2 Theoretical Concepts

Highlights of this chapter: Derivation of general concepts of electrodynamics, different ways to represent electromagnetic fields, discussion of Gaussian beams, introduction of evanescent fields, theory of image formation and resolution limits.

2.1 "Wrap-up" of classical electrodynamics

2.1.1 Maxwell's Equations

At the very beginning of any problem in optics there are Maxwell's equations. We use SI units unless stated otherwise:

$$\nabla \times E(r,t) = -\frac{\partial B(r,t)}{\partial t} \tag{1}$$

$$\nabla \times H(r,t) = \frac{\partial D(r,t)}{\partial t} + j(r,t) \tag{2}$$

$$\nabla \cdot D(r,t) = \rho(r,t) \tag{3}$$

$$\nabla \cdot B(r,t) = 0 \tag{4}$$

where E is the electric field, D the electric displacement, H the magnetic field and B the magnetic induction. The initial problem to find 16 unknown vector components can be highly simplified (depending on the medium).

Example: In an isotropic, homogeneous, source-free Medium, the electromagnetic field is described by *two scalar fields* only.

Conservation of charge is implicitly contained in Maxwell's equation. It is possible to derive the continuity equation:

$$\nabla \cdot j(r,t) + \frac{\partial \rho(r,t)}{\partial t} = 0 \tag{5}$$

Electromagnetic properties of (macroscopic) matter are included via the polarization P and the magnetization M:

$$\begin{aligned} D(r,t) &= \epsilon_0 E(r,t) + P(r,t) && (6) \\ H(r,t) &= \mu_0^{-1} B(r,t) - M(r,t) && (7) \end{aligned}$$

These equations do not depend on the medium and are always valid!

2.1.2 Wave equation

By inserting equation 6 in Maxwell's equation two inhomogeneous wave equations are obtained:

$$\begin{aligned} \nabla \times \nabla \times E(r,t) + \frac{1}{c^2}\frac{\partial^2 E(r,t)}{\partial t^2} &= -\mu_0 \frac{\partial}{\partial t}\left(j + \frac{\partial P}{\partial t} + \nabla \times M\right) && (8) \\ \nabla \times \nabla \times H(r,t) + \frac{1}{c^2}\frac{\partial^2 H(r,t)}{\partial t^2} &= \nabla \times j + \nabla \times \frac{\partial P}{\partial t} + \mu_0 \frac{\partial M}{\partial t} && (9) \end{aligned}$$

Also these equations are valid irrespective of the specific properties of the medium.

2.1.3 Linear media

In linear, isotropic media, there is the following relation between D and E and H and B, respectively:

$$\begin{aligned} D &= \epsilon_0 \epsilon E \qquad (P = \epsilon_0 \chi_e E) && (10) \\ B &= \mu_0 \mu H \qquad (M = \chi_m H) && (11) \\ j_c &= \sigma E && (12) \end{aligned}$$

Generally, ϵ and μ may depend on the spatial coordinate (inhomogeneous media) and/or on the frequency (dispersive media).

2.1.4 Time-dependent fields

The spectrum of an arbitrary electric field $E(r,t)$ is defined by the Fourier transform $\widehat{E}(r,\omega)$:

$$\widehat{E}(r,\omega) = \frac{1}{\sqrt{2\pi}} \int_{\infty}^{-\infty} E(r,t) e^{i\omega t} dt \qquad (13)$$

Maxwell's equation can be Fourier-transformed to find:

$$\begin{aligned}
\nabla \times \widehat{E}(r,\omega) &= i\omega \widehat{B}(r,\omega) && (14)\\
\nabla \times \widehat{H}(r,\omega) &= -i\omega \widehat{D}(r,\omega) + \widehat{j}(r,\omega) && (15)\\
\nabla \cdot \widehat{D}(r,\omega) &= \widehat{\rho}(r,\omega) && (16)\\
\nabla \cdot \widehat{B}(r,\omega) &= 0 && (17)
\end{aligned}$$

From the solution of the Fourier-transformed equations it is possible to find $E(r,t)$ via inverse Fourier transform:

$$E(r,t) = \frac{1}{\sqrt{2\pi}} \int_{\infty}^{-\infty} \widehat{E}(r,\omega) e^{-i\omega t} dt \qquad (18)$$

In the *special case of a monochromatic field*

$$E(r,t) = \mathrm{Re}\left\{E(r)e^{-i\omega t}\right\} = 1/2\left(E(r)e^{-i\omega t} + E^{*}(r)e^{i\omega t}\right) \qquad (19)$$

one finds the following Maxwell's equations:

$$\begin{aligned}
\nabla \times E(r) &= i\omega B(r) && (20)\\
\nabla \times H(r) &= -i\omega D(r) + j(r) && (21)\\
\nabla \cdot D(r) &= \rho(r) && (22)\\
\nabla \cdot B(r) &= 0 && (23)
\end{aligned}$$

This is equivalent to 14. Therefore, the solution $E(r)$ is equivalent to the solution $\widehat{E}(r,\omega)$ for one frequency component of an arbitrary time-dependent field.

2.1.5 Boundary conditions

At the boundary between two media the material properties change in a discontinuous way. As an example we take the boundary of two media indexed with i, j. Within each medium i or j the *inhomogeneous Helmholtz equations* are valid:

$$\begin{aligned}
\left(\nabla^2 + k_i^2\right) E_i &= -i\omega\mu_0\mu_i j_i + \frac{\nabla\rho_i}{\epsilon_0\epsilon_i} && (24)\\
\left(\nabla^2 + k_i^2\right) H_i &= \nabla \times j_i && (25)
\end{aligned}$$

These can easily be derived from Maxwell's equations taking into account the identity $\nabla \times \nabla\times = -\nabla^2 + \nabla\nabla\cdot$. Here we define the wavenumbers $k_i = (\omega/c)\sqrt{\mu_i\epsilon_i}$.

Often the medium is source-free and the inhomogeneous equation reduce to the *homogeneous Helmholtz equations.*

In order to derive the boundary conditions it is convenient to write Maxwell's equation in integral form:

$$\int_{\partial S} E(r,t)\ ds = -\int_S \frac{\partial}{\partial t} B(r,t) \cdot n_S\ da \tag{26}$$

$$\int_{\partial S} H(r,t)\ ds = \int_S \left[j(r,t) + \frac{\partial}{\partial t} D(r,t) \right] \cdot n_S\ da \tag{27}$$

$$\int_{\partial V} D(r,t) \cdot n_S\ da = \int_V \rho(r,t)\ dV \tag{28}$$

$$\int_{\partial V} B(r,t) \cdot n_S\ da = 0 \tag{29}$$

Here V, S and $\partial V, \partial S$ are the volumes and surfaces, or their borders, respectively. Applied to a sufficiently small area Maxwell's equation provide boundary conditions for the tangential components (with the surface current density K)

$$n \times (E_i - E_j) = 0 \qquad on\ \partial D_{ij} \tag{30}$$

$$n \times (H_i - H_j) = K \qquad on\ \partial D_{ij} \tag{31}$$

and the normal components:

$$n \cdot (D_i - D_j) = \sigma_i \qquad on\ \partial D_{ij} \tag{32}$$

$$n \cdot (B_i - B_j) = 0 \qquad on\ \partial D_{ij} \tag{33}$$

The following figure 3 illustrates the geometry:

2.1.6 Energy conservation

From Maxwell's equation one finds after a simple transformation:

$$\int_{\partial V} (E \times H) \cdot n\ da = -\int_V \left[H \cdot \frac{\partial B}{\partial t} + E \cdot \frac{\partial D}{\partial t} + j \cdot E \right]\ dV \tag{34}$$

For a linear medium we find by inserting

$$\int_{\partial V} (E \times H) \cdot n\ da + \frac{1}{2} \frac{\partial}{\partial t} \int_V \left[D \cdot E + B \cdot H \right]\ dV = \tag{35}$$

$$-\int_V j \cdot E\ dV - \frac{1}{2} \int_V \left[E \cdot \frac{\partial P}{\partial t} - P \cdot \frac{\partial E}{\partial t} \right]\ dV - \frac{\mu_0}{2} \int_V \left[H \cdot \frac{\partial M}{\partial t} - M \cdot \frac{\partial H}{\partial t} \right]\ dV$$

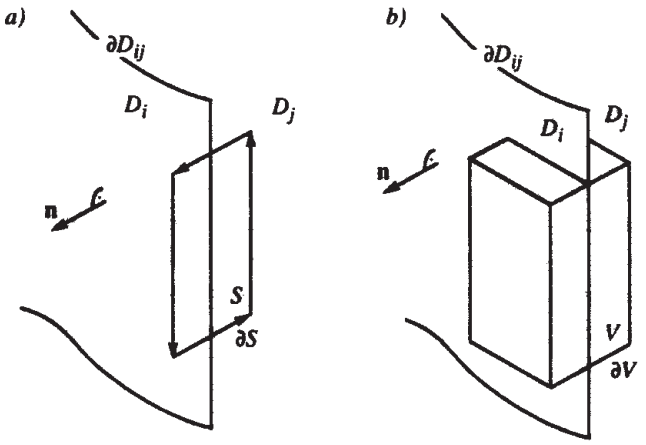

Figure 3: Sketch to derive boundary conditions.

The *Poynting Theorem* is an interpretation of this equation, in a sense that the first term on the left side is the energy flux density (Poynting-Vector $S = E \times H$) describing the flux out of the volume V, whereas the second term on the left side is the change of the energy inside the volume V and the term on the right side is the dissipation within V.

An important quantity is the time average of the Poynting-Vector:

$$\langle S \rangle = \frac{1}{2} \operatorname{Re} \{ E \times H^* \} \tag{36}$$

In the far field the electric and magnetic fields are transversal. Therefore:

$$\langle S \rangle_{dist} = \frac{1}{2} \sqrt{\frac{\epsilon_0 \epsilon}{\mu_0 \mu}} |E|^2 n_r \tag{37}$$

2.2 Representations of electromagnetic fields

2.2.1 Dyadic Green's function

General Concept:
Green's functions provide a method to find solution of inhomogeneous linear differential equation of the general form:

$$\mathcal{L}\mathbf{A}(r) = \mathbf{B}(r) \tag{38}$$

where $\mathcal{L}$ is a linear operator and $A(r)$ and $B(r)$ are vector fields. In electromagnetism $B(r)$ can be represented as a source field and $A(r)$ an unknown response. A solution of the equation above is a solution of the homogeneous equation A_0 for $B = 0$ and an arbitrary particular solution.

The idea of the Green's function ansatz is to first find a solution for a special δ-like problem:

$$\mathcal{L}\overleftrightarrow{\mathbf{G}}(r, r') = \overleftrightarrow{\mathbf{I}}\,\delta(r - r') \tag{39}$$

In this equation $\mathcal{L}$ acts on each column of $\overleftrightarrow{\mathbf{G}}(r, r')$ independently. $\overleftrightarrow{\mathbf{G}}(r, r')$ is the *Dyadic Green's function.*

A solution of the general inhomogeneous equation can then be found as

$$\mathbf{A}(r) = \int_V \overleftrightarrow{\mathbf{G}}(r, r')\mathbf{B}(r')\, dV' \tag{40}$$

Remark:
Care has to be taken with singular points. Often $\vec{G}(r, r')$ is singular at $r = r'$. In this case an infinitesimal volume has to excluded from the integration and treated separately. For now, we assume that the field points we consider are outside the source volume.

