NASA TECHNICAL REPORT

NASA TR R-319

NASA TR R-319

CASE FILE COPY

FOURIER TRANSFORM REPRESENTATION OF AN IDEAL LENS IN COHERENT OPTICAL SYSTEMS

by Gerald J. Grebowsky

Goddard Space Flight Center

Greenbelt, Md.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • JANUARY 1970

NASA TR R-319

FOURIER TRANSFORM REPRESENTATION OF AN IDEAL LENS IN COHERENT OPTICAL SYSTEMS

By Gerald J. Grebowsky

Goddard Space Flight Center
Greenbelt, Md.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

For sale by the Clearinghouse for Federal Scientific and Technical Information
Springfield, Virginia 22151 – Price $3.00

ABSTRACT

This document presents a mathematical analysis of the approximations required to obtain the Fourier transform representation of an ideal lens. An attempt is made throughout the paper to demonstrate the physical significance of the approximations, and the variations from ideal results, produced by neglected terms in the mathematical formulation. The approximations involved are considered in terms of the output signals in optical spectrum analyzer, optical imaging, and optical correlator systems.

CONTENTS

FOURIER TRANSFORM REPRESENTATION OF AN IDEAL LENS IN COHERENT OPTICAL SYSTEMS

by
Gerald J. Grebowsky
Goddard Space Flight Center

INTRODUCTION

In recent years there has been a growing interest in the application of optical imaging techniques for the purpose of processing data signals. These efforts are largely based on the interpretation of optical imaging systems as spatial filters (Reference 1). By introducing Fourier transform methods the relation between an object and its image has the same form as the relation between the input and output signals of an electrical system. Comparing the transform of the object and image, the imaging process (unity magnification is assumed here) can be described by

$$I(\omega_x, \omega_y) = T(\omega_x, \omega_y)\, O(\omega_x, \omega_y) ,$$

or

$$\text{Image spectrum} = \text{Transfer function} \times \text{Object spectrum} .$$

This expression has the same form as that of an electrical network except that the optical spectrums are two-dimensional. Since optical objects and images are two-dimensional, a Fourier transform must be taken with respect to two spatial coordinates instead of the single time coordinate which appears in electrical signals.

It is the transfer function relation given above that leads to the spatial filter interpretation of optical imaging systems. The optical transfer function, $T(\omega_x, \omega_y)$, is a characteristic of the optical elements in a system. An ideal imaging system would have a transfer function that is constant over the object frequency range. In such an ideal system the image would be an exact replica of the object. The corresponding electrical system would have a flat frequency response over the bandwidth of the input signal.

It may appear at first thought that the transfer-function notation is nothing more than an arbitrary selection of notation. However, in optical imaging systems, an object represented by a sinusoidal spatial variation of light amplitude is imaged as such—even in the presence of aberrations (Reference 2). Aberration effects appear as reduced contrast and lateral shift of the sinusoidal test image. Thus, using sinusoidal test gratings, it is theoretically possible to experimentally determine the transfer function for a given optical system. In general, the optical transfer function can have complex values. The magnitude is related to the reduction in contrast, and the phase is related to the lateral shift of the image. In an actual system, the transfer function will not have the constant amplitude and phase of the ideal imaging system described above.

Since the optical transfer function is determined by comparing the light amplitude variations of the output image to that of the original object input, the introduction of any additional element into the optical system to vary the amplitude and/or phase transmission properties will change the optical transfer function. To readily utilize an optical system as a spatial filter in a fairly direct manner, it is necessary to know what amplitude and phase variations should be inserted and where they should be inserted. Otherwise, obtaining a particular transfer function for a spatial filter application would be a trial and error proposition. This implementation problem is solved for many cases of practical importance by the optical Fourier transform representation discussed in this report.

Within certain limitations, the light amplitude distribution in the back focal plane of a lense is proportional to the two-dimensional Fourier transform of the light amplitude distribution of a two-dimensional object inserted on the front side of the lens. Within the range of validity for this optical Fourier transform representation, the transfer function is varied by a multiplicative factor represented by the amplitude and phase transmission properties of an element inserted into the back focal plane of the lens. For example, to set the transfer function equal to zero for a particular frequency component, the light passing through the corresponding point in the back focal plane of the lens is simply blocked.

The mathematical development of the optical Fourier transform representation presented in this report is intended to clarify the limitations and interpretation of the Fourier transform operation of a lens. The derivation is based on the Rayleigh-Sommerfeld diffraction formula and on optical paths defined by geometrical ray tracing. As each limitation is introduced, an attempt is made to describe the effects on the accuracy of the optical Fourier transform representation. Such detailed consideration has been found lacking in available treatments of the derivation (Reference 3) and is the main purpose for the development presented in this report.

FOCAL PROPERTIES OF A LENS

To derive the formula for a focussed diffraction pattern, the properties of an ideal lens are defined. Our discussion is restricted to the case of an ideal lens; the effects of lens aberrations and diffraction at the edge of the lens are ignored. These effects are assumed to be taken into account by modifying the end result or by restricting the range of variables to a region in which the ideal assumptions are valid within experimental accuracies.

The definition of an ideal lens is based on the geometrical focussing properties shown in Figure 1. The properties assumed can be stated as follows:

1. The lens can be represented by a plane, L, perpendicular to the optical axis, with all refraction taking place at this plane. This is a thin lens approximation that neglects the thickness of the lens.

2. The rays passing through the point O (intersection of the optical axis and the lens plane L) are called principal rays; they are not deviated.

3. All incident rays parallel to a principal ray will be focussed at the point where the principal ray intersects the back focal plane F′. That is, the light reaching a point in the back focal plane, F′, at a distance, $\rho = f \tan\theta$ (where f is the focal length—the distance between the planes L and F′) from the optical axis, is contributed by a principal ray making an angle θ with the optical axis plus all rays initially parallel to this principal ray.

4. If a plane P is constructed perpendicular to a bundle of parallel incident rays, the optical path length will be the same for any of the parallel rays moving from P to the common point of focus in F′.

For any set of parallel rays, Figure 1 represents the projection of the parallel rays onto the plane through the optical axis and the principal ray as shown in Figure 2. In Figure 1 the distances ρ and ρ_1 are measured from the optical axis in the plane of the figure. As shown in Figure 2, these distances represent different quantities in the planes F and F′ respectively. The distance ρ in the back focal plane F′ is the distance from the optical axis to a point (x, y). In the plane F, ρ_1 is *not* the distance from the optical axis to a point (x_1, y_1); ρ_1 is the projection of this distance onto the axis defined by the intersection of the plane F with the plane through the optical axis and a point (x, y) in the plane F′. Since the orientation of the ρ_1 axis will depend on the position (x, y) in F′, ρ_1 will be a function of x, y, x_1, and y_1 whereas ρ depends only on x and y. The difference in meaning of ρ and ρ_1 also appears when the algebraic sign is considered. In Figures 1 and 2, ρ is the distance from the optical axis to the point of focus and is a positive quantity regardless of where the point is located. On the other hand, ρ_1 is a coordinate of the intersection of a ray with the F plane and we will use

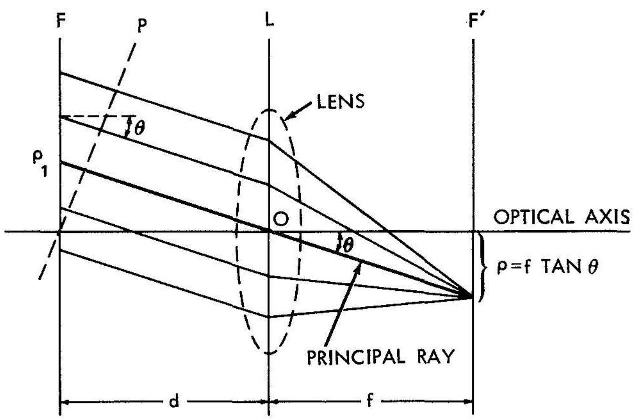

Figure 1—Ideal focusing of parallel rays.

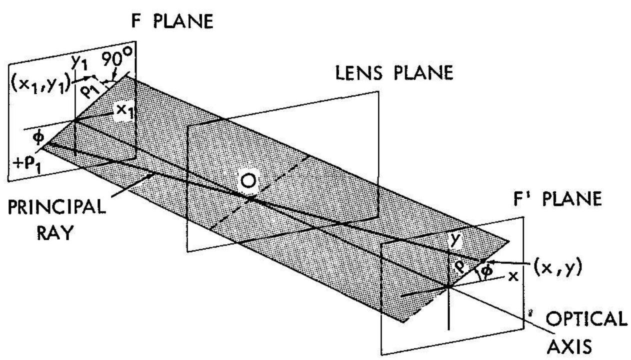

Figure 2—Difference between ρ and ρ_1.

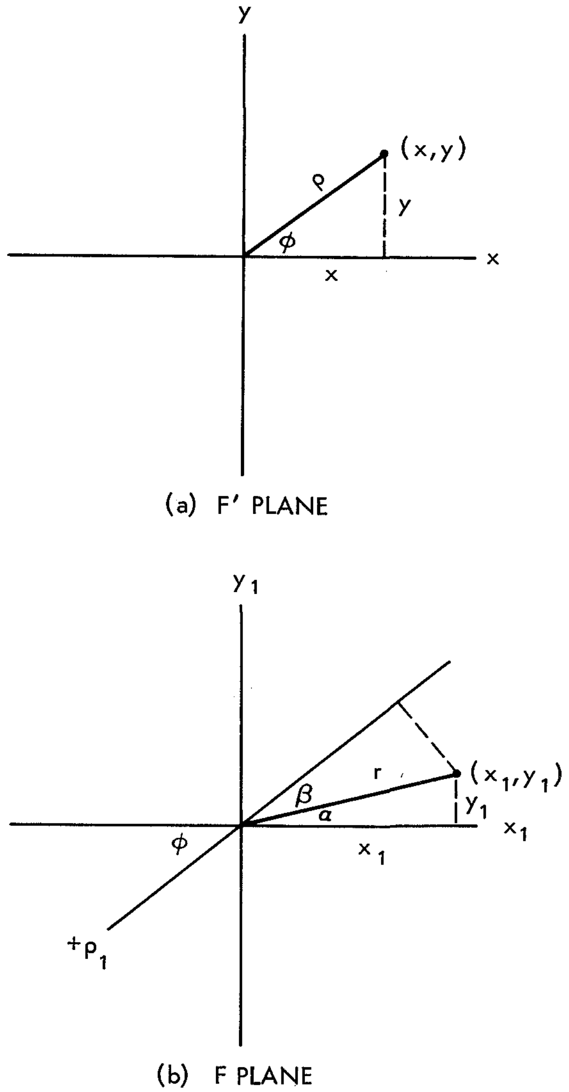

Figure 3—Geometry for ρ and ρ_1.

the sign convention that ρ_1 is positive above the optical axis and negative below it when drawn as in Figure 1 (rays sloping down to the right). In reference to a point (x, y) in F′, the positive ρ_1 axis will lie in the quadrant opposite to that of the point (x, y).

Figure 3 shows the geometry of the various lines in planes F and F′ and demonstrates the sign convention for ρ_1. From the geometry of Figure 3(a), the following relations are found:

$$\cos\phi = \frac{x}{\rho}, \qquad \sin\phi = \frac{y}{\rho},$$

$$\rho = \left(x^2 + y^2\right)^{1/2}. \tag{1}$$

From the geometry of Figure 3(b), ρ_1 can be determined for any point (x_1, y_1) as follows:

$$\phi = \alpha + \beta, \qquad \cos\alpha = \frac{x_1}{r},$$

$$\sin\alpha = \frac{y_1}{r}, \qquad r = \left(x_1^2 + y_1^2\right)^{1/2},$$

$$\rho_1 = -r\cos\beta = -r\cos(\phi - \alpha)$$
$$= -r\{\cos\phi\cos\alpha + \sin\phi\sin\alpha\},$$

or

$$\rho_1 = -x_1\cos\phi - y_1\sin\phi.$$

Substituting the relations for $\cos\phi$ and $\sin\phi$ obtained from Figure 3(a), the expression for ρ_1 can be rewritten as

$$\rho_1 = \frac{-xx_1 - yy_1}{\rho} = \frac{-xx_1 - yy_1}{\left(x^2 + y^2\right)^{1/2}}. \tag{2}$$

The dependence of ρ on x and y and of ρ_1 on x, y, x_1, and y_1 is explicit in Equations 1 and 2 respectively. These results agree with the discussion in the last paragraph.

Figure 4 is now used to determine an expression for the optical path length from a point (x_1, y_1) in F to a point (x, y) in F′. Figure 4 shows a principal ray AA′ and a representative parallel ray (dotted). The plane P is perpendicular to the incident rays. Since the thickness of the lens is neglected, the optical path length r for the principal ray AA′ will be given by the geometrical length:

$$r = \ell_1 + \ell_2 + \ell_3 . \tag{3}$$

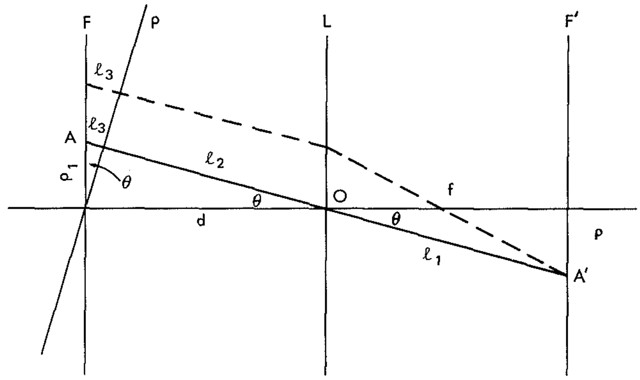

Figure 4—Geometry for optical path length.

From the geometry of Figure 4, ℓ_1 is given by the expression

$$\ell_1 = \left(f^2 + \rho^2\right)^{1/2} . \tag{4}$$

By similar triangles, ℓ_2 is given by

$$\frac{\ell_2}{d} = \frac{f}{\ell_1} \qquad \text{or} \qquad \ell_2 = \frac{df}{\ell_1} . \tag{5}$$

For the principal ray, AA′, the optical path length from the plane P to the point A′ in the back focal plane F′ is $\ell_1 + \ell_2$. Using Equations 4 and 5 this length is given as

$$\ell_1 + \ell_2 = \ell_1 + \frac{df}{\ell_1} = \frac{\ell_1^2 + df}{\ell_1} = \frac{f^2 + \rho^2 + df}{\left(f^2 + \rho^2\right)^{1/2}} .$$

By the fourth assumption for an ideal lens, the optical path length from the plane P to the focus point A′ is the same for all the rays parallel to AA′ (note that $\ell_1 + \ell_2$ does not depend on ρ_1). Therefore the expression found for $\ell_1 + \ell_2$ holds for every parallel ray and the expression for the optical path length r can be written

$$r = \ell_1 + \ell_2 + \ell_3 = \frac{f^2 + \rho^2 + df}{\left(f^2 + \rho^2\right)^{1/2}} + \ell_3 . \tag{6}$$

The term ℓ_3 remains to be determined. By comparison for the principal ray and the representative ray in Figure 4 it should be obvious that this term will not be the same for all parallel rays. From the right triangle with ℓ_3 as a leg, ℓ_3 can be expressed as

$$\ell_3 = \rho_1 \sin\theta .$$

However, $\sin\theta = \rho/(f^2 + \rho^2)^{1/2}$, so the expression for ℓ_3 can be rewritten as

$$\ell_3 = \frac{\rho\rho_1}{(f^2 + \rho^2)^{1/2}} . \tag{7}$$

Substituting for ℓ_3 in the expression for the optical path length r, then

$$r = \frac{f^2 + df + \rho^2 + \rho\rho_1}{(f^2 + \rho^2)^{1/2}} . \tag{8}$$

Previously derived expressions for ρ and ρ_1 are given by Equations 1 and 2. Substituting for ρ and ρ_1 in Equation 8 we obtain

$$r = \frac{f^2 + x^2 + y^2 + df - xx_1 - yy_1}{(f^2 + x^2 + y^2)^{1/2}} . \tag{9}$$

This expression gives the optical path length from any point (x_1, y_1) in a plane F (a distance d in front of the lens) to a point (x, y) in the back focal plane F′. To facilitate further discussion, this expression can be written

$$r = R(x, y) - \alpha x_1 - \beta y_1 , \tag{10}$$

where

$$R(x, y) = \frac{f^2 + df + x^2 + y^2}{(f^2 + x^2 + y^2)^{1/2}} ,$$

$$\alpha = \frac{x}{(f^2 + x^2 + y^2)^{1/2}} ,$$

and

$$\beta = \frac{y}{(f^2 + x^2 + y^2)^{1/2}} .$$

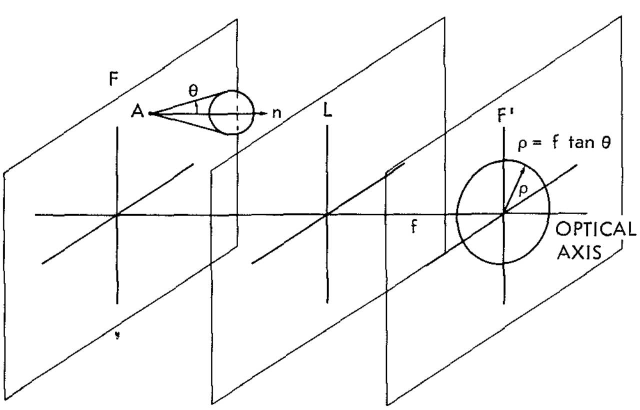

Figure 5—Focussing of a cone of light.

To summarize our results at this point, consider a point A in the plane F as shown in Figure 5. Assuming that light is radiated in all directions from the point A, consider the portion of light propagated in directions at an angle θ with respect to the normal, n, to the plane F. The light rays representing these directions will form the surface of a cone of half angle θ as shown in Figure 5. For each of these rays a parallel principal ray can be drawn and each principal ray will make an angle θ with the optical axis. Thus for each ray at an angle θ with respect to the normal to F at the point A, the results derived above can be applied. That is, each ray is focussed to a point on the ring of radius $\rho = f \tan\theta$ where the corresponding principal ray intersects the back focal plane F'. The optical path length from A to each point on the ring is given by Equation 10. Since this holds for any point A, the general focal properties of the ideal lens can be stated:

1. The light radiated from all points on F in directions at an angle θ with respect to the normal to F is focussed into a ring of radius $\rho = f \tan\theta$ in the back focal plane F'.
2. The optical path length r from any point (x_1, y_1) on the plane F to any point (x, y) in the back focal plane F' is given by Equation 10.

In these statements the plane F is a plane perpendicular to the optical axis at a distance d in front of the lens.

Anticipating the derivations in the next section, the relation between ρ and θ, specified by the first focal property above, can also be expressed in terms of $\cos\theta$ as

$$\cos\theta = \frac{f}{\left(f^2 + \rho^2\right)^{1/2}} = \frac{f}{\left(f^2 + x^2 + y^2\right)^{1/2}} . \tag{11}$$

This expression can be derived from the geometry of the figures in this section or from the equation, $\rho = f \tan\theta$, in terms of $\tan\theta$ by applying trigonometric identities.

FOCUSED DIFFRACTION PATTERN

Since only light distributions on plane surfaces are being considered, the Rayleigh-Sommerfeld diffraction formula can be used (see Appendix A). In rectangular coordinates this diffraction formula has the form:

$$A(x, y, z) = \frac{-1}{2\pi} \iint A'\left(x_1, y_1\right) \frac{e^{ikr}}{r} \left[ik - \frac{1}{r}\right] \cos\theta \, dx_1 \, dy_1 . \tag{12}$$

This formula gives the complex light amplitude A(x, y, z) at any point in space ($z > 0$) caused by a monochromatic coherent light distribution $A'(x_1, y_1)$ given for every point (x_1, y_1) in a plane F. Referring to Figure 6, the terms in the diffraction formula are defined as

1. $A'(x_1, y_1)$ is the complex amplitude of monochromatic light given for all points (x_1, y_1) in a plane F located at $z = 0$.
2. A(x, y, z) is the complex amplitude of light produced by $A(x_1, y_1)$ at a point (x, y, z) in space ($z \geq 0$).

