Chapter 2
Wave Optics

2.1 Huygen's Principle and Diffraction Integral

The geometrical treatment of the propagation of light presented in the preceding chapter is only an approximate description. It does not take the spatial amplitude distribution of the electromagnetic wave into account. Geometrical optics can be applied as long as the wavelength is small compared to the lateral extent of the wave. This restriction, as already discussed, is equivalent to a large Fresnel number N. The exact description of the propagation of light is obtained by utilizing Maxwell's equations to derive the wave equations for the electric and the magnetic fields. If we neglect the vector properties of the field, the wave equation for the electric field E in homogeneous, isotropic, loss-free, dielectric media reads [1.1,1.3,1.63]:

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + \frac{\partial^2 E}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2} = 0 \tag{2.1}$$

with c being the speed of light in the medium. This equations holds for each of the three components of the field. In general an infinite number of solutions to this wave equation exists since all fields E for which the relations:

$$E(x,y,z,t) = E(\omega t - k_x x - k_y y - k_z z) \tag{2.2}$$

$$|k|^2 = k_x^2 + k_y^2 + k_z^2 = \frac{\omega^2}{c^2} \tag{2.3}$$

hold are solutions of the wave equation (2.1). One well-known electromagnetic field is the plane wave, which reads in the real presentation:

$$E_r(x,y,z,t) = E_0 \cos(\omega t - k_x x - k_y y - k_z z) \tag{2.4}$$

with:

$\omega = 2\pi\nu$	angular frequency
$\nu = c/\lambda$	light frequency
$k = (k_x, k_y, k_z)$	wave vector with $\|k\| = 2\pi/\lambda$

The wave vector $\boldsymbol{k}$ points into the direction of propagation of the wave and is perpendicular to the planes of constant phase. Another common solution is the spherical wave:

$$E(r,t) = E_0 \frac{\lambda}{r} \cos(\omega t - kr) \qquad r>\lambda \tag{2.5}$$

with r being the distance from the source. Although both waves are solutions of the wave equation, they cannot be realized since no spatial confinement is allowed. Without lateral confinement the waves are infinite in extent and contain infinite power. Therefore, infinite waves do not make physical sense. However, they can be considered as very good approximations applicable to the description of many optical phenomena.

Before we discuss more realistic electromagnetic waves, we will now introduce the complex notation for the electric field which is very useful to simplify analytical derivations. Since it is much more convenient to replace the cosine by the complex exponentials, the plane wave (2.4) can be rewritten in the form:

$$E_r(x,y,z,t) = E_0 \frac{1}{2}\left[\exp[i(\omega t - k_x x - k_y y - k_z z)] + \exp[-i(\omega t - k_x x - k_y y - k_z z)]\right] = \frac{1}{2}\left[E + E^*\right.$$

where, as usual, the asterisk indicates a complex conjugate. It is sufficient to consider only one of the two complex electric fields on the right hand side. The real, physical field has been replaced by a complex field. In a similar way, one obtains the complex field for the spherical wave (2.5):

$$E = E_0 \frac{\lambda}{r} \exp[i(\omega t - kr)] \tag{2.6}$$

The real physical field E_r is linked to the complex field E through:

$$E_r = \frac{1}{2}\left[E + E^*\right] \tag{2.7}$$

and the time-averaged intensity is given by:

$$I = \frac{1}{2} c\epsilon_0 EE^*$$

Note that throughout this book we will always present the electric field in its complex notation. For any optical system only one electric field out of the infinite number of solutions of the wave equation describes the light propagation. This unique wave is determined by the boundary conditions of the optical system. This means that at a certain time the electric field in a certain area is given and the propagation in space and time is calculated by using the wave equation. In the following we will describe this time-space

development of electric fields. We will find a mathematical way to determine the electric field at any point in space for an initial field distribution given in a confined area (e.g. a resonator mirror). The propagation of a given confined electric field into space is called *diffraction*. The propagation of electric fields can be evaluated by using Huygen's Principle: *An electric field distribution E(x,y) given on a surface propagates into space in such a way that each point (x,y) can be considered as the source of a spherical wave with amplitude E(x,y). At a point P=(x,y,z) the resulting electric field is the superposition of all of these spherical waves.* As shown in Fig. 2.1 we have to subdivide the starting plane into points $(x_i, y_i, 0)$. Without lack of generality, we assume in the following that the points are equally spaced in both directions. At point P the spherical wave originating from one point on the source plane reads:

$$E_i(P) = C\, E(x_i y_i)\, \frac{\exp[i(\omega t - kr_i)]}{r_i}\, \cos\theta_i\, \Delta x \Delta y \tag{2.8}$$

with r_i being the distance from point $(x_i, y_i, 0)$ to P and $\Delta x, \Delta y$ denoting the point spacing in the x- and y-direction. The term $\cos\theta_i$ accounts for the radiation pattern of a Lambertian source: the energy flow is maximum in the direction of the surface normal ($\cos\theta = 1$) and is equal to zero tangential to the surface. The constant C is as yet an unknown proportionality factor which ensures that the total energy is conserved. In order to attain the resulting field at point P we have to take the sum over all starting points. For a monochromatic source the summation over N points yields

$$E(P) = \sum_{i=1}^{N} E_i(P) = C \exp(i\omega t) \sum_{i=1}^{N} E(x_i y_i)\, \frac{\exp(-ikr_i)}{r_i}\, \cos\theta_i\, \Delta x \Delta y \tag{2.9}$$

This calculation of the field in point P will, of course, become more and more accurate as the number of points is increased. In the limit $N \dashrightarrow \infty$, the summation is replaced by the integration over the starting surface. This discussion of diffraction is purely phenomenological. However, compared to the mathematical derivation of the diffraction integral, this empirical approach provides a better feeling for the physical meaning of the diffraction integral.

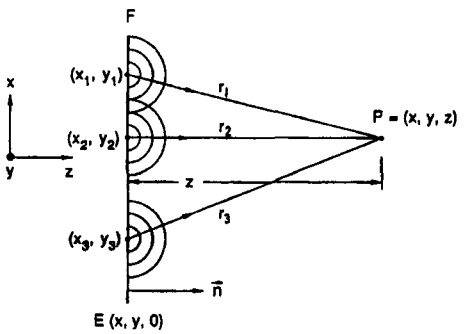


Fig. 2.1 The propagation of an electric field distribution given on a plane at $z=0$ can be dealt with by adding the spherical waves emerging from all points on the plane.

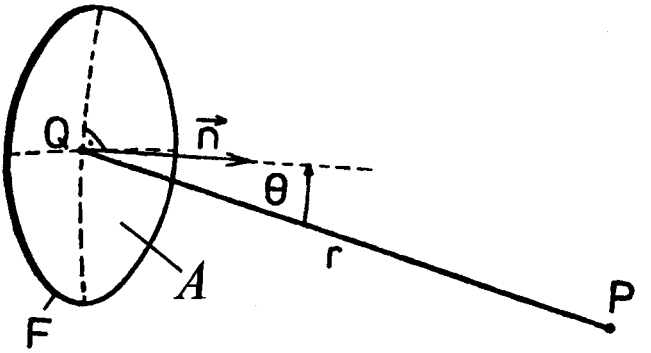


Fig. 2.2 The electric field E at a point P in front of a mirror can be calculated by applying the Kirchhoff integral (2.10) to the field distribution E_1 on the mirror. The paraxial approximation ($\cos\theta \approx 1$) must hold and the diameter of the mirror needs to be much greater than the wavelength.

A mathematical derivation*[1.1,1.4,1.23,1.63] yields for the electric field at point P:

$$E(P) = \frac{i}{\lambda} \exp(i\omega t) \int_A E_1(Q) \frac{\exp(-ikr)}{r} dA \tag{2.10}$$

This surface-integral is referred to as the *Kirchhoff integral.*

Comparing the integral with (2.9), we notice that the cosine term has disappeared. This stems from the paraxial approximation ($\cos\theta \approx 1$) used in the mathematical derivation. A second approximation made is that the extent of the surface A, defined as the square root of the surface area, has to be much greater than the wavelength. As far as optical resonators are concerned both restrictions are easily satisfied. Mirror size and spacing are generally large enough to ensure the validity of both approximations. The paraxial approximation also assumes that the distance r does not change considerably during the integration. It is for this reason that the term $1/r$ is commonly placed outside the integral.

*A strict mathematical derivation of the Kirchhoff-integral is not possible since a fundamental theorem of mathematics is violated. In spite of this mathematical inconsistency, the integral is used in the presented form since the experimental results are in agreement with its theoretical predictions. An exact treatment of diffraction was performed by Arnold Sommerfeld [1.4]. This exact diffraction integral differs from the Kirchhoff integral only for small distances from the surface or for large distances from the optical axis [1.27]. With the above-mentioned restrictions that apply to the Kirchhoff integral, both diffraction integrals provide the same results.

2.2 Diffraction

In the following, we apply the Kirchhoff integral to two common diffraction geometries and discuss the properties of the diffracted field at different distances from the aperture. The time-dependent exponential in (2.10) will be dropped for convenience.

2.2.1 Rectangular Aperture

We consider a rectangular aperture with width $2a$ and height $2b$, as depicted in Fig. 2.3, and we calculate the field distribution $E_2(x_2,y_2)$ for different distances L from the aperture by applying the Kirchhoff integral (2.10) to the field distribution $E_1(x_1,y_1)$ inside the aperture. The distance r between two points in plane 1 and plane 2 reads:

$$r = L\sqrt{1 + \frac{(x_2-x_1)^2 + (y_2-y_1)^2}{L^2}} \tag{2.11}$$

The field $E_2(x_2,y_2)$ is given by the Kirchhoff integral (2.10), which reads:

$$E_2(x_2,y_2) = \frac{i}{\lambda L}\int_{-b}^{b}\int_{-a}^{a} E_1(x_1,y_1)\,\exp\left[-ikL\sqrt{1 + [(x_2-x_1)^2+(y_2-y_1)^2]/L^2}\right] dx_1 dy_1 \tag{2.12}$$

The $1/r$ term in the integrand of (2.10) was replaced by $1/L$ outside of the integral due to the validity of the paraxial approximation. This integral cannot be solved analytically in its present form. However, if we restrict ourselves to distances large compared to the aperture dimensions $(L \gg x_1,y_1)$ and assume that the field distribution does not spread too fast laterally $(x_2/L, y_2/L \ll 1)$, the square root in the exponential of (2.12) can be expanded into a series:

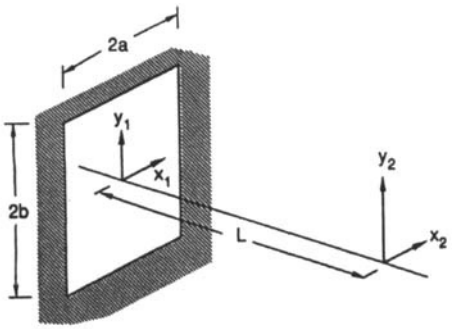


Fig. 2.3 Geometry of a rectangular aperture. The electric field at a plane separated by a distance L from the aperture plane is being calculated.

$$r = L\left[1 + \frac{1}{2}\left(\frac{x_2-x_1}{L}\right)^2 + \frac{1}{2}\left(\frac{y_2-y_1}{L}\right)^2 - \frac{1}{8}\left(\frac{x_2-x_1}{L}\right)^4 - \frac{1}{8}\left(\frac{y_2-y_1}{L}\right)^4 + \ldots\right] \tag{2.13}$$

The following three areas can be distinguished:

a) Fraunhofer-Diffraction

For distances L very large compared to the aperture size, the quadratic terms x_1^2/L, y_1^2/L become negligibly small and the expression (2.13) for the distance r can be reduced to:

$$r = L - \frac{x_1x_2+y_1y_2}{L} + \frac{x_2^2+y_2^2}{2L} \tag{2.14}$$

The quadratic term of (2.14) represents the phase curvature of a spherical wave with radius of curvature L. This term is usually disregarded since it vanishes when the intensity of the field is calculated. By using the angular coordinates $\theta_x=x_2/L$ and $\theta_y=y_2/L$, the normalized coordinates $x_1^*=x_1/a$ and $x_2^*=x_2/b$, and Eq. (2.14), the Kirchhoff integral (2.12) reads:

$$E_2(\theta_x,\theta_y) = i\frac{ab}{\lambda L}\exp(-i\phi)\int_{-1}^{1}\int_{-1}^{1} E_1(x_1^*,y_1^*)\exp[ik(a\theta_x x_1^* + b\theta_y y_1^*)]\,dx_1^*dy_1^* \tag{2.15}$$

with $$\phi = kL + \frac{x_2^2+y_2^2}{2L}$$

This integral equation can be separated into two equations, one for the x-direction and one for the y-direction, if the field E_1 can be written as a product:

$$E_1(x_1^*,y_1^*) = u_1(x_1^*)\,v_1(y_1^*)$$

The resulting field $E_2(\theta_x,\theta_y) = u_2(\theta_x)\,v_2(\theta_y)$ is given by the two integral equations:

$$u_2(\theta_x) = \exp\left[-i\left(\frac{\phi}{2}-\frac{\pi}{4}\right)\right]\sqrt{N_x}\int_{-1}^{1} u_1(x_1^*)\exp[ikax_1^*\theta_x]\,dx_1^* \tag{2.16}$$

$$v_2(\theta_y) = \exp\left[-i\left(\frac{\phi}{2}-\frac{\pi}{4}\right)\right]\sqrt{N_y}\int_{-1}^{1} v_1(y_1^*)\exp[ikby_1^*\theta_y]\,dy_1^* \tag{2.17}$$

with $N_x=a^2/\lambda L$ and $N_y=b^2/\lambda L$ being the Fresnel numbers in x-direction and y-direction, respectively. The field distributions (2.16) and (2.17) are referred to as the far field of the field distribution $E_1(x,y)$. The angular coordinates θ_x, θ_y denote the angles under which the field diverges into space, also called the divergence angles. If the far field is observed at the back focal plane of a lens, the distance L has to be replaced by the focal length f. Mathematically, these equations correspond to the Fourier transformation.

Example: Homogeneously Illuminated Aperture

For a homogeneous field distribution $E_1(x,y)=const.=E$, the integral equation (2.16) and (2.17) can be solved analytically. This yields for the field intensity in x-direction (excluding a proportionality factor):

$$I_x(\theta_x) = N_x \frac{\sin^2(ka\theta_x)}{(ka\theta_x)^2} \tag{2.18}$$

and the corresponding expression for the y-direction. The intensity distribution at a sufficiently large distance from the aperture is thus given by:

$$I(\theta_x,\theta_y) = I_0 N_x N_y \frac{\sin^2(ka\theta_x)}{(ka\theta_x)^2} \frac{\sin^2(kb\theta_y)}{(kb\theta_y)^2} \tag{2.19}$$

with I_0 being the intensity across the aperture. Figure 2.4 shows the one-dimensional intensity distribution (2.18) in the x-direction. The intensity minima are determined by:

$$\theta_x = n\frac{\lambda}{2a} \qquad \theta_y = m\frac{\lambda}{2b} \tag{2.20}$$

with n,m being integer numbers greater than zero.

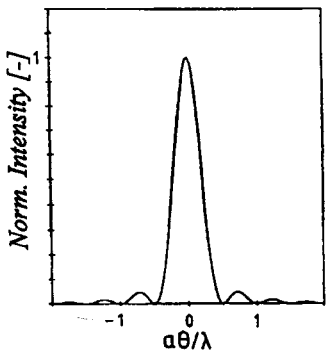


Fig. 2.4 One-dimensional intensity distribution in the far field of a homogeneously illuminated rectangular aperture.

The full width half maximum (FWHM) diameters of the central lobe of the intensity distribution (2.19) are:

$$\Delta\theta_x = 0.44\,\frac{\lambda}{a} \qquad \Delta\theta_y = 0.44\,\frac{\lambda}{b} \tag{2.21}$$

In conclusion, the field distribution at a sufficiently large distance L from the aperture is given by the Kirchhoff integral with the Fraunhofer approximation (2.15) (often referred to as the Fraunhofer integral). This field distribution, which remains constant as the distance L is increased, is called the far field. The far field is usually expressed in terms of the divergence angles θ_x and θ_y. The lateral position of the field minima and maxima can be determined by multiplying the divergence angles by the distance L. The Fraunhofer approximation can be applied if the Fresnel numbers $N_{x,y}$ are much smaller than 1; a "safe" limit is $N_{x,y}<0.2$.

Example: For λ=500nm and an aperture width of $2a$=2mm, the Fraunhofer-approximation can be used if the distance L is greater than 10m. At a distance of L=40m the position of the first intensity minimum with respect to the optical axis is x=10mm. Propagation over a distance of L=4km results in a minimum position of x=1m. At both distances, the shapes of the intensity distributions are the same.

b) Fresnel-Diffraction

If at least one of the Fresnel numbers $N_{x,y}$ is greater than one (which means that we get closer to the aperture), the linear approximation (2.14) of the distance r is not applicable. We have to add the quadratic terms of x_1 and y_1 in order to be able to calculate the field distribution. Thus, we insert the second order series of the distance:

$$r = L - \frac{x_1x_2+y_1y_2}{L} + \frac{x_1^2+y_1^2}{2L} + \frac{x_2^2+y_2^2}{2L} \tag{2.22}$$

into the integral (2.12) and we obtain:

$$E_2(x_2,y_2) = i\,\exp[-ikL]\,\sqrt{N_xN_y} \quad \bullet \tag{2.23}$$

$$\int_{-1}^{1}\int_{-1}^{1} E_1(x_1^*,y_1^*)\,\exp[-i\pi N_x(x_1^{*2}+x_2^{*2}-2x_1^*x_2^*)\;-i\pi N_y(y_1^{*2}+y_2^{*2}-2y_1^*y_2^*)]\;dx_1^*dy_1^*$$

with the normalized coordinates $x^*=x/a$ and $y^*=y/b$. The Fresnel approximation can be applied if the Fresnel number N is smaller than 100. Similar to the Fraunhofer integral

(2.15), the Fresnel integral can be separated into two one-dimensional integrals if the electric field can be expressed as a product of one-dimensional fields. Unfortunately, there is no analytical solution to the Fresnel integral and the electric field distribution has to be computed numerically. In the case of a homogeneously illuminated rectangular aperture the values of the Fresnel integral can be found in mathematical handbooks such as [1.28]. Figures 2.5 and 2.6 show numerically calculated and experimentally recorded intensity distributions for different Fresnel numbers for a homogeneously illuminated square aperture. In the experiment a metal aperture of side length 2mm was inserted into a collimated HeNe-laser beam (λ=632.8nm) with a beam diameter expanded to 15mm by means of a telescope. Note the agreement between the calculated and the recorded intensity distributions and how the intensity distributions approach the intensity distribution in the far field as the Fresnel number decreases.

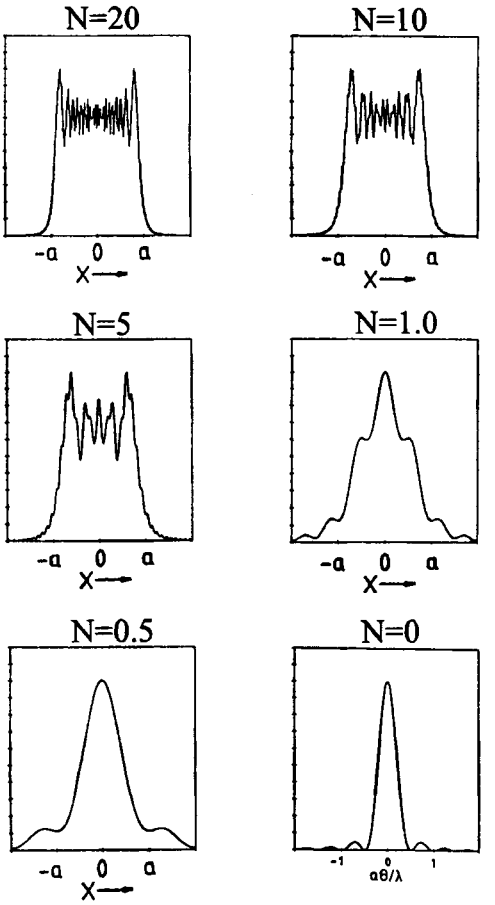


Fig. 2.5 Calculated normalized one-dimensional intensity distributions for a homogeneously illuminated aperture with width $2a$ and different Fresnel numbers N. The lower right graph shows the far field intensity distribution.

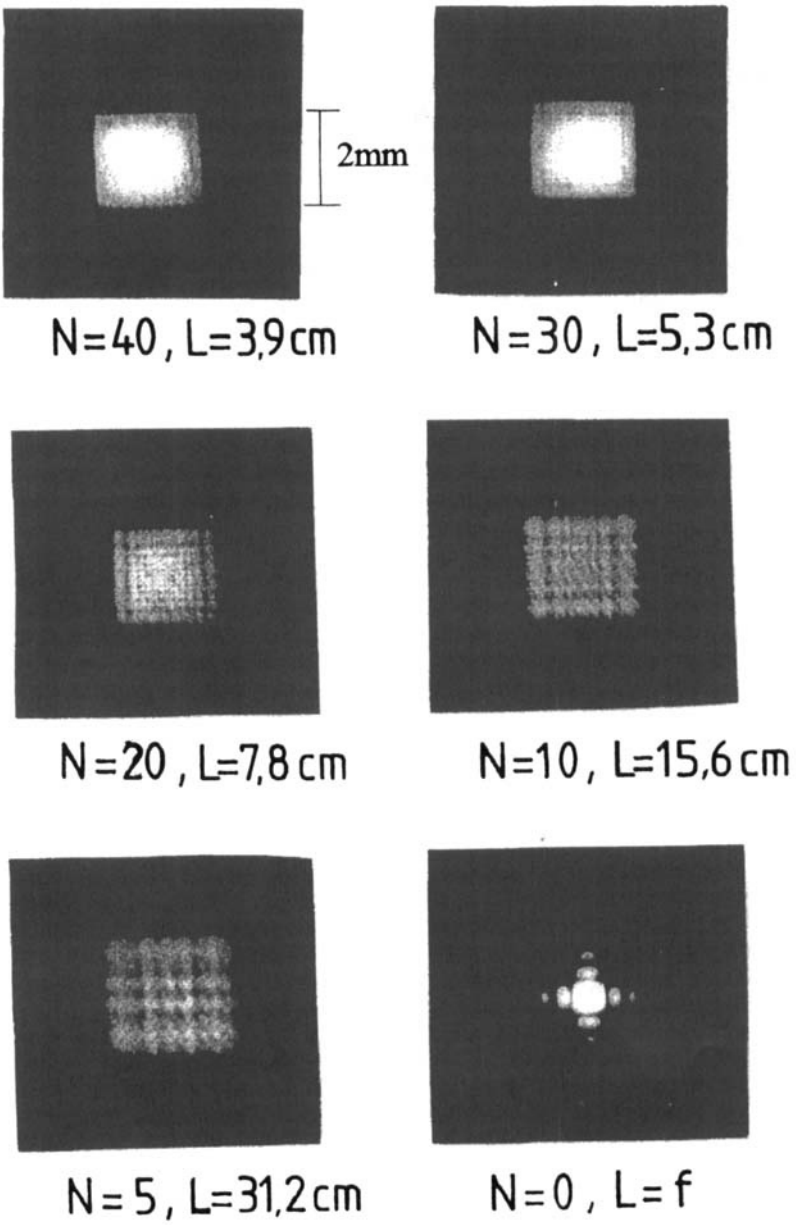


Fig. 2.6 Intensity distributions photographed at different distances L behind a square aperture with side length 2mm. The aperture was homogeneously illuminated with a HeNe-laser beam (λ=632.8nm) and the intensity distributions were recorded by means of a CCD-camera. The lower right photograph shows the far-field, recorded in the focal plane of a focusing lens. Here the CCD camera was overexposed to show the side lobes more clearly.

c) Geometrical Optics

If we approach the aperture even closer, so that the Fresnel numbers N_x, N_y become greater than 100, the quadratic series (2.22) for the distance r is not accurate enough. Unfortunately, simply adding the next term (fourth power) to the expansion, does not solve the problem. The Kirchhoff integral (2.10) is not applicable to such high Fresnel numbers since we are too close to the aperture and the paraxial approximation *($\cos\theta=1$)* does not hold anymore. A detailed investigation, however, indicates that the field distributions do not change considerably for large Fresnel numbers. Additionally, the shape of the intensity distribution is almost identical to that of the initial distribution across the aperture (see *$N=40$* in Fig.2.6). Field propagation for Fresnel numbers greater than 100 can be calculated by using the laws of geometrical optics as long as one is not interested in fine structures near the aperture edges, or by using the place wave approximation (see Sec. 2.9).

2.2.2 Circular Aperture

If the electric field is diffracted by a circular aperture with radius R (Fig. 2.7), it is convenient to express the Kirchhoff integral in terms of polar coordinates r, Φ with

$$x_i = r_i \cos\Phi_i , \qquad y_i = r_i \sin\Phi_i , \qquad 0 \le r_i \le R$$

In the Fraunhofer approximation one obtains with $\theta = r_2/L$: (2.24)

$$E_2(\theta,\Phi_2) = i\, N \exp[-ikL] \int_0^{2\pi}\int_0^{1} E_1(r_1^*,\Phi_1) \exp[-ikR\theta r_1^* \cos(\Phi_2-\Phi_1)]\; r_1^*\, dr_1^*\, d\Phi_1$$

and the Fresnel integral reads: (2.25)

$$E_2(r_2^*,\Phi_2) = iN \exp[-ikL] \int_0^{2\pi}\int_0^{1} E_1(r_1^*,\Phi_1) \exp[-i\pi N(r_1^{*2}+r_2^{*2}-2r_1^* r_2^* \cos(\Phi_2-\Phi_1))]\; r_1^*\, dr_1^*\, d\Phi_1$$

with $N = R^2/(\lambda L)$: Fresnel number

$r_i^* = r/R$: normalized radial coordinate

For field distributions E_i that can be expressed as products of a purely radial function $u(r)$ and an oscillating angular term:

$$E_i(r_i^*,\Phi_i) = u_i(r_i^*) \exp[-im\Phi_i] , \qquad m = 0,1,2,\ldots \; ; \; i = 1,2$$

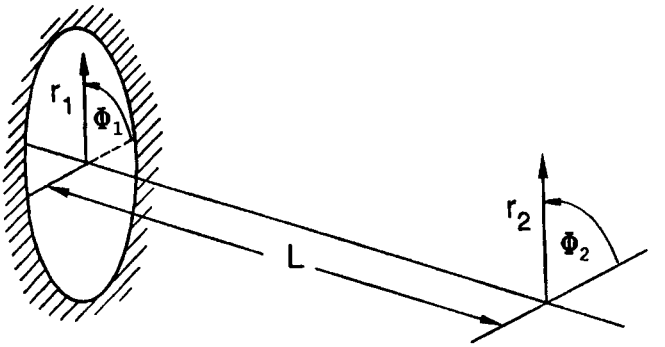


Fig. 2.7 Diffraction at a circular aperture. The field distributions at both planes are expressed in terms of polar coordinates r, Φ.

both integrals can be simplified by performing the angular integration. The final result with the Fraunhofer approximation is given by:

$$u_2(\theta) = -(-i)^{m+1}\, 2\pi N \exp[-ikL] \int_0^1 u_1(r_1^*)\, J_m\left(\frac{2\pi}{\lambda} R\theta r_1^*\right)\, r_1^* dr_1^* \tag{2.26}$$

and the Fresnel integral is given by:

$$u_2(r_2^*) = -(-i)^{m+1}\, 2\pi N \exp[-ikL] \int_0^1 u_1(r_1^*) \exp[-i\pi N(r_1^{*2}+r_2^{*2})]\, J_m(2\pi N r_1^* r_2^*)\, r_1^* dr_1 \tag{2.27}$$

with $J_m(x)$: Bessel function of order m [1.28].