Green's function for the electric field:
We start with expressing the time periodic (!) electric and magnetic field $\mathbf{E}(r)$ and $\mathbf{H}(r)$ via a scalar and vector potential $\phi(r)$ and $\mathbf{A}(r)$, respectively:

$$\mathbf{E}(r) = i\omega\mathbf{A}(r) - \nabla\phi(r) \tag{41}$$

$$\mathbf{H}(r) = \frac{1}{\mu_0\mu}\nabla \times \mathbf{A}(r) \tag{42}$$

With the Lorentz gauge condition

$$\nabla \cdot \mathbf{A}(r) = i\omega\mu_0\mu\epsilon_0\epsilon\phi(r) \tag{43}$$

inserting of the potentials into Maxwell's equations results in the inhomogeneous Helmholtz equations

$$\left[\nabla^2 + k^2\right]\mathbf{A}(r) = -\mu_0\mu\mathbf{j}(r) \tag{44}$$

$$\left[\nabla^2 + k^2\right]\phi(r) = -\rho(r)/\epsilon_0\epsilon \tag{45}$$

A scalar Green's function $G_0(r, r')$ can now be found as solution of the scalar problem:

$$\left[\nabla^2 + k^2\right]G_0(r, r') = -\delta(r - r\prime) \tag{46}$$

For the two potentials one finds

$$\mathbf{A}(r) = \mu_0\mu\int_V \mathbf{j}(r')G_0(r, r')\,dV' \tag{47}$$

$$\phi(r) = \frac{1}{\epsilon_0\epsilon}\int_V \rho(r')G_0(r, r')\,dV' \tag{48}$$

In free space the only physical solution for the scalar problem 46 is:

$$G_0(r, r') = \frac{\exp\left(\pm ik|r - r'|\right)}{4\pi|r - r'|} \tag{49}$$

The solution correspond to spherical waves propagating out of (+ sign) or into (- sign) the origin.

In order to find the Green's function for the electric field it is convenient to start with the wave equation in a homogeneous field:

$$\left[(\nabla \times \nabla \times) - k^2\right] \mathbf{E}(r) = i\omega\mu_0\mu \mathbf{j}(r) \tag{50}$$

From this we can formulate the general definition of the *dyadic Green's function for the electric field*:

$$\left[(\nabla \times \nabla \times) - k^2\right] \overleftrightarrow{\mathbf{G}}(r, r') = \overleftrightarrow{\mathbf{I}} \delta(r - r') \tag{51}$$

Note, that $\overleftrightarrow{\mathbf{G}}(r, r')$ is a tensor!

Then, we find the general solution for the electric field from the particular solution

$$\mathbf{E}(r) = i\omega\mu_0\mu \int_V \overleftrightarrow{\mathbf{G}}(r, r') \mathbf{j}(r)\, dV' \tag{52}$$

together with an arbitrary homogeneous solution $\mathbf{E}_0(r)$:

$$\mathbf{E}(r) = \mathbf{E}_0(r) + i\omega\mu_0\mu \int_V \overleftrightarrow{\mathbf{G}}(r, r') \mathbf{j}(r)\, dV' \quad r \notin V \tag{53}$$

and similar for the magnetic field:

$$\mathbf{H}(r) = \mathbf{H}_0(r) + \int_V \left[\nabla \times \overleftrightarrow{\mathbf{G}}(r, r')\right] \mathbf{j}(r)\, dV' \quad r \notin V \tag{54}$$

The equations above are called the *volume integral equations* and are the starting point of various other formalisms.

In order to solve the equations above for a specific distributions of currents, an explicit expression for $\overleftrightarrow{\mathbf{G}}(r, r')$ has to be found. We can start with expressing the electric field via the vector potential in Lorentz gauge. From equation 42 one finds:

$$\mathbf{E}(r) = i\omega \left[1 + \frac{1}{k^2} \nabla\nabla\cdot\right] \mathbf{A}(r) \tag{55}$$

The first column vector of $\overleftrightarrow{\mathbf{G}}$, defined in equation 51, i.e. $\mathbf{G}_x$, is simply the electric field due to a point source current $\mathbf{j} = (i\omega\mu_0)^{-1}\delta(r - r')\mathbf{n}_x$. The vector potential originating from this source is

$$\mathbf{A} = (i\omega)^{-1} G_0(r, r')\mathbf{n}_x \tag{56}$$

Inserting this one finds

$$\mathbf{G}_x(r, r') = \left[1 + \frac{1}{k^2}\nabla\nabla\cdot\right] G_0(r, r')\mathbf{n}_x \tag{57}$$

with similar expressions for $\mathbf{G}_y$ and $\mathbf{G}_z$.
Finally, with the definition of $\nabla \cdot \left[G_0 \overleftrightarrow{I}\right] = \nabla G_0$ one finds:

$$\overleftrightarrow{\mathbf{G}}(r, r') = \left[\overleftrightarrow{I} + \frac{1}{k^2}\nabla\nabla\cdot\right] G_0(r, r') \tag{58}$$

We will use the Green's function approach in some of the following chapters.

2.2.2 Angular spectrum representation

Electromagnetic fields can be represented in various ways, e.g., in different bases. It depends on the specific problem or geometry to decide which representation is most appropriate.

The *angular spectrum representation* is most convenient if propagating beams in homogeneous media or at boundaries between homogeneous media are considered. In this context we understand evanescent waves (with a complex component of the k-vector) as plane waves as well.

We consider a general problem. i.e. the coherent scattering of light by an arbitrary object. The scattered fields may be collected at a screen or detector. Figure 4 illustrates the geometry.

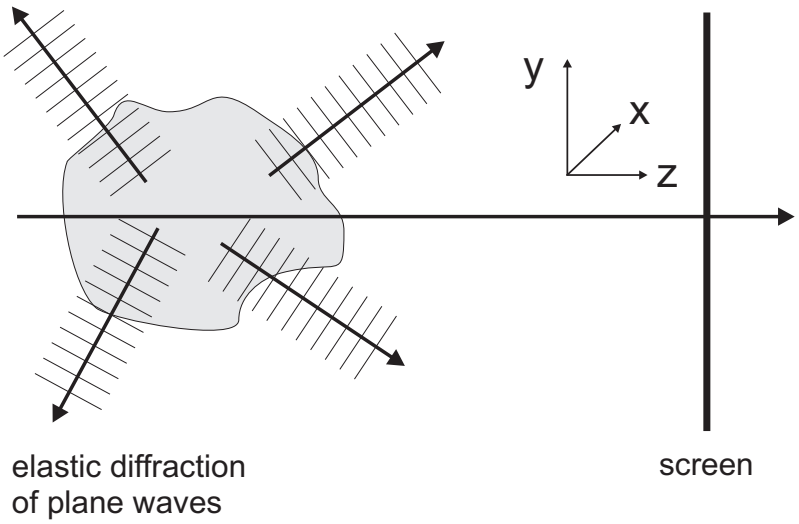

Figure 4: Schematic of plane wave diffraction on an arbitrary object.

First, the arbitrary field $E(x, y, z)$ is written with the help of its Fourier components in the $x - y$-plane. The z-direction is the direction of propagation:

$$E(x, y, z) = \frac{1}{2\pi} \iint \widehat{E}(k_x, k_y, z) e^{i(k_x x + k_y y)} \, dk_x dk_y \tag{59}$$

In a homogeneous, isotropic, source-free medium the homogeneous Helmholtz equation is valid:

$$(\nabla^2 + k^2) E(x, y, z) = 0 \tag{60}$$

Inserting leads to:

$$\widehat{E}(k_x, k_y, z) = \widehat{E}(k_x, k_y, 0) e^{ik_z} \tag{61}$$

Thus:

$$E(x, y, z) = \frac{1}{2\pi} \iint \widehat{E}(k_x, k_y, 0) e^{i(k_x x + k_y y + k_z z)} \, dk_x dk_y \tag{62}$$

This representation is the angular spectrum representation. It is straightforward to

evaluate $E(x,y,z)$ from $E(x,y,0)$ by inserting the Fourier components:

$$E(x,y,z) = \tag{63}$$
$$\frac{1}{4\pi^2}\iint_{-\infty}^{\infty} E(x',y',0)e^{-i(k_x x'+k_y y')}\,dx'dy'e^{i(k_x x+k_y y+k_z z)}\,dk_x dk_y$$
$$= \tag{64}$$
$$\frac{1}{4\pi^2}\iint_{-\infty}^{\infty} E(x',y',0)\iint_{-\infty}^{\infty} e^{i(k_x(x-x')+k_y(y-y')+k_z z)}\,dk_x dk_y dx'dy'$$

$E(x,y,z)$ is a convolution of $E(x,y,0)$ with the *optical transfer function*

$$H(x-x',y-y') = \iint_{-\infty}^{\infty} e^{i(k_x(x-x')+k_y(y-y')+k_z z)}\,dk_x dk_y \tag{65}$$

$H(x-x',y-y')$ can also be expressed in reciprocal space after Fourier transformation:

$$\widehat{E} = \widehat{E'}\cdot\widehat{H} \tag{66}$$

Obviously:

$$\widehat{H} = e^{i(k_x x+k_y y+k_z z)} \tag{67}$$

With

$$k_z = \sqrt{k^2-k_x^2-k_y^2} = 2\pi\sqrt{1/\lambda^2-\rho^2} \tag{68}$$

it is apparent that the spatial Fourier components propagate as *plane waves* or as *evanescent waves* (with a complex k_z). Information about spatial frequencies with

$$\Delta x < 1/k = \frac{\lambda}{2\pi n} \tag{69}$$

decay exponentially!

This information is only available in the **optical near-field**!

As already mentioned the angular spectrum representation is convenient when calculating the transmission or reflection at interfaces or boundaries. Here, the Fresnel coefficients for transmission and reflections can be utilized.

Reminder: The Fresnel coefficients describe the reflection (r) and transmission (t) for $s-$ and $p-$ polarization for a wave traveling from the medium i to the medium

t under the incidence (transmission) angle ϑ_i (ϑ_t):

$$r_s = \frac{\frac{n_i}{\mu_i}\cos\vartheta_i - \frac{n_t}{\mu_t}\cos\vartheta_t}{\frac{n_i}{\mu_i}\cos\vartheta_i + \frac{n_t}{\mu_t}\cos\vartheta_t} \tag{70}$$

$$t_s = \frac{2\frac{n_i}{\mu_i}\cos\vartheta_i}{\frac{n_i}{\mu_i}\cos\vartheta_i + \frac{n_t}{\mu_t}\cos\vartheta_t} \tag{71}$$

$$r_p = \frac{\frac{n_t}{\mu_t}\cos\vartheta_i - \frac{n_i}{\mu_i}\cos\vartheta_t}{\frac{n_i}{\mu_i}\cos\vartheta_i + \frac{n_t}{\mu_t}\cos\vartheta_t} \tag{72}$$

$$t_p = \frac{2\frac{n_i}{\mu_i}\cos\vartheta_i}{\frac{n_i}{\mu_i}\cos\vartheta_t + \frac{n_t}{\mu_t}\cos\vartheta_i} \tag{73}$$

An arbitrary problem of light incident on an interface of two media is illustrated in the following figure 5.

