Journal of Modern Optics
2010, 1–13, iFirst

Taylor & Francis
Taylor & Francis Group

On the modeling of optical systems containing elements of different scales

H.P. Urbach*, O.T.A. Janssen, S. van Haver and A.J.H. Wachters

Optics Research Group, Delft University of Technology, P.O. Box 5046, 2600 GA Delft, The Netherlands

(*Received 13 September 2010; final version received 25 October 2010*)

In optical systems often components of widely varying scales occur. The scale, expressed in units of wavelength, determines what kind of model is needed for the component. For objects with size of the order of the wavelength, rigorous electromagnetic models are required. We discuss several home-made rigorous methods such as the Finite Element Method, the Finite Difference Time Domain Method and the Rigorous Coupled Wave Method and describe our experience with them. For lenses of high numerical aperture, the Extended Nijboer Zernike method is considered. An important issue is how to propagate the field from the exit plane of one component to the entrance plane of the other. We explain a numerical integration method by which the diffraction integral can be computed with an error that is independent of the propagation distance.

Keywords: rigorous models; diffraction integrals

1. Introduction

In modern optical systems elements of widely different sizes can occur. Examples are optical recording, high resolution (confocal) microscopy, optical lithography and scatterometry for wafer inspection. Macroscopic elements like lenses occur together with microscopic objects such as small grooves in a diffraction grating, or nano-sized particles and nano-antennas for lighting or sensing applications.

It is of course impossible and also unnecessary to use Maxwell's equations throughout such an optical system. Instead, it is much more efficient to combine different models with different degrees of rigour. For example, in major parts of the optical system Fresnel's approximation of the Rayleigh–Sommerfeld diffraction integral is often sufficiently accurate to propagate the field from the exit plane of one element to the input plane of the next. Using relatively simple and therefore computationally cheap models is in particular essential when one wants to optimize certain properties of the system, because optimization requires the evaluation of many similar systems.

When different models are used for different parts of an optical system, the issue arises how to link them. If for some small element Maxwell's equations have to be used to achieve sufficient accuracy, then one is obliged to keep track of the propagation of the different components of the electric field throughout the system.

In this paper we will give an overview of work that has been done in the recent past on the modeling of complicated optical systems which consists of several elements with widely different scales. We shall summarize our own experience with different models and will point out difficulties we have encountered and how we have solved these, hoping that these experiences may be valuable to others.

We shall mainly focus on time-harmonic fields, although in the example at the end of the paper we also briefly address the case of (partially) coherent illumination. Furthermore, all materials are assumed to be linear.

The content of the paper is as follows. In the next section we report on our experience with different methods for rigorously solving Maxwell's equations for structures with feature sizes of the order of the wavelength and smaller. In particular, we discuss and compare the Finite Element Method (FEM), the Finite Difference Time Domain Method (FDTD), the Volume Integral Equation (VIE), the Surface Integral Equation (SIE) and the Rigorous Coupled Wave Analysis (RCWA). We shall mainly consider three-dimensional problems because for two-dimensional problems memory and CPU are not bottle necks and hence with all methods one can obtain sufficiently accurate results by increasing the number of unknowns. In [1] a number of rigorous solvers made by different groups were bench marked on a two-dimensional problem in

*Corresponding author. Email: h.p.urbach@tudelft.nl

ISSN 0950–0340 print/ISSN 1362–3044 online

DOI: 10.1080/09500340.2010.538736
http://www.informaworld.com

which a plane wave of 852 nm was diffracted by a slit in a Ag layer. All methods gave similar results, be it not at the same computational costs. But three-dimensional problems are still challenging.

In practice many different configurations occur, e.g. isolated scatterers or emitters and structures that are periodic in one or two directions. Very often, the structures are embedded in a multilayer. Widely differing materials occur, such as metals with negative real part of their permittivity at the frequency of interest, anisotropic materials with real symmetric or complex Hermitian permittivity tensor or inhomogeneous materials whose properties vary with position, such as in a bleached photo-resist. Furthermore, widely different sources can occur. For example in the modeling of the emission of a light emitting diode (LED), the sources are inside the device which are moreover spatially incoherent. In optical recording, a focused laser spot is incident on the optical disk to read or write a pit pattern. A useful rigorous solver should be able to deal with all these types of sources.

In Section 3 we will discuss the propagation of light between different optical elements through a homogeneous material. We shall in particular consider cases when Fresnel's approximation is not sufficiently accurate and one of the Rayleigh–Sommerfeld diffraction integrals have to be used. An efficient numerical integration method that is accurate for any propagation distance is explained.

In the fourth section we consider the modeling of a lens of high numerical aperture for which the rotation of polarization cannot be neglected. To express the field in the focal region in the field in the entrance pupil of the lens, the pupil field is expanded in Zernike polynomials and the field in the focal region is written in cylindrical coordinates in terms of a Bessel series. One of the benefits of this analytical method is that it is very suitable for computing the focused field inside a multilayer. Furthermore, because the images of the Zernike polynomials in the pupil can be computed and stored beforehand, this so-called Extended Nijboer–Zernike (ENZ) method is very fast in optimization problems.

In the final section we explain how the different models can be used together to simulate the behavior of complicated optical systems. As an example we discuss the model of the imaging of a mask in photolithography for integrated circuits.

2. Rigorous Maxwell models

2.1. FEM

In solving Maxwell's equations for time-harmonic fields, one can either eliminate the electric or the magnetic field to obtain a second-order differential equation, namely the vector Helmholtz equation. If the magnetic field is eliminated we get for the complex electric field amplitude $\boldsymbol{E}$:

$$\omega^2\epsilon_0\mu_0\boldsymbol{\epsilon}_r\boldsymbol{E} - \boldsymbol{\nabla}\times\boldsymbol{\mu}_r^{-1}\boldsymbol{\nabla}\times\boldsymbol{E} = -\mathrm{i}\omega\mu_0\boldsymbol{J}, \tag{1}$$

where ω is the frequency, ϵ_0 and μ_0 are the permittivity and permeability of vacuum, $\boldsymbol{\epsilon}_r$ and $\boldsymbol{\mu}_r$ are the relative permittivity and permeability tensors, respectively and $\boldsymbol{J}$ is the current density. Far away from the scatterers, the field radiated by the currents satisfies Sommerfeld's outgoing radiation conditions, or equivalently, the Silver–Müller radiation condition [2]. An incident field $\boldsymbol{E}^{\mathrm{i}}$, emitted by sources at very large distance may also be given and in that case it is the scattered field $\boldsymbol{E}-\boldsymbol{E}^{\mathrm{i}}$ that satisfies Sommerfeld's radiation condition.

We first discuss the case of one or several isolated scatterers (objects) embedded in a multilayer with interfaces parallel to the (x,y)-plane. This case obviously includes the case of scatterers surrounded by a single homogeneous material as a special case. We define as computational domain (CD) a rectangular block $x_l<x<x_u,\ y_l<y<y_u,\ z_l<z<z_u$ which contains the isolated scatterers in its interior and which is such that $z>z_u$ and $z<z_l$ both consist of a *single* homogeneous material.