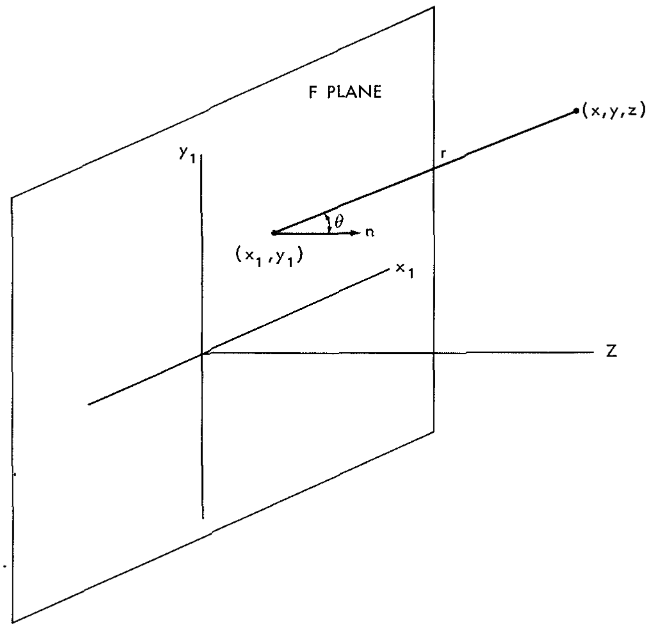

Figure 6—Relation between points (x, y, z) and (x_1, y_1) in diffraction formula.

3. r is the distance from a point (x_1, y_1) in plane F to the point (x, y, z).

4. θ is the angle between r and n, where r is directed from (x_1, y_1) to (x, y, z) and n is the normal to the plane F at (x_1, y_1) in the direction of the positive z axis. The term $\cos\theta$ is usually referred to as the obliquity factor.

5. $k = 2\pi/\lambda$, where λ is the wavelength of the monochromatic light.

In general each of the three vector components of the electromagnetic field representing the light distribution must be determined by the diffraction formula. This discussion considers the scalar light amplitude distribution which requires only one equation (Reference 4). In practice this is permissible if polarization affects can be neglected. Thus the light amplitude can be defined such that the square of its absolute magnitude gives the intensity, which is a measurable quantity.

The diffraction formula can be immediately simplified by considering the relative magnitudes of the terms inside the brackets:

$$\left[ik - \frac{1}{r}\right] = i\,\frac{2\pi}{\lambda} - \frac{1}{r}\,.$$

For wavelengths λ as long as 100 microns (far infrared) the first term is relatively large (600) while for r larger than 1 cm the second term is less than one. For visible light λ is much less than 100 microns (0.4 to 0.7 microns), and k is of the order of 10^5. Since this discussion (as in most cases in optics) will deal only with r greater than one centimeter, the $1/r$ term is negligible and can be dropped without any appreciable effect on accuracy. The diffraction formula, Equation 12, can therefore be written

$$A(x, y, z) = \frac{-i}{\lambda}\iint A'(x_1, y_1)\,\frac{e^{ikr}}{r}\,\cos\theta\,dx_1\,dy_1\,, \tag{13}$$

where the constant factor, ik, has been taken outside the integral.

The obliquity factor, $\cos\theta$, is a weighting factor that accounts for the difference in the amount of light radiated in different directions. Since $\cos\theta$ has a maximum value of one when θ equals zero, this factor has a maximum value of one for light contributions propagated normal to the signal plane and drops off as the angle with respect to the surface normal increases. Referring to Figure 6, if the light from a point (x_1, y_1) contributing to the light at the point (x, y, z) is assumed to travel the straight line r, this line is a light ray at an angle θ to the normal n. In the previous section, it was shown that through the focal property of an ideal lens, this angle is a constant for all light contributing to a point (x, y) in the back focal plane and that x, y, and θ are related by the expression

$$\cos\theta = \frac{f}{\left(f^2 + x^2 + y^2\right)^{1/2}} . \qquad (11)$$

In other words, an ideal lens focuses light of constant obliquity factor into a ring of radius $\left(x^2 + y^2\right)^{1/2}$ specified for a given θ by the above expression.

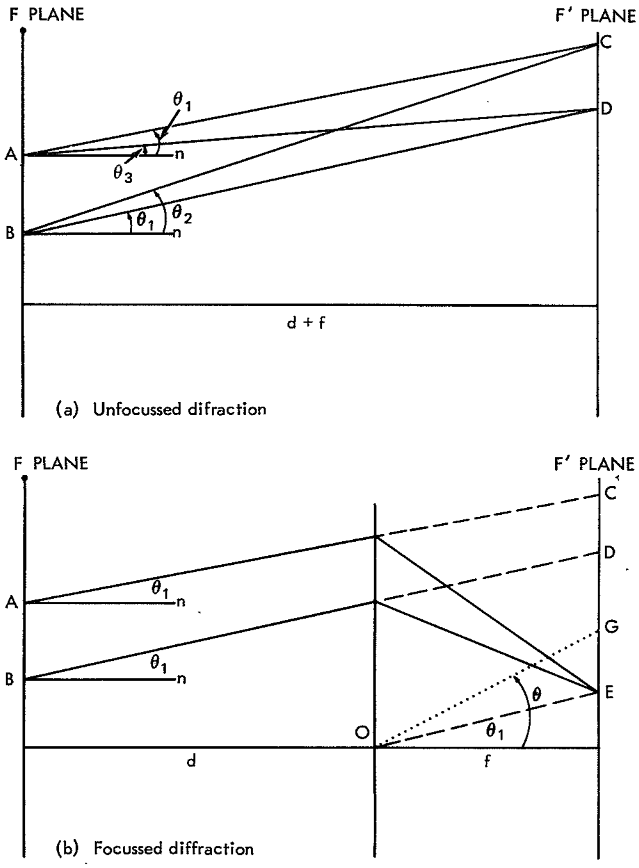

Figure 7—Comparison of diffraction with and without focussing.

The significance of this focal effect can be seen by comparing the two diagrams in Figure 7. In Figure 7(a), the points A and B are sample points in the F plane, and the points C and D are sample points in a parallel plane at $z = d + f$. Considering the point C, it is noted that the paths AC and BC have obliquity factors of $\cos\theta_1$ and $\cos\theta_2$ respectively. From this example it is obvious that for a point such as C (or D) the obliquity factor will depend on the location of the contributing point (x_1, y_1) in F. Likewise, considering the point A in F, it is noted that the paths AC and AD have obliquity factors of $\cos\theta_1$ and θ_3 respectively. This indicates that the obliquity factor also depends on the location of the point (x, y, z). Since determining the obliquity factor is included in the derivation of the diffraction formula, it is given here without proof for the case $z = d + f$ as shown in Figure 7(a):

$$\cos\theta = \frac{d + f}{\left[\left(x - x_1\right)^2 + \left(y - y_1\right)^2 + (d + f)^2\right]^{1/2}} . \qquad (14)$$

This expression includes the coordinates of both the point (x_1, y_1), in the source plane, and the point (x, y, z), at which the diffracted light amplitude is to be found. The expression for the general case will have a z in place of the $(d + f)$ used for the special case of Figure 7(a). In the diffraction formula, the obliquity factor appears under the integral sign since x_1 and y_1 are the variables of integration and appear in the obliquity factor as given by Equation 14.

Let us now consider the case of focussed diffraction as illustrated in Figure 7(b). In Figure 7(b), only the rays AC and BD of Figure 7(a) are considered, and as indicated by the angle θ_1, AC and BD are parallel rays. The dotted portion of these rays indicates the path of light followed in Figure 7(a). Because of refraction by the lens, these paths are changed to those focussed to the point E. Now, considering a point such as E, then the obliquity factor $\cos\theta_1$ is the same for points A and B and therefore independent of the coordinates (x_1, y_1) of the point in F. Considering any other point G in F′, it is recalled that to contribute to a point G a ray must be parallel to the principal ray OG. Rays parallel to OG will have an obliquity factor $\cos\theta$ different from $\cos\theta_1$ for the point E. Thus the obliquity factor does depend on the location of the point (x, y) in the back focal plane. The obliquity factor for the case of a focussed diffraction pattern is given by Equation 11 as $\cos\theta = f/(f^2 + x^2 + y^2)^{1/2}$ and does not depend on the coordinates x_1 and y_1.

Since the obliquity factor for the focussed diffraction pattern is independent of the integration variables x_1 and y_1, this factor can be taken outside the integral and Equation 13 can be written as

$$A(x, y) = \frac{-if}{\lambda\left(f^2 + x^2 + y^2\right)^{1/2}} \iint A'\left(x_1, y_1\right) \frac{e^{ikr}}{r}\, dx_1\, dy_1\,, \tag{15}$$

where $A(x, y)$ now represents the complex light amplitude at a point (x, y) in the back focal plane of a lens.

To complete the discussion, now consider the term r which was defined as the distance from the contributing point to the point of interest. In Figure 6 this distance is measured along the straight line from (x_1, y_1) to (x, y, z). In Figure 7(b), the light traveling from A to E does not follow a straight line because of refraction at the lens plane L. Assuming that the effects of the length of the refracted path are the same as traveling an equivalent distance in a straight line, then the r in the diffraction formula can be interpreted as the optical path length determined in the previous section. Thus e^{ikr} represents the change in phase over an optical length r, and $1/r$ is an attenuation factor that decreases the amplitude contribution as the optical path length increases.

Substituting the optical path length expression for r, as given by Equations 9 and 10, Equation 15 becomes

$$A(x, y) = \frac{-if}{\lambda\left(f^2 + x^2 + y^2\right)^{1/2}} \iint \frac{A'\left(x_2, y_1\right) e^{ik[R(x,y) - \alpha x_1 - \beta y_1]}}{\left[\dfrac{f^2 + x^2 + y^2 + df - xx_1 - yy_1}{\left(f^2 + x^2 + y^2\right)^{1/2}}\right]}\, dx_1\, dy_1\,. \tag{16}$$

The terms $\left(f^2 + x^2 + y^2\right)^{1/2}$ cancel and $e^{ikR(x,y)}$ can be taken outside the integral to give

$$A(x, y) = \frac{-if}{\lambda} e^{ikR(x,y)} \iint \frac{A'\left(x_1, y_1\right) e^{-i2\pi[\alpha x_1/\lambda + \beta y_1/\lambda]}}{f^2 + x^2 + y^2 + df - xx_1 - yy_1}\, dx_1\, dy_1\,. \tag{17}$$

Now introducing the new variables p and q defined as

$$p = \frac{\alpha}{\lambda} = \frac{x}{\lambda\left(f^2 + x^2 + y^2\right)^{1/2}}, \tag{18a}$$

and

$$q = \frac{\beta}{\lambda} = \frac{y}{\lambda\left(f^2 + x^2 + y^2\right)^{1/2}}, \tag{18b}$$

and factor $f^2 + df$ from the denominator of the integral of Equation 17, then

$$A(x, y) = \frac{-i}{\lambda(f + d)} e^{ikR(x,y)} \iint \frac{A'\left(x_1, y_1\right) e^{-i2\pi(px_1 + qy_1)}}{1 + \frac{x\left(x - x_1\right) + y\left(y - y_1\right)}{f(f + d)}} dx_1\, dy_1 . \tag{19}$$

FOURIER TRANSFORM APPROXIMATION

The algebraic identity,

$$\frac{1}{1 + \frac{M}{N}} = 1 - \frac{1}{1 + \frac{N}{M}},$$

can be used to obtain

$$\frac{1}{1 + \frac{x\left(x - x_1\right) + y\left(y - y_1\right)}{f(f + d)}} = 1 - \frac{1}{1 + \frac{f(f + d)}{x\left(x - x_1\right) + y\left(y - y_1\right)}} .$$

Introducing this identity, Equation 19 can be written as

$$A(x, y) = \frac{-i}{\lambda(f + d)} e^{ikR(x,y)} \iint \left[1 - \frac{1}{1 + \frac{f(f + d)}{x\left(x - x_1\right) + y\left(y - y_1\right)}}\right] A'\left(x_1, y_1\right) e^{-i2\pi(px_1 + qy_1)} dx_1\, dy_1 . \tag{20}$$

Equation 20 can be rewritten with an integral for each term in the bracket so that

$$A(x, y) = \frac{-i}{\lambda(f + d)} e^{ikR(x,y)} \iint A'\left(x_1, y_1\right) e^{-i2\pi(px_1 + qy_1)} dx_1\, dy_1$$

$$+ \frac{i}{\lambda(f + d)} e^{ikR(x,y)} \iint \frac{A'\left(x_1, y_1\right) e^{-i2\pi(px_1 + qy_1)}}{1 + \frac{f(f + d)}{x\left(x - x_1\right) + y\left(y - y_1\right)}} dx_1\, dy_1 . \tag{21}$$

By restricting the maximum values (aperture limits) of x, y, x_1, and y_1, the second integral of Equation 21 can be made negligible compared to the first since the denominator of the integrand can be made large. This approximation will be discussed in more detail later; here it is simply assumed that it is possible to neglect the second integral.

The diffraction formula can then be written approximately as

$$A(x, y) = \frac{-i\, e^{ikR(x,y)}}{\lambda(f + d)} F(p, q) , \tag{22}$$

where $F(p, q)$ is the two-dimensional Fourier transform of $A'\left(x_1, y_1\right)$ given as

$$F(p, q) = \iint A'\left(x_1, y_1\right) e^{-i2\pi(px_1+qy_1)}\, dx_1\, dy_1 .$$

To measure this light distribution, it is the intensity which is the square of the magnitude of the complex amplitude $A(x, y)$, which is measured. This intensity is given as

$$I(x, y) = A(x, y)\, A^*(x, y) = \frac{1}{\left[\lambda(f + d)\right]^2} F(p, q)\, F^*(p, q) , \tag{23}$$

or

$$I(x, y) = \frac{|F(p, q)|^2}{\left[\lambda(f + d)\right]^2} .$$

Thus the *intensity in the back focal plane* of a lens (within the limits to be determined for the approximation made) is given by the square of the magnitude of the Fourier transform of the *light amplitude in the plane F.*

If the aperture restrictions in the back focal plane limit the maximum values of x and y so that the phase variations caused by the exponential term in front of the integral in Equation 22 can be considered constant, then

$$A(x, y) = KF(p, q) , \tag{24}$$

where K is a complex constant given by

$$K = \frac{-i\, e^{ikR(x,y)}}{\lambda(f + d)} .$$

Thus, within the range of x and y for which $e^{ikR(x,y)}$ can be assumed constant (i.e. negligible phase variation), the amplitude distribution in the back focal plane is proportional to the Fourier transform of the light amplitude distribution in the plane F. This relationship requires tighter restrictions on x and y than our previous approximation. In terms of spectrum analysis in the back focal plane, this relation is not important since only intensity can be measured. However, in cascaded lens systems, the Fourier transform relation between amplitudes allows each pair of lenses to be accounted for by a double Fourier transform operation. The advantages of such a relation will be demonstrated in a later section.

FOURIER COMPONENTS (p, q) AND FOCAL PLANE COORDINATES (x, y)

In the previous discussions, the amplitude distribution $A(x, y)$ has been expressed in terms of the Fourier transform of $A'(x_1, y_1)$ (refer to Equation 22). However, the Fourier transform coordinates are p and q and are defined by Equations 18 as

$$p = \frac{x}{\lambda\left(f^2 + x^2 + y^2\right)^{1/2}} , \tag{18a}$$

$$q = \frac{y}{\lambda\left(f^2 + x^2 + y^2\right)^{1/2}} . \tag{18b}$$

Substituting for p and q in Equation 22, an expression for $A(x, y)$ can be obtained in terms of x and y; however, this result is somewhat complicated by the fact that p and q are each dependent on both x and y. When the above expressions are substituted for p and q, the optical Fourier transform is given by

$$F(p, q) = F\left(\frac{x}{\lambda\left(f^2 + x^2 + y^2\right)^{1/2}} , \frac{y}{\lambda\left(f^2 + x^2 + y^2\right)^{1/2}}\right) ,$$

$$= \iint A'\left(x_1, y_1\right) e^{-i2\pi\left[xx_1/\lambda\left(f^2+x^2+y^2\right)^{1/2} + yy_1/\lambda\left(f^2+x^2+y^2\right)^{1/2}\right]} dx_1\, dy_1 . \tag{25}$$

A more desirable relationship would exist if p were directly proportional to x and if q were directly proportional to y. Then the light amplitude at a particular value of x would be related to a particular value of p and a similar relation would exist between q and y. As given by Equation 18, p and x (also q and y) are not so simply related since p also depends on y (q also depends on x). In a later discussion the importance of a linear relation between the transform coordinates, p, q, and the spatial coordinates, x, y, will be shown.

To obtain a linear relation between p and x, consider the series expansion of Equation 18a.

$$p = \frac{x}{\lambda f}\left[1 - \frac{x^2 + y^2}{2f^2} + \frac{3}{8}\left(\frac{x^2 + y^2}{f^2}\right)^2 \cdots\right].$$

Restricting our analysis to an area of the focal plane such that

$$\frac{x^2 + y^2}{2f^2} \ll 1 ,$$

then all but the first term in the brackets can be neglected to obtain the approximation

$$p = \frac{x}{\lambda f} . \tag{26}$$

Similarly, an approximation of q can be obtained

$$q = \frac{y}{\lambda f} . \tag{27}$$

Thus, within a restricted area of the back focal plane, the Fourier transform expression can be written as

$$F(p, q) = F\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right) = \iint A'\left(x_1, y_1\right) e^{-i2\pi(xx_1/\lambda f + yy_1/\lambda f)}\, dx_1\, dy_1 . \tag{28}$$

The coordinates x and y in the back focal plane are then scaled representations of the frequencies p and q respectively. That is, light contributions corresponding to a spatial frequency p in the x_1 direction appear at the coordinate $x = \lambda f p$ in the back focal plane. (Similarly, contributions of spatial frequency q in the y_1 direction appear at $y = \lambda f q$.)

The actual restriction to be imposed on x and y for the above approximation will depend on how accurate a Fourier frequency value is required in a particular application. The error in the approximate frequency of Equations 26 and 27 as a fraction of the exact value given by Equation 18 is

$$E_f = \frac{\left(\frac{x}{\lambda f}\right)}{\left(\frac{x}{\lambda\left(f^2 + x^2 + y^2\right)^{1/2}}\right)} - 1 = \left(1 + \frac{x^2 + y^2}{f^2}\right)^{1/2} - 1 . \tag{29}$$

Letting $r^2 = x^2 + y^2$ (r is the radius of a circle in the x, y plane), then r can be expressed in terms of multiples of the focal length f such that

$$r = af \,. \tag{30}$$

Substituting $r^2 = x^2 + y^2 = a^2 f^2$ into Equation 29,

$$E_f = \left(1 + a^2\right)^{1/2} - 1 \,. \tag{31}$$

The curve of Figure 8 gives the percent error ($100\ E_f$) of the linear approximation of frequency as a function of a. For a less than 0.14, the error will be less than 1 percent. Thus the linear approximations

$$p = \frac{x}{\lambda f} \,, \tag{26}$$

and

$$q = \frac{y}{\lambda f} \,, \tag{27}$$

are accurate to within 1 percent for values of x and y satisfying the restriction

$$\left(x^2 + y^2\right)^{1/2} = r \leqq 0.14f \,.$$

For accuracies better than 1 percent, smaller values of a must be imposed as given by Equation 31 and Figure 8.

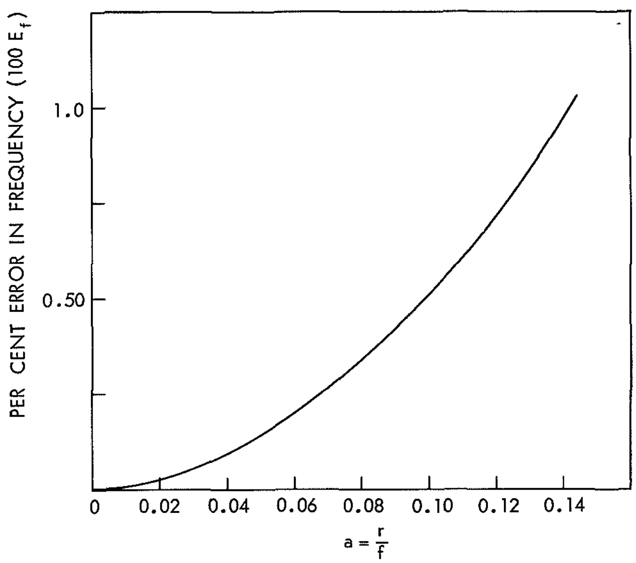

Figure 8—Percent error in linear frequency approximation.

