Chapter 4

Fourier Optics

Contents

4.1 Basic principles of scalar diffraction theory

4.1.1 Introduction

Diffraction can be defined as "any deviation of a light ray from rectilinear propagation, which is not caused by reflection nor refraction". It was already known for centuries that light rays, passing through a small aperture in an opaque screen do not form a sharp shadow on a distant screen. That smooth transition from light to shadow could only be explained by assuming that light has a $wavelike$ character. Diffraction theory has been further developed by Huygens, Fresnel, Kirchhoff and Sommerfeld; the latter was the first to find an exact solution for diffraction of a plane wave at a semi-infinite thin conducting plate. In this chapter we will limit ourselves to the approximation of a $scalar$ theory: here only one single component of the electric or magnetic field vector is considered. This also means that we neglect the (possible) coupling between electric and magnetic fields. By comparing this approximation with exact theories, and also with experiments, it turns out that this scalar diffraction theory is good whenever:

- the diffracting aperture is large compared with the wavelength of the light;
- the diffracting field is calculated at a large distance from the aperture.

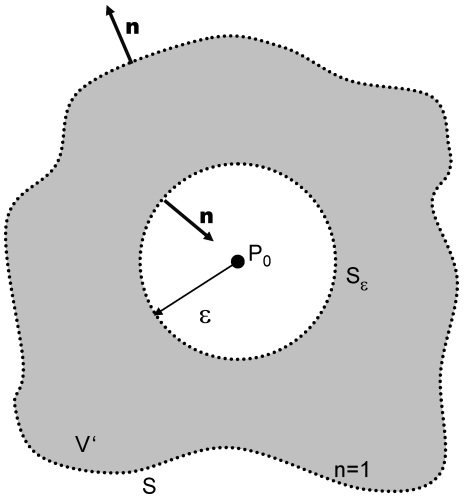


Figure 4.1: Volume V enclosed by surface S

4.1.2 Integral theorem of Helmholtz and Kirchhoff

Suppose one wants to calculate the electric field in a point of observation P_0. Consider then an arbitrary closed surface S surrounding P_0, and enclosing a volume V (figure 4.1). We also assume that the space is homogeneous, with an index of refraction $n = 1$ (free space). [The theory can easily be extended to media with a real index n, simply by replacing the vacuum wavelength λ by λ/n in all equations.] We moreover assume that there are no light sources nor light traps within V, and that the light is monochromatic with frequency f (or angular frequency $\omega = 2\pi f$ or wavelength $\lambda = c/f$). This implies that each field can be represented by its complex phasor, which gives the amplitude and phase. Within the volume V both the electric field E and the magnetic field H obey a Helmholtz equation:

$$\begin{aligned} \nabla^2\mathbf{E} + k^2\mathbf{E} &= 0 \\ \nabla^2\mathbf{H} + k^2\mathbf{H} &= 0 \end{aligned} \tag{4.1}$$

In the scalar approximation each component of the electric field or the magnetic field obeys the same equation. We will, from now on, represent this component by the symbol U, and we will call this the $field$. It satisfies:

$$\nabla^2 U + k^2 U = 0 \tag{4.2}$$

It is clear that $|U|^2$ is proportional to the irradiance, which gives the power density. In order to calculate the field U in the point of observation P_0 one starts from Green's theorem

$$\iiint_V \left(G\nabla^2 U - U\nabla^2 G\right) dv = \iint_S \left(G\frac{\partial U}{\partial n} - U\frac{\partial G}{\partial n}\right) ds \tag{4.3}$$

which applies whenever the field function U and Green's function G, together with their first and second derivatives are single-valued and continuous within and on S. For Green's function one chooses here a unit-amplitude spherical wave expanding about the point P_0

$$G = \frac{e^{-jk|\mathbf{r}-\mathbf{r}_0|}}{|\mathbf{r}-\mathbf{r}_0|}, \text{ with } k = \frac{2\pi}{\lambda} \tag{4.4}$$

Because this function is not continuous in P_0 we have to exclude this point from V. Therefore a small sphere with surface S_ϵ and radius ϵ around P_0 is excluded from the volume V. Green's theorem is now applied in the volume V' lying between S en S_ϵ with enclosing surface $S' = S + S_\epsilon$. It is clear that G, being a spherical wave, also obeys a Helmholtz equation

$$\nabla^2 G + k^2 G = 0 \tag{4.5}$$

Hence the left-hand side of Green's equation reduces to:

$$\int\int_{V'}\int \left(G\nabla^2 U - U\nabla^2 G\right) dv = \int\int_{V'}\int \left(GUk^2 - UGk^2\right) dv = 0 \tag{4.6}$$

and consequently

$$\int\int_{S'} \left(G\frac{\partial U}{\partial n} - U\frac{\partial G}{\partial n}\right) ds = 0 \tag{4.7}$$

or

$$\int\int_{S} \left(G\frac{\partial U}{\partial n} - U\frac{\partial G}{\partial n}\right) ds = -\int\int_{S_\varepsilon} \left(G\frac{\partial U}{\partial n} - U\frac{\partial G}{\partial n}\right) ds \tag{4.8}$$

Note that for a general point P_1 on S', one has

$$G(P_1) = \frac{e^{-jk|\mathbf{r}_1 - \mathbf{r}_0|}}{|\mathbf{r}_1 - \mathbf{r}_0|} \tag{4.9}$$

$$\frac{\partial G}{\partial n}(P_1) = -\cos(\mathbf{n}, \mathbf{r}_1 - \mathbf{r}_0)\left(jk + \frac{1}{|\mathbf{r}_1 - \mathbf{r}_0|}\right)\frac{e^{-jk|\mathbf{r}_1 - \mathbf{r}_0|}}{|\mathbf{r}_1 - \mathbf{r}_0|} \tag{4.10}$$

If P_1 lies on S_ϵ, then

$$G(P_1) = \frac{e^{-jk\varepsilon}}{\varepsilon} \tag{4.11}$$

and

$$\frac{\partial G}{\partial n}(P_1) = \left(jk + \frac{1}{\varepsilon}\right)\frac{e^{-jk\varepsilon}}{\varepsilon} \tag{4.12}$$

Letting ϵ now become arbitrary small, the continuity of U and its derivative around P_0 allows us to write:

$$\begin{aligned}\int\int_{S_\varepsilon} \left(G\frac{\partial U}{\partial n} - U\frac{\partial G}{\partial n}\right) ds &= 4\pi\varepsilon^2\left[\frac{\partial U(P_0)}{\partial n}\frac{e^{-jk\varepsilon}}{\varepsilon} - U(P_0)\frac{e^{-jk\varepsilon}}{\varepsilon}\left(jk + \frac{1}{\varepsilon}\right)\right] \\ &= -4\pi U(P_0)\end{aligned} \tag{4.13}$$

which, after substitution in (4.8) gives

$$U(P_0) = \frac{1}{4\pi}\int\int_{S} \left(\frac{\partial U}{\partial n}\frac{e^{-jk|\mathbf{r} - \mathbf{r}_0|}}{|\mathbf{r} - \mathbf{r}_0|} - U\frac{\partial}{\partial n}\frac{e^{-jk|\mathbf{r} - \mathbf{r}_0|}}{|\mathbf{r} - \mathbf{r}_0|}\right) ds \tag{4.14}$$

This formula allows for the field at an arbitrary point P_0 to be expressed in terms of its boundary values on any closed surface surrounding that point. It is known as the *integral theorem of Helmholtz and Kirchhoff*.

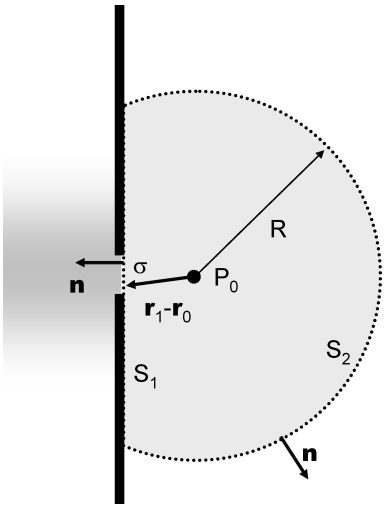


Figure 4.2: Diffraction trough an aperture in a screen

4.1.3 Diffraction through an aperture in a planar screen

Consider now the diffraction of light by an aperture in a screen (figure 4.2). The light wave is assumed to impinge from the left, and the field at P_0 is to be calculated. The previous integral theorem can be used, on condition that the surface of integration S is carefully chosen. Following Kirchhoff, we choose the surface S to consist of two parts: a plane surface S_1 lying directly behind the diffracting screen, joined and closed by a large spherical cap S_2 of radius R and centered at the observation point P_0. Applying the integral theorem of Helmholtz and Kirchhoff gives:

$$U(P_0) = \frac{1}{4\pi} \int\int_{S_1+S_2} \left(\frac{\partial U}{\partial n} G - U \frac{\partial G}{\partial n} \right) ds \tag{4.15}$$

where, as before

$$G = \frac{e^{-jk|\mathbf{r}-\mathbf{r}_0|}}{|\mathbf{r}-\mathbf{r}_0|} \tag{4.16}$$

Note here that Green's function is defined for the complete space, and not only in a half one. Moreover, one can prove that the function U, because it satisfies the Helmholtz equation, also satisfies the *Sommerfeld's radiation condition*:

$$\lim_{R\to\infty} R \left(\frac{\partial U}{\partial n} + jkU \right) = 0 \tag{4.17}$$

which implies that the integral (4.15) over S_2 will vanish when R becomes arbitrary large. Next we need to know the values of U and its derivative on the surface S_1. Here we follow the assumptions which Kirchhoff adopted, and which are still known as *Kirchhoff's boundary conditions*:

1. across the aperture σ, the field U and its derivative $\partial U/\partial n$ are exactly the same as they would be in the absence of a screen
2. over that portion of S_1 which differs from $sigma$, we set $U = 0$ and $\partial U/\partial n = 0$

Although these assumptions seem intuitively reasonable, they are mathematically inconsistent! Indeed: when a solution of a 3-dimensional wave equation is zero, together with its derivative, on a finite surface, then it has to be zero everywhere. Nevertheless, it turns out that Kirchhoff's boundary conditions yields results which agree very well with experiments, at least when the approximations of section 4.1.1 are satisfied. The field in the point of observation P_0 is consequently:

$$U(P_0) = \frac{1}{4\pi} \int\int_{\sigma} \left(\frac{\partial U}{\partial n} G - U \frac{\partial G}{\partial n} \right) ds \tag{4.18}$$

As we calculate the field U only in observation points P_0 at large distances from the aperture, it follows that

$$|\mathbf{r} - \mathbf{r}_0| \gg \lambda \tag{4.19}$$

$$k \gg \frac{1}{|\mathbf{r} - \mathbf{r}_0|} \tag{4.20}$$

and (4.10) becomes :

$$\frac{\partial G}{\partial n}(P_1) = -jk \cos(\mathbf{n}, \mathbf{r}_1 - \mathbf{r}_0) \frac{e^{-jk|\mathbf{r}_1 - \mathbf{r}_0|}}{|\mathbf{r}_1 - \mathbf{r}_0|} \tag{4.21}$$

Substituting this in (4.18) gives:

$$U(P_0) = \frac{1}{4\pi} \int\int_{\sigma} \frac{e^{-jk|\mathbf{r}_1 - \mathbf{r}_0|}}{|\mathbf{r}_1 - \mathbf{r}_0|} \left(\frac{\partial U}{\partial n} + jkU \cos(\mathbf{n}, \mathbf{r}_1 - \mathbf{r}_0) \right) ds_1 \tag{4.22}$$

This formula is known as the *Fresnel-Kirchhoff diffraction formula*.

