From: Sheppard CJR (2011) Diffraction optics, Handbook of Biomedical Optics, DA Boas, C Pitris, N Ramanujam, eds. CRC Press, pp. 11-31. ISBN: 1420090364, 978-142009036946.

5. Diffraction optics

5.1. Huygens' Principle

According to Huygens, each point on a wavefront serves as the source of a spherical secondary wavelet with the same frequency as the primary wave. The amplitude at any point is the superposition of these wavelets. Note that Huygens' principle considers diffraction as a summation of spherical waves, not as a summation of plane waves, as we will consider in Section 5.7. This theory gives a simple qualitative description of diffraction but needs to be adapted to give good agreement with more exact formulations that will be shown later, in Section 5.6.

Consider the propagation of a plane wave. Each point in the wavefront can be considered as a source of secondary waves (Fig. 5.1). This describes the propagating wave correctly, but suggests the possibility that the wave can equally well propagate backwards. This is one reason the model has to be modified to agree with exact theories.

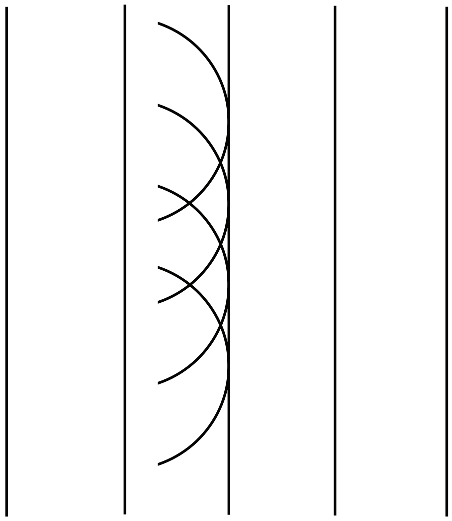

Fig. 5.1 Huygens’ principle for the propagation of a plane wave.

5.2. Fraunhofer and Fresnel Diffraction

Consider an opaque screen illuminated with a plane wave. The light spreads as a result of diffraction. If the observation screen is far enough away from the aperture, the diffraction pattern does not change in structure, but merely changes in size, as the distance is further increased. This situation is called Fraunhofer diffraction (Fig. 5.2). Closer to the aperture the diffraction pattern does change with distance. This is called Fresnel diffraction. Calculation of Fresnel diffraction is based on an approximation, which eventually breaks down: closer to the aperture more advanced theories are required. We shall discuss thse regions later. The Fraunhofer diffraction pattern is obtained at a very large distance from the aperture, but using a lens, an image of it can be formed at a finite distance. In this case, the regions either closer or further from the lens give Fresnel diffraction.

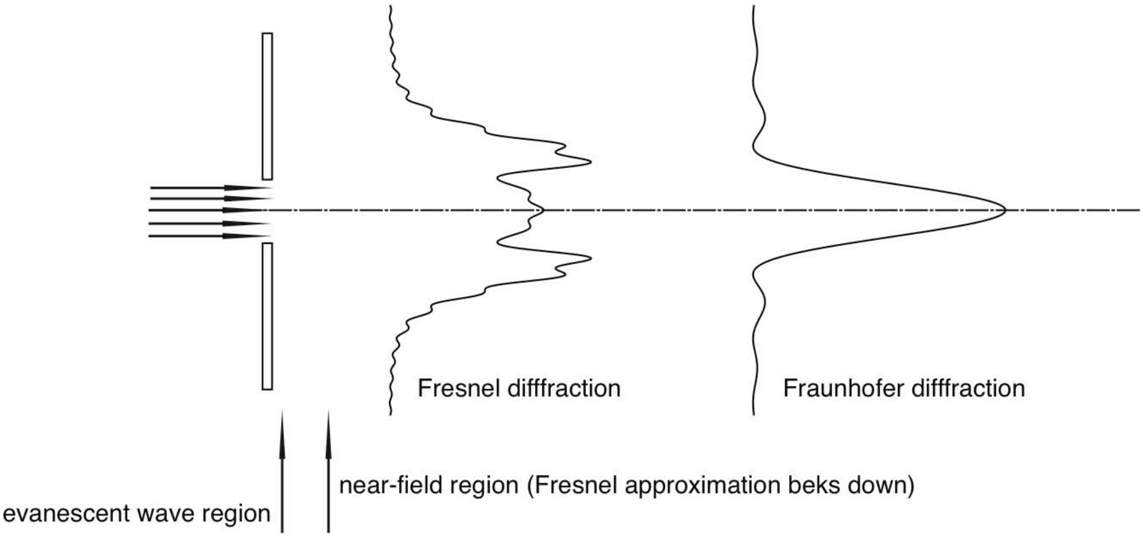

Fig. 5.2 Regimes of diffraction.

5.3. Huygens' diffraction formula

According to Huygens' principle the amplitude at *P* is obtained by integrating the contribution from the points in the surface *S*. If the source and observation point are quite close to the axis, for unpolarized light, the field vectors can be represented by scalars because they are almost normal to the axis. We find that

$$U\left(P\right)=-\frac{i}{\lambda}\iint_{S}\frac{e^{ikr}}{r}U\left(Q\right)dS\,. \tag{5.1}$$

This form of the equation is in a slightly different form compared with that of Hecht (Hecht 1987), for example. *U*(*Q*) gives the strength of the illumination at *Q*, e^{ikr}/r represents a spherical wave emanating from *Q*, and the factor $-i/\lambda$ results from the fact that *Q* is a driven dipole. The far-field of a dipole is $\pi/2$ out of phase with the forcing function, similar to the behaviour of resonance.

This expression has been developed from a semi-qualitative model: it is not strictly correct but gives a good prediction if, first, we are not too close to the aperture, and, second, the aperture is large compared with the wavelength. In optics these two conditions are usually, but not always, true. The more rigorous Kirchhoff diffraction formula we will derive later, but usually this is not much of an improvement because it still does not take account of the fact that the illumination of the aperture is changed by the presence of the screen, and in addition vector effects are also neglected.

If the aperture is illuminated with a spherical wave from a point a distance *s* away (Fig.5.3), then

$$U(Q) = A\frac{e^{iks}}{s} \quad ,$$

and

$$U(P) = -A\frac{i}{\lambda}\iint_S \frac{e^{ik(r+s)}}{rs}\,dS\,. \tag{5.2}$$

This is a form of the Huygens-Fresnel diffraction formula.

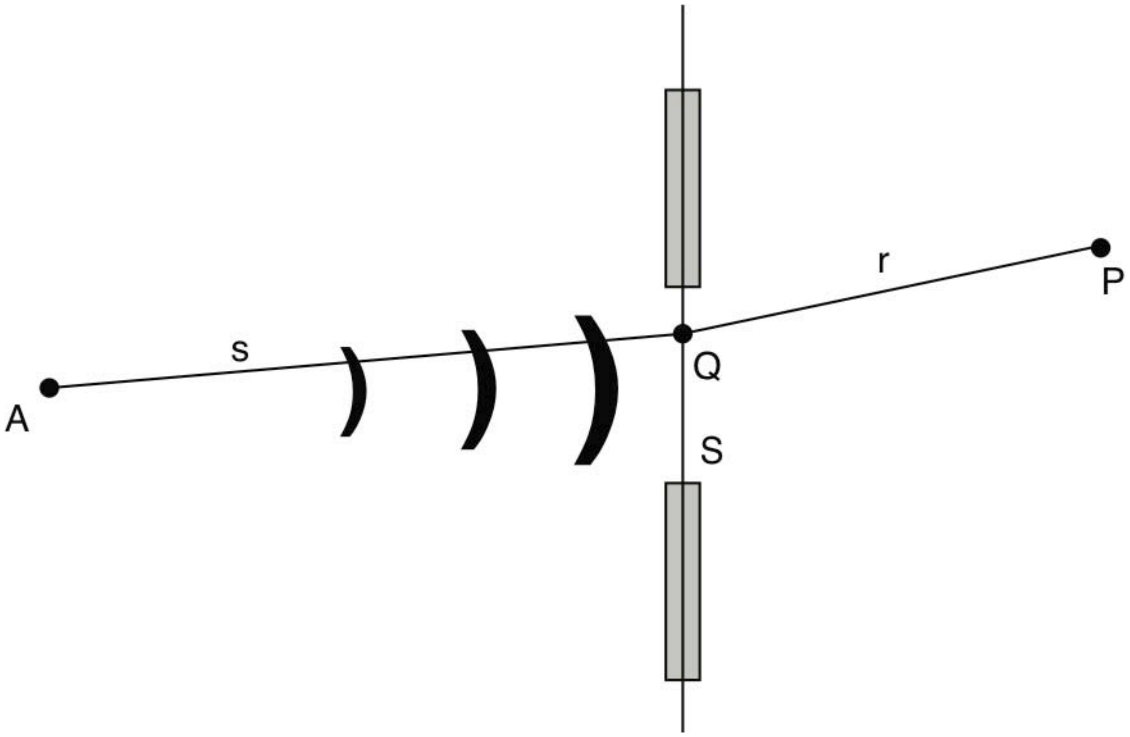

Fig. 5.3 Geometry of diffraction.

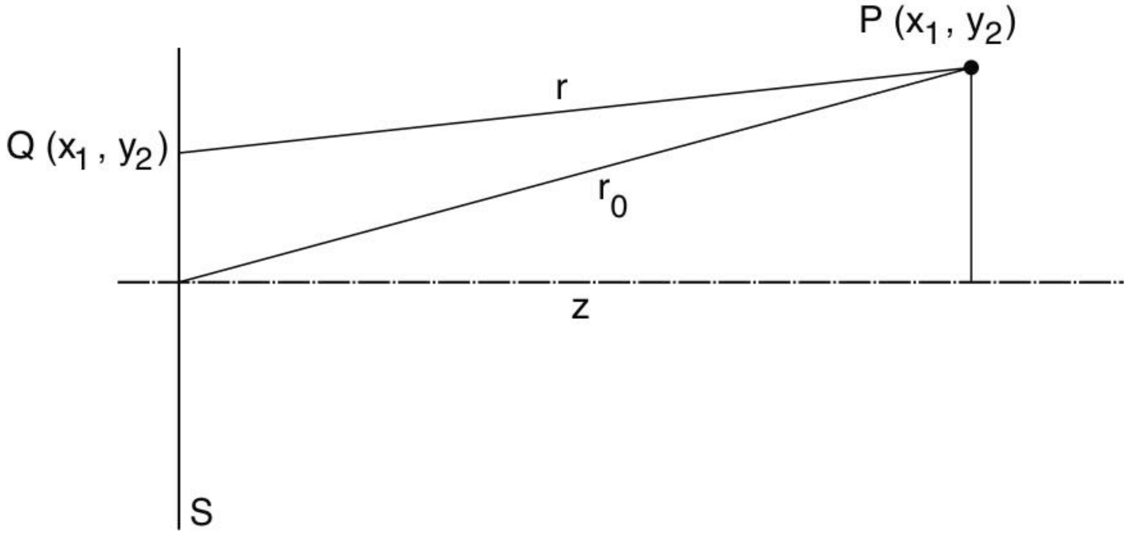

Fig. 5.4 Geometry of Fraunhofer diffraction.

5.4. Fraunhofer diffraction

In Eq. (5.1), r does not change very much as Q varies over S. Although it is still necessary to take account of the changes in the exponent, because this produces a multiplicative factor in the denominator, we can assume r is constant at a value r_0 :

$$U_2(P) = -\frac{i}{\lambda r_0} \iint_S e^{ikr} U_1(Q)\, dS. \tag{5.3}$$

Now, we have (Fig. 5.4)

$$\begin{aligned} r^2 &= (x_2 - x_1)^2 + (y_2 - y_1)^2 + z^2 \\ &= r_0^2 + (x_1^2 + y_1^2) - 2(x_1 x_2 + y_1 y_2). \end{aligned} \tag{5.4}$$

Taking the square root of both sides, if $x_1, y_1 << r_0$, we can neglect the second term and expand the square root to give

$$r = r_0 - \frac{(x_1 x_2 + y_1 y_2)}{r_0}. \tag{5.5}$$

This is one form of the Fraunhofer approximation. Sometimes the expansion is made in terms of z, rather than r_0, but expansion in r_0 is valid for larger values of x_2, y_2. So for our case

$$U_2(P) = -\frac{i}{\lambda r_0} \exp(ikr_0) \iint_S U_1(x_1, y_1) \exp\left[-\frac{ik}{r_0}(x_1 x_2 + y_1 y_2)\right] dx_1 dy_1. \tag{5.6}$$

The condition $x_1, y_1 << r_0$ can be written

$$r_0 >> \frac{k\left(x_1^{\ 2} + y_1^{\ 2}\right)_{\max}}{2}. \tag{5.7}$$

Introducing the Fresnel number N,

$$N = \frac{\left(x_1^{\ 2} + y_1^{\ 2}\right)_{\max}}{\lambda r_0}, \tag{5.8}$$

we have

$$N << \frac{1}{\pi}. \tag{5.9}$$

As an example, for $x_1^{\,2} + y_1^{\,2} = 1\ \text{mm}^2$, that is the width is 2mm, then for the Fraunhofer condition to be valid, $r_0 >> 6\text{m}$, which will not be the case in a laboratory experiment. But for $x_1 = 100\mu\text{m}, r_0 >> 60\text{mm}$, which we can observe easily.

If we define the Fourier transform of $f(x)$ as

$$F\left[f(m)\right] = \int_{-\infty}^{+\infty} f(x) e^{-2\pi imx} dx, \tag{5.10}$$

Eq. 5.6 is recognized as saying that

$$U_2(P) = \text{ const.} \times F\left[U_1(x_1, y_1)\right]. \tag{5.11}$$

In these expressions $U(x_1, y_1)$ is taken as the value of the incident field in the aperture, and zero outside of the aperture. This is called the Kirchhoff boundary condition.

5.4.1. Examples of Fraunhofer diffraction - Single Slit

For a long slit, length $2l$, width $2a$, in the plane $y_2 = 0$ illuminated uniformly

$$U_2(P) = -\frac{i}{\lambda r_0} \exp(ikr_0) \iint_S \exp\left(-\frac{2\pi i x_1 x_2}{\lambda r_0}\right) dx_1 dy_1$$

$$= -\frac{2\ell i}{\lambda r_0} \exp(ikr_0) \int_{-a}^{a} \exp\left(-\frac{2\pi i x_1 x_2}{\lambda r_0}\right) dx_1$$

$$= -\frac{2\ell i}{\lambda r_0} \exp(ikr_0) 2a \left[\frac{\sin\left(\dfrac{2\pi a x_2}{\lambda r_0}\right)}{\dfrac{2\pi a x_2}{\lambda r_0}}\right],$$

so

$$I = \frac{A^2}{\ell^2 r_0^2} \left[\frac{\sin(ka\sin\theta)}{ka\sin\theta}\right]^2, \tag{5.12}$$

where A is the area of the aperture.

5.4.2. Rectangular aperture

Consider now a rectangular slit, sides $2a$ and $2b$, illuminated with a plane wave. The diffracted amplitude is

$$U(P) = -\frac{i}{\lambda r_0}\exp(ikr_0)\int_{-a}^{a}\exp\left(-\frac{2\pi i x_1 x_2}{\lambda r_0}\right)dx_1 \int_{-b}^{b}\exp\left(-\frac{2\pi i y_1 y_2}{\lambda r_0}\right)dy_1. \qquad (5.13)$$

so that the intensity is

$$I = \left(\frac{A^2}{\lambda^2 r_0^2}\right)^2 \left[\frac{\sin\left(\frac{kax_2}{r_0}\right)}{\left(\frac{kax_2}{r_2}\right)}\right]^2 \left[\frac{\sin\left(\frac{kby_2}{r_0}\right)}{\left(\frac{kby_2}{r_2}\right)}\right]^2 . \qquad (5.14)$$

There are zeros along a series of perpendicular lines in the diffraction pattern (Fig. 5.5).