By using Stratton–Chu's integral formula the scattered field in the exterior of the CD can be expressed in its the tangential components on the boundary of the CD:

$$\begin{aligned}\boldsymbol{E}^{\mathrm{s}}(\boldsymbol{r}) = &+\iint_{\partial CD}\hat{\boldsymbol{n}}\times\boldsymbol{H}^{\mathrm{s}}(\boldsymbol{r}')\mathbf{G}_E(\boldsymbol{r},\boldsymbol{r}')\mathrm{d}^2\boldsymbol{r}'\\ &+\iint_{\partial CD}\hat{\boldsymbol{n}}\times\boldsymbol{E}^{\mathrm{s}}(\boldsymbol{r}')\mathbf{G}_H(\boldsymbol{r},\boldsymbol{r}')\mathrm{d}^2\boldsymbol{r}',\end{aligned} \tag{2}$$

$$\begin{aligned}\boldsymbol{H}^{\mathrm{s}}(\boldsymbol{r}) = &-\frac{\varepsilon_m}{\mu_0}\iint_{\partial CD}\hat{\boldsymbol{n}}\times\boldsymbol{E}^{\mathrm{s}}(\boldsymbol{r}')\mathbf{G}_E(\boldsymbol{r},\boldsymbol{r}')\mathrm{d}^2\boldsymbol{r}'\\ &+\iint_{\partial CD}\hat{\boldsymbol{n}}\times\boldsymbol{H}^{\mathrm{s}}(\boldsymbol{r}')\mathbf{G}_H(\boldsymbol{r},\boldsymbol{r}')\mathrm{d}^2\boldsymbol{r}',\end{aligned} \tag{3}$$

where $\hat{\boldsymbol{n}}$ is the outward pointing unit normal on the boundary ∂CD of the CD and $\mathbf{G}_E$, $\mathbf{G}_H$ are dyadic tensors of the electromagnetic field which are defined such that for any constant unit vector $\hat{\boldsymbol{v}}$:

$$\boldsymbol{\nabla}\times\mathbf{G}_E\hat{\boldsymbol{v}} = \mathrm{i}\omega\mu_0\mathbf{G}_H\hat{\boldsymbol{v}}, \tag{4}$$

$$\boldsymbol{\nabla}\times\mathbf{G}_H\hat{\boldsymbol{v}} = -\mathrm{i}\omega\epsilon_m\mathbf{G}_E\hat{\boldsymbol{v}} - \hat{\boldsymbol{v}}\delta(\boldsymbol{r}-\boldsymbol{r}'), \tag{5}$$

where ϵ_m is the permittivity in the absence of the nontrivial scatterers. Using the Stratton–Chu integral formula and the continuity of the tangential components of the total electric and total magnetic field across the boundary ∂CD, one can derive coupled

integral equations for these tangential components on ∂CD. These integral equations can be used as (rigorous) boundary conditions on ∂CD to terminate the computational domain in a rigorous way. The integrals in Stratton–Chu's formula are however highly singular and have to be interpreted as Cauchy Principal value integrals. This implies that their numerical approximation is difficult. Furthermore, the thus obtained boundary conditions are nonlocal in the sense that the tangential component of the electric field in a given point on the boundary is related to the tangential components of the magnetic field in all other points of the boundary and vice versa. This leads to many nonzeros in the system matrix that is obtained after discretization.

The integral equations are therefore almost never used. In most cases the Perfectly Matched Layer (PML) [3] is instead applied to terminate the CD. In using the PML, one defines an extended computational domain (ECD) which is a rectangle that contains the CD in its interior (see Figure 1). The PML is the region inside the ECD and outside of the CD. Since the ECD is larger than the CD, the PML method requires more unknowns than the integral equations on the boundary, but this is compensated for by a much simpler implementation. Inside the CD one can use as unknown field either the total field or the scattered field. When there is no multilayer, i.e. the scatterer is surrounded by a single material, the scattered field is simply the difference between the total field and the (given) incident field. Inside the PML the scattered field is used as an unknown field. The unknown fields inside the CD and the PML are linked by the requirement on the interface between the CD and the PML that the tangential components of the total electric and magnetic fields should be continuous.

The scattered field inside the PML is absorbed and at the outer boundary of the PML (i.e. the boundary of the ECD) the tangential electric or tangential magnetic field components are set equal to zero.

When there is a multilayer background, special precautions have to be taken to define the unknown field in the PML. We first define the field $\boldsymbol{E}^0$ as the field that would be excited in the multilayer by the incident field $\boldsymbol{E}^{\mathrm{i}}$ and the current density $\boldsymbol{J}$ in the case that there are no isolated scatterers. Then we define the scattered field as the difference between the total field and this field $\boldsymbol{E}^0$, i.e.

$$\boldsymbol{E}^{\mathrm{s}} = \boldsymbol{E} - \boldsymbol{E}^0. \tag{6}$$

When there is no multilayer but a homogeneous background, the field $\boldsymbol{E}^0$ is equal to the incident field and the scattered field are both as defined previously.

Inside the PML complex stretching coordinates are defined by which the ordinary curl operator in Maxwell's equations for the scattered field in the PML are modified and the scattered field is absorbed. The permittivity and permeability inside the PML are simply copied from the values in the underlying multilayer. The most elegant and uniform formulation of the differential equation inside the PML is however obtained by applying an additional complex transformation [4,5] by which the modified curl operator becomes again the ordinary curl operator. Due to this transformation, the scattered field is transformed in a modified scattered field and the permittivity and permeability in the PML transform in permittivity and permeability tensors $\tilde{\boldsymbol{\epsilon}}_r$, $\tilde{\boldsymbol{\mu}}_r$. These tensors depend on the material properties of the underlaying multilayer and the complex stretching coordinates and they are complex diagonal tensors. Hence, the artificial material in the PML is always anisotropic, even when the materials in the multilayer are all isotropic. Furthermore, the tensors in the PML do not correspond to real materials but are artificial metamaterials. Since the PML is a non-reflecting absorbing layer, it is not visible from within the CD. This invisibility has inspired a large research effort to realize artificial materials with the properties of the PML to obtain an optical clock [6]. Note that there can be different layers of the multilayer background under the PML. One then simply takes the local permittivity and transforms it to a tensor using the complex stretching coordinate transformation. The vector Helmholtz equation for the (transformed) scattered field $\tilde{\boldsymbol{E}}^{\mathrm{s}}$ in the PML is:

$$\omega^2\epsilon_0\mu_0\tilde{\boldsymbol{\epsilon}}_r\tilde{\boldsymbol{E}}^{\mathrm{s}} - \boldsymbol{\nabla}\times\tilde{\boldsymbol{\mu}}_r^{-1}\boldsymbol{\nabla}\times\tilde{\boldsymbol{E}}^{\mathrm{s}} = \boldsymbol{0}. \tag{7}$$

As stated above, the PML absorbs the scattered field and one can use as boundary condition on the outer boundary of the PML that the tangential electric field (Dirichlet condition) or the tangential magnetic field (Neumann condition) is zero. For this to be sufficiently accurate the PML has to be chosen thick enough and the imaginary part of the material tensors in the PML must be sufficiently large to guarantee a sufficient fast decrease of the field. A large absorption requires that sufficiently many grid points are used in the discretization of the PML. Roughly speaking the product of the thickness and the absorption coefficient in the PML should be constant and approximately five grid points are needed over the thickness of the PML for sufficient accuracy. In general, using the Neumann condition gives slightly more accurate results than the Dirichlet condition.

Although the PML is a very convenient and robust method to terminate the computational

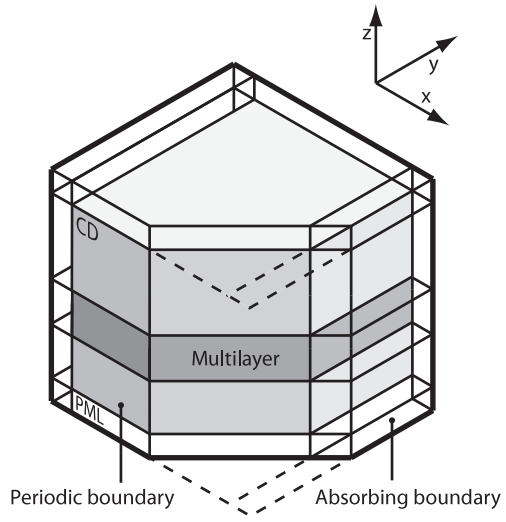

Figure 1. Computational domain with PML and periodic boundaries, embedded in a multilayer.

domain, there are some cases where its application is difficult. An example is when a strong wave is generated which propagates parallel to the interface between the PML and the CD. This happens when a plasmon surface wave is generated at the interface between a metal and a dielectric and when this interface is parallel to the interface between the PML and the CD [7]. The plasmon has a very long tail in general in the dielectric which is hardly damped inside the PML because it does not propagate perpendicular to the PML [7]. In such a case one has to rotate the computational domain so that the wave does not propagate parallel to the interface between the PML and the CD, or else a very thick PML with a fine discretization is needed and the Neumann condition should be used.

Summarizing, the application of the PML yields a boundary value problem for the vector Helmholtz equation in the cubic region ECD, with as unknown the transformed scattered field $\tilde{\boldsymbol{E}}^{\mathrm{s}}$ in the PML and the total electric field $\boldsymbol{E}$ in the CD, respectively. These two fields are linked by the requirement that the tangential components of the total electric and total magnetic fields are continuous across the interface between the CD and the PML. Finally, on the outer boundary we impose as boundary condition for the transformed scattered electric field the Neumann condition.