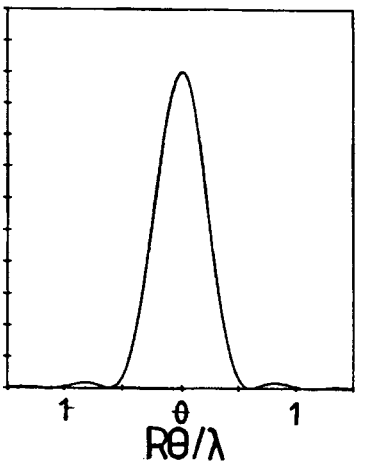
Example: Far field of a homogeneously illuminated circular aperture with radius R. By using equation (2.26) with $m=0$ and $u(r)=E_0=const.$ one obtains:

$$u_2(\theta) = 2\pi\, N\, E_0\, \frac{J_1(2\pi R\theta/\lambda)}{2\pi R\theta/\lambda} \tag{2.28}$$

The corresponding intensity distribution is shown in Fig. 2.8 together with the photograph of a far field pattern of a HeNe-laser diffracted by a round aperture with radius $R=1mm$. The far field intensity distribution of a circular aperture is called the Airy pattern. It looks very similar to the far field of a rectangular aperture in one dimension. The position of the intensity minima, height of the intensity maxima, and the energy content in the central lobe are, however, quite different as Table 2.1 indicates.

Table 2.1 Comparison of far field properties of rectangular and circular apertures. Positions and diameters are given in terms of divergence. *The height of the central lobe is normalized to 1.0.

	rectangular aperture width 2a	circular aperture diameter 2R
FWHM diameter	$0.44\lambda/a$	$0.52\lambda/R$
Position 1. minimum	$0.50\lambda/a$	$0.61\lambda/R$
Position 2. minimum	$1.00\lambda/a$	$1.12\lambda/R$
Height of 1. maximum*	0.04718	0.01753
Height of 2. maximum*	0.01694	0.00522
Diameter containing 86% of total power	$1.05\ \lambda/a$	$1.61\ \lambda/R$
Power content of central peak	81.5%	84%
Peak intensity of central peak I_{max}/I_0	$(4ab/\lambda f)^2$	$(\pi\rho^2/\lambda f)^2$



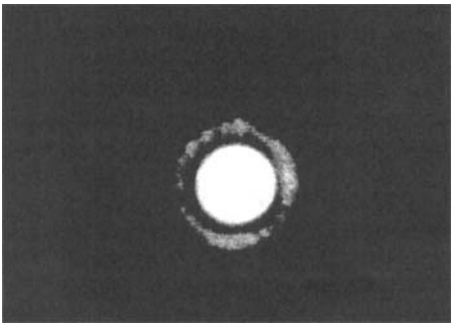


Fig 2.8 Calculated and measured far field intensity distributions for a circular aperture with radius *a*. For the photograph a pinhole with radius 1mm was inserted into an expanded and collimated HeNe- laser beam (λ=632.8nm). The far field was generated in the focal plane of a positive lens.

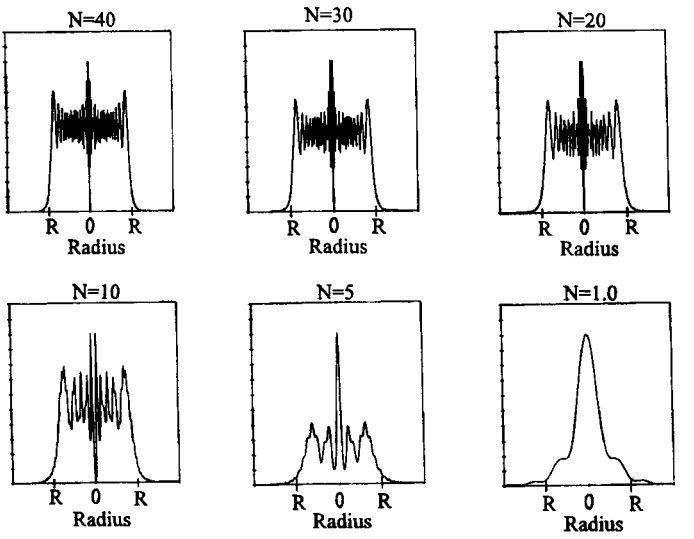


Fig. 2.9 Calculated radial intensity distributions for different Fresnel numbers N of a homogeneously illuminated circular aperture of radius R. With decreasing Fresnel number the distributions become less spiky and for Fresnel numbers of less than 1, the far field intensity distribution is approached.

For Fresnel numbers greater than about 0.2, the Fresnel integral should be used to calculate the field distribution. Similar to the rectangular symmetry, this can only be accomplished by using numerical methods. Figure 2.9 shows the calculated radial intensity distributions for a homogeneously illuminated circular aperture with radius R for different Fresnel numbers N. The diffraction patterns are similar to those of a rectangular aperture (Fig. 2.5), but now the rotational symmetry results in a high intensity peak on the optical axis. In high power lasers having circular output beams this so called "Spot of Arago" can lead to the destruction of optical components. For both rectangular and circular apertures, the far field depends only on the product of the aperture diameter and the far field angle θ (assuming a constant wavelength λ). Decreasing the aperture width thus results in the increase of the beam divergence. Again we observe that the beam parameter product is a constant of the optical system. This is quite a satisfying result since it is in agreement with our discussion of the phase space transformations (see Sec. 1.2.2). Let us define the far field half angle of divergence θ via the 86% power content (see Table 2.1) and the beam radius w in a similar way ($w=0.86a$ for the rectangular aperture and $w=0.93R$ for the circular aperture). We then get for the beam parameter products in the two geometries:

$$\textit{rectangular}: \quad w\theta = 1.418\,\frac{\lambda}{\pi} \qquad\qquad \textit{circular}: \quad w\theta = 2.337\,\frac{\lambda}{\pi}$$

This is already close to the minimum beam parameter product of λ/π.

2.3 Collins Integral

2.3.1 One-Dimensional Optical Systems

We have so far only discussed the free propagation of the electric field without any optical elements like lenses and mirrors being in the path. We have found that the electric field after propagation is linked to the initial field via the Kirchhoff integral. It is possible to incorporate any optical system with parabolic surfaces into the propagation and derive a corresponding diffraction integral. This generalized Kirchhoff integral is referred to as the Collins integral. The derivation of the Collins integral is beyond the scope of this book, the interested reader is referred to the original paper [1.34]. The basic principle of the derivation is to find the shortest distance between points on two separated planes where arbitrary optical elements may be located between those planes. By applying Fermat's principle one can derive an expression for the distance r which depends on the elements of the ray transfer matrix $\boldsymbol{M}$. We will first restrict the discussion to one dimensional optical systems, which means optics that exhibit rotational symmetry and can thus be described by a 2x2 ray transfer matrix. As already shown in the last chapter, the generalization to two dimensions can be performed by replacing the matrix elements with the corresponding submatrices.

If the electric field propagates from a plane 1 to a plane 2 and an optical system with a ray transfer matrix $\boldsymbol{M}$ is located between the planes, the Kirchhoff integral in the Fresnel approximation reads (Fig.2.10):

$$E_2(x_2,y_2) =$$

$$= \frac{i}{\lambda B}\exp[-ikL]\int\int E_1(x_1,y_1)\exp[-i\frac{\pi}{\lambda B}(Ax_1^2+Dx_2^2-2x_1x_2 + Ay_1^2+Dy_2^2-2y_1y_2)]\,dx_1dy_1 \tag{2.29}$$

with A,B,D being the elements of the ray transfer matrix $\boldsymbol{M}$ and L is the optical path along the optical axis.

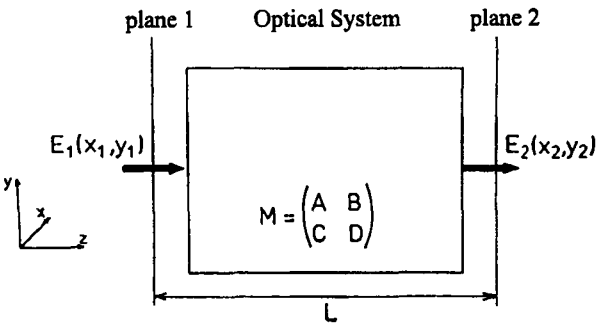


Fig. 2.10 Propagation of an electric field through an optical system. The fields at plane 1 and plane 2 are linked to each other via the Collins integral (2.29).

This integral is referred to as the Collins integral. It does not differ from the Kirchhoff integral with the Fresnel approximation, except for the fact that the shortest distance between two points is already incorporated with the ray transfer matrix elements being system parameters. Similar to the Kirchhoff integral the derivation is based on the applicability of the paraxial approximation. This means that the Collins integral is valid for Fresnel numbers lower than 100.

Examples:

1) Free Space Propagation:
We consider the simplest case of an empty optical system with length L. By inserting the ray matrix elements $A=1$, $B=L$ and $D=1$ into (2.29) we obtain, as to be expected, the Fresnel integral (2.23).

2) Field Distribution at the Focal Plane of a Lens:
A field distribution E_1 is given at a plane in front of focusing lens with focal length f. The distance from the plane to the focusing lens is d. We want to calculate the resulting field distribution at the focal plane of the lens. The ray transfer matrix elements for the propagation to the focal plane of the lens are found to be $A=0$, $B=f$ and $D=1-d/f$. Insertion into (2.29) yields:

$$E_2(\theta_x,\theta_y) = i\,\frac{\exp(-ik\delta)}{\lambda f}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E_1(x_1,y_1)\,\exp[ik(\theta_x x_1 + \theta_y y_1)]\,dx_1 dy_1 \tag{2.30}$$

with $$\delta = d + f + \frac{1}{2}(\theta_x^2 + \theta_y^2)(f-d)$$

and $$\theta_x = \frac{x_2}{f}\ ,\ \theta_y = \frac{y_2}{f}\ ,\ k = \frac{2\pi}{\lambda}$$

A comparison with (2.15) indicates that this integral is equivalent to the Kirchhoff integral in the Fraunhofer approximation. This means that at the focal plane of a focusing lens the far field intensity distribution of the electric field E_1 can be measured. This is the common experimental technique used to determine angles of divergence of laser beams. In the special case of $d=f$, the parabolic phase term in front of the integral disappears and the exact Fourier transform of the field is observed in the back focal plane. The integral in (2.30) is referred to as a Fourier transform. Fourier transforms play an important role in diffraction theory not only because of their linkage with the far field, but also because they can provide a better understanding of the properties of electric fields and their propagation. We shall discuss the Fourier transforms and the mathematics involved in more detail in Sec. 2.4.2.

2.3.2 Two-Dimensional Optical Systems

The Collins integral formalism can be generalized for arbitrary optical systems provided they have parabolic interfaces or parabolic index profiles in paraxial approximation. As was discussed in the preceding chapter, these optics are described by 4x4 ray transfer matrices $\boldsymbol{M}$. Propagation of rays through these optical systems can be written in the form:

$$\begin{pmatrix} \boldsymbol{r}_2 \\ \boldsymbol{\gamma}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix} \begin{pmatrix} \boldsymbol{r}_1 \\ \boldsymbol{\gamma}_1 \end{pmatrix} \tag{2.31}$$

with $\boldsymbol{r}_i = (x_i, y_i)$, $\boldsymbol{\gamma}_i = (\alpha_i, \beta_i)$, $i=1,2$

and $\boldsymbol{A},\boldsymbol{B},\boldsymbol{C},\boldsymbol{D}$ being the 2x2 matrices presented in Sec. 1.2.4.

By using the same generalization we have already applied to the matrix determinant (1.61), we can find the generalized Collins integral for two-dimensional optical systems:

$$E_2(x_2,y_2) = \frac{i\ \exp(-ikL)}{\lambda\sqrt{\det \boldsymbol{B}}} \int\int E_1(x_1,y_1)\ \exp[-i\frac{\pi}{\lambda}(\boldsymbol{r}_1\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{r}_1 + \boldsymbol{r}_2\boldsymbol{D}\boldsymbol{B}^{-1}\boldsymbol{r}_2 - 2\boldsymbol{r}_1\boldsymbol{B}^{-1}\boldsymbol{r}_2)]\ dx_1 dy_1 \tag{2.32}$$

with $\boldsymbol{B}^{-1}$ being the inverse of matrix $\boldsymbol{B}$ and $det\boldsymbol{B}$ denoting its determinant.

The beauty of this equation lies in the fact that we can easily memorize the diffraction integral for two-dimensional systems due to its similarity to the one-dimensional equation (2.29). The generalized Collins integral (2.32) enables us to calculate even complex diffraction problems like the propagation through rotated and tilted optical elements as the following examples will show:

Examples:

1) Phase Space Beam Analyzer

We use the Phase Space Beam Analyzer as discussed in Sec. 1.2.4 and calculate the light propagation from the entrance plane 1 to the recording plane 4 (see Fig. 1.23). We neglect the presence of the second slit and determine the phase term $\exp[-i\pi\Phi/\lambda]$ in the integrand of (2.32). By using the resulting ray transfer matrix (1.78) we get (note that the coordinates in the recording plane are denoted as x_4,y_4):

$$\Phi = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \begin{pmatrix} P & 0 \\ 0 & P \end{pmatrix} \begin{pmatrix} 0 & 1/S \\ 1/S & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} x_4 \\ y_4 \end{pmatrix} \begin{pmatrix} 1/P & f_0/f \\ f_0/f & 1/P \end{pmatrix} \begin{pmatrix} 0 & 1/S \\ 1/S & 0 \end{pmatrix} \begin{pmatrix} x_4 \\ y_4 \end{pmatrix}$$

$$- 2 \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \begin{pmatrix} 0 & 1/S \\ 1/S & 0 \end{pmatrix} \begin{pmatrix} x_4 \\ y_4 \end{pmatrix} \tag{2.33}$$

which can be rewritten as:

$$\Phi = \frac{2P}{S}x_1y_1 + \frac{2}{PS}x_4y_4 - \frac{2}{S}(x_1y_4+x_4y_1) + \frac{f_0}{fS}(x_4^2+y_4^2) \tag{2.34}$$

The diffraction integral (2.32) then reads with $detB=-S^2$: (2.35)

$$E_4(x_4,y_4) = \frac{\exp[-ik(a+b)]}{S} \cdot$$
$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} E_1(x_1,y_1)\exp\left[-i\frac{\pi}{\lambda}\left[\frac{2P}{S}x_1y_1 + \frac{2}{PS}x_4y_4 - \frac{2}{S}(x_1y_4+x_4y_1) + \frac{f_0}{fS}(x_4^2+y_4^2)\right]\right] dx_4dy_4$$

This integral can only be solved numerically, except for Gaussian input beams.

2) Far Field Measurement with Tilted Lens

We want to record the far field intensity distribution in the focal plane of a focusing lens with focal length f. The electric field incident on the lens is E_1. Unfortunately, the surface normal of the lens is tilted by an angle α around the x-axis with the z-axis being the optical axis (see Fig. 1.20). By using (1.84) and (1.85), the ray transfer matrix for the passage through the lens and propagation to the focal plane can be approximated by:

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix} = \begin{pmatrix} 1-\cos\alpha & 0 & f & 0 \\ 0 & 1-\dfrac{1}{\cos\alpha} & 0 & f \\ \dfrac{-\cos\alpha}{f} & 0 & 1 & 0 \\ 0 & \dfrac{-1}{f\cos\alpha} & 0 & 1 \end{pmatrix} \tag{2.36}$$

With $\sqrt{\det\boldsymbol{B}} = f$ and $\boldsymbol{B}^{-1} = \begin{pmatrix} 1/f & 0 \\ 0 & 1/f \end{pmatrix}$, we get from (2.32):

$$E_2(x_2,y_2) = \frac{i}{\lambda f}\exp[-ik(f+\frac{1}{2}(x_2^2+y_2^2))] \cdot$$
$$\int\int E_1(x_1,y_1)\exp[-i\frac{\pi}{\lambda f}((1-\cos\alpha)x_1^2+(1-1/\cos\alpha)y_1^2-2(x_1x_2+y_1y_2))]\, dx_1dy_1 \tag{2.37}$$

By introducing the variables $\theta_x=x_2/f$ and $\theta_y=y_2/f$ we can rewrite this as:

$$E_2(\theta_x,\theta_y) = i\,\frac{\exp(ik\delta)}{\lambda f}\iint E_1^+(x_1,y_1)\,\exp[-ik(\theta_x x_1 + \theta_y y_1)]\,dx_1dy_1 \tag{2.38}$$

with $$E_1^+(x_1,y_1) = E_1(x_1,y_1)\,\exp[-i\frac{\pi}{\lambda f}((1-\cos\alpha)x_1^2+(1-1/\cos\alpha)y_1^2)]$$

and $$\delta = f + \frac{f}{2}(\theta_x^2+\theta_y^2)$$

A comparison with (2.30) indicates that we measure the far field intensity distribution of the field E_1^+ which is the product of the original field and the parabolic phase terms induced by the lens tilt. Fortunately, there are transformation rules available that simplify the analytical calculation of Fourier integrals like this. This will be shown in the following section.

2.4 Collins Integral and Vanishing Ray Matrix Elements

2.4.1 Imaging Condition (B=0)

We know from our discussion of ray transfer matrices in the preceding chapter that a vanishing B-component represents an imaging optical system. The matrix component A represents the magnification by which the object is scaled in the image plane. In terms of wave optics, imaging means that the initial field distribution $E_1(x_1,y_1)$ at the input plane 1 is reproduced at plane 2. Following relation then holds for the field $E_2(x_2,y_2)$ at plane 2:

$$E_2(x_2,y_2) = C\,\frac{1}{A}\,E_1(\frac{x_2}{A},\frac{y_2}{A}) \tag{2.39}$$

The constant C can be a complex function of x_2 ,y_2 with $CC^*=1$, and the factor $1/A$ is necessary to conserve the total power. It is possible to prove that the Collins integral (2.29) yields the imaging condition (2.39) in the limit $B\text{--->}0$ [1.72]. In order to show this we rewrite Eq. (2.29) as:

$$E_2(x_2,y_2) = \frac{i}{\lambda B}\,\exp[-ikL]\,\exp[-i\frac{\pi}{\lambda B}(D-\frac{1}{A})(x_2^2+y_2^2)]\;\bullet \tag{2.40}$$

$$\iint E_1(x_1,y_1)\,\exp[-i\frac{\pi}{\lambda B}[A(x_1-\frac{x_2}{A})^2 + A(y_1-\frac{y_2}{A})^2]\,dx_1dy_1$$

By using one of the definitions of the Dirac delta function:

$$\delta(z) = \lim_{B\to 0} \sqrt{\frac{i}{2\pi B}} \exp\left[-i\frac{z^2}{2B}\right] \tag{2.41}$$

$$\int_{-\infty}^{+\infty} f(z)\delta(z-z_o)dz = f(z_0) \tag{2.42}$$

the Collins integral (2.40) in the limit $B \to 0$ can be written as:

$$E_2(x_2,y_2) = \frac{-2\pi}{\lambda} \exp[-ikL] \exp[-i\frac{\pi}{\lambda B}(D-\frac{1}{A})(x_2^2+y_2^2)] \cdot$$

$$\cdot \int\int E_1(x_1,y_1)\ \delta(i\sqrt{kA}(x_1-\frac{x_2}{A}))\delta(i\sqrt{kA}(y_1-\frac{y_2}{A}))\ dx_1dy_1$$

Performing the integration and applying the relationship $\delta(az)=\delta(z)/|a|$ yields the final result:

$$E_2(x_2,y_2) = \exp[-ikL] \exp[-i\frac{\pi}{\lambda B}(D-\frac{1}{A})(x_2^2+y_2^2)]\ \frac{1}{A}\ E_1(\frac{x_2}{A},\frac{y_2}{A})$$

Equation (2.39) holds for $B\to 0$ (imaging and the original plane at $z=0$) if the paraxial approximation is fulfilled.

2.4.2 Fourier Transformation (A=0)

We have already seen that in the focal plane of a focusing lens, the Collins integral becomes proportional to the Fourier transform of the incident field E_1. This means that the intensity distribution in the focal plane equals the intensity distribution in the far field. For an arbitrary optical system, the general condition for propagation to a focal plane is given by $A=0$. In this case the Collins integral (2.29) reads:

$$E_2(\theta_x,\theta_y) = \frac{i}{\lambda B} \exp[-ik(L+\frac{DB}{2}(\theta_x^2+\theta_y^2))] \int\int E_1(x_1,y_1) \exp[ik(x_1\theta_x+y_1\theta_y)]\ dx_1dy_1$$

$$= \frac{i}{\lambda B} \exp[-ik(L+\frac{DB}{2}(\theta_x^2+\theta_y^2))]\ \mathcal{F}\{E_1(x_1,y_1)\} \tag{2.43}$$

which is equivalent to the Kirchhoff integral in Fraunhofer approximation and thus represents the far field distribution. If the optical system is chosen such that in addition to A, both A and D are equal to 0, the far field amplitude is proportional to the Fourier transform. The useful feature of the Fourier transform is the fact that a light source is completely characterized either by its electric field E or by the Fourier transform $\boldsymbol{F(E)}$ of the field. It is for this reason that Fourier transforms and the knowledge of their transformation properties are more important in optics than their basic relationship to the far field. Diffraction calculations can be considerably simplified if the light is represented by its Fourier transform rather than by the electric field itself. This will become more clear when we apply the mathematics behind Fourier transforms to diffraction problems [1.30].

Properties of Fourier Transforms

The Fourier Transform $\boldsymbol{F}$ of a function $f(x)$ is defined by:

$$A(\theta) = \boldsymbol{F}(f(x)) = \frac{k}{2\pi}\int_{-\infty}^{+\infty} f(x)\ \exp[ik\theta x]\ dx\ ,\quad k\in\mathbb{R} \tag{2.44}$$

$A(\theta)$ is called the Fourier spectrum or Fourier transform of $f(x)$. For the Fourier transform to exist, the function $f(x)$ has to meet certain requirements:

a) $f(x)$ is continuous and its derivative df/dx exists at all but a finite number of points,

b) $f(x)$ is square-integrable, which means that it has to decrease faster than $1/x^2$.

The inverse Fourier transformation, $\boldsymbol{F}^{-1}$, restores the original function $f(x)$:

$$f(x) = \boldsymbol{F}^{-1}(A(\theta)) = \int_{-\infty}^{+\infty} A(\theta)\ \exp[-ik\theta x]\ d\theta \tag{2.45}$$

If $A(\theta)$, $B(\theta)$ are the Fourier transforms of the complex functions $f(x)$ and $g(x)$, respectively, and a,b are complex numbers, the following rules apply [1.1]:

Function	Fourier Spectrum	Name of Operation
$af(x)+bg(x)$	$aA(\theta)+bB(\theta)$	Linearity
$f(ax)$	$A(\theta/a)\ 1/a$	Reciprocity
$f(x-x_0)$	$A(\theta)\ exp[i\theta x_0]$	Shifting
$f(x)\ exp[-i\theta_0 x]$	$A(\theta-\theta_0)$	Shifting
$f(x)g(x)$	$A(\theta)\odot B(\theta)$	Convolution
$f(x)\odot g(x)$	$A(\theta)B(\theta)$	Convolution
$f(x)\otimes g(x)$	$A(\theta)B^*(\theta)$	Correlation

Convolution and correlation are defined as:

$$\textit{Convolution:} \quad A(\theta)\odot B(\theta) = \int_{-\infty}^{+\infty} A(\phi)\, B(\theta-\phi)\, d\phi \tag{2.46}$$

$$\textit{Correlation:} \quad f(x)\otimes g(x) = \int_{-\infty}^{+\infty} f(x')\, g^{*}(x'-x)\, dx' \tag{2.47}$$

Table 2.2 Common Fourier transforms (a,b,and k are real and c is a complex number) [1.1]:

f(x)	range	A(θ)=*F*(f(x))
c	for all x	c δ(θ)
c δ(x)		c
c	for a<x<b	
0	else	-ic[exp(ikbθ)-exp(ikaθ)]/(2πθ)
exp[ikθ$_0$x]	for all x	δ(θ-θ$_0$)
cos(kθ$_0$x)	for all x	[δ(θ-θ$_0$)+δ(θ+θ$_0$)]/2
sin(kθ$_0$x)	for all x	[δ(θ-θ$_0$)-δ(θ+θ$_0$)]/(2i)
exp(-π(x/a)^2)	for all x	k|a|/(2π) exp[-π(kaθ/(2π))^2]
exp(-x/a)	for x>0	
0	else	k/(2π/(1/a+ikθ))

Examples:

1) Far Field of a Sinusoidal Amplitude Grating

We consider a slit of infinite height in the y-direction and width $2a$ in the x-direction which is homogeneously illuminated by a plane wave with wavelength λ. Thus, we have an electric field $E(x)$ which is equal to a constant amplitude E_0 for $|x| \le a$ and exactly zero for $|x|>a$. The amplitude transmission $t(x)$ inside the slit varies sinusoidally between 0 and 1, according to

$$t(x) = \frac{1}{2}(1+\cos[N2\pi x/a]) = \frac{1}{2}(1+g(x)), \qquad N=0,1,2,...$$

and $$g(x) = \cos[N2\pi x/a]$$

The transmission always exhibits maxima at the edges and center of the slit while N controls the number of oscillations in between. The far field $E_F(\theta)$ is proportional to the Fourier transform of the product $E(x)t(x)$ (with $k=2\pi/\lambda$ in (2.44)):

$$E_F(\theta) = c\ F[E(x)t(x)] = c\ \frac{1}{2}F[E(x)(1+g(x))]$$

By using the convolution theorem and the linearity theorem we can rewrite this as:

$$E_F(\theta) = c\ \frac{1}{2}(F[E(x)]+F[E(x)]\odot F[g(x)])$$

We can look up both Fourier transforms in the Table 2.2 (with $\theta_0=N\lambda/a$) and together with (2.46) we obtain:

$$E_F(\theta) = \frac{cE_0}{\lambda}\left[\frac{\sin(ka\theta)}{k\theta} + \int_{-\infty}^{+\infty}\frac{\sin(ka\phi)}{k\phi}\ \frac{1}{2}(\delta(\theta-\phi-\theta_0)+\delta(\theta-\phi+\theta_0))\ d\phi\right]$$

$$= \frac{caE_0}{\lambda}\left[\frac{\sin(ka\theta)}{ka\theta} + \frac{1}{2}\frac{\sin(ka(\theta-\theta_0))}{ka(\theta-\theta_0)} + \frac{1}{2}\frac{\sin(ka(\theta+\theta_0))}{ka(\theta+\theta_0)}\right]$$

The sinusoidal transmission generates peaks in the far field at angles $\theta=\pm\theta_0=\pm N\lambda/a$.

2) Collins Integral Written In Terms of Fourier Transforms

We can rewrite the Collins-Integral (2.29) in terms of Fourier transforms and inverse Fourier transforms of the electric field and the parabolic phase factors. This rearrangement is very useful for a fast numerical calculation of diffraction integrals since time-efficient algorithms for the computation of Fourier transforms exist (such as the Fast Fourier Transform Algorithm). The basic idea is to transform the Collins integral into a convolution integral and apply the Fourier transform convolution theorem. A straightforward rearrangement of the Collins integral (2.29) results in:

$$E_2(x_2,y_2) = \frac{i}{\lambda B}\exp[-ikL]\ \exp[-i\frac{\pi}{\lambda B}(D-1)(x_2^2+y_2^2)]\ \bullet$$

$$\int\int E_1(x_1,y_1)\ \exp[-i\frac{\pi}{\lambda B}(A-1)(x_1^2+y_1^2)]\ \ \exp[-i\frac{\pi}{\lambda B}[(x_1-x_2)^2+(y_1-y_2)^2]]\ dx_1dy_1$$

$$\text{|----------}f(x_1,y_1)\text{----------| |---------}g(x_1-x_2,y_1-y_2)\text{--------|}$$

which is a two-dimensional convolution integral.

