

3

Propagation and focusing of optical fields

光场的传播与聚焦

In this chapter we use the angular spectrum representation outlined in Section 2.12 to discuss field distributions in strongly focused laser beams. The same formalism is applied to understand how the fields in a given reference plane are mapped to the far-field. The theory is relevant for the understanding of confocal and multiphoton microscopy, single-emitter experiments, and the understanding of resolution limits. It also defines the framework for different topics to be discussed in later chapters.

3.1 Field propagators

In Section 2.12 we have established that, in a homogeneous space, the spatial spectrum $\hat{\mathbf{E}}$ of an optical field \mathbf{E} in a plane $z = \text{const.}$ (image plane) is uniquely defined by the spatial spectrum in a different plane $z = 0$ (object plane) according to the linear relationship

$$\hat{\mathbf{E}}(k_x, k_y; z) = \hat{H}(k_x, k_y; z) \hat{\mathbf{E}}(k_x, k_y; 0) . \quad (3.1)$$

where \hat{H} is the so-called *propagator in reciprocal space*

$$\hat{H}(k_x, k_y; z) = e^{\pm i k_z z} , \quad (3.2)$$

also referred to as the *optical transfer function* (OTF) in free space. Remember that the longitudinal wavenumber is a function of the transverse wavenumber, i.e. $k_z = [k^2 - (k_x^2 + k_y^2)]^{1/2}$, where $k = n k_0 = n \omega / c = n 2\pi / \lambda$. The \pm sign indicates that the field can propagate in positive and/or negative z direction. Equation (3.1) can be interpreted in terms of linear response theory: $\hat{\mathbf{E}}(k_x, k_y; 0)$ is the input, \hat{H} is a filter function, and $\hat{\mathbf{E}}(k_x, k_y; z)$ is the output. The filter function describes the propagation of an arbitrary spectrum through space. \hat{H} can also be regarded as the response function because it describes the field at z due to a point source at $z = 0$. In this sense, it is directly related to the Green's function $\vec{\mathbf{G}}$.

The filter \hat{H} is an oscillating function for $(k_x^2 + k_y^2) < k^2$ and an exponentially decreasing function for $(k_x^2 + k_y^2) > k^2$. Thus, if the image plane is sufficiently separated from the object plane, the contribution of the decaying parts (evanescent waves) is zero and the integration can be reduced to the circular area $(k_x^2 + k_y^2) \leq k^2$. In other words, the image at z is a *low pass filtered* representation of the original field at $z = 0$. The spatial frequencies $(k_x^2 + k_y^2) > k^2$ of the original field are filtered out during propagation and the information on high spatial variations gets lost. Hence, there is always a loss of information on propagating from near- to far-field and only structures with lateral dimensions larger than

$$\Delta x \approx \frac{1}{k} = \frac{\lambda}{2\pi n} \quad (3.3)$$

can be imaged with sufficient accuracy. Here, n is the index of refraction. This equation is qualitative and we will provide a more detailed discussion in Chapter 4. In general, higher resolution can be obtained by a higher index of refraction of the embodying system (substrate, lenses, etc.) or by shorter wavelengths. Theoretically, resolutions down to a few nanometers can be achieved by using far-ultraviolet radiation or X-rays. However, X-rays do cause damage to many samples. Furthermore, they are limited by the poor quality of lenses and do not provide the wealth of information of optical frequencies. The central idea of *near-field optics* is to increase the bandwidth of spatial frequencies by retaining the evanescent components of the source fields.

Let us now determine how the fields themselves evolve. For this purpose we denote the transverse coordinates in the object plane at $z = 0$ as (x', y') and in the image plane at $z = \text{const.}$ as (x, y) . The fields in the image plane are described by the angular spectrum (2.111). We just have to express the Fourier spectrum $\hat{\mathbf{E}}(k_x, k_y; 0)$ in terms of the fields in the object plane. Similarly to Eq. (2.105) this Fourier spectrum can be represented as

$$\hat{\mathbf{E}}(k_x, k_y; 0) = \frac{1}{4\pi^2} \iint_{-\infty}^{\infty} \mathbf{E}(x', y', 0) e^{-i[k_x x' + k_y y']} dx' dy'. \quad (3.4)$$

After inserting into Eq. (2.111) we find the following expression for the field \mathbf{E} in the image plane $z = \text{const.}$

$$\begin{aligned} \mathbf{E}(x, y, z) &= \frac{1}{4\pi^2} \iint_{-\infty}^{\infty} \mathbf{E}(x', y'; 0) \iint_{-\infty}^{\infty} e^{i[k_x (x-x') + k_y (y-y') \pm k_z z]} dx' dy' dk_x dk_y \\ &= \mathbf{E}(x, y; 0) * H(x, y; z). \end{aligned} \quad (3.5)$$

This equation describes an invariant filter with the following impulse response (*propagator in direct space*)

$$H(x, y; z) = \iint_{-\infty}^{\infty} e^{i[k_x x + k_y y \pm k_z z]} dk_x dk_y. \quad (3.6)$$

H is simply the inverse Fourier transform of the propagator in reciprocal space \hat{H} (3.2). The field at $z = \text{const.}$ is represented by the convolution of H with the field at $z = 0$.

3.2 Paraxial approximation of optical fields

In many optical problems the light fields propagate along a certain direction z and spread out only slowly in the transverse direction. Examples are laser beam propagation or optical waveguide applications. In these examples the wavevectors $\mathbf{k} = (k_x, k_y, k_z)$ in the angular spectrum representation are almost parallel to the z -axis and the transverse wavenumbers (k_x, k_y) are small compared to k . We can then expand the square root of Eq. (2.109) in a series as

$$k_z = k \sqrt{1 - (k_x^2 + k_y^2)/k^2} \approx k - \frac{(k_x^2 + k_y^2)}{2k}. \quad (3.7)$$

This approximation is called the paraxial approximation and it considerably simplifies the analytical integration of the Fourier integrals. In the following we shall apply the paraxial approximation to find a description for weakly focused laser beams.

3.2.1 Gaussian laser beams

We consider a fundamental laser beam with a linearly polarized, Gaussian field distribution in the beam waist

$$\mathbf{E}(x', y', 0) = \mathbf{E}_0 e^{-\frac{x'^2 + y'^2}{w_0^2}}, \quad (3.8)$$

where \mathbf{E}_0 is a constant field vector in the transverse (x, y) plane. We have chosen $z = 0$ at the beam waist. The parameter w_0 denotes the beam waist radius. We can calculate the spatial Fourier spectrum at $z = 0$ as¹

$$\begin{aligned} \hat{\mathbf{E}}(k_x, k_y; 0) &= \frac{1}{4\pi^2} \iint_{-\infty}^{\infty} \mathbf{E}_0 e^{-\frac{x'^2 + y'^2}{w_0^2}} e^{-i[k_x x' + k_y y']} dx' dy' \\ &= \mathbf{E}_0 \frac{w_0^2}{4\pi} e^{-(k_x^2 + k_y^2) \frac{w_0^2}{4}}, \end{aligned} \quad (3.9)$$

¹ $\int_{-\infty}^{\infty} \exp(-ax^2 + ibx) dx = \sqrt{\pi/a} \exp(-b^2/4a)$ and $\int_{-\infty}^{\infty} x \exp(-ax^2 + ibx) dx = ib\sqrt{\pi} \exp(-b^2/4a)/$
($2a^{3/2}$)

which is again a Gaussian function. We now insert this spectrum into the angular spectrum representation Eq. (2.111) and replace k_z by its paraxial expression in Eq. (3.7)

$$\mathbf{E}(x, y, z) = \mathbf{E}_0 \frac{w_0^2}{4\pi} e^{ikz} \iint_{-\infty}^{\infty} e^{-(k_x^2 + k_y^2) \left(\frac{w_0^2}{4} + \frac{iz}{2k} \right)} e^{i[k_x x + k_y y]} dk_x dk_y, \quad (3.10)$$

This equation can be integrated and gives as a result the familiar paraxial representation of a Gaussian beam

$$\mathbf{E}(x, y, z) = \frac{\mathbf{E}_0 e^{ikz}}{(1 + 2iz/kw_0^2)} e^{-\frac{(x^2 + y^2)}{w_0^2} \frac{1}{(1 + 2iz/kw_0^2)}}. \quad (3.11)$$

To get a better feeling for a paraxial Gaussian beam we set $\rho^2 = x^2 + y^2$, define a new parameter z_0 as

$$z_0 = \frac{k w_0^2}{2}, \quad (3.12)$$

and rewrite Eq. (3.11) as

$$\mathbf{E}(\rho, z) = \mathbf{E}_0 \frac{w_0}{w(z)} e^{-\frac{\rho^2}{w^2(z)}} e^{i[kz - \eta(z) + k\rho^2/2R(z)]} \quad (3.13)$$

with the following abbreviations

$$\begin{aligned} w(z) &= w_0(1 + z^2/z_0^2)^{1/2} && \text{beam radius} \\ R(z) &= z(1 + z_0^2/z^2) && \text{wavefront radius} \\ \eta(z) &= \arctan z/z_0 && \text{phase correction} \end{aligned} \quad (3.14)$$

The transverse size of the beam is usually defined by the value of $\rho = \sqrt{x^2 + y^2}$ for which the electric field amplitude has decreased to a value of $1/e$ of its center value

$$|\mathbf{E}(x, y, z)| / |\mathbf{E}(0, 0, z)| = 1/e. \quad (3.15)$$

It can be shown that the surface defined by this equation is a hyperboloid whose asymptotes enclose an angle

$$\theta = \frac{2}{k w_0} \quad (3.16)$$

with the z -axis. From this equation we can directly find the correspondence between the numerical aperture ($\text{NA} = n \sin \theta$) and the beam angle as $\text{NA} \approx 2n/kw_0$. Here we used the fact that in the paraxial approximation, θ is restricted to small beam angles. Another property of the paraxial Gaussian beam is that close

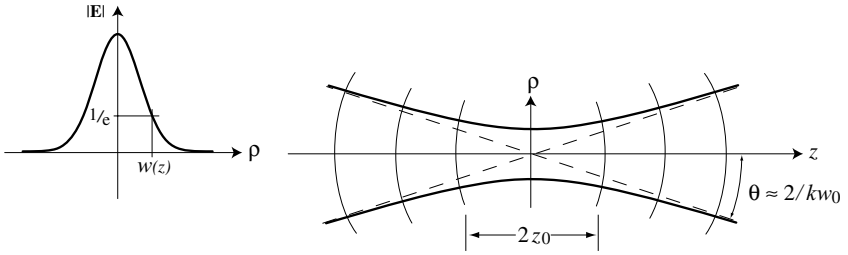


Figure 3.1 Illustration and main characteristics of a paraxial Gaussian beam. The beam has a Gaussian field distribution in the transverse plane. The surfaces of constant field strength form a hyperboloid along the z -axis.

to the focus, the beam stays roughly collimated over a distance $2z_0$. z_0 is called the *Rayleigh range* and denotes the distance from the beam waist to where the beam radius has increased by a factor of $\sqrt{2}$. It is important to notice that along the z -axis ($\rho = 0$) the phases of the beam deviate from those of a plane wave. If at $z \rightarrow -\infty$ the beam was in phase with a reference plane wave, then at $z \rightarrow +\infty$ the beam will be exactly out of phase with the reference wave. This phase shift is called *Gouy phase shift* and has practical implications in nonlinear confocal microscopy [1]. The 180° phase change happens gradually as the beam propagates through its focus. The phase variation is described by the factor $\eta(z)$ in Eq. (3.14). The tighter the focus the faster the phase variation will be.

A qualitative picture of a paraxial Gaussian beam and some of its characteristics are shown in Fig. 3.1 and more detailed descriptions can be found in other textbooks [2, 3]. It is important to notice that once the paraxial approximation is introduced, the field \mathbf{E} fulfills no longer Maxwell's equations. The error becomes larger the smaller the beam waist radius w_0 is. When w_0 becomes comparable to the reduced wavelength λ/n we have to include higher-order terms in the expansion of k_z in Eq. (3.7). However, the series expansion converges very badly for strongly focused beams and one needs to find a more accurate description. We shall return to this topic at a later stage.

Another important aspect of Gaussian beams is that they do not exist, no matter how rigorous the theory that describes them! The reason is that a Gaussian beam profile demands a Gaussian spectrum. However, the Gaussian spectrum is infinite and contains evanescent components that are not available in a realistic situation. Thus, a Gaussian beam must always be regarded as an approximation. The tighter the focus, the broader the Gaussian spectrum and the more contradictory the Gaussian beam profile will be. Hence, it actually does not make much sense to include higher-order corrections to the paraxial approximation.

3.2.2 Higher-order laser modes

A laser beam can exist in different transverse modes. It is the laser cavity that determines which type of transverse mode is emitted. The most commonly encountered higher beam modes are Hermite–Gaussian and Laguerre–Gaussian beams. The former are generated in cavities with rectangular end mirrors whereas the latter are observed in cavities with circular end mirrors. In the transverse plane, the fields of these modes extend over larger distances and have sign variations in the phase.

Since the fundamental Gaussian mode is a solution of a linear homogeneous partial differential equation, namely the Helmholtz equation, any combinations of spatial derivatives of the fundamental mode are also solutions to the same differential equation. Zauderer [4] pointed out that Hermite–Gaussian modes \mathbf{E}_{nm}^H can be generated from the fundamental mode \mathbf{E} according to

$$\mathbf{E}_{nm}^H(x, y, z) = w_0^{n+m} \frac{\partial^n}{\partial x^n} \frac{\partial^m}{\partial y^m} \mathbf{E}(x, y, z), \quad (3.17)$$

where n, m denote the order and degree of the beam. Laguerre–Gaussian modes $\mathbf{E}_{n,m}^L$ are derived in a similar way as

$$\mathbf{E}_{nm}^L(x, y, z) = k^n w_0^{2n+m} e^{ikz} \frac{\partial^n}{\partial z^n} \left(\frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right)^m \{ \mathbf{E}(x, y, z) e^{-ikz} \}. \quad (3.18)$$

Thus, any higher-order modes can be generated by simply applying Eqs. (3.17) and (3.18). It can be shown that Laguerre–Gaussian modes can be generated as a superposition of a finite number of Hermite–Gaussian modes and vice versa. The two sets of modes are therefore not independent. Note that the parameter w_0 only represents the beam waist for the Gaussian beam and that for higher-order modes the amplitude E_0 does not correspond to the field at the focal point. Figure 3.2 shows the fields in the focal plane ($z = 0$) for the first four Hermite–Gaussian modes. As indicated by the arrows, the polarizations of the individual maxima are either in phase or 180° out of phase with each other.

The commonly encountered doughnut modes with a circular intensity profile can be described by a superposition of Hermite–Gaussian or Laguerre–Gaussian modes. Linearly polarized doughnuts are simply defined by the fields \mathbf{E}_{01}^L or \mathbf{E}_{11}^L . An azimuthally polarized doughnut mode is a superposition of two perpendicularly polarized \mathbf{E}_{01}^H fields and a radially polarized doughnut mode is a superposition of two perpendicularly polarized \mathbf{E}_{10}^H fields.