We briefly describe the modifications that are needed when the configuration is periodic with respect to the coordinates x and y. Let p_x and p_y be the periods in the x- and y-directions. Then $\boldsymbol{\epsilon}_r$ and $\boldsymbol{\mu}_r$ are p_x- and p_y-periodic with respect to x and y, respectively. Let $\boldsymbol{k}_{\parallel} = k_x\hat{\boldsymbol{x}} + k_y\hat{\boldsymbol{y}}$. Then if the current density and the incident field are quasi-periodic in the sense that

$$\boldsymbol{J}(x + p_x, y + p_y, z) = \boldsymbol{J}(x, y, z)\exp[\mathrm{i}(k_x p_x + k_y p_y)], \tag{8}$$

$$\boldsymbol{E}^{\mathrm{i}}(x + p_x, y + p_y, z) = \boldsymbol{E}^{\mathrm{i}}(x, y, z)\exp[\mathrm{i}(k_x p_x + k_y p_y)], \tag{9}$$

the total and the scattered field are quasi-periodic as well. Quasi-periodic means here 'periodic apart from a phase shift'. Functions with this property are also called Bloch waves. An arbitrary field or function can always be written as an integral over Bloch waves, with $\boldsymbol{k}_{\parallel}$ in the Brillouin zone [5,8,9]. Hence, if the current density is localized or there is a non-periodic incident field such as a focused spot, the current and the incident field and the total electric field that is generated by them are all written as integrals over the Brillouin zone of Bloch fields and every quasi-periodic electric field in the expansion is obtained by solving a quasi-periodic boundary value problem on a unit cell of the structure with quasi-periodic boundary conditions given by

$$\hat{\boldsymbol{x}} \times \boldsymbol{E}(x + p_x, y, z) = \hat{\boldsymbol{x}} \times \boldsymbol{E}(x, y, z)\exp(\mathrm{i}k_x p_x), \tag{10}$$

$$\hat{\boldsymbol{y}} \times \boldsymbol{E}(x, y + p_y, z) = \hat{\boldsymbol{y}} \times \boldsymbol{E}(x, y, z)\exp(\mathrm{i}k_x p_x), \tag{11}$$

and similarly for the magnetic field $\boldsymbol{H} = (-\mathrm{i}/\omega)\boldsymbol{\mu}_r^{-1}\nabla \times \boldsymbol{E}$. These quasi-periodic boundary conditions follow from the continuity of the electric and magnetic field components.

The CD is now a rectangular region which is equal to one cell of the periodic structure in x- and y-directions. By extending the CD in only the positive and negative z-direction, i.e. only in the direction in which the configuration is not periodic, the ECD is obtained. Hence, the PML now consists of two unconnected rectangular regions. When the configuration is only periodic with respect to one coordinate, say the x-coordinate, then the rectangular CD is equal to one cell in the x-direction and so large in the y- and z-direction that all non-trivial scatterers are contained in the CD. The ECD is then obtained by extending the CD in the positive and negative y- and z-direction. All fields are expanded in fields that are quasi-periodic with respect to the x-coordinate and for every k_x in the Brillouin zone a boundary value problem is solved on the ECD.

The weak formulation of the boundary value problem, whether for a quasi-periodic configuration or for a non-periodic case, is derived in the usual way. The ECD is meshed by a tetrahedral or hexahedral mesh which is constructed such that a tetrahedral or

hexahedral never crosses an interface between different media. In general hexahedral meshes give more accurate results for the same number of unknowns. Since in many cases of interest there are oblique and even curved interfaces, meshing the configuration such that an element never crosses an interface is not easy and some mesh generators may yield meshes of low quality. This is true in particular for tetrahedral meshes, hexahedral meshes give in general for the same number of unknowns more accurate results. For a tetrahedral mesh of good quality, the minimum ratio of the volume and the third power of the diameter of the tetrahedra is sufficiently large.

On the mesh so-called curl-conforming finite elements must be used because they assure that the tangential field components are continuous. Furthermore, the curl-conforming elements satisfy the condition on the divergence of the field which follows from Maxwell's equations in a weak sense. This is important to assure convergence of the numerical solutions when the grid size is decreased [4,10].

The curl-conforming elements are more difficult to implement than classical nodal elements, but classical nodal elements should not be used for 3D problems because the tangential components of the field are in general not continuous across interfaces and furthermore unphysical, so-called spurious solutions may occur in some cases. There exist curl-conforming elements of every order $p \geq 1$. The restrictions of the elements of order p to a mesh element is a polynomial of degree $\leq p$ in all three coordinates. The convergence of the numerical solution to the actual physical field measured in the energy norm of the electric field, is for elements of degree p equal to h^p, i.e. we have

$$\left(\iiint_{CD} |\boldsymbol{E}(\boldsymbol{r}) - \boldsymbol{E}_h(\boldsymbol{r})|^2 \, \mathrm{d}^3\boldsymbol{r}\right)^{1/2} \leq Ch^p \left(\iiint_{CD} |\boldsymbol{E}(\boldsymbol{r})|^2 \, \mathrm{d}^3\boldsymbol{r}\right)^{1/2}, \qquad (12)$$

where C is a constant which depends on the configuration but is independent of the current source and incident field as well as of the total volume (or equivalently, of the wavelength) [4]. This is in marked contrast with classical nodal elements for which the convergence rate of elements of order p is $p+1$. For in particular the lowest order curl-conforming elements (degree 1), the convergence rate is only linear in the mesh size h, which means that the convergence is *very* slow indeed. Although the lowest order elements are (in contrast with the higher order elements) easy to programme, the linear convergence rate makes them inappropriate for use. One should use elements that are at least of second order. We have compared second- and third-order elements and found that the second-order are most satisfactory. With third-order elements much more CPU is needed to build the matrix and to solve the system because the system matrix has more non-zeros. When second-order elements are used, it is in general sufficient to use a grid with 10–20 points per wavelength (measured in the material) to obtain a very accurate solution. To achieve the same accuracy, the number of grid points per wavelength has to be increased when the computational domain is increased. This is due to the fact that the constant C on the right-hand side of error estimate (12) increases when the computational domain becomes bigger.

A major issue is how to solve the system of equations numerically for 3D problems. For 2D problems the system is generally so small that one can often use a direct method like Gaussian elimination. For 3D problems, Gaussian elimination requires far too much storage and is also much too slow. Therefore, one has to resort to an iterative method [11]. However, except for the uninteresting case of very short wavelengths for which the problem is almost static, the system matrix is not positive definite and hence an iterative solver does not converge without so-called preconditioning of the system. Preconditioning means that the system is changed in an equivalent system that is positive definite so that an iterative solver does converge. A famous preconditioning method is the so-called ILUTP decomposition which was devised by Saad [12]. First the elements of the system matrix that are smaller than a certain threshold are replaced by zeros. Then an incomplete LU-decomposition is computed. 'Incomplete' means that all elements of the decomposed matrix that are outside of a band of width chosen by the user, are set to zero. This is done to save memory. But only when the band is large enough, the ILUTP-decomposition will be sufficiently close to the inverse of the system matrix and the iterative solver such as BICGSTAB [13] will converge. Applying the Schatz method [4] or multi-grid methods may speed up convergence, but these methods are very difficult to implement. The problem with ILUTP and other preconditioning methods is that the user has to guess appropriate values for certain parameters, such as the mentioned threshold and the width of the band on non-zeros. If the guess is wrong, the iterative solver fails to converge. Convergence is in general only achieved when the ILUTP decomposition is similar to the inverse of the system matrix. But it takes long times to compute such a decomposition. An example was described in [5] where the scattering of a focused spot by a pit structure on an optical disc was considered. The CD consisted of a Si disc with a groove containing three cylindrical pits in a row with a diameter of 200 nm and with two neighboring grooves without pits. The wavelength was 405 nm at

which the refractive index of silicon is $n = 5.42 + 0.329\mathrm{i}$. The CD is approximately $650 \times 650 \times 100\,\mathrm{nm}^3$. The number of unknowns is approximately one million and it took 20 h on a HP DL585 2.4 GHz processor with 48 Gbytes to solve the system. 90% of the CPU was spend in computing the ILUTP. Lowest order elements were used which meant that, even with this huge number of unknowns, the error measured in the energy norm exceeded 5%.

