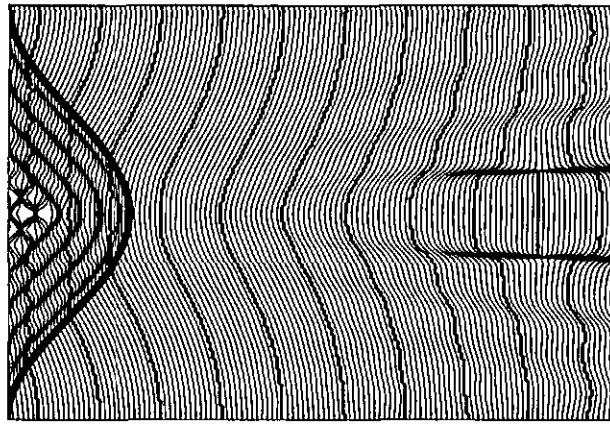
Peter Blattner

1477

Band 367

Light fields emerging from periodic optical microstructures



UFO Atelier für Gestaltung & Verlag

Université de Neuchâtel
Institut de Microtechnique

Light fields emerging from periodic optical microstructures

Thèse

Présentée à la Faculté des sciences
pour obtenir le grade de docteur ès sciences
par

Peter Blattner

Neuchâtel, janvier 1999

IMPRIMATUR POUR LA THÈSE

Light fields emerging from periodic optical microstructures

de M. Peter Blattner

UNIVERSITÉ DE NEUCHÂTEL

FACULTÉ DES SCIENCES

La Faculté des sciences de l'Université de Neuchâtel sur le rapport des membres du jury,

MM. R. Dändliker (directeur de thèse), H.-P. Herzig, D. Courjon (Besançon) et J.J. Stamnes (Bergen, N)

autorise l'impression de la présente thèse.

Neuchâtel, le 21 janvier 1999

Le doyen:

F. Stoeckli

Abstract

Until the beginning of the eighties, optical systems were composed of classical elements such as lenses, mirrors, and beamsplitters. Thanks to spectacular improvements in the microelectronics manufacturing, it became possible to realize structured surfaces with feature sizes comparable to the optical wavelength. This enables the realization of elements with novel properties. Therefore, the accurate modeling of the interaction of light with such optical microstructures became necessary. The present work gives an overview of the author's contributions to the fields of electromagnetic optics, micro-optics, diffractive optics, near field optics, and optical metrology.

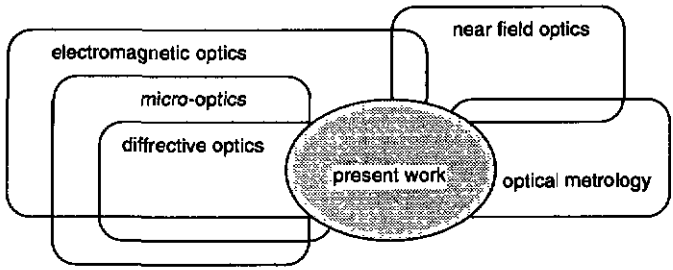
Contents

Chapter 1

Introduction

Until the beginning of the eighties, optical systems were composed of classical elements such as lenses, mirrors, and beamsplitters. Thanks to spectacular improvements in the microelectronic manufacturing, it became possible to realize structured surfaces with feature sizes comparable to the optical wavelength. This enables the realization of elements with novel properties. Therefore, the accurate modeling of the interaction of light with such optical microstructures became necessary.

The present work gives an overview of the author's contributions to the fields of electromagnetic optics, micro-optics, diffractive optics, near field optics, and optical metrology.



Part of the work has already been published. The following papers are in the appendix of this thesis:

[A.I] P. Blattner, H. P. Herzig, and R. Dändliker. "Scanning near-field optical microscopy: transfer function and resolution limit". Opt. Comm., **155**, 245–250, (1998).

[A.II] P. Blattner and H. P. Herzig. "Rigorous diffraction theory applied to microlenses". J. Mod. Opt., **45**(7), 1395–1403, (1998).

[A.III] P. Blattner, H. P. Herzig, K. J. Weible, J. M. Teijido, H. J. Heimbeck, E. Langenbach, and J. Rogers. "Diffractive optics for compact space communication terminals". J. Mod. Opt., **43**(7), 1473–1484, (1996).

[A.IV] P. Blattner, H. P. Herzig, and S. S. H. Naqvi. "Scanning spot metrology for testing of photolithographic masks". Opt. Eng., **23**, 2425–2427, (1995).

[A.V] P. Blattner, S. S. H. Naqvi, H. P. Herzig, and P. Ehbets. "Diffractive structures for testing nano-meter technology". Microelectr. Eng., **27**, 543–546, (1995).

The thesis is structured as follows:

Chapter 2 provides a short review of the mathematical treatment of optical fields in general. The specific results will be used in later chapters. Chapter 3 outlines the grating diffraction problem. I show that it is important to use rigorous diffraction theories, if the structures are in the order of the optical wavelength. I present the Fourier modal method as a suitable tool for solving the grating diffraction problem. Furthermore, I extend the existent Rayleigh method and the Waterman approach to dielectric surface relief structures. Introducing approximate theories helps to understand the complex problem of light-matter interaction. Chapter 4 introduces the concept of transfer functions for linear systems. With this concept, complicated optical systems can be modelled. As an example, a novel approach to understand the optical behavior of the scanning near field optical microscope (SNOM) is presented. I emphasize that for a good SNOM, the transfer functions representing the capability to convert evanescent waves into propagating waves has to be optimized, rather than the overall throughput. Chapter 5 applies the methods introduced in the previous chapters to different important cases. In particular, the near field and far field of microlenses and slit apertures are investigated. Furthermore, we discuss properties of phase singularities created by optical microstructures in detail. Phase singularities are isolated points of zero amplitude, hence the phase is not determined at these points. These points are well localized. Therefore, measuring their position can give information about the object (shape, index of refraction, etc.). Chapter 6 gives examples of the application of optical microstructures in the field of intersatellite laser communication systems and in optical lithography. Chapter 7 deals with the metrology of optical microstructures. We introduce two fast, non-destructive methods to characterize surface

relief gratings. Furthermore, we review the concept of resolution and explain apparent superresolution in interference microscopy. I present empirically that electromagnetic waves exhibit similar behavior than those observed in fluids: For smooth and thin structures, the phase field shows a laminar flow. If we increase the thickness, phase singularities appear and the system becomes turbulent. Chapter 8, finally, summarizes the conclusions of this thesis.

Chapter 2

Theoretical background

2.1 Three dimensional vector fields

2.1.1 Maxwell eqnations

The behavior of electromagnetic field is totally described by the *Maxwell equations*, viz.

$$\nabla \times \mathbf{E} + \dot{\mathbf{B}} = 0\,, \tag{2.1}$$

$$\nabla \times \mathbf{H} = \mathbf{j} + \dot{\mathbf{D}}\,, \tag{2.2}$$

$$\nabla \mathbf{D} = \rho\,, \tag{2.3}$$

$$\nabla \mathbf{B} = 0\,. \tag{2.4}$$

Equations (2.1) and (2.2) are vector equations, 2.3 and 2.4 scalar relations. $\mathbf{E}$ is the electric field, $\mathbf{H}$ the magnetic field, $\mathbf{D}$ the electric displacement, $\mathbf{B}$ the magnetic induction, $\mathbf{j}$ the electric current density, and ρ the electric charge density. Maxwell equations describe the interrelation between these different fields. The standard units of the physical quantities are resumed in Tab. 2.1

2.1.2 Material equations

For linear isotropic materials the general relation between the electric field and the electric displacement is

$$\mathbf{D} = \epsilon_0 \epsilon \mathbf{E}\,, \qquad \text{with} \qquad \epsilon_0 = 8.85 \cdot 10^{-12}\ AsV^{-1}m^{-1}\,, \tag{2.5}$$

$$[E] = V\ m^{-1} \quad , \quad [D] = As\ m^{-2} \quad , \quad [j] = A\ m^{-2}\,,$$
$$[H] = A\ m^{-1} \quad , \quad [B] = Vs\ m^{-2} \quad , \quad [\rho] = As\ m^{-3}\,.$$

Table 2.1: Units of the different physical quantities

where ϵ is the *electric permittivity* of the medium. A similar relation holds for the magnetic field and the magnetic induction

$$\mathbf{B} = \mu_0 \mu \mathbf{H}\,, \qquad \text{with} \qquad \mu_0 = 4\pi \cdot 10^{-7}\ VsA^{-1}m^{-1}\,, \tag{2.6}$$

μ is the *magnetic permeability* and equal to 1 for non-magnetic materials.

2.1.3 Boundary conditions at a surface of discontinuity

Maxwell's equations were only stated for regions of space throughout which the physical properties of the medium (i.e. ϵ and μ) are continuous. In micro-optics a lot of elements contain abrupt boundaries between two media. The properties of the field vectors across such discontinuities are described by boundary conditions. Let $\mathbf{n}_{12}$ be a unit vector normal to the interface separating two media. The following conditions are valid [1]:

$$\mathbf{n}_{12} \cdot (\mathbf{B}_2 - \mathbf{B}_1) = 0\,, \tag{2.7}$$

i.e. *the normal component of the magnetic induction is continuous across the surface of discontinuity.*

$$\mathbf{n}_{12} \cdot (\mathbf{D}_2 - \mathbf{D}_1) = \rho_S\,, \tag{2.8}$$

i.e. *in the presence of a layer of surface charge density ρ_S on the surface, the normal component of the electric displacement changes abruptly across the surface, by an amount equal to ρ_S.*

$$\mathbf{n}_{12} \times (\mathbf{E}_2 - \mathbf{E}_1) = 0\,, \tag{2.9}$$

i.e. *the tangential component of the electric vector is continuous across the surface.*

$$\mathbf{n}_{12} \times (\mathbf{H}_2 - \mathbf{H}_1) = \mathbf{j}_S\,, \tag{2.10}$$

i.e. *in the presence of a surface current density $\mathbf{j}_S$, the tangential component of the magnetic field vector changes abruptly by $\mathbf{j}_S \times \mathbf{n}_{12}$.*

We are mainly interested in dielectric materials ($\rho = 0$, $\mathbf{j} = 0$). Thus all right hand terms of Eqs. (2.7) - (2.10) are zero. The corresponding components of the electric and magnetic field vectors are continuous through the interface.

2.1.4 Poynting vector and energy density

If we form the scalar product of $\mathbf{H}$ with the first Maxwell equation Eq. 2.1and of $\mathbf{E}$ with the second Maxwell equation Eq. 2.2 and subtract both results we obtain

$$\mathbf{H} \cdot (\nabla \times \mathbf{E}) - \mathbf{E} \cdot (\nabla \times \mathbf{H}) = -\mathbf{E} \cdot \dot{\mathbf{D}} - \mathbf{H} \cdot \dot{\mathbf{B}}\,, \tag{2.11}$$

this gives

$$\nabla(\mathbf{E}\times\mathbf{H}) = -\tfrac{1}{2}\frac{\partial}{\partial t}\left(\mathbf{E}\cdot\mathbf{D}+\mathbf{H}\cdot\mathbf{B}\right). \tag{2.12}$$

The unit of the scalar products inside the parenthesis of the right side of Eq. (2.12) are *Joules per cubic meters* ($[ED] = [BD] = Jm^{-3}$). We therefore write

$$\nabla\,\mathbf{S} + \frac{\partial}{\partial t}\,w = 0\,, \tag{2.13}$$

where w is the energy inside a unit volume associated to an electromagnetic field. Written in terms of the electric energy density w_e and the magnetic energy density w_m

$$w = w_e + w_m\,, \qquad \text{with} \tag{2.14}$$

$$w_e = \tfrac{1}{2}\,\mathbf{E}\cdot\mathbf{D}\,, \qquad \text{and} \qquad w_m = \tfrac{1}{2}\,\mathbf{B}\cdot\mathbf{H}\,. \tag{2.15}$$

The power flow in electromagnetic fields is described by the *Poynting vector*

$$\mathbf{S} = \mathbf{E}\times\mathbf{H}\,. \tag{2.16}$$

The physical interpretation of $|\mathbf{S}|$ according Eq. (2.13) represents the amount of energy which crosses per second an unit area ($[S] = Js^{-1}m^{-2}$). By definition $\mathbf{S}$ is always perpendicular to $\mathbf{E}$ and $\mathbf{H}$.

2.1.5 Wave equations

Using some elementary vector analysis it is possible to decouple the field relations described by the Maxwell equations. For linear, isotropic, but (dielectric-) inhomogeneous media ($\epsilon = \epsilon(\mathbf{r})$) it results

$$\Delta\mathbf{E} - \epsilon_0\epsilon(\mathbf{r})\mu_0\mu\ddot{\mathbf{E}} = -\nabla\left(\mathbf{E}\frac{\nabla\epsilon(\mathbf{r})}{\epsilon(\mathbf{r})}\right), \tag{2.17}$$

and

$$\Delta\mathbf{H} - \epsilon_0\epsilon(\mathbf{r})\mu_0\mu\ddot{\mathbf{H}} = -\frac{\nabla\epsilon(\mathbf{r})}{\epsilon(\mathbf{r})}\times(\nabla\times\mathbf{H})\,. \tag{2.18}$$

Equations (2.17) and (2.18) are the *inhomogeneous* wave equations for $\mathbf{E}$ and $\mathbf{H}$. The different field components are *coupled* through the inhomogeneity of the material ($\nabla\epsilon(\mathbf{r}) \neq 0$).

2.1.6 Time harmonic fields

We are mostly interested in time harmonic, monochromatic fields. For this purpose, the complex vectors $\mathbf{E}_\omega$ and $\mathbf{H}_\omega$ are usually introduced. The *physical* fields are then the real

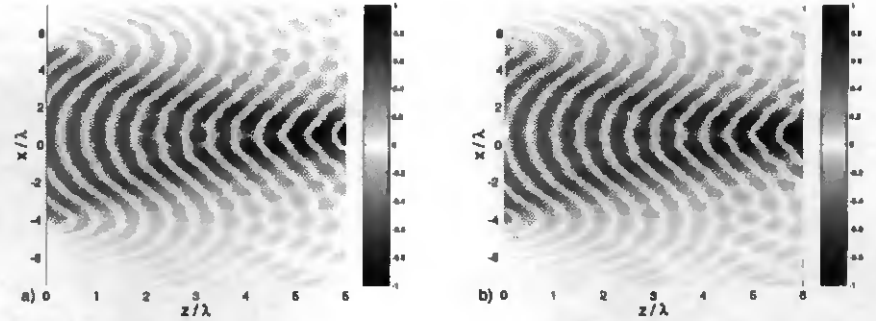


Figure 2.1: Electric field distribution of an converging wave at a) $t = 0$ and b) $t = \frac{\pi}{2}\omega^{-1}$.

parts of complex vectors, namely

$$\mathbf{E}(\mathbf{r},t) = \mathrm{Re}\left\{\mathbf{E}_\omega(\mathbf{r})e^{-i\omega t}\right\}, \tag{2.19}$$

$$\mathbf{H}(\mathbf{r},t) = \mathrm{Re}\left\{\mathbf{H}_\omega(\mathbf{r})e^{-i\omega t}\right\}. \tag{2.20}$$

In the complex vector notation the time propagation is therefore included in the *phase* of the vector, i.e. it is sufficient to add ωt to the phase of the calculated $\mathbf{E}_\omega$ to find the field $\mathbf{E}(\mathbf{r},t)$.

For illustration, let us consider one component of a converging electric wave created by an optical microstructures (for more details see Sec. 5). Figure 2.1 a) shows the calculated electric field at $t = 0$. The field at $t = \frac{\pi}{2}\omega^{-1}$ (Fig. 2.1 b) is then found by simply applying Eq. (2.19) the calculated complex field of Fig. 2.1 a).

With the introduction of the complex field vectors, the first two Maxwell equations Eqs. 2.1 and 2.2 simplify to

$$\nabla \times \mathbf{E}_\omega = -i\omega\mu_0\mu\mathbf{H}_\omega, \tag{2.21}$$

$$\nabla \times \mathbf{H}_\omega = +i\omega\epsilon_0\epsilon\mathbf{E}_\omega. \tag{2.22}$$

The third and the fourth Maxwell equation Eqs. 2.3 and 2.4 are now also fulfilled by the Eqs. (2.21) and (2.22), because of $\nabla(\nabla\times\,) \equiv 0$. Furthermore, this two equations imply that the magnetic field can be deduced from the imaginary part of the complex electric field, and visa versa.

$$\mathbf{H}(\mathbf{r},t) = \frac{-1}{\mu_0\mu\omega}\nabla \times \mathrm{Im}\left\{\mathbf{E}_\omega(\mathbf{r})e^{i\omega t}\right\}, \tag{2.23}$$

$$\mathbf{E}(\mathbf{r},t) = \frac{1}{\epsilon_0\epsilon\omega}\nabla \times \mathrm{Im}\left\{\mathbf{H}_\omega(\mathbf{r})e^{i\omega t}\right\}. \tag{2.24}$$

The *time averaged* energy densities are related to the square of the magnitude of the complex fields,

$$< w_e > = \frac{1}{4}\epsilon_0\epsilon|\mathbf{E}_\omega|^2 , \qquad \text{and} \qquad < w_m > = \frac{1}{4}\mu_0\mu|\mathbf{H}_\omega|^2 . \tag{2.25}$$

And finally, it can be shown that the *time averaged* Poynting vector is expressed as

$$< \mathbf{S} > = \tfrac{1}{2}\,\mathrm{Re}\,\{\mathbf{E}_\omega \times \mathbf{H}_\omega^*\} . \tag{2.26}$$

Using Eq. (2.13), it follows that the divergence of the time averaged Poynting vector is zero, viz.

$$\nabla\ < \mathbf{S} > \equiv 0 . \tag{2.27}$$

The beauty of describing a time harmonic field by a complex vector is that the time averaged energy is the absolute value of the vector, that the time (and space) propagation is included in the phase of the vector, that the physical field is the real part of the vector, and that the complementary field can be deduced from the imaginary part of the vector.

2.2 Two dimensional geometry

In the present work, we will only treat two dimensional geometries. The grating vector is chosen to be parallel to the surface. Furthermore, the media and the fields shall not vary in the $\hat{y}$-direction. For these geometries we can distinguish two different fundamental cases where the inhomogeneous wave equations Eqs. (2.17) and (2.18) become particularly simplified.

If the electric field is parallel to the $\hat{y}$-axis (which is called *TE-polarization*), the vector wave equation (Eq. (2.17)) for the electric field becomes a scalar equation. Furthermore, the scalar product on the right hand side of Eq. (2.17) is zero. Applying the complex representation introduced in Sec. 2.1.6 yields

$$\Delta U(\mathbf{r}) + k_0^2\epsilon(\mathbf{r})U(\mathbf{r}) = 0 , \qquad \text{with} \qquad k_0^2 = \epsilon_0\mu_0\mu\omega^2 = (2\pi/\lambda)^2 , \tag{2.28}$$

where $U = \mathbf{E}_\omega \cdot \hat{y}$ is the scalar representation of the (complex) electric field and λ is the free space wavelength. Equation (2.28) is the inhomogeneous *Helmholtz equation* for the electric field in TE-polarization.

If the magnetic field is parallel to the $\hat{y}$-axis (*TM-polarization*) the situation is somewhat more complicated. Due to the dielectric inhomogeneity, the right hand side becomes different from zero, viz.

$$\Delta V(\mathbf{r}) + k_0^2\epsilon(\mathbf{r})V(\mathbf{r}) = \epsilon(\mathbf{r})^{-1}\ \nabla\epsilon(\mathbf{r})\ \nabla V(\mathbf{r}) , \tag{2.29}$$

where $V = \mathbf{H}_\omega \cdot \hat{y}$ is the scalar representation of (complex) magnetic field in TM-polarization. The right hand side of Eq. (2.29) can give rise to some problems, especially if the gradient of the permittivity is singular (for example in the case of a dielectric surface relief structure).

2.2.1 Poynting vector in two dimensional geometry

We have already seen that the time averaged Poynting vector $< \mathbf{S} >$ is the real part of the vector product of the complex electric field and the complex magnetic field (Eq. (2.26)). In addition, the magnetic field is contained in the imaginary part of the complex electric field (Eq. (2.23)). In two dimensional geometries these relations allow to express $< \mathbf{S} >$ for the TE-polarization as

$$< \mathbf{S} >_{TE} = \frac{-i}{2\omega\mu_0\mu} \mathrm{Re}\{U(\mathbf{r})\nabla U^*(\mathbf{r})\} \ . \tag{2.30}$$

Similar considerations are valid for the TM-polarization, thus

$$< \mathbf{S} >_{TM} = \frac{i}{2\omega\epsilon_0\epsilon} \mathrm{Re}\{V(\mathbf{r})\nabla V^*(\mathbf{r})\} \ . \tag{2.31}$$

Writing the complex field in amplitude and phase,

$$U = A_{TE}e^{i\phi_{TE}} \ , \quad \text{and} \quad V = A_{TM}e^{i\phi_{TM}} \ , \tag{2.32}$$

and introducing it into Eq. (2.30) and Eq. (2.31), respectively, yields

$$< \mathbf{S} >_{TE} = \frac{-1}{2\omega\mu\mu_0} A_{TE}^2 \nabla\phi_{TE} \ , \qquad \text{for TE} \ , \tag{2.33}$$

$$< \mathbf{S} >_{TM} = \frac{1}{2\omega\epsilon\epsilon_0} A_{TM}^2 \nabla\phi_{TM} \qquad \text{for TM} \ . \tag{2.34}$$

The Poynting vector is therefore always *orthogonal* to the phase contours of the complex fields. The absolute value of $< \mathbf{S} >$ is proportional to the square of the amplitude of the fields and the gradient of the phase. In ray optics, $\mathcal{L}(\mathbf{r}) = k^{-1}\phi(\mathbf{r})$ is called *Eikonal*. The gradient of the Eikonal $\nabla\mathcal{L}$ indicates the local propagation direction of the individual rays. Equations (2.33) and (2.33) confirm that the time averaged Poynting vector (and therefore the energy flow) is actually parallel to the gradient of the Eikonal.

The electric energy densities in TE and the magnetic energy density in TM become

$$< w_e >_{TE} = \frac{1}{4}\epsilon\epsilon_0 A_{TE}^2 \ , \qquad \text{and} \qquad < w_m >_{TM} = \frac{1}{4}\mu\mu_0 A_{TM}^2 \ . \tag{2.35}$$

For illustration, let us consider one part of the field of Fig. 2.1. Figure 2.2 a) shows the square of the amplitude of the complex electric field A_{TE}^2. In Fig. 2.2 b) the phase and

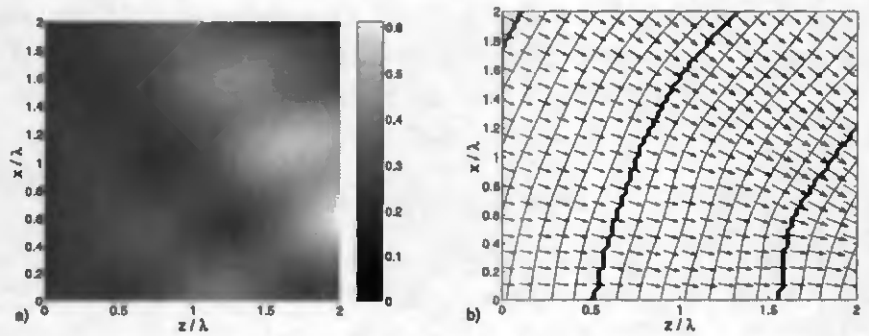


Figure 2.2: a) Square of the amplitude of a complex electric field. b) Phase and Poynting vector of the same field.

the Poynting vector are presented. The lines correspond the a contour phase plot, the distance between two neighbor lines is $\pi/4$ and between the thick lines 2π. The Poynting vectors (arrows) are always orthogonal to the phase lines. Their length is proportional to the local intensity (energy flow density).

Note that it is possible that the length of the Poynting vector is zero at a certain point even if the electric energy density is not zero. At this points the phase gradient is zero, the field amplitude however not. We will discuss this interesting point later when we are considering phase dislocations (Sec. 5.3).

Furthermore, Eq. (2.30) and Eq. (2.31) are in general more appropriated for the determination of the Poynting vector than Eq. (2.33) and Eq. (2.34). The latter equations include the phase of the complex fields. The determination of the (unwrapped) phase may be not trivial (or even impossible in the case of phase singularities).

Finally, in two dimensional geometry, the *curl* of the Poynting vector is always zero, $\nabla \times \mathbf{S} \equiv 0$, whereas the *curl* of the *time averaged Poynting vector is usually different from zero.* It can be shown that in two dimension geometry this vector points always in the $\hat{y}$-direction, viz.

$$\begin{aligned} \nabla \times <\mathbf{S}>_{TE} &= \frac{A_{TE}}{\omega \mu_0 \mu} \left[\partial_z \phi_{TE}\, \partial_x A_{TE} - \partial_x \phi\, \partial_z A_{TE}\right] \hat{y}\,, \\ \nabla \times <\mathbf{S}>_{TM} &= \frac{A_{TM}}{\omega \epsilon_0 \epsilon} \left[\partial_z \phi_{TM}\, \partial_x A_{TM} - \partial_x \phi\, \partial_z A_{TM}\right] \hat{y}\,. \end{aligned} \tag{2.36}$$

The *curl* of the time averaged Poynting vector contains the information about how *turbulent* a field is. It is an alternative mathematical description of the topography of a time harmonic electromagnetic field. We will see that $\nabla \times <\mathbf{S}>$ is different from zero

at phase dislocations, whereas it is always zero at saddle points (see Sec. 5.3). To our knowledge, this important difference between the *curl* of the Poynting vector and the *curl* of the time averaged Poynting vector has not yet been discussed in open literature.

2.3 Homogeneous media, plane waves

For homogeneous regions the Helmholtz equations simplify to

$$\Delta U(\mathbf{r}) + k_0^2 c U(\mathbf{r}) = 0\,, \tag{2.37}$$

$$\Delta V(\mathbf{r}) + k_0^2 c V(\mathbf{r}) = 0\,. \tag{2.38}$$

There exist an important set of solutions: the plane waves. They are characterized by their wave vector $\mathbf{k} = (k_x, k_z)$ indicating the direction of propagation of the wave. For a two dimensional geometry, plane waves are represented by

$$U(\mathbf{r}) = u\, e^{\mathrm{i}\mathbf{k}\mathbf{r}} = u \exp[\mathrm{i}k_x x + \mathrm{i}k_z z]\,, \tag{2.39}$$

Introducing Eq. (2.39) into the Helmholtz equation Eq. (2.38) imposes for the projections of the wave vector the relation

$$k_x^2 + k_z^2 = |\mathbf{k}|^2 = k_0^2 c\,. \tag{2.40}$$

Equation (2.40) implies that the length of the wave vector is fixed by the permittivity of the medium and the wavelength. Thus, the wave vector has to be on a sphere (resp. a circle in two dimensional geometry) of radius $k_0 n$, where $n = \sqrt{c}$ is the refractive index of the homogeneous medium. This sphere is referred to as the *Ewald sphere* [2]. Only one of the two components of the wave vector can be chosen independently. One of the consequences is, that there exist two different kinds of plane waves. If

$$k_x^2 < k_0^2 c\,, \tag{2.41}$$

the second component k_z is real. We get a *homogeneous* or *propagating* plane wave which has a constant amplitude u. If

$$k_x^2 > k_0^2 c\,, \tag{2.42}$$

the second component k_z becomes imaginary. Equation (2.39) rewrites as

$$U(\mathbf{r}) = u \exp[\mathrm{i}k_x x] \exp[\pm\gamma z]\,, \qquad \text{with} \quad \gamma = \mathrm{i}k_z = \sqrt{k_x^2 - k_0^2 c}\,. \tag{2.43}$$

We get *inhomogeneous* or *evanescent* waves. The physically meaningful sign in the exponent of Eq. (2.43) is determined by the geometry of the problem. In any case the field at

infinity has to be finite. Evanescent waves are characterized by an exponential change of their amplitude. The *penetration depth*

$$z_{eff} \stackrel{\text{def}}{=} 2\pi/\gamma = 2\pi\,(k_x^2 - k_0^2\epsilon)^{-1/2}\,, \tag{2.44}$$

is proportional to the inverse of γ. For large k_x the penetration depth z_{eff} is in the order of $2\pi\,k_x^{-1}$.

Note that the Poynting vector associated with an evanescent wave is *always parallel to the $\hat{x}$-axis* ($\nabla\phi = (k_x, 0, 0)$), i.e. there is no energy transport in the $\hat{z}$-direction. Evanescent waves play an important role in the interaction of light with optical microstructures.

2.4 Discussion

In this chapter, we established the basic notions for the mathematical treatment of optical fields. We outlined the advantage of introducing complex vectors to describe time harmonic fields. We showed that the inhomogeneous wave equation is more complicated for TM-polarization than for TE-polarization, especially if there are discontinuities in the permittivity function. We introduced the *curl* of the time averaged Poynting vector and showed that it indicates how turbulent a time harmonic electromagnetic field is. And finally, we showed that transverse field variations that are smaller than the wavelength give raise to exponentionally damped, and non-propagating (evanescent) waves.

Chapter 3

Interaction theories

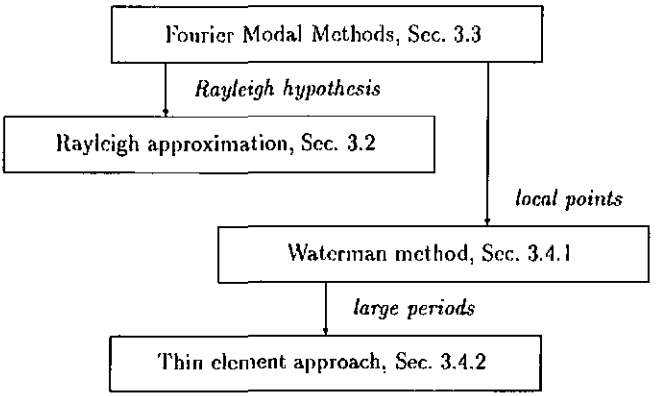


Figure 3.1: Different interaction models discussed in this chapter.

This chapter treats different theories of interaction of light with periodic optical microstructures (Fig. 3.1). A good overview of different integral and differential methods applied to grating structure has been presented by PETIT [3].

Here, we discuss in the first section some basic properties of all surface relief gratings. In Sec. 3.2 the well establish Rayleigh method is reviewed and applied to dielectric gratings. In Sec. 3.3 two different rigorous diffraction models are presented. Both are differential methods. The last section of this chapter (Sec. 3.4) outlines different approximated methods including the Waterman method, the thin element approach and a second order effective medium theory.

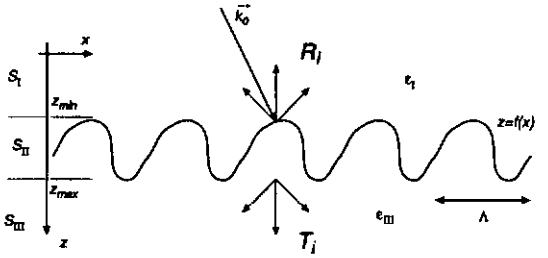


Figure 3.2: Setup for the grating diffraction problem.

3.1 Grating diffraction problem

The grating diffraction problem can be formulated as follows: *Given a periodic surface, delimiting the space into two homogeneous regions. What is the light field emerging from the structure if the surface is illuminated by a plane wave?* To answer this question different approaches have been proposed in the open literature. Let us first resume some basic properties common to all surface relief grating structures.

The grating surface profile is described by a periodic function $z = f(x) = f(x + \Lambda)$.

For commodity, we separate the space into three different regions (see Fig. 3.2): Two homogeneous region S_I and S_{III}, characterized by a permittivity ϵ_I and ϵ_{III}, respectively, and an inhomogeneous region which includes the periodic surface modulation for $z_{min} < z < z_{max}$. The field in the first region may be written as a sum of the incident field $E_0(\mathbf{r})$ and the reflected field $E_r(\mathbf{r})$. The field in the third region is the transmitted field $E_t(\mathbf{r})$, viz.

$$E_I(\mathbf{r}) = E_0(\mathbf{r}) + E_r(\mathbf{r}), \qquad \text{and} \qquad E_{III}(\mathbf{r}) = E_t(\mathbf{r}). \tag{3.1}$$

Let us suppose that the incident field is an infinite plane wave

$$E_0 = e^{i\alpha_0 x + i\beta_0 z}. \tag{3.2}$$

Equation (3.2) and the periodicity of the geometry suggest that the reflected and the transmitted fields are also periodic in the $\hat{x}$-direction, i.e. they can be expressed as Fourier series

$$E_r(\mathbf{r}) = e^{i\alpha_0 x} \sum_n v_n(z) e^{-inKx} = \sum_n v_n(z) e^{i\alpha_n x}, \qquad K = \frac{2\pi}{\Lambda}, \tag{3.3}$$

$$E_t(\mathbf{r}) = e^{i\alpha_0 x} \sum_n w_n(z) e^{-inKx} = \sum_n w_n(z) e^{i\alpha_n x}, \tag{3.4}$$

with

$$\alpha_n = \alpha_0 - nK. \tag{3.5}$$

If we insert Eq. (3.3) into the Helmholtz equation Eq. (2.28) we find that

$$\frac{\mathrm{d}v_n}{\mathrm{d}z} + (k^2 c_I - \alpha_n^2) v_n = 0\,, \quad \text{and} \quad \frac{\mathrm{d}w_n}{\mathrm{d}z} + (k^2 c_{III} - \alpha_n^2) w_n = 0\,. \tag{3.6}$$

The general solutions of Eq. (3.6) are

$$v_n(z) = A_n e^{\mathrm{i}\beta_{1,n} z} + B_n e^{-\mathrm{i}\beta_{1,n} z}\,, \quad \text{with} \quad \beta_{1,n} = \sqrt{k^2 c_I - \alpha_n^2}\,, \tag{3.7}$$

$$w_n(z) = C_n e^{\mathrm{i}\beta_{1,n} z} + D_n e^{-\mathrm{i}\beta_{1,n} z}\,, \quad \text{with} \quad \beta_{2,n} = \sqrt{k^2 c_{III} - \alpha_n^2}\,. \tag{3.8}$$

We see that the reflected and the transmitted field consists of superposition of plane waves. The direction of propagation are given by the grating period. From the physical point of view the reflected field can only consist of backward traveling waves and the transmitted of forward traveling waves, i.e. $A_n \equiv 0$ and $D_n \equiv 0$. We finally find the general expansion of the reflected and transmitted field (called *Rayleigh expansion*), defining $R_n \equiv B_n$ and $T_n \equiv C_n$,

$$E_r(\mathbf{r}) = \sum_n R_n e^{\mathrm{i}\alpha_n x - \mathrm{i}\beta_{1,n} z}\,, \qquad \text{for } z < \min f(x)\,, \tag{3.9}$$

$$E_t(\mathbf{r}) = \sum_n T_n e^{\mathrm{i}\alpha_n x + \mathrm{i}\beta_{2,n} z}\,, \qquad \text{for } z > \max f(x)\,. \tag{3.10}$$

Considering Eq. (3.9) and Eq. (3.10) we recognize that solving the grating diffraction problem means determining the complex amplitudes of the reflected and transmitted waves R_n and T_n.

In most cases, we are interested in the *diffraction efficiency*, that is the energy flux associated to every diffraction order. Based on the definition of the Poynting vector, Eq. (2.16), we define the diffraction efficiencies of the nth reflected and transmitted wave as

$$\eta_{R,n} = \mathrm{Re}\{\beta_{1,n}/\beta_0\}|R_n|^2\,, \tag{3.11}$$

$$\eta_{T,n} = C\,\mathrm{Re}\{\beta_{2,n}/\beta_0\}|T_n|^2\,, \tag{3.12}$$

where

$$C = \begin{cases} 1 & \text{for TE-polarization} \\ \dfrac{c_I}{c_{III}} & \text{for TM-polarization} \end{cases}\,. \tag{3.13}$$

The real part of the z-projection is taken to exclude evanescent waves. They do not contribute to the diffraction efficiency. It is obvious that for non-absorbing materials (i.e. dielectric materials) the sum of all diffraction efficiencies should be equal to 1, viz.

$$\sum_n \eta_{R,n} + \eta_{T,n} = 1\,. \tag{3.14}$$

Equation (3.14) expresses the fundamental law of energy conservation. It is also useful, or even necessary, to verify the validity of Eq. (3.14) for the different diffraction theories.

3.2 The Rayleigh method

Let us turn back to the grating diffraction problem. Historically, Lord Rayleigh introduced a first approach at the beginning of the century for reflection gratings [4]. During almost 60 years no contradiction to this approach has been found. With the introduction of modern numerical computation tools failure of the method has been found especially for deep grating structures [5]. Different papers discuss the validity of the Rayleigh method for sinusoidal [6] and general periodic [7, 8, 9, 10, 11, 12, 13, 14, 15] surfaces. Different methods have been proposed to improve the convergence [16]. It has generally been accepted the Rayleigh method is neither always wrong nor always valid. Of course this constraint limits the application of this method. On the other hand, the method is very easy to understand and to implement. We present it from a numerical analysis point of view. We apply it to the case of dielectric surface relief gratings, which is not treated to our knowledge in literature.

The Rayleigh method is based on the following hypothesis: The Rayleigh expansions, Eq. (3.9) and Eq. (3.10), are not only valid in the two homogeneous region S_I and S_{III} but also *inside* the modulated region. Therefore, we can write

$$E_{I+} = e^{i\alpha_0 x + i\beta_0 z} + \sum_n R_n e^{i\alpha_n x - i\beta_{1,n} z}, \qquad \text{for} \quad z < f(x), \tag{3.15}$$

$$E_{III+} = \sum_n T_n e^{i\alpha_n x + i\beta_{2,n} z}, \qquad \text{for} \quad z > f(x). \tag{3.16}$$

Note that the only difference to Eq. (3.9) and Eq. (3.10) is in the domain of definition of the functions. On the grating surface the fields have to satisfy the boundary conditions

$$E_{I+}(x,z) - E_{III+}(x,z) = 0, \qquad \forall x, z \in \mathcal{P}, \tag{3.17}$$

$$\frac{\mathrm{d}}{\mathrm{d}n} E_{I+}(x,z) - C \frac{\mathrm{d}}{\mathrm{d}n} E_{III+}(x,z) = 0, \qquad \forall x, z \in \mathcal{P}, \tag{3.18}$$

where C is again given by Eq. (3.13). The grating diffraction problem can now be reformulated as follows: *How do we have to choose the amplitudes of the reflected and transmitted waves (Eq. (3.16) and Eq. (3.16)) in order to best satisfy the boundary conditions (Eq. (3.17) and (Eq. (3.18)).* The grating diffraction problem is therefore equivalent to a parameter optimization. The merit functions to be minimized is the errors made in the boundary conditions, viz.

$$\chi_1^2 = \int_{\mathcal{P}} |E_{I+}(x,z) - E_{III+}(x,z)|^2 \mathrm{d}l, \tag{3.19}$$

$$\chi_2^2 = \int_{\mathcal{P}} \left| \frac{\mathrm{d}}{\mathrm{d}n} E_{I+}(x,z) - C \frac{\mathrm{d}}{\mathrm{d}n} E_{III+}(x,z) \right|^2 \mathrm{d}l. \tag{3.20}$$

For numerical implementation the boundary is sampled at M different points. Furthermore, we limit the number of reflected and transmitted waves R_n and T_n to N (including

propagating and evanescent diffraction orders). We therefore get M equations of type Eq. (3.17) and M equations of type Eq. (3.18) with $2N$ unknowns. The nice thing is that these equations are *linear* in R_n and T_n, hence, we can write the boundary conditions in matrix form

$$\mathbf{M}^I\,\mathbf{R} + \mathbf{M}^{III}\,\mathbf{T} = \mathbf{b}_1\,, \tag{3.21}$$

$$\mathbf{N}^I\,\mathbf{R} + \mathbf{N}^{III}\,\mathbf{T} = \mathbf{b}_2\,. \tag{3.22}$$

In the first equation the matrices $\mathbf{M}^I$ and $\mathbf{M}^{III}$ are according the first boundary condition Eq. (3.17) given by

$$\mathbf{M}^I_{n,m} = \exp[\mathrm{i}\alpha_n x_m - \mathrm{i}\beta_{1,n} f(x_m)]\,, \tag{3.23}$$

$$\mathbf{M}^{III}_{n,m} = -\exp[\mathrm{i}\alpha_n x_m - \mathrm{i}\beta_{2,n} f(x_m)]\,, \tag{3.24}$$

and the vector $\mathbf{b}_1$ by

$$\{\mathbf{b}_1\}_m = -\exp[\mathrm{i}\alpha_0 x_m + \mathrm{i}\beta_0 f(x_m)]\,. \tag{3.25}$$

The second equation includes the field derivative normal to the boundary surface. It is not difficult to show that for a general plane wave $u = \exp(\mathrm{i}\alpha x + \mathrm{i}\beta z)$ this field derivative can be expressed as

$$\frac{\mathrm{d}u}{\mathrm{d}n} = \frac{-\alpha f' + \beta}{\sqrt{1+f'^2}}\;u\,. \tag{3.26}$$

The matrices $\mathbf{N}^I$ and $\mathbf{N}^{III}$ are therefore given by

$$\mathbf{N}^I_{n,m} = (-\alpha_n f'(x_m) + \beta_{1,n})\,\exp[\mathrm{i}\alpha_n x_m - \mathrm{i}\beta_{1,n} f(x_m)]\,, \tag{3.27}$$

$$\mathbf{N}^{III}_{n,m} = (\alpha_n f'(x_m) - \beta_{2,n})\,\exp[\mathrm{i}\alpha_n x_m - \mathrm{i}\beta_{2,n} f(x_m)]\,, \tag{3.28}$$

and the excitation vector $\mathbf{b}_2$ by

$$\{\mathbf{b}_2\}_m = (\alpha_0 f'(x_m) + \beta_0)\,\exp[\mathrm{i}\alpha_0 x_m + \mathrm{i}\beta_0 f(x_m)]\,. \tag{3.29}$$

Due to the linearity of the problem, the optimization problem is simplified. In numerical analysis, problems of this type is referred to as the *general linear least squares method* [17]. The general form is

$$y(x) = \sum_{k=1}^{M} a_k X_k(x)\,, \tag{3.30}$$

where a_k are the parameters to be determined, $X_1(x), \ldots, X_M(x)$ are the basis functions. In our case

$$a_k = \begin{bmatrix} R_n \\ T_n \end{bmatrix}\,, \quad \text{and} \quad X_k(x_i) = \begin{bmatrix} \mathbf{M}^I & \mathbf{M}^{III} \\ \mathbf{N}^I & \mathbf{N}^{III} \end{bmatrix}\,. \tag{3.31}$$

The figure of merit for this linear model is

$$\chi^2 = \sum_{i=1}^{N} \left[y_i - \sum_{k=1}^{M} a_k X_k(x_i) \right]^2 . \tag{3.32}$$

Equation (3.32) is nothing else than the approximation of the integrals in Eqs. (3.19) and (3.20). It can be shown that the best set of parameters in the sense of χ^2 is given by [17]

$$a_j = \sum_{k=1}^{M} C_{jk} \left[\sum_{i=1}^{N} y_i X_k(x_i) \right] , \tag{3.33}$$

where C_{jk} is the covariance matrix defined by

$$C_{jk} = [\alpha]^{-1}_{jk} , \quad \text{with} \quad [\alpha]_{jk} = \sum_{i=1}^{N} X_j(x_i) X_k(x_i) . \tag{3.34}$$

In most modern high-level computer languages (such as MATLAB™) the general linear least square method is a *single line command*, namely

```
>> a = X \ y
```

That is the reason why the numerical implementation of the Rayleigh method is straightforward.