3.2.3 Longitudinal fields in the focal region

The paraxial Gaussian beam is a transverse electromagnetic (TEM) beam, i.e. it is assumed that the electric and magnetic fields are always transverse to the

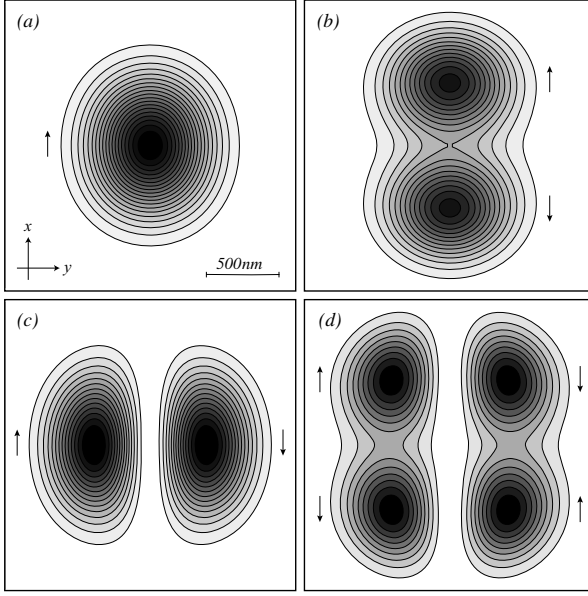


Figure 3.2 Intensity ($|E|^2$) in the focal plane ($z = 0$) of the first four Hermite-Gaussian modes. (a) (00) mode (Gaussian mode), (b) (10) mode, (c) (01) mode, and (d) (11) mode. The wavelength and beam angle are $\lambda = 800$ nm and $\theta = 28.65^\circ$, respectively. The arrows indicate the polarization direction of the individual lobes. A linear scaling is used between contour lines.

propagation direction. However, in free space the only true TEM solutions are infinitely extended fields such as plane waves. Therefore, even a Gaussian beam must possess field components polarized in the direction of propagation. In order to estimate these longitudinal fields we apply the divergence condition $\nabla \cdot \mathbf{E} = 0$ to the x -polarized Gaussian beam, i.e.

$$E_z = - \int \left[\frac{\partial}{\partial x} E_x \right] dz. \quad (3.19)$$

E_z can be derived using the angular spectrum representation of the paraxial Gaussian beam Eq. (3.10). In the focal plane $z = 0$ we obtain

$$E_z(x, y, 0) = -i \frac{2x}{kw_0^2} E_x(x, y, 0), \quad (3.20)$$

where E_x corresponds to the Gaussian beam profile defined in Eq. (3.8). The prefactor shows that the longitudinal field is 90° out of phase with respect to the transverse field and that it is zero on the optical axis. Its magnitude depends on the tightness of the focus. Figures 3.3 and 3.4 show the calculated total

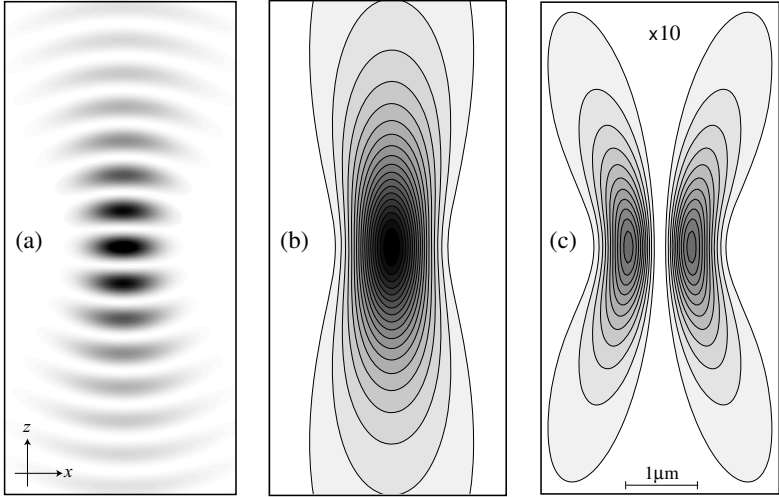


Figure 3.3 Fields of the Gaussian beam depicted in the polarization plane (x, z) . The wavelength and beam angle are $\lambda = 800$ nm and $\theta = 28.65^\circ$, respectively. (a) Time dependent power density; (b) total electric field intensity ($|\mathbf{E}|^2$); (c) longitudinal electric field intensity ($|\mathbf{E}_z|^2$). A linear scaling is used between contour lines.

and transverse electric field distribution for the Gaussian beam and the Hermite–Gaussian (10) beam, respectively. While the longitudinal electric field of the fundamental Gaussian beam is always zero on the optical axis it shows two lobes to the sides of the optical axis. Displayed on a cross-section through the beam waist, the two lobes are aligned along the polarization direction. The longitudinal electric field of the Hermite–Gaussian (10) mode, on the other hand, has its maximum at the beam focus with a much larger field strength. This longitudinal field qualitatively follows from the 180° phase difference and the polarization of the two corresponding field maxima in Fig. 3.2, since the superposition of two similarly polarized plane waves propagating at angles $\pm\varphi$ to the z -axis with 180° phase difference also leads to a longitudinal field component. It has been proposed to use the longitudinal fields of the Hermite–Gaussian (10) mode to accelerate charged particles along the beam axis in linear particle accelerators [5]. The longitudinal (10) field has also been applied to image the spatial orientation of molecular transition dipoles [6, 7]. In general, the (10) mode is important for all experiments that require the availability of a longitudinal field component. We shall see in Section 3.6 that the longitudinal field strength of a strongly focused higher-order laser beam can even exceed the transverse field strength.

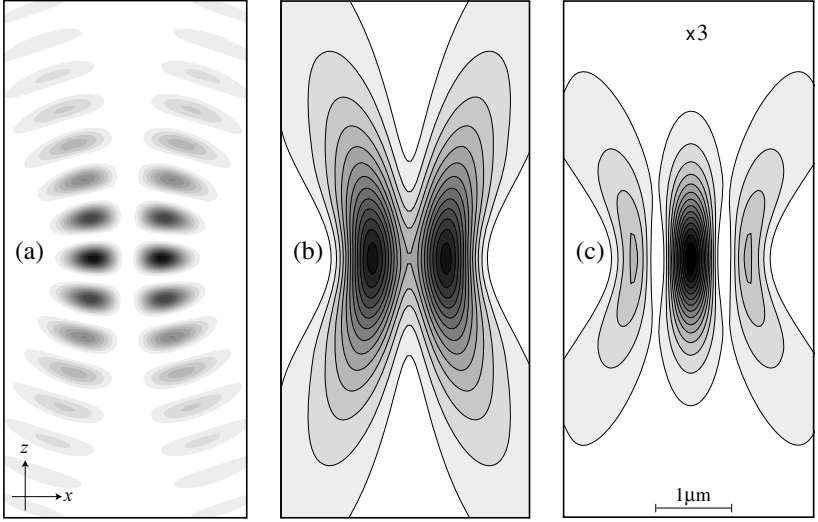


Figure 3.4 Fields of the Hermite–Gaussian (10) mode. Same scaling and definitions as in Fig. 3.3.

3.3 Polarized electric and polarized magnetic fields

If we send an optical beam through a polarizer, we eliminate one of the two transverse field components. The transmitted field is then called *polarized electric*.

In fact, any propagating optical field can be split into a polarized electric (PE) and a polarized magnetic (PM) field

$$\mathbf{E} = \mathbf{E}^{\text{PE}} + \mathbf{E}^{\text{PM}} . \quad (3.21)$$

For a PE field, the electric field is linearly polarized when projected on the transverse plane. Similarly, for a PM field the magnetic field is linearly polarized when projected on the transverse plane. Let us first consider a PE field for which we can choose $\mathbf{E}^{\text{PE}} = (E_x, 0, E_z)$. Requiring that the field is divergence free ($\nabla \cdot \mathbf{E}^{\text{PE}} = 0$) we find that

$$\hat{E}_z(k_x, k_y; 0) = -\frac{k_x}{k_z} \hat{E}_x(k_x, k_y; 0) , \quad (3.22)$$

which allows us to express the fields $\mathbf{E}^{\text{PE}}, \mathbf{H}^{\text{PE}}$ in the form

$$\mathbf{E}^{\text{PE}}(x, y, z) = \iint_{-\infty}^{\infty} \hat{E}_x(k_x, k_y; 0) \frac{1}{k_z} [k_z \mathbf{n}_x - k_x \mathbf{n}_z] e^{i[k_x x + k_y y \pm k_z z]} dk_x dk_y , \quad (3.23)$$

$$\mathbf{H}^{\text{PE}}(x, y, z) = Z_{\mu\epsilon}^{-1} \iint_{-\infty}^{\infty} \hat{E}_x(k_x, k_y; 0) \frac{1}{kk_z} [-k_x k_y \mathbf{n}_x + (k_x^2 + k_z^2) \mathbf{n}_y - k_y k_z \mathbf{n}_z] e^{i[k_x x + k_y y \pm k_z z]} dk_x dk_y, \quad (3.24)$$

where \mathbf{n}_x , \mathbf{n}_y , \mathbf{n}_z are unit vectors along the x , y , z axes. To derive \mathbf{H}^{PE} we used the relations in Eq. (2.113).

To derive the corresponding PM fields we require that $\mathbf{H}^{\text{PM}} = (0, H_y, H_z)$. After following the same procedure as before one finds that in the PM solution the expressions for the electric and magnetic fields are simply interchanged

$$\mathbf{E}^{\text{PM}}(x, y, z) = Z_{\mu\epsilon} \iint_{-\infty}^{\infty} \hat{H}_y(k_x, k_y; 0) \frac{1}{kk_z} [(k_y^2 + k_z^2) \mathbf{n}_x - k_x k_y \mathbf{n}_y + k_x k_z \mathbf{n}_z] e^{i[k_x x + k_y y \pm k_z z]} dk_x dk_y, \quad (3.25)$$

$$\mathbf{H}^{\text{PM}}(x, y, z) = \iint_{-\infty}^{\infty} \hat{H}_y(k_x, k_y; 0) \frac{1}{k_z} [k_z \mathbf{n}_y - k_y \mathbf{n}_z] e^{i[k_x x + k_y y \pm k_z z]} dk_x dk_y. \quad (3.26)$$

It is straightforward to demonstrate that in the paraxial limit the PE and PM solutions are identical. In this case they become identical with a TEM solution.

The decomposition of an arbitrary optical field into a PE and a PM field has been achieved by setting one transverse field component to zero. The procedure is similar to the commonly encountered decomposition into transverse electric (TE) and transverse magnetic (TM) fields for which one longitudinal field component is set to zero (see Problem 3.2).

3.4 Far-fields in the angular spectrum representation

In this section we will derive the important result that Fourier Optics and Geometrical Optics naturally emerge from the angular spectrum representation.

Consider a particular (localized) field distribution in the plane $z = 0$. The angular spectrum representation tells us how this field propagates and how it is mapped onto other planes $z = z_0$. Here, we ask what the field will be in a very remote plane. Vice versa, we can ask what field will result when we focus a particular far-field onto an image plane. Let us start with the familiar angular spectrum representation of an optical field

$$\mathbf{E}(x, y, z) = \iint_{-\infty}^{\infty} \hat{\mathbf{E}}(k_x, k_y; 0) e^{i[k_x x + k_y y \pm k_z z]} dk_x dk_y. \quad (3.27)$$

We are interested in the asymptotic far-zone approximation of this field, i.e. in the evaluation of the field in a point $\mathbf{r} = \mathbf{r}_\infty$ at infinite distance from the object plane.

The dimensionless unit vector \mathbf{s} in the direction of \mathbf{r}_∞ is given by

$$\mathbf{s} = (s_x, s_y, s_z) = \left(\frac{x}{r}, \frac{y}{r}, \frac{z}{r} \right), \quad (3.28)$$

where $r = (x^2 + y^2 + z^2)^{1/2}$ is the distance of \mathbf{r}_∞ from the origin. To calculate the far-field \mathbf{E}_∞ we require that $r \rightarrow \infty$ and rewrite Eq. (3.27) as

$$\mathbf{E}_\infty(s_x, s_y, s_z) = \lim_{kr \rightarrow \infty} \iint_{(k_x^2 + k_y^2) \leq k^2} \hat{\mathbf{E}}(k_x, k_y; 0) e^{ikr [\frac{k_x}{k} s_x + \frac{k_y}{k} s_y \pm \frac{k_z}{k} s_z]} dk_x dk_y. \quad (3.29)$$

Because of their exponential decay, evanescent waves do not contribute to the fields at infinity. We therefore rejected their contribution and reduced the integration range to $(k_x^2 + k_y^2) \leq k^2$. The asymptotic behavior of the double integral as $kr \rightarrow \infty$ can be evaluated by the method of *stationary phase*. For a clear outline of this method we refer the interested reader to Chapter 3.3 of Ref. [3]. Without going into details, the result of Eq. (3.29) can be expressed as

$$\mathbf{E}_\infty(s_x, s_y, s_z) = -2\pi i k s_z \hat{\mathbf{E}}(k s_x, k s_y; 0) \frac{e^{ikr}}{r}. \quad (3.30)$$

This equation tells us that the far-fields are entirely defined by the Fourier spectrum of the fields $\hat{\mathbf{E}}(k_x, k_y; 0)$ in the object plane if we replace $k_x \rightarrow k s_x$ and $k_y \rightarrow k s_y$. This simply means that the unit vector \mathbf{s} fulfills

$$\mathbf{s} = (s_x, s_y, s_z) = \left(\frac{k_x}{k}, \frac{k_y}{k}, \frac{k_z}{k} \right), \quad (3.31)$$

which implies that only one plane wave with the wavevector $\mathbf{k} = (k_x, k_y, k_z)$ of the angular spectrum at $z = 0$ contributes to the far-field at a point located in the direction of the unit vector \mathbf{s} . The effect of all other plane waves is cancelled by destructive interference. This beautiful result allows us to treat the field in the far-zone as a collection of rays with each ray being characterized by a particular plane wave of the original angular spectrum representation (Geometrical optics). Combining Eqs. (3.30) and (3.31) we can express the Fourier spectrum $\hat{\mathbf{E}}$ in terms of the far-field as

$$\hat{\mathbf{E}}(k_x, k_y; 0) = \frac{ir e^{-ikr}}{2\pi k_z} \mathbf{E}_\infty(k_x, k_y), \quad (3.32)$$

keeping in mind that the vector \mathbf{S} is entirely defined by k_x, k_y . This expression can be substituted into the angular spectrum representation (Eq. 3.27) as

$$\mathbf{E}(x, y, z) = \frac{ir e^{-ikr}}{2\pi} \iint_{(k_x^2 + k_y^2) \leq k^2} \mathbf{E}_\infty(k_x, k_y) e^{i[k_x x + k_y y \pm k_z z]} \frac{1}{k_z} dk_x dk_y. \quad (3.33)$$

Thus, as long as evanescent fields are not part of our system then the field \mathbf{E} and its far-field \mathbf{E}_∞ form essentially a Fourier transform pair at $z = 0$. The only deviation is given by the factor $1/k_z$. In the approximation $k_z \approx k$, the two fields form a perfect Fourier transform pair. This is the limit of *Fourier Optics*.

As an example consider the diffraction at a rectangular aperture with sides $2L_x$ and $2L_y$ in an infinitely thin conducting screen which we choose to be our object plane ($z = 0$). A plane wave illuminates the aperture at normal incidence from the back. For simplicity we assume that the field in the object plane has a constant field amplitude \mathbf{E}_0 whereas the screen blocks all the field outside of the aperture. The Fourier spectrum at $z = 0$ is then

$$\begin{aligned}\hat{\mathbf{E}}(k_x, k_y; 0) &= \frac{\mathbf{E}_0}{4\pi^2} \int_{-L_y}^{+L_y} \int_{-L_x}^{+L_x} e^{-i[k_x x' + k_y y']} dx' dy' \\ &= \mathbf{E}_0 \frac{L_x L_y}{\pi^2} \frac{\sin(k_x L_x)}{k_x L_x} \frac{\sin(k_y L_y)}{k_y L_y},\end{aligned}\quad (3.34)$$

With Eq. (3.30) we now determine the far-field as

$$\mathbf{E}_\infty(s_x, s_y, s_z) = -iks_z \mathbf{E}_0 \frac{2L_x L_y}{\pi} \frac{\sin(ks_x L_x)}{ks_x L_x} \frac{\sin(ks_y L_y)}{ks_y L_y} \frac{e^{ikr}}{r}, \quad (3.35)$$

which, in the paraxial limit $k_z \approx k$, agrees with Fraunhofer diffraction.

