

# Finite-element model for three-dimensional optical scattering problems

Xiuhong Wei

*Optical Research Group, Department of Imaging Science and Technology, Delft University of Technology,  
P.O. Box 5046, 2600 GA Delft, The Netherlands*

Arthur J. H. Wachters and H. Paul Urbach

*Philips Laboratories, Professor Holstlaan 4, 5656 AA Eindhoven, The Netherlands*

Received December 1, 2005; revised August 14, 2006; accepted August 16, 2006;  
posted September 20, 2006 (Doc. ID 66174); published February 14, 2007

We present a three-dimensional model based on the finite-element method for solving the time-harmonic Maxwell equation in optics. It applies to isotropic or anisotropic dielectrics and metals and to many configurations such as an isolated scatterer in a multilayer, bi-gratings, and crystals. We discuss the application of the model to near-field optical recording. © 2007 Optical Society of America

OCIS codes: 260.2110, 000.3860, 000.4430, 210.4770, 240.7040.

## 1. INTRODUCTION

In micro-optics, light interacts with structures with size of the order of the wavelength. The modeling of these optical systems is very important in photonics research, and rigorous models based on Maxwell's equations have to be applied when accurate results are needed.

The configurations and materials that can occur are very diverse. Sometimes the scatterer is a single object, such as an optical tweezer that is illuminated by a focused spot. But scattering structures can also be periodic in one (gratings), two (bi-gratings), or three directions (crystals). The scatterers are often illuminated by an external source, but there can also be sources inside the structure such as in biosensors. The materials that occur can be (absorbing) dielectrics or metals. Many materials are isotropic, but crystals often are optically anisotropic. Sometimes the scatterer is optically inhomogeneous. An example is the readout of bit patterns in a phase-change optical disc. The bits on these discs are small optically inhomogeneous amorphous regions in a crystalline background, which have been written by locally heating the disc by using a laser spot. These examples show that a general and flexible model is desirable with which many configurations and materials can be studied.

In many problems of micro-optics the scattering objects are not surrounded by homogeneous materials but are embedded in a planar multilayer. The interference effects due to light that has reflected at the interfaces of the multilayer can lead to strong modulations of the field amplitude, and therefore it must be possible to incorporate the multilayer background in the model.

The computational modeling of optical diffraction problems using Maxwell's equations is a difficult subject. Even though in micro-optics the regions of interest usually have sizes of the order of the wavelength, the amounts of memory and CPU that are required are often huge for

three-dimensional (3D) configurations. The literature on rigorous computational models is extensive. Often models are specific; i.e., they apply to a certain configuration and to a certain class of materials. Important configurations are one-dimensional (1D) diffraction gratings and bi-gratings. An often-used computational model for gratings is the rigorous-coupled wave method or the Fourier method.<sup>1–4</sup> The Fourier method is specific for periodic structures. Methods that solve integral equations or differential equations are more general and can be applied to periodic as well as nonperiodic structures. The advantage of volume integral equations (VIEs) and surface integral equations (SIEs) is that the radiation condition for the scattered field is automatically satisfied. The Green's tensor that occurs in the integral equations has a strong singularity. The SIE method is attractive because it reduces a 3D scattering problem to a relatively small problem for which the unknowns are tangential field components on surfaces and/or interfaces. However, the SIE is difficult to implement and cannot be used when there is an inhomogeneous material. Because the matrix of the discretized volume integral is full, the VIE often requires a lot of storage and CPU. By using a regular grid, storage and CPU requirements can be reduced considerably, but then interfaces that are not parallel to faces of the grid are approximated by a so-called staircase. A famous numerical method to solve integral equations is the fast multipole method, which was introduced by Rokhlin and co-workers<sup>5,6</sup> to reduce the computational costs.<sup>7</sup> See Ref. 8 for an interesting review of integral equation methods.

The finite-difference time domain (FDTD) and the finite-element method (FEM) are methods that solve the differential equations directly. In contrast with the integral equation methods, truncation of the computational domain (CD) is needed, and this must be done such that unphysical reflections are negligible. The FDTD<sup>9–11</sup> was

introduced by Yee. It is the most widely used method for solving Maxwell's equations in optics and also for monochromatic light. Berenger's perfectly matched layer<sup>12</sup> (PML) is a good method for truncating the domain without causing reflections. Although quite a number of additional unknowns are needed inside the PML, it is more accurate than traditional absorbing boundary conditions. The FDTD can also be applied to metals with permittivities with negative real parts by taking dispersion into account.<sup>13,14</sup>

In the FEM the mesh used is usually unstructured so that many configurations can be modeled adequately. The FEM is most useful when a relatively complicated configuration has to be modeled accurately. For example, in the optical recording problem studied in this paper, the sidewalls of the grooves and pits on the optical disc have oblique angles. Approximating such a configuration by a regular grid would cause errors. In the FEM this configuration can readily be meshed by using tetrahedra without any modeling error. Furthermore, the use of unstructured meshes in the FEM permits the mesh to be adapted to the material properties and configuration. The FEM is also very flexible regarding the kind of materials that can be modeled. Anisotropic, inhomogeneous materials and metals with permittivities that have negative real parts can all be taken into account without special precautions. The FEM has been applied to optical problems by several authors.<sup>15,16</sup> To guarantee convergence and the absence of spurious modes in the computed solution, special so-called curl-conforming Nédélec–Mur edge elements<sup>17–19</sup> should be used. These elements are more difficult to implement than standard nodal elements that are normally used in the FEM. This is an important reason why the FEM is less frequently used in optics than in other fields. Because FEM have been studied extensively in the mathematical literature, a lot is known about their mathematical properties.<sup>20–22</sup> By using edge elements of higher order, more accurate solutions can be obtained for the same amount of memory.<sup>23–25</sup>

We shall present in this paper a FEM model with which all materials and configurations of micro-optics that have been mentioned above can be studied. The different geometric configurations lead to different boundary conditions. Sometimes the configuration is periodic in a particular direction, whereas the electromagnetic field is not. An example is the scattering of a focused spot by a 1D grating. This type of problem will be treated by decomposing the electromagnetic field into quasi-periodic, so-called Bloch waves and solving the corresponding quasi-periodic boundary-value problems on a cell of the grating. We will use a rather abstract formalism by which all cases can be dealt with and which leads to an efficient implementation in computer code such that the major part of the code is shared by all configurations.