We know that the Fourier transform of a convolution integral equals the product of the individual Fourier transforms $F(f)$ and $F(g)$. The inverse Fourier transform of this product is therefore equivalent to the convolution integral itself:

$$E_2(x_2,y_2) = \frac{i}{\lambda B} \exp[-ikL] \exp[-i\frac{\pi}{\lambda B}(D-1)(x_2^2+y_2^2)] \; F^{-1}\left[F(f(x_1,y_1)) \; F(g(x_1,y_1))\right]$$

The Fourier transform of $g(x_1,y_1)$ can be calculated analytically and the final form reads:

$$E_2(x_2,y_2) = \frac{i}{\lambda B} \exp[-ikL] \exp[-i\frac{\pi}{\lambda B}(D-1)(x_2^2+y_2^2)] \; F^{-1}\left[F(f(x_1,y_1)) \exp[-i\frac{\pi B}{\lambda}(\theta_x^2+\theta_y^2)]\right]$$

2.5 Gaussian Beams

2.5.1 Gaussian Beams in One-Dimensional Optical Systems

The Collins integral describes the changes in field structure as the electromagnetic field propagates through an optical system. In general the amplitude as well as the shape of the electric field distribution varies during propagation. We can, however, find a special class of fields for which the shape of the distribution remains constant and only the amplitude and lateral extent of the field is changed. These fields are referred to as eigensolutions of the diffraction integral. In the case of an unconfined electric field (which means that the integration is performed from -∞ to +∞), the eigensolutions of the Collins integral can be found analytically. Again, we first deal with one-dimensional optics which can be described by 2x2 ray transfer matrices and which exhibit rotational symmetry. An electric field $E_1(x,y)$ is an eigensolution of the Collins integral (2.29) if the following relation holds:

$$E_2(x_2,y_2) = \gamma \; E_1(\frac{x_2}{\sigma},\frac{y_2}{\sigma}) = \tag{2.48}$$

$$\frac{i}{\lambda B} \exp[-ikL] \int\int E_1(x_1,y_1) \exp[-i\frac{\pi}{\lambda B}(Ax_1^2+Dx_2^2-2x_1x_2 + Ay_1^2+Dy_2^2-2y_1y_2)] \; dx_1 dy_1$$

The form of the integral leads us to the assumption that a Gaussian field distribution

$$E_1(x_1,y_1) = E_0 \exp\left[\frac{-ik}{2q_1}(x_1^2+y_1^2)\right] \tag{2.49}$$

with q_1 being a complex number, might be a solution of the integral equation (2.48).

Insertion of (2.49) into (2.48) yields:

$$E_2(x_2,y_2) = \frac{E_0}{A-B/q_1} \exp\left[\frac{-ik}{2q_2}(x_2^2+y_2^2)\right] \tag{2.50}$$

with $$q_2 = \frac{Aq_1 + B}{Cq_1 + D} \tag{2.51}$$

This fundamental eigensolution of the diffraction integral is called the *Gaussian beam.* A Gaussian beam stays a Gaussian beam as the field propagates through an optical system with parabolic surfaces or parabolic index profiles. Only the amplitude and the characteristic parameter q, called the beam parameter, are changed. The transformation rule (2.51) is called the ABCD law of Gaussian optics. This ABCD law can be considered as a generalization of the geometrical optics ABCD law (1.90) already discussed in Section 1.2.6. In the limit $\lambda \to 0$, the geometrical optics ABCD law is obtained from (2.51) and all imaging characteristics of Gaussian beams can then be described by geometrical optics, as will be discussed in further detail below.

Free Space Propagation of Gaussian Beams

In the simplest case, the field distribution in plane 1 (at $z=0$) is a real Gaussian distribution with a beam radius w_0 (Fig. 2.11):

$$E_1(x_1,y_1) = E_0 \exp\left[-\frac{x_1^2+y_1^2}{w_0^2}\right] \tag{2.52}$$

A comparison with (2.49) indicates that the Gaussian beam parameter q_1 is given by:

$$q_1 = i\frac{\pi w_0^2}{\lambda} = iz_0 \tag{2.53}$$

Propagation over a distance z will change the beam parameter according to the ABCD law (2.51). By using the ray transfer matrix for free space propagation (1.10), the new beam parameter at a distance z is found to be:

$$q(z) = q_1 + z \tag{2.54}$$

In order to obtain the new Gaussian beam according to (2.50), the term $1/q(z)$ has to be determined. A straightforward calculation yields:

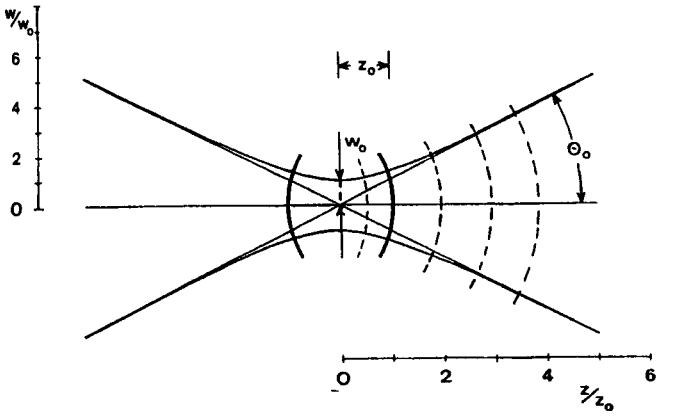


Fig. 2.11 Free space propagation of a Gaussian beam.

$$\frac{1}{q(z)} = \frac{1}{z_0(z/z_0 + z_0/z)} - \frac{i\lambda}{\pi w_0^2[1+(z/z_0)^2]} \tag{2.55}$$

The real and imaginary parts of $1/q(z)$ determine the phase and the beam radius of the Gaussian beam, respectively. To show this, we define two new quantities,

$$R(z) = z_0\left[\frac{z}{z_0} + \frac{z_0}{z}\right], \qquad w(z) = w_0\sqrt{1 + \left(\frac{z}{z_0}\right)^2} \tag{2.56}$$

(2.57)

We can then write the beam parameter $q(z)$ as:

$$\frac{1}{q(z)} = \frac{1}{R(z)} - \frac{i\lambda}{\pi w(z)^2} \tag{2.58}$$

The Gaussian beam at a distance z from plane 1 thus reads:

$$E_2(x_2,y_2) = \frac{E_0}{1-iz/z_0}\exp\left[-\frac{ik}{2}\frac{x_2^2+y_2^2}{R(z)}\right]\exp\left[-\frac{x_2^2+y_2^2}{w(z)^2}\right] \tag{2.59}$$

We see that the $R(z)$ represents the radius of curvature of the phase front and $w(z)$ is the beam radius, defined by a $1/e^2$-decrease of the intensity with respect to the on- axis value. The Gaussian beam diverges as it propagates along the z-axis according to (2.57). A characteristic parameter of the Gaussian beam is the Rayleigh range z_0 with:

$$z_0 = \frac{\pi w_0^2}{\lambda} \tag{2.60}$$

which is also referred to as the depth of field, or sometimes as the confocal parameter. The Rayleigh range denotes the distance from the origin (location of the beam waist w_0) at which the beam radius has increased by a factor of $\sqrt{2}$. At distances z much greater than the Rayleigh range the Gaussian beam exhibits the behavior of a spherical wave. In this case, the imaginary part of $1/q(z)$ becomes negligibly small and the radius of curvature $R(z)$ increases linearly with the distance z. The Gaussian ABCD law (2.51) will then be transformed into the geometrical ABCD law (1.90). The same behavior can be found if the wavelength λ is increased, since this is equivalent to decreasing the Rayleigh range.

For smaller distances z, however, the propagation behavior is completely different than that of spherical waves in geometrical optics. The radius of curvature $R(z)$ of a Gaussian beam shows a minimum value of $2z_0$ at $z=z_0$, whereas a spherical wave increases its phase curvature proportionally to the distance from the origin. This is why Gaussian beams exhibit imaging properties that are quite different from those of geometrical optics.

For large distances $z \gg z_0$ the beam radius $w(z)$ approaches an asymptote at a finite angle θ_0. This angle, known as the divergence angle, reads:

$$\theta_0 = \lim_{z\to\infty} \frac{w(z)}{z} = \frac{w_0}{z_0} = \frac{\lambda}{\pi w_0} \tag{2.61}$$

The product of the divergence angle and the beam waist is called the *beam parameter product*. All Gaussian beams have the same beam parameter product of:

$$w_0\,\theta_0 = \frac{\lambda}{\pi} \tag{2.62}$$

which represents the minimum value possible for any field distribution if a suitable definition for the beam radii is used (see Sec. 2.6). The beam parameter product is a constant of the beam and is not changed by propagation through an optical system, provided that the optical elements can be described by ray transfer matrices.

Example: For waist w_0 of 1mm and a wavelength of λ=500nm we get:

Rayleigh range:	6.283 m
Divergence angle:	0.159 mrad
Distance z with $w(z)$=1m:	6.283 m.

Transformation Rules for Gaussian Beams

In the general case of an arbitrary optical system, the relationships for the beam radius and the radius of curvature become more complicated. If the ray transfer matrix of the optical system is given by:

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

the new beam radius and phase curvature can be calculated by applying the ABCD law (2.51). The Gaussian beam entering the optical system is described by the q-parameter $q=z+iz_0$, and at the exit plane a new Gaussian beam emerges with the beam parameter $q'=-z'+iz_0'$ (Fig. 2.12). The distances z and z' denote the location of the beam waist of the initial and the transformed Gaussian beams, respectively. The ABCD law yields for the new waist location and the new Rayleigh range:

$$z' = \frac{(Az+B)(Cz+D) - ACz_0^2}{C^2z_0^2 + (Cz+D)^2} \qquad \text{if } C \neq 0 \tag{2.63}$$

$$z' = -\frac{Az+B}{D} \qquad \text{if } C=0 \tag{2.64}$$

$$z_0' = z_0 \frac{Cz'+A}{Cz+D} \qquad \text{with } z' \text{ from (2.63/2.64)} \tag{2.65}$$

Note that the new beam waist is located left of the exit plane if z' is negative. A positive value of z' refers to a beam waist location to the right of the optical system, as it is the case in Fig. 2.12.

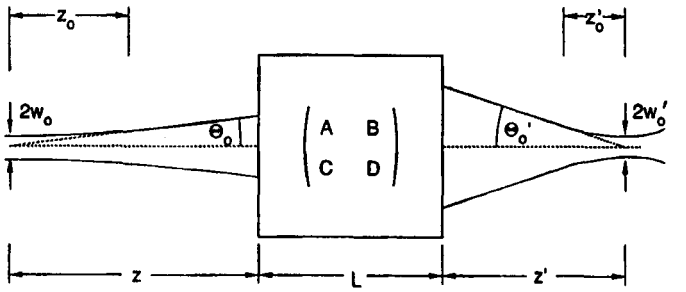


Fig. 2.12 The propagation through an optical system of length L and ray transfer matrix M transforms the Gaussian beam into a new Gaussian beam with beam waist location z', Rayleigh range z_0', and divergence angle θ_0'.

The new angle of divergence θ'_0 and the new beam waist w'_0 can be calculated from:

$$\theta'_0 = \sqrt{\frac{\lambda}{\pi z'_0}} \quad , \qquad w'_0 = \sqrt{\frac{\lambda z'_0}{\pi}}$$

which means that the beam parameter product $\theta'_0 w'_0$ remains constant ($=\lambda/\pi$).

A different approach to calculating the Gaussian beam transformation is to link the beam radii and angles of divergence at the exit plane of the optical system to those at the entrance plane. If w_1, R_1, and θ_{01} denote the beam radius, radius of curvature, and divergence angle at the entrance plane, respectively, then the beam radius w_2, the radius of curvature R_2, and divergence angle θ_{02} at the exit plane are given by:

$$w_2^2 = A^2 w_1^2 + 2\,ABw_1^2/R_1 + B^2\theta_{01}^2 \tag{2.66}$$

$$\theta_{02}^2 = C^2 w_1^2 + 2\,CDw_1^2/R_1 + D^2\theta_{01}^2 \tag{2.67}$$

$$R_2 = \frac{w_2^2}{\sqrt{w_2^2\theta_{02}^2 - (\lambda/\pi)^2}} \tag{2.68}$$

The new beam waist location z', the Rayleigh range z'_0, and the beam waist radius w'_0 can be found by using the relations:

$$z' = \frac{-1}{R_1}\,\frac{1}{\dfrac{1}{R_1^2} + \dfrac{\lambda^2}{\pi^2 w_2^4}} \quad , \qquad z'_0 = \frac{\lambda}{\pi\theta_{02}^2} \quad , \qquad w'_0 = \frac{\lambda}{\pi\theta_{02}}$$

In the following we will present the transformation rules for common optical systems:

a) Plane Dielectric Slab

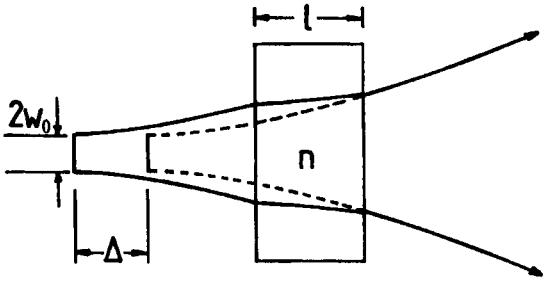


Fig. 2.13 Propagation of a Gaussian beam through a plane dielectric slab with refractive index n.

The ray transfer matrix of the plane slab (see Fig. 1.5) yields with (2.63)-(2.65):

$$z'_0 = z_0 \qquad z' = -z - \frac{l}{n}$$

$$\theta'_0 = \theta_0 \qquad w'_0 = w_0$$

If the refractive index of the slab were $n=1$, we would get the new waist location $z' = -z-l$, which is just the location of the original beam waist. For indexes other than unity, the new beam waist is shifted to the right (as seen looking back into the slab) by:

$$\Delta = l\left(\frac{n-1}{n}\right)$$

but the Gaussian beam remains unchanged.

b) Propagation Through a Telescope

Telescopes are widely used in laser systems to magnify the beam and thus decrease the angle of divergence (Fig. 2.14). A lower divergence results in a smaller spot size in the focal plane of a focusing lens. With (2.63)-(2.65) we get:

$$z' = -z\left[\frac{f_2}{f_1}\right]^2 + \frac{f_2}{f_1}(f_1+f_2)$$

$$w'_0 = w_0\frac{f_2}{f_1}, \qquad \theta'_0 = \theta_0\frac{f_1}{f_2} \qquad z'_0 = z_0\left[\frac{f_2}{f_1}\right]^2$$

A telescope with magnification $M=f_2/f_1$ decreases the divergence by a factor $1/M$ and increases the beam radius by M. Again, the beam parameter product stays constant.

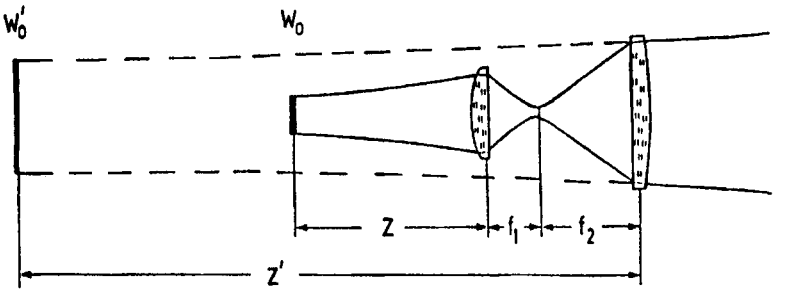


Fig. 2.14 Gaussian Beam propagation inside a telescope (not to scale).

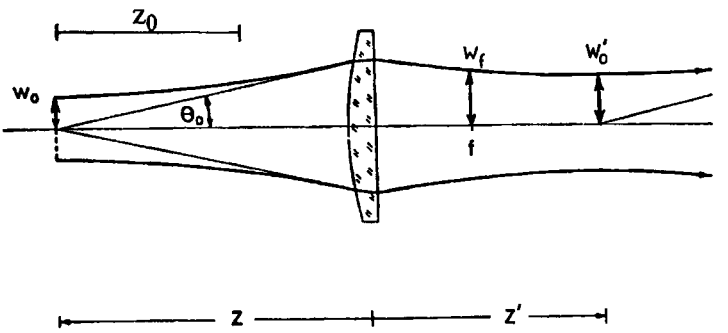


Fig. 2.15 Gaussian beam propagation through a focusing lens. An image of the beam waist is generated at a distance z' behind the lens.

c) Imaging/Focusing with a Lens

We are looking for the location and the size of the new beam waist when the Gaussian beam is transformed by a lens. For a lens with positive focal length f, the Gaussian beam is focused as depicted in Fig. 2.15. Since the focusing of Gaussian beams plays an important role in laser optics, we will discuss this beam propagation problem in more detail.

Note that we use the sign convention of geometrical optics for the distances z and z': both distances are positive if the object (beam waist w_0) and the image (beam waist w'_0) are located to the left and to the right of the lens, respectively. The imaging condition of geometrical optics:

$$\frac{1}{z} + \frac{1}{z'} = \frac{1}{f}$$

does not apply to the waist of Gaussian beams since it is only valid for spherical waves whose radii of curvatures increase proportionally to the distance z. For Gaussian beams the different phase front propagation behavior generates an additional term in the imaging condition. By using the ABCD law (2.51) and setting the radii of curvature to infinity at both planes, the imaging condition can be found to be:

$$\frac{1}{z} + \frac{1}{z'} = \frac{1}{f} + \frac{z_0^2}{z(z^2+z_0^2-zf)} \tag{2.69}$$

with z' being the distance of the new beam waist location from the lens. If z' is positive the beam waist is located at the right hand side of the lens.

The following relations hold for the parameters depicted in Fig. 2.15:

$$w'_0 = w_0 \frac{f}{\sqrt{z_0^2 + (z-f)^2}} \tag{2.70}$$

$$z'_0 = z_0 \frac{f^2}{z_0^2 + (z-f)^2} \tag{2.71}$$

$$w_f = f\,\theta_0 \tag{2.72}$$

Figures 2.16 and 2.17 show the graphic presentation of (2.69) and (2.70), respectively. The imaging condition for geometrical optics is reached in the limit $z_0/z=0$ which means that the distance of the object beam waist has to be much greater than the Rayleigh range. This is to be expected because for large distances from the beam waist, the Gaussian beam behaves like a spherical wave as far as propagation is concerned.

The reader should note that there are major differences in the imaging of the waist of Gaussian beams as compared to geometrical optics. One example worth mentioning is the beam waist of the Gaussian beam located in the front focal plane $(z=f)$. According to (2.69), the image waist is found in the back focal plane $(z'=f)$. This is quite surprising since from our experience in geometrical optics we would expect that placing a source in the front focal plane would produce a collimated beam (image at infinity)! This focusing experiment can serve as a quick check to determine whether a laser beam can be described as a Gaussian beam or as a geometrical optics light source (at the exit aperture of a fiber for example).

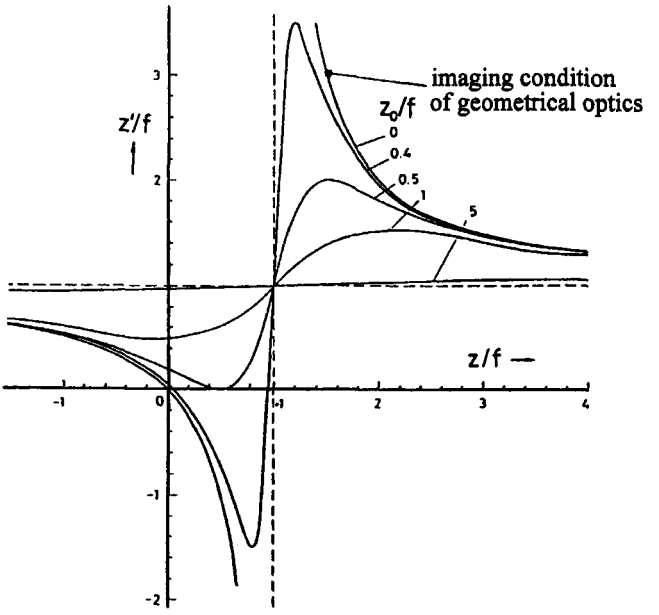


Fig. 2.16 Imaging of Gaussian beams according to (2.69).

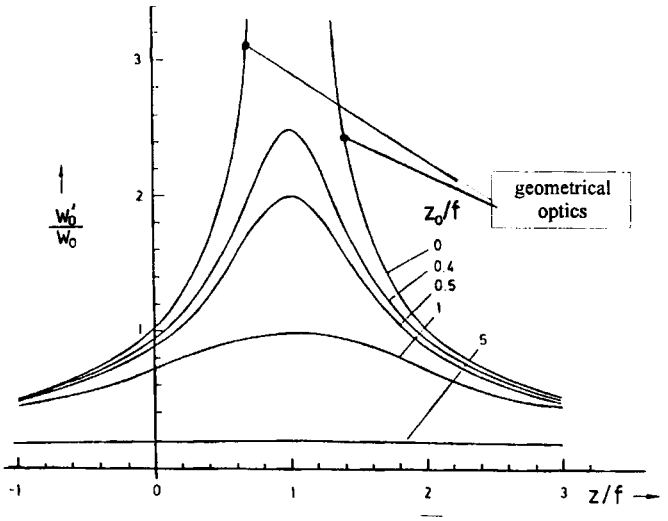


Fig. 2.17 Imaging of Gaussian beams. Shown is the relationship between the magnification and the distance z of the beam waist from the focusing lens according to (2.70).

Equations (2.70) and (2.71) indicate that we cannot simultaneously attain a small focal spot radius w'_0 and a large Rayleigh range z'_0. Division of (2.70) by (2.71) yields:

$$\frac{w'^2_0}{z'_0} = \frac{w_0^2}{z_0} = \frac{\lambda}{\pi} = w_0\theta_0 \tag{2.73}$$

In addition to the beam parameter product $w_0\theta_0$, the ratio of focal spot size to Rayleigh range is a constant of the beam. If we try to decrease the focal spot radius (by choosing a lens with a shorter focal length or by increasing the Rayleigh range z_0 of the object beam with a telescope), we will reduce the Rayleigh range proportional to the area of the focal spot. Equation (2.73) holds only for Gaussian beams. Laser beams usually have beam parameter products that are larger than λ/π (by a factor 3-100, depending on the type of laser and the output power). Equation (2.73) then reads:

$$\frac{w'^2_0}{z'_0} = \frac{w_0^2}{z_0} = M^2\frac{\lambda}{\pi} = M^2\, w_0\theta_0 \tag{2.74}$$

with M^2 being the beam propagation factor. It is for this reason that in laser material processing it is difficult to generate thin, deep cuts into a work piece. The quality of a laser cut is commonly defined by the aspect ratio which is the depth of the cut over the width. In order to get a high aspect ratio it is a prerequisite to have a laser beam with a low beam parameter product. Laser beam quality can thus be characterized by the beam parameter product. High beam quality is equivalent to a low beam parameter product (beam propagation factor, M^2, near unity) or, described in other words, by a large Rayleigh range

for a given focal spot size. It is interesting that the beam radius in the focal plane (2.72) depends only on the divergence angle of the Gaussian beam in front of the lens. This is a direct consequence of the fact that the Fourier transform of the beam is generated in the focal plane.

In most cases the beam waist w_0 and the Rayleigh range z_0 of the beam are not known and only the beam radius w_L at the lens can be determined and (2.70) cannot be used to calculate the focal spot radius w'_0. Fortunately, if the object distance z is large compared to the focal length f ($z>5f$ is a good rule of thumb), we can derive an approximate formula for the spot size w'_0. According to (2.57) the beam radius on the lens is given by:

$$w_L = w_0 \sqrt{1+(z/z_0)^2}$$

By using this expression and the assumption $z \gg f$, we can rewrite (2.70) and (2.71):

$$w'_0 = \frac{w_0}{z_0} \frac{f}{\sqrt{1+((z-f)/z_0)^2}} = \frac{\lambda f}{\pi w_0} \frac{1}{\sqrt{1+((z-f)/z_0)^2}} \approx \frac{\lambda f}{\pi w_L} \tag{2.75}$$

$$z' \approx f \tag{2.76}$$

Example: A Nd:YAG laser whose output beam is a Gaussian beam with $w_0=0.582mm$, $z_0=1m$, and $\lambda=1.064\mu m$ is focused by means of a $50mm$ lens. The lens is at $z=1m$ in front of the beam waist. Using (2.69) and (2.70) we get for the position and the size of the focal spot:

z' = 51.25mm $\quad$ $w'_0 = 0.0211$ mm

The approximated equations (2.75) and (2.76) provide the following solutions:

z' = 50.00mm $\quad$ $w'_0 = 0.0206$ mm

d) Penetration into a Medium with a Higher Index of Refraction

We have a Gaussian beam hitting a dielectric planar interface with refractive index n. The radius of curvature and beam radius of the beam at the interface are R_1 and w_1, respectively. The ray transfer matrix for the refraction at the interface reads:

$$\boldsymbol{M} = \begin{pmatrix} 1 & 0 \\ 0 & 1/n \end{pmatrix}$$

The ABCD law (2.51) yields for the new q-parameter q_2:

$$\frac{1}{q_2} = \frac{1}{nq_1}$$

Separation into the radius of curvature R and the beam radius w according to (2.58) yields:

$$R_2 = n\, R_1 \quad , \qquad w_2^2 = w_1^2$$

The refraction leads to an increase in curvature, but the beam radius at the interface remains unchanged.

2.5.2 Elliptical Gaussian Beams

So far one-dimensional Gaussian beams or Gaussian beams of rotational symmetry have been discussed. They are completely characterized by the complex beam parameter q with:

$$\frac{1}{q} = \frac{1}{R} - \frac{i\lambda}{\pi w^2}$$

By measuring the real part and the imaginary part of q at a position z in the reference frame, the beam is completely determined. Equations (2.56) to (2.58) provide the position z_w of the waist and the Rayleigh range z_0. If it is known that the beam is a TEM_{00} mode with rotational symmetry, the waist radius w_0 is obtained from (2.60), and (2.61) provides the far field divergence angle θ_0. Such a beam is called a *stigmatic fundamental mode*. It exhibits a circular spot and a spherical wave front at every distance z. The curves of constant intensity (isophotes) and the lines of constant phase (isophases) in the x-y plane are always circles.

However, in some lasers, such as diode lasers and slab lasers, the fundamental mode exhibits an elliptical cross section. Let us assume a stigmatic TEM_{00} mode, generated by a HeNe laser, passing through a thin cylinder lens with refractive power D_y and the focal line along the x-axis (see Fig. 1.18b). The spot size directly behind the lens is unchanged and the radius of curvature in the x-direction is unaffected. The radius of curvature in the y-direction now reads:

$$\frac{1}{R_y} = \frac{1}{R} - D_y$$

The isophases are now ellipses. Due to the focusing in y-direction, the beam radius w_y will differ from w_x with increasing distance z from the lens (Fig. 2.18). The isophotes, which are circles right behind the lens, become elliptical. Such a beam is called simple astigmatic. The ellipses of constant intensity and constant phase have the same orientation (along the principal axes of the cylinder lens) as the beam propagates.

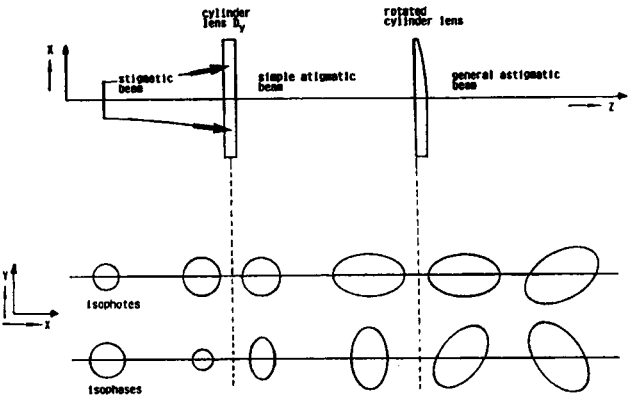


Fig. 2.18 Transformation of a stigmatic beam into a simple astigmatic beam and into a general astigmatic beam. For a general astigmatic beam, the isophotes and the isophases have different orientation and the ellipses rotate as the beam propagates.

The two ellipses are determined by 4 parameters plus one additional parameter for the orientation of the ellipses with respect to the reference frame. The field of a simple astigmatic beam can still be separated in the form:

$$E(x,y) = E(x)\ E(y)$$

The Gaussian beam propagation law (2.51) can be applied separately in the x- and the y-direction for the q- parameters q_x and q_y, respectively. If such a beam is focused with a spherical lens, two line foci will appear at different distances from the lens.

After a distance L the simple astigmatic beam passes a rotated cylinder lens. The ellipse of constant intensity is not affected by this, but the ellipse of constant phase is rotated. The two ellipses now have different orientations. Such a beam is called a general astigmatic beam. The main axes of the two ellipses form an angle to each other and the orientation does not remain constant during the propagation. A general astigmatic beam is characterized by six parameters, three for each ellipse (two for the main axes and one for the orientation).