Since the approximation requires limiting consideration to the area within a circle of radius $r_{max} = 0.14f$ (for accuracy within 1 percent), the maximum value of $x^2 + y^2$ is specified by

$$\left(x^2 + y^2\right)_{max} = r_{max}^2 = 0.02f^2 \,.$$

Squaring the approximate expressions for p and q (Equations 26 and 27) and adding,

$$p^2 + q^2 = \frac{x^2 + y^2}{f^2 \lambda^2} \,. \tag{32}$$

Applying the restrictions on $x^2 + y^2$ to Equation 32,

$$p^2 + q^2 \leq \frac{0.02}{\lambda^2}.$$

The maximum allowed value of p occurs when $q = 0$, and the maximum q occurs when $p = 0$

$$p_{max} = \frac{0.14}{\lambda} \quad \text{for} \quad q = 0 \,(\text{i.e., } y = 0)\,,$$

$$q_{max} = \frac{0.14}{\lambda} \quad \text{for} \quad p = 0 \,(\text{i.e., } x = 0)\,.$$

As an example, consider green light of wavelength 5461×10^{-8} cm. In this case the simplified expressions for p and q are accurate to within 1 percent for frequencies in the range

$$\left(p^2 + q^2\right)^{1/2} \leq \frac{0.14}{\lambda} = \frac{0.14 \times 10^8}{5461} \approx 2560 \text{ cycles/cm} \,.$$

On the x axis ($y = 0$, $q = 0$), the maximum frequency will be

$$p_{max} = 2560 \text{ cycles/cm} \,.$$

On the y axis ($x = 0$, $p = 0$), the maximum frequency is likewise

$$q_{max} = 2560 \text{ cycle/cm} \,.$$

In practice, the limitation of available techniques for controlling the input light distribution, $A'\left(x_1, y_1\right)$, restricts maximum spatial frequencies to values below the 2560 cycles/centimeter restriction imposed above. Therefore, the linear approximations of the frequency components p and q (Equations 26 and 27) are applicable for practical systems and Equations 22 and 23 can be expressed as

$$A(x, y) = KF\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right), \tag{32}$$

$$I(x, y) = \frac{\left|F\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right)\right|^2}{\left[\lambda(f + d)\right]^2}. \tag{33}$$

The accuracy of p and q used in Equations 32 and 33, and as defined by Equation 18, is determined by Equation 31, as shown in Figure 8 for values of $a = r/f$. Thus for a given focal length, f, the restriction on the maximum value of r determines the accuracy of the linear approximation introduced here. Of course, these equations also include the approximation assumed earlier in neglecting terms other than the $F(p, q)$ term. The next section considers that approximation and whether the restriction $x^2 + y^2 \leqq 0.02 f^2$ is a sufficient restriction to assure the validity of neglecting terms other than $F(p, q)$.

APPROXIMATION LIMITS FOR THE FOURIER TRANSFORM REPRESENTATION

Returning to the focussed diffraction formula given by Equation 20,

$$A(x, y) = \frac{-i e^{ikR(x,y)}}{\lambda(f+d)} \iint \left[1 - \frac{1}{1 + \dfrac{f(f+d)}{x(x-x_1) + y(y-y_1)}}\right] A'(x_1, y_1)\, e^{-i2\pi(px_1+qy_1)}\, dx_1\, dy_1\,, \tag{20}$$

consider the limitations required to obtain the Fourier transform approximation given by Equation 22. To obtain the form of a Fourier transform of $A'(x_1, y_1)$, the bracketed term must be approximated by a constant. This term can be assumed equal to one, if the range of x, y, x_1, and y_1 is restricted to satisfy the inequality

$$\frac{1}{1 + \dfrac{f(f+d)}{x^2 + y^2 - (xx_1 + yy_1)}} \ll 1\,.$$

Referring back to Equation 21, this approximation corresponds to making the second integral negligible compared to the first integral, which has the form of a Fourier transform. The complete term inside the brackets of Equation 20 is effectively a weighting factor which varies the contribution from each point (x_1, y_1) to the point (x, y). This factor represents the effect of the obliquity factor and path length attenuation. It is usually assumed that these effects are negligible and that the inequality condition is satisfied. In the following analysis, a more detailed quantitative discussion of this approximation is presented.

Neglecting the variable term, when it satisfies the inequality condition given above, is an approximation of the light amplitude contribution from each point (x_1, y_1) to a point (x, y). That is, the contribution $dA(x, y)$ at a point (x, y) from an infinitesimal region $dx_1\, dy_1$ about the point (x_1, y_1) is given exactly by

$$dA(x, y) = K\left[1 - \frac{1}{1 + \dfrac{f(f+d)}{x(x-x_1) + y(y-y_1)}}\right] A'(x_1, y_1)\, e^{-i2\pi(px_1+qy_1)}\, dx_1\, dy_1\,;$$

and applying the approximation of neglecting the variable term inside the bracket, then

$$dA(x, y) = K A'(x_1, y_1)\, e^{-i2\pi(px_1+qy_1)}\, dx_1\, dy_1\,. \tag{34}$$

The total light amplitude $A(x, y)$ at a point (x, y) is obtained by integrating over the range of x_1 and y_1. The integration of Equation 34 will yield an approximation for the total light amplitude $A(x, y)$ at least as accurate as the worse case of Equation 34. That is, the greatest possible error would be given by the maximum value of the neglected term.

Thus, to determine the limitations to be imposed, consider the maximum error introduced by neglecting the variable term to obtain Equation 34. Denoting the error by the fraction, E_A, given by the ratio of the neglected term to the exact factor within the brackets of Equation 20, then

$$E_A = \frac{x^2 + y^2 - (xx_1 + yy_1)}{f(f+d)} . \tag{35}$$

To simplify the discussion, x, y, x_1, and y_1 are expressed in terms of polar coordinates r, ϕ, r_1, and ϕ_1. The relations between these coordinates are given by

$$\begin{aligned} r^2 &= x^2 + y^2 , & r_1^2 &= x_1^2 + y_1^2 , \\ x &= r\cos\phi , & x_1 &= r_1 \cos\phi_1 , \\ y &= r \sin\phi , & y_1 &= r_1 \sin\phi_1 . \end{aligned}$$

Substituting in Equation 35

$$E_A = \frac{r^2 - rr_1 (\cos\phi\cos\phi_1 + \sin\phi\sin\phi_1)}{f(f+d)} . \tag{36}$$

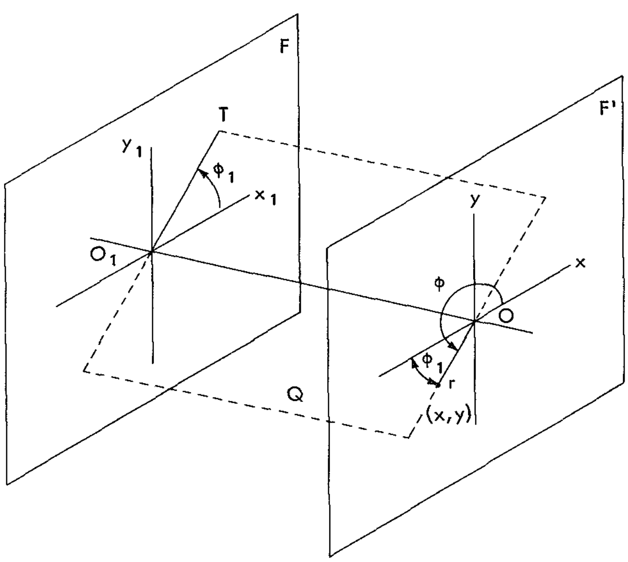

Figure 9—Diagram of conditions for maximum E_A.

Using the identity $\cos(\phi - \phi_1) = \cos\phi\cos\phi_1 + \sin\phi\sin\phi_1$, Equation 36 can be rewritten as

$$E_A = \frac{r^2 - rr_1 \cos(\phi - \phi_1)}{f(f+d)} . \tag{37}$$

The cosine term can take values between plus and minus one. Since the maximum error is of interest, consider the case for $\cos(\phi - \phi_1) = -1$. Equation 37 can be rewritten for cosine equal to minus one as

$$E_A = \frac{r^2 + rr_1}{f(f+d)} . \tag{38}$$

Figure 9 shows the relative positions of points (x_1, y_1) and (x, y) for the case $\cos(\phi - \phi_1) = -1$.

As shown by the figure, the maximum error defined by Equation 38 applies to the light contributions from points (x_1, y_1) located on the line of intersection $O_1 T$ between the planes F and Q. The plane Q is a plane containing the optical axis and the point (x, y) in the back focal plane F'. The points (x_1, y_1) are further restricted to the portion of the line of intersection of F and Q that is on the side of the optical axis opposite from the point (x, y). From the geometry of the figure it is clear that the angle ϕ is equal to $\pi + \phi_1$. Thus $\phi - \phi_1$ is equal to π and cosine $(\phi - \phi_1) = \cos \pi = -1$, as required for the maximum E_A given by Equation 38. For any point in the F plane which does not fall on the line $O_1 T$, the cosine term will be greater than -1 and the value of E_A will be less than that given by Equation 38.

Examination of Equation 38 shows that the error E_A increases as r and r_1 increase. Therefore, to determine the maximum value of E_A as a function of r and r_1, only the maximum values of r and r_1 need be specified. Conversely, if E_A is to be restricted to a value of less than or equal to a specified value, the maximum values of r and r_1 must satisfy Equation 38 for that particular value of E_A.

In order to analyze the relation between maximum E_A, r, and r_1 we must consider the interdependence of the maximum values of r and r_1 caused by the limitations of a finite lens aperture. In the discussion, it is assumed that diffraction effects at the end of the lens aperture are negligible. Figure 10 shows the extreme rays which can pass through a lens aperture of radius R_L to reach the points at the distance r_{max} from the optical axis. It should be apparent that any ray parallel to but above the upper extreme ray, or parallel to but below the lower extreme ray, will be outside the lens aperture and will not pass through the lens. Thus any signal point outside the ray defined by r_{1max} in Figure 10 cannot contribute to both of the points $+r_{max}$ and $-r_{max}$. For example, if the signal area were extended upward beyond the r_{1max} limit, the additional signal interval cannot contribute to the spectral point at $+r_{max}$ since the necessary light path will fall outside of the lens aperture. Under these circumstances the amplitude at the spectral point at $+r_{max}$ will not correspond to the entire signal but only to the interval below the $+r_{1max}$ limit. From this example it is apparent that the r_{1max} limit given by Figure 10 defines the maximum signal interval over which every point contributes to the spectral points at $\pm r_{max}$:

The dashed lines in Figure 10 represent the extreme rays to a spectral point at a distance r that is less that r_{max}. It is seen that the extreme rays for such a case define a maximum signal interval longer than that obtained for r_{max}. This means that the signal interval defined by r_{1max} increases as the spectral range of interest, defined by r_{max}, decreases. Thus it is seen that for a given maximum frequency (i.e., r_{max}), the maximum value of the signal interval r_{1max} is limited by the lens aperture.

To derive an expression defining the relation between r_{max} and r_{1max}, the geometry of Figure 11 is used. This figure represents the upper extreme ray and the principal ray contributing to the spectral point at $+r_{max}$. Since the extreme ray must be parallel to the principal ray, the angles θ are equal and because of a similar-triangles relationship

$$\frac{r_{max}}{f} = \frac{r_{1max}}{a} = \frac{R_L}{a + d} .$$

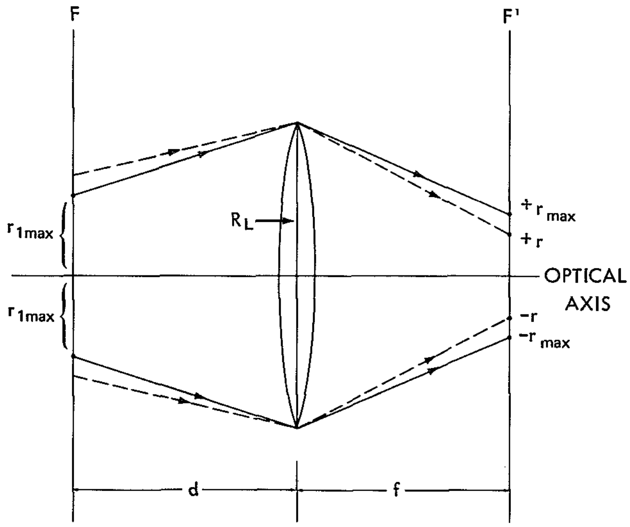

Figure 10—Lens aperture limitations on r_{max} and r_{1max}.

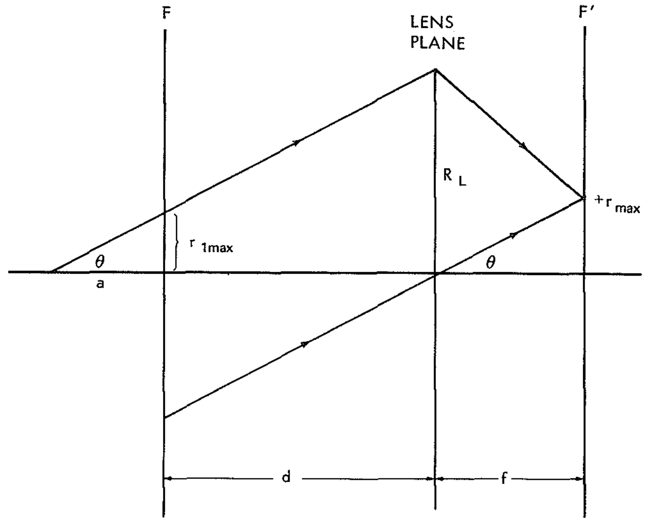

Figure 11—Geometry for relation between r_{max} and r_{1max}.

From these relations, two equations for a can be obtained:

$$a = \frac{r_{1max}}{\left(\frac{r_{max}}{f}\right)}, \quad \text{and} \quad a = \frac{R_L}{\left(\frac{r_{max}}{f}\right)} - d .$$

Since the right hand sides of these equations must be equal, then

$$r_{1max} = R_L - d\,\frac{r_{max}}{f} . \tag{39}$$

A lens is usually specified by its F stop which is defined as

$$F = \frac{f}{2R_L} . \tag{40}$$

Dividing both sides of Equation 39 by f,

$$\frac{r_{1max}}{f} = \frac{R_L}{f} - \frac{d}{f}\,\frac{r_{max}}{f} . \tag{41}$$

From Equation 40 it is found that $R_L/f = 1/2F$; substituting into Equation 41

$$\frac{r_{1max}}{f} = \frac{1}{2F} - \frac{d}{f}\,\frac{r_{max}}{f} . \tag{42}$$

It is obvious from Equation 42, as well as from Figure 10, that r_{1max} cannot be greater than the lens aperture radius R_L. Equation 42 defines the maximum allowed signal aperture radius r_{1max} because of the limitations of the lens aperture. In practice, the size of the signal aperture is specified by either physical consideration or a desired size format. The terms in Equation 42 can be rearranged to define the maximum spectral term r_{max} as

$$\frac{r_{max}}{f} = \frac{\frac{1}{2F} - \frac{r_{1max}}{f}}{\frac{d}{f}} . \tag{43}$$

Equation 43 defines the maximum allowable r for a given r_{1max} as determined by the restriction of a lens aperture. By rearranging terms in Equation 38, a second expression can be obtained specifying the limitations on r_{max} required for an allowed error E_A:

$$\frac{r_{max}}{f} = \frac{1}{2} \frac{r_{1max}}{f} \left[\left(1 + \frac{4E_A \left(1 + \frac{d}{f}\right)}{\left(\frac{r_1}{f}\right)^2} \right)^{1/2} - 1 \right] . \tag{44}$$

To demonstrate the application of Equations 43 and 44, consider the case for $r_{1max}/f = 1/5$ (e.g., for $f = 100$mm, $r_1 = 20$mm). Figure 12 is a plot of r_{max}/f as a function of d/f for the specified input aperture, $r_{1max}/f = 1/5$. The curves labeled $F = 1.4$ and $F = 2$ correspond to Equation 43 for the specified values of F. The curves labeled $E_A = 0.02$, $E_A = 0.01$, and $E_A = 0.005$ correspond to Equation 44 for the specified values of E_A. The F curves specify the upper limit on r/f due to the lens aperture, and the E_A curves specify the upper limit for a given accuracy of the approximation. For a chosen value of d/f, the value of r_{max}/f must be below both the F and E_A curves which apply to the particular system being considered.

For example, consider the case in which a lens with $F = 2$ is to be used and the maximum error to be allowed is $E_A = 0.02$. The greatest value allowed for r/f corresponds to the point A which is the intersection of the $F = 2$ curve and the $E_A = 0.02$ curve. The value of $d/f = 0.5$ would be selected to obtain the value $r_{max}/f = 0.1$. For any other value of d/f, the limit on r_{max}/f would be less than the maximum

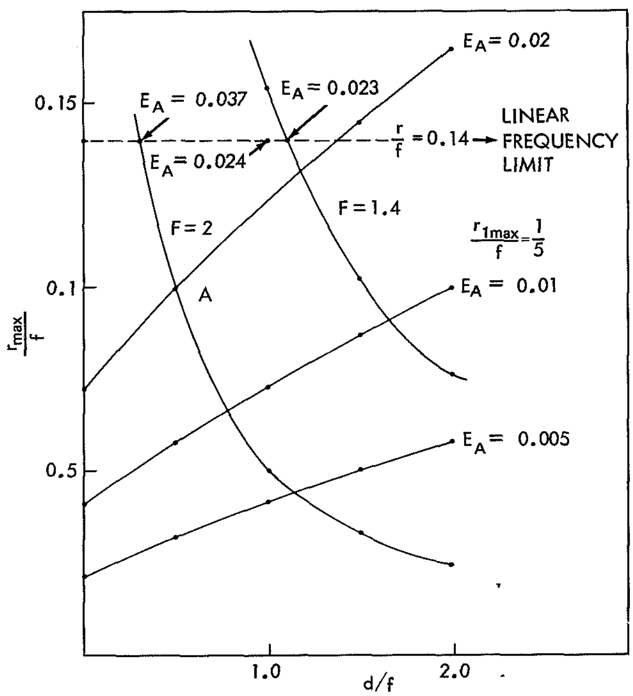

Figure 12—Limitation on maximum spectral term r_{max}/f.

at point A. For d/f less than 0.5, the $E_A = 0.02$ curve specifies a tighter limit on r_{max}/f, while for d/f greater than 0.5 the $F = 2$ curve limits r_{max}/f. Of course any combination of r_{max}/f and d/f corresponding to a point below the curves is allowed; the curves only define the upper limit on r_{max}/f for a given value of d/f.

The specification of a desired value of the maximum error E_A is not readily determined in practice. Since the error in the contribution from (x_1, y_1) varies from point to point, the total effect of the error cannot be determined unless the integration of the exact expression given by Equation 20 can be evaluated. To circumvent this difficulty, a more or less logical selection of parameters is considered, and the maximum E_A specified by these parameters is determined. The value of E_A determined will then specify the maximum error in the approximation be considered.

In Figure 12, the dashed line corresponding to $r_{max}/f = 0.14$ represents the limit determined in the last section for an accuracy of better than 1 percent in the linear relation between spectral frequency and back focal-plane coordinates. From the previous discussion of Equation 38 it is noted that the error E_A decreases as d/f increases. Therefore, if it is not desired to lower the previous limit of $r_{max}/f = 0.14$, the error, E_A, can be improved (decreased) by using the largest possible value of d/f. Referring to Figure 12, note that for a lens with $F = 2$, the maximum value of d/f is 0.3, which is given by the intersection of the dashed line $r/f = 0.14$ and the $F = 2$ curve. As noted in Figure 12, the value of E_A at this point is 0.037 (from Equation 38). If a lens with $F = 1.4$ was considered, Figure 12 shows that d/f can be increased to a value of 1.1 and the error to $E_A = 0.023$ can be reduced. Thus the usual result is obtained that the lens of lower F provides the better characteristics (lower F implies larger lens aperture for given focal length). In addition to having the larger error E_A, the $F = 2$ lens restriction of $d/f = 0.3$ presents practical problems—the lens mount and input aperture mount must be designed to allow for a small spacing ($d = 3$ cm for $f = 10$ cm).