3.2.1 Validity of the Rayleigh method

Once we have found the best set of parameters (in our case the best fitted complex amplitudes of the plane waves) it is important to check its validity. For the Rayleigh approach, there exist two straightforward methods for this purpose. First, from the physics point of view, it is clear that the energy has to be conserved for dielectric materials, see Eq. (3.14). We therefore define an error in the energy conservation law as

$$\epsilon_n = \left| 1 - \sum_n (\eta_{R,n} + \eta_{T,n}) \right| . \tag{3.35}$$

Second, from the mathematical point of view we have to verify the *goodness-of-fit* of the model. For this purpose we introduce [17]

$$Q = \gamma(\nu/2, \chi^2/2) , \tag{3.36}$$

where $\gamma(a, x)$ is the *incomplete gamma function* and ν the degrees of freedom in the fitting model (i.e. $\nu = N - M$). If Q is close to zero for a set of data, then it is most probable that the model is correct. In terms of Rayleigh method this means that the plane wave expansion of the fields inside the modulation is correct. If Q is close to 1, the mathematical model is likely to be wrong.

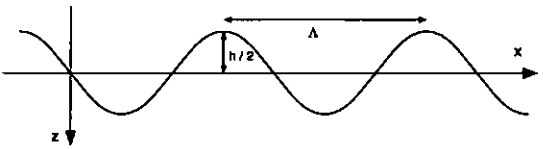


Figure 3.3: Sinusoidal grating.

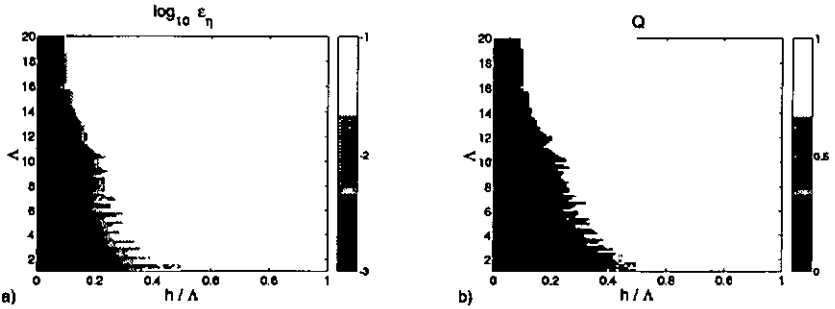


Figure 3.4: Region of validity of the Rayleigh approach for a sinusoidal surface relief grating. a) the energy conservation error and b) the goodness-of-fit are reported.

Example Let us look at a sinusoidal surface relief grating of period Λ and height h (Fig. 3.3). The refractive index of the first medium is $n_1 = 1.5$ and of the second medium $n_2 = 1$. The grating is illuminated by a perpendicular incident plane wave of TE-polarization. We calculate Q and ϵ_η for different grating periods and grating heights. In Fig. 3.4 a) the energy conservation error ϵ_η, Fig. 3.4 b) the goodness-of-fit Q is reported. There is an excellent agreement between the two graphs. We therefore conclude that for certain types of grating configurations (light regions) the Rayleigh approach is mathematically *and* physically wrong.

The inverse statement is less trivial: How can we show that in the dark region (i.e. in regions with energy conservation) the Rayleigh approach is correct? One possibility is to compare the diffraction efficiency calculated by the Rayleigh method with rigorous calculations. Section 3.3 will introduce different rigorous methods. Here, we only show calculations for a sinusoidal grating of period $\Lambda = 3\lambda$ (Fig. (3.5)). The diffraction efficiency versus the grating height is reported. In the region with $Q = 0$ a good agreement exists between the rigorous and the Rayleigh method.

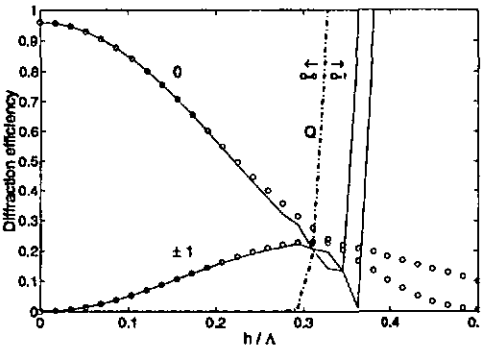


Figure 3.5: Diffraction efficiency versus grating height for a sinusoidal grating with $\Lambda = 3\lambda$. The lines correspond to the Rayleigh approach, the circles to the rigorous calculated points. Q is the goodness-of-fit of the Rayleigh approach.

3.3 Rigorous methods

We have seen that there are cases in which the Rayleigh method gives wrong results. Especially for deep grating structures, the field representation inside the modulated region is not valid anymore. Therefore, we have to base the field expansion inside the modulated region on the wave equations. For this purpose different rigorous approaches have been reported. Normally, they are classified in four different groups: analytical, integral, finite element, and differential methods. Unfortunately, analytic expressions of the rigorous solutions for the diffraction are rare. In principle, they exist only for the case where the discontinuity of the permittivity function coincide with the axis of a *elementary* coordinate system (i.e.flat surface, circular, or elliptical interface[1]). Integral and finite element are mostly used to model non-periodic structures. For period structures, methods solving the differential wave equation are generally used. Usually, they are based on a *modal* expansion of the field. The initial work on binary surface relief structures has been made by BURCKHARDT [19],KASPAR [20], and KNOP [21]. Therefore, the method is sometimes called BKK-method. Later MOHARAM applied it to general surface relief structures [22] by slicing the general surface into a stack of lamellar gratings. He called the method *rigorous coupled wave approach*, RCWA. In practice, it can be shown that the method leads to the same mathematical differential equation system than the BKK method. For

[1]There exist a method (C-method) introduced by CHANDEZON [18] which makes a transformation of the coordinates in a way that the surface profile becomes flat. Instead of having a simple wave equation and complicate boundaries, the C-method has simple boundaries but a complicate wave equation.

this reason we prefer to use the term *eigenmode method [23]* or better *Fourier modal method, FMM* (in contrast to other modal approaches as presented by MORF [24]). The FMM has also been successfully applied to cross-gratings [25] and gratings of anisotropic materials [26].

Since the introduction of the method in the early 70's, different mathematical improvements have been proposed [27, 28, 29, 30]. The main problem is still the poor convergence of the FMM in the case of TM-polarized plane waves diffracted by a general surface relief grating structure of non-infinite conducting metals.

In this section, we first introduce the concept of the coupled-wave analysis. The method is based on a physical approach and is therefore quite intuitive [31]. The solution is a superposition of individual waves. The waves are coupled through the Fourier coefficients of the permittivity expansion of the surface modulation. We will see that this approach leads to a second order differential equation system with *non-constant* terms. The situation can be handled by reformulating the wave approach by a modal approach. Second, we outline the rigorous eigenmode approach. The Helmholtz equation, a partial differential equation, is separated in x and z. The two resulting (ordinary) differential equations are easily solved.

3.3.1 Slicing the surface relief

The first step, common to all modal approaches (including RCWA, FMM, Morf), is to remove the z dependency of the permittivity function $\epsilon(x, z)$. This is done by slicing the surface relief structure into a stack of lamellar layers (see Fig. 3.6 a). Inside each layer, the permittivity is supposed to be constant in z direction, $\epsilon(x, z_i)$. It is now the *binary permittivity function*, (Fig. 3.6 a), and not the binary surface function, which is expanded into a Fourier series (in RCWA,FMM), resp. polynomials (Morf, [24]). Inside each layer a field ansatz is made and at the end the different fields are matched through the boundary conditions.

We will now study the field expansion for the RCWA, resp. the FMM in more detail. For simplicity of notation we will omit the numbering of the individual layers. We consider therefore the field inside one slice.

3.3.2 Coupled-wave analysis

The general concept of coupled waves is used to describe the propagation of waves in a media which can be characterized by any kind of perturbation. The solution of the perturbational case is found by linearly superposing non-perturbational solutions. This

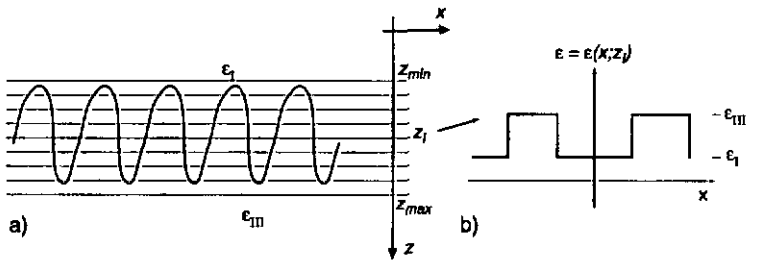


Figure 3.6: The z dependency of the permittivity function $\epsilon(x,z)$ is removed by dividing the surface relief into a stack of lamellar layers, a). The permittivity function (b) of each layer is then expand into a Fourier series.

method has been extensively applied to volume holograms which have small perturbation terms. It is however also possible to apply the method for surface relief gratings.

Coupled wave ansatz

Let S_{II} be the modulated region. If there is no perturbation S_{II} is a transparent, homogeneous and isotropic media, characterized by a dielectric constant ϵ_{II}. As a solution of the homogeneous Helmholtz equation we choose plane waves

$$E_j(x,z) = e^{-i\alpha_j x - i\beta_j z}\,, \tag{3.37}$$

with

$$\alpha_j^2 + \beta_j^2 = k^2\epsilon_{II}\,. \tag{3.38}$$

Assuming an infinite surface-relief grating, the perturbation may be written as a Fourier series

$$p(\mathbf{r}) = k^2 E(x,z)\sum_l \epsilon_l \exp(-ilKx)\,, \qquad \text{with} \qquad K = 2\pi/\Lambda\,. \tag{3.39}$$

The coupled wave approach assumes now, that the electric field $E(x,z)$ in the perturbational medium can be obtained by a linear superposition of a certain number of plane waves, namely

$$E(x,z) = \sum_j A_j(z)E_j(x,z) = \sum_j A_j(z)e^{-i\alpha_j x - i\beta_j z}\,, \tag{3.40}$$

In the concept of perturbation theory, the inhomogeneous Helmholtz equation writes as [31]

$$\Delta U(\mathbf{r}) + k_0^2\epsilon(\mathbf{r})U(\mathbf{r}) = p(\mathbf{r})\,, \qquad \text{with} \qquad k_0^2 = \epsilon_0\mu_0\mu\omega^2 = (2\pi/\lambda)^2\,. \tag{3.41}$$

Introducing Eq. (3.40) into Eq. (3.41) yields

$$\sum_j \left[A_j \Delta E_j + 2\frac{\mathrm{d}A_j}{\mathrm{d}z}\frac{\partial E_j}{\partial z} + \frac{\mathrm{d}^2 A_j}{\mathrm{d}z^2} E_j + k^2 \epsilon_{II} A_j E_j \right] = p(\mathbf{r}) \,. \tag{3.42}$$

Because the individual plane waves satisfy the homogeneous wave equation, Eq. (3.42) is reduced to

$$\sum_j \left[2\frac{\mathrm{d}A_j}{\mathrm{d}z}\frac{\partial E_j}{\partial z} + \frac{\mathrm{d}^2 A_j}{\mathrm{d}z^2} E_j \right] = \sum_j \left[-\mathrm{i}2\beta_j \frac{\mathrm{d}A_j}{\mathrm{d}z} + \frac{\mathrm{d}^2 A_j}{\mathrm{d}z^2} \right] E_j = p(\mathbf{r}) \,. \tag{3.43}$$

Orthogonality of the plane waves

If we multiply Eq. (3.43) by a plane wave E_m and integrate then over the whole space, we can take advantage of the orthogonality

$$\int_{-\infty}^{\infty} \mathrm{d}x E_j(x,z) E_m^*(x,z) = e^{-\mathrm{i}(\beta_j - \beta_m)z} \int_{-\infty}^{\infty} \mathrm{d}x e^{-\mathrm{i}(\alpha_j - \alpha_m)x} = \delta_{jm} \,, \tag{3.44}$$

of the individual plane waves. Using this orthogonality relation Eq. (3.43) becomes

$$\frac{\mathrm{d}^2 A_m}{\mathrm{d}z^2} - 2i\beta_m \frac{\mathrm{d}A_m}{\mathrm{d}z} = e^{\mathrm{i}\beta_m z} \int_{-\infty}^{\infty} \mathrm{d}x e^{\mathrm{i}\alpha_m x} p(\mathbf{r}) \,. \tag{3.45}$$

Introducing now the perturbation term given by Eq. (3.39) into Eq. (3.45)

$$\frac{\mathrm{d}^2 A_m}{\mathrm{d}z^2} - 2i\beta_m \frac{\mathrm{d}A_m}{\mathrm{d}z} = k^2 e^{\mathrm{i}\beta_m z} \sum_l \epsilon_l \int_{-\infty}^{\infty} \mathrm{d}x e^{\mathrm{i}(\alpha_m - lK)x} E(x,z) \,. \tag{3.46}$$

The right hand side of Eq. (3.46) becomes

$$k^2 \sum_j \sum_l e^{\mathrm{i}(\beta_m - \beta_j)z} A_j \epsilon_l \int_{-\infty}^{\infty} \mathrm{d}x e^{\mathrm{i}(\alpha_m - lK - \alpha_j)x} \,. \tag{3.47}$$

Non-vanishing integrals are only obtained if

$$\alpha_j = \alpha_m - lK \,, \tag{3.48}$$

which is recognized as the condition for diffraction by the grating. We see that *the condition for diffraction is included in the concept of coupled waves.* The set of plane waves is discrete and the spacing in the α-space (or frequency space) is K. Equation (3.48) mutually relates the x-projections of the individual waves. The offset value (i.e. α_0) is given by the x-projection of the incident wave plane wave

$$\alpha_0 = \sqrt{\epsilon_I} \sin\phi_0 \,, \tag{3.49}$$

where ϕ_0 is the angle of incidence.

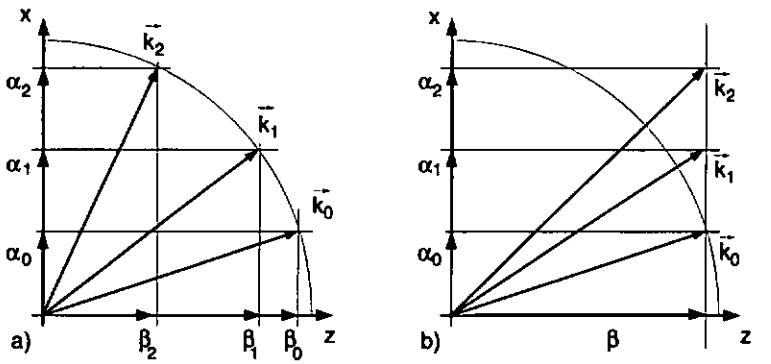


Figure 3.7: Different possible choices for the beta-projection. a) beta value leads to differential equations with non-constant coefficients, b) K-vector closure is used in the rigorous classical coupled wave approach, (RCWA).

Coupled wave equations

By replacing the right hand side of Eq. (3.46) with Eq. (3.47) the coupled wave equations become

$$\frac{\mathrm{d}^2 A_m}{\mathrm{d}z^2} - 2i\beta_m \frac{\mathrm{d}A_m}{\mathrm{d}z} = k^2 \sum_j e^{i(\beta_m - \beta_j)z} \epsilon_{m-j} A_j \,, \tag{3.50}$$

Equation (3.50) is a system of coupled differential equations of non-constant terms. This formulation is called *beta-value* approach. The different plane waves are coupled through the Fourier harmonics of the perturbation expansion ϵ_{m-j}. If the perturbation is weak (in the sense that there are only smooth transition) only few coefficients are necessary to express the solution. For a sinusoidal index grating (like a volume hologram) only one Fourier coefficient is necessary. If one takes only two waves and neglects the second order derivative, the solution proposed by KOGELNIK [32] is found. For surface relief structures, however, there is normally a sharp jump of the permittivity inside the modulation. In this situation, a large number of coupling coefficients have to be taken into account. This may lead to difficulties for the numerical implementation of Eq. (3.50) because of the non-constant coefficients. One possibility to overcome this problem is to choose another set of elementary waves.

Alternative choice of the individual waves

In our previous formulation of the coupled waves, the individual waves were chosen in the following way: First we said that they are solutions of the non-perturbational wave

equation. This requirement relates the projection α_j and β_j of the wave vectors trough Eq. (3.38) (see Fig. 3.7 a). We have seen that this leads coupled wave equations with non-constant coefficients. The non-constant term ($e^{i(\beta_m-\beta_j)}$) can be avoided if we choose all the β_j *constant*:

$$\beta_j = \beta = \sqrt{k^2\epsilon_{II} - \alpha_0^2}\,, \qquad \text{i.e.} \quad E_j(x,z) = e^{-i\alpha_j x - i\beta z}\,. \tag{3.51}$$

Note that $t\beta$ has no index anymore, because it is the same for all the terms (Fig. 3.7 b). Note also that the individual "waves" do not satisfy the wave equation anymore. We better call them *space harmonics*. From Eq. (3.42) we get

$$\frac{\mathrm{d}^2 A_m}{\mathrm{d}z^2} - 2i\beta\frac{\mathrm{d}A_m}{\mathrm{d}z} + (\alpha_m^2 - \beta^2)\delta_{mj}A_m = k^2\sum_j \epsilon_{m-j}A_j \tag{3.52}$$

instead of Eq. (3.50). This corresponds to the rigorous coupled wave approach (RCWA) presented by Moharam and Gaylord [22]. Setting $A_{1,m} = A_m$ and $A_{2,m} = \mathrm{d}A_m x/\mathrm{d}z$, the second order differential equation system is reduced to a first order system, viz.

$$\begin{pmatrix} \dfrac{\mathrm{d}A_{1,m}}{\mathrm{d}z} \\ \dfrac{\mathrm{d}A_{2,m}}{\mathrm{d}z} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{M} & 2i\beta\mathbf{I} \end{pmatrix} \begin{pmatrix} A_{1,m} \\ A_{2,m} \end{pmatrix}, \tag{3.53}$$

where $\mathbf{0}$ is a zero matrix, $\mathbf{I}$ an unit matrix, and $\mathbf{M}$ a matrix with $M_{mj} = k^2\epsilon_{m-j} - (\alpha_m^2 - \beta^2)\delta_{mj}$. The four submatrices $(\mathbf{0},\mathbf{I},\mathbf{M},2i\beta\mathbf{I})$ form together the coupled wave matrix $\mathbf{S}$. The resulting system ($\dot{\mathbf{A}} = \mathbf{SA}$) can be solved by making for the A_m the eigenvalue ansatz

$$A_m(z) = \sum_p a_p\, w_{p,m}\, e^{\kappa_p z}\,, \tag{3.54}$$

where κ_p and $w_{p,m}$ are the eigenvalues and eigenvectors of $\mathbf{S}$. The values a_p are found through the boundary conditions (see Sec. 2.1.3).

The eigenvalues and eigenvectors of $\mathbf{S}$ are essentially determined by the matrix $\mathbf{M}$. Actually, it is not difficult to show that for each eigenvalue μ of $\mathbf{M}$ there exist *two* eigenvalues, determined by the second order equation

$$\kappa^2 - 2i\beta\kappa - \mu = 0\,. \tag{3.55}$$

The eigenvector w_m is then given by

$$w_m = \begin{pmatrix} v_m/\kappa \\ v_m \end{pmatrix}, \tag{3.56}$$

where v_m is an eigenvector of $\mathbf{M}$. Therefore, it is sufficient to search the eigenvalues and eigenvectors of $\mathbf{M}$; the eigenvalues and eigenvectors of $\mathbf{S}$ are found through Eqs. (3.55) and (3.56).

The numerical implementation is straightforward, that is the reason why the coupled wave method is so popular.

3.3.3 Fourier Modal Method (FMM)

Starting point of the Fourier Modal Method (FMM) is the Helmholtz equation (Eq. (2.38)). Separating the variables $E(x,z) = X(x)Z(z)$ leads to two ordinary differential equations

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}X(x) + \left[k^2\epsilon(x) - \gamma^2\right]X(x) = 0\,, \tag{3.57}$$

and

$$\frac{\mathrm{d}^2}{\mathrm{d}z^2}Z(z) + \gamma^2 Z(z) = 0\,, \tag{3.58}$$

where γ^2 is the separation constant. The solution of Eq. (3.58) can be expressed as

$$Z(z) = a\exp(\mathrm{i}\gamma z) + b\exp(-\mathrm{i}\gamma z)\,. \tag{3.59}$$

To solve Eq. (3.57) the complex relative permittivity is usually expressed in terms of a Fourier series

$$\epsilon(x) = \sum_{p=-\infty}^{\infty} \epsilon_p \exp(\mathrm{i}2\pi p x/\Lambda)\,. \tag{3.60}$$

The periodicity of ϵ implies (through the *Floquet-Bloch theorem*) a pseudo-periodic solution, i.e. $X(x+\Lambda) = \exp(\mathrm{i}\alpha_0\Lambda)X(x)$. Therefore, we expand $X(x)$ as follows

$$X(x) = \sum_{m=-\infty}^{\infty} P_m \exp(\mathrm{i}\alpha_m x)\,, \tag{3.61}$$

where

$$\alpha_m = \alpha_0 + 2\pi m/\lambda\,, \quad \text{and} \quad \alpha_0 = k_0\sin(\theta_0)\,. \tag{3.62}$$

The complete solution of Eq. (2.38) is given by

$$E_y(x,z) = \sum_{l=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} P_{ln}\exp(\mathrm{i}\alpha_l x)\left\{a_n\exp(\mathrm{i}\gamma_n z) + b_n\exp(-\mathrm{i}\gamma_n z)\right\}\,, \tag{3.63}$$

where P_{ln} and γ_n are found by inserting Eq. (3.63) and Eq. (3.60) in Eq. (2.38). We obtain

$$\sum_{m=-\infty}^{\infty}\sum_{p=-\infty}^{\infty}\left[(\alpha_m^2+\gamma^2)\exp(\mathrm{i}\alpha_p x)\delta_{mp} - k^2\epsilon_p\exp(\mathrm{i}\alpha_{m+p}x)\right]P_m = 0\,. \tag{3.64}$$

Multiplication with $\exp(-i\alpha_l x)$ and integration over the period d gives a system of equations

$$\sum_{m=-\infty}^{\infty} (k^2 \epsilon_{l-m} - \alpha_m^2 \delta_{lm}) P_m = \gamma^2 P_l \,, \quad \text{or} \quad \mathbf{Q} \cdot \mathbf{P} = \gamma^2 \mathbf{P} \,. \tag{3.65}$$

Thus, P_n and γ_n^2 are given by the eigenvalues and eigenvectors of $\mathbf{Q}$ with

$$Q_{lm} = k^2 \epsilon_{l-m} - \alpha_m^2 \delta_{lm} \,. \tag{3.66}$$

Note that the coupling matrix $\mathbf{Q}$ of the Fourier modal method is very close to the matrix $\mathbf{M}$ in the rigorous coupled wave approach, Eq. (3.53), namely

$$\mathbf{M} - \beta^2 \mathbf{I} = \mathbf{Q} \,. \tag{3.67}$$

Thus, the eigenvalues of $\mathbf{M}$ and $\mathbf{Q}$ are related by

$$\mu - \beta^2 = \gamma^2 \,. \tag{3.68}$$

Although the starting point is different, the rigorous coupled wave approach (RCWA) and the Fourier modal method (FMM) lead to the same numerical implementation. The term RCWA is probably more used in open literature than FMM. We prefer "Fourier modal method" because it describes better the physical aspect (inside the grating structure the field is expanded into *modes*).

3.3.4 Boundaries

Once we have found the fields inside each layer, we have to match them at the boundaries in order to determine the unknowns field amplitudes R_i, and T_i of the Rayleigh expansion Eqs. (3.9) and (3.10), as well as the amplitudes a_n and b_n of the modes in side each layer, Eq. (3.63). The boundary conditions (see Sec. 2.1.3) implies that for TE-polarization the field and its derivatives in the z direction are continuous across the boundaries. We will therefore get a large system of linear equations, which can be easily solved by using standard methods (for example Gauss elimination). The size of the matrix is $2(N+1)M \times 2(N+1)M$, where N is the number of waves retained in the analysis and M the number of layers used to describe the general surface relief structures. The factor two is due to the fact that there are always $2N$ unknowns inside each slab. The detailed description of the boundary problem is well documented in literature, see for example [23, Sec. 2.2.5].

3.4 Approximate methods

Although the numerical implementation of the FMM, or the RCWA, is not difficult, simplification of the theory is in many situations desirable. If the ratio between the grating period and the wavelength grows, the number of modes to be retained in the analysis increases. Thus, the CPU-time and the computer memory will also increase. We will therefore study different approximations. First, if the Rayleigh approach is valid (i.e. if the gratings are not too deep), it can be shown that the complex amplitudes of the different diffraction orders are given by a simple integral expression. Second, if the grating period is much larger than the optical wavelength, the integral expression is simplified and we will find the classical thin element approach. And third, if the grating period is much smaller than the wavelength, an analytical expression of the eigenvalues in the rigorous method exists.

3.4.1 Waterman method

The *Waterman method* has been introduced by WATERMAN [33] in the field of acoustic waves. Later MAYSTRE [34] applied the formalism to the grating diffraction problem for perfect conducting metallic surface relief structures. In this case, the boundary conditions are simplified. The field, for TE polarization, or the normal derivative of the field, for TM polarization, is equal to zero on the boundary. The consequence is that the amplitude of the reflected waves (there are only reflected waves for perfect metals) are given on the boundary by the normal derivative, for TE-polarization, or the field itself, for TM-polarization. Here, we try to establish a similar approach for dielectrics. The boundary conditions for two homogeneous dielectric region impose only the continuity of the field and its derivative across the border. We therefore get some kind of mixed expressions. The amplitudes are given by the field *and* the normal derivative of the field on the boundary for both polarizations. To our knowledge this approach has not yet been reported in open literature. In this work, we will not discuss the general validity of the method. The interested reader is referred to a recent paper of BAGIEU and MAYSTRE [35], discussing the validity of Waterman method for perfect metal gratings. The extension to dielectric gratings is straightforward.

Starting point of the Waterman approach is the Rayleigh expansion

$$E_r(\mathbf{r}) = \sum_n R_n e^{i\alpha_n x - i\beta_{1,n} z}, \tag{3.69}$$

$$E_t(\mathbf{r}) = \sum_n T_n e^{i\alpha_n x + i\beta_{2,n} z}, \tag{3.70}$$

Eq. (3.9) and Eq. (3.10). We know that (by construction) E_r and E_t satisfy the Helmholtz

equations

$$\Delta E_r + k_1^2 E_r = 0\,, \qquad \text{with} \quad k_1^2 = k_0^2 \epsilon_I\,, \tag{3.71}$$

$$\Delta E_t + k_2^2 E_t = 0\,, \qquad \text{with} \quad k_2^2 = k_0^2 \epsilon_{III}\,. \tag{3.72}$$

For further considerations, we introduce two auxiliary functions

$$U_1(x,z) = \exp(\mathrm{i}\alpha_n x - \mathrm{i}\beta_{1,n} z)\,, \quad \text{and} \quad U_2(x,z) = \exp(\mathrm{i}\alpha_n x + \mathrm{i}\beta_{2,n} z)\,. \tag{3.73}$$

The two function satisfy the Helmholtz equations, thus also their conjugates

$$\Delta U_1^* + k_1^2 U_1^* = 0\,, \tag{3.74}$$

$$\Delta U_2^* + k_2^2 U_2^* = 0\,. \tag{3.75}$$

Multiplying the first pair of Helmholtz equations Eq. (3.71) and Eq. (3.72) by U_1^* and U_2^*, respectively, the second pair of Helmholtz equations Eq. (3.74) and Eq. (3.75) by E_r and E_t, respectively, and subtracting the first from the second yield

$$E_r \Delta U_1^* - U_1^* \Delta E_r = 0\,, \tag{3.76}$$

$$E_t \Delta U_2^* - U_2^* \Delta E_t = 0\,. \tag{3.77}$$

By integrating the above two equations over the hatched areas in Fig. 3.8 and using the second Green identity, it can be deduced that

$$\int\int_{S_1} (E_r \nabla^2 U_1^* - U_1^* \nabla^2 E_r)\,\mathrm{d}S = \int_{\partial S_1} (E_r \frac{\mathrm{d}U_1^*}{\mathrm{d}n} - U_1^* \frac{\mathrm{d}E_r}{\mathrm{d}n})\,\mathrm{d}l = 0\,, \tag{3.78}$$

$$\int\int_{S_2} (E_t \nabla^2 U_2^* - U_2^* \nabla^2 E_t)\,\mathrm{d}S = \int_{\partial S_2} (E_t \frac{\mathrm{d}U_2^*}{\mathrm{d}n} - U_2^* \frac{\mathrm{d}E_t}{\mathrm{d}n})\,\mathrm{d}l = 0\,, \tag{3.79}$$

where ∂S_1 and ∂S_2 are the boundaries of the hatched areas. Because of the periodicity of the problem, the paths parallel to the z-axis (i.e. PS'' and $R''Q$, respectively, $S'P$ and QR') are opposite and cancel each other.

The line integrals become therefore

$$\int_{\mathcal{P}} (E_r \frac{\mathrm{d}U_1^*}{\mathrm{d}n} - U_1^* \frac{\mathrm{d}E_r}{\mathrm{d}n})\,\mathrm{d}l = \int_{S'R'} (E_r \frac{\mathrm{d}U_1^*}{\mathrm{d}z} - U_1^* \frac{\mathrm{d}E_r}{\mathrm{d}z})\,\mathrm{d}x\,, \tag{3.80}$$

$$\int_{\mathcal{P}} (E_t \frac{\mathrm{d}U_2^*}{\mathrm{d}n} - U_2^* \frac{\mathrm{d}E_t}{\mathrm{d}n})\,\mathrm{d}l = \int_{S''R''} (E_t \frac{\mathrm{d}U_2^*}{\mathrm{d}z} - U_2^* \frac{\mathrm{d}E_t}{\mathrm{d}z})\,\mathrm{d}x\,. \tag{3.81}$$

The right hand side of Eq. (3.80) and Eq. (3.80) is further transformed. Because of the orthogonality of the plane waves, the individual amplitudes can be separated. After some cumbersome calculations we find that

$$\begin{aligned} R_n &= \tfrac{1}{2\mathrm{i}\Lambda\beta_{1,n}} \int_0^\Lambda \exp[-\mathrm{i}\alpha_n x - \mathrm{i}\beta_{1,n} f(x)]\phi_1(x)\mathrm{d}x\,, \qquad \text{with} \\ \phi_1(x) &= [1 + f'(x)^2]^{1/2} \tfrac{\mathrm{d}E_r}{\mathrm{d}n} + \mathrm{i}[-\beta_{1,n} - \alpha_n f'(x)]E_r\,, \end{aligned} \tag{3.82}$$

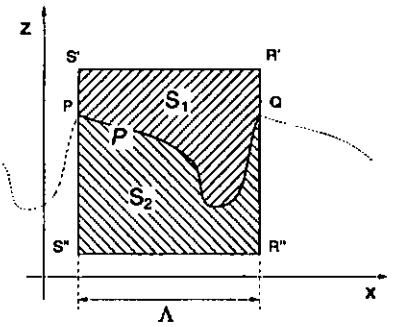


Figure 3.8: For the Waterman method, the second Greens identity is applied to S_1 and S_2.

and

$$T_n = \frac{1}{2i\Lambda\beta_{2,n}} \int_0^\Lambda \exp[-i\alpha_n x - i\beta_{2,n} f(x)]\phi_2(x)\mathrm{d}x\,, \qquad \text{with}$$
$$\phi_2(x) = [1 + f'(x)^2]^{1/2} \frac{\mathrm{d}E_t}{\mathrm{d}n} + i[-\beta_{2,n} - \alpha_n f'(x)]E_t\,. \tag{3.83}$$

We are now in the following situation: *The amplitude of the reflected and transmitted field expansion are given by an integral expression of the field and its normal derivative on the boundary*. Unfortunately, these two integral equations are coupled. Therefore we introduce the local point approximation. We consider the border consisting of locally independent points. The field at a certain point on $\mathcal{P}$ is then the same as the field obtained by replacing $\mathcal{P}$ by a perfect interface *tangent* to $\mathcal{P}$ at the same point. The same is valid for the normal derivative of the field: The normal derivative of the field at a certain point on $\mathcal{P}$ is the same as the normal derivative of the field obtained by replacing $\mathcal{P}$ by a perfect interface *tangent* to $\mathcal{P}$ at the same point. This statement is by the way related to the Rayleigh method (Sec. 3.2). The Rayleigh method assumes that the Rayleigh expansion is not only valid in the homogeneous region but also in the modulated region, and especially also on the boundary. The points on the boundary are therefore *independent* of each other.

The transmittance and reflectivity of a wave at a flat interface is described by the *Fresnel formulae* $r(x)$ and $t(x)$. We are therefore able to replace E_r and E_t on the boundary by the incident field E_i, viz.

$$E_r(x, f(x)) = r(x)E_i(x, f(x))\,, \tag{3.84}$$
$$E_t(x, f(x)) = t(x)E_i(x, f(x))\,, \qquad \text{with} \qquad E_i(x, f(x)) = e^{i\alpha_0 x + i\beta_0 f(x)}\,. \tag{3.85}$$

The Fresnel formula depends on the local slope of the grating border ($f'(x)$) and the polarization of the incident wave. After introducing Eqs. (3.84) and (3.85) in to Eqs. (3.82)

and (3.83) the complex amplitudes of the transmitted and reflected fields are found to be

$$R_n = \frac{1}{2}\int_0^\Lambda \psi_{1,n}(x)\exp[-\mathrm{i}(\beta_1-\beta_{1,n})f(x)]\exp(-inKx)\mathrm{d}x\,, \quad \text{with} \quad (3.86)$$

$$\psi_{1,n}(x) = \frac{1}{2i\beta_{1,n}}\{r'(x)f'(x)+\mathrm{i}r(x)[f'(x)(\alpha_0-\alpha_n)-(\beta_0-\beta_{1,n})]\}\,, \quad (3.87)$$

and

$$T_n = \frac{1}{2}\int_0^\Lambda \psi_{2,n}(x)\exp[-\mathrm{i}(\beta_1-\beta_{2,n})f(x)]\exp(-inKx)\mathrm{d}x\,, \quad \text{with} \quad (3.88)$$

$$\psi_{2,n}(x) = \frac{1}{2i\beta_{2,n}}\{t'(x)f'(x)+it(x)[f'(x)(\alpha_0-\alpha_n)-(\beta_0-\beta_{2,n})]\}\,. \quad (3.89)$$

The obtained equations are very interesting: The coefficients in the Rayleigh expansion are given by the *Fourier transform* of a complex function. The amplitude of the complex function $\psi_{j,n}$ is more or less the transmission of the interface t, plus the variation of the transmission $t'f'$ (which can be negligible in most cases). The phase of the complex function is the accumulated phase through the element. The expression is very close to the thin element approach, except that each diffraction order has its own accumulated phase, i.e. *the thickness of the grating is different for each diffraction order*. Geometrically this means that only the $\hat{z}$-projection of the wavevector (thus $\beta_{2,n}$) "sees" the element. For large diffraction angles (hence, for higher diffraction orders, or small grating periods), $\beta_{2,n}$ is small, therefore the element is "thinner" than for small diffraction angles.

3.4.2 Thin element approach

It is evident that for large grating periods the propagation constants $\beta_{2,n}$ of the central diffraction orders are equal. For perpendicular incidence ($\beta_1 = k_0 n_1$) Eq. 3.88 rewrites as

$$T_n = \int_0^\Lambda \exp[-ik_0(n_1-n_2)f(x)]\exp(-inKx)\mathrm{d}x = \mathcal{FT}\{\exp[i\phi(x)]\}\,, \quad (3.90)$$

which is the classical expression for the thin element approach. The element acts as a phase only element. The phase-delay

$$\phi(x) = -ik_0(n_1-n_2)f(x) \quad (3.91)$$

is proportional to the element thickness. The diffraction efficiencies are directly obtain by taking the Fourier transform of the complex transmission function.

The thin element approach is usually introduced through the concept of ray optics. The optical field is represented by independent rays. Crossing a thin element, each ray is delayed proportionally to the element thickness. This "intuitive" approach leads directly to Eq. 3.90. Here, we based the thin element approach directly on the wave equation.

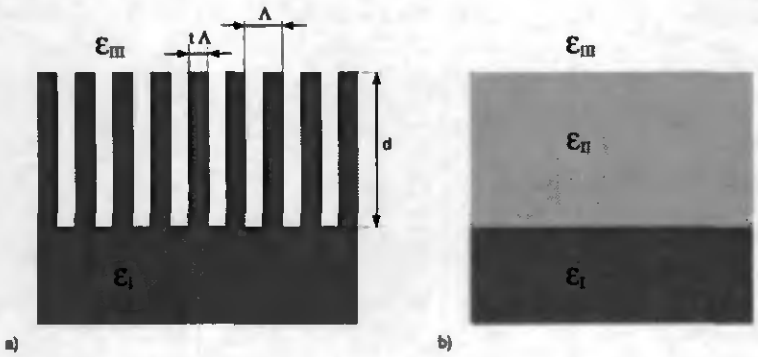


Figure 3.9: A subwavelength grating a) can be replaced by a homogeneous anisotropic layer b).

Often Eq. (3.90) is referred to as the scalar *diffraction* model (see for example [36, Sec. 10.4]), in contrast to the rigorous vector theories. From our point of view the notion *scalar diffraction* should be used for describing the *propagation* of a scalar wave, and not for the *interaction*.

3.4.3 Subwavelength gratings

It is well known that in the quasi static limit (i.e. the ratio between the grating period and the wavelength approaches zero), the optical properties of the grating are equivalent to a homogeneous birefringent layer (Fig. 3.9). These gratings are therefore used as artificial index structures [37], antireflection layers [38], and polarizing components [39]. The classical theory of subwavelength gratings expresses the homogeneous birefringent layer with two different electric permittivity [40]. In the first approach these are independent of the ratio between the grating period and the wavelength (ratio Λ/λ). It is possible to derive the expressions of the electric permittivity directly from the coupling matrix of the eigenmode approach (Eq. (3.66)), retaining only one wave in the analysis [41]. Here, we establish a second order effective index model by retaining three waves in the coupling matrix. We will compare the results with existing higher order theories reported in literature [38]. New to our knowledge is the expression for the amplitude of the first evanescent diffraction order of a subwavelength grating, Eq. (3.101). This allows to extend the validity of the effective index model to the near field of the element. We limit the analysis to TE-polarization and perpendicular incidence.