The concept of the complex beam parameter can still be applied if we introduce a complex beam matrix $\boldsymbol{Q}^{-1}$ with [1.35,1.36]:

$$\boldsymbol{Q}^{-1} = \begin{pmatrix} \frac{1}{q_{xx}} & \frac{1}{q_{xy}} \\ \frac{1}{q_{xy}} & \frac{1}{q_{yy}} \end{pmatrix} \tag{2.77}$$

Table 2.3 Properties of Gaussian beams

	constant intensity	constant phase	free parameters
stigmatic	$\frac{x^2}{w^2(z)} + \frac{y^2}{w^2(z)} = C$	$\frac{kx^2}{R(z)} + \frac{ky^2}{R(z)} = C$	2
simple astigmatic	$\frac{x^2}{w_x^2(z)} + \frac{y^2}{w_y^2(z)} = C$	$\frac{kx^2}{R_x(z)} + \frac{ky^2}{R_y(z)} = C$	5
general astigmatic	$\frac{x^2}{w_x^2(z)} + \frac{xy}{w_{xy}^2(z)} + \frac{y^2}{w_y^2(z)} = C$	$\frac{kx^2}{R_x(z)} + \frac{kxy}{R_{xy}(z)} + \frac{ky^2}{R_y(z)} = C$	6

The general astigmatic fundamental mode is given by:

$$E = E_0 \exp\left[\frac{-ik}{2} \boldsymbol{r}\boldsymbol{Q}^{-1}\boldsymbol{r}\right] \quad , \quad \boldsymbol{r} = (x,y) \tag{2.78}$$

with

$$\boldsymbol{r}\boldsymbol{Q}^{-1}\boldsymbol{r} = \frac{x^2}{q_{xx}} + \frac{2xy}{q_{xy}} + \frac{y^2}{q_{yy}} \tag{2.79}$$

The beam is determined by the three different complex parameters of the $\boldsymbol{Q}^{-1}$ matrix. The propagation of such a beam through ABCD-type optical systems can be evaluated by applying the Collins integral (2.32). A straightforward but troublesome calculation yields the generalized ABCD law [1.53]:

$$\boldsymbol{Q}_2^{-1} = (\boldsymbol{C} + \boldsymbol{D}\boldsymbol{Q}_1^{-1})(\boldsymbol{A} + \boldsymbol{B}\boldsymbol{Q}_1^{-1})^{-1} \tag{2.80}$$

with $\boldsymbol{A},\boldsymbol{B},\boldsymbol{C},\boldsymbol{D}$ being the 2x2 submatrices introduced in Sec. 1.2.4.

Examples:

1) Transformation of a general astigmatic Gaussian beam by an astigmatic thin lens. By using (2.80) with:

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \boldsymbol{I} \; , \; \boldsymbol{B} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \boldsymbol{0} \; , \; \boldsymbol{C} = \begin{pmatrix} -1/f_x & 0 \\ 0 & -1/f_y \end{pmatrix} \; , \; \boldsymbol{D} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \boldsymbol{I}$$

we obtain: $\boldsymbol{Q}_2^{-1} = \boldsymbol{C} + \boldsymbol{Q}_1^{-1}$ (2.81)

This relation is equivalent to the three equations:

$$\frac{1}{q_{xx,2}} = -\frac{1}{f_x} + \frac{1}{q_{xx,1}}$$

$$\frac{1}{q_{xy,2}} = \frac{1}{q_{xy,1}}$$

$$\frac{1}{q_{yy,2}} = -\frac{1}{f_y} + \frac{1}{q_{yy,1}}$$

The real parts of $1/q_{xx}$ and $1/q_{yy}$, which represent the radii of curvatures, are changed.

2) Propagation of a general astigmatic Gaussian beam in free space over a distance z. The four submatrices are $\boldsymbol{A}=\boldsymbol{I}$, $\boldsymbol{B}=z\boldsymbol{I}$, $\boldsymbol{C}=\boldsymbol{0}$, $\boldsymbol{D}=\boldsymbol{I}$. The propagation law (2.80) yields:

$$\frac{1}{q_{xx,2}} = \frac{1}{N}\left[\frac{1}{q_{xx,1}}\left(1+\frac{z}{q_{yy,1}}\right) - \frac{z}{q_{xy,1}^2}\right]$$

$$\frac{1}{q_{xy,2}} = \frac{1}{N}\,\frac{1}{q_{xy,1}}$$

$$\frac{1}{q_{yy,2}} = \frac{1}{N}\left[\frac{1}{q_{yy,1}}\left(1+\frac{z}{q_{xx,1}}\right) - \frac{z}{q_{xy,1}^2}\right]$$

with $$N = \left(1 + \frac{z}{q_{xx,1}}\right)\left(1 + \frac{z}{q_{yy,1}}\right) - \frac{z^2}{q_{xy}^2}$$

This complicated set of equations reflects the fact that the two ellipses are rotating while the beam propagates in the z-direction. The simple case of the stigmatic beam is immediately obtained for $1/q_{xx,1}=1/q_{yy,1}=1/q_1$, $1/q_{xy,1}=0$ and the upper equation results in (2.51).

A stigmatic beam can be transformed into a general astigmatic beam by a suitable set of lenses as shown in Fig. 2.18. The inverse transformation can also be applied, but only for Gaussian beams, as was shown in [1.36]. First, the matrix $\boldsymbol{Q}^{-1}$ has to be diagonalized by a rotation. The angle of rotation is given by:

$$\tan(2\phi) = \frac{2/q_{xy}}{1/q_{xx} - 1/q_{yy}} \qquad (2.82)$$

Since this equation generally yields a complex angle, a simple astigmatic lens cannot transform a general astigmatic beam into a stigmatic one. These transformation properties were first investigated in detail by Arnaud [1.32]. It is interesting to note that the complex angle ϕ is constant for free space propagation. This can be easily proved by inserting the three equations of example 2 into (2.82).

The transformation of a general astigmatic Gaussian beam into a stigmatic one has to be done in two steps. First we generate a simple astigmatic beam by a rotation and in a second step the simple astigmatic beam is transformed into a stigmatic one. The ellipse of constant intensity is rotated with respect to the reference frame by an angle θ_w, given by:

$$\tan(2\theta_w) = \frac{2\ Im(1/q_{xy})}{Im(1/q_{xx} - 1/q_{yy})} = \frac{2/w_{xy}^2}{1/w_{xx}^2 - 1/w_{yy}^2} \tag{2.83}$$

and the ellipse of constant phase is rotated by θ_p, with:

$$\tan(2\theta_p) = \frac{2\ Re(1/q_{xy})}{Re(1/q_{xx} - 1/q_{yy})} = \frac{2/R_{xy}}{1/R_{xx} - 1/R_{yy}} \tag{2.84}$$

If θ_p is equal to θ_w, we already have a simple astigmatic beam, and by rotating the reference frame by this angle we obtain the beam in the main axes presentation. If the two angles are not equal, we have to rotate the axis of the phase ellipse by a suitable bifocal or cylindrical lens. In both cases we obtain a simple astigmatic beam. If this beam is focused, two foci appear at different distances from the focusing lens. By using a set of cylindrical lenses the foci can be imaged into one circular waist [1.57].

Note that all the equations discussed above hold only for Gaussian beams. For other beams such as higher order modes, these relations are not valid! The case of general fields will be discussed in the next section.

2.6 Intensity Moments and Beam Propagation

2.6.1 Stigmatic and Simple Astigmatic Beams

We have seen in the last section how to calculate the beam radius w and the divergence angle θ of Gaussian beams in any plane within an optical system by using the ABCD law for the beam parameter q. If we start at the plane of the beam waist w_0 and propagate through an optical system with ray matrix elements A, B, C, D, the beam radius w at the new plane and the divergence angle θ of the new Gaussian beam read according to (2.66) and (2.67)

$$w^2 = A^2 w_0^2 + B^2 \theta_0^2 \tag{2.85}$$

$$\theta^2 = C^2 w_0^2 + D^2 \theta_0^2 \tag{2.86}$$

assuming the simplest case of a stigmatic Gaussian beam (circular symmetry). For simple astigmatic beams the above relations can be applied separately for the two main axes. Beam radius and divergence angle are both defined by the $1/e^2$-decrease of the intensity. For free space propagation over a distance z with $A=1$, $B=z$, $C=0$, and $D=1$, we obtain the familiar result:

$$w = w_0\sqrt{1 + (z/z_0)^2} \quad , \qquad \theta = \theta_0$$

or $$w^2 = w_0^2 + z^2\theta_0^2$$

The majority of laser beams, however, cannot be described by Gaussian beams. The application of the above shown propagation rules is restricted to lasers working in fundamental mode operation like HeNe lasers or single emitter diode lasers. Fortunately, the ABCD law can be generalized to arbitrary, partially coherent beams by defining beam sizes via the intensity moments of the beam [1.53,1.58,1.60,1.62,1.64]. In this section we will discuss simple astigmatic beams. These are elliptical beams whose main axes of the phase and the intensity ellipses coincide with the x,y reference frame, the equivalent to the simple astigmatic Gaussian beam of Sec. 2.5.2. The field can then be factorized into a product of functions depending on x or y. The propagation of the x and the y terms can be treated separately. The stigmatic beam with circular symmetry is a special case of the astigmatic beam. In the following only the x terms are considered. By replacing x with y the corresponding formulas for the y-direction are obtained. The general case in which a separation in x and y terms is not possible will be discussed in Sec. 2.6.2.

First Order Moments

The first order moments are defined as:

$$\langle x \rangle = \frac{1}{P_N} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x|E(x,y)|^2 dxdy \tag{2.87}$$

$$\langle \theta_x \rangle = \frac{1}{P_N} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \theta_x |E_F(\theta_x,\theta_y)|^2 d\theta_x d\theta_y \tag{2.88}$$

with P_N being a normalization factor, related to the total power in the beam:

$$P_N = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} |E(x,y)|^2 dxdy = \lambda^2 \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} |E_F(\theta_x,\theta_y)|^2 d\theta_x d\theta_y \tag{2.89}$$

The Fourier transform in (2.88) can be replaced by the first derivative of the field. Application of (2.44) and (2.45) yields:

$$\langle\theta_x\rangle = \frac{-i\lambda}{2\pi P_N} \int_{-\infty}^{+\infty} E(x,y)\frac{\delta E^*(x,y)}{\delta x} dxdy \tag{2.90}$$

This equation is sometimes more convenient for analytical calculations. The above definitions require that the field decreases sufficiently fast with x, otherwise (2.87) will diverge. Furthermore, in order to get a finite value in (2.88) the field must be continuous. Note that although the integrals have infinite limits, the paraxial approximation has to hold for the electric field. This means that the far field exists only for small values of θ_x.

The first moments provide the center of gravity of the near field and the propagation direction (far field). These moments can be measured by recording the intensity distributions with a CCD camera and processing the image mathematically using image processing software. It is interesting to look at the propagation of the first moments. By applying the Collins integral, (2.87) and (2.88) yield after some troublesome but straightforward calculations:

$$\langle x_2\rangle = A\langle x_1\rangle + B\langle\theta_{x1}\rangle \tag{2.91}$$

$$\langle\theta_{x2}\rangle = C\langle x_1\rangle + D\langle\theta_{x1}\rangle \tag{2.92}$$

and the corresponding expressions for the y-direction. These equations are identical with those we obtained for the propagation of rays in geometrical optics, as discussed in Sec. 1.2. The center of gravity of an arbitrary field propagates through optical systems in the same way as rays, a very simple and satisfying result.

Second Order Moments

For simple astigmatic beams the following relations can be applied separately for the x- and the y-direction. Beam radius and divergence angle are defined by:

$$\langle w_x^2\rangle = \frac{4}{P_N} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^2 |E(x,y)|^2 dxdy \tag{2.93}$$

$$\langle\theta_x^2\rangle = \frac{4\lambda^2}{P_N}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\theta_x^2|E_F(\theta_x,\theta_y)|^2 d\theta_x d\theta_y \tag{2.94}$$

Equation (2.94) can again be transformed by using the Fourier transform of (2.87):

$$\langle\theta_x^2\rangle = \frac{\lambda^2}{\pi^2 P_N}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left|\frac{\delta E(x,y)}{\delta x}\right|^2 dxdy \tag{2.95}$$

A third, mixed moment of second order exists which is defined by:

$$\langle w_x\theta_x\rangle = \frac{-i\lambda}{\pi P_N}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x\,[E(x,y)\frac{\delta E^*(x,y)}{\delta x} - E^*(x,y)\frac{\delta E(x,y)}{\delta x}]\,dxdy \tag{2.96}$$

The factor 4 in the above equations is introduced to adapt the results to the waist radius and the divergence angle of a Gaussian beam. If the electric field of a Gaussian beam, which may be astigmatic but aligned, is inserted one obtains:

$$\langle w_{x,y}^2\rangle = w_{x,y}^2$$

$$\langle\theta_{x,y}^2\rangle = \theta_{0\,x,y}^2$$

$$\langle w_{x,y}\theta_{x,y}\rangle = \frac{w_{x,y}^2}{R_{x,y}}$$

The second intensity moments of an arbitrary field can be identified with the beam radius, the angle of divergence, and the radius of curvature. It is assumed that the origin of the reference frame coincides with the centers of gravity, which means that the first moments of the field vanish. At the beam waist the radius of curvature is infinite. For an arbitrary field the waist can thus be defined by the condition:

$$\langle w_x\theta_x\rangle = 0$$

If the beam and the optical system both exhibit rotational symmetry (a one-dimensional optical system with a 2x2 ray transfer matrix), the second moment definitions of the beam radii read:

$$\langle w_r^2\rangle = \frac{4\pi}{P_N}\int_0^{\infty} r^3|E(r)|^2 dr \tag{2.97}$$

$$\langle\theta^2\rangle = \frac{4\pi\lambda^2}{P_N}\int_0^\infty \theta^3 |E_F(\theta)|^2 d\theta \tag{2.98}$$

$$\langle w_r\theta_r\rangle = \frac{-i\lambda}{P_N}\int_0^\infty r^2 \left[E\frac{\delta E^*}{\delta r} - E^*\frac{\delta E}{\delta r}\right] r\, dr \tag{2.99}$$

Using these definitions of the beam radii and divergence angles the generalized ABCD law for one-dimensional optics can be derived by applying the Collins integral or the Wigner function presented in Sec.2.7. The ABCD law holds in rectangular symmetry as well as for the radial components:

$$\langle w_2^2\rangle = A^2\langle w_1^2\rangle + 2AB\langle w_1\theta_1\rangle + B^2\langle\theta_1^2\rangle \tag{2.100}$$

$$\langle w_2\theta_2\rangle = AC\langle w_1^2\rangle + (AD+BC)\langle w_1\theta_1\rangle + BD\langle\theta_1^2\rangle \tag{2.101}$$

$$\langle\theta_2^2\rangle = C^2\langle w_1^2\rangle + 2CD\langle w_1\theta_1\rangle + D^2\langle\theta_1^2\rangle \tag{2.102}$$

It is convenient to refer to the beam waist with $\langle w_1^2\rangle = \langle w_{01}^2\rangle$ *and* $\langle w_1\theta_1\rangle = 0$. The above equations then reduce to:

$$\langle w_2^2\rangle = A^2\langle w_{01}^2\rangle + B^2\langle\theta_1^2\rangle \tag{2.103}$$

$$\langle w_2\theta_2\rangle = AC\langle w_{01}^2\rangle + BD\langle\theta_1^2\rangle \tag{2.104}$$

$$\langle\theta_2^2\rangle = C^2\langle w_{01}^2\rangle + D^2\langle\theta_1^2\rangle \tag{2.105}$$

We can define a Rayleigh range z_0 by generalizing the expression from Gaussian optics:

$$\langle z_0^2\rangle = \langle w_0^2\rangle / \langle\theta_0^2\rangle \tag{2.106}$$

Be aware that the Rayleigh range can be different for the x- and the y-direction if the beam is elliptical. By using this definition, the generalized law for free space propagation of an arbitrary field over a distance z has the same form as for Gaussian beam propagation:

$$\langle w^2\rangle = \langle w_0^2\rangle \left[1 + \frac{z^2}{\langle z_0^2\rangle}\right] \tag{2.107}$$

The product of the waist $\sqrt{\langle w_0^2 \rangle}$ and the divergence $\sqrt{\langle \theta^2 \rangle}$ is called the beam parameter product. It characterizes the beam quality. In the best case of a Gaussian beam it results in:

$$\left[\langle w_0^2 \rangle \langle \theta^2 \rangle\right]_{TEM_{00}} = \frac{\lambda^2}{\pi^2}$$

For higher order modes (see Sec. 5.2.1) one obtains in rectangular symmetry:

$$\left[\langle w_0^2 \rangle \langle \theta^2 \rangle\right]_{TEM_{m}} = (2m+1)^2 \frac{\lambda^2}{\pi^2}$$

and in circular symmetry:

$$\left[\langle w_0^2 \rangle \langle \theta^2 \rangle\right]_{TEM_{pl}} = (2p+\ell+1)^2 \frac{\lambda^2}{\pi^2}$$

Arbitrary fields, especially high power laser beams, are a mixture of modes and the beam parameter product is defined by:

$$\left[\langle w_0^2 \rangle \langle \theta^2 \rangle\right]_{general} = M^4 \frac{\lambda^2}{\pi^2} \tag{2.108}$$

with M^2 being the beam propagation factor. With the exception of the fundamental mode (Gaussian beam, $M^2=1$), the beam propagation factor is always larger than one. M^2 is a constant of the beam and does not change during propagation through ABCD-type optical systems.

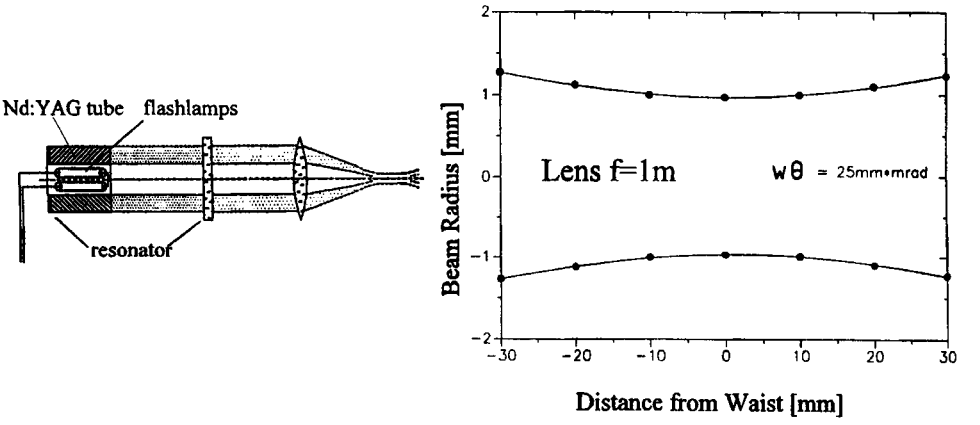


Fig. 2.19 Second moment beam radii in the vicinity of the beam waist of a focused Nd:YAG tube laser beam. A parabolic fit using (2.107) yields a beam parameter product of 25 mm mrad.

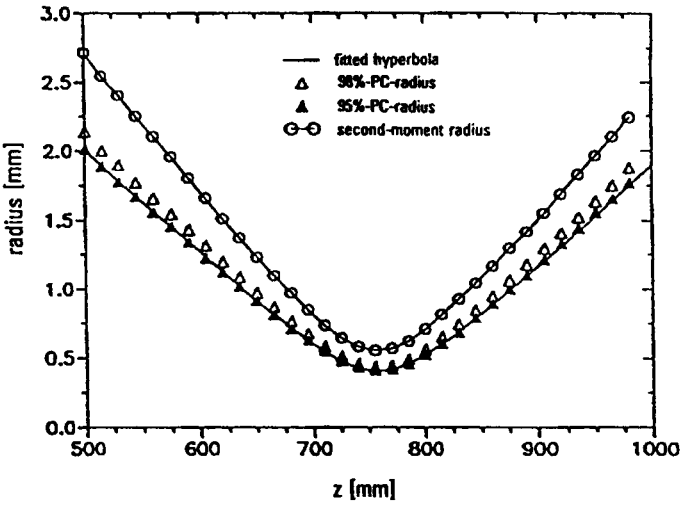


Fig. 2.20 Measured beam radii of a focused Nd:YAG laser beam (stable resonator in multimode operation) using the second moment definition (2.97) and two different power content (PC) definitions of the beam radius [1.70] (© SPIE 1994).

Figure 2.19 shows an experimental application of the generalized propagation law (2.103). The beam from an Nd:YAG tube laser (outer beam diameter at the output coupler: 55mm, inner beam diameter: 35mm) was focused by means of a focusing lens with focal length f=1m. The second moment beam radii were determined in the vicinity of the focal spot according to (2.97) by using a CCD camera and an image processing system. The solid line in the right hand graph represents the parabola, fitted according to (2.107). This parabolic fit enables one to determine the waist radius $\sqrt{\langle w_0^2 \rangle}$ and the Rayleigh range $\sqrt{\langle z_0^2 \rangle}$. The knowledge of these two parameters is sufficient to calculate the divergence angle and the beam parameter product [1.66]. It should be noted that the second moment beam radii do not exist for intensity distributions that decrease slower than $1/x^2$. This is the reason second order moment radius definitions cannot be applied to beams diffracted by hard edge apertures (the second moment diverges in the far field).

In practice, beam radii and divergence angles are defined via the confinement of 86.5% of the total beam power. In most cases beam radii defined by the second order intensity moment are larger than those containing 86.5% of the total power [1.70] (see Fig. 2.20). The generalized ABCD law, therefore, is a powerful tool to understand the characteristics of electromagnetic fields and define the beam quality of laser beams [1.64,1.71,1.73,1.86,1.107], but it does only provide limited information the focusability of laser beams in material processing applications.

The Complex Beam Parameter q

As was shown for the Gaussian beam, the mixed moment $\langle w\theta \rangle$ is directly related to the radius of curvature. For arbitrary fields a generalized radius of curvature can be defined:

$$R_g = \frac{\langle w^2 \rangle}{\langle w\theta \rangle} \tag{2.109}$$

By using the generalized q-parameter [1.60]:

$$\frac{1}{q_g} = \frac{1}{R_g} - \frac{i\lambda M^2}{\pi \langle w^2 \rangle}, \tag{2.110}$$

we can rewrite (2.100)-(2.102) into the convenient form of the well-known ABCD law:

$$q_{g,2} = \frac{Aq_{g,1} + B}{Cq_{g,1} + D} \tag{2.111}$$

Unfortunately, this generalized ABCD law cannot be applied to general two-dimensional beams as will be discussed in the next section.

2.6.2 Generalized Astigmatic Beams

In Sec. 2.5.2 the generalized Gaussian beam was discussed, characterized by different ellipses for constant intensity and constant phase. Additionally, these ellipses are rotated with respect to each other and with respect to the reference frame. This most general Gaussian beam is determined by six free parameters. In the case of an arbitrary beam with general astigmatism, we will have more parameters since the divergence is no longer related to the beam waist, but is given by (2.95).

First Order Moments

Two-dimensional beams are characterized by four first order moments which transform like rays as discussed in Sec. 1.2.4. The center of gravity of arbitrary fields propagates through two-dimensional optics on a trajectory that is defined by the ray transformation:

$$\begin{pmatrix} \langle x_2 \rangle \\ \langle y_2 \rangle \\ \langle \theta_{x2} \rangle \\ \langle \theta_{y2} \rangle \end{pmatrix} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix} \begin{pmatrix} \langle x_1 \rangle \\ \langle y_1 \rangle \\ \langle \theta_{x1} \rangle \\ \langle \theta_{y1} \rangle \end{pmatrix} \tag{2.112}$$

With the ray transfer matrix:

$$\boldsymbol{M} = \begin{pmatrix} A_{xx} & A_{xy} & B_{xx} & B_{xy} \\ A_{yx} & A_{yy} & B_{yx} & B_{yy} \\ C_{xx} & C_{xy} & D_{xx} & D_{xy} \\ C_{yx} & C_{yy} & D_{yx} & D_{yy} \end{pmatrix} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix}$$

Second Order Moments

They can be defined in the same way as given by (2.93), (2.94), and (2.96). But now ten second order moments exist which are conveniently arranged in a 4x4 matrix $\boldsymbol{P}$, called the second order moment matrix or the variance matrix [1.71]:

$$\boldsymbol{P} = \begin{pmatrix} \langle w_x^2 \rangle & \langle w_x w_y \rangle & \langle w_x \theta_x \rangle & \langle w_x \theta_y \rangle \\ \langle w_x w_y \rangle & \langle w_y^2 \rangle & \langle w_y \theta_x \rangle & \langle w_y \theta_y \rangle \\ \langle w_x \theta_x \rangle & \langle w_y \theta_x \rangle & \langle \theta_x^2 \rangle & \langle \theta_x \theta_y \rangle \\ \langle w_x \theta_y \rangle & \langle w_y \theta_y \rangle & \langle \theta_x \theta_y \rangle & \langle \theta_y^2 \rangle \end{pmatrix} \tag{2.113}$$

with the newly introduced second moments:

$$\langle w_x w_y \rangle = \frac{4}{P_N} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy |E(x,y)|^2 dx dy \tag{2.114}$$

$$\langle w_x \theta_y \rangle = \frac{-i\lambda}{\pi P_N} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x \left[E(x,y) \frac{\delta E^*(x,y)}{\delta y} - E^*(x,y) \frac{\delta E(x,y)}{\delta y} \right] dx dy \tag{2.115}$$

$$\langle \theta_x \theta_y \rangle = \frac{4}{P_N} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \theta_x \theta_y |E_F(\theta_x, \theta_y)|^2 d\theta_x d\theta_y \tag{2.116}$$

By using the Fourier transform, the last equation can also be written as:

$$\langle \theta_x \theta_y \rangle = \frac{\lambda^2}{\pi^2 P_N} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\delta E(x,y)}{\delta x} \frac{\delta E^*(x,y)}{\delta y} dx dy \tag{2.117}$$

The physical meaning of the mixed second order moments are as follows:

$\langle w_x w_y \rangle$	:	characterizes the rotation of the near field intensity ellipse with respect to the reference frame
$\langle \theta_x \theta_y \rangle$	:	characterizes the rotation of the far field intensity ellipse with respect to the reference frame
$\langle w_x \theta_y \rangle, \langle w_y \theta_x \rangle$	:	orbital angular momentum of the field (twist)
$\langle w_{x,y} \theta_{x,y} \rangle = \langle w^2_{x,y} \rangle / R_{x,y}$	:	characterizes the radius of curvature in x-and y-direction, respectively

In total sixteen second order moments exist which, due to symmetry relations, reduce to a maximum of ten different moments. In the most general case, an arbitrary radiation is thus characterized by ten different parameters. These parameters have to be determined experimentally by intensity measurements. By using the two-dimensional ABCD-propagation law in free space and by measuring $\langle w_x^2(z) \rangle$, $\langle w_y^2(z) \rangle$ and $\langle w_x w_y(z) \rangle$ at a reasonable number of positions z along the focal region of the beam, the following nine parameters can be determined: $\langle w_x^2(0) \rangle$, $\langle w_y^2(0) \rangle$, $\langle w_x w_y(0) \rangle$, $\langle w_x \theta_x(0) \rangle$, $\langle w_y \theta_y(0) \rangle$, $\langle \theta_x^2 \rangle$, $\langle \theta_x \theta_y \rangle$, $\langle \theta_y^2 \rangle$, and the sum $\langle w_x \theta_y(0) \rangle + \langle w_y \theta_x(0) \rangle$. One additional measurement is required to separate the sum. This can be done by using a cylindrical lens. Details are discussed in [1.71,1.125].

Special Beams

The different types of beams discussed in Sec. 2.5.2 appear again in the general case. The most simple one is the beam of circular symmetry, the stigmatic beam, with beam radius w and half angle of divergence θ. Its second order moment matrix reads:

$$\boldsymbol{P_S} = \begin{pmatrix} \langle w^2 \rangle & 0 & \langle w\theta \rangle & 0 \\ 0 & \langle w^2 \rangle & 0 & \langle w\theta \rangle \\ \langle w\theta \rangle & 0 & \langle \theta^2 \rangle & 0 \\ 0 & \langle w\theta \rangle & 0 & \langle \theta^2 \rangle \end{pmatrix} \tag{2.118}$$

The stigmatic beam is characterized by three parameters. This is one more than for the stigmatic Gaussian beam, because the waist radius w_0 and the half angle of divergence are no longer related by (2.62) but by:

$$w_0 \theta = M^2 \frac{\lambda}{\pi}$$

The simple astigmatic beam, which is aligned and can be factorized, is described by the second order matrix:

$$\boldsymbol{P}_{SA} = \begin{pmatrix} \langle w_x^2 \rangle & 0 & \langle w_x \theta_x \rangle & 0 \\ 0 & \langle w_y^2 \rangle & 0 & \langle w_y \theta_y \rangle \\ \langle w_x \theta_x \rangle & 0 & \langle \theta_x^2 \rangle & 0 \\ 0 & \langle w_y \theta_y \rangle & 0 & \langle \theta_y^2 \rangle \end{pmatrix} \tag{2.119}$$

Now six parameters are required to characterize this beam, two more than for a Gaussian beam with simple astigmatism. The beam propagation factor can be different in the x- and the y-direction.