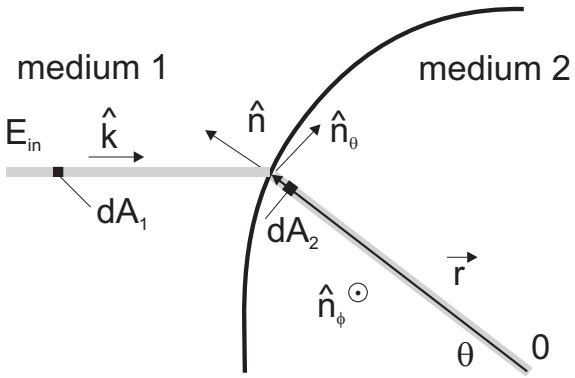

Figure 5: Arbitrary problem of light incident on an interface of two media.

The field in the medium 2 right after the interface can be written as:

$$E_2(x,y,z) = \left[t_s(\vartheta_n)\left|E_2^{(s)}\right| n_\varphi + t_p(\vartheta_n)\left|E_2^{(p)}\right| n_\vartheta\right]\sqrt{\frac{n_1}{n_2}}\,(\cos\vartheta)^{1/2} \tag{74}$$

where

$$E_2^{(s)} = (E_{inc}\cdot\hat{n}_\phi)\hat{n}_\phi \quad E_2^{(p)} = E_{inc}\cdot\hat{n}_\Theta)\hat{n}_\Theta \tag{75}$$

and

$$\cos\vartheta = \left(\widehat{k}\widehat{n}(x,y,z)\right) \tag{76}$$

The term after the bracket [ ] preserves the energy flux.

2.2.3 Evanescent fields

We have already encountered evanescent fields in the previous subsection. Evanescent fields often play a crucial role in the field of nanooptics where the interaction on a length scale far below the wavelength of light is of interest.

Evanescent fields are described as a plane wave of the form

$$E(r,t) = E_0 e^{i(kr-\omega t)} \tag{77}$$

where*at least one component of the wavevector* k is *imaginary.*

Evanescent waves occurs, e.g., at interfaces between homogeneous media. The following figure 6 shows an interface between two media with ε_1, μ_1 und ε_2, μ_2.

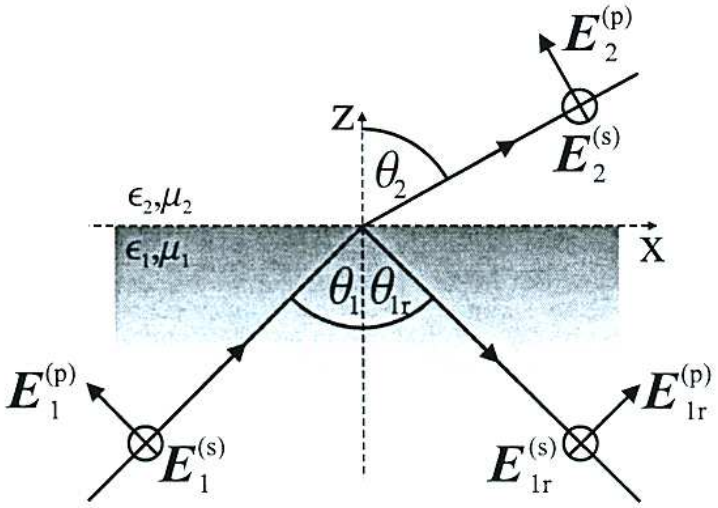

Figure 6: Transmission and reflection of a wave on a planar boundary.

Reflection is described by Snell's law and the Fresnel coefficients.

$$\sqrt{\epsilon_1\mu_1}\sin\vartheta_1 = \sqrt{\epsilon_2\mu_2}\sin\vartheta_2 \tag{78}$$

The transmitted field is

$$E_2 = \begin{pmatrix} -t^p E_1^p \cos\vartheta_2 \\ t^s E_1^s \\ t^p E_1^p \sin\vartheta_2 \end{pmatrix} e^{ik_{2,x}x} e^{ik_{2,z}z} e^{-i\omega t} \tag{79}$$

where $t^{s,p}$ are the Fresnel coefficients for s- and p-polarization, respectivey.

The relation between the components of $k_1 = 2\pi\sqrt{\epsilon_1\mu_1}/\lambda$ und $k_2 = 2\pi\sqrt{\epsilon_2\mu_2}/\lambda$ is:

$$\begin{aligned} k_{2,x} &= k_2 \sin\vartheta_2 = k_1 \sin\vartheta_1 \\ k_{2,z} &= k_2 \cos\vartheta_2 = k_2\sqrt{1 - \frac{\epsilon_1\mu_1}{\epsilon_2\mu_2}\sin^2\vartheta_1} \end{aligned} \tag{80}$$

Or:

$$k_{2,z} = (\epsilon_2\mu_2 k^2 - k_{||}^2)^{1/2} = k\sqrt{\epsilon_2\mu_2 - \epsilon_1\mu_1\sin^2\vartheta_1} \qquad with\ k = 2\pi/\lambda \tag{81}$$

There is a critical angle of incidence, if the root becomes negativ:

$$\vartheta_c = \arcsin\sqrt{\frac{\epsilon_2\mu_2}{\epsilon_1\mu_1}} \tag{82}$$

Example: glass/air $\epsilon_2 = 1$, $\epsilon_1 = 2.25$, $\mu_1 = \mu_2 = 1$ somit $\vartheta_c = 41,8°$

For $\vartheta_1 > \vartheta_c$ $k_{2,z}$ becomes imaginary. For the field we find:

$$E_2 = \begin{pmatrix} -it^p E_1^p \sqrt{\frac{\epsilon_1\mu_1}{\epsilon_2\mu_2}\sin^2\vartheta_1 - 1} \\ t^s E_1^s \\ t^p E_1^p \sqrt{\frac{\epsilon_1\mu_1}{\epsilon_2\mu_2}}\sin\vartheta_1 \end{pmatrix} e^{ik_{2,x}x} e^{-\alpha z} e^{-i\omega t} \tag{83}$$

with the decay constant

$$\alpha = k_2\sqrt{\frac{\epsilon_1\mu_1}{\epsilon_2\mu_2}\sin^2\vartheta_1 - 1} \tag{84}$$

The following figure 7 plots the decay of an evanescent field for two different incident angles.

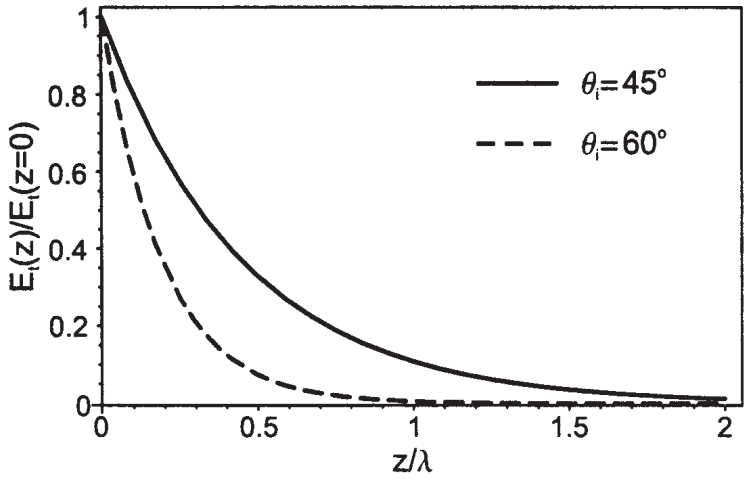

Figure 7: Decay of an evanescent field for two different incident angles.

Evanescent fields can be detected via a probe which is brought closer to the surface than $\approx \lambda/2$. This is the basis for *Scanning-Tunneling-Optical-Microscopy (STOM).*

Energy transport via evanescent fields:

If the incident angle is larger than the critical angle, then there is total internal reflection. No energy is transported to the optically thinner medium. This can be verified by calculating the Poynting vector:

$$\langle S \rangle_{2,z} = \frac{1}{2} \operatorname{Re}(E_{2,x} H^*_{2,y} - E_{2,y} H^*_{2,x}) = 0 \tag{85}$$

However, there is an energy transport parallel to the interface:

$$\langle S \rangle_{2,x} \propto e^{-\alpha z} \neq 0 \tag{86}$$

Frustrated total internal reflection:
Evanescent waves can be transferred into propagating waves when they interact with matter. This is the basic idea behind optical near-field microscopy since in this way evanescent fields can be detected in the far-field by an ordinary detector.
The following figure 8 illustrates reflection on a system consisting of three layers with different dielectric constants (ϵ_i, μ_i und $i \in \{1, 2, 3\}$):

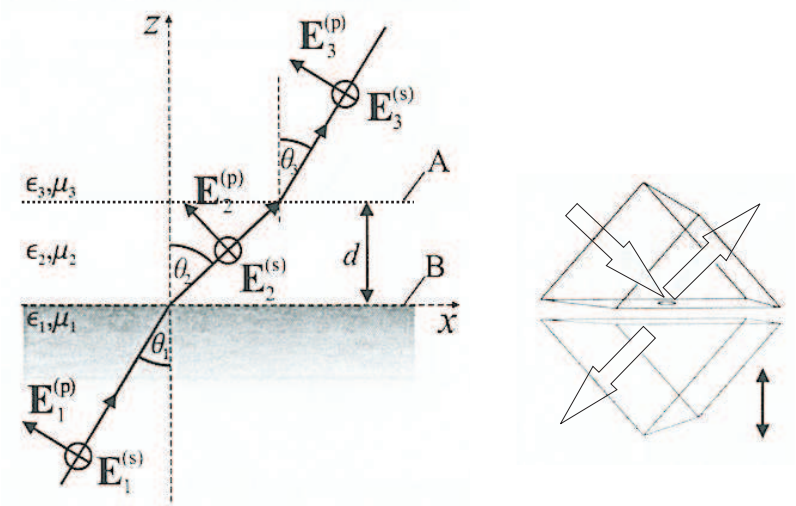

Figure 8: Left: Geometry of reflection on a stack of three layers. Right: Frustrated total internal reflection with a two-prism system.

The wave vector parallel to the interface is:

$$k_{j,z} = (\epsilon_i \mu_i k - k_{||}^2)^{1/2} = k(\epsilon_i \mu_i - \epsilon_1 \mu_1 \sin \vartheta_1)^{1/2} \qquad i \in \{1, 2, 3\} \tag{87}$$

For $n_2 < n_3 < n_1$ there are three possible cases:

1. $\vartheta_1 < \arcsin(n_2/n_1)$ or $k_{||} < n_2 k$: The fields propagate in each of the three media. A decetor in the far-field registers only a weak dependence (actually a modulation) of the intensity as a function of the thickness of layer 2.

2. $\arcsin(n_2/n_1) < \vartheta_1 < \arcsin(n_3/n_1)$ or $n_2 k < k_{||} < n_3 k$: The field can only propagate in medium 1 and medium 3. There is a strong dependence of the detected intensity as a function of the thickness of layer 2.

3. $\vartheta_1 > \arcsin(n_3/n_1)$ or $k_{||} > n_3 k$: Waves are evanescent, both in medium 2 and 3.