The inconsistency of Kirchhoff's boundary conditions was removed by Sommerfeld by choosing an alternative Green's function. He considered

$$G' = \frac{e^{-jk|\mathbf{r} - \mathbf{r}_0|}}{|\mathbf{r} - \mathbf{r}_0|} - \frac{e^{-jk|\mathbf{r} - \mathbf{r}'_0|}}{|\mathbf{r} - \mathbf{r}'_0|} \tag{4.23}$$

where P'_0 is the mirror image of P_0 on the opposite side of the screen (figure 4.3). The derivative now becomes:

$$\begin{aligned} \frac{\partial G'}{\partial n} = & -\cos(\mathbf{n}, \mathbf{r} - \mathbf{r}_0) \left(jk + \tfrac{1}{|\mathbf{r} - \mathbf{r}_0|} \right) \tfrac{e^{-jk|\mathbf{r} - \mathbf{r}_0|}}{|\mathbf{r} - \mathbf{r}_0|} \\ & + \cos(\mathbf{n}, \mathbf{r} - \mathbf{r}'_0) \left(jk + \tfrac{1}{|\mathbf{r} - \mathbf{r}'_0|} \right) \tfrac{e^{-jk|\mathbf{r} - \mathbf{r}'_0|}}{|\mathbf{r} - \mathbf{r}'_0|} \end{aligned} \tag{4.24}$$

For each point P_1 on the screen S_1 one has

$$|\mathbf{r}_1 - \mathbf{r}_0| = \left|\mathbf{r}_1 - \mathbf{r}'_0\right| \tag{4.25}$$

$$\cos(\mathbf{n}, \mathbf{r}_1 - \mathbf{r}_0) = -\cos(\mathbf{n}, \mathbf{r}_1 - \mathbf{r}'_0) \tag{4.26}$$

hence

$$G'(P_1) = 0 \tag{4.27}$$

$$\frac{\partial G'}{\partial n}(P_1) = -2 \cos(\mathbf{n}, \mathbf{r}_1 - \mathbf{r}_0) \left(jk + \frac{1}{|\mathbf{r}_1 - \mathbf{r}_0|} \right) \frac{e^{-jk|\mathbf{r}_1 - \mathbf{r}_0|}}{|\mathbf{r}_1 - \mathbf{r}_0|} \tag{4.28}$$

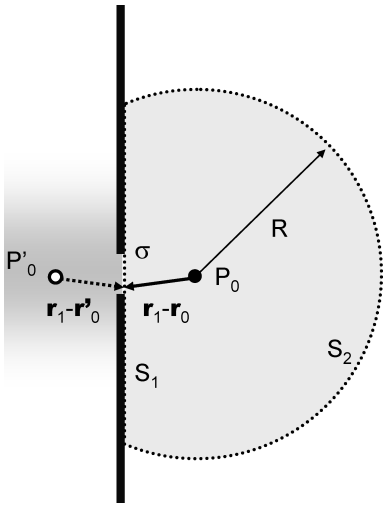


Figure 4.3: Diffraction through an aperture in a screen

Because G' is zero on the complete surface S_1, equation (4.15) reduces to

$$U(P_0) = \frac{1}{4\pi} \int\int_{S_1} \left(-U \frac{\partial G'}{\partial n} \right) ds \tag{4.29}$$

which is sometimes called the *first Rayleigh-Sommerfeld diffraction formula*. This expression also shows that it is not necessary to choose a value for the derivative of U on S_1; the knowledge of the field U suffices. One sets $U \equiv 0$ on that portion of S_1 which differs from the aperture σ, whereas across the aperture, U is exactly the same as it would be in the absence of the screen. So no boundary conditions need to be chosen for the derivative of U, and the inconsistencies of the Kirchhoff's boundary conditions have been removed. This finally leads to:

$$U(P_0) = \frac{-1}{j\lambda} \int\int_{\sigma} U(P_1) \cos(\mathbf{n}, \mathbf{r}_1 - \mathbf{r}_0) \frac{e^{-jk|\mathbf{r}_1 - \mathbf{r}_0|}}{|\mathbf{r}_1 - \mathbf{r}_0|} ds_1 \tag{4.30}$$

or

$$U(P_0) = \int\int_{\sigma} h(P_0, P_1) U(P_1) ds_1 \tag{4.31}$$

if we set

$$h(P_0, P_1) = \frac{-1}{j\lambda} \cos(\mathbf{n}, \mathbf{r}_1 - \mathbf{r}_0) \frac{e^{-jk|\mathbf{r}_1 - \mathbf{r}_0|}}{|\mathbf{r}_1 - \mathbf{r}_0|} \tag{4.32}$$

in which $h(P_0, P_1)$ is a weighting factor that is applied to the field $U(P_1)$ in order to synthesize the field in P_0 . Formulas (4.30) and (4.32) are known as the *Rayleigh-Sommerfeld diffraction formulas*. They show that the field is a superposition (=integral) of spherical waves starting from each point in the aperture, each with an appropriate amplitude and obliquity factor. This is called the *Huygens - Fresnel principle*, because it is an extension of the intuitive concept of secondary wavelets, formulated by Huygens already in 1678. Although the Sommerfeld formulation removes the inconsistencies in Kirchhoff's theory, in practical applications both formulas give essentially the same solutions, provided the aperture is much larger than the wavelength. Nevertheless one generally chooses to use the first Rayleigh Sommerfeld solution because of its simplicity.

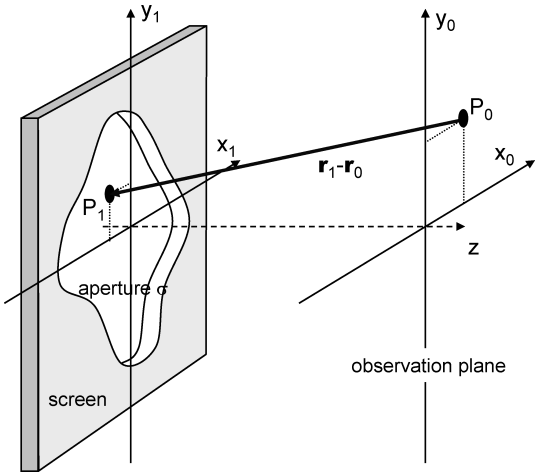


Figure 4.4: Transmission through an aperture

4.2 Fresnel and Fraunhofer diffraction

4.2.1 Fresnel diffraction formula

Assume now that the diffracting aperture lies in the (x_1, y_1) plane and is illuminated from the left by a monochromatic wave U (figure 4.4). The field is calculated in the plane of observation (x_0, y_0), parallel to the (x_1, y_1) plane, but a distance z to the right. The field in the point P_0 is given by (4.32) which we rewrite as

$$U(P_0) = \int\int_{\sigma} h(x_0, y_0, x_1, y_1) U(x_1, y_1) dx_1 dy_1 \tag{4.33}$$

with

$$h(x_0, y_0, x_1, y_1) = \frac{-1}{j\lambda} \cos(\mathbf{n}, \mathbf{r}_1 - \mathbf{r}_0) \frac{e^{-jk|\mathbf{r}_1 - \mathbf{r}_0|}}{|\mathbf{r}_1 - \mathbf{r}_0|} \tag{4.34}$$

and

$$|\mathbf{r}_1 - \mathbf{r}_0| = \sqrt{z^2 + (x_0 - x_1)^2 + (y_0 - y_1)^2} \tag{4.35}$$

Suppose now that the axial distance z is much larger than the transverse dimensions. Then

$$\cos(\mathbf{n}, \mathbf{r}_1 - \mathbf{r}_0) \cong 1 \tag{4.36}$$

The error is smaller than 5%, when the angle between $\mathbf{n}$ and $\mathbf{r}_1 - \mathbf{r}_0$ is smaller than $18°$. Also the expression $|\mathbf{r}_1 - \mathbf{r}_0|$ in the denominator of (4.34) may be replaced by z:

$$h(x_0, y_0, x_1, y_1) \cong \frac{-1}{j\lambda z} e^{-jk|\mathbf{r}_1 - \mathbf{r}_0|} \tag{4.37}$$

Furthermore one can develop the exponential in a binomial expansion, retaining only the first two terms:

$$\begin{aligned} |\mathbf{r}_1 - \mathbf{r}_0| &= \sqrt{z^2 + (x_1 - x_0)^2 + (y_1 - y_0)^2} \\ &\cong z \left[1 + \frac{1}{2} \left(\frac{x_1 - x_0}{z} \right)^2 + \tfrac{1}{2} \left(\frac{y_1 - y_0}{z} \right)^2 \right] \end{aligned} \tag{4.38}$$

This gives

$$h(x_0, y_0, x_1, y_1) \cong \frac{-e^{-jkz}}{j\lambda z} e^{-\frac{jk}{2z}\left[(x_1-x_0)^2+(y_1-y_0)^2\right]} \tag{4.39}$$

We can now replace the integration over the aperture σ by an integration over the entire plane, if we put

$$U(x_1, y_1) \equiv 0 \tag{4.40}$$

outside the aperture σ. This finally gives

$$U(x_0, y_0) = \frac{-e^{-jkz}}{j\lambda z} \int\int_{-\infty}^{+\infty} U(x_1, y_1) e^{-\frac{jk}{2z}\left[(x_1-x_0)^2+(y_1-y_0)^2\right]} dx_1 dy_1 \tag{4.41}$$

or

$$U(x_0, y_0) = \frac{-e^{-jkz}}{j\lambda z} e^{-\frac{jk}{2z}\left[x_0^2+y_0^2\right]} \int\int_{-\infty}^{+\infty} U(x_1, y_1) e^{-\frac{jk}{2z}\left[x_1^2+y_1^2\right]} e^{j\frac{2\pi}{\lambda z}[x_0 x_1 + y_0 y_1]} dx_1 dy_1 \tag{4.42}$$

This result is called the *Fresnel diffraction integral*. It clearly shows that the field $U(x_0, y_0)$ in the observation plane is the 2-dimensional Fourier transform of the field in the object plane

$$U(x_1, y_1) e^{-\frac{jk}{2z}\left[x_1^2+y_1^2\right]} \tag{4.43}$$

where the spatial frequencies are defined by:

$$f_x = -\frac{x_0}{\lambda z} \text{ en } f_y = -\frac{y_0}{\lambda z} \tag{4.44}$$

As this result is valid close to the aperture, it is called the *near-field approximation* ; one sometimes speaks of the *Fresnel diffraction regime*. This approximation is not valid too close to the aperture; however it is not easy to calculate exactly the limits of validity. A sufficient condition is that the higher-order term in the expansion be small, but this is not a necessary condition. Indeed. it suffices that they do not change the value of the integral too much *after integration*, and this also depends on the function U. In regions where the exponential varies only slowly ("stationary-phase" regime) the contributions of the higher-order terms may often be neglected, even for large values of $k/2z$. The general conclusions of deeper analyses is that the accuracy of the Fresnel approximation is extremely good to distances that are very close to the aperture.

4.2.2 Fraunhofer approximation

When the distance z between the two planes is so large that

$$z >> \frac{k\left(x_1^2 + y_1^2\right)_{\max}}{2} \tag{4.45}$$

then a further simplification is possible: the quadratic phase term can also be neglected, giving:

$$U(x_0, y_0) = \frac{-e^{-jkz}}{j\lambda z} e^{-\frac{jk}{2z}\left[x_0^2+y_0^2\right]} \int\int_{-\infty}^{+\infty} U(x_1, y_1) e^{j\frac{2\pi}{\lambda z}[x_0 x_1 + y_0 y_1]} dx_1 dy_1 \tag{4.46}$$

This shows that the field in the image plane is the Fourier transform of the field in the aperture, when the spatial frequencies are set to $f_x = -x_0/\lambda z$ en $f_y = -y_0/\lambda z$. The region where this approximation is valid is called the *far field* or the *Fraunhofer diffraction regime*. For example for a HeNe laser with a wavelength of about $6.10^{-7}m$ and an aperture of $1mm$ the far field starts at about $z > 5m$.

Important remark: In the previous sections we calculated the diffraction of a field incident upon an aperture. This aperture limited the transverse extent of the field, a condition necessary to develop the theoretical model. When, on the other hand, the incident field is smaller than the aperture itself (example: a point source, or a laser beam), then the aperture has no influence on the (diffraction of the) field. This implies that the theory remains valid for describing diffraction of any field with $finite$ transverse dimensions.

4.2.3 Examples of Fraunhofer diffraction patterns

Rectangular aperture

Consider a rectangular aperture in a screen. The amplitude transmittance of that screen can then be written as:

$$t(x_1, y_1) = \text{rect}\left(\frac{x_1}{\ell_x}\right)\text{rect}\left(\frac{y_1}{\ell_y}\right) \tag{4.47}$$

in which ℓ_x and ℓ_y are the width and height of the aperture. Suppose this aperture is illuminated by a plane monochromatic wave of unit amplitude, normally incident on the screen, then $U(x_1, y_1) = t(x_1, y_1)$. Formula (4.46) then gives the Fraunhofer diffraction pattern:

$$U(x_0, y_0) = -\frac{e^{-jkz}}{j\lambda z}e^{-\frac{jk}{2z}(x_0^2+y_0^2)}F_{2D}\left\{U(x_1, y_1)\right\} \tag{4.48}$$

with

$$f_x = -\frac{x_0}{\lambda z} \text{ and } f_y = -\frac{y_0}{\lambda z} \tag{4.49}$$

Because

$$F_{2D}\left\{U\left(x_1,y_1\right)\right\} = \ell_x\ell_y\text{sinc}\left(\pi\ell_x f_x\right)\text{sinc}\left(\pi\ell_y f_y\right) \tag{4.50}$$

with

$$\text{sinc}(x) = \frac{\sin x}{x} \tag{4.51}$$

one finds

$$U\left(x_0, y_0\right) = \frac{-e^{-jkz}}{j\lambda z}e^{-j\frac{k}{2z}(x_0^2+y_0^2)}\ell_x\ell_y\text{sinc}\left(\frac{\pi\ell_x x_0}{\lambda z}\right)\text{sinc}\left(\frac{\pi\ell_y y_0}{\lambda z}\right) \tag{4.52}$$

The irradiance I of the diffraction pattern is consequently:

$$I(x_0, y_0) = \left(\frac{\ell_x\ell_y}{\lambda z}\right)^2\text{sinc}^2(\frac{\pi\ell_x x_0}{\lambda z})\text{sinc}^2(\frac{\pi\ell_y y_0}{\lambda z}) \tag{4.53}$$

Figure 4.5 shows the pattern along the x_0 axis, figure 4.6 shows an experimental far field diffraction pattern of a rectangular aperture [**?**]. Remark: in optics text books the sinc-function is usually defined as $sinc(x) = (sin\pi x)/\pi x)$.