The next, more complicated, beam is the rotated simple astigmatic beam. The corresponding matrix is obtained by applying the rotation matrix $\boldsymbol{R}(\alpha)$ of (1.71) and (1.72). This results in:

$$\boldsymbol{P}_{SA}(\alpha) = \boldsymbol{R}^{-1}(\alpha)\, \boldsymbol{P}_{SA}\, \boldsymbol{R}(\alpha) \tag{2.120}$$

Now all elements of the rotated second order matrix differ from zero, but only seven independent parameters exist. In addition to the six parameters of the simple astigmatic beam, the rotation angle α is a new parameter. In contrast to more complicated beams, the matrix $\boldsymbol{P}_{SA}(\alpha)$ exhibits the following restrictions:

$$\langle w_x \theta_y \rangle = \langle w_y \theta_x \rangle$$

$$\frac{\langle xy \rangle}{\langle w_x^2 \rangle - \langle w_y^2 \rangle} = \frac{\langle w_x \theta_y \rangle}{\langle w_x \theta_x \rangle - \langle w_y \theta_y \rangle} = \frac{\langle \theta_x \theta_y \rangle}{\langle \theta_x^2 \rangle - \langle \theta_y^2 \rangle} = \frac{1}{2}\tan(2\alpha)$$

The beam quality, characterized for simple astigmatic beams by the propagation factors in the x- and the y-direction, now becomes more complicated. As was shown by Nemes [1.71], the generalized beam propagation factor is a constant of propagation:

$$M^4 = (\pi/\lambda)^2 \sqrt{\det \boldsymbol{P}} \geq 1 \tag{2.121}$$

For the simple astigmatic beam, it can be written as the product of the beam propagation factors in the x- and the y-direction:

$$M^4 = M_x^2 M_y^2 \tag{2.122}$$

$$M_x^2 = \frac{k}{2}\sqrt{\langle w_x^2\rangle\langle\theta_x^2\rangle - \langle w_x\theta_x\rangle^2} \geq 1 \tag{2.123}$$

$$M_y^2 = \frac{k}{2}\sqrt{\langle w_y^2\rangle\langle\theta_y^2\rangle - \langle w_y\theta_y\rangle^2} \geq 1 \tag{2.124}$$

A second interesting invariant quantity is the astigmatism factor T given by:

$$T = \frac{\pi^2}{2\lambda^2}\left([\langle w_x^2\rangle\langle\theta_x^2\rangle - \langle w_x\theta_x\rangle^2] + [\langle w_y^2\rangle\langle\theta_y^2\rangle - \langle w_y\theta_y\rangle^2] + \right.$$
$$\left. + 2[\langle w_x w_y\rangle\langle\theta_x\theta_y\rangle - \langle x\theta_y\rangle\langle y\theta_x\rangle]\right) \tag{2.125}$$

The astigmatism factor characterizes the astigmatism of the beam. For stigmatic beams, T is equal to the generalized beam propagation factor:

$$T_S = M^4$$

and for simple astigmatic beams one gets:

$$T_{SA} = \frac{1}{2}\left(M_x^4 + M_y^4\right)$$

Rotated and general astigmatic beams are characterized by the parameters T and M^4.

Beam Propagation

The beam propagation of a general astigmatic beam cannot be described by the generalized ABCD law (2.111) since the generalized q-parameter q_g does not provide all information on the beam properties. The transformation rule of the second order moment matrix for propagation from plane 1 to plane 2 through an optical system described by a 4x4 ray transfer matrix $\boldsymbol{M}$ according to (2.112) can be derived using the Wigner function (see Sec. 2.7) [1.109,1.110]:

$$\boldsymbol{P}_2 = \boldsymbol{M}\boldsymbol{P}_1\boldsymbol{M}^T \tag{2.126}$$

with $\boldsymbol{M}^T$ being the transpose of $\boldsymbol{M}$. This equation can be written in a form similar to the linear case given by (2.100)-(2.102). The first two of the ten equations read:

$$\langle w_{x2}^2\rangle = \langle(A_{xx}w_{x1} + A_{xy}w_{y1} + B_{xx}\theta_{x1} + B_{xy}\theta_{y1})^2\rangle$$

$$\langle w_{x2} w_{y2} \rangle = \langle (A_{xx} w_{x1} + A_{xy} w_{y1} + B_{xx} \theta_{x1} + B_{xy} \theta_{y1}) \cdot$$

$$(A_{yx} w_{x1} + A_{yy} w_{y1} + B_{yx} \theta_{x1} + B_{yy} \theta_{y1}) \rangle$$

In the most general case, each parameter of the beam depends on all ten parameters of the incident beam. The equation arises now as to whether a general astigmatic beam can be transformed into a stigmatic one by using suitable ABCD optics. This is a delicate problem which is beyond the scope of this book [1.67]. Under certain conditions this transformation is possible, but not in general.

2.6.3 Beam Quality

We have seen that the propagation of partially coherent fields can be characterized by the second order intensity moments. If the intensity $I(x,y)$ in the beam waist is known, the beam radii at any distance z from the waist as well as the angle of divergence in the far field can be calculated using the generalized ABCD law (2.103/105). Laser beams can, in general, be described by partially coherent fields. Therefore, we can apply the generalized propagation laws to investigate the focusing properties of laser beams and define a measure for laser beam quality.

As shown in Fig. 2.21, the focusing of a laser beam with beam parameter product $w\theta$ is equivalent to generating a beam waist with a small waist radius w_0. Note that we define beam radii of the laser beam by the squareroot of the second intensity moments, according to (2.93) or (2.97). At a distance z from the plane of this focal spot, the beam radius $w(z)$ is given by the generalized ABCD law:

$$\langle w^2(z) \rangle = \langle w_0^2 \rangle \left[1 + \frac{z^2}{\langle z_0^2 \rangle} \right] \tag{2.127}$$

At the distance of one Rayleigh range z_0 from the focal spot, the cross sectional area has increased by a factor of 2. The Rayleigh range and the beam waist radius are related to the divergence angle via:

$$\langle \theta_0^2 \rangle = \frac{\langle w_0^2 \rangle}{\langle z_0^2 \rangle} \tag{2.128}$$

We are interested in a small focal spot size and a large Rayleigh range. This leads us to the following ratio for the definition of laser beam quality:

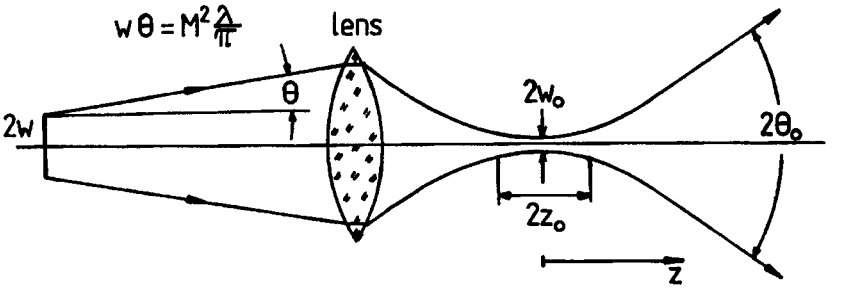


Fig. 2.21 Focusing of a laser beam with propagation factor M^2 by means of lens. Good beam quality is characterized by a small beam cross sectional area in the focal plane and a large Rayleigh range z_0.

$$\frac{w_0^2}{z_0} = w_0\theta_0 = w\theta = M^2\frac{\lambda}{\pi} \tag{2.129}$$

A measure for the beam quality is a low beam parameter product $w\theta$ which is equivalent to a low beam propagation factor M^2. The knowledge of the beam propagation factor is sufficient to characterize the beam quality of the laser. Once M^2 is measured for the laser beam (see Part VI for experimental techniques), the beam quality is well defined no matter what kind of optics are used for beam transformation and focusing (M^2 stays constant in ABCD-optical systems!).

Size and position of the beam waist w_0 can be calculated by using the Gaussian beam imaging conditions (2.69) and (2.70). Both equations also apply to partially coherent fields if the beam radii and the divergence angles are defined by the second intensity moments. To a good approximation, w_0 is determined by the divergence angle θ of the beam in front of the lens and the focal length f:

$$w_0 \approx f\,\theta \tag{2.130}$$

Since we have defined all beam dimensions via the second intensity moments, Eqs. (2.103)-(2.106) exactly hold for all laser beams. Definition and measurement of laser beam quality was recently standardized in ISO-procedures using this second order intensity moment approach [1.86]. Unfortunately, the second moments do not provide consistent information in regards to the power contained within the beam diameter. For Gaussian beams, the second order intensity moment provides a beam diameter that is equal to the Gaussian beam diameter with a power content of 86.5%. For other beams, the power contained in the second order intensity diameter is in general higher. Therefore, for some beams, the beam

quality definition via the second order intensity moments may not reflect the focusing properties in materials processing applications. This is especially true for unstable resonators, whose beams exhibit side lobes in the far field (Fig. 2.22). If a hard edged output coupler is used, the second intensity moment in the far field is not defined (integral is infinite).

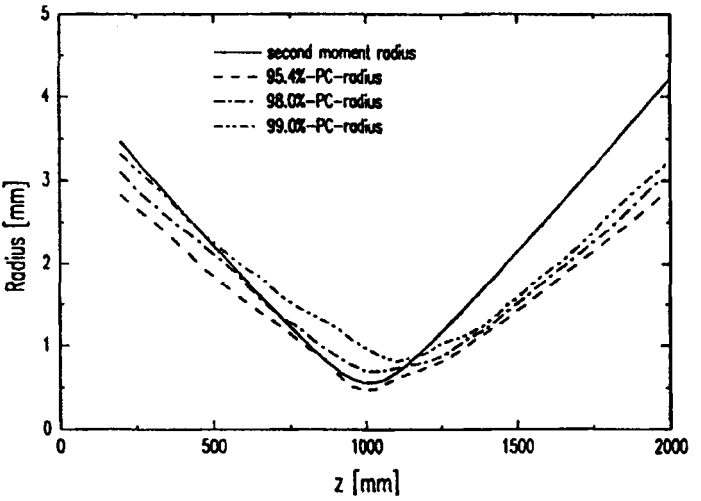


Fig. 2.22 Calculated beam radii of a focused unstable resonator laser beam The position of the beam waist is changed and the beam caustic becomes non-hyperbolic if the beam radii are defined by power contents (PC-radii) [1.70] (© SPIE 1994).

2.7 Coherence

A common definition of coherence is: "Coherence is a fixed relationship between the phase of waves in a beam of single-frequency radiation. Two waves of light are coherent, when the phase difference between their waves is constant; they are non-coherent if there is a random or changing phase relationship. Stable interference pattern are formed only by radiation emitted by coherent (or partially coherent) light". A monochromatic beam with a beam propagation factor of $M^2=1$ is a TEM_{00}-Gaussian beam and always coherent. But a beam with $M^2>1$ can be a coherent higher order mode, or a superposition of several modes with different frequencies. A coherent mode with $M^2>1$ can, in principle, be converted into a fundamental mode by a suitable phase plate, as was shown theoretically and experimentally [1.75,1.91,1.101], whereas a superposition of modes cannot. Therefore, the characterization of a beam only by M^2 is not sufficient. At least one additional number, the global degree of coherence K is required.

Coherence is discussed extensively in the well known book by Born and Wolf [1.2], starting from first principles, but neglecting the photon character of light. A more fundamental approach, taking into account the quantized field, can be found in [1.74]. In the following section a simplified version, adapted to the requirements of laser radiation, will be presented. For an easier understanding, temporal coherence will be briefly summarized.

2.7.1 Temporal Coherence

Normal light sources emit a statistical sequence of light bursts of duration τ with no fixed phase relation between each other. This is also true for high power lasers, with the only difference that the duration τ is much larger. Even the emission of stabilised single frequency-lasers is not a precise sinusoidal wave since the phase fluctuates statistically with a time constant in the range of milliseconds to seconds. Coherence characteristics of some lasers are shown in Tab. 2.4.

Table 2.4 Temporal coherence parameters of some light sources. All numbers depend on the special mode of operation and are rough estimates. ($\Delta t_{coh}=\ell_{coh}/c$, $\Delta t_{coh}\,\Delta\omega=0.5$).

Source	temporal coherence length ℓ_{coh} [m]	coherence time Δt_{coh} [s]	bandwidth $\Delta\omega$ [Hz]
torch light	5×10^{-7}	1.5×10^{-15}	3.34×10^{14}
solid-state laser	10^{-2}	3×10^{-11}	1.67×10^{12}
CO_2 laser	1	3×10^{-9}	1.67×10^{8}
He-Ne laser	10	3×10^{-8}	1.67×10^{7}
He-Ne laser (stabilised)	1.5×10^{8}	0.5	1

A wave train of limited duration τ can be described by:

$$E(\boldsymbol{r},t) = A(\boldsymbol{r})\, f(t/\tau)\, \exp[i\omega t - ikz] \tag{2.131}$$

The time τ is a measure for the duration of the wave. If such a field is superimposed with the same field delayed by Δt, an interference pattern will appear only if the time delay is in the range of τ or smaller. An example for time-limited radiation is the spontaneous emission of atoms. The amplitude decays exponentially in time:

$$f(t) = \begin{bmatrix} 0 & for\ t<t_0 \\ \exp[-(t-t_0)/\tau + i\phi_n] & for\ t \geq t_0 \end{bmatrix}$$

where amplitude A, phase ϕ_n and time of emission τ are random. A gas discharge of high temperature emits Gaussian-like wave-trains, due to Doppler-broadening:

$$f(t) = \exp\left[-\left(\frac{t-t_n}{\tau}\right)^2 + i\phi_n\right]$$

The emission of a cw-laser is nearly stable in amplitude, but fluctuates in phase:

$$f(t) = \exp[i\phi_n(t/\tau)]$$

where the time constant τ is roughly the inverse bandwidth. For a more quantitative description of temporal coherence, an experimental set-up, the Michelson-interferometer has to be discussed.

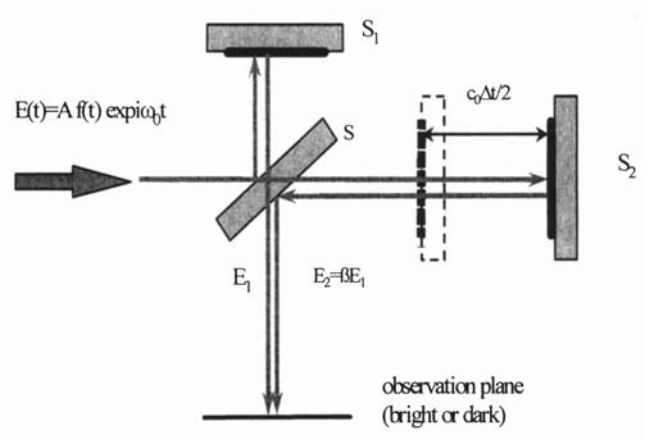


Fig. 2.23 The Michelson-interferometer to determine temporal coherence.

The Michelson Interferometer

In Fig. 2.23 a schematic Michelson-interferometer is shown. The incoming beam is split by the mirror S into two beams. Each beam is reflected by mirrors S_1, S_2, respectively, and both are recombined in the observation plane. By moving mirror S_2, a time delay Δt between the two beams can be generated. Using expression (2.131), the resulting field in the observation plane reads with (2.131):

$$E_{obs}(t) = E_1(t) + E_2(t) = A_1 f(t)\exp[i\omega_0 t] + A_2 f(t+\Delta t)\exp[i\omega_0(t+\Delta t)] \qquad (2.132)$$

The intensity is recorded with a detector (CCD-camera, film), which squares and integrates $E_{obs}(t)$. The resulting signal I is:

$$I = \frac{1}{2}c_0\epsilon_0\frac{1}{T}\int_0^T EE^* dt = \frac{1}{2}c_0\epsilon_0 \langle EE^* \rangle \qquad (2.133)$$

with T being the observation time. It is assumed that T is large compared to the time constant τ of the fluctuations. The bracket denotes the time average. Insertion of (2.132) into (2.133) results in:

$$I = \frac{1}{2}c_0\epsilon_0 \left[|A_1|^2\langle f(t)f^*(t)\rangle + |A_2|^2\langle f(t+\Delta t)f^*(t+\Delta t)\rangle + A_1A_2^*\exp[i\omega_0\Delta t]\langle f(t)f^*(t+\Delta t)\rangle + cc.\right]$$

$$\Leftrightarrow \qquad I = I_1 + I_2 + \Gamma_{11} + \Gamma_{11}^* \qquad (2.134)$$

I_1, I_2 are the intensities of the individual wave trains and Γ_{11} is called the auto-correlation function, because the field is correlated with itself. The normalised value of Γ_{11} is the complex degree of coherence $\gamma_{11}(\Delta t)$:

$$\gamma_{11} = \frac{\Gamma_{11}(\Delta t)}{\sqrt{I_1 I_2}} = \frac{A_1A_2^*\exp[-i\omega_0\Delta t]\langle f(t)f^*(t+\Delta t)\rangle}{\sqrt{I_1 I_2}} \qquad (2.135)$$

Using Equation (2.135), (2.134) can be written as:

$$I = I_1 + I_2 + 2|\gamma_{11}|\sqrt{I_1 I_2}\cos(\omega_0\Delta t + \phi) \qquad (2.136)$$

The intensity is modulated with the delay-time Δt. The intensity contrast is referred to as visibility V which is derived from (2.136):

$$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} = 2|\gamma_{11}|\frac{\sqrt{I_1 I_2}}{I_1 + I_2} \tag{2.137}$$

For a coherent beam the complex degree of coherence will be modulated, but the modulus is constant and equal one. In all other cases $|\gamma_{11}|$ will decrease with increasing delay Δt. An example is shown in Fig. 2.24.

Coherence Time

The parameter τ in (2.131) is a measure for the coherence time. If the delay Δt is large compared with τ, the contrast will be low and vice versa. It is convenient to define the coherence time by an average value, weighted with the auto-correlation function. By convention, the following definition is used for the coherence time [1.2]:

$$\Delta t_{coh}^2 = \frac{\int \Delta t^2 |\Gamma_{11}(\Delta t)|^2 d(\Delta t)}{\int |\Gamma_{11}(\Delta t)|^2 d(\Delta t)} \tag{2.138}$$

Some examples are given in Tab.2.5. Note that the coherence time is not equal the pulse duration, as will be shown in the next section.

Table 2.5 Characteristic parameters of different pulses

amplitude $E(t)/A_0$	spectrum $E(\omega)/A_0\tau$	coherence function $\Gamma(\Delta t)/\lvert A_0\rvert^2\tau$	coherence length $\Delta t_{coh}/\tau$	time-bandwidth product $\Delta t_{coh}\Delta\omega$
rectangular pulse				
$\begin{cases}\exp(i\omega_0 t) & for\ \lvert t\rvert<\tau \\ 0 & for\ \lvert t\rvert\geq\tau\end{cases}$	$\sqrt{\frac{2}{\pi}}\frac{\sin[(\omega-\omega_0)\tau]}{(\omega-\omega_0)\tau}$	$\left[2-\frac{\lvert\Delta t\rvert}{\tau}\right]\exp(i\omega_0\Delta t)$ $\lvert\Delta t\rvert \leq 2\tau$	$\sqrt{0.4}$	0.91
exponential decay				
$\begin{cases}\exp(i\omega_0 t - t/\tau) & for\ t>0 \\ 0 & for\ t\leq 0\end{cases}$	$\frac{1/\sqrt{2\pi}}{1-i(\omega-\omega_0)\tau}$	$\frac{1}{2}\exp(i\omega_0\Delta t - \lvert\Delta t\rvert/\tau)$	$\frac{1}{\sqrt{2}}$	0.71
Gaussian pulse				
$\exp\left(i\omega_0 t - (t/\tau)^2\right)$	$\exp-[(\omega-\omega_0)\tau/2]^2$	$\sqrt{\frac{\pi}{2}}\exp\left[i\omega_0\Delta t - \frac{1}{2}(\Delta t/\tau)^2\right]$	$\frac{1}{\sqrt{2}}$	0.5

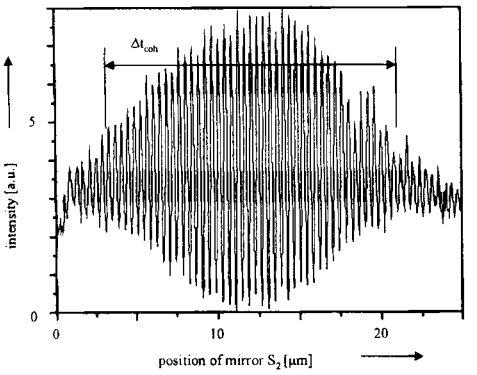


Fig. 2.24 Measurement of the temporal coherence of a beam generated by a short-pulse Ti-Sapphire-laser with $\Delta t_{coh} = 5.8 \times 10^{-14}$ s [1.85].

Spectrum

The amplitude spectrum $E(\omega)$ and the power spectrum $G(\omega)$ of the field are given by the Fourier transform of the field:

$$E(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} E(t)\exp[i\omega t]dt \qquad G(\omega) = E(\omega)E^*(\omega) \tag{2.139}$$

A very useful relation between the auto-correlation function and the power spectrum is the Wiener-Khinchin theorem:

$$G(\omega) = \int \Gamma_{11}(\Delta t)\exp[i\omega\Delta t]d(\Delta t) \tag{2.140}$$

which is easy to prove [1.2]. The Fourier spectrum of the auto-correlation function is equal to the power spectrum. With the Wiener-Khinchin theorem, it is possible to determine the power spectrum of a light source using a Michelson-interferometer.

Spectral width

The power spectrum can be used to define the center frequency $\hat{\omega}$ and the spectra width $\Delta\omega$ of the field by its first order and second order moments:

$$\hat{\omega} = \int_{-\infty}^{+\infty} \omega G^2(\omega)d\omega \Big/ \int_{-\infty}^{+\infty} G^2(\omega)d\omega \tag{2.141}$$

$$\Delta\omega^2 = \int_{-\infty}^{+\infty}(\omega-\hat{\omega})^2 G^2(\omega)d\omega \Big/ \int_{-\infty}^{+\infty} G^2(\omega)d\omega \tag{2.142}$$

Fourier Relation

Spectral width $\Delta\omega$ and coherence time Δt_{coh} are related by the Fourier relation:

$$\Delta t_{coh}\ \Delta\omega \geq \frac{1}{2} \tag{2.143}$$

This relation holds for arbitrary fields, if the two parameters are defined according to (2.138) and (2.142) and the two weighting function Γ, G are related by the Fourier transform. The equality in (2.143) holds for Gaussian shaped fields [1.2].

Pulse width and Chirp

The width of a single pulse is defined by its second intensity moment:

$$\Delta t_p^2 = 4\int_{-\infty}^{+\infty} t^2 I^2(t)dt \Big/ \int_{-\infty}^{+\infty} I^2(t)dt \tag{2.144}$$

The factor 4 is arbitrary, but guarantees that the pulse width is larger than the coherence length. In general, the pulse width will differ from the coherence length. A well known example is the chirped Gaussian pulse, given by:

$$E = A\ \exp[i\omega_0 t - \delta t^2] \qquad\qquad \delta = a - ib$$

where δ is a complex constant. This special pulse is produced by non-linear interaction of high intensity pulses with Kerr-active media (glass, liquids). The instantaneous frequency is the time-derivative of the phase and increases (in this case) linearly with time. The constant b is called the chirp:

$$\omega_{inst} = \frac{d}{dt}(\omega_0 t + bt^2) = \omega_0 + 2bt \tag{2.145}$$

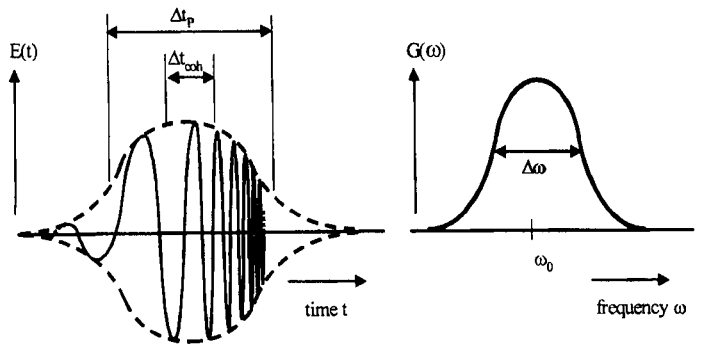


Fig. 2.25 Field *E(t)* and power spectrum *G(ω)* of a chirped Gaussian pulse. The oscillation in the left figure is not to scale.

amplitude spectrum

$$E(\omega) = \frac{A}{\delta\sqrt{2}} \exp\left[-\frac{(\omega-\omega_0)^2}{4\delta}\right]$$

power spectrum

$$G(\omega) = \frac{|A|^2}{2(a^2+b^2)} \exp\left[-\frac{(\omega-\omega_0)^2}{2}\,\frac{a}{a^2+b^2}\right]$$

center frequency

$$\hat{\omega} = \omega_0$$

spectral width

$$\Delta\omega = \sqrt{\frac{a^2+b^2}{2a}}$$

coherence time

$$\Delta t_{coh} = \sqrt{\frac{a}{2(a^2+b^2)}}$$

pulse duration

$$\Delta t_p = \frac{1}{\sqrt{2a}}$$

The following relations hold for a chirped Gaussian pulse:

$$\Delta t_{coh} = \frac{\Delta t_p}{\sqrt{1+b^2/a^2}} \le \Delta t_p \qquad \Delta t_{coh}\,\Delta\omega = \frac{1}{2}$$

It is important to know that the Michelson-interferometer can measure the spectrum or the temporal coherence length, but not the pulse width. To determine the pulse width correctly, non-linear methods have to be used such as frequency conversion or two-photon fluorescence [1.90].

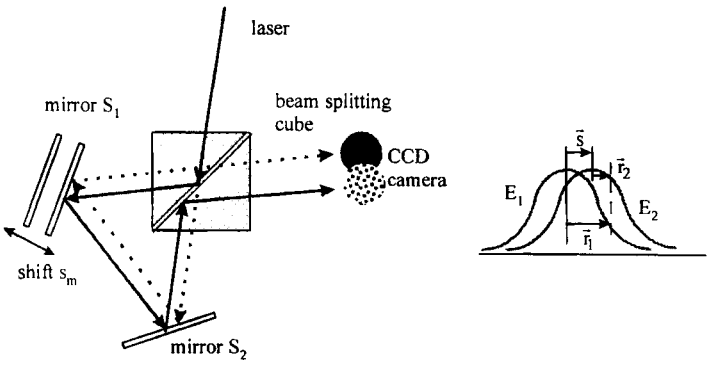


Fig. 2.26 The shear interferometer to determine spatial coherence [1.87].

2.7.2 Spatial Coherence

Spatial coherence is of interest for many applications such as holography, lithography, or laser etching of gratings. Whether high or low spatial coherence is required depends on the application. Whereas temporal coherence is determined via a temporal shift of the field, transverse coherence is measured by a transverse shift between the two fields.

The Shear Interferometer

It was originally used to measure lens aberrations [1.22,1.33]. A schematic set-up is shown in Fig. 2.26. The beam under investigation is divided by a beam splitter into two beams, which are recombined by the mirrors S_1, S_2 in the observation plane. By shifting mirror S_2, a transverse shift $\vec{s}$ of the two beams is produced. The resulting field reads.

$$E = E_1(\boldsymbol{r}_1,t) + E_2(\boldsymbol{r}_2,t)$$

For quasi-monochromatic fields the time difference Δt between the two beams can be neglected. The intensity is the time average, squared value. After some manipulation, the same relation as in the case of temporal coherence (2.134) is obtained:

$$I(\boldsymbol{r}_1,\boldsymbol{r}_2) = I_1(\boldsymbol{r}_1) + I_2(\boldsymbol{r}_2) + \Gamma_{12}(\boldsymbol{r}_1,\boldsymbol{r}_2) + \Gamma_{12}^*(\boldsymbol{r}_1,\boldsymbol{r}_2)$$

where Γ_{12} now denotes the cross-correlation function, which in general is a four-dimensional function.

$$\Gamma_{12}(\boldsymbol{r}_1,\boldsymbol{r}_2) = \frac{1}{2}c_0\epsilon_0\langle E_1(\boldsymbol{r}_1)E_2^*(\boldsymbol{r}_2)\rangle \qquad (2.146)$$

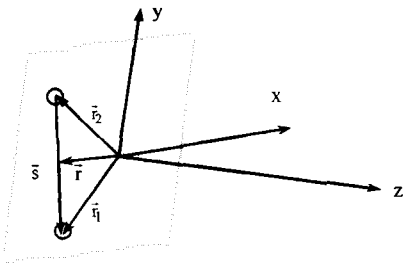


Fig. 2.27 The vectors r,s identifying the two points of the field, which interfere.

