of a finite-diameter lens is to reduce and otherwise modify the amplitude of the Fourier plane distribution for points sufficiently far off axis, or, in other words, for object spatial frequencies that are sufficiently high. Consider an object Fourier component of spatial frequency $f_o$. In the plane of the virtual Fourier transform distribution there appears to be a corresponding point of light a distance $-X_o$ off axis, where, in the usual small-angle approximation, $X_o$ is related to $f_o$ by $X_o = \lambda d f_o$. As indicated by shading in Fig. 5-17, some of the light that appears, as viewed from the right-hand side of the object, to come from that point may be blocked by the finite aperture of the lens. The image of that point, a distance $X_o' = (z_2/z_1)X_o$ off axis in the (real) Fourier transform plane to the right of the lens, will not be as bright as it would be were the lens sufficiently large. Such a partial obscuration of the source plane and accompanying attenuation of off-axis portions of the Fourier transform distribution is another example of vignetting by the finite lens aperture. Again there are upper and lower spatial frequencies $\rho_1$ and $\rho_2$ that describe the ranges for which there is no vignetting, partial vignetting, and total obscuration by the lens aperture. The calculation of $\rho_1$ and $\rho_2$, though straight-forward, is sufficiently complicated that it is not reproduced here. Figure 5-18 illustrates the effect of vignetting on the spectrum of a Ronchi ruling. This figure should be compared with Fig. 5-14, which shows the spectrum of a Ronchi ruling in the absence of vignetting.





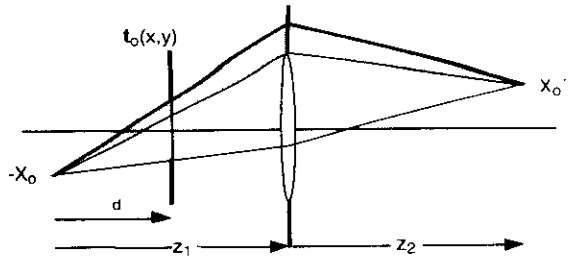



Fig. 5-17. Vignetting by the finite aperture of the lens. Light traveling ray paths in the shaded region are blocked by the aperture and do not reach the Fourier transform plane.

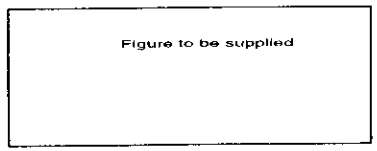



Fig. 5-18. Effect of vignetting on Fourier plane distribution with Ronchi ruling object

### 5.3.3 Effect of Non-Monochromatic Light

It has been assumed throughout the preceding analyses that the light illuminating the object is monochromatic. Such a condition is not, in fact, necessary, though for non-monochromatic illumination the effects of different wavelengths must be taken into account. If the point-source illumination is provided by the focused beam of a laser whose output is in a single spectral line, the bandwidth is sufficiently small that the wavelength dependency of the Fourier plane distributions can generally be ignored. If, however, the illumination is from a non-narrowband point source, it is necessary to integrate over the spectral distribution of the source to calculate the optical intensity in the Fourier plane. Thus, for the case of the object transparency behind the lens, the optical intensity is given by

$$I_{FT}(x,y) = \mathcal{K}\int_0^\infty S(\lambda)\left(\frac{1}{\lambda d}\right)^2\left|\mathbf{T}_o\left(\frac{x}{\lambda d},\frac{y}{\lambda d}\right)\right|^2 d\lambda\ , \qquad (5.56)$$

where $\mathcal{K}$ is a proportionality factor and where $S(\lambda)$ is the power spectral density of the illuminating source. The effect of spectral spreading of the Fourier transform distribution is immediately evident to a viewer peering through a grating at a distant incandescent or mercury-arc street lamp at night. Since such a street lamp radiates light with a large spectral bandwidth, the Fourier intensity pattern appears in multiple colors, each color pattern being scaled by its

wavelength. Consistent with the λ dependency of the diffraction phenomenon, the red-light Fourier distribution is spreads out over a larger area than does the blue-light distribution.

### 5.3.4 Fourier Transform Modules

Because of their use to us later, we define as *Fourier transform modules* the single- and two-lens systems of Fig. 5-19. Both lenses of the second system, which is analyzed in a homework problem, have focal length $f$ and are assumed to satisfy the thin-lens assumption, i.e., to have negligible thickness. The relationship between the complex amplitude of an input wave field $U_o(x,y)$ and the complex amplitude in the Fourier plane is given for both cases by

$$U_{FT}(x,y) = \frac{1}{\lambda f}\hat{U}_o\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right), \tag{5.57}$$

where $\hat{U}_o(u,v) = \mathcal{F}\{U_o(x,y)\}$. These particular lens systems are given special attention because the Fourier-plane distributions are free of multiplicative quadratic phase factors. As discussed earlier, the systems not only map plane waves into points of light but also points of light into plane waves. The two-lens module is impractical in the form presented because its use requires that the object be in intimate contact with the first lens and that the Fourier transform distribution be in intimate contact with the second lens. This deficiency can be removed by a slight redesign of the module, a point also discussed in a homework problem.

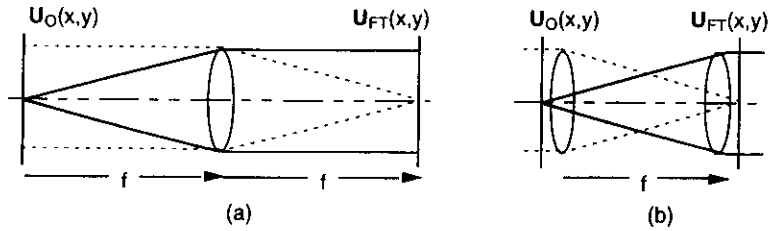



Fig. 5-19. Fourier transform modules: (a) single lens module; (b) two-lens module (lenses identical, each with focal length f). Solid ray paths show collimation of light from point (point into plane wave); dotted ray paths show focusing of collimated light (plane wave into point).

## 5.4 THICK AND COMPOUND LENSES

The thin lens concept is an idealization: real lenses have finite thickness. Indeed, compound, or multi-element lenses may involve many individual lens sections operating together. Multi-element lenses generally perform better than simple thin lenses in exacting applications, for with multiple elements it is possible to compensate for a variety of aberrations. Figure 5-20 illustrates a typical thick lens. Front and back focal points are defined in the figure. A light ray from the front focal point of the lens is brought parallel to the optical axis by the lens. Similarly, a ray incident from the left and parallel to the optical axis passes through the back focal point of the lens. Note that a light ray from the front focal point is refracted at both surfaces of the lens. If the in-air parts of that ray are continued, as indicated by dotted lines in the figure, their intersection is at the *front principal*





*surface* of the lens. Tangent to that surface at the optical axis is the *front principal plane*. The distance between the front focal point and the front principal plane is the *effective focal length* of the lens, denoted $f$. An ideal thin lens of focal length $f$ placed in the front principal plane has the same effect on light coming from the front focal point, i.e., it will cause that light to be collimated.

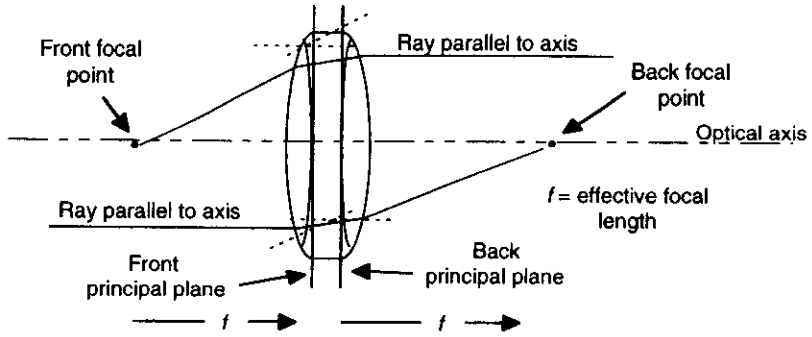



Fig. 5-20. Thick lens.

The *back principal surface* is defined similarly, but for rays from the left that pass through the back focal point. Tangent to the back principal surface is the *back principal plane*. The distance between the back principal plane and the back focal point is the same effective focal length $f$. An ideal thin lens of focal length $f$ placed in the back principal plane will bring on-axis collimated light to a focus at the back focal point.

A collimated beam of light incident on the lens from the left at some arbitrary small angle is brought to a focus in the back focal plane of the lens at a point determined by passing a line parallel to the beam through the point of intersection of the back principal plane and the optical axis. Similar results are obtained for light originating at an off-axis point in the front focal plane, except now the front principal plane must be used.

So long as we work with the effective focal distance, measuring from the principal planes, the action of a real lens for Fourier transforming and imaging is essentially the same as for an ideal thin lens: light from points is still mapped into plane waves, and ligh. from plane waves is still mapped into points. From a modeling standpoint, there is simply a little additional distance involved, the separation between the principal planes. That distance (sometimes referred to as the *ellipsis* of the lens) can easily be ignored in laying out conceptual diagrams for optical systems, by effectively squeezing it to zero. Of course, when a system is actually built, the physical thickness of the lens may be an important consideration because of mounting requirements.

## 5.5 ABERRATIONS AND OPTIMUM LENS ARRANGEMENTS

The system of Fig. 5-12(a) has certain practical advantages over other systems for producing the Fourier transform of a planar object. To begin with, high-quality, low-cost lenses are available that are optimized for imaging an on-axis point to an on-axis point over reasonable distances and with negligible aberration. The systems of Figs. 5-12(b) and 5-9, on the other hand, require more expensive lenses, because the lenses must work equally well with light coming from and/or going off-axis as well as on-axis. In addition, the arrangement of Fig. 5-12(a) prevents vignetting by the

lens aperture: so long as the entire object is satisfactorily illuminated by the converging light waves, every part of it is Fourier transformed equally well for all spatial frequencies. Such operation is not observed with the system of Fig. 5-12(b) or that of Fig. 5-9. Of course the system of Fig. 5-12(a) introduces a quadratic phase factor in the Fourier transform distribution. However, that factor should not be considered a blemish but, rather, simply a manifestation of the basic nature of the system. Furthermore, as noted previously, the presence of such a phase factor has no effect on the Fourier plane intensity distribution. If cascaded Fourier transform operations are to be performed, such as in the next chapter, transform lens systems of the form explored in problem 5.10 are sometimes used. These lens systems effectively eliminate vignetting, and like the ideal two-lens Fourier transform module, they produce transform-plane distributions free of quadratic phase factors. However, unlike the ideal two-lens module, they provide some relief or separation between the object and Fourier transform planes and the lens elements themselves.