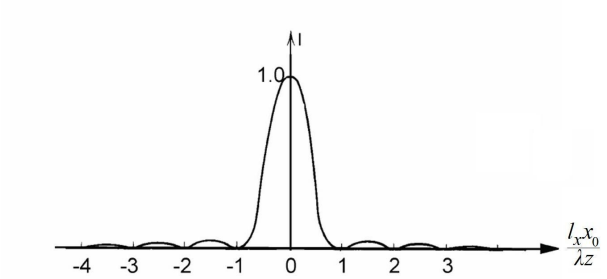


Figure 4.5: Fraunhofer diffraction pattern of a rectangular aperture

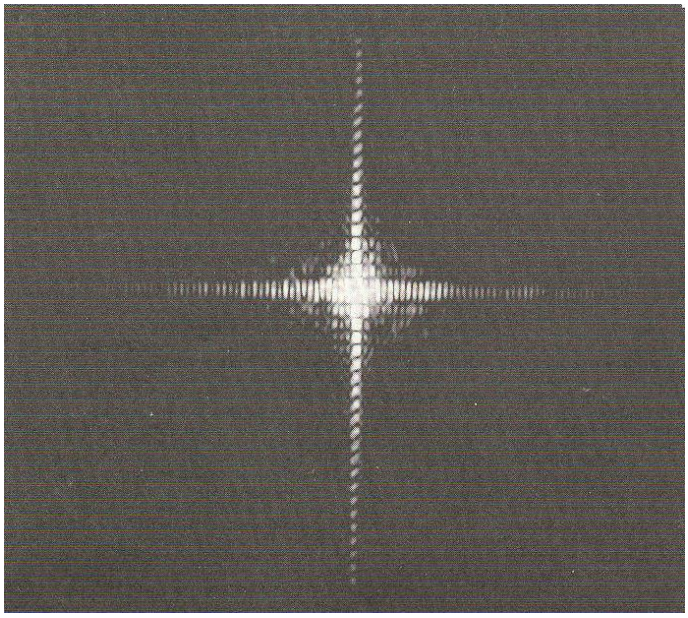


Figure 4.6: An experimental Fraunhofer diffraction pattern of a rectangular aperture (from [**?**])

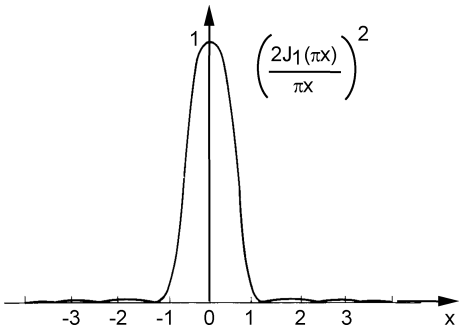


Figure 4.7: Fraunhofer diffraction pattern of a circular aperture or Airy pattern

Circular aperture

The amplitude transmittance t of a circular aperture with diameter l is given by

$$t(r_1) = \text{circ}\left[\frac{r_1}{\ell/2}\right] \tag{4.54}$$

Because of the circular symmetry, the Fourier transform in formula (4.52) reduces to a Fourier-Bessel transform B:

$$B\{f\} = 2\pi \int_0^{+\infty} r f(r) J_0(2\pi f r)\, dr \tag{4.55}$$

If, once again, the illuminating wave is a plane monochromatic wave, of unit amplitude, normally incident on the screen, then $U(r_1) = t(r_1)$.

Because

$$B\left\{\text{circ}\left(\frac{r_1}{\ell/2}\right)\right\} = \left(\frac{\ell}{2}\right)^2 \frac{J_1(\pi \ell f)}{\ell f/2} \tag{4.56}$$

one finds

$$U(r_0) = \frac{e^{-jkz}}{j\lambda z} e^{-j\frac{kr_0^2}{2z}} \left[\left(\frac{\ell}{2}\right)^2 \frac{J_1\left(\frac{\pi \ell r_0}{\lambda z}\right)}{\frac{\ell r_0}{2\lambda z}}\right] \tag{4.57}$$

or

$$U(r_0) = e^{-jkz} e^{-j\frac{kr_0^2}{2z}} \frac{k\ell^2}{j8z} \left[2\frac{J_1\left(\frac{k\ell r_0}{2z}\right)}{\frac{k\ell r_0}{2z}}\right] \tag{4.58}$$

and the irradiance becomes:

$$I(r_0) = \left(\frac{k\ell^2}{j8z}\right)^2 \left[2\frac{J_1\left(\frac{k\ell r_0}{2z}\right)}{\frac{k\ell r_0}{2z}}\right]^2 \tag{4.59}$$

This light distribution is called an $Airy$ pattern, after G.B. Airy, an astronomer who first derived it.

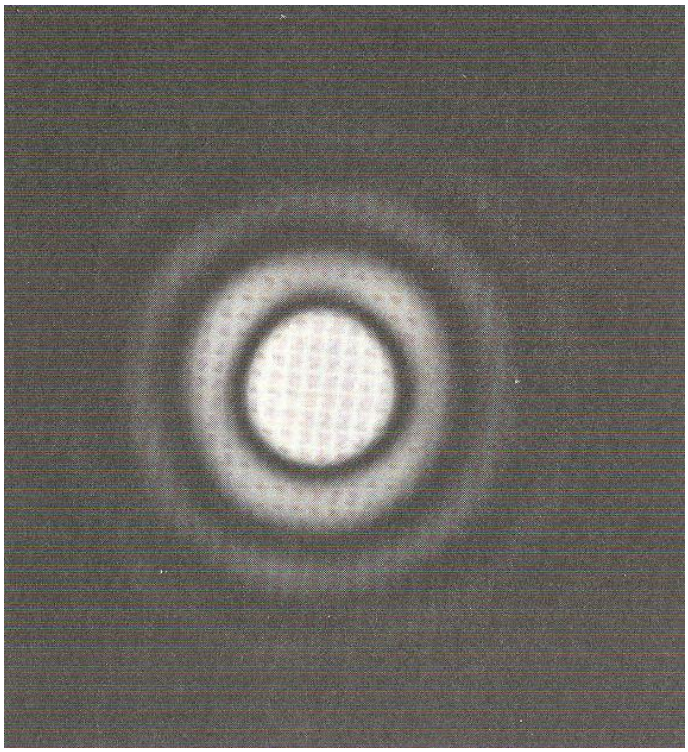


Figure 4.8: An experimental Fraunhofer diffraction pattern of a circular aperture, from [**?**]

From figure 4.7 one can see that the distance r_0 to the first zero equals $r_0 = 1.22\lambda z/\ell$. This is also the radius of the circular Airy-spot. Figure 4.8 shows an experimental far field diffraction pattern of a circular aperture [**?**].

Gaussian beam

As third example we consider a Gaussian beam with its waist in the (x_1, y_1)-plane. Hence it has a plane wavefront, and an amplitude distribution:

$$U(x_1, y_1) = e^{-\frac{x_1^2+y_1^2}{w^2}} \tag{4.60}$$

This amplitude does not go to zero at a finite distance from the axis; hence the previous formulas do not apply, at least in theory. However the exponential decreases so fast that we can still use the diffraction formulas. Because the Fourier transform of Gaussian is again a Gaussian:

$$F_{2D}\left\{e^{-\frac{x^2+y^2}{w^2}}\right\} = \pi w^2 e^{-\pi^2 w^2\left(f_x^2+f_y^2\right)} \tag{4.61}$$

we find

$$U(x_0, y_0) = -\pi\frac{w^2 e^{-jkz}}{j\lambda z} e^{\frac{-jk}{2z}\left(x_0^2+y_0^2\right)} e^{-\pi^2 w^2\left(f_x^2+f_y^2\right)} \tag{4.62}$$

with

$$f_x = -\frac{x_0}{\lambda z} \text{ and } f_y = -\frac{y_0}{\lambda z} \tag{4.63}$$

Hence in the far field a Gaussian beam behaves as a spherical wave, with half-angle:

$$\theta = \frac{\lambda}{\pi w} \tag{4.64}$$

This is exactly the same result we obtained (in another way) when studying Gaussian beams in the lectures on lasers. Condition (4.45) means that the Fraunhofer regime starts after a few times the Rayleigh range.

4.2.4 Fresnel diffraction at a square aperture

Calculations of Fresnel diffraction patterns are much more complicated than Fraunhofer ones, simply because one can not use the well-known Fourier transform formulas. We illustrate this with the simple example of a square aperture with side ℓ, normally illuminated with a monochromatic plane wave of unit amplitude. The field in the object plane is then

$$U\left(x_1, y_1\right) = \operatorname{rect}\left(\frac{x_1}{\ell}\right) \operatorname{rect}\left(\frac{y_1}{\ell}\right) \tag{4.65}$$

Application of formula (4.41) gives

$$\begin{aligned} U(x_0, y_0) &= \frac{-e^{-jkz}}{j\lambda z} \int_{-\ell/2}^{+\ell/2} e^{-\frac{jk}{2z}\left[(x_1-x_0)^2+(y_1-y_0)^2\right]} dx_1 dy_1 \\ &= \frac{-e^{-jkz}}{j\lambda z} \int_{-\ell/2}^{+\ell/2} e^{-\frac{jk}{2z}(x_1-x_0)^2} dx_1 \int_{-\ell/2}^{+\ell/2} e^{-\frac{jk}{2z}(y_1-y_0)^2} dy_1 \end{aligned} \tag{4.66}$$

Introducing new variables u and v by

$$u = \sqrt{\frac{k}{\pi z}}\left(x_1 - x_0\right) \text{ and } v = \sqrt{\frac{k}{\pi z}}\left(y_1 - y_0\right) \tag{4.67}$$

gives

$$U(x_0, y_0) = \frac{-e^{-jkz}}{j\lambda z} \frac{\pi z}{k} \int_{u_1}^{u_2} e^{-\frac{j\pi u^2}{2}} du \int_{v_1}^{v_2} e^{-\frac{j\pi v^2}{2}} dv \tag{4.68}$$

with

$$\begin{aligned} u_1 &= \sqrt{\frac{k}{\pi z}}\left(-\frac{\ell}{2} - x_0\right), \; u_2 = \sqrt{\frac{k}{\pi z}}\left(+\frac{\ell}{2} - x_0\right) \\ v_1 &= \sqrt{\frac{k}{\pi z}}\left(-\frac{\ell}{2} - x_0\right), \; v_2 = \sqrt{\frac{k}{\pi z}}\left(\frac{\ell}{2} - x_0\right) \end{aligned} \tag{4.69}$$

These integrals can be expressed in terms of the Fresnel integrals:

$$C\left(\alpha\right) = \int_0^{\alpha} \cos\frac{\pi t^2}{2} dt \text{ and } S\left(\alpha\right) = \int_0^{\alpha} \sin\frac{\pi t^2}{2} dt \tag{4.70}$$

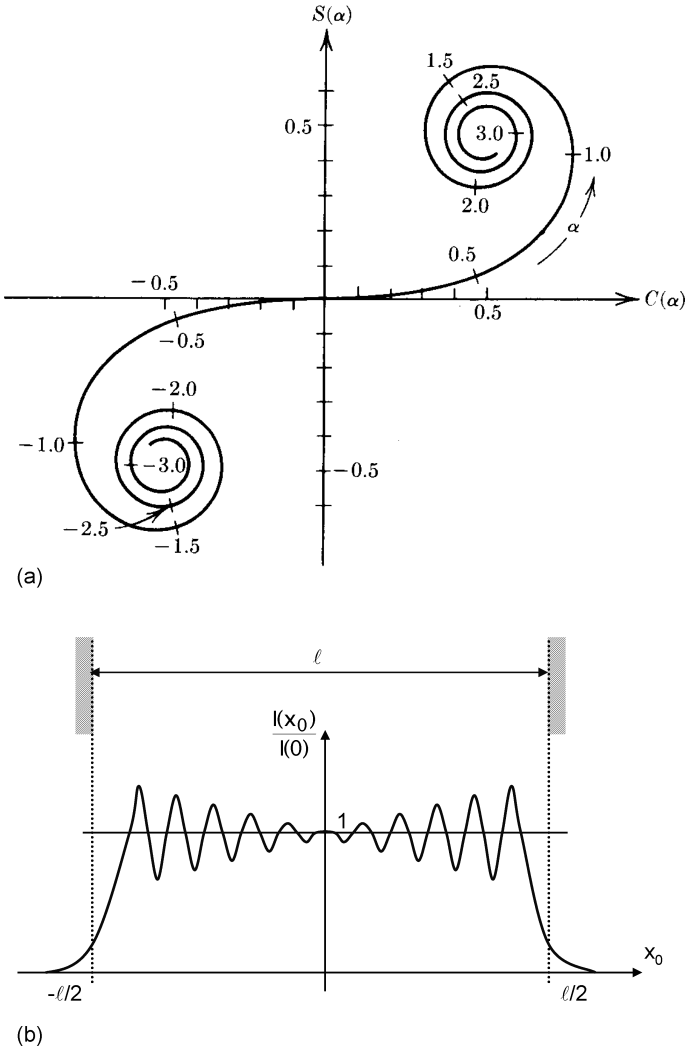


Figure 4.9: Calculation of Fresnel diffraction. (a) Cornu-spiral, (b) diffraction pattern of a slit