The following figure 9 illustrates the scenarios:

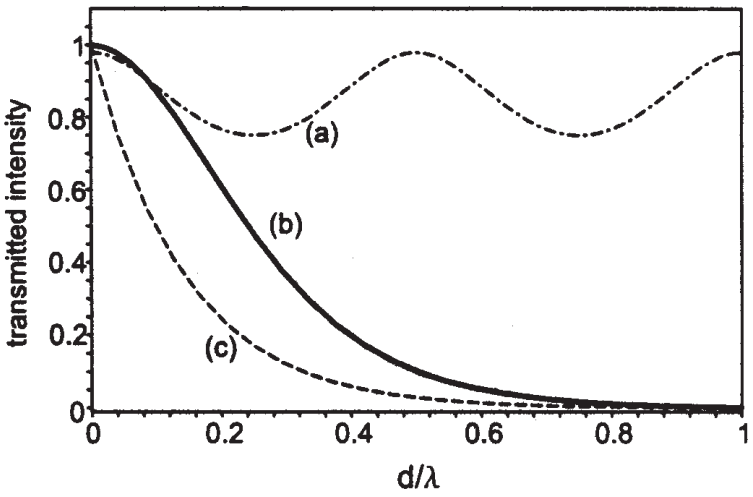

Figure 9: transmission of a p-polarized wave through a stack of three layers. a) and b) corresponds to the cases discussed in the text. c) is a purely exponential decay with the parameters from b), but without a third medium present.

Remark: The dependence of the intensity in case 2 differs from a pure exponential decay in case of the absence of medium 3. This is a first indication that the situation in the near-field may strongly depend on the probe!

2.2.4 Gaussian beams

In order to describe beam propagation in a more realistic way taking into account the finite transversal extension and the wave front curvature, the representation of Gaussian beams is most suitable. Here we would like to summarize their basic properties and discuss focusing issues.
In order to derive Gaussian beams one first starts with a Gaussian distribution of the electric field in an (x'-y'-)plane at $z' = 0$:

$$E(x', y', 0) = E_0 \exp\left(-\frac{x'^2 + y'^2}{w_0^2}\right) \tag{88}$$

The Fourier transfor also has Gaussian shape:

$$\widehat{E}(k_x, k_y, 0) = \widehat{E}_0 \frac{w_0^2}{2} \exp\left(-(k_x^2 + k_y^2) w_0^2/4\right) \tag{89}$$

Inserting into the angular spectrum representation results in:

$$E(x, y, 0) = E_0 \frac{w_0^2}{4\pi} e^{ik_z z} \iint e^{-(k_x^2+k_y^2)(w_0^2/4+iz/2k)} e^{i(k_x x + k_y y)} \, dk_x dk_y \tag{90}$$

Here, we used the *paraxial approximation*:

$$k_z = k\sqrt{1 - \frac{k_x^2 + k_y^2}{k^2}} \approx k - \frac{k_x^2 + k_y^2}{2k} \tag{91}$$

The integral can be evaluated and one finds the field of the fundamental Gaussian mode:

$$E(\rho, z) = E_0 \frac{w_0}{w(z)} e^{-\frac{\rho^2}{w^2(z)}} e^{i\left[kz - \eta(z) + k\rho^2/2R(z)\right]} \tag{92}$$

with

$$w(z) = w_0(1 + z^2/z_0^2)^{1/2} \qquad \text{beam waist} \tag{93}$$

$$R(z) = z(1 + z^2/z_0^2) \qquad \text{wavefront radius} \tag{94}$$

$$\eta(z) = \arctan(z/z_0) \qquad \text{phase correction (Gouy phase)} \tag{95}$$

with $\rho^2 = x^2 + y^2$ and the *Rayleigh length* $z_0 = kw_0^2/2$. Within the Rayleigh length the focussed spot increases by $\sqrt{2}$.

Asymptotically, one finds:

$$\lim_{z\to\infty}\left(\frac{w(z)}{z}\right) = \frac{2}{kw_0} = \frac{\lambda}{\pi w_0} \quad \text{or} \tag{96}$$

$$\tag{97}$$

$$NA = 2n/kw_0 \tag{98}$$

Remark:
The product

$$NA \cdot w_0 = 2n/k \tag{99}$$

is a constant.

The following figure 10 illustrates the properties of a Gaussian beam:

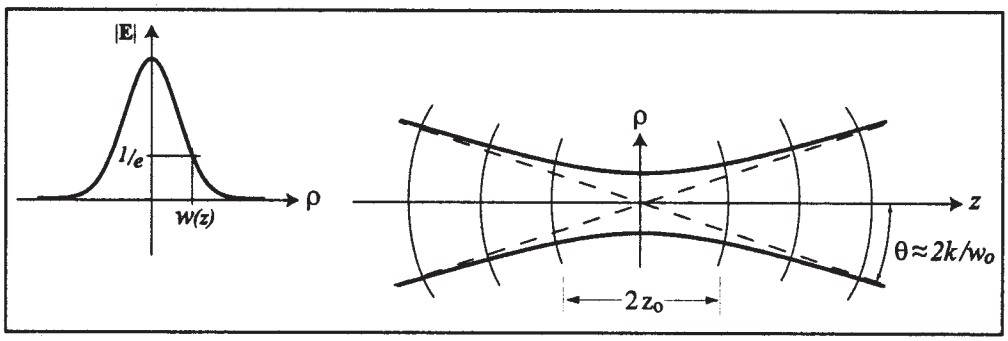

Figure 10: Properties of a Gaussian beam. From [Novotny and Hecht, "Principles of Nano-Optics"]

If the Gaussian beam passes a lens with focal length f at a distance z_l from the focus, then the beam transforms according to:

$$z_0' = \frac{f^2 z_0}{(z_l - f)^2 + z_0^2} \tag{100}$$

Higher order Gaussian modes
Higher order modes can be derived from the fundamental mode:

- Hermite-Gaussian-Modes (Eigenmodes in resonators with rectangular geometry):

$$E_{nm}^H(x,y,z) = w_0^{n+m} \frac{\partial^n}{\partial x^n} \frac{\partial^m}{\partial y^m} E(x,y,z) \tag{101}$$

- Laguerre-Gaussian modes (Eigenmodes in resonators with spherical geometry):

$$E_{nm}^L(x,y,z) = k^n w_0^{2n+m} e^{ikz} \frac{\partial^n}{\partial x^n} \left(\frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right)^m \left\{ E(x,y,z) e^{-ikz} \right\} \tag{102}$$

The following figure 11 shows the intensity distribution of different Hermite-Gaussian modes in the focal plane:

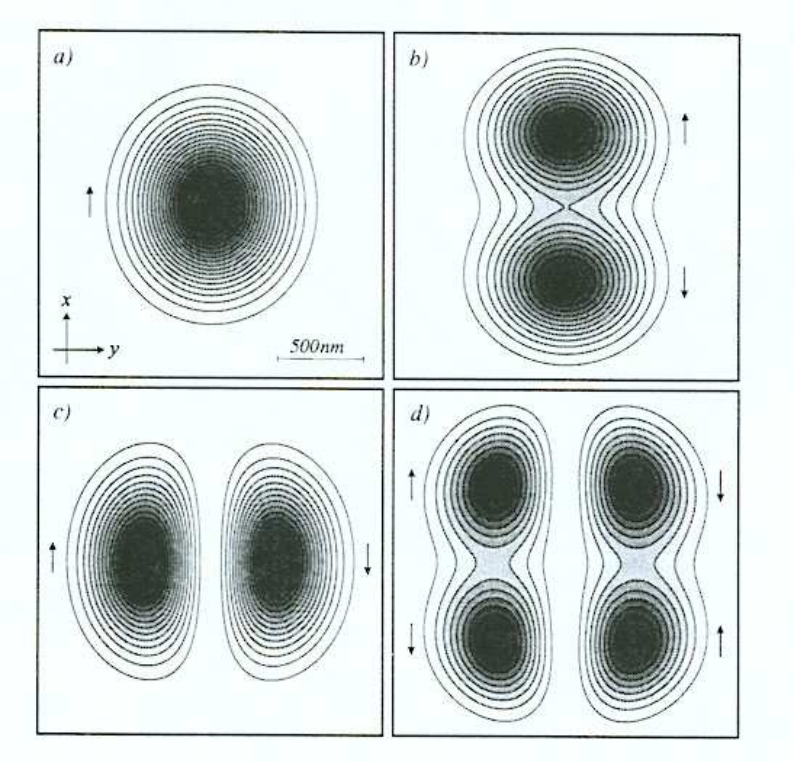

Figure 11: Intensity distribution of Hermite-Gaussian modes in the focal plane: a) 00-mode, b) 10-mode, c) 01-mode, d) 11-mode. From [Novotny and Hecht, "Principles of Nano-Optics"]

- In paraxial approximation Gaussian modes are TEM modes, i.e., the E- and B-field have the same form and the longitudinal component of E and B vanishes.
- If the paraxial approximation is not valid (e.g., under strong focusing with w_0 being very small), then longitudinal components of E and B have to be taken into account (see figure 12).
- The longitudinal E-field of the fundamental (00) Gaussian mode always vanishes on the optical axis. The longitudinal E-field of the Hermite-Gaussian (10)-mode has a maximum on the optical axis (see figure 12).

Remark: It has been suggested to use the longitudinal field of the 10-mode of strong laser pulses for particle acceleration.

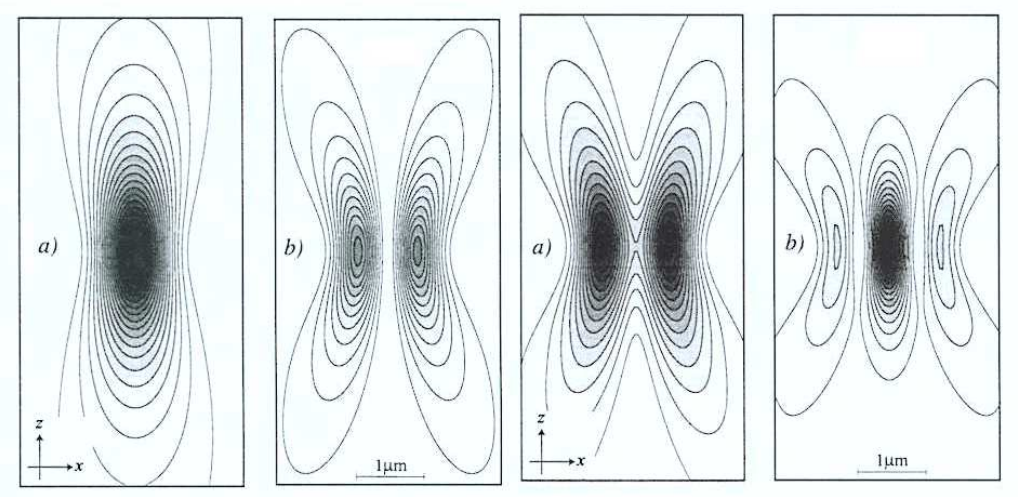

Figure 12: Absolute value of the electric field strength (a) and of the longitudinal component (b) for a 00-mode (left) and a 10-mode (right) under strong focusing (no paraxial approximation). From [Novotny and Hecht, "Principles of Nano-Optics"]

2.2.5 Focusing of Gaussian beams

Strongly focussed laser beams are of paramount importance in fluorescence microscopy, in confocal microscopy, in optical traps, and many other applications. Here, the paraxial approximation is not valid. With the help of the angular spectrum representation it is straightforward to calculate the field of strongly focussed Gaussian beams.

The following figure 13 shows the geometry of the problem.