Equation (3.30) is an important result. It links the near-fields of an optical problem with the corresponding far-fields. While in the near-field a rigorous description of fields is necessary, the far-fields are well approximated by the laws of Geometrical Optics.

3.5 Focusing of fields

The limit of classical light confinement is achieved with highly focused laser beams. Such beams are used in fluorescence spectroscopy to investigate molecular interactions in solutions and the kinetics of single molecules on interfaces [6]. Highly focused laser beams also play a key role in confocal microscopy and optical data storage, where resolutions on the order of $\lambda/4$ are achieved. In optical tweezers, focused laser beams are used to trap particles and to move and position them with high precision [8]. All these fields require a theoretical understanding of strongly focused light.

The fields of a focused laser beam are determined by the boundary conditions of the focusing optical element and the incident optical field. In this section we will study the focusing of a paraxial optical field by an aplanatic optical lens as shown in Fig. 3.5. In our theoretical treatment we will follow the theory established by Richards and Wolf [9, 10]. The fields near the optical lens can be formulated by the rules of Geometrical Optics. In this approximation the finiteness of the optical

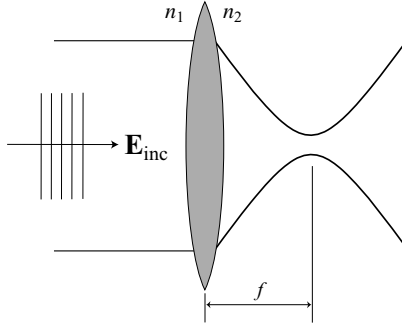


Figure 3.5 Focusing of a laser beam by an aplanatic lens.

wavelength is neglected ($k \rightarrow \infty$) and the energy is transported along light rays. The average energy density is propagated with the velocity $v = c/n$ in the direction perpendicular to the geometrical wavefronts. To describe an aplanatic lens we need two rules: (1) the sine condition and (2) the intensity law. These rules are illustrated in Fig. 3.6. The *sine condition* states that each optical ray which emerges from or converges to the focus F of an aplanatic optical system intersects its conjugate ray on a sphere of radius f (Gaussian reference sphere), where f is the focal length of the lens. By conjugate ray, one understands the refracted or incident ray that propagates parallel to the optical axis. The distance h between the optical axis and the conjugate ray is given by

$$h = f \sin(\theta) , \quad (3.36)$$

θ being the divergence angle of the conjugate ray. Thus, the sine condition is a prescription for the refraction of optical rays at the aplanatic optical element. The *intensity law* is nothing but a statement of energy conservation: the energy flux along each ray must remain constant. As a consequence, the electric field strength of a spherical wave has to scale as $1/r$, r being the distance from the origin. The intensity law ensures that the energy incident on the aplanatic lens equals the energy that leaves the lens. We know that the power transported by a ray is $P = (1/2)Z_{\mu\epsilon}^{-1/2}|\mathbf{E}|^2 dA$, where $Z_{\mu\epsilon}$ is the wave impedance and dA is an infinitesimal cross-section perpendicular to the ray propagation. Thus, as indicated in the figure, the fields before and after refraction must fulfill

$$|\mathbf{E}_2| = |\mathbf{E}_1| \sqrt{\frac{n_1}{n_2}} \sqrt{\frac{\mu_2}{\mu_1}} \cos^{1/2}\theta . \quad (3.37)$$

Since in practically all media the magnetic permeability at optical frequencies is equal to one ($\mu = 1$), we will drop the term $\sqrt{\mu_2/\mu_1}$ for the sake of more convenient notation.

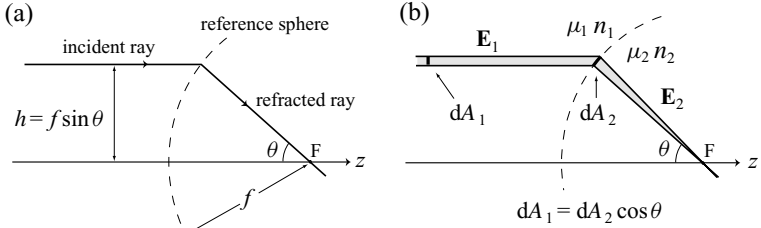


Figure 3.6 (a) Sine condition of Geometrical Optics. The refraction of light rays at an aplanatic lens is determined by a spherical surface with radius f . (b) Intensity law of Geometrical Optics. The energy carried along a ray must stay constant.

Using the sine condition, our optical system can be represented as shown in Fig. 3.7. The incident light rays are refracted by the reference sphere of radius f . We denote an arbitrary point on the surface of the reference sphere as $(x_\infty, y_\infty, z_\infty)$ and an arbitrary field point near the focus by (x, y, z) . The two points are also represented by the spherical coordinates (f, θ, ϕ) and (r, ϑ, ϕ) , respectively.

To describe refraction of the incident rays at the reference sphere we introduce the unit vectors \mathbf{n}_ρ , \mathbf{n}_ϕ , and \mathbf{n}_θ , as shown in Fig. 3.7. \mathbf{n}_ρ and \mathbf{n}_ϕ are the unit vectors of a cylindrical coordinate system, whereas \mathbf{n}_θ together with \mathbf{n}_ϕ represent unit vectors of a spherical coordinate system. We recognize that the reference sphere transforms a cylindrical coordinate system (incoming beam) into a spherical coordinate system (focused beam). Refraction at the reference sphere is most conveniently calculated by splitting the incident vector \mathbf{E}_{inc} into two components denoted as $\mathbf{E}_{\text{inc}}^{(s)}$ and $\mathbf{E}_{\text{inc}}^{(p)}$. The indices (s) and (p) stand for s-polarization and p-polarization, respectively. In terms of the unit vectors we can express the two fields as

$$\mathbf{E}_{\text{inc}}^{(s)} = [\mathbf{E}_{\text{inc}} \cdot \mathbf{n}_\phi] \mathbf{n}_\phi, \quad \mathbf{E}_{\text{inc}}^{(p)} = [\mathbf{E}_{\text{inc}} \cdot \mathbf{n}_\rho] \mathbf{n}_\rho. \quad (3.38)$$

As shown in Fig. 3.7 these two fields refract at the spherical surface differently. While the unit vector \mathbf{n}_ϕ remains unaffected, the unit vector \mathbf{n}_ρ is mapped into \mathbf{n}_θ . Thus, the total refracted electric field, denoted by \mathbf{E}_∞ , can be expressed as

$$\mathbf{E}_\infty = \left[t^s [\mathbf{E}_{\text{inc}} \cdot \mathbf{n}_\phi] \mathbf{n}_\phi + t^p [\mathbf{E}_{\text{inc}} \cdot \mathbf{n}_\rho] \mathbf{n}_\theta \right] \sqrt{\frac{n_1}{n_2}} (\cos \theta)^{1/2}. \quad (3.39)$$

For each ray we have included the corresponding transmission coefficients t^s and t^p as defined in Eqs. (2.50). The factor outside the brackets is a consequence of the intensity law to ensure energy conservation. The subscript ∞ was added to indicate that the field is evaluated at a large distance from the focus $(x, y, z) = (0, 0, 0)$.

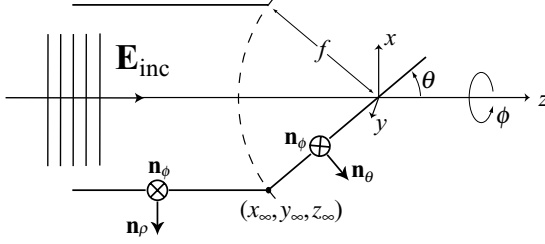


Figure 3.7 Geometrical representation of the aplanatic system and definition of coordinates.

The unit vectors \mathbf{n}_ρ , \mathbf{n}_ϕ , \mathbf{n}_θ can be expressed in terms of the Cartesian unit vectors \mathbf{n}_x , \mathbf{n}_y , \mathbf{n}_z using the spherical coordinates θ and ϕ defined in Fig. 3.7.

$$\mathbf{n}_\rho = \cos \phi \mathbf{n}_x + \sin \phi \mathbf{n}_y, \quad (3.40)$$

$$\mathbf{n}_\phi = -\sin \phi \mathbf{n}_x + \cos \phi \mathbf{n}_y, \quad (3.41)$$

$$\mathbf{n}_\theta = \cos \theta \cos \phi \mathbf{n}_x + \cos \theta \sin \phi \mathbf{n}_y - \sin \theta \mathbf{n}_z. \quad (3.42)$$

Inserting these vectors into Eq. (3.39) we obtain

$$\begin{aligned} \mathbf{E}_\infty(\theta, \phi) = & t^s(\theta) \left[\mathbf{E}_{\text{inc}}(\theta, \phi) \cdot \begin{pmatrix} -\sin \phi \\ \cos \phi \\ 0 \end{pmatrix} \right] \begin{pmatrix} -\sin \phi \\ \cos \phi \\ 0 \end{pmatrix} \sqrt{\frac{n_1}{n_2}} (\cos \theta)^{1/2} \\ & + t^p(\theta) \left[\mathbf{E}_{\text{inc}}(\theta, \phi) \cdot \begin{pmatrix} \cos \phi \\ \sin \phi \\ 0 \end{pmatrix} \right] \begin{pmatrix} \cos \phi \cos \theta \\ \sin \phi \cos \theta \\ -\sin \theta \end{pmatrix} \sqrt{\frac{n_1}{n_2}} (\cos \theta)^{1/2}, \end{aligned} \quad (3.43)$$

which is the field in Cartesian vector components just to the right of the reference sphere of the focusing lens. We can also express \mathbf{E}_∞ in terms of the spatial frequencies k_x and k_y by using the substitutions

$$k_x = k \sin \theta \cos \phi, \quad k_y = k \sin \theta \sin \phi, \quad k_z = k \cos \theta. \quad (3.44)$$

The resulting far-field on the reference sphere is then of the form $\mathbf{E}_\infty(k_x, k_y)$ and can be inserted into Eq. (3.33) to rigorously calculate the focal fields. Thus, the field \mathbf{E} near the focus of our lens is entirely determined by the far-field \mathbf{E}_∞ on the reference sphere. All rays propagate from the reference sphere towards the focus $(x, y, z) = (0, 0, 0)$ and there are no evanescent waves involved.

Due to the symmetry of our problem it is convenient to express the angular spectrum representation Eq. (3.33) in terms of the angles θ and ϕ instead of k_x and k_y . This is easily accomplished by using the substitutions in Eq. (3.44) and

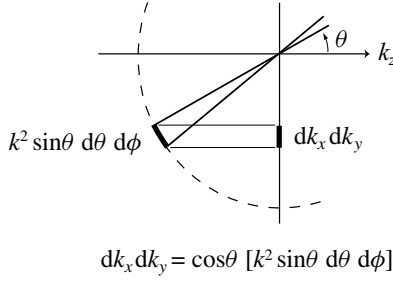


Figure 3.8 Illustration of the substitution $(1/k_z) dk_x dk_y = k \sin \theta d\theta d\phi$. The factor $1/k_z = 1/(k \cos \theta)$ ensures that the differential areas on the plane and the sphere stay equal.

expressing the transverse coordinates (x, y) of the field point as

$$x = \rho \cos \varphi, \quad y = \rho \sin \varphi. \quad (3.45)$$

In order to replace the planar integration over k_x, k_y by a spherical integration over θ, ϕ we must transform the differentials as

$$\frac{1}{k_z} dk_x dk_y = k \sin \theta d\theta d\phi, \quad (3.46)$$

which is illustrated in Fig. 3.8. We can now express the angular spectrum representation of the focal field (Eq. 3.33) as

$$\mathbf{E}(\rho, \varphi, z) = \frac{ikf e^{-ikf}}{2\pi} \int_0^{\theta_{\max}} \int_0^{2\pi} \mathbf{E}_{\infty}(\theta, \phi) e^{ikz \cos \theta} e^{ik\rho \sin \theta \cos(\phi - \varphi)} \sin \theta d\phi d\theta.$$

(3.47)

We have replaced the distance r_{∞} between the focal point and the surface of the reference sphere by the focal length f of the lens. We have also limited the integration over θ to the finite range $[0 \dots \theta_{\max}]$ because any lens will have a finite size. Furthermore, since all fields propagate in the positive z -direction we retained only the $+$ sign in the exponent of Eq. (3.33). Equation (3.47) is the central result of this section. Together with Eq. (3.43), it allows us to calculate the focusing of an arbitrary optical field \mathbf{E}_{inc} by an aplanatic lens with focal length f and numerical aperture

$$\text{NA} = n \sin \theta_{\max}, \quad (0 < \theta_{\max} < \pi/2), \quad (3.48)$$

where $n = n_2$ is the index of refraction of the surrounding medium. The field distribution in the focal region is entirely determined by the far-field \mathbf{E}_{∞} . As we

shall see in the next section, the properties of the laser focus can be engineered by adjusting the amplitude and phase profile of \mathbf{E}_∞ .

3.6 Focal fields

Typically, the back-aperture of a microscope objective is a couple of millimeters in diameter. In order to make use of the full NA of the objective, the incident field \mathbf{E}_{inc} has to fill or overfill the back-aperture. Thus, because of the large diameter of the incident beam it is reasonable to treat it in the paraxial approximation. Let us assume that \mathbf{E}_{inc} is entirely polarized along the x -axis, i.e.

让我们假设 \mathbf{E}_{inc} 沿着 x 轴完全偏振

$$\mathbf{E}_{\text{inc}} = E_{\text{inc}} \mathbf{n}_x. \quad (3.49)$$

Furthermore, we assume the waist of the incoming beam coincides with the lens so it hits the lens with a planar phase front. For simplicity we also assume that we have a lens with good antireflection coating so we can neglect the Fresnel transmission coefficients

$$t_\theta^s = t_\theta^p = 1. \quad (3.50)$$

With these assumptions the far-field \mathbf{E}_∞ in Eq. (3.43) can be expressed as

$$\begin{aligned} \mathbf{E}_\infty(\theta, \phi) &= E_{\text{inc}}(\theta, \phi) [\cos \phi \mathbf{n}_\theta - \sin \phi \mathbf{n}_\phi] \sqrt{n_1/n_2} (\cos \theta)^{1/2} \\ &= E_{\text{inc}}(\theta, \phi) \frac{1}{2} \begin{bmatrix} (1 + \cos \theta) & -(1 - \cos \theta) \cos 2\phi \\ -(1 - \cos \theta) \sin 2\phi & -2 \cos \phi \sin \theta \end{bmatrix} \sqrt{\frac{n_1}{n_2}} (\cos \theta)^{1/2}, \end{aligned} \quad (3.51)$$

where the last expression is represented in Cartesian vector components. To proceed we need to specify the amplitude profile of the incoming beam E_{inc} . We will concentrate on the three lowest Hermite–Gaussian modes displayed in Fig. 3.2. The first of these modes corresponds to the fundamental Gaussian beam and the other two can be generated according to Eq. (3.17) of Section 3.2.2. Expressing the coordinates $(x_\infty, y_\infty, z_\infty)$ in Fig. 3.7 by the spherical coordinates (f, θ, ϕ) we find

$(0, 0)$ mode:

$$E_{\text{inc}} = E_0 e^{-(x_\infty^2 + y_\infty^2)/w_0^2} = E_0 e^{-f^2 \sin^2 \theta / w_0^2} \quad (3.52)$$

$(1, 0)$ mode:

$$E_{\text{inc}} = E_0 (2x_\infty/w_0) e^{-(x_\infty^2 + y_\infty^2)/w_0^2} = (2E_0 f/w_0) \sin \theta \cos \phi e^{-f^2 \sin^2 \theta / w_0^2} \quad (3.53)$$