In Section 2 the geometric configurations, the material properties, and the sources and incident fields are described. The (CD) is defined, and the incident field and the source are decomposed in Bloch waves. In Section 3 a quasi-periodic boundary-value problem is formulated on the CD. Exact boundary conditions on the nonperiodic boundaries of the CD could be formulated by using the analogue of the Dirichlet-to-Neumann map (also called

the Caldéron map). This yields a nonlocal boundary condition that is rigorously equivalent to the radiation condition for the scattered field and that can easily be incorporated in the variational formulation of the boundary value problem. This procedure leads to what is called the mixed FEM–BEM approach.<sup>26</sup> Although this formulation of the nonperiodic boundary conditions is rigorous and therefore deserves to get preference over approximate methods, we have used the PML in the present setup because it is easier to implement and since it gives, in general, satisfactory results. By using complex stretched coordinates,<sup>27</sup> a modified version of the vector Helmholtz equation is obtained in the PML, which causes damping of the scattered field without generating nonphysical reflections. The disadvantage of this formulation is that the differential equation in the PML contains a modified curl operator that requires an adaptation of the edge elements. We therefore redefine the scattered field in the PML such that the differential equation contains only the classical curl operator so that the standard curl-conforming elements can be applied throughout the computational domain.

In Section 4 the variational formulation of the boundary-value problem for the vector Helmholtz equation on the extended domain is derived. By the redefinition of the scattered field in the PML, the differential equation inside the CD is very similar to that in the PML. However, the unknowns in the CD and the PML are chosen to be different, namely, the total and the scattered fields, respectively. One could choose the scattered field as unknown throughout the extended computational domain (ECD). As will be explained in Section 4, this more simple formulation has some disadvantages when the scatterers are embedded in a multilayer as they usually are. In deriving the variational formulation, the differential equation is multiplied by a test vector field and partially integrated. In this way the highest- (second-) order derivatives disappear from the problem and hence need not be approximated. The periodic boundary conditions of Dirichlet type are satisfied by the test vector fields, and those of Neumann type are imposed implicitly in the weak sense in the variational formulation. The (electric or magnetic) field that is to be computed is then approximated by a linear combination of finite-element basis functions on the given mesh, with coefficients that are to be determined. This approximation is substituted in the weak formulation, and the test vector fields in the variational formulation are chosen in the space of base functions. This then yields a discretized system of equations for the unknown coefficients. The linear system is huge but sparse because the finite-element basis functions are nonzero only on a few elements.

A major problem is solving the large sparse system of equations obtained after discretization of the variational formulation. One distinguishes direct and iterative methods. Direct methods are variants of Gaussian elimination and are very robust. However, the amount of memory and the number of operations required by direct methods scale badly with the number of unknowns, especially for 3D problems. Because in many optical problems the number of unknowns may easily exceed  $10^5$ , a direct solver cannot be used, and one of the, less robust, iterative

methods must be applied. Iterative methods often converge very slowly due to a small eigenvalue or because there are eigenvalues with negative and positive real parts.<sup>28</sup> For Helmholtz type of problems with transmission boundary conditions such as occur in optics, acoustics, etc., the differential operator usually is indefinite, and therefore the system matrix often has eigenvalues with positive and negative real parts.<sup>29</sup> The system must then be transformed into an equivalent system with more favorable properties for convergence of the iterative solver. This transformation is called preconditioning. Often the preconditioning consists of multiplying the system matrix by a matrix that approximates the inverse matrix. The thus-obtained transformed system matrix becomes sufficiently similar to the identity matrix for the iterative solver to converge rapidly. Due to the preconditioning, the system usually becomes (much) less sparse; hence more memory is required for storing the preconditioned matrix than for the original sparse matrix. We briefly discuss the main features of the iterative method and the preconditioner that we have used in Section 4. More details will appear in a separate paper.

Often, the scattered field has to be determined in the exterior of the CD. It may, for example, be desirable to compute the scattered far field. In Section 5 it is in particular explained how the scattered far field can be computed efficiently by using a Fourier-transformed version of the Stratton–Chu formula.

In Section 6 we describe the application of the FEM model to the study of resolution enhancement in optical recording by using a solid immersion lens (SIL). The SIL is kept at a distance of approximately 25 nm above the rotating disc by using actuators. The waves of high spatial frequency are tunneled through the air gap between the SIL and the disc, causing a smaller scanning spot than with conventional far-field optics. We present results of computations of the scattered near and far fields. In contrast with conventional recording, the computed detected reflected intensity is largest when the spot is focused at the center of a pit instead of between the pits.

## 2. SCATTERING CONFIGURATIONS

We shall consider 3D configurations that either consist of isolated scatterers (Fig. 1) or are periodic with respect to one (Fig. 2), two (bi-gratings, Fig. 3), or three directions (photonic crystals). All configurations will be dealt with in one formalism.

In the following  $(x_1, x_2, x_3)$  will be a Cartesian coordinate system. Except for the case where the system is pe-

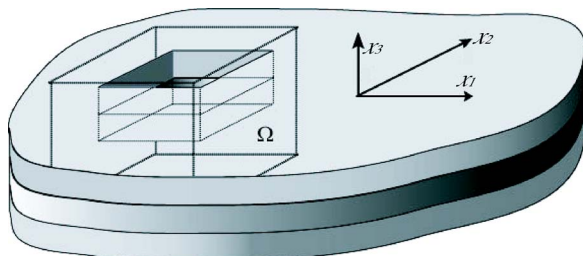


Fig. 1. (Color online) Example of a geometry of the CD  $\Omega$  for a nonperiodic configuration.

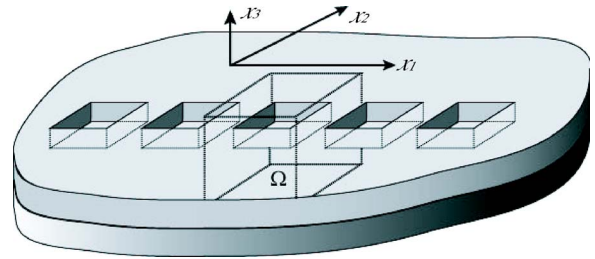


Fig. 2. (Color online) Example of a configuration that is periodic in the  $x_1$  direction. The CD  $\Omega$  has width along the  $x_1$  direction equal to the period.