A much better method for solving the large system of equations has been devised and implemented in the free package Pardiso [14]. Pardiso is a clever mixture of an iterative and a direct solver which does not require the user to set parameter values which are difficult to choose and which is moreover very reliable, fast and memory efficient. Only since we started to use Pardiso and began to implement second-order curl conforming elements, has the FEM model worked very satisfactorily. With Pardiso applied to 3D problems, memory requirements are typically a factor of 5 less than for ILUTP in combination with BICGSTAB. Computation times with Pardiso are even more than a factor of 10 shorter than with the old method and require approximately five times less memory!

Often the scattered far field is of interest. Different methods exist to deduce the far field from the scattered near field computed inside the CD.

- One can write the scattered field in the exterior of the CD as a volume integral over an effective dipole density which is proportional to the difference between the permittivity in the CD and the local value of the permittivity in the underlaying multilayer, multiplied by the local total field:

$$\boldsymbol{E}^{\mathrm{s}}(\boldsymbol{r}) = -\iiint \omega^2 \Delta\boldsymbol{\epsilon}_r(\boldsymbol{r}')\mu_0 \mathbf{G}_E(\boldsymbol{r},\boldsymbol{r}')\boldsymbol{E}(\boldsymbol{r}')\mathrm{d}^3\boldsymbol{r}', \qquad (13)$$

  where $\mathbf{G}_E(\boldsymbol{r},\boldsymbol{r}')$ is the Green's tensor for the electric field in the multilayer as defined above in (4) and (5). In the case that the configuration is periodic and a quasi-periodic boundary value problem is considered, the Green's tensor should be quasi-periodic as well for the same $\boldsymbol{k}_\parallel$ in the Brillouin zone. This means that for a constant unit vector $\hat{\boldsymbol{v}}$, $\mathbf{G}_E\hat{\boldsymbol{v}}$ is the electric field of a periodic row of electric dipoles with dipole moments that differ in phase according to the phase factor $\exp[\mathrm{i}(k_x p_x + k_y p_y)]$.

  In the Fraunhofer approximation, the field of a dipole and hence the dyadic Green's tensor can be approximated by plane waves and the volume integral (13) then simplifies to an integral over plane waves.
- Instead of a volume integral over the CD, the scattered field in the exterior of the CD can also be expressed in terms of the tangential scattered electric and magnetic fields on the boundary of the CD using Stratton–Chu's formula (2). In the far field region, the Green tensors can again be approximated by plane waves.

Note that in a multilayer, closed formulae can only be derived for the Fourier transforms of the Green tensors, i.e. for the expansions of the Green tensors in plane waves. The field itself can only be obtained from a back Fourier transformation in planes parallel to the interfaces of the multilayer (i.e. in planes $z = \text{constant}$). Hence, in the case of a multilayer background, the Fraunhofer scattered far field is much easier to compute than the scattered field at closer distances to the CD.

2.2. FDTD

A standard reference for the FDTD method is [15]. Our FDTD implementation applies to the same general set of configurations, materials, sources and incident fields as our FEM code. Again a rectangular CD is defined and a PML is used in the directions in which the configuration is not periodic. When the configuration is periodic with respect to some of the coordinates, the same expansion of the sources and fields in terms of quasi-periodic fields is used as described above.

The FDTD is more general than the FEM in that it can be used for sources and incident fields that have a finite frequency band, i.e. which are not time-harmonic. In fact, even in the case that the sources and incident fields are time-harmonic, the time-harmonic field is computed in the FDTD method by solving the Maxwell equations as an initial value problem. The computation is continued until the fields have become time-harmonic, i.e. until the transient has disappeared. In the FDTD, the electric and the magnetic field are computed together but on a staggered rectangular grid.

We list some important features of the FDTD.

(1) The implementation is relatively simple, in particular for non-dispersive media.
(2) Memory requirements are very low because explicit time stepping is used and hence only the electric and magnetic fields at the previous and current times have to be kept in memory. In particular, no system matrix has to be stored.

(3) Convergence of the FDTD is guaranteed for non-dispersive dielectrics provided that the time-step is chosen sufficiently small compared to the mesh size [15].

It is often said that the convergence rate is second order for the lowest order version of the FDTD. This compares favorably with the FEM where the lowest order elements give only a linear convergence rate. However, quadratic convergence holds only when the mesh is uniform and for a single homogeneous material. In the neighborhood of an interface between different materials, the convergence is not second order due to the fact that the grid is staggered. However, by a clever averaging of the material parameters on both sides of the interface, second-order convergence can nevertheless be achieved [16]. In spite of this improvement, it remains a disadvantage of the FDTD compared to the FEM that the rectangular FDTD mesh can not approximate oblique and curved interfaces as accurately as a FEM mesh generator using tetrahedra can.

For frequencies for which the real part of the permittivity of a metal is negative, the FDTD is unstable. This is due to the fact that when the initial value problem is solved for such a material, causality is violated. The solution to this problem is to restore causality by taking dispersion into account. This can be done at little additional computational costs by solving at every time step not only for the electric and magnetic fields but also for the polarization density. Hence, the polarization density is stored at the previous and current time in addition to the electric and the magnetic field. For time-harmonic problems it is not important which dispersion model is used (the Lorentz model, Drude's model or another model) as long as the value of the permittivity at the frequency of interest is correct and the dispersion relation satisfies the Kramers–Kronig relations.

In the FDTD, the PML is dealt with slightly differently as in the FEM. Due to the fact that in the FDTD a general time-dependent field occurs, the PML is formulated in the time-domain and contains an operator that is a convolution in time [17]. A negative real part of the permittivity is allowed by implementing a so-called auxiliary differential equation (ADE) technique [18] for dispersive media. Quasi-periodic boundary conditions can be implemented as well. We used the so-called cosine-sinus method [19] to compute the field in a periodic structure for a given incident plane wave (or more generally an incident quasi-periodic field).

Anisotropic materials are also not a problem for the FDTD. Furthermore, as in the FEM case, the scattered near field can be propagated outside in the exterior of the computational domain by using one of the methods (volume integral or Stratton–Chu's formula) that were mentioned above.

In summary it can be said that for many problems the FDTD is a very convenient and robust computational method which requires only little memory. It is easy to implement. However, if one wants to apply the method to oblique interfaces and to metals with negative real part of the permittivity, additional precautions have to be taken. Furthermore, the staircase mesh is not accurate for curved interfaces. For time-harmonic problems more accurate solutions can be obtained with the FEM (provided curl-conforming elements of at least second order and the Pardiso solver are used). But for problems that are not time-harmonic, FDTD is the only practical method. Also when not enough storage for the FEM is available, FDTD is a good alternative.

2.3. VIE and SIE

In the volume integral method, the scattered electric field is written as an integral over electric fields of electric dipoles inside the volume of the non-trivial scatterers [20]. When the scatterer(s) are embedded in a multilayer, the multilayer is incorporated in the field of the dipoles (i.e. in the dyadic Green's tensor). The dipole strengths in the integral are then only the additional strengths compared to what they otherwise would be when there are no non-trivial scatterers. The additional induced dipole density $\boldsymbol{P}$, is proportional to the difference $\Delta\boldsymbol{\epsilon}_r$ between the permittivity tensor inside the scatterer and the background, multiplied by the total electric field $\boldsymbol{E}$: $\boldsymbol{P}=\epsilon_0\Delta\boldsymbol{\epsilon}_r\boldsymbol{E}$. By substituting this in the volume integral over all dipole fields, one obtains (13). By using $\boldsymbol{E}=\boldsymbol{E}^{\mathrm{s}}+\boldsymbol{E}^0$, where $\boldsymbol{E}^0$ is again the field that would exists in the multilayer background when the non-trivial scatterers would be absent, we obtain a volume integral equation (VIE) for the total electric field $\boldsymbol{E}$.