## REFERENCES


5-1 F. A. Jenkins and H. E. White, *Fundamentals of Optics*, 4th ed., McGraw-Hill, New York, 1976.

5-3 D. C. O'Shea, *Elements of Modern Optical Design*, Wiley, New York, 1985.

5-4 J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics*, John Wiley and Sons, New York (1978). Sect. 10-6 and Appendix A2.

5-5 J. W. Goodman, *Introduction to Fourier Optics*, 2nd ed., McGraw-Hill, New York (1996); Sects. 5.1 and 5.2.


## PROBLEMS

5.1 Figure 5-6 illustrates a simple geometrical method for determining the location of the image point conjugate to a given object point. Two rays are projected from the object point. One passes undeviated through the center of the lens; the other, originally parallel to the optical axis, is bent by the lens so that it passes through the back focal point. Where the two rays meet specifies the location of the image point.

(a) Confirm the validity of this procedure by constructing to scale the ray diagram for the case $d_o = 3$ cm, $f = 2$ cm, $h = 1$ cm, where $h$ is is the distance of the object point from the optical axis. Confirm your result by application of Eqs. (5.1) and (5.2)

(b) Use the same procedure but now with $d_o = 1$ cm. In this case, the image point is *virtual*, light appearing to come from a point on the same side as the lens.

5.2 A monochromatic wave field given by complex amplitude

$$\mathbf{U}_o(x,y) = \frac{1}{2}\mathrm{cyl}\left(\frac{r}{w}\right)[1+\cos 2\pi f_o x]$$

is impressed across the front focal plane of a converging lens of focal length 50 cm and diameter 5 cm. Find and sketch the x-axis optical intensity distribution in the back focal plane of the lens, assuming that $f_o = 20$ cycles/mm, $w = 20$ mm, and $\lambda = 633$ nm. Label the distance separating diffraction components, distances to nulls, etc. Be sure to check for vignetting.





5.3 A Fourier transforming lens of diameter $D_L$ is used to produce the Fourier transform associated with an object distribution $\mathbf{U}_o(x,y)$ of diameter $D_O$ a distance $d$ in front of it.

(a) Assuming monochromatic, normally incident plane wave illumination, show that Eqs. (5.38) and (5.39) correctly describe, respectively, (i) the maximum spatial frequency $\rho_1$ for which the back focal plane distribution accurately represents the Fourier spectrum of the object wave and (ii) the spatial frequency $\rho_2$ above which no light reaches the Fourier plane.

(b) For what spatial frequency, in cycles/mm, does vignetting begin if $D_L = 4$ cm, $D_O = 2$ cm, $d = 50$ cm, and $\lambda = 600$ nm?

5.4 A thin positive lens of focal length $f = 30$ cm is illuminated by a normally incident monochromatic plane wave at wavelength $\lambda = 633$ nm. Immediately behind the lens is placed a sinusoidal amplitude grating with fundamental frequency of 200 cycles/inch. The Fourier transform distribution is observed in the back focal plane of the lens. How far off axis are the diffraction components? Express your answer in millimeters.

5.5 An object transparency with complex amplitude transmittance

$$\mathbf{t}_a(x,y) = \left[\frac{1}{4} + \frac{1}{2}\cos 2\pi f_o x\right]\mathrm{rect}\left(\frac{x}{w},\frac{y}{w}\right)$$

is placed in the front focal plane of a positive spherical lens of focal length f=25 cm and illuminated by a normally incident plane wave of wavelength $\lambda = 500$ nm. Assume that the lens is masked by a square aperture of width $D = 3$ cm and that $w = 1$ cm. Show the effect of vignetting by the lens by sketching the x-axis cross-section of the optical intensity distribution as it appears in the back focal plane for (a) $f_o = 60$ mm$^{-1}$, (b) $f_o = 120$ mm$^{-1}$, and (c) $f_o = 180$ mm$^{-1}$.

5.6 A structure know as a *Gabor zone plate* is characterized by complex wave amplitude transmittance

$$\mathbf{t}_a(x,y) = \frac{1}{2}\left[1+\cos\left(\alpha r^2\right)\right],$$

where $r^2 = x^2 + y^2$.

(a) Show that this structure exhibits simultaneously the characteristics of a flat piece of glass, a positive lens, and a negative lens.

(b) Find the focal length of the positive lens of part (a).

5.7 A pinhole spatial filter consists of a short-focal-length microscope objective with a pinhole placed in the back focal plane. It is used to remove noise-like high spatial frequency components of a laser beam that are produced by dust and scratches on the laser output window. The microscope objective Fourier transforms the Gaussian-profile laser beam to produce a compact Gaussian-profile spot of light, which is incident on the pinhole. If the pinhole is of the correct diameter, it passes virtually the entire Gaussian-shaped light distribution but blocks most light scattered at large angles by dust and scratches. Assume a

wavelength $\lambda$ of 633 nm, a diameter for the incident laser beam of 1.0 mm, as measured between the $1/e^2$ intensity points on the beam, and an effective focal length for the microscope objective of 9.0 mm. How large should the pinhole be to pass 99% of the power in the focused Gaussian beam? (Note: a plot or table of the error function, erf($\xi$), may be of use to you in calculating your answer.)

5.8 Verify that for sufficiently large values of $z_1$ the factor Md in Eq. (5.52) approaches $f$ as a limit, with Eq. (5.55) resulting.

5.9 Show that the two-lens Fourier transform module illustrated in Fig. 5-19(b) yields the complex wave amplitude $\mathbf{U}_f(x,y) = \frac{1}{\lambda f}\hat{\mathbf{U}}_o\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right)$ in the output Fourier transform plane if object amplitude $\mathbf{U}_o(x,y)$ is impressed across the object plane.

5.10 The two-lens Fourier transform module of Fig. 5-19(b) is important conceptually. However, it is impractical because there is no relief, or separation, between the lenses and the input and output planes. Fig. P5.10 shows a modified form of the two-lens module that provides some space between the lenses and the object and Fourier transform planes. As suggested by the ray paths, light from a plane wave in the input plane (represented by the parallel rays) is mapped into a point in the output plane, and light from a point is mapped into a plane wave—the two key characteristics of a "quadratic-phase-free" Fourier transform module. Using the imaging equation for thin lenses [Eq. (5.1)], find the required lens separation S if the system of Fig. P5.10 is to have the desired characteristics, given that f=30 cm and d=2 cm.

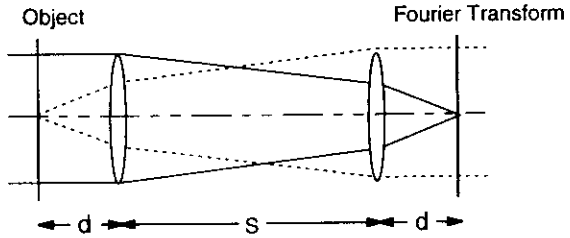



Fig. P5.10. Two-lens Fourier transform module with separation $d$ between the lenses and the object and Fourier transform planes.

5.11 A planar transparent object 1.0 cm in diameter is placed a distance $2f$ in front of a lens of diameter 2.5 cm and focal length 10 cm and illuminated by a normally incident plane wave at wavelength $\lambda = 0.5$ μm. Find the maximum object frequency for which vignetting by the lens can be ignored. Express your answer in $mm^{-1}$.

5.12 A circular positive lens is illuminated by a normally incident plane wave. Find the relationship between the cone angle of the light focused by the lens and the f-number of the lens.





5.13 Show that if the object transparency is right up against the lens, the two canonical forms of illumination treated in Section 5.3.2 lead to the same results.

5.14 A monochromatic object distribution with complex amplitude $U_{obj}(x,y)$ is positioned 10 cm in front of a lens of focal length $f = 5$ cm, producing a unit-magnification image 10 cm behind the lens. Assume that the object has a diameter of 1 cm, that $\lambda = 633$ nm, and that $U_{obj}(x,y)$ contains no spatial frequency components above $\rho_{max} = 75$ $mm^{-1}$. Use reasoning similar to that used in the vignetting analysis of Section 5.3.1 to show that the infinite-lens impulse response $h_\infty(x,y;\xi,\eta)$ can be used if the lens is at least approximately 2 cm in diameter.

**New problems (to be ordered)**

5.15 Sketch a ray diagram, to scale, similar to that of Fig. 5-2(c), assuming the following:

- The diameter of the lens is 2 cm
- The object is positioned 9 cm in front of the lens
- The focal length of the lens is 5 cm

Show clearly each of the three rays described immediately following the boxed material on signed distances. Label all the distance in your figure. Use single-headed arrows to indicate the signs associated with the different distance quantities.

5.16 A planar object is positioned in the object plane of the system of Fig. 5-3(b) but with a slight longitudinal position error $\Delta d_o$.

(a) Find the corresponding shift $\Delta d_i$ in the longitudinal position of the image plane. Assume that $|\Delta d_o| << 1$.

(b) Find the general expression for $\Delta d_i/\Delta d_o$, for arbitrary $d_o$, again assuming that $|\Delta d_o| << 1$. This expression gives the longitudinal magnification of the imaging system.

(c) Relate the longitudinal magnification found in part (b) to the transverse magnification given by Eq. (5.2).

## INDEX - CHAPTER 5

# 6. Imaging Spatially Coherent Wave Fields

In this chapter we investigate the formation of images with monochromatic or spatially coherent quasimonochromatic light. Such image formation is referred to as *coherent* image formation. The use of systems consisting of a cascade of two Fourier-transform lens modules is considered first because of their conceptual simplicity. The object-image relationship for these systems is given by a simple space-invariant convolution expression. Single-lens and other more complicated lens systems are then analyzed, including systems that are fundamentally space-variant. Results of the monochromatic case are extended to include the case of quasimonochromatic coherent wave fields, and the resolution characteristics of such systems are discussed.