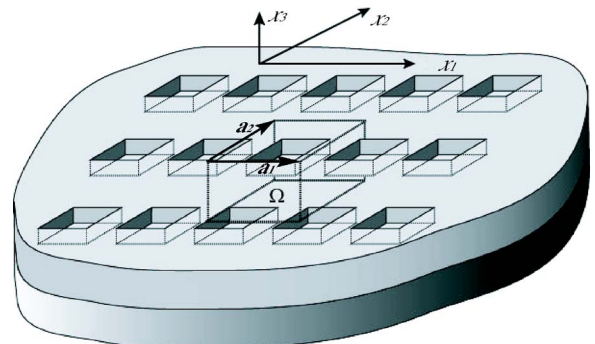


Fig. 3. (Color online) Example of a configuration that is periodic in two directions in the  $(x_1, x_2)$  plane. The vectors  $\mathbf{a}_1$  and  $\mathbf{a}_2$  span the unit cell  $\mathcal{A}$  in the  $x_3=0$  plane and  $\Omega = \mathcal{A} \times (x_3^s, x_3^l)$ .

riodic in three directions, the  $x_3$  axis is always a direction in which the structure is bounded and not periodic. When there is an optical system with an optical axis, the  $x_3$  axis coincides with the optical axis.

The sources of the electromagnetic fields may be inside or outside the scatterers. The sources inside the scattering structure are specified by a current density. In contrast, the currents of the sources that are outside the scatterers are often not specified. Instead, the radiated field that is incident on the scatterers is then assumed known. This incident field can, for example, be an arbitrarily polarized plane wave or a spot focused by an imaging system.

In optical problems, the scatterers are often inside a planar multilayer, and it is thus essential to incorporate such a multilayer in the model. The multilayer is often referred to as the background medium.

For all mentioned configurations, a computational domain (CD)  $\Omega$  will be defined in which the electromagnetic field is computed. The CD is truncated by using a PML. The union of the CD  $\Omega$  and the PML will be called the extended computational domain (ECD)  $\tilde{\Omega}$ . The domains will be considered in Subsection 2.B, and the PML will be described in Subsection 3.B.

When the configuration is periodic in at least one direction, the scattering problem can be solved only when the sources and incident fields are quasi-periodic; i.e., after a translation over one period, they are the same except for a (constant) change of phase. When the sources and/or incident fields are more general, they must first be expanded in quasi-periodic so-called Bloch functions. After solving the scattering problem for every quasi-periodic source and incident field, the total field is obtained by coherently

summing the solutions of the separate quasi-periodic scattering problems. This procedure will be explained in Subsection 2.C.

All sources and fields in this manuscript are time harmonic and are taken to depend on time  $t$  by the factor  $\exp(-i\omega t)$  for some  $\omega > 0$ . This factor will be omitted from all formulas.

### A. Multilayer Background

The multilayer background consists of  $N+1$  planar layers with interfaces at  $x_3 = x_3^{(j)}$ ,  $j=0, \dots, N-1$ , parallel to the  $(x_1, x_2)$  plane. The numbering is such that

$$x_3^{(0)} > x_3^{(1)} > \dots > x_3^{(N-1)}. \quad (1)$$

Hence  $x_3 > x_3^{(0)}$  and  $x_3 < x_3^{(N)}$  are half-spaces. We define for convenience  $x_3^{(-1)} = +\infty$  and  $x_3^{(N)} = -\infty$ . The following assumption is made:

Every layer  $x_3^{(j)} < x_3 < x_3^{(j+1)}$  (i.e., including the half-spaces) consists of homogeneous isotropic material with relative dielectric permittivity  $\epsilon_r^{(j)}$  and relative magnetic permeability  $\mu_r^{(j)}$ . When layer  $j$  is absorbing,  $\epsilon_r^{(j)}$  and/or  $\mu_r^{(j)}$  are complex numbers with nonnegative imaginary parts.

The assumption that all layers of the multilayer are isotropic is made only for simplicity. In principle, anisotropic materials could be allowed. It will be convenient to define piecewise-constant functions of  $x_3$  that give the relative permittivity and the relative permeability in each layer of the multilayer:

$$\left. \begin{aligned} \tilde{\epsilon}_r(x_3) &= \epsilon_r^{(j)} \\ \tilde{\mu}_r(x_3) &= \mu_r^{(j)} \end{aligned} \right\} \text{ for } x_3^{(j)} < x_3 < x_3^{(j+1)}, \quad j = 0, \dots, N. \quad (2)$$

### B. Isolated and Periodic Scatterers

Suppose that the configuration is periodic with respect to  $n_p$  directions, for some  $n_p \in \{0, 1, 2, 3\}$ :

$n_p=0$ . In this case the configuration is not periodic; hence it consists of isolated scatterers embedded in a multilayer background.

$n_p=1$ . In this case the configuration is periodic in one direction, which is assumed to be the  $x_1$  direction. With respect to the basis  $(x_1, x_2, x_3)$ , we define the vector  $\mathbf{a}_1 = (p, 0, 0)^T$ , where  $p$  is the period of the configuration.

$n_p=2$ . In this case there are two linearly independent vectors  $\mathbf{a}_1$  and  $\mathbf{a}_2$  in the  $x_3=0$  plane such that the configuration is invariant under translations over multiples of these two vectors. These structures are called bi-gratings or two-dimensional photonic crystals.

$n_p=3$ . Then there are three linearly independent vectors  $\mathbf{a}_1$ ,  $\mathbf{a}_2$ , and  $\mathbf{a}_3$  such that the configuration is invariant under translations over multiples of these three vectors. Examples of these structures are 3D photonic crystals.

When  $1 \leq n_p \leq 3$ , we define the  $n_p$ -dimensional lattice by the set

$$\mathcal{L}_{n_p} = \left\{ \ell = \sum_{i=1}^{n_p} \ell_i \mathbf{a}_i; \quad \text{for integer } \ell_i \right\}. \quad (3)$$

The fundamental cell of this lattice is

$$\mathcal{A} = \left\{ \mathbf{r} = \sum_{i=1}^{n_p} y_i \mathbf{a}_i; \quad -1/2 < y_1, \dots, y_{n_p} < 1/2 \right\}. \quad (4)$$

The reciprocal lattice consists of the vectors

$$\mathbf{m} = \sum_{i=1}^{n_p} m_i \mathbf{b}_i, \quad (5)$$

where  $m_i$  are integers and vectors  $\mathbf{b}_i$ ,  $i=1, \dots, n_p$ , are such that

$$\mathbf{b}_i \cdot \mathbf{a}_j = 2\pi \delta_{ij}, \quad \text{for } 1 \leq i, j \leq n_p,$$

$$\mathbf{b}_i \cdot \hat{\mathbf{e}}_j = 0, \quad \text{for } 1 \leq i \leq n_p \text{ and for } n_p + 1 \leq j \leq 3, \quad (6)$$

where  $\hat{\mathbf{e}}_j$  is the unit vector along the  $x_j$  axis.