We have seen in Sec. 3.3.3 that the field inside a grating structure can be represented as

a superposition of different modes. The modes are coupled though the Fourier coefficients of the permittivity expansion. These coefficients are the off-diagonal elements in the coupling matrix **Q**. The diagonal elements represent the eigenmodes of the *unperturbated* media. This homogeneous medium is characterized by the zeroth Fourier coefficient ϵ_0 of the permittivity expansion. If the grating period is smaller than the wavelength (for perpendicular incidence) only the zeroth diffraction order is propagating. The elements on the diagonal becomes important since they are proportional to the square of the spatial frequencies of the different diffraction orders $\alpha_m^2 = (mK)^2 \sim (m/\Lambda)^2$. For very small grating periods, the coupling coefficient (i.e. the off diagonal elements) have no influence on the eigenmodes anymore. That is the raison why it is possible to limit the number of considered waves to three. If we retain only the coupling between next neighbors the coupling matrix becomes

$$Q_{lm} = \begin{pmatrix} k^2\epsilon_0 - K^2 & k^2\epsilon_1^* & 0 \\ k^2\epsilon_1 & k^2\epsilon_0 & k^2\epsilon_1^* \\ 0 & k^2\epsilon_1 & k^2\epsilon_0 - K^2 \end{pmatrix} . \tag{3.92}$$

It is possible to show that the eigenvalues of Eq. (3.92) are approximately

$$\gamma_1^2 = k^2\epsilon_0 - K^2 , \tag{3.93}$$

$$\gamma_2^2 = k^2\epsilon_0 - K^2 - k^2\tilde{\epsilon} , \tag{3.94}$$

$$\gamma_3^2 = k^2\epsilon_0 + k^2\tilde{\epsilon} , \tag{3.95}$$

with

$$\tilde{\epsilon} = \frac{2k^2\epsilon_1\epsilon_1^*}{K^2} = 2\left|\epsilon_1\right|^2 \left(\frac{\Lambda}{\lambda}\right)^2 . \tag{3.96}$$

The square root of this eigenvalues (thus $\gamma_{1,2,3}$) are the propagation constant of the modes in $\hat{z}$ direction. It is evident that for subwavelength gratings the first two propagation constant γ_1 and γ_2 are pure imaginary numbers ($k^2\epsilon_0 < K^2$). These modes are attenuated exponentially and can be neglected if the *grating thickness d is large compared to the penetration depth of the non-propagating modes.* Inside the grating structure the field is then mainly described by *one mode* (see Eq. (3.63)). The subwavelength grating acts therefore as a homogeneous layer. The permittivity of this layer is then given by Eqs. (3.95) and (3.96). It is not difficult to express the first two Fourier coefficients for a binary permittivity function analytically

$$\epsilon_0 = t\epsilon_I + (1-t)\epsilon_{III} , \tag{3.97}$$

and

$$\epsilon_1 = \frac{1}{\pi}\sin(\pi t)\left[\epsilon_I - \epsilon_{III}\right] , \tag{3.98}$$

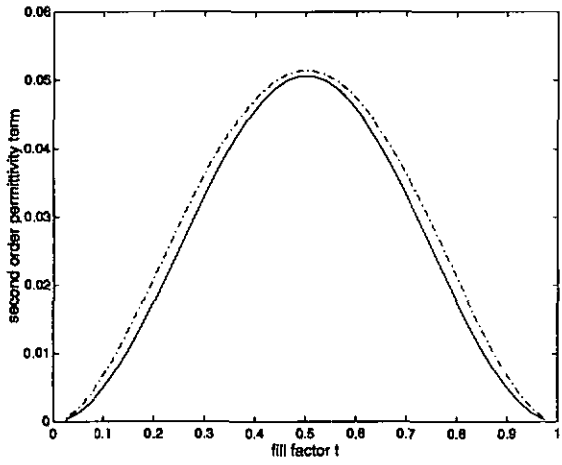


Figure 3.10: Second order term $\tilde{c}$ calculated from Eq. (3.99) (line) and from the work of Raguin [38, Eq. (3)] (dotted-line). The calculations are made for a grating of period $\Lambda/\lambda = 0.4$, $n_1 = 1.5$ and $n_3 = 1$, TE-polarization, and for different fill factors t.

where t is the normalized *fill factor* of the binary function ($0 < t < 1$). The artificial permittivity becomes

$$\epsilon_{II,TE} = \epsilon_0 + \tilde{\epsilon} = t\epsilon_I + (1-t)\epsilon_{II} + \frac{2}{\pi^2}\sin^2(\pi t)\,[\epsilon_I - \epsilon_{II}]^2 \left(\frac{\Lambda}{\lambda}\right)^2 . \qquad (3.99)$$

The first term ϵ_0 is exactly the same as for the classical zero-order effective index models [38]. The second term depends on the ratio Λ/λ.

Equation (3.99) is different from the second order effective index theory reported by other authors [38, Eq. (3)]. The difference is however quite small (see Fig. 3.10). Figure 3.11 shows the phase of the zero diffraction order of a fine binary grating for different ratios Λ/λ. The zero order effective index theory is only valid for very small grating periods. The second order theories (of the present work and of the work by Raguin [38]) show good agreement with the rigorously calculated values up to the appearance of the first propagating diffraction order.

In the effective index model presented above the subwavelength grating is replaced by a thin layer of artificial index. The artificial index is given by Eq. (3.99). In this representation the field in the output region is given by *one* propagating plane wave. The higher order evanescent waves are neglected. That is the reason why this approach is only valid in the far field of the element. The limitation is of course given by the *penetration depth* of the higher order evanescent waves. The amplitudes of these evanescent waves

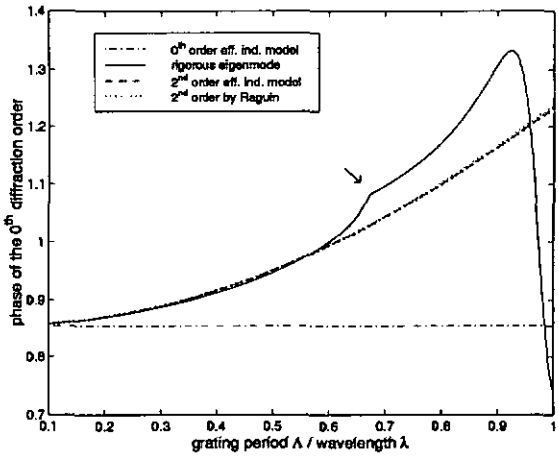


Figure 3.11: Phase of the zeroth diffraction order of a binary grating for different grating periods using several different models. The calculations were made for TE-polarization, $n_1 = 1.5, n_2 = 1, h = \lambda/2$, and $t = 0.5$. The arrow indicates where the first (reflected) diffraction order appears ($\Lambda = \lambda/1.5$).

(i.e. of the first and the minus first diffraction order) can easily by deduced. For this purpose we consider the *eigenvector* P_{l3} associated with the propagating mode (γ_3) inside the grating structure. It can be shown that

$$P_{l3} = \begin{pmatrix} k^2 c_1^* \\ K^2 \\ k^2 c_1 \end{pmatrix} . \tag{3.100}$$

The different coefficients of the eigenvectors indicate the strength of the associated modes. Therefore, we conclude that the amplitude of the zero diffraction order is $k^2 c_1^* / 2K^2$ larger than the amplitude of the first order. The factor two is due to the fact that the energy is distributed between the first and minus first order. Hence,

$$|T_1| = |T_{-1}| = \frac{\sin(\pi t)}{2\pi} [c_I - c_{III}] \left(\frac{\Lambda}{\lambda}\right)^2 |T_0| , \tag{3.101}$$

where $T_{-1,0,1}$ are the amplitudes of the minus first, zero, and first diffraction order. Although the derivation of Eq. (3.101) was somewhat intuitive, the resulting field amplitudes agree well with the rigorous calculations (see Fig. 3.12 as an example).

Using Eq. (3.101) the near field of the subwavelength structure can be easily calculated, as close to the structure as the penetration depth of the next higher order evanescent

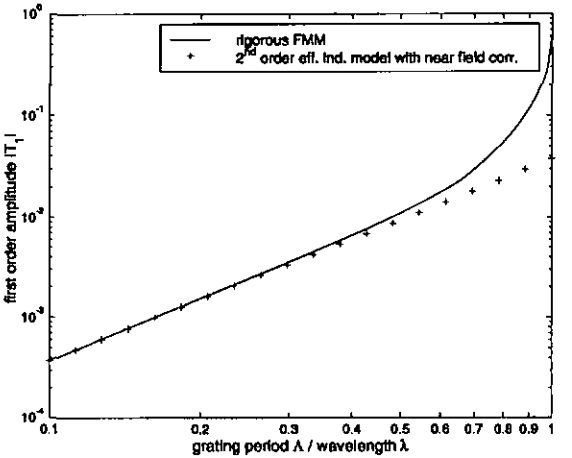


Figure 3.12: Amplitude of the first diffraction order of a subwavelength binary grating ($n_1 = 1.5, n_2 = 1, h = \lambda/2, t = 0.5$,TE-polarization) calculated by the rigorous Fourier Modal Method (FMM) (line), and the approximate effective index model with near field correction (crosses).

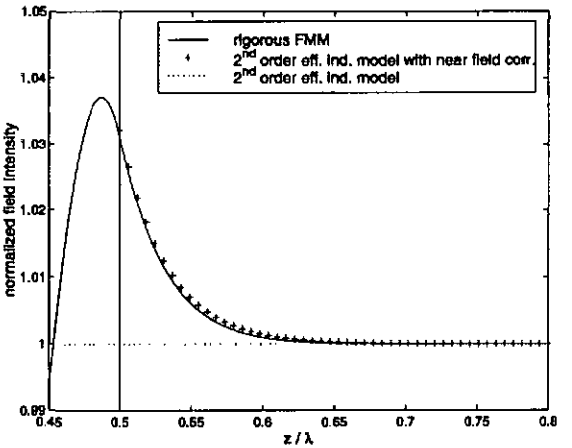


Figure 3.13: Normalized amplitude of the near field of a subwavelength binary grating ($n_1 = 1.5, n_2 = 1, h = 0.5/\lambda, \Lambda = 0.2\lambda, t = 0.5$, TE-polarization) along the propagation direction $\hat{z}$. The field has been calculated with different methods (see figure legend).

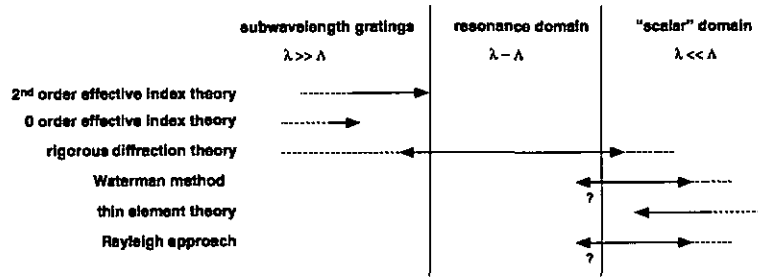


Figure 3.14: Overview of the validity of different grating theories (not complete). The arrows indicate the range commonly used by the corresponding theory.

wave, which is practically as close as a fraction of the wavelength. An example is shown in Fig. 3.13.

3.5 Discussion

In this chapter, we presented different theories to solve the grating interaction problem (see Fig. 3.14). For gratings having geometrical dimensions in the order of the optical wavelength, the solutions have to be based on Maxwell's equations. The problem of all *rigorous* theories is that the computational requirements increases for larger ratio between the grating period and the wavelength. In differential methods, as presented here, the number of waves or modes retained in to analysis is typically proportional to the ratio Λ/λ. In integral methods the number of sampling point used to numerically calculate the integral is *also* proportional to this ratio. The same is valid for finite element methods, increasing the geometric sizes will increase the number of mesh point used in the analysis. Increasing the ratio Λ/λ will *in any case* increase the need for computer memory and calculation time if rigorous solutions are needed. It is therefore necessary to simplify the models if possible, being aware of the validity of the approximations.

In particular, we have presented a second order effective index model which is valid as soon as the grating period is smaller than half of the wavelength. The method is more accurate for *thick gratings* than for thin ones (because of the neglected evanescent waves inside the grating structure). In the near field of the structure the first and minus first diffraction order can be expressed analytically.

If the grating period is much larger than the wavelength, several models exist, including the thin element approach, the Waterman or the Rayleigh methods. These methods work best if the structures are *thin*. The region of validity of the Waterman and the Rayleigh

method is still a current research topic [35].

Chapter 4

Spectral response and transfer function

4.1 System response

In Chapter 3 we considered the interaction of light with optical microstructures. In most of the applications these structures are parts of a larger optical system. For linear materials Maxwell's equations are linear. Therefore, we first study linear systems in more details. We then apply the theory to some important cases. The notion of linear system will be useful for the discussion of scanning near field optical microscope (SNOM).

One of the basic properties of linear systems (or even the definition) is the *superposition principle*: The sum of the response to two impulses of a linear systems is the same as the response to the sum of the two impulses. In a more general form, this is written as

$$U_2(x) = \int_{-\infty}^{\infty} t(x, x')\, U_0(x')\, \mathrm{d}x' \,, \tag{4.1}$$

where $t(x, x')$ is the impulse response, Fig. 4.1.

Equation (4.1) is often used in Fourier optics. Another approach is to consider the

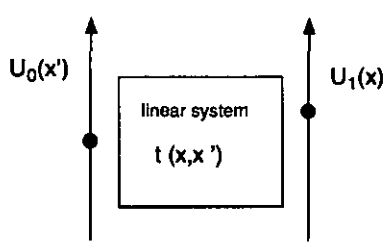


Figure 4.1: Impulse response

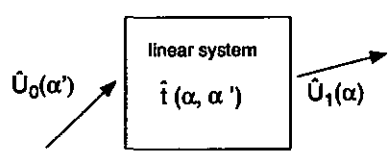


Figure 4.2: Spectral response

	space invariant	frequency invariant
spectral response	$\hat{t}(\alpha,\alpha') = \hat{t}(\alpha)\delta(\alpha-\alpha')$	$\hat{t}(\alpha,\alpha') = \hat{t}(\alpha-\alpha')$
frequency space	$\hat{U}_1(\alpha) = \hat{t}(\alpha)\hat{U}_0(\alpha)$	$\hat{U}_1(\alpha) = \hat{t}(\alpha) * \hat{U}_0(\alpha)$
impulse response	$t(x,x') = t(x-x')$	$t(x,x') = t(x)\delta(x-x')$
direct space	$U_2(x) = t(x) * U_1(x)$	$U_2(x) = t(x)U_1(x)$
	$\hat{t}(\alpha)$: transfer function	$t(x)$: transmission function
Examples	ideal imaging system scanning system planar interface free space propagation	thin element (parax.)

Table 4.1: Linear system: terminology and relations

response of one *spatial frequency*. The superposition principle is of course also applicable in this case: the sum of the response for two frequencies of a linear system is the same as the response for the sum of the two frequencies. Therefore, one may also write

$$\hat{U}_2(\alpha) = \int_{-\infty}^{\infty} \hat{t}(\alpha,\alpha')\,\hat{U}_0(\alpha')\,\mathrm{d}\alpha' \,, \tag{4.2}$$

where $\hat{t}(\alpha,\alpha')$ is the spectral response, Fig. 4.2. Equation (4.2) describes how a plane wave traveling at spatial frequency α' is transferred into a plane wave traveling at frequency α.

The spectral response function is somewhat similar to the scattering matrix approach (or S-matrix approach), see for example the work of NIETO-VESPERINAS [42, Chap.5]. However, in contrast to the S-matrix approach, we will include also *evanescent* waves.

If the impulse only depends on the difference of the position of the input and the output (and therefore not on the absolute position), the system is said to by *space invariant*. As an example, an ideal imaging system is space invariant. Such systems play an important role in classical optics. Similar to this, one can define *frequency invariant* systems. The spectral response of a frequency invariant system only depends on the difference of the input and output frequency. We will see that the thin element introduced in Sec.3.4.2 is a frequency invariant system.

The different terms used in this chapter are summarized in Tab. 4.1. Theirs definitions and properties are discussed in the following sections.

4.2 Space invariant systems

By definition, the impulse response of a space invariant system is

$$t(x,x') = t(x-x') \,. \tag{4.3}$$

The superposition integral, Eq. (4.1) is therefore the convolution,

$$U_2(x) = \int_{\infty}^{\infty} t(x - x')U_1(x')\mathrm{d}x' . \tag{4.4}$$

A convolution in the space domain corresponds to a multiplication in the frequency domain. The spectral response of a space invariant system is therefore described as (c.f. Table 4.1)

$$\hat{U}_2(\alpha) = \hat{t}(\alpha)\hat{U}_1(\alpha) , \tag{4.5}$$

where $\hat{t}(\alpha)$ is the Fourier transform of the impulse response $t(x)$. Equation (4.5) describes the the change of each frequency. $\hat{t}(\alpha)$ is therefore the *frequency transmission* of the system (in analogous to the *complex amplitude transmission* in frequency invariant systems). There are several important cases of space invariance.

- an ideal imaging system (Fig. 4.3 a) is completely described by a space invariant impulse response (the *point-spread function).* The associated (coherent) transfer function is called *amplitude transfer function, ATF.* The notion of space invariance allows the treatment of imaging system in a very elegant way (see for example GOODMAN [43])

- systems that are scanned (in $\hat{x}$-direction) and that are not changing properties during the scan are also space invariant, Fig. 4.3 b). Examples of such systems are scanning detectors and scanning light sources. The transfer function is the *detector acceptance function*, or the *radiation pattern* of the source.

- if a plane wave hits a *planar interface* the amplitude of the wave is changed, however not the (spatial) frequency.

- In free space propagation a plane wave changes the phase (propagating wave) or the amplitude (evanescent wave) , however not the frequency.

The imaging systems and the scanning systems are well documented in the open literature. Here, we focus the discussion on the planar interface and the free space propagation.

4.2.1 Flat interface

It is well known how an infinite plane wave is reflected and refracted on a flat interface. It is described by the Fresnel reflection formulas [1, Sec. 1.5], usually expressed for different incident angles. Here, we are considering spatial frequencies. Since the frequency response functions are defined from $-\infty < \alpha < \infty$, we are not only including propagating plane waves but also evanescent waves. Let n_1 and n_2 by the index of refraction of the first and

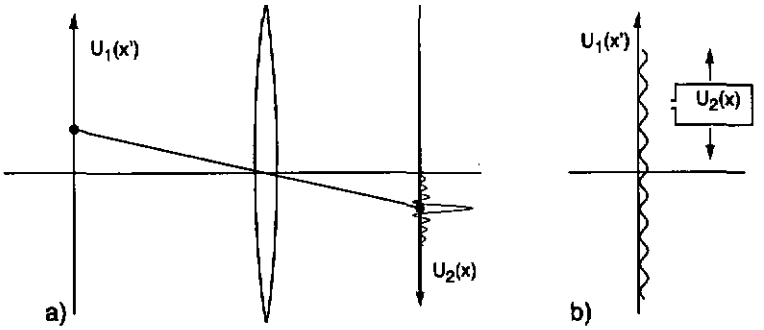


Figure 4.3: Space invariant systems: a) ideal imaging system b) ideal scanning detection system

the second region. The length of the wave vector is then given as $k_1 = n_1 k$ and $k_2 = n_2 k$. The incident, reflected and refracted waves are

$$E_0 = e^{i\alpha x + i\beta_1 z}, \qquad E_r = r\, e^{i\alpha x - i\beta_1 z}, \qquad E_t = t\, e^{i\alpha x + i\beta_2 z}; \tag{4.6}$$

where r and t are the amplitudes of the reflected and transmitted wave. The z-projections β_1 and β_2 are imposed by the constant length of the wave vectors k_1 and k_2, namely

$$\beta_1^2 = k_1^2 - \alpha_1^2, \qquad \text{and} \qquad \beta_2^2 = k_2^2 - \alpha_1^2. \tag{4.7}$$

Applying the ordinary boundary conditions to the flat interface problem yield for TE-polarization

$$r_{TE} = \frac{\beta_1 - \beta_2}{\beta_1 + \beta_2}, \qquad t_{TE} = \frac{2\beta_1}{\beta_1 + \beta_2}, \tag{4.8}$$

and for TM-polarization

$$r_{TM} = \frac{n_2^2\beta_1 - n_1^2\beta_2}{n_2^2\beta_1 + n_1^2\beta_2}, \qquad t_{TM} = \frac{2n_1 n_2 \beta_1}{n_2^2\beta_1 + n_1^2\beta_2}. \tag{4.9}$$

Now, we distinguish the following three cases (see Fig 4.4).

1. $\alpha < \min(k_1, k_2)$: this corresponds to the classical case of reflection and refraction. All waves are propagating plane waves.

2. $\min(k_1, k_2) < \alpha < \max(k_1, k_2)$: this corresponds to the situation where one wave is evanescent and the others propagating. If the second media is denser ($k_2 > k_1$), $t(\alpha)$ describes the *evanescent wave to propagating wave conversion capability* of a flat interface (case 2a). If the first media is denser ($k_1 > k_2$) , $t(\alpha)$ describes the *propagating wave to evanescent wave conversion capability*, corresponding to the total reflection (case 2b).

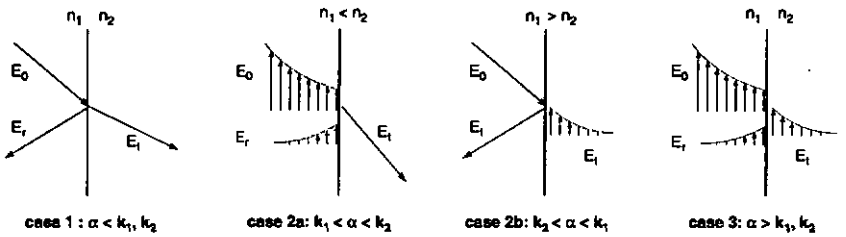


Figure 4.4: Refraction and reflection of propagating and evanescent plane waves at a planar interface.

3. $\alpha > \max(k_1, k_2)$: this corresponds to the situation where all the waves are evanescent. It is interesting to study the "reflection" and "refraction" of evanescent waves.

Case 1 and case 2b are well known and described in textbooks (see for example [1, Sec. 1.5] or [44, Sec. 6.2]). The other cases are less common. Figure 4.5 shows two examples of planar interface refraction, indicating the three different cases. The square of the amplitude of the transmitted wave versus the normalized spatial frequency (α/k) is reported for a) $n_1 < n_2$ and b) $n_1 > n_2$. An evanescent incident wave can be realized by either a medium of index before the incident structure or by a grating structure which generates evanescent waves of high spatial frequency.

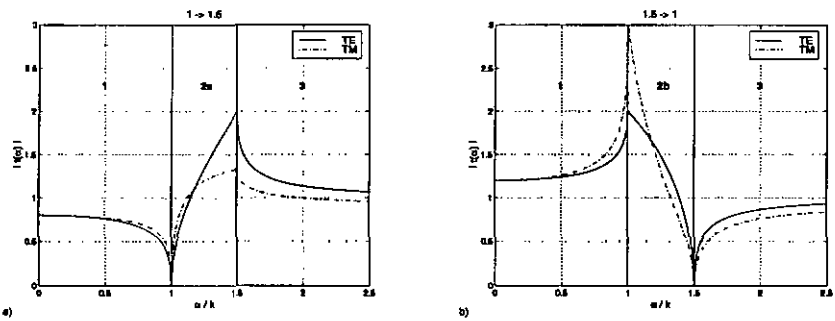


Figure 4.5: Spectral response of a planar interface for two different geometries: a) $n_1 < n_2$ and b) $n_1 > n_2$.

Reflection and refraction of evanescent waves If all the waves are evanescent, β_1 and β_2 can be replaced by $i\gamma_1$ and $i\gamma_2$, where γ_1, γ_2 are positive real numbers. The expression for the spectral responses remain basically the same as Eqs. (4.8) and (4.9).

The spectral responses are real functions, indicating that all the waves have the same phase. Note that for large values of α, the interface becomes transparent for the TE-polarized evanescent wave, however not for the TM-polarization

$$\lim_{\alpha\to\infty} t_{TE}(\alpha) = 1\,, \qquad \text{and} \qquad \lim_{\alpha\to\infty} t_{TM}(\alpha) = \frac{2n_1n_2}{n_2^2+n_1^2}\,. \tag{4.10}$$

Evanescent – propagating wave conversion Let us limit our consideration to the evanescent to propagating wave conversion in TE-polarization. The TM-polarization and the propagating to evanescent wave can be deduced in a similar way.

If we choose $k_1 < \alpha < k_2$, the propagation constant β_1 is complex. We therefore define

$$\beta_1 = \mathrm{i}\gamma_1\,, \tag{4.11}$$

the spectral responses become now also complex

$$r_{TE}(\alpha) = \frac{\mathrm{i}\gamma_1 - \beta_2}{\mathrm{i}\gamma_1 + \beta_2}\,, \qquad \text{and} \qquad t_{TE}(\alpha) = \frac{2i\gamma_1}{\mathrm{i}\gamma_1 + \beta_2}\,. \tag{4.12}$$

It is not difficult to show that the real part and the imaginary part of r and t form a *circle*

$$\mathrm{Re}^2\{r\} + \mathrm{Im}^2\{r\} = 1\,, \qquad \text{and} \qquad \mathrm{Re}^2\{t-1\} + \mathrm{Im}^2\{t\} = 1\,. \tag{4.13}$$

This indicates that *there is always a phase change associated with evanescent wave to propagating wave conversion.* The same is valid for the inverse case, that is, *there is always a phase change for propagating wave to evanescent wave conversion.* The second case corresponds to the total reflection. The phase change associated with it can easily by measured [1, Sec. 1.5.4].

Let us summarize this section by illustrating the reflection and refraction on a flat interface by an unusual representation. In Fig. 4.6 the imaginary and the real part of $t(\alpha)$ for different values of α, is reported (TE-polarization, $k_1 < k_2$). The different cases can easily be identified. First, the transmission function is purely real (corresponding to the classical refraction case). Then, the real part and imaginary part form a circle (indicating the phase change associated with the evanescent to propagating wave conversion). And finally, the transfer function is real again, and approaches unity.

4.2.2 Free space propagation

Usually, free space propagation is introduced through the Rayleigh-Sommerfeld or the Kirchhoff diffraction integrals (see for example [45, Chap.4]). Here, we will use the concept of transfer functions.

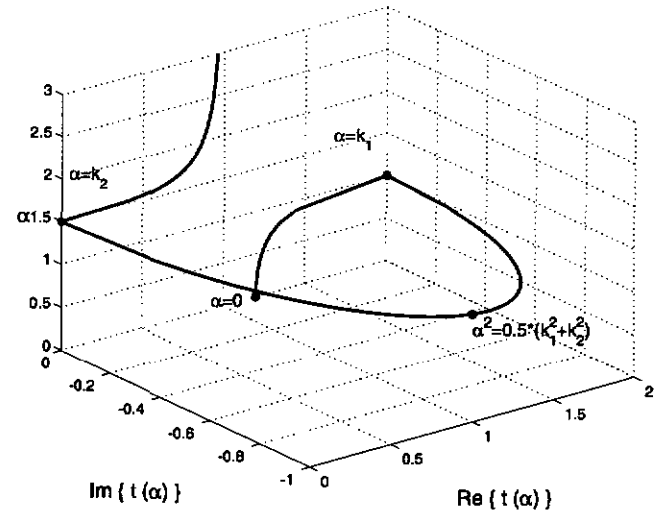


Figure 4.6: Planar interface: amplitude of the transmitted wave illustrated in the complex plane, TE-polarization. $n_1 = 1$ and $n_2 = 1.5$

It is well know how plane waves behave in free space (see. Sec. 2.3). Propagating plane waves change the phase, evanescent waves change the amplitude (see Fig. 4.7). In terms of transfer function this can be written as

$$\hat{t}(\alpha) = \begin{cases} \exp[i\beta z] & \text{with} \quad \beta = \sqrt{\epsilon_0 k_0^2 - \alpha^2} & \text{for} \quad \epsilon_0 k_0^2 > \alpha^2 \\ \exp[\pm\gamma z] & \text{with} \quad \gamma = \sqrt{\alpha^2 - \epsilon_0 k_0^2} & \text{for} \quad \epsilon_0 k_0^2 < \alpha^2 \end{cases} \tag{4.14}$$

The consequence is that if we want to detect a field which has variations smaller than the wavelength, i.e. spatial frequencies larger than nk, the detection has to be very close to the position of generation, because of the exponential decay of the waves. Based on this

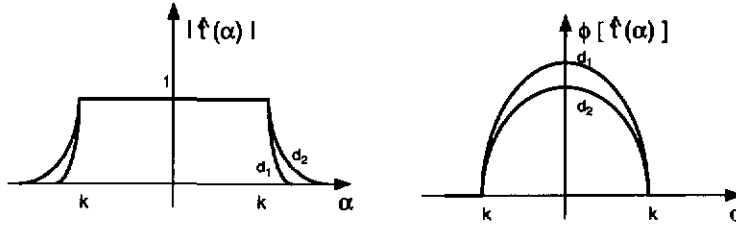


Figure 4.7: Free space transfer function

principle it is possible to overcome the classical resolution limit (see Sec. 4.4 and [46, 47]).

A second consequence of Eq. (4.14) is a *phase delay* between different spatial frequencies during propagation. The higher frequencies are more delayed than the lower frequencies. This phase delay is responsible that a field distribution will spread out during propagation[1]. If the spectrum of the given field distribution is limited to the central frequencies, the spread is less pronounced than if the whole spectrum is used, i.e. the diffraction pattern of a large aperture is smaller than the diffraction pattern of a small aperture.

Equation (4.14) describes how light propagation behaves in the spatial frequency space. On the other hand, it is possible to calculate the field distribution in another plane without passing through the frequency space. The field in the output plane is the field in the input plane convoluted with the *free space impulse response*. This impulse response is the Fourier transform of the free space transfer function which is

$$t(x;z) = \frac{1}{2\pi}\int_{\infty}^{\infty} \exp[\mathrm{i}(\alpha x + \beta z)]\mathrm{d}\alpha = \frac{-kz}{2ir} H_1^{(1)}(kr)\,, \tag{4.15}$$

where $H_1^{(1)}(t)$ is the first-order Hankel function of the first kind and $r = (x^2 + z^2)^{1/2}$. Figure 4.8 shows the intensity of the free space transfer function versus x and z. The derivation of Eq. (4.15) can be found in [45, Sec. 5.2]. For small arguments the first-order Hankel function behaves like r^{-1} (Fig 4.9). For larger arguments $H_1^{(1)}(t)$ is proportional to $r^{-1/2}$ (see [48, Chap. 9]). These approximations allows to write for large distances (here large means larger than the wavelength)

$$t(x;z) = \frac{e^{-\mathrm{i}\pi/4}}{\sqrt{\lambda r}}\,\frac{z}{r}\,e^{\mathrm{i}kr}\,, \qquad \text{for} \qquad kr > 1\,. \tag{4.16}$$

Equation (4.16) is nothing else than a *cylindrical* wave. Further approximations for large propagation distances (mainly the replacement of kr in the exponential function of Eq. (4.16) by a quadratic term) leads to the *Fresnel* impulse response [44, Sec. 4.1] or [43, Chapt. 4], representing a two dimensional parabolic wave.

$$t(x;z) = \frac{e^{-\mathrm{i}\pi/4}}{\sqrt{\lambda z}}\,e^{\mathrm{i}kz}\,e^{-\mathrm{i}\pi x^2/\lambda z}\,. \tag{4.17}$$

For very large distances, it is not allowed to further simplify the impulse response of Eq. (4.17) directly. The system is not space invariant anymore, $t(x,x') \neq t(x-x')$. The superposition integral of Eq. (4.1) becomes a *Fourier integral*. The diffraction pattern does not change the shape but only the scale while propagation. This situation corresponds to the *Fraunhofer* diffraction.

[1] The spread of the field can be compared to the spread of a pulse in a dispersive medium [44, p. 188].

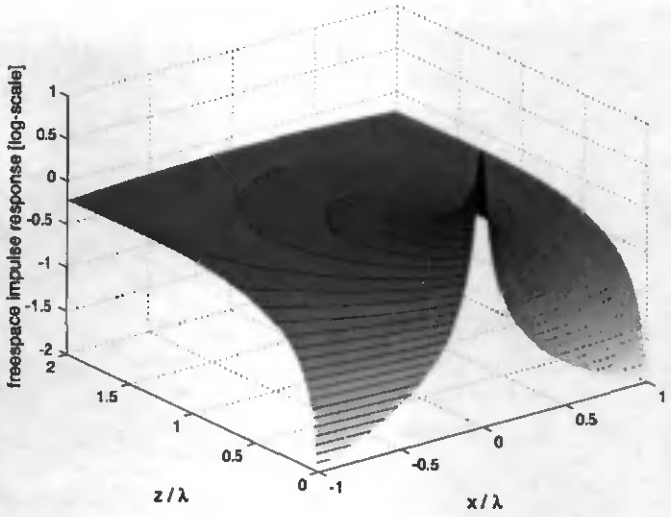


Figure 4.8: Two dimensional free space impulse response

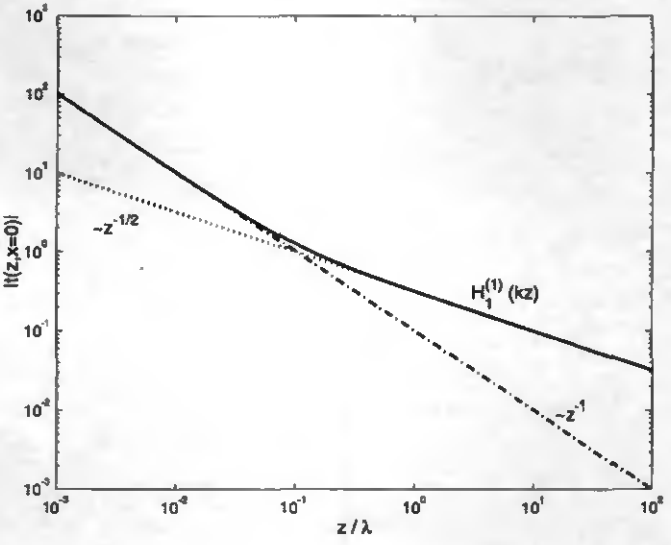


Figure 4.9: Behavior of the two dimensional free space impulse response along the propagation axes

4.3 Frequency invariant response

The spectral response of a frequency invariant system depends only on the difference of the output frequencies at the input (c.f. Table 4.1), therefore

$$\hat{t}(\alpha, \alpha') = \hat{t}(\alpha - \alpha') \,. \tag{4.18}$$

The superposition integral, Eq. (4.2) becomes a convolution in α. In the direct space this convolution becomes the multiplication

$$U_1(x) = t(x) U_0(x) \,, \tag{4.19}$$

where $t(x)$ is the Fourier transformation of the spectral response. Equation (4.19) describes how the complex amplitude of the field is changed by the system at each point. $t(x)$ is therefore the *complex amplitude transmission function* (or simply the *transmission function*) of the system. In the thin element approach (Sec.3.4.2) the elements are described by a transmission function (like in Eq. (4.19)). The system is therefore frequency invariant in this approach.

4.3.1 Thin element approach

In Sec. 3.4.2 we have already encountered transmission functions describing the interaction of light with gratings in the thin element approach. This interaction is simply described by the complex transmission function

$$U_2(x) = \exp[\mathrm{i}\phi(x)]\, U_1(x) \,, \quad \text{with} \quad \phi(x) = \exp[-\mathrm{i}k_0(n_1 - n_2) f(x)] \,, \tag{4.20}$$

where $f(x)$ is the surface profile of the element.

We have also seen that the thin element approach is only an approximation of the rigorous grating theory. In Fig. 4.10 we illustrate the spectral response of a binary grating structure. For frequencies smaller than $\simeq 0.25k$ the elements behaves as a frequency invariant systems. The range of validity of the frequency invariance depends of course on the structure. It is interesting to observe that the amplitude of the individual diffraction orders becomes maximal at the point of total reflection, similar to the case of a flat interface (see Fig. 4.5 b).

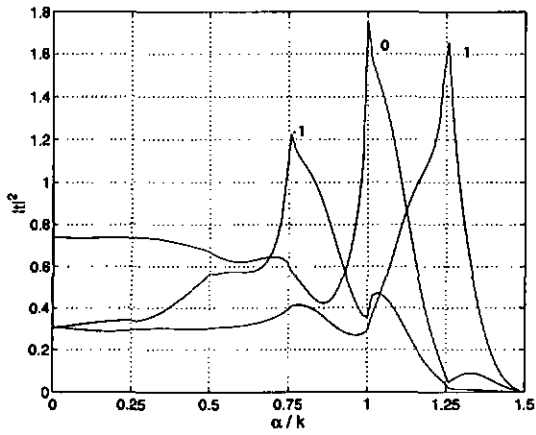


Figure 4.10: Rigorously calculated spectral impulse response of a 4λ-period binary grating. Only the $1, 0, -1$ diffraction orders are shown.

4.4 SNOM represented by transfer functions (A.I)

Up to now, we used the spectral response as a representation of one element. We now are interest how to assemble the different components to a larger system. The most important observation is that in frequency invariant systems, the different transmission functions are simply multiplied, and in space invariant system the transfer functions are multiplied. An example is shown in Fig. 4.11 where this method has been applied to the scanning near field optical microscope, SNOM.

In [49] (see appendix A.I), we consider the near field detection system as an optical instrument characterized by a transfer function. For this purpose, the measurement system is separated into different parts as shown in Fig. 4.11. Note that in [49] the spatial frequencies α are written as k_x, whereas the propagation β constants are called k_z. The object is illuminated by an incident wave and generates a field distribution in the output plane. In general, the interaction of light with sub-wavelength structures has to be computed rigorously with methods described in Sec.3.3. We define a spectral response function H_{obj} as being the action of the object on the incident wave for a certain *spatial* frequency. The resulting field is detected by the scanning probe at a constant distance from the sample. The radiating part of the field propagates in free space, whereas the evanescent part is attenuated. This effect is described by the free space transfer function P, corresponding to Eq. (4.14). The detection is made by scanning a tip over the sample. If the property of the detection system does not change during the scan the linear system

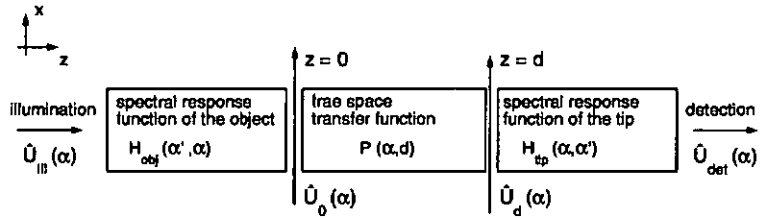


Figure 4.11: The scanning near field optical microscope is considered as an optical instrument characterized by spectral response and transfer functions.

is space invariant. Therefore, the detected signal can be expressed as a convolution

$$U_{\rm det}(x) = \int_{-\infty}^{\infty} t(x-x')U_d(x')\mathrm{d}x' = t_{\rm tip} * U_d\,, \tag{4.21}$$

where $t(x-x')$ is the impulse response of the detector. In frequency space Eq. (4.21) becomes a multiplication

$$\hat{U}_{\rm det}(\alpha) = T(\alpha)\hat{U}_d(\alpha)\,, \tag{4.22}$$

where $T(\alpha)$ is the transfer function of the detector. For linear systems the transfer function is expressed as a linear superposition integral

$$T(\alpha) = \int_{-\infty}^{\infty} H_{\rm tip}(\alpha,\alpha')\ D(\alpha')\ \mathrm{d}\alpha'\,, \tag{4.23}$$

where $H_{\rm tip}(\alpha,\alpha')$ is the spectral response of the tip and $D(\alpha')$ the detector acceptance function. The acceptance function of the tip depends on the chosen geometry. For a classical single-mode fiber tip SNOM (Fig. 4.12 a) only one frequency is guided, $D(\alpha') = \delta(\alpha')$ and the transfer function becomes particularly simple, namely

$$T(\alpha) = H_{\rm tip}(\alpha,0)\,. \tag{4.24}$$

In the case of a cantilever tip based SNOM, the acceptance function of the detector is equal to the *coherent amplitude transfer function* (ATF) [43] of the detection system. The different frequencies are now integrated over the numerical aperture of the system. For an ideal imaging system the coherent amplitude transfer function is a binary function, Eq. (4.23) rewrites as

$$T(\alpha) = \int_{-k\,\mathrm{NA}}^{k\,\mathrm{NA}} H_{\rm tip}(\alpha,\alpha')\ \mathrm{d}\alpha'\,, \tag{4.25}$$

where NA is the numerical aperture of the imaging system. The aim of the tip-detector is to provide sub-wavelength resolution. For this purpose, the geometrical dimensions of the

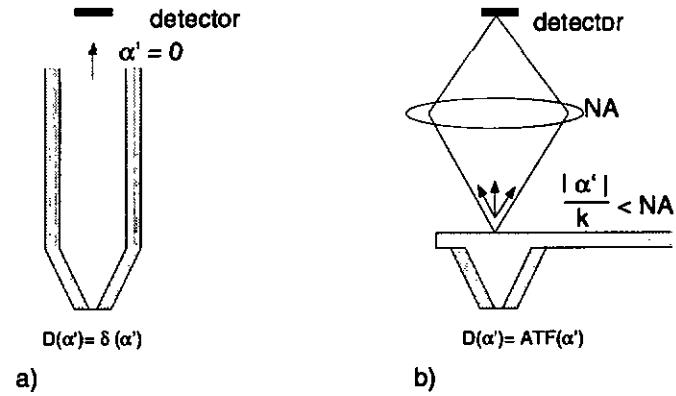


Figure 4.12: Acceptance function $D(\alpha')$ for different geometries. a) fiber tip detection, b) SNOM based on a cantilever tip .

tip have to be in the sub-wavelength range. Thus, the spectral response $H_{\text{tip}}(\alpha, \alpha')$ also has to be calculated with a rigorous model. In Fig. 4.13 the spectral response $H_{\text{tip}}(\alpha, 0)$ of a coated and uncoated dielectric fiber tip is shown. The calculations have been made by the same algorithm as for the grating structure (FMM), by considering periodically arranged tips (see also comments in the introduction of Chap. 5). The investigations were limited to two-dimensional geometries. The uncoated tips have higher coupling efficiencies for low spatial frequencies than the coated tips. On the other hand, the electromagnetic field is less confined, i.e. the angular spectrum is narrower. The capability to convert evanescent waves to propagating waves is higher for the coated tip than for the uncoated tip. The signal to noise ratio, and therefore the resolution, is better for the coated tip, although the overall throughput of the uncoated tip is higher.