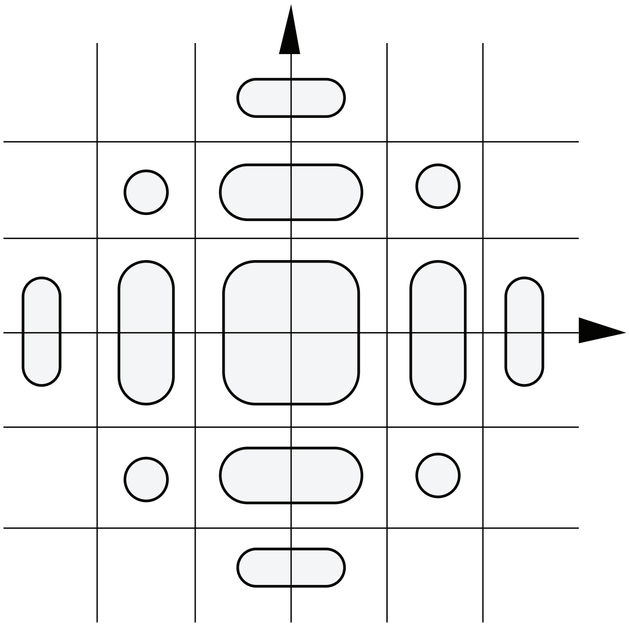

Fig. 5.5 Fraunhofer diffraction by a rectangular aperture.

5.4.3. Circular aperture

Consider diffraction by a circular aperture, radius *a*. Introducing polar coordinates (Fig. 5.6)

$$x_1 = R_1\cos\phi,\; x_2 = R_2\cos\varphi,$$
$$y_1 = R_1\sin\phi,\; y_2 = R_2\sin\varphi, \qquad (5.15)$$

the diffracted field is

$$U_2(\mathrm{P}) = -\frac{i}{\lambda r_0}\exp(ikr_0)\int_0^{2\pi}\int_0^a \exp\left[-\frac{ikR_1R_2\cos(\phi-\varphi)}{r_0}\right]R_1\,dR_1\,d\phi\,. \qquad (5.16)$$

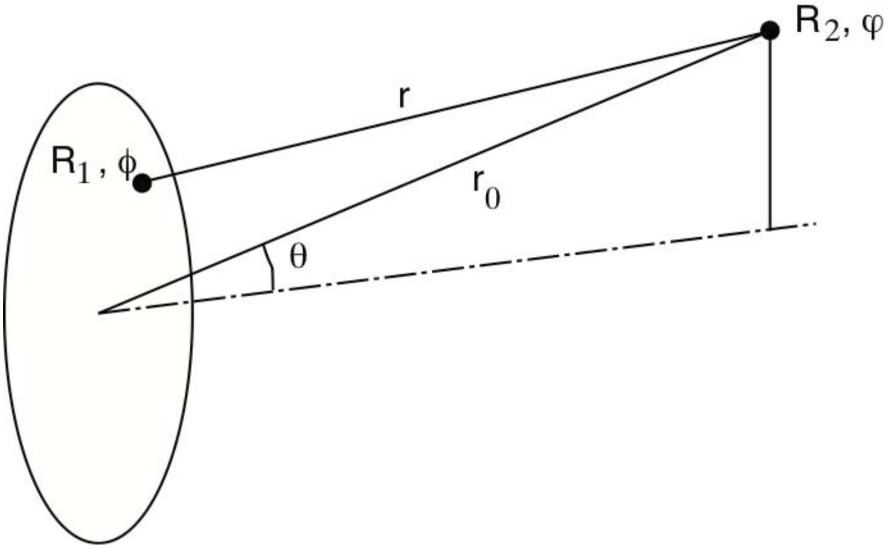

Fig. 5.6 Diffraction by a circular aperture.

As the answer, by symmetry, is independent of φ, we can assume without loss of generality that $\varphi = 0$. The integral in ϕ is thus

$$\int_0^{2\pi}\exp\left(-\frac{ikR_1R_2\cos\phi}{r_0}\right)d\phi\,. \qquad (5.17)$$

But we know that

$$J_0(z) = \frac{1}{2\pi}\int_0^{2\pi}\exp(iz\cos\phi)d\phi \qquad (5.18)$$

as this is the definition of the Bessel function of the first kind of order zero, so

$$U_2(\mathrm{P}) = -\frac{2\pi i}{\lambda r_0}\exp(ikr_0)\int_0^a J_0\left(\frac{kR_1R_2}{r_0}\right)R_1\,dR_1\,. \qquad (5.19)$$

This integral can be solved by use of the recurrence relationship (Abramowitz and Stegun 1965) Eq. 9.1.20

$$\frac{1}{z}\frac{d}{dz}[zJ_1(z)] = J_0(z)\,. \qquad (5.20)$$

We have

$$\int J_0(z)z\,dz = zJ_1(z), \tag{5.21}$$

so the diffracted amplitude is

$$U(R_2) = -\frac{i}{\lambda r_0}\exp(ikr_0)\frac{r_0^2}{k^2R_2^2}\int_0^{kaR_2/r_0} J_0(\xi)\xi d\xi$$

$$= -\frac{i}{\lambda R_2^2}\left[\xi J_1(\xi)\right]_0^{kaR_2r_0} = \frac{i\pi a^2}{r_0\lambda}\exp(ikr_0)\left[\frac{2J_1\left(\frac{kaR_2}{r_0}\right)}{\left(\frac{kaR_2}{r_0}\right)}\right]$$

$$= -\frac{iA}{\lambda r_0}\exp(ikr_0)\left[\frac{2J_1(ka\sin\theta)}{(ka\sin\theta)}\right], \tag{5.22}$$

where A is the area of the aperture, or

$$I(\theta) = \left(\frac{A}{\lambda r_0}\right)^2\left[\frac{2J_1(ka\sin\theta)}{(ka\sin\theta)}\right]^2. \tag{5.23}$$

This is illustrated in Fig. 5.7. Note that we have written the equation in this form as $2J_1(v)/v \to 1$ for $v \to 0$.

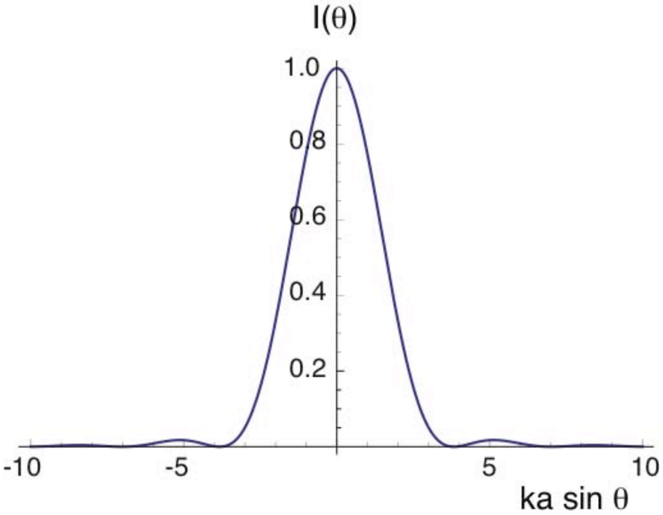

Fig. 5.7 Fraunhofer diffraction by a circular aperture.

5.4.4. Annular Aperture

For an annular aperture (Fig. 5.8), in Eq. (5.6) we can put

$$U_1(R_1) = \mathrm{circ}\left(\frac{R_1}{a}\right) - \mathrm{circ}\left(\frac{R_1}{\varepsilon a}\right), \tag{5.24}$$

where $\mathrm{circ}(x) = 1, x < 1; = 0, x > 1$. Therefore

$$U(R_2) = -\frac{i\pi a^2}{\lambda r_0}\exp(ikr_0)\left\{\left[\frac{2J_1\left(\frac{kaR_2}{r_0}\right)}{\left(\frac{kaR_2}{r_0}\right)}\right] - \varepsilon^2\left[\frac{2J_1\left(\frac{k\varepsilon aR_2}{r_0}\right)}{\left(\frac{k\varepsilon aR_2}{r_0}\right)}\right]\right\}. \tag{5.25}$$

As the value of ε is increased (Fig. 5.9), the central peak becomes narrower, but the side lobes become stronger. For the limiting case when $\varepsilon \to 1$,

$$\begin{aligned} U(R_2) &= -\frac{i}{\lambda r_0}\exp(ikr_0)\int_0^\infty \delta(R_1 - a)\, J_0\left(\frac{2\pi R_1 R_2}{\lambda r_0}\right) 2\pi R_1\, dR_1 \\ &= -\frac{i}{\lambda r_0}\exp(ikr_0)\, J_0\left(\frac{2\pi R_2 a}{\lambda r_0}\right). \end{aligned} \tag{5.26}$$

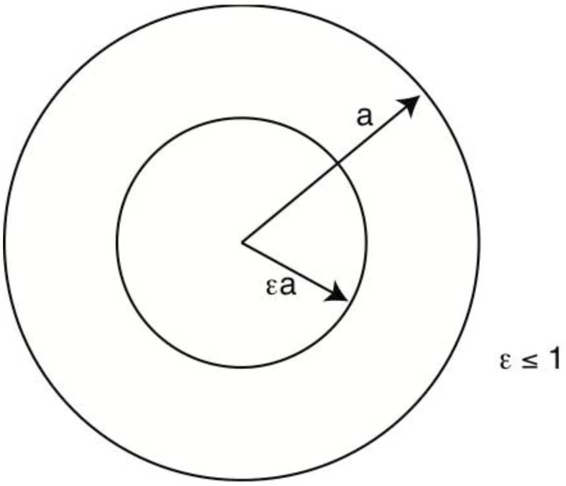

Fig. 5.8 An annular aperture.

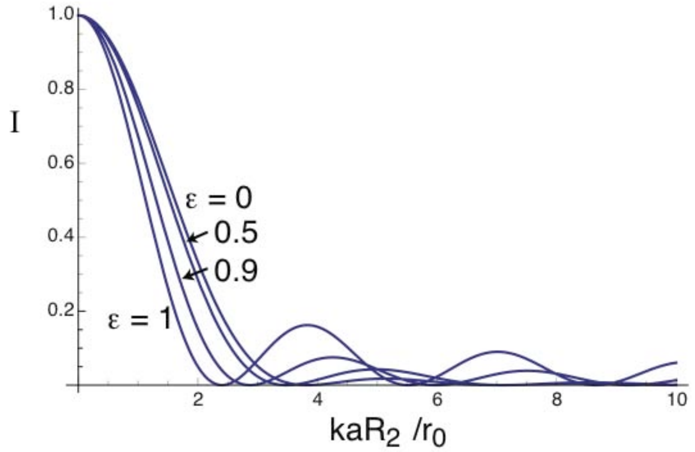

Fig. 5.9 Normalized intensity in the Fraunhofer diffraction pattern of an annular aperture.

5.5. Fresnel Diffraction

5.5.1. Introduction

Knowing the field in a plane x_1, y_1 we can calculate the field in the plane x_2, y_2 using Eq.5.3.

Using the binomial theorem to expand the square root, Eq.5.5 now becomes (without neglecting the second term)

$$r = r_0 + \frac{x_1^2 + y_1^2}{2r_0} - \frac{x_1 x_2 + y_1 y_2}{r_0}. \tag{5.27}$$

So

$$U_2(P) = -\frac{i}{\lambda r_0}\exp(ikr_0)\iint_S U_1(x_1,y_1)\exp\left[-\frac{ik}{r_0}(x_1x_2 + y_1y_2)\right]\exp\left[\frac{ik}{2r_0}\left(x_1{}^2 + y_1{}^2\right)\right]dx_1\,dy_1 \tag{5.28}$$

.

5.5.2. Circular aperture

Consider a circular aperture illuminated with a plane wave. Using Eq 5.18 in Eq.5.28 we have

$$U_2(R_2, r_0) = -\frac{ik}{r_0}\exp(ikr_0)\int_0^a J_0\left(\frac{kR_1R_2}{r_0}\right)\exp\left(\frac{ikR_1^2}{2r_0}\right)R_1 dR_1. \tag{5.29}$$

Unfortunately this integral cannot be solved in terms of elementary functions, though it can be solved in terms of Lommel functions.(Born and Wolf 1975) This approach is not particularly useful, and it is usually easier to solve it numerically.

But, along the axis it reduces to a simple form:

$$U_2(0,r_0) = -\frac{ik}{r_0}\exp(ikr_0)\int_0^a \exp\left(\frac{ikR_1^2}{2r_0}\right)R_1 dR_1$$

$$= -\frac{ik}{r_0}\exp(ikr_0)\left[\exp\left(-\frac{ikR_1^2}{2r_0}\right)\right]_0^a \qquad (5.30)$$

$$= -2i\exp(ikr_0)\exp\left(-\frac{ika^2}{4r_0}\right)\sin\left(\frac{ka^2}{4r_0}\right).$$

or

$$I_2(0,r_0) = 4\sin^2\left(\frac{ka^2}{4r_0}\right). \qquad (5.31)$$

So we obtain a series of maxima and minima (zeros) in intensity along the axis (Fig. 5.10). This can be explained in terms of Fresnel's zones in the following way. Contributions from successive zones (Fig. 5.11) tend to cancel as the phases are different by 180°. If the number of zones is N, then $a^2 = n\lambda r$ and

$$N = \frac{a^2}{\lambda r}, \qquad (5.32)$$

that is it is equal to the Fresnel number, defined earlier (Eq.5.8).

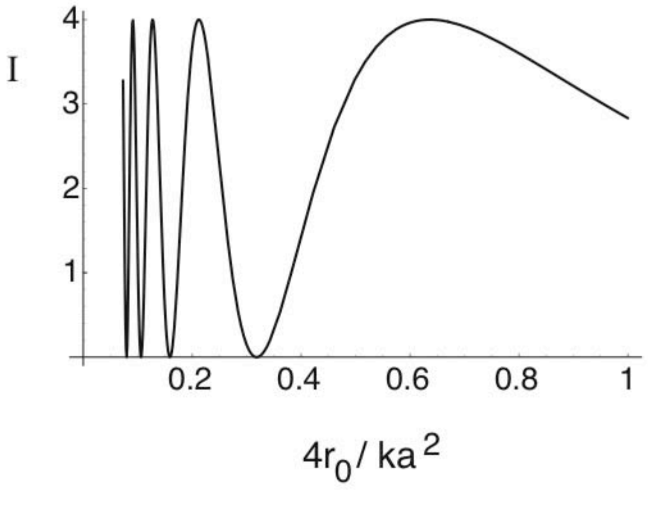

Fig. 5.10 Intensity along the axis of a circular aperture according to Fresnel diffraction theory.

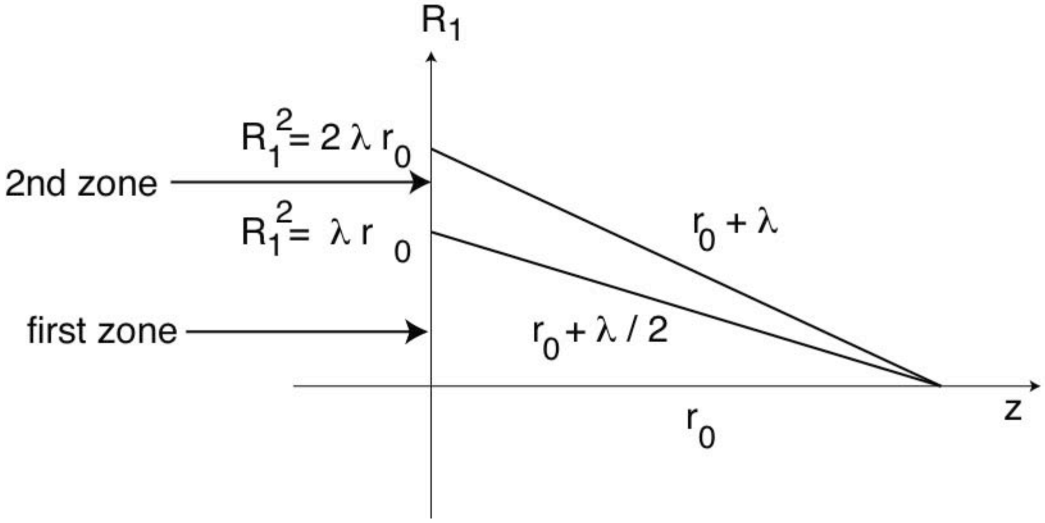

Fig. 5.11. The Fresnel half-period zone construction.