## 6.1 IMAGING WITH FOURIER-TRANSFORM MODULES: THE COHERENT IMPULSE RESPONSE AND COHERENT TRANSFER FUNCTION

The formation of images of monochromatic wave fields can in certain cases be conveniently described in terms of a 2-D Fourier transform followed by a second Fourier transform. The system of Fig. 6-1, which consists of a cascaded pair of single-lens Fourier transform modules of different lengths, serves to illustrate.

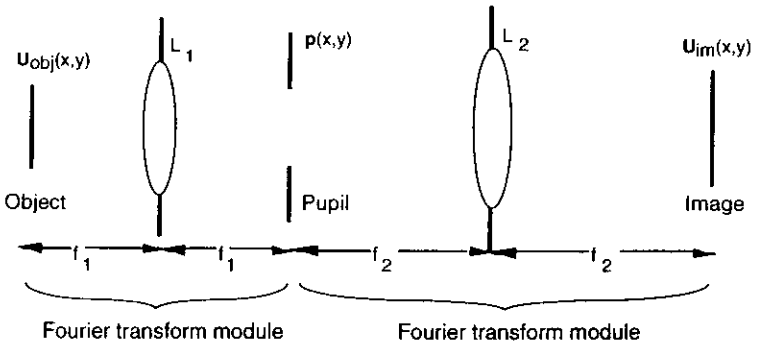



Fig. 6-1. Imaging system consisting of two single-lens Fourier transform modules. Lens $L_1$ has focal length $f_1$; lens $L_2$ has focal length $f_2$. The object distribution is assumed to be monochromatic.

Assume an object wave field $\mathbf{U}_{obj}(x,y)$ as the input to the system. So long as vignetting at lens $L_1$ can be ignored, the complex wave amplitude in the back focal plane of the first lens, $\mathbf{U}_f(x,y)$, is given through Eq. (5.33) by a scaled version of the Fourier transform of $\mathbf{U}_{obj}(x,y)$:





$$\mathbf{U}_f(x,y) = \frac{1}{\lambda f_1}\hat{\mathbf{U}}_{obj}\left(\frac{x}{\lambda f_1}, \frac{y}{\lambda f_1}\right), \tag{6.1}$$

where $\hat{\mathbf{U}}_{obj}(u,v) = \mathcal{F}\{\mathbf{U}_{obj}(x,y)\}$. Between the two modules is an aperture characterized by pupil function $\mathbf{p}(x,y)$. The input to the second Fourier transform module, denoted $\mathbf{U}'_f(x,y)$, is thus the product of $\mathbf{U}_f(x,y)$ with $\mathbf{p}(x,y)$:

$$\mathbf{U}'_f(x,y) = \frac{1}{\lambda f_1}\hat{\mathbf{U}}_{obj}\left(\frac{x}{\lambda f_1}, \frac{y}{\lambda f_1}\right)\mathbf{p}(x,y). \tag{6.2}$$

Again using Eq. (5.33) we can express the wave amplitude distribution in the output image plane by a Fourier transform:

$$\mathbf{U}_{im}(x,y) = \frac{1}{\lambda f_2}\mathcal{F}\{\mathbf{U}'_f(x,y)\}\Big|_{\substack{u=x/\lambda f_2 \\ v=y/\lambda f_2}}. \tag{6.3}$$

Substituting for $\mathbf{U}'_1(x,y)$ and applying the similarity theorem yields

$$\mathbf{U}_{im}(x,y) = \frac{1}{\lambda f_2}\left[\lambda f_1 \mathbf{P}(u,v) ** \mathbf{U}_{obj}(-\lambda f_1 u, -\lambda f_1 v)\right]\Big|_{\substack{u=x/\lambda f_2 \\ v=y/\lambda f_2}}, \tag{6.4}$$

where $\mathbf{P}(u,v) = \mathcal{F}\{\mathbf{p}(x,y)\}$, the Fourier transform of the pupil function. The minus signs in the second term of the convolution result from the sequential application of two forward Fourier transforms rather than a forward transform followed by an inverse transform. Writing out the convolution integral and substituting for u and v yields

$$\mathbf{U}_{im}(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{f_1}{f_2}\mathbf{P}(\xi,\eta)\mathbf{U}_{obj}\left[-\lambda f_1\left(\frac{x}{\lambda f_2} - \xi\right), -\lambda f_1\left(\frac{y}{\lambda f_2} - \eta\right)\right]d\xi d\eta$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{f_1}{f_2}\mathbf{P}(\xi,\eta)\mathbf{U}_{obj}\left(\frac{x-\lambda f_2\xi}{-f_2/f_1}, \frac{y-\lambda f_2\eta}{-f_2/f_1}\right)d\xi\, d\eta. \tag{6.5}$$

To simplify Eq. (6.5) denote by M the *magnification ratio* $f_2/f_1$,

$$M = \frac{f_2}{f_1}, \tag{6.6}$$

and introduce new variables $\mu$ and $\zeta$ given by

$$\mu = \lambda f_2 \xi, \qquad \zeta = \lambda f_2 \eta. \tag{6.7}$$

On substitution of these parameters, Eq. (6.5) becomes

$$\mathbf{U}_{im}(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{M}\mathbf{P}\left(\frac{\mu}{\lambda f_2}, \frac{\zeta}{\lambda f_2}\right)\mathbf{U}_{obj}\left(\frac{x-\mu}{-M}, \frac{y-\zeta}{-M}\right)\left(\frac{1}{\lambda f_2}\right)^2 d\mu\, d\zeta. \tag{6.8}$$

To simplify further, two new functions, *the geometrical optics image* $\mathbf{U_g}(x,y)$ and the *complex wave amplitude impulse response for coherent space-invariant imaging* $\mathbf{h}_{coh}(x,y)$, often referred to as the *coherent impulse response* or *coherent point spread function*, are defined in accord with

$$\mathbf{U}_g(x,y) = \frac{1}{M}\mathbf{U}_{obj}\left(\frac{-x}{M}, \frac{-y}{M}\right) \tag{6.9}$$

and

$$\mathbf{h}_{coh}(x,y) = \left(\frac{1}{\lambda f_2}\right)^2 \mathbf{P}\left(\frac{x}{\lambda f_2}, \frac{y}{\lambda f_2}\right). \tag{6.10}$$

The wave amplitude in the image plane then assumes the form of a simple convolution:

$$\mathbf{U}_{im}(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{h}(\mu,\zeta)\mathbf{U_g}(x-\mu, y-\zeta)\, d\mu\, d\zeta, \tag{6.11a}$$

or

$$\mathbf{U}_{im}(x,y) = \mathbf{h}_{coh}(x,y) ** \mathbf{U}_g(x,y). \tag{6.11b}$$

The imaging system input-output relationship just derived can be expressed by the cascade of two operations. First, the object wave field $\mathbf{U}_{obj}(x,y)$ is scaled to produce the geometrical optics image $\mathbf{U_g}(x,y)$. The scaled function $\mathbf{U_g}(x,y)$ is then convolved with $\mathbf{h}_{coh}(x,y)$, which is a scaled version of the Fourier transform of the pupil function. This cascade of operations is illustrated in Fig. 6-2. Note that $\mathbf{h}_{coh}(x,y)$ is not the impulse response of the overall system but only of the post-scaling portion. Because the image distribution cannot be expressed directly in terms of a convolution of $\mathbf{U}_{obj}(x,y)$ with some impulse response, the overall imaging system is, in a strict mathematical sense, space-variant. However, since the departure from strict space invariance involves only the scaling operation, such imaging is commonly described as being space invariant.

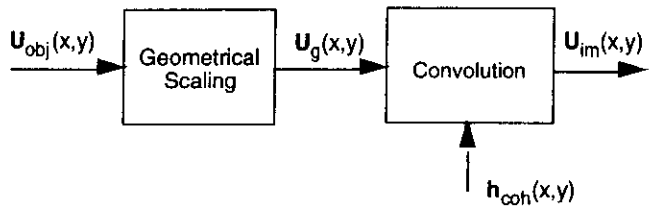



Fig. 6-2. Block diagram representing object-image relationship for coherent imaging system of Fig. 6-1.





Note that the system of Fig. 6-1 is accurately described by Eq. (6.11) only to the extent that vignetting by the lens apertures can be ignored. As illustrated by a homework problem, it is ultimately the relationship between the size of the mid-plane aperture, the size of the object, and the sizes of the lenses that determines the validity of this description.

The function $\mathbf{U_g}(x,y)$ represents the wave field that would be predicted by geometrical optics. To see that this is so we let the wavelength $\lambda$ approach zero. Diffraction effects, which scale with $\lambda$, disappear. By inspection, as $\lambda \to 0$, $\mathbf{h}_{coh}(x,y)$ grows narrower in width and stronger in amplitude, in the limit assuming the characteristics of a delta function. It is shown in a homework problem that, so long as the pupil function satisfies the condition $\mathbf{p}(0,0) = 1$, this delta function has unit volume. The image obtained in the zero-wavelength limit is thus given by $\mathbf{U}_{im}(x,y) = \mathbf{U_g}(x,y) ** \delta(x,y)$, or by $\mathbf{U_g}(x,y)$ itself—a replica of the object wave, scaled in size and amplitude and inverted through the origin. For finite values of $\lambda$ $\mathbf{U_g}(x,y)$ is in general smoothed by convolution with $\mathbf{h}_{coh}(x,y)$. Modification of the Fourier transform of the object wave by the aperture of the imaging system is thus seen to correspond to a *spatial filtering* of $\mathbf{U_g}(x,y)$. The result is analogous to putting an electrical signal through a linear filter network.