In [49] we make the following conclusions: In the representation of transfer functions, the limitation of the scanning near field microscope can be clearly discussed, although back reflections from the tip are not taken into account. Sub-wavelength signals have to overcome the exponential attenuation due to propagation of the evanescent waves and the low-pass filter characteristic of the probe tip. The spectral response of the tip expresses the capability of conversing evanescent to propagating waves. It determines therefore the resolution power of the SNOM.

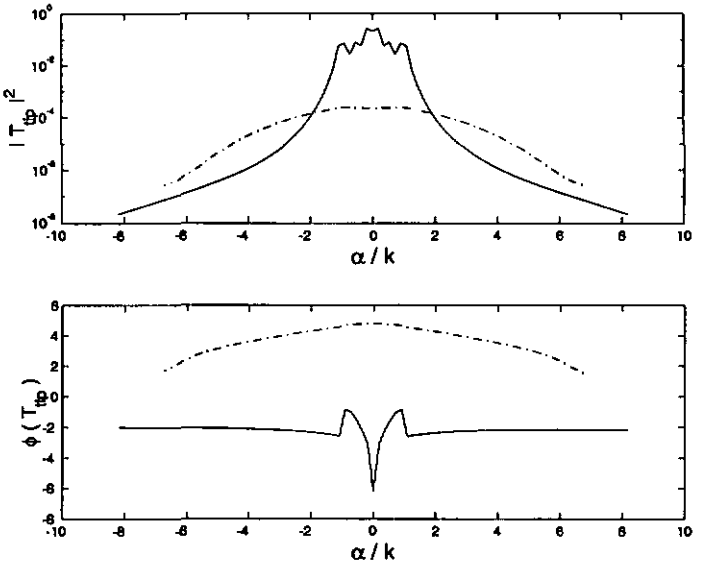


Figure 4.13: Calculated intensity and phase of the complex spectral response function of a one-dimensional dielectric probe tip : uncoated (solid line) and coated (dashed line). The coated tip has a clear slit aperture of $\lambda/10$.

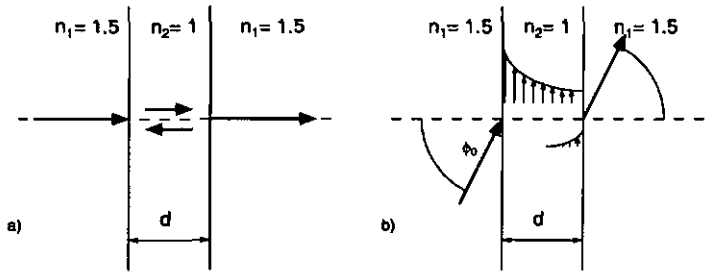


Figure 4.14: 3 homogeneous regions: "sandwich" structure. a) for a perpendicular plane wave a a backward and forward plane wave is generated inside the structure. For a large incident angle the field inside the structure is evanescent.

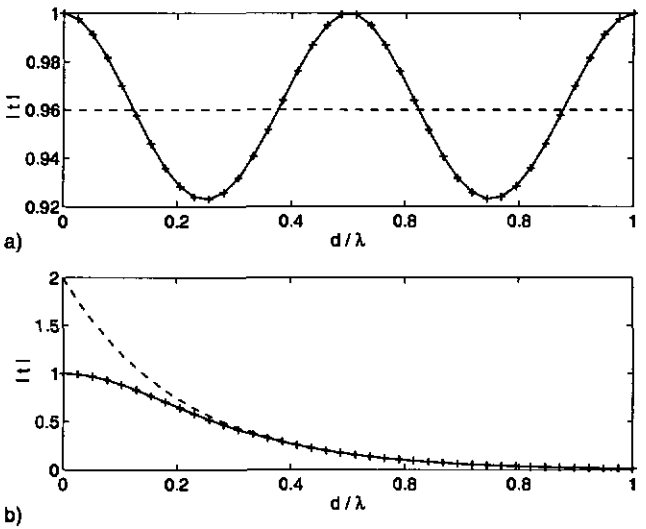


Figure 4.15: Transmitted intensity of the sandwich structure versus the thickness d/λ. a) corresponds to Fig. 4.14 a) and b) to Fig. 4.14 b), respectively. The dashed lines were calculated by the transfer function method. The solid lines have been calculated by a self consistent approach. The points '+' correspond to the improved transfer function method (Sec. 4.5.1).

4.5 Limitation of the transfer function concept

In some cases, like the flat interface, not only forward propagating waves are generated but also backward waves. Depending on the reflected waves, the overall behavior of the system may change. A trivial example of this is the sandwich configuration shown in Fig. 4.14 . In the first case (Fig. 4.14 a) a perpendicular incident plane wave is refracted (and reflected) on a first flat interface after some free-space propagation the field is again refracted and reflected by a second interface. The reflected wave travels again backward and forward, and so on. In Fig. 4.15 a) the intensity of the transmitted wave for different thickness is shown. The dashed line is the calculation based on transfer functions, and the solid line corresponds to the self consistent calculation (made by matching all the waves at once). The self consistent model shows the typical periodic behavior as a Fabry-Pérot resonator. Depending on the thickness of the structure, constructive and destructive interferences occur. In this situation, multiplication of the three transmission functions (first interface, free space, second interface) gives wrong values. If the geometry (resp.

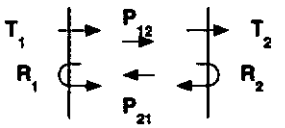


Figure 4.16: Improved transfer function concept: also the reflections are taken into account

the incident angle) is chosen in a way that inside the sandwich structure there are two *evanescent* waves (Fig. 4.14 b), the method based on the transfer function is accurate *as soon as the influence of the back reflected evanescent wave disappears* (Fig. 4.15 b).

4.5.1 Improvement of the transfer model in the case of back propagating waves

The problems described above can be removed by a refinement of the model. The idea is to describe also the *reflections* with a transfer function and then iteratively include them in the system. In Fig. 4.16 an example of such as system is shown. T_1 and T_2 are the transfer functions of the transmitted waves, R_1 and R_2 the transfer functions of the reflected waves at the first and the second interface, respectively. P_{12} and P_{21} describe the system in between, depending on the direction of propagation, usually $P_{12} = P_{21} = P$.

If we do not include the back-reflections, the transfer function of the overall system is

$$H_0 = T_2\, P\, T_1\,. \tag{4.26}$$

The index 0 stands for the 0th iteration step. After going once forth and back, the transfer function is

$$H_1 = H_0 + T_2\, P\, R_1\, P\, R_2\, P\, T_1 = T_2\,(1 + P\, R_1\, P\, R_2)\, P\, T_1\,, \tag{4.27}$$

and after N iterations

$$H_N = T_2 \left(\sum_{n=0}^{N} [P\, R_1\, P\, R_2]^n\right) P\, T_1\,. \tag{4.28}$$

We are considering only *passive* systems, i.e. there is no energy creation inside the structure. The sum of the geometric progression in Eq. (4.28) converges, thus the transfer function of the overall system including the back-reflections becomes

$$H = \lim_{N\to\infty} H_N = T_2 \frac{1}{1-A}\, P\, T_1\,, \qquad \text{with} \qquad A = P\, R_1\, P\, R_2\,. \tag{4.29}$$

Using this approach for the sandwich example in Fig. 4.14 gives the expected exact results (see the crosses in Fig. 4.14).

We believe that this improved method is not limited to the case of only *one* backward and forward traveling wave, but can be applied in more general cases, such as stacked grating structures, or more accurate models for the scanning near field optical microscope (i.e. to include also the back-reflections from the tip). Note that in a more general case, the transfer functions are described by matrices, and not scalars as in the example of Sec. 4.5. The sum of the geometric progression cannot be expressed analytically, and therefore also the limit $N \to \infty$ cannot be represented by Eq. (4.29).

Chapter 5

Near field and far field of microstructures

In chapter 3 we learned different techniques to compute the interaction of light with periodic microstructures. These methods allow to calculate the near field and far field of different grating structures. The term *near field* and *far field* have different signification depending on the field of utilizations. In classical optics, and especially in Fourier optics [43], the far field is considered as the region where the different spatial frequencies do not interfere anymore, i.e. the diffraction pattern do not change the shape during propagation but only the scale (Fraunhofer diffraction). The near field is then defined as "everything closer than the far field" [43, p. 67] (for example the Fresnel region). In near field optics, the near field is then defined as region where the field is mainly influenced by evanescent waves, i.e. the near field distances are smaller than the wavelength. And the far field region is "everything further away than the near field" [1]. Here, we are using the term *near field* in a less strict way. It is the region close to the element, without specifying *how close* it really is.

We will now apply the interaction theory of chapter 3 to different optical microstructures. Although the theories were developed for *periodic* structure, we will also apply them to *single* elements. For this purpose, we introduce a large grating period to reduce the influence of neighbor elements. Further details on this approach are presented in [50] (see appendix A.11).

[1] Sometimes the term *intermediate zone* is used in near field optics, especially in total internal reflection mode (STOM). It describes the penetration depth of the zero (evanescent) diffraction order.

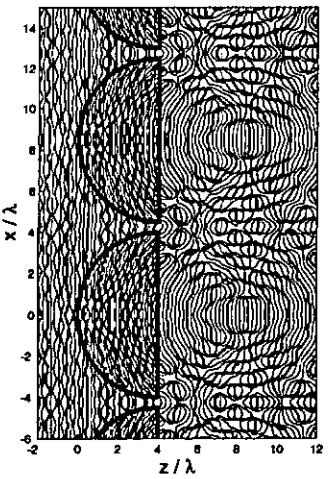


Figure 5.1: Calculated near field phase contour plot of a one-dimensional microlens array. The elements have a diameter of 8λ, a height of 4λ and a refractive index of $n = 1.5$. The array is illuminated by a TE-polarized plane wave with perpendicular incidence.

5.1 Near field and far field of microlenses (A.II)

In this section, we discuss the behavior of small cylindrical microlenses, arranged in an one-dimensional array or as single element. For this purpose, we apply one of the standard rigorous diffraction theory (FMM) introduced in Chapter 3. Figure 5.1 shows an example for the near field an array of small microlenses.

We have investigated the coupling effect between the elements. Furthermore, we have computed the complex transmission function by rigorous diffraction theory and compared them with classical theories (combined raytracing, thin element approach). Finally, we have discussed the focal properties of microlenses in the rigorous regime. The detailed work can be found in [50] (see appendix A.II).

It turned out that it is possible to simulate single element diffraction with the aid of a rigorous grating diffraction method by carefully choosing the grating period. First, the grating period should not be a multiple of the wavelength, and second, the period should be larger than twice the diameter of the element. Furthermore, we have presented a comparison between three different methods to calculate the interaction of light with small sized lenses. For thin elements, all three methods predict well their behavior, even for very small lenses. This is mainly due to the fact that thin lenses are dominated by the edge diffraction which influences the focusing behavior of the lens. We have shown

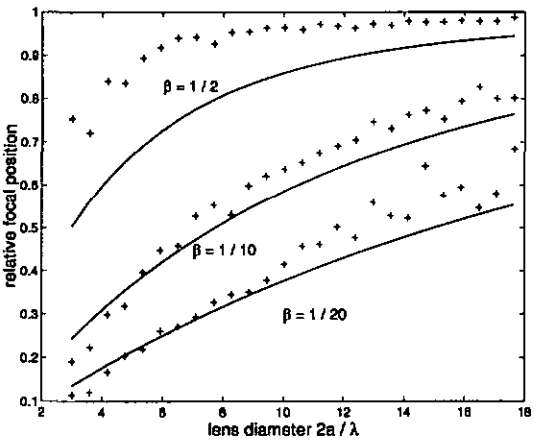


Figure 5.2: Illustration of the focal shift. The focal position relative to the paraxial focal position versus the diameter of the lens is plotted for different aspect ratios $\beta = h / 2a$. The lines correspond to the focal positions calculated by the thin element approach and the crosses to the rigorously calculated focus positions.

that there is a good agreement of the classically predicted focal shift compared to rigorous computations.

For thick lenses the combined raytracing is more precise than the thin element approach. For such elements, the focal shift turned out to be less pronounced than predicted by the thin element approach (see Fig. 5.2). In any case, it has been illustrated that also very small lenses exhibit some focusing behavior. Such small sized microlenses may be used in optical lithography. The idea is to replace mask and projection lens by one single element (smart mask, see Sec. 6.2).

5.2 Near field and far field of slit apertures

In A.II we have shown that it is possible to use a rigorous grating diffraction to model single elements, if the grating period is carefully chosen. Here, we will use this method to calculate the diffraction pattern of single slit apertures. In the thin element approach, the transmission of a slit is just a binary amplitude function. The diffraction pattern is proportional to a $sinc^2$-function. This result is independent of the ratio between the slit diameter and wavelength. Figure 5.3 shows the rigorously calculated diffraction pattern (logarithmic intensity scale) for two different slit apertures. The calculations have been

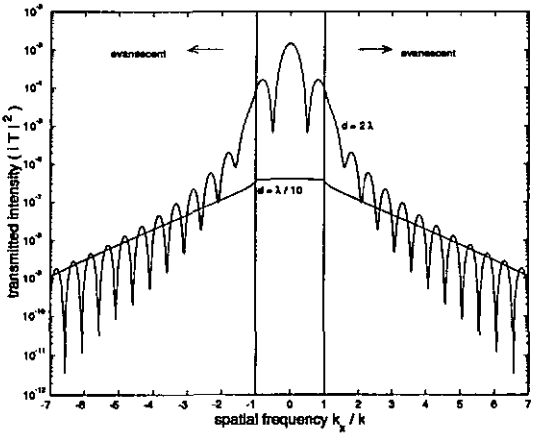


Figure 5.3: Rigorously calculated diffraction spectrum of a slit of aperture d. The slit is illuminated by a perpendicularly incident plane wave with TE-polarization.

made for a perpendicularly incident plane wave with TE-polarization.

The larger slit has a diameter of $d = 2\lambda$. The diffraction pattern is $sinc^2$-like with a first minimum at $k_x/k = 1/2$. The smaller slit has an aperture of $d = \lambda/10$. The diffraction pattern is completely different from the large aperture case. The difference to the $sinc^2$-function has the following reasons: In thin element approach the field after the aperture jumps sharply from 0 to 1. Such amplitude jumps in fractions of the wavelength are not physical. For the rigorous calculation the opaque region consists of a thin layer ($h = \lambda/10$) of a *finite conducting* metal. The field change smoothly its amplitude.

It is interesting to see that the two different types of plane waves behave completely different. The amplitude of the propagating waves are almost constant all over the region of propagation, giving raise to a spherical like wave. The amplitude of the evanescent waves are exponentially decreasing as a function of the spatial frequency.

The behavior of the two different types of plane waves has some consequences in near field optics. The aim of near field optics is to measure high spatial frequencies, hence the evanescent part of the field. Figure 5.3 shows how the different field components are transmitted through a slit. In terms of spectral response functions this corresponds to $t(0, x')$; it shows how an evanescent wave is *converted* into a propagating wave. Although the total throughput of the large slit is much higher, the smaller slit is more interesting. As shown in A.1 the signal noise ratio of the evanescent to propagating waves conversion is much better for the small slit. These considerations are fundamental and therefore

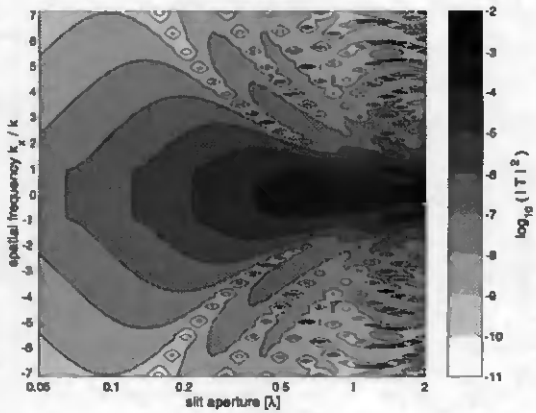


Figure 5.4: Rigorously calculated diffraction spectrum versus the slit aperture d.

indispensable for near field optics.

In Fig. 5.4 we present the diffraction spectrum as a function of the slit aperture. The transition between from $sinc^2$-like spectrum to the flat spectrum occurs for slit diameters smaller than about half the wavelength.

5.3 Phase dislocations and saddle points

5.3.1 Phase dislocations in homogeneous regions

An electromagnetic field can be described by an amplitude and a phase (and a polarization state). Interaction of light with micro-optical structures will changes the amplitude and the phase. Measuring the amplitude and the phase gives some information about the microstructures. Unfortunately, only the time averaged intensity of the waves can be measured. For phase measurements the interference of an object wave with the reference wave is needed. That is probably the reason why historically many of investigations (including discussions on resolution, imaging properties, etc) are based on the intensity distribution. In this section, we are discussing some basic properties of phase distributions. Our main interest are *phase dislocations*.

Phase dislocations or phase singularities are *isolated* points where the time averaged intensity of the field is zero. At this point, the phase is not determined. Mathematically this means that the real part and the imaginary part of the complex field amplitude are

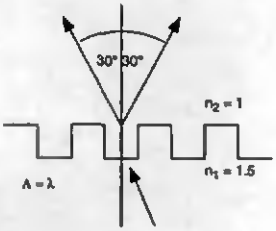


Figure 5.5: Configuration used to illustrate phase singularities.

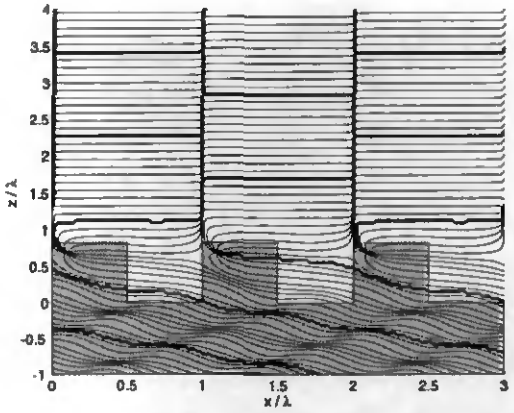


Figure 5.6: Contour plot of the phase of the optical near field of a binary structure.

zero at these points, viz.

$$\mathrm{Re}\{U(\mathbf{r})\} = \mathrm{Im}\{U(\mathbf{r})\} \equiv 0\,. \tag{5.1}$$

Phase dislocations can be observed in the near field and the far field of optical microstructures, in speckles patterns etc. As an example, Fig. 5.6 shows a phase contour plot of a binary grating structure corresponding to the configuration of Fig. 5.5. The phase singularities are located at about $z = 0.7\lambda$ and $x = 0.1\lambda + m\lambda$. To our knowledge, the first work on phase dislocations in optical fields was presented by NYE [51] at the beginning of the 70's. More recently the interest in this field has increased [52, 53, 54, 55]. Here, we would like to make some basic considerations based on the plane wave expansion of (periodic) optical fields.

Phase singularities can easily be realized. The interference of two plane waves of equal amplitudes generates fringes , i.e. lines on which the time averaged intensity is equal to zero. An *isolated* phase dislocations can then be produced by adding a third plane wave (Fig. 5.7).

In a more general case, a necessary condition that a phase singularities generated by the interference of N-infinite plane wave is that

$$\sum_{m \neq n} |a_m| \geq |a_n| \quad \forall \mathrm{n}\,, \tag{5.2}$$

where a_m are the (complex) amplitudes of the propagating plane waves of the electromag-

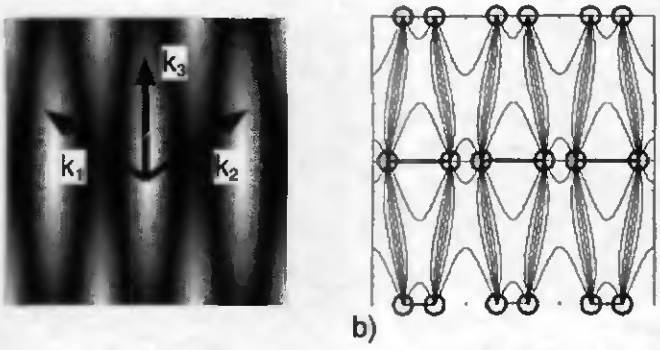


Figure 5.7: The interference of 3 plane waves shows phase dislocations: a) intensity, b) phase. The phase dislocations are indicated by circles.

netic field

$$U(\mathbf{r}) = \sum_m a_m \exp[i\mathbf{k}_m\mathbf{r}]\,. \tag{5.3}$$

The proof of Eq. 5.2 is straightforward. Remember that a phase dislocation occurs if the complex amplitude is zero, hence

$$\sum_m a_m \exp[i\mathbf{k}_m\mathbf{r}] \equiv 0\,. \tag{5.4}$$

Subtracting a_n on the left and right side of Eq. (5.4) yields

$$\sum_{m\neq n} a_m \exp[i\mathbf{k}_m\mathbf{r}] = -a_n \exp[i\mathbf{k}_n\mathbf{r}]\,, \tag{5.5}$$

taking the absolute value,

$$\left|\sum_{m\neq n} a_m \exp[i\mathbf{k}_m\mathbf{r}]\right| = |a_n|\,, \tag{5.6}$$

and finally applying the triangular inequality to the left side of Eq. (5.6) gives Eq. (5.2).

Let us look closer at the case of the interference of three plane waves. We choose the wave vectors $\mathbf{k}_m$ such that they form an equilateral triangle. The field is

$$U(\mathbf{r}) = a_1 \exp[i\mathbf{k}_1\mathbf{r}] + a_2 \exp[i\mathbf{k}_2\mathbf{r}] + a_3 \exp[i\mathbf{k}_3\mathbf{r}]\,. \tag{5.7}$$

The conditions to have a phase singularity at $\mathbf{r} = 0$ implies that the sum of the complex amplitudes becomes zero, viz.

$$a_1 + a_2 + a_3 \equiv 0\,. \tag{5.8}$$

This equation has of course different solutions depending on the amplitudes and the phases of the individual waves. Two possible solutions are shown in Fig. 5.8 a) and c).

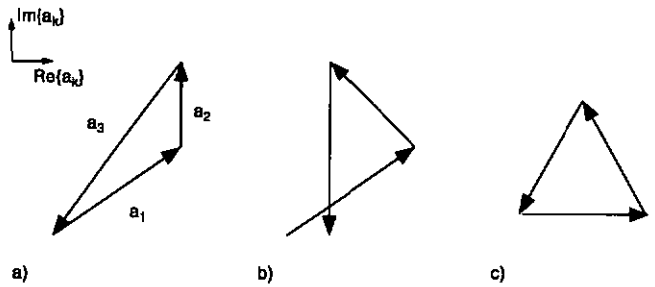


Figure 5.8: Sum of three complex values in the complex plane. A phase singularity exists at $\mathbf{r} = (0, 0)$ if the sum is equal to zero, as in case a) and c).

If Eq. (5.8) is not satisfied, but only Eq. (5.2), the phase dislocation is shifted away from $\mathbf{r} = 0$. Therefore, the precise localization measurement of phase singularities gives some sensitive information about the complex amplitudes of the individual plane waves. If Eq. (5.3) is the Rayleigh expansion of a field generated by a grating structure, the measurement of the positions of phase dislocations can give some information about the grating structure (see Sec. 7.3.2).

It is interesting to examine a phase dislocation in more details. For this purpose we choose the $\mathbf{k}_m$ such that they form an equilateral triangle (and $\mathbf{k}_1 = (0, 1)$). Figure 5.9 a) shows the time averaged intensity of the sum of the three waves. The phase contour plot (Fig. 5.9 b) indicates well the position of the phase dislocation. At this point, the time averaged Poynting vectors is zero. The energy flows around the phase singularity, (Fig. 5.9 c). The *curl* of the time averaged Poynting vectors is different from zero, as illustrated in Fig. 5.9 d), see also comments at the end of Sec. 2.2.1 . Figure 5.9 b) to d) show a "spectacular" aspect of optical light fields: Similar to air or water turbulencies, light fields can exhibits whirls and vortices.

Note, that the phase *always* jumps by π while passing through a phase dislocation, independent from which direction the scan is made. The phase jump itself is very sharp, theoretically with an infinite slope. This is not astonishing, because in terms of the real amplitude, this means that the sign changes.

5.3.2 Saddle points

In the previous section, we discussed the behavior of an optical field at points with time averaged intensities equal to zero. We have seen that also the time averaged Poynting

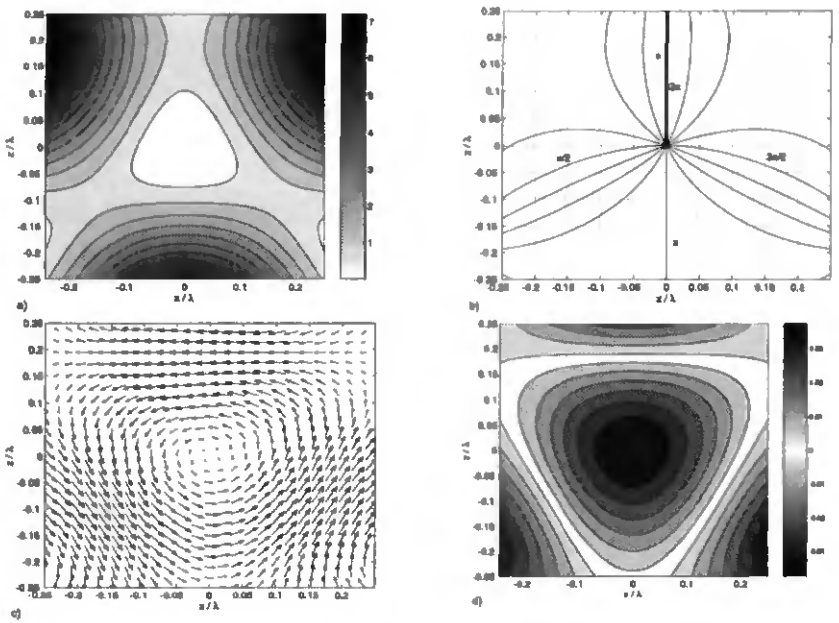


Figure 5.9: Phase singularity generated by three plane waves. a) the time averaged intensity b) the phase contour c) the time averaged Poynting vector and d) the curl of the time averaged Poynting vector is shown.

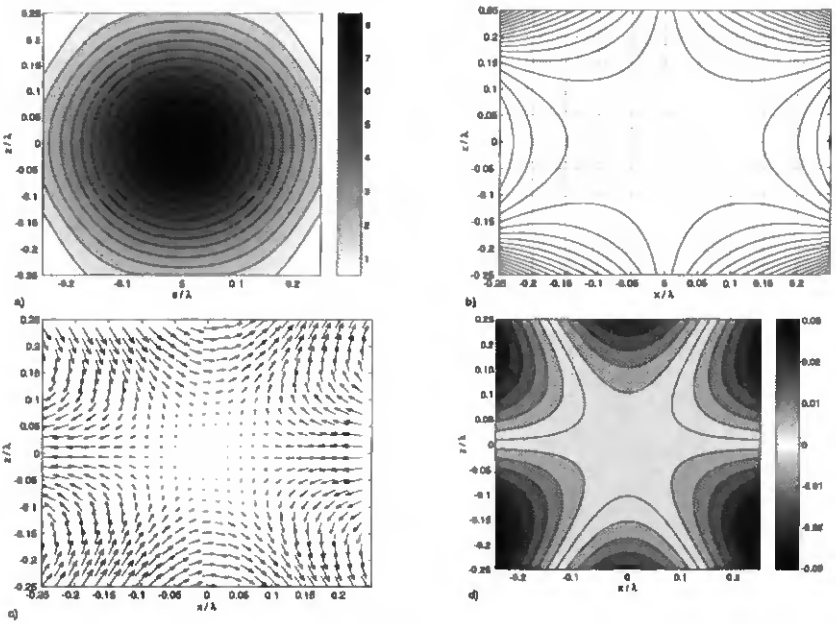


Figure 5.10: Saddle points generated by three plane waves. a) the time averaged intensity b) the phase contour c) the time averaged Poynting vector and d) the curl of the time averaged Poynting vector.

vector

$$< \mathbf{S} > = \frac{-1}{2\omega\mu} A_e^2 \, \nabla\phi_e \tag{5.9}$$

is zero, c.f. Eq. (2.33). Other interesting points are locations having an intensity different from zero but a zero gradient of the phase. We call these positions *saddle points*. The energy flow is obviously zero at these points. For illustration, we consider again the case of the interference of three plane waves. The configuration is the same as in Fig. 5.9, except that the phase of the complex amplitudes are chosen differently ($a_{1,2,3} = 1$).

The resulting field is illustrated in Fig. 5.10. Note that, for this specific case, the intensity is maximal at the saddle points. This is, however, in general not the case. The necessary condition for a saddle point is that the gradient of the phase is equal to zero.

5.3.3 Phase singularities generated by optical microstructures

In Sec. 5.3.1 phase singularities are generated by the superposition of (homogeneous) plane waves. Of course, it is possible to observe phase dislocations also in other fields. In Fig. 5.6

we have shown phase dislocations inside a modulated (hence non-homogeneous) region. There, the field is typically represented by a superposition of modes (Eq. (3.63)). If phase singularities are used for the metrology of grating structure, we need to know the relation between the *location* of the singularity and the *geometry* of the structure. Unfortunately, no trivial relation exists, at least not for small sized structures, and rigorous computations are necessary.

We would like to illustrate this with the following example. A binary dielectric grating ($n_1 = 1.5$, $n_2 = 1$) of period $\Lambda = 4\lambda$ is illuminated by a perpendicularly incident plane wave with TE-polarization. Let us look at the phase behavior in the near field of such a structure for different grating depths d. Figure 5.11 shows a series of phase contour plots. In Fig. 5.11 a) to d) the contour lines are approaching each other at around $x/\lambda = 0.47, z/\lambda = 0.49$. Increasing the grating depth gives rise to *two* phase dislocations, Fig. 5.11 e) and f). The rotation is opposite from each other. By increasing the grating depth even more, the location of the phase singularities changes, Fig. 5.11 g) and h). For very large grating depths, these two specific phase singularities will disappear (not shown in Fig. 5.11).

Actually, it is possible the separate the *life-cycle* of the phase singularities into three parts. This can be best seen if wee look at the local minimum of the intensities. In the *pre-birth* phase, corresponding to region **A** in Fig. 5.12, the intensity of the local minimum is different from zero. Approaching the point of birth (at $d = 0.42\lambda$), the intensity drops sharply to zero (In Fig. 5.12 the "zero" intensity is at 10^{-9}, corresponding the the numerical noise level of the simulations). The intensity stays then at zero (region **B**) until $d = 1.68\lambda$. Increasing the depth the phase dislocation disappears (region **C**).

Figure 5.13 gives another view for the live cycle of a phase singularity. Here, the position of the phase dislocation for different grating depths is shown. The configuration is still the same as used for Fig. 5.12. At the point of birth ($d = 0.42\lambda$) *two* phase dislocations are generated. Similarly, at the point of death ($d = 1.68\lambda$) the two phase dislocations recombine.

It is possible to conclude from Fig. 5.13 that there is a relationship between the edge location and the position of the phase dislocation. However, it is evident that this relationship is *not trivial*. Hence, measuring the phase in the near field gives more information about the structure, but the reconstruction of the object is not trivial.

5.3.4 Stability of phase dislocation against spatial filtering

The observation of phase dislocations is not limited to the near field of a structure. An imaging system transforms the near field through a low-pass filter. First of all, the

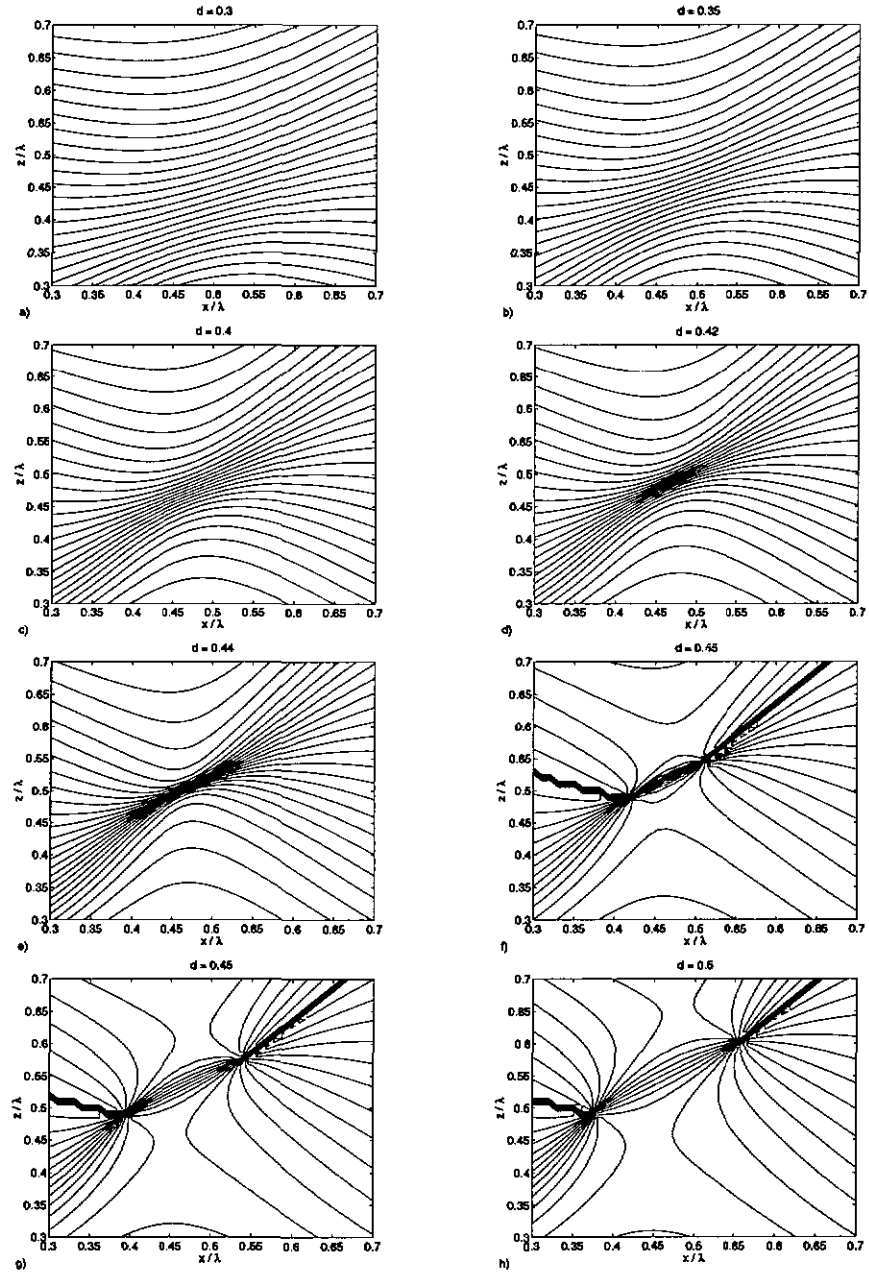


Figure 5.11: *Birth* of a phase singularity: The phase contour of a binary grating structure is shown for different grating depths d.

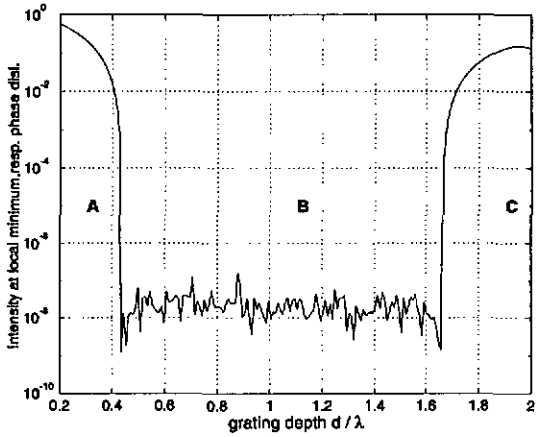


Figure 5.12: Intensity of the local minimum intensity is plotted in function of the grating depth. A phase singularity exists in region **B**.

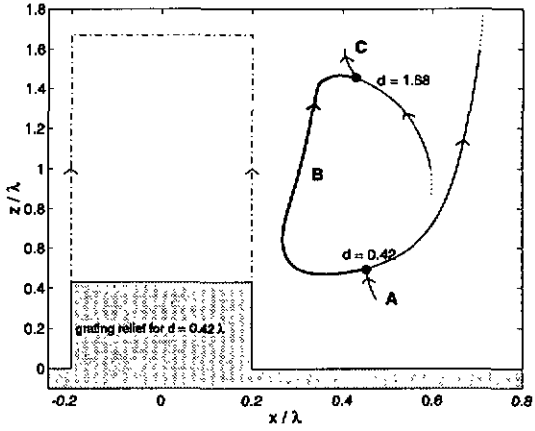


Figure 5.13: Location of the local minimum of intensity for different grating depths.

evanescent waves do not contribute to the far field diffraction pattern. Furthermore, some diffraction orders are cut away. The resulting superposition of waves can be quite different from the near field case. But still it is possible that the far field contains phase dislocations. This is illustrated in Fig. 5.14, where the phase contour in the near field and in the image plane of a binary grating are plotted for different numerical apertures NA. The grating period is $\Lambda = 4.5\lambda$, corresponding to the case of 9 propagating diffraction orders ($n_1 = 1.5$). The height of the grating is $h = \lambda$. In Fig. 5.14 a) the rigorously calculated near field is shown. In Fig. 5.14 b) the phase contour plot in the imaging plane of an ideal (aberration free) 1:1 imaging system is reported. The numerical aperture is NA=*1*, corresponding to the case where all propagating diffraction orders are retained, however the evanescent waves are cut away because they do not contribute to the far field. It is interesting to see that the phase dislocation which is close to the surface (at $x = 0.4\lambda$) has changed position. The other three, which are outside the penetration depth of the evanescent wave, have of course not changed. For Fig. 5.14 c) to f) the numerical aperture has been successively reduced, and therefore also the number of interfering waves. For NA=*0.65* (Fig. 5.14 f) all phase dislocations have disappeared.

Let us summarize the most important properties of phase dislocations

1. Phase dislocations are isolated points with intensity equal to zero. At this point the phase is not determined. They can be created by the interference of at least 3 plane waves of appropriate amplitude.

2. In a scan through a phase dislocation the phase changes always by π, independent of the scan direction

3. The *curl* of the time averaged Poynting vector is always different from zero at a phase dislocation.

And finally, let us mention that phase dislocations are also observed in other scientific research fields. Figure 5.15 shows a *cotidal* map of the north see. The lines correspond to the location of the same tidal phase. The two large points in the middle of the north sea are *amphidromic* points, i.e. isolated points of zero (tide -) amplitude.

This considerations on phase singularities will help us to discuss the "superresolution" in phase images (see Sec. 7.3.2). The next section will show a consequence of the fact that the phase changes always by π through a phase dislocation.