$(0, 1)$ mode:

$$E_{\text{inc}} = E_0 (2y_\infty/w_0) e^{-(x_\infty^2 + y_\infty^2)/w_0^2} = (2E_0 f/w_0) \sin \theta \sin \phi e^{-f^2 \sin^2 \theta / w_0^2} \quad (3.54)$$

The factor $f_w(\theta) = \exp(-f^2 \sin^2 \theta / w_0^2)$ is common to all modes. The focal field \mathbf{E} will depend on how much the incoming beam is expanded relative to the size of

聚焦场 \mathbf{E} 将取决于入射光束相对于透镜的多少

透镜的半径等于 $f \sin(\theta_{\max})$

62

Propagation and focusing of optical fields

the lens. Since the *aperture radius* of our lens is equal to $f \sin \theta_{\max}$ we define the *filling factor* f_0 as

填充因子 f_0

$$f_0 = \frac{w_0}{f \sin \theta_{\max}}, \quad (3.55)$$

which allows us to write the exponential function in Eqs. (3.52)–(3.54) in the form

$$f_w(\theta) = e^{-\frac{1}{f_0^2} \frac{\sin^2 \theta}{\sin^2 \theta_{\max}}}. \quad (3.56)$$

This function is called the *apodization function* and can be viewed as a pupil filter. We now have all the necessary ingredients to compute the field \mathbf{E} near the focus. With the mathematical relations

$$\begin{aligned} \int_0^{2\pi} \cos n\phi e^{ix \cos(\phi-\varphi)} d\phi &= 2\pi (i^n) J_n(x) \cos n\varphi \\ \int_0^{2\pi} \sin n\phi e^{ix \cos(\phi-\varphi)} d\phi &= 2\pi (i^n) J_n(x) \sin n\varphi, \end{aligned} \quad (3.57)$$

we can carry out the integration over ϕ analytically. Here, J_n is the n th-order Bessel function. The final expressions for the focal field now contain a single integration over the variable θ . It is convenient to use the following abbreviations for the occurring integrals:

$$I_{00} = \int_0^{\theta_{\max}} f_w(\theta) (\cos \theta)^{1/2} \sin \theta (1 + \cos \theta) J_0(k\rho \sin \theta) e^{ikz \cos \theta} d\theta \quad (3.58)$$

$$I_{01} = \int_0^{\theta_{\max}} f_w(\theta) (\cos \theta)^{1/2} \sin^2 \theta J_1(k\rho \sin \theta) e^{ikz \cos \theta} d\theta \quad (3.59)$$

$$I_{02} = \int_0^{\theta_{\max}} f_w(\theta) (\cos \theta)^{1/2} \sin \theta (1 - \cos \theta) J_2(k\rho \sin \theta) e^{ikz \cos \theta} d\theta \quad (3.60)$$

$$I_{10} = \int_0^{\theta_{\max}} f_w(\theta) (\cos \theta)^{1/2} \sin^3 \theta J_0(k\rho \sin \theta) e^{ikz \cos \theta} d\theta \quad (3.61)$$

$$I_{11} = \int_0^{\theta_{\max}} f_w(\theta) (\cos \theta)^{1/2} \sin^2 \theta (1 + 3 \cos \theta) J_1(k\rho \sin \theta) e^{ikz \cos \theta} d\theta \quad (3.62)$$

$$I_{12} = \int_0^{\theta_{\max}} f_w(\theta) (\cos \theta)^{1/2} \sin^2 \theta (1 - \cos \theta) J_1(k\rho \sin \theta) e^{ikz \cos \theta} d\theta \quad (3.63)$$

$$I_{13} = \int_0^{\theta_{\max}} f_w(\theta) (\cos \theta)^{1/2} \sin^3 \theta J_2(k\rho \sin \theta) e^{ikz \cos \theta} d\theta \quad (3.64)$$

$$I_{14} = \int_0^{\theta_{\max}} f_w(\theta) (\cos \theta)^{1/2} \sin^2 \theta (1 - \cos \theta) J_3(k\rho \sin \theta) e^{ikz \cos \theta} d\theta \quad (3.65)$$

where the function $f_w(\theta)$ is given by Eq. (3.56). Notice, that these integrals are functions of the coordinates (ρ, z) , i.e. $I_{ij} = I_{ij}(\rho, z)$. Thus, for each field point we have to numerically evaluate these integrals. Using these abbreviations we can now express the focal fields of the various modes as

(0, 0) mode:

$$\mathbf{E}(\rho, \varphi, z) = \frac{ikf}{2} \sqrt{\frac{n_1}{n_2}} E_0 e^{-ikf} \begin{bmatrix} I_{00} + I_{02} \cos 2\varphi \\ I_{02} \sin 2\varphi \\ -2i I_{01} \cos \varphi \end{bmatrix} \quad (3.66)$$

$$\mathbf{H}(\rho, \varphi, z) = \frac{ikf}{2Z_{\mu\varepsilon}} \sqrt{\frac{n_1}{n_2}} E_0 e^{-ikf} \begin{bmatrix} I_{02} \sin 2\varphi \\ I_{00} - I_{02} \cos 2\varphi \\ -2i I_{01} \sin \varphi \end{bmatrix}$$

(1, 0) mode:

$$\mathbf{E}(\rho, \varphi, z) = \frac{ikf^2}{2w_0} \sqrt{\frac{n_1}{n_2}} E_0 e^{-ikf} \begin{bmatrix} iI_{11} \cos \varphi + iI_{14} \cos 3\varphi \\ -iI_{12} \sin \varphi + iI_{14} \sin 3\varphi \\ -2I_{10} + 2I_{13} \cos 2\varphi \end{bmatrix} \quad (3.67)$$

$$\mathbf{H}(\rho, \varphi, z) = \frac{ikf^2}{2w_0 Z_{\mu\varepsilon}} \sqrt{\frac{n_1}{n_2}} E_0 e^{-ikf} \begin{bmatrix} -iI_{12} \sin \varphi + iI_{14} \sin 3\varphi \\ i(I_{11} + 2I_{12}) \cos \varphi - iI_{14} \cos 3\varphi \\ 2I_{13} \sin 2\varphi \end{bmatrix}$$

(0, 1) mode:

$$\mathbf{E}(\rho, \varphi, z) = \frac{ik f^2}{2w_0} \sqrt{\frac{n_1}{n_2}} E_0 e^{-ikf} \begin{bmatrix} i(I_{11} + 2I_{12}) \sin \varphi + iI_{14} \sin 3\varphi \\ -iI_{12} \cos \varphi - iI_{14} \cos 3\varphi \\ 2I_{13} \sin 2\varphi \end{bmatrix} \quad (3.68)$$

$$\mathbf{H}(\rho, \varphi, z) = \frac{ik f^2}{2w_0 Z_{\mu\epsilon}} \sqrt{\frac{n_1}{n_2}} E_0 e^{-ikf} \begin{bmatrix} -iI_{12} \cos \varphi - iI_{14} \cos 3\varphi \\ iI_{11} \sin \varphi - iI_{14} \sin 3\varphi \\ -2I_{10} - 2I_{13} \cos 2\varphi \end{bmatrix}.$$

For completeness, we have also listed the magnetic fields for the three modes. They can be derived in the same way by using the corresponding paraxial input fields H_∞ with the magnetic field axis along the y -axis. Notice that only the zero-order Bessel function possesses a non-vanishing value at its origin. As a consequence, only the (1, 0) mode has a longitudinal electric field (E_z) at its focus.

In the limit $f_w = 1$ the fields for the (0, 0) mode are identical with the solutions of Richards and Wolf [10]. According to Eq. (3.56), this limit is reached for $f_0 \rightarrow \infty$, which corresponds to an infinitely overfilled back-aperture of the focusing lens. This situation is identical with a plane wave incident on the lens. Figure 3.9 demonstrates the effect of the filling factor f_0 on the confinement of the focal fields.

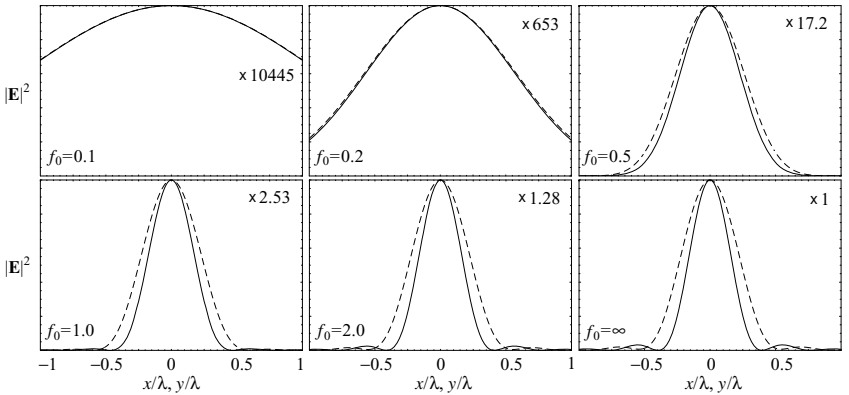


Figure 3.9 Influence of the filling factor f_0 of the back-aperture on the sharpness of the focus. A lens with $\text{NA} = 1.4$ is assumed and the index of refraction is 1.518. The figure shows the magnitude of the electric field intensity $|\mathbf{E}|^2$ in the focal plane $z = 0$. The dashed curves have been evaluated along the x -direction (plane of polarization) and the solid curves along the y -direction. All curves have been scaled to an equal amplitude. The scaling factor is indicated in the figures. The larger the filling factor is, the bigger is the deviation between the solid and dashed curve, indicating the importance of polarization effects.

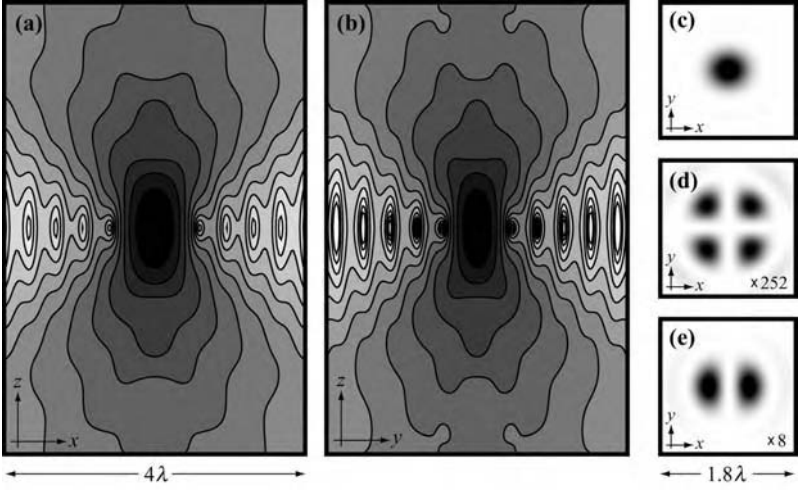


Figure 3.10 (a, b) Contour plots of constant $|\mathbf{E}|^2$ in the focal region of a focused Gaussian beam ($\text{NA} = 1.4$, $n = 1.518$, $f_0 = 1$); (a) plane of incident polarization (x, z), (b) plane perpendicular to plane of incident polarization (y, z). A logarithmic scaling is used with a factor of 2 between adjacent contour lines. (c, d, e) show the magnitude of the individual field components $|\mathbf{E}_x|^2$, $|\mathbf{E}_y|^2$, and $|\mathbf{E}_z|^2$ in the focal plane ($z=0$), respectively. A linear scale is used.

In these examples we used an objective with numerical aperture of 1.4 and an index of refraction of 1.518, which corresponds to a maximum collection angle of 68.96° . It is obvious that the filling factor is important for the quality of the focal spot and thus for the resolution in optical microscopy. It is important to notice that with increasing field confinement at the focus the focal spot becomes more and more elliptical. While in the paraxial limit the spot is perfectly circular, a strongly focused beam has a spot that is elongated in the direction of polarization. This observation has important consequences: as we aim towards higher resolutions by using spatially confined light we need to take the vector nature of the fields into account. Scalar theories become insufficient. Figure 3.10 shows field plots for the electric field for a filling factor of $f_0 = 1$ and a $\text{NA} = 1.4$ objective lens. The figure depicts the total electric field intensity \mathbf{E}^2 in the plane of incident polarization (x, z) and perpendicular to it (y, z). The three images to the side show the intensity of the different field components in the focal plane $z=0$. The maximum relative values are $\text{Max}[E_y^2]/\text{Max}[E_x^2] = 0.003$, and $\text{Max}[E_z^2]/\text{Max}[E_x^2] = 0.12$. Thus, an appreciable amount of the electric field energy is in the longitudinal field.

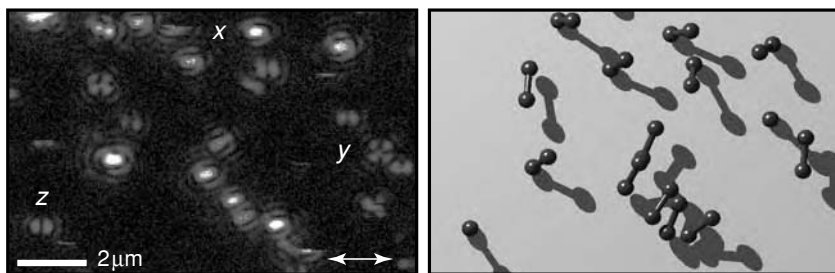


Figure 3.11 Single molecule excitation patterns. A sample with isolated single molecules is raster scanned in the focal plane of a strongly focused laser beam. For each pixel, the fluorescence intensity is recorded and encoded in the color scale. The excitation rate in each pixel is determined by the relative orientation of local electric field vector and molecular absorption dipole moment. Using the known field distribution in the laser focus allows the dipole moments to be reconstructed from the recorded patterns. Compare the patterns marked x , y , and z with those in the previous figure.

How can we experimentally verify the calculated focal fields? An elegant method is to use a single dipolar emitter, such as a single molecule, to probe the field. The molecule can be embedded into the surrounding medium with index n and moved with accurate translators to any position $\mathbf{r} = (x, y, z) = (\rho, \varphi, z)$ near the laser focus. The excitation rate of the molecule depends on the vector product $\mathbf{E} \cdot \boldsymbol{\mu}$, with $\boldsymbol{\mu}$ being the transition dipole moment of the molecule. The excited molecule then relaxes with a certain rate and probability by emitting a fluorescence photon. We can use the same aplanatic lens to collect the emitted photons and direct them onto a photodetector. The fluorescence intensity (photon counts per second) will be proportional to $|\mathbf{E} \cdot \boldsymbol{\mu}|^2$. Thus if we know the dipole orientation of the molecule, we can determine the field strength of the exciting field at the molecule's position. For example, a molecule aligned with the x -axis will render the x -component of the focal field. We can then translate the molecule to a new position and determine the field at this new position. Thus, point by point we can establish a map of the magnitude of the electric field component that points along the molecular dipole axis. With the x -aligned molecule we should be able to reproduce the pattern shown in Fig. 3.10(c) if we scan the molecule point by point in the plane $z = 0$. This has been demonstrated in various experiments and will be discussed in Chapter 9.