Note that for a general axially-symmetric U_1, along the axis the amplitude is

$$U_2(0,r_0) = -\frac{ik}{r_0}\exp(ikr_0)\int_0^a U_1(R_1)\exp\left(\frac{ikR_1^2}{2r_0}\right)R_1\,dR_1\,. \qquad (5.33)$$

Let $R_1^2/a^2 = t$, and put $U_1(R_1) = V_1(t)$. We have

$$2R_1dR_1 = a^2dt\,,$$

and thus

$$U_2(0,r_0) = -i\pi N\exp(ikr_0)\int_0^1 V_1(t)\exp(i\pi Nt)\,dt\,. \qquad (5.34)$$

So the intensity is given by

$$I_2(N) = (\pi N)^2\left|FT\left[V_1(t)\right]\right|^2. \qquad (5.35)$$

5.5.3. Rectangular Aperture

If a rectangular aperture, sides $2a$, $2b$, is illuminated by a plane wave, from Eq.5.28,

$$U_2\left(x_2,y_2\right)=-\frac{ie^{ikr_0}}{\lambda r_0}\int_{-a}^{+a}\exp\left(\frac{ikx_1^2}{2r_0}\right)\exp\left[-\frac{ik\left(x_1x_2\right)}{r_0}\right]dx_1$$

$$\times\int_{-b}^{+b}\exp\left(\frac{iky_1^2}{2r_0}\right)\exp\left[-\frac{ik\left(y_1y_2\right)}{r_0}\right]dy_1. \tag{5.36}$$

Now, consider the integral

$$I=\int_{-a}^{+a}\exp\left(\frac{ikx_1^2}{2r_0}\right)\exp\left[-\frac{ik\left(x_1x_2\right)}{r_0}\right]dx_1$$

$$=\int_{-a}^{+a}\exp\left[\frac{ik\left(x_1^2-2x_1x_2\right)}{2r_0}\right]dx_1. \tag{5.37}$$

By completing the square

$$I=\exp\left(-\frac{ikx_2^2}{2r_0}\right)\int_{-a}^{+a}\exp\left[\frac{ik\left(x_2-x_1\right)^2}{2r_0}\right]dx_1.$$

$$=\sqrt{\frac{r_0\lambda}{2}}\exp\left(-\frac{ikx_2^2}{2r_0}\right)\int_{-w_1}^{+w_1}\exp\left(\frac{i\pi w^2}{2}\right)dw, \tag{5.38}$$

where

$$w=\sqrt{\frac{2}{r_0\lambda}}\left(x_2-x_1\right). \tag{5.39}$$

We now introduce the Fresnel integrals, defined

$$\begin{aligned} C\left(w\right)&=\int_0^w\cos\left(\frac{\pi w'^2}{2}\right)dw' \\ S\left(w\right)&=\int_0^w\sin\left(\frac{\pi w'^2}{2}\right)dw' \quad . \\ F\left(w\right)&=C+iS \end{aligned} \tag{5.40}$$

So

$$I = 2\exp\left(-\frac{ikx_2^2}{2r_0}\right)\frac{F(w_2) - F(w_1)}{\sqrt{\frac{2}{r_0\lambda}}} \tag{5.41}$$

$$= 2\exp\left(-\frac{ikx_2^2}{2r_0}\right)\frac{F\left[\sqrt{\frac{2}{r_0\lambda}}(x_2 + a)\right] - F\left[\sqrt{\frac{2}{r_0\lambda}}(x_2 - a)\right]}{\sqrt{\frac{2}{r_0\lambda}}} \tag{5.42}$$

and thus

$$U_2(x_2, y_2) = -2i\exp(ikr_0)\exp\left(\frac{ikR_2^2}{2r_0}\right)\left[F(w_{x_2}) - F(w_{x_1})\right]\left[F(w_{y_2}) - F(w_{y_1})\right]. \tag{5.43}$$

The Fresnel integrals are tabulated. They have the important properties

$$\left.\begin{aligned} C(w) &= -C(-w) \\ S(w) &= -S(-w) \end{aligned}\right\} \text{ odd function }, \tag{5.44}$$

$$C(0) = S(0) = 0, \qquad F(0) = 0, \tag{5.45}$$

$$C(\infty) = S(\infty) = \frac{1}{2}, \qquad F(-\infty) = -\frac{1}{2}(1-i), \tag{5.46}$$

$$C(-\infty) = S(-\infty) = -\frac{1}{2}, \qquad F(-\infty) = -\frac{1}{2}(1-i). \tag{5.47}$$

Let us consider first what happens as the aperture becomes very large. Then from Eq.5.43, at $x_2 = y_2 = 0$

$$U_2(0,0) = -\frac{i}{2}\exp(ikr_0)(1+i)^2 = \exp(ikr_0). \tag{5.48}$$

So it reproduces exactly what is expected for a plane wave! This is remarkable, because the assumptions we have made such as the Fresnel approximation, and that $r \approx r_0$ breakdown for points on the aperture that are far from the axis. It works because the contributions from these off-axis points are weak because they are far away. So now it is established that the method can be applied even for an infinitely large aperture, this gives us some confidence so that we can look at some other cases.

5.5.4. Single Slit

Here $b \to \infty$ and we can take $y_2 = 0$ without loss of generality, so that

$$U_2(x_2) = \frac{(1-i)}{2}\exp(ikr_0)\exp\left(-\frac{ikx_2^2}{2r_0}\right)\left\{F\left[\sqrt{\frac{2}{r_0\lambda}}(x_2+a)\right] - F\left[\sqrt{\frac{2}{r_0\lambda}}(x_2-a)\right]\right\}. \qquad (5.49)$$

For slits we get intensity distributions such as those shown in Fig.5.12.

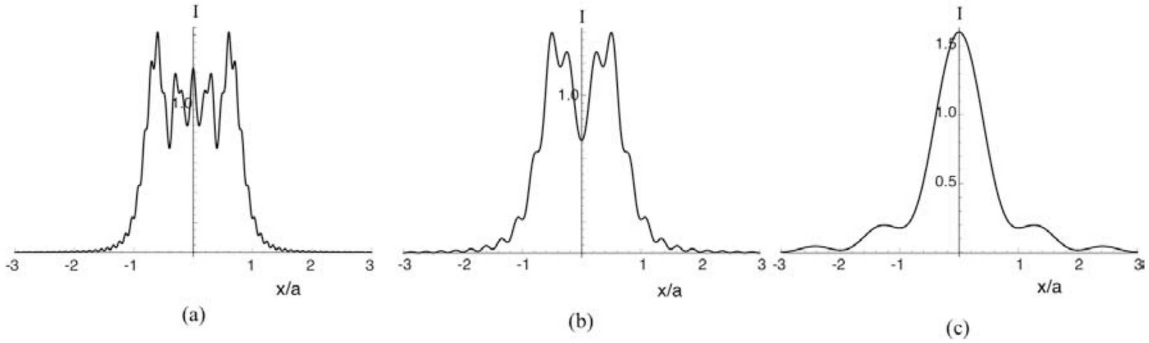

Fig. 5.12 Fresnel diffraction by a slit: (a) $N = a^2/\lambda r_0 = 5$, (b) $N = 2$, (c) $N = 0.5$.

Along the axis, $x_2 = 0$, and

$$U_2(0) = -(1+i)\exp(ikr_0)F\left(\sqrt{\frac{2}{r_0\lambda}}a\right). \qquad (5.50)$$

The intensity along the axis (Fig. 5.13) should be compared with that for a circular aperture. It should be noted that the minima are not zeros this time. The maxima correspond approximately to the case when the Fresnel number $N = a^2/\lambda r_0$ is

$$N = 2n + \tfrac{3}{4}, \qquad (5.51)$$

and the minima approximately to the case when

$$N = 2n + \tfrac{5}{4}, \qquad (5.52)$$

where n is zero or a positive integer.

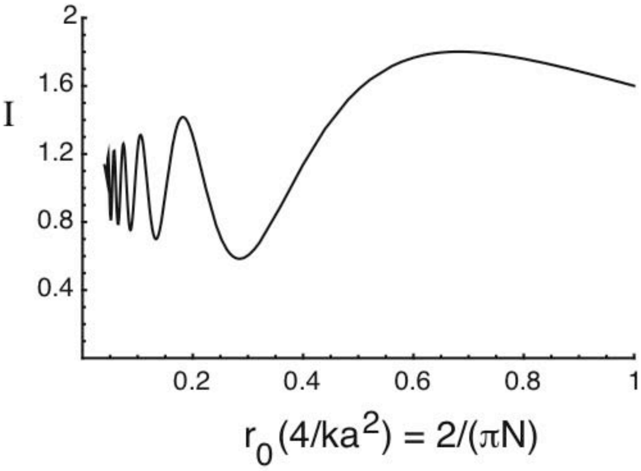

Fig. 5.13 Intensity along the axis for a slit aperture according to Fresnel diffraction.

5.5.5. Half-plane

For diffraction by an edge, we can put in Eq.5.42

$$w_2 = \sqrt{\frac{2}{r_0\lambda}}x_2,\, w_1 \to \infty\,, \tag{5.53}$$

so that the intensity is

$$I = \frac{1}{2}\left\{\left[\frac{1}{2} - C\left(\sqrt{\frac{2}{r_0\lambda}}x_2\right)\right]^2 + \left[\frac{1}{2} - S\left(\sqrt{\frac{2}{r_0\lambda}}x_2\right)\right]^2\right\}. \tag{5.54}$$

Note that the diffraction pattern is the same at any distance, but of course it scales with distance (Fig. 5.14). This is in contrast to the diffraction pattern for a slit, which changes with distance.

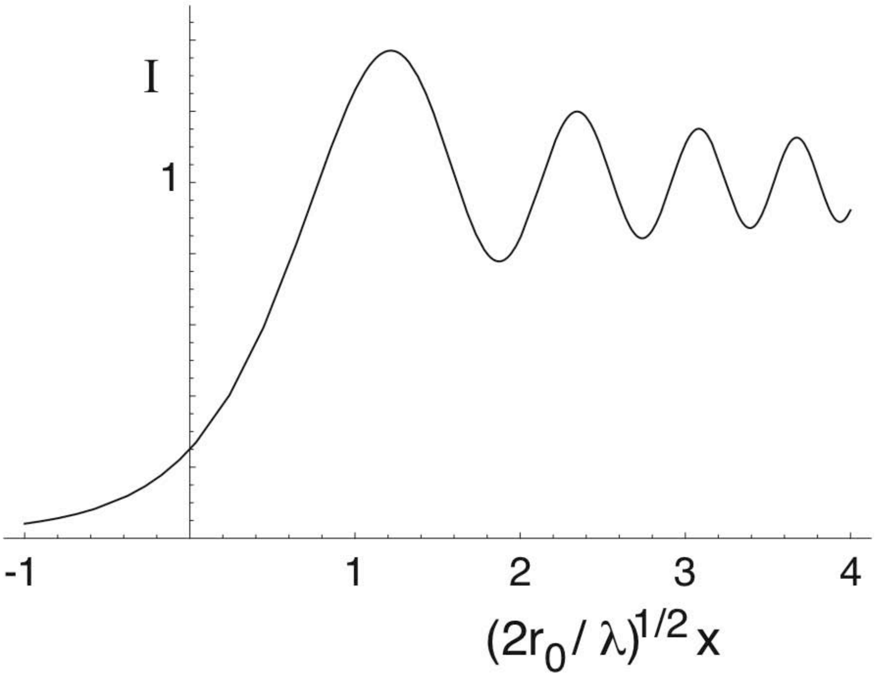

Fig. 5.14 Fresnel diffraction by an edge.

5.5.6. Circular Obstruction

Consider a circular obstruction, radius *a*. Then

$$
\begin{aligned}
U_2(0,r_0) &= \frac{ik}{r_0}\exp(ikr_0)\int_a^\infty \exp\left(-\frac{ikR_1^2}{2r_0}\right)R_1 dR_1 \\
&= \frac{ik}{r_0}\exp(ikr_0)\int_0^\infty \exp\left(-\frac{ikR_1^2}{2r_0}\right)R_1 dR_1 - \frac{ik}{r_0}\exp(ikr_0)\int_0^a \exp\left(-\frac{ikR_1^2}{2r_0}\right)R_1 dR_1 .
\end{aligned}
\tag{5.55}
$$

The first term represents a plane wave, and the second the diffracted field for a circular aperture, rather than an obstruction. This is a particular case of Babinet's principle, which states that the sum of the fields for two complementary screens is equal to the unobstructed disturbance.

The first term we evaluated in Eq.5.48: it gave just $\exp(ikr_0)$. So

$$
\begin{aligned}
U_2 &= -\exp(ikr_0) + \left\{\exp(ikr_0)\left[\exp\left(\frac{ika^2}{2r_0}\right)+1\right]\right\} \\
&= \exp(ikr_0)\exp\left(\frac{ika^2}{2r_0}\right).
\end{aligned}
\tag{5.56}
$$

The intensity is thus constant along the axis (the approximations break down when we get too close to the obstruction). This is the Poisson (or Arago) spot. It can be regarded as being caused by light scattered from the edge of the disc. Comparing with the case of the circular aperture, Eq. 5.30 represents interference between the edge-diffracted wave and the undiffracted wave ($R_1 = 0$). This is the principle of the boundary diffraction wave concept introduction by Young.

5.6. Kirchhoff Diffraction Integral

We now return to the problem of deriving the diffraction integral starting from the wave equation

$$\left(\nabla^2 + k^2\right)U = 0. \tag{5.57}$$

Consider a closed surface *S* with inward normal **n** (Fig. 5.15)**.** Then we can show using Green's theorem that at any point inside *S*

$$\iint_S \left(U \frac{\partial U'}{\partial n} - U' \frac{\partial U}{\partial n} \right) dS = 0 , \tag{5.58}$$

for any U' which satisfies also the wave equation. We can get different solutions by appropriate choice of the so-called Green function U' that satisfies the wave equation.

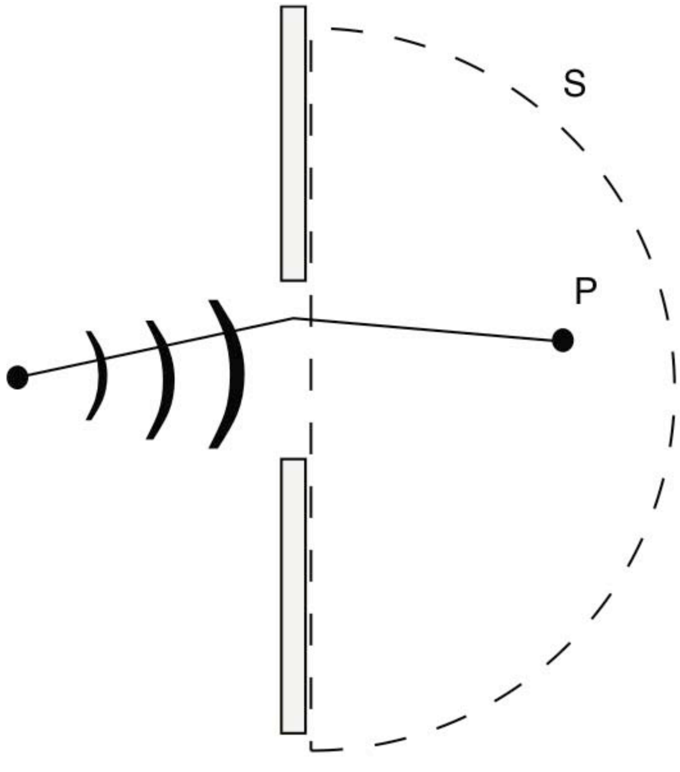

Fig. 5.15 Diffraction according to Kirchhoff diffraction theory.