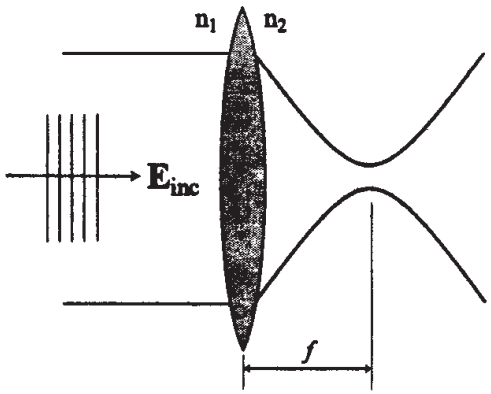

Figure 13: Geometry of focusing a beam with a lens. From [Novotny and Hecht, "Principles of Nano-Optics"]

As outline before the Fresnel coefficients can be utilized in the angular spectrum representation to solve the problem. The following figure 14 determines the notation:

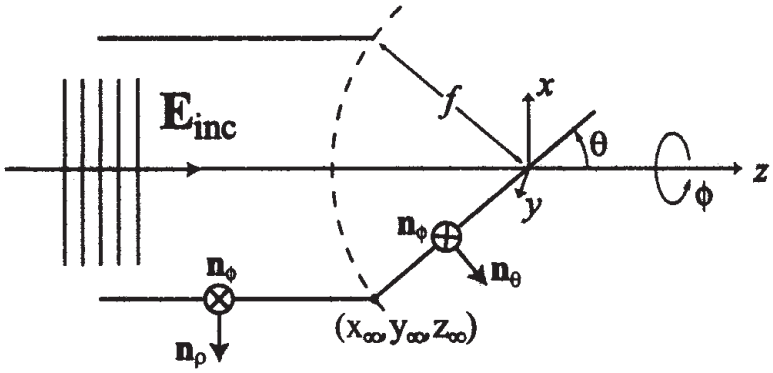

Figure 14: Notation in the focusing problem. From [Novotny and Hecht, "Principles of Nano-Optics"]

The field right after transmission through the lens can be derived from the incoming field as follows:

$$E_\infty = [t_s(E_{inc} n_\varphi) n_\varphi + t_p(E_{inc} n_\rho) n_\vartheta] \sqrt{\frac{n_1}{n_2}} (\cos\vartheta)^{1/2} \tag{103}$$

The focal point $(0,0,0)$ is far away from the lens (distance large compared to the optical wavelength). Inserting of the unitary vector n_φ and n_ϑ result in:

$$E_\infty = t_s \left[E_{inc}(\vartheta,\varphi) \begin{pmatrix} -\sin\varphi \\ \cos\varphi \\ 0 \end{pmatrix} \right] \begin{pmatrix} -\sin\varphi \\ \cos\varphi \\ 0 \end{pmatrix} \sqrt{\frac{n_1}{n_2}} (\cos\vartheta)^{1/2} \tag{104}$$

$$+t_p \left[E_{inc}(\vartheta,\varphi) \begin{pmatrix} \cos\varphi \\ \sin\varphi \\ 0 \end{pmatrix} \right] \begin{pmatrix} \cos\varphi\cos\theta \\ \sin\varphi\cos\theta \\ -\sin\theta \end{pmatrix} \sqrt{\frac{n_1}{n_2}} (\cos\vartheta)^{1/2} \tag{105}$$

As a special case we assume that (i) the field is polarized in x-direction,

$$E_{inc} = |E_{inc}|\, n_x \tag{106}$$

(ii) the lens is anti-reflection coated ($t_s = t_p = 1$), and (iii) the "beam waist" coincides with the lens. Then one finds:

$$E_\infty = |E_{inc}|\,(\vartheta,\varphi)\frac{1}{2} \begin{pmatrix} (1+\cos\vartheta) - (1-\cos\vartheta)\cos 2\varphi \\ -(1-\cos\vartheta)\sin 2\varphi \\ -2\cos\varphi\sin\theta \end{pmatrix} \sqrt{\frac{n_1}{n_2}} (\cos\vartheta)^{1/2} \tag{107}$$

Finally, we can derive the angular spectrum representation for the field at (ρ,φ,z) (polar coordinates) from a field $E_\infty(\vartheta,\varphi)$ at a point far away ($\lim_{kr\to\infty}$). In polar coordinats one finds after some math:

$$E(\rho,\varphi,z) = \frac{ikfe^{-ikf}}{2\pi} \int_0^{\theta_{\max}} \int_0^{2\pi} E_\infty(\vartheta,\phi) e^{ikz\cos\vartheta} e^{ik\rho\sin\vartheta\cos(\phi-\varphi)} \sin\vartheta \; d\vartheta d\phi \tag{108}$$

This equation is a central result. It allows calculation of focussing of an arbitrary incoming field E_{inc} by a lens with focal length f and numerical aperture $NA = n\sin\theta_{max}$. The field in the focal region is entirely determined by the far-field E_∞
For example, we can plug-in different Gaussian modes for the incident field $|E_{inc}|$ and calculate the field at the focus.

Focusing of Hermite-Gaussian modes:

It is possible to write the Hermite-Gaussian modes in the coordinates (f, ϑ, φ) as follows:
00-mode:

$$E_{inc} = E_0 e^{-f^2 \sin^2 \vartheta / w_0^2} \tag{109}$$

10-mode:

$$E_{inc} = (2E_0 f / w_0) \sin \vartheta \cos \varphi e^{-f^2 \sin^2 \vartheta / w_0^2} \tag{110}$$

01-mode:

$$E_{inc} = (2E_0 f / w_0) \sin \vartheta \sin \varphi e^{-f^2 \sin^2 \vartheta / w_0^2} \tag{111}$$

An important parameter, the **filling factor** f_0 is the ratio of the diameter of the incident beam compared to the diameter of the lens.

$$f_0 = \frac{w_0}{f \sin \vartheta_{\max}} \tag{112}$$

The analytic results for the electric field in the focus are:
00-mode:

$$E(\rho, \vartheta, \varphi) = \frac{ikf}{2} \sqrt{\frac{n_1}{n_2}} E_0 e^{-ikf} \begin{bmatrix} I_{00} + I_{02} \cos 2\varphi \\ I_{02} \sin 2\varphi \\ -2iI_{01} \cos 2\varphi \end{bmatrix} \tag{113}$$

10-mode:

$$E(\rho, \vartheta, \varphi) = \frac{ikf^2}{2w_0} \sqrt{\frac{n_1}{n_2}} E_0 e^{-ikf} \begin{bmatrix} iI_{11} \cos \varphi + iI_{14} \cos 3\varphi \\ -iI_{12} \sin \varphi + iI_{14} \sin 3\varphi \\ -2I_{10} + 2I_{13} \cos 2\varphi \end{bmatrix} \tag{114}$$

01-mode:

$$E(\rho, \vartheta, \varphi) = \frac{ikf^2}{2w_0} \sqrt{\frac{n_1}{n_2}} E_0 e^{-ikf} \begin{bmatrix} i(I_{11} + 2I_{12}) \sin \varphi + iI_{14} \sin 3\varphi \\ -iI_{12} \cos \varphi - iI_{14} \cos 3\varphi \\ 2I_{13} \sin 2\varphi \end{bmatrix} \tag{115}$$

The terms I_{ij} stand for specific integrals over Bessel functions which are provided for completeness at the end of this subsection. (Only the 10-Mode has a longitudinal field at the focal position!)

The following figure 15 demonstrates the influence of the filling factor f_0 on the size of the focus. Note, that the focus is not symmetric!

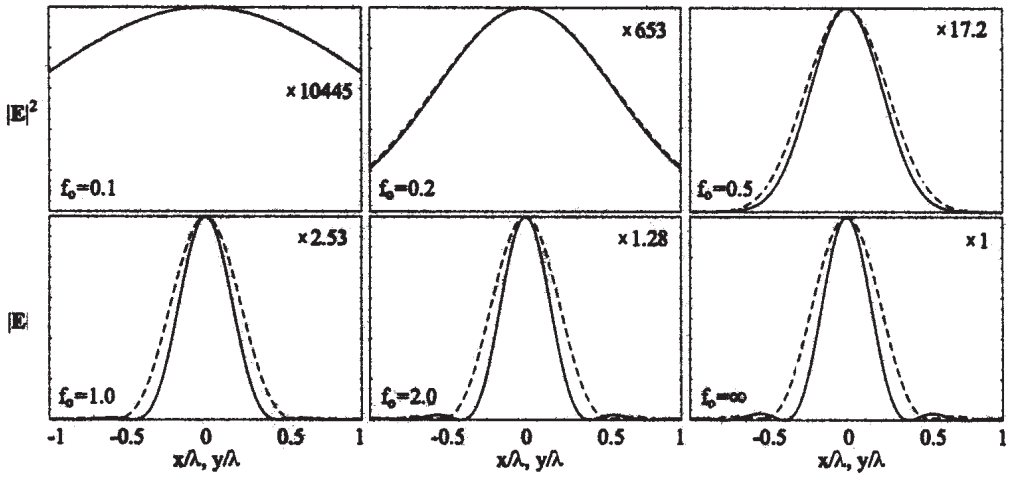

Figure 15: Influence of the filling factor on the size of the focus. The absolute value of the electric field is plotted along the x-axis (dashed, also direction of polarization) and the y-axis (solid) for different f_0. From [Novotny and Hecht, "Principles of Nano-Optics"]

Focusing of Hermite-Gaussian Doughnut modes:

The so-called Doughnut (DM) modes are of particular practical interest. Their intensity distribution reminds at the shape of a doughnut. DM modes can be produced by a superposition from Hermite-Gaussian (HG) modes:

$$\text{linear polarized DM} \quad : \quad LP = HG_{10}n_x + iHG_{01}n_x \tag{116}$$

$$\text{radial polarized DM} \quad : \quad RP = HG_{10}n_x + iHG_{01}n_y \tag{117}$$

$$\text{azimuthal polarized DM} \quad : \quad AP = -HG_{10}n_x + iHG_{01}n_y \tag{118}$$

The field in the focus of a DM is:
RP-mode:

$$E(\rho,\vartheta,\varphi) = \frac{ikf^2}{2w_0}\sqrt{\frac{n_1}{n_2}}E_0 e^{-ikf}\begin{bmatrix} i(I_{11}-I_{12})\cos\varphi \\ i(I_{11}-I_{12})\sin\varphi \\ -4I_{10} \end{bmatrix} \tag{119}$$

AP-mode:

$$E(\rho,\vartheta,\varphi) = \frac{ikf^2}{2w_0}\sqrt{\frac{n_1}{n_2}}E_0 e^{-ikf}\begin{bmatrix} i(I_{11}+3I_{12})\sin\varphi \\ -i(I_{11}+3I_{12})\cos\varphi \\ 0 \end{bmatrix} \tag{120}$$

In the following figure 16 the ratio of the field strength of the longitudinal and transversal component (RP-DM mode) is plotted as a function of the numerical aperture of the lens.

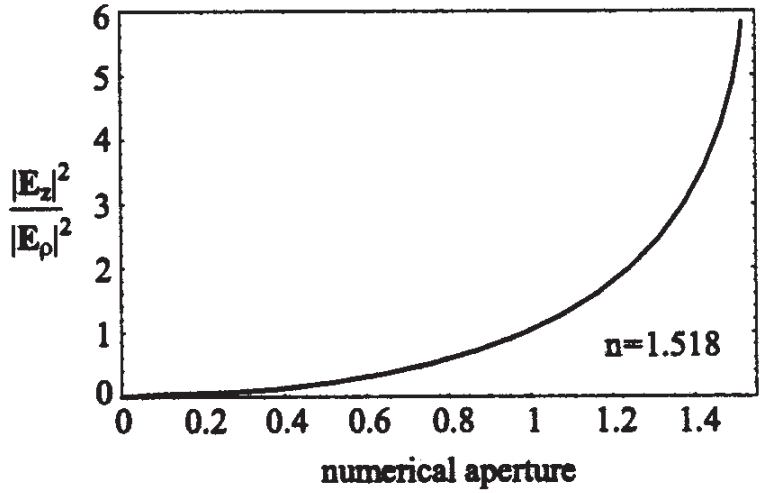

Figure 16: Ratio of the field strength of the longitudinal and transversal component (RP-DM mode) versus the numerical aperture of the lens. From [Novotny and Hecht, "Principles of Nano-Optics"]

One can use phase plates to create DM-mode, for example a (10)-mode with a 180° phase plate as depicted in figure 17. Contributions of higher modes can be eliminated with spatial filters. More advanced are programmable (e.g. liquid crystal) phase plates.