The VIE can be solved numerically using curl-conforming elements. An advantage is that only the region occupied by the non-trivial scatterers has to be meshed, hence the number of unknowns is less than in the case of FEM and the FDTD. For curl-conforming elements, the solution obtained with the VIE is more accurate than with the FEM when meshes with similar grid spacings are used. Therefore, in contrast with the FEM, the lowest order curl-conforming elements can give already quite accurate results. However, the system matrix of the VIE is full and therefore memory requirements can be large. Since this matrix is not positive definite, an iterative solver cannot be used without preconditioning. An alternative is to

replace the original system matrix by its product with its Hermitian conjugate. This product is positive definite and the resulting system of equations can be solved iteratively by e.g. the conjungate gradient method [21]. Another iterative method was proposed by Martin et al. [22]. It is based on increasing the volume of the scatterers iteratively and computing the Green tensor in the configuration with the scatterer. However, this method is in general inefficient because if one counts the number of operations that are needed, it turns out to be as costly as a direct method. Martin et al.'s method may be useful however when the solutions for a large set of sources and/or incident fields are required in the same configuration.

Instead of using volume integral equations one can in certain cases also use integral equations on interfaces between homogeneous media [23]. The unknown fields are the tangential components of the electric or magnetic fields. Solving the surface integral equations (SIE) requires the least amount of memory of all rigorous computational methods discussed, because the three-dimensional scattering problem is formulated as a two-dimensional numerical problem. The system of linear equations can be solved using the fast multipole techniques [24,25]. However, the SIE is very difficult to implement, in particular in a multi-layer background and for anisotropic materials and can not be used at all when some of the materials are inhomogeneous.

2.4. RCWA

The Rigorous Coupled Wave Analysis (RCWA) is a very popular computational method for periodic structures. It belongs to the broader class of modal expansion methods to which many different authors have contributed, notably Kogelnik [26] and Knop [27]. A version of the RCWA for so-called conical and classical mounting was published by Gaylord and and co-workers [28,29]. Li [30] contributed a lot to the mathematical understanding of the method. Although the RCWA seems by nature to be restricted to periodic configurations, it was successfully extended by using a PML with either Dirichlet of Neumann boundary conditions on the outer boundary, to non-periodic problems as well [31].

For certain simple 1D periodic structures with interfaces parallel to the coordinate axes, the RCWA seems to be the fastest method. A grating with oblique interfaces is not so well suited because in the RCWA the configuration is approximated by slices. However, to a structure that is simply titled, a coordinate transformation can be applied as explained by Gaylord and Moharam [29].

For structures that are periodic in two directions, the RCWA is rather slow and certainly not the fastest method.

3. On the numerical integration of diffraction integrals

Consider an optical system with the z-axis as optical axis. Suppose U is a field component that is given in some plane $z=z_1$ and that is emitted by sources in $z<z_1$. The field at a point $\boldsymbol{r}_2$ with $z_2>z_1$ can be computed by using the Rayleigh–Sommerfeld (RS) diffraction formula:

$$U(\boldsymbol{r}_2) = -\iint_{\Omega} U(\boldsymbol{r}_1)\frac{2(z_2-z_1)}{\|\boldsymbol{r}_2-\boldsymbol{r}_1\|}\left(\mathrm{i}k-\frac{1}{\|\boldsymbol{r}_2-\boldsymbol{r}_1\|}\right) \times \frac{\exp[\mathrm{i}k\|\boldsymbol{r}_2-\boldsymbol{r}_1\|]}{\|\boldsymbol{r}_2-\boldsymbol{r}_1\|}\,\mathrm{d}x_1\,\mathrm{d}y_1, \tag{14}$$

where k is the wave number in the medium through which the field is propagating and Ω is the region where the field is non-zero (or non-negligible). This formula is rigorous for all propagation distances z_2-z_1. The RS formula is the rigorous mathematical formulation of Huygens' Principle which can be applied to every field component separately. Alternatively, one can use (14) to compute E_x and E_y and then deduce E_z from:

$$E_z(\boldsymbol{r}_2) = -\frac{1}{4\pi^2}\iint \frac{\left\{\begin{array}{l} k_x\mathcal{F}(E_x)(k_x,k_y,z_2) \\ \quad +k_y\mathcal{F}(E_y)(k_x,k_y,z_2)\end{array}\right\}}{k_z}\,\mathrm{d}k_x\,\mathrm{d}k_y, \tag{15}$$

where $\mathcal{F}$ is the Fourier transformation with respect to x, y. This formula follows by Fourier transformation of $\nabla\cdot\boldsymbol{E}=0$. Usually the plane wave spectrum has numerical aperture <1 so that $k_z>0$ for all k_x, k_y for which the plane wave amplitudes are not zero. Only when the propagation distance z_2-z_1 is very small, contributions of evanescent waves may be important. In that case care has to be taken that the singularity corresponding to $k_z=0$ is numerically treated appropriately. In any case, the singularity is an integrable singularity.

As one can propagate the electric field components individually using (14), one may wonder what the benefits are of Stratton–Chu's formula. Stratton–Chu's formula is indispensable when one requires a rigorous propagation of a field that is given on a curved surface, for example on (part of) a sphere. One then needs the electric and magnetic field components that are tangential to the surface. Also when the medium in which the field is propagated is anisotropic, Stratton–Chu's formula should be used. But when

we propagate from a given plane that is perpendicular to the optical axis in an isotropic medium, there is no advantage in using Stratton–Chu's formula over the scalar formula. The numerical approach described here for the Rayleigh–Sommerfeld diffraction integral can also be used for Stratton–Chu's integral, even when the integration is over a curved surface. For simplicity we omit the details.

If the distance $z_2 - z_1$ is so large that the spherical waves can be approximated by plane waves or parabolic waves, the Fraunhofer and Fresnel approximations are obtained, respectively. Fresnel's approximation satisfies the paraxial wave equation and conversely every solution of the paraxial wave equation can be written in terms of a Fresnel diffraction integral. Hence, for paraxial fields such as unapodized laser beams, Fresnel's approximation is also valid for short propagation distances. For apodized fields the distance to the initial plane should be large enough.

In some cases Fresnel's approximation is not accurate enough and diffraction integral (14) has to be solved numerically. This is not an easy task because when k is large, the phase term $k\|\boldsymbol{r}_2 - \boldsymbol{r}_1\|$ can be fast oscillating when $\boldsymbol{r}_1$ runs through the region of integration. We describe here a numerical integration method which gives a uniform error *independent of the value of k*.

The boundary of Ω can have any shape in principle, for example it can be a circle or a polygon and the numerical integration method that we will describe applies to any shape. But for simplicity we shall consider here the case of a rectangle. When there is no aperture in the $z = z_1$-plane so that the field is not apodized in this plane, Ω is usually chosen to be a sufficiently large rectangle.

We first choose polar coordinates centred on the projection of the point of observation (x_2, y_2, z_2) on the $z = z_1$-plane, so that

$$x_1 = r_1 \cos\varphi_1, \qquad y_1 = r_1 \sin\varphi_1. \tag{16}$$

Then we define the variable ϱ_1 by

$$\varrho_1 = (r_1^2 + (z_2 - z_1)^2)^{1/2} - (z_2 - z_1),$$

and change the integration variables from x_1, y_1 to ϱ_1, φ_1. The integrand becomes then simple but the boundary of the region in terms of the new coordinates is difficult to describe. But polygonal and circular boundaries can be dealt with using the new coordinates.