We have,  
if  $n_p=1$ ,

$$\mathbf{b}_1 = (2\pi/p, 0, 0)^T; \quad (7)$$

if  $n_p=2$ ,

$$\mathbf{b}_1 = 2\pi \frac{\mathbf{a}_2 \times \hat{\mathbf{e}}_3}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \hat{\mathbf{e}}_3)}, \quad (8)$$

$$\mathbf{b}_2 = 2\pi \frac{\mathbf{a}_1 \times \hat{\mathbf{e}}_3}{\mathbf{a}_2 \cdot (\mathbf{a}_1 \times \hat{\mathbf{e}}_3)}; \quad (9)$$

if  $n_p=3$ ,

$$\mathbf{b}_1 = 2\pi \frac{\mathbf{a}_2 \times \mathbf{a}_3}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)}, \quad (10)$$

$$\mathbf{b}_2 = 2\pi \frac{\mathbf{a}_1 \times \mathbf{a}_3}{\mathbf{a}_2 \cdot (\mathbf{a}_1 \times \mathbf{a}_3)}, \quad (11)$$

$$\mathbf{b}_3 = 2\pi \frac{\mathbf{a}_1 \times \mathbf{a}_2}{\mathbf{a}_3 \cdot (\mathbf{a}_1 \times \mathbf{a}_2)}. \quad (12)$$

The fundamental cell of the reciprocal lattice is the Brillouin zone of dimension  $n_p$  defined by

$$\mathcal{B} = \left\{ \mathbf{k} = \sum_{i=1}^{n_p} \eta_i \mathbf{b}_i; \quad -1/2 \leq \eta_1, \dots, \eta_{n_p} < 1/2 \right\}. \quad (13)$$

The nontrivial scatterers may be arbitrary inhomogeneous and/or anisotropic dielectrics, conductors, or magnetic materials. The relative permittivity and permeability are, in general, complex two-tensors  $\underline{\epsilon}_r(\mathbf{r})$ , which are written as



$$\underline{\epsilon}_r(\mathbf{r}) = \text{Re}[\underline{\epsilon}_r(\mathbf{r})] + i \text{Im}[\underline{\epsilon}_r(\mathbf{r})], \quad (14)$$

$$\underline{\mu}_r(\mathbf{r}) = \text{Re}[\underline{\mu}_r(\mathbf{r})] + i \text{Im}[\underline{\mu}_r(\mathbf{r})], \quad (15)$$

where  $\text{Re}(\underline{\epsilon}_r)$ ,  $\text{Im}(\underline{\epsilon}_r)$ ,  $\text{Re}(\underline{\mu}_r)$ , and  $\text{Im}(\underline{\mu}_r)$  are real-valued tensors. The following assumptions are made:

1.  $\underline{\epsilon}_r$  and  $\underline{\mu}_r$  are periodic on the lattice; i.e., for every  $\ell$  in  $\mathcal{L}_{n_p}$ , see Eq. (3):

$$\underline{\epsilon}_r(\mathbf{r} + \ell) = \underline{\epsilon}_r(\mathbf{r}), \quad \underline{\mu}_r(\mathbf{r} + \ell) = \underline{\mu}_r(\mathbf{r}). \quad (16)$$

2.  $\underline{\epsilon}_r - \tilde{\epsilon}_r$  and  $\underline{\mu}_r - \tilde{\mu}_r$  have bounded support as functions of the coordinates with respect to which the configuration is not periodic, i.e., with respect to  $x_i$  for  $i = n_p + 1, \dots, 3$ . Hence for every  $i = n_p + 1, \dots, 3$ , if  $|x_i|$  is sufficiently large, then  $\underline{\epsilon}_r(\mathbf{r}) = \tilde{\epsilon}_r(x_3) \underline{1}$  and  $\underline{\mu}_r(\mathbf{r}) = \tilde{\mu}_r(x_3) \underline{1}$ , where  $\underline{1}$  is the unit tensor.

3.  $\text{Re}(\underline{\epsilon}_r)$  and  $\text{Im}(\underline{\mu}_r)$  are symmetric tensors, and, since we do not consider optically active materials and because the time dependence of the fields is assumed to be given by the factor  $\exp(-i\omega t)$  for some positive  $\omega$ , the imaginary parts  $\text{Im}(\underline{\epsilon}_r)$  and  $\text{Im}(\underline{\mu}_r)$  are nonnegative symmetric tensors. Note that for metals the real part of the permittivity is often negative in the optical domain; hence  $\text{Re}(\underline{\epsilon}_r)$  is, in general, not positive definite.

4. The components of the tensors  $\underline{\epsilon}_r$  and  $\underline{\mu}_r$  and of their inverse  $\underline{\epsilon}_r^{-1}$  and  $\underline{\mu}_r^{-1}$  are essentially bounded functions, i.e.,

$$\underline{\epsilon}_r, \underline{\mu}_r, \underline{\epsilon}_r^{-1}, \underline{\mu}_r^{-1} \in L^\infty(\mathbb{R}^3). \quad (17)$$

We shall introduce a notation for the CD by which all cases of periodicity can be dealt with in one formalism. By using this notation, we do not have to list the periodic boundary conditions, and this simplifies the problem statement considerably. We first introduce the parameterization

$$\mathbf{r} \mapsto (\exp(i\mathbf{b}_1 \cdot \mathbf{r}), \dots, \exp(i\mathbf{b}_{n_p} \cdot \mathbf{r})), \quad (18)$$

which maps the fundamental lattice cell  $\mathcal{A}$  onto the  $n_p$ -dimensional torus:  $T^{n_p} = \prod_{i=1}^{n_p} S^1$ , where  $S^1$  is the unit circle in the complex plane. By considering fields on the unit cell as fields defined on the torus, the fields are automatically periodic. For example, for  $n_p = 1$  the configuration is periodic in the  $\mathbf{a}_1$  direction but not in the directions of  $\mathbf{a}_2$  and  $\mathbf{a}_3$ . We then use the circle  $T^1 = S^1$  instead of the line  $y_1 \mathbf{a}_1$ ,  $-1/2 < y_1 < 1/2$ , and all fields are functions of  $y_1$  considered to be defined on the circle so that they are automatically periodic with respect to  $y_1$ .