In practice lower values of r_{1max}/f and r/f may be satisfactory. In such cases, the error E_A would be less than the 0.023 determined here, and higher-F lenses (smaller lenses) may be used. An extreme case has been considered here and what amounts to an extreme error, $E_A = 0.023$ (or 2.3 percent), has been determined. This extreme value of the error, introduced by the Fourier transform approximation, is quite reasonable and should be sufficient justification for using the transform approximation in most applications.

The following sections consider the phase of the light distribution in the back focal plane F'. It will be shown that selecting $d/f = 1$ has advantages in reducing the phase factors not associated with the Fourier transform. In Figure 12 the point corresponding to $d/f = 1$ and $r_{max}/f = 0.14$ is shown to have an error value $E_A = 0.024$. Thus it is seen that for the extreme case considered in the discussion above, reducing the value of d/f by one tenth increases the error by 0.001. Such a slight increase in the error E_A is quite reasonable in terms of the advantage gained in the phase approximation treated in the next section. In addition, since the location of a plane at a value of d/f can never be completely accurate, the displacement from the $F = 1.4$ curve allows a safety margin of +10 percent allowable error in the location specified by $d/f = 1$, without exceeding the limitations imposed by the lens aperture.

Thus it has been shown how Equations 38, 43, and 44 can be used to determine and/or specify the parameter limits and the accuracy of the Fourier transform representation:

$$A(x, y) = \frac{-ie^{ikR(x,y)}}{\lambda(f+d)} \iint A'(x_1, y_1)\, e^{-i2\pi(px_1+qy_1)}\, dx_1\, dy_1 \; . \tag{45}$$

It has been shown in particular that for the maximum values $r_{max}/f = 0.14$ and $r_{1max}/f = 1/5$, and the desirable choice of $d/f = 1$, the worse possible error in the amplitude values given by Equation 38 is 2.4 percent. Since the term neglected in Equation 20 is negative, the approximate amplitude given by Equation 45 will be higher than the exact values by no more than 2.4 percent.

FOURIER TRANSFORM REPRESENTATION OF OPTICAL IMAGING

By limiting the area of consideration in the input plane,

$$\left(\frac{r_{1max}}{f} < \frac{1}{5}\right) ,$$

and in the back focal plane,

$$\left(\frac{r_{max}}{f} < 0.14\right) ,$$

it has been shown that the light amplitude distribution, $A(x, y)$, in the back focal plane of a lens is given with reasonable accuracy by

$$A(x, y) = \frac{-ie^{ikR(x,y)}}{\lambda(f+d)} F(p, q) \; , \tag{46}$$

where

$$p = \frac{x}{\lambda f} , \qquad q = \frac{y}{\lambda f} , \tag{47}$$

$$R(x, y) = \frac{f^2 + df + x^2 + y^2}{\left(f^2 + x^2 + y^2\right)^{1/2}} = \frac{f^2 + df + r^2}{\left(f^2 + r^2\right)^{1/2}} , \tag{48}$$

and

$$F(p, q) = \iint A'(x_1, y_1)\, e^{-i2\pi(px_1+qy_1)}\, dx_1\, dy_1 \; . \tag{49}$$

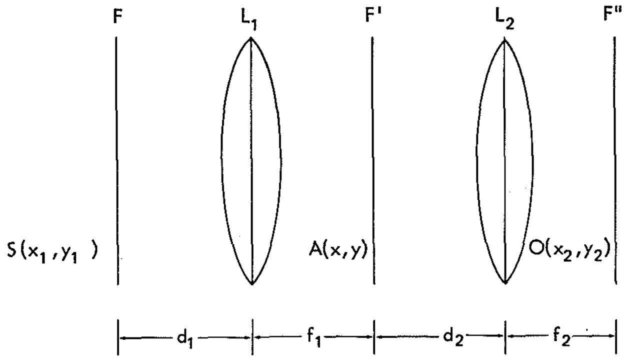

Figure 13—Two lens optical imaging system.

As given by Equation 49, $F(p, q)$ is the two-dimensional Fourier transform of the light amplitude distribution $A'(x_1, y_1)$ in a plane perpendicular to the optical axis and at a distance d in front of the lens.

As pointed out in the discussion of Equations 22, 23 and 24, the phase term $e^{ikR(x,y)}$ is of concern only when a second lens is introduced to produce an image as shown in Figure 13.*

To simplify our notation, K is introduced and defined as

$$K = \frac{-i e^{ikR(x,y)}}{\lambda(f + d)}, \tag{50}$$

and Equation 46 is written

$$A(x, y) = K F(p, q). \tag{51}$$

In Figure 13, the light amplitude distribution $A'(x_1, y_1)$ in the input plane F is given as $S(x_1, y_1)$. Using Equations 49, 50, and 51, the light amplitude distribution $A(x, y)$ in the plane F' (back focal plane of lens L_1) is given as

$$A(x, y) = K_1 \iint S(x_1, y_1)\, e^{-i2\pi(px_1+qy_1)}\, dx_1\, dy_1, \tag{52}$$

where

$$p = \frac{x}{\lambda f_1}, \qquad q = \frac{y}{\lambda f_1}, \tag{53}$$

$$K_1 = \frac{-i e^{ikR_1(x,y)}}{\lambda(f_1 + d_1)}, \tag{54}$$

and

$$R_1(x, y) = \frac{f_1^2 + d_1 f_1 + r^2}{\left(f_1^2 + r^2\right)^{1/2}}. \tag{55}$$

*Only conventional optical systems are considered here. If holographic techniques are considered, the phase factor in Equation 46 would determine the form of the interference pattern produced by A(x, y) and a reference signal.

Similarly, $A(x, y)$ is the input signal to the lens L_2; the light distribution, $O(x_2, y_2)$, in the output plane F'' (back focal plane L_2) can be written as

$$O(x_2, y_2) = K_2 \iint A(x, y)\, e^{-i2\pi(p'x+q'y)}\, dx\, dy \;, \tag{56}$$

where

$$p' = \frac{x_2}{\lambda f_2} \;, \qquad q' = \frac{y_2}{\lambda f_2} \;, \tag{57}$$

$$K_2 = \frac{-i e^{ikR_2(x_2, y_2)}}{\lambda\left(f_2 + d_2\right)} \;, \tag{58}$$

and

$$R_2(x_2, y_2) = \frac{f_2^2 + f_2 d_2 + r_2^2}{\left(f_2^2 + r_2^2\right)^{1/2}} \,. \tag{59}$$

Substituting Equation 52 into Equation 56, an expression can be obtained for the output image $O(x_2, y_2)$ in terms of the input image $S(x_1, y_1)$:

$$O(x_2, y_2) = K_2 \iint e^{-i2\pi(p'x+q'y)}\, dx\, dy \left\{ K_1 \iint S(x_1, y_1)\, e^{-i2\pi(px_1+qy_1)}\, dx_1\, dy_1 \right\} . \tag{60}$$

Assuming that the function $S(x_1, y_1)$ allows the order of integration to be reversed, then Equation 60 can be rewritten as

$$O(x_2, y_2) = K_2 \iint S(x_1, y_1)\, dx_1\, dy_1 \left\{ \iint K_1\, e^{-i2\pi(px_1+qy_1+p'x+q'y)}\, dx\, dy \right\} . \tag{61}$$

The integral within the brackets is complicated by the presence of the factor K_1, which contains an exponential dependent upon x and y. Limiting the values of x and y (i.e., r/f) so that the phase variation in K_1 can be considered negligible, then the K_1 factor can be taken outside the integrals so that

$$O(x_2, y_2) = K_2 K_1 \iint S(x_1, y_1)\, dx_1\, dy_1 \left\{ \iint e^{-i2\pi(px_1+qy_1+p'x+q'y)}\, dx\, dy \right\} . \tag{62}$$

Now consider the integral within the brackets and substitute for p, q, p' and q' the values from Equations 53 and 57 so that

$$\iint e^{-i2\pi(px_1+qy_1+p'x+q'y)}\,dx\,dy = \int e^{-i2\pi(x_1/\lambda f_1+x_2/\lambda f_2)x}\,dx \int e^{-i2\pi(y_1/\lambda f_1+y_2/\lambda f_2)y}\,dx\,. \tag{63}$$

Up to this point limits of integration have not been mentioned. Because of the presence of an aperture in an optical system, the signals exist only over a finite range of the aperture coordinates. However, since the signals are zero outside this range (e.g., $S(x_1, y_1) = 0$ outside the aperture in the F plane), the contribution to the integral beyond the aperture limits will also be zero. Thus the limits of integration are from $-\infty$ to $+\infty$. These limits are in agreement with the Fourier transform integrals.

Now, the Dirac delta function can be defined by the integral equation

$$\delta(x-a) = \int_{-\infty}^{\infty} e^{-i2\pi\mu(x-a)}\,d\mu\,. \tag{64}$$

Comparing each of the integrals on the right side of Equation 63 with the integral in Equation 64, then

$$\begin{aligned}\iint e^{-i2\pi(px_1+qy_1+p'x+q'y)}\,dx\,dy &= \delta\left(\frac{x_1}{\lambda f_1}+\frac{x_2}{\lambda f_2}\right)\delta\left(\frac{y_1}{\lambda f_1}+\frac{y_2}{\lambda f_2}\right), \\ &= \lambda^2 f_1^2\,\delta\left(x_1+\frac{f_1}{f_2}x_2\right)\delta\left(y_1+\frac{f_1}{f_2}y_2\right).\end{aligned} \tag{65}$$

The second step of Equation 65, uses the identity

$$\delta(ax) = \frac{1}{|a|}\,\delta(x)\,.$$

Equation 65 is valid only if x and y range from $-\infty$ to $+\infty$. In optical systems this is not the case since the range of the coordinates x and y is limited as demonstrated in the previous discussion. However, Equation 65 is assumed valid to simplify the discussion. Substituting Equation 65 into Equation 62 gives

$$O(x_2, y_2) = K_2K_1\lambda^2 f_1^2 \iint S(x_1, y_1)\,\delta\left(x_1+\frac{f_1}{f_2}x_2\right)\delta\left(y_1+\frac{f_1}{f_2}y_2\right)dx_1\,dy_1\,. \tag{66}$$

Now, the sifting property of the Dirac delta function which is defined by

$$\int F(x)\,\delta(x+a)\,dx = F(-a)\ ,$$

can be made use of. Applying this property of the delta function to Equation 66, it becomes

$$O\left(x_2, y_2\right) = K_2 K_1 \lambda^2 f_1^{\,2}\, S\left(-\frac{f_1}{f_2}\, x_2, -\frac{f_1}{f_2}\, y_2\right) . \tag{67}$$

In deriving Equation 67 it was assumed that K_1 was approximately constant. The factor K_2 is variable only in phase, as seen by referring to Equation 58. Since only intensity is seen or measured, the phase variations of K_2 can be ignored, and Equation 67 can be interpreted as giving the output image, $O\left(x_2, y_2\right)$, in terms of a proportionality factor $\left(K_2 K_1 \lambda^2 f_1^{\,2}\right)$ which multiplies the original input signal, S, expressed in the new coordinates

$$\left(-\frac{f_1}{f_2}\, x_2,\ -\frac{f_1}{f_2}\, y_2\right) .$$

To clarify the significance of these new coordinates, consider the relation between the two sides of

$$S\left(x_1, y_1\right) \rightarrow S\left(-\frac{f_1}{f_2}\, x_2,\ -\frac{f_1}{f_2}\, y_2\right) . \tag{68}$$

Since the two sides of Equation 68 correspond point for point, then the $\left(x_1, y_1\right)$ and $\left(x_2, y_2\right)$ coordinates are related by

$$x_1 = -\frac{f_1}{f_2}\, x_2 \quad \text{and} \quad y_1 = -\frac{f_1}{f_2}\, y_2 . \tag{69}$$

Equation 69 represents the fact that a signal point which was originally at the coordinates x_1 and y_1 will be imaged to the point at

$$x_2 = -\frac{f_2}{f_1}\, x_1 \quad \text{and} \quad y_2 = -\frac{f_2}{f_1}\, y_1 .$$

The magnification in an optical image is defined as the ratio of the imaged coordinate of a point to the original coordinate such that

$$m_x = \frac{x_2}{x_1} = -\frac{f_2}{f_1}\ ,$$

and

$$m_y = \frac{y_2}{y_1} = -\frac{f_2}{f_1}. \tag{70}$$

Equations 70 were written separately although it is apparent that here the magnification is the same in any direction. In some cases it is possible to obtain different magnifications in different directions (e.g., cylindrical lens system). Equations 67 and 70 show that the output image is proportional to the input image with a change in scale. The minus sign, which appears in Equation 67 and 70, represents an inversion of the image.

In many applications, there is no requirement for a magnified image. In such cases, lenses of equal focal length $f_1 = f_2$ could be used, and a magnification of $m = -1$ would be obtained.

For the case $f_1 = f_2 = f$, Equation 67 becomes

$$o(x_2, y_2) = K_2 K_1 \lambda^2 f^2 S(-x_2, -y_2). \tag{71}$$

Thus, for equal focal length lenses, the output image $o(x_2, y_2)$ is proportional to an inverted replica of the input signal.

It is for this case, $f_1 = f_2 = f$, that the optical imaging process can be described as consecutive Fourier transforms. This can be shown by replacing f_2 by f in Equation 57 and substituting for p' and q' in Equation 56 to obtain

$$o(x_2, y_2) = K_2 \iint A(x, y)\, e^{-i2\pi(x_2 x/\lambda f + y_2 y/\lambda f)}\, dx\, dy. \tag{72}$$

By replacing f_1 by f in Equation 53, then

$$p = \frac{x}{\lambda f}, \qquad q = \frac{y}{\lambda f}, \qquad dx = \lambda f\, dp, \qquad dy = \lambda f\, dq. \tag{73}$$

Substituting Equation 73 into Equation 72

$$o(x_2, y_2) = K_2 \lambda^2 f^2 \iint A(x, y)\, e^{-i2\pi(px_2 + qy_2)}\, dp\, dq. \tag{74}$$

Using Equations 51, 52, 71 and 74, the two step process of optical imaging can be expressed as

$$A(x, y) = K_1 F(p, q) = K_1 \iint S(x_1, y_1)\, e^{-i2\pi(px_1+qy_1)}\, dx_1\, dy_1 , \tag{75}$$

and

$$O(x_2, y_2) = K_2 K_1 \lambda^2 f^2 S(-x_2, -y_2) = K_2 K_1 \lambda^2 f^2 \iint F(p, q)\, e^{-i2\pi(px_2+qy_2)}\, dp\, dq , \tag{76}$$

where

$$p = \frac{x}{\lambda f} , \qquad q = \frac{y}{\lambda f}$$

$$K_1 = \frac{-ie^{ikR_1(x,y)}}{\lambda(f + d_1)} , \qquad K_2 = \frac{-ie^{ikR_2(x_2,y_2)}}{\lambda(f + d_2)} ,$$

$$R_1(x, y) = \frac{f^2 + fd_1 + r^2}{(f^2 + r^2)^{1/2}} , \qquad R_2(x_2, y_2) = \frac{f^2 + fd_2 + r_2^2}{(f^2 + r_2^2)^{1/2}} ,$$

$$r^2 = x^2 + y^2 , \qquad \text{and} \qquad r_2^2 = x_2^2 + y_2^2 .$$

The last expression in Equation 76 assumes that the factor K_1 can be considered constant in phase over the range of the values p and q (i.e., x and y). This approximation is the subject about to be considered; shown here are the advantages of the resulting expression. To appreciate the significance of Equations 75 and 76, consider the standard Fourier transform equations in the notation used here:

$$F(p, q) = \iint S(x_1, y_1)\, e^{-i2\pi(px_1+qy_1)}\, dx_1\, dy_1 , \tag{77}$$

$$S(x_1, y_1) = \iint F(p, q)\, e^{+i2\pi(px_1+qy_1)}\, dp\, dq , \tag{78}$$

and

$$S(-x_1, -y_1) = \iint F(p, q)\, e^{-i2\pi(px_1+qy_1)}\, dp\, dq . \tag{79}$$

Equation 77 is usually referred to as the Fourier transform while Equation 78 is the inverse Fourier transform. Note that the exponent of Equation 78 is positive and that of Equation 79 is negative. Since the optical transform produced by a lens has a negative exponent, the inverse transform defined by Equation 78 never appears in optical systems. The second lens in an optical system, such as that of Figure 13, produces a Fourier transform of a Fourier transform as represented by Equation 79. Note that the inversion, or change of sign of the coordinate, is introduced by the second Fourier transform, whereas an inverse transform would not invert the signal. Thus, comparing Equations 75 and 76 with Equations 77 and 79, it is noted that the optical imaging process of two lenses is described by two successive Fourier transforms relations. Except for determining the absolute amplitudes involved, the constants in front of the integrals of Equations 75 and 76 do not affect the form of the variations. In most cases only the relative amplitudes are of interest, and the constants are dropped.

The advantage of the Fourier transform representation described by Equations 75 and 76 can be shown by considering the introduction of a filter in the F' plane. If the transmission characteristics of the filter are known, then a function $M(p, q)$ can be determined which represents the fraction of incident light amplitude passed at each coordinate corresponding to the values of p and q. The filtered output $O_f(x_2, y_2)$ is then given by Equation 76, if $F(p, q)$ is replaced by $M(p, q)\,F(p, q)$

$$O_f(x_2, y_2) = K_2 K_1 \lambda^2 f^2 \iint M(p, q)\,F(p, q)\,e^{-i2\pi(px_2+qy_2)}\,dp\,dq \,. \tag{80}$$

Thus the specification of a filter for a particular application can be determined uniquely when the Fourier transform representation is used.

It is noted at this point that the Fourier transform representation of Equations 75 and 76 require the use of lenses of equal focal length. The specification of a filter for the case of unequal focal lengths is exactly the same; however, the Fourier transform relation of Equation 76 is modified by introducing a factor of f_1/f_2 in the exponent to account for the magnification.

This additional f_1/f_2 results in a magnified filtered image which contains the same information as the filtered image, $O_f(x_2, y_2)$, given by Equation 80. The only difference (neglecting the phase of K_2) is in the scale. Throughout the remaining part of this report, the special case of equal focal length lenses is considered to simplify the analysis.

ELIMINATION OF UNDESIRABLE PHASE VARIATIONS

Having now seen the significance of the Fourier transform representation of optical imaging, it is possible to consider the approximation involved in the derivation of Equation 76. The actual relation corresponding to Equation 76 can be written as

$$O(x_2, y_2) = K_2 \lambda^2 f^2 \iint K_1\,F(p, q)\,e^{-i2\pi(px_2+qy_2)}\,dp\,dq \,. \tag{81}$$

The term K_1 appearing in the integral was defined as

$$K_1 = \frac{-ie^{ikR_1(x,y)}}{\lambda\left(f + d_1\right)} , \tag{82}$$

where

$$R_1 (x, y) = \frac{f^2 + fd_1 + r^2}{\left(f^2 + r^2\right)^{1/2}} , \qquad \left(r^2 = x^2 + y^2\right) . \tag{83}$$

Comparing Equations 81 and 76, it is apparent that Equation 76 is valid only if the phase variation of K_1 can be neglected over the range of the integration variables p and q (or x and y). This requirement is now analysed.