If we choose $U' = \exp(ikr)/r$, which satisfies the wave equation, and excluding the region around its singularity at $r = 0$, we obtain the Kirchhoff diffraction integral:

$$U_p = \frac{1}{4\pi} \iint \left[U \frac{\partial}{\partial n}\left(\frac{e^{ikr}}{r} \right) - \left(\frac{e^{ikr}}{r} \right) \frac{\partial U}{\partial n} \right] dS . \tag{5.59}$$

It must be stressed that this expression is rigorously correct (in the scalar approximation).

Consider now diffraction by a planar screen with an aperture illuminated with a wave emanating from point *S*. We assume that the field in the aperture is the same as in the screen were absent. So in the aperture

$$U = \frac{U_0 e^{iks}}{s}, \frac{\partial U}{\partial n} = \frac{U_0 e^{iks}}{s}\left(ik - \frac{1}{s} \right) \cos(n,s) , \tag{5.60}$$

where $\cos(n,s)$ is the cosine of the angle between the **n** and **s** directions, whereas on the screen

$$U = 0, \frac{\partial U}{\partial n} = 0. \tag{5.61}$$

These are called the Kirchhoff boundary conditions.

So, neglecting $1/s$ and $1/r$ in comparison to *k*,

$$U(P) = -\frac{iU_0}{\lambda} \iint_{\text{Aperture}} \frac{e^{ik(r+s)}}{rs} \left[\frac{\cos(n,\text{s}) - \cos(n,r)}{2} \right] dS. \tag{5.62}$$

Here we have assumed that the contribution from the spherical part of the surface *S* vanishes as its radius is made increasingly large. The factor in square brackets is called the obliquity factor. Apart from this, the expression is identical to Eq. 5.22. Eq. 5.62 is the Kirchhoff diffraction formula. Note that the surface of integration is arbitrary.

Although the Kirchhoff diffraction integral is exact, the assumed boundary conditions are not. So the Kirchhoff diffraction formula is also not exact. In particular it fails to reproduce the assumed field in the aperture. This is because the choice of *U* and $\partial U / \partial n$ in the aperture is inconsistent.

It can be shown that in fact, for the special case of integration over a planar region,

$$U_p = \frac{1}{2\pi}\iint U\frac{\partial}{\partial n}\left(\frac{e^{ikr}}{r}\right)dS \tag{5.63}$$

and

$$U_p = -\frac{1}{2\pi}\iint\left(\frac{e^{ikr}}{r}\right)\frac{\partial U}{\partial n}dS \tag{5.64}$$

These are called the Rayleigh-Sommerfeld I and II diffraction formulae. They have the advantage of being consistent as we need only assume U or $\partial U/\partial n$ in the aperture, but are not in practice any more accurate as we do not usually know either U or $\partial U/\partial n$ accurately. In fact, Eq. 5.59 can be thought of as the average of Eq.5.63 and Eq.5.64 and has been claimed to give a better prediction of the field at P.

5.7. Angular spectrum of plane waves

Consider a single plane wave propagating at angle θ, for which

$$E = E_0\exp(-ikx\sin\theta)\exp(-ikz\cos\theta). \tag{5.65}$$

In the plane $z=0$

$$E = E_0\exp(-ikx\sin\theta),$$

which is a harmonic variation with spatial wavelength $\lambda/\sin\theta$. So if we alter θ, we alter the spatial wavelength in the x direction. According to Fourier synthesis, *any* field in the plane $z=0$ can be represented by a sum of Fourier components:

$$E_y(x) = \int_{-\infty}^{+\infty} G(\theta)\exp(-ikx\sin\theta)d(\sin\theta). \tag{5.66}$$

$G(\theta)$ is the strength of a plane wave component travelling in direction θ. Here $G(\theta)$ is *complex* to account for the relative phase of the components. As a simple example, for the field

$$\begin{aligned}E_y(x) &= A\cos\left(\frac{2\pi x}{\Lambda}\right)\\ &= \frac{A}{2}\exp\left(\frac{2\pi ix}{\Lambda}\right)+\frac{A}{2}\exp\left(-\frac{2\pi ix}{\Lambda}\right),\end{aligned} \tag{5.67}$$

we have immediately $\sin\theta = \pm\lambda/\Lambda$, representing two plane waves In this case, the diffraction pattern of the cosine grating in the far field is two bright spots.

5.8. Evanescent waves

Note that

$$E(x) = \int_{-\infty}^{+\infty} G(\theta)\exp(-ikx\sin\theta)d(\sin\theta) \tag{5.68}$$

has limits ±∞. But $|\sin\theta| \le 1$ for θ to be real. So $|\sin\theta| > 1$ corresponds to waves travelling at a complex angle. We have

.

$$k_x^2 + k_z^2 = k^2, \tag{5.69}$$

so that

$$k_z^2 = k^2 - k_x^2. \tag{5.70}$$

If $k_x > k$, we can put

$$k_z = \pm i\sqrt{k_x^2 - k^2},$$

which is imaginary. Therefore the electric field is

$$E = E_0 \exp(-ik_x x)\exp\left[\pm\left(z\sqrt{k_x^2 - k^2}\right)\right], \tag{5.71}$$

representing a wave travelling in the *x* direction, but with an exponential decay in the *z* direction. Note that we take the positive root to give a physical solution for z ≥ 0.

This is an *evanescent* wave. The integral integrates over all propagating *and evanescent waves*. In the far field, the evanescent waves have decayed, so they make no contribution: only the propagating waves remain.

5.9 Diffraction by a phase screen

Suppose we have a screen that has an amplitude transmittance $t(x,y)$. This is complex, to account for amplitude and phase effects. Then if it is illuminated by a plane wave, the field immediately after the screen is $U_0 t$.

A thin lens can be thought of as such a screen. It consists of an amplitude term $P(x_1,y_1)$ that is called the pupil function of the lens, which is unity in the aperture and zero outside, and a phase term $e^{i\Phi(x,y)}$.

$$t(x_1,y_1) = P(x_1,y_1)e^{i\Phi(x,y)}, \tag{5.72}$$

or in polar coordinates

$$t\left(r_1,\theta_1\right)=P\left(r_1,\theta_1\right)e^{i\Phi\left(r_1,\theta_1\right)}. \tag{5.73}$$

In general we can expand Φ as a power series in r and also as a series in $\cos n\theta$:

$$\begin{aligned}\Phi &= \sum_{n,m} a_{mn} r^m \cos n\theta \\ &= a_{00}+a_{20}r^2+a_{40}r^4+\ldots+a_{11}r\cos\theta+a_{21}r^2\cos\theta+\ldots+a_{22}r^2\cos 2\theta+\ldots,\end{aligned} \tag{5.74}$$

or, collecting all the terms except the squared term,

$$\Phi = a_{20}r^2+\Phi'\left(r,\theta\right). \tag{5.75}$$

So the amplitude transmittance of the lens can be written

$$t\left(r_1,\theta_1\right)=P\left(r_1\right)\exp\left(ia_{20}r_1^2\right)e^{i\Phi'}. \tag{5.76}$$

After the lens, neglecting diffraction for the present,

$$\begin{aligned}U\left(r\right) &= P\left(r\right)\exp\left(ia_{20}r^2\right)e^{i\Phi'}e^{ikz} \\ &= P\left(r_1\right)e^{i\Phi'}\exp\left[ik\left(\frac{a_{20}r^2}{k}+z\right)\right].\end{aligned} \tag{5.77}$$

Neglecting Φ', which represents aberrations, the phase front through the origin is the paraboloid with equation (Fig. 5.16)

$$\frac{a_{20}r^2}{k}+z=0. \tag{5.78}$$

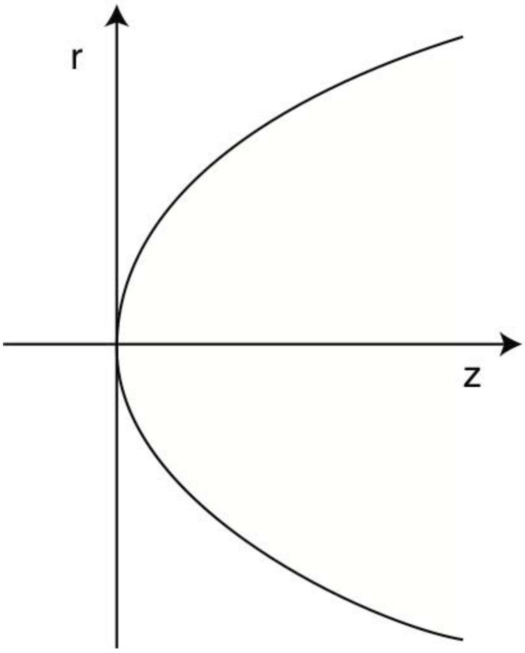

Fig. 5.16 Phase front of a lens.

Consider a sphere, radius *f*, centred on $z = f$. Then

$$r^2 + (z-f)^2 = f^2,$$

and using the binomial and assuming $r << f$

$$(z-f) = \sqrt{f^2 - r^2} = \pm f\left(1 - \frac{r^2}{2f}\right). \tag{5.79}$$

Taking the negative root and comparing Eq.5.78 and Eq.5.79

$$\frac{a_{20}}{k} = -\frac{1}{2f}. \tag{5.80}$$

So the lens, including the aberration term, can be taken as

$$t(r_1,\theta_1) = P(r_1)\exp\left(-\frac{ikr_1^2}{2f}\right)e^{i\Phi'}, \tag{5.81}$$

where *f* is the focal length of the lens.

5.10. Thin Lens

For a thin lens, the field after the lens is (Goodman 1968)

$$U_2(x,y) = U_1(x,y)\,P(x,y)\exp\left[ik(n-1)\Delta(x,y)\right] \tag{5.82}$$

where Δ is the thickness of the glass, n is its refractive index (Fig. 5.17). In the paraxial approximation, $x / R_1 << 1$ and higher powers can be neglected, so

$$\Delta(x,y) = \Delta_0 - \frac{x^2+y^2}{2}\left(\frac{1}{R_1} - \frac{1}{R_2}\right). \tag{5.83}$$

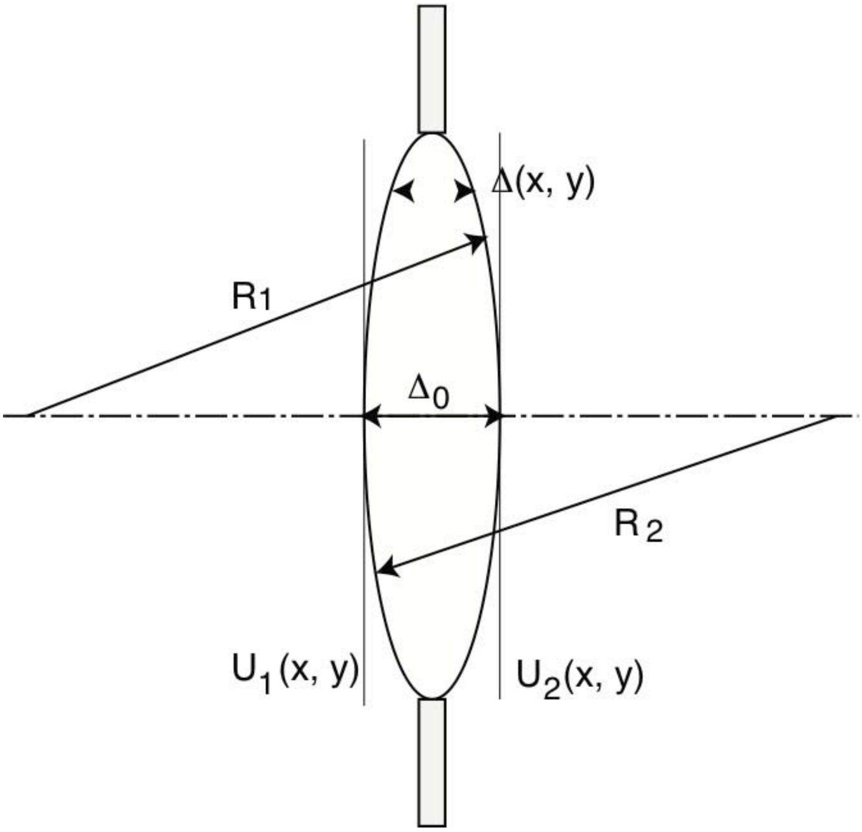

Fig. 5.17 The thin lens.

Thus

$$U_2(x,y) = U_1(x,y)P(x,y)\exp\left[ik(n-1)\Delta_0\right]\exp\left[-ik(n-1)\left(\frac{x^2+y^2}{2}\right)\left(\frac{1}{R_1} - \frac{1}{R_2}\right)\right]. \tag{5.84}$$

Putting

$$\frac{1}{f} = (n-1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right), \tag{5.85}$$

we then have

$$U_2(x,y) = U_1(x,y)P(x,y)\exp[ik(n-1)\Delta_0]\exp\left[-\frac{ik}{2f}\left(x^2+y^2\right)\right]. \qquad (5.86)$$

The term $\exp[ik(n-1)\Delta_0]$ is a constant phase term, which we neglect.

5.11. Focus of a Lens

We assume the lens is illuminated by a plane wave so that

$$U_2(x,y) = P(x,y)\exp\left[-\frac{ik\left(x^2+y^2\right)}{2f}\right]. \qquad (5.87)$$

This represents a spherical wave convergent on the point F (Fig. 5.18). But

$$U_3(x_3,y_3) = -\frac{ie^{ikf}}{\lambda f}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} U_2(x,y)\exp\left\{\frac{ik}{2f}\left[(x_3-x)^2+(y_3-y)^2\right]\right\}dx\,dy$$

$$(5.88)$$

as

$$r^2 = f^2 + (x_3-x)^2 + (y_3-y)^2,$$

so that if $(x_3-x) << f$.

$$r \approx f + \frac{(x_3-x)^2}{2f} + \frac{(y_3-y)^2}{2f}.$$

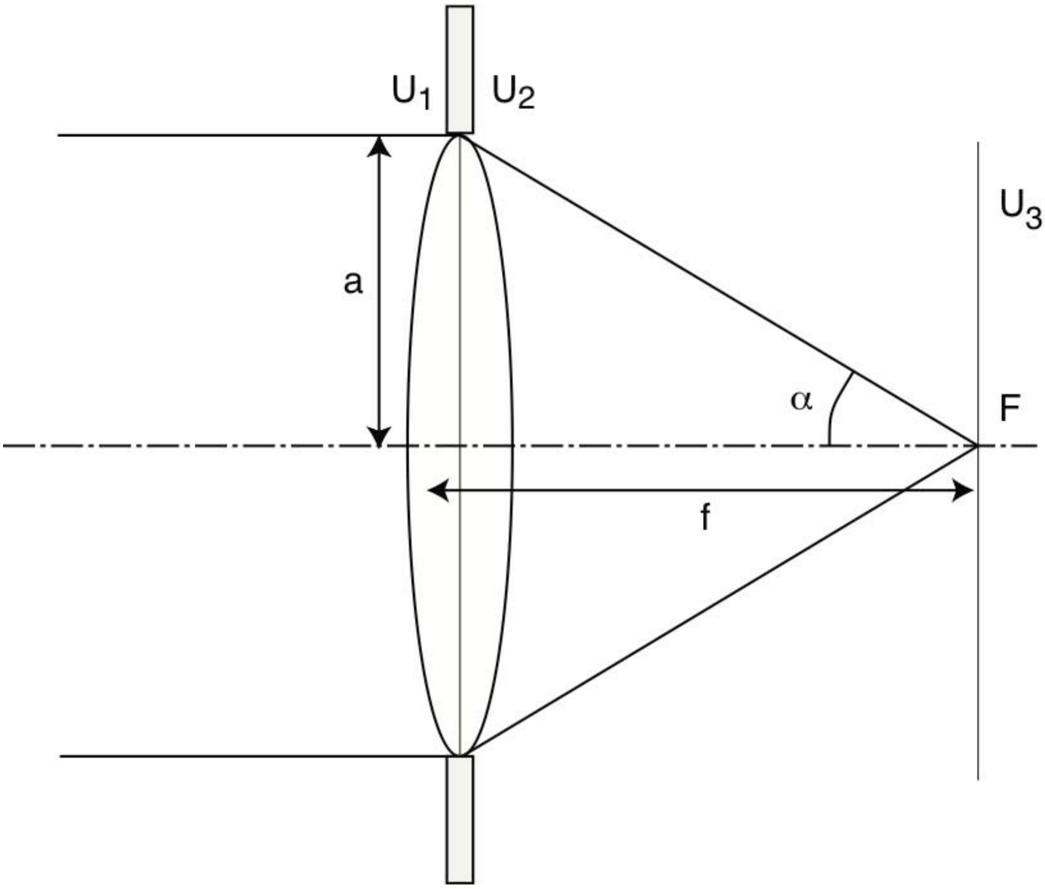

Fig. 5.18 Focusing by a lens.