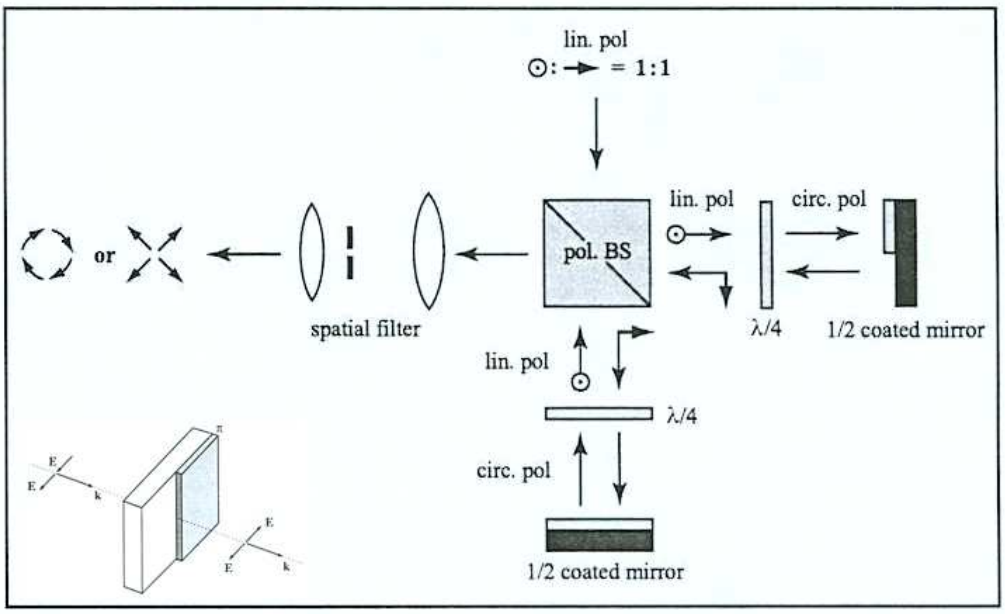

Figure 17: Simple setup to produce a Doughnut-mode. RP or AP polarization are controlled by the position of the mirrors. The inset shows a detail of the half coated mirror. From [Novotny and Hecht, "Principles of Nano-Optics"]

For completeness we list in the following the integrals used in the equations above:

$$I_{00} = \int_0^{\vartheta_{\max}} f_w(\vartheta)(\cos\vartheta)^{1/2}\sin\vartheta(1+\cos\vartheta)J_0(k\rho\sin\vartheta)e^{ikz\cos\vartheta}\,d\vartheta \quad (121)$$

$$I_{01} = \int_0^{\vartheta_{\max}} f_w(\vartheta)(\cos\vartheta)^{1/2}\sin^2\vartheta J_1(k\rho\sin\vartheta)e^{ikz\cos\vartheta}\,d\vartheta \quad (122)$$

$$I_{02} = \int_0^{\vartheta_{\max}} f_w(\vartheta)(\cos\vartheta)^{1/2}\sin\vartheta(1-\cos\vartheta)J_2(k\rho\sin\vartheta)e^{ikz\cos\vartheta}\,d\vartheta \quad (123)$$

$$I_{10} = \int_0^{\vartheta_{\max}} f_w(\vartheta)(\cos\vartheta)^{1/2}\sin^3\vartheta J_0(k\rho\sin\vartheta)e^{ikz\cos\vartheta}\,d\vartheta \quad (124)$$

$$I_{11} = \int_0^{\vartheta_{\max}} f_w(\vartheta)(\cos\vartheta)^{1/2}\sin^2\vartheta(1+3\cos\vartheta)J_1(k\rho\sin\vartheta)e^{ikz\cos\vartheta}\,d\vartheta \quad (125)$$

$$I_{12} = \int_0^{\vartheta_{\max}} f_w(\vartheta)(\cos\vartheta)^{1/2}\sin^2\vartheta(1-\cos\vartheta)J_1(k\rho\sin\vartheta)e^{ikz\cos\vartheta}\,d\vartheta \quad (126)$$

$$I_{13} = \int_0^{\vartheta_{\max}} f_w(\vartheta)(\cos\vartheta)^{1/2}\sin^3\vartheta J_2(k\rho\sin\vartheta)e^{ikz\cos\vartheta}\,d\vartheta \quad (127)$$

$$I_{14} = \int_0^{\vartheta_{\max}} f_w(\vartheta)(\cos\vartheta)^{1/2}\sin^2\vartheta(1-\cos\vartheta)J_3(k\rho\sin\vartheta)e^{ikz\cos\vartheta}\,d\vartheta \quad (128)$$

with

$$f_w(\vartheta) = \exp\left(-\frac{1}{f_0^2}\frac{\sin^2\vartheta}{\sin^2\vartheta_{\max}}\right) \quad (129)$$

2.3 Image formation and resolution

In this subsection we will describe the problem of image formation which lies at the heart of microscopy in a more abstract way. A limit is often set by the optical resolution which we address at the end of this subsection.

2.3.1 Imaging of coherent objects

The general problem of imaging is to find a relation between the electric fields at object points to the electric fields at image points. Figure 18 introduces our notation:

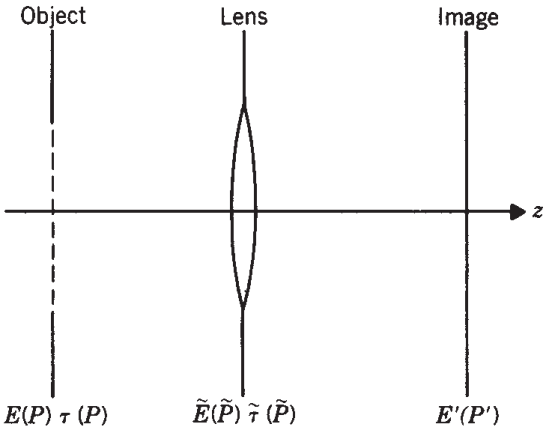

Figure 18: Notation and general problem of imaging.

In order to solve the problem of imaging one can start from the basic equations of diffraction theory. We will not review the theory of diffraction here in all detail, but refer to books on optics (e.g. Klein and Furtak, Optics, Wiley).

We start here with the Fresnel-Kirchhoff formula of diffraction:

$$E'(P') = const. \oint_S \tau(\widetilde{P}) \frac{e^{-ik(R+R')}}{RR'} d\widetilde{x} d\widetilde{y} \tag{130}$$

There is a straightforward interpretation of equation 130, as illustrated in figure 19: The field at the image point P' is the field of a spherical (elementary) wave with origin at $\widetilde{P}$. This wave however is "induced" by the electric field of the spherical (elementary) wave with origin at P in the object plane. The integral (over S) sums up

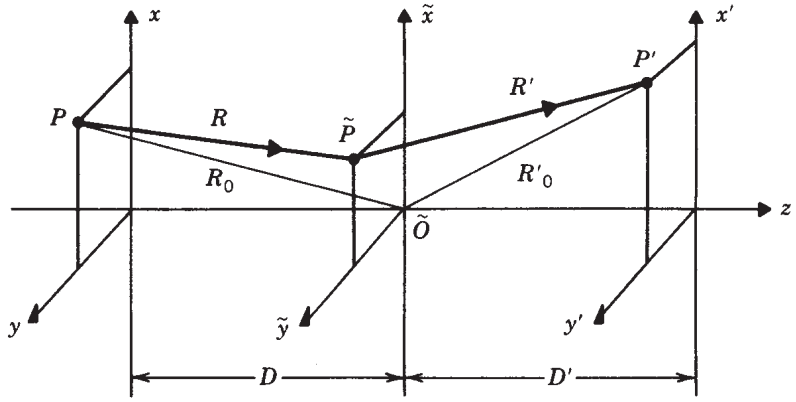

Figure 19: Notation and interpretation of the Fresnel-Kirchhoff diffraction formula. From [Klein and Furtak "Optics"].

over all waves from the intermediate plane (Here we assume that we collect mainly waves emitted in forward direction from the intermediate plane.).

The *transmission function* $\tau(\widetilde{P})$ accounts for a modification of free space transmission, i.e. a phase plate, lens, aperture, etc. in the intermediate plane.

In the formula above one can identify the *propagation kernel* $h(\widetilde{P} \to P')$:

$$h(\widetilde{P} \to P') = \frac{e^{-ik(R+R')}}{RR'} \tag{131}$$

In *Fresnel diffraction* it is possible to find a second order approximation for $h(\widetilde{P} \to P')$ (details can be found in optics books):

$$\begin{aligned} h(P \ \to \ P') = \left(\frac{i}{\lambda}\right)^2 \frac{e^{-ik(D+D')}}{DD'} \oint\limits_S \tau(\widetilde{P}) \cdot \\ \exp\left\{-ik\left[\frac{(\widetilde{x}-x)^2+(\widetilde{y}-y)^2}{2D} + \frac{(\widetilde{x}-x')^2+(\widetilde{y}-y')^2}{2D'}\right]\right\} d\widetilde{x}d\widetilde{y} \end{aligned} \tag{132}$$

Now, let us consider a specific example, imaging via a lens. Then, the transmission function of a (thin) lens has to be used:

$$\tau_L(\widetilde{r}) = |\tau_L(\widetilde{r})| \exp\left[\frac{-i2\pi}{\lambda}(n-1)d(\widetilde{r})\right] \tag{133}$$

Here, n and $d(\widetilde{r})$ are the index of refraction and thickness of the lens, respectively.

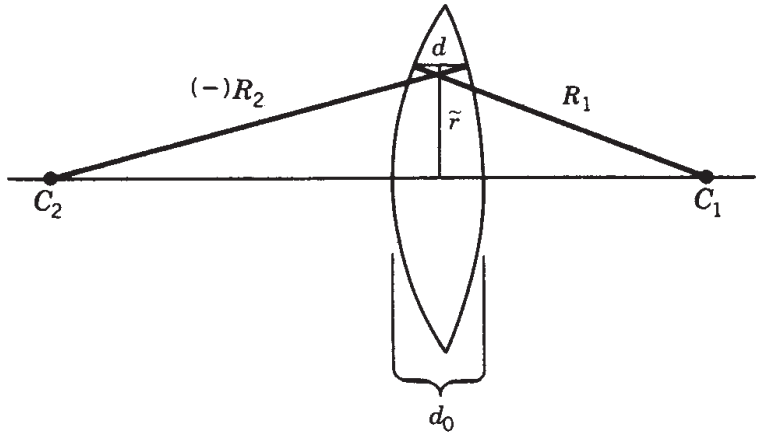

Figure 20: Special case of imaging with a biconvex lens. From [Klein and Furtak "Optics"].