Often the variables $\boldsymbol{r},\boldsymbol{s}$ are used instead of $\boldsymbol{r}_1$, $\boldsymbol{r}_2$ (Fig. 2.27). These are defined as:

$$\boldsymbol{r} = (\boldsymbol{r}_1 + \boldsymbol{r}_2)/2 \qquad \boldsymbol{s} = \boldsymbol{r}_1 - \boldsymbol{r}_2 \tag{2.147}$$

The bracket again represents the time average. It can be expanded into a series of coherent functions, which is called modal-expansion [1.42,1.45,1.93,1.97]. For $s=0$ and $\boldsymbol{r}_1=\boldsymbol{r}_2$, the intensity and the power are directly related to Γ_{12}:

$$I(\boldsymbol{r}) = \Gamma_{12}(\boldsymbol{r}, s=0) \tag{2.148}$$

$$P = \int\int \Gamma_{12}(\boldsymbol{r}, s=0)\, d\boldsymbol{r} \qquad \text{with} \quad d\boldsymbol{r} = dxdy \tag{2.149}$$

The complex degree of coherence is the normalized cross-correlation function:

$$\gamma_{12}(\boldsymbol{r}_1,\boldsymbol{r}_2) = \frac{\Gamma_{12}(\boldsymbol{r}_1,\boldsymbol{r}_2)}{\sqrt{I_1(\boldsymbol{r}_1)I_2(\boldsymbol{r}_2)}} = |\gamma_{12}(\boldsymbol{r}_1,\boldsymbol{r}_2)|\exp[i\Phi_{12}(\boldsymbol{r}_1,\boldsymbol{r}_2)] \tag{2.150}$$

where $\Phi_{12}(\boldsymbol{r}_1,\boldsymbol{r}_2)$ represents the mutual phase of the two fields (phase difference) in the two points $\boldsymbol{r}_1,\boldsymbol{r}_2$. The intensity can now be written as:

$$I = I_1 + I_2 + 2|\gamma_{12}|\sqrt{I_1 I_2}\cos[\Phi_{12}] \tag{2.151}$$

which has a from analogous to the temporal coherence case (2.136). If the field is completely coherent the modulus of γ_{12} is equal to one everywhere. The modulation of the intensity is given by the mutual phase Φ_{12}. If the two fields in the observation plane have a plane wave front, the phase Φ_{12} is constant. For non-planar waves the phase difference depends on the shape of the wave front. An example is given in Fig. 2.28, which shows the shear interferogram of a fundamental mode (Gaussian beam).

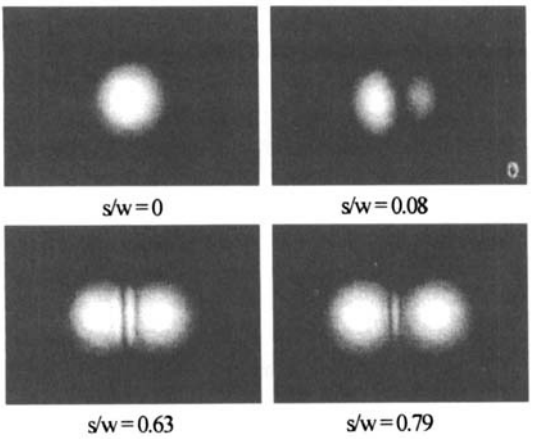


Fig. 2.28 Shear interferogram of a fundamental mode with M^2=1.2 of a He/Ne-laser [1.87]. s/w: shear normalized to the beam radius.

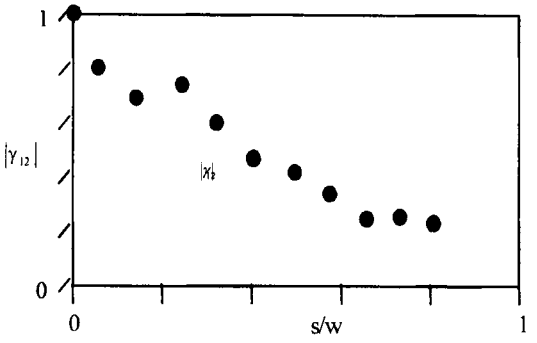


Fig. 2.29 $|\gamma_{12}|$ vs the normalized shear s for a He/Ne-laser with M^2 =1.2 [1.87]. w: beam radius.

Outside the waist the mode exhibits a spherical wave front, and consequently interference patterns appear. From the intensity pattern the wave front can be reconstructed, using (2.147) and (2.148). If the beam is not completely coherent ($|\gamma_{12}|<1$), two measurements are necessary to determine $|\gamma_{12}|$ and Φ_{12}. This can be done by inserting a quarter wave plate in one beam and performing a second measurement. Equation (2.151) then reads:

$$I = I_1 + I_2 + 2|\gamma_{12}|\sqrt{I_1 I_2}\sin[\Phi_{12}] \tag{2.152}$$

The result of such a measurement is shown in Fig. 2.29. The degree of coherence $|\gamma_{12}|$ is plotted versus the shear for a beam with $M^2 = 1.2$. To obtain reliable results, the optical elements must be very homogeneous and free of stress induced birefringence. The shear-interferometer has the advantage that for each value s, the CCD-camera delivers the complete set of r parameters. A useful device to measure the two-dimensional spatial coherence is shown in Fig. 2.30, where a rotating Dove-prism is used to scan the full transverse plane. Some experimental results are compiled in Fig. 2.31.

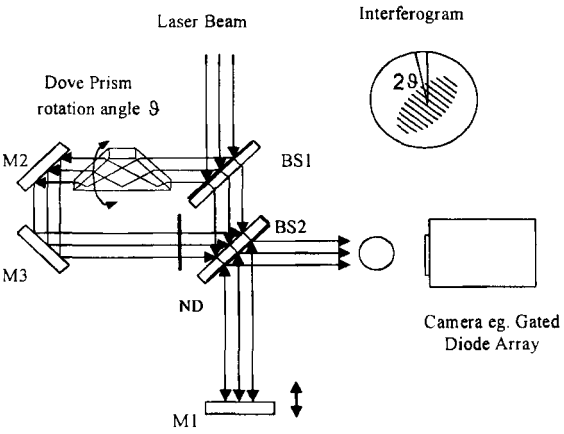


Fig. 2.30 A shearing interferometer with a rotating Dove prism [1.79] (© OSA 1998).

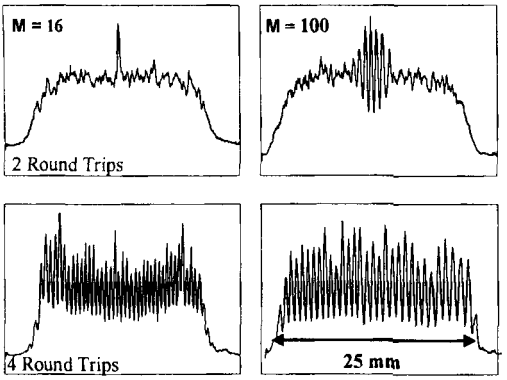


Fig. 2.31 Shear interferogram of a copper-vapour laser beam. Parameters are the magnification M of the unstable resonator and the number of round trips [1.79] (© OSA 1998).

Young's Double-slit Interferometer

One of the most famous experiments in quantum optics is Young's double slit experiment, because it proved the wave theory of light. The set-up is in principle very simple (Fig. 2.32). A light source is illuminating a double slit, or two pin-holes. From the two pin holes separated by the distance s two spherical fields E_1, E_2 emanate and are collimated with a lens. An interference pattern is generated in the observation plane, which is magnified and recorded by a CCD-camera. The two pinholes are realized by two masks with cross slits as shown in Fig. 2.33. The cross-correlation function for this special example is also given in Fig.2.33. Using the cross-correlation function all beam parameters such as phase front, mode-content, beam width, beam divergence, global degree of coherence and transverse coherence length can be evaluated. In an alternate embodiment, the two pin holes can be replaced by two single mode fibers [1.88].

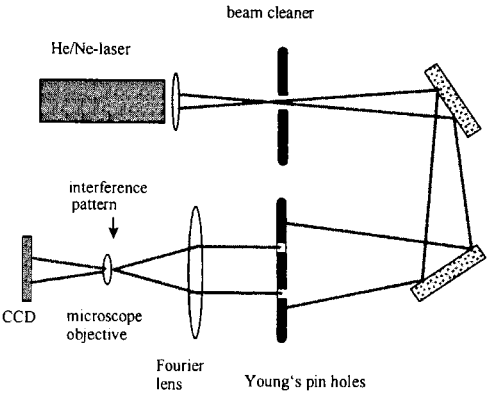


Fig. 2.32 Young's interferometer to determine the cross-correlation function of a multimode He/Ne-laser [1.55] (© Springer-Verlag 1989).

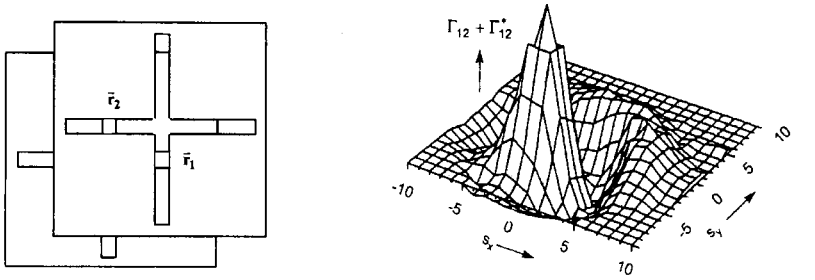


Fig. 2.33 The two pin holes r_1, r_2 of Young's interferometer and a measured cross correlation function [1.55] (© Springer-Verlag 1989).

The determination of the beam parameters by Young's interferometer is very time consuming, because for each value r the parameter s has to be scanned point by point over the complete two-dimensional transverse plane.

Some useful definitions:

Visibility

The visibility or contrast is defined in the same way as for temporal coherence:

$$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} = 2|\gamma_{12}|\frac{\sqrt{I_1 I_2}}{I_1 + I_2} \tag{2.153}$$

The decrease of V with increasing shift can be caused by poor coherence or by different intensities of the two beams.

Global Degree of Coherence

The modulus of the cross-correlation function will decay with increasing shear and approaches asymptotically zero. To eliminate the influence of the intensity decay, the cross-correlation function has to be normalised. A convenient definition of the global degree of coherence is [1.102,1.48]:

$$K^2 = \frac{\iiiint |\Gamma_{12}(\boldsymbol{r}_1,\boldsymbol{r}_2)|^2 d\boldsymbol{r}_1 d\boldsymbol{r}_2}{\iint \Gamma_{12}(\boldsymbol{r}_1,\boldsymbol{r}_1)d\boldsymbol{r}_1 \iint \Gamma_{12}(\boldsymbol{r}_2,\boldsymbol{r}_2)d\boldsymbol{r}_2} = \frac{1}{P^2}\iiiint |\Gamma_{12}(\boldsymbol{r}_1,\boldsymbol{r}_2)|^2 d\boldsymbol{r}_1 d\boldsymbol{r}_2 \qquad (2.154)$$

If the field is completely coherent with:

$$E_{1,2} = A_{1,2}(\boldsymbol{r}_{1,2}) \exp[i(\omega t - kz)]$$

the cross-correlation function becomes

$$\Gamma_{12} = 0.5\, c_0 \epsilon_0 A_1 A_2^*$$

and (2.154) results in $K^2=1$. In all other cases, K^2 will be smaller than unity. A relation between the global coherence and the beam propagation factor exists [1.90,1.48]:

$$\sqrt{K^2}\, M^2 \geq 8/9 \qquad (2.155)$$

This relation is useful to check on experimental results. If the cross-correlation function can be factorized, e.g. in the case of rectangular symmetry:

$$\Gamma_{12}(\boldsymbol{r}_1,\boldsymbol{r}_2) = \Gamma_x(x_1,x_2)\Gamma_y(y_1,y_2) \quad or \quad \Gamma_{12}(\boldsymbol{r}_1,\boldsymbol{r}_2) = \Gamma_x(x,s_x)\Gamma_y(y,s_y)$$

the global degree of coherence can be simplified to yield (with $P^2 = P_x^2 P_y^2$):

$$K^2 = \frac{1}{P_x^2}\iint |\Gamma_x(x_1,x_2)|^2 dx_1 dx_2 \; \frac{1}{P_y^2}\iint |\Gamma_y(y_1,y_2)|^2 dy_1 dy_2 = K_x^2 K_y^2 \qquad (2.156)$$

Transverse Coherence Length

For a fixed position r_1 the degree of coherence will decrease with increasing s, but it can be modulated, depending on the beam structure. Therefore, the local coherence length is defined as an average value, weighted with the modulus of the cross-correlation function:

$$\ell_{loc}^2(r_1) = 8\,\frac{\iint (r_1-r_2)^2|\Gamma_{12}(r_1,r_2)|^2 dr_2}{\iint |\Gamma_{12}(r_1,r_2)|^2 dr_2}$$

Using (2.147), the global coherence length now reads:

$$\ell_{glob}^2 = 8\frac{\iiint\iint (r_1-r_2)^2|\Gamma_{12}(r_1,r_2)|^2 dr_1 dr_2}{\iiint\iint |\Gamma_{12}(r_1,r_2)|^2 dr_1 dr_2} = \frac{8}{K^2P^2}\iiint\iint s^2|\Gamma_{12}(r,s)|^2 ds dr \qquad (2.157)$$

Similar to the temporal case, the factor 8 is arbitrary, but guarantees that the coherence length is always smaller or equal the beam diameter, defined by (2.93). It is important to notice that the coherence length is a measure for the decay of the contrast, which can be caused by different intensities or by poor coherence. Even a coherent beam has a limited coherence length given by the beam diameter. In the case of rectangular symmetry the transverse degree of coherence reads:

$$\ell_{glob}^2 = \ell_{glob,x}^2 + \ell_{glob,y}^2 \qquad (2.158)$$

with:

$$\ell_{glob,x}^2 = \frac{8}{K_x^2P_x^2}\iint s_x^2|\Gamma_x(x,s_x)|^2 dx ds_x\,, \quad \ell_{glob,y}^2 = \frac{8}{K_y^2P_y^2}\iint s_y^2|\Gamma_y(y,s_y)|^2 dy ds_y \qquad (2.159)$$

The far field angle of coherence is defined by

$$\Theta_{coh,x,y}^2 = \lim_{z\to\infty}\left[\frac{\ell_{glob,x,y}}{z}\right]^2 \qquad (2.160)$$

Examples

To demonstrate the practical aspects of these definitions, let us discuss some examples [1.102], without going into the mathematical details. In the cases below, beam radius w_{Rx} and coherence length $\ell_{glob,x}$ are normalized to the fundamental mode radius w_{0x}, which propagates according to (2.107).

Pure TEM_{m0}

In rectangular symmetry, the field is given by (5.8) and the cross-correlation function reads:

$$\Gamma_x(x_1,x_2) = \frac{1}{2}c_0\epsilon_0|E_0|^2 \exp\left[-\frac{x_1^2+x_2^2}{w_x^2}\right] H_0^2 \; H_m\left(\frac{\sqrt{2}x_1}{w_x}\right) \; H_m\left(\frac{\sqrt{2}x_2}{w_y}\right)$$

Beam radius and divergence are defined by the second moments (2.93, 2.94, 5.12):

$$w_{Rx}(z) = w_{0x}(z)\sqrt{2m+1} = w_{0x}(0)\sqrt{1+z^2/z_0^2}/\sqrt{2m+1}$$

$$\Theta_{Rx} = \Theta_{0x}\sqrt{2m+1}$$

The beam radius scales according to (2.107). The coherence parameters result from the above equations:

beam propagation factor: $M_x^2 = 2m+1$

degree of coherence: $|\gamma_{12}| = 1$

global degree of coherence: $K_x = 1$

coherence length: $\ell_{glob.x}(z) = 2w_{0x}(z)\sqrt{2m+1}$

The coherence length is the same as the beam diameter and scales hyberbolically with z.

b) Two oscillating modes TEM_{00} +TEM_{01}

The fundamental mode and the next higher order mode may oscillate in a rectangular one-dimensional resonator. Then (5.8) holds and the resulting field reads:

$$E = E_0 \exp\left[-\frac{x^2}{w_{0x}^2}\right]\left[b_0 \exp(i\omega_{00}t) + b_1 \exp(i\omega_{10}t) \; H_1\left(\frac{\sqrt{2}x}{w_{0x}}\right)\right]$$

In non-degenerated resonators the eigenfrequencies ω_{00}, ω_{10} are different and the cross-terms vanish when averaged over time. The cross-correlation function reduces to:

$$\Gamma_x(x_1,x_2) = \frac{1}{2}c_0\epsilon_0|E_0|^2 \exp\left[-\frac{x_1^2+x_2^2}{w_{0x}^2}\right]\left(a_0 + a_1 \; H_1\left(\frac{\sqrt{2}x_1}{w_{0x}}\right) \; H_1\left(\frac{\sqrt{2}x_2}{w_{0x}}\right)\right)$$

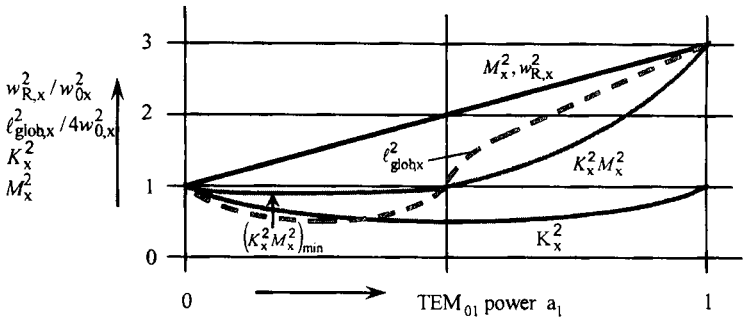


Fig. 2.34 Coherence parameters of the two-mode oscillation vs the power a_1 of the TEM_{10}-mode.

Equation (2.149) yields a condition for the mode expansion coefficients:

$$a_0 + a_1 = |b_0|^2 + |b_1|^2 = 1 \qquad 0 \le a_0, a_1 \le 1$$

The beam parameters can be evaluated analytically and depend on the fractional powers a_0, a_1 of the modes:

beam radius: $$w_{Rx}(z) = w_{0x}(z)\sqrt{1+2a_1}$$

beam propagation factor: $$M_x^2 = 1+2a_1$$

degree of coherence: $$|\gamma_{12,x}| = \frac{|a_0+4a_1x_1x_2/w_{0x}^2|}{\sqrt{a_0+4a_1x_1^2}\sqrt{a_0+4a_1x_2^2}}$$

global coherence: $$K_x^2 = a_0^2+a_1^2$$

coherence length: $$\ell_{glob,x}(z) = 2w_{0x}(z)\sqrt{1+2a_1(a_1-a_0)/K_x^2}$$

It is interesting to note that the coherence length has a minimum for $a_1=1/3$ (Fig. 2.34).

The Gauss-Schell beam [1.47,1.68,1.74]

This is the most general partially coherent beam with respect to the second order moments and a Gaussian intensity distribution. The general Gauss-Schell-beam, which is often used to describe multimode lasers, is characterized by ten independent parameters [1.68]. We will

restrict the discussion to the one-dimensional beam with its two parameters w_{Rx}, σ, which describe beam radius and degree of coherence, respectively. The Gauss-Schell beam is only defined by its cross-correlation function. It reads in the one-dimensional case:

$$\Gamma_x(x,s_x,z) = \frac{1}{2}c_0\epsilon_0|E_0|^2 \exp\left[-\frac{2x^2}{w_{Rx}^2}-\frac{s_x^2}{2w_{Rx}^2\sigma^2}-\frac{2ixs_x}{w_{Rx}^2\sigma}\frac{z}{z_{Rx}}\right] \tag{2.161}$$

$$z_{Rx} = \sigma z_0 \quad : \quad \textit{Rayleigh range}$$

The beam intensity is given by (2.148) and results from the above equation for s_x=0:

$$I(x,z) = \frac{1}{2}c_0\epsilon_0|E_0|^2 \exp\left[-\frac{2x^2}{w_{Rx}^2}\right]$$

with $$w_{Rx}(z) = w_{Rx}(0)\sqrt{1+(z/z_{Rx})^2}$$

The intensity is Gaussian and remains Gaussian when the beam is propagating in free space or through parabolic optical systems. The divergence, however, is higher for the same beam waist and the Rayleigh range shorter compared to a Gaussian TEM_{00}-mode. The parameter σ characterizes the degree of coherence with $0 \le \sigma \le 1$. For complete coherence σ is equal to 1.0, an incoherent beam exhibits $\sigma = 0$. The local transverse coherence length of this specific beam does not depend on the position x. Some relevant parameters are:

beam radius: $w_{Rx}(z)$

beam divergence: $\Theta_{Rx} = \Theta_{0x}/\sigma$

beam propagation factor: $M_x^2 = 1/\sigma$

degree of coherence: $$|\gamma_{12,x}(x_1,x_2)| = \exp\left[-\frac{1}{2}\left(\frac{x_1-x_2}{w_{Rx}}\right)^2\left(\frac{1}{\sigma^2}-1\right)\right]$$

global coherence: $K_x^2 = \sigma$

coherence length: $\ell_{glob,x}(z) = 2w_{0x}(z)\sigma$

with w_{0x} and θ_{0x} being the values of the fundamental mode.
For coherent light (σ =1), the parameters of the fundamental mode are obtained. The Gauss-Schell beam can be realized by an incoherent superposition of squared Gauss-Hermite polynomials with suitable coefficients [1.42,1.48].

Determination of Phase and Coherence by Non-interferometric Measurements

For sake of simplicity, fields with only one transverse coordinate $E(x,z)$ are discussed. The two-dimensional case is not straight forward, as will be briefly pointed out in the next section. The cross correlation function of $E(x,z)$ is a two-dimensional function $\Gamma_x(x_1,x_2)$ or $\Gamma_x(x,s_x)$, using (2.147). By measuring Γ_x in one transverse plane (e.g. $z=0$) all parameters of interest such as beam width, divergence, phase front, and coherence can be evaluated. On the other hand, the propagation of the field and the intensity depend on the phase and the amplitude distribution in a z-plane and is also determined by the cross-correlation function Γ_x. The full information on the field is contained in Γ_x as well as in the intensity structure. It is possible to reconstruct the two-dimensional cross-correlation function $\Gamma_x(x,s_x)$ from the two-dimensional intensity $I(x,z)$ by using the ambiguity function Z_A , as will be demonstrated.

The Ambiguity-function Z_A

The propagation of the cross-correlation function $\Gamma_x(x_1,x_2,z)$ in free space is given by the Fresnel-integral in the paraxial approach (2.23). However, Γ_x is the product of two fields resulting in a complicated diffraction integral. It is much easier to use the ambiguity-function, which is the Fourier-transform of the cross-correlation function with respect to x [1.25,1.38]:

$$\square_{A,x}(s_x,\alpha_x) = \frac{1}{\lambda}\int_{-\infty}^{+\infty}\Gamma_x(x,s_x)\exp[ik\alpha_x x]dx \tag{2.162}$$

The ambiguity function is a representation of the field in the s_x,α_x space. The propagation of this function is determined by the ABCD-law [1.38], which for propagation in free space reads:

$$Z_{A,x}(s_x,\alpha_x,z) = Z_{A,x}(s_x-\alpha_x z,\ \alpha_x,\ 0) \tag{2.163}$$

The propagating function remains the same, if s_x is replaced by $s_x-\alpha_x z$. This coordinate transformation means a shearing of the function by the angle φ, as plotted in Fig. 2.35, with $\tan\varphi = z/z_0$ where z_0 is the Rayleigh length and φ the Guoy-shift (see (5.39) and (5.40)). Let us discuss some properties of this new function. For $\alpha_x = s_x = 0$, (2.162) and (2.149) immediately deliver:

$$Z_{A,x}(0,0) = \frac{1}{\lambda}\int_{-\infty}^{+\infty}\Gamma_x(x,s_x=0)dx = \frac{1}{\lambda}P_x$$

This special value of the function is equal the square root of the power. According Parseval's theorem, the following relationship holds for $|Z_{A,x}|^2$:

$$\int\int |Z_{A,x}(s_x,\alpha_x)|^2 ds_x d\alpha_x = \frac{1}{\lambda}\int\int |\Gamma_x(x,s_x)|^2 dx ds_x = \frac{1}{\lambda}\int\int |\Gamma_x(x_1,x_2)|^2 dx_1 dx \qquad (2.164)$$

which according to (2.136) is the x-part of the global degree of coherence.

$$K_x^2 = \frac{\lambda}{P_x^2}\int\int |Z_{A,x}(s_x,\alpha_x)|^2 ds_x d\alpha_x$$

The transverse coherence length is given by (2.159). Replacing Γ_x by $Z_{A,x}$ and using (2.162) results in:

$$\ell_{glob,x}^2 = \frac{8}{K_x^2 P_x^2 \lambda}\int\int s_x^2 |Z_{A,x}(s_x,\alpha_x)|^2 ds_x d\alpha_x \qquad (2.165)$$

The second moment of s_x, weighted with the squared modulus of the ambiguity function is the transverse coherence length as shown in Fig.2.35. The second moment of α_x does not depend on z and is the far field angle of coherence as defined by (2.160):

$$\Theta_{coh,x}^2 = \frac{8}{K_x^2 P_x^2 \lambda}\int\int \alpha_x^2 |Z_{A,x}(s_x,\alpha_x)|^2 ds_x d\alpha_x \qquad (2.166)$$

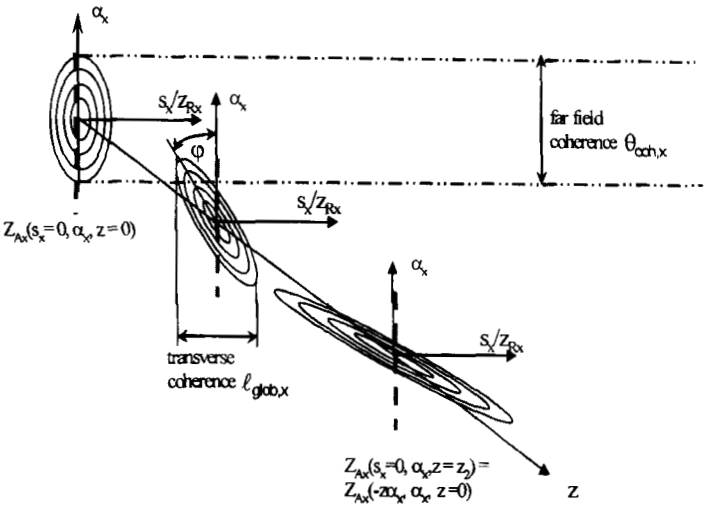


Fig.2.35 The ambiguity function is sheared by the angle $\tan\varphi = z/z_0$, when propagating in free space. It is convenient to normalise the s_x-parameter to the corresponding Rayleigh length z_{Rx}. The Fourier transform of the intensity is given by the dotted lines.