We shall now define the CD. We first mention that when  $n_p = 0$ , i.e., when the configuration is not periodic, the CD will be a rectangular block, and therefore  $T^{n_p} \times \prod_{i=n_p+1}^3 (x_i^s, x_i^l)$  and  $\mathbb{R}^{n_p} \times \prod_{i=n_p+1}^3 (x_i^s, x_i^l)$  must, for  $n_p = 0$ , be identified with  $\prod_{i=1}^3 (x_i^s, x_i^l)$ . When  $n_p = 3$ , i.e., when the configuration is periodic in all three directions, these sets should be identified with the sets  $T^3$  and  $\mathbb{R}^3$ , respectively.

*Definition of the CD:*

$$\Omega \equiv T^{n_p} \times \prod_{i=n_p+1}^3 (x_i^s, x_i^l), \quad (19)$$

with  $x_i^s$  and  $x_i^l$  for  $i = n_p + 1, \dots, 3$ , chosen such that the exterior of the set

$$\mathbb{R}^{n_p} \times \prod_{i=n_p+1}^3 (x_i^s, x_i^l) \quad (20)$$

consists only of planar layers of the multilayer, i.e.,  $\underline{\epsilon}_r = \tilde{\epsilon}_r$  and  $\underline{\mu}_r = \tilde{\mu}_r$  in the exterior of expression (20). Furthermore,  $x_3^s$  and  $x_3^l$  satisfy

$$x_3^s \leq x_3^{(N-1)}, \quad x_3^l \geq x_3^{(0)}; \quad (21)$$

hence, each of the half-spaces  $x_3 < x_3^s$  and  $x_3 > x_3^l$  consists of homogeneous matter.

*Remarks:*

1. When  $n_p = 3$ , we have  $\Omega = \mathcal{A}$ ; hence  $\Omega$  is then identical to the fundamental cell of the lattice.

2. Properties (21) are necessary for applying the PML.

3. For a perfect conductor (PEC), the imaginary part of the permittivity is infinite; hence expression (17) then does not hold for  $\underline{\epsilon}_r$ . A PEC can nevertheless easily be taken into account in the FEM. As will be explained below, the FEM is applied to a boundary-value problem for the vector Helmholtz equation for either the electric or the magnetic field. The region occupied by the PEC is excluded from the CD  $\Omega$  and therefore is not meshed. On the boundary of the PEC the tangential components of the electric field are set equal to zero. Hence, when the boundary-value problem for the electric field is solved, this boundary condition is a Dirichlet condition. It is explicitly incorporated in the finite elements used in the approximation. In contrast, when one solves the boundary-value problem for the magnetic field, the boundary condition on the PEC is of Neumann type and hence is taken account of implicitly in the variational formulation. The presence of a PEC thus causes no fundamental problems and is in fact advantageous because there are then fewer unknowns. In contrast, a good conductor that is not considered a PEC needs a fine mesh because of the skin effect. Metals with high but finite conductivity now receive a lot of attention in optics because of the interest in plasmon waves generated at metallic surfaces. We stress that there are no fundamental computational problems when the FEM is applied to good conductors. One needs only more computing power. For simplicity we shall not mention the case of a PEC separately in the remainder of this paper.

The boundary of the CD is

$$\partial\Omega = T^{n_p} \times \left[ \prod_{i=n_p+1}^3 [x_i^s, x_i^l] \setminus \prod_{i=n_p+1}^3 (x_i^s, x_i^l) \right]. \quad (22)$$

Note that  $[x_i^s, x_i^l]$  is the closed interval  $x_i^s \leq x_i \leq x_i^l$ , whereas  $(x_i^s, x_i^l)$  is the open interval (i.e., without end points)  $x_i^s < x_i < x_i^l$ . For example, when  $n_p = 2$ , the boundary consists of the set of all points  $(x_1, x_2, x_3)$ , with  $(x_1, x_2)$  on the torus  $T^2$  and with  $x_3 = x_3^s$  or  $x_3 = x_3^l$ . When  $n_p = 3$ , the boundary of  $\Omega$  is empty:  $\partial\Omega = \emptyset$ .

### Definition of the ECD

$$\tilde{\Omega} \equiv T^{n_p} \times \prod_{i=n_p+1}^3 (x_i^s - \delta_i^s, x_i^l + \delta_i^l), \quad (23)$$

for some  $\delta_i^s > 0$  and  $\delta_i^l > 0$ .

Hence the ECD is an extension of the CD in the direction of the coordinates  $x_i$ ,  $i = n_p + 1, \dots, 3$ , i.e., the coordinates with respect to which the structure is not periodic. The set  $\tilde{\Omega} \setminus \Omega$  is the PML. In Section 3 we will formulate a boundary-value problem on  $\tilde{\Omega}$  whose solution gives an approximation of the solution of the scattering and radiation problem. Note that when the structure is periodic in all three directions ( $n_p = 3$ ), then  $\tilde{\Omega} = \Omega$ ; i.e., the ECD is identical to the CD, and a PML is not needed.

### C. Sources and Incident Fields

An arbitrary incident field  $\mathbf{E}^i$ ,  $\mathbf{H}^i$  may exist in one or both of the half-spaces  $x_3 < x_3^{(N)}$  and  $x_3 > x_3^{(0)}$ . There may also be imposed current sources present, described by a current density  $\mathbf{J}(\mathbf{r})$ .

We will assume that these currents have the following properties:

1. When  $n_p \geq 1$ , the current either is quasi-periodic on the lattice or has compact support with respect to the coordinates  $x_i$ ,  $i = 1, \dots, n_p$ .

2. The current density must have compact support with respect to the remaining coordinates  $x_i$ ,  $i = n_p + 1, \dots, 3$ , i.e., the coordinates with respect to which the configuration is not periodic.

Note that the currents do not have to be confined to  $\Omega$ .

The ECD will be used as the CD. However, when  $n_p \geq 1$ , a boundary-value problem can only be formulated on  $\tilde{\Omega}$  for an incident field and a current density that are quasi-periodic in the sense defined next. In the following,  $\mathbf{k}$  will be a vector such that, with respect to the basis  $(x_1, x_2, x_3)$  and for the given  $n_p$ , we have

$$\mathbf{k} = (k_1, k_2, k_3)^T, \quad \text{with } k_i = 0 \text{ for all } n_p + 1 \leq i \leq 3. \quad (24)$$

Then a vector field  $\mathbf{V}_{\mathbf{k}}$  is called  $\mathbf{k}$  quasi-periodic when there exists a vector field  $\mathbf{V}$  that is periodic on the  $n_p$ -dimensional lattice, such that

$$\mathbf{V}_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k} \cdot \mathbf{r}} \mathbf{V}(\mathbf{r}), \quad (25)$$

for all  $\mathbf{r}$ . Property (25) is equivalent to

$$\mathbf{V}_{\mathbf{k}}(\mathbf{r} + \ell) = e^{i\mathbf{k} \cdot \ell} \mathbf{V}_{\mathbf{k}}(\mathbf{r}), \quad (26)$$

for all lattice vectors  $\ell$ .