To obtain a better physical sense of why the system impulse response should be given by the Fourier transform of the pupil function, consider placing a point source—the physical equivalent to $\delta(x,y)$—on axis in the object plane of the system of Fig. 6-1. This source launches a spherical wave, which is converted to a plane wave by lens $L_1$. That plane wave illuminates the aperture, and lens $L_2$ Fourier transforms the resulting transmitted wave field, which is proportional to $\mathbf{p}(x,y)$. Thus, for example, a square aperture characterized by pupil function

$$\mathbf{p}(x,y) = \mathrm{rect}\left(\frac{x}{w}, \frac{y}{w}\right) \tag{6.12}$$

results in a coherent point spread function that has the form of a 2-D sinc function:

$$\mathbf{h}_{coh}(x,y) = \left(\frac{w}{\lambda f_2}\right)^2 \mathrm{sinc}\left(\frac{x}{\lambda f_2 / w}, \frac{y}{\lambda f_2 / w}\right). \tag{6.13}$$

The factor $(1/\lambda f_2)^2$ in Eq. (6.10) is consistent with energy conservation, as is demonstrated by a homework problem.

The spatial filtering operation represented by the convolution of Eq. (6.11) can be expressed in the spatial frequency domain. Taking the Fourier transform of both sides yields

$$\hat{\mathbf{U}}_{im}(u,v) = \hat{\mathbf{U}}_g(u,v)\mathbf{H}_{coh}(u,v), \tag{6.14}$$

where $\hat{\mathbf{U}}_{im}(u,v) = \mathcal{F}\{\mathbf{U}_{im}(x,y)\}$ and $\hat{\mathbf{U}}_g(u,v) = \mathcal{F}\{\mathbf{U}_g(x,y)\}$. The function $\mathbf{H}_{coh}(u,v)$ is given by

$$\mathbf{H}_{coh}(u,v) = \mathcal{F}\{\mathbf{h}_{coh}(x,y)\}. \tag{6.15}$$

This function is the *complex wave amplitude transfer function* of the imaging system, often referred to as the *coherent transfer function*. Substituting for $\mathbf{h}_{coh}(x,y)$ from Eq. (6.10) yields

$$\mathbf{H}_{coh}(u,v) = \mathbf{p}(-\lambda f_2 u, -\lambda f_2 v). \tag{6.16}$$

The coherent transfer function is seen to be a scaled version of the system pupil function. The minus signs, which result from applying a forward—as opposed to an inverse—Fourier transform to **P**, can often be ignored. In many cases of practical importance, in fact, **p**(x,y) is symmetric, and the minus signs have no effect whatsoever.

The relationship between the pupil function and the coherent transfer function is of fundamental importance to an understanding of the imaging of monochromatic wave fields. As an example, consider again the case where the aperture is square and of width $w$, with **p**(x,y) being given again by Eq. (6.12). The coherent transfer function for the system is then given by

$$\mathbf{H}_{coh}(u,v) = \mathrm{rect}\left(\frac{-\lambda f_2 u}{w}, \frac{-\lambda f_2 v}{w}\right)$$

$$= \mathrm{rect}\left(\frac{u}{w/\lambda f_2}, \frac{v}{w/\lambda f_2}\right). \tag{6.17}$$

This function is illustrated in Fig. 6-3(a). It is evident from Eq. (6.17) and the figure that the highest, or cutoff, spatial frequencies of the image amplitude distribution in the x and y directions, $u_c$ and $v_c$, are given by

$$u_c = v_c = \frac{w}{2\lambda f_2}. \tag{6.18}$$

Through the relationship between spatial frequency and the quantity $(1/\lambda)\sin\theta$ [See Eq. (4.77)], this upper limit on image frequency is seen to correspond to a maximum tilt angle for any plane wave components making up the image distribution.

It should be stressed that $w/2\lambda f_2$ is the x and y cutoff frequency in the *image* amplitude distribution. If it is recalled that the image and object distributions differ in scale by a factor M, it should be clear that this frequency corresponds to an *object* spatial frequency given by M times $w/2\lambda f_2$. If it is assumed that w = 20 mm, $f_2$ = 100 mm, and $\lambda$ = 1 μm = $10^{-3}$ mm, then $w/2\lambda f_2$ = 100 $mm^{-1}$. If $f_1$ = 50 mm, the imaging system magnifies by a factor M=2. A smaller spatial period (higher spatial frequency) in the object plane is thus mapped into a larger spatial period (lower spatial frequency) in the image plane. Equivalently, a larger wave tilt angle in the object plane is mapped into a smaller wave tilt angle in the image plane. The maximum spatial frequency of 100 cycles per millimeter observed in the image plane corresponds to a spatial frequency of 200 cycles per millimeter in the object plane.

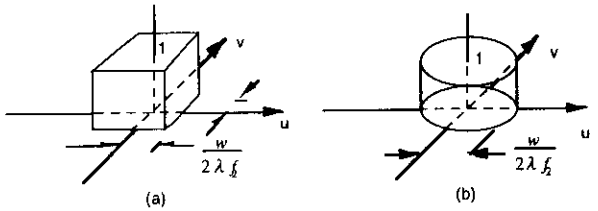







Fig. 6-3. Coherent transfer functions for system of Fig. 6-1 with (a) square aperture of width w, (b) circular aperture of diameter w.

Typically, the aperture of an imaging system is circular, and the pupil function and, hence, the coherent transfer function of the system are represented by cylinder or circ functions. In the case of a circular aperture of diameter $w$ the coherent transfer function for the imaging system of Fig. 6-1 is given by

$$\mathbf{H}_{coh}(\rho) = \mathrm{cyl}\left(\frac{\rho}{w/\lambda f_2}\right). \tag{6.19}$$

The cutoff frequency in any direction, $\rho_c$, is given by

$$\rho_c = \frac{w}{2\lambda f_2}. \tag{6.20}$$

This function is illustrated in Fig. 6-3(b). Again, all spatial frequency components of the object, up to a cutoff frequency determined by the diameter of the aperture, are transmitted to the image plane without change in amplitude or phase. Beyond that cutoff frequency, no spatial frequency components are passed. The optical system is seen to perform an ideal low-pass spatial filtering operation in imaging the object.

It must be kept in mind that it is complex wave *amplitude* distributions, not intensity distributions, that are filtered. The effect of the filtering on the corresponding optical intensity distributions will have a nonlinear dependence through the squared-modulus operation relating amplitude to intensity. This point is explored later in the chapter.

The operation of the two-lens system of Fig. 6-1 is duplicated by that of the four-lens system of Fig. 6-4(a), since from an analytical standpoint the single-lens and two-lens Fourier transform modules perform the same operation. The four-lens system has an advantage over the two-lens system in that the lenses introduce no vignetting. The object-image relationship is again expressed by Eq. (6.11) with $\mathbf{h}_{coh}$(x,y) given by Eq. (6.10) and $\mathbf{U}_g$(x,y) given by Eq. (6.9).

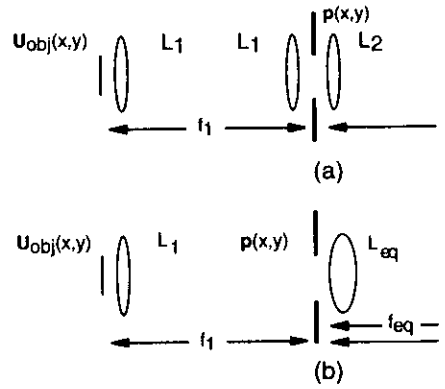



Fig. 6-4. Coherent imaging systems: (a) four-lens space-invariant imaging system and (b) related two-lens system that produces the same image intensity distribution; (c) system of (b) relabeled.

For many applications, the four lenses of the system of Fig. 6-4(a) can be reduced to two. An important result is a reduction in the number of glass surfaces, which can produce undesired reflections and collect light-scattering dust. To begin with, so long as only the optical *intensity* of the image distribution is of interest, the fourth and final lens can be removed, for its removal merely contributes a quadratic phase factor to the image wave amplitude distribution that disappears upon evaluation of the intensity. In a further simplification, the two lenses positioned at the central aperture can be combined into a single, more powerful lens. To see this, we combine terms in the transmittance function associated with the combination of the two central lenses and the aperture:

$$\exp\left[-j\frac{k}{2f_1}\left(x^2+y^2\right)\right]\mathbf{p}(x,y)\exp\left[-j\frac{k}{2f_2}\left(x^2+y^2\right)\right]=\exp\left[-j\frac{k}{2f_{eq}}\left(x^2+y^2\right)\right]\mathbf{p}(x,y), \quad (6.21)$$

where $f_{eq}$, the focal length of the equivalent combined lens, is given by

$$\frac{1}{f_{eq}}=\frac{1}{f_1}+\frac{1}{f_2}. \quad (6.22)$$

It is seen that the powers of the two lenses—their reciprocal focal distances—add in producing the equivalent lens.

Both simplifications taken together result in the system illustrated in Fig. 6-4(b). The object-image wave amplitude relationship is given by

$$\mathbf{U}_{im}(x,y)=\left[\mathbf{U}_{\mathbf{g}}(x,y)**\mathbf{h}_{coh}(x,y)\right]\exp\left[j\frac{k}{2f_2}\left(x^2+y^2\right)\right], \quad (6.23)$$

where $\mathbf{U_g}(x,y)$ and $\mathbf{h_{coh}}(x,y)$ are given again by Eqs. (6.9) and (6.10). The quadratic phase factor results from the removal of the fourth and final lens. The optical intensity of the image is given by

$$I_{im}(x,y)=\left|\mathbf{U}_{\mathbf{g}}(x,y)**\mathbf{h}_{coh}(x,y)\right|^2, \quad (6.24)$$

which is the same as is obtained with the system of Fig. 6-1. A revised system diagram appropriate for this imaging system is shown in Fig. 6-5. Despite the geometrical scaling and the post-convolution multiplication by a quadratic phase factor, coherent imaging systems characterized by this block diagram are nonetheless often referred to as being space-invariant.

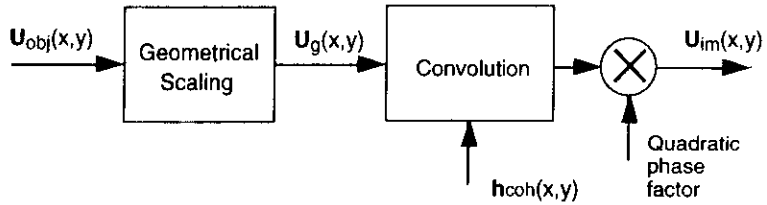



Fig. 6-5. Block diagram representing object-image relationship for coherent imaging system of Fig. 6-4(c).