From now on we shall not write explicitly the limits ±∞ on the integrals. The field after the lens is

$$U_3(x_3,y_3) = -\frac{ie^{ikf}}{\lambda f}\iint P(x,y)\exp\left[-\frac{ik\left(x^2+y^2\right)}{2f}\right]\exp\left[\frac{ik\left(x^2+y^2\right)}{2f}\right]$$

$$\times\exp\left[\frac{ik\left(x_3^2+y_3^2\right)}{2f}\right]\exp\left[-\frac{ik}{f}\left(xx_3+yy_3\right)\right]dx\,dy$$

$$= -\frac{ie^{ikf}}{\lambda f}\exp\left[\frac{ik\left(x_3^2+y_3^2\right)}{2f}\right]\iint P(x,y)\exp\left[-\frac{ik}{f}\left(xx_3+yy_3\right)\right]dx\,dy. \tag{5.89}$$

We now consider some important special cases for the pupil function.

5.12. Circularly Symmetric Aperture

Most real optical systems have circular symmetry, so we can put

$$P(x,y) = P(r)$$

and

$$U_3(r_3) = -\frac{ie^{ikf}}{\lambda f}\exp\left(\frac{i\pi r_3^2}{\lambda f}\right)\int_0^\infty P(r)J_0\left(\frac{2\pi r r_3}{\lambda f}\right)2\pi r dr\,. \tag{5.90}$$

In particular, for a circular aperture of radius *a*

$$U_3(r_3) = -\frac{ie^{ikf}}{\lambda f}\exp\left(\frac{i\pi r_3{}^2}{\lambda f}\right)\pi a^2\left[\frac{2J_1\left(\frac{2\pi\, r_3 a}{\lambda f}\right)}{\frac{2\pi\, r_3 a}{\lambda f}}\right]. \tag{5.91}$$

Let us define the *numerical aperture* of the lens:

$$NA = \sin\alpha = \frac{a}{f}\,. \tag{5.92}$$

We define a normalized optical coordinate

$$v = kr_3\sin\alpha \approx \frac{2\pi\, r_3 a}{\lambda f}. \tag{5.93}$$

and also put

$$\rho = \frac{r}{a}. \tag{5.94}$$

Then in general

$$U_3(v) = -\frac{ie^{ikf}}{\lambda f}\exp\left(\frac{iv^2\lambda f}{4\pi a^2}\right)2\pi a^2\int_0^1 P(\rho)J_0(v\rho)\rho d\rho\,, \tag{5.95}$$

or introducing the Fresnel number

$$N = \frac{a^2}{\lambda f}, \tag{5.96}$$

and the field is

$$U_3(v) = -i\pi Ne^{ikf}\exp\left(\frac{iv^2}{4\pi N}\right)\int_0^1 2P(\rho)J_0(v\rho)\rho d\rho\,. \tag{5.97}$$

For a plain circular aperture

$$U_3(v) = -i\pi N e^{ikf} \exp\left(\frac{iv^2}{4\pi N}\right)\left[\frac{2J_1(v)}{v}\right]. \tag{5.98}$$

This is called the amplitude point spread function or impulse response. Note that

$$N = \frac{a^2}{\lambda f} = \frac{a \sin\alpha}{\lambda} = \frac{f \sin^2\alpha}{\lambda},$$

i.e. for a big lens (compared with the wavelength), N is very large and the exponential term is close to unity. The intensity is the modulus squared of the amplitude, giving the Airy disc:

$$I(v) = \pi^2 N^2 \left[\frac{2J_1(v)}{v}\right]^2. \tag{5.99}$$

The intensity variation is as shown in Fig. 5.7. In Fig.5.9 the effect of a central obstruction is also shown. As the obstruction ratio ε increases, the point spread function becomes narrower, but with larger side-lobes.

5.13. Effect of defocus

We now consider the field on a defocused plane, $z = f + \delta z$ (Fig. 5.19). The field is

$$U_3(x_3,r_3) = -\frac{ie^{ikz}}{\lambda z}\iint P(x,y)\exp\left[-\frac{ik}{2f}\left(x^2+y^2\right)\right]\exp\left\{\frac{ik}{2z}\left[\left(x_3-x\right)^2+\left(y_3-y\right)^2\right]\right\}dxdy$$

$$= -\frac{ie^{ikz}}{\lambda z}\exp\left[\frac{ik}{2z}\left({x_3}^2+{y_3}^2\right)\right]\iint P(x,y)\exp\left[-\frac{ik}{2f}\left(x^2+y^2\right)\right]$$

$$\times\exp\left[\frac{ik}{2z}\left(x^2+y^2\right)\right]\exp\left[-\frac{2\pi i}{\lambda z}\left(xx_3+yy_3\right)\right]dxdy.$$

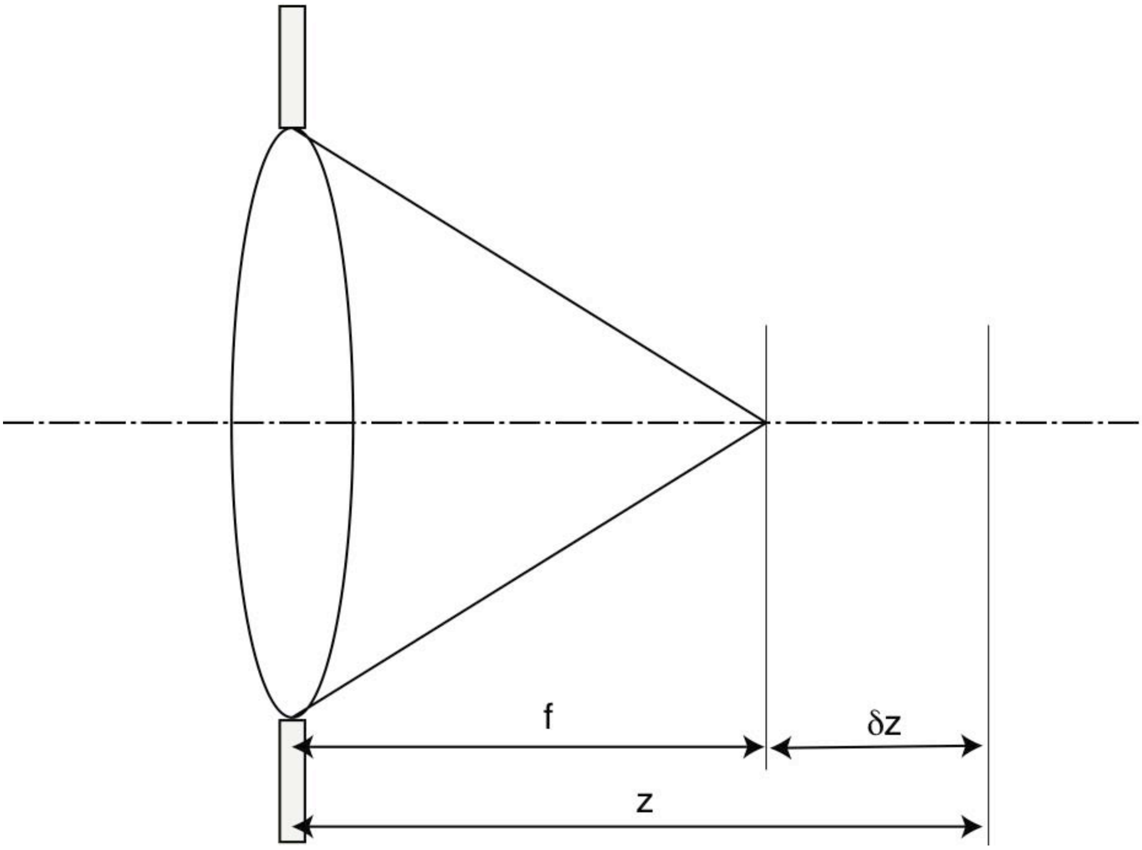

Fig. 5.19 Geometry of defocus.

For a radially symmetric system

$$U_3\left(r_3\right)=-\frac{ie^{ikz}}{\lambda z}\exp\left(\frac{i\pi r_3^2}{\lambda z}\right)\int P\left(r\right)\exp\left[-\frac{ikr^2}{2}\left(\frac{1}{f}-\frac{1}{z}\right)\right]J_0\left(\frac{2\pi rr_3}{\lambda z}\right)2\pi rdr\ . \tag{5.100}$$

For a circular aperture, again of radius a, put

$$v=\frac{2\pi ar_3}{\lambda z}=\frac{2\pi ar_3}{\lambda f} \tag{5.101}$$

$$u=\frac{2\pi a^2}{\lambda}\left(\frac{1}{f}-\frac{1}{z}\right)\approx\frac{2\pi a^2}{\lambda f^2}\delta z \tag{5.102}$$

if $\delta z << f$. So

$$U_3\left(v,u\right)=-i\pi Ne^{ikz}\exp\left(\frac{iv^2}{4\pi N}\right)\int_0^1 2J_0\left(v\rho\right)\exp\left(-\frac{iu\rho^2}{2}\right)\rho d\rho\,. \tag{5.103}$$

Along the axis, $v=0$, the field is

$$U_3(v,u) = -i\pi N e^{ikz} \exp\left(\frac{iv^2}{4\pi N}\right) \int_0^1 2\exp\left(-\frac{iu\rho^2}{2}\right)\rho d\rho$$

$$= -i\pi N e^{ikz} \exp\left(\frac{iv^2}{4\pi N}\right)\left[-\frac{2}{iu}\exp\left(-\frac{iu\rho^2}{2}\right)\right]_0^1 \qquad (5.104)$$

$$= -i\pi N e^{ikz} \exp\left(-\frac{iu}{4}\right)\left[\frac{\sin\left(\frac{u}{4}\right)}{\frac{u}{4}}\right].$$

So the intensity along the axis is

$$I(0,u) = \pi^2 N^2 \left[\frac{\sin\left(\frac{u}{4}\right)}{\frac{u}{4}}\right]^2, \qquad (5.105)$$

.

The field at a general point can be calculated from Lommel functions or by numerical integration.

It is interesting to consider also the *annular* aperture $P(\rho) = \delta(\rho - 1)$. Then from Eq. 5.100

$$U_3(v,u) = -2i\pi N e^{ikz} \exp\left(\frac{iv^2}{4\pi N}\right)\exp\left(-\frac{iu}{2}\right)J_0(v) \qquad (5.106)$$

or

$$I(v,u) = 4\pi^2 N^2 J_0^2(v). \qquad (5.107)$$

Note that the intensity does not change with *u* (within the range of validity of the equation). This represents a Bessel beam (a so-called diffraction-free beam), which is the subject of much research at present. Actually it is very well known as a mode of free space in cylindrical coordinates, for example, of a circular waveguide. Power diffracts outwards, but also *inwards* from the large side-lobes, to achieve a dynamic equilibrium.

5.14. Image formation

Before the lens (Fig. 5.20)

$$U_2(x_2,y_2) = -\frac{ie^{ikd_1}}{\lambda d_1}\iint U_1(x_1,y_1)\exp\left\{\frac{ik}{2d_1}\left[(x_2-x_1)^2+(y_2-y_1)^2\right]\right\}dx_1\,dy_1\,.$$

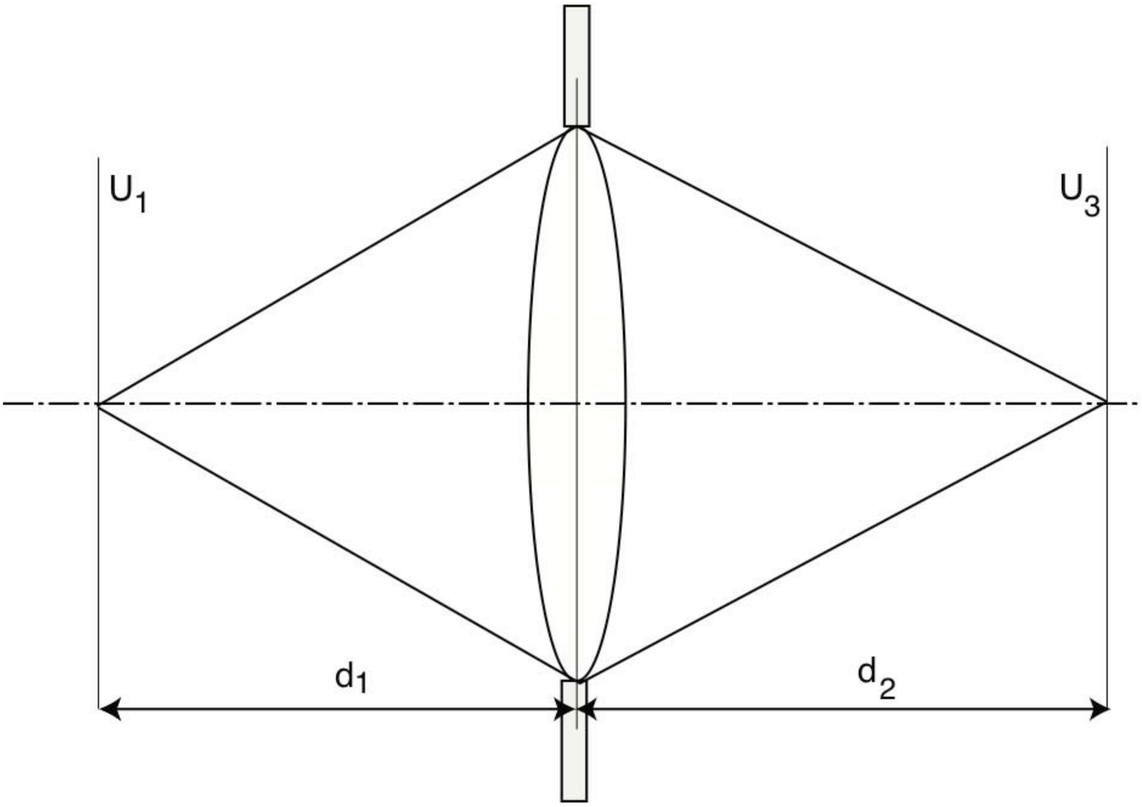

Fig. 5.20 Imaging by a lens.