An arbitrary incident field  $\mathbf{E}^i$  can always be expanded in terms of quasi-periodic fields. To show this, we introduce the Fourier transform of  $\mathbf{E}^i$  with respect to  $x_1, \dots, x_{n_p}$ . Let  $\mathbf{k}$  be as in Eq. (24), then

$$\mathcal{F}(\mathbf{E}^i)(\mathbf{k}, x_{n_p+1}, \dots, x_3) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \mathbf{E}^i(\mathbf{r}) e^{-i\mathbf{k} \cdot \mathbf{r}} dx_1 \dots dx_{n_p}. \quad (27)$$

The inverse transform can be written as follows:

$$\begin{aligned} \mathbf{E}^i(\mathbf{r}) &= \left( \frac{1}{2\pi} \right)^{n_p} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \mathcal{F}(\mathbf{E}^i) \\ &\quad \times (\mathbf{k}, x_{n_p+1}, \dots, x_3) e^{i\mathbf{k} \cdot \mathbf{r}} dk_1 \dots dk_{n_p} \\ &= \left( \frac{1}{2\pi} \right)^{n_p} \sum_{\mathbf{m}} \int_{\mathcal{B}+\mathbf{m}} \mathcal{F}(\mathbf{E}^i) \\ &\quad \times (\mathbf{k}, x_{n_p+1}, \dots, x_3) e^{i\mathbf{k} \cdot \mathbf{r}} dk_1 \dots dk_{n_p} \\ &= \int_{\mathcal{B}} \mathbf{E}_{\mathbf{k}}^i(\mathbf{r}) dk_1 \dots dk_{n_p}, \end{aligned} \quad (28)$$

where the sum is over all reciprocal-lattice vectors  $\mathbf{m}$  and  $\mathbf{E}_{\mathbf{k}}^i$  is the  $\mathbf{k}$ -quasi-periodic field given by

$$\mathbf{E}_{\mathbf{k}}^i(\mathbf{r}) = \left( \frac{1}{2\pi} \right)^{n_p} \sum_{\mathbf{m}} \mathcal{F}(\mathbf{E}^i)(\mathbf{k} + \mathbf{m}, x_{n_p+1}, \dots, x_3) e^{i(\mathbf{k}+\mathbf{m}) \cdot \mathbf{r}}. \quad (29)$$

An arbitrary current density can, of course, be written in the same way:

$$\mathbf{J}(\mathbf{r}) = \int_{\mathcal{B}} \mathbf{J}_{\mathbf{k}}(\mathbf{r}) dk_1 \dots dk_{n_p}, \quad (30)$$

with  $\mathbf{J}_{\mathbf{k}}(\mathbf{r})$  given by

$$\mathbf{J}_{\mathbf{k}}(\mathbf{r}) = \left( \frac{1}{2\pi} \right)^{n_p} \sum_{\mathbf{m}} \mathcal{F}(\mathbf{J})(\mathbf{k} + \mathbf{m}, x_{n_p+1}, \dots, x_3) e^{i(\mathbf{k}+\mathbf{m}) \cdot \mathbf{r}}. \quad (31)$$

The incident electric field and the imposed current density have thus been written as integrals over the  $n_p$ -dimensional Brillouin zone  $\mathcal{B}$  of quasi-periodic incident fields and current densities. For every  $\mathbf{k} \in \mathcal{B}$  the solution of the scattering problem with incident field  $\mathbf{E}_{\mathbf{k}}^i$  and current density  $\mathbf{J}_{\mathbf{k}}$  is  $\mathbf{k}$  quasi-periodic. The total electric field  $\mathbf{E}$  is obtained by integrating the quasi-periodic solutions  $\mathbf{E}_{\mathbf{k}}$  over the Brillouin zone:

$$\mathbf{E}(\mathbf{r}) = \int_{\mathcal{B}} \mathbf{E}_{\mathbf{k}}(\mathbf{r}) dk_1 \dots dk_{n_p}. \quad (32)$$

In the sequel we shall mainly discuss the scattering problem for quasi-periodic incident fields and current densities.

*Remarks.* For different vectors in the Brillouin zone, the matrices of the linear systems obtained after discretization of the variational formulations of the quasi-periodic boundary-value problems are different. When the periods are small, the CD is correspondingly small, but the Brillouin zone and therefore the number of quasi-periodic problems are large. The decomposition is then practical only when the code is run in parallel on many processors or when the same preconditioning matrix can be used for several quasi-periodic fields with similar  $\mathbf{k}$ . For the scattering of a focused incident spot by a structure that is periodic in one ( $n_p = 1$ ) or two ( $n_p = 2$ ) directions, one could instead define the CD so large that the field is negligible outside  $\Omega$ . Then only one, relatively big, diffraction problem must be solved. The CD should be

chosen such that not only the incident field but also the scattered field are negligible outside it. Since the scattered field decreases only with the first power of the reciprocal distance, the CD may have to be chosen quite large for sufficient accuracy.

### 3. FORMULATION OF A BOUNDARY-VALUE PROBLEM ON THE EXTENDED COMPUTATIONAL DOMAIN

In this section we shall derive, for a given  $\mathbf{k}$  in the Brillouin zone, a boundary-value problem for  $\mathbf{k}$ -quasi-periodic fields on the ECD  $\tilde{\Omega}$ .

#### A. Vector Helmholtz Equation

For given  $\mathbf{k}$ -quasi-periodic incident field  $\mathbf{E}_\mathbf{k}^i$ ,  $\mathbf{H}_\mathbf{k}^i$  and current density  $\mathbf{J}_\mathbf{k}$ , let  $\mathbf{E}_\mathbf{k}$ ,  $\mathbf{H}_\mathbf{k}$  be the total quasi-periodic field that satisfies Maxwell's equations with relative permittivity tensor  $\underline{\epsilon}_r$  and the relative permeability tensor  $\underline{\mu}_r$  having the properties mentioned in Subsection 2.B. The field  $\mathbf{E}_\mathbf{k}$ ,  $\mathbf{H}_\mathbf{k}$  is called the total field to distinguish it from the incident and the scattered fields, but it should be noted that it is the field for only one vector  $\mathbf{k}$  in the Brillouin zone.