In the case of a biconvex lens (see figure 20) one can approximate:

$$d(\widetilde{r}) \approx d_0 - \frac{\widetilde{r}^2}{2}\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \tag{134}$$

And:

$$\tau_L(\widetilde{r}) = |\tau_L(\widetilde{r})|\, e^{i\phi(d_0)} \exp\left[\frac{i\pi}{\lambda f}\widetilde{r}^2\right] \tag{135}$$

because

$$1/f = (n-1)\,[1/R_1 - 1/R_2] \tag{136}$$

In the following we consider imaging between *conjugated planes* (see figure 21):

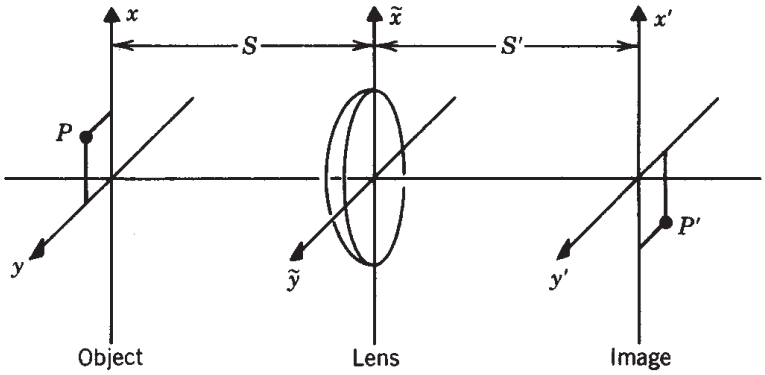

Figure 21: Imaging between conjugated planes. From [Klein and Furtak "Optics"].

Again we denote the propagation kernel from a point P to a point P' as follows:

$$h(P \rightarrow P') = \left(\frac{i^2}{\lambda^2}\right) \frac{e^{-ik(S+S')}}{SS'} \oint_S \tau_L(\widetilde{x},\widetilde{y}) \cdot \tag{137}$$
$$\exp\left\{-ik\left[\frac{(\widetilde{x}-x)^2+(\widetilde{y}-y)^2}{2S} + \frac{(\widetilde{x}-x')^2+(\widetilde{y}-y')^2}{2S'}\right]\right\} d\widetilde{x}d\widetilde{y}$$

Defining the differences of the *spatial frequencies* $\Delta u = u' - u$ und $\Delta v = v' - v$ as

$$\Delta u = u' - u = -\left(\frac{x}{S\lambda} + \frac{x'}{S'\lambda}\right) \tag{138}$$

$$\Delta v = v' - v = -\left(\frac{y}{S\lambda} + \frac{y'}{S'\lambda}\right) \tag{139}$$

one finds after short calculation of the expression in the exponent of all exponentials

$$\{\quad\} = -ik(R_0 + R_0') - \frac{i\pi}{\lambda}\widetilde{r}^2\left(\frac{1}{S} + \frac{1}{S'} - \frac{1}{f}\right) - i2\pi(\Delta u\widetilde{x} + \Delta v\widetilde{y}) \tag{140}$$

where R_0 and R_0' are defined as in figure 19.

If there is an imaging between conjugated planes (as assumed), then the quadratic term vanishes!

Finall, one derives:

$$h(P \to P') = \left(\frac{i}{\lambda}\right)^2 \frac{1}{SS'} e^{-ik(R_0+R_0')} \oint\limits_S |\tau_L(\widetilde{x},\widetilde{y})| \exp\left\{-i2\pi(\Delta u\widetilde{x} + \Delta v\widetilde{y})\right\} d\widetilde{x}d\widetilde{y}$$

The propagation kernel between conjugated planes is proportional to the Fourier transform of the transmission function of the lens:

$$h(P \to P') = \left(\frac{i}{\lambda}\right)^2 \left(\frac{1}{SS'}\right) e^{-ik(R_0+R_0')} T_L(\Delta u, \Delta v) \tag{141}$$

with

$$T_L = F\left[|\tau_L|\right] \tag{142}$$

It is now easy to find the electric field in the image plane by integration:

$$E'(P') = \left(\frac{i^2}{SS'\lambda^2}\right) e^{-ikR_0'} \oint\limits_S E(x,y)\tau_L(x,y) e^{-ikR_0} T_L(\Delta u, \Delta v) dxdy \tag{143}$$

The expression can be simplified by collecting terms that depend only on the source variables and introducing new variables for the integration:

$$\begin{aligned} E'(x',y') &= i^2 \left(\frac{S}{S'}\right) e^{-ikR_0'} \oint\limits_S \mathcal{E}(S\lambda u, S\lambda y) T_L(u'-u, v'-v)\, dudv & (144)\\ &= e^{-ikR_0'}/m\left\{\mathcal{E}(S\lambda u, S\lambda y) \circledast T_L(u,v)\right\} & (145) \end{aligned}$$

where $m = -S'/S$ the transversal magnification and $\circledast$ the mathematical operation of convolution.

The field distribution in the image plane is a convolution of the source function $\mathcal{E}$ with the Fourier transform of the transmission function of the lens.

The special case of the **field distribution in the image plane of a point source** is derived by inserting a delta-function as source function:

$$E_{point}(x',y') = \frac{A\lambda}{i} e^{-i\phi_o} \left(\frac{m}{S'^2\lambda^2}\right) T_L\left[\left(\frac{-1}{\lambda S'}(x' - mx_0\right), \left(\frac{-1}{\lambda S'}(y' - my_0\right)\right] \tag{146}$$

where A is the strength of the source field.
Instead of a single image point at $x_0' = mx_0, y_0' = my_0$ (as in ray optics description

of imaging) we find a broader field distribution (due to diffraction).

In case of a spherical lens with

$$|\tau_L(\widetilde{r})| = 1 \quad \text{if } \widetilde{r} < \widetilde{r}_0 \tag{147}$$
$$|\tau_L(\widetilde{r})| = 0 \quad \text{elsewhere} \tag{148}$$

one finds via Fourier transformation:

$$E(x', y') = E(x'_0, y'_0) \left(\frac{2J_1\left(\frac{2\pi\widetilde{r}_0 \Delta r'}{\lambda S'}\right)}{\left(\frac{2\pi\widetilde{r}_0 \Delta r'}{\lambda S'}\right)} \right) \tag{149}$$

$$S(x', y') = S(x'_0, y'_0) \left(\frac{2J_1\left(\frac{2\pi\widetilde{r}_0 \Delta r'}{\lambda S'}\right)}{\left(\frac{2\pi\widetilde{r}_0 \Delta r'}{\lambda S'}\right)} \right)^2 \tag{150}$$

with

$$\Delta r' = \left[(x' - x'_0)^2 + (y' - y'_0)^2\right]^{1/2} \tag{151}$$

This is the well known Airy disc as plotted in figure 22

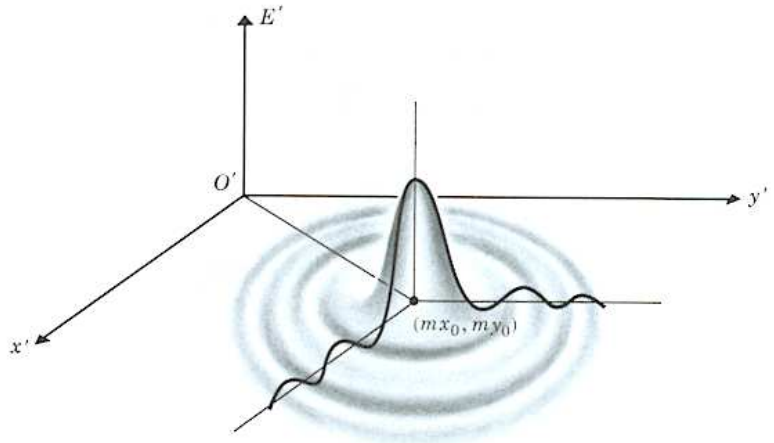

Figure 22: Intensity distribution of a point source imaged by a spherical lens of finite aperture. From [Klein and Furtak "Optics"].

Brief summary of some aspects of imaging:

There is no simple one-to-one relation when imaging an object point P with a finite (otherwise perfect) lens. Instead the following relation holds:

$$P \xrightarrow{h} P' \quad \text{with} \quad h \propto F\left[|\tau_L(\widetilde{r})|\right] \tag{152}$$

The electric field in the image plane is a convolution

$$E(x', y') = \{\mathcal{E}(S\lambda u, S\lambda y) \circledast T_L(u, v)\} \tag{153}$$

or after an additional Fourier transformation

$$\widehat{E}(k_{x'}, k_{y'}) = \widehat{\mathcal{E}}(k_{x'}, k_{y'}) \cdot |\tau_L| \tag{154}$$

The transmission function of a finite lens cuts off spatial frequencies with $k_x > \frac{\widetilde{x}}{S\lambda}$. It thus acts as a low-pass filter.

2.3.2 Imaging of incoherent objects

The same mathematical formalism as derived above can be used when describing imaging of incoherent objects:

- **coherent objects** (e.g. objects illuminated by coherent sources) $\Longrightarrow$ summing up **amplitudes** of diffracted elementary waves
- **incoherent objects** (self illuminating or fluorescent objects) $\Longrightarrow$ summing up **intensities** of diffracted elementary waves

An important quantity in the theory of imaging of incoherent objects is the **"Point-Spread-Function" (PSF)**. It describes the intensity distribution resulting from imaging a point-like object through a specific imaging system.

We showed in the theory of imaging of coherent objects: The amplitude distribution of the image of a point source is proportional to the Fourier transform of the transmission function of the lens. Obviously, for the intensity $S(P')$ it holds:

$$S(P') = \frac{\Phi}{\sigma_L} \frac{1}{S'^2\lambda^2} \left| T_L\left[\left(\frac{-1}{\lambda S'}(x' - x'_0)\right), \left(\frac{-1}{\lambda S'}(y' - y'_0)\right)\right]\right|^2 \tag{155}$$

$$= \Phi\mathcal{O}(x' - x'_0, y' - y'_0; P') \tag{156}$$

Here Φ is the total flux and σ_L the area of the lens.

$\mathcal{O}$ is the Point-Spread-Function. In a more realistic case T_L is the Fourier Transform of the transmission function of the lens *including* deviations from the idealized spherical phase factor ($\exp(\frac{i\pi}{\lambda f}\widetilde{r}^2)$, i.e. abberations should be included. This is already indicated as an explicit dependance of $\mathcal{O}$ on P'.

Typically, the PSF $\mathcal{O}$ is normalized to 1 ($\mathcal{O}_0$).

In case of a lens without an abberations it is :

$$\mathcal{O}_0(x'-x'_0, y'-y'_0) = \frac{\pi\widetilde{r}_0^2}{\lambda^2 S'^2}\left|\frac{2J_1(w)}{w}\right|^2 \tag{157}$$

with

$$w = \frac{2\pi\widetilde{r}_0}{\lambda S}\left[(x'-x'_0)^2+(y'-y'_0)^2\right]^{1/2} \tag{158}$$

The following figure 23 shows the image of a point source positioned off the optical axis. In this case well known lens aberrations (coma) are visible.