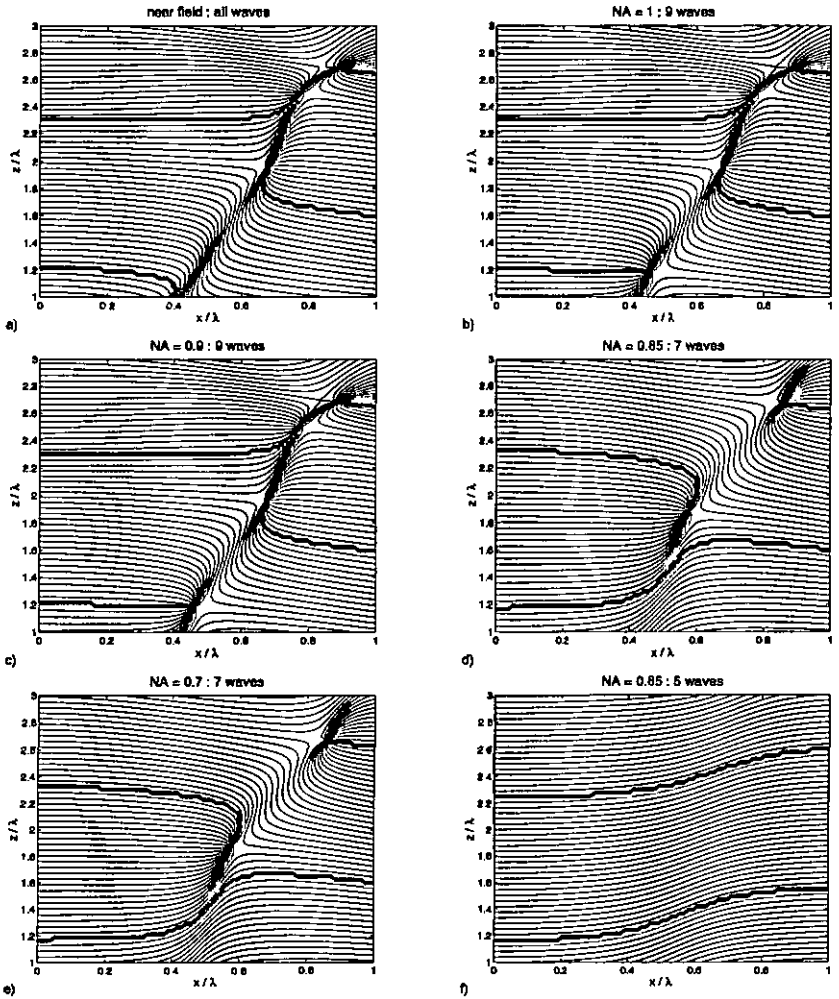


Figure 5.14: Rigorously calculated phase contour plots of the a binary grating structure ($\Lambda = 4.5\lambda$, $t = 0.4$, $n_1 = 1.5$, $n_2 = 1$), illuminated by a perpendicular incident plane wave (TE-polarization). In a) the near field is shown. In b) to f) the far field is reported, obtained by low pass filtering with different cut-off frequencies. This corresponds to an aberration free, coherent imaging system of numerical aperture NA.

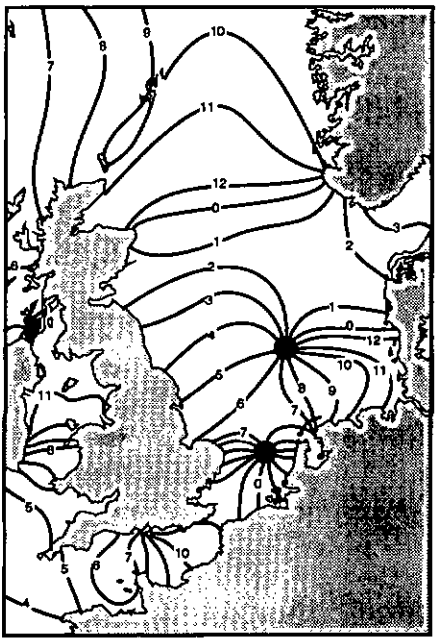


Figure 5.15: Cotidal map of the north sea. The lines correspond to locations of same tidal phase. The spacing between two lines is one hour. The two large points in the middle of the north sea are amphidromic points, i.e. isolated points of zero tide amplitude.

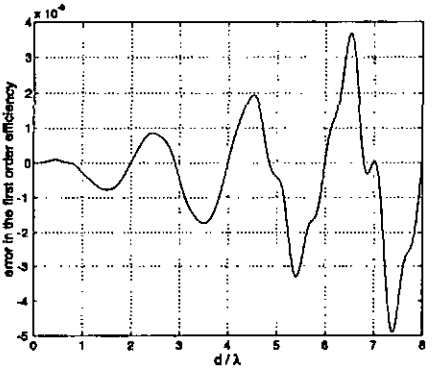


Figure 5.16: Difference of the rigorously calculated first order intensity and the thin element prediction. The calculations are made for a grating of large grating period $\Lambda = 20\lambda$ and different grating depths d.

5.3.5 Consequence of phase dislocation for the far field: an example

It is well know that the diffraction efficiency of a binary structure is about 40.5% in the thin element approach. We have seen that the thin element predicts wrong values for the diffraction efficiencies especially for small grating periods. From the discussion on phase singularities we know now that the phase jumps are always π. The thin element approach, however, predicts phase jumps different to π. The phase is just proportional to the surface profile. Hence, if the surface height jump is different from a multiple of π we admit a *non-physical* field. This behavior can be illustrate by the comparison between rigorous calculations and thin element calculations for different surface heights. Figure 5.16 shows this difference for the first order diffraction efficiency of a binary phase grating versus the structure depth. The index of refraction are $n_1 = 1.5$ and $n_2 = 1$. The grating period is $\Lambda = 20\lambda$. The Fresnel reflection losses are compensated in the case of the thin element approach. The grating is in the thin element region, this means that the overall error is small (less than 0.2%). However, it is interesting to see that the error is maximum for $d = (m + 0.5)\pi$ corresponding to phase jumps of $\phi = (m + \pi/2)$, which are the most *non-physical*. For $d = m\lambda$ the error is minimum.

Chapter 6

Applications of optical microstructures

In Sec. 5.1 we discussed the behavior of microlenses. We have seen that for small diameters the edge diffraction effects become more important than the refraction on the curved surface. There are many other elements where diffraction and refraction plays an important role. A general discussion about refractive versus diffractive elements is out of the scope of the present work. The interested reader is referred to [56]. Here, we only mention that planar diffractive optical elements (DOEs) [57] are an alternative to classical elements (lenses, mirrors, beamsplitters). Relying on diffraction and interference rather than on reflection and refraction (Fig 6.3), elements with unique and novel properties can be realized. Almost any structure shape, including non–rotationally symmetric aspherics, can be manufactured, providing all degrees of freedom for the design. Other interesting aspects of DOEs are low weight, strong dispersion, and the possibility to make segmented elements, large arrays of elements, beamsplitters, and polarizers. These properties are useful for many applications of DOEs, including filters for image data processing [58],

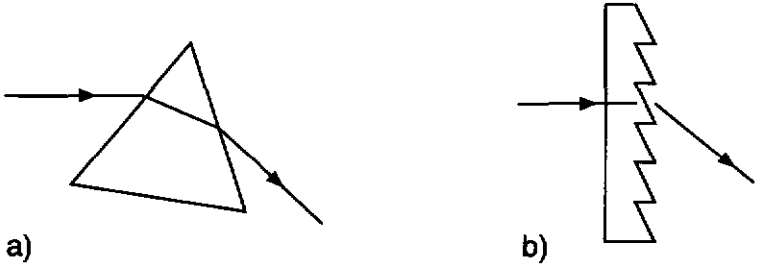


Figure 6.1: Beam deflection made by a) refraction, b) by diffraction.

beam shaping [59, 60], and antireflection structures [61, 38]. Furthermore, the combination of refractive and diffractive surfaces (hybrid elements) offers new possibilities for optical design. The negative dispersion of DOEs can be used to compensate the chromatic aberrations of refractive lenses [62, 63]. Hybrid elements can also be used to compensate the temperature induced variations of their mounting system. Some examples of athermalized hybrid lenses have been discussed in the literature [64, 65], and [66] (see appendix A.III).

6.1 Far field beam shaping (A.III)

The objective of this section is to demonstrate the potential of diffractive optical elements (DOEs) for the design of optical and opto-electronic systems for advanced laser intersatellite communication terminals. Diffractive optical elements for space applications must comply with a number of requirements, including mechanical, thermal and optical stability [63]. Suitable techniques for realizing the microstructures in space qualified materials are based on a variety of high resolution lithographic and optical processes [56]. We discuss possible applications for DOEs in compact optical communication terminals. The design and experimental realization of three different elements are presented. These elements include a ring pattern generator allowing tracking and data transmission within one optical element , an athermalized hybrid achromat, and a hybrid lens system.

An example of the elements studied in [66] (see appendix A.III) is shown in Fig. 6.2. A single (diffractive) element is used to split the incoming beam into a focused point and a ring pattern. The ring pattern is used for the tracking between the two satellites. Changes of the relative positions of the sending and receiving communication terminals are directly translated into a shift of the ring intensity pattern on the tracking sensor, which can be measured by means of the 4-quadrant detector. The focused point is used for the receiver signal. Figure 6.3 shows the measured intensity distribution in the tracking detector plane, compared with the theoretical predictions. There is an excellent agreement between both curves.

6.2 Near field smart masks

As shown in Sec. 5.1 even small microlenses can focus light down to spots of a few wavelength in diameter. They can be used in photolithography, or more particularly, to simplify the mask imaging system be including the projection system in the mask.

There are a number of applications where it is required to image an array of dots, lines,

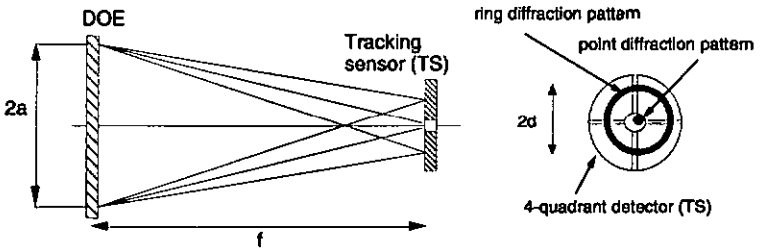


Figure 6.2: Schematic view of the principles of the tracking/receiving system of a intersatellite laser communication terminal. The incoming beam is splitted into a focused point (used for the receiver signal) and a ring pattern (used for tracking).

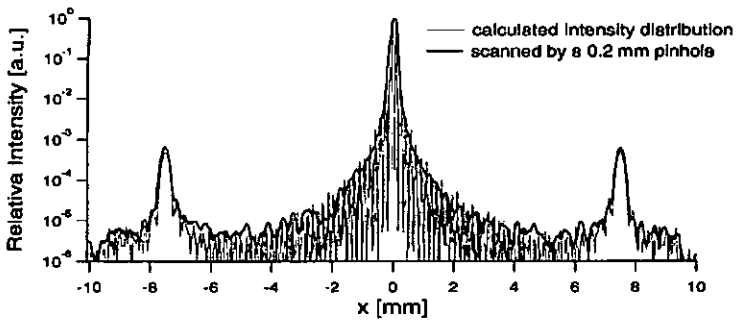


Figure 6.3: Measured and calculated diffraction pattern of a ring-point generator used for tracking and receiving in a space communication terminal.

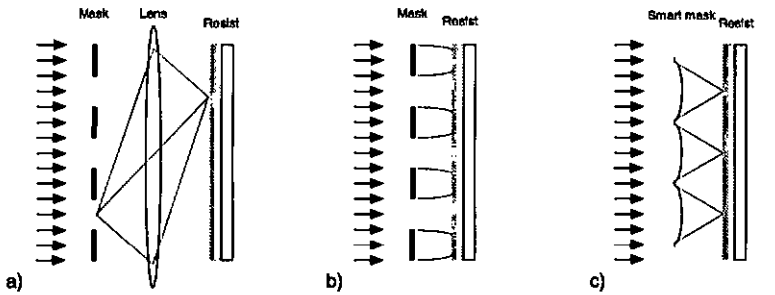


Figure 6.4: Different types of photolithography: a) projection lithography, b) proximity printing, c) smart mask lithography.

or other shapes onto a photoresist coating. One example is in the formation of arrays of field emitters for flat panel displays. Normally, light passes through small apertures in the photomask and is transferred onto the photoresist coating by a projection lens (Fig. 6.4 a), or by free space propagation (proximity printing, Fig. 6.4 b). In both cases an important fraction of the light is blocked by the chromium mask, which makes the process inefficient. We propose an improved photomask arrangement, which is substantially more efficient [67, 68], Fig. 6.4 c). In our system, each element on the photomask includes its own focusing system that concentrates the light onto a a point, a line or an other area of a photoresist layer.

In Fig. 6.5 the near field of a microlens is shown with and without the presence of the photosensitive layer. The calculations have been made with the Fourier modal method FMM (Sec. 3.3.3). It is evident that the back reflection from the photoresist layer disturbs the near field of the lens, but the light confinement is still adequate for photolithographic printing.

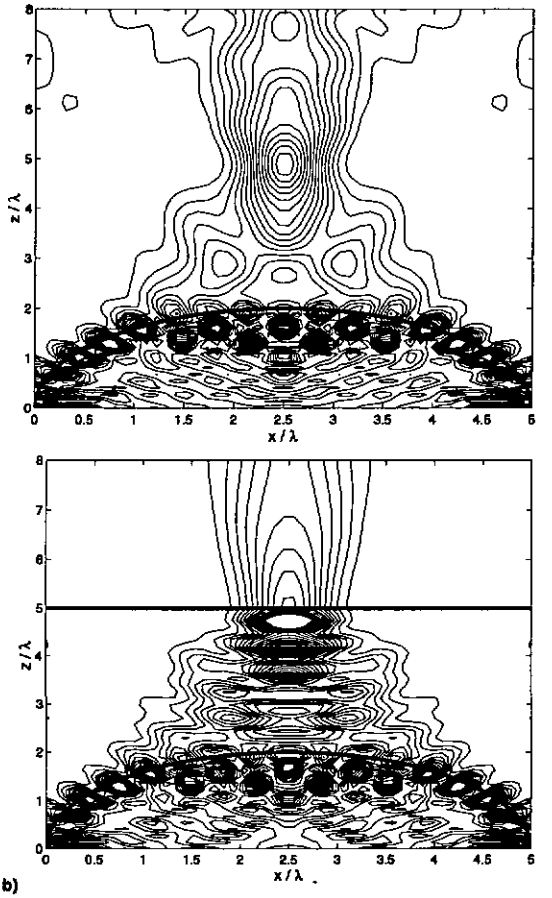


Figure 6.5: Rigorously calculated intensity distribution in the near field of a microlens (diameter $d = 5\lambda$, height $h = 2\lambda$, $n = 1.5$, TE-polarization), a) free space, b) presence of a dielectric layer ($n = 1.5$) at $z = 5\lambda$.

Chapter 7

Characterization of grating structures

The testing of grating structures typically includes the measurement of diffraction efficiencies, resolving power, or spectral purity [36, Chap. 11], hence the optical propriety of the element. Here, we are more interested to study the *geometry of the microstructure itself* by optical methods. The basic question is : Is it possible to reconstruct the geometry of a grating structure by measuring its optical response? By geometry, we mean the shape of the structure: grating period, line-width, grating depth, slant angle, or more general the surface function $f(x)$.

In chapter 3, we learned about the complicated interaction of light with such structures. We have seen that the object and the field generated by the object are not the same. We will therefore distinguish between the *object* and the *object field*. We estimate that *in general, it is not possible the solve the inverse interaction problem*, but there are a number of important cases where it is still possible to get information about the structure itself. If the element is large and thin (hence, in the 'scalar" domain) there is a *linear* relation between the object geometry and the phase of the emerging field. In this case, the characterization of the geometry is referred to the characterization of the field, which is not too difficult using standard optical characterization methods. In a more general formulation it is possible to state that the inverse-scattering problem can be directly solve as long as the Rayleigh hypothesis is valid (see Sec. 3.2.1). There are different papers treating this problem in the near field [69, 70] and in the far field [42, Chap.9].

If the element is in the resonance domain (i.e. the grating period is in the order of the wavelength) the situation is more delicate. The general approach is to make a forward optimization (see Fig. 7.1). This includes the following steps: 1. The grating structure is illuminated by an electromagnetic field of a certain state (spatial frequency spectrum k_x,

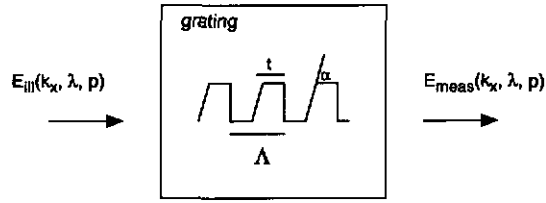


Figure 7.1: Characterization of the *geometry* of periodic optical microstructures.

wavelength spectrum λ, polarization state p). 2. The response of the grating is measured (i.e. the state of the emerging field is determined). 3. The measurements are then fitted to theoretical (rigorous) calculations based on a parameterized geometry (Λ,t,α,etc). In order to refine the results, the measurement and calculations are repeated for different states.

Based on forward optimization, different methods have been reported in the literature. Probably the most complete approach has been made by Naqvi and coworkers [71, 72, 73, 74]. The basic idea of his method is to measure the first order diffraction efficiency in function of different incident angles. The obtained measurement is some kind of a "fingerprint" of the grating structure. Later, the method has been refined to include different polarization states and wavelengths; also higher diffraction orders were measured. Because several grating periods are illuminated at the same time, this method allows only to extract information about the mean value of the different parameters (mean line-width, mean slant angle, etc).

In this chapter, we present two different methods to characterize the geometry of the grating structure. The first method (scanning spot metrology) enables the extraction of information on the *local* line-widths, the second method (laser diffractometer) measures the mean grating period with picometer accuracy ($\Delta\Lambda \leq 5pm$).

At the end of the chapter we will discuss some aspects of *superresolution*. It is clear, that the question on how well the shape of an object can be determined from measurements implies directly the question on how good is the resolution of the method. We will therefore review some historical aspects on resolution and superresolution.

7.1 Scanning spot method (A.IV, A.V)

Advancement in the areas of phase shift optical lithography, e-beam lithography, and x-ray lithography have enabled the realization of very fine relief structures in the nanometer-micrometer range. The ability to measure the relief parameters and the absolute position

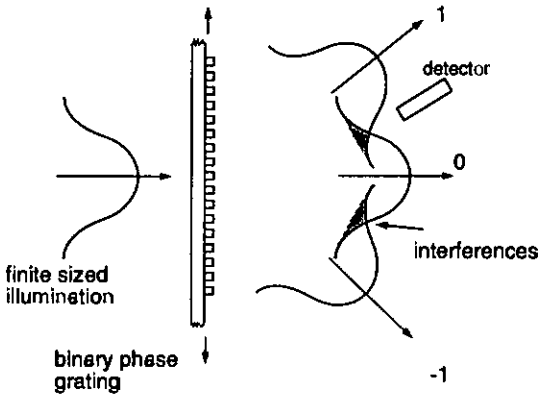


Figure 7.2: Principle of the scanning spot metrology: A grating structure is illuminated by a focused laser beam. A detector is measuring the interference while the structure is scanned through the beam.

accuracy of these structures is of obvious importance, not only to determine if the desired structure has been realized, but also to optimize the fabrication process. The wish list for metrology includes non-destructive testing, testing of large areas, and testing of large aspect ratio structures having sub-μm lateral dimensions. Current techniques for measurement are optical microscopy (including conventional, confocal and near field techniques) [75], scanning electron microscopy [76] (SEM), and scanning probe microscopy [77] (SPM). None of these techniques, however, is capable of providing rapid accurate sub-μm measurements over larger areas. Optical testing methods have been applied to address these requirements. The relevant parameters of a lithographic process, i.e. the depth and width of a grating like test structure, can be determined by scattering an incident laser beam at the structures and measuring the far field intensity distribution. A fabrication technology can be best characterized, if specific structures with sensitive and easy-to-measure properties are realized. Gratings provide ideal test elements, since they split an illuminating laser beam into a discrete array of diffraction orders. Previous published results have demonstrated the capability of far field measurements to rapidly and very accurately measure large aspect ratio sub-μm feature over larger areas. The main drawback is that these techniques only determine the average parameters, and are unable to determine local errors such as single line defects. Therefore, we have investigated a scanning spot method which provides accurate information about edge locations of opaque structures on chrome masks. The method involves illuminating the structure with a small

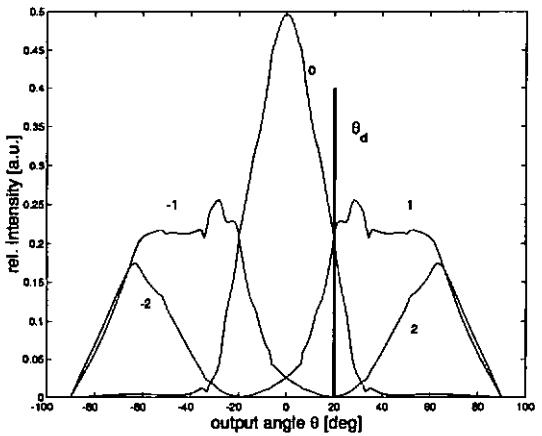


Figure 7.3: Rigorously calculated diffraction pattern of a binary grating structure of $\Lambda = 1.5\mu m$, illuminated by a Gaussian beam of diameter $2\omega_0 = 1.5\mu m$. The thickness of the grating is $d = 0.7\mu m$, the incident wave is TE-polarized and has a wavelength of $\lambda = 0.632\mu m$.

spot size focused laser beam and evaluating the diffraction pattern as the structure is scanned. It has been applied to amplitude gratings, realized typically by opaque and transparent line structures on chrome masks.

The total transmitted power gives information about the edge locations of the lines [78] (see appendix A.IV). In the case of phase gratings a different approach is used. The focused laser beam generates enlarged diffraction orders. If the beam size is sufficiently small, i.e. has a large angular spectrum, the diffraction pattern of the diffracted orders overlap and interfere (Fig.7.2). Scanning the mask introduces a (linear) phase change of the higher diffraction orders. The resulting interference modulation depends on the ratio between the grating period and the focus spot size. This modulation can be used to extract the edge locations with high accuracy [79].

An example of a rigorous calculated diffraction pattern of a binary dielectric grating structure illuminated by a Gaussian beam is shown in Fig. 7.3, the contributions of the different diffraction order being reported separately. In this situation, the best place to put the detector is at $\theta_d = 20°$. There, the first diffraction and zero diffraction order are equal, and therefore a maximal modulation depth is obtained.

Based on this principle, we were able to determine local line-widths of $1\mu m$ gratings with a precision of $10nm$. The detailed work on amplitude gratings can be found in [78]

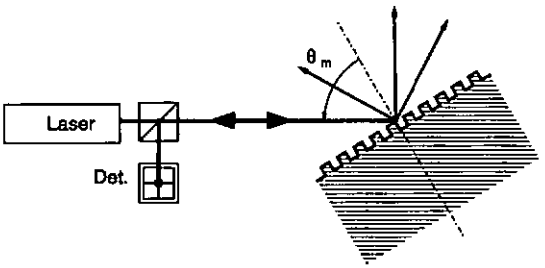


Figure 7.4: Littrow diffraction angle measurement system.

and [80] (see appendix A.IV and A.V).

7.2 Laser diffractometer

The laser diffractometer is a well-known instrument to determine the period of a grating by measuring the diffraction angles. We will show that this method is highly accurate. The basic principle is illustrated in Fig. 7.4. A grating structure splits an incoming beam into different diffraction orders. Measuring the Littrow diffraction angles θ_m requires only one rotating stage to determine the mean grating period. The direction of the diffraction orders in the Littrow mount is related to the grating period Λ by

$$n_{\text{air}} \sin\theta_m = m\lambda/2\Lambda\,, \tag{7.1}$$

where n_{air} is the refractive index of the air, m is the diffraction order and λ is the wavelength.

If the laser diffractometer is placed in a temperature and humidity controlled environment, high accuracy can be achieved. The index of the air has to be taken into account with the same precision as the measurement accuracy ($|\Delta\Lambda/\Lambda| = |\Delta n_{\text{air}}/n_{\text{air}}|$).

Based on this principle, a laser diffractometer has been built and characterized. Figures 7.6 and 7.5 show a schematic view, respectively a picture of the system. The angular precision of the rotary stage is in the order of $0.1'' \approx 0.5\mu rad$. This instrument allows the rapid and precise determination of the grating period of reflection gratings ($\Lambda = 400nm \ldots 10\mu m$) with an accuracy of better than 10^{-5}. Therefore, it is possible to determine the mean grating period of a $1\mu m$ grating to better than 10 picometer!

The method has been successfully applied to determine the precise grating periods and also the orthogonality of cross-gratings. Figure 7.7 a) shows an atomic force microscope (AFM) measurement of such a structure. The orthogonality of the gratings is determined

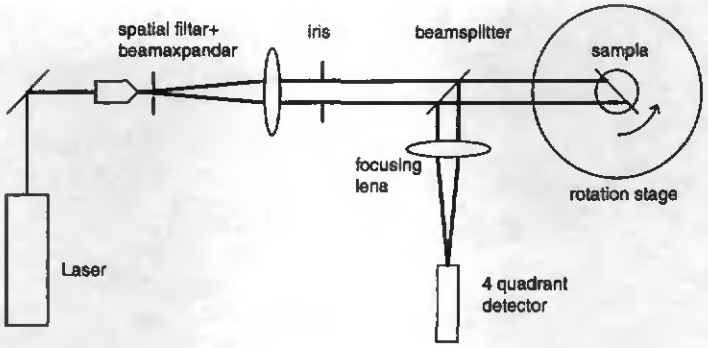


Figure 7.5: Schematic view of the laser diffractometer.

Figure 7.6: Picture of the laser diffractometer built up at the Swiss Federal Institute of Metrology.

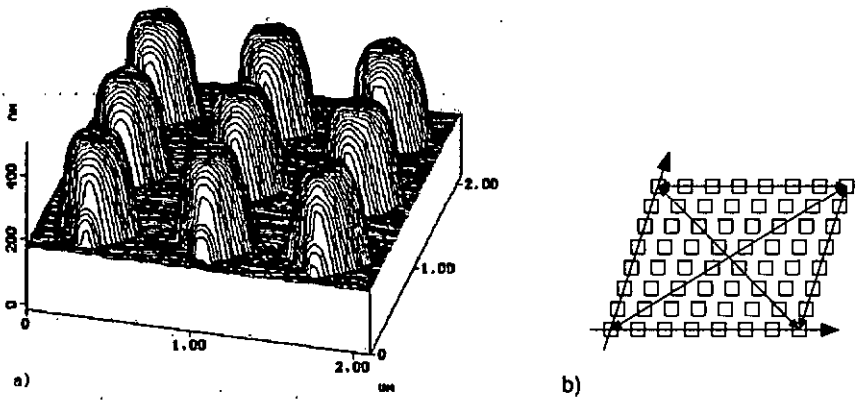


Figure 7.7: Cross grating used as test object for the laser diffractometer. a) AFM measurement, b) extracting information on the orthogonality of the gratings.

by the measurement of the grating period in diagonal directions (Fig. 7.7 b) and applying some elementary trigonometrical formulas.

Table 7.1 summarizes the comparison of measurements made by the laser diffractometer and the AFM. The agreement between the two methods is very good. The difference in the absolute value of the grating period is less than 10^{-4}. The AFM measurements (itself a *local* method) were obtained by taking the mean value of a large range scan over several hundreds of grating periods. The translation stage was interferometrically controlled. The main advantage of the laser diffractometer over the AFM measurements is the fast,

	AFM $[nm]$	diffraction $[nm]$	difference rel
x-pitch	700.967	700.953	$2.0E-5$
y-pitch	700.981	700.928	$7.5E-5$
diagonal 1 - pitch	989.297	989.248	$6.2E-5$
diagonal 2 - pitch	993.265	993.327	$6.2E-5$
$\rightarrow$ orthogonality	90.2290 °	90.236 °	$7.1E-5$

Table 7.1: Comparison between AFM and laser diffractometer measurements.

contactless and accurate determination of the mean grating period. The "mathematical operation" of taking the average over several hundreds of grating periods is realized optically by the diffraction of the light beam in free-space.

7.3 Superresolution

7.3.1 Early work on the degree of freedom of an image

Early considerations on optical resolution are due to LORD RAYLEIGH [81], who showed that for a classical optical telescope the resolution limit is $x_R = \lambda f/D$, where f and D are the focal length and the aperture diameter of the instrument, respectively. TORALDO DI FRANCIA [82, 83] derived the intuitive conclusion that the image is completely determined by giving the complex amplitudes at a square lattice of points spaced $\lambda/2NA$ apart (NA being the numerical aperture). Therefore, the image has

$$S = X/x_R = DX \;/\; \lambda f \tag{7.2}$$

degrees of freedom, where X is the size of the object field, S is also called the Shannon number. As a consequence, since the object field can have an infinite number of degrees of freedom, it follows that many different objects fields can have one and the same image. The observer must necessarily make use of some *a priori* information[1].

13 years later Toraldo di Francia modified his claims [84] and proved that his first conclusions were mathematically wrong. For, if the object field size has a finite size, the Fourier transform on the pupil is an analytic function; thus, knowing this function over a finite interval is sufficient to reconstruct the complete function. This mathematical argument induces the following interesting question: *Why this possibility of superresolution has not yet been transferred into physical reality, with a real and substantial gain compared with conventional instruments?* Toraldo di Francia showed that the practical limitation of the number of degrees of freedom is a mathematical phenomenon. He analyzed the difficulties that are inherent in the retrieval of the missing information. Later GORI [85] generalized the concept of degree of freedom to optical imaging system in the presence of aberrations. The argumentation is based on the following idea:

The relation between the object field and the image field is expressed by the superposition integral

$$U_i(x) = \int_\Sigma U_o(x')\; K(x, x')\; \mathrm{d}x' \,, \tag{7.3}$$

where $K(x, x')$ is the impulse response of the imaging system (see also Chap. 4) and Σ is the object field domain. The object field $U_o(x)$ and the image field $U_i(x')$ are expanded using the same set of basis functions

$$U_o(x') \;=\; \sum_m a_m\, \psi_m(x') \,, \quad \text{and} \tag{7.4}$$

[1] Note that again we make a distinction between the *object* and the *object field* (that is the field created by the object).

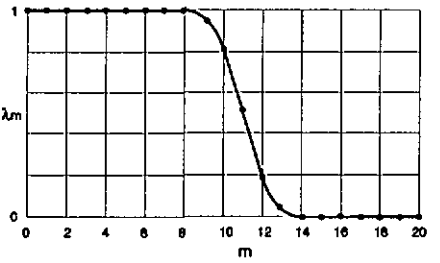


Figure 7.8: Positive real number λ_m. The graph indicates how a basis function of the object field is transferred into the basis function of the image for a coherent and aberration free optical system. The configuration corresponds to with a Shannon number $S = 11.5$.

$$U_i(x) = \sum_m b_m \, \psi_m(x) \,, \tag{7.5}$$

with

$$a_m = \int_\Sigma U_o(x') \, \psi_m^*(x') \, \mathrm{d}x' \,, \tag{7.6}$$

$$b_m = \int_\Sigma U_o(x) \, \psi_m^*(x) \, \mathrm{d}x \,. \tag{7.7}$$

The basis functions in Eqs. (7.4) and (7.5) are chosen to be the eigenfunctions of the impulse response, viz.

$$\int_\Sigma \psi(x') \, K(x, x') \, \mathrm{d}x' = \lambda \, \psi(x) \,. \tag{7.8}$$

For a perfect and coherent imaging system this eigenfunctions $\psi_m(x)$ are the *prolate spheroidal wave functions* [86, 87, 88].

Introducing Eqs. (7.4) and (7.5) into Eq. (7.3), interchanging sum and integral, and taking advantage of Eq. (7.3) results in

$$b_m = \lambda_m \, a_m \,. \tag{7.9}$$

Hence, λ_m indicates how a basis function of the object field is transferred into the basis function of the image field. Therefore, suppose we know the image $U_i(x)$, the coefficients b_m of the image expansion are found by Eq. (7.7). The coefficients a_m of the object field are then determined by Eq. (7.9) and the object field can be reconstructured by Eq. (7.4).

However, practically the eigenvalue λ_m is one for $m < S$ and zero for $m > S$. As an example, the behavior of λ_m as a function of m is shown in Fig. 7.8 for $S = 11.5$. The behavior is similar to that of a step function. It is clear that for a real imaging system, characterized by a certain noise level, the higher order terms of the object field expansion can not be determined anymore. Therefore, the resolution limit based on the sampling

theorem is a good estimation for the resolution of a classical optical system, although physically there is no *sharp* limit.

All this considerations were made to study the reconstruction of an *unknown* object field. If we have some *a priori* information about the object or about the object field, the situation is completely different. It has already been noted by TORALDO DI FRANCIA [84] (and unfortunately forgotten nowadays by many authors) that if we have some *a priori* information about the object or the object field, the term *resolution* loses its signification and the problem is transferred to a parameter optimization, as discussed in the introduction of this chapter.

7.3.2 Superresolution in phase images

At the beginning of the nineties, TYSCHINSKY [89], and TYSCHINSKY and VELZEL [90] presented spectacular images, reactivating the discussion on superresolution. Figure 7.9 shows one of theirs measurements [90, Fig. 9] of a semiconductor test object with an etched structure in the Si-layer. The measurements have been made with a high resolution interference microscope (see schematic view in Fig 7.10). The phase image has details well beyond the classical resolution limit. Unfortunately, there is no satisfying explanation in their publications.

Only recently, TOTZECK [91, 92] gives full explanation of the apparent superresolution in phase images: the phase measurements show scans through phase dislocations, hence sharp jumps are expected. As we have seen in Sec. 5.3, such phase dislocations can be easily generated by the superposition of plane waves. The phase measurements do not contradict the classical resolution criteria. The term "superresolution" in the context of phase images is therefore *not suitable*.

The main problem to solve in the measurement of λ-sized structures is the *light interaction problem*. Since the elements are in the resonance domain, there is *no linear relation between the object and the phase anymore*. As an example, we present in Fig. 7.11 a) the near field and Fig. 7.11 b) the scan through two phase dislocations for a similar element as Tychinsky had used. Unfortunately, we do not know all the parameters (wavelength, polarization state, spatial frequency spectrum) of Tychinsky's work. That is the reason why it is not possible to make a quantitative comparison between the two graphs. Qualitatively, the phase measurement of Fig. 7.9 agrees well with the calculated values of Fig. 7.11 b).

To solve the inverse problem, that is to find the object from the image field, we need forward parameter optimizations as discussed at the beginning of this chapter. The imaging system (hence, the low-pass filtering) can directly by included in the optimization.

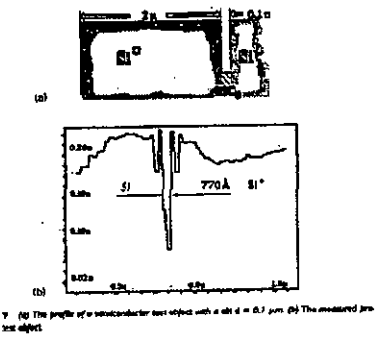


Figure 7.9: Phase measurement made by Tyschinsky and Velzel [90, Fig. 9].

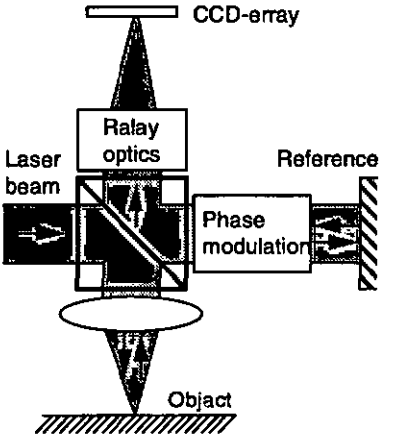


Figure 7.10: Schematic view of a high resolution interference microscope (Linnik type).

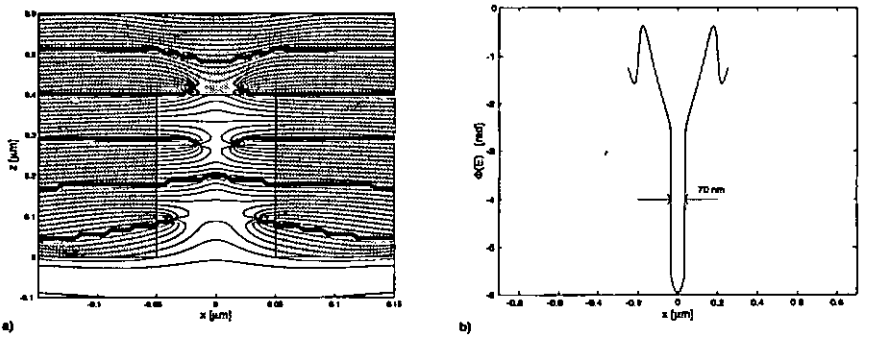


Figure 7.11: a) Rigorously calculated phase contour plot of the near field of an etched $100nm \times 400nm$ groove in a Si-layer. The structure is illuminated by a perpendicularly incident plane wave of TE-polarization ($\lambda = 488nm$). b) Phase plot of a scan close to two phase dislocations. The phase changes rapidly within $70nm$.

7.4 Discussion

We would like to conclude this brief considerations on characterization and superresolution by the following statements:

- First of all, one has to be aware that the field generated by a λ-sized object has not the same shape as the object itself. The interaction of light with these kind of structures has to be modelled using a *rigorous interaction* theory (see for example Sec. 3.3.3).

- Therefore, it makes no sense to develop complicated theories on superresolution, or diffraction theories (in the sense of *propagation*), if the starting point is not a *physically realistic* field; e.g. to make investigation on the propagation of a binary amplitude function with line-widths *below* the wavelength, as it has been made for example in [93], or to admit a π-phase jump without the associated amplitude change.

- Since there does not exist (at least not to our knowledge) a general rigorous *inverse* interaction model, the extraction of geometric parameters of λ-sized microstructures has to be made through forward optimization (c.f. Chapt. 7). For this purpose, it is necessary to acquire as much information about the object as possible. Hence, it is advisable to measure *amplitude* and *phase* of the emerging optical field.

- And, finally, if we have some *a priori* information about the object, the notion of *resolution looses its signification* and the problem is reduced to a parameter optimization.

Chapter 8

Conclusions

The objective of this work was to show different aspects of the interaction of light with periodic optical microstructures. In a first part, we discussed different theoretical models. We showed that it is important to use *rigorous* diffraction theories, if the structures are in the order of the optical wavelength. We presented the Fourier modal method as a suitable tool for solving the grating diffraction problem. Furthermore, we extended the existent Rayleigh method and the Waterman approach to dielectric surface relief structures. Introducing approximate theories helps to understand the complex problem of light-matter interaction.

For optical systems consisting of several elements, we introduced the concept of *spectral response and transfer function*. New in this approach is, that also evanescent waves are taken into considerations. This opens new possibilities in the modelling of modern complex optical systems. As an example, we presented the scanning near field optical microscope (SNOM) as an optical instrument, characterized by transfer functions and spectral responses. We emphasized that for a good SNOM, the transfer functions representing the capability to convert evanescent waves into propagating waves has to be optimized, rather than the overall throughput.

We described how *phase singularities* can be generated. Phase singularities are isolated points of zero amplitude, hence the phase is not determined at these points. These points are well localized. Therefore, measuring their position can give information about the object (shape, index of refraction, etc).

Finally, we evoked the question if it is possible to reconstruct the geometry of a grating structure by measuring its optical response. We found that, in general, we will have to use forward parameter optimization to reconstruct the object. We discussed different points on superresolution. In our opinion, there is no superresolution observed in phase images (if we relate the term resolution to the degrees of freedom of the image).

In summary, the present work shows some aspects of the interaction of light with optical microstructures. Such a work cannot be complete, a lot of points were not mentioned (surface plasmon generation, partial coherence, conical incidence, etc). During this work not only solutions were found to existing problems, but also *new questions* were evoked. Especially, it would be interesting to study phase dislocations in more details, to find conditions for their creation. We observed that small changes of one geometrical parameter can totally alter the phase topography (see Fig. 8.1). For smooth and thin structures, the phase field is also smooth and without singularities. If the thickness of the structure changes, phase singularities appear and the phase topography becomes *turbulent*[1]. It is similar to hydrodynamics, where two different flow state exists: a laminar flow and a turbulent flow. It would be interesting to apply hydrodynamical principles to optical systems (or viz versa, optical principles to hydrodynamical systems...).