It is useful to relabel the system of Fig. 6-4(b) as in Fig. 6-4(c). The distances $f_1$ and $f_2$ become object and image distances $d_o$ and $d_i$, respectively, and the relationship between $d_o$, $d_i$, and the focal length of the equivalent combined lens, $f_{eq}$, is then the standard single-lens imaging condition:

$$\frac{1}{f_{eq}}=\frac{1}{d_o}+\frac{1}{d_i}. \quad (6.25)$$

The magnification factor M is now given by

$$M=\frac{d_i}{d_o}, \quad (6.26)$$

$\mathbf{h_{coh}}(x,y)$ is given by

$$\mathbf{h}_{coh}(x,y)=\left(\frac{1}{\lambda d_i}\right)^2\mathbf{P}\left(\frac{x}{\lambda d_i},\frac{y}{\lambda d_i}\right), \quad (6.27)$$

and the final multiplicative quadratic phase factor is given by

$$\exp\left[j\frac{k}{2d_i}\left(x^2+y^2\right)\right]. \quad (6.28)$$

It is important to note that Lens $L_1$ in the system of Fig. 6-4(b) cannot be removed without destroying the convolution relationships of Eqs. (6.23) and (6.24). Without lens $L_1$ there is no Fourier transform distribution $\hat{\mathbf{U}}_{obj}$ incident on the pupil and, as a consequence, no transfer function characterizing the imaging operation! As a consequence, the system shown in Fig. 6-6(a), even though it satisfies the geometrical optics condition for imaging ($1/f = 1/d_o + 1/d_i$), is *space variant* in the way it images complex wave amplitude distributions. The cause of the space-variance, intimately related to the vignetting discussed in Chapt. 5, is illustrated in Fig. 6-6(b). For the on-axis region of the object, labeled A, only low spatial frequency content of the object distribution—corresponding to light rays traveling more or less parallel to the z-axis—passes through the lens to be imaged: high spatial frequency content from that region is blocked by the

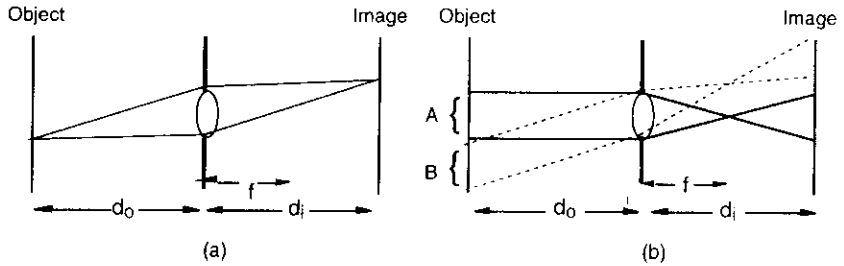



Fig. 6-6. Space-variant single-lens imaging system. From a geometrical optics viewpoint, the single lens provides point-to-point imaging, as illustrated in (a). However, from a wave optics standpoint, as illustrated in (b), only low spatial frequency content in region A is imaged, whereas only high spatial frequency content in region B is.

aperture. For region B, on the other hand, only light corresponding to high spatial frequency content of the object passes through the aperture to reach the image plane, whereas light corresponding to low spatial frequency content is blocked. This system is analyzed in Sec. 6.3 in such a way that the space variance is brought out clearly.

## 6.2 SPACE-INVARIANT SINGLE-LENS IMAGING SYSTEM

It is possible for a single-lens imaging system to be made space invariant if the object is smaller than the lens and if an aperture, suitably small, is placed in the back focal plane of the lens. Such a system is illustrated in Fig. 6-7. Although this system is not made up of Fourier transform modules, the input-output relationship can nevertheless be expressed, to within a quadratic phase factor, in terms of a convolution. Let $U_{obj}(x,y)$ again denote the input object wave amplitude distribution. If the lens is sufficiently large that vignetting can be ignored, then, from Eq. (5.30) the wave amplitude incident on the aperture in the back focal plane is given by

$$\mathbf{U}_f(x,y) = \frac{1}{\lambda f}\exp\left[j\frac{k}{2f}\left(1-\frac{d_o}{f}\right)\left(x^2+y^2\right)\right]\hat{\mathbf{U}}_{obj}\left(\frac{x}{\lambda f},\frac{y}{\lambda f}\right). \tag{6.29}$$

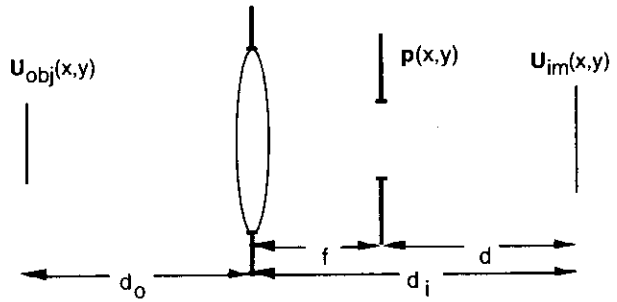



Fig. 6-7. Space-invariant single-lens imaging system. Limiting aperture is in back focal plane of lens.

In order to find the wave amplitude in the image plane we take the wave field immediately behind the back focal-plane aperture,

$$\mathbf{U}_p(x,y) = \mathbf{U}_f(x,y)\mathbf{p}(x,y), \tag{6.30}$$

and calculate the effects of propagation through the distance d using the Fresnel diffraction formula, Eq. (4.31):

$$\mathbf{U}_{im}(x,y) = \frac{1}{\lambda d}\exp\left[j\frac{k}{2d}\left(x^2+y^2\right)\right]\mathcal{F}\left\{\mathbf{U}_p(x,y)\exp\left[j\frac{k}{2d}\left(x^2+y^2\right)\right]\right\}\Bigg|_{\substack{u=x/\lambda d\\ v=y/\lambda d}}, \tag{6.31}$$

where from the figure





$$d = (d_i - f). \tag{6.32}$$

In general this calculation does not lead to an imaging relationship. However, if the distance d satisfies the relationship

$$\frac{1}{d} = -\frac{1}{f}\left(1-\frac{d_o}{f}\right), \tag{6.33}$$

the quadratic phase factor in the Fourier integral portion of Eq. (6.31) is canceled by the one in Eq. (6.29), with the greatly simplified result

$$\mathbf{U}_{im}(x,y) = \frac{1}{\lambda d}\exp\left[j\frac{k}{2d}\left(x^2+y^2\right)\right]\mathcal{F}\left\{\frac{1}{\lambda f}\hat{\mathbf{U}}_{obj}\left(\frac{x}{\lambda f},\frac{y}{\lambda f}\right)\mathbf{p}(x,y)\right\}\Bigg|_{\substack{u=x/\lambda d\\ v=y/\lambda d}}. \tag{6.34}$$

Upon substitution for u and v and some rearranging of terms, Eq. (6.34) can be written in the form

$$\mathbf{U}_{im}(x,y) = \left[\mathbf{U}_g(x,y)**\mathbf{h}_{coh}(x,y)\right]\exp\left[j\frac{k}{2d}\left(x^2+y^2\right)\right], \tag{6.35}$$

where $\mathbf{U}_g(x,y)$ is again given by Eq. (6.9), d = Mf, and $\mathbf{h}_{coh}(x,y)$ now has the form

$$\mathbf{h}_{coh}(x,y) = \left(\frac{1}{\lambda d}\right)^2\mathbf{P}\left(\frac{x}{\lambda d},\frac{y}{\lambda d}\right). \tag{6.36}$$

The magnification factor M is given by

$$M = \frac{d_i}{d_o} = \frac{d}{f} = \frac{d_i - f}{f} = \frac{f}{d_o - f}. \tag{6.37}$$

The condition expressed by Eq. (6.33) can be combined with Eq. (6.32) to produce the more familiar imaging condition,

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}. \tag{6.38}$$

If desired, the residual quadratic phase factor of Eq. (6.35) can be canceled by placing a positive lens of focal length d in the image plane. This phase factor is of course inconsequential in the normally encountered case when only the image intensity is of interest.

It should again be emphasized that the input-output relationship for this system is correctly described by a convolution only if vignetting by the lens can be ignored. As discussed in a homework problem, this will be the case only if the lens diameter L satisfies the condition

$$L \geq W + \frac{d_o w}{f}, \tag{6.39}$$

where $W$ is the object diameter and w is the diameter of the aperture in the focal plane. Under these conditions, we say that the aperture in the Fourier transform plane is the *limiting aperture* of the system. If an aperture of an imaging system is the limiting aperture, then by definition enlarging any other aperture in the system—in this case, e.g., the size of the lens—will not change the wave field in the output plane of the system.

It is useful at this stage to review what we have done so far. Three basic systems have been analyzed, those of Figs. 6-1 [and therefore 6-4(a)], 6-4(c), and 6-7. In all cases, the image amplitude distribution is obtained, as shown in the block diagram of Fig. 6-5, by first scaling the object distribution, convolving the result with $\mathbf{h}_{coh}(x,y)$, and, when appropriate, multiplying the result by a quadratic phase factor. The optical intensity associated with the image distribution is obtained by evaluating the associated squared modulus. The block diagram of Fig. 6-5 has great importance, for it correctly describes *any* coherent imaging system for which the limiting aperture is in the plane where the Fourier transform of the object appears. That condition holds for the systems noted above. It does *not* hold for the single-lens imaging system illustrated in Fig. 6-6, where the limiting aperture is in the plane of the lens rather than in the plane where the object transform appears.