Let $r = r_1^\ell$, $\ell = 1, \ldots, N$ be the circles with center (x_2, y_2) which are either tangential to the rectangle or intersect it in (at least) one of its corner points (see Figure 2). For every strip $\varrho_1^\ell < \varrho_1 < \varrho_1^{\ell+1}$ there are

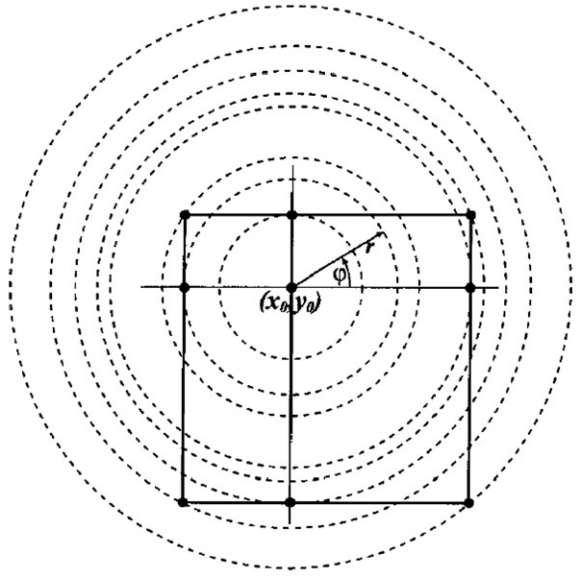

Figure 2. Rectangular region with polar coordinate system centred at (x_0, y_0) and with the circles $r = r_1^\ell$, $\ell = 1, \ldots, N$.

a fixed number of angular sectors M_ℓ such that for every ϱ_1 in the strip the integral over φ_1 is a sum over M_ℓ sectors $\varphi_1^j(\varrho_1) < \varphi_1 < \varphi_1^{j+1}(\varrho_1)$, $j = 1, \ldots, M_\ell$. Using $\Delta z = z_2 - z_1$ and

$$\mathrm{d}x_1\,\mathrm{d}y_1 = r_1\,\mathrm{d}r_1\,\mathrm{d}\varphi_1 = \frac{\mathrm{d}r_1}{\mathrm{d}\varrho_1}\,\mathrm{d}\varrho_1\,\mathrm{d}\varphi_1 = (\varrho_1 + \Delta z)\mathrm{d}\varrho_1\,\mathrm{d}\varphi_1,$$

$$\|\boldsymbol{r}_2 - \boldsymbol{r}_1\| = \varrho_1 + \Delta z,$$

the integral can be written as a sum over all strips:

$$U(\boldsymbol{r}_2) = \sum_{\ell=1}^{N} \int_{\varrho_1^\ell}^{\varrho_1^{\ell+1}} F(\varrho_1) \exp(\mathrm{i}k\varrho_1)\mathrm{d}\varrho_1, \tag{17}$$

with, for $\varrho_1^\ell < \varrho_1 < \varrho_1^{\ell+1}$, $F(\varrho_1)$ a sum over integrals over sectors $\varphi_1^j(\varrho_1) < \varphi_1 < \varphi_1^{j+1}(\varrho_1)$:

$$F(\varrho_1) = \sum_{j=1}^{M_\ell} \int_{\varphi_1^j(\varrho_1)}^{\varphi_1^{j+1}(\varrho_1)} f(\varrho_1, \varphi_1)\mathrm{d}\varphi_1, \tag{18}$$

with

$$\begin{aligned} f(\varrho_1, \varphi_1) = {} & \frac{\mathrm{i}\Delta z \exp(\mathrm{i}k\Delta z)}{\lambda} \\ & \times U\big(x_1 + [(\varrho_1 + \Delta z)^2 - (\Delta z)^2]^{1/2} \cos\varphi_1, \\ & y_1 + [(\varrho_1 + \Delta z)^2 - (\Delta z)^2]^{1/2} \sin\varphi_1, z_1\big) \\ & \times \left[1 + \frac{\mathrm{i}}{k(\varrho_1 + \Delta z)}\right] \frac{1}{\varrho_1 + \Delta z}. \end{aligned} \tag{19}$$

Note that f is a relatively slowly varying function of the integration variables ϱ_1, φ_1, the fast oscillations occur only due to the factor $\exp(\mathrm{i}k\varrho_1)$. The field $U(x_1, y_1, z_1)$

is now approximated by a bi-cubic spline. For a ϱ_1 of interest the corresponding integrals over the angular sectors $\varphi_1^j(\varrho_1) < \varphi_1 < \varphi_1^{j+1}(\varrho_1)$ are computed by Gaussian integration. It was shown [32] that the number of integration points needed in the Gaussian integration to achieve a relative error <1% was in a number of typical cases ≤20. The numerical computation of integral ℓ in (17) over ϱ_1 requires a rather fine mesh when $\varrho_1 = \varrho_1^\ell$ or $\varrho_1 = \varrho_1^{\ell+1}$ (or both) are tangential to the rectangle. The reason is that when e.g. the circle $\varrho_1 = \varrho_1^\ell$ is tangential to the rectangle, we have for small $\varrho_1 - \varrho_1^\ell > 0$

$$F(\varrho_1) = a_0 + a(\varrho_1 - \varrho_1^\ell)^{1/2} + \cdots, \tag{20}$$

hence the derivative of F is infinite for $\varrho_1 = \varrho_1^\ell$. We therefore write

$$\begin{aligned}
&\int_{\varrho_1^\ell}^{\varrho_1^{\ell+1}} F(\varrho_1)\exp(\mathrm{i}k\varrho_1)\mathrm{d}\varrho_1 \\
&= \int_{\varrho_1^\ell}^{(\varrho_1^\ell+\varrho_1^{\ell+1})/2} F(\varrho_1)\exp(\mathrm{i}k\varrho_1)\mathrm{d}\varrho_1 \\
&\quad + \int_{(\varrho_1^\ell+\varrho_1^{\ell+1})/2}^{\varrho_1^{\ell+1}} F(\varrho_1)\exp(\mathrm{i}k\varrho_1)\mathrm{d}\varrho_1 \\
&= \exp(\mathrm{i}k\varrho_1^\ell)\int_0^{[(\varrho_1^{\ell+1}-\varrho_1^\ell)/2]^{1/2}} 2sF(\varrho_1^\ell + s^2)\exp(\mathrm{i}ks^2)\mathrm{d}s \\
&\quad + \exp(\mathrm{i}k\varrho_1^{\ell+1})\int_0^{[(\varrho_1^{\ell+1}-\varrho_1^\ell)/2]^{1/2}} 2sF(\varrho_1^\ell - s^2)\exp(\mathrm{i}ks^2)\mathrm{d}s.
\end{aligned} \tag{21}$$

The functions $s \mapsto 2sF(\varrho_1^\ell \pm s^2)$ are approximated again by cubic splines. For a cubic spline $\sigma(s)$ the integral

$$\int_0^{[(\varrho_1^{\ell+1}-\varrho_1^\ell)/2]^{1/2}} \sigma(s)\exp(\mathrm{i}ks^2)\mathrm{d}s,$$

can be expressed in Fresnel integrals and derivatives of Fresnel integrals, i.e. analytic expressions can be used. It can be shown [32] that for a mesh size $h>0$ used in the definition of the spline, the error made in the integral by replacing the actual integrand by the spline is of the order of $h^2/k^{1/2}$ and is independent of the distance Δz.

4. Modeling of a lens of high numerical aperture

To characterize a lens for a certain object plane, it is sufficient to consider its corresponding point-spread function, i.e. the images of point sources in the object plane. For high numerical aperture lenses, the change of polarization can not be neglected. The point source that is to be imaged is then an electric dipole in some point of the object plane. It clearly suffices to choose the dipole vector parallel to the three axes of the cartesian coordinate system. We start from the reasonable assumption that the field in image space is similar to that of a time-reversed electric dipole situated at the Gaussian image point. Provided that the Gaussian image point is far from the lens so that the Fresnel number is large, the field on a Gaussian reference sphere through the exit pupil of the lens and with as centre the Gaussian image point, is similar to that of a time-reversed electric dipole field in the far field approximation. This field is tangential to the sphere and satisfies

$$\hat{\boldsymbol{R}} \times \boldsymbol{H} = \left(\frac{\mu_0}{\epsilon}\right)^{1/2} \boldsymbol{E}, \tag{22}$$

where $\hat{\boldsymbol{R}}$ is the unit vector pointing from a point $\boldsymbol{r}_Q$ on the sphere to its centre. The field in observation points $\boldsymbol{r}_P$ not too far from the Gaussian image point, can then be obtained by integrating the field over the part of the Gaussian reference sphere that is inside the exit pupil using Stratton–Chu's formulae (2), (3) in which $\mathbf{G}_E(\boldsymbol{r}_P, \boldsymbol{r}_Q)$ and $\mathbf{G}_H(\boldsymbol{r}_P, \boldsymbol{r}_Q)$ occur. With (22), the magnetic field can be eliminated and by expanding the exponent in the factor $\exp(k|\boldsymbol{r}_P - \boldsymbol{r}_Q|)$ in the dyadic Green tensors $\mathbf{G}_E(\boldsymbol{r}_P, \boldsymbol{r}_Q)$ and $\mathbf{G}_H(\boldsymbol{r}_P, \boldsymbol{r}_Q)$ by assuming that the Fresnel number is large i.e. $(R+z_P)/\lambda$ is large, the Debye integral is finally obtained [33,34]. Using generalized Jones calculus and energy conservation along the rays, the polarization and amplitude of the electric field on the Gaussian reference sphere can be expressed in terms of the electric field of the electric dipole in the entrance pupil of the lens.