Because

$$\begin{aligned}\int_{u_1}^{u_2} e^{-\frac{j\pi u^2}{2}}\,du &= \int_0^{u_2}\left(\cos\frac{\pi}{2}u^2 + j\sin\frac{\pi}{2}u^2\right)du - \int_0^{u_1}\left(\cos\frac{\pi}{2}u^2 + j\sin\frac{\pi}{2}u^2\right)du \\ &= [C(u_2) - C(u_1)] - j\,[S(u_2) - S(u_1)]\end{aligned} \tag{4.71}$$

we find

$$\begin{aligned}U(x_0, y_0) = -\frac{e^{-jkz}}{2j} \quad &([C(u_2) - C(u_1)] - j\,[S(u_2) - S(u_1)])\,. \\ &([C(v_2) - C(v_1)] - j\,[S(v_2) - S(v_1)])\end{aligned} \tag{4.72}$$

and the irradiance distribution I is:

$$I(x_0) = \frac{1}{4}\left([C(u_2) - C(u_1)]^2 + [S(u_2) - S(u_1)]^2\right)\left([C(v_2) - C(v_1)]^2 + [S(v_2) - S(v_1)]^2\right) \tag{4.73}$$

The meaning of this expression can be easily understood with the help of figure 4.9a which gives $C(u)$ on the horizontal and $S(u)$ on the vertical axis as a function of the parameter u. This graph

is known as the Cornu-spiral. One can prove that the irradiance $I(x_0, y_0)$ is proportional to the length of a line segment connecting two points on the spiral. When the image point shifts, the representative points run through the spiral. Hence the irradiance oscillates strongly, as illustrated in figure 4.9b . This graph shows the Fresnel diffraction pattern of a one-dimensional slit of length l, measured at a given distance z from the slit. When the observation plane approaches the plane of the slit, the shape of the diffraction pattern approaches the shape of the slit itself. On the other hand, at larger distances, the diffraction pattern becomes much wider than the slit, ending up (at very large distances) with a Fraunhofer diffraction pattern.

4.2.5 Fresnel diffraction and spatial frequencies

Formulas (4.34) and (4.41) can also be written as:

$$U(x_0, y_0) = \int\int_S h(x_0 - x_1, y_0 - y_1) U(x_1, y_1) dx_1 dy_1 \tag{4.74}$$

with

$$h(x_0 - x_1, y_0 - y_1) \cong \frac{-e^{-jkz}}{j\lambda z} e^{-\frac{jk}{2z}\left[(x_1 - x_0)^2 + (y_1 - y_0)^2\right]} \tag{4.75}$$

You immediately recognize a convolution structure:

$$f * g = \int_{-\infty}^{+\infty} f(\tau) g(t - \tau) \, d\tau \tag{4.76}$$

Hence the image $U(x_0, y_0)$ is a two-dimensional convolution between $U(x_1, y_1)$ and

$$h(x_0, y_0) = \frac{-e^{-jkz}}{j\lambda z} e^{-\frac{jk}{2z}\left[x_0^2 + y_0^2\right]} \tag{4.77}$$

If we define the following Fourier transforms:

$$\begin{array}{ll} F_0(f_x, f_y) & = F_{2D}\{U(x_0, y_0)\} \\ F_1(f_x, f_y) & = F_{2D}\{U(x_1, y_1)\} \\ H(f_x, f_y) & = F_{2D}\{h(x_0, y_0)\} \end{array} \tag{4.78}$$

then the Fourier transform of (4.74) gives

$$F_0(f_x, f_y) = F_1(f_x, f_y) \cdot H(f_x, f_y) \tag{4.79}$$

This means that $H(f_x, f_y)$ is a *transfer function* which describes the evolution of the spatial spectrum of the light within the Fresnel diffraction regime. That transfer function $H(f_x, f_y)$ is nothing else than the two-dimensional Fourier-transform of (4.77):

$$\begin{array}{ll} H(f_x, f_y) & = F_{2D}\{h(x_0, y_0)\} \\ & = -\dfrac{e^{-jkz}}{j\lambda z} F_x\left\{e^{-\frac{jk}{2z}x_0^2}\right\} F_y\left\{e^{-\frac{jk}{2z}y_0^2}\right\} \end{array} \tag{4.80}$$

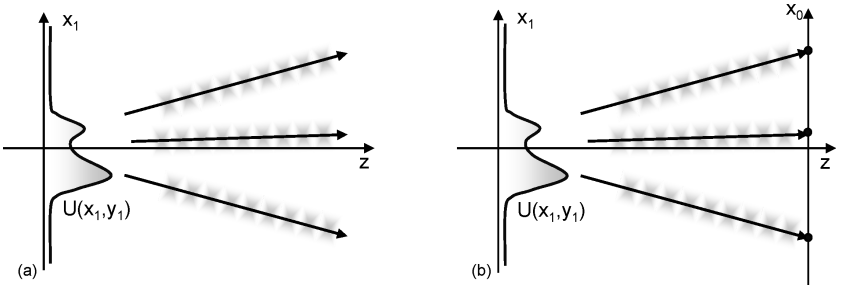


Figure 4.10: Plane-wave spectrum of an aperture

With the formula:

$$F\left\{e^{j\alpha t^2}\right\} = \sqrt{\frac{\pi}{\alpha}} e^{j\frac{\pi}{4}} e^{-j\frac{(2\pi f)^2}{4\alpha}} \tag{4.81}$$

this becomes

$$\begin{aligned} H\left(f_x, f_y\right) &= -\frac{e^{-jkz}}{j\lambda z}\sqrt{\frac{-2\pi z}{k}} e^{j\frac{\pi}{4}} e^{-j\frac{4\pi^2 f_x^2}{-4k/2z}} \sqrt{\frac{-2\pi z}{k}} e^{j\frac{\pi}{4}} e^{-j\frac{4\pi^2 f_y^2}{-4k/2z}} \\ &= -\frac{e^{-jkz}}{j\lambda z}\frac{-2\pi z}{k} j e^{-j\pi\lambda z f_x^2} e^{-j\pi\lambda z f_y^2} \\ &= e^{-jkz} e^{j\pi\lambda z\left(f_x^2+f_y^2\right)} \end{aligned} \tag{4.82}$$

4.2.6 The angular spectrum of plane waves

In this section we will give a physical explanation of the previous conclusions. Let us rewrite (4.78):

$$F_1\left(f_x, f_y\right) = \int\int_{-\infty}^{+\infty} U\left(x_1, y_1\right) e^{-j2\pi(x_1 f_x + y_1 f_y)} dx_1 dy_1 \tag{4.83}$$

and its inverse

$$U\left(x_1, y_1\right) = \int\int_{-\infty}^{+\infty} F_1\left(f_x, f_y\right) e^{+j2\pi(x_1 f_x + y_1 f_y)} df_x df_y \tag{4.84}$$

The latter formula says that the field U in the (x_1, y_1)-plane can be considered as a superposition of fields $\exp(j2\pi(x_1 f_x + y_1 f_y))$, each with its own amplitude $F_1(f_x, f_y)$. Now you know that $\exp(j2\pi(x_1 f_x + y_1 f_y))$, is nothing else than the intersection of the (x_1, y_1)-plane, with a plane wave of which the propagation **k**-vector has components:

$$k_x = 2\pi f_x, \qquad k_y = 2\pi f_y \quad \text{en} \quad k_z = \sqrt{k^2 - k_x^2 - k_y^2} \tag{4.85}$$

In other words: the Fourier decomposition of the field $U(x_1, y_1)$ is a decomposition in plane waves. This is illustrated in figure 4.10a.

This explains why the function $F_1(f_x, f_y)$ is called the *angular spectrum of plane waves*. When propagating over a distance z in a homogeneous space, each of those plane waves acquires a phase increase of $\exp(-j2\pi f_z z)$, with

$$f_z = \sqrt{\frac{1}{\lambda^2} - f_x^2 - f_y^2} \tag{4.86}$$

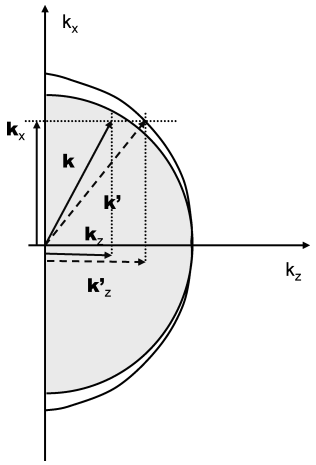


Figure 4.11: Paraboloid approximation of a spherical wavefront

The field in an arbitrary (x_0, y_0)-plane, after propagating over a distance z, is found by adding the new plane waves together:

$$U(x_0, y_0) = \int\int_{-\infty}^{\infty} F_1(f_x, f_y)\, H(f_x, f_y)\, e^{j2\pi(x_o f_x + y_o f_y)} df_x df_y \tag{4.87}$$

with

$$H(f_x, f_y) = e^{-j2\pi\sqrt{\frac{1}{\lambda^2} - f_x^2 - f_y^2}\, z} \tag{4.88}$$

For paraxial fields:

$$f_x \ll f_z \text{ en } f_y \ll f_z \tag{4.89}$$

hence

$$f_z \approx \frac{1}{\lambda}\left(1 - \frac{(\lambda f_x)^2 + (\lambda f_y)^2}{2}\right) \tag{4.90}$$

and consequently:

$$H(f_x, f_y) = e^{-jkz} e^{j\pi\lambda z\left(f_x^2 + f_y^2\right)} \tag{4.91}$$

which is exactly the same result as (4.82).

The paraxial approximation is illustrated in figure 4.11. In principle the **k**-vector lies on a circle (for a given wavelength); in the paraxial approximation it lies on a paraboloid.

We have already seen that the field at a large distance (Fraunhofer regime) is, up to a proportionality constant and a phase curvature, the Fourier transform of the original field (when the spatial frequencies f_x en f_y are expressed as function of the position x_0 en y_0). This can create some confusion, because we have now seen that this Fourier transform is in fact nothing else than a decomposition in plane waves. Those two models are not contradictory, when each of those waves are replaced by a light ray, starting at the $z = 0$ plane. The field at position (x_0, y_0, z) is then created by the ray with k-vector:

$$\begin{aligned} k_x &= 2\pi f_x = -\tfrac{2\pi}{\lambda}\tfrac{x_0}{z} \\ k_y &= 2\pi f_x = -\tfrac{2\pi}{\lambda}\tfrac{x_0}{z} \end{aligned} \tag{4.92}$$

These are exactly the same relations as in the Fraunhofer formula. This is illustrated in figure 4.10b. In reality the interpretation of the Fraunhofer formula is a little bit more complicated, because a plane wave extends to infinity, and is not a simple ray. In the next section we show mathematically how the Fresnel-field (= superposition of plane waves at each position) transforms into the Fraunhofer-field (= one single plane wave at each position). This will show that the ray model can indeed be used.

4.2.7 Transition from Fresnel to Fraunhofer regime

The transition from the Fresnel to the Fraunhofer diffraction regime can best be understood with the angular-spectrum description. We start with the angular spectrum (= decomposition in plane waves) in the $z = 0$ plane. In order to calculate the field at position z we use the propagator (4.82):

$$H(f_x, f_y) = e^{-jkz} e^{j\pi\lambda z\left(f_x^2+f_y^2\right)} \tag{4.93}$$

For simplicity we will limit ourselves here to *two* dimensions. The field at position z is then :

$$\begin{aligned} U(x,z) &= e^{-jkz} \int_{-\infty}^{+\infty} F_1(f_x)\; e^{j\pi\lambda z f_x^2}\; e^{j2\pi f_x x}\; df_x \\ &= e^{-jkz} \int_{-\infty}^{+\infty} F_1(f_x)\; e^{j\pi\left(\sqrt{\lambda z}\, f_x + \frac{x}{\sqrt{\lambda z}}\right)^2} e^{-j\pi\frac{x^2}{\lambda z}}\; df_x \end{aligned} \tag{4.94}$$

For obtaining the Fraunhofer regime, z has to be large enough. The leading term in the integrand is the exponential one:

$$e^{j\pi\left(\sqrt{\lambda z}\, f_x + \frac{x}{\sqrt{\lambda z}}\right)^2} \tag{4.95}$$

For large z values, this function oscillates very fast in f_x, except when:

$$\sqrt{\lambda z} f_x \approx -\frac{x}{\sqrt{\lambda z}} \tag{4.96}$$

The fast oscillating part does not contribute to the integral, hence:

$$\begin{aligned} U(x,z) &= e^{-jkz}\; e^{-j\pi\frac{x^2}{\lambda z}}\; F_1\left(f_x = \frac{-x}{\lambda z}\right) \int_{-\infty}^{+\infty} e^{j\pi\lambda z f_x^2}\; df_x \\ &= e^{-jkz}\; e^{-j\pi\frac{x^2}{\lambda z}}\; F_1\left(f_x = \frac{-x}{\lambda z}\right) \frac{2}{\sqrt{2\lambda z}} \int_0^{+\infty} e^{j\pi\frac{t^2}{2}} dt \quad \text{with} \quad t = \sqrt{2\lambda z}\; f_x \end{aligned} \tag{4.97}$$

Because:

$$\int_0^{+\infty} e^{j\pi\frac{t^2}{2}} dt = C(\infty) + jS(\infty) = \frac{1}{2}(1+j) \tag{4.98}$$

the field at position z (z large) is then:

$$U(x,z) = \frac{e^{-jkz}}{\sqrt{2\lambda z}}\ (1+j)\ F_1\left(f_x = \frac{-x}{\lambda z}\right)\ e^{-j\pi\frac{x^2}{\lambda z}} \tag{4.99}$$

In a similar way one can prove the general expression:

$$U(x,y,z) = \frac{e^{-jkz}}{2\lambda z}\ (1+j)^2\ F_1\left(f_x = \frac{-x}{\lambda z}, f_y = \frac{-y}{\lambda z}\right)\ e^{\frac{-j\pi}{\lambda z}(x^2+y^2)} \tag{4.100}$$

which is similar to (4.46).