It should be emphasized that the block diagram of Fig. 6-5 represents an image-space interpretation of the image filtering operation, meaning that the convolution comes after the geometrical scaling operation. An alternative interpretation of a coherent imaging operation is suggested by the block diagram of Fig. 6-8, which puts the filtering operation in object space. This interpretation allows one to think of an *effective* object distribution $\mathbf{U}_{eff}(x,y)$, given by $\mathbf{U}_{eff}(x,y) = \mathbf{U}_{obj}(x,y) ** \tilde{\mathbf{h}}_{coh}(x,y)$, which is imaged *perfectly* by the system and then combined with an appropriate quadratic phase factor. Note that the convolution kernel $\tilde{\mathbf{h}}_{coh}(x,y)$ is an object-space kernel rather than an image-space kernel, i.e., it operates directly on $\mathbf{U}_{obj}(x,y)$ rather than on its image-space counterpart $\mathbf{U}_g(x,y)$. For the system of Fig. 6-1 it can be shown that $\tilde{\mathbf{h}}_{coh}(x,y) = (1/\lambda f_1)^2 \mathbf{P}(x/\lambda f_1, y/\lambda f_1)$. Note that the scale factor involves the focal length of lens $L_1$ rather than that of lens $L_2$. The corresponding transfer function is given by $\tilde{\mathbf{H}}_{coh}(u,v) = \mathbf{p}(-\lambda f_1 u, -\lambda f_1 v)$.

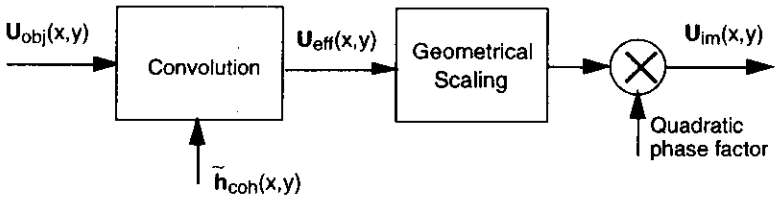



Fig. 6-8. Modified block diagram puts filtering in object space rather than in image space.

## 6.3 SPACE-VARIANT SINGLE-LENS IMAGING

All of the imaging systems analyzed thus far in this chapter are consistent in operation with the block diagram of Fig. 6-5, in some cases with the quadratic phase factor equal to unity. Consider now the single-lens imaging system illustrated in Fig. 6-9(a), which, as noted earlier, is not. This system is easily analyzed by means of an artifice suggested by Fig. 6-9(b). In that figure the single





lens of Fig. 6-9(a) has been replaced by an equivalent pair of lenses straddling the aperture. The focal lengths of these two lenses equal the distances $d_o$ and $d_i$, as indicated, consistent with Eq. (6.38). In addition, a pair of lenses with focal lengths $-d_o$ and $+d_o$ are placed in cascade in the object plane and another pair of lenses with focal lengths $+d_i$ and $-d_i$ are placed in cascade in the image plane. The lenses in each pair cancel one another and have no net effect optically. However, if lens $L_1'$ is associated with the object distribution and lens $L_2'$ with the image distribution, the remaining system, consisting of the four internal lenses and aperture, is seen to constitute the space invariant four-lens imaging system of Fig. 6-4(a). Exploiting the results of the analysis of that system, it is easily shown that the complex amplitude in the image plane is given by

$$\mathbf{U}_{im}(x,y) = \left( \left\{ \frac{1}{M} \mathbf{U}_{obj}\left(\frac{-x}{M}, \frac{-y}{M}\right) \exp\left[ j \frac{k}{2d_o}\left( \frac{x^2}{M^2} + \frac{y^2}{M^2} \right) \right] \right\} ** \mathbf{h}_{coh}(x,y) \right) \exp\left[ j \frac{k}{2d_i}\left(x^2 + y^2\right) \right]. \quad (6.40)$$

The term in braces plays the role of $\mathbf{U}_g(x,y)$, being obtained by multiplying the object wave field by the transmittance function of the negative lens $L'_1$ and then scaling and inverting the product.

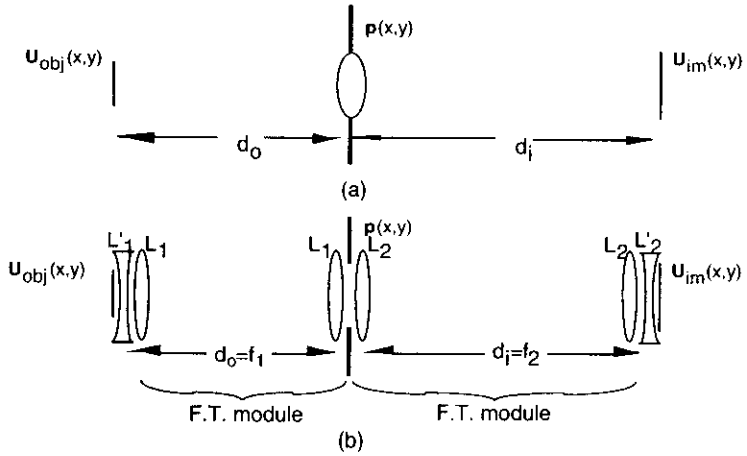



Fig. 6-9. Space-variant single-lens imaging system: (a) actual system; (b) equivalent system containing space-invariant core system consisting of two two-lens Fourier transform modules. Lenses labeled $L_1$ have a focal lengths equal to $d_o$; those labeled $L_2$ have focal lengths equal to $d_i$. The negative lenses $L'_1$ and $L'_2$ have focal lengths $-d_o$ and $-d_i$, respectively.

Although Eq. (6.40) involves a convolution, because $\mathbf{U}_{obj}(x,y)$ is multiplied by the negative lens factor the quasi space invariance we have associated with systems studied earlier in this chapter does not apply. A mathematically more elegant object-image relationship is in fact obtained if the space-variance of the system is incorporated directly into the analytical description of the system. Specifically, through a suitable substitution of variables, Eq. (6.40) can be rewritten in the form

$$\mathbf{U}_{im}(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{U}_{obj}(\xi,\eta)\mathbf{h}_{ss}(x,y;\xi,\eta)\,d\xi\,d\eta, \tag{6.41}$$

where the *space-variant* kernel $\mathbf{h}_{ss}(x,y;\xi,\eta)$ is given by

$$\mathbf{h}_{ss}(x,y;\xi,\eta) = \exp\left[j\frac{k}{2d_o}\left(\xi^2+\eta^2\right)\right]\frac{1}{\lambda^2 d_o d_i}\mathbf{P}\left(\frac{x}{\lambda d_i}+\frac{\xi}{\lambda d_o},\frac{y}{\lambda d_i}+\frac{\eta}{\lambda d_o}\right)\exp\left[j\frac{k}{2d_i}\left(x^2+y^2\right)\right]. \tag{6.42}$$

The core of this expression is the term involving, once again, the Fourier transform of the pupil function. It is multiplied, however, by two quadratic phase factors, one a function of the object plane coordinates ($\xi$,$\eta$) and the other a function of the image plane coordinates (x,y). Note that the input to the superposition integral of Eq. (6.41) is the *object* distribution itself rather than the geometrical optics image. Magnification of the image is accounted for by the scaling of x and y by $d_i$ and of $\xi$ and $\eta$ by $d_o$ in the pupil-transform factor of the kernel, and inversion of the image is accounted for by the plus signs in place of minus signs. The corresponding system diagram is shown in Fig. 6-10. This block diagram is completely general and can be used to represent not only the systems studied thus far (including, as special cases, those that are space invariant) but any coherent imaging system consisting of lenses and apertures . The basic framework for determining such a representation is presented in the following section.


Fig. 6-10. General space-variant system diagram for coherent imaging system.

## 6.4 GENERAL FRAMEWORK FOR ANALYZING COHERENT IMAGING SYSTEMS

If it is to apply to an arbitrary coherent imaging system, the kernel of Eq. (6.42) must be rewritten with more general parameters $R_1$ and $R_2$ replacing $d_o$ and $d_i$ in the quadratic phase factors and with a more general *system pupil function* transform $\mathbf{P}_s$ replacing $\mathbf{P}$:

$$\mathbf{h}_{ss}(x,y;\xi,\eta) = \exp\left[j\frac{k}{2R_1}\left(\xi^2+\eta^2\right)\right]\frac{1}{\lambda^2 d_o d_i}\mathbf{P}_s\left(\frac{\xi}{\lambda d_o}+\frac{x}{\lambda d_i},\frac{\eta}{\lambda d_o}+\frac{y}{\lambda d_i}\right)$$
$$\times\exp\left[j\frac{k}{2R_2}\left(x^2+y^2\right)\right] \tag{6.43}$$

A ray-optics-based procedure for determining $R_1$, $R_2$, $d_i$, $d_o$, and system pupil function $p_s(x,y)$ is illustrated with the help of the example imaging system shown in Fig. 6-11. Although this system is rather specific, it nevertheless contains all elements necessary to allow it to serve as a general example. It is assumed that the focal lengths of the two lenses are known and that the locations of the object and image planes have been determined using geometrical optics methods [Fig. 6-11(a)]. Furthermore, it is assumed that all apertures in the system are circular and centered





on the optical axis. The first step of the procedure is to find the so-called *aperture stop* of the system. The aperture stop is the same as the limiting aperture discussed in Sec. 6-2. It is found, as shown in Fig. 6-11(b), by tracing a ray from the point on the optical axis in the object plane—such a ray is called an *axial ray*—and increasing the angle it makes with the axis until the ray first strikes an aperture somewhere in the system. That aperture is, by definition, the aperture stop of the system. An axial ray that just grazes the aperture stop is called a *marginal* axial ray. The aperture stop might be a specially-positioned stop, as is the case in the example, or it might be the edge of one of the lenses, as would be the case with the single-lens system of Fig. 6-9.

Once the aperture stop has been found, the locations of the *entrance and exit pupils* of the system must be determined. The entrance pupil is the aperture stop *as seen from the object plane*. As shown in Fig. 6-12(a), in the example system the entrance pupil takes the form of a virtual image of the aperture stop. Similarly, the exit pupil is the aperture stop *as seen from the image plane*. In the example, as shown in Fig. 6-12(b), the exit pupil, like the entrance pupil, also takes the form of a virtual image of the aperture stop. Having located the entrance and exit pupils it is possible to specify the parameters $R_1$ and $R_2$ in Eq. (6.43), for they are, respectively, the distance from the object to the entrance pupil and the distance from the exit pupil to the image. In the example, both $R_1$ and $R_2$ are finite. In the case of the systems of Figs. 6-1, 6-4, and 6-7, however, $R_1$ and, in some cases at least, $R_2$ are infinite. The implication of this latter condition will be addressed shortly.