It should be noted that when the expansion of the exponent mentioned above can not be used because the Fresnel number is not sufficiently large, the integral over the Gaussian reference sphere can be computed numerically using a method that is analogous to the one explained in Section 3.

Suppose that we use solid angle $\mathrm{d}\Omega = R\sin\vartheta\,\mathrm{d}\vartheta\,\mathrm{d}\varphi$ to integrate the Debye integral over the Gaussian reference sphere. If the electric dipole is on the optical axis, the integrals with respect to φ can be computed analytically. The remaining integrals over ϑ have to be computed numerically. Alternatively, the electric field in the entrance (or exit) pupil of the lens can be expanded in Zernike polynomials and the field close to the Gaussian image plane can be written as a sum over Bessel functions using the ENZ theory [35]. This approach is by far superior over the FFT (i.e. Hopkins method) with respect to accuracy of the field values that can be obtained in the focal region. When accurate simulations of the (vectorial) point spread function of a high NA lens are required, the ENZ is the ideal method. It is also relatively easy to compute the point spread function inside a multilayer [36,40]. The ENZ is

furthermore applicable to aberration retrieval using through focus intensity measurements. Compared to the FFT, the ENZ is much slower when the field is to be determined in only one plane. But since the basis functions in the expansion depend only on the numerical aperture, when different object fields are imaged by the same optical system ENZ becomes very fast because the basis functions can then be reused.

5. A complicated optical system: the imaging of a mask in photolithography for integrated circuits

In advanced lithography the transmissive (or for EUV reflective) structures on the mask that have to be imaged in the photo-resist are of the order of the wavelength. Usually the source is spatially incoherent. This implies that the source has to be divided into sufficiently many point sources and that for every point source, the image of the mask in the photo-resist has to be computed. The total light intensity is the sum of the intensities due to all point sources. 'Sufficiently many point sources' means that the extended source has to be split up into regions which are so small that in the image plane of the source, they cannot be distinguished. Often (but not always) the illumination is Köhler illumination which means that every point source induces a plane wave illumination of the mask.

Because of the sub-wavelength features in modern masks, Kirchhoff boundary conditions to model the transmission by the mask are not accurate enough and a rigorous electromagnetic model such as the FEM or the FDTD has to be used [37]. The illumination optics has usually low NA and therefore the illuminating field at the mask of every point source can be computed by geometrical optics. As the source is unpolarized, we need in the Köhler illumination case to compute the field transmitted (or reflected) by the mask twice for every incident plane wave (corresponding to every point source), namely for two orthogonal polarizations. The computed transmitted (or reflected) fields are object fields for imaging by the high NA objective into the photo-resist. The latter is done using the vectorial diffraction ENZ model described in Section 4. In fact, first the far field version of Stratton–Chu's formula is used to propagate the computed near field towards the entrance pupil of the lens. The thus obtained field is expanded in Zernike polynomials and mapped onto a series of Bessel fields in image space close to the Gaussian image plane using the ENZ theory. The total intensity in image space is obtained by summing the intensities corresponding to both orthogonal polarizations and all point sources of the extended spatially incoherent source. In case the multilayer of the photo-resist and substrate has non-flat interfaces, or when the change of refractive index of the resist due to the induced bleaching is important, the field inside the resist has to be computed using a rigorous solver and adding the intensities afterwards.

One of the main problems of the imaging of a mask in lithography is the very large field of view which in general requires large computational volumes when the illumination of the mask is simulated. One way of dealing with this problem is to use a so-called partition of unity of the mask region [38]. A partition of unity is a finite collection of *smooth* functions $\{\psi_j\}$ defined on a plane $z = \text{constant}$ (the z-axis is the optical axis) such that $0 \le \psi_j(x, y) \le 1$, for every j the set V_j of points (x, y) where ψ_j is non-zero is bounded, $\psi_j(x, y) = 1$ for all (x, y) in a set $U_j \subset V_j$ and such that

$$\sum_j \psi_j(x, y) = 1$$

for all (x, y) in the region of the mask. For a given incident field $\boldsymbol{E}^{\mathrm{i}}$ we can thus write

$$\boldsymbol{E}^{\mathrm{i}}(x, y, z) = \sum_j \psi_j(x, y)\boldsymbol{E}^{\mathrm{i}}(x, y, z),$$

and use $\psi_j(x, y)\boldsymbol{E}^{\mathrm{i}}(x, y, z)$ as the incident field in order to isolate the part of the mask where ψ_j is 1, i.e. U_j. The advantage of using smooth functions ψ_j instead of discontinuous step functions is that in this way more physically realistic incident fields are simulated. A discontinuous incident field obtained by simply restricting the incident field to a sub-domain would cause a strong diffraction (spreading) of the field when transmitted by the mask. In Figure 3 an example of a lithographic system is shown with Köhler illumination. The wavelength is 992 nm. The mask is made of 80 nm thick chromium on a SiO_2 substrate with so-called double hammerhead contact holes (Figure 4 (top)). These features are often used as test objects because they closely resemble the interconnects between different layers of a chip. The holes have a diameter of 180 nm and a center-to-center distance of 720 nm. The permittivity of the chromium and substrate at 192 nm is $-3.32 + 3.42\mathrm{i}$ and 2.25, respectively. The field transmitted through the mask is computed using the FDTD. Then the far field version of Stratton–Chu's method is used to compute the field components in the entrance pupil of the objective (Figure 4 (center)). This field is decomposed using Zernike polynomials using the ENZ theory. In Figure 4 (bottom) the aerial intensity in focus (center) and at a distance of a quarter wavelength in front (left) and behind the focal plane (right), are shown for a numerical aperture of 0.95 and a magnification of 0.25.

An important problem is the optimization of the mask design. The shape of the holes in the mask are

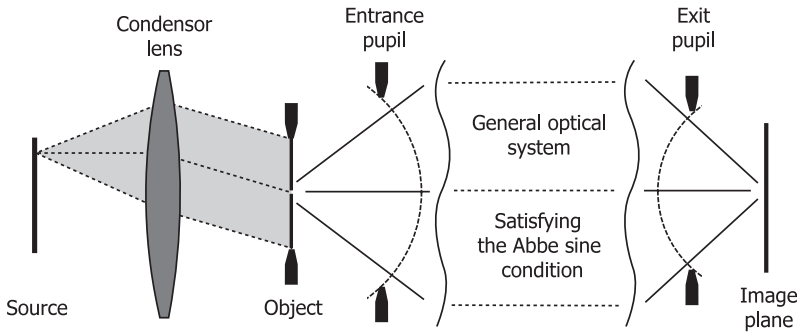

Figure 3. Optical system for lithography.