3.7 Focusing of higher-order laser modes

So far, we have discussed focusing of the fundamental Gaussian beam. What about the (10) and (01) modes? We have calculated those in order to synthesize *doughnut*

modes with arbitrary polarization. Depending on how we superimpose those modes we obtain

Linearly polarized doughnut mode:

$$LP = HG_{10} \mathbf{n}_x + i HG_{01} \mathbf{n}_x \quad (3.69)$$

Radially polarized doughnut mode:

$$RP = HG_{10} \mathbf{n}_x + HG_{10} \mathbf{n}_y \quad (3.70)$$

Azimuthally polarized doughnut mode:

$$AP = -HG_{01} \mathbf{n}_x + HG_{01} \mathbf{n}_y \quad (3.71)$$

Here, $HG_{ij} \mathbf{n}_l$ denotes a Hermite–Gaussian (ij) mode polarized along the unit vector \mathbf{n}_l . The linearly polarized doughnut mode is identical with the Laguerre–Gaussian (01) mode defined in Eq. (3.18) and it is easily calculated by adding the fields of Eqs. (3.67) and (3.68) with a 90° phase delay. To determine the focal fields of the other two doughnut modes we need to derive the focal fields for the y -polarized modes. This is easily accomplished by rotating the existing fields in Eqs. (3.67) and (3.68) by 90° around the z -axis. The resulting focal fields turn out to be

Radially polarized doughnut mode:

$$\mathbf{E}(\rho, \varphi, z) = \frac{ik f^2}{2w_0} \sqrt{\frac{n_1}{n_2}} E_0 e^{-ikf} \begin{bmatrix} i(I_{11} - I_{12}) \cos \varphi \\ i(I_{11} - I_{12}) \sin \varphi \\ -4I_{10} \end{bmatrix} \quad (3.72)$$

$$\mathbf{H}(\rho, \varphi, z) = \frac{ik f^2}{2w_0 Z_{\mu\epsilon}} \sqrt{\frac{n_1}{n_2}} E_0 e^{-ikf} \begin{bmatrix} -i(I_{11} + 3I_{12}) \sin \varphi \\ i(I_{11} + 3I_{12}) \cos \varphi \\ 0 \end{bmatrix}$$

Azimuthally polarized doughnut mode:

$$\mathbf{E}(\rho, \varphi, z) = \frac{ik f^2}{2w_0} \sqrt{\frac{n_1}{n_2}} E_0 e^{-ikf} \begin{bmatrix} i(I_{11} + 3I_{12}) \sin \varphi \\ -i(I_{11} + 3I_{12}) \cos \varphi \\ 0 \end{bmatrix} \quad (3.73)$$

$$\mathbf{H}(\rho, \varphi, z) = \frac{ik f^2}{2w_0 Z_{\mu\epsilon}} \sqrt{\frac{n_1}{n_2}} E_0 e^{-ikf} \begin{bmatrix} i(I_{11} - I_{12}) \cos \varphi \\ i(I_{11} - I_{12}) \sin \varphi \\ -4I_{10} \end{bmatrix}.$$

With the definition of the following integrals

$$I_{\text{rad}} = I_{11} - I_{12} = \int_0^{\theta_{\text{max}}} f_w(\theta) (\cos \theta)^{3/2} \sin^2 \theta J_1(k\rho \sin \theta) e^{ikz \cos \theta} d\theta \quad (3.74)$$

$$I_{\text{azm}} = I_{11} + 3I_{12} = \int_0^{\theta_{\text{max}}} f_w(\theta) (\cos \theta)^{1/2} \sin^2 \theta J_1(k\rho \sin \theta) e^{ikz \cos \theta} d\theta \quad (3.75)$$

we see that to describe the focusing of radially polarized and azimuthally polarized doughnut modes we need to evaluate totally two integrals. The radial and azimuthal symmetries are easily seen by transforming the Cartesian field vectors into cylindrical field vectors as

$$\begin{aligned} E_\rho &= \cos \varphi E_x + \sin \varphi E_y, \\ E_\phi &= -\sin \varphi E_x + \cos \varphi E_y, \end{aligned} \quad (3.76)$$

and similarly for the magnetic field. While the radially polarized focused mode has a rotationally symmetric longitudinal electric field E_z , the azimuthally polarized focused mode has a rotationally symmetric longitudinal magnetic field H_z . As shown in Fig. 3.12 the longitudinal field strength $|E_z|^2$ increases with increasing numerical aperture. At a numerical aperture of $\text{NA} \approx 1$ the magnitude of $|E_z|^2$ becomes larger than the magnitude of the radial field $|E_\rho|^2$. This is important for applications that require strong longitudinal fields. Figure 3.13 shows field plots for the focused radially polarized beam using the same parameters and settings as in Fig. 3.10. More detailed discussions of the focusing of radially and azimuthally polarized beams are presented in Refs. [11–13]. The field distribution in the beam focus has been measured using single molecules as probes [7] and the knife-edge method [13].

Although laser beams can be adjusted to a higher mode by manipulating the laser resonator, it is desirable to convert a fundamental Gaussian beam into a higher-order mode *externally* without perturbing the laser characteristics. Such a conversion can be realized by inserting phase plates into different regions in the beam cross-section [14]. As shown in Fig. 3.14, the conversion to a Hermite–Gaussian (10) mode is favored by bisecting the fundamental Gaussian beam with the edge of a thin phase plate which shifts the phase of one half of the beam by 180° . The incident beam has to be polarized perpendicular to the edge of the phase plate and subsequent spatial filtering has to be performed to reject higher-order modes. A related approach makes use of half-coated mirrors to delay one half of the laser beam. In this case, the beam passes twice through the bisected part and hence the thickness of the coated part must be $\lambda/4$. Other mode-conversion schemes make use of external four-mirror ring cavities or interferometers [15, 16]. The approach

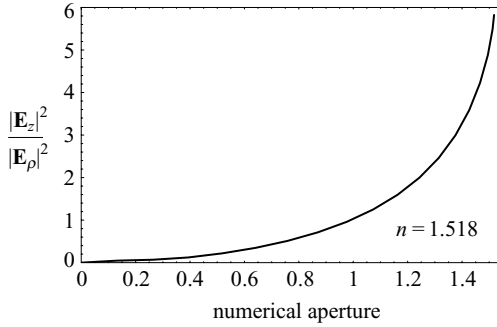


Figure 3.12 Ratio of longitudinal and transverse electric field intensity $|\mathbf{E}_z|^2/|\mathbf{E}_\rho|^2$ of a radially polarized doughnut mode as a function of numerical aperture ($f_0 = 1$, $n = 1.518$). $|\mathbf{E}_\rho|^2$ has its maximum on a ring in the plane $z = 0$ whereas the maximum of $|\mathbf{E}_z|^2$ is at the origin $(x, y, z) = (0, 0, 0)$. According to the figure, the maximum longitudinal electric energy density can be more than five times larger than the maximum transverse electric energy density.

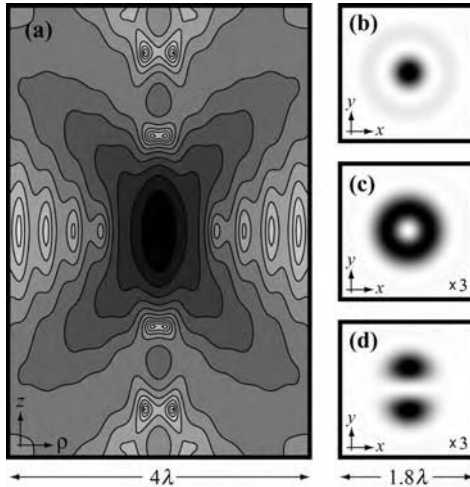


Figure 3.13 (a) Contour plots of constant $|\mathbf{E}|^2$ in the focal region of a focused radially polarized doughnut mode ($\text{NA} = 1.4$, $n = 1.518$, $f_0 = 1$) in the (ρ, z) plane. The intensity is rotationally symmetric with respect to the z -axis. A logarithmic scaling is used with a factor of 2 between adjacent contour lines. (b, c, d) show the magnitude of the individual field components $|\mathbf{E}_z|^2$, $|\mathbf{E}_\rho|^2$, and $|\mathbf{E}_y|^2$ in the focal plane ($z = 0$), respectively. A linear scale is used.

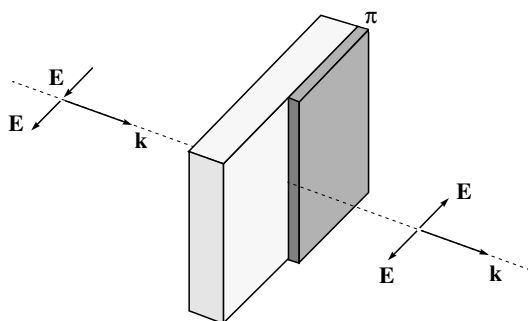


Figure 3.14 Generation of a Hermite–Gaussian (10) beam. A fundamental Gaussian beam is bisected at the edge of a 180° phase plate. The polarization of the incident beam is perpendicular to the edge of the phase plate. The arrangement delays one half of the beam by 180° and therefore favors the conversion to the Hermite–Gaussian (10) mode. A subsequent spatial filter rejects any modes of higher order than the (10) mode.

shown in Fig. 3.15(a) has been developed by Youngworth and Brown to generate azimuthally and radially polarized beams [11, 12]. It is based on a Twyman–Green interferometer with half-coated mirrors. The polarization of the incoming Gaussian beam is adjusted to 45° . A polarizing beamsplitter divides the power of the beam into two orthogonally polarized beams. Each of the beams passes a $\lambda/4$ phase plate which makes the beams circularly polarized. Each beam then reflects from an end mirror. One half of each mirror has a $\lambda/4$ coating which, after reflection, delays one half of the beam by 180° with respect to the other half. Each of the two reflected beams passes through the $\lambda/4$ plate again and becomes converted into equal amounts of orthogonally polarized Hermite–Gaussian (10) and (01) modes. Subsequently, one of these modes will be rejected by the polarizing beamsplitter whereas the other will be combined with the corresponding mode from the other interferometer arm. Whether a radially polarized mode or an azimuthally polarized mode is generated depends on the positioning of the half-coated end mirrors. To produce the other mode one needs to simply rotate the end mirrors by 90° . The two modes from the different interferometer arms need to be in phase, which requires adjustability of the path length. The correct polarization can always be verified by sending the output beam through a polarizer and by selectively blocking the beam in one of the two interferometer arms. Since the mode conversion is not 100% efficient one needs to spatially filter the output beam to reject any undesired modes. This is accomplished by focusing the output beam on a pinhole with adjusted diameter. Although the pinhole also transmits the fundamental mode, higher order modes have larger lateral extent and are rejected by the pinhole.

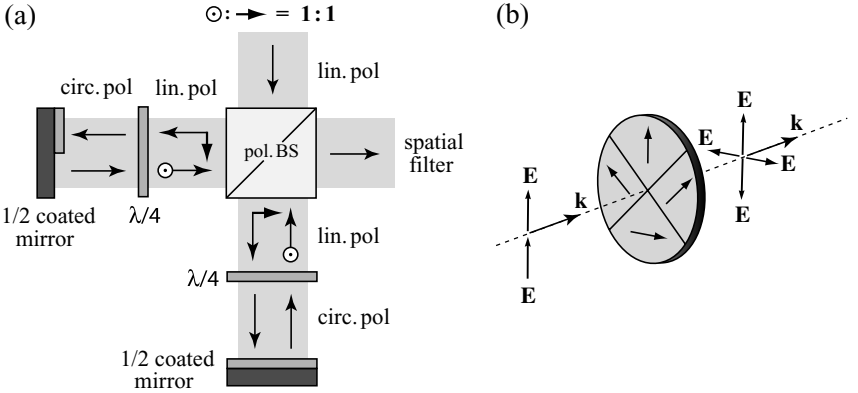


Figure 3.15 Two different mode conversion schemes for the generation of radially and azimuthally polarized modes. (a) Using a Twyman–Green interferometer. The incident beam is polarized at 45° and is split by a polarizing beamsplitter into two orthogonally polarized beams of equal power. Each beam is then turned circularly polarized and reflected off a half-coated end mirror. (b) Using a “composite waveplate” consisting of four quadrants with different optical axes. Each segment is oriented such that the field is rotated to point in the radial direction. In both schemes, the outgoing beam needs to be spatially filtered to reject unwanted higher-order modes. See text for details.

To avoid noise and drift sensitive interferometers, Dorn *et al.* have implemented a single-path mode conversion scheme for radially and azimuthally polarized beams [13]. As shown in Fig. 3.15(b), a laser beam is sent through a $\lambda/2$ waveplate consisting of four segments. The optical axis of each segment is oriented such that the field is rotated to point in the radial direction. Subsequent spatial filtering extracts the desired mode with very high purity. A phase plate as shown in Fig. 3.15(b) can be fabricated by cutting two $\lambda/2$ plates into four quadrants each, and then assembling the pieces into two new phase plates. This mode-conversion principle can be generalized to waveplates with many elements such as liquid crystal spatial light modulators. It can be expected that programmable spatial light modulators will be able to convert an incoming beam to any desired outgoing laser mode.

3.8 Limit of weak focusing

Before we proceed to the next section we need to verify that our formulas for the focused fields render the familiar paraxial expressions for the limit of small θ_{\max} . In this limit we may do the approximations $\cos \theta \approx 1$ and $\sin \theta \approx \theta$. However,

for the phase factor in the exponent of the integrals $I_{00} \dots I_{14}$ we need to retain the second-order term, i.e. $\cos \theta \approx 1 - \theta^2/2$, because the first-order term alone would cancel the θ dependence. For small arguments x , the Bessel functions behave like $J_n(x) \approx x^n$. Using these approximations, a comparison of the integrals $I_{00} \dots I_{14}$ shows that the integral I_{00} is of lowest order in θ , followed by I_{11} and I_{12} . Whereas I_{00} defines the paraxial Gaussian mode, the other two remaining integrals determine the paraxial Hermite–Gaussian (1, 0) and (0, 1) modes. In principle, the integration of I_{00} , I_{10} and I_{11} can now be carried out analytically. However, since the results lead to inconvenient Lommel functions we reduce our discussion to the focal plane $z = 0$. Furthermore, we assume an overfilled back-aperture of the lens ($f_0 \gg 1$) so that the apodization function $f_w(\theta)$ can be considered constant. Using the substitution $x = k\rho\theta$ we find

$$I_{00} \approx \frac{2}{k\rho} \int_0^{k\rho\theta_{\max}} x J_0(x) dx = 2\theta_{\max}^2 \frac{J_1(k\rho\theta_{\max})}{k\rho\theta_{\max}}. \quad (3.77)$$

The paraxial field of the focused Gaussian beam in the focal plane turns out to be

$$\mathbf{E} \approx ik f \theta_{\max}^2 E_0 e^{-ikf} \frac{J_1(k\rho\theta_{\max})}{k\rho\theta_{\max}} \mathbf{n}_x. \quad (3.78)$$

This is the familiar expression for the point-spread function in the paraxial limit. Abbe's and Rayleigh's definitions of the resolution limit are closely related to the expression above as we shall see in Section 4.1. The focal fields of the (1, 0) and (0, 1) modes in the paraxial limit can be derived in a similar way as

(1, 0) mode:

$$\mathbf{E} \propto \theta_{\max}^3 [J_2(k\rho\theta_{\max})/(k\rho\theta_{\max})] \cos \varphi \mathbf{n}_x, \quad (3.79)$$

(0, 1) mode:

$$\mathbf{E} \propto \theta_{\max}^3 [J_2(k\rho\theta_{\max})/(k\rho\theta_{\max})] \sin \varphi \mathbf{n}_x. \quad (3.80)$$

In all cases, the radial dependence of the paraxial focal fields is described by Bessel functions and *not* by the original Gaussian envelope. After passing through the lens the beam shape in the focal plane becomes oscillatory. These spatial oscillations can be viewed as diffraction lobes and are a consequence of the boundary conditions imposed by the aplanatic lens. We have assumed $f_0 \rightarrow \infty$ and we can reduce the oscillatory behavior by reducing f_0 . However, this is at the expense of the spot size. The fact that the spot shape is described by an Airy function and not by a Gaussian function is very important. In fact, there are no free propagating Gaussian beams! The reason is, as outlined in Section 3.2.1, that a Gaussian profile has a Gaussian Fourier spectrum which is never zero and only asymptotically approaches zero as $k_x, k_y \rightarrow \infty$. Thus, for a Gaussian profile we need to include

evanescent components, even if their contribution is small. The oscillations in the Airy profile arise from the hard cut-off at high spatial frequencies. The smoother this cut-off the less oscillatory the beam profile will be.