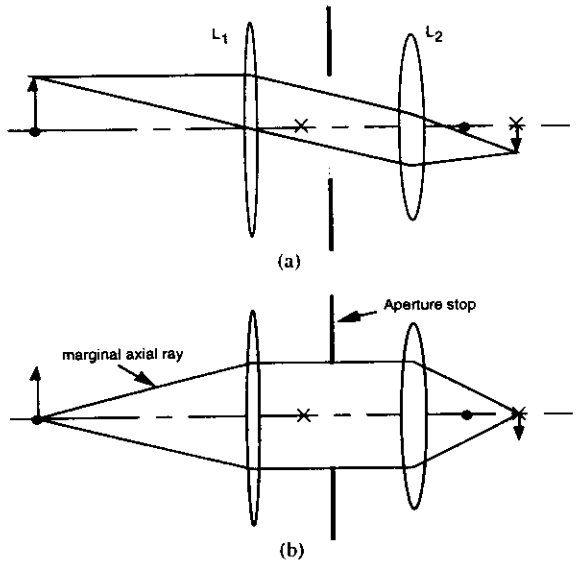



Fig. 6-11. Example of general coherent imaging system: (a) basic system showing object, image, and lens: associated focal planes are denoted by • for $L_1$ and × for $L_2$; (b) locating aperture stop by increasing angle of axial ray until first it strikes an aperture.

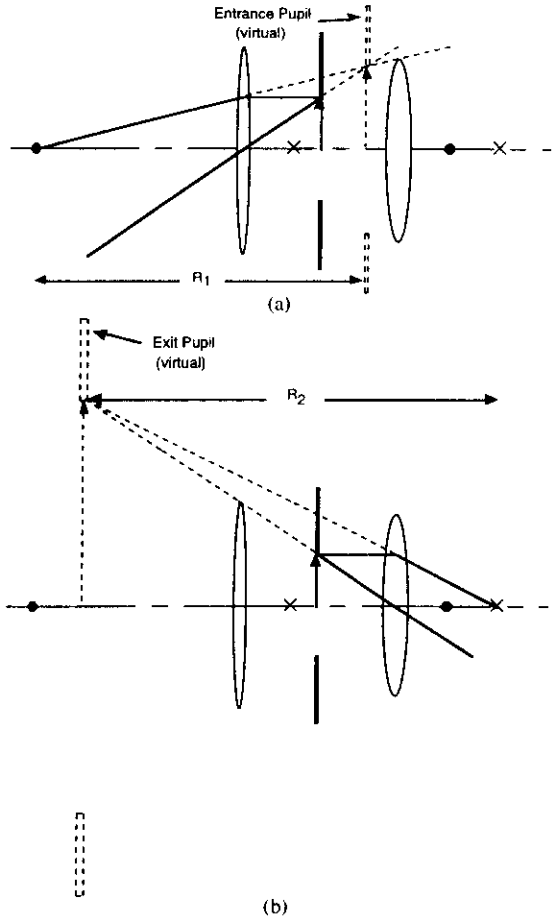



Fig. 6-12. Determining locations of entrance and exit pupils: (a) entrance pupil is determined by locating image of aperture stop as seen from object space; (b) exit pupil is determined by locating image of aperture stop as seen from image space.

The next step in the analysis is to determine a suitable system pupil function $\mathbf{p}_S(x,y)$ and associated distances $d_o$ and $d_i$.* We do this by identifying the plane in which a single lens can,

* The system pupil function and the two associated distances are in fact not uniquely specified for a given imaging system but depend on what method is used to determine them. One method is introduced in this section; another is presented in a homework problem. In either case, $\mathbf{p}_S(x,y)$ will be a scaled version of the aperture stop





*from a ray optics standpoint*, perform the same imaging operation. The location of the equivalent-lens plane is determined by tracing marginal axial rays from the object to the image, as shown in Fig. 6-13(a), projecting their object- and image-space segments back toward the center of the imaging system, and noting where they intersect. The plane of intersection is the plane of the equivalent single lens. The distances from that plane to the object and image planes are, respectively, $d_o$ and $d_i$. If the equivalent lens is to perform the desired imaging operation, it must have focal length f satisfying the condition $1/f = 1/d_o + 1/d_i$. Since the rays locating the lens plane are marginal axial rays, their intersection defines a scaled version of the aperture stop. It is this scaled version of the aperture stop that serves as the system pupil function. It has the same functional form as the aperture stop pupil function (e.g., *circ*), but its size depends on the geometry of the imaging system. The equivalent single-lens system is shown in Fig. 6-13(b). It is important to note that this system is equivalent only in a ray-optics sense. The results of Sect. 6.3 indicate clearly that the quadratic phase factors correct for the single-lens system are different from those for the system of Fig. 6-11.

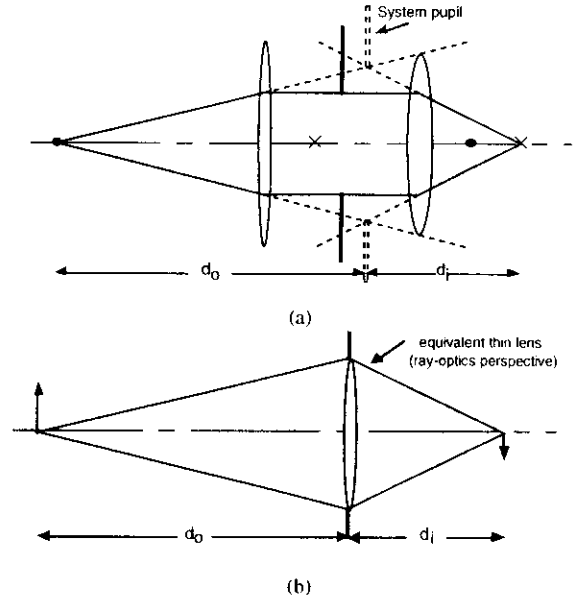



Fig. 6-13: Determining equivalent single-lens imaging system and, thereby, distances $d_o$ and $d_i$. Location of equivalent lens is determined by projecting marginal axial rays from object and image space and noting where they intersect, as shown in (a). The system pupil is defined by the intersection of the collection of all such marginal ray projections. The equivalent single thin-lens system is shown in (b).

pupil function and will be characterized by a diameter w. Of great importance: the ratios $w:d_o:d_i$ are characteristic of the specific system and is the same regardless of which method is chosen.

Once $R_1$, $R_2$, $d_i$, $d_o$, and $\mathbf{p}_s(x,y)$ have been found it is advisable to perform one additional step to determine the maximum allowable object size for which there is no vignetting by other apertures in the system. To perform this step, begin with the marginal axial ray determined in step 1 [refer to Fig. 6-11(b)] and, while keeping the ray pinned to the edge of the aperture stop, change its angle until it begins to be cut off by the edge of a component—in the case of the example by lens $L_2$. Figure 6-14 illustrates. Where that ray crosses the object plane determines the diameter of the largest allowable object if vignetting is to be avoided.

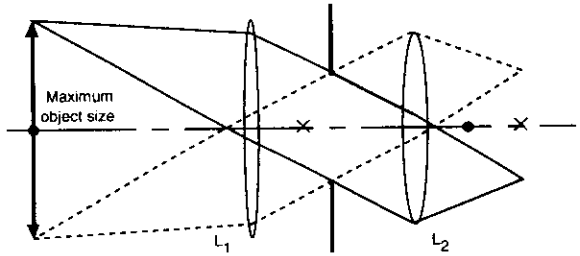



Fig. 6.14. Determination of maximum allowable object size.

## 6.5 NONMONOCHROMATIC COHERENT IMAGE FORMATION

Throughout the preceding analyses it has been assumed that the waves in the image forming process are monochromatic. If the object wave is not monochromatic, calculation of the image plane distribution can be much more complicated. Fortunately, the added complexity is not great if the system satisfies the quasimonochromatic condition, as specified by conditions (4.72) and (4.73). Under this condition, as noted in Sec. 4.7, the wave propagation theory developed for monochromatic waves can be applied with no modification other than the replacement of the complex wave amplitude with an appropriate time-varying phasor and the substitution of $\bar{\lambda}$ for $\lambda$. As a consequence, imaging by space invariant systems of the type analyzed in this chapter is still described by equations of the form given by Eq. (6.35), but with the complex amplitudes explicitly time-varying:

$$\mathbf{U}_{im}(x,y,t) = \left[\mathbf{U}_{\mathbf{g}}(x,y,t)**\mathbf{h}_{coh}(x,y)\right]\exp\left[j\frac{\bar{k}}{2R}\left(x^2+y^2\right)\right], \tag{6.44}$$

where $\bar{k}$ denotes the average value of the wave number, i.e., $\bar{k} = 2\pi/\bar{\lambda}$. The time-varying geometrical optics image amplitude $\mathbf{U}_g(x,y,t)$ is simply a scaled version of the time-varying object amplitude $\mathbf{U}_{obj}(x,y,t)$. Consistent with Eq. (3.47), the corresponding optical intensity is given by

$$I_{im}(x,y,t) = \left\langle\left|\mathbf{U}_{im}(x,y,t)\right|^2\right\rangle = \left\langle\left|\mathbf{U}_{\mathbf{g}}(x,y,t)**\mathbf{h}_{coh}(x,y)\right|^2\right\rangle, \tag{6.45}$$

the time average being taken over an appropriate interval.

If the object distribution is spatially coherent, $\mathbf{U}_{obj}(x,y,t)$ can be written in the form (see Sec. 3.5)





$$\mathbf{U}_{obj}(x,y,t) = \mathbf{U}_{obj_s}(x,y)\mathbf{B}(t), \tag{6.46}$$

where $\mathbf{B}(t)$, assuming it is referenced to the origin, is given by

$$\mathbf{B}(t) = \frac{\mathbf{U}_{obj}(0,0,t)}{\sqrt{\left\langle\left|\mathbf{U}_{obj}(0,0,t)\right|^2\right\rangle}}. \tag{6.47}$$

It is easily shown that under these conditions

$$\mathbf{U}_{im}(x,y,t) = \left\{\left[\mathbf{U}_{\mathbf{g}_s}(x,y)**\mathbf{h}_{coh}(x,y)\right]\exp\left[j\frac{\bar{k}}{2R}\left(x^2+y^2\right)\right]\right\}\mathbf{B}(t), \tag{6.48}$$

where $\mathbf{U}_{\mathbf{g}_s}(x,y)$ is the geometrical optics image associated with the spatial-function part of $\mathbf{U}_{obj}(x,y,t)$, i.e., with $\mathbf{U}_{obj_s}(x,y)$. Substitution of this expression in Eq. (6.45) shows that under the assumed conditions the image intensity is given by

$$I_{im}(x,y) = \left|\mathbf{U}_{g_s}(x,y)**\mathbf{h}_{coh}(x,y)\right|^2, \tag{6.49}$$

which is effectively the same as the expression obtained for monochromatic wave fields.