Let  $\mathbf{E}_\mathbf{k}^0$ ,  $\mathbf{H}_\mathbf{k}^0$  be the field that would be caused by the given incident field  $\mathbf{E}_\mathbf{k}^i$ ,  $\mathbf{H}_\mathbf{k}^i$  and the imposed current density  $\mathbf{J}_\mathbf{k}$  when only the multilayer is present (i.e., there are no other scatterers). The field  $\mathbf{E}_\mathbf{k}^0$ ,  $\mathbf{H}_\mathbf{k}^0$  can be computed by a standard expansion in plane-wave Fourier components. Note that when the configuration is periodic in three directions there is no multilayer, and the field  $\mathbf{E}_\mathbf{k}^0$ ,  $\mathbf{H}_\mathbf{k}^0$  is then set equal to zero.

The scattered field is defined by

$$\mathbf{E}_\mathbf{k}^s = \mathbf{E}_\mathbf{k} - \mathbf{E}_\mathbf{k}^0, \quad \mathbf{H}_\mathbf{k}^s = \mathbf{H}_\mathbf{k} - \mathbf{H}_\mathbf{k}^0. \quad (33)$$

It satisfies

$$\nabla \times \mathbf{E}_\mathbf{k}^s = i\omega\mu_0\underline{\mu}_r\mathbf{H}_\mathbf{k}^s + i\omega\mu_0(\underline{\mu}_r - \tilde{\underline{\mu}}_r)\mathbf{H}_\mathbf{k}^0, \quad (34)$$

$$\nabla \times \mathbf{H}_\mathbf{k}^s = -i\omega\epsilon_0\underline{\epsilon}_r\mathbf{E}_\mathbf{k}^s + -i\omega\epsilon_0(\underline{\epsilon}_r - \tilde{\underline{\epsilon}}_r)\mathbf{E}_\mathbf{k}^0, \quad (35)$$

$$\mathbf{E}_\mathbf{k}^s \text{ and } \mathbf{H}_\mathbf{k}^s \text{ satisfy the outgoing radiation conditions.} \quad (36)$$

The fact that the electromagnetic energy is finite in every bounded region of space is equivalent to the mathematical statement that the electric and magnetic fields are locally square integrable:

$$\mathbf{E}_\mathbf{k} \text{ and } \mathbf{H}_\mathbf{k} \text{ are in } \mathbf{L}_{loc}^2(\mathbb{R}^3), \quad (37)$$

where  $\mathbf{L}^2(\mathbb{R}^3)$  is the space of all square-integrable vector fields  $\mathbf{V}: \mathbb{R}^3 \rightarrow \mathbb{C}^3$ . The fields  $\mathbf{E}_\mathbf{k}^0$ ,  $\mathbf{H}_\mathbf{k}^0$  and  $\mathbf{E}_\mathbf{k}^s$ ,  $\mathbf{H}_\mathbf{k}^s$  are also locally square integrable, of course.

By eliminating the magnetic field we get the vector Helmholtz equations for the electric fields:

$$\omega^2\epsilon_0\mu_0\underline{\epsilon}_r\mathbf{E}_\mathbf{k} - \nabla \times (\underline{\mu}_r^{-1} \nabla \times \mathbf{E}_\mathbf{k}) = -i\omega\mu_0\mathbf{J}_\mathbf{k}, \quad (38)$$

$$\begin{aligned} & \omega^2\epsilon_0\mu_0\underline{\epsilon}_r\mathbf{E}_\mathbf{k}^s - \nabla \times (\underline{\mu}_r^{-1} \nabla \times \mathbf{E}_\mathbf{k}^s) \\ & = -\omega^2\epsilon_0\mu_0(\underline{\epsilon}_r - \tilde{\underline{\epsilon}}_r)\mathbf{E}_\mathbf{k}^0 - i\omega\mu_0 \nabla \times [(1 - \underline{\mu}_r^{-1}\tilde{\underline{\mu}}_r)\mathbf{H}_\mathbf{k}^0]. \end{aligned} \quad (39)$$

By eliminating the electric fields we get similar equations for  $\mathbf{H}_\mathbf{k}$  and  $\mathbf{H}_\mathbf{k}^s$ . Each of these equations is a special case of the following general equation:

$$k_0^2\alpha\mathbf{U}_\mathbf{k} - \nabla \times (\underline{\beta} \nabla \times \mathbf{U}_\mathbf{k}) = \mathbf{F}_\mathbf{k} + \nabla \times \mathbf{G}_\mathbf{k}, \quad (40)$$

where  $k_0 = \omega\sqrt{\epsilon_0\mu_0}$  is the wavenumber in vacuum;  $\alpha$  and  $\beta$  are complex-valued tensors,

$$\alpha = \text{Re}(\alpha) + i \text{Im}(\alpha), \quad (41)$$

$$\beta = \text{Re}(\beta) + i \text{Im}(\beta), \quad (42)$$

where  $\text{Re}(\alpha)$  and  $\text{Re}(\beta)$  are symmetric tensors and  $\text{Im}(\alpha)$  and  $\text{Im}(\beta)$  are real symmetric nonnegative tensors; and  $\mathbf{F}_\mathbf{k}$  and  $\mathbf{G}_\mathbf{k}$  are given quasi-periodic fields.

Since  $\mathbf{U}_\mathbf{k}$ ,  $\mathbf{F}_\mathbf{k}$ , and  $\mathbf{G}_\mathbf{k}$  are  $\mathbf{k}$  quasi-periodic, we have

$$\begin{aligned} \mathbf{U}_\mathbf{k}(\mathbf{r}) &= e^{i\mathbf{k}\cdot\mathbf{r}}\mathbf{U}(\mathbf{r}), & \mathbf{F}_\mathbf{k}(\mathbf{r}) &= e^{i\mathbf{k}\cdot\mathbf{r}}\mathbf{F}(\mathbf{r}), & \mathbf{G}_\mathbf{k}(\mathbf{r}) \\ &= e^{i\mathbf{k}\cdot\mathbf{r}}\mathbf{G}(\mathbf{r}), \end{aligned} \quad (43)$$

where  $\mathbf{U}$ ,  $\mathbf{F}$ , and  $\mathbf{G}$  are periodic on lattice  $\mathcal{L}_{n_p}$ . We have

$$\nabla \times \mathbf{U}_\mathbf{k} = e^{i\mathbf{k}\cdot\mathbf{r}}(i\mathbf{k} + \nabla) \times \mathbf{U} = e^{i\mathbf{k}\cdot\mathbf{r}}\nabla_\mathbf{k} \times \mathbf{U}, \quad (44)$$

where

$$\nabla_\mathbf{k} = i\mathbf{k} + \nabla. \quad (45)$$

Hence, Eq. (40) can equivalently be written in terms of periodic fields as follows:

$$k_0^2\alpha\mathbf{U} - \nabla_\mathbf{k} \times (\underline{\beta}\nabla_\mathbf{k} \times \mathbf{U}) = \mathbf{F} + \nabla_\mathbf{k} \times \mathbf{G}. \quad (46)$$