For reasons that will become evident, $R_1 (x, y)$ is expressed in the form

$$\begin{aligned} R_1 (x, y) &= \left(f + d_1\right) + P\left(f - d_1\right) + Q f , \\ &= f(1 + P + Q) + d_1 (1 - P) . \end{aligned} \tag{84}$$

Equation 83 can be rewritten as

$$R_1 (x, y) = \frac{f^2 + r^2}{\left(f^2 + r^2\right)^{1/2}} + \frac{d_1 f}{\left(f^2 + r^2\right)^{1/2}} = f\left(1 + \frac{r^2}{f^2}\right)^{1/2} + d_1 \left(1 + \frac{r^2}{f^2}\right)^{-1/2} . \tag{85}$$

Equating the coefficients of f and d_1 in the final forms of Equations 84 and 85, then

$$1 - P = \left(1 + \frac{r^2}{f^2}\right)^{-1/2} ,$$

and

$$1 + P + Q = \left(1 + \frac{r^2}{f^2}\right)^{1/2} .$$

Solving for P and Q

$$P = 1 - \left(1 + \frac{r^2}{f^2}\right)^{-1/2} ,$$

$$Q = \left(1 + \frac{r^2}{f^2}\right)^{1/2} + \left(1 + \frac{r^2}{f^2}\right)^{-1/2} - 2 = 2\left[\frac{1 + \dfrac{r^2}{2f^2}}{\left(1 + \dfrac{r^2}{f^2}\right)^{1/2}} - 1\right].$$

Substituting for P and Q, Equation 84 can be written

$$R_1(x, y) = (f + d_1) + (f - d_1)\left[1 - \frac{1}{\left(1 + \dfrac{r^2}{f^2}\right)^{1/2}}\right] + 2f\left[\frac{1 + \dfrac{r^2}{2f^2}}{\left(1 + \dfrac{r^2}{f^2}\right)^{1/2}} - 1\right]. \tag{86}$$

The K_1 term given by Equation 82 can be rewritten using Equation 86 so that

$$K_1 = \left[\frac{-ie^{ik(f+d_1)}}{\lambda(f+d_1)}\right] e^{ik(f-d_1)[1-(1+r^2/f^2)^{-1/2}]}\ e^{i2kf[(1+r^2/2f^2)(1+r^2/f^2)^{-1/2}-1]}. \tag{87}$$

The terms grouped within the first brackets of Equation 87 are constant and therefore can be taken out from under the integral sign in Equation 81. The remaining exponentials in Equation 87 are phase factors which depend on the variables of integration. The exponential of the remaining terms must be limited so that the phase variations can be considered negligible.

Since restrictions must be considered on the value of r/f, so that the phase terms can be considered negligible, then the exponentials of Equation 87 must be simplified by expanding in power series of (r^2/f^2) while the first terms of the expansions are all dropped. Expanding the exponent of the first exponential gives

$$e^{ik(f-d_1)[1-(1+r^2/f^2)^{-1/2}]} = e^{ik(f-d_1)\left[1-\{1-(r^2/2f^2)+3/8(r^2/f^2)^2\cdots\}\right]},$$

$$\simeq e^{ik(f-d_1)(r/f)^2/2}.$$

Similarly, by expanding the second exponential

$$e^{i2kf[(1+r^2/2f^2)(1+r^2/f^2)^{-1/2}-1]} = e^{i2kf\left[-1+(1+r^2/2f^2)\{1-(r^2/2f^2)+3/8(r^2/f^2)^2\cdots\}\right]},$$

$$\simeq e^{ikf(r/f)^4/4}.$$

Substituting these approximate terms in Equation 87 gives

$$K_1 = \frac{-ie^{ik(f+d_1)}}{\lambda(f+d_1)}\ e^{ik(f-d_1)(r/f)^2/2}\ e^{ikf(r/f)^4/4}. \tag{88}$$

For values of (r/f) less than our previous limit of 0.14, the approximations in each of the phase terms is accurate to within 2 percent of its exact values. It should be noted that neglecting terms in the exponentials is valid only since phase variations less than one cycle are to be considered.

Since the desire is to eliminate the phase variations of K_1, Equation 88 indicates that the optimum choice of the distance d_1 equal to f eliminates the first phase term. Thus for the case when d_1 is chosen equal to f, Equation 88 can be reduced to

$$K_1 = \left[\frac{-i e^{i2kf}}{2\lambda f}\right] e^{ikf(r/f)^4/4} , \qquad \left(\text{for } d_1 = f\right) . \tag{89}$$

If Equation 89 is substituted for K_1 in Equation 81, then

$$O\left(x_2, y_2\right) = \frac{-iK_2 \lambda f e^{i2kf}}{2} \iint e^{ikf(r/f)^4/4} F(p, q) e^{-i2\pi(px_2+qy_2)} dp dq . \tag{90}$$

When the value of (r/f) is limited so that the phase factor appearing under the integral of Equation 90 can be neglected, the output image $O\left(x_2, y_2\right)$ is given by

$$O\left(x_2, y_2\right) = K_1 K_2 \lambda^2 f^2 \iint F(p, q) e^{-i2\pi(px_2+qy_2)} dp dq , \tag{91}$$

where

$$K_1 = \frac{-ie^{i2kf}}{2\lambda f} .$$

Equation 91 is identical to Equation 76 which has been shown to be the desired form for the Fourier transform representation of optical imaging.

To derive a specification for the maximum limit on (r/f) which allows the variable term in K_1 to be neglected, consider the effect of the phase term for a particular $F(p, q)$:

$$F(p, q) = \delta(q) \left[A_0 \delta(p) + \frac{B}{2}\left\{\delta\left(p - p_0\right) + \delta\left(p + p_0\right)\right\}\right] . \tag{92}$$

The locations of the frequency terms contained in Equation 92 are diagrammed in Figure 14. Since, by definition, the delta function $\delta(q)$ is equal to zero for q unequal to zero, Equation 92 represents the spectrum of a signal which varies in only one dimension. That is, there are no frequency components in the y direction; therefore, the signal is constant with respect to the y coordinate. Rather than interpret Equation 92 as the spectrum of a particular signal, it can also be assumed that only

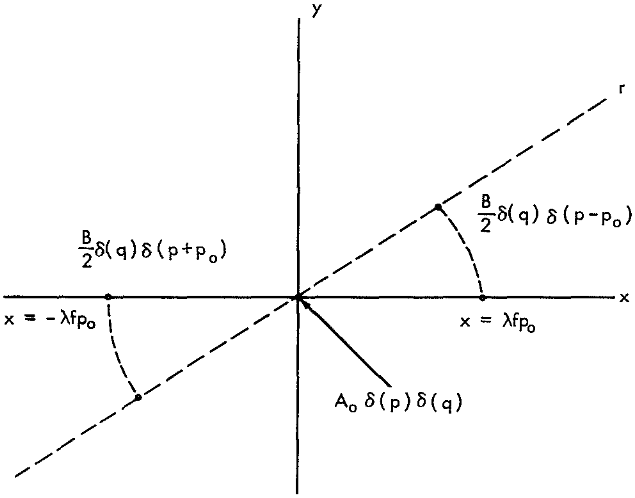

Figure 14—Location of frequency terms in the spectrum plane.

three sample points of a more general spectrum are being considered. Since there is nothing to single out the x direction in an optical system, the analysis will apply to a set of spectral points along any radial axis in the frequency plane as indicated by the r axis in Figure 14. This is obvious if it is considered that arbitrarily selection of any orientation can be made for the x- and y-coordinate axes. It can also be shown that the terms used to specify a maximum limit on (r/f) also apply to the general case. Thus the results are simply interpreted for the special case of Equation 92 as a general criteria for neglecting the undesired phase factor in Equation 90.

Substituting for $F(p, q)$ as given by Equation 92 and applying the interpretation of the discussion above, Equations 90 and 91 can be simplified to

$$\underline{O(x_2)} = K \int e^{ikf(x/f)^4/4} \left[A_0\,\delta(p) + \frac{B}{2}\left\{\delta(p-p_0) + \delta(p+p_0)\right\}\right] e^{-i2\pi p x_2}\,dp\ , \tag{93}$$

$$O(x_2) = K \int \left[A_0\,\delta(p) + \frac{B}{2}\left\{\delta(p-p_0) + \delta(p+p_0)\right\}\right] e^{-i2\pi p x_2}\,dp\ . \tag{94}$$

The new factor K in Equations 93 and 94 is defined as

$$K = \frac{-iK_2\,\lambda\, f\, e^{i2kf}}{2}\ . \tag{95}$$

The integral with respect to q was taken by applying the sifting property of the delta function $\delta(q)$. The y_2 dependence has been dropped on the left side of Equations 93 and 94 since the output image varies only with respect to the x_2 coordinate. The left side of Equation 93 is underlined so that we can identify the ensuing results of Equations 93 and 94 throughout the remainder of our discussion.

The first exponential in Equation 93 can be rewritten using the definition $p = x/\lambda f$, while the integration with respect to p is performed simply by applying the sifting property of the delta function:

$$\underline{O(x_2)} = K\int e^{ikf(\lambda p)^4/4}\left[A_0\,\delta(p) + \frac{B}{2}\left\{\delta(p+p_0) + \delta(p-p_0)\right\}\right] e^{-i2\pi p x_2}\,dp\ ,$$

$$= K\left[A_0 + \frac{B}{2}\,e^{ikf(\lambda p_0)^4/4}\left\{e^{i2\pi p_0 x_2} + e^{-i2\pi p_0 x_2}\right\}\right]\ ,$$

$$\underline{O(x_2)} = K\left[A_0 + B\,e^{ikf(\lambda p_0)^4/4}\cos 2\pi p_0 x_2\right]. \tag{96}$$

The result after integrating Equation 94 is similar to Equation 96 except for the exponential which appears in Equation 96:

$$O(x_2) = K\left[A_0 + B\cos 2\pi p_0 x_2\right]. \tag{97}$$

Equations 96 and 97 represent the amplitude distribution of the output image produced by the frequency-plane distribution $F(p, q)$ given by Equation 92. Equation 96 represent the output $\underline{O(x_2)}$ when the phase factor is considered and Equation 97 represent the output $O(x_2)$ when the phase factor is neglected. Comparing Equation 96 and 97, a criteria for neglecting the phase factor is still not very apparent since the significance of the exponential is not very clear.

If the observation of the output image is considered, then the intensity, rather than the amplitude as given by Equations 96 and 97, must be dealt with. The intensities are given by the relations

$$\underline{I(x_2)} = \underline{O(x_2)}\,\underline{O^*(x_2)} = \left|\underline{O(x_2)}\right|^2\ , \tag{98}$$

and

$$I(x_2) = O(x_2)\,O^*(x_2) = \left|O(x_2)\right|^2. \tag{99}$$

The starred terms in Equation 98 and 99 represent the complex conjugate of the unstarred terms. Using Equations 96 and 97 in Equations 98 and 99 respectively

$$\underline{I(x_2)} = |K|^2\left[A_0^2 + B^2\cos^2 2\pi p_0 x_2 + 2A_0 B\cos\frac{kf}{4}(\lambda p_0)^4\cos 2\pi p_0 x_2\right]\ , \tag{100}$$

and

$$I(x_2) = |K|^2 \left[A_0^2 + B^2 \cos^2 2\pi p_0 x_2 + 2A_0 B \cos 2\pi p_0 x_2\right] . \tag{101}$$

Now comparing Equations 100 and 101, it is found that the phase factor introduces a cosine factor which attenuates the $\cos 2\pi p_0 x_2$ component in the observed image. In the general case, the amplitude B of any one component will be considerably less than the component A_0. Therefore, the third term in Equations 100 and 101 represent the larger of the two x_2-dependent terms in the image intensity. It is desirable to limit the maximum value of p_0 to obtain a value of $\cos (kf/4) (\lambda p_0)^4$ as near to one as possible so that the Fourier-transform representation used in deriving Equation 101 can be considered a good approximation.

The cosine term can be expressed as a function of x by the definition $p = x/\lambda f$, and since the results will apply to the general case, x can be replaced by the more general notation r. Thus the cosine term can be expressed

$$\cos \frac{kf}{4} (\lambda p_0)^4 = \cos \frac{kf}{4} \left(\frac{r}{f}\right)^4 . \tag{102}$$

In Equation 102 the frequency p_0 is given the general interpretation of a spatial frequency in the direction of an r axis (see Figure 14), and p_0 and r are related by $p_0 = r/\lambda f$. That is, Equation 102 applies to the general case of a spectrum along any radial axis r in the back focal plane F'. The term (r/f) can be expressed as a multiple of $(4/kf)^{1/4}$ by defining a factor m by the relation

$$\frac{r}{f} = m\left(\frac{4}{kf}\right)^{1/4} . \tag{103}$$

Substituting for (r/f) in Equation 102,

$$\cos \frac{kf}{4} \left(\frac{r}{f}\right)^4 = \cos m^4 . \tag{104}$$

The ultimate limit on m can be recognized by noting that for $m = (\pi/2)^{1/4}$ the cosine term as given by Equation 104 is zero (i.e., $\cos \pi/2 = 0$). For this value of m the Fourier transform result of Equation 101 is completely in error with respect to the third term, since the cosine term present in Equation 100 is zero and the third term is eliminated. Thus for $m = (\pi/2)^{1/4}$, the ideal two step Fourier transform representation yields a term which does not exist in the actual image given by Equation 100.* For values of m less than $(\pi/2)^{1/4}$, the cosine of Equation 104 has nonzero values

*This accounts for the frequent neglecting of phase terms, only when less than $\pi/2$ which appears in many references dealing with approximate solutions of diffraction problems.

as shown by the curve of Figure 15. Selecting a limit on m is based on specifying how accurately the third term of Equations 100 and 101 should agree. A value of $m = 0$ is necessary to have complete agreement between Equations 100 and 101; however, $m = 0$ corresponds to a frequency $p_0 = 0$ which corresponds to a dc term only. Thus a compromise limit must be established between the limits $m = 0$ and $m = (\pi/2)^{1/4}$.

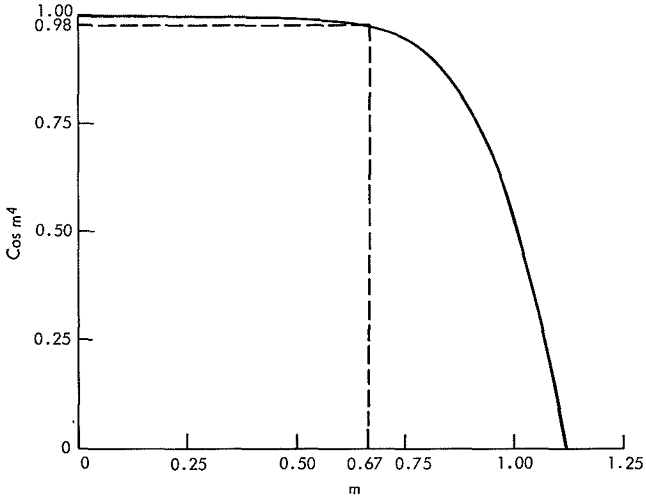

Figure 15—Graph of $\cos m^4$ vs. m.

To determine the limitation on m, it is necessary to specify the desired accuracy of the third term in Equation 101 as compared to the third term in Equation 100. Again, the accuracy of the approximation can be given in terms of a fractional error E_ϕ defined as

$$E_\phi = \frac{1 - \cos m^4}{\cos m^4} = \frac{1}{\cos m^4} - 1 . \tag{105}$$

The percent error in the third term of Equation 101 is then $+100\ E_\phi$ percent compared to the exact term in Equation 100. Note that the error E_ϕ does not represent a fraction of the total image intensity. The error E_ϕ corresponds only to a particular term in the image intensity. In the general case there would be a series of such terms and the maximum E_ϕ would be determined by Equation 106. This E_ϕ would represent the maximum error in terms of the form of the third term in Equation 101 and would correspond to the term involving the highest frequency of interest.

Specifying the maximum allowable error is rather arbitrary and will usually depend on the particular application considered. However, for an example a limit of 2% accuracy can be specified for the approximation, i.e., $E_\phi = 0.02$. Substituting in Equation 105 the condition for the maximum value of m is obtained:

$$\cos m^4 \geq \frac{1}{1.02} = 0.98 . \tag{106}$$

From Figure 15 it is found that Equation 106 requires values of m less than or equal to 0.67. Using the maximum value $m = 0.67$, Equation 103 becomes

$$\left(\frac{r}{f}\right)_{max} = 0.67 \left(\frac{4}{kf}\right)^{1/4} . \tag{107}$$

To obtain a numerical result for comparison with the previous limit, $r/f_{max} = 0.14$, a wavelength $\lambda = 5461 \times 10^{-8}$ cm and focal length $f = 10$ cm is considered. Substituting in Equation 107 it is found that

$$\left(\frac{r}{f}\right)_{max} = 0.67\left(\frac{4\lambda}{2\pi f}\right)^{1/4} = 0.67\left(\frac{4\times 5461\times 10^{-8}}{2\pi\times 10}\right)^{1/4},$$

$$\left(\frac{r}{f}\right)_{max} \simeq 0.03\,. \tag{108}$$

Equation 108 specifies the aperture limit in the frequency plane F' to assure an error limit of less than 2 percent caused by neglecting phase variations. Notice that this phase limit restricts the frequency aperture to approximately one fifth of the previous value of $r/f_{max} = 0.14$, which was sufficient for the linearization and amplitude approximations. The corresponding frequency limit is given as

$$p_{max} = \frac{1}{\lambda}\left(\frac{r}{f}\right)_{max} \simeq \frac{0.03}{5461\times 10^{-8}} = 550 \text{ cycles/cm}\,. \tag{109}$$

Thus for spatial frequencies less than 550 cycles/cm, neglecting the phase term to obtain the Fourier transform representation of Equation 91 introduces an error of no more than 2 percent in terms of the form of the third term in the image intensity of Equation 101. It is again pointed out that for most practical cases, the frequency capability of present input techniques restricts the possible frequencies to a lower value than that specified by Equation 109.

It has been shown how Equations 103 and 105 are used to determine the error E_ϕ for any frequency plane aperture with a radius defined by (r_{max}/f). It has further been shown for a particular case ($\lambda = 5461$ Å and $f = 10$ cm) that the limit, $(r/f)_{max} = 0.03$, provides an accuracy within 2 percent for the terms in which the phase variation appears. It has also been noted that this phase approximation requires a tighter restriction on the maximum frequency terms. In fact, for examples used, the maximum frequency is one fifth of that allowed for an accurate amplitude approximation. Of course, this further restriction of the frequency range of interest will also improve the accuracy of the amplitude approximation.

Figure 12 can be used to consider the amplitude error E_A for the values $d/f = 1$ and $r_{max}/f = 0.03$ assuming $r_{1max}/f = 1/5$. The point corresponding to $d/f = 1$ and $r_{max}/f = 0.03$ is located below the curve corresponding to $E_A = 0.005$. Therefore, the further restriction on r_{max}/f required for the phase approximation reduces the error in the amplitude approximation to a value less than 0.5 percent. This result shows that the restriction considered in this section not only provides a Fourier transform relation, which is accurate in phase, but also improves the accuracy of the amplitude approximations previously considered.

Within the limits presented in this section, the two-lens optical imaging process can be described by the equations

$$A(x, y) = \left[\frac{-ie^{i2kf}}{2\lambda f}\right] F(p, q) = \left[\frac{-ie^{i2kf}}{2\lambda f}\right]\iint S(x_1, y_1)\, e^{-i2\pi(px_1+qy_1)}\, dx_1\, dy_1\,, \tag{110}$$

and

$$O(x_2, y_2) = e^{ikR_2}\left[\frac{-fe^{i2kf}}{2(f+d_2)}\right]\iint F(p, q)\, e^{-i2\pi(px_2+qy_2)}\, dp\, dq\ . \tag{111}$$

In practice only the variations in amplitude are of interest and the constant factors within the brackets are dropped, so that

$$A(x, y) = F(p, q) = \iint S(x_1, y_1)\, e^{-i2\pi(px_1+qy_1)}\, dx_1\, dy_1\ , \tag{112}$$

$$O(x_2, y_2) = e^{ikR_2}\, S(-x_2, -y_2) = e^{ikR_2}\iint F(p, q)\, e^{-i2\pi(px_2+qy_2)}\, dp\, dq\ . \tag{113}$$

Equations 112 and 113 represent the form of the optical Fourier transform representation commonly used. These equations describe the relative amplitude and phase variations of spectrum $A(x, y)$ and image $O(x_2, y_2)$. Note that the phase term e^{ikR_2} is retained in Equation 113. This factor has no effect on the image intensity since multiplication by the complex conjugate eliminates this term. However, if the image $O(x_2, y_2)$ is to be processed further by another lens, the effect of the phase factor e^{ikR_2} must be considered. In such cases, the criterion for neglecting the variation in phase due to the factor e^{ikR_1} must also be reevaluated since the criterion developed above was based on image intensity effects.