The transverse coherence length is given by (2.159). Replacing Γ_x by $Z_{A,x}$ and using (2.162) results in:

$$\ell^2_{glob,x} = \frac{8}{K_x^2 P_x^2 \lambda} \int\int s_x^2 |Z_{A,x}(s_x,\alpha_x)|^2 ds_x d\alpha_x \tag{2.165}$$

The second moment of s_x, weighted with the squared modulus of the ambiguity functior is the transverse coherence length as shown in Fig. 2.35. The second moment of α_x does not depend on z and is the far field angle of coherence as defined by (2.160):

$$\Theta^2_{glob,x} = \frac{8}{K_x^2 P_x^2 \lambda} \int\int \alpha_x^2 |Z_{A,x}(s_x,\alpha_x)|^2 ds_x d\alpha_x \tag{2.166}$$

The relevant coherence parameters can be evaluated, if $Z_{A,x}$ is known. To understand how it is obtained from intensity measurements, let us start with (2.162) at z=0, which for s_x = 0 together with (2.148) delivers:

$$Z_{A,x}(s_x=0,\alpha_x) = \int I(x,z=0)\ \exp[ik\alpha_x x]\ dx \tag{2.167}$$

Propagation by a distance z is described by (2.163):

$$Z_{A,x}(s_x,\alpha_x) = \int I(x,z)\ \exp[ik\alpha_x x]\ dx \qquad with \quad z = -s_x/\alpha_x \tag{2.168}$$

In a measurement, the one-dimensional intensity distribution *I(x,z)* is Fourier-transformed to generate the ambiguity function with s_x=-zα_x. z is running from -∞ to +∞. In reality it is sufficient to determine the intensity in 10 to 20 positions around the focal range, where it is varying strongly. This simplifies the determination of the coherence parameters, because the transverse intensity structure can be easily measured with a CCD-camera. Details of the measurement technique were published in [1.80]. The procedure is as follows

- measure *I (x,z)* at 10-20 different positions around the focal region
- calculate $Z_{A,x}$ using (2.168) and replace *z* by $-s/\alpha$
- With $Z_{A,x}$, global coherence, coherence length and far-field divergence can be calculated according to (2.165) and (2.166)
- Fourier transform of $Z_{A,x}$ generates the cross-correlation function (2.162), from which amplitude and phase front can be calculated.
- beam width, divergence, and beam propagation factor can be calculated from *I(x,z)* using the second order intensity moments (see Sec. 2.6)

Figure 2.36 shows experimental results for an off-axis diode-end-pumped Nd:YAG laser. By moving the pump spot laterally with respect to the optical axis of the stable resonator, Gauss-Hermite modes of a desired order can be excited. The measured global coherence versus the diode adjustment is shown in Fig. 2.36. The global degree of coherence always has a maximum when a pure TEM_{om} mode oscillates, but never reaches $K_x^2 = 1$, which means that other modes oscillate as well. Table 2.6 compares the results of this non-interferometric measurement with the ones obtained with a Young-interferometer. The theoretical values are based on the assumption that only two modes are oscillating, using the formulas of the two-mode-case.

Table 2.6 Comparison of beam propagation factor M^2, global degree of coherence K_x^2, and global transverse coherence length $\ell_{coh,x}$, obtained by Young's interferometer and by non-interferometric measurements [1.108]. a_0, a_1 are the relative power contents of TEM_{00} and TEM_{01} mode, respectively.

$a_0 : a_1$	M_x^2		K_x^2			$\ell_{glob,x} / w_{Rx}$		
	theory	exp.	theory	Young	non-interf.	theory	Young	non-interf.
1.00 : 0.00	1.0	1.01	1.0	0.96 - 1.0	1.0	1.0	0.99	1.01
0.75 : 0.25	1.75	1.71	0.79	0.70 - 0.75	0.74	0.63	0.63	0.60
0.50 : 0.50	2.0	2.04	0.71	0.66 - 0.73	0.71	0.71	0.74	0.71
0.25 : 0.75	2.5	2.39	0.79	0.72 - 0.77	0.75	0.94	0.89	0.87
0.00 : 1.00	3.0	3.04	1.0	0.89 - 0.99	0.95	1.0	0.97	0.96

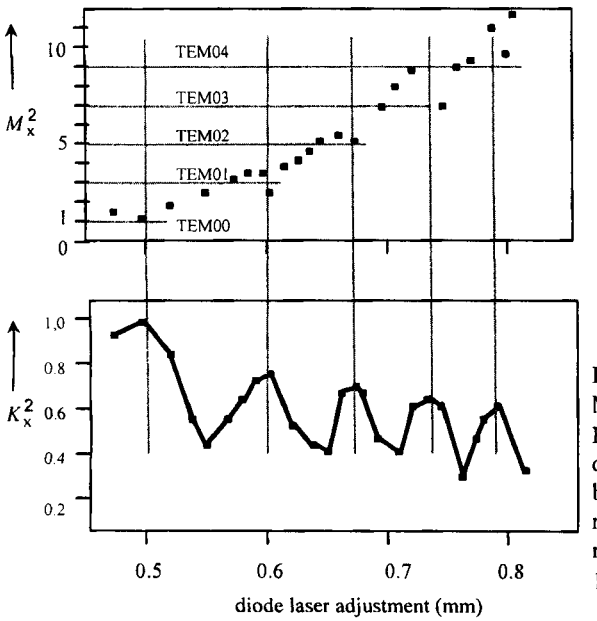


Fig. 2.36 Beam propagation factor M_x^2 and global degree of coherence K_x^2. The modes were generated in a diode end-pumped Nd-YAG-laser, by laterally shifting the diode with respect to the optical axis of the resonator [1.76] (© Elsevier Science 1996).

If the field cannot be separated according to $E(x,y) = E_0 f(x) \cdot g(y)$, it exhibits two transverse dimensions, resulting in a four-dimensional cross-correlation function. The corresponding ambiguity functions are also four-dimensional. However, since the intensity $I(x,y,z)$ is only a three-dimensional function, the missing information is hidden in the azimuthal phase structure. By inserting a cylinder lens and rotating it by an angle β, the additional information can be recovered. Now the intensity becomes four-dimensional, $I(x,y,\beta,z)$ and can be used to determine all coherence characteristics and beam parameters, but it is a time consuming method.

Wigner-function

Another function often used in Optics is the Wigner-function $h(\boldsymbol{r},\boldsymbol{\theta})$, which is defined as the Fourier transform of the cross-correlation function with respect to $\boldsymbol{s}=\boldsymbol{r}_1-\boldsymbol{r}_2$:

$$h(\boldsymbol{r},\theta) = \frac{1}{\lambda^2}\int \Gamma_{12}(\boldsymbol{r},\boldsymbol{s}) \exp[-i\boldsymbol{k}\theta\boldsymbol{s}]\, d\boldsymbol{s} \tag{2.169}$$

where $\boldsymbol{k}$ is the wave vector (k_x,k_y), $\boldsymbol{r}$ is the position vector and $\boldsymbol{\theta}$ the far field divergence vector:

$$\boldsymbol{r} = \begin{pmatrix} x \\ y \end{pmatrix} \qquad \boldsymbol{\theta} = \begin{pmatrix} \theta_x \\ \theta_y \end{pmatrix}$$

This function was first introduced by Wigner [1.24] for the description of quantum mechanical phenomena in the phase space and later used by Walther [1.30] and Wolf [1.45]. The properties of this function and its application to beam propagation is summarized in several papers of Bastiaans [1.39,1.40,1.44]. The special feature of coherence is discussed in detail by Eppich [1.78,1.80].

The Wigner-function as well as the ambiguity-function and the cross-correlation function completely characterize a radiation field and all three functions contain the same information. The cross-correlation function is obtained directly from the interference experiments as discussed in the beginning of this section. Ambiguity and Wigner-function have to be evaluated from intensity measurements, which are easier to perform but require sophisticated numerical approaches. The Wigner-function is related to the radiance of geometrical optics, being approximately (but not correctly) the power per area and angle [1.30]. The advantage of using the Wigner-function is its simple propagation law through first order optical elements. The same relation holds as for the ambiguity function:

$$h(\boldsymbol{r}_2,\theta) = h(\boldsymbol{M}^{-1}\boldsymbol{r}_2\ ,\ z) \tag{2.170}$$

where M is the ABCD-matrix of Sec. 2.4.2. For propagation in free space, h is sheared the same way as shown for the ambiguity function in Fig. 2.35.

Without mathematical derivations, some results are summarized:

beam power
$$P = \int h(r,\theta)drd\theta \tag{2.171}$$

near field intensity
$$I(r) = \int h(r,\theta)d\theta \tag{2.172}$$

far field intensity
$$I(\theta) = \int h(r,\theta)dr \tag{2.173}$$

beam width
$$\left.\begin{matrix}\langle x^2\rangle \\ \langle y^2\rangle \\ \langle xy\rangle\end{matrix}\right\} = \frac{1}{P}\int\begin{pmatrix}x^2 \\ y^2 \\ xy\end{pmatrix} h(r,\theta)drd\theta \tag{2.174}$$

beam divergence
$$\left.\begin{matrix}\langle \theta_x^2\rangle \\ \langle \theta_y^2\rangle \\ \langle \theta_x\theta_y\rangle\end{matrix}\right\} = \frac{1}{P}\int\begin{pmatrix}\theta_x^2 \\ \theta_y^2 \\ \theta_x\theta_y\end{pmatrix} h(r,\theta)drd\theta \tag{2.175}$$

A look at Section 2.6.2 shows that the Wigner-function delivers all beam parameters. Unfortunately, it can not be measured directly, but for one-dimensional transverse fields it can be evaluated from the intensity by a suitable transformation [1.78]. For two-dimensional fields it is also possible to retrieve the Wigner-function from intensity-measurements, but the procedure is much more complicated [1.78]. Therefore h is more of a mathematical tool, that can be used to calculate the propagation of light, also through non-parabolic optical systems, including slits and arbitrary phase plates. In order to gain a better feeling for this concept, let us apply the Wigner-function theory to a one-dimensional Gaussian beam with a field distribution at the waist position of:

$$E(x,z=0) = E_0 \exp\left[-\frac{x^2}{w_0^2}\right]$$

The cross-correlation function is given by (2.150/2.151):

$$\Gamma(x_1,x_2,z=0) = \frac{1}{2}c_0\epsilon_0|E_0| \exp\left[-\frac{x_1^2+x_2^2}{w_0^2}\right] = \frac{1}{2}c_0\epsilon_0|E_0| \exp\left[-\frac{2x^2+s^2/2}{w_0^2}\right] \tag{2.176}$$

Application of the Fourier transform (2.169) results in the Wigner-function at $z = 0$:

$$h_0(x,\theta,z=0) = \sqrt{2\pi}\frac{w_0}{\lambda}\frac{1}{2}c_0\epsilon_0|E_0|^2 \exp\left[-2\frac{x^2}{w_0^2} - \frac{k^2\theta^2 w_0^2}{2}\right] \qquad (2.177)$$

Free space propagation of the Wigner-function is performed by applying the ABCD-law (2.170), replacing x by $x-\theta z$:

$$h_0(x,\theta,z) = \sqrt{2\pi}\frac{w_0}{\lambda}\frac{1}{2}c_0\epsilon_0|E_0|^2 \exp\left[-2\frac{(x-\theta z)^2}{w_0^2} - \frac{k^2\theta^2 w_0^2}{2}\right] \qquad (2.178)$$

which again is a Gaussian shaped function in x- and θ- direction. This function and its propagation in free space are shown in Fig. 2.37. Starting at the waist, the projection onto the x-axis is the near field intensity with the waist-radius $\langle x^2\rangle = w_0^2/4$.The projection on the θ-axis is the far field intensity with the divergence $\langle\theta^2\rangle = \theta_0^2/4$. With increasing distance z, the beam radius increases according to (2.57), whereas the far field divergence remains constant.

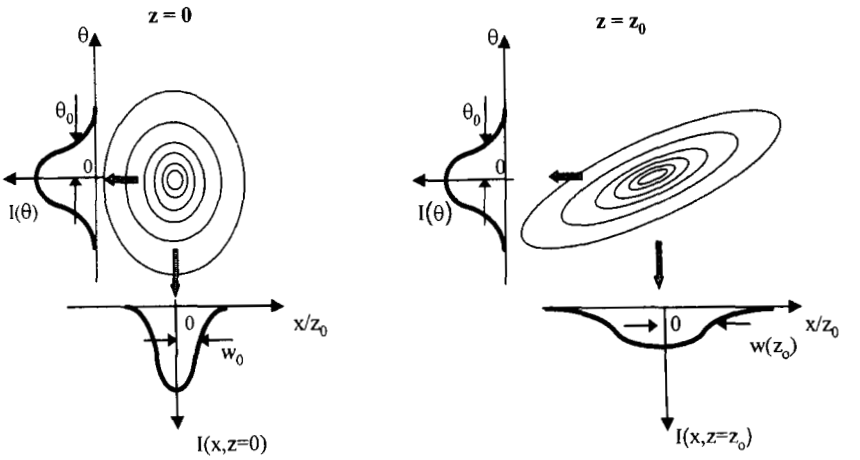


Fig. 2.37 The Wigner-function of the Gaussian beam and its propagation in free space. The x coordinate is normalized to the Rayleigh-length $z_0=\pi w_0^2/\lambda$.

2.8 Diffraction Theory of Optical Resonators

2.8.1 Integral-Equation for the Electric Field Distribution

By using the formalism of the Collins integral introduced previously, it is now a straightforward task to develop a diffraction theory for optical resonators [1.26,1.29, 1.31]. We are basically interested in the steady-state electric field distribution on the two resonator mirrors. Let us consider an optical resonator with rotational symmetry, as depicted in Fig. 2.38, with a field distribution $E_1(x,y)$ on mirror 1. What does the electric field look like after one round trip inside the resonator? In Sec. 1.3 we calculated the ray transfer matrix for a round trip, assuming an empty resonator:

$$M = \begin{pmatrix} G & 2Lg_2 \\ \dfrac{G^2-1}{2Lg_2} & G \end{pmatrix}, \qquad G = 2g_1g_2-1 \tag{2.179}$$

By inserting the matrix elements into the Collins-Integral (2.29) we get the field after the round trip:

$$E_2(x_2,y_2) = i\frac{\exp[-i2kL]}{2Lg_2\lambda}\int\int E_1(x_1,y_1)\exp\left[\frac{-i\pi}{2Lg_2\lambda}\left(G(x_1^2+y_1^2+x_2^2+y_2^2)-2(x_1x_2+y_1y_2)\right)\right]dx_1dy_1 \tag{2.180}$$

whereby the integration is performed over the surface of mirror 1.

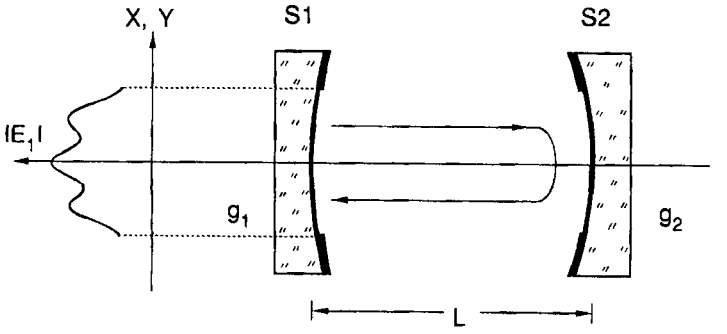


Fig. 2.38 After one round trip inside the resonator, the field distribution $E_1(x,y)$ has reproduced itself. In steady state operation the field amplitude can change but the shape of the field distribution has to remain constant.

The steady state oscillation is defined by the constancy of the field distribution: field E_2 is identical to the starting field E_1, except for a proportionality factor γ. The following relation must therefore hold:

$$E_2(x_2,y_2) = \gamma\ E_1(x_2,y_2) \tag{2.181}$$

The complex proportionality constant γ is called the eigenvalue of the field distribution. Replacing the left hand side of (2.180) with (2.181) yields an integral equation for the field distribution $E_1(x,y)$. Once this distribution is calculated (in general this is done numerically), the electric field on any other plane inside or outside the resonator can be determined by applying the corresponding Collins integral [1.26,1.29,1.31].

It should be kept in mind that the Collins-integral can only be used strictly to calculate the beam propagation between two apertures. If both of the resonator mirrors are limited in size, as is the case in Fig. 2.38, one must not use the Collins integral for the complete round trip. The two transits between the mirrors have to be dealt with separately by using two Collins integrals with the corresponding ray transfer matrices for each transit.

The physical meaning of the eigenvalue γ becomes clear when we compare the total power P_2 hitting the mirror after the round trip with the initial power P_1. We get:

$$P_2 = \frac{1}{2}c\epsilon_0 \int\int E_2(x_2,y_2)E_2^*(x_2,y_2)dx_2dy_2 = \gamma\gamma^* \frac{1}{2}c\epsilon_0 \int\int E_1(x_2,y_2)E_1^*(x_2,y_2)dx_2dy_2$$

$$= \gamma\gamma^*\ P_1 \tag{2.182}$$

where we have used (2.181). The integration is again performed over the surface of mirror 1. It is obvious that the factor $\gamma\gamma^*$ represents the power fraction that hits the mirror surface again after one round trip. This factor is referred to as the loss factor or the diffraction loss factor V. In Fig. 2.38 only the fraction V of the initial power falls into the aperture and gets reflected by the mirror, whereas the remainder $1\text{-}V$ hits the aperture and gets absorbed or scattered. The fractional power loss is called the diffraction loss $\Delta V=1\text{-}V$ of the field distribution. For a laser in steady-state oscillation, the diffraction loss as well as other losses generated by output coupling and scattering are exactly compensated by the gain of the active medium. In general an infinite number of self-reproducing field distributions that fulfill (2.180) and (2.181) can be found for a giv en resonator. Each of these field distributions represents an eigenmode of the resonator with a well defined eigenvalue γ. Which of these eigenmodes will oscillate inside a laser resonator depends on the boundary conditions like aperture location and size and the gain of the active medium. All of those eigenmodes whose losses are compensated by the gain can oscillate. Since different eigenmodes experience different loss, the number of oscillating eigenmodes is decreased as the aperture size or the gain of the active medium are reduced. By using this mode selection technique one can force only the lowest-loss mode, also referred to as the fundamental mode, to oscillate.

2.8.2 The Gaussian Beam as a Fundamental Resonator Mode

One special solution of the eigenvalue problem (2.180/2.181) is the Gaussian beam, which represents the fundamental mode in a stable resonator. In the following, we will discuss the integral equation for an arbitrary optical resonator by using the Collins integral (2.29) expressed in terms of a general round trip ray transfer matrix $\boldsymbol{M}$. The mirror surfaces are assumed to extend to infinity and no apertures are located inside the resonator (Fig. 2.39). In reality, this means that the transverse dimension of the field is small compared to the sizes of all apertures.

We consider a resonator with an arbitrary number of optical elements located inside and the ray transfer matrix $\boldsymbol{M}$ represents the matrix for a resonator round trip starting on a given, but arbitrarily chosen, reference plane. The field distribution on the reference plane is a solution of the Collins integral equation:

$$\gamma E(x_2,y_2) = i\frac{\exp[-ik2L]}{\lambda B}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} E(x_1,y_1)\exp\left[\frac{-i\pi}{\lambda B}\left(A(x_1^2+y_1^2)+D(x_2^2+y_2^2)-2(x_1x_2+y_1y_2)\right)\right]dx_1dy_1$$

We know already that a Gaussian beam with beam parameter q_1 will transform into a Gaussian beam with beam parameter q_2 in this integral equation, with:

$$q_2 = \frac{Aq_1 + B}{Cq_1 + D} \tag{2.183}$$

Since we are looking for a steady-state solution, $q_2=q_1$ must hold. By replacing the left hand side of (2.183) with q_1, we can determine q_1 to be:

$$\frac{1}{q_1} = \frac{D-A}{2B} - \frac{i}{2B}\sqrt{4-(A+D)^2} \tag{2.184}$$

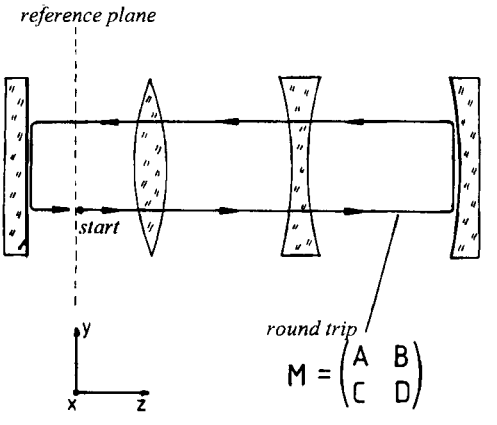


Fig. 2.39 Determination of the self-consistent Gaussian beam parameter on the reference plane for a round trip in an optical resonator.

Comparison with (2.58) yields the radius of curvature and the beam radius of the Gaussian beam on the reference plane:

$$R_1 = \frac{2B}{D-A} \quad , \quad w_1^2 = \frac{\lambda}{\pi}\frac{2B}{\sqrt{4-(A+D)^2}} \tag{2.185a,b}$$

Equation (2.185b) reveals that a Gaussian beam can only be a solution if:

$$|A+D| < 2 \tag{2.186}$$

holds. Optical resonators for which (2.186) holds are called *stable resonators*. It should be added that this result always holds no matter which reference plane in the resonator is chosen. The Gaussian beam represents the fundamental mode inside a stable optical resonator.

Example: Linear two-mirror resonator without internal optical elements
As already discussed in Sec. 1.3 the ray transfer matrix for the roundtrip starting on spherical mirror 1 reads:

$$M = \begin{pmatrix} G & 2Lg_2 \\ \dfrac{G^2-1}{2Lg_2} & G \end{pmatrix} \qquad \textit{with} \quad G = 2g_1g_2-1$$

The stability condition (2.186) now reads:

$$|G| < 1 \quad \Rightarrow \quad 0 < g_1g_2 < +1 \tag{2.187}$$

By using (1.185a,b), the radius of curvature and the beam radius on mirror 1 are:

$$R_1 = \infty \quad , \quad w_1^2 = \frac{2Lg_2\lambda}{\pi\sqrt{1-G^2}}$$

An infinite radius of curvature on the mirror means that the mirror surface represents the wave front of the Gaussian beam (remember that in Sec. 1.3 we have replaced the mirror by two lenses and our reference plane is located in between). The Gaussian beam parameters on mirror 2 can be obtained by switching the indices in (2.187).

Beam radii and radii of curvature at any other plane inside or outside the resonator can be easily obtained by applying the ABCD law for Gaussian beams. However, for propagation outside the resonator one has to incorporate one 'replacement lens' into the corresponding ray transfer matrix.

2.9 Plane Wave Representation of Diffraction

An exact solution of the wave equation (2.1) is the infite, monochromatic plane wave, which, omitting the frequency term, reads in the complex notation:

$$\boldsymbol{E}(x,y,z) = \boldsymbol{A}_0 \exp[-i\boldsymbol{kr}] \tag{2.188}$$

with $\boldsymbol{E}=(E_x,E_y,E_z)$: electric field vector
$\boldsymbol{k}=(k_x, k_y, k_z)$: wave vector with $|\boldsymbol{k}|=2\pi/\lambda=\omega/c$
$\boldsymbol{r}=(x,y,z)$: position vector

Any superposition of plane waves travelling in different directions $\boldsymbol{k}$ is also a solution of (2.1):

$$\boldsymbol{E}(x,y,z) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} A(k_x,k_y) \exp[-i\boldsymbol{kr}]\, dk_x dk_y \tag{2.189}$$

$A(k_x,k_y)$ is called the angular spectrum of the field $\boldsymbol{E}$. For a single plane wave propagating in direction $\boldsymbol{k}=(k_{x0},k_{y0})$, the spectrum is a delta peak:

$$A(k_{x0},k_{y0}) = A_0\delta(k_x-k_{x0},k_y-k_{y0}) \qquad with \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \delta(k_x-k_{x0},k_y-k_{y0})dk_x dk_y = 1$$

Since $|\boldsymbol{k}|$ is constant, (2.189) can be rewritten into the form: (2.190)

$$\boldsymbol{E}(x,y,z) = \int_{k_x^2+k_y^2\le k^2} \int A(k_x,k_y) \exp\left[-i\left(k_x x+k_y y+z\sqrt{k^2-k_x^2-k_y^2}\right)\right] dk_x dk_y$$

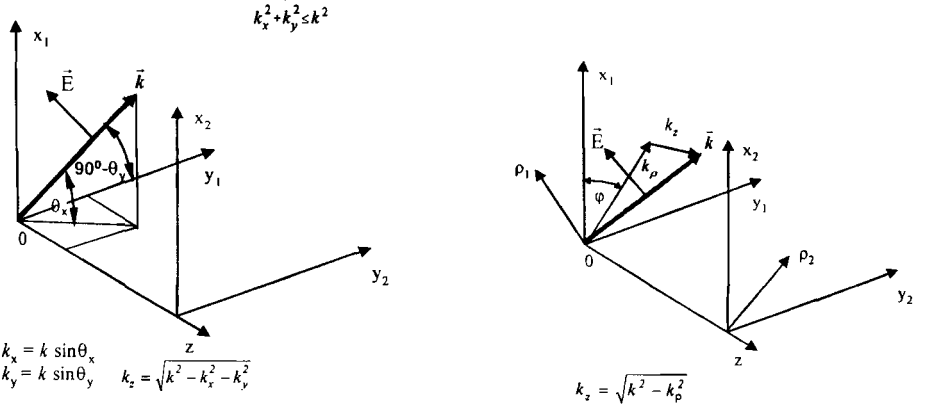


Fig. 2.40 A plane wave propagation in the direction of $\boldsymbol{k}$. Left: rectangular symmetry, right: circular symmetry.

The exponent in (2.190) becomes real for $k_x^2+k_y^2>k^2$ which means that the field amplitude decays very fast with $exp[-2\pi z/\lambda]$. This evanescent term does not contribute to the propagating field and can be neglected. The integration is limited to values $k_x^2+k_y^2 \le k^2$ or angles $|\theta_x|, |\theta_y| \le \pi/2$ and (2.190) holds for $z>\lambda$. If the field $\boldsymbol{E}(x,y,z)$ is given at the plane z=0, (2.190) results in:

$$\boldsymbol{E}(x_1,y_1,0) = \iint \boldsymbol{A}(k_x,k_y)\, \exp\left[-i(k_x x_1 + k_y y_1)\right]\, dk_x dk_y$$

$$\Leftrightarrow \qquad \boldsymbol{E}(x_1,y_1,0) = \boldsymbol{F}^{-1}[\boldsymbol{A}(k_x,k_y)]$$

$\boldsymbol{E}(x_1,y_1,0)$ is the inverse Fourier transform $\mathbf{F}^{-1}$ of the angular spectrum, as discussed in Sec. 2.4.2, and accordingly, the angular spectrum $\boldsymbol{A}(k_x,k_y)$ is the Fourier transform of the field $\boldsymbol{E}$. Now (2.190) can be rewitten as:

$$\boldsymbol{E}(x_2,y_2,z) = \iint \boldsymbol{F}[\boldsymbol{E}(x_1,y_1,0)]\, \exp\left[-i\left(k_x x_2 + k_y y_2 + z\sqrt{k^2-k_x^2-k_y^2}\right)\right]\, dk_x dk \qquad (2.191)$$

If the field is given at the plane $z=0$, its propagation in free space is determined by (2.191).This so called plane wave representation is a better approach to diffraction than the Kirchhoff/Fresnel-integrals [1.2,1.77]. In addition, it automatically incorporates the vector character of the electric field. The plane wave presentation is useful for the evaluation of beam propagation in anisotropic media, higher order approximation of beam propagation and for the evaluation of diffraction free beams.

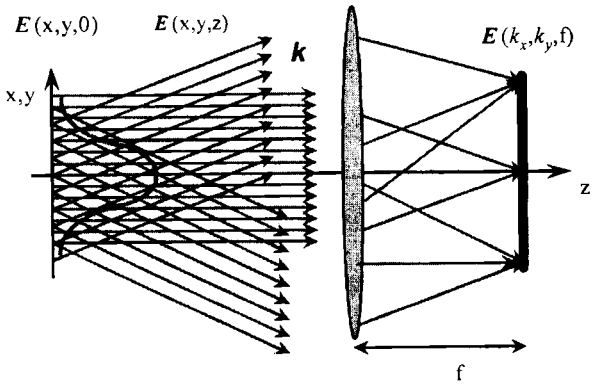


Fig. 2.41 The field distribution in the plane z=0 is decomposed into plane waves. The plane wave spectrum is the Fourier transform of the field in a given plane (z=0) and it can be observed in the far field or in the focal plane of a lens. The propagation of the plane waves is described by (2.188), and their superposition delivers the propagating field $\boldsymbol{E}$(x,y,z).