3.9 Focusing near planar interfaces

Many applications in optics involve laser beams that are strongly focused near planar surfaces. Examples are confocal microscopy where objective lenses with $\text{NA} > 1$ are used, optical microscopy or data storage based on solid immersion lenses, and optical tweezers where laser light is focused into a liquid to trap tiny particles. The angular spectrum representation is well suited to solve for the fields since the planar interface is a constant coordinate surface. For simplicity we assume that we have a single interface between two dielectric media with indices n_1 and n_2 (see Fig. 3.16). The interface is located at $z = z_0$ and the focused field \mathbf{E}_f illuminates the interface from the left ($z < z_0$). While the spatial frequencies k_x and k_y are the same on each side of the interface, k_z is not. Therefore, we specify k_z in the domain $z < z_0$ by k_{z1} defined by $k_{z1} = (k_1^2 - k_x^2 - k_y^2)^{1/2}$. Similarly we define $k_{z2} = (k_2^2 - k_x^2 - k_y^2)^{1/2}$ for the domain $z > z_0$. The wavenumbers are determined by $k_1 = (\omega/c)n_1$ and $k_2 = (\omega/c)n_2$, respectively.

The interface leads to reflection and transmission. Therefore, the total field can be represented as

$$\mathbf{E} = \begin{cases} \mathbf{E}_f + \mathbf{E}_r & : z < z_0 \\ \mathbf{E}_t & : z > z_0 \end{cases} \quad (3.81)$$

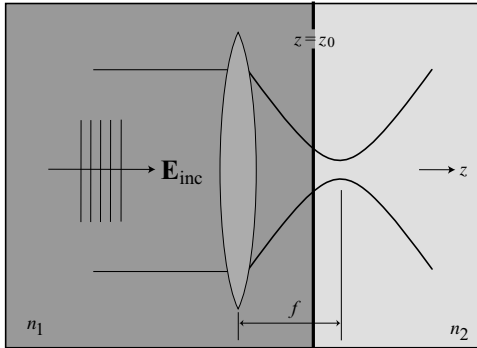


Figure 3.16 Focusing of a laser beam near an interface at $z = z_0$ between two dielectric media with refractive indices n_1 and n_2 .

where \mathbf{E}_r and \mathbf{E}_t represent the reflected and transmitted fields, respectively. The refraction of plane waves at planar interfaces is described by Fresnel reflection coefficients (r^s , r^p) and transmission coefficients (t^s , t^p), which were defined in Chapter 2 (Eqs. (2.49) and (2.50)). As indicated by the superscripts, these coefficients depend on the polarization of the field. We therefore need to split each plane wave component in the angular spectrum representation of the field \mathbf{E} into an s-polarized part and a p-polarized part

$$\mathbf{E} = \mathbf{E}^{(s)} + \mathbf{E}^{(p)}. \quad (3.82)$$

$\mathbf{E}^{(s)}$ is parallel to the interface while $\mathbf{E}^{(p)}$ is perpendicular to the wavevector \mathbf{k} and $\mathbf{E}^{(s)}$. The decomposition of the incoming focused field \mathbf{E}_f into s- and p-polarized fields has already been done in Section 3.5. According to Eq. (3.39) we obtain the s- and p-polarized fields by projecting \mathbf{E}_f along the unit vectors \mathbf{n}_θ and \mathbf{n}_ϕ , respectively. Equation (3.43) represents the refracted far-field as a sum of s- and p-polarized fields expressed in terms of θ and ϕ . Using the substitutions of Eq. (3.44) we are able to express the far-field in terms of the spatial frequencies k_x and k_y .

In the case where \mathbf{E}_f originates from a paraxial beam polarized in the x -direction we can express the far-field as (cf. Eq. (3.51))

$$\mathbf{E}_\infty = E_{\text{inc}}(k_x, k_y) \begin{bmatrix} k_y^2 + k_x^2 k_{z_1}/k_1 \\ -k_x k_y + k_x k_y k_{z_1}/k_1 \\ 0 - (k_x^2 + k_y^2) k_x/k_1 \end{bmatrix} \frac{\sqrt{k_{z_1}/k_1}}{k_x^2 + k_y^2}, \quad (3.83)$$

where the first terms in the bracket specify the s-polarized field and the second ones the p-polarized field. Notice, that according to Fig. 3.16 we consider a lens with the same medium on both sides, i.e. $n_1 = n = n'$. \mathbf{E}_∞ is the asymptotic far-field in the direction of the unit vector $\mathbf{s} = (k_x/k, k_y/k, k_{z_1}/k)$ and corresponds to the field on the surface of the reference sphere of the focusing lens. With \mathbf{E}_∞ the angular spectrum representation of the incident focused beam is given by (c.f. Eq. (3.33))

$$\mathbf{E}_f(x, y, z) = \frac{if e^{-ik_1 f}}{2\pi} \iint_{k_x, k_y} \mathbf{E}_\infty(k_x, k_y) \frac{1}{k_{z_1}} e^{i[k_x x + k_y y + k_{z_1} z]} dk_x dk_y. \quad (3.84)$$

To determine the reflected and transmitted fields (\mathbf{E}_r , \mathbf{E}_t) we define the following angular spectrum representations

$$\mathbf{E}_r(x, y, z) = \frac{if e^{-ik_1 f}}{2\pi} \iint_{k_x, k_y} \mathbf{E}_r^\infty(k_x, k_y) \frac{1}{k_{z_1}} e^{i[k_x x + k_y y - k_{z_1} z]} dk_x dk_y, \quad (3.85)$$

$$\mathbf{E}_t(x, y, z) = \frac{if e^{-ik_1 f}}{2\pi} \iint_{k_x, k_y} \mathbf{E}_t^\infty(k_x, k_y) \frac{1}{k_{z_2}} e^{i[k_x x + k_y y + k_{z_2} z]} dk_x dk_y. \quad (3.86)$$

Notice that in order to ensure that the reflected field propagates in the backward direction we had to change the sign of k_{z_1} in the exponent. We also made sure that the transmitted wave propagates with the longitudinal wavenumber k_{z_2} .

In the next step we invoke the boundary conditions at $z = z_0$, which leads to explicit expressions for the yet undefined far-fields \mathbf{E}_r^∞ and \mathbf{E}_t^∞ . Using the Fresnel reflection or transmission coefficients we obtain

$$\mathbf{E}_r^\infty = -E_{\text{inc}}(k_x, k_y) e^{2i k_{z_1} z_0} \begin{bmatrix} -r^s k_y^2 + r^p k_x^2 k_{z_1}/k_1 \\ r^s k_x k_y + r^p k_x k_y k_{z_1}/k_1 \\ 0 + r^p (k_x^2 + k_y^2) k_x/k_1 \end{bmatrix} \frac{\sqrt{k_{z_1}/k_1}}{k_x^2 + k_y^2}, \quad (3.87)$$

$$\mathbf{E}_t^\infty = E_{\text{inc}}(k_x, k_y) e^{i(k_{z_1} - k_{z_2})z_0} \begin{bmatrix} t^s k_y^2 + t^p k_x^2 k_{z_2}/k_2 \\ -t^s k_x k_y + t^p k_x k_y k_{z_2}/k_2 \\ 0 - t^p (k_x^2 + k_y^2) k_x/k_2 \end{bmatrix} \frac{k_{z_2}}{k_{z_1}} \frac{\sqrt{k_{z_1}/k_1}}{k_x^2 + k_y^2}. \quad (3.88)$$

These equations together with Eqs. (3.83)–(3.86) define the solution of our problem. They hold for an interface between two materials characterized by constant E_i and μ_i . This is straight-forward to verify by evaluating the boundary conditions at $z = z_0$ (Problem 3.7). We are now able to evaluate the field distribution near a plane interface illuminated by a strongly focused laser beam. The field depends on the amplitude profile $E_{\text{inc}}(k_x, k_y)$ of the incident paraxial beam (cf. Eqs. (3.52)–(3.54)) and on the defocus z_0 . The defocus essentially introduces a phase factor into the expressions for \mathbf{E}_r^∞ and \mathbf{E}_t^∞ . Although we concentrated on a single interface, the results are easily adapted to a multiply layered interface by introducing generalized Fresnel reflection/transmission coefficients that account for the total structure (cf. Ref. [17]).

In the next step, we can use the relations Eq. (3.44) to perform a transformation to spherical coordinates. As before, we are able to reduce the double integrals to single integrals by involving Bessel functions. We avoid going into further details and instead discuss some important aspects that result from this theory.

In the example of Fig. 3.17 a Gaussian beam is focused by an aplanatic objective lens of $\text{NA} = 1.4$ on a glass/air interface at $z_0 = 0$. The most characteristic features in the field plots are the standing wave patterns in the denser medium. These standing wave patterns occur at angles θ beyond the critical angle of total internal reflection θ_c . To understand this let us have a look at a single plane wave in the angular spectrum representation of the incident focused field \mathbf{E}_f . This plane wave is characterized by the two transverse wavenumbers k_x, k_y , its polarization and complex amplitude given by the Fourier spectrum $\hat{\mathbf{E}}_f$. The transverse wavenumbers are the same on each side of the interface, but the longitudinal wavenumbers k_z are not

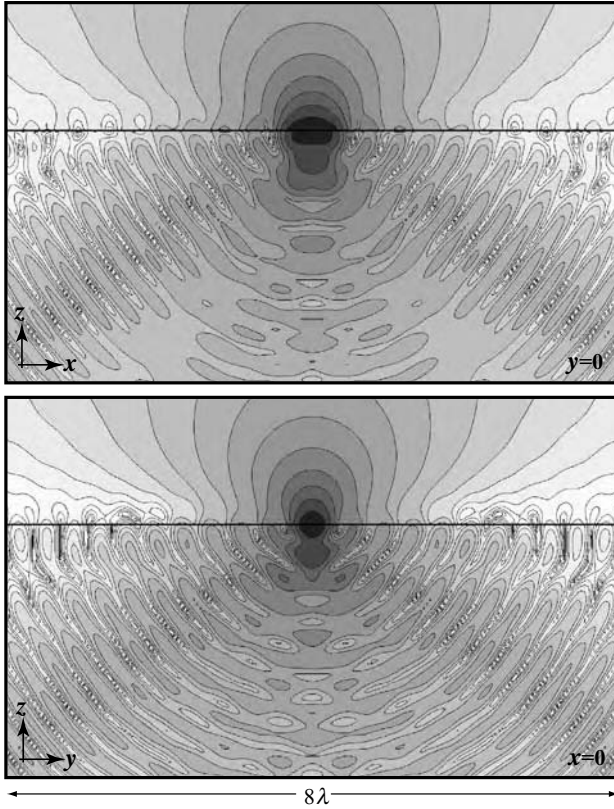


Figure 3.17 Contour plots of constant $|\mathbf{E}|^2$ in the focal region of a Gaussian beam ($\text{NA} = 1.4$, $n = 1.518$, $f_0 = 2$) focused on a glass/air interface ($n_1 = 1.518$, $n_2 = 1$). A logarithmic scaling is used with a factor of 2 between adjacent contour lines. The critical angle for total internal reflection is $\theta_c = 41.2^\circ$. All plane wave components incident from angles larger than θ_c are totally reflected at the interface and interfere with the incoming waves.

since they are defined as

$$k_{z_1} = \sqrt{k_1^2 - (k_x^2 + k_y^2)}, \quad k_{z_2} = \sqrt{k_2^2 - (k_x^2 + k_y^2)}. \quad (3.89)$$

Eliminating k_x, k_y we obtain

$$k_{z_2} = \sqrt{k_{z_1}^2 + (k_2^2 - k_1^2)}. \quad (3.90)$$

Let θ denote the angle of incidence of the plane wave so that

$$k_{z_1} = k_1 \cos \theta . \quad (3.91)$$

Equation (3.90) can then be written as

$$k_{z_2} = k_2 \sqrt{1 - \frac{k_1^2}{k_2^2} \sin^2 \theta} . \quad (3.92)$$

It follows that k_{z_2} can be either real or imaginary, depending on the sign of the expression under the square root. This in turn depends on the angle θ . We find that for angles larger than

$$\theta_c = \arcsin \frac{n_2}{n_1} \quad (3.93)$$

k_{z_2} is imaginary. Thus, for $\theta > \theta_c$ the considered plane wave is totally reflected at the interface giving rise to an evanescent wave on the other side of the interface. The standing wave patterns seen in Fig. 3.17 are a direct consequence of this phenomenon: all the supercritical ($\theta > \theta_c$) plane wave components of the incident focused field are totally reflected at the interface. The standing wave pattern is due to the equal superposition of incident and reflected plane wave components. Due to total internal reflection an appreciable amount of laser power is reflected at the interface. The ratio of reflected to transmitted power can be further increased by using a larger filling factor or a higher numerical aperture. For example, in applications based on solid immersion lenses with numerical apertures of 1.8 ... 2 over 90% of the beam power is reflected at the interface.

An inspection of the focal spot reveals that the interface further increases the ellipticity of the spot shape. Along the polarization direction (x) the spot is almost twice as big as in the direction perpendicular to it (y). Furthermore, the interface enhances the strength of the longitudinal field component E_z . At the interface, just outside the focusing medium ($z > -z_0$), the maximum relative intensity values for the different field components are $\text{Max}[E_y^2]/\text{Max}[E_x^2] = 0.03$ and $\text{Max}[E_z^2]/\text{Max}[E_x^2] = 0.43$. Thus, compared with the situation where no interface is present (cf. Fig. 3.10), the longitudinal field intensity is roughly four times stronger. How can we understand this phenomenon? According to the boundary conditions at the interface, the transverse field components E_x, E_y have to be continuous across the interface. However, the longitudinal field scales as

$$E_{z_1} \varepsilon_2 = E_{z_2} \varepsilon_1 . \quad (3.94)$$

With $\varepsilon_2 = 2.304$ we find that E_z^2 changes by a factor of 5.3 from one side to the other side of the interface. This qualitative explanation is in reasonable agreement with the calculated values. In the focal plane, the longitudinal field has its two maxima just to the side of the optical axis. These two maxima are aligned along

the polarization direction and give rise to the elongated spot size. The relative magnitude of $\text{Max}[E_y^2]$ is still small but it is increased by a factor of 10 by the presence of the interface.

In order to map the dipole orientation of arbitrarily oriented single molecules it is desirable that all three excitation field components (E_x , E_y , E_z) in the focus are of comparable magnitude. It has been demonstrated that this can be achieved by annular illumination for which the center part of the focused laser beam is suppressed [18]. This can be achieved by placing a central obstruction such as a circular disk in the excitation beam. In this situation, the integration of plane wave components runs over the angular range $[\theta_{\min} \dots \theta_{\max}]$ instead of, as before, over the full range $[0 \dots \theta_{\max}]$. By using annular illumination we reject the plane wave components with propagation directions close to the optical axis, thereby suppressing the transverse electric field components. As a consequence, the longitudinal field components in the focus will be enhanced as compared to the transverse components. Furthermore, the local polarization of the interface due to the longitudinal fields gives rise to a strong enhancement of the E_y fields. Hence, strong longitudinal fields are a prerequisite for generating strong E_y fields close to interfaces. It is possible to prepare the annular beam such that the three patterns Fig. 3.10(c–e) are of comparable magnitude.

3.10 Reflected image of a strongly focused spot

It is interesting to further investigate the properties of the reflected field \mathbf{E}_r given by Eq. (3.85) and Eq. (3.87). The image of the reflected spot can be experimentally recorded as shown in Fig. 3.18. A 45° beamsplitter reflects part of the incoming beam upwards where it is focused by a high NA objective lens near a planar interface. The distance between focus ($z = 0$) and interface is designated by z_0 . The reflected field is collected by the same lens, transmitted through the beamsplitter and then focused by a second lens onto the image plane. There are four different media involved and we specify them with the refractive indices defined in Fig. 3.18. We are interested in calculating the resulting field distribution in the image plane. It will be shown that, for the case where the beam is incident from the optically denser medium, the image generated by the reflected light is strongly aberrated.