OPTICAL CORRELATOR SYSTEMS

Considered now is a three lens optical system as shown in Figure 16. In this system the signal plane F is assumed to be in the front focal plane of lens L_1, and each of the other lenses (L_2 and L_3) is located so that its front focal plane coincides with the back focal plane of the preceding lens. With this configuration the amplitude distribution corresponding to the input signal to each lens is the output signal in the back focal plane of the preceding lens and is a focal length in front of the lens. This location of the signal planes provides the advantage of eliminating the phase terms dependent upon the distance from the lens to the input plane as discussed in relation to Equations 88 and 89. The optical system in Figure 16 consists of a two-lens imaging system as discussed in the preceding section followed by a third lens that produces a Fourier transform of the light amplitude

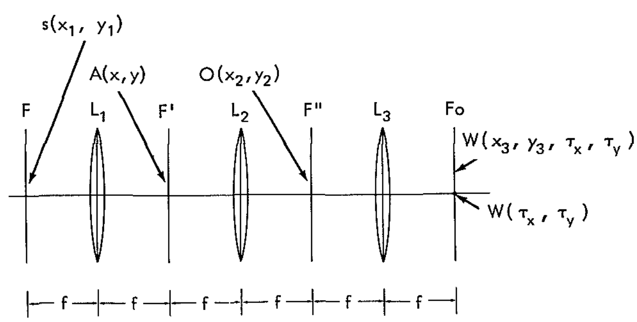

Figure 16—Optical correlator system.

distribution of the image. As pointed out at the end of the last section, the processing of the image $O(x_2, y_2)$ by an additional lens involves its amplitude rather than the intensity; therefore, the phase effects of each lens will be considered. Throughout this section it is assumed that the aperture limitations are sufficiently restrictive so that the linear frequency and amplitude approximations developed earlier are valid. The focal lengths of the three lenses are assumed equal to simplify the analysis; in the general case, unequal focal lengths would introduce magnification or demagnification.

In the optical system of Figure 16, lens L_1 produces a light amplitude distribution in its back focal plane F' which is proportional to the Fourier transform of the input signal, $S(x_1, y_1)$, except for a multiplicative phase factor. As discussed in reference to Equations 112 and 113, all constant factors are dropped and only terms which vary with respect to the coordinates in the four signal planes of interest (F, F', F'', F_0) are retained. Using only the variable exponential in Equation 89 for K_1, the amplitude in the F' plane is given by Equation 75 which can be written

$$A(x, y) = e^{ikf(r/f)^4/4} F(p, q) = e^{ikf(r/f)^4/4} \iint S(x_1, y_1)\, e^{-i2\pi(px_1+qy_1)}\, dx_1\, dy_1 . \tag{114}$$

For our development of a correlator it is advantageous to introduce notation for the signal $S(x_1, y_1)$ which accounts for displacement of the signal from some reference position. Referring to Figure 17, signal point A can be considered at the new position A′. If A is a point of the signal $S(x_1, y_1)$, the light amplitude at A is $S(x_A, y_A)$. Since A′ is the same signal point as A (it has only been moved), the light amplitude at A′ must also be $S(x_A, y_A)$. The coordinates of the point A′ are $x_1 = x_A + \tau_x$ and $y_1 = y_A + \tau_y$. Thus the notation for the signal must be such that if the coordinates x_1, y_1 are substituted for the point A′, then $S(x_A, y_A)$ is obtained. The required notation is $S(x_1 - \tau_x, y_1 - \tau_y)$ as can be seen by substituting the values of x_1 and y_1 for each of the points A and A′. In either case the signal amplitude is $S(x_A, y_A)$. Using this new notation for a signal, Equation 114 can be rewritten as

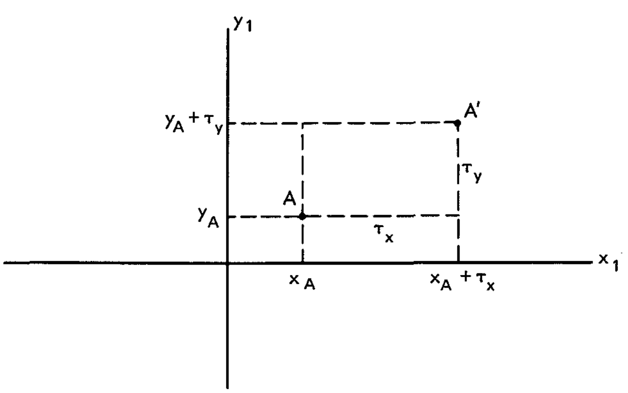

Figure 17—Displacement of a signal point in the input plane F.

$$\begin{aligned} A(x, y) &= e^{ikf(r/f)^4/4} F(p, q) \\ &= e^{ikf(r/f)^4/4} \iint S(x_1 - \tau_x, y_1 - \tau_y)\, e^{-i2\pi(px_1+qy_1)}\, dx_1\, dy_1 . \end{aligned} \tag{115}$$

The displacements τ_x and τ_y are positive when the displacement is in both the positive x_1 and positive y_1 directions.

In Equation 115 the function $F(p, q)$ represents the Fourier transform of the displaced function $S(x_1 - \tau_x, y_1 - \tau_y)$. Thus the $F(p, q)$ in

Equation 115 includes all the information regarding the signal including its displacement. From Fourier transform theory, the transform corresponding to a displaced signal such as in Equation 115 differs from the transform $F(p, q)$ of the undisplaced signal of Equation 114 by an exponential phase term, $e^{-i2\pi(\tau_x p+\tau_y q)}$. This principle need not be of any further concern; it is pointed out only to emphasize that the $F(p, q)$ in Equation 115 corresponds to the displaced signal $S(x_1 - \tau_x, y_1 - \tau_y)$.

The amplitude distribution given by Equation 115 appears in the plane F' and represents the input signal to the lens L_2. Lens L_2 performs a Fourier transform operation on $A(x, y)$, and the image amplitude in the plane F'' is given by

$$O(x_2, y_2) = \phi(x_2, y_2)\iint e^{ikf(r/f)^4/4} F(p, q)\, e^{-i2\pi(px_2+qy_2)}\, dp\, dq\,. \tag{116}$$

Equation 116 corresponds to Equation 81 except that only the variable terms of K_1 and K_2 have been retained. The exponential appearing in the integrand corresponds to the variable term in K_1 as discussed above. The function $\phi(x_2, y_2)$ in front of the integral represents the variable part of K_2. From the definition of K_2 given under Equation 76 the variable part of K_2 is obtained from the term e^{ikR_2} where

$$R_2 = \frac{f^2 + fd_2 + r_2^2}{\left(f^2 + r_2^2\right)^{1/2}}\,, \quad \text{and} \quad r_2^2 = x_2^2 + y_2^2\,. \tag{117}$$

Since R_2 has the same form as R_1, R_2 can be expanded in the form of Equation 86

$$R_2(x_2, y_2) = (f + d_2) + (f - d_2)\left[1 - \frac{1}{\left(1 + \frac{r_2^2}{f^2}\right)^{1/2}}\right] + 2f\left[\frac{1 + \frac{r_2^2}{2f^2}}{\left(1 + \frac{r_2^2}{f^2}\right)^{1/2}} - 1\right]. \tag{118}$$

The first term and the -1 term in the last bracket of Equation 118 are constant and can be dropped since only the variable part is of interest. The second term vanishes since $d_2 = f$ in the system being considered. Thus the only variable term in Equation 118 is the fraction in the brackets of the third term. The function $\phi(x_2, y_2)$ is therefore given by

$$\phi(x_2, y_2) = e^{i2kf\left[(1+r_2^2/2f^2)(1+r_2^2/f^2)^{-1/2}\right]}\,. \tag{119}$$

The variable part of K_2 given by Equation 119 was derived from the complete expansion of R_2 rather than from an approximate expansion analogous to Equations 88 and 89, since the aperture restrictions necessary for the validity of Equations 88 or 89 would require a signal and image aperture much smaller than that normally desired in optical systems. For an image aperture defined by $(r_2/f)_{max} = 0.14$, and $f = 10$ cm, and $\lambda = 5461 \times 10^{-8}$ cm, the phase term $\phi(x_2, y_2)$ can introduce phase shifts as great as 38π radians (19 cycles). It was pointed out that the phase approximation of Equation 89 was accurate with 2 percent. For the image aperture considered here, this phase inaccuracy can be of the order of 0.4 cycles. This magnitude of phase error may not be negligible, and, therefore, the more complete exponential was used in defining $\phi(x_2, y_2)$ by Equation 119.

Returning to the image amplitude distribution $O(x_2, y_2)$ given by Equation 116, the notation will be changed to take into account the possibility of image displacement corresponding to the signal displacement previously considered. In the last two sections it was pointed out that the imaged amplitude, $O(x_2, y_2)$, corresponds to an inverted replica of the input signal $S(x_1, y_1)$. This inverted property of the image applies to the image motion as well. That is, if the signal is displaced in the positive x_1 and y_1 direction, the image is displaced in the negative x_2 and y_2 directions. Thus, if $O(x_2, y_2)$ corresponds to the inverted image of $S(x_1, y_1)$, the displaced image corresponding to $S(x_1 - \tau_x, y_1 - \tau_y)$ is obtained simply by reversing the sign of the displacement to obtain $O(x_2 + \tau_x, y_2 + \tau_y)$. Using the displacement notation for the image amplitude distribution, Equation 116 can be rewritten as

$$O(x_2 + \tau_x, y_2 + \tau_y) = \phi(x_2, y_2) \iint e^{ikf(r/f)^4/4} F(p, q)\, e^{-i2\pi(px_2 + qy_2)}\, dp\, dq \,. \tag{120}$$

The final lens, L_3, in Figure 16 operates on the light amplitude distribution appearing in its input plane F''. For an optical correlator operation, a reference signal $R(x_2, y_2)$ is inserted into the plane F'' in the form of an amplitude transmission function of a photographic transparency. In this case the light amplitude distribution operated on by lens L_3 is that which appears on the output side of the reference transparency. This light amplitude is given by the product of the incident light amplitude, $O(x_2 + \tau_x, y_2 + \tau_y)$, and the reference transmission function $R(x_2, y_2)$. Thus the light amplitude distribution W in the output plane F_0 is given by

$$W(x_3, y_3, \tau_x, \tau_y) = e^{ikf(r_3/f)^4/4} \iint R(x_2, y_2)\, O(x_2 + \tau_x, y_2 + \tau_y)\, e^{-i2\pi(sx_2 + ty_2)}\, dx_2\, dy_2 \,, \tag{121}$$

where

$$s = \frac{x_3}{\lambda f} \,, \qquad t = \frac{y_3}{\lambda f} \,, \text{ and } r_3^2 = x_3^2 + y_3^2 \,.$$

Since the system being considered terminates at the F_0 plane, the intensity will be detected, measured, or recorded in the F_0 plane. The intensity in the output plane is given by the product of Equation 121 and its complex conjugate. The complex conjugate product of the exponential in front of the integral results in the cancellation of the exponential. Thus, the exponential in Equation 121 can be dropped since it will not affect the detected intensity output. Equation 121, therefore, can be simplified to

$$W(x_3, y_3, \tau_x, \tau_y) = \iint R(x_2, y_2) O(x_2 + \tau_x, y_2 + \tau_y)\, e^{-i2\pi(sx_2 + ty_2)}\, dx_2\, dy_2 \,. \tag{122}$$

Finally, if only the point located at the intersection of the optical axis with the plane F_0 (back focal point of L_3) is considered, $s = t = 0$ (i.e., $x_3 = y_3 = 0$), and Equation 122 reduces to

$$W(\tau_x, \tau_y) = \iint R(x_2, y_2)\, O(x_2 + \tau_x, y_2 + \tau_y)\, dx_2\, dy_2 \,, \tag{123}$$

where

$$W(\tau_x, \tau_y) = W(x_3 = 0, y_3 = 0, \tau_x, \tau_y) \,.$$

Equation 123 corresponds to a two-dimensional correlation function that implies that the light amplitude at the back focal point ($x_3 = y_3 = 0$) of the lens L_3 is given by the cross correlation of the reference, $R(x_2, y_2)$, and the image amplitude, $O(x_2, y_2)$. Thus as the input signal is displaced, the variation of the light amplitude $W(\tau_x, \tau_y)$ corresponds to the variation of the correlation function with respect to the displacements τ_x and τ_y. Note that the correlation function defined by Equation 123 involves the image amplitude $O(x_2, y_2)$ which is inverted with respect to the input signal $S(x_1, y_1)$. Therefore, if $R(x_2, y_2)$ is not a symmetrical function, it must be oriented correctly with respect to the image $O(x_2, y_2)$ rather than with respect to the input signal $S(x_1, y_1)$.

Considered briefly now is the implication of the steps from Equation 122 to Equation 123. This step in the derivation was accomplished by stating that only the single point in the output plane F_0 which lies on the optical axis (i.e., $x_3 = y_3 = 0$) could be considered. In practice it is physically impossible to isolate a single point. The best attempt that can be made is to restrict the light measurement or detection to a small area about the selected point. The light amplitude at points within this area (except for the one point on the optical axis) is given by Equation 122 rather than by Equation 123. The light amplitude distribution will not be uniform over the finite area of measurement because of the phase variation involved in the integral of Equation 122. For example, using a pinhole aperture 10 microns in diameter to define the detection area, the phase term in Equation 122 can vary as much as 4π radians (2 cycles) over the range of the image aperture (assuming $r_{2max} \sim 0.14f$, $f = 10$ cm, $\lambda = 5461 \times 10^{-8}$ cm). The effects of the phase term in Equation 122 is to reduce the light amplitude at points off axis since the contributions to the integral are not in phase. Therefore, the actual light available through a pinhole aperture located in the F_0 plane at $x_3 = y_3 = 0$ will be less than that found by assuming the light amplitude given by Equation 123 appears at all points within the pinhole aperture. This problem will not be considered any further here since the analysis would depend on the type of photodetector or measurement technique used. It will be assumed that the variations involved are small enough so that any measurement will yield values proportional to the square of the amplitude given by Equation 123.

As pointed out above, the correlation function defined by Equation 123 involves the image amplitude $O(x_2, y_2)$ rather than the single amplitude $S(x_1, y_1)$. As defined by Equation 120 the image amplitude contains phase terms not present in the signal. A correlation operation can be performed based on the image as given by Equation 123; however, the reference signal $R(x_2, y_2)$ would have to be selected in terms of the image $O(x_2, y_2)$ including the phase terms. The correlation function obtained would correspond to a distorted signal rather than the actual signal $S(x_1, y_1)$. The presence of distortion caused by the phase terms in Equation 120, therefore, complicates the analysis and determination of the correlation process. For example, the image amplitude $O(x_2, y_2)$ will be complex (phase variation as well as amplitude), and for complete correlation a complex reference signal is required. Such reference transparencies are difficult to produce. The phase distortions are commonly neglected and a reference signal is selected on the basis of an ideal image (no distortion) of the input signal. Analysis is now made of such a system to determine the effects of the undesirable phase terms present in our equations.

Consider a signal which would produce an ideal image amplitude defined by

$$O(x_2 + \tau_x) = \sum_n B_n \cos 2\pi p_n (x_2 + \tau_x) . \tag{124}$$

Equation 124 defines a signal image composed of a series of cosine harmonics in one dimension. A one-dimensional signal has been chosen to simplify the analysis. Referring to Equation 120 it is found that each frequency term in the image has a phase term $e^{ikf(\lambda p_n)^4/4}$ associated with it, and the image also has a phase term $\phi(x_2, y_2)$ associated with it. Thus, for the actual image,

$$O(x_2 + \tau_x) = \phi(x_2) \sum_n B_n e^{ikf(\lambda p_n)^4/4} \cos 2\pi p_n (x_2 + \tau_x) . \tag{125}$$

A reference signal without phase can be considered such that

$$R(x_2) = \sum_m R_m \cos 2\pi p_m x_2 . \tag{126}$$

The reference signal $R(x_2)$ defined by Equation 126 has been selected to have the same cosine harmonics $(m = n)$ as the imaged signal being considered. Note that the reference signal defined by Equation 126 does not contain the phase terms present in Equation 125. The product of reference and image for the ideal image of Equation 124 is given by

$$R(x_2)\,O(x_2+\tau_x) = \sum_{n,m} B_n R_m \cos 2\pi p_m x_2 \cos 2\pi p_n (x_2+\tau_x) \,. \tag{127}$$

The product of reference and image for the actual image of Equation 125 is given by

$$R(x_2)\,O(x_2+\tau_x) = \phi(x_2)\sum_{n,m} B_n R_m e^{ikf(\lambda p_n)^4/4} \cos 2\pi p_m x_2 \cos 2\pi p_n (x_2+\tau_x) \,. \tag{128}$$

Substituting Equations 127 and 128 into Equation 123, the result for the correlation function of the ideal image is

$$W(\tau_x) = \sum_{n,m} B_n R_m \int \cos 2\pi p_m x_2 \cos 2\pi p_n (x_2+\tau_x)\, dx_2 \,, \tag{129}$$

and for the correlation functions of the actual image:

$$W(\tau_x) = \sum_{n,m} B_n R_m e^{ikf(\lambda p_n)^4/4} \int \phi(x_2) \cos 2\pi p_m x_2 \cos 2\pi p_n (x_2+\tau_x)\, dx_2 \,. \tag{130}$$

Comparing Equations 129 and 130 it is found that the term $e^{ikf(\lambda p_n)^4/4}$ affects the phase of each term in the double summation. From the previous discussion of frequency limitations with respect to this phase term it can be seen that applying the limitation developed for imaged intensity limits the phase variation of this term to approximately 7 degrees. In summing terms which are not in phase, the result will be less than summing the same terms in amplitude only. Thus the presence of the phase term $e^{ikf(\lambda p_n)^4/4}$ has the effect of reducing the value of $W(\tau_x)$ in Equation 130 as compared to Equation 129. However, since the maximum phase will be about 7 degrees, the difference due to this term will be small. The phase term $\phi(x_2)$ appears in the integral of each term in the sum and has the same effect on the integral (can be considered as summation) as the phase term discussed above had on the summation. However, as discussed above, the phase variations of $\phi(x_2)$ ranges over 19 cycles and the effect on the value of the integral will be correspondingly greater. The actual magnitude of the reduction in $W(\tau_x)$ because of these phase terms is difficult to evaluate in general since the reduction will depend on the form of the signals involved. However, from the discussion here it is apparent that the actual correlation function observed will be smaller in amplitude than that predicted using an ideal image. This result is obvious if it is considered that the presence of the phase terms in the actual image produce a mismatch between the signal and reference and therefore reduce the correlation. The effects of these phase terms can be reduced by further restricting the frequency range p_{max} (or r_{max}), and signal, and image aperture size which would limit the variation of the phase terms. An analysis of the required limitations will not be begun since it will depend to a large extent on the type of signals involved and the correlation results desired. Here, the equations necessary for such an evaluation have been developed and hopefully the significance of the various effects that appear in an optical system have been pointed out.

PHASE CORRECTIONS

In the last few sections were discussed the effects of undesirable phase terms in optical systems and it was demonstrated that these effects can be minimized by restricting the size of signal apertures and the spectral range of the signals. An alternative approach can be pursued by inserting phase corrections into the optical system. Such phase corrections can be implemented by inserting sheets or plates of transparent materials whose thickness or index of refraction has variations that introduce phase terms opposite in sign to those introduced by the system.