4.2.8 Coherent versus incoherent fields

Up to now we always considered all fields to be monochromatic: in each point in space the field oscillates sinusoidal, with a well-known amplitude and phase. The phase *difference* between two different points in space and time is then constant. We call this kind of fields *coherent* ones. For these coherent fields the Fourier transform can be interpreted as a decomposition in plane waves, and each spatial frequency corresponds with one single direction of propagation. But light sources can also be very broadband, as for instance when white light is used. When the field is not monochromatic, things are completely different. Now there is no fixed phase relation anymore between two different points, and the fields are said to be *incoherent*. Of course it is still possible to Fourier transform that field, but this has not anymore the meaning of a decomposition in plane waves. It turns out to be practical to consider incoherent light as a superposition of monochromatic, coherent contributions. This can be illustrated by calculating the far field light distribution of a rectangular aperture. When illuminated with coherent light (eg with a laser) the image is a sinc-distribution, as we have calculated in section 2.3. If this same aperture is now illuminated with white light, the image is a very diffuse spot. This can however be written as an incoherent superposition of many sinc-contributions, each one with its own phase and own width.

4.3 Fourier transforming properties of optical systems

4.3.1 Phase transformation by a lens

When light passes through a lens, it undergoes a phase transform (figure 4.12); in this section we will calculate it for a monochromatic plane wave. A lens is composed of material (e.g. glass) in which light travels *slower* than in air; this is described by the index of refraction n. Here we assume that the lens is a *thin* lens, which implies that a light ray will leave the tangent plane of the lens T at the same transverse position (x,y) as on entering. The only effect of the lens is a retardation over a time delay which is proportional to the local thickness $\Delta(x,y)$ of the lens. If we write Δ_0 for the maximal thickness of the lens (the thickness in the middle), then the phase retardation between both tangent planes σ and σ' is given by

$$\phi(x,y) = \underbrace{kn\Delta(x,y)}_{\downarrow\ lens} + \underbrace{k\left[\Delta_0 - \Delta(x,y)\right]}_{\downarrow\ air} \tag{4.101}$$

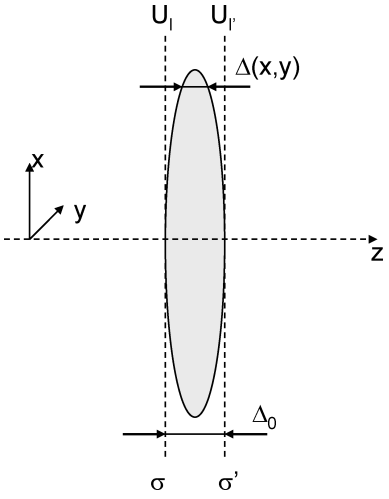


Figure 4.12: A thin lens

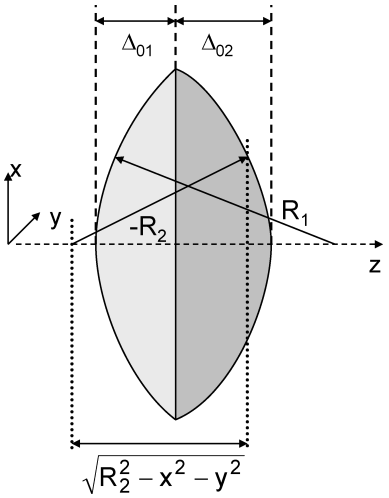


Figure 4.13: Parameters of a thin lens

Let us write $U_l(x, y)$ for the field incident on the first tangent plane; the field leaving the lens is then:

$$U_{l'}(x, y) = t_l(x, y)U_l(x, y) \tag{4.102}$$

with

$$t_l(x, y) = e^{-jk\Delta_0}e^{-jk(n-1)\Delta(x,y)} \tag{4.103}$$

For calculating $\Delta(x, y)$ we split the lens in two parts (figure 4.13) such that

$$\Delta(x, y) = \Delta_1(x, y) + \Delta_2(x, y) \tag{4.104}$$

The radius of curvature R is positive for a concave surface, and negative for a convex one. This gives

$$\begin{aligned}\Delta_1(x,y) &= \Delta_{01} - \left(R_1 - \sqrt{R_1^2 - x^2 - y^2}\right) \\ &= \Delta_{01} - R_1\left(1 - \sqrt{1 - \frac{x^2+y^2}{R_1^2}}\right)\end{aligned} \tag{4.105}$$

and

$$\begin{aligned}\Delta_2(x,y) &= \Delta_{02} - \left(-R_2 - \sqrt{R_2^2 - x^2 - y^2}\right) \\ &= \Delta_{02} + R_2\left(1 - \sqrt{1 - \frac{x^2+y^2}{R_2^2}}\right)\end{aligned} \tag{4.106}$$

Consequently

$$\Delta(x,y) = \Delta_0 - R_1\left(1 - \sqrt{1 - \frac{x^2+y^2}{R_1^2}}\right) + R_2\left(1 - \sqrt{1 - \frac{x^2+y^2}{R_2^2}}\right) \tag{4.107}$$

with

$$\Delta_0 = \Delta_{01} + \Delta_{02} \tag{4.108}$$

4.3.2 The paraxial approximation

The expression for $\Delta(x,y)$ can be simplified in the paraxial approximation; then :

$$\begin{aligned}\sqrt{1 - \frac{x^2+y^2}{R_1^2}} &\cong 1 - \frac{x^2+y^2}{2R_1^2} \\ \sqrt{1 - \frac{x^2+y^2}{R_2^2}} &\cong 1 - \frac{x^2+y^2}{2R_2^2}\end{aligned} \tag{4.109}$$

hence

$$\Delta(x,y) = \Delta_0 - \frac{x^2+y^2}{2}\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \tag{4.110}$$

Substitution of (4.110) in (4.103) gives

$$t_l(x,y) = e^{-jkn\Delta_0} e^{jk(n-1)\frac{x^2+y^2}{2}\left[\frac{1}{R_1} - \frac{1}{R_2}\right]} \tag{4.111}$$

We note that the physical parameters of the lens n, R_1 and R_2 can be combined in one single parameter f (which we call the "focal distance")

$$\frac{1}{f} = (n-1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \tag{4.112}$$

The phase transformation of the lens now becomes:

$$t_l(x,y) = e^{-jkn\Delta_0} e^{j\frac{k}{2f}\left(x^2+y^2\right)} \tag{4.113}$$

The sign conventions we have adopted for the radii of curvature R_1 and R_2 for a double convex lens in figure 4.15 can also be used for other types of lenses, see figure 4.14. You can control that for

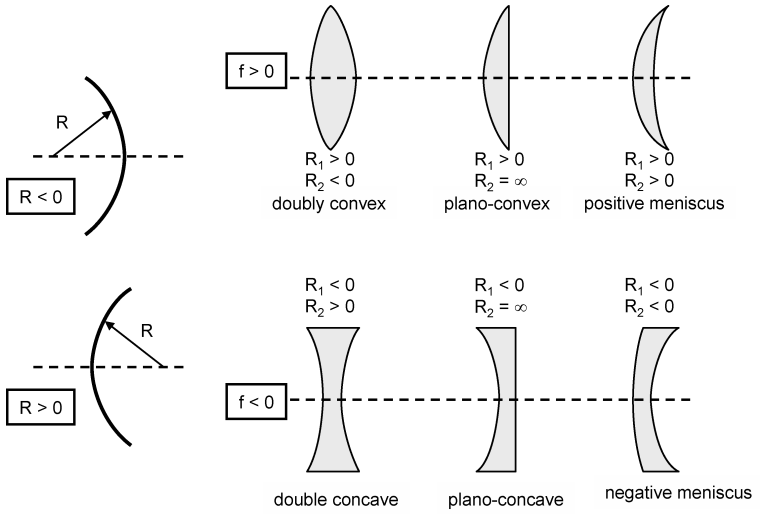


Figure 4.14: Convex and concave lenses

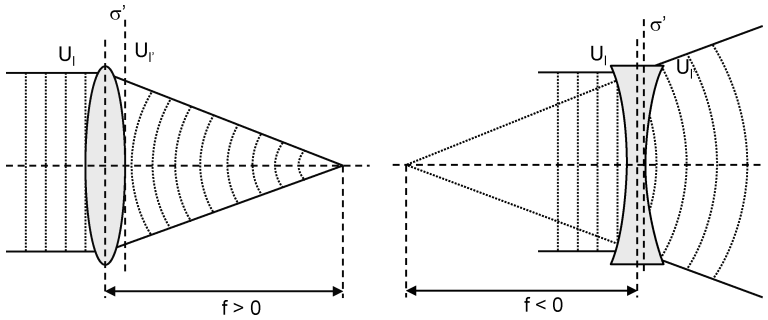


Figure 4.15: Wavefront at convex and concave lenses

the upper row in figure 4.14 the foci f are positive, whereas for the lower row they are negative. When a plane unit-amplitude wave is incident perpendicular on the lens, the field leaving it is given by:

$$U_{l'}(x,y) = e^{-jkn\Delta_0} e^{j\frac{k}{2f}\left(x^2+y^2\right)} \tag{4.114}$$

The first part gives simply a constant phase retardation, whereas the second part describes a curvature of the wavefront converging towards a point on the z-axis at a distance f from the lens (figure 4.15a).

The field in the plane σ' is indeed given by:

$$e^{jk\sqrt{f^2+x^2+y^2}} = e^{jkf\sqrt{1+\frac{x^2+y^2}{f^2}}} \tag{4.115}$$

or in the paraxial approximation:

$$e^{jkf\sqrt{1+\frac{x^2+y^2}{f^2}}} \cong e^{jkf} e^{j\frac{k}{2f}\left(x^2+y^2\right)} \tag{4.116}$$

and this is the same as in (4.114).

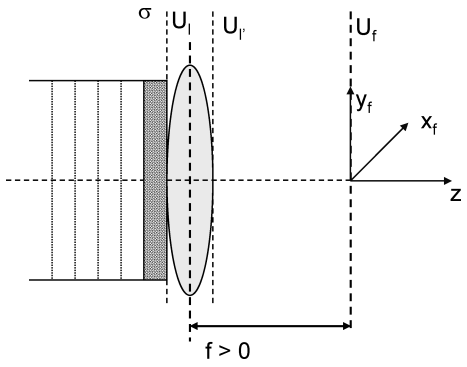


Figure 4.16: A mask placed against a lens

If the focal distance is negative, then the wave is a divergent spherical wave, the origin of which lies a distance f in front of the lens. Note that these conclusions are only valid in the paraxial approximation. If this approximation is not valid, then the wave leaving the lens is not a spherical one, and all kinds of aberrations will show up.

4.3.3 Fourier-transforming properties of a mask placed against a lens

Suppose now we position a (gray-scale) mask with amplitude transmittance $t_0(x, y)$ in a plane σ just in front of a lens (figure 4.16) which is supposed to be larger than the mask. This is now uniformly illuminated by a normally incident, monochromatic plane wave of amplitude A, propagating along the $+z$ axis.