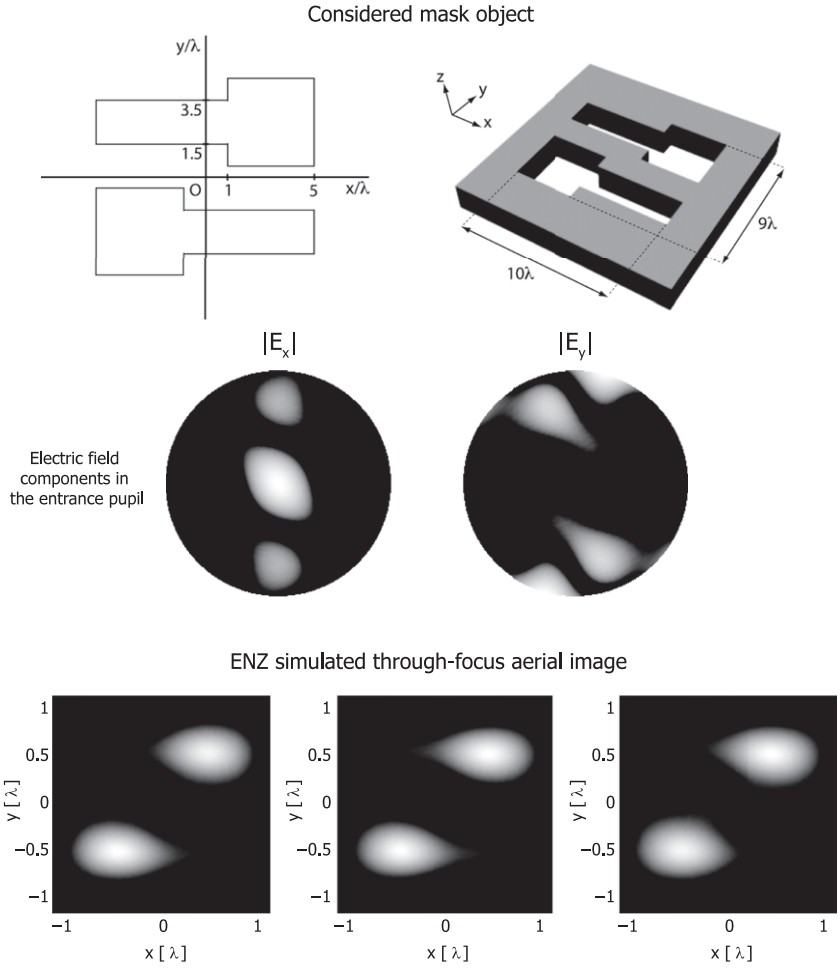

Figure 4. Top: configuration of the mask. Center: electric field components in the entrance pupil of the objective. Bottom: aerial image (intensity) through focus; the focal distance is ± a quarter wavelength. The source is assumed to be a point source on the optical axis (coherent case).

optimized using so-called optical proximity correction (OPC) while the objective lens is kept fixed. In this case in particular, the ENZ theory is very efficient because the images of the Zernike pupil polynomials can be computed and stored once and reused for every new mask design.

In this model the bulk of the computational resources is needed to compute the transmission by the mask using the FDTD [37]. Compared to this the resources for the ENZ expansion are negligible. The computation for every effective point source in the extended source and for every polarization can be done parallel.

Optical lithography for integrated circuits is an important example of a complicated optical system in which several models have to be combined to obtain reliable results at affordable computational resources.

References

[1] Besbes, M.; Hugonin, J.P.; Lalanne, P.; van Haver, S.; Janssen, O.T.A.; Nugrowati, A.M.; Xu, M.; Pereira, S.F.; Urbach, H.P.; van de Nes, A.S.; Bienstman, P.; Granet, G.; Moreau, A.; Helfert, S.; Sukharev, M.; Seideman, T.; Baida, F.I.; Guizal, B.; Van Labeke, D. *J. Eur. Opt. Soc.* **2007,** *2,* 07022.

[2] Colton, D.; Kress, R. *Integral Equation Methods in Scattering Theory*; Krieger: Florida, 1992.

[3] Berenger, J.P. *J. Comput. Phys.* **1994,** *114,* 185–200.

[4] Monk, P. *Finite Element Methods for Maxwell's Equations*; Oxford University Press: Oxford, UK, 2003.

[5] Wei, X.; Wachters, A.J.; Urbach, H.P. *J. Opt. Soc. Am. A* **2007,** *29,* 866–881.

[6] Pendry, P. *Science* **2006,** *312,* 1780–1782.

[7] Janssen, O.T.A.; Urbach, H.P.; 't Hooft, G. *Opt. Express* **2006,** *14,* 11823–11832.

[8] Jannopoulos, J.D.; Johnson, S.G.; Winn, J.N.; Meade, R.D. *Photonic Crystals*; Princeton University Press: Princeton, NJ, 2008.

[9] Janssen, O.T.A.; Wachters, A.J.H.; Urbach, H.P. *Opt. Express* **2010,** *18,* 24522–24535.

[10] Nedelec, J.C. *Numer. Math.* **1980,** *35,* 315–341.

[11] van der Vorst, H.A. *SIAM J. Sci. Stat. Comput.* **1992,** *13,* 631–644.

[12] Saad, Y. *Iterative Methods for Sparse Linear Systems*; SIAM: Philadelphia, 2003.

[13] Erlangga, Y.A.; Vuik, C.; Oosterlee, C.W. *Appl. Numer. Math.* **2004,** *50,* 409–425.

[14] Schenk, O.; Gärtner, K. *Future Gener. Comput. Syst.* **2004,** *20,* 475–487.

[15] Taflove, A.; Hageness, S.C. *Computational Electrodynamics*; Artech House: London, 1995.

[16] Deinega, A.; Valuev, I. *Opt. Lett.* **2007,** *32,* 3429–3431.

[17] Roden, J.; Gedney, S.D. *Microwave Opt. Technol. Lett.* **2009,** *27,* 334–339.

[18] Fuji, M.; Tahara, M.; Sakagami, I.; Freude, W.; Russer, P. *IEEE J. Quantum Electron.* **2004,** *40,* 175–182.

[19] Harms, P.; Mittra, R.; Ko, W. *IEEE Tr. Ant. Prop.* **1994,** *42,* 1317–1324.

[20] Van de Nes, A.S. Rigorous Electromagnetic Field Simulations of Advanced Optical Systems. Thesis, Delft University of Technology, Delft, 2005.

[21] Abubakar, A.; van den Berg, P.M. *J. Comput. Phys.* **2004,** *195,* 236–262.

[22] Martin, O.J.F.; Girard, C.; Dereux, A. *Phys. Rev. Lett.* **1995,** *74,* 526–529.

[23] Ylä-Oijala, P.; Taskinen, M.; Järvenpää, S. *Radio Sci.* **2005,** *40,* RS6002, doi: 10.1029/2009RS003169.

[24] Song, J.M.; Chew, W.C. *Microwave Opt. Technol. Lett.* **2007,** *10,* 14–19.

[25] Carpentieri, B.; Duff, I.S.; Giraud, L.; Sylvand, G. *SIAM J. Sci. Comput.* **2005,** *27,* 774–792.

[26] Kogelnik, H. *Bell Syst. Tech. J.* **1969,** *48,* 2909–2947.

[27] Knop, K. *J. Opt. Soc. Am.* **1978,** *68,* 1206–1210.

[28] Moharam, M.G.; Grann, E.B.; Pommet, D.A.; Gaylord, T.K. *J. Opt. Soc. Am. A* **1983,** *73,* 1105–1112.

[29] Gaylord, T.K.; Moharam, M.G. *Proc. IEEE* **1985,** *73,* 894–937.

[30] Li, L. *J. Opt. Soc. Am. A* **1993,** *10,* 2581–2591.

[31] Lalanne, P.; Silberstein, E. *Opt. Lett.* **2000,** *25,* 1092–1094.

[32] Veerman, J.A.C.; Rusch, J.A.; Urbach, H.P. *J. Opt. Soc. Am. A* **2005,** *22,* 636–646.

[33] Sheppard, C.J.R.; Török, P. *Optik* **1997,** *104,* 175–177.

[34] Török, P. *Opt. Lett.* **2000,** *25,* 1463–1466.

[35] Braat, J.J.M.; Van Haver, S.; Janssen, A.J.E.M.; Dirksen, P. *Prog. Opt.* **2008,** *51,* 349–466.

[36] Van Haver, S.; Braat, J.J.M.; Dirksen, P.; Janssen, A.J.E.M. *J. Eur. Opt. Soc. Rapid Publ.* **2006,** *1,* 06004.

[37] Janssen, O.T.A.; Van Haver, S.; Janssen, A.; Braat, J.J.M.; Urbach, H.P.; Pereira, S.F. *Proc. SPIE* **2010,** *6924,* 692410.

[38] Adams, R.A. *Sobolev Spaces*; Academic Press: New York, 1975.

[39] Nugrowati, A.M. Vectorial Diffraction of Extreme Ultraviolet Light and Ultrashort Light Pulses. Thesis, Delft University of Technology, Delft, 2008.

[40] Braat, J.J.M.; van Haver, S.; Janssen, A.J.E.M.; Pereira, S.F. *J. Eur. Opt. Soc. Rapid Publ.* **2009,** *4,* 09048.