The basic equation representing the Fourier transform operation of a lens was given by Equation 45 as

$$A(x, y) = \frac{-i e^{ikR(x,y)}}{\lambda(f+d)} \iint A'\left(x_1, y_1\right) e^{-i2\pi(px_1+qy_1)} dx_1\, dy_1 \ . \tag{45}$$

Rewriting this equation and retaining only the variable part of the terms outside the integral

$$A(x, y) = \phi(x, y) \iint A'\left(x_1, y_1\right) e^{-i2\pi(px_1+qy_1)} dx_1\, dy_1 \ , \tag{131}$$

where

$$\phi(x, y) = e^{i2kf\left[(1+r^2/2f^2)(1+r^2/f^2)^{-1/2}\right]}, \text{ assuming } \quad d = f \ ,$$

as derived in Equation 119. As pointed out in all discussions above, the phase term $\phi(x, y)$ destroys the simple Fourier transform representation of lens focussing properties since the integral part of Equation 131 corresponds to a Fourier transform by itself. Thus consider inserting a phase correction plate into the back focal plane of a lens with transmission properties given by

$$P(x, y) = A_0\, e^{iC} \phi^*(x, y) \ , \tag{132}$$

where

$$A_0 = \text{constant} \ ,$$

$$C = \text{constant} ,$$

and

$$\phi^* (x, y) = e^{-i2kf[(1+r^2/2f^2)(1+r^2/f^2)^{-1/2}]} .$$

The light amplitude distribution appearing at the output side of the plate will be

$$A(x, y) P(x, y) = \iint A' (x_1, y_1) e^{-i2\pi(px_1+qy_1)} dx_1 dy_1 , \tag{133}$$

where the constant term $A_0 e^{iC}$ has been dropped and $\phi(x, y) \phi^* (x, y) = 1$. Thus by inserting a phase plate with transmission properties given by Equation 132 in the back focal plane of each lens in an optical system, the phase terms are eliminated. From the definition given by Equation 132 it is found that the phase correction depends on the focal length f of the lens and the wavelength $\lambda (k = 2\pi/\lambda)$ of the light. The phase correction is not dependent on the signal used, and therefore a phase plate can be made for the lens and wavelength to be used in the system. Of course, the correction of phase by this method requires an accurate technique for producing the phase plate and positioning the plate in the optical system. In any case, it has been shown that the elimination of undesirable phase terms is possible at least in theory. Any inaccuracies in production or location of the phase plate may be acceptable as long as the phase terms are appreciably less than before the plate was introduced. Assuming the phase plate is an accurate representation of the transmission function of Equation 132, the Fourier transform relation of Equation 133 will be valid. With the relationship given by Equation 133, the operation of spectrum-analyzer, imaging, and optical-correlation systems can be described by the ideal cases used in the respective discussions, and no undesirable phase terms appear in the equations.

PHASE TERMS WHEN $d \neq f$

In the consideration of phase terms was considered the special case of an input plane coincident with the front focal plane of a lens ($d = f$). This special case was chosen to eliminate the phase effects of a term proportional to $(f - d)$. From Equation 118 can be gotten a complete expression for the variable part of the exponential term e^{ikR} as

$$\phi(x, y) = e^{ik(d-f)(1+r^2/f^2)^{-1/2}} e^{i2kf(1+r^2/2f^2)(1+r^2/f^2)^{-1/2}} . \tag{134}$$

Equation 134 reduces to the form of Equation 119 when $d = f$. Restricting our consideration to a rather limited range in the back focal plane of a lens, it has been shown that Equation 134 can be found to a good approximation in the form of Equation 88 such that

$$\phi(x, y) = e^{ik(f-d)(r/f)^{2/2}} e^{ikf(r/f)^{4/4}} . \tag{135}$$

Equation 135 can be considered as a representation of the phase in a back focal plane containing a frequency spectrum, while Equation 134 is a more accurate representation that applies in a back focal plane containing an image of the input signal. This application of Equation 134 and 135 is based on the relatively larger apertures commonly used in the signal and image planes.

In the systems that have been considered here, the complete phase variations, as given by Equation 134, appear only in the correlator system. This can be seen by noting the presence of $\phi(x_2, y_2)$ in Equation 120. Since this phase factor is expressed in terms of the coordinates x_2 and y_2 of the image phase, a very restrictive aperture limitation cannot be used without severaly affecting our signal handling capability. Therefore, the approximation of Equation 135 will not be valid and $\phi(x_2, y_2)$ in Equation 120 will have the form of Equation 134 with d_2 and r_2 replacing d and r respectively:

$$\phi(x_2, y_2) = e^{ik(d_2-f)(1+r_2^2/f^2)^{-1/2}} e^{i2kf(1+r_2^2/2f^2)(1+r_2^2/f^2)^{-1/2}} , \tag{136}$$

where $r_2^2 = x_2^2 + y_2^2$ and d_2 is the distance from the spectrum plane F' to the lens L_2 (see Figure 16). In the sample correlation function of Equation 130 it can be seen that the additional phase term dependent on $(d_2 - f)$ will increase the effect of $\phi(x_2)$ on the integrals. In practice, a system would be specified on the basis of locating lens L_2 so that $d_2 = f$. However, the exact positioning of the lenses in an optical system is obviously a practical impossibility. Thus the additional phase term containing $(d_2 - f)$ represents the phase distortion introduced by inaccuracies in the implementation of the system. Since the quantity $(d_2 - f)$ represents an inaccuracy, its value will usually be undetermined. Therefore, the first term in Equation 136 represents an undetermined phase error in the optical correlator system. If a guess or estimate of the tolerances in the system can be made, this error term can be used to determine the maximum distortion of the correlation function by analysis similar to that implied by Equation 130.

Since the variation of the phase term containing $(d_2 - f)$ in Equation 136 is not known specifically, the elimination of this term by a phase correction plate is not possible. Thus in a system containing phase correction plates, only the second term of Equation 136 can be eliminated. In such systems, $\phi(x_2, y_2)$ is completely given by the position error term

$$\phi(x_2, y_2) = e^{ik(d_2-f)(1+r_2^2/f^2)^{-1/2}} . \tag{137}$$

The distortion of the correlation function in a phase corrected system is, therefore, completely dependent upon the positioning errors. Again referring to the sample of Equation 130, $\phi(x_2)$ would be given in the form of Equation 137. The phase term in front of the integral of Equation 130 would also be replaced by an error term from an expression such as Equation 135 as will be discussed below.

The phase term given by Equation 135 represents the variable part of Equation 88. The exponential dependent on $(f - d_1)$ represents an error term due to lens positioning. To account for this error the complete phase approximation of Equation 135 must be used in place of the K_1 exponential of Equation 89. Thus the error phase term will appear throughout the previous analysis wherever the K_1 term has been used.

The effect of the K_1 phase term has been considered on the image intensity and on the correlation function in earlier sections. In the correlator discussion, the variable phase term of K_1 appears in the integral used to define the image amplitude distribution in Equation 120. To account for errors in placement of lens L_1 (see Figure 16) the exponential $e^{ikf(r/f)^4/4}$ in Equation 120 must be replaced by a phase term of the form of Equation 135, which can be written

$$\phi(x, y) = e^{ik(f-d_1)(r/f)^2/2} \, e^{ikf(r/f)^4/4} \, . \tag{138}$$

Referring to the sample correlation of Equation 130, the phase term given by Equation 138 will replace the $e^{ikf(r/f)^4/4}$ term in front of the integral. The error phase term has the affect of adding an additional variation to the phase of the terms in the summation. For a phase corrected system, the $e^{ikf(r/f)^4/4}$ term is eliminated and the undesirable phase difference of terms in the summation will be dependent only on the accuracy of the system implementation.

In the discussion of imaging systems, a factor m was defined by Equation 103 and a method for determining the accuracy of the image intensity was developed based on this parameter. To extend this method to include the case for $d_1 \neq f$, m can merely be redefined by

$$m^4 = \frac{k}{2}\left(\frac{r}{f}\right)^2 \left[\left(f - d_1\right) + \frac{f}{2}\left(\frac{r}{f}\right)^2\right] , \tag{139}$$

which is obtained from the exponents of the terms in Equation 138. For d_1 less than f, the limits on m defined in the previous discussion will apply to Equation 139 for the maximum value of r. It is noted that since $(f - d_1)$ is a positive quantity when d_1 is less than f, the required limit on r_{max}/f will be less than that determined for the case $d_1 = f$. When d_1 is greater than f, $(f - d_1)$ is a negative quantity which would imply that the value of r_{max}/f can be greater than that for the case $d_1 = f$. This is true except for cases in which d_1 is sufficiently greater than f so that for some value of r/f less than r_{max}/f the value of Equation 139 is greater in absolute value than for r_{max}/f. That is, since $(f - d_1)$ is negative, the right side of Equation 139 is zero at $r = 0$ and becomes negative as r increases until it reaches a maximum negative value and then increases to positive values. Depending on the value of d_1 and the limit r_{max}/f, it is possible that the phase at the maximum negative value is greater than that at the aperture limit r_{max}/f. In such cases, the maximum negative value must be considered rather than the end value at r_{max}/f, and the aperture may have to be restricted to values below this maximum. In any case, the brackets on the right side of Equation 139 must be considered as an absolute-value symbol when the quantity within is negative so that m will have real values. In other words, the concern is with the magnitude of the phase variations and not the sign.

For systems with phase correction, only the first term of Equation 138 will remain and Equation 139 will be simplified to

$$m^4 \;\square\; \frac{k}{2}\left(\frac{r}{f}\right)^2 \left(f - d_1\right). \tag{140}$$

Equation 140 can then be used with the image intensity criterion developed earlier to consider the effects of positioning error for lens L_1 in phase corrected systems.

In most of the literature the phase corrected form of Equation 138,

$$\phi(x, y) = e^{ik(f-d_1)(r/f)^2/2}, \tag{141}$$

is used even though phase correction techniques may not be employed. This application of Equation 141 requires that the frequency limitation be sufficient so that the $e^{ikf(r/f)^4/4}$ term can be neglected. This application also implies that the term $\left(f - d_1\right)$ is much greater than the maximum value of $f/2\,(r/f)^2$. If this condition does not hold, the neglected term will contribute a phase comparable to that of Equation 141 which would then be in error. Conversely, if $\left(f - d_1\right)$ is not greater than $f/2\,(r/f)^2$, and the $e^{ikf(r/f)^4/4}$ term is considered negligible, than the term given by Equation 141 is also negligible since it is comparable to the neglected term.

In this section has been outlined the procedure for taking into account the additional phase term arising from inaccuracies in the positioning of lenses. It was pointed out that since these terms are caused by inaccuracies, they are generally not specified completely. The worse case, however, can be specified by estimating the maximum error in the position of a lens. From this extreme estimate the necessary aperture limitation, or the evaluation of errors in the desired optical outputs, can be determined for a worse-case analysis. Unfortunately, because of the undetermined nature of these terms, phase correction cannot be used to eliminate their effects.

SUMMARY

The derivation presented in this report demonstrates that the Fourier transform representation of a focussed diffraction pattern is a reasonable approximation for describing the operation of coherent optical systems with lenses. The basic assumptions consisted of the ideal focal properties of a lens and the use of perfectly coherent light. Except for undesirable phase effects, it was demonstrated that the Fourier transform representation is obtained as a good approximation by imposing limitations on the size and frequency content of signals allowed. The phase terms can also be eliminated by aperture limitations; however, the restrictions are more severe. Depending on the application, a trade-off must be made between the limitations required for elimination of undesired terms and the desired signal size and frequency content.

Techniques for evaluating the effects of the various approximations and for analyzing the operation of ideal optical systems have been presented. For specified signals and applications these expressions can be used to determine the theoretical errors in assuming ideal operations as is commonly practiced. The analysis presented is by no means complete; however, it is hoped that it is sufficiently detailed to provide a clear insight into the required approximations.

This report represents an initial step in the development of a detailed analysis of the capabilities of optical processing systems. Further studies are required to formulate complete criteria and analysis techniques for practical optical systems. Some of the important areas which must be considered include:

1. Lens aberrations
2. Coherence
3. Transmission properties of modulation media
4. Band-limited signal approximations

These areas were not treated in the analysis presented here since the initial study was restricted to ideal systems. The complexity of the mathematical formulation of optical patterns can be simplified somewhat by using the notation of communication theory (Reference 5). Such methods are becoming quite useful in modern optics studies. The development of these techniques provides a means for avoiding the complicated mathematical formulations inherent in diffraction problems. However, any new formulations such as these must be considered in terms of the more rigorous formulation since the various approximations are basically the same in both formulations.

Goddard Space Flight Center
National Aeronautics and Space Administration
Greenbelt, Maryland, March 28, 1969
125-23-02-05-51

REFERENCES

1. O'Neill, E. L., "Introduction to Statistical Optics," New York: Addison-Wesley, 1963.

2. Linfoot, E. H., "Fourier Methods in Optical Image Evaluation," p. 12, Focal Press, 1964.

3. Rhodes, J. E., Jr., "Analysis and Synthesis of Optical Images," *American Journal of Physics*, 21:337-343, May 1953.

4. Born, M., and Wolf, E., "Principles of Optics," New York: Pergamon Press, 1959.

5. Levi, L., "Applied Optics, A Guide to Optical System Design," vol. 1, New York: John Wiley, 124-154, 1968.

6. Sommerfeld, A., "Lectures on Theoretical Physics," vol. 4, New York: Academic Press, 1964.

Appendix A

Derivation of Diffraction Formula

This discussion is restricted to the somewhat special case of diffraction at an infinite plane surface as diagrammed in Figure A1. The shaded area in the figure represents a cross-section of an infinite slab. The basic problem is to determine the electric field at any point P in the diffraction region, which includes all points to the right of the plane boundary surface indicated in Figure A1. When the plane slab is not present, the electric field at any point P can be found simply by substituting the coordinates of P into the mathematical expression describing the light propagating from whatever light source may be present. In itself, finding a mathematical representation for a given light source is not a simple problem. The light radiated by a square is dependent upon the mechanism generating the light as well as the geometry of the source. In many cases it is assumed that a good approximation is obtained by considering ideal light sources which radiate spherical waves (point sources) or plane waves (point sources at an infinite distance).

Inserting a plane surface into the path of the light waves as shown in Figure A1 complicates our problem. Since the presence of the plane effects the propagation characteristics in space, the electric field at any point will now depend on the characteristics of both the light source and the plane. The characteristics of the plane depend on the type of material of which it is made and these characteristics usually vary from point to point in the surface. Thus the problem is that of determining the electric field in the presence of a surface which can have widely varying electrical properties from point to point. As the reader may already know, problems of this nature are very difficult and, in fact, very few diffraction problems have been solved rigorously. Fortunately, in many cases of practical interest, results within experimental accuracy can be obtained by less rigorous techniques.

In order to implement a discussion of diffraction problems, a formula must be derived for diffraction at a plane surface. This result was first derived by Sommerfeld[6] in 1896 and, as will be demonstrated, is effectively a mathematical representation of Huygen's principle for the special case of a plane diffraction surface. The basic assumption we will start with is that the components of the electric field are known at every point on the right hand boundary

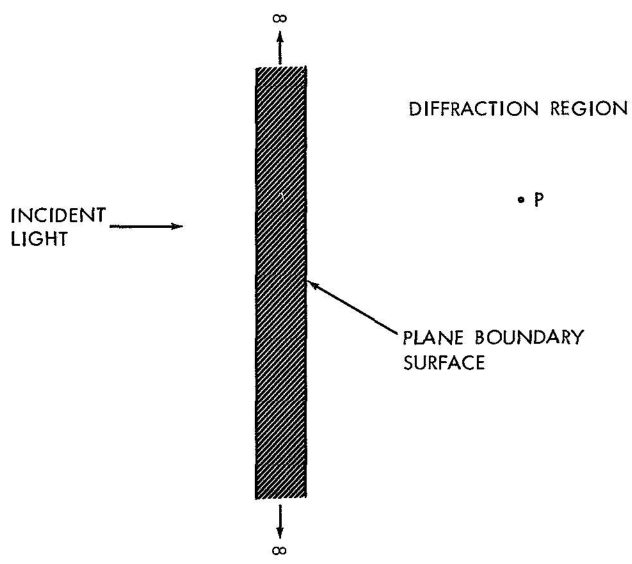

Figure A1—Diagram of diffraction configuration.

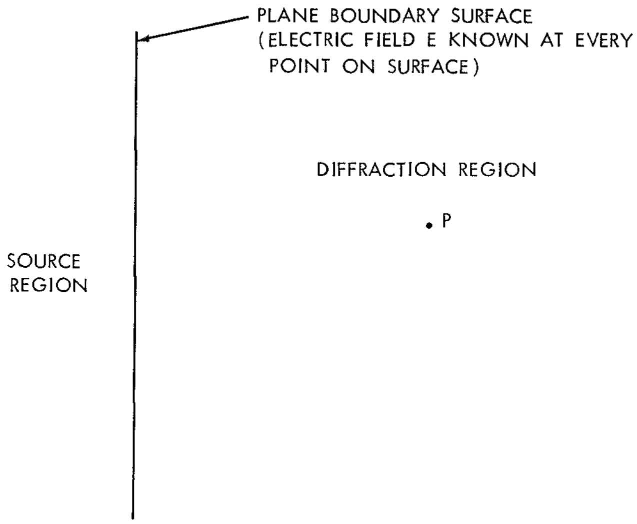

Figure A2—Outline of diffraction problem.

surface of the plane slab (refer to Figure A1). The methods for determining these field values are of no importance at this point; however, in many cases of interest, the assumption of a multiplying factor representing the transmission properties of a thin material provides results in close agreement with experiment. For the present purpose, it is simply assumed that the value of the electric field at every point on the plane boundary surface is known (i.e., can be found easily).

Referring to Figure A2, the problem to be solved can be stated as follows:

> Given the electric field at every point on an infinite plane boundary, what is the electric field at any point P in the diffraction region?

As shown in Figure A2, the diffraction region is defined to be all space on the right side of the boundary surface (note that this region does not contain any light sources). In Figure A2, it is assumed that all light sources are to the left of the boundary surface and that the diffraction region includes all points to the right of the boundary. It can be assumed that the diffraction region is in free space (velocity of light is c in all directions) and that there are no electric currents or charges present in this region. Since the electric field is a vector quantity, its direction at any point is as important as its magnitude. As in the case of any vector, the electric field can be considered in terms of its components in the x, y, and z directions. To simplify the discussion it is assumed that the light waves are monochromatic, or are constant in time at a single frequency. When necessary this discussion can be extended to the general case of nonmonochromatic waves by considering each separate frequency component, as described here, and summing up all components.

Each component E (in x, y, z components) of the electric field of a monochromatic wave will satisfy the Helmhotz equation (time-independent wave equation) at every point P in free space which contains no electrical sources:

$$\left(\nabla^2 + k^2\right)E = 0 \ , \tag{A1}$$

where

$$E = \text{x, y, or z component of the electric field} \ ,$$

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \ ,$$

$$k = \frac{\omega}{c} = \frac{2\pi}{\lambda} ,$$

ω = angular frequency ,

c = speed of light ,

λ = wavelength of light .

Using the values for E on the plane boundary surface and the fact that E must satisfy Equation A1 at every point P in our diffraction region, a formula can be derived for the electric field at P in terms of the values given on the surface.

An arbitrary function V which also satisfies Helmholtz equation can be introduced

$$\left(\nabla^2 + k^2\right)V = 0 . \tag{A2}$$

There are many functions which will satisfy Equation A2; however, we will continue to use the symbol V and reserve the selection of a specific function until a few additional characteristics of V which will allow us to accomplish our derivation have been determined. In terms of E and V, two vectors $\vec{F}_1$ and $\vec{F}_2$ can be defined as

$$\vec{F}_1 = E \nabla V , \tag{A3}$$

and

$$\vec{F}_2 = V \nabla E , \tag{A4}$$

where

$$\nabla = \vec{i}\frac{\partial}{\partial x} + \vec{j}\frac{\partial}{\partial y} + \vec{k}\frac{\partial}{\partial z} ,$$

and ∇V and ∇E denote the gradient of V and E respectively. Gauss's theorem is introduced as

$$\iiint\limits_{\text{volume}} \nabla \cdot \vec{F}\, dv = \iint\limits_{\text{surface}} \vec{F} \cdot d\vec{s} , \tag{A5}$$

where the volume integral on the left can be taken over any volume that does not contain discontinuities of the divergence of $\vec{F}(\nabla \cdot \vec{F})$ and the surface integral on the right is over the surface which encloses the volume ($\vec{F}$ must be continuous on the surface so that the surface integral can be found).