The field U incident on the lens is then

$$U_l\left(x,y\right) = At_0\left(x,y\right) \tag{4.117}$$

Formula (4.113) gives the field immediately behind the lens:

$$U_{l'}(x,y) = U_l(x,y)e^{-jkn\Delta_0}e^{j\frac{k}{2f}\left(x^2+y^2\right)} \tag{4.118}$$

This field propagates further along the z-axis. The field at a distance $z = f$ (i.e. in the back focal plane) can be calculated with the Fresnel diffraction formula (4.42).

$$U_f(x_f,y_f) = -\frac{e^{-jkf}}{j\lambda f}e^{-j\frac{k}{2f}\left(x_f^2+y_f^2\right)}\int\limits_{-\infty}^{+\infty}\int U_{l'}(x,y)e^{-j\frac{k}{2f}\left(x^2+y^2\right)}e^{j\frac{2\pi}{\lambda f}\left(xx_f+yy_f\right)}dxdy \tag{4.119}$$

After substituting U_l one finds

$$U_f(x_f,y_f) = -\frac{e^{-jkf}}{j\lambda f}e^{-j\frac{k}{2f}\left(x_f^2+y_f^2\right)}e^{-jkn\Delta_0}A\int\limits_{-\infty}^{+\infty}\int t_0(x,y)e^{j\frac{2\pi}{\lambda f}\left(xx_f+yy_f\right)}dxdy \tag{4.120}$$

This implies that the field in the back focal plane is proportional to the two-dimensional Fourier transform of the transmittance function $t_0(x, y)$ of the object/mask (or, in general, the field incident

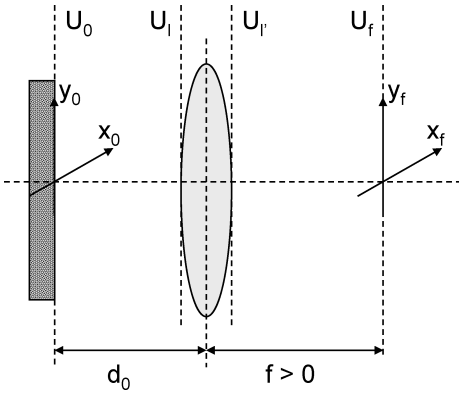


Figure 4.17: A mask at a distance d_o in front of a lens

on the lens), in which the spatial frequencies (f_x, f_y) and the positions in the focal plane (x_f, y_f) are related by:

$$\begin{aligned} f_x &= -\frac{x_f}{\lambda f} \\ f_y &= -\frac{y_f}{\lambda f} \end{aligned} \tag{4.121}$$

This Fourier-transform relation is not an exact one, because of the quadratic phase factor

$$e^{-j\frac{k}{2f}\left(x_f^2+y_f^2\right)} \tag{4.122}$$

which is not constant in the focal plane. However, this phase factor disappears when calculating the power distribution in the focal plane. Consequently the power distribution in the focal plane is exactly given by the *power spectrum* of the mask. This means that without a lens the far field (i.e. the plane wave decomposition) is found at large distance, but with the lens the far field is found in the lens focal plane (obviously with a different scaling factor).

4.3.4 Fourier-transform properties of a mask placed a distance in front of a lens

The same mask is now placed a distance d_0 in front of the lens and illuminated in the same way (figure 4.17).

We assume that d_o is large enough, so that we may use the Fresnel diffraction formula. This means that we can apply formulas (4.79) and (4.82).

Let us call

$$F_0(f_x, f_y) = F_{2D}\left\{At_0\right\} \tag{4.123}$$

and

$$F_l(f_x, f_y) = F\left\{U_l\right\} \tag{4.124}$$

then

$$F_l(f_x, f_y) = F_0(f_x, f_y)e^{j\pi\lambda d_0\left(f_x^2+f_y^2\right)}e^{-jkd_0} \tag{4.125}$$

Expression (4.120) can now be rewritten as

$$U_f(x_f, y_f) = -\frac{e^{-jkf}}{j\lambda f}e^{-jkn\Delta_0}e^{-j\frac{k}{2f}\left(x_f^2+y_f^2\right)}F_l\left(-\frac{x_f}{\lambda f}, -\frac{y_f}{\lambda f}\right) \tag{4.126}$$

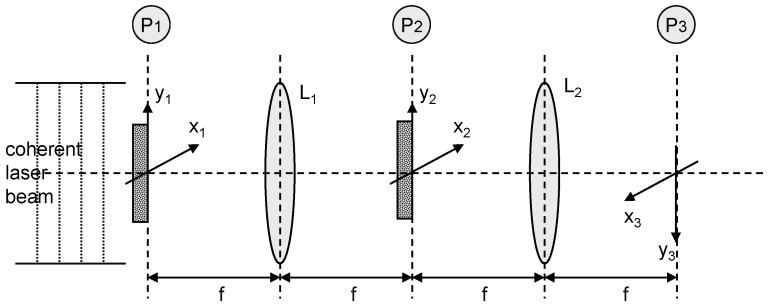


Figure 4.18: The optical convolution processor

Substituting (4.125) in (4.126) gives, after deleting a constant phase factor:

$$U_f(x_f, y_f) = -\frac{1}{j\lambda f} e^{-j\frac{k}{2f}\left(x_f^2+y_f^2\right)} e^{j\frac{\pi\lambda d_0}{\lambda^2 f^2}\left(x_f^2+y_f^2\right)} F_0\left(-\frac{x_f}{\lambda f}, -\frac{y_f}{\lambda f}\right) \tag{4.127}$$

or

$$U_f(x_f, y_f) = -\frac{A}{j\lambda f} e^{-j\frac{k}{2f}\left(1-\frac{d_0}{f}\right)\left(x_f^2+y_f^2\right)} \int_{-\infty}^{+\infty} \int t_0(x,y).e^{+j\frac{2\pi}{\lambda f}(xx_f+yy_f)} dxdy \tag{4.128}$$

By choosing $d_0 = f$ the exponential in front of the integral disappears, and we end up with an exact Fourier transform between the transmittance t_0 and the field in the focal plane. In these calculations we neglected the finite transverse dimensions of the lens; this is allowed whenever the mask is small as compared to the lens.

4.3.5 Optical convolution processor

It is possible to realize, in an optical set up, a convolution between two functions

$$\left| \int_{-\infty}^{+\infty} g\left(\xi, \eta\right) h\left(x-\xi, y-\eta\right) d\xi d\eta \right| \tag{4.129}$$

The set up is shown in figure 4.18. The "input"-function g is realized as a mask with transmittance function $g(x_1, y_1)$ and placed in plane P_1, which is the first focal plane of a lens L_1. It is normally illuminated with a monochromatic plane wave. In the second focal plane P_2 of this lens one obtains the Fourier transform $k_1 G(-x_2/\lambda f, -y_2/\lambda f)$ of g (in which k_1 = complex constant). In this plane P_2 one puts a second mask with transmittance function

$$t(x_2, y_2) = k_2 H\left(\frac{-x_2}{\lambda f}, \frac{-y_2}{\lambda f}\right) \tag{4.130}$$

with

$$H\left(f_x, f_y\right) = F_{2D}\left\{h\left(x, y\right)\right\} \tag{4.131}$$

Behind this mask, the field is proportional to $G.H$; hence in the focal plane P_3 of the second lens L_2 one finds the following irradiance:

$$I(x_3, y_3) = K \left| \int\int_{-\infty}^{+\infty} g(\xi, \eta)\, h(x_3 - \xi, y_3 - \eta)\, d\xi d\eta \right|^2 \tag{4.132}$$

This kind of convolution processor can be practically realized by inserting in the planes P_1 and P_2 a LC-SLM (Liquid Crystal Spatial Light Modulator). This is a two-dimensional transmission screen, with which one can realize an arbitrary two-dimensional function with the help of a computer. In plane P_3 one puts a CCD (Charged Coupled Device) Camera which transforms the optical information in an electrical one. An advantage of this processor is the fact that the "calculations" are done immediately (the speed is only limited by the input and output speed of the LC-SLM and CCD). Disadvantage (as compared to the calculations with a digital computer) is the analog character of the calculations, with its inaccuracy, the non-linearity in the LC-SLM and CCD, and also the aberrations in the lenses. Moreover there is also a technological problem: the function H is usually a complex function. This implies that in plane P_2 one needs an SLM in which not only the amplitude, but also the phase transmittance should be simulated electronically; and this is a very difficult technological problem.

4.4 Resolving power of an optical system

In the previous section we always assumed that the lenses we used were much larger than the transparencies; moreover we only calculated the field in the $focal$ plane. In this section we look for the field or image in an $image$ plane; moreover we take the finite dimension of the lens into account. A perfect imaging system has an infinite resolving power. In real systems, the resolving power is limited, mainly because of two reasons. First: geometrical aberrations limit the sharpness of the image, and consequently fine details get lost. In principle those aberrations can be reduced, almost as much as one wishes. The second reason is the presence of *diffraction*. Indeed, as every optical system has only finite transverse dimensions, one always finds a diffraction spot in the image. This spot can be reduced by increasing the diameter of the optical components, but this invariably increases the geometrical aberrations; hence it is not clear whether the overall quality of the image will improve.

The resolving power is usually defined by considering two points in the object, and calculating the image of them, to see whether or not those images overlap. This method is similar to the analysis of linear systems with an impulse response technique: indeed, each object can be considered as being a collection of points; the image is then the collection of all image points or spots.

It turns out that the resolving power of an optical system differs whether the object is coherently or incoherently illuminated, because diffraction depends on the degree of coherence. For perfect coherent illumination, the image is build up by adding the *complex* amplitudes of all the diffraction patterns of all points of the object. With incoherent illumination on the other hand, one has to add all the irradiances of the image-points. It should be clear that the latter always gives smoother images: a complex amplitude can indeed be negative, giving ripples due to destructive interference. But it is less clear what influence this has on the resolving power of the optical system. It turns out that this depends on the actual optical system: depending on the phase relations in the object

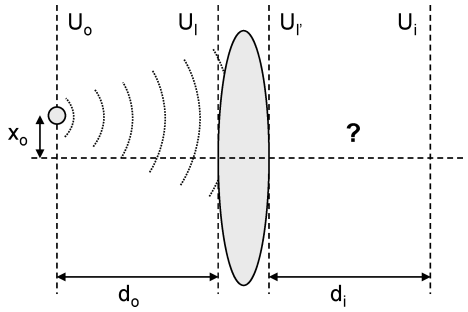


Figure 4.19: Image of a point source through a thin lens

field, the coherent image can either look sharper or just more hazy than the incoherent one! This is for example important in microscopy, where a correct choice of the illumination can dramatically increase the quality of the final image.

We first start with a simple set up: consider an object composed of one single point, and an optical system with one single aberration-free thin lens. What is the image? The image is described with an index i ("image": coordinates x_i, y_i); in the object plane we have an index o ("object": coordinates x_o, y_o). Due to the linearity of the wave-propagation phenomenon,we can always write:

$$U_i\left(x_i, y_i\right) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h\left(x_i, y_i; x_o, y_o\right) U_o\left(x_o, y_o\right) dx_o dy_o \tag{4.133}$$

for perfect coherent illumination, and

$$\left|U_{\,i}\left(x_i, y_i\right)\right|^2 = \kappa \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left|h\left(x_i, y_i; x_o, y_o\right)\right|^2 \left|U_o\left(x_o, y_o\right)\right|^2 dx_o dy_o \tag{4.134}$$

for an incoherent object, with

$$\kappa = \frac{1}{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left|h(0,0;x,y)\right|^2 dxdy} \tag{4.135}$$

The function $h(x_i, y_i; x_o, y_o)$ is called the *point spread function* or PSF; it is the image of a unit-amplitude point source at position (x_o, y_o). This is the function we are looking for. The incident wave is a spherical one, paraxially described by :

$$h_l\left(x_l, y_l; x_o, y_o\right) = \frac{-e^{-jkd_o}}{j\lambda d_o} \exp\left(-j\frac{k}{2d_o}\left[\left(x_l - x_o\right)^2 + \left(y_l - y_o\right)^2\right]\right) \tag{4.136}$$

Behind the lens we have (apart from a constant phase factor)

$$h_{l'}\left(x_{l'}, y_{l'}; x_o, y_o\right) = h_l\left(x_{l'}, y_{l'}; x_o, y_o\right) P\left(x_{l'}, y_{l'}\right) \exp\left(j\frac{k}{2f}\left[x_{l'}^2 + y_{l'}^2\right]\right) \tag{4.137}$$

The function $P(x, y)$ is the pupil function (it is 1 inside the lens, 0 outside).