[1] Examples of animated fields can be found on the Internet pages of the Institute of Microtechnology, University of Neuchâtel : http://www-optics.unine.ch/research/Microopt/RigDiffraction/main.html

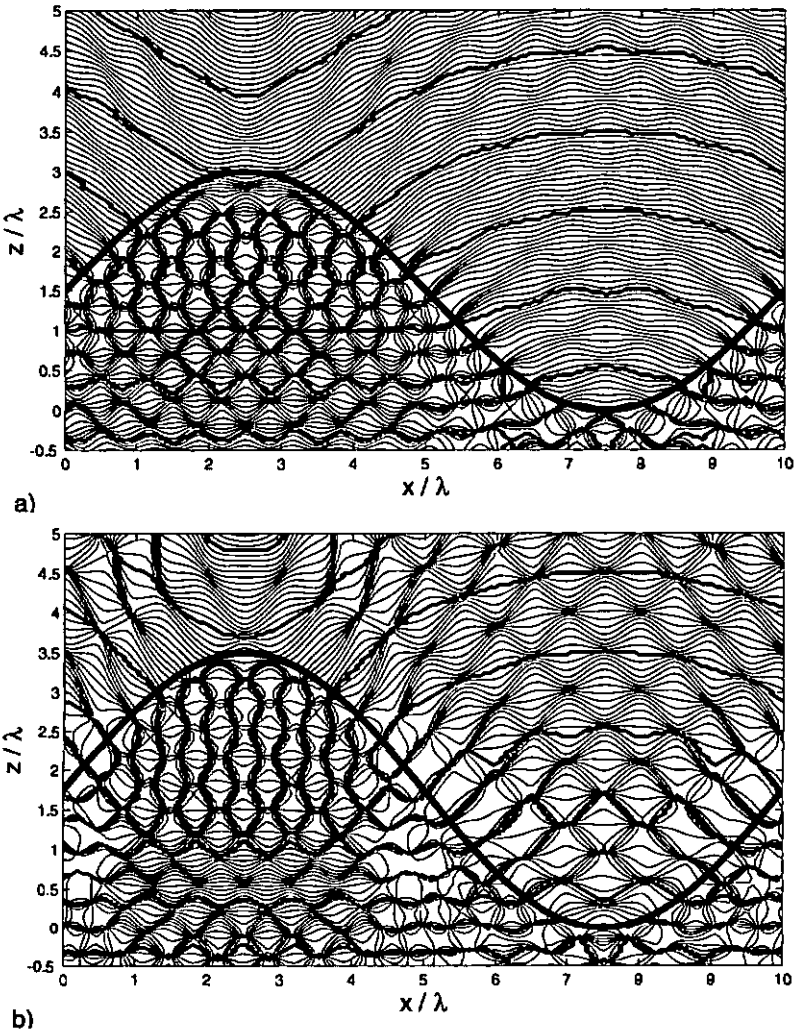


Figure 8.1: Rigorously calculated phase contour plots of a sinusoidal surface relief grating. a) For a thin grating ($h = 3\lambda$) the phase topography is smooth. b) Increasing the grating thickness to $h = 3.5\lambda$ changes dramatically the phase topography, phase singularities are generated and turbulencies appear.

Acknowledgments

I would like to thank

- Prof. René Dändliker, director of this thesis. Thanks to him I started my studies in electronic physics at the University of Neuchâtel more than 10 years ago. I'm very grateful that after my studies he accepted me as PHD-student in his research group. During all the years he was ready for spending some of his precious time for discussions, but he also give me enough freedom allowing me to follow my own scientific interests.

- Prof. Hans Peter Herzig, I really enjoyed the collaboration. He always had a lot of ideas and suggestions. Actually, he contributed to most of the concepts I have used in this thesis. I'm also grateful for critical reviewing my thesis work.

- Prof. Jakob J. Stamnes and Prof. Daniel Courjon for their interest in my thesis as members of the jury.

- Dr. Peter Ehbets, Dr. Ken Weible, and Dr. Damien Prongue, for initiating me into the field of diffractive optics.

- my Finnish colleague Dr. Markku Kuittinen. He brought the Fourier Modal Method to Neuchâtel.

- Micheal Ansorge, Prof. O. Besson for discussions on linear systems and numerical analysis.

- Dr. Ruedi Thalmann, Dr. Felix Meli from the Swiss Federal Office of Metrology, for the assistance in the laser-diffractometer project.

- Dr. Michael Totzeck for several conversations on phase dislocations.

- Dr. Reinhard Czichy, Dr. Thomas Weigel, Dominic Doyle and Falk Draheim for their support during my stay at the European Space Agency in Noordwijk, NL.

- Prof. Sohail S. H. Naqvi and Silke Traut for setting up the scanning spot experiment.
- Prof. Fabienne Weible-Marquis for interesting discussions on practical aspects of near field optics.
- my former office colleague Dr. Urs Vokinger for several hot debates on many optical and non-optical problems.
- Dr. Reinhard Voelkel for many polito-scientific discussions.
- Dr. Christoph Berger, Dr. Manuel Teijido, Marcel Groccia for their computer support.
- Philippe Nussbaum for his support on practical problems in the lab (how to set up an optical system...)
- Antonello Nesci and Carsten Ruckstuhl for doing the experimental work in the sub-wavelength project.
- Jean-Christoph Roulet and Yves-Alain Peter, the only Neuchâtel indigenes in the Applied Optics group, for introducing me to local customs and traditions.
- Dr. Toralf Scharf, Dr. Yves Salvadé, Dr. Ali Pourzand and Etienne Rochat, for introducing me to other field in optics.
- Dr. Peter Kipfer and Andreas Schilling, for bringing some Teutonical practices to Neuchâtel (auf deutsch: Bier).
- Dr. Wolfgang Singer, for many interesting discussions about optics and oenology during his short stay at the IMT.
- Dr. Peter Torres, Dr. Natalie Beck, Alex Los, Beny Guldimann for making easier the daily life during the studies in electronic physics ...
- the other groups of the Institute of Microtechnology as well as the secretarial staff for the friendly and open environment.

Finally, my greatest thanks to my wife Eveline for her patience, support and encouragements.

Bibliography

[1] M. Born and E. Wolf. *Principles of optics*. Pergamon Press, Oxford, 1993.

[2] N. W. Ashcroft and N. D. Mermin. *Solid state physics*. Holt, Rinehart and Winston, Philadelphia, 1976.

[3] R. Petit, editor. *Electromagnetic theory of gratings*, volume 22 of *Topics in current physics*. Springer-Verlag, Berlin, 1980.

[4] Lord Rayleigh. "On the dynamical theory of gratings". Proc. Roy. Soc. Ser. A, **79**, 399-416, (1907).

[5] R. F. Millar. "On the Rayleigh assumption in scattering by a periodic surface". Proc. Camb. Phil. Soc., **65**, 773–791, (1969).

[6] M. Nevière and M. Cadilhac. "Sur la validite du developpement de Rayleigh". Opt. Comm., **2**(5), 235–238, (1970).

[7] N. R. Hill and V. Celli. "Limits of convergence of the Rayleigh method for surface scattering". Phys. Rev. B, **17**(6), 2478-2481, (1978).

[8] H. A. Kalhor. "Numerical evalnation of Rayleigh Hypothesis for analyzing scattering from corrugated gratings - TE polarization". IEEE Trans., **AP-24**, 884–889, (1976).

[9] P. M. van den Berg and J. T. Fokkema. "The Rayleigh hypothesis in the theory of reflection hy a grating". J. Opt. Soc. Am., **69**(1), 27–31, (1979).

[10] A. Wrigin. "Reflection from a corrugated surface". J. Acoust. Soc. Am., **68**(2), 692–699, (1980).

[11] A. Wirgin. "On Rayleigh's theory of sinusoidal diffraction gratings". Opt. Act., **27**(12), 1671-1692, (1980).

[12] J. H. Hugonin, R. Petit, and M. Cadilhac. "Plane-wave expansions used to describe the field diffracted by a grating". J. Opt. Soc. Am., **71**(5), 593–598, (1981).

[13] J. A. DeSanto. "Scattering from a perfectly reflecting arbitrary periodic surface: An exact theory". Radio Science, **16**(6), 1315–1326, (1981).

[14] B. V. Apel'tsin and A. G. Kyurkchan. "Rayleigh's Hypothesis and the analytic properties of wave fields". Soviet J Comm. Tech. Elec., **30**, 97–111, (1985).

[15] D. Maystre and M. Cadilhac. "Singularities of the continuation of fields and validity of Rayleigh's hypothesis". J. Math. Phys., **26**(9), 2201–2205, (1985).

[16] H. Ikuno and K. Yasuura. "Improved Point-Matching Method with application to scattering from a periodic surface". IEEE Trans., **AP-21**(5), 657–662, (1973).

[17] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in Pascal*. Cambridge University Press, New York, 1992.

[18] J. Chandezon, D. Maystre, and G. Raoult. "A new theoretical method for diffraction gratings and its numerical applications". J. Optics, **11**, 235–241, (1980).

[19] G. B. Burckhardt. "Diffraction of a plane wave at a sinusoidally stratified dielectric grating". J. Opt. Soc. Am., **56**(11), 1502–1509, (1966).

[20] F. G. Kaspar. "Diffraction by thick, periodically stratified gratings with complex dielectric constant". J. Opt. Soc. Am., **63**(1), 37–45, (1973).

[21] K. Knop. "Rigorous diffraction theory for transmission gratings with deep rectangular grooves". Journal of the Optical Society of America, **68**, 1206–1210, (1978).

[22] M. G. Moharam and T. K. Gaylord. "Diffraction analysis of dielectric surface-relief gratings". J. Opt. Soc. Am., **72**(10), 1385–1392, (1982).

[23] J. Turunen. "Diffraction theory of microrelief gratings". In H. P. Herzig, editor, *Micro-optics*. Taylor & Francis, Inc, 1997.

[24] R. H. Morf. "Exponentially convergent and numerically efficient solution of Maxwell's equations for lamellar gratings". J. Opt. Soc. Am. A, **12**(5), 1043–1056, (1995).

[25] L. Li. "New formulation of the Fourier modal method for crossed surface-relief gratings". J. Opt. Soc. Am. A, **14**(10), 2758–2767, (1997).

[26] L. Li. "Reformulation of the Fourier modal method for surface-relief gratings made with anisotropic materials". J. Mod. Opt., **45**(7), 1313–1334, (1998).

[27] M. G. Moharam, E. B. Grann, and D. A. Pommet. "Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings". J. Opt. Soc. Am. A, **12**(5), 1068ff, (1995).

[28] P. Lalanne. "Improved formulation of the coupled-wave method for two-dimensional gratings". J. Opt. Soc. Am. A, **14**(7), 1592-1598, (1997).

[29] P. Lalanne. "Convergence performance of the coupled-wave and the differential methods for thin gratings". J. Opt. Soc. Am. A, **14**(7), 1583-1591, (1997).

[30] Lifeng Li. "Use of Fourier series in the analysis of discontinuous periodic structures". J. Opt. Soc. Am. A, **13**(9), 1870-1876, (1996).

[31] R. Dändliker. "Coupled waves: a powerful concept in modern optics". In *Fifth international topical meeting on education and training in optics*, SPIE 3130, pages 279-288, Delft, The Netherlands, 1997.

[32] H. Kogelnik. "Coupled wave theory for thick hologram gratings". Bell Sys. Techn. J., **48**(9), 2909-2947, (1969).

[33] P. C. Waterman. "Scattering by periodic surfaces". J. Acoust. Soc. Am., **57**, 791-802, (1975).

[34] D. Maystre. "Rigorous vector theories of diffraction gratings". In E. Wolf, editor, *Progress in Optics*, volume XXI, pages 1-64. Elsevier, Amsterdam, 1984.

[35] M. Bagieu and D. Maystre. "Waterman and Rayleigh methods for diffraction grating problems: extension of the convergence domain". J. Opt. Soc. Am. A, **15**(6), 1566-1576, (1998).

[36] E. G. Loewen and E. Popov. *Diffraction gratings and applications*. Marcel Dekker, Inc, New York, 1997.

[37] H. Haidner, P. Kipfer, W. Stork, and N. Streibl. "Zero-order gratings used as an artifical distributed index medium". Optik, **89**(3), 107-112, (1992).

[38] D. H. Raguin and G. M. Morris. "Antireflection structured surfaces for the infrared spectral region". Appl. Opt., **32**(7), 1154-1167, (1993).

[39] Ch.W. Haggans, L. Li, T. Fujita, and R. K. Kostuk. "Lamellar gratings as polarization components for specularly reflected beams". J. Mod. Opt., **40**(4), 675-686, (1993).

[40] P. Yeh, A. Yariv, and C. Hong. "Electromagnetic propagation in periodic stratified media. I General theory". J. Opt. Soc. Am., **67**(4), 423–438, (1977).

[41] M. Kuittinen, J. Turunen, and P. Vahimaa. "Subwavelength-structured elements". In J. Turunen and F. Wyrowski, editors, *Diffractive optics for industrial applications*, pages 303–323. Akademie Verlag GmbH, Berlin, 1997.

[42] M. Nieto-Vesperinas. *Scattering and diffraction in physical optics*. John Wiley and Sons, Inc., New York, 1991.

[43] J. W. Goodman. *Introduction to Fourier optics*. McGraw-Hill, New York, 1996.

[44] B. E. A. Saleh and M. C. Teich. *Fundamentals of photonics*. John Wiley & Sons, Inc., New York, 1991.

[45] J. J. Stamnes. *Waves in focal regions*. Adam Hilger, Bristol, 1986.

[46] D. W. Pohl, W. Denk, and M. Lanz. "Optical stethoscopy: Image recording with resolution $\lambda/20$". Appl. Phys. Lett., **44**(7), 651–653, (1984).

[47] U. Duerig, D. W. Pohl, and F. Rohner. "Near-field optical-scanning microscopy". J. Appl. Phys., **59**(20), 3318–3327, (1986).

[48] M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, volume AMS-55. National Bureau of Standards, Washington,DC, 1972.

[49] P. Blattner, H. P. Herzig, and R. Dändliker. "Scanning near-field optical microscopy: transfer function and resolution limit". Opt. Comm., **155**, 245–250, (1998).

[50] P. Blattner and H. P. Herzig. "Rigorous diffraction theory applied to microlenses". J. Mod. Opt., **45**(7), 1395–1403, (1998).

[51] J. F. Nye and M. V. Berry. "Dislocations in wave trains". Proc. R. Soc. London Ser. A, **63**, 165, (1974).

[52] V. Basistiy, M. S. Soskin, and M. V. Vasnetsov. "Optical wavefront dislocations and their properties". Opt. Comm., **119**, 604–612, (1995).

[53] J. F. Nye. "Unfolding of higher-order wave dislocations". J. Opt. Soc. Am. A, **15**(5), 1132–1138, (1998).

[54] M. V. Berry. "Wave dislocation reactions in nonparaxial Gaussian beams". J. Mod. Opt., **45**(9), 1845–1848, (1998).

[55] L. V. Kreminskaya, M. S. Soskin, and A. I. Khizhnyak. "The Gaussian lenses give birth to optical vortices in laser beams". Opt. Comm., **145**, 377–384, (1998).

[56] H. P. Herzig, editor. *Micro-optics*. Taylor & Francis Ltd, London, 1997.

[57] J. Turunen and F. Wyrowski, editors. *Diffractive optics for industrial and commercial applications*. Akademie Verlag GmbH, Berlin, 1997.

[58] F. Wyrowski. "Digital phase-encoded inverse filter for optical pattern recognition". Appl. Opt., **30**(32), 4650–4657, (1991).

[59] M. T. Eismann, A. M. Tai, and J. N. Cederquist. "Iterative design of a holographic beamformer". Appl. Opt., **28**(13), 2641, (1989).

[60] P. Ehbets, H. P. Herzig, R. Dändliker, P. Regnault, and I. Kjelberg. "Beam shaping of high-power laser diode arrays by continuous surface-relief elements". J. Mod. Opt., **40**(4), 635–645, (1993).

[61] Y. Ono, Y. Kimura, Y. Ohta, and N. Nishida. "Antireflection effect in ultrahigh spatial-frequency holographic relief gratings". Appl. Opt., **26**(6), 1142–1146, (1987).

[62] T. Stone and N. George. "Hybrid diffractive-refractive lenses and achromats". Appl. Opt., **27**(14), 2960–2971, (1988).

[63] R. H. Czichy. *Hybrid optics for space applications*, volume SP-1158. ESA Publication Division, Noordwijk, The Netherlands, 1993.

[64] G. P. Behrmann and J. P. Bowen. "Influence of temperature on diffractive lens performance". Appl. Opt., **32**(14), 2483–2489, (1993).

[65] C. Londoño, W. T. Plummer, and P. P. Clark. "Athermalization of a single component lens". Appl. Opt., **32**(13), 2295–2302, (1993).

[66] P. Blattner, H. P. Herzig, K. J. Weible, J. M. Teijido, H. J. Heimbeck, E. Langenbach, and J. Rogers. "Diffractive optics for compact space communication terminals". J. Mod. Opt., **43**(7), 1473–1484, (1996).

[67] R. Völkel, H. P. Herzig, Ph. Nussbaum, P. Blattner, R. Dändliker, E. Cullmann, and W. B. Hugle. "Microlens lithography and smart masks". Microelectronic Engineering, **35**, 513–516, (1997).

[68] R. Völkel, H. P. Herzig, and P. Blattner. "Smart Masks", 1997.

[69] N. Garcia and M. Nieto-Vesperinas. "Near-field optics inverse-scattering reconstruction of reflective surfaces". Opt. Lett., **18**(24), 2090–2092, (1993).

[70] N. Garcia and M. Nieto-Vesperinas. "Direct solution to the inverse scattering problem for surfaces from near-field intensities without phase retrieval". Opt. Lett., **20**(9), 949–951, (1995).

[71] S. S. H. Naqvi, S. Gaspar, K. Hickman, K. Bishop, and J. R. McNeil. "Linewidth measurement of gratings on photomasks: a simple technique". Appl. Opt., **31**(10), 1377–1384, (1992).

[72] S. S. H. Naqvi, R. H. Krukar, J. R. McNeil, J. E. Franke, T. M. Niemczyk, D. M. Haaland, R. A. Gottscho, and A. Kornblit. "Etch depth estimation of large-period silicon gratings with multivariate calibration of rigorously simulated diffraction profiles". J. Opt. Soc. Am. A, **11**(9), 2485–2493, (1994).

[73] Ch. J. Raymond, M. R. Murnane, S. L. Prins, S. S. H. Naqvi, and J. E. McNeil. "Multiparameter grating metrology using scatterometry". J. Vac. Sci. Technol. B, **15**(2), 361–368, (1997).

[74] S. A. Coulombe, B. K. Minhas, C. J. Raymond, S. S. H. Naqvi, and J. R. McNeil. "Scatterometry measurement of sub-0.1 μm linewidth gratings". J. Vac. Sci. Technol. B, **16**(1), 80–87, (1998).

[75] E. Betzig, J. K. Trautman, T.D. Harris, and J. S. Weiner. "Breaking the diffraction barrier: optical microscopy on a nanometric scale". Science, **251**, 1468, (1991).

[76] M. Postek and D. Joy. "Submicrometer microelectronics dimensional metrology: Scanning electron microscopy". NBS J. of Res., **92**(3), 205–228, (1987).

[77] J. E. Griffith and D. A. Grigg. "Dimensional metrology with scanning probe microscopes". J. Appl. Phys., **74**(9), R83–R109, (1993).

[78] P. Blattner, H. P. Herzig, and S. S. H. Naqvi. "Scanning spot metrology for testing of photolithographic masks". Opt. Eng., **23**, 2425–2427, (1995).

[79] P. Blattner and H. P. Herzig. "Optical testing of fine grating structures". In *Optical Inspection and Micromeasurements*, SPIE 2782, pages 628–634, Besancon, 1996.

[80] P. Blattner, S. S. H. Naqvi, H. P. Herzig, and P. Ehbets. "Diffractive structures for testing nano-meter technology". Microelectr. Eng., **27**, 543–546, (1995).

[81] Lord Rayleigh. "On the theory of optical images, with special reference to the microscope". Phil. Mag., **54**(5), 167–195, (1896).

[82] G. ToraldoDiFrancia. "Resolving power and information". J. Opt. Soc. Am., **45**(7), 497–501, (1955).

[83] G. ToraldoDiFrancia. "Directivity, super-gain and information". Trans. IRE, **AP-4**, 473–478, (1956).

[84] G. ToraldoDiFrancia. "Degrees of freedom of an image". J. Opt. Soc. Am., **59**(7), 799–805, (1969).

[85] F. Gori, S. Paolucci, and L. Ronchi. "Degrees of freedom of an optical image in coherent illumination, in the presence of aberrations". J. Opt. So. Am., **65**(5), 495–501, (1975).

[86] D. Slepian and H. O. Pollack. "Prolate spheroidal wave functions, Fourier analysis and uncertainly - I". Bell System Tech. J., **40**, 43–63, (1961).

[87] H. J. Landau and H. O. Pollack. "Prolate spheroidal wave functions, Fourier analysis and uncertainly - II". Bell System Tech. J., **40**, 65–84, (1961).

[88] D. Slepian and E. Sonnenblick. "Eigenvalues associated with Prolate Spheroidal Wave Functions of zero order". Bell System Tech. J., **44**, 1745–1759, (1965).

[89] V. Tychinsky. "Wavefront dislocations and registering images inside the Airy disk". Opt. Comm., **81**, 131–137, (1991).

[90] V. P. Tychinsky and C. H. F. Velzel. "Super-resolution in microscopy". In J. C. Dainty, editor, *Current trends in optics*, pages 255–268. Academic Press, London, 1994.

[91] M. Totzeck and H. J. Tiziani. "Phase-singularities in 2D diffraction fields and interference microscopy". Opt. Comm., **138**, 365–382, (1997).

[92] M. Totzeck and H. J. Tiziani. "Interference microscopy of sub-lambda structures: A rigorous computation method and measurements". Opt. Comm., **136**, 61–74, (1997).

[93] M. W. Kowarz. "Homogeneous and evanescent contributions in scalar near-field diffraction". Appl. Opt., **34**(7), 3055–3063, (1995).

Appendix A

Publications

[A.I] P. Blattner, H. P. Herzig, and R. Dändliker. "Scanning near-field optical microscopy: transfer function and resolution limit". Opt. Comm., **155**, 245-250, (1998).

[A.II] P. Blattner and H. P. Herzig. "Rigorous diffraction theory applied to microlenses". J. Mod. Opt., **45**(7), 1395-1403, (1998).

[A.III] P. Blattner, H. P. Herzig, K. J. Weible, J. M. Teijido, H. J. Heimbeck, E. Langenbach, and J. Rogers. "Diffractive optics for compact space communication terminals". J. Mod. Opt., **43**(7), 1473-1484, (1996).

[A.IV] P. Blattner, H. P. Herzig, and S. S. H. Naqvi. "Scanning spot metrology for testing of photolithographic masks". Opt. Eng., **23**, 2425-2427, (1995).

[A.V] P. Blattner, S. S. H. Naqvi, H. P. Herzig, and P. Ehbets. "Diffractive structures for testing nano-meter technology". Microelectr. Eng., **27**, 543-546, (1995).

ELSEVIER

Optics Communications 155 (1998) 245–250

OPTICS COMMUNICATIONS

Scanning near-field optical microscopy: transfer function and resolution limit

P. Blattner [1], H.P. Herzig, R. Dändliker

Institute of Microtechnology, University of Neuchâtel, Rue A.-L. Breguet 2, 2000 Neuchâtel, Switzerland

Received 8 September 1997; revised 8 June 1998; accepted 14 July 1998

Abstract

We present scanning near-field optical microscopy as an optical instrument characterized by a transfer function. This approach gives some theoretical guidelines for the design of near-field optical measurement systems. We emphasize that it is important to distinguish between the resolution for the optical field and the resolution for the object. In addition, to solve the general inverse diffraction problem the measurement of phase and amplitude of the electromagnetic field is necessary.

Keywords: Scanning near-field optical microscope; Rigorous diffraction theory; Sub-wavelength resolution

1. Introduction

Since the introduction of near-field optical microscopy, many different configurations of near-field measurements have been proposed [1]. Parallel to the technological advances, several efforts have been made in the theoretical modelling of near-field optical detection systems. A recent overview of near-field diffraction theories has been published by Girard et al. [2]. The paper includes different scattering theories, and theories based on the plane wave, respective multipole expansions of the electromagnetic field. The basic question to be answered by any theory is the resolution limit of near-field optical microscopy. Discussions about the resolution of near-field optical microscopes can be found in different papers [3–6]. Vigoureux [6] presents a historical overview of the resolution limit of classical microscopy (mainly the Rayleigh criterion), and analyzes the consequences for near-field microscopy. The main conclusion (which is in the meantime well known) is that information on sub-wavelength details of electromagnetic fields is contained in the non-radiative (or evanescent) components of the field. Therefore, it is important to collect the near-field information.

In our present paper, the near-field detection system is considered as an optical instrument characterized by a transfer function. For this purpose, the measurement system is separated into different parts as shown in Fig. 1.

The object is illuminated by an incident wave and generates a field distribution in the output plane. In general, the interaction of light with sub-wavelength structures has to be computed rigorously. We define a spectral response function H_{obj} as being the action of the object on the incident wave of a certain *spatial* frequency. The resulting field is detected by the scanning probe at a constant distance from the sample. The radiating part of the field propagates in free space, whereas the evanescent part is attenuated. This effect is described by the free space transfer function P. The probe itself is again characterized by a spectral response function H_{tip}. This approach can of course also be applied to the inverse case, where the tip is used to illuminate the structure.

Compared to a classical optical instrument, the term *resolution* has to be defined differently in near field optics.

[1] E-mail: peter.blattner@imt.unine.ch

0030-4018/98/$ - see front matter
PII: S0030-4018(98)00401-5

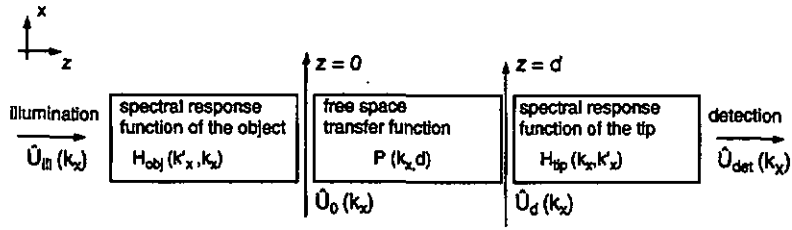


Fig. 1. The scanning near-field optical microscope is considered as an optical instrument characterized by spectral response and transfer functions.

In classical optics, the field is directly related to the object. Thus, the resolution for the field and the resolution for the object are the same. In near-field optics, due to the complicated interaction problem, the resolution is related to the ability to reconstruct the electromagnetic field in the output plane (or output space) of the object. In consequence, sub-micrometer resolution of a measured field does not imply that the object can be determined with the same accuracy.

We limit the analysis to non-elastic interaction, 2D geometry, and to TE-polarization. Furthermore, no back reflections from the tip are taken into account. This means that the theory presented here is not a self-consistent approach. However, it allows us to discuss the field resolution problem of near-field optical microscopy in a comprehensive way. Furthermore, the reconstruction of the object from the field data is not treated in this paper.

2. Spectral response function of the object

Most of the near-field papers describe the interaction problem by a macroscopic perturbation theory [7,8]. This approach writes the field as a superposition of the zero order field plus a first order perturbation term. This perturbation is proportional to the incident field and to the Fourier transform of the surface profile. However, the perturbation theory is limited to smooth surfaces, i.e. the surface roughness is typically small compared to the wave-

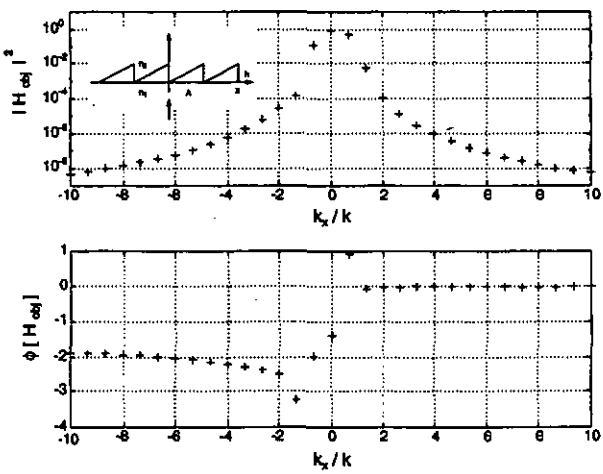


Fig. 2. Spectral reponse of a blazed grating structure (intensity and phase) in the output plane of the element for a perpendicular illumination. The spectrum has been calculated with the rigorous eigenmode method [12]. The parameters of the setup are $n_1 = 1.5$, $n_2 = 1$, grating period $\Lambda = 1.5\lambda$, height $h = \lambda$, and TE-polarization.

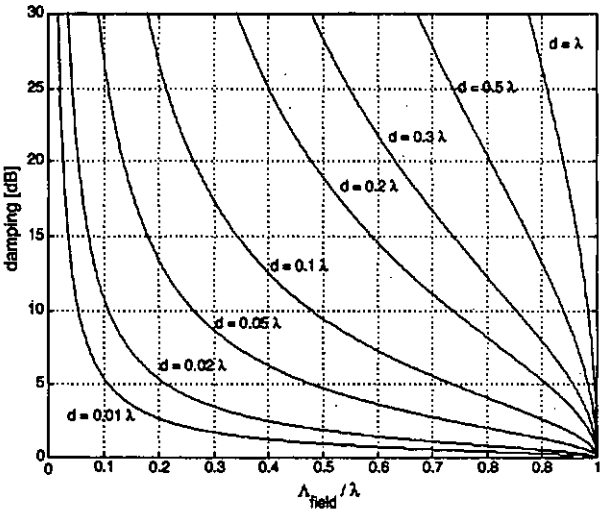


Fig. 3. Damping of evanescent waves versus the field size Λ/λ for different scanning distances.

length. For deep structures (e.g. typical surface relief gratings) this approach is not valid anymore and the solution of the interaction problem has to rigorously fulfil Maxwell's equations. In the case of periodic objects, most of the theories are based on the differential representation of the inhomogeneous wave function [9–12]. These approaches directly compute the complex amplitudes of the diffracted (propagating and evanescent) waves. Hence, it is possible to compute the light distribution in the near-field of any periodic structure [13].

For linear materials and coherent illumination, it is possible to define a spectral response function $H_{obj}(k_x', k_x)$ which relates the output spatial frequency spectrum $\hat{U}_0(k_x)$ to the incident excitation $\hat{U}_{ill}(k_x')$ by a linear superposition

$$\hat{U}_0(k_x) = \int_{-\infty}^{\infty} \hat{U}_{ill}(k_x') H_{obj}(k_x', k_x)\, dk_x'. \qquad (1)$$

This function describes the action of the object on an incident wave. In its most general form the entire spectrum is taken into account, i.e. $-\infty < k_x, k_x' < \infty$. In the case of infinite plane wave illumination ($\hat{U}_{ill}(k_x') = A_0\, \delta(k_x' - k_{x0})$), Eq. (1) becomes

$$\hat{U}_0(k_x) = A_0 H_{obj}(k_x, k_{x0}). \qquad (2)$$

Eq. (2) is valid for the classical scanning near-field optical microscope (SNOM) and the photon tunneling microscope (PSTM/STOM). Fig. 2 shows the calculated plane wave spectrum of a blazed grating structure in a SNOM configuration.

The calculations were performed with the rigorous eigenmode method [12]. In general, the spectral response is a complex function including an amplitude and a phase distribution. It follows that, similar to the holographic reconstruction process, the amplitude *and* the phase have to be measured to reconstruct the original object field [2]. Different approaches for measuring the phase by near-field interferometry have already been presented in the open literature [16–20].

3. Free space transfer function

The radiating part of the generated object field is propagating in free space, whereas the evanescent part is attenuated. In the plane wave spectrum this transfer can be written as a filter function

$$U_d(k_x) = P(k_x, d) U_0(k_x), \qquad (3)$$

where $P(k_x, d)$ is the free space transfer function and d is the distance from the tip to the sample. In particular, the complex free space transfer function is given by

$$P(k_x, d) = \begin{cases} \exp\left[i d \sqrt{k^2 - k_x^2} \right] & \text{for } |k_x| < k, \\ \exp\left[-d \sqrt{k_x^2 - k^2} \right] & \text{for } |k_x| \geq k, \end{cases} \qquad (4)$$

where $k = 2\pi/\lambda$ and λ is the free space wavelength. Thus, evanescent waves are affected by an amplitude

[2] In holography, the original wave is reconstructed from the recorded phase and amplitude. Note that, for the case of weak perturbations, it is possible to reconstruct the surface profile without the necessity for phase retrieval [14,15].

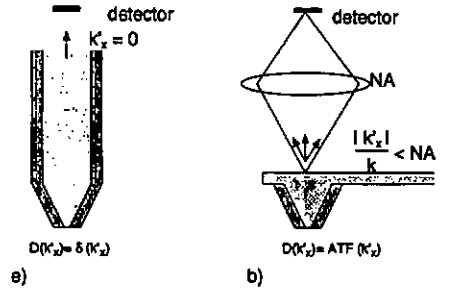


Fig. 4. Acceptance function $D(k'_x)$ for different geometries. (a) Fiber tip detection, (b) SNOM based on a cantilever tip.

change, propagating waves by a phase change. The local sub-wavelength field information is contained in the evanescent waves. The exponential damping of this evanescent waves has therefore some crucial consequences for the scanning tip distance. The damping factor is shown in Fig. 3 as a function of the field period Λ_{field} for different scanning distances. It is obvious that the scanning distance should be as small as possible.

In the complete system the damping effect of the higher diffraction orders is even increased by the low-pass filter characteristic of the probe tip.

4. Spectral response of the tip

The detection is made by scanning a tip over the sample. If the property of the detection system does not change during the scan the linear system is space invariant. Therefore, the detected signal can be expressed as a convolution

$$U_{det}(x) = \int_{-\infty}^{\infty} t(x-x')\, U_d(x')\, \mathrm{d}x' = t_{tip} * U_d\,, \qquad (5)$$

where $t(x-x')$ is the impulse response of the detector. In frequency space Eq. (5) becomes a multiplication

$$\hat{U}_{det}(k_x) = T(k_x)\hat{U}_d(k_x)\,, \qquad (6)$$

where $T(k_x)$ is the transfer function of the detector. For linear systems the transfer function is expressed as a linear superposition integral,

$$T(k_x) = \int_{-\infty}^{\infty} H_{tip}(k_x, k'_x)\, D(k'_x)\, \mathrm{d}k'_x\,, \qquad (7)$$

where $H_{tip}(k_x, k'_x)$ is the spectral response of the tip and $D(k'_x)$ the detector acceptance function. The acceptance

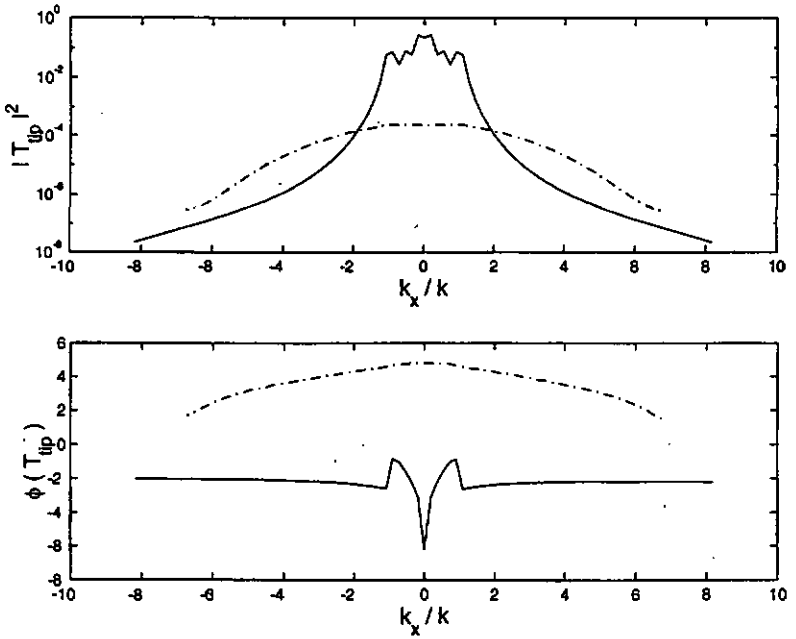


Fig. 5. Calculated intensity and phase of the complex spectral response function of a one-dimensional dielectric probe tip: uncoated (solid line) and coated (dashed line). The coated tip has a clear slit aperture of $\lambda/10$.

function of the tip depends on the chosen geometry. For a classical single-mode fiber tip SNOM (Fig. 4a) only one frequency is accepted, $D(k'_x) = \delta(k'_x)$ and the transfer function becomes particularly simple,

$$T(k_x) = H_{\mathrm{tip}}(k_x, 0). \tag{8}$$

In the case of a cantilever tip based SNOM, the acceptance function of the detector is equal to the *coherent amplitude transfer function* (ATF) [21] of the detection system. The different frequencies are now integrated over the numerical aperture of the system. For an ideal imaging system the coherent amplitude transfer function is a binary function, Eq. (7) can be rewritten as

$$T(k_x) = \int_{-k\,\mathrm{NA}}^{k\,\mathrm{NA}} H_{\mathrm{tip}}(k_x, k'_x)\,\mathrm{d}k'_x \tag{9}$$

where NA is the numerical aperture of the imaging system. The aim of the tip-detector is to provide sub-wavelength resolution. For this purpose, the geometrical dimensions of the tip have to be in the sub-wavelength range. Thus, the spectral response $H_{\mathrm{tip}}(k_x, k'_x)$ also has to be calculated with a rigorous model. In Fig. 5 the spectral response of a coated and uncoated dielectric fiber tip is shown. The calculations have been made by the same algorithm as for the grating structure, by considering periodically arranged tips. The uncoated tips have higher coupling efficiencies for low spatial frequencies than the coated tips. On the other hand, the electromagnetic field is less confined, i.e. the angular spectrum is more narrow.

5. Field reconstruction, resolution

In its simplest configuration (plane wave illumination, fiber tip detection) the spectrum of the detected signal is obtained by multiplying the spectral response of the object field $H_{\mathrm{obj}}(0, k_x)$ with the free space transfer functions $P(k_x, d)$ (Eq. (4)) and spectral response of the tip $H_{\mathrm{tip}}(k_x, 0)$ (Eq. (6)),

$$\hat{U}_{\mathrm{det}}(k_x) = \hat{H}_{\mathrm{tip}}(k_x, 0)\, P(k_x, d)\, \hat{U}_0(k_x), \tag{10}$$

$$\hat{U}_0(k_x) = \hat{U}_{\mathrm{in}}(k_x)\, \hat{H}_{\mathrm{tip}}(0, k_x). \tag{11}$$

Inversely, it is theoretically possible to determine the object field by deconvoluting the measured field with the impulse response of the system. For this purpose, the transfer function of the probe and the distance have to be known exactly. One the other hand, one should be aware that the process of deconvolution is in general quite sensitive to noise. A real measurement system is characterized by a certain signal to noise ratio (SNR). If the system is shot-noise limited, the SNR is proportional to the *power spectrum density* and the total number of collected photoelectrons $\bar{n}_e$ during the scan (total integrating time). Hence

$$\mathrm{SNR}(k_x) = \frac{|\hat{U}_{\mathrm{det}}(k_x)|^2}{\sum_m |\hat{U}_{\mathrm{det}}(k_{x,m})|^2}\, \bar{n}_e, \tag{12}$$

where $\hat{U}_{\mathrm{det}}(k_{x,m})$ is the sampled spectrum of the field.

As a numerical example, we consider the case of resolving a field variation at the object surface of 50 nm at a wavelength of 500 nm. For a scanning distance of 10 nm, the field is attenuated by 10 dB (Fig. 3). At a scanning distance of $d = 25$ nm the field is already damped by 25 dB. The low-pass filter characteristic of a coated tip, having an aperture of 50 nm, adds another 50 dB to the attenuation with respect to the central frequency. The situation is even worse for the uncoated tip (see Fig. 5), namely 80 dB instead of 50 dB. In this numerical examples, the signals are mainly decreased by the limited resolution of the tips. Considering a resolution of 100 nm, the response of the tips is much better. The coated tip attenuates the signal by 15 dB, the uncated by 60 dB. In practice, it is necessary to determine the spectral response of the tip experimentally.

6. Conclusions

The near field microscope can be considered as an optical instrument defined by a transfer function. In this representation, the limitation of the microscope can be clearly discussed, although backreflections from the tip are not taken into account. Sub-wavelength signals have to overcome the exponential attenuation due to propagation of the evanescent waves (Fig. 3) and the low-pass filter characteristic of the probe tip (Fig. 5). The spectral response of the tip expresses the capability of conversing evanescent to propagating waves. The system is characterized by a signal to noise ratio which depends on the spatial frequency. The signals have to be above the noise level of the instrument. This limitations are fundamental for any type of near-field microscopy. Furthermore, it is important to realize that in a general case, similar to holography, amplitude *and* phase measurements are needed in order to solve the reconstruction problem. We emphasize that there is a difference between the object resolution and the object *field* resolution. The resolution limit usually discussed in the open literature is the object *field* resolution. The statement how good the object is resolved has lost its classical meaning in sub-wavelength optics. The object reconstruction itself (its geometry and its dielectric behavior) needs some a priori information about the object, the detection system and the illumination.