The reflected far-field \mathbf{E}_r^∞ before it is refracted by the first lens has been calculated in Eq. (3.87). It is straightforward to refract this field at the two lenses and refocus it onto the image plane. The two lenses perform transformations between spherical and cylindrical systems. In Section 3.5 it has been shown that the lens refracts the unit vector \mathbf{n}_ρ into the unit vector \mathbf{n}_θ , or vice versa, whereas the unit vector \mathbf{n}_ϕ remains unaffected. In order to oversee the entire imaging process we follow the light path from the beginning. The incoming field \mathbf{E}_{inc} is an x -polarized,

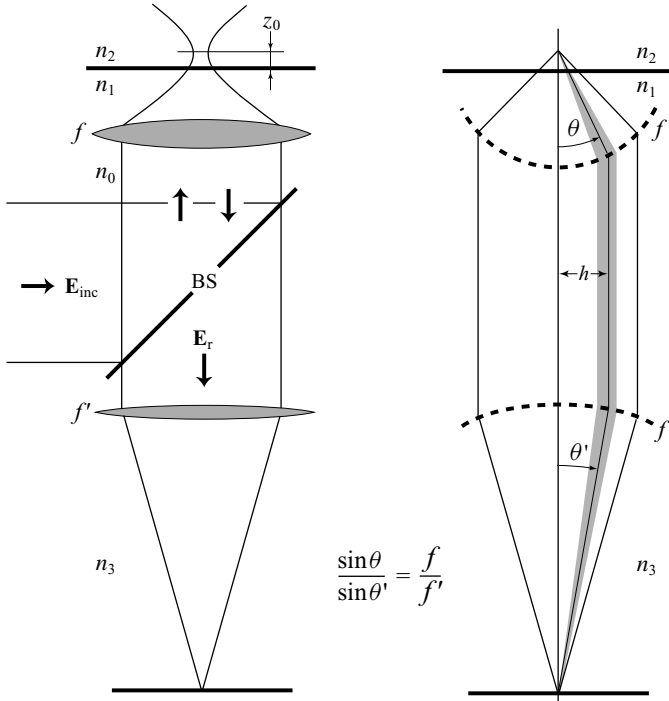


Figure 3.18 Experimental set-up for the investigation of the reflected image of a diffraction-limited focused spot. A linearly polarized beam is reflected by a beam-splitter (BS) and focused by a high NA objective lens with focal radius f onto an interface between two dielectric media n_1, n_2 . The reflected field is collected by the same lens, transmitted through the beamsplitter and refocused by a second lens with focal radius f' .

paraxial beam defined as (Eq. (3.49))

$$\mathbf{E}_{\text{inc}} = E_{\text{inc}} \mathbf{n}_x, \quad (3.95)$$

where E_{inc} is an arbitrary beam profile. Expressed in cylindrical coordinates the field has the form

$$\mathbf{E}_{\text{inc}} = E_{\text{inc}} [\cos \phi \mathbf{n}_\rho - \sin \phi \mathbf{n}_\phi]. \quad (3.96)$$

After refraction at the first lens f it turns into

$$\mathbf{E} = E_{\text{inc}} [\cos \phi \mathbf{n}_\theta - \sin \phi \mathbf{n}_\phi] \sqrt{\frac{n_0}{n_1}} (\cos \theta)^{1/2}. \quad (3.97)$$

The field is now reflected at the interface. The Fresnel reflection coefficient r^p accounts for the reflection of \mathbf{n}_θ -polarized fields whereas r^s accounts for the reflection of \mathbf{n}_ϕ -polarized fields. We obtain for the reflected field

$$\mathbf{E} = E_{\text{inc}} e^{2ik_{z_1} z_0} [-\cos \phi r^p \mathbf{n}_\theta - \sin \phi r^s \mathbf{n}_\phi] \sqrt{\frac{n_0}{n_1}} (\cos \theta)^{1/2}, \quad (3.98)$$

where z_0 denotes the defocus (cf. Eq. (3.87)). Next, the field is refracted by the same lens f as

$$\mathbf{E} = E_{\text{inc}} e^{2ik_{z_1} z_0} [-\cos \phi r^p \mathbf{n}_\rho - \sin \phi r^s \mathbf{n}_\phi], \quad (3.99)$$

and propagates as a collimated beam in the negative z -direction. Expressed in Cartesian field components the field reads as

$$\mathbf{E}_r^\infty = -E_{\text{inc}} e^{2ik_{z_1} z_0} [\cos^2 \phi r^p - \sin^2 \phi r^s] \mathbf{n}_x + \sin \phi \cos \phi [r^p + r^s] \mathbf{n}_y. \quad (3.100)$$

This is the field immediately after refraction at the reference sphere f . For an incident field focused on a perfectly reflecting interface located at $z_0 = 0$ the reflection coefficients are $r^p = 1$ and $r^s = -1$.² In this case we simply obtain $\mathbf{E}_{\text{ref}}^\infty = -E_{\text{inc}} \mathbf{n}_x$, which is, besides the minus sign, identical with the assumed input field of Eq. (3.49). The difference in sign indicates that the reflected field is “upside down”.

In order to calculate the reflected collimated beam anywhere along the optical axis we have to substitute $\sin \theta = \rho/f$ and $\cos \theta = [1 - (\rho/f)^2]^{1/2}$, where ρ denotes the radial distance from the optical axis (see Problem 3.8). This allows us to plot the field distribution in a cross-sectional plane through the collimated reflected beam. We find that the Fresnel reflection coefficients modify the polarization and amplitude profile of the beam, and, more importantly, also its phase profile. For no defocus ($z_0 = 0$) phase variations only arise at radial distances $\rho > \rho_c$ for which the Fresnel reflection coefficients become complex numbers. The critical distance corresponds to $\rho_c = f n_2/n_1$ and is the radial distance associated with the critical angle of total internal reflection ($\theta_c = \arcsin(n_2/n_1)$). Since $\rho_c < f$ there are no aberrations if $n_2 > n_1$.

We now proceed to the refraction at the second lens f' . Immediately after refraction the reflected field reads as

$$\mathbf{E} = E_{\text{inc}} e^{2ik_{z_1} z_0} [-\cos \phi r^p \mathbf{n}_{\theta'} - \sin \phi r^s \mathbf{n}_\phi] \sqrt{\frac{n_0}{n_3}} (\cos \theta')^{1/2}, \quad (3.101)$$

where we introduced the new azimuth angle θ' as defined in Fig. 3.18. The field now corresponds to the far-field \mathbf{E}_r^∞ that we need in Eq. (3.33) to calculate the field

² Notice that the reflection coefficients r^s, r^p for a plane wave at normal incidence differ by a factor of -1 , i.e. $r^s(\theta=0) = -r^p(\theta=0)$.

distribution in the image space. We express this field in Cartesian field components using the relations in Eqs. (3.41)–(3.42) for $\mathbf{n}_{\theta'}$ and \mathbf{n}_{ϕ} and obtain

$$\mathbf{E}_r^\infty = -E_{\text{inc}} e^{2ik_{z_1} z_0} \begin{bmatrix} r^p \cos \theta' \cos^2 \phi - r^s \sin^2 \phi \\ r^p \cos \theta' \sin \phi \cos \phi + r^s \sin \phi \cos \phi \\ -r^p \sin \theta' \cos \phi + 0 \end{bmatrix} \sqrt{\frac{n_0}{n_3}} (\cos \theta')^{1/2}. \quad (3.102)$$

This far-field can now be introduced into Eq. (3.47), which, after being adapted to the current situation, reads as

$$\mathbf{E}(\rho, \varphi, z) = \frac{ik_3 f' e^{-ik_3 f'}}{2\pi} \int_0^{\theta'_{\text{max}}} \int_0^{2\pi} \mathbf{E}_r^\infty(\theta', \phi) e^{-ik_3 z \cos \theta'} e^{ik_3 \rho \sin \theta' \cos(\phi - \varphi)} \sin \theta' d\phi d\theta'. \quad (3.103)$$

Notice that we had to change the sign in one of the exponents in order to ensure that the field propagates in the negative z -direction. To proceed, we could express the longitudinal wavenumbers k_{z_1} and k_{z_2} in terms of the angle θ' . This would also make the reflection and transmission coefficients functions of θ' . However, it is more convenient to work with θ and transform the integral in Eq. (3.105) correspondingly.

As indicated in Fig. 3.18 the angles θ and θ' are related by

$$\frac{\sin \theta}{\sin \theta'} = \frac{f'}{f}, \quad (3.104)$$

which allows us to express the new longitudinal wavenumber k_{z_3} in terms of θ as

$$k_{z_3} = k_3 \sqrt{1 - (f/f')^2 \sin^2 \theta}. \quad (3.105)$$

With these relationships we can perform a substitution in Eq. (3.105) and represent the integration variables by θ and ϕ . The Fresnel reflection coefficients $r_s(\theta)$, $r_p(\theta)$ are given by Eqs. (2.49) together with the expressions for the longitudinal wavenumbers k_{z_1} and k_{z_2} in Eqs. (3.91) and (3.92). For the lowest three Hermite–Gaussian beams, explicit expressions for $E_{\text{inc}}(\theta, \phi)$ have been stated in Eqs. (3.52)–(3.54) and the angular dependence in ϕ can be integrated analytically by using Eq. (3.57). Thus, we are now able to calculate the field near the image focus.

In practically all optical systems the second focusing lens has a much larger focal length than the first one, i.e. $f/f' \ll 1$. We can therefore reduce the complexity of the expressions considerably by making the approximation

$$[1 \pm (f/f')^2 \sin^2 \theta]^{1/n} \approx 1 \pm \frac{1}{n} \left(\frac{f}{f'} \right)^2 \sin^2 \theta. \quad (3.106)$$

If we retain only the lowest orders in f/f' , the image field can be represented by

$$\begin{aligned} \mathbf{E}(\rho, \varphi, z) = & \frac{ik_3 f' e^{-ik_3 f'}}{2\pi} \frac{f^2}{f'^2} \int_0^{\theta_{\max}} \int_0^{2\pi} \mathbf{E}_r^\infty(\theta, \phi) e^{(i/2)k_3 z (f/f')^2 \sin^2 \theta} \\ & \times e^{ik_3 \rho (f/f') \sin \theta \cos(\phi - \varphi)} \sin \theta \cos \theta d\phi d\theta, \end{aligned} \quad (3.107)$$

where \mathbf{E}_r^∞ reads as

$$\mathbf{E}_r^\infty(\theta, \phi) = -E_{\text{inc}}(\theta, \phi) e^{2ik_1 z_0 \cos \theta} \begin{bmatrix} r^p \cos^2 \phi - r^s \sin^2 \phi \\ \sin \phi \cos \phi (r^p + r^s) \\ 0 \end{bmatrix} \sqrt{\frac{n_0}{n_3}}. \quad (3.108)$$

In order to keep the discussion in bounds we will assume that the incident field E_{inc} is a fundamental Gaussian beam as defined in Eq. (3.52). Using the relations in Eq. (3.57) we can integrate the ϕ dependence and finally obtain

$$\mathbf{E}(\rho, \varphi, z) = E_0 \frac{k_3 f^2}{2 f' i} e^{-ik_3 (z+f')} \sqrt{\frac{n_0}{n_3}} \left[(I_{0r} + I_{2r} \cos 2\varphi) \mathbf{n}_x - I_{2r} \sin 2\varphi \mathbf{n}_y \right], \quad (3.109)$$

with

$$\begin{aligned} I_{0r}(\rho, z) = & \int_0^{\theta_{\max}} f_w(\theta) \cos \theta \sin \theta [r_p(\theta) - r_s(\theta)] J_0(k_3 \rho \sin \theta f/f') \\ & \times \exp[(i/2)k_3 z (f/f')^2 \sin^2 \theta + 2ik_1 z_0 \cos \theta] d\theta, \end{aligned} \quad (3.110)$$

$$\begin{aligned} I_{2r}(\rho, z) = & \int_0^{\theta_{\max}} f_w(\theta) \cos \theta \sin \theta [r_p(\theta) + r_s(\theta)] J_2(k_3 \rho \sin \theta f/f') \\ & \times \exp[(i/2)k_3 z (f/f')^2 \sin^2 \theta + 2ik_1 z_0 \cos \theta] d\theta, \end{aligned} \quad (3.111)$$

where f_w is the apodization function defined in Eq. (3.56). We find that the spot depends on the Fresnel reflection coefficients and the defocus defined by z_0 . The latter simply adds for each plane wave component an additional phase delay. If the upper medium n_2 is a perfect conductor we have $r_p = -r_s = 1$ and the integral I_{2r} vanishes. In this case the reflected spot is linearly polarized and rotationally symmetric.

In order to discuss the field distributions in the image plane we choose $n_1 = 1.518$ for the object space, $n_3 = 1$ for the image space, and a numerical aperture of 1.4 ($\theta_{\max} = 67.26^\circ$) for the objective lens. For the ideally reflecting interface, the images in the lower row of Fig. 3.19 depict the electric field intensity $|\mathbf{E}_r|^2$ as a function of slight defocus. It is evident that the spot shape and size are not significantly

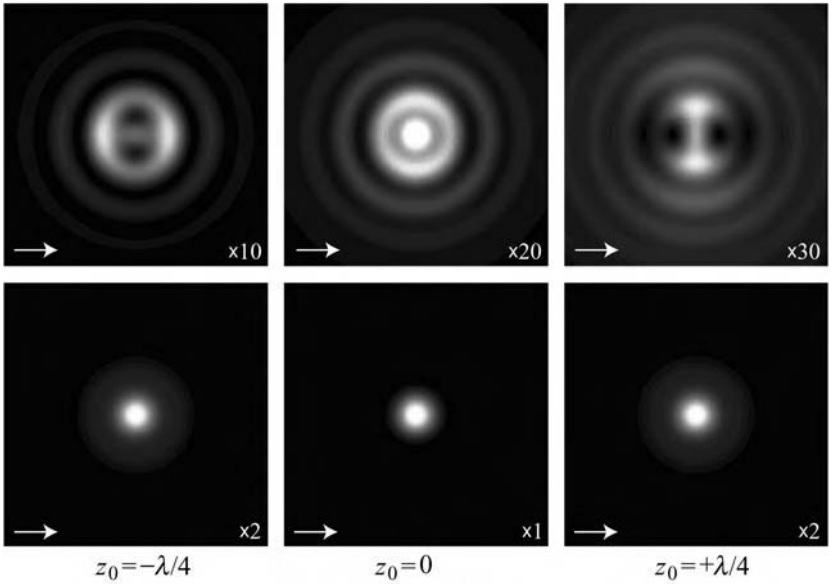


Figure 3.19 Reflected images of a diffraction-limited focused spot. The spot is moved in steps of $\lambda/4$ across the interface. z_0 is positive (negative) when the focus is below (above) the interface. The primary focusing objective lens has a numerical aperture of 1.4. The index of refraction is $n_1 = 1.518$ and the filling factor $f_0 = 2$. The upper row shows the situation for a glass/air interface ($n_2 = 1$) and the lower row for a glass/metal interface ($\varepsilon_2 \rightarrow -\infty$). Large aberrations are observed in the case of the glass/air interface because the totally internally reflected plane wave components generate a second virtual focus above the interface. The arrow indicates the direction of polarization of the primary incoming beam.

affected by the defocus. However, as shown in the upper row in Fig. 3.19 the situation is very different if the medium beyond the interface has a lower index than the focusing medium, i.e. if $n_2 < n_1$. In this case, the reflected spot changes strongly as a function of defocus. The spot shape deviates considerably from a Gaussian spot and resembles the spot of an optical system with axial astigmatism. The overall size of the spot is increased and the polarization is not preserved since I_{0r} and I_{2r} are of comparable magnitude. The patterns displayed in Fig. 3.19 can be verified in the laboratory. However, some care has to be applied when using dichroic beamsplitters since they have slightly different characteristics for s- and p-polarized light. In fact, the patterns in Fig. 3.19 depend sensitively on the relative magnitudes of the two superposed polarizations. Using a polarizer in the reflected beam path allows us to examine the two polarizations separately as shown in Fig. 3.20. Notice that the focus does not coincide with the interface when the intensity of the reflected

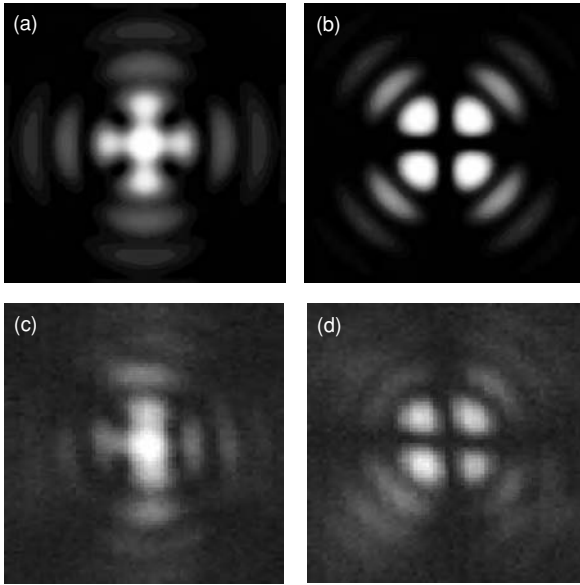


Figure 3.20 Decomposition of the in-focus reflected image (center image of Fig. 3.19) into two orthogonal polarizations. (a), (c) polarization in direction of incident polarization (\mathbf{n}_x); (b), (d) polarization perpendicular to incident polarization (\mathbf{n}_y). (a), (b) are calculated patterns and (c), (d) are experimental patterns.

pattern is maximized. The focus coincides with the interface when the *center* of the reflected pattern ($I_0(\rho, z)$) has maximum intensity. The images in Figs. 3.19 and 3.20 display the electric energy density, which is the quantity that is detected by optical detectors such as a CCD. On the other hand, the *total* energy density, and the magnitude of the time-averaged Poynting vector, render rotationally symmetric patterns.