Application of the Fresnel diffraction formula gives:

$$h\left(x_i,y_i;x_o,y_o\right) = \\ \frac{-e^{-jkd_i}}{j\lambda d_i}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h_{l'}\left(x_{l'},y_{l'};x_o,y_o\right)\exp\left(-j\frac{k}{2d_i}\left[\left(x_i-x_{l'}\right)^2+\left(y_i-y_{l'}\right)^2\right]\right)dx_{l'}dy_{l'} \tag{4.138}$$

Combining those expressions we find :

$$\begin{aligned} h\left(x_i,y_i;x_o,y_o\right) = \quad & \frac{-e^{-jk(d_o+d_i)}}{\lambda^2 d_o d_i}\exp\left(-j\frac{k}{2d_i}\left[x_i^2+y_i^2\right]\right)\exp\left(-j\frac{k}{2d_o}\left[x_o^2+y_o^2\right]\right) \\ & \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} P\left(x,y\right)\exp\left(-j\frac{k}{2}\left(\frac{1}{d_o}+\frac{1}{d_i}-\frac{1}{f}\right)\left(x^2+y^2\right)\right) \\ & \exp\left(+jk\left[\left(\frac{x_o}{d_o}+\frac{x_i}{d_i}\right)x+\left(\frac{y_o}{d_o}+\frac{y_i}{d_i}\right)y\right]\right)\,dxdy \end{aligned} \tag{4.139}$$

The quadratic phase terms in front of the integral can be neglected for object and image locations close to the axis and will be omitted hereafter; moreover we know that, in imaging:

$$\frac{1}{d_o}+\frac{1}{d_i}=\frac{1}{f} \tag{4.140}$$

So we find:

$$h\left(x_i,y_i;x_o,y_o\right) = \\ \frac{-e^{-jk(d_o+d_i)}}{\lambda^2 d_o d_i}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} P\left(x,y\right)\exp\left(+jk\left[\left(\frac{x_o}{d_o}+\frac{x_i}{d_i}\right)x+\left(\frac{y_o}{d_o}+\frac{y_i}{d_i}\right)y\right]\right)dxdy \tag{4.141}$$

Because the lateral magnification M of a thin lens equals $-d_i/d_o$, we can write:

$$h\left(x_i,y_i;x_o,y_o\right) = \\ \frac{-e^{-jk(d_o+d_i)}}{\lambda^2 d_o d_i}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} P\left(x,y\right)\exp\left(+j\frac{2\pi}{\lambda d_i}\left[\left(x_i-Mx_o\right)x+\left(y_i-My_o\right)y\right]\right)dxdy \tag{4.142}$$

This implies that the impulse response is nothing else than the Fraunhofer diffraction pattern (= Fourier transform) of the exit pupil, centered on the image coordinates $x_i = Mx_o$ and $y_i = My_o$. Further on we see that the impulse response only depends on the combined variables $x_i - Mx_o$ and $y_i - My_o$; in other words : on the relative position with respect to the ideal geometrical image point. If we now introduce new variables:

$$x_o^{'} = Mx_o \quad \text{and} \quad y_o^{'} = My_o \tag{4.143}$$

and

$$x' = \frac{x}{\lambda d_i} \quad \text{and} \quad y' = \frac{y}{\lambda d_i} \tag{4.144}$$

then we finally find

$$h\left(x_i,y_i;x_o^{'},y_o^{'}\right) = \\ -e^{-jk(d_o+d_i)}\frac{d_i}{d_o}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} P\left(\lambda d_i x^{'},\lambda d_i y^{'}\right)\exp\left(+j2\pi\left[\left(x_i-x_o^{'}\right)x^{'}+\left(y_i-y_o^{'}\right)y^{'}\right]\right)dx^{'}dy^{'} \tag{4.145}$$

Because this impulse response is space-invariant, the integral in (4.133) turns out to be a convolution:

$$U_i(x_i, y_i) = \frac{1}{M^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h\left(x_i - x_o^{'}, y_i - y_o^{'}\right) U_o\left(\frac{x_o^{'}}{M}, \frac{y_o^{'}}{M}\right) dx_o^{'} dy_o^{'} \tag{4.146}$$

If we now define:

$$h' = \frac{1}{M^2} h \tag{4.147}$$

then we obtain:

$$U_i(x_i, y_i) = h'(x_i, y_i) * U_o\left(\frac{x_i}{M}, \frac{y_i}{M}\right) \tag{4.148}$$

The function $U_o\left(\frac{x_i}{M}, \frac{y_i}{M}\right)$ is the *perfect* image, as found in paraxial geometrical optics. The final image is then the convolution of this perfect geometrical image with the function h', given by:

$$h'(x_i, y_i) = e^{-jk(d_o+d_i)} \frac{1}{M} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} P\left(\lambda d_i x', \lambda d_i y'\right) \exp\left(j2\pi \left[x_i x' + y_i y'\right]\right) dx' dy' \tag{4.149}$$

This is in fact the inverse Fourier transform of the pupil function. The factor $1/M$ in this formula shows that the image, which is M times larger than the object (a consequence of formula 4.148), is also M times weaker; the power concentration in two dimensions is consequently M^2 times weaker. This is of course the law of conservation of energy.

In the previous calculations we considered a very simple optical system: namely one single thin lens. More complicated optical systems can always be reduced to a single thick lens, at least in the paraxial approximation, when they are aberration-free. It can then be shown that the main conclusions remain valid, on condition that the aperture is chosen to be the exit pupil of the system. When the pupil $P(x, y)$ is a slit, then h' is a sinc-function. Mostly, however, the aperture is circular; then $h'(x, y)$ is the 2-dimensional Fourier transform of this function, which gives the well known Airy-function for $|h'|^2$. Consequently:

$$\left|h'(r)\right|^2 = \left(\frac{kD}{8d_i}\right)^2 \left[2\frac{J_1(kDr/2z)}{kDr/2z}\right]^2 \tag{4.150}$$

In this formula D is the diameter of the exit pupil, d_i is the distance between this pupil and the image plane, k is the wavenumber and r is the radial coordinate in the image plane. The first zero of the Airy pattern lies at a distance

$$r = 1.22\frac{\lambda d_i}{D} \tag{4.151}$$

There exist different conventions for defining the resolving power of an optical system. One of the possibilities is the so called *Rayleigh criterion*, according to which two points are just resolved when the center of one of the Airy patterns coincides with the first zero of the other one. This implies that the expression above gives the distance in the image plane between two just-resolved points. The distance in object plane can then be calculated using the transverse magnification M of the optical system.

One also knows that the NA (= Numerical Aperture) of a lens, as seen from object space, is approximated by

$$NA_o = \frac{D}{2d_o} \tag{4.152}$$

The corresponding NA in image space is then

$$NA_i = \frac{D}{2d_i} \tag{4.153}$$

and consequently:

$$\begin{aligned} \text{resolution in object plane} &= 0.61\frac{\lambda}{NA_o} \approx \frac{\lambda}{NA_o} \\ \text{resolution in image plane} &= 0.61\frac{\lambda}{NA_i} \approx \frac{\lambda}{NA_i} \end{aligned} \tag{4.154}$$

For a telescope (in which the object lies at $-\infty$), it is better to work with the $angular$ resolution in object space. It is:

$$\text{angular resolution} = 1.22\frac{\lambda}{D} \approx \frac{\lambda}{D} \tag{4.155}$$

All those formulas clearly show that a good resolution is only possible when using optical systems with a large diameter D.

By taking profit of the convolution structure, it is also possible to derive the resolution from a transfer concept in the spatial frequency domain. Indeed, from

$$U_i\left(x_i, y_i\right) = h'\left(x_i, y_i\right) * U_o\left(\frac{x_i}{M}, \frac{y_i}{M}\right) \text{ (coherent)} \tag{4.156}$$

and

$$\left|U_i\left(x_i, y_i\right)\right|^2 = \kappa\left|h'\left(x_i, y_i\right)\right|^2 * \left|U_o\left(\frac{x_i}{M}, \frac{y_i}{M}\right)\right|^2 \text{ (incoherent)} \tag{4.157}$$

one concludes, after Fourier transformation:

$$U_i^F\left(f_x, f_y\right) = M^2 H\left(f_x, f_y\right) U_o^F\left(Mf_x, Mf_y\right) \text{ (coherent)} \tag{4.158}$$

and

$$I_i^F\left(f_x, f_y\right) = M^2 O\left(f_x, f_y\right) I_o^F\left(Mf_x, Mf_y\right) \text{ (incoherent)} \tag{4.159}$$

In these formulas the functions U_o^F and H are, in the coherent case, the Fourier transforms of the functions U_o and $h^{'}$; whereas in the incoherent case I_o^F and O are the Fourier transforms of $|U_o|^2$ and $\kappa|h^{'}|^2$ respectively. $H(f_x)$ is the coherent transfer function, and $O(f_x, f_y)$ is called the OTF or the optical transfer function. The absolute value of $O(f_x, f_y)$ is the *modulation transfer function* (MTF), and the phase of it is phase transfer function. It is possible to demonstrate the following general relations:

$$\begin{aligned} &O(f_x, f_y) \leq O(0) = 1 \\ &O(f_x, f_y) = -O(-f_x, f_y) = -O(f_x, -f_y) = O(-f_x, -f_y) \end{aligned} \tag{4.160}$$

With the autocorrelation theory one can further prove that:

$$O(f_x, f_y) = \frac{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} H(f_x^{'}, f_y^{'})\, H^*(f_x^{'} - f_x, f_y^{'} - f_y) df_x^{'} df_y^{'}}{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left|H(f_x^{'}, f_y^{'})\right|^2 df_x^{'} df_y^{'}} \tag{4.161}$$

in which H is the Fourier transform of $h^{'}$. A simple change in variables finally gives:

$$O(f_x, f_y) = \frac{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} H(f_x^{'} + f_x/2, f_y^{'} + f_y/2)\, H^*(f_x^{'} - f_x/2, f_y^{'} - f_y/2) df_x^{'} df_y^{'}}{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left|H(f_x^{'}, f_y^{'})\right|^2 df_x^{'} df_y^{'}} \tag{4.162}$$

This formula shows the relation between the coherent transfer function $H(f_x, f_y)$ and the incoherent optical transfer function $O(f_x, f_y)$. Applying the theorem of Parseval (a property of Fourier transforms) shows that

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left|H(f_x^{'}, f_y^{'})\right|^2 df_x^{'} df_y^{'} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} |h(0,0,x,y)|^2 \, dxdy = \frac{1}{\kappa} \tag{4.163}$$

It is possible to measure the MTF of an optical system simply by incoherently illuminating a mask on which a sinusoidal transmittance pattern is written. This pattern has only one single spatial frequency (the function $I_o^F(f_x, f_y)$ is Dirac-distribution). One measures then the amplitude of the sinusoidal image; this gives already 1 point of the MTF. Repeating this procedure for different spatial frequencies finally gives the complete MTF.

There is a simple relation between the coherent transfer function H and the pupil function P. This relation can be found by comparing the inverse fourier transform of h

$$h'(x_i, y_i) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} H(f_x, f_y) \exp\left(j2\pi\left[x_i f_x + y_i f_y\right]\right) df_x df_y \tag{4.164}$$

with formula 4.149 to find that (the constant phase term can be neglected.) :

$$H(f_x, f_y) = \frac{1}{M} P(\lambda d_i f_x, \lambda d_i f_y) \tag{4.165}$$

Consequently the coherent transfer function is proportional to the pupil function itself (after an appropriate change of variables). The incoherent transfer function is then

$$O(f_x, f_y) = \frac{\int_{-\infty}^{+\infty} P\left(x' + \frac{\lambda d_i f_x}{2}, y' + \frac{\lambda d_i f_y}{2}\right) P\left(x' - \frac{\lambda d_i f_x}{2}, y' - \frac{\lambda d_i f_y}{2}\right) dx'dy'}{\int_{-\infty}^{+\infty} |P(x', y')|^2 \, dx'dy'} \tag{4.166}$$

The pupil function equals 1 inside and 0 outside the aperture. Consequently the denominator in the expression above is nothing else than the area of the exit pupil. The numerator is the area of overlap between the aperture and its own shifted image, after shifting over $\lambda d_i f_x$ in the x-

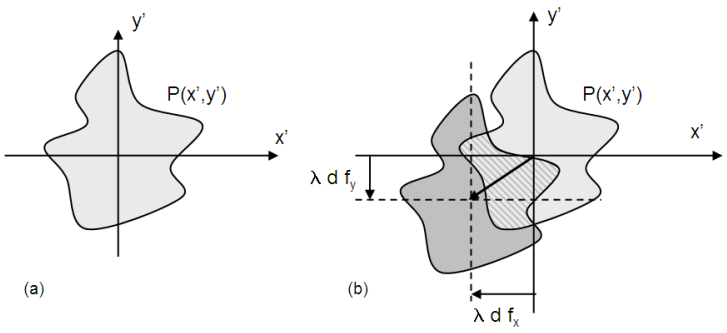


Figure 4.20: Calculation of the overlap of an aperture with itself

and $\lambda d_i f_y$ in the y-direction. This is shown in figure 4.20. From this geometrical interpretation it follows that the OTF of an aberration-free system is always real and non-negative (and hence the OTF equals the MTF). The function is not necessarily monotonically decreasing.

Let us now have a closer look at the coherent and incoherent transfer function for a diffraction-limited optical system with a square resp. circular exit pupil. The square pupil has a pupil function:

$$P(x,y) = \text{rect}\left(\frac{x}{D}\right)\text{rect}\left(\frac{y}{D}\right) \tag{4.167}$$

Which gives a coherent transfer function:

$$H\left(f_x,f_y\right) = \text{rect}\left(\frac{\lambda d_i f_x}{D}\right)\text{rect}\left(\frac{\lambda d_i f_y}{D}\right) \tag{4.168}$$

This has cutoff frequencies in x- and y-direction given by:

$$f_c = \frac{D}{2\lambda d_i} \tag{4.169}$$

The incoherent transfer function is easy to calculate. One obtains:

$$O\left(f_x,f_y\right) = \Lambda\left(\frac{f_x}{2f_c}\right)\Lambda\left(\frac{f_y}{2f_c}\right) \tag{4.170}$$

Λ is the triangle function (with a value of 1 when the argument equals 0, 0 when the argument is larger than or equal to 1, and linear in between).