Acknowledgements

This work was partially supported by the Swiss National Science Foundation and the Priority Program Optique II.

References

[1] D.W. Pohl, D. Courjon, Near Field Optics, Kluwer, Dordrecht, 1993.
[2] C. Girard, A. Dereux, Rep. Prog. Phys. 59 (1996) 657.
[3] J.M. Vigoureux, F. Depasse, C. Girard, Appl. Optics 31 (1992) 3036.
[4] M.W. Kowarz, Appl. Optics 34 (1995) 3055.
[5] M. Xiao, Optics Comm. 132 (1996) 403.
[6] J.M. Vigoureux, D. Courjon, Appl. Optics 31 (1992) 3170.
[7] D. Van Labeke, D. Barchiesi, J. Opt. Soc. Am. A 9 (1992) 732.
[8] D. Barchiesi, D. Van Labeke, Ultramicroscopy 57 (1995) 196.
[9] M.G. Moharam, T.K. Gaylord, J. Opt. Soc. Am. 72 (1982) 1385.
[10] L. Li, J. Mod. Optics 40 (1993) 553.
[11] R.H. Morf, J. Opt. Soc. Am. A 12 (1995) 1043.
[12] J. Turunen, Diffraction theory of microrelief gratings, in: H.P. Herzig (Ed.), Micro-optics, Taylor & Francis, 1997.
[13] P. Blattner, H.P. Herzig, Rigorous diffraction theory applied to microlens arrays, in: J. Turunen, P. Wyrowski (Eds.), EOS Topical Meeting on Diffractive Optics, Savonlinna, 1997, pp. 36–38.
[14] N. Garcia, M. Nieto-Vesperinas, Optics Lett. 20 (1995) 949.
[15] J.J. Greffet, A. Sentenac, R. Carminati, Optics Comm. 116 (1995) 20.
[16] C. Bainier, D. Courjon, F. Baida, C. Girard, J. Opt. Soc. Am. A 13 (1996) 267.
[17] H. Masuda, Y. Suzui, A. Takayanagi, N. Umeda, Development of heterodyne photon scanning tunneling microscope, in: International Workshop on Interferometry, Satellite Meeting for ICO-17, Wako, 1996, pp. 141, 142.
[18] S. Pilevar, W.A. Atia, C.C. Davis, Reflection near-field scanning optical microscopy: an interferometric approach, in: Near Field Optics-3, BRNO, Czech Republic, 1995, EOS Topical Meetings Digest, Vol. 8, pp. 69, 70.
[19] M. Vaez-Iravani, R. Toledo-Crow, Appl. Phys. Lett. 62 (1993) 1044.
[20] D. Charraut, C. Bainier, D. Courjon, C. Girard, Pure Appl. Optics 6 (1997) 491.
[21] J.W. Goodman, Introduction to Fourier Optics, McGraw-Hill, New York, 1996.

JOURNAL OF MODERN OPTICS, 1998, VOL. 45, NO. 7, 1395–1403

Rigorous diffraction theory applied to microlenses

P. BLATTNER† and H. P. HERZIG

Institute of Microtechnology, University of Neuchâtel, Rue Breguet 2, 2000 Neuchâtel, Switzerland

(Received 4 November 1997; revision received 13 January 1998)

Abstract. In this paper, we discuss the behaviour of small cylindrical microlenses, arranged in one-dimensional arrays and as single elements. For this purpose, we apply a standard rigorous diffraction theory, commonly used for diffraction gratings. We investigate the coupling effect between the elements. It turns out that single elements behave like periodic elements if the spacing is chosen correctly. Furthermore, we compute the complex transmission function by rigorous diffraction theory and compare them with classical theories (combined ray tracing and the thin-element approach). Finally, we discuss the focal properties of microlenses in the rigorous regime.

1. Introduction

Today's technology allows the realization of small microlenses which have geometrical dimensions of the order of a few wavelengths (figure 1). These lenses are typically realized by the reflow technique [1] and have perfect spherical surfaces. In standard lens design, mainly two approaches are applied to describe the behaviour of such elements. One method is ray tracing and the other is the thin-element approach. In its simplest version, ray tracing applies the laws of refraction and reflection to individual rays. No diffraction effects are taken into account. However, it is possible to include the optical path length through the optical elements (phase-sensitive ray tracing). This allows one to determine the field in the plane of the exit pupil of an optical system. In a second step, the free-space propagation of the field can be computed by diffraction theory (angular spectrum, Rayleigh–Sommerfeld or Fresnel). This method is referred to as *combined* ray tracing [2]. In the thin-element approach, the plane of the entrance pupil and the plane of the exit pupil coincide. In this case, the element alters only the phase of the incident wave, that is the wave is delayed proportionally to the surface relief of the element.

The classical theories lose their validity if the geometrical dimensions of the element are of the order of the wavelength. For this purpose, different rigorous diffraction theories have been investigated in the past. They are based on either the integral or the differential representation of the wave equation. In the literature, only a few examples are presented for the rigorous computation of microlenses. Most of them are valid for diffractive microlenses [3, 4]. Rigorous computation of refractive microlenses based on intergral boundary methods and applied to have been presented by Tabbara [5] and more recently by Wang and Prata [6].

† email: peter.blattner@imt.unine.ch

Figure 1. Scanning electron microscopy picture of an array of very small microlenses (diameter, about 3 μm) [1].

However, these calculations are cumbersome and have been restricted to single elements. On the other hand, there exist efficient rigorous theories valid for periodic structures only [7–9]. They are mainly used to predict the diffraction efficiency of high-frequency gratings.

In this paper, we apply the rigorous eigenmode method to microlens arrays [10]. We restrict the analysis to arrays of continuous relief lenses illuminated with a plane wave (two dimensions and transverse electric polarization). Furthermore, we analyse the diffraction at single elements by introducing a large spacing between the elements and we discuss the coupling effects between the elements. We compare the transfer function of microlenses calculated by rigorous diffraction theory and two classical theories (combined ray tracing and the thin-element approach). Finally, we investigate the focusing properties of microlenses.

2. Rigorous diffraction theory

Illuminating periodically arranged elements by a plane wave results in a discrete plane-wave spectrum. Therefore, the field can be written as

$$U_T(x, z) = \sum_{m=-\infty}^{\infty} T_m \exp\left[i(k_{xm}x + k_{zm}z)\right], \tag{1}$$

where k_{xm} and k_{zm} are the projections of the associate plane wave wave-vector $\mathbf{k}_m$ onto the x and z axis respectively. The diffraction problem consists in finding the amplitudes T_m of the individual plane waves (or diffraction orders). In the rigorous eigenmode method [10] the periodically modulated surface structure or, in our case, the periodically arranged scatters are divided into a stack of thin films. The total number of layers depends on the thickness-to-wavelength ratio. We typically chose $25 \times h/\lambda$ layers, where h is the thickness of the structure and λ the wavelength. The dielectric constant distribution $\epsilon(x, z) = \epsilon_n(x, z_n)$ of each film is expanded into a Fourier series, and the electric field into eigenmode functions. Introducing these expansions in the wave equation results in an eigenvalue system that can be easily solved. The dimension of the eigenvalue system depends on the

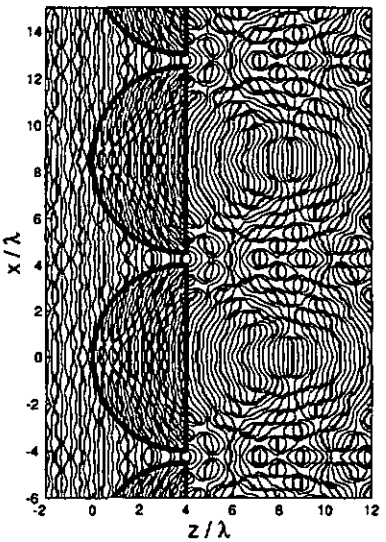


Figure 2. Calculated near-field phase contour plot of a microlens array. The elements have a diameter of 8λ, a height of 4λ and a refrative index $n = 1.5$. The array is illuminated by a TE-polarized plane wave, with perpendicular incidence.

grating period-to-wavelength ratio. It has been shown that for dielectric structures all propagating diffraction orders and some few evanescent diffraction orders (typically ten) have to be retained in the analysis [10]. Once the eigenvalues are determined, the amplitudes of the eigenmode functions are found by matching the electromagnetic boundary conditions.

With this method, it is possible to find the rigorous complex transmission function of any periodic structure, and in particular also the light distribution in the near field of microlens arrays (figure 2).

In the following, we show how the method can be applied to modelling a single microlens. There are mainly two points to consider when comparing microlens arrays with single elements. First, there are interference effects during the free-space propagation between the elements (e.g. the self-imaging effect or the Talbot effect).

Secondly, since the geometrical dimensions are small, the individual lenses couple with each other (intrinsic coupling). The coupling influence can be illustrated by comparing the rigorous transmission function of periodically arranged elements with the rigorous transmission function of a single element. Unfortunately, there exist only a few analytic (rigorous) solutions for the interaction of light with single elements. One case is the diffraction at a cylinder (figure 3). In that case, it is possible to express the field outside the cylinder by a superposition of Bessel and Hankel functions (for example [11])

$$U_{\text{single}}(r, \theta) = U_0 \sum_{n=-\infty}^{\infty} (-\mathrm{i})^n \exp(\mathrm{i}n\theta)[J_n(kr)A_n + H_n^{(2)}(kr)B_n]. \tag{2}$$

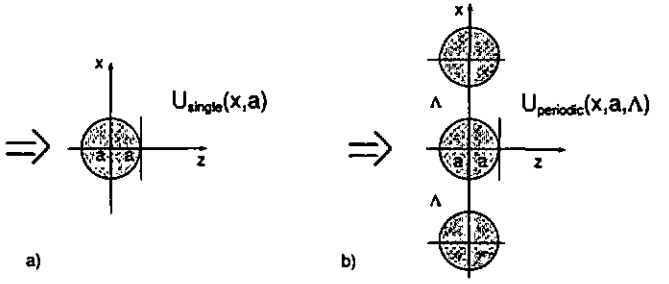


Figure 3. Configuration used to illustrate the coupling effect between elements. (*a*) The calculated field U_{single} of a cylinder illuminated with a plane wave is compared with (*b*) the field of periodically arranged cylinders $U_{periodic}$.

A_n and B_n are coefficients which depend on the diameter of the cylinder and the indices of refraction of the cylinder and the surrounding area. In parallel to that, we compute the field generated by periodically arranged cylinders with the method presented before. We denote this field by $U_{periodic}(x, y) = U_T$. The coupling parameter χ is now defined as the difference between the fields U_{single} and $U_{periodic}$ in the output plane, integrated over a fixed interval of one wavelength:

$$\chi(\Lambda) = \int_{x=-\lambda/2}^{\lambda/2} |U_{single}(x, a) - U_{periodic}(x, a, \Lambda)|^2 \, dx, \tag{3}$$

where Λ is the grating period of the microlens array. Figure 4 illustrates the coupling coefficient χ in function of the grating period Λ for different diameters of the cylinders. It it interesting to observe that a resonance occurs if the grating period is a multiple of the wavelength. Furthermore, the coupling effect is larger for small cylinders than for large cylinders. For very small structures the coupling may occur even if the elements are spaced by several wavelengths. To understand this behaviour, we compute the spectrum of (single) cylinders of different diameters. If the size of the cylinders are large (figure 5), the spectrum is mainly given by the propagating frequencies, that is $|k_x| < k = 2\pi/\lambda$. For small cylinders (figure 6), the spectrum is mainly given by the spatial frequencies around $\pm k$, that is they correspond to evanescent waves (or surface waves) of wavelength λ, also called critical evanescent waves. If the elements are arranged periodically, the spectrum becomes discrete. Depending on the grating period it is possible that the sampling falls directly on the critical frequencies $k_x = \pm k$. In this case, there is maximum coupling between the elements. In addition the diffraction pattern becomes dominated by the two surface waves (in the positive and negative directions). This is due to the large penetration depth† of the evanescent waves.

It turned out that it is possible to simulate the diffraction of a single element with the help of a grating diffraction theory, if the grating period is chosen correctly. First, the grating period should not be a multiple of the wavelength

† The penetration depth of an evanescent wave is defined as $z_{eff} = 1/\gamma$, with $\gamma = (k_x^2 - k^2)^{1/2}$. For a critical evanescent wave ($k_x = k$), z_{eff} becomes infinite!

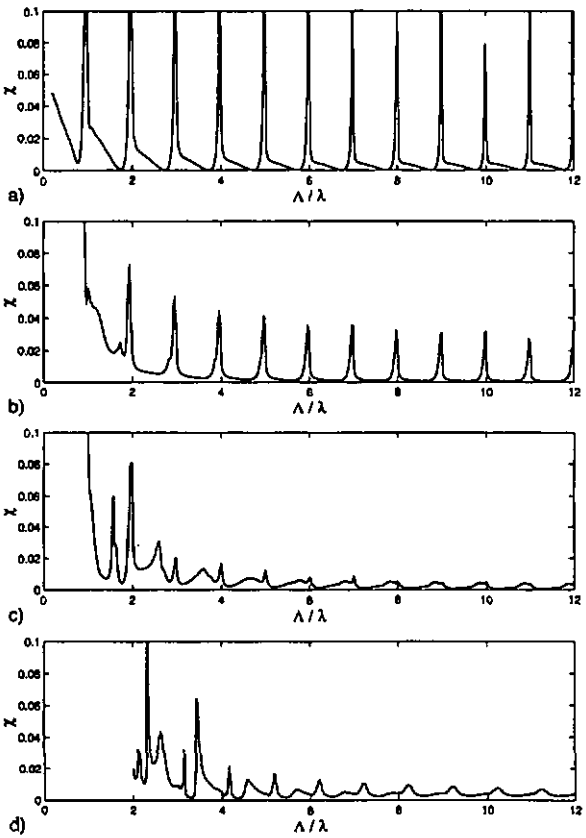


Figure 4. Coupling coefficient χ (equation (3)) against grating period Λ for different diameters 2a of the cylinders: (*a*) $2a = \lambda/5$; (*b*) $2a = \lambda/2$; (*c*) $2a = \lambda$; (*d*) $2a = 2\lambda$. The coupling coefficient shows a resonance if the grating period is a multiple of the wavelength. The smaller cylinders in (*a*) show larger coupling effects than the larger cylinders in (*b*) do.

and, second, the period should be larger than twice the diameter of the element. This result is justified by figure 4.

3. Comparison of classical and rigorous theories

In this section we compare classical and rigorous diffraction theory applied to two different microlenses: a thick ($h = 2\lambda$) and thin ($h = \lambda/5$) cylinder lens of the same diameter $2a = 4\lambda$. In this context, the terms *thick* and *thin* are referred to paraxial and non-paraxial elements, that is to small and large deflection angles respectively. In our case the deflection angles at the border of the element are 41.8°

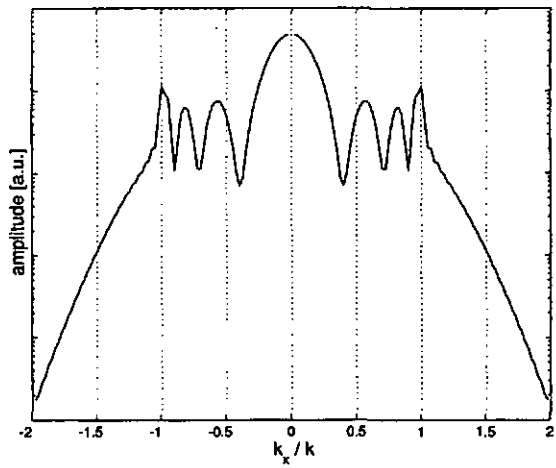


Figure 5. Spectrum of a cylinder of diameter $2a = 3\lambda$ (a.u., arbitrary units). Frequencies $|k_x|$ larger than $k = 2\pi/\lambda$ correspond to evanescent waves. The spectrum of such a large cylinder is mainly given by the propagating waves ($|k_x| < k$).

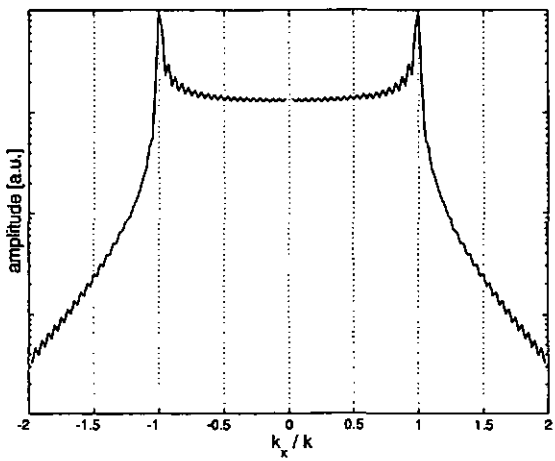


Figure 6. Spectrum of a cylinder of diameter $2a = \lambda/2$ (a.u., arbitrary units). The spectrum of such a small cylinder is mainly given by the critical evanescent waves of frequency $k_x = \pm k$.

(thick) and 7.5° (thin). It turns out that the spectrum of thin lenses (figure 7) calculated by the classical theories (combined ray tracing and the thin-element approach) agree well with the rigorous computations. In this case, the influence of the thin element on the wave front is small and the diffraction pattern is mainly

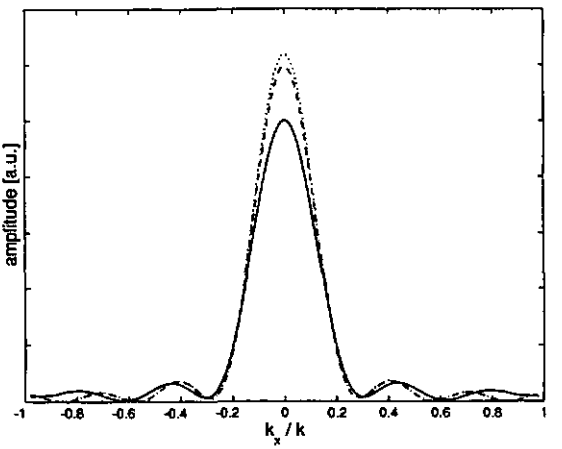


Figure 7. Calculated spectrum of a thin lens (a.u., arbitrary units): (——), rigorous computation; (– – –), ray tracing; (· · · · ·), the thin-element approach. The lens has a diameter of $2a = 4\lambda$ and a thickness of $h = \lambda/5$.

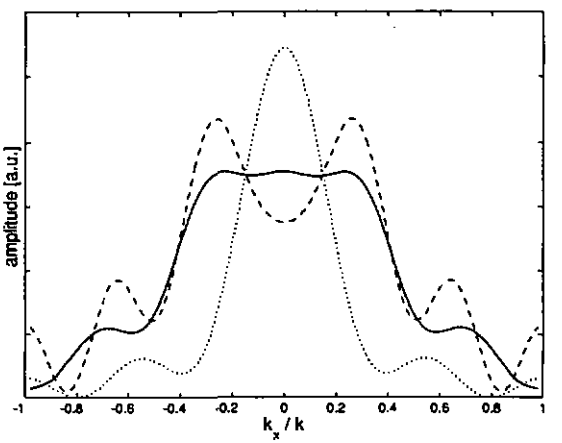


Figure 8. Calculated spectrum of a thick lens (a.u., arbitrary units): (——), the rigorous computaiton; (– – –), ray tracing; (· · · · ·), the thin-element appraoch. The lens has a diameter of $2a = 4\lambda$ and a thickness of $h = 2\lambda$.

given by the aperture, that is the spectrum has a *sinc* form. The only difference between the theories is in the absolute intensities. The thin-element approach and ray tracing show higher intensities, because there are no losses in the transmission functions (only the phase is affected).

In the case of the thick element (figure 8), the prediction made by the thin-element approach is wrong. The theory does not predict the bending of the border rays through the element. Therefore, the spectrum of the transmission function is

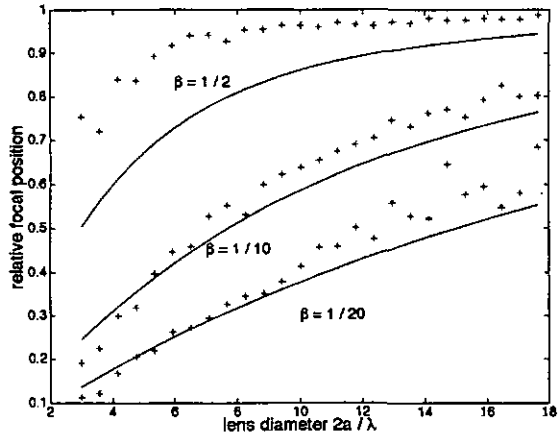


Figure 9. Illustration of the focal shift, where the focal position relative to the paraxial focal position in function of the diameter of the lens is plotted for different aspect ratios $\beta = h/2a$: (——), focal positions calculated by the thin-element approach; (+), the rigorous calculated focus positions.

too small! On the other hand, combined ray tracing shows smaller differences from the rigorous computations. One can conclude that combined ray tracing is more accurate than the thin-element approach and that the method is valid for elements with diameters down to few wavelengths.

4. Focal properties of microlenses

The thin element approach predicts a focal shift for elements having low Fresnel numbers [12]. The Fresnel number is defined as $N = a^2/\lambda f$, where f is the (paraxial) focal length of the lens. Microlenses, which have a diameter $2a$ of a few wavelengths are especially affected by this effect, because the minimum achievable focal length is $f_{\min} = a/(n-1) \approx 2a$, n being the refractive index of the lens. Thus, the maximum achievable Fresnel number is of the order of

$$N_{\max} \approx \frac{a}{2\lambda}. \tag{4}$$

This means that, owing to the small size, their properties are dominated by diffraction at the aperture. It has been shown in section 3 that the thin-element approach does not describe thick lenses sufficiently well. Hence, we expect a different behaviour of the rigorous computed focal shift than prediced by thin theory. Figure 9 illustrates the focal shift in function of the diameter of the lens for different aspect ratios β. The aspect ratio β is defined as the ratio of the lens thickness to the lens diameter, that is $\beta = h/2a$. For small aspect ratios the computations based on the thin-element approach agree well with the rigorous computations. For high aspect ratios (and as consequence also *thick* elements) there is quite a large difference between the two models. It turns out that the focal shift of thick lenses is less severe than predicted by the thin-element approach.

5. Conclusions

We have shown that it is possible to simulate single-element diffraction with the aid of a rigorous grating diffraction method by carefully choosing the grating period. Furthermore, we presented a comparison between three different methods to calculate the interaction of light with small lenses. For thin elements, all three methods predict well their behaviour, even for very small lenses. This is mainly because thin lenses are dominated by the edge diffraction which influences the focusing behaviour of the lens. We have shown that there is a good agreement between the classically predicted focal shift and rigorous computations.

For thick lenses, combined ray tracing is more precise than the thin-element approach. For such elements, the focal shift turned out to be less pronounced than predicted by the thin-element approach. In any case, it hs been illustrated that very small lenses also exhibit some focusing behaviour. This result is in a good agreement with the results presented by Wang and Prata [6].

Acknowledgment

This work has been supported by the Swiss priority program Optique II.

References

[1] Völkel, R., Herzig, H. P., Nussbaum, Ph., Blattner, P., Dändliker, R., Cullmann, E., and Hugle, W. B., 1997, *Microelectron. Engng*, **35**, 513.
[2] Stamnes, J. J., 1986, *Waves in Focal Regions* (Bristol: Adam Hilger).
[3] Hirayama, K., Glytsis, E. N., Gaylord, T. K., and Wilson, D. W., 1996, *J. opt. Soc. Am.* A, **13**, 2219.
[4] Schmitz, M., and Bryngdahl, O., 1997, *J. opt. Soc. Am.* A, **14**, 901.
[5] Tabbara, W., 1973, *J. opt. Soc. Am.*, **63**, 17.
[6] Wang, A., and Prata, A., Jr, 1995, *J. opt. Soc. Am.* A, **12**, 1161.
[7] Moharam, M. G., and Gaylord, T. K., 1982, *J. opt. Soc. Am.*, **72**, 1385.
[8] Li, L., 1993, *J. mod. Optics*, **40**, 553.
[9] Morf, R. H., 1995, *J. opt. Soc. Am.* A, **12**, 103.
[10] Turunen, J., 1997, *Micro-optics*, edited by H. P. Herzig (London: Taylor & Francis).
[11] Kozaki, S., 1982, *J. appl. Phys.*, **53**, 7195.
[12] Li, Y., and Wolf, E., 1981, *Optics Commun.*, **39**, 211.

JOURNAL OF MODERN OPTICS, 1996, VOL. 43, NO. 7, 1473–1484

Diffractive optics for compact space communication terminals

P. BLATTNER, H. P. HERZIG, K. J. WEIBLE, J. M. TEIJIDO
Institute of Microtechnology, University of Neuchâtel,
rue A.-L. Breguet 2, 2000 Neuchâtel, Switzerland

H. J. HEIMBECK, E. LANGENBACH
Fisba Optik AG, Rohrschacherstrasse 268, 9016 St Gallen, Switzerland

and J. ROGERS
Leica Heerbrugg, 9435 Heerbrugg, Switzerland

(*Received 16 October 1995; revision received 14 February 1996*)

Abstract. Free-space laser communication links with data rates between 10 and 500 Mbits s^{-1} are required to cover the large amount of communication needs between low-orbit satellites, geostationary satellites and ground stations. The objective of this paper is to demonstrate the potential of diffractive optical elements for the design of optical and optoelectronic systems for advanced laser communication terminals. Three different examples have been realized: a ring pattern generator, an athermalized and achromatic hybrid collimator system, and a hybrid beacon system.

1. Introduction

The design of first-generation free-space laser communication systems is based on laser diodes with output powers of the order of 100 mW [1]. The data rate transmission is of the order of 100 Mbits s^{-1}. This leads to terminals with large transmitter and receiver telescope diameters and, consequently, to high terminal mass and dimensions. The optical systems are usually designed with refractive lenses and reflective mirrors. Alternatives are planar diffractive optical elements (DOEs). By relying on diffraction and interference rather than on reflection and refraction, unique and novel properties can be realized. Almost any structure shape, including non-rotationally symmetric aspherics, can be manufactured, which provides all degrees of freedom for the design. Other interesting aspects of DOEs are their low weight, their strong dispersion, and the possibility of making segmented elements, large arrays of elements, beam splitters and polarizers. These properties are useful for many applications of DOEs in space, including filters for image data processing [2], beam shaping [3, 4] and antireflection structures [5, 6]. Furthermore, the combination of refractive and diffractive surfaces (hybrid elements) offers new possibilities for optical design. The negative dispersion of DOEs can be used to compensate the chromatic aberrations of refractive lenses [7, 8]. Hybrid elements can also be used to compensate the

temperature-induced variations in their mounting system. Some examples of athermalized hybrid lenses have been discussed in the literature [9, 10]. DOEs for space applications must comply with a number of requirements, including mechanical, thermal and optical stability [8]. Suitable techniques for realizing the microstructures in space-qualified materials are based on a variety of high-resolution lithographic and optical processes [11].

In this paper, we discuss possible applications for DOEs in compact optical communication terminals. The design and experimental realization of three different elements are presented. These elements include a ring pattern generator allowing tracking and data transmission within one optical element, an athermalized hybrid achromat and a hybrid lens system.

2. Free-space laser communication terminals

Figure 1 shows as an example the design of the optical head of the very small optical user terminal (VSOUT) discussed in [1]. The head consists of two units: an optical block and a mirror pointing assembly attached to the optics block. The optics block is made up of the laser transmitter assembly, the acquisition sensor, the tracking–receiver combination, the telescope and the associated beam-forming optics. A beacon system (generating a large divergent beam) is mounted at the outer coarse pointing mirror housing and is connected to a high-power laser package with a fibre. Many potential applications for DOEs exist in such a system, as follows.

(1) *Laser diode transmitters.* High-power light sources allow long communication distance; a beam-shaping DOE could be used to collimate the output beams of a high-power laser diode array.
(2) *Telescope system.* Complex telescope systems may be simplified using DOEs; the DOEs allow one to reduce the number of elements; furthermore, all lenses could be fabricated with the same material.

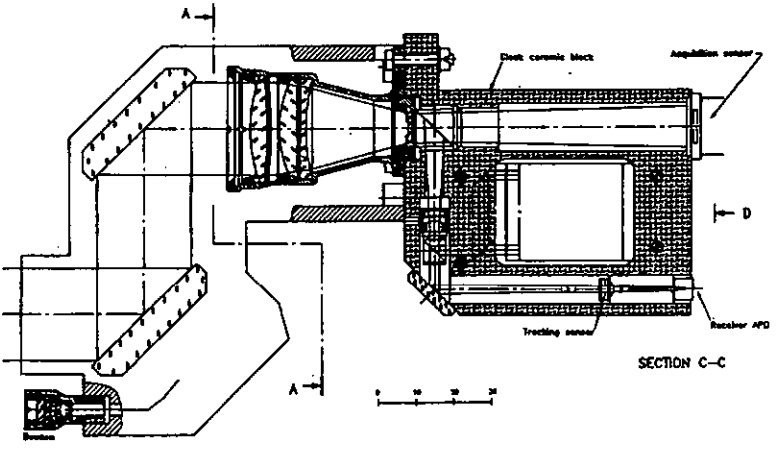


Figure 1. Very small optical user terminal (VSOUT): optical head design.

(3) *Tracking–receiver system.* DOEs allow multitasking with a single element; the tracking and receiving function could be realized by a segmented element.
(4) *Collimator of the beacon head.* The system can be simplified by hybrid (refractive–diffractive) elements.
(5) *Beam splitter.* Multiple beam splitting can be done with one DOE; the number of beam-splitting elements and their weight could be reduced.

In the following, a detailed description of the design of the DOEs used in the tracking–receiver system, an athermalized hybrid collimator, as well as the hybrid beacon system is given. All three elements have been realized by lithography and subsequent etching. The procedure is to generate a mask by electron beam or laser beam lithography. Then, the mask is transferred into an eight-level surface-relief structure by dry etching in fused silica or optical glass.

2.1. *Tracking–receiver system*

A schematic diagram of the tracking–receiver system [12] of the VSOUT optical head is shown in figure 2. The tracking sensor consists of a four-quadrant detector with a central hole which acts as a field stop for the subsequent receiver optics. The specific task of the diffractive optical element, situated at the entrance of the tracking–receiver system, consists of focusing the incoming beam into the central hole (receiver signal) while generating a ring intensity pattern in the tracking detector plane (tracking signal). Changes in the relative positions of the sending and receiving communication terminals are directly translated into a shift of the ring intensity pattern on the tracking aensor, which can be measured by means of the four-quadrant detector.

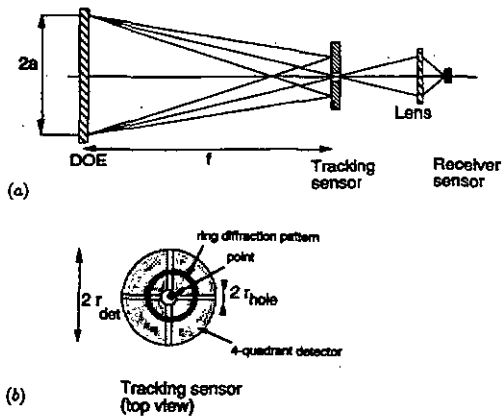


Figure 2. Principles of the tracking–receiving system: (*a*) side view; (*b*) top view. Changes in the relative positions of the sending and receiving communication terminals are directly translated into a shift of the ring intensity pattern on the tracking sensor, which can be measured by means of the four-quadrant detector.

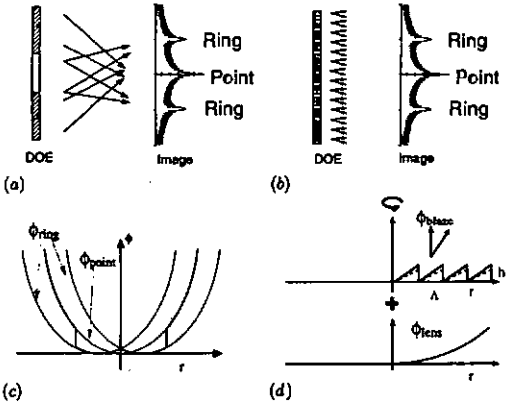


Figure 3. Two different design approaches for the tracking–receiver element: (*a*) segmented aperture design; (*b*) common aperture design; (*c*) the phase function of the segmented element; (*d*) the phase function of the common aperture element.

Different design methods exist for the tracking–receiver DOE. The two specific tasks may be made by a segmented element, which means that the surface of the DOE is divided into two parts: a simple lens function in one part for the focusing point, and a radially shifted lens function in the other part for generation of the ring pattern (segmented aperture design). An alternative technique is to generate the ring pattern by the first diffraction order of a rotational blazed grating structure combined with a focusing function. The focal point in the centre is then realized by the zero order of the grating structure. In this case, both functions are generated by the whole element and thus have the same aperture (common aperture design). The two design strategies are shown in figures 3 (*a*) and (*b*).

DOEs are mainly described by their phase function. For the segmented elements, each segment has its own phase function. In the case of the tracking–receiver DOE the phase function ϕ is then given by

$$\phi(r) = \begin{cases} -\dfrac{k}{2f} r^2, \\ -\dfrac{k}{2f}(r - r_{\text{ring}})^2, \end{cases} \quad \text{for} \begin{cases} r_{\text{seg}} < r < a, \\ 0 < r < r_{\text{seg}}, \end{cases} \tag{1}$$

where k is the wave-vector, f the focal length of the tracking system, r_{ring} the radius of the ring intensity pattern on the four-quadrant detector, $2a$ the diameter of the DOE, and r_{seg} the radius of separation of the two different segments. The function $\phi(r)$ is shown in figure 3 (*c*). The advantage of this concept is that the ratio of the energy used for tracking (ring) to the energy used for the receiver (focal point) is directly given by the ratio of the surfaces of the two segments (i.e. $(r_{\text{seg}}/a)^2$). Furthermore, the two optical functions can be tested and characterized independently by covering one of the segments. However, a drawback of this concept might be that the system is not shift invariant, that is a change in the input

intensity distribution directly affects the output functions. Also, the radius of the diffraction spot of the focus and the width of the ring are given by the aperture of the corresponding segments. This can lead to design problems. In particular, if much more energy has to be in the central point than in the ring, the surface of the segment used to generate the ring becomes small; thus the width of the ring becomes very large owing to diffraction. On the other hand, if much more energy has to be in the ring than in the point, the point becomes very large. Depending on the design specifications the segmented approach may not be satisfying. The above-mentioned problem can be avoided by the common aperture design.

The phase function of the common aperture design is composed of a lens function and a blazed grating structure defined over the entire aperture:

$$\phi(r) = -\frac{k}{2f}r^2 + h \bmod\left(\frac{r}{\Lambda}, 1\right), \tag{2}$$

where Λ is the period of the grating structure, h is the phase depth of the grating structure and mod (a, b) is the modulo function (also referred to the fractional part of the division a/b). The phase function $\phi(r)$ is shown in figure 3 (d). The grating period Λ determines the beam deflection of the first diffraction order and thus the radius of the ring, while the phase depth h of the grating structure determines the energy distribution in the different diffraction orders. The diffraction efficiency of the first order of a blazed phase grating is given by

$$\eta_1 = \sin^2\left(\frac{h}{4}\right). \tag{3}$$

For a 2π-deep blaze structure, only the first diffraction order has energy; only a ring structure without a central peak is generated. If the phase depth differs from 2π, then the zero order will also gain some energy, but also higher orders appear, that is rings of radius $2r_{\text{ring}}$, $3r_{\text{ring}}$, etc., are created. However, these higher-order rings do not affect the functionality of the tracking system. In this approach the diameter of the diffraction spot is given by the aperture of the whole DOE and is therefore smaller than in the first approach.

Based on the design of the VSOUT a ring generator has been fabricated. The specifications of the element are as follows: the design wavelength is 818 nm, the focal length of the element is 20 mm, the diameter of the DOE is 2 mm, and the diameter of the central hole in the tracking detector is 0·1 mm. Furthermore, 75% of the energy has to be inside the central hole of the tracking detector used for the receiver. The resulting eight-phase level element has smallest features of the order of 2 μm. The DOE was realized in fused silica. In addition, a set of visible light demonstrators (633 nm) for the ring–point generation has been made, with a focal length of 400 mm and a diameter of the generated ring of 15 mm. For these demonstrators, both design approaches have been tested. Figure 4 (*a*) shows the diffraction pattern of the common aperture design. Scans through the focus in the image plan, together with the calculated function are presented in figures 4 (*b*) and (*c*) for the segment aperture and the common aperture design respectively. The measured intensity distribution shows excellent agreement with the expected theoretical data. The difference between theory and measurements can be explained by etch depth errors in the fabrication.

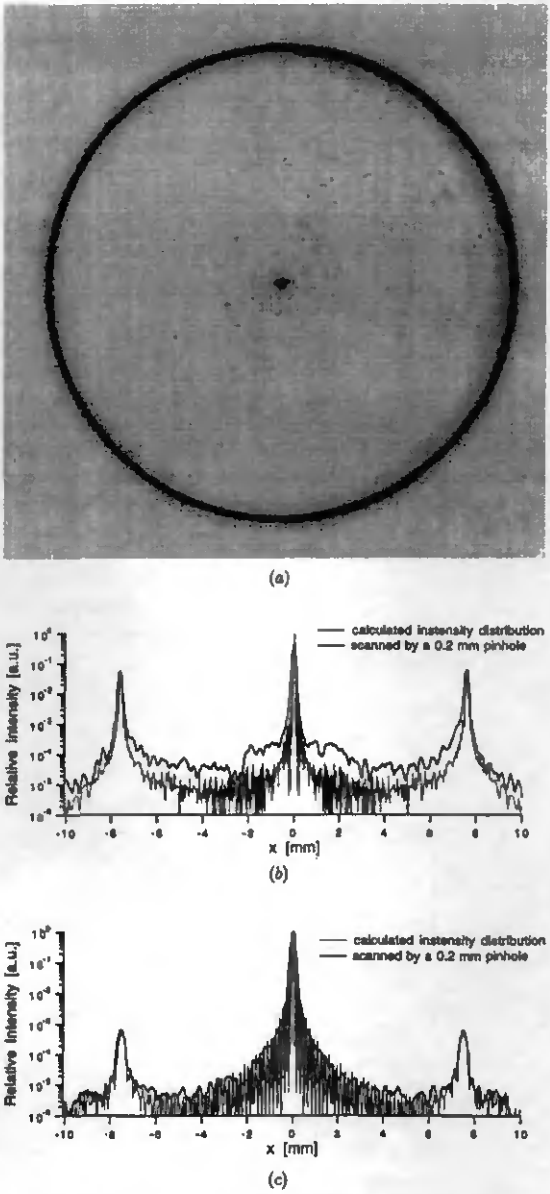


Figure 4. (*a*) The diffraction pattern of the common aperture design. (*b*), (*c*) Scans through the focus in the image plan, together with the calculated curves (*b*) for the segment aperture and (*c*) for the common aperture design (a.u., arbitrary units).

2.2. *Athermalized and achromatic hybrid collimator system*

Space environment introduces severe conditions for the thermal behaviour on the used elements. For this purpose an athermalized achromatic system has been designed, which keeps the position of the focal point within a small region when operating temperature and illumination wavelength is changing. The negative dispersion associated with the DOE is used to compensate the chromatic aberrations of the refractive lens, while the thermal properties of the optical material are exploited to compensate the thermally induced expansion of the supporting mount.

In particular, the dispersion of the DOE is negative, stronger and material independent compared with the refractive element. An achromat can therefore be designed with two focusing elements, contrary to a traditional achromatic doublet, where one of the two refractive elements needs negative power for the chromatic compensation. Consequently, the focusing power of the other refractive element has to be higher; thus the magnitude of the higher-order aberrations are larger in the classical design. Furthermore, the hybrid achromat can be designed with the same material which simplifies also the design of a thermally compensated system.