Multiplying by the pupil function, the amplitude after the lens is

$$U_2(x_2,y_2) = -\frac{ie^{ikd_1}}{\lambda d_1}P(x_2,y_2)\exp\left[-\frac{ik}{2f}\left(x_2^2+y_2^2\right)\right]$$

$$\iint U_1(x_1,y_1)\exp\left\{\frac{ik}{2d_1}\left[(x_2-x_1)^2+(y_2-y_1)^2\right]\right\}dx_1\,dy_1.$$

So finally the image amplitude is

$$
\begin{aligned}
U_3(x_3,y_3) = & -\frac{1}{\lambda^2 d_1 d_2}\exp[ik(d_1+d_2)]\int\int\int\int P(x_2,y_2)U_1(x_1,y_1) \\
& \times\exp\left\{\frac{ik}{2d_1}\left[(x_2-x_1)^2+(y_2-y_1)^2\right]\right\} \\
& \times\exp\left\{\frac{ik}{2d_2}\left[(x_3-x_2)^2+(y_3-y_2)^2\right]\right\} \\
& \times\exp\left[-\frac{ik}{2f}\left(x_2^2+y_2^2\right)\right]dx_1dy_1dx_2dy_2. \\
= & -\frac{1}{\lambda^2 d_1 d_2}\exp[ik(d_1+d_2)]\int\int\int\int P(x_2,y_2)U_1(x_1,y_1) \\
& \times\exp\left[\frac{ik}{2d_1}\left(x_1^2+y_1^2\right)\right]\exp\left[\frac{ik}{2d_2}\left(x_3^2+y_3^2\right)\right] \\
& \times\exp\left[\frac{ik}{2}\left(\frac{1}{d_1}+\frac{1}{d_2}-\frac{1}{f}\right)\left(x_2^2+y_2^2\right)\right] \\
& \times\exp\left\{-ik\left[x_2\left(\frac{x_1}{d_1}+\frac{x_3}{d_2}\right)+y_2\left(\frac{y_1}{d_1}+\frac{y_3}{d_2}\right)\right]\right\}dx_1dy_1dx_2dy_2.
\end{aligned}
\tag{5.108}
$$

We now look at some special cases.

5.14.1 Special cases

The first case that we consider is when the lens law is satisfied, $1/d_1+1/d_2=1/f$. Then, $d_2=Md_1$, where M is magnification. The field is

$$U_3(x_3,y_3) = -\frac{1}{\lambda^2 M d_1}\exp\left[ikd_1(1+M)\right]\int\int\int\int P(x_2,y_2)U_1(x_1,y_1)$$

$$\times\exp\left[\frac{ik}{2d_1}\left(x_1^2+y_1^2\right)\right]\exp\left[\frac{ik}{2Md_1}\left(x_3^2+y_3^2\right)\right]$$

$$\times\exp\left\{-\frac{ik}{d_1}\left[x_2\left(x_1+\frac{x_3}{M}\right)+y_2\left(y_1+\frac{y_3}{M}\right)\right]\right\}dx_1dy_1dx_2dy_2. \qquad (5.109)$$

Performing the integrals in terms of x_2,y_2 we have

$$h(x,y) = \iint P(x_2,y_2)\exp\left[\frac{ik}{d_1}(x_2x+y_2y)\right]dx_2dy_2, \qquad (5.110)$$

which is the point spread function, given by the Fourier transform of the pupil function. Then

$$U_3(x_3,y_3) = -\frac{1}{\lambda^2 M d_1^2}\exp\left[ikd_1(1+M)\right]\exp\left[\frac{ik}{2Md_1}\left(x_3^2+y_3^2\right)\right]\iint U_1(x_1,y_1)$$

$$\times\exp\left[\frac{ik}{d_1}\left(x_1^2+y_1^2\right)\right]h\left(x_1+\frac{x_3}{M},y_1+\frac{y_3}{M}\right)dx_1dy_1. \qquad (5.111)$$

Now, for good imaging h falls off quickly, i.e. x_1+x_3/M is small, or $x_1\approx -x_3/M$, so that (Goodman 1968)

$$U_3(x_3,y_3) = -\frac{1}{\lambda^2 M d_1^2}\exp\left[ikd_1(1+M)\right]\exp\left[-\frac{ik}{2Md_1}\left(x_3^2+y_3^2\right)\left(1+\frac{1}{M}\right)\right]$$

$$\times\iint U_1(x_1,y_1)h\left(x_1+\frac{x_3}{M},y_1+\frac{y_3}{M}\right)dx_1dy_1.$$

(5.112)

The intensity can thus be written

$$I_3(x_3,y_3) = \frac{1}{\left(\lambda^2 M d_1^2\right)^2}\left|\left(U_1\otimes h\right)\right|^2, \qquad (5.113)$$

where ⊗ represents the convolution operation. We can show that this is also valid with defocus: then h is then the defocused point spread function and

$$u = \frac{2\pi a^2}{\lambda}\left[\frac{1}{f} - \left(\frac{1}{d_1} + \frac{1}{d_2}\right)\right]. \tag{5.114}$$

So for coherent imaging, for an object $t(x,y)$, the intensity in the image is

$$I_3(x_3,y_3) = \frac{1}{\left(\lambda^2 M d_1^2\right)^2}\left|(t \otimes h)\right|^2. \tag{5.115}$$

Imaging is *linear* in amplitude, and space invariant that is the convolution means that each point of the object results in a distribution of *amplitude* in the image given by the *amplitude* point spread function. Finally, the *intensity* in the image is given by finding the modulus squared of the amplitude.

Note that Eq.(5.112) shows that the image of a point (x_1,y_1) in the object occurs at a point $(x_1 + x_3 / M = 0, y_1 + y_3 / M = 0)$, the center of the point spread function *h*. That is at $x_3 = -Mx_1$, $y_3 = -My_1$: thus the image is *inverted* and magnified by a factor *M*.

Next we consider the defocused case, $1/d_1 + 1/d_2 \neq 1/f$. We take

$$\frac{1}{d_1} + \frac{1}{d_2} - \frac{1}{f} = \frac{1}{d_0}. \tag{5.116}$$

From Eq.5.116 we can see that Eq.5.117 is still valid if we put $P_{eff}(x_2,y_2)$, given by

$$P_{eff}(x_2,y_2) = P(x_2,y_2)\exp\left[-\frac{ik}{2d_0}\left(x_2^{\ 2} + y_2^{\ 2}\right)\right]. \tag{5.118}$$

This effective pupil function is called the defocused pupil function. It is a complex quantity, given by multiplying the ordinary pupil function by a quadratic phase variation. Eq.5.118 is only true for small defocus, as the approximation in Eq.5.112 is otherwise not valid. This is because the point spread function becomes broader with defocus, i.e. the *intensity* point spread function behaves as shown in Fig. 5.21.

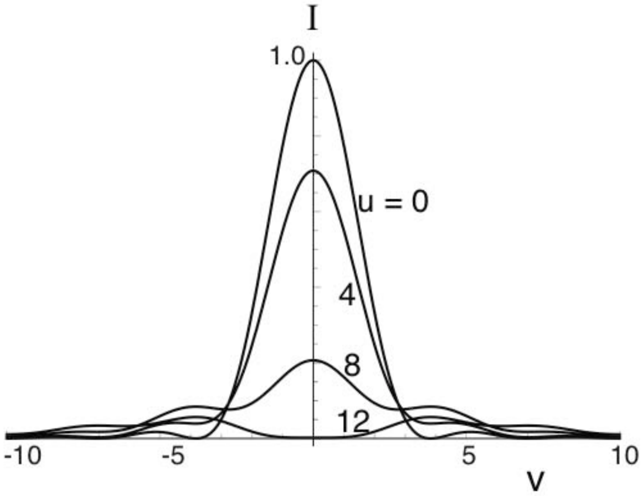

Fig. 5.21 Defocused image of a point.

As the system is defocused, the peak intensity decreases, and the pattern spreads out. By conservation of energy the total energy in the pattern, $\iint |h(x,y)|^2 dxdy$, must be constant. The zeros in the pattern also disappear with defocus. Note that the *amplitude* in the point spread function is complex for the defocused case.

Next we now look at an example when defocus is not small. We consider the special case when $d_1 = d_2 = f$. So

$$\frac{1}{d_0} = \frac{1}{d_1} + \frac{1}{d_2} - \frac{1}{f} = \frac{1}{f}, \tag{5.118}$$

and M = 1. To solve this case we return to Eq.5.108. As it stands the expression does not simplify much! The reason is that there is Fresnel diffraction by the object, which is then truncated by the pupil (Fig. 5.22). So it is quite a complicated problem. However, it is soluble if we consider the pupil to be very big, so there is negligible truncation. Then Eq.5.108 becomes

$$\begin{aligned}U_3(x_3,y_3) = &-\frac{1}{\lambda^2 f^2}\exp(ikf)\int\int\int\int U_1(x_1,y_1)\\ &\times\exp\left[\frac{ik}{2f}\left(x_1^2+y_1^2\right)\right]\exp\left[\frac{ik}{2f}\left(x_3^2+y_3^2\right)\right]\exp\left[-\frac{ik}{2f}\left(x_2^2+y_2^2\right)\right]\\ &\times\exp\left\{-\frac{ik}{f}\left[x_2(x_1+x_3)+y_2(y_1+y_3)\right]\right\}dx_1dy_1dx_2dy_2\end{aligned}$$

(5.119)

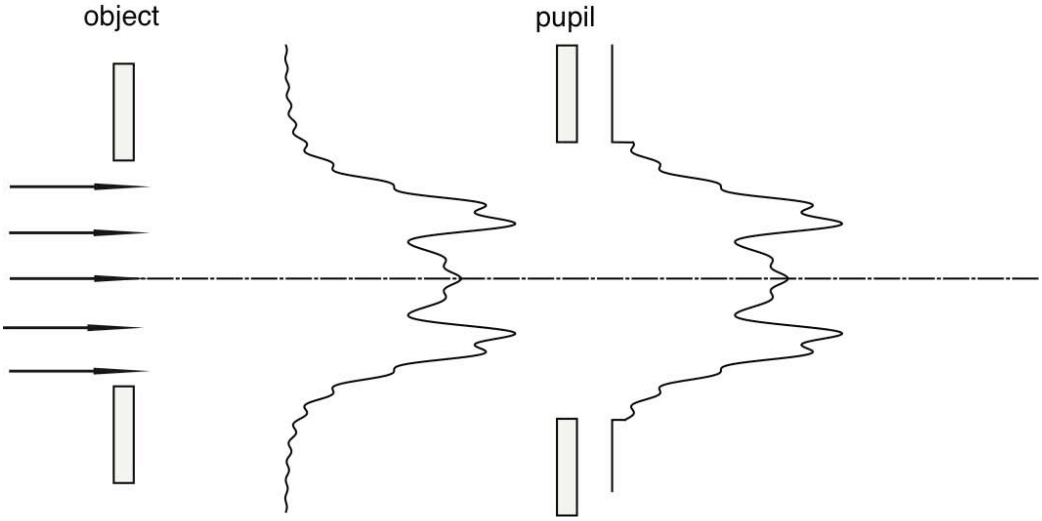

Fig. 5.22 Truncation of the beam by the pupil.

So far this is not much simpler! But we can do the integrals in x_2 and y_2 now. We have

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left[-\frac{ik}{2f}(x_2^2+y_2^2)\right]\exp\left\{-\frac{ik}{f}[x_2(x_1+x_3)+y_2(y_1+y_3)]\right\}dx_2dy_2. \qquad (5.120)$$

This is just the Fourier transform of a Gaussian (albeit of imaginary argument). You can get this from tables of Fourier transforms, or it can be evaluated using the properties of the Fresnel integrals. After putting $x = x_1 + x_3$, $y = y_1 + y_3$, it is

$$-\frac{2\pi f}{ik}\exp\left[-\frac{ik}{2f}\left(x^2+y^2\right)\right]. \qquad (5.121)$$

The important features to notice are that it is independent of x_2, y_2, and when you put it back in Eq.5.119, the quadratic terms in x_1, x_3, y_1 and y_3, all cancel to give

$$U_3(x_3,y_3) = -\frac{ie^{2ikf}}{\lambda f}\iint U_1(x_1,y_1)\exp\left[-\frac{ik}{f}(x_1x_3+y_1y_3)\right]dx_1dy_1. \qquad (5.122)$$

Compare this with Eq.5.89. Again we have the 2D Fourier transform, but now the parabolic phase factor of Eq.5.89 is no longer present (Fig.5.23). This is a very important result, which is the basis of most Fourier optics systems. It allows us to perform a 2D Fourier transform almost instantaneously (Fig. 5.24), in the time for light to travel a distance 2*f*.

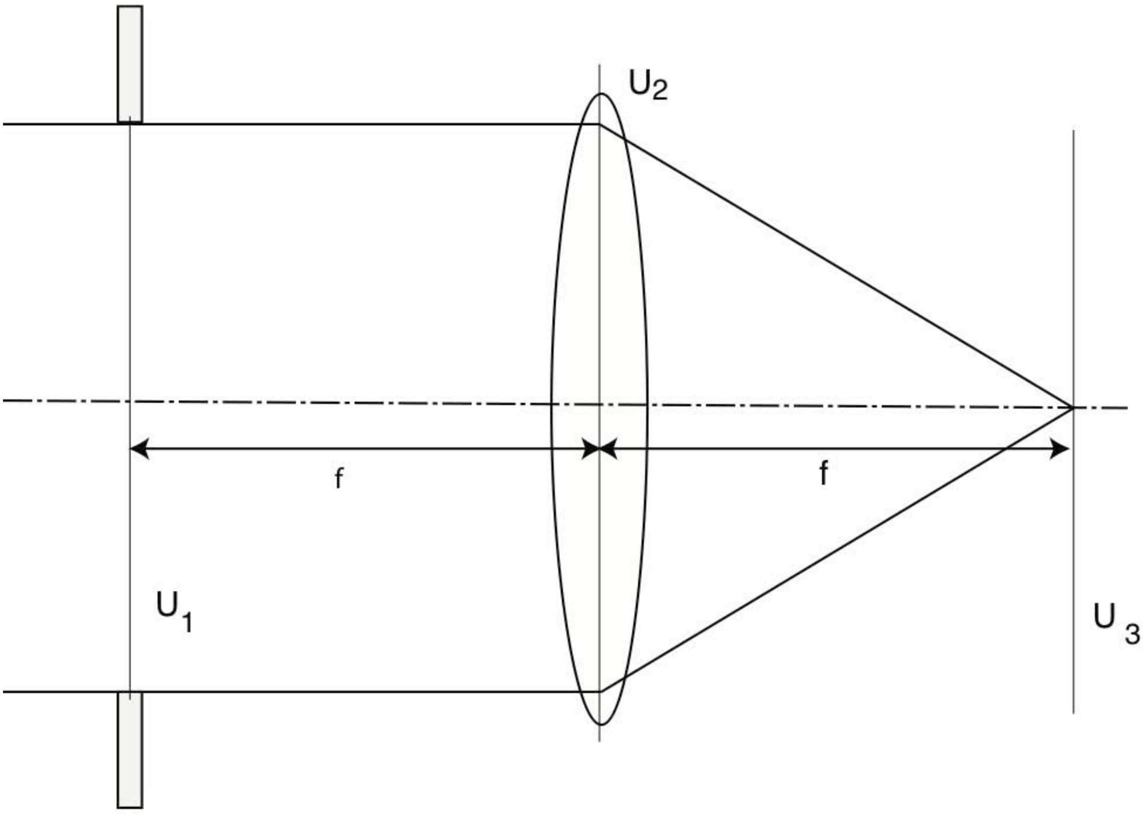

Fig. 5.23 Imaging of an object in the front focal plane.