## 6.6 RESOLUTION IN COHERENT IMAGING

In many cases we are interested in the ability of an optical system to faithfully image the detail in an object. The ability of the system to resolve small detail is directly dependent on the spatial frequency bandwidth of the imaging operation. As noted in Sect. 6.1, the spatial frequency cutoff for coherent imaging with the system of Fig. 6-1, assuming a circular aperture, is given by $\rho_c = w/2\lambda f_2$, where w is the diameter of the aperture. This frequency is the maximum spatial frequency that can be observed in the *image* plane wave amplitude distribution. The corresponding object spatial frequency is given by $w/2\lambda f_1$. Object wave amplitude structures containing spatial frequencies up to $w/2\lambda f_1$ can thus be faithfully imaged. Recall that spatial frequencies for complex wave amplitudes correspond directly to the tilt angles associated with plane-wave components of the distributions. A plane wave component with a higher tilt angle corresponds to a point of light focused farther off axis in the pupil plane of the system. Plane wave components of the object wave field with sufficiently large tilt angles are focused to points of light so far off axis that they are blocked by the aperture, and the light never reaches the image plane.

It is informative to consider resolution in coherent imaging systems in connection with the imaging of sampled wave amplitude distributions. Assume a bandlimited complex wave amplitude object distribution $\mathbf{U}_0(x,y)$ with cutoff frequency $f_o$. This distribution can be Nyquist sampled with a sample spacing equaling $1/2f_o$ in the $x$- and $y$-directions. As was established in Chapt. 2, $\mathbf{U}_0(x,y)$ can be recovered by subjecting the sampled distribution to an ideal lowpass filtering operation, where the cutoff frequency of the filter equals $f_o$. This operation corresponds to imaging the sampled object distribution with a circular-aperture imaging system whose aperture diameter $w$ satisfies the condition $w/\lambda f_1 = 2f_o$. Equivalently, $\lambda f_1/w$ must equal the sample spacing. (The aperture could also be square with width $w$. See homework problem.)

## REFERENCES

## PROBLEMS

6.1 Verify that the impulse response of Eq. (6.10) approaches $\delta(x,y)$ in the geometrical optics limit, assuming the pupil-plane aperture to be clear on axis. How can you explain this result physically?

6.2 An imaging system of the kind illustrated in Fig. 6-1 has lenses of focal length $f_1 = f_2 = 10$ cm. In the pupil plane of the system is placed the triangular aperture shown in Fig. P6.2. Assuming monochromatic object distribution

$$\mathbf{U}_{obj}(x,y) = \frac{1}{2}\left[1 + \cos(2\pi f_o x)\right]$$

at wavelength $\lambda = 500$ nm, find the maximum spatial frequency $f_o$, in cycles per millimeter, for which the image distribution $\mathbf{U}_{im}(x,y)$ equals the geometrical optics image $\mathbf{U}_{\mathbf{g}}(x,y)$.

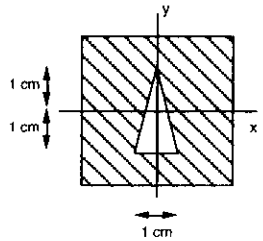



Fig. P6.2. Triangular aperture.

6.3 Show that Eq. (6.39) gives the correct condition for vignetting by the lens in coherent imaging performed by the single-lens system of Fig. 6-6.

6.4 The following specifications apply to the 2-lens imaging system shown in Fig. 6-1: object diameter $D = 2$ cm, pupil diameter $d = 3$ cm, $f_1 = 30$ cm, $f_2 = 40$ cm, and $\lambda = 500$ nm.

(a) Assuming no vignetting by the lenses, find the highest spatial frequency present in the image plane wave amplitude distribution. Express your answer in $mm^{-1}$.

(b) With reference to part (a), what is the corresponding object spatial frequency?

(c) How large must (i) lens $L_1$ and (ii) lens $L_2$ be to avoid vignetting? Express your answer in centimeters.

6.5 The flow of radiant power (measured, for example, in watts) through a plane is proportional to the spatial integral of the optical intensity incident on the plane,





$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_{inc}(x,y)\,dx\,dy.$$

Show that the radiant power flow through the pupil plane aperture of the imaging system of Fig. 6-1 equals the radiant power flow at the image plane, assuming for convenience that $M = 1$. Ignore vignetting.

6.6 Assuming the transfer function of Eq. (6.17), find the highest image spatial frequency that can be found in the image and specify the direction of that spatial frequency.

6.7 The object-image relationship for the system of Fig. 6-1 can be expressed in the form given in Eq. (6.41).

(a) Find the corresponding space-variant kernel $\mathbf{h}_a(x,h;\xi,\eta)$.

(b) Find the form of the kernel in the limit $\lambda \to 0$.

6.8 With reference to step (4) of the procedure of Sec. 6.3, show that if Uobj(x,y)=1, Eq. (6.40) reduces to $\mathbf{U}_{im}(x,y) = \text{const.} \times \exp\left[j\frac{k}{2R}(x^2+y^2)\right]$.

6.9 The imaging system can be analyzed following the procedure outlined in Sec. 6.4, find the following parameters of the imaging system illustrated in Fig. 6-7. Assuming the lens to be infinite in diameter:

(a) Find the distance from the object plane to the entrance pupil, $R_1$.

(b) Find the distance from the exit pupil to the image plane, $R_2$.

(c) Determine the location of the system pupil and the associated distance $d_o$ and $d_i$.

(d) Show by substitution in Eq. (6.43) and appropriate change in variables that Eq. (6.41) can be recast, for this system, in the form given by Eq. (6.35)—i.e., show that the two methods to analyzing the system of Fig. 6-7, that of Sec. 6.2 and that of 6.4, lead to essentially the same results.

(e) Assuming that the pupil function $\mathbf{p}(x,y)$ is given by rect(x/w,y/w) and that the system magnification is unity ($d_i/d_o=1$), find an appropriate expression for the system pupil function $\mathbf{p}_s(x,y)$.

6.10 In the following problem, refer to the system of Fig. 6-7 and assume the following parameters: $f = 20$ cm, $d_o = d_i = 40$ cm. Assume further that the lens is sufficiently large in diameter that there is no vignetting and that the limiting aperture is characterized by pupil function p(x,y) = rect(x/w,y/w), where $w = 1$ cm. The object consists of two monochromatic point sources of wavelength $\lambda = 500$ nm positioned about the origin on the x-axis and separated a distance S. (Such an object could be produced by illuminating a two-pinhole mask with a normally incident monochromatic plane wave.) The complex amplitude of the image is therefore given by the sum of two sinc functions, perhaps with associated quadratic phase factors.

(a) For what point-source separation are the two sinc functions separated by their peak-to-first-null distance?

(b) Assume that the phase of one of the object point sources in part (a) is shifted by 180° (e.g., by placing a half-wave retarder in front of one of the pinholes). Show by means of an x-axis cross-section sketch the appearance of the image plane *intensity*.

6.11 Assume that a point object, represented by $\mathbf{U}_{obj}(x,y) = \delta(x-\xi,y-\eta)$, is located at coordinates $(\xi,\eta)$ in the object plane of the system of Fig. 6-9(a). Use Eqs. (4.26a) and (5.1) to show that the complex amplitude distribution in the conjugate image plane is given (to within a constant phase factor that does not depend on x, y, $\xi$, or $\eta$) by the expression of Eq. (6.42).

6.12 The imaging system illustrated in Fig. 6.4(c) is used to image a monochromatic object with complex wave amplitude given by $\mathbf{U}_{obj}(x,y) = 1 + \cos 2\pi f_o x$. Object distance $d_o$, wavelength $\lambda$, spatial frequency $f_o$, and focal length $f_{eq}$ are given by

$$d_o = 20cm,\ \lambda = 633nm,\ f_o = 400mm^{-1},\ f_{eq} = 8cm.$$

On axis in the pupil plane of the system is placed a square aperture of width $w$.

(a) Sketch the spatial frequency spectrum of the object amplitude distribution, $\hat{\mathbf{U}}_{obj}(u,v)$. Scale the axes in $mm^{-1}$.

(b) Find the image distance $d_i$, in cm.

(c) Find the imaging magnification M.

(d) Find the complex amplitude of the geometrical optics image, $\mathbf{U}_g(x,y)$.

(e) Sketch the spatial frequency spectrum of the geometrical optics image amplitude distribution, $\hat{\mathbf{U}}_g(u,v)$. Scale the axes in $mm^{-1}$.

(f) Find the optical intensity of the geometrical optics image, $I_g(x,y)$.

(g) Sketch the spatial frequency spectrum of the geometrical optics image intensity distribution, $\hat{I}_g(u,v)$. Scale the axes in $mm^{-1}$.

(h) Find the coherent imaging complex amplitude transfer function $\mathbf{H}_{coh}(u,v)$.

(i) Sketch the appearance in the Fourier transform domain of $\mathbf{U}_g(x,y)$ and $\mathbf{H}_{coh}(u,v)$. Show clearly the relationship between the spatial frequency components of the geometrical optics image and the coherent transfer function cutoff frequencies in the u and v directions.

(j) What is the minimum allowable aperture size w, in mm, if the actual image distribution, $\mathbf{U}_{im}(x,y)$, is to be the same as the geometrical optics image?

(k) Assume that the aperture remains the same size but that it is rotated through 45°. How much larger may $f_o$ be in this case if $\mathbf{U}_{im}(x,y)$ is still to equal $\mathbf{U}_g(x,y)$?





## INDEX - CHAPTER 6