Equation (2.191) automatically predicts that laterally confined electric field distributions will develop a longitudinal field component in direction of the beam propagation. The field of the single plane wave, travelling with an angle with respect to the z-axis, has a longitudinal component E_z, and the superposition of all fields will also have a resulting longitudinal component. This is a general property of fields with a transverse structure propagating in charge free media. Only the infinite plane wave, propagating in z-direction has no longitudinal component, which is why light is often described as a transverse electro-magnetic field. In the paraxial approach with $k_x, k_y \ll k$, the square root can be expanded. In first order approximation, (2.191) can then be transformed into:

$$\boldsymbol{E}(x_2,y_2,z) \propto \exp[-ikz] \int\limits_{k^2+k^2\le k^2} \int \boldsymbol{F}[E(x_1,x_2,0)] \exp\left[-i\left(k_x x_2 + k_y y_2 + \frac{z}{2k}(k_x^2+k_y^2)\right)\right] dk_x dk_y \tag{2.192}$$

By inserting the Fourier-integral:

$$\boldsymbol{F}[\boldsymbol{E}(x_1,y_1,0) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \boldsymbol{E}(x_1,x_1,0)] \exp[i(k_x x_1 + k_y y_1)] \, dx_1 dy_1 \tag{2.193}$$

and integrating (2.192) with respect to k_x, k_y, the well known Fresnel-integral is obtained:

$$\boldsymbol{E}(x_2,y_2,z) = i\frac{k}{2\pi z}\exp[-ikz]\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \boldsymbol{E}(x_1,x_1,0)] \exp\left[-\frac{ik}{2z}[(x_2-x_1)^2+(y_2-y_1)^2]\right] dx_1 dy_1 \tag{2.194}$$

Example:
A Gaussian field distribution with the waist radius w_0 is given at the plane z=0:

$$\boldsymbol{E}(x,y,0) = \boldsymbol{E}_0 \exp\left[-\frac{x^2+y^2}{w_0^2}\right]$$

The Fourier Transform reads:

$$\boldsymbol{A}(k_x,k_y,0) = \boldsymbol{F}[\boldsymbol{E}(x,y,0)] = \boldsymbol{E}_0 \int\int \exp\left|-\frac{x^2+y^2}{w_0^2}\right| \exp[i(k_x x + k_y y)] \, dxdy$$

The integral can be solved solved and delivers the angular spectrum of the Gaussian shaped field:

$$A(k_x,k_y,0) = E_0\frac{w_0^2}{4\pi}\exp\left[-\frac{w_0^2}{4}(k_x^2+k_y^2)\right]$$

The paraxial approximation of the plane wave presentation (2.192) then yields for the electric field at the plane z:

$$E(x,y,z) = \frac{E_0}{1+iz/z_0}\exp\left[-\frac{x^2+y^2}{w_0^2(1+iz/z_0)}\right] \qquad k_x,k_y \ll k \tag{2.195}$$

with $z=\pi w_0^2/\lambda$ being the Rayleigh range. For large values of k_x,k_y, or large angles θ_x,θ_y the correct equation (2.189) has to be solved. For an x-polarized beam, an approximate second order solution reads [1.38,1.41]:

$$E(x,y,z) = \frac{E_0}{1+iz/z_0}\exp\left[-\frac{x^2+y^2}{w_0^2(1+iz/z_0)}\right]\left(1-2i\frac{z/z_0}{kz_0(1+iz/z_0)}\ ;\ 0\ ;\ -i\frac{x/z_0}{1+iz/z_0}\right) \tag{2.196}$$

The x-component is corrected by an additional term. In the far field ($z\to\infty$), this term approaches $-2/kz_0=-\theta_0^2$ where θ_0 is the half angle of divergence of the beam. This term becomes relevant for highly divergent fields. Moreover, a z-component appears, as already discussed. The real part of E_z is the projection of the tilted E-vector, which is tangential to the curved wave front, as shown in Fig. 2.42.

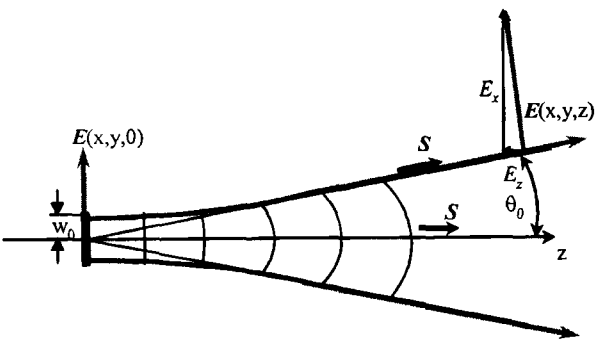


Fig. 2.42 The orientation of the electric field vector of a Gaussian beam

The z-component can be written as:

$$Re\left(\frac{E_z}{E_x}\right) = -\frac{x}{R(z)}$$

with $R(z)$ being the radius of curvature of the phase front. The longitudinal field component vanishes in the waist, where the field is plane and perpendicular to the optical axis. The imaginary part of E_z is shifted by 90° with respect to E_x, which means that the field is slightly elliptically polarized in the x-z plane.

Example:
At the output of a diode laser, the waist radius is about 1 μm for a wavelength of $\lambda=0.8\ \mu m$. The half angle of divergence is $\theta_0=\lambda/(\pi w0) = 0.25\ mrad$, the Rayleigh range is $z_0=\pi w_0^2/\lambda=3.9\ \mu m$ and the z-component of the electric field E_z at the waist $x=y=w_0$ becomes $E_z(w_0,w_0,0)= - \theta_0/e^2 = -i\ 0.025\ E_0$. In the far field, the correction of the x-component is $-0.0625\ E_0$. Even for diode lasers, the correction terms are small.

Circular Symmetry
If the electric field exhibits circular symmetry and does not depend on the azimuthal angle ϕ, the same procedure as shown in Sec. 2.2.2 can be applied. The plane wave presentation in circular symmetry is given by:

$$\boldsymbol{E}(r_2,z) = 2\pi\int_0^k \boldsymbol{A}(k_r)\ J_0(k_r r_2)\ \exp\left[-iz\sqrt{k^2-k_r^2}\right] k_r dk_r \qquad (2.197)$$

Replacing the angular spectrum by its Fourier Transform results in:

$$\boldsymbol{E}(r_2,z) = \int_0^k\int_0^\infty \boldsymbol{E}(r_1,0)\ J_0(k_r r_2)J_0(k_r r_1)\ \exp\left[-iz\sqrt{k^2-k_r^2}\right] r_1 dr_1 k_r dk_r \qquad (2.198)$$

J_0 is the zero-order Bessel function. In the paraxial approach ($|k_r| \ll |k|$), the Fresnel-integral for circular symmetry of Sec. 2.2.2. is obtained.

2.10 Diffraction-Free Beams

Normally the structure of the electromagnetic field changes while propagating in free space. These changes are described by the wave equation (2.1) or by one of the diffraction integrals (2.12) and (2.191). However, special solutions of the diffraction integral exist, the eigensolutions, for which the shape of the field remains constant. Only the amplitude and the transverse dimension vary in such a way that the total power remains constant. One eigensolution is the Gaussian beam; others are discussed in Sec. 5.1. All of these beams experience diffractive spreading which means that their amplitude decreases during propagation. Fortunately, diffraction-free solutions of the wave equation exist [1.51,1.52,1.106]. One is the infinite plane wave as discussed in the previous Section. Normally, the field distribution generated by a superposition of plane waves will change during propagation. But there are special superpositions that do not change their field distributions.

A set-up for the realization of a beam generated by a superposition of plane waves is shown in Fig. 2.43. In the front focal plane of a lens with focal length f is a coherently illuminated ring with radius R and width ΔR. The field starting from the differential area $R\Delta R\Delta\phi$ is collimated by the lens resulting in a plane wave:

$$\Delta E(x,y,z,t) = E_0 \exp[-i(k_x x + k_y y + k_z z - \omega t)]\ \Delta\phi \tag{2.199}$$

with the wave vector components:

$$k_x = \frac{2\pi}{\lambda} \sin\phi\ \sin\theta_P \qquad k_y = \frac{2\pi}{\lambda} \cos\phi\ \sin\theta_P \qquad k_z = \frac{2\pi}{\lambda} \cos\phi\ \cos\theta_P$$

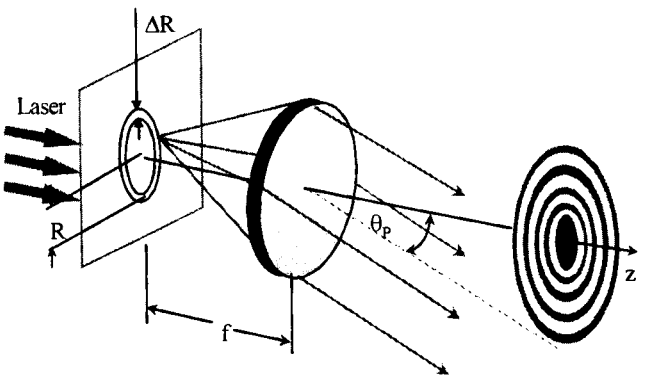


Fig. 2.43 Schematic set-up for generating a diffraction-free Bessel beam. In order to illuminate the lens completely, the width ΔR of the ring must be smaller than $\lambda f/w$., with w being the lens radius.

The resulting field behind the lens is obtained by integrating over all waves with ϕ varying between 0 and 2π. Equation (2.199) yields:

$$E(x,y,z,t) = E_0 \exp[-i(k_z z - \omega t)] \int_0^{2\pi} \exp[-ia(x \sin\phi + y \cos\phi)] d\phi \qquad (2.200)$$

with: $a = \frac{2\pi}{\lambda} \sin\theta_P$

The angle θ_P is determined by the focal length f of the lens and the radius R of the ring:

$$\tan\theta_P = R/f \qquad (2.201)$$

The above integral can be solved analytically and yields the Bessel function of zero order:

$$E(r,z,t) = E_0 \exp[i(k_z z - \omega t)] \, J_0(ar) \qquad (2.202)$$

and the intensity distribution:

$$I(r) = I_0 \, J_0^2(ar) \qquad (2.203)$$

with r being the radial coordinate (Fig. 2.44). This is an exact solution of the wave equation, as long as the field is not limited transversally. The intensity distribution does not depend on z, meaning the structure of the beam remains constant during propagation (Fig. 2.45). Be aware, that (2.202) only holds directly behind the lens. Due to the diameter $2w$ of the lens, the plane wave is limited and will spread due to diffraction. However, diffraction effects can be neglected for large Fresnel numbers $N=w^2/(\lambda L)$.

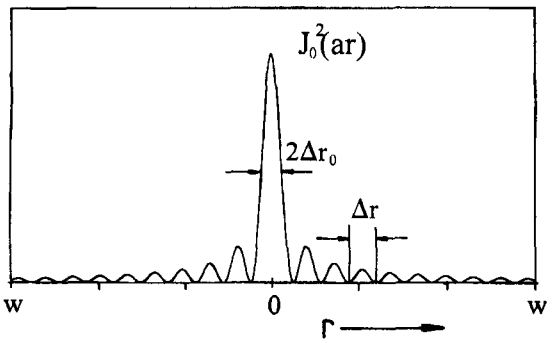


Fig. 2.44 Radial intensity distribution of the diffraction-free Bessel beam.

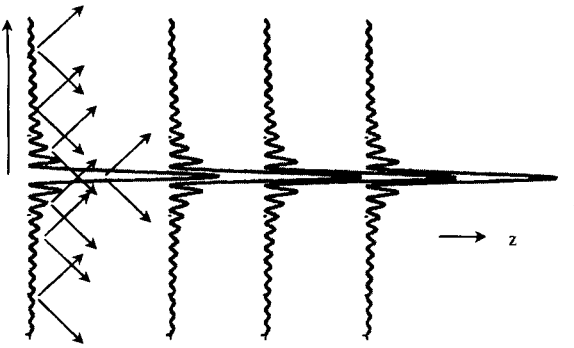


Fig. 2.45 The intensity distribution of a Bessel beam remains constant during propagation, even though power is constantly flowing transversally as indicated by the arrows.

The transverse structure of the intensity is plotted in Fig. 2.44. A central peak with a half maximum radius of:

$$\Delta r_0 \cong \frac{1}{a} \tag{2.204}$$

is surrounded by side lobes with peak intensities decaying with $1/r$. The distance between two consecutive intensity zeroes is almost constant:

$$\Delta r \cong \frac{\pi}{a} \tag{2.205}$$

The power in each individual ring is equal to that in the central peak. By using these approximations, the power P_0 in the central peak can be written as:

$$P_0 = P_{tot} \frac{\Delta r}{w} \tag{2.206}$$

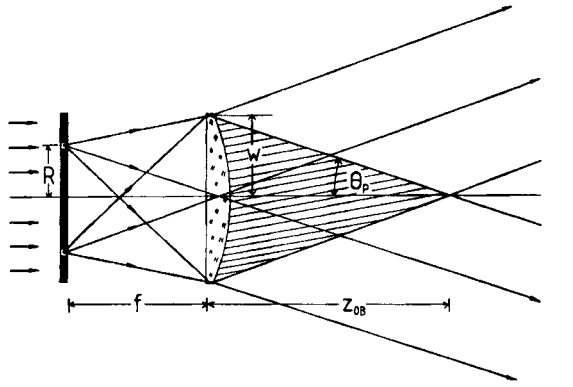


Fig. 2.46 The Rayleigh range z_{0B} of the Bessel beam. The shaded area is the interference range of the plane waves.

with P_{tot} being the total beam power at the lens. The sharp central peak is generated by the interference of the plane waves. It will propagate in the z-direction without broadening as long as the plane waves overlap. This diffraction-free propagation distance z_{0B} is given by (Fig. 2.46):

$$z_{0B} = \frac{w}{\tan\theta_P} \tag{2.207}$$

which again is the well-known Rayleigh range, now for large angles θ_P.

The set-up in Fig. 2.43 exhibits low efficiency, because only a small portion of the laser beam is used. A more efficient set-up using an axicon with radius R is shown in Fig. 2.47 [1.106]. Again the resulting field is a superposition of plane waves which all have the same slope $\sin\theta_p = k_{r0}/k$, $\tan\theta_p = (n-1)\gamma$, where n is the index of refraction of the axicon and γ its apex angle. Let us apply the plane wave presentation of diffraction in circular symmetry (2.197) to calculate the resulting intensity distribution. The angular spectrum is a delta peak:

$$E(k_{r0},0) = E_0\ \delta(k_r - k_{r0}) \qquad with \quad \int_0^\infty \delta(k_r - k_{r0})\ k_r dk_r = 1$$

and (2.197) right away delivers for field and intensity::

$$E_B(r,z) = E_0\ J_0(ar)\ \exp[-ikz] \qquad with \quad a = k\ \sin\theta_P \tag{2.208}$$

$$I(r) = I(0)\ J_0^2(ar) \tag{2.209}$$

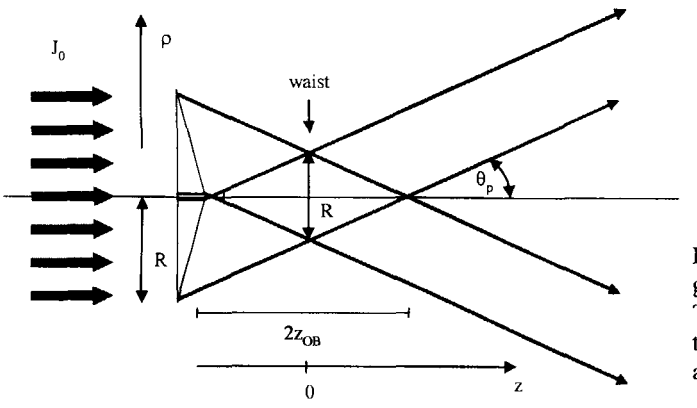


Fig. 2.47 Bessel beam generation using an axicon. The radius of the waist of the Bessel beam is $w=R/2$ at z=0.

Again a central peak with a half maximum radius of:

$$\Delta r_0 \cong \frac{1}{a} \tag{2.210}$$

and a power content of:

$$P_0 = \pi \, \Delta r_0^2 \, I(0) \tag{2.211}$$

is surrounded by side lobes. For $ar > 5\pi/2$, the Bessel function is approximated by:

$$J_0(ar) = \sqrt{\frac{2}{\pi ar}} \cos(ar - \pi/4)$$

which delivers for the distance between two consecutive intensity zeroes:

$$\Delta r \cong \frac{\pi}{a} \tag{2.212}$$

The power in each ring is equal to that in the central peak. By using these approximations, the power P_0 in the central peak for the set-up in Fig. 2.46 can be written as:

$$P_0 = P_{tot} \frac{\pi}{Ra} \tag{2.213}$$

with P_{tot} being the total beam power incident on the axicon with radius R. The sharp central peak is generated by the interference of the plane waves. It will propagate in the z-direction without broadening as long as the plane waves overlap. This diffraction-free propagation distance z_{0B} is approximately half the overlap-region in Fig. 2.47:

$$z_{0B} = \frac{R}{2\tan\theta_P} = \frac{R\pi}{\lambda a} \tag{2.214}$$

which is consistent with the usual definition of the Rayleigh range, the ratio of waist radius to divergence. Compared to the Rayleigh range of a Gaussian beam with the same waist radius Δr_0:

$$z_{0G} = \frac{\pi \Delta r_0^2}{\lambda} = \frac{\pi}{\lambda a^2} = \frac{z_{0B}}{Ra} \tag{2.215}$$

the Rayleigh range of the Bessel beam is larger by a factor *Ra*. Unfortunately, the power content in the main peak is reduced by the same factor.

The beam parameter product of a diffraction-free beam can only be evaluated numerically, but a rough estimation is easily done. The waist is located in the center of the overlap region at $z=0$. The waist radius containing 86.5% of the total power is given by:

$$w_0 = 0.865\ R/2$$

The far field divergence is θ_P since plane waves propagating at an angle θ_P are considered. For infinite plane waves, the far field is given by a delta ring $\delta(\theta-\theta_P)$. Since the near field is limited to a radius of $R/2$, a broadening occurs with $\Delta\theta_P \approx 2\lambda/R$. The ring, containing 86.%% of the power has a divergence of approximately $\theta_P + 0.865\Delta\theta_P \approx \theta_P$. The beam parameter product thus reads:

$$w_0\theta_P = \frac{0.865}{4}\frac{Ra\lambda}{\pi}$$

and the beam propagation factor results in:

$$M^2 = w_0\theta_P \frac{\pi}{\lambda} \approx 0.216\ Ra$$

which holds for $Ra \gg 1$.

Example 1 (annular apperture):
A circular slit with radius $R=2mm$ and width 20μm is placed in front of a lens with focal length $f=200mm$ and diameter $2w=20mm$. A 1W laser beam with a beam radius of 3mm (we assume that the intensity profile is homogeneous) and a wavelength of $\lambda=1\mu m$ generates a Bessel beam with the following parameters:

Divergence angle of the plane waves:	θ_P	= 10 mrad
Width of central peak:	$2\Delta r_0$	= 31.8 μm
Total power at the lens:	P_{tot}	= 8.89 mW
Power in central peak:	P_0	= 42.4 μW
Rayleigh range:	z_{0B}	= 1 m

A Gaussian beam with the same waist radius w_0 has a Rayleigh range of 794 μm.

This example clearly indicates that there are some disadvantages to the generation of diffraction-free beams:

- only a small part of the laser beam can be transformed into a Bessel beam using the simple annular aperture. This drawback can be overcome by using an axicon, by amplifying the Bessel beam, directly or holographically [1.59,1.106] or by using special optical resonators [1.82,1.106].

- generation of a small spot radius Δr_0 with a Rayleigh range that is $R/\Delta r_0$ larger than the conventional Rayleigh range requires a total beam power that is larger by the same factor.

- the enhancement of the Rayleigh range by the factor $R/\Delta r_0$ reduces the power in the central peak by the same factor.

- if the central peak is used only by cutting off the outer diffraction peaks with a pinhole, the Rayleigh range will be immediately reduced to the normal value. However, the central peak can be used to attain higher efficiencies in nonlinear optics and spectroscopy [1.89,1.92,1.95,1.96]. In this case, the central peak is reduced during propagation, but the Rayleigh range is not reduced in the same way.

Example 2 (axicon) [1.106]:
A ZnSe axicon (n=2.4) with radius $R=9.35mm$ and apex angle of 0.5° is homogeneously illuminated with a 1kW CO_2 laser beam at a wavelength of $\lambda=10.6\mu m$ The generated Bessel beam exhibits the following parameters:

Divergence angle of the plane waves:	θ_P	= 12.2 mrad
Width of central peak:	$2\Delta r_0$	= 277 μm
Power in central peak:	P_0	= 45.7 W
Rayleigh range:	z_{0B}	= 383 mm

A Gaussian beam with the same waist radius w_0 has a Rayleigh range of 5.69 mm.
This example indicates, that it is possible to get a considerable fraction of the total power into the central peak of the Bessel beam.

2.11 Beam Propagation in Anisotropic Crystals

In the previous sections of this chapter the diffraction and the propagation of scalar fields was discussed. This is correct for isotropic media, but not for birefringent media and anisotropic crystals. Now the situation becomes more complicated, and we have to deal with uni- and biaxial crystals. The physics of anisotropic crystals are well treated in the textbooks of Optics [1.2,1.6]. Here, only a very simple example will be discussed qualitatively.

A uniaxial crystal is characterized by two principal refractive indices, n_o and n_{eo}. They depend on the orientation of the polarization with respect to the c-axis of the crystal, as shown in Fig. 2.48. Inside the crystal the beam is characterized by the field $\boldsymbol{E}$ and the electric displacement $\boldsymbol{D}$. In the most general case, the correlation between these vectors is determined by the dieletric tensor ϵ:

$$\begin{pmatrix} D_x \\ D_y \\ D_z \end{pmatrix} = \epsilon_0 \, \epsilon \, \boldsymbol{E} = \epsilon_0 \begin{pmatrix} \epsilon_{xx} & \epsilon_{xy} & \epsilon_{xz} \\ \epsilon_{yx} & \epsilon_{yy} & \epsilon_{yz} \\ \epsilon_{zx} & \epsilon_{zy} & \epsilon_{zz} \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} \tag{2.216}$$

For loss-free crystals, the dielectric tensor is symmetric and can be diagonalized by choosing the three principal axes for which $\boldsymbol{D}$ and $\boldsymbol{E}$ are parallel. For a uniaxial crystal, (2.216) is then reduced to:

$$\begin{pmatrix} D_x \\ D_y \\ D_z \end{pmatrix} = \epsilon_0 \begin{pmatrix} n_o^2 E_x \\ n_o^2 E_y \\ n_{oe}^2 E_z \end{pmatrix} \tag{2.217}$$

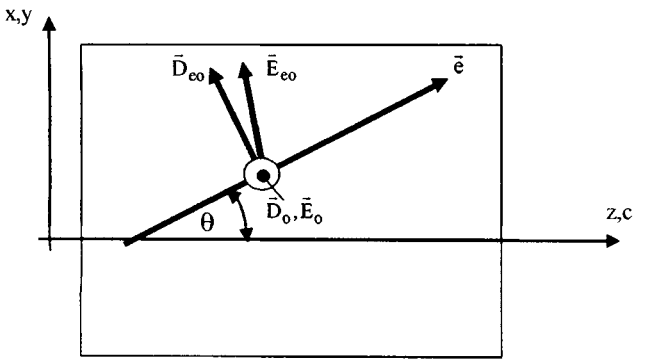


Fig. 2.48 Ordinary and extraordinary beams in a uniaxial crystal.

For uniaxial crystals, the wave equation delivers two polarized waves, which propagate inside the crystal without changing the polarization:

- the ordinary beam $\boldsymbol{D}_o$, $\boldsymbol{E}_o$ with $\boldsymbol{D}_o$ perpendicular to the c-axis of the crystal and perpendicular to the propagation vector $\boldsymbol{e}$ of the phase fronts.

- the extraordinary beam $\boldsymbol{D}_{eo}$, $\boldsymbol{E}_{eo}$ with $\boldsymbol{D}_{eo}$ perpendicular to $\boldsymbol{D}_o$ and $\boldsymbol{e}$. $\boldsymbol{D}_{eo}$ and $\boldsymbol{E}_{eo}$ are in general not collinear because of the anisotropy.

The two beams are described by:

$$\boldsymbol{D}_o = \boldsymbol{D}_{o,0} \exp[i(\omega t - n_o k_o \boldsymbol{er})] \tag{2.218}$$

$$\boldsymbol{D}_{eo} = \boldsymbol{D}_{eo,0} \exp[i(\omega t - n_{eo}(\theta) k_o \boldsymbol{er})] \tag{2.219}$$

where the refractive index n_{eo} of the extraordinary beam depends on the angle θ:

$$\frac{1}{n_{eo}^2(\theta)} = \frac{\cos^2\theta}{n_o^2} + \frac{\sin^2\theta}{n_{eo}^2} \tag{2.220}$$

The ordinary and the extraordinary beam propagate with different phase velocities.

Now let us discuss the situation of a Gaussian beam, propagating in z-direction, parallel to the crystal axis c, and polarized in x direction. The $\boldsymbol{E}$ field in the y-z plane (shaded in Fig. 2.49) is perpendicular to the c-axis and to the propagation vector $\boldsymbol{e}$. It is an ordinary beam $\boldsymbol{E}_o$, $\boldsymbol{D}_o$. But the field in the x-z-plane is perpendicular to $\boldsymbol{D}_o$ and $\boldsymbol{e}$, and therefore an extraordinary beam. Inside the crystal ordinary and extraordinary parts of the Gaussian beam propagate with different phase velocities.

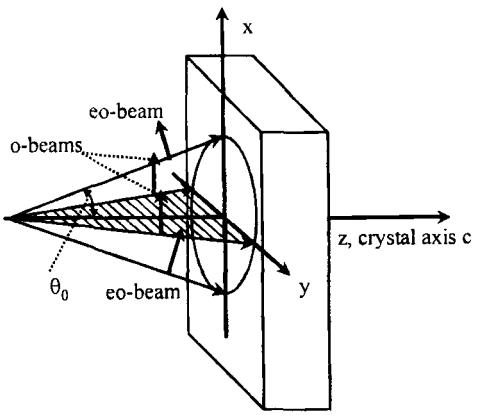


Fig. 2.49 The Gaussian beam incident on a uniaxial crystal consists of ordinary and extraordinary parts [1.104] (© OSA 2002).

These phase shifts, depending on the transverse coordinates, produce interference pattern and distort the Gaussian beam. Additionally the x-polarized beam is partly converted into y-polarization. The theory is straight forward, but tiring. The plane wave method in the paraxial approach is used as discussed in Sec. 2.9 [1.103,1.104]. Some results are presented in Figs. 2.50 and 2.51. The theory can be extended to higher order modes and modes of cylindrical symmetry [1.105].

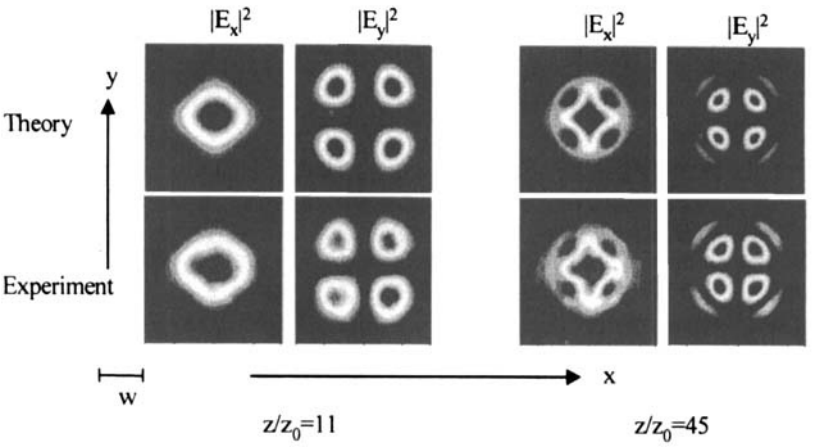


Fig. 2.50 The distortion of an x-polarized TEM_{00} beam propagating in a Calcite-crystal for two different propagation distances z/z_0 with z_0 being the Rayleigh range. The transverse coordinates are normalized to the beam radius *w(z)* [1.104] (© OSA 2002).

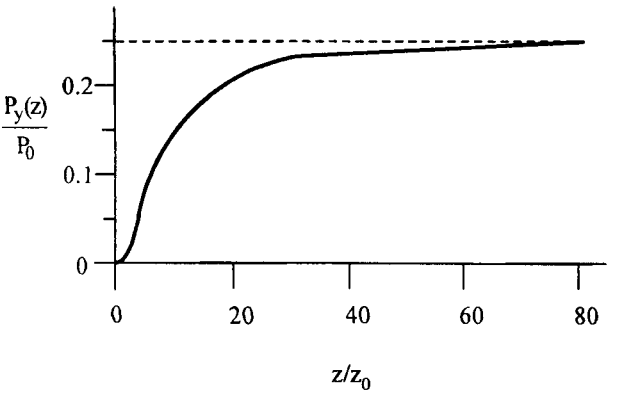


Fig. 2.51 The power $P_y(z)$ of the y-component versus the normalized propagation length z/z_0. In the limit of $z >> z_0$, 25% of the total power is converted into the y-component [1.104] (© OSA 2002).