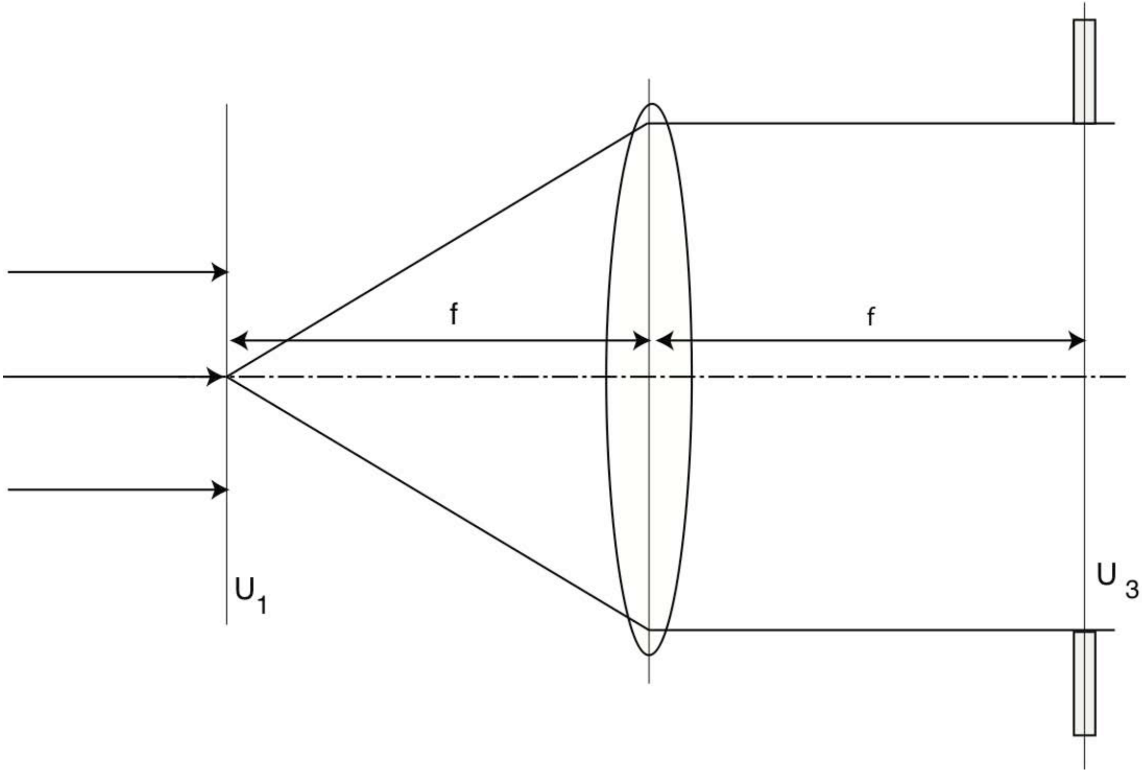

Fig. 5.24 Fourier transformation by a lens.

Note that if U_1 is a constant, assuming the pupil P is very big, U_1 is a delta-function. If the pupil is finite we get the point spread function. If U_1 is a circular function, a constant value within a circle, smaller than the pupil *P*, then the final amplitude is its Hankel function, that is the Airy disk $2J_1(v)/v$.

We have seen that if U_1 is a delta function then U_1 is a constant. Two of these units may be coupled together, as in Fig. 5.25. But the Fourier transform of the Fourier transform of $U_1(x)$ is $U_1(-x)$. So we just form an inverted, unity magnification image of U_1. This is called a 4*f* optical system.

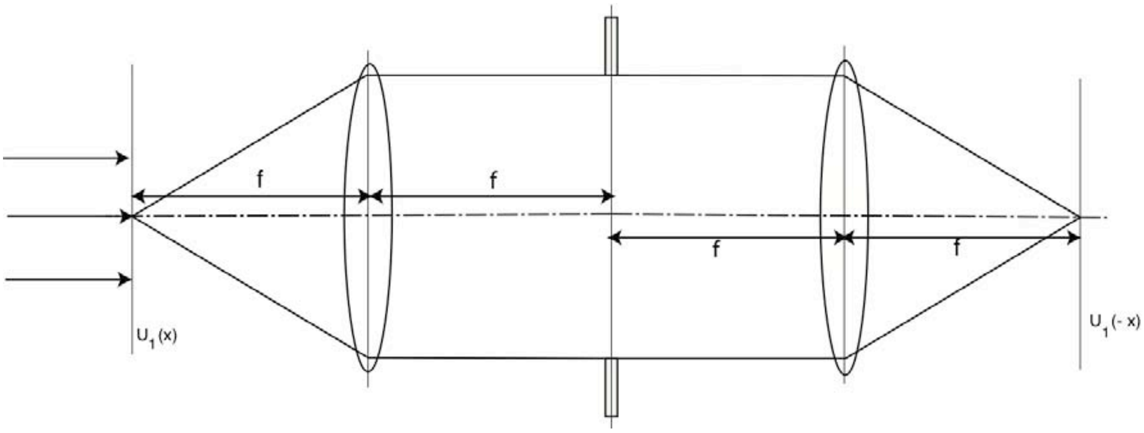

Fig. 5.25 A 4*f* system.

5.15. Coherent Transfer Function

We have looked at two imaging systems, shown in Figs. 5.20 and 5.25. For the system in Fig. 5.20, Eq.5.89 shows that the image amplitude is multiplied by a parabolic phase factor, but for the system in Fig. 5.25 there is no phase factor, and the field can be written simply as

$$U_3(x_3, y_3) = (U_1 \otimes h)(x_3, y_3). \qquad (5.123)$$

In each case the amplitude point spread function *h* is given by the 2D Fourier transform of the pupil function.

Considering the system in Fig. 5.25, we see that if the pupils are very big then U_3 is a perfect image of U_1. If we think of U_1 as a grating it produces various diffraction orders, and these are combined by the second lens to produce an image. However, the pupil *P* cuts off some of the diffraction orders and hence a perfect image is not formed in practice. *P* can therefore be thought of as having the effect of a coherent transfer function, a low pass filter. We resolve U_1 into gratings, and some of these orders get through the system. The strength of the spatial frequency components is multiplied by *P* to give their strength in the image. This is the principle of the Abbe theory of image formation in a microscope.

Mathematically, we introduce the Fourier transform of the object amplitude $U_{1,}$ given by $\tilde{U}_1$

$$\tilde{U}_1(m,n)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}U_1(x_1,y_1)\exp[-2\pi i(mx_1+ny_1)]dx_1\,dy_{1.} \qquad (5.124)$$

Inverting, we get

$$U_1(x_1,y_1)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\tilde{U}_1(m,n)\exp[2\pi i(mx_1+ny_1)]dm\,dn\,. \qquad (5.125)$$

But from Eq.5.123, and neglecting the multiplying constants we had previously

$$U_3(x_3,y_3)=\iint U_1(x_1,y_1)h\left(x_1+\frac{x_3}{M},y_1+\frac{y_3}{M}\right)dx_1\,dy_1. \qquad (5.126)$$

Substituting Eq.5.125 in Eq.5.136:

$$U_3(x_3,y_3)=\int\int\int\int\tilde{U}_1(m,n)h\left(x_1+\frac{x_3}{M},y_1+\frac{y_3}{M}\right)$$

$$\times\exp[2\pi i(mx_1+ny_1)]dm\,dn\,dx_1\,dy_1. \qquad (5.127)$$

But we also have

$$h(x,y)=\iint P(x_2,y_2)\exp\left[-\frac{ik}{f_1}(x_2x+y_2y)\right]dx_2\,dy_{2\,,} \qquad (5.128)$$

$$P(x_2,y_2)=\iint h(x,y)\exp\left[\frac{ik}{f_1}(x_2x+y_2y)\right]dx\,dy\,. \qquad (5.129)$$

So doing the integrals in x_1,y_1 in Eq.5.127, and putting $x=x_1+x_3/M$ and so on,

$$\iint h_1(x,y)\exp\left\{2\pi i\left[m\left(x-\frac{x_3}{M}\right)+n\left(y-\frac{y_3}{M}\right)\right]\right\}dx\,dy$$

$$=P(m\lambda f_1,n\lambda f_1)\exp\left[-2\pi i\left(\frac{mx_3}{M}+\frac{ny_3}{M}\right)\right].$$

So Eq.5.127 can be written

$$U_3(x_3,y_3)=\iint\tilde{U}_1(m,n)P(m\lambda f_1,n\lambda f_1)\exp\left[-2\pi i\left(\frac{mx_3}{M}+\frac{ny_3}{M}\right)\right]dm\,dn\,. \qquad (5.130)$$

So if pupil function is as shown in Fig. 5.26 (a), the coherent transfer function is as shown in Fig.5.26 (b). For an object that is only a function of x, $n = 0$ and the corresponding coherent transfer function is given by a section through the two-dimensional transfer function (Fig. 5.27). Note that in general there will be both positive and negative spatial frequencies. The imaging system behaves as a low-pass filter, that is it transmits spatial frequencies less than $(\sin\alpha)/\lambda$. Note the sudden cut-off, corresponding to the edge of the pupil, as compared with transfer functions of electrical systems, which roll off smoothly.

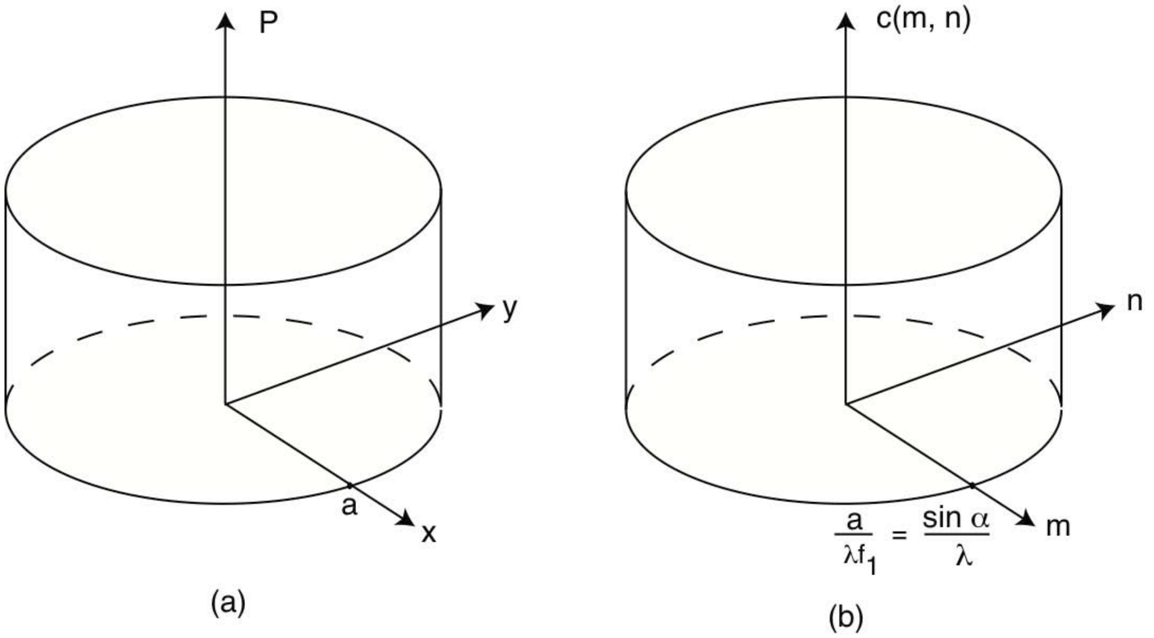

Fig. 5.26 (a) The pupil function and (b) the coherent transfer function.

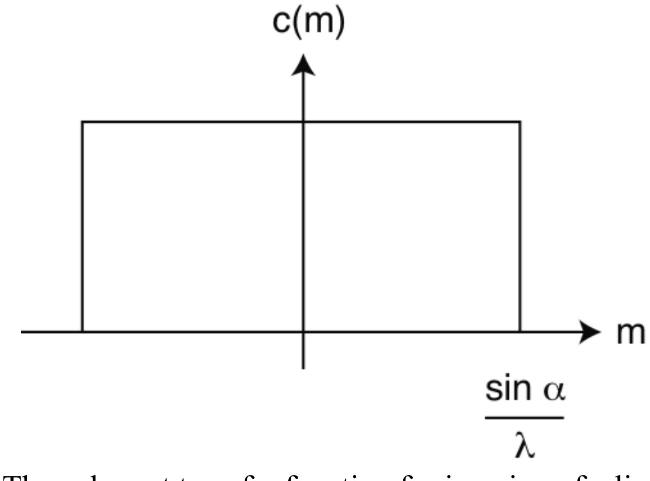

Fig. 5.27 The coherent transfer function for imaging of a line structure.

5.15.1 A grating object

Consider as an example an amplitude transmittance that is a cosinusoidal variation on a constant background (Fig. 5.28 (a)):

$$U_1 = 1 + a\cos(2\pi \nu x)\,. \tag{5.131}$$

The period of grating is $\Lambda = 1/\nu$. The spectrum of the object is the Fourier transform of U_1. As

$$U_1 = 1 + \frac{a}{2}e^{2\pi i \nu x} + \frac{a}{2}e^{-2\pi i \nu x}$$

we have for the object spectrum (Fig. 5.28 (b))

$$T(m) = \tilde{U}_1 = \left[\delta(m) + \frac{a}{2}\delta(m-\nu) + \frac{a}{2}\delta(m-\nu)\right]\delta(n)\,. \tag{5.132}$$

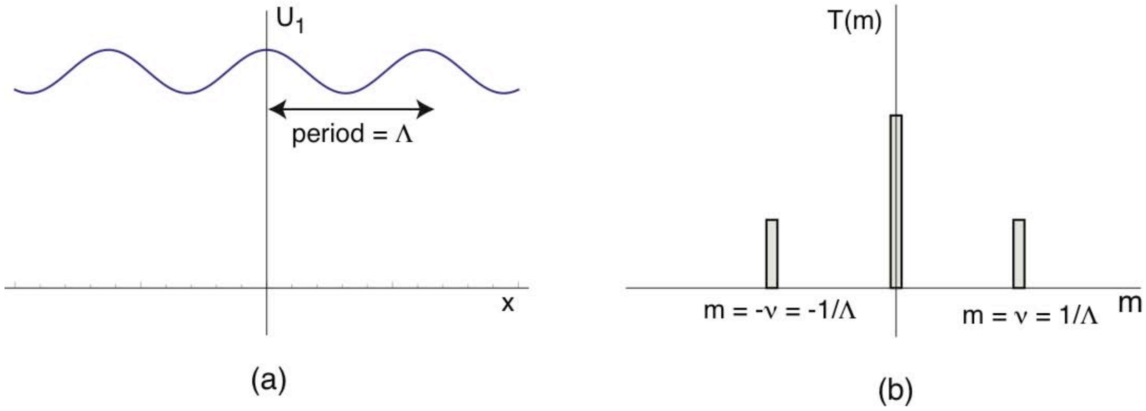

Fig. 5.28 (a) A grating object, and (b) its spatial frequency content

So in this case the image amplitude is

$$U_3(x_3, y_3) = \iint \left[\delta(m) + \frac{a}{2}\delta(m-\nu) + \frac{a}{2}\delta(m+\nu)\right] c(m,n)\delta(n)$$

$$\times \exp\left[-\frac{2\pi i}{M}(m x_3 + n y_3)\right] dm\, dn$$

$$= c(0) + \frac{a}{2}c(\nu)\exp\left(-\frac{2\pi i \nu x_3}{M}\right) + \frac{a}{2}c(-\nu)\exp\left(\frac{2\pi i \nu x_3}{M}\right).$$

For a circular pupil, *c* is even so $c(-\nu) = c(+\nu)$, and we obtain for the image amplitude

$$U_3(x_3, y_3) = c(0) + a c(\nu)\cos\left(\frac{2\pi \nu x_3}{M}\right).$$

In this case, if $\nu < a/\lambda f_1$ and $c(\nu) = 1$, the image is the same as the object but magnified by *M*.

The image *intensity* is thus

$$I_3\left(x_3,y_3\right)=\left|c\left(0\right)+ac\left(v\right)\cos\left(\frac{2\pi v x_3}{M}\right)\right|^2.$$

For the moment take *a* as *real*, and also *c* as real, so then

$$I_3\left(x_3,y_3\right)=c^2\left(0\right)+2c\left(0\right)c\left(v\right)a\cos\left(\frac{2\pi v x_3}{M}\right)+a^2c^2\left(v\right)\cos^2\left(\frac{2\pi v x_3}{M}\right). \qquad (5.133)$$

Using the identity $\cos(2\theta)=\cos^2\theta-1$, we obtain

$$I_3\left(x_3,y_3\right)=\left[c^2\left(0\right)+\frac{a^2}{2}c^2\left(v\right)\right]+2c\left(0\right)c\left(v\right)a\cos\left(\frac{2\pi v x_3}{M}\right)+\frac{a^2}{2}c^2\left(v\right)\cos\left(\frac{4\pi v x_3}{M}\right). \qquad (5.134)$$

Note that a second harmonic term is introduced by the squaring operation, but this can be neglected if *a* is small.