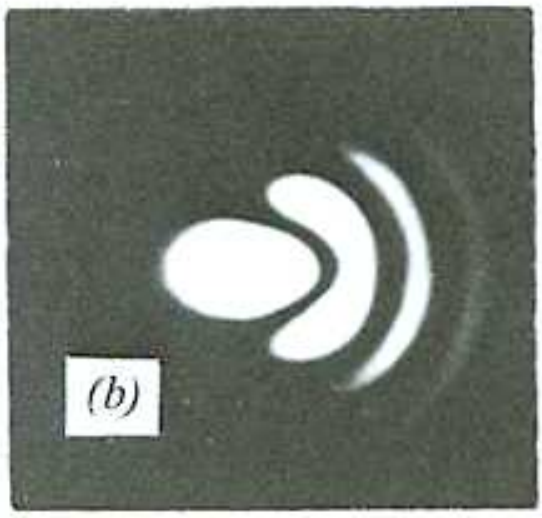

Figure 23: Image of a point source positioned off the optical axis. From [Klein and Furtak "Optics"].

Extended sources:
In case of an extended source one can find the image, i.e., the intensity distribution in the image plane via integration:

$$S'(x',y') = \int\!\!\int_{-\infty}^{\infty} \mathcal{O}(x'-x'_0, y'-y'_0; P') S'_P(x'_0, y'_0)\; dx'_0 dy'_0 \tag{159}$$

where $S'_P(x'_0, y'_0)$ corresponds to the image of an ideal ray-optical theory (one-to-one corresondance between object and image point).

The convolution integral has a simple form after additional Fourier transformation:

$$\widehat{S}' = \widehat{\mathcal{O}} \cdot \widehat{S}'_P \tag{160}$$

$$\widehat{\mathcal{O}} = F[\mathcal{O}] = \left|\widehat{\mathcal{O}}\right| e^{i\Psi} \tag{161}$$

The functions above are denoted as:

$$\begin{aligned} \widehat{\mathcal{O}} &= \textbf{Optical Transfer Function (OTF)} \\ \left|\widehat{\mathcal{O}}\right| &= \textbf{Modulation Transfer Function (MTF)} \\ \Psi &= \textbf{Phase Transfer Function (PTF)} \end{aligned}$$

Similar as in coherent imaging theory the physical meaning of the various functions become apparent when decomposing the image $S'_P(x'_0, y'_0)$ in Fourier components:

- MTF modulates the amplitude of each Fourier component
- PTF modulates the phase of each Fourier component

The inverse Fourier transform of the modulated components results in the image.

The experimental determination or a priori calculation of the Optical Transfer Function is of enormous importance in practical optics.

It is possible to estimate an upper bound for the spatial frequencies transmitted by an ideal lens:

$$\lambda_{\min}/2 = \frac{0.61\lambda S'}{\widetilde{r}_0} \tag{162}$$

$$f_{\max} \approx \frac{\widetilde{r}_0}{\lambda S'} \tag{163}$$

where $\widetilde{r}_0$ is the radius of the lens, S' is the distance to the image (S' and f_{max} is the focal position of infinitely far objects.

The following figure 24 show calculated MTFs:

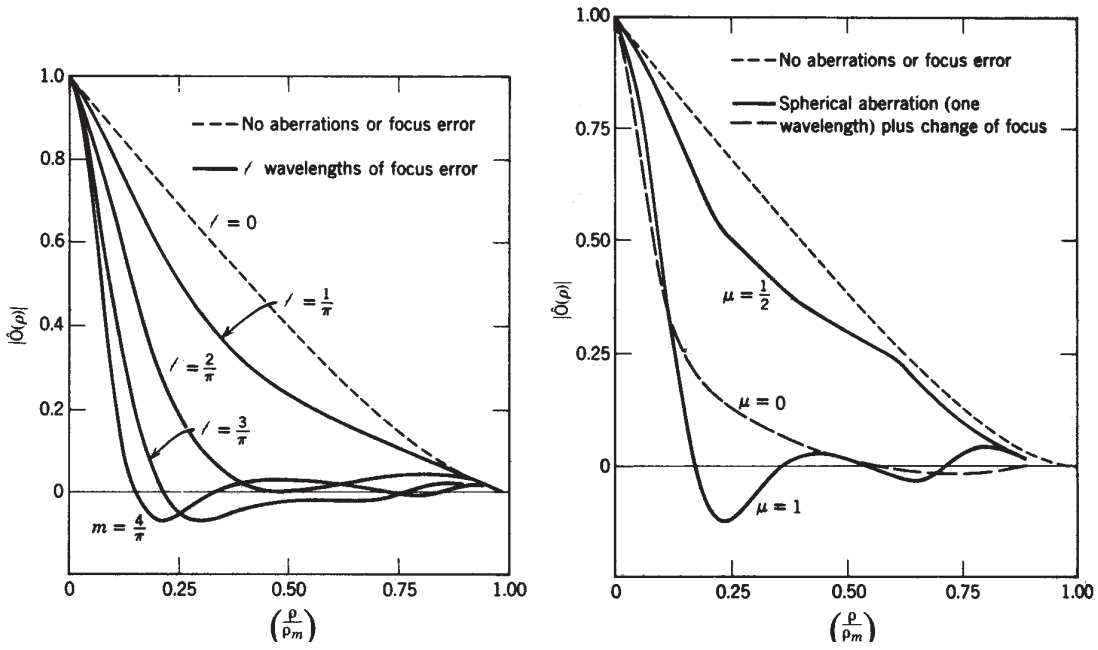

Figure 24: Example of a calculated Modulation Transfer Function. l is a parameter to quantify a focus error (in units of the wavelength), μ is the distance between paraxial and marginal focus. Negative values of the MTF indicate a phase jump of π corresponding to an inversion of the contrast. From [Klein and Furtak "Optics"].

Figure 25 shows the measured OTF of a camera lens:

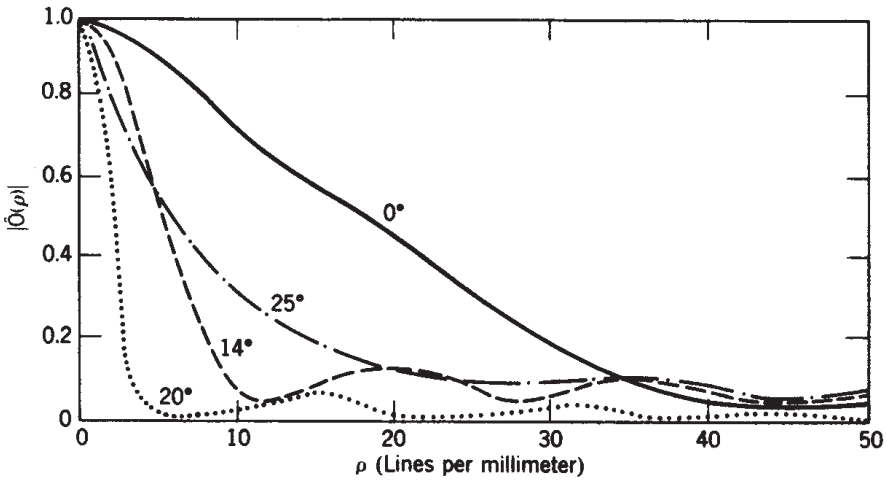

Figure 25: Measured OTFs of a camera lens. From [Klein and Furtak "Optics"].

2.3.3 Optical resolution

- An optical system cannot resolve two objects which are closer together than the spatial cut-off frequency of the system.

  A first criterion for optical resolution was formulated by Lord Rayleigh (John William Strutt, 1842-1919).

  The **Rayleigh-Kriterium:**
  Two point-like objects can be resolved if the maximum of one Airy disc (diffraction pattern of one object) coincides with the minimum of the other Airy disc (diffraction pattern of the other object). This is the case if:

$$r_{\min} = 0.61 \frac{\lambda}{NA} \tag{164}$$

  Here $NA = r_0/f$ is the replaced by the **numerical aperture**

$$NA = n \sin \alpha \approx n \frac{r_0}{f} \tag{165}$$

  Remarks:

- the Rayleigh-criterion is a definition.
- In theory the center of an Airy disc can be located with arbitrary precision, i.e. the separation of two objects is always possible.
- The signal-to-noise ratio (SNR) limits the resolution in practise (detector noise, finite signal) and also fundamentally (quantum noise).
- If the two objects have different properties (wavelength of fluorescence, polarization, etc.), then resolution can be improved.

The following figure 26 shows the measured intensity distribution of two neighboring objects:

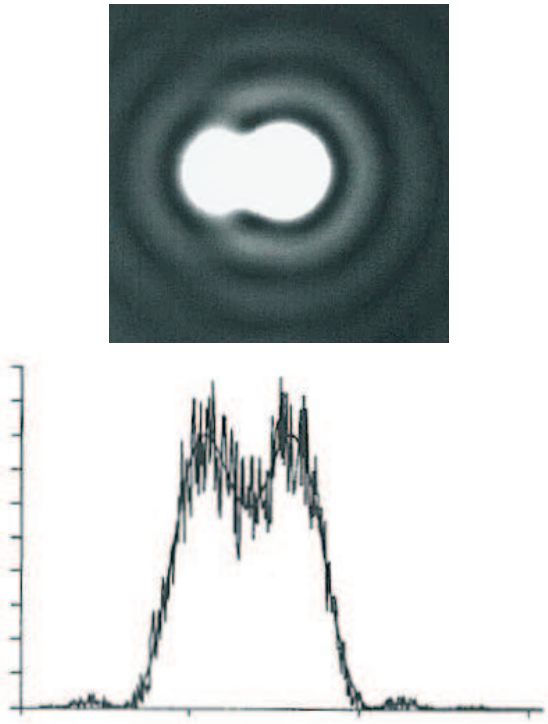

Figure 26: Measured intensity distribution of to neighboring objects. Top: camera image. Bottom: cross section. [R. H. Webb, Rep. Prog. Phys. 59, 427 (1996)]

It is possible to define other criteria for the optical resolution:

Sparrow-criterion:
Distance where the minumum between two Airy discs vanishes.

Abbe-criterium (resolution of a grating):

$$d > \frac{\lambda}{2 \sin \alpha} \tag{166}$$

where d is the grating period and α the angle to the first diffraction maximum. the following figure 27 illustrates the transition from Rayleigh- to Sparrow-criterion.

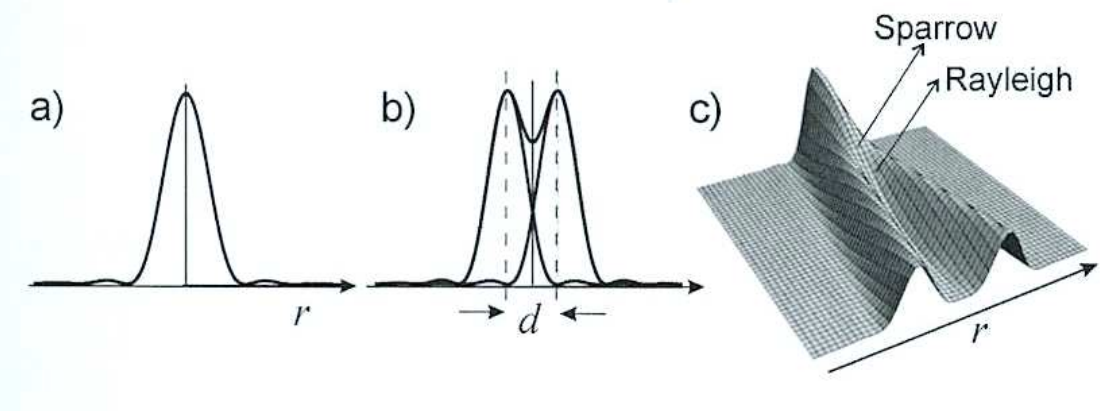

Figure 27: Transition from Rayleigh to Sparrow criterion.