How can we understand the appearance of the highly aberrated spot in the case of a glass air interface? The essence lies in the nature of total internal reflection. All plane wave components with angles of incidence in the range $[0 \dots \theta_c]$, θ_c being the critical angle of total internal reflection ($\approx 41.2^\circ$ for a glass air interface), are partly transmitted and partly reflected at the interface. Both reflection coefficients r_s and r_p are real numbers and there are no phase shifts between incident and reflected waves. On the other hand, the plane wave components in the range $[\theta_c \dots \theta_{\max}]$ are totally reflected at the interface. In this case the reflection coefficients become complex valued functions imposing a phase shift between incident and reflected waves. This can be viewed as an additional path difference between incident and

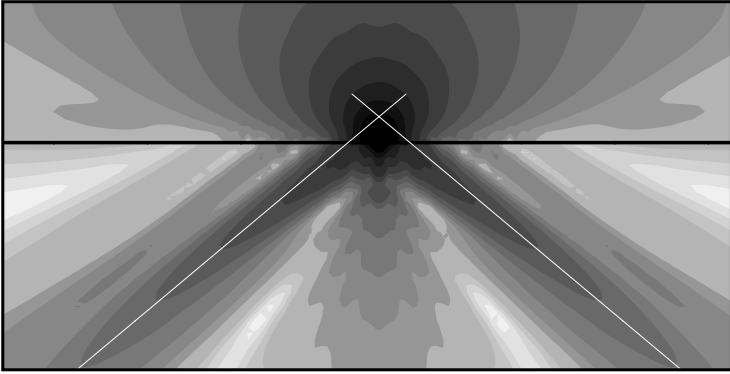


Figure 3.21 Scattered radiation (reflected and transmitted) of a laser focused on a glass/air interface. Same parameters as in Fig. 3.17. The lines indicate the apparent direction of radiation as seen by an observer in the far-field. The lines intersect in a *virtual focus* located $\approx 0.5\lambda$ above the interface. While all plane wave components in the angular range $[0 \dots \theta_c]$ originate from the focal point on the interface, the supercritical plane wave components emerge from an apparent spot above the interface giving rise to the aberrations in Fig. 3.19. Image size: $16\lambda \times 31\lambda$.

reflected waves similar to the Goos Hanchen shift [19]. It displaces the apparent reflection point beyond the interface thereby creating a second, virtual focus [20]. In order to visualize this effect we plot in Fig. 3.21 only the scattered field (transmitted and reflected) of Fig. 3.17. If we detected this radiation on the surface of an enclosing sphere with large radius, the direction of radiation would appear as indicated by the two lines which obviously intersect above the interface. Although all reflected radiation originates at the interface, there is an apparent origin above the interface. If we follow the radiation maxima from the far-field towards the interface we see that close to the interface the radiation bends towards the focus to ensure that the origin of radiation comes indeed from the focal spot.

We thus find the important result that the reflected light associated with the angular range $[0 \dots \theta_c]$ originates from the real focal point on the interface, whereas the light associated with $[\theta_c \dots \theta_{\max}]$ originates from a virtual point located above the interface. To be correct, the “virtual” point above the interface is not really a geometrical point. Instead, it is made of many points distributed along the vertical axis. The waves that emanate from these points have different relative phases and give rise to a conically shaped wavefront similar to the Mach cone in fluid dynamics. The resulting toroidal aberration was first investigated by Maeker and Lehman [21].

The observation of the aberrations in the focal point's reflected image has important consequences for reflection-type confocal microscopy and data sampling. In these techniques the reflected beam is focused onto a pinhole in the image plane. Because of the aberrations of the reflected spot, most of the reflected light is blocked by the pinhole destroying the sensitivity and resolution. However, it has been pointed out that this effect can dramatically increase the contrast between metallic and dielectric sample features [20] because the reflected spot from a metal interface appears to be aberration free. Finally, it has to be emphasized that the real focal spot on the interface remains greatly unaffected by the interface; the aberrations are only associated with the reflected image. The understanding of the patterns in Figs. 3.19 and 3.20 proves to be very valuable for the alignment of an optical system, for example to ensure that the focal plane of a laser coincides with the glass/air interface (object plane)

Problems

- 3.1 The paraxial Gaussian beam is not a rigorous solution of Maxwell's equations. Its field is therefore not divergence free ($\nabla \cdot \mathbf{E} \neq 0$). By requiring $\nabla \cdot \mathbf{E} = 0$ one can derive an expression for the longitudinal field E_z . Assume that $E_y = 0$ everywhere and derive E_z to lowest order for which the solution is non-zero. Sketch the distribution of $|E_z|^2$ in the focal plane.
- 3.2 Determine the decomposition of an arbitrary optical field into transverse electric (TE) and transverse magnetic (TM) fields. The longitudinal field E_z vanishes for the TE field, whereas H_z vanishes for the TM field.
- 3.3 Consider the fields emerging from a truncated hollow metal waveguide with a square cross-section and with ideally conducting walls. The side length a_0 is chosen in such a way that only the lowest order TE_{10} mode polarized in the x -direction is supported. Assume that the fields are not influenced by the edges of the truncated side walls.
 - (a) Calculate the spatial Fourier spectrum of the electric field in the exit plane ($z = 0$).
 - (b) Calculate and plot the corresponding far-field ($\mathbf{E} \cdot \mathbf{E}^*$).
- 3.4 Verify that energy is conserved for a strongly focused Gaussian beam as described in Section 3.6. To do this, compare the energy flux through transverse planes on both sides of the optical lens. It is an advantage to choose one plane at the origin of the focus ($z = 0$). The energy flux is calculated most conveniently by evaluating the z -component of the time-averaged Poynting vector $\langle S_z \rangle$ and integrating it over the area of the transverse plane. Hint: You will need the Bessel function closure relation

$$\int_0^\infty J_n(a_1 b x) J_n(a_2 b x) x dx = \frac{1}{a_1 b^2} \delta(a_1 - a_2). \quad (3.112)$$

Check the units!

- 3.5 Consider a small circular aperture with radius a_0 in an infinitely thin and ideally conducting screen which is illuminated by a plane wave at normal incidence and polarized along the x -axis. In the long wavelength limit ($\lambda \gg a_0$) the electric field in

the aperture ($z=0$, $x^2+y^2 \leq a_0^2$) has been derived by Bouwkamp [22] as

$$\begin{aligned} E_x(x, y) &= -\frac{4ik E_0}{3\pi} \frac{2a_0^2 - x^2 - 2y^2}{\sqrt{a_0^2 - x^2 - y^2}}, \\ E_y(x, y) &= -\frac{4ik E_0}{3\pi} \frac{xy}{\sqrt{a_0^2 - x^2 - y^2}}, \end{aligned} \quad (3.113)$$

where E_0 is the incident field amplitude. The corresponding spatial Fourier spectrum has been calculated by Van Labeke *et al.* [23] as

$$\hat{E}_x(k_x, k_y) = \frac{2ik a_0^3 E_0}{3\pi^2} \left[\frac{3k_y^2 \cos(a_0 k_\rho)}{a_0^2 k_\rho^4} - \frac{(a_0^2 k_x^4 + 3k_y^2 + a_0^2 k_x^2 k_y^2) \sin(a_0 k_\rho)}{a_0^3 k_\rho^5} \right] \quad (3.114)$$

$$\hat{E}_y(k_x, k_y) = -\frac{2ik a_0^3 E_0}{3\pi^2} \left[\frac{3k_x k_y \cos(a_0 k_\rho)}{a_0^2 k_\rho^4} - \frac{k_x k_y (3 - a_0^2 k_\rho^2) \sin(a_0 k_\rho)}{a_0^3 k_\rho^5} \right], \quad (3.115)$$

with $k_\rho = (k_x^2 + k_y^2)^{1/2}$ being the transverse wavenumber.

- (a) Derive the Fourier spectrum of the longitudinal field component E_z .
- (b) Find expressions for the field $\mathbf{E} = (E_x, E_y, E_z)$ at an arbitrary field point (x, y, z) .
- (c) Calculate the far-field and express it in spherical coordinates (r, ϑ, φ) and spherical vector components $\mathbf{E} = (E_r, E_\vartheta, E_\varphi)$. Expand in powers of ka_0 and retain only the lowest orders. What does this field look like?

- 3.6 The reflected image of a laser beam focused on a dielectric interface is given by Eqs. (3.109)–(3.111). Derive these equations starting from Eq. (3.100) which is the collimated reflected field. Notice that the fields propagate in the negative z -direction.
- 3.7 Show that the field \mathbf{E} defined through \mathbf{E}_f , \mathbf{E}_r , and \mathbf{E}_t in Section 3.9 fulfills the boundary conditions at the interface $z = z_0$. Furthermore, show that the Helmholtz equation and the divergence condition are fulfilled in each of the two half-spaces.
- 3.8 In order to correct for the aberrations introduced by the reflection of a strongly focused beam from an interface we design a pair of phase plates. By using a polarizing beamsplitter, the collimated reflected beam (cf. Fig. 3.18 and Eq. (3.100)) is split into two purely polarized light paths. The phase distortion in each light path is corrected by a phase plate. After correction, the two light paths are recombined and refocused on the image plane. Calculate and plot the phase distribution of each phase plate if the incident field is a Gaussian beam ($f_0 \rightarrow \infty$) focused by an NA = 1.4 objective on a glass air interface ($z_0 = 0$) and incident from the optically denser medium with $n_1 = 1.518$. What happens if the focus is displaced from the interface ($z_0 \neq 0$)?

References

- [1] M. Muller, J. Squier, K. R. Wilson, and G. J. Brakenhoff, “3D microscopy of transparent objects using third-harmonic generation,” *J. Microsc.* **191**, 266–274 (1998).
- [2] A. E. Siegman, *Lasers*. Mill Valley, CA: University Science Books (1986).

- [3] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics*, New York: Cambridge University Press (1995).
- [4] E. Zauderer, "Complex argument Hermite-Gaussian and Laguerre-Gaussian beams," *J. Opt. Soc. Am. A* **3**, 465–469 (1986).
- [5] E. J. Bochove, G. T. Moore, and M. O. Scully, "Acceleration of particles by an asymmetric Hermite-Gaussian laser beam," *Phys. Rev. A* **46**, 6640–6653 (1992).
- [6] X. S. Xie and J. K. Trautman, "Optical studies of single molecules at room temperature," *Annu. Rev. Phys. Chem.* **49**, 441–480 (1998).
- [7] L. Novotny, M. R. Beversluis, K. S. Youngworth, and T. G. Brown, "Longitudinal field modes probed by single molecules," *Phys. Rev. Lett.* **86**, 5251–5254 (2001).
- [8] A. Ashkin, J. M. Dziedzic, J. E. Bjorkholm, and S. Chu, "Observation of a single-beam gradient force optical trap for dielectric particles," *Opt. Lett.* **11**, 288–290 (1986).
- [9] E. Wolf, "Electromagnetic diffraction in optical systems. I. An integral representation of the image field," *Proc. Roy. Soc. A* **253**, 349–357 (1959).
- [10] B. Richards and E. Wolf, "Electromagnetic diffraction in optical systems. II. Structure of the image field in an aplanatic system," *Proc. Roy. Soc. A* **253**, 358–379 (1959).
- [11] K. S. Youngworth and T. G. Brown, "Focusing of high numerical aperture cylindrical-vector beams," *Opt. Express* **7**, 77–87 (2000).
- [12] K. S. Youngworth and T. G. Brown, "Inhomogeneous polarization in scanning optical microscopy," *Proc. SPIE* **3919**, 75–85 (2000).
- [13] R. Dorn, S. Quabis, and G. Leuchs, "Sharper focus for a radially polarized light beam," *Phys. Rev. Lett.* **91**, 233901 (2003).
- [14] L. Novotny, E. J. Sanchez, and X. S. Xie, "Near-field optical imaging using metal tips illuminated by higher-order Hermite-Gaussian beams," *Ultramicroscopy* **71**, 21–29 (1998).
- [15] M. J. Snadden, A. S. Bell, R. B. M. Clarke, E. Riis, and D. H. McIntyre, "Doughnut mode magneto-optical trap," *J. Opt. Soc. Am. B* **14**, 544–552 (1997).
- [16] S. C. Tidwell, D. H. Ford, and D. Kimura, "Generating radially polarized beams interferometrically," *Appl. Opt.* **29**, 2234–2239 (1990).
- [17] W. C. Chew, *Waves and Fields in Inhomogeneous Media*. New York: Van Nostrand Reinhold (1990).
- [18] B. Sick, B. Hecht, and L. Novotny, "Orientational imaging of single molecules by annular illumination," *Phys. Rev. Lett.* **85**, 4482–4485 (2000).
- [19] J. D. Jackson, *Classical Electrodynamics*, New York: John Wiley & Sons, 3rd edn. (1998).
- [20] K. Karrai, X. Lorenz, and L. Novotny, "Enhanced reflectivity contrast in confocal solid immersion lens microscopy," *Appl. Phys. Lett.* **77**, 3459–3461 (2000).
- [21] H. Mauecker and G. Lehmann, "Die Grenze der Totalreflexion. I-III," *Ann. Physik* **10**, 115–128, 153–160, and 161–166 (1952).
- [22] C. J. Bouwkamp, "On Bethe's theory of diffraction by small holes," *Philips Res. Rep.* **5**, 321–332 (1950).
- [23] D. Van Labeke, D. Barchiesi, and F. Baida, "Optical characterization of nanosources used in scanning near-field optical microscopy," *J. Opt. Soc. Am. A* **12**, 695–703 (1995).