Now let us consider what happens if *a* and *c* are complex. Now Eq.5.133 becomes

$$I_3\left(x_3,y_3\right)=\left|c\left(0\right)\right|^2+2c\left(0\right)\mathrm{Re}\left[ac\left(v\right)\right]\cos\left(\frac{2\pi v x_3}{M}\right). \qquad (5.135)$$

If we consider an object whose *thickness* l varies in a cosinusoidal fashion

$$l=l_0+l_1\cos(2\pi v x),$$

Then if its refractive index is *n*, the phase change on passing through it is nl, and so the amplitude on the far side if it is illuminated with a plane wave is

$$\begin{aligned} t&=\exp\left\{ik\left[l_0+l_1\cos\left(2\pi v x\right)\right]\right\}\\ &=\exp\left(ikl_0\right)\exp\left[ikl_1\cos\left(2\pi v x\right)\right]. \end{aligned} \qquad (5.136)$$

The first part of this is just a constant phase term and hence can be ignored. We then expand the second part into its Fourier components. Actually this is identical to frequency modulation (FM) in communication theory, and the strengths of the various harmonics are given by Bessel Functions. Let us just look at the much simpler case when $kl_1<<1$ however. Then

$$t=1+ikl_1\cos\left(2\pi v x\right). \qquad (5.137)$$

We can see that this is the same as Eq.5.131 with *a* given by an imaginary quantity. Thus a phase object is represented by imaginary *a*. We see from Eq.5.135, that if $c(v)$ is *real* then there is no image, and we just see a constant intensity. Note that this is not in general true for a strong phase object where terms of strength a^2 cannot be ignored.

Previously we introduced the concept of the defocused pupil function. From Eqs. 5.100 or 5.117 we have

$$P_{eff}(\rho) = P(\rho)\exp\left(-\frac{iu}{2}\rho^2\right), \qquad (5.138)$$

so that now we introduce the defocused transfer function $c(m,u)$, as shown in Fig. 5.29. For $u = 0$, that is the focused case, $c(m)$ is purely real. As u is increased, the imaginary part increases in strength. For $u > \pi$, the real part starts to go negative which is not good for imaging of amplitude information. The imaginary part images phase information, but this also starts to go negative for $u > 2\pi$. So we only get good imaging of phase information if $u \le 2\pi$. If u is negative the imaginary part of $c(m,u)$ also becomes positive, so contrast is reversed. For a phase object, there is no contrast at the focus, and positive or negative contrast on either side of focus. Defocusing of the microscope was often used to image weak phase structures, as for example in biological samples, before more modern methods of phase imaging were invented.

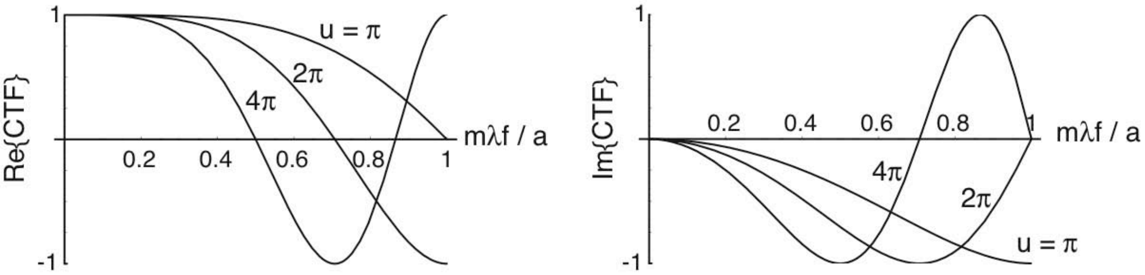

Fig. 5.29 Defocused coherent transfer function.

5.15.2. Square Wave Object

Consider a square wave object. Then

$$U_1(x) = \frac{1}{2} + \frac{2}{\pi}\left[\cos(2\pi\nu x) - \frac{1}{3}\cos(6\pi\nu x) + \frac{1}{5}\cos(10\pi\nu x) - \ldots\right]. \qquad (5.139)$$

The harmonics are transmitted up to the cut-off frequency. By making $1/\nu$ very large, we can calculate the image of a single edge. Interestingly, although the number of terms then is very large we still get an overshoot as in the image of a square wave object. The image of an edge can be expressed in terms of Fresnel integrals. The fact that the wiggles do not disappear for a large number of terms is a consequence of the Gibbs phenomenon.

5.16. Spatial Filtering

The principle of spatial filtering is to alter the strength in the image of the Fourier components of the object by putting in an appropriate mask: For example suppose we change the phase everywhere except at $m = 0$ by $\pi / 2$. So the transfer function is

$$c(v) = \begin{cases} -i & v \neq 0 \\ 1 & v = 0. \end{cases} \tag{5.140}$$

Then from Eq.5.143

$$\begin{aligned} I_3 &= 1 + 2\,\mathrm{Re}(-ai)\cos\left(\frac{2\pi v x_3}{M}\right) \\ &= 1 + 2\,\mathrm{Im}(a)\cos\left(\frac{2\pi v x_3}{M}\right). \end{aligned} \tag{5.141}$$

So again we have managed to image the phase information. This method is called Zernike phase contrast, for the invention of which Zernike received the Nobel prize. In practice it is easier to change the phase just of the direct beam $m = 0$, rather than vice-versa, but the result is exactly the same. Note that changing the phase by $-\pi / 2$ rather than by $+\pi / 2$ reverses the contrast. Also, if we make the filter

$$c(v) = \begin{cases} -bi, & v \neq 0, \\ 1, & v = 0, \end{cases} \tag{5.142}$$

we then have

$$I_3 = 1 + 2b\,\mathrm{Im}(a)\cos\left(\frac{2\pi v x_3}{M}\right). \tag{5.143}$$

If $b > 1$, we also enhance the contrast by a factor *b,* making very weak phase objects visible.

5.17. Incoherent Imaging

Most of what we have said so far is applicable to *coherent* optical systems. This requires that the point spread function of the imaging system is small in extent compared with the lateral spatial coherence of the illumination. The opposite condition, where the point spread function is large compared with the spatial coherence, results in *incoherent* imaging. Incoherent imaging is in everyday life a much more common phenomenon. For example photography, or just seeing, is normally an incoherent process. Fluorescence also results in incoherent image formation. The in-between case, where the point spread function is about the size of the spatial coherence of the illumination, is *partially coherent.* This case is much more complicated, and arises for example in the theory of microscope imaging. In incoherent imaging, because there is no coherent interference between neighbouring points, we have to sum the intensities for these points, rather than the complex amplitudes. We return to the geometry of Section 5.14.

Each point in the object results in a diffraction blur in the image. To obtain the total image we sum over the contributions from the *intensities* of the individual points.

For *coherent* imaging, we had in Eq.5.108

$$I_3(x_3,y_3) = \frac{1}{\left(\lambda^2 M d_1^2\right)^2}\left| t \otimes h \right|^2,$$

but now for *incoherent* imaging, we have

$$I_3(x_3,y_3) = \frac{1}{\left(\lambda^2 M d_1^2\right)^2}\left(\left| t \right|^2 \otimes \left| h \right|^2 \right). \tag{5.144}$$

That is we must convolve the object *intensity* $\left| t \right|^2$ with the *intensity* point spread function $\left| h \right|^2$. Note that the term point spread function can refer to either the amplitude or the intensity point spread function in the literature.

We can derive Eq.5.144 properly by considering a single point in the object U_1 at (x,y). Its image from Eq.5.111 is

$$U_3(x_3,y_3) = \frac{1}{\lambda^2 M d_1^2}\exp\left[ikd_1(1+M)\right]\exp\left[\frac{ik}{2Md_1}\left(x_3^2+y_3^2\right)\right]$$

$$\times\iint \delta(x_1-x)\delta(y_1-y)\exp\left[\frac{ik}{d_1}\left(x_1^2+y_1^2\right)\right]h\left(x_1+\frac{x_3}{M},y_1+\frac{y_3}{M}\right)dx_1dy_1$$

$$= \frac{1}{\lambda^2 M d_1^2}\exp\left[ikd_1(1+M)\right]\exp\left[\frac{ik}{2Md_1}\left(x_3^2+y_3^2\right)\right]$$

$$\times\exp\left[\frac{ik}{d_1}\left(x^2+y^2\right)\right]h\left(x+\frac{x_3}{M},y+\frac{y_3}{M}\right).$$

So the intensity in the image of a single point is simply

$$I_3(x_3,y_3) = \frac{1}{\left(\lambda^2 M d_1^2\right)^2}\left| h\left(x+\frac{x_3}{M},y+\frac{y_3}{M}\right)\right|^2. \tag{5.145}$$

Finally, adding up for the points of the object

$$I_3(x_3,y_3)=\frac{1}{\left(\lambda^2 M d_1^2\right)^2}\iint\left|h\left(x+\frac{x_3}{M},y+\frac{y_3}{M}\right)\right|^2\left|t(x,y)\right|^2 dx\,dy \qquad (5.146)$$

which represents Eq.5.144.

5.17.1 Two-point object

One of the most important theoretical objects is two bright points in a dark background (e.g. two pinholes in an opaque screen, or two stars). The image is then, for different normalized separations $2v_0$, as shown in Fig. 5.30 (a). For $v_0 = 2.5$ the points are well resolved. For $v_0 = 1.5$ they are not: they just almost look like one point. It is traditional to say that the points are just resolved when the maximum of one point spread function is placed over the zero of the other. This is called the Rayleigh criterion, and occurs when $v_0 = 1.92$. It is found that for a circular pupil aperture, the intensity in the middle is then 0.735 times the intensity at the points themselves. Note that the Rayleigh criterion applies for incoherent imaging. In general, though, we may introduce the *generalized Rayleigh criterion*, which states that the points are just resolved if the intensity at the centre is 0.735 times that at the points. The intensity at the points need not be the same as the intensity at the maximum: some published papers have got this wrong! By putting in the values for the Bessel function we obtain, for the incoherent case

$$(\Delta x)_{\min} = 0.61\frac{\lambda}{\sin\alpha}\ . \qquad (5.147)$$

This is often written as $1.22 f\lambda / D$ in terms of the diameter D of the pupil, rather than the radius. A typical value for a high power microscope is about 0.5 μm. For the coherent case, similar plots are shown in Fig. 5.30 (b). We find

$$(\Delta x)_{\min} = 0.82\frac{\lambda}{\sin\alpha}, \qquad (5.148)$$

that is the resolution is not as good.

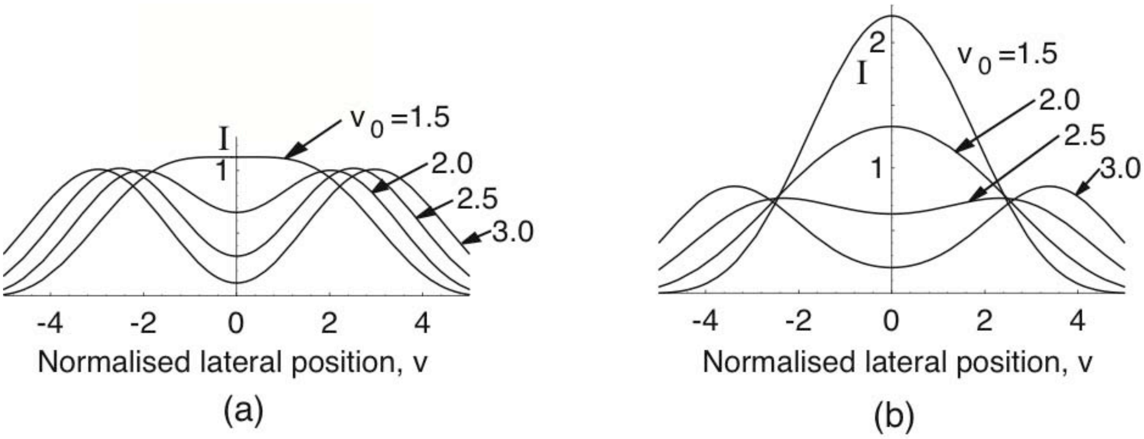

Fig. 5.30 Images of a two-point object: (a) incoherent imaging, (b) coherent imaging.

5.17.2. Optical transfer function

Eq.5.146 is linear in intensity. This means we can introduce a transfer function, usually called the OTF (optical transfer function). Note that this operates on *intensities*, rather than the amplitudes for the coherent transfer function described earlier. So they are not strictly comparable.

We introduce the object *intensity* spectrum, given by the Fourier transform of its intensity, which is in contrast with Eq.5.124,

$$V_1(m,n)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left|U_1(x_1,y_1)\right|^2\exp\left[-2\pi i\left(mx_1+ny_1\right)\right]dx_1dy_1. \qquad (5.149)$$

Then we can show by exactly the same method as in Section 5.16 that

$$I(x_3,y_3)=\iint V_1(m,n)C(m,n)\exp\left[-2\pi i\left(\frac{mx_3}{M}+\frac{ny_3}{M}\right)\right]dm\,dn \qquad (5.150)$$

where the OTF, $C(m,n)$, is given by the Fourier transform of the intensity point spread function $|h|^2$. As for the coherent case we had $P(m\lambda f,n\lambda f)=F(h)$, where $F(.)$ represents the Fourier transform, now we have

$$\begin{aligned}C(m,n)&=F\left(|h|^2\right)\\&=P(m\lambda f,n\lambda f)\otimes P^*(m\lambda f,n\lambda f),\end{aligned} \qquad (5.151)$$

This 2D convolution represents the area of overlap of the two pupils. We can show that the area of overlap for two circles is

$$\frac{2}{\pi}\left[\cos^{-1}(\tilde{m})-\tilde{m}\sqrt{1-\tilde{m}^2}\right], \qquad (5.152)$$

which is shown in Fig. 5.31. The cut-off frequency is twice that for a coherent system.

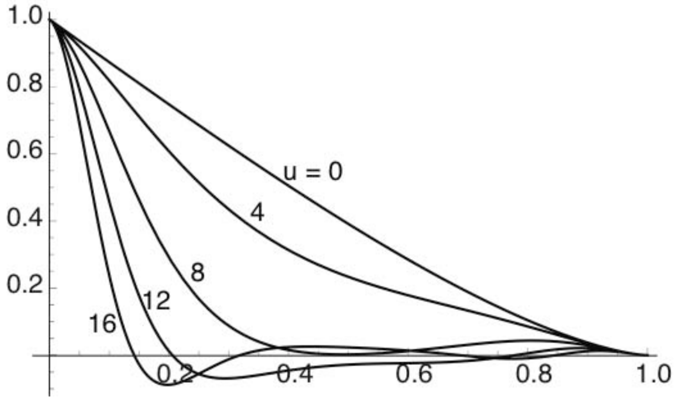

Fig. 5.31 Defocused optical transfer function for a circular pupil.

If the system is defocused, we must integrate over the area of overlap taking into account the phase of the pupil. This cannot be done analytically, but can be expressed in terms of a single integral.(Hopkins 1955) The result is also shown in Fig. 5.31. The response drops off with defocus, that is imaging of these particular spatial frequency components is worse. It is the mid-spatial frequencies that are most strongly affected, resulting in poorer imaging. Note that the OTF must always be purely real. An important feature is that the OTF can go negative with defocus. This means that some spatial frequency components have their contrast reversed. This results in optical artifacts, which means that you can see something that is not really in the object.

References

Abramowitz, M. and Stegun, I. A. 1965. *Handbook of Mathematical Functions*. New York, Dover.

Born, M. and Wolf, E. 1975. *Principles of Optics*. Oxford, Pergamon Press.

Goodman, J. W. 1968. *Introduction to Fourier Optics*. New York, McGraw-Hill.

Hecht, E. 1987. *Optics*. Reading, Addison-Wesley.

Hopkins, H. H. 1955. The frequency response of a defocused optical system. *Proceedings of the Royal Society of London Series A* 231: 91-103.