For such an athermalized system the variation of the mounting has to be compensated by the focal length variation of the optical system (figure 5). The focal length variations are described by an optothermal expansion coefficient, the mechanical variations in the mounting material by a thermal expansion coefficient. The general approach for the design of an athermalized system is to choose correctly the glasses and the mounting materials so that these expansion coefficients are equal [13]. Once the appropriate glass and mounting material has been determined, the optical design is made. In this design the DOE element is consdered as an ultra-thin lens with a very high refractive index, referred to as the high-refractive-index (HRI) model [8, 13]. The attractive festure of this model

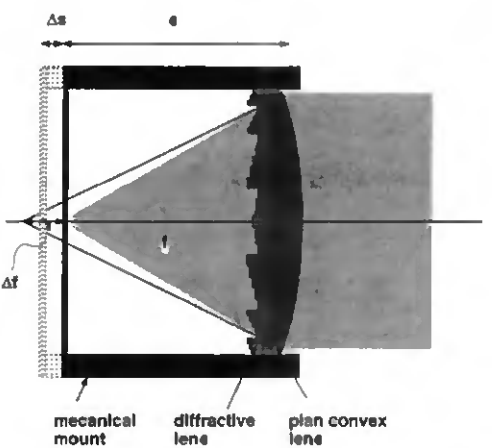


Figure 5. Athermalized achromat, together with its mounting mechanics. For an athermalized element the displacement Δs introduced by the supporting mount has to be compensated by the thermal variation Δf in the focal length.

is that DOEs can be represented in any lens design software as arbitrary aspherics, which in turn are realized directly by standard DOE manufacturing processes.

A hybrid collimator of 5·2 mm diameter and 8 mm focal length has been designed and fabricated. The system is thermally compensated for a range from −20 to 40°C and has a working wavelength range 630–670 nm. The mount material is a combination of aluminium and steel which has a thermal expansion coefficient of $26{\cdot}4 \times 10^{-6}°C^{-1}$. It has been found that FK51, which has an optothermal expansion coefficient of $25 \times 10^{-6}°C^{-1}$, is an appropriate glass in order to satisfy the athermalization condition. The length variation of the mount is 11 μm over the whole temperature range; the calculated shift Δf of the focal length of the hybrid system is 12·6 μm. Hence, the absolute change in the focal point position is less than 2 μm. The designed DOE has been realized in FK51 by photolithography and ion beam milling. The resulting eight-phase level element has smallest features in the order of 3 μm. A good criterion to judge the performance of the collimator is to measure the wave-front aberrations of the system. Figure 6 (*a*) shows the expected optical path difference for the central design wavelength of 650 nm and the measurement wavelength of 633 nm and figure 6 (*b*) shows a double-pass interferogram of the hybrid collimator at 633 nm. The central fringe of the measured double-pass interferogram behaves exactly as the calculated values. The expected maximal optical path difference is of the order of $\lambda/4$ which corresponds nicely to the shift of $\lambda/2$ in the measured double-pass interferogram.

2.3. *Hybrid beacon system*

The third selected application for DOEs is the collimator for the beacon system. The task of the beacon system is to generate a large divergent beam, which is used in a pre-operational phase of the communication link in order to locate the opposite communication terminal. In the classical VSOUT design the beacon system consists of four refractive lenses (figure 7 (*a*)) of different glass types. The benefit of a hybrid design is the reduction in the number of elements, namely from four elements of different materials to three elements of the same material, which simplifies the system and lowers its weight. The hybrid system is shown in figure 7 (*b*). The DOE is located at surface 3. Note that the planar back side of the DOE substrate (surface 2) is used as protection of the optical system and can be cleaned easily. As described in section 2.2, the design is based on the HRI model. The system is optimized for an operating wavelength of 850 nm, a focal length of 6·25 mm, and a numerical aperture of 0·4. The resulting aspheric phase function is shown in figure 8.

The hybrid beacon system and the classical beacon system have been designed and realized in order to compare the two different approaches. The DOEs were manufactured by a multistep lithographic process in fused silica for a working wavelength of 850 nm. The smallest structure for an eight-level element is in the order of 6 μm. Figure 9 shows the calculated wave-front aberrations for both systems. It can be seen that the hybrid design shows even less aberrations than the classical design. The wave front of the classical and the hybrid design were measured and compared at 633 nm. The expected peak-to-valley wave-front error at this wavelength turns out to be $0{\cdot}2\lambda$ for the classical and $1{\cdot}6\lambda$ for the hybrid system. The measured values are 0.32λ and $1{\cdot}29\lambda$. Thus the measurements are in

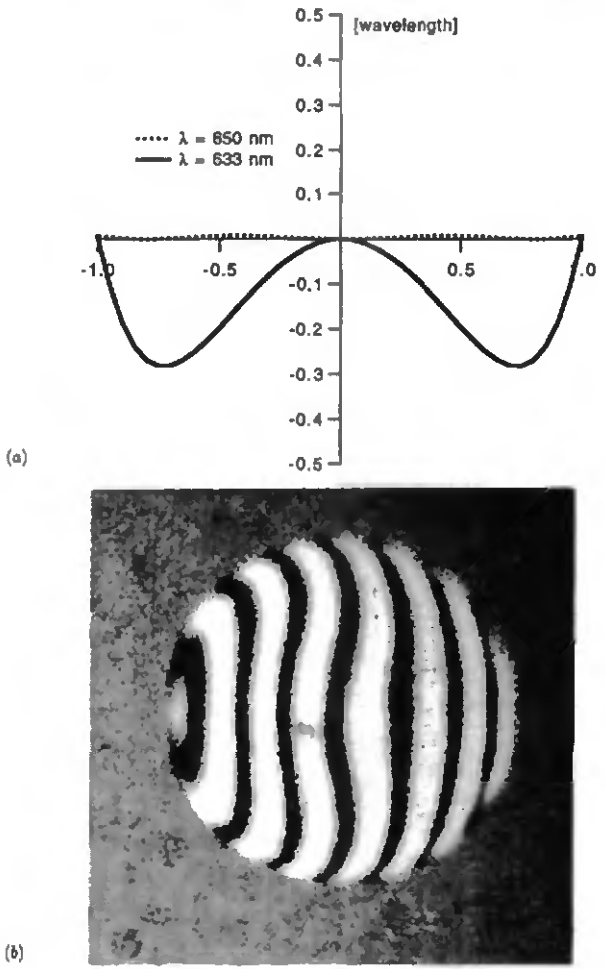
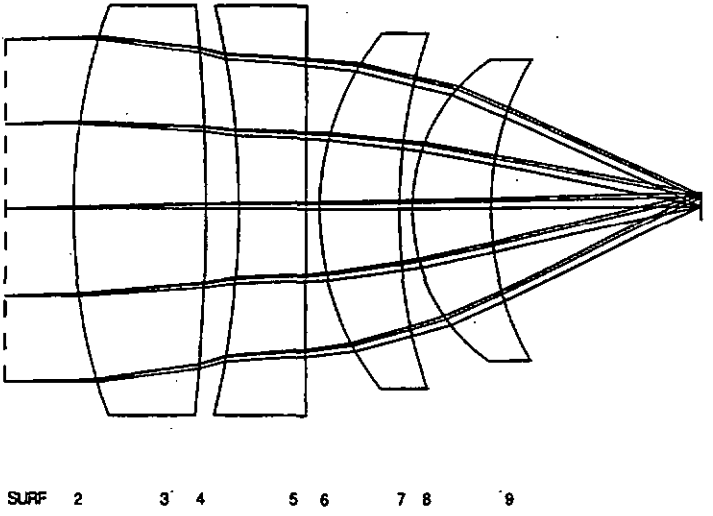


Figure 6. (*a*) Expected optical path difference of the achromatic system for $\lambda = 633$ nm. (*b*) Double-pass interferogram of the realized system at this wavelength. The measurement shows exactly the calculated behaviour.

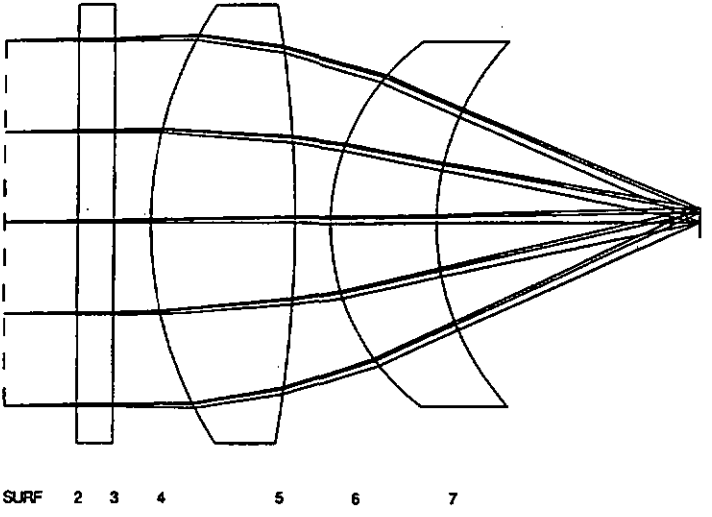
agreement with the theoretical values, the differences are explained by misalignment errors.

3. Conclusions

It has been shown that there is a large potential for diffractive optical elements in optical space communication terminals. The system design with the help of



(*a*)



(*b*)

Figure 7. Beacon system: (*a*) classical design; (*b*) hybrid design.

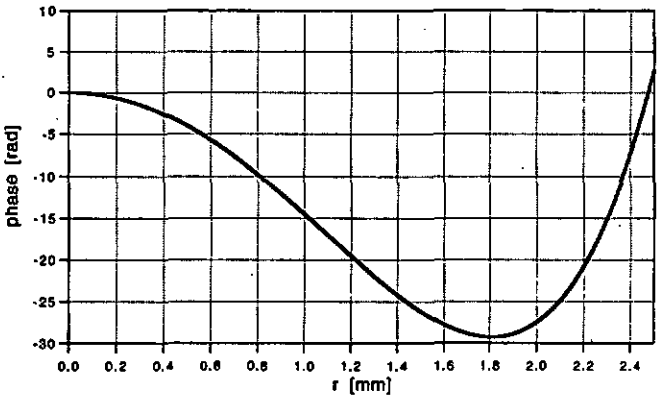
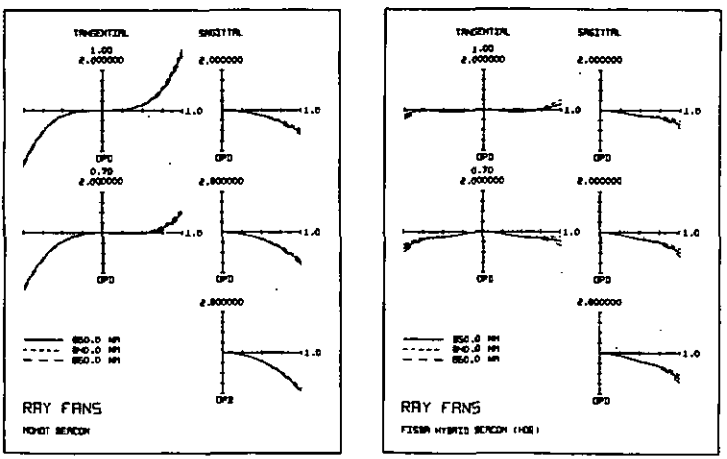


Figure 8. Phase function of the DOE for an optimized beacon collimator with a numerical aperture of 0·4 and a focal length of 6·25 mm.



(*a*) (*b*)

Figure 9. Expected wave-front aberration of the beacon collimator for (*a*) a classical design and (*b*) a hybrid design.

DOEs benefits from all the main advantages of these elements in order to reduce complexity, weight and costs. Three different systems using diffractive optics have been selected at the example of the VSOUT. First, a tracking–receiver system has been discussed where both functionalities were performed by a single DOE. The task of this DOE is to generate a point–ring pattern. For this purpose, two different design strategies have been investigated. One is based on a segmented aperture, and the other on a common aperture approach. The realized elements showed good agreement with the theory for both types of design. However, it

turned out that the segmented aperture-based system may give some problems for specific point–ring energy ratios. Secondly, the design and realization of an athermalized and achromatic hybrid collimator system have been presented. The performance of this system is in excellent agreement with the expected theoretical values. Third, a hybrid beacon system has been presented consisting of one diffractive and two refractive elements. It has been shown that it is possible to reduce the complexity of a classical system by introducing DOEs while keeping the same optical performances. All elements have satisfied the given severe specifications for space applications, which encourages the use of DOEs in other systems in the near future.

Acknowledgments

Parts of this work have been done in the framework of the ESA project No. Col0566/93/NL/JV and of the Swiss priority program 'Optique'.

References

[1] Czichy, R. H., and Wittig, M., 1993, *Proceedings of the Conference on Holographic Systems, Components and Applications*, Neuchâtel, 1993, IEE Conference Publication No. 379 (London: Institution of Electrical Engineers), pp. 255–259.
[2] Wyrowski, F., 1991, *Appl. Optics*, 30, 4650–4657.
[3] Eisenmann, M. T., Tai, A. M., and Cederquist, J. N., 1989, *Appl. Optics*, 28, 2641–2650.
[4] Ehbets, P., Herzig, H. P., Dändliker, R., Regnault, P., and Kjelberg, I., 1993, *J. mod. Optics*, 40, 637–645
[5] Ono, Y., Kimura, Y., Ohta, Y., and Nishida, N., 1987, *Appl. Optics*, 26, 1142–1146.
[6] Raguin, D., and Morris, G. M., 1993, *Appl. Optics*, 32, 2582–2598.
[7] Stones, T., and George, N., 1988, *Appl. Optics*, 27, 2960–2971.
[8] Czichy, R. H., 1993, *Hybrid Optics for Space Applications*, ESA Scientific/Technical Monographs No. SP-1158 (Noordwijk: ESA Publications Division).
[9] Behrmann, G. P., and Bowen, J. P., 1993, *Appl. Optics*, 32, 2483–2489.
[10] Londoño, C., Plummer, W. T., and Clark, P. P., 1993, *Appl. Optics*, 32, 2295–2302.
[11] Herzig, H. P., Gale, M. T., Lehmann, H. W., and Morf. R., 1993, *Perspectives for Parallel Interconnects*, edited by Ph. Lalanne and P. Chavel (Berlin: Springer).
[12] Czichy, R. H., 1992, ESA/PAT/274.
[13] Jamieson, T. H., 1981, *Opt. Engng*, 20, 156–160.
[14] Sweatt, W. C., 1979, *J. opt. Soc. Am.*, 69, 486–487.

Scanning spot metrology for testing of photolithographic masks

Peter Blattner
Hans Peter Herzig
University of Neuchâtel
Institute of Microtechnology
Rue A.-L. Breguet 2
CH-2000 Neuchâtel, Switzerland
E-mail: blattnar@imt.unine.ch

S. Sohail H. Naqvi, MEMBER SPIE
University of New Mexico
Center for High Technology Materials
Department of Electrical and Computer Engineering
Albuquerque, New Mexico 87131

Abstract. We investigated an optical method for characterizing submicrometer structures of photolithographic masks, enabling fast and nondestructive testing over large areas. The scanning spot metrology provides accurate information about edge locations of opaque structures on chrome masks. Algorithms for the extraction of edge locations from the detector signal are discussed and applied to the characterization of a modulated grating mask. Local fabrication errors of the order of 10 to 50 nm can be detected.

Subject terms: optical metrology; knife-edge method; submicrometer structures; photolithographic masks.

Optical Engineering 34(8), 2425–2427 (August 1995).

1 Introduction

Advancement in the areas of phase shift optical lithography, e-beam lithography, and x-ray lithography have enabled the realization of very fine relief structures in the nanometer to micrometer range. The ability to measure the relief parameters and the absolute position accuracy of these structures is of obvious importance, not only to determine if the desired structure has been realized but also to optimize the fabrication process. The wish list for metrology includes nondestructive testing, testing of large areas, and testing of large-aspect-ratio structures having submicrometer lateral dimensions. Current techniques for measurement are optical microscopy[1] (including conventional, confocal, and near-field techniques), scanning electron microscopy[2] (SEM), and scanning-force microscopy.[3] None of these techniques, however, is capable of providing rapid accurate submicrometer measurements over larger areas.

Optical testing methods have been applied to address these requirements. The relevant parameters of a lithographic process, i.e., the depth and width of a gratinglike test structure, can be determined by scattering an incident laser beam at the structures and measuring the far-field intensity distribution. Previously published results have demonstrated the capability of far-field measurements to rapidly and very accurately measure large-aspect-ratio submicrometer feature over larger areas.[4] The main drawback is that these techniques determine only the average parameters, and they are unable to determine local errors such as single-line defects.

Therefore, we investigated scanning spot metrology, which provides accurate information about edge locations of opaque structures on chrome masks. The method involves illuminating the mask with a small-spot-size focused laser beam and measuring the total transmitted power as the mask is scanned. In the following, we introduce the basic principle and discuss algorithms for extraction of edge locations from the detector signal. Finally, the scanning spot metrology is applied to the characterization of a modulated grating mask having an average grating period of 8 μm.

Paper SWI-54 received Nov. 15, 1994; revised manuscript received Jan. 19, 1995; accepted for publication Feb. 13, 1995.

2 Scanning Spot Metrology

2.1 *Principle*

In contrast to far-field diffraction metrology, scanning spot metrology yields information about local fabrication errors of lithographic masks. The experimental setup is shown in Fig. 1. A test structure is illuminated with a focused laser beam and the total transmitted power is measured as the mask is scanned. An integrating sphere with a large entrance aperture is used to measure the total transmitted power. If the laser spot size and the feature sizes on the mask are of comparable dimensions, the detector signal will alternate between high and low values as alternating lines and spaces are illuminated. The edge position information can then be extracted using appropriate signal processing algorithms.

Scanning spot metrology is essentially based on the concept of knife-edge scanning of a laser beam.[5] Consider the case of a 1-D knife-edge illuminated with a focused Gaussian laser beam. The intensity distribution of the incoming beam is given by

$$I(x,y)=\frac{2P_0}{\pi w_0^2}\exp\left(-2\frac{x^2+y^2}{w_0^2}\right)=I_1(x)I_2(y)\ , \qquad (1)$$

where w_0 is the beam radius at $1/e^2$ and P_0 is the incident power. Note, that $I(x,y)$ is separable in x and y. Using a coordinate system where the beam is centered at the location $u=x$ and the edge is located at $u=a_0$, the total transmitted power when the mask is scanned can be written in terms of the complementary error function[6] as

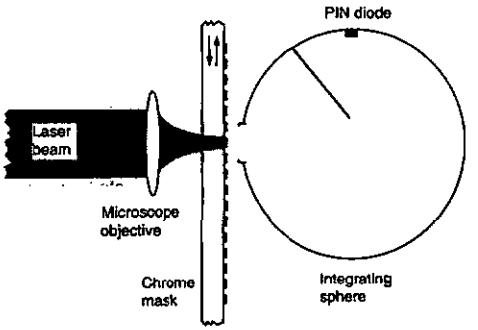


Fig. 1 Experimental arrangement of the scanning spot metrology.

$$P_T(x) = \int_{x_0}^{\infty} I_1(x)\,dx \int_{-\infty}^{\infty} I_2(y)\,dy = \frac{P_0}{2}\,\mathrm{erfc}\left(\frac{a_0 - x}{w_0/\sqrt{2}}\right). \quad (2)$$

Thus, when a knife-edge scan is made, the unknown quantities are, in the most general case, the incident power, the beam radius, and the edge location. They are found by iteratively fitting the experimental data to the function given in Eq. (2). Figure 2 displays experimental data for a knife-edge scan together with their best fits, which are obtained using 20× and 100× microscope objectives. Beam radii of 3.02 and 1.08 μm, respectively, are obtained. Smaller illumination spots can be obtained by optimizing the incident beam diameter with respect to the aperture of the microscope objective.

Applying the same concept but replacing the knife edge by an amplitude grating consisting of a series of opaque and transparent zones, the transmitted power can directly be expressed by

$$P_T(x) = \frac{P_0}{2}\sum_{i=0}^{N}\left[\mathrm{erfc}\left(\frac{a_i - x}{w_0/\sqrt{2}}\right) - \mathrm{erfc}\left(\frac{b_i - x}{w_0/\sqrt{2}}\right)\right], \quad (3)$$

where a_i and b_i are the edge locations of the grating mask. Whereas the knife-edge technique is mainly used to determine beam profiles, the grating-edge technique enables measuring the location of the edges of the grating.

2.2 *Edge Extraction Algorithm*

The ability to extract edge information from $P_T(x)$ in the case of multiple edges greatly depends on the radio of the beam radius w_0 and the feature sizes s on the mask. Figure 3 shows the contrast of the function $P_T(x)$ as a function of the ratio s/w_0 for a grating with 50% duty cycle. For the example of a beam radius of 1 μm (Fig. 2, 100× objective), submicrometer structures with feature sizes down to 0.5 μm could be characterized.

For modulated grating structures, the local maximum and minimum values of $P_T(x)$ are no longer constant and change as the linewidth or space width varies, according to the contrast function in Fig. 3. The information on the edge position can be obtained by determining the local maximum and minimum of the measured transmission function. The accuracy of the edge position detection, however, is then limited directly by the accuracy of the translation stage used to scan the mask. The metrology issue, therefore, is to determine the linewidth and space widths of each line/space pair by fitting, as in the case of knife-edge metrology, the theoretical total transmitted power function, given by Eq. (3), to the measured values. In the ideal case, an optimization over all parameters would be done. The parameters to be determined include the incident power P_0, the beam radius w_0, and all edge locations (a_i, b_i).

To obtain a good initial guess of the incident power, the data are normalized to have a mean value of 0.5, corresponding to the average fill factor of the modulated grating structure. The initial guess of the edge locations of the mask are the positions for which the normalized transmission is 0.5. The start-up value of the beam radius is determined using the knife-edge technique at the first edge of the grating structure. In any multiparameter optimization problem the computer time required rapidly increases as the dimension of the parameter space is increased. Consideration of all edges simultaneously may not be necessary, however, because the response $P_T(x)$ primarily depends on a few neighboring edges. The edge extraction algorithm is therefore limited to local optimizations of a maximum of only two parameters at the time. During optimization, not only are the edge locations varied, but the estimates of the beam radius and the total incident power are also readapted. Because the initial guesses

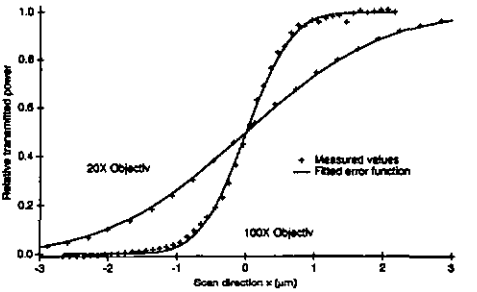


Fig. 2 Measured intensity while scanning a knife edge through the focus of two different laser spots generated by 20× and 100× microscope objectives.

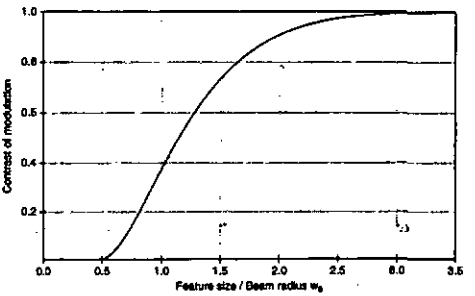


Fig. 3 Contrast of modulation versus the linewidth-to-beam-radius ratio.

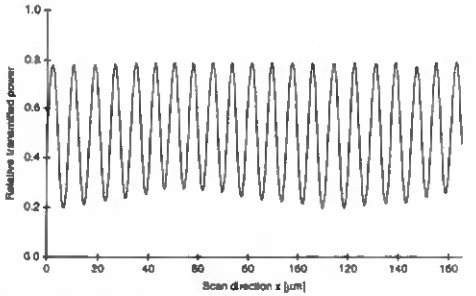


Fig. 4 Measured relative transmitted power as a function of the illuminated location on the mask.

of the parameters are quite good, the optimization algorithm converges rapidly and good repeatability of the edge location algorithm is obtained.

2.3 *Results*

The scanning spot metrology technique was experimentally applied to a modulated grating structure on a normal reflectivity chrome mask. The grating consists of a series of opaque and transparent zones with average period of 8 µm and lateral modulations up to 500 nm. Light from a HeNe laser was focused on the structure using a 20× microscope objective. The light was linearly polarized (TE), however no polarization effects are expected in the presented measurement technique. Figure 4 shows the measured and normalized transmitted power $P_T(x)$, obtained when the mask is scanned on a high-precision translation stage with 25-nm step resolution.

Note that the maximum intensity remains almost constant, indicating that the space widths are hardly varying, whereas the minimum levels change because of variation of the linewidth, which corresponds exactly to the used encoding technique of the grating structure.[7] Applying the edge extraction algorithm, the linewidth and space width can be determined. Figure 5 shows the linewidth extracted from the transmitted power measurements compared with the theoretical data as used to generate the mask. Local fabrication errors between 10 and 50 nm can be easily detected. The origins are deviation errors of the e-beam during writing, but also the result of a difference between the grid of the e-beam writer and of the theoretical data. The errors are measured within one scan field of the e-beam writer, therefore no stitching errors are observed.

3 Conclusions

We introduced scanning spot metrology of lithographic masks that enables a fast estimation of local errors over large scanning areas. Algorithms for the extraction of edge locations from the detector signal were discussed and applied to the characterization of a modulated grating mask with a typical structure size of 4 µm. Local fabrication errors of the order of 10 to 50 nm were detected. With an illumination beam focused down to 1 µm spot size, submicrometer structures can also be tested. The limited depth of focus, however, will then require an autofocus system to control the spot diameter.

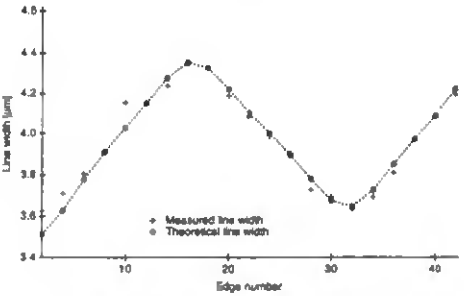


Fig. 5 Linewidth for each edge number.

Acknowledgment

This work was supported by the Swiss Priority Program, Optique.

References

1. E. Betzig, J. K. Trautman, T. D. Harris, and J. S. Weiner, *Science* **251**, 1468–1470 (1991).
2. M. Postek and D. Joy, "Submicrometer microelectronics dimensional metrology: scanning electron microscopy," *NBS J. Res.* **92**, 205–228 (1987).
3. M. R. Rodgers and K. M. Monahan, "Using the atomic force microscope to measure submicron dimension of integrated circuit devices and processes," in *Integrated Circuit Metrology, Inspection, and Process Control V*, W. H. Arnold, Ed., *Proc. SPIE* **1464**, 358–366 (1991).
4. S. S. H. Naqvi, R. H. Krukar, J. R. McNeil, J. E. Franke, T. M. Niemczyk, D. M. Haaland, R. A. Gottscho, and A. Kornblit, "Etch depth estimation of large-period silicon gratings with multivariate calibration of rigorously simulated diffraction profiles," *J. Opt. Soc. Am. A* **11**(9), 2485–2493 (1994).
5. Y. Suzaki and A. Tachibana, "Measurement of the µm sized radius of Gaussian laser beam using the scanning knife-edge," *Appl. Opt.* **14**(12), 2809–2810 (1975).
6. M. T. Gale and H. Meier, "Rapid evaluation of submicron Laser spots," *RCA Rev.* **46**, 56–69 (1985).
7. A. W. Lohmann and D. P. Paris, "Binary Fraunhofer holograms, generated by computer," *Appl. Opt.* **6**(10), 1739–1748 (1967).

Peter Blattner received the diploma in physical electronics from the University of Neuchâtel, Switzerland, in 1992. In 1993, he was a young graduate trainee at the European Space Agency, where he worked in the field of metrology of diffractive optical elements. Since 1994, he has been a research assistant in applied optics at the Institute of Microtechnology at the University of Neuchâtel, Switzerland.

Hans Peter Herzig: Biography and photograph appear with the paper "High-carrier-frequency fan-out gratings fabricated by total internal reflection holographic lithography" in this issue.

S. Sohail H. Naqvi received the BSEE with highest distinction, the MSEE, and PhD degrees in 1984, 1986, and 1988, respectively, all from Purdue University. He joined the faculty of the University of New Mexico after graduation, where he is now with the Center for High Technology Materials and an associate professor in the Department of Electrical and Computer Engineering. His research interests are in diffractive optics, electromagnetic wave diffraction theory with application toward development of diffraction-based semiconductor metrology techniques, and optical communications.

ELSEVIER

Microelectronic Engineering 27 (1995) 543-546

MICROELECTRONIC ENGINEERING

Diffractive structures for testing nano-meter technology

P. Blattner[a], S. S. H. Naqvi[b], P. Ehbets[a], and H. P. Herzig[a]

[a]Institute of Microtechnology, University of Neuchâtel, Rue A. -L. Breguet 2, CH-2000 Neuchâtel, Switzerland

[b]Center for High Technology Materials, Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM 87131, USA

We investigated two optical methods for characterizing submicron structures. Average errors of a few nanometers can be determined by the far-field diffraction metrology utilizing diffractive structures having enhanced sensitivity to fabrication errors. The scanning spot metrology is well suited for analyzing lithographic masks.

1. INTRODUCTION

Advancement in the areas of pbase shift optical lithography, e-beam lithography, and x-ray lithography have enabled the realization of very fine relief structures in the nm-μm regime. The ability to measure the relief parameters and absolute position accuracy of these structures is of obvious importance, not only to determine if the desired structure has been realized but also to optimize the fabrication process. The wish list for metrology includes non-destructive testing, testing of large areas, and testing of large aspect ratio structures having sub-μm lateral dimensions. Current techniques for measurement are optical microscopy (including conventional, confocal and near-field techniques)[1], scanning electron microscopy[2] (SEM), and scanning probe microscopy[3] (SPM). None of these techniques, however, is capable of providing rapid accurate sub-μm measurements over larger areas.

To address these requirements. we investigated the use of far-field diffraction metrology utilizing diffractive structures having enhanced sensitivity to fabrication errors in linewidth and etch depth. An important point to note bere is that this technique determines the average error, and is unable to determine local errors such as single line defects.

Secondly, we introducc scanning spot metrology of chrome masks which involves illuminating the mask with a small spot size focused laser beam and measuring the total transmitted power as the mask is scanned. Algorithms for extraction of edge locations from the detector signal are discussed, and applied to the characterization of a modulated grating mask.

2. FAR - FIELD DIFFRACTION METROLOGY

A fabrication technology can be best characterized, if specific structures with sensitive and easy-to-measure properties are realized. Grating structures provide ideal test elements for far-field diffraction metrology, since they split an illuminating laser beam into a discrete array of diffraction orders. Gratings with small periodicities p in the order of the optical wavelength λ produce only a very limited number n of propagating diffraction orders ($n < p/\lambda$). If the grating period becomes smaller than the read-out wavelength, only the zero order will be generated and can be used for the characterization. For regular gratings, the inteosity distribution in the far-field depends strongly on the relief shape, the illumination angle and the wavelength.

The basic principle of far-field metrology is to measure the intensity distribution of the diffraction orders as a function of the incidence angle or the wavelength and then, to determine the relief shape and depth by numerical modelization. Previous published results bave demonstrated the capability of this technique to rapidly and very accurately measure

SSDI 0167-9317(94)00163-4

large aspect ratio sub-μm feature over larger areas[5,6]. Problems with this technique arise, if the lateral position accuracy of the lithographic process has to be determined. In the case of a regular grating, position errors produce a continuous blur of the diffraction orders, which is difficult to measure. For this task, we have investigated modulated grating structures. Position sensitive elements are obtained by designing pulse-position modulated gratings which generate a well defined intensity function[4] in the first diffraction order of the carrier grating. An appropriate measurement signal is obtained by encoding a fan-out function[5] which produces a discrete set of equally intense light spots in the first order. Positioning errors affect directly the fan-out function and therefore, introduce uniformity errors in the far-field. The average position error within the illuminated field is determined by adding a statistical Gaussian distributed position error to the ideal modulated grating structure and calculating the corresponding uniformity error in the far-field that matches the uniformity measurement.

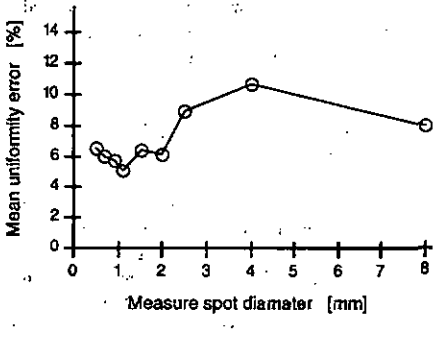


Fig. 1. Measured uniformity error in function of the beam diameter.

Experimental results have been obtained by characterizing a high resolution e-beam written mask with a carrier grating period equal to 1 μm. The modulation of the 9 beam fan-out function has a period equal to 125 μm and introduces locally position modulation up to 50 nm. We have measured the uniformity error as a function of the illumination diameter (see Fig. 1). For small beam diameters, the uniformity error decreases and reaches a minimum of 5% for a diameter equal to 1 mm. This minimum value corresponds to the case, when one scan field of the e-beam writer is illuminated. The decrease of the uniformity error with the beam diameter can be explained by the fact that the statistical errors are better averaged out.

The remaining 5 % uniformity error is essentially due to systematic errors inside one scan field. The simulation has shown that the 5 % error corresponds to a mean lateral position error equal to 50 nm inside one scan field. For larger beam diameters, the uniformity error increases. In this case, more than one scan field are illuminated and larger stitching errors between the scan fields change the statistics. As a consequence, stitching errors in the order of 100 nm have been determined. If a sufficient number of scan fields is illuminated, the stitching errors are also averaged out and again a low uniformity error equal to 8 % is reached. A more accurate characterization of the systematic errors inside one scan-field is possible, if not only the intensity but also the relative phase of the fan-out beams is measured.

3. SCANNING SPOT METROLOGY

In contrast to the far-field diffraction metrology, the scanning spot metrology yields information about local fabrication errors of lithographic masks. The experimental setup involves illuminating the structure with a small spot size laser beam and measuring the total transmitted power as the mask is scanned. If the laser spot size and the minimum feature size on the mask are of comparable dimensions, the detector signal will alternate between high and low values as alternating lines and spaces are illuminated. The edge position information can then be extracted using appropriated signal processing algorithms.

The scanning spot metrology is essentially based on the concept of knife edge scanning of a laser beam[6]. Consider the case of a one-dimensional knife-edge illuminated with a focused Gaussian laser beam. Using a coordinate system where the beam is centered at the location $u = x$, and the edge is located at $u = a_0$, the total transmitted power as the mask is scanned, can be written in terms of the complementary error function[7]

$$P_T(x) = \frac{P_0}{2}\,\mathrm{erfc}\left(\frac{a_0 - x}{w_0\sqrt{2}}\right), \qquad (1)$$

where w_0 is the beam radius at $1/e^2$, and P_0 is the incident power. Thus, when a knife edge scan is made, the unknown quantities are, in the most general case, the incident power, the beam radius and the edge location. They are found by iteratively fitting the experimental data to the function given in Eq. (1). Figure 2 displays experimental data obtained using 20X and 100X microscope objectives together with their best fits. A beam radius of 1.08 μm and 3.02 μm is respectively obtained.

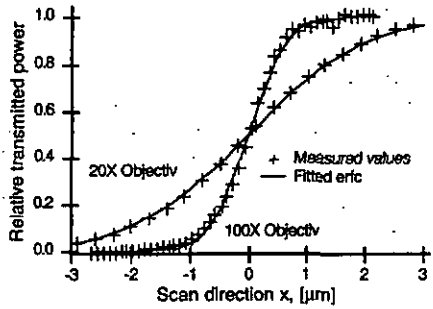


Fig. 2. Measured intensity while scanning a knife edge through the focus of two different laser spots.

Applying the concept of knife-edge scanning to a grating mask, which consists of several opaque and transparent zones, the transmitted power can directly be expressed by

$$P_T(x) = \frac{P_0}{2}\sum_{i=0}^{N}\left[\mathrm{erfc}\left(\frac{a_i - x}{w_0\sqrt{2}}\right) - \mathrm{erfc}\left(\frac{b_i - x}{w_0\sqrt{2}}\right)\right], \qquad (2)$$

where a_i and b_i are the edge locations of the grating mask. As before, estimates of the incident power, the beam radius, and the edge location can be obtained, in principal, by fitting the above function to the experimental transmitted power data.

The ability to extract edge informations from $P_T(x)$ in the case of multiple edges greatly depends on the relationship of the beam radius w_0 to the minimum feature size of the mask being examined. This is illustrated in Fig. 3 which shows the contrast of the function $P_T(x)$ as the period to beam radius ratio is varied for an equal line/space grating illuminated with a Gaussian laser beam.

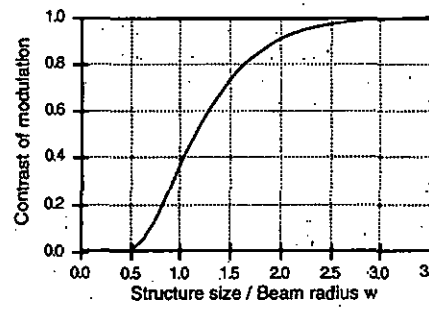


Fig. 3. Contrast of modulation versus the period to beam radius ratio.

For modulated grating structures the local maximum and minimum values of $P_T(x)$ are no longer constant, and change as the linewidth or space width varies. The metrology issue, therefore, is to accurately and precisely determine the linewidth and space widths of each line/space pair.

In the ideal case an optimization over all parameters would be done. The parameters to be determined include, the incident power P_0, the beam radius w_0, and all the edge locations (a_i, b_i). In any multi-parameter optimization problem the computer time required rapidly increases as the dimension of the parameter space is increased. However, consideration of all edges simultaneously may not be necessary, since the response $P_T(x)$ primarily depends on a few neighbor edges. We have therefore decided to limit the edge extraction algorithm to only contain local optimizations, which also results in increased computational efficiency.

To obtain a good initial guess of the edge locations, the data were normalized to have a mean value of 0.5, corresponding to an average fill factor of the modulated grating structure. The initial guess of the edge locations of the mask are then obtained from the intersection of the normalized measurements with a line drawn at the 0.5 level. Initial guess of the beam radius and total incident power are made using knife edge tecbnique at the first edge of the grating structure.

The scanning spot metrology technique was experimentally applied to a modulated grating

structure on a chrome mask. The carrier grating has a period of 8 μm and modulations up to 500 nm. Light from a He-Ne laser was focused on the structure using a microscope 20X objective. Figure 4 shows the measured normalized transmitted power $P_T(x)$ obtained as the mask was scanned on a high precision translation stage.

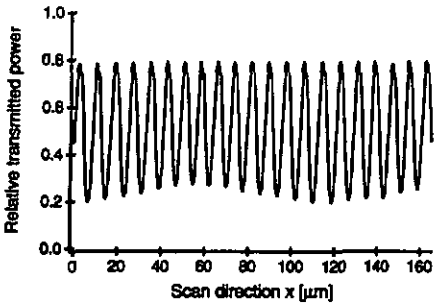


Fig. 4. Measured relative transmitted power as function of the illuminated location on the mask.

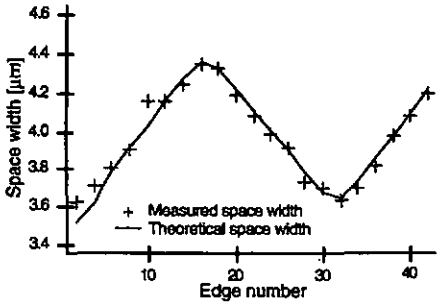


Fig. 5. Space width for each edge number.

Note that the maximum intensity remains almost constant, indicating that the space widths are hardly varying while the minima levels change due to variations in linewidth, which corresponds exactly to the encoding technique of the grating structure. Applying the edge extraction algorithm, the linewidth and space width can be determined. Figure 5 shows the space width extracted from the transmitted power measurements along with theoretical data used to generate the mask. Local fabrication errors between 10 nm - 50 nm can be easily detected.

4. CONCLUSIONS

The ability to measure the relief parameters and absolute position accuracy of very fine relief structures in the nm-μm regime is of importance in modern micro-engineering. The wish list for metrology includes non-destructive testing, testing of large areas, and testing of large aspect ratio structures having sub-μm lateral dimensions.

To address these requirements, we investigated the use of far-field diffraction metrology utilizing diffractive structures having enhanced sensitivity to fabrication errors in linewidth and etch depth. Average fabrication errors of opaque and transparent structures in the order of a few nanometers can be determined.

Furthermore, we have introduced scanning spot metrology of lithographic masks. Algorithms for extraction of edge locations from the detector signal are discussed. In contrast to the far-field diffraction metrology, this methods yields information about local errors.

REFERENCES

1. E. Betzig, J.K. Trautman, T.D. Harris, and J.S. Weiner, Science 251, 1468 (1991).
2. M. Postek and D. Joy, NBS J. of Res. 92, 205 (1987).
3. J.E. Griffith and D.A. Grigg, J. Appl. Phys. 74, R83 (1993).
4. E. Noponen and J. Turunen, JOSA A 11, 1097 (1994).
5. M.T. Gale, M. Rossi, H. Schütz, P. Ehhets, H.P. Herzig, and D. Prongué, Appl. Opt. 32, 2526 (1993).
6. J.A. Arnaud, W.M. Hubbard, G.D. Mandeville, B. de la Claviere, E.A. Franke, and J.M. Franke, Appl. Opt. 12, 2275 (1971).
7. M.T. Gale, H. Meier, RCA Review 46, 56 (1985).