The circular aperture gives:

$$\begin{aligned} P(x,y) &= \text{circ}\left(\frac{\sqrt{x^2+y^2}}{D/2}\right) \\ H\left(f_x,f_y\right) &= \text{circ}\left(\frac{2\lambda d_i\sqrt{f_x^2+f_y^2}}{D}\right) \\ f_c &= \frac{D}{2\lambda d_i} \end{aligned} \tag{4.171}$$

The OTF is, once again, easy to present in a graph, but rather difficult to calculate analytically. One finds:

$$\begin{aligned} O(f) &= \frac{2}{\pi}\left[\arccos\left(\tfrac{f}{2f_c}\right) - \tfrac{f}{2f_c}\sqrt{1-\left(\tfrac{f}{2f_c}\right)^2}\right] \quad \text{when } f \leq 2f_c \\ O(f) &= 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \text{when } f \geq 2f_c \end{aligned} \tag{4.172}$$

with $f = \sqrt{f_x^2 + f_y^2}$.

Figure 4.23 shows both OTF's as function of the f-coordinate. It is interesting to compare the maximum frequency that can be resolved by the system in the coherent and incoherent case. In the coherent case (equation (4.168)), frequencies up to f_c can be resolved. In the incoherent case (equation (4.170)), the maximum resolvable frequency equals $2f_c$! To understand this intuitively, consider figure 4.21. As explained in 4.2.8, fields can be decomposed into plane waves, where each spatial frequency corresponds to one single direction of propagation. Since the lens has finite

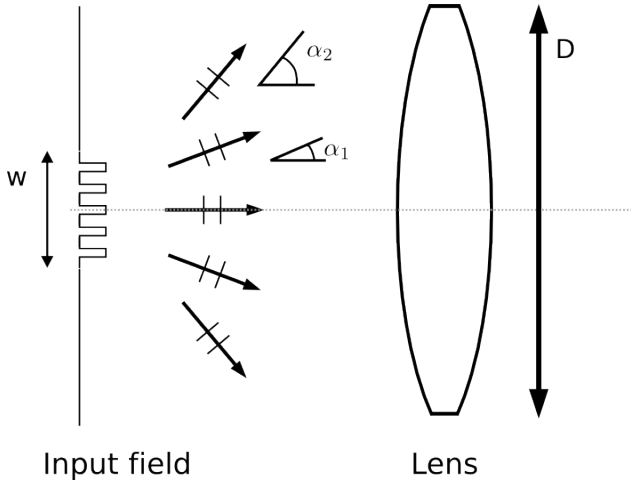


Figure 4.21: Decomposition of a periodic image into plane waves. The lens has a finite diameter D.

dimensions, it cannot capture all incoming angles (in figure 4.21 for example, α_2 is not imaged anymore). When even the angle α_1 is no more captured by the lens, we have reached the cut-off frequency, and only the zeroth-order component is captured. We also saw that incoherent light can be considered as a superposition of coherent contributions, whereby all possible phase relationships between different parts of the field should be taken into account. If we now look at the intensity distribution depicted in 4.22, and consider two specific field distributions -as shown in the figure- from the ensemble of all possible field distributions with the same intensity distribution, we see that, from a coherent point of view, the frequency of the second field is only half the frequency of the original intensity profile. If the frequency is halved, the angles are halved, and as such more frequencies are captured by the lens. In other words, an intensity distribution with frequency f can be considered to be the result of the superposition of an ensemble of field distributions, including a field distribution with frequency f and another with frequency f/2. We now showed the two most extreme cases, all other contributions lie somewhere in between.

In the discussion above we completely neglected possible geometrical aberrations in the optical system. In principle it is not allowed to use the concept of transfer function when there are aberrations, simply because the optical system is not space invariant anymore; at best, one could use a

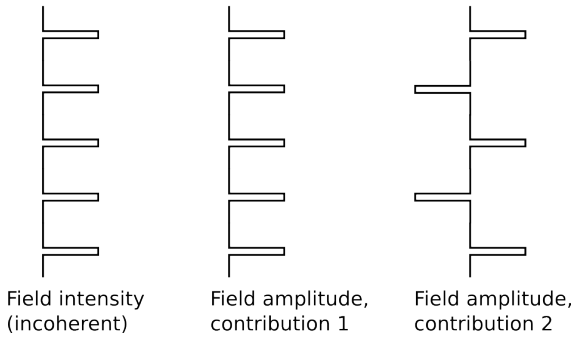


Figure 4.22: The incoherent field can be seen as a sum of coherent contributions. The two pictures on the right show two such contributions. In the second case, the odd rows are phase-shifted by π.

"local" transfer function. Nevertheless, in practical situations, one still continues to use them. The MTF of an optical system with aberrations is always worse than in a system without aberrations: for each spatial frequency its value is lower than in an aberration-free system. This is because the wavefront leaving the optical system is no longer a spherical one, but shows phase aberrations, which implies that the OTF now becomes complex. We can model this by assuming the pupil is still illuminated by a perfect spherical wave, but that a plate inside the aperture shifts the phase. The plate induces an effective path-length error of $kW(x,y)$. The pupil function now becomes complex, and is referred to as the generalized pupil function:

$$P_{gen}(x,y) = P(x,y)e^{jkW(x,y)} \tag{4.173}$$

To illustrate this, we consider the simple case of a defocussing error ϵ:

$$\frac{1}{d_i} + \frac{1}{d_o} - \frac{1}{f} = \epsilon, \tag{4.174}$$

If we look back to (4.139), and keeping in mind the new generalized pupil function, W(x,y) can easily be found to be:

$$W(x,y) = \frac{\epsilon\left(x^2+y^2\right)}{2} \tag{4.175}$$

After some math, one finds the OTF for the system with defocussing ϵ and a rectangular aperture of width D:

$$\begin{aligned} O\left(f_x,f_y\right) &= \Lambda\left(\frac{f_x}{2f_c}\right)\Lambda\left(\frac{f_y}{2f_c}\right) \\ &\quad \operatorname{sinc}\left[\frac{\epsilon D^2}{\lambda}\left(\frac{f_x}{2f_c}\right)\left(1-\frac{|f_x|}{2f_c}\right)\right]\operatorname{sinc}\left[\frac{\epsilon D^2}{\lambda}\left(\frac{f_y}{2f_c}\right)\left(1-\frac{|f_y|}{2f_c}\right)\right] \end{aligned} \tag{4.176}$$

For ϵ=0, this equation simplifies to (4.170). Note that the OTF is real. More in general, if $W(x,y)$ is centrosymmetric, the OTF will be real (this can easily be seen from the definition of the OTF).

As one can see, for some frequencies, the MTF can become zero or even negative. Those frequencies give a good resolving power, but the image is reversed. At the zeros of the MTF there is no resolving power. This is illustrated in figure 4.24 and figure 4.25.

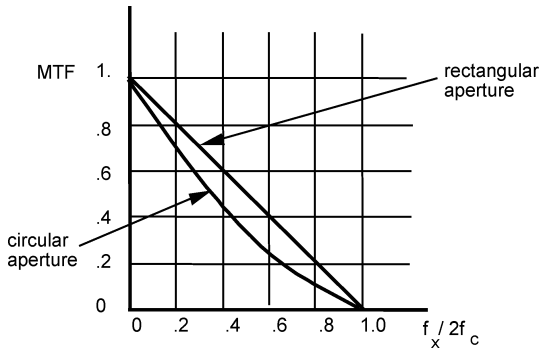


Figure 4.23: MTF of a rectangular and circular aperture.

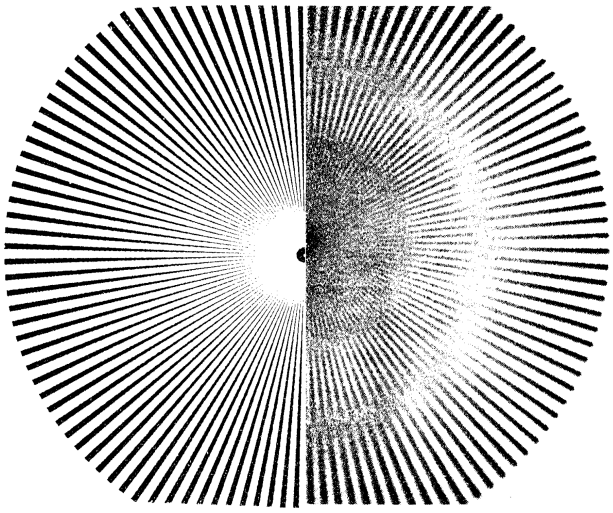


Figure 4.24: Radial test pattern (left) and image through an optical system (right)

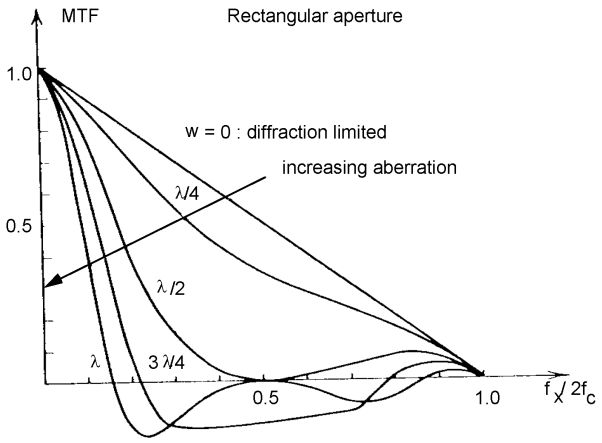
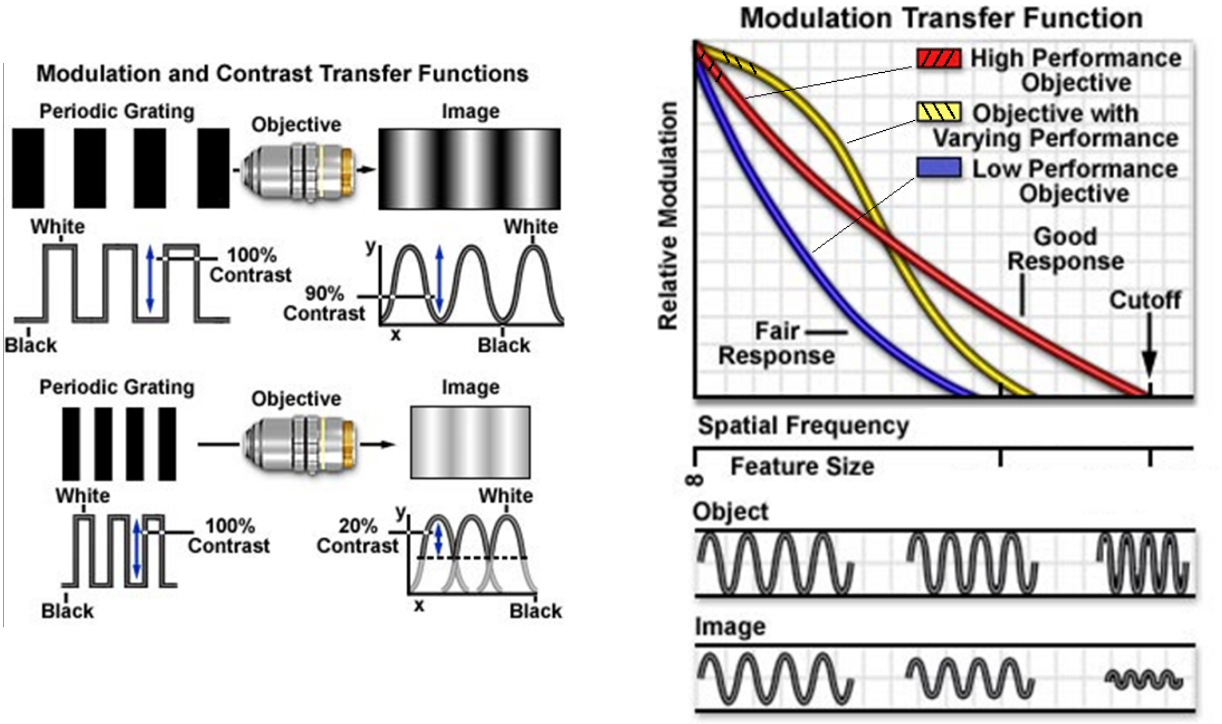


Figure 4.25: MTF of a rectangular aperture with aberrations

Figure 4.24 shows on the left a test pattern which is presented to the optical system; on the right its image. One clearly sees the part with a reversed image.

Figure 4.25 shows the MTF of an optical system with rectangular aperture for different degrees of aberrations; these aberrations were realized by displacing the image plane slightly with respect to the correct position.

Figure 4.26(a) shows how to measure the MTF in practice. As the spatial frequency increases, the modulation depth decreases. When reaching the cut-off frequency, the image becomes gray, and the relative modulation (see 4.26(b)), becomes zero.



(a) (b)

Figure 4.26: Measuring the MTF. Both figures show how the modulation depth decreases as the frequency increases.