From Equations A3 and A4,

$$\nabla \cdot \vec{F}_1 = \nabla \cdot E \nabla V = \nabla E \cdot \nabla V + E \nabla^2 V , \tag{A6}$$

and

$$\nabla \cdot \vec{F}_2 = \nabla \cdot V \nabla E = \nabla V \cdot \nabla E + V \nabla^2 E . \tag{A7}$$

From above it is noted that Equations A6 and A7 must not have discontinuities within the volume of integration; therefore, E and V must have continuous first and second derivatives. Since diffraction region free of electrical current and charge is under consideration, E will meet this requirement for any volume in the diffraction region. Since a specific V has not yet been selected, we will note this requirement and be sure to satisfy it when selecting V. Thus Equation 5A can be written by substituting Equations 3A, 4A, 6A, and 7A:

$$\iiint \nabla \cdot \vec{F}_1 \, dv = \iiint \{\nabla E \cdot \nabla V + E\nabla^2 V\} \, dv = \iint E \nabla V \cdot d\vec{s} , \tag{A8}$$

$$\iiint \nabla \cdot \vec{F}_2 \, dv = \iiint \{\nabla V \cdot \nabla E + V\nabla^2 E\} \, dv = \iint V \nabla E \cdot d\vec{s} . \tag{A9}$$

Equation A9 can be subtracted from Equation A8, and noting that $\nabla E \cdot \nabla V - \nabla V \cdot \nabla E = 0$, then

$$\iiint \{E\nabla^2 V - V\nabla^2 E\} \, dv = \iint \{E\nabla V - V\nabla E\} \cdot d\vec{s} . \tag{A10}$$

This is Green's theorem and holds for any functions E and V which have continuous first and second derivations in the region of integration. From Equations A1 and A2 it is known that $\nabla^2 E = -k^2 E$ and $\nabla^2 V = -k^2 V$ and therefore the bracket on the left hand side of Equation A10 gives

$$E\nabla^2 V - V\nabla^2 E = -k^2 EV - \left(-k^2 EV\right) = 0 . \tag{A11}$$

Since the integrand is zero as given in Equation A11, the volume integral on the left hand side of Equation A10 is zero and Equation A10 can be rewritten as

$$\iint \{E\nabla V - V\nabla E\} \cdot d\vec{s} = 0 . \tag{A12}$$

The surface integral in Equation A12 is to be taken over any closed surface which does not enclose discontinuous points. For the purposes here, the surface is chosen as indicated in Figure A3. The outer surface consists of the infinite boundary plane, A, on which the electric field is known, and a hemisphere, C, of infinite radius which connects the ends of the plane at infinity. An inner surface, Σ, is defined as a sphere centered at P with radius ϵ. Taking A, C, and Σ as the closed surface, the volume between the sphere Σ and the outer surface, A - C, is defined. If the limit is taken as ϵ goes to zero, the point P will be the only point in the diffraction region outside the surface of integration. Thus the sphere Σ isolates the point P where we want to find the electric field. The surface integral of Equation A12 can be written as the sum of the integrals over A, C, and Σ:

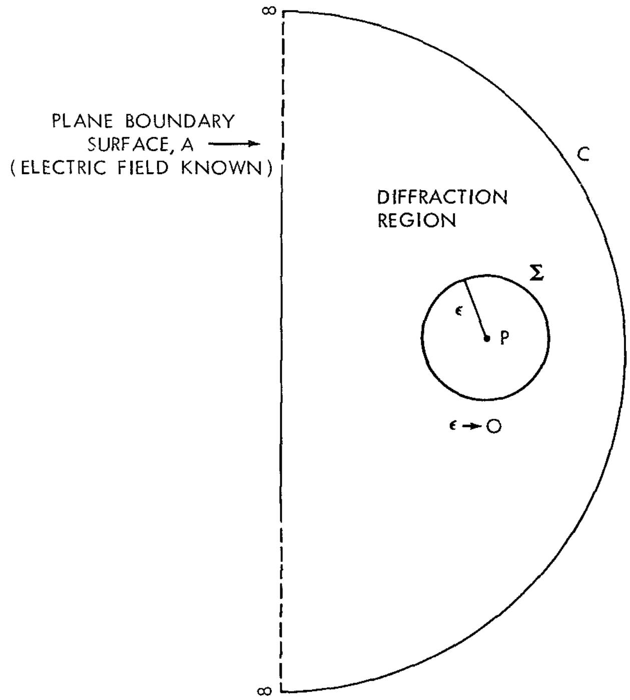

Figure A3—Boundary surface enclosing all points except P.

$$\iint_C \{E\nabla V - V\nabla E\} \cdot d\vec{s} + \iint_A \{E\nabla V - V\nabla E\} \cdot d\vec{s} + \iint_\Sigma \{E\nabla V - V\nabla E\} \cdot d\vec{s} = 0 . \tag{A13}$$

The surface integral over the hemisphere at infinity can be eliminated through a physical argument given by Born and Wolf (Reference 4). In practice a light wave starts at some time, and since it propagates at a finite velocity (c in free space) must have an end. We can imagine the infinite hemisphere continually expanding in front of the light waves. In this way, the contribution of the wave on the hemisphere is zero since the light waves never reach the hemisphere. The integral over the surface C will therefore be zero and Equation A13 can be written

$$\iint_A \{E\nabla V - V\nabla E\} \cdot d\vec{s} + \iint_\Sigma \{E\nabla V - V\nabla E\} \cdot d\vec{s} = 0 . \tag{A14}$$

Advantage is now taken of our freedom to select a function V in order to simplify Equation A14. In the integral over the plane surface A it is noted that the term $V\nabla E$ requires the values of ∇E on the boundary surface. Since ∇E is not known on A, V must be required to be zero on the boundary surface to eliminate this term. Equation A14 can then be written

$$\iint_A E\nabla V \cdot d\vec{s} + \iint_\Sigma \{E\nabla V - V\nabla E\} \cdot d\vec{s} = 0 . \tag{A15}$$

To consider the integral over the sphere Σ, the surface element $d\vec{s}$ is expressed in polar coordinates

$$d\vec{s} = -\epsilon^2 \sin\theta \, d\theta \, d\phi \, \hat{r} \, ,$$

where $\hat{r}$ is a unit vector in a radial direction away from the center at P and the minus sign is required by the convention that a surface normal is directed away from the enclosed volume. The integral over Σ can be written as

$$-\int_0^{\pi} \sin\theta \, d\theta \int_0^{2\pi} d\phi \, \epsilon^2 \, (E\nabla V - V\nabla E) \cdot \hat{r} \, , \tag{A16}$$

By vector identity $(E\nabla V - V\nabla E) \cdot \hat{r} = E(\partial V/\partial r) - V(\partial E/\partial r)$, and Equation A16 can be written

$$-\int_0^{\pi} \sin\theta \, d\theta \int_0^{2\pi} d\phi \left[E\left(\epsilon^2 \frac{\partial V}{\partial r}\right) - \epsilon^2 V \frac{\partial E}{\partial r}\right] . \tag{A17}$$

But the desire is to take the limit as ϵ goes to zero. The term $\epsilon^2 V(\partial E/\partial r)$ in Equation A17 can be eliminated if V is required to satisfy the condition

$$\lim_{\epsilon \to 0} \epsilon^2 V = 0 \, , \tag{A18}$$

at any point P in the diffraction region.

Since E has continuous first derivations $\partial E/\partial r$ will be finite and the conditioning given by Equation A18 will give

$$\lim_{\epsilon \to 0} \epsilon^2 V \frac{\partial E}{\partial r} = 0 \, . \tag{A19}$$

Expression A17 is then given as

$$-\int_0^{\pi} \sin\theta \, d\theta \int_0^{2\pi} d\phi \left\{\lim_{\epsilon \to 0} E\left(\epsilon^2 \frac{\partial V}{\partial r}\right)\right\} . \tag{A20}$$

Since E is the value of the field on the surface Σ, and the surface Σ reduces to the point P when ϵ goes to zero, the limit in Equation A20 can be written

$$\lim_{\epsilon \to 0} E \epsilon^2 \frac{\partial V}{\partial r} = E(P) \lim_{\epsilon \to 0} \epsilon^2 \frac{\partial V}{\partial r} \, , \tag{A21}$$

where $E(P)$ is the electric field at P. We will require that V satisfy the condition

$$\lim_{\epsilon \to 0} \epsilon^2 \frac{\partial V}{\partial r} = 1 , \tag{A22}$$

at any point P in the diffraction region. The limit in the brackets of Equation A20 is then simply $E(P)$ and Equation A20 gives

$$-\int_0^{\pi} \sin \theta \, d\theta \int_0^{2\pi} d\phi \, E(P) = -4\pi E(P) . \tag{A23}$$

Substituting the result of Equation A23 into Equation A15 for the surface integral over Σ,

$$\iint_A E \nabla V \cdot d\vec{s} - 4\pi E(P) = 0 . \tag{A24}$$

Rearranging terms, Equation A24 can be written

$$E(P) = \frac{1}{4\pi} \iint_A E \nabla V \cdot d\vec{s} . \tag{A25}$$

By vector identity, $\nabla V \cdot d\vec{s} = (\partial V/\partial n)\, ds$, where $\partial/\partial n$ represents a partial derivative with respect to a coordinate perpendicular to the plane boundary surface in a direction out from the enclosed region. Equation A25 can be written without vectors as

$$E(P) = \frac{1}{4\pi} \iint_A E \frac{\partial V}{\partial n} \, ds . \tag{A26}$$

Except for the selection of a function V which satisfies all the conditions that have been used here, Equation A26 has the required form. The left side is just the field at a point P, and since the integral on the right side is on the plane boundary surface, A the E in the integrand assume the given values on the surface. Thus the field at any point P is given in terms of the given values on the surface A by Equation A26.

In deriving Equation A26, restrictions have been imposed which must be satisfied by the function V. The requirements to be satisfied by V are:

1. V must have continuous first and second derivations within the region inside the boundaries shown in Figure A3.
2. $(\nabla^2 + k^2) V = 0$.

3. $V = 0$ on boundary surface.

4. $\nabla V \neq 0$ on boundary surface.

5. $\lim_{\epsilon \to 0} \epsilon^2 V = 0$ at any point P in diffraction region.

6. $\lim_{\epsilon \to 0} \epsilon^2 (\partial V/\partial r) = 1$ at any point P in diffraction region.

Fortunately there is a function which meets all these requirements:

$$V = \frac{e^{ikr'}}{r'} - \frac{e^{ikr}}{r} , \tag{A27}$$

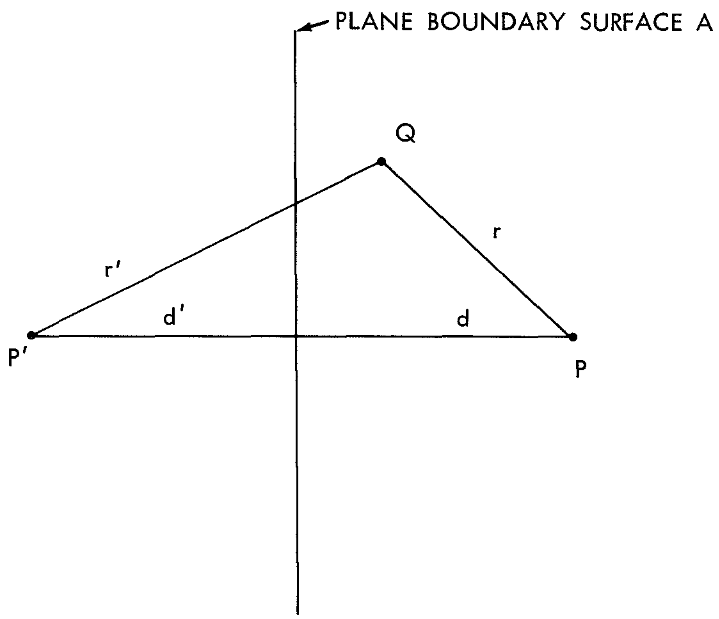

Figure A4—Geometry for definition of V.

where r and r′ are defined by Figure A4 as

r = distance from P to any point Q ,

r' = distance from P′ to any point Q,

P' = mirror image of P (i.e., PP′ is perpendicular to the boundary A and $d = d'$),

and

$$k = \frac{\omega}{c} = \frac{2\pi}{\lambda} .$$

That the function V given by Equation A22 does in fact satisfy all requirements, can be proved by direct substitution into the expressions listed above. Here we will only discuss the continuity requirements (item 1 above). As defined by Equation A27, V has discontinuities at $r = 0$ and at $r' = 0$. These discontinuities appear at the points P and P′ respectively; P′ lies outside the diffraction region, and P was separated out of the integration region by the sphere Σ. Thus the only points at which discontinuities appear are outside the region specified in the continuity requirement, and V given by Equation A27 does satisfy this requirement.

Returning to Equation A26, it is noted that $\partial V/\partial n$ on the surface A is required, rather than V itself. Coordinates can be selected so that the z axis is perpendicular to surface A as shown in Figure A5 so that $\partial V/\partial n$ becomes $-(\partial V/\partial z)$ (minus sign appears since the positive Z direction is opposite to the positive $\vec{n}$ direction). From the geometry of Figure A5, r and r′ are given as

$$r = \left[x^2 + y^2 + (d - z)^2\right]^{1/2} , \qquad \text{(A28)}$$

and

$$r' = \left[x^2 + y^2 + (d + z)^2\right]^{1/2} . \qquad \text{(A29)}$$

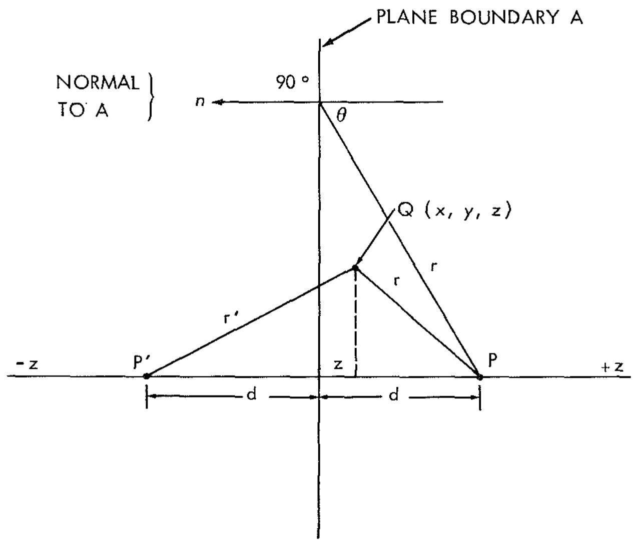

Figure A5—Geometry for definition of terms in V.

Substituting Equations A28 and A29 into Equation A27, it is found that

$$\frac{\partial V}{\partial n} = -\frac{\partial V}{\partial z} = -\frac{\partial}{\partial z}\left\{\frac{e^{ik\left[x^2+y^2+(d+z)^2\right]^{1/2}}}{\left[x^2+y^2+(d+z)^2\right]^{1/2}} - \frac{e^{ik\left[x^2+y^2+(d-z)^2\right]^{1/2}}}{\left[x^2+y^2+(d-z)^2\right]^{1/2}}\right\} ,$$

$$= \left\{\frac{-ik(d+z)}{\left[x^2+y^2+(d+z)^2\right]} + \frac{d+z}{\left[x^2+y^2+(d+z)^2\right]^{3/2}}\right\} e^{ik\left[x^2+y^2+(d+z)^2\right]^{1/2}}$$

$$+ \left\{\frac{-ik(d-z)}{\left[x^2+y^2+(d-z)^2\right]} + \frac{d-z}{\left[x^2+y^2+(d-z)^2\right]^{3/2}}\right\} e^{ik\left[x^2+y^2+(d-z)^2\right]^{1/2}} . \qquad \text{(A30)}$$

Now in Equation A26, integration was over the surface A, so that the value of $-(\partial V/\partial z)$ on A which is obtained by setting $z = 0$ in Equation A30 is needed:

$$\left.\frac{\partial V}{\partial n}\right|_A = -\left.\frac{\partial V}{\partial z}\right|_{z=0} = \frac{-2e^{ik\left[x^2+y^2+d^2\right]^{1/2}}}{\left[x^2+y^2+d^2\right]^{1/2}} \frac{d}{\left[x^2+y^2+d^2\right]^{1/2}} \left[ik - \frac{1}{\left[x^2+y^2+d^2\right]^{1/2}}\right] . \qquad \text{(A31)}$$

Equation A31 can be simplified by noting from Equation A28 and the geometry of Figure A5* that

$$r = \left[x^2 + y^2 + d^2\right]^{1/2} , \quad \text{when} \quad z = 0 ; \qquad \text{(A32)}$$

$$\cos\theta = \frac{d}{\left[x^2 + y^2 + d^2\right]^{1/2}} . \qquad \text{(A33)}$$

*Note that in Figure A5, P and P′ were chosen as point on the z axis to obtain Equations A32 and A33. In general x would be replaced by $(x - x_0)$ and y would be replaced by $(y - y_0)$ where x_0, y_0 define the x, y coordinates of the points P and P′. In the derivation, the z coordinate of the point P was represented by d to avoid confusion with the coordinates of the point Q. In general the d in Equations A32 and A33 is replaced by z. Otherwise the general results have the same form as found above.

Substituting Equations A32 and A33 into Equation A31, then

$$\left.\frac{\partial V}{\partial n}\right|_A = -\left.\frac{\partial V}{\partial z}\right|_{z=0} = \frac{-2e^{ikr}}{r}\left[ik - \frac{1}{r}\right]\cos\theta \; . \tag{A34}$$

Now Equation A34 can be substituted into Equation A26 so that the diffraction formula can be completed as

$$E(P) = \frac{-1}{2\pi}\iint_A E_A \frac{e^{ikr}}{r}\left[ik - \frac{1}{r}\right]\cos\theta \, ds \; , \tag{A35}$$

where (refer to Figure A5)

$E(P)$ = Electric field at a point P in the diffraction region,

E_A = electric field on the plane boundary A,

r = distance from P to a point on A,

θ = angle between r and normal to plane A,

and

$$k = \frac{\omega}{c} = \frac{2\pi}{\lambda} \; .$$

Given the electric field E at every point on the surface A, Equation A35 can be used to determine the electric field $E(P)$ at any point P in the diffraction region. This statement depends, of course, on whether the integration indicated in Equation A35 can be performed. If the integration cannot be performed analytically, it can be assumed that a numerical solution to any desired accuracy can be obtained using a computer. In many problems of interest, satisfactory results can be obtained by approximating Equation A35 using the geometry of the specific problem. For example, for small angles θ such that $\cos\theta \simeq 1$, and at great distances r such that $1/r \ll k$, Equation A35 can be approximated as

$$E(P) = \frac{1}{i\lambda}\iint_A E_A \frac{e^{ikr}}{r}\, ds \; . \tag{A36}$$

Equation A36 represents Huygens' principle since the contribution from each point on the boundary surface is given by $E_A\, e^{ikr}/r$ which describes the spatial variation of a spherical wave. Thus the diffracted field as given by Equation A36 can be interpreted as the summation (integral) of spherical waves radiating from each point on the diffraction boundary.

NASA-Langley, 1970 —— 23

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
WASHINGTON, D.C. 20546
OFFICIAL BUSINESS

FIRST CLASS MAIL

POSTAGE AND FEES PAID
NATIONAL AERONAUTICS AN
SPACE ADMINISTRATION

POSTMASTER: If Undeliverable (Section 158 Postal Manual) Do Not Return

"The aeronautical and space activities of the United States shall be conducted so as to contribute . . . to the expansion of human knowledge of phenomena in the atmosphere and space. The Administration shall provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof."

— NATIONAL AERONAUTICS AND SPACE ACT OF 1958

NASA SCIENTIFIC AND TECHNICAL PUBLICATIONS

TECHNICAL REPORTS: Scientific and technical information considered important, complete, and a lasting contribution to existing knowledge.

TECHNICAL NOTES: Information less broad in scope but nevertheless of importance as a contribution to existing knowledge.

TECHNICAL MEMORANDUMS: Information receiving limited distribution because of preliminary data, security classification, or other reasons.

CONTRACTOR REPORTS: Scientific and technical information generated under a NASA contract or grant and considered an important contribution to existing knowledge.

TECHNICAL TRANSLATIONS: Information published in a foreign language considered to merit NASA distribution in English.

SPECIAL PUBLICATIONS: Information derived from or of value to NASA activities. Publications include conference proceedings, monographs, data compilations, handbooks, sourcebooks, and special bibliographies.

TECHNOLOGY UTILIZATION PUBLICATIONS: Information on technology used by NASA that may be of particular interest in commercial and other non-aerospace applications. Publications include Tech Briefs, Technology Utilization Reports and Notes, and Technology Surveys.

Details on the availability of these publications may be obtained from:

SCIENTIFIC AND TECHNICAL INFORMATION DIVISION

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Washington, D.C. 20546