The fact that the electric and magnetic fields are periodic and locally square integrable is equivalent to the statement

$$\mathbf{U} \in \mathbf{H}_{loc}(\text{curl}, T^{n_p} \times \mathbb{R}^{3-n_p}), \quad (47)$$

where  $\mathbf{H}_{loc}(\text{curl}, T^{n_p} \times \mathbb{R}^{3-n_p})$  is the space of vector fields, which, together with their curl, are locally square integrable on one period of the geometry and which are periodic on the lattice.

The meaning of  $\alpha$ ,  $\beta$ ,  $\mathbf{F}$ , and  $\mathbf{G}$  for the different cases when  $\mathbf{U}$  is an electric or magnetic field are listed in Table 1.

**Table 1. Values of  $\alpha$ ,  $\beta$ ,  $\mathbf{F}$ , and  $\mathbf{G}$  When the Total or Scattered Electric or Magnetic Field Is Calculated**

Parameter	$\mathbf{u} = \mathbf{E}$	$\mathbf{U} = \mathbf{E}^s$	$\mathbf{U} = \mathbf{H}$	$\mathbf{U} = \mathbf{H}^s$
$\alpha$	$\underline{\epsilon}_r$	$\underline{\epsilon}_r$	$\underline{\mu}_r$	$\underline{\mu}_r$
$\beta$	$\underline{\mu}_r^{-1}$	$\underline{\mu}_r^{-1}$	$\underline{\epsilon}_r^{-1}$	$\underline{\epsilon}_r^{-1}$
$\mathbf{F}$	$-i\omega\mu_0\mathbf{J}$	$-\omega^2\epsilon_0\mu_0(\underline{\epsilon}_r - \tilde{\underline{\epsilon}}_r)\mathbf{E}^0$	$\mathbf{0}$	$-\omega^2\epsilon_0\mu_0(\underline{\epsilon}_r - \tilde{\underline{\mu}}_r)\mathbf{H}^0$
$\mathbf{G}$	$\mathbf{0}$	$-i\omega\mu_0 \times (1 - \underline{\mu}_r^{-1}\tilde{\underline{\mu}}_r)\mathbf{H}^0$	$-\underline{\epsilon}_r^{-1}\mathbf{J}$	$i\omega\epsilon_0 \times (1 - \underline{\epsilon}_r^{-1}\tilde{\underline{\epsilon}}_r)\mathbf{E}^0$

### B. Perfectly Matched Layer

The idea of a PML was first published by Berenger.<sup>12</sup> We do not use Berenger's modification of Maxwell's equations but apply instead the formulation in terms of complex stretching coordinates introduced by Chew and Weedon<sup>27</sup> because it is very suitable for applying the FEM.

Let the operators  $\nabla^\zeta$  and  $\nabla_{\mathbf{k}}^\zeta$  be defined by

$$\nabla^\zeta = \left( \frac{\partial}{\zeta_1 \partial x_1}, \frac{\partial}{\zeta_2 \partial x_2}, \frac{\partial}{\zeta_3 \partial x_3} \right), \quad (48)$$

$$\nabla_{\mathbf{k}}^\zeta = i\mathbf{k} + \nabla^\zeta. \quad (49)$$

Here  $\zeta = (\zeta_1, \zeta_2, \zeta_3)$  is a vector field defined on the entire ECD  $\tilde{\Omega}$  but set equal to (1,1,1) inside  $\Omega$ . Inside the PML,  $\zeta_j$  is a complex-valued function of  $x_j$  by which the complex stretching is specified. Write

$$\underline{D}(\zeta) = \begin{bmatrix} \zeta_1 & 0 & 0 \\ 0 & \zeta_2 & 0 \\ 0 & 0 & \zeta_3 \end{bmatrix}; \quad (50)$$

then we have

$$\nabla_{\mathbf{k}}^\zeta \times \mathbf{V} = i\mathbf{k} \times \mathbf{V} + \frac{1}{\det(\underline{D}(\zeta))} \underline{D}(\zeta) \nabla \times [\underline{D}(\zeta) \mathbf{V}]. \quad (51)$$

Hence, for  $\mathbf{V} \in L^2(\tilde{\Omega})$ , we have

$$\nabla_{\mathbf{k}}^\zeta \times \mathbf{V} \in L^2(\tilde{\Omega}) \Leftrightarrow \nabla \times \underline{D}(\zeta) \mathbf{V} \in L^2(\tilde{\Omega}). \quad (52)$$

In the vector Helmholtz equation for the scattered field,  $\nabla_{\mathbf{k}}$  is replaced by  $\nabla_{\mathbf{k}}^\zeta$ :

$$k_0^2 \underline{\alpha} \mathbf{U}^s - \nabla_{\mathbf{k}}^\zeta \times (\underline{\beta} \nabla_{\mathbf{k}}^\zeta \times \mathbf{U}^s) = \mathbf{F}^s + \nabla_{\mathbf{k}} \times \mathbf{G}^s, \quad \text{in } \tilde{\Omega}. \quad (53)$$

The superscript  $s$  is used to emphasize that the scattered field is considered. Note that, according to Table 1,  $\mathbf{F}^s = \mathbf{G}^s = \mathbf{0}$  inside the PML.

The component of  $\zeta$  in the direction of the normal on  $\partial\Omega$  is set equal to a complex function or a complex constant (e.g.,  $15 + 15i$ ) to ensure that the scattered field is damped inside the PML. The tensors  $\underline{\alpha}$  and  $\underline{\beta}$  in the PML get the same value as in the underlying multilayer.

Inside the PML the modified curl  $\nabla_{\mathbf{k}}^\zeta \times \mathbf{U}^s$  differs from the classical curl  $\nabla_{\mathbf{k}} \times \mathbf{U}^s$ . This is not convenient because the finite elements have to be adapted so that they become conforming with respect to the modified curl instead of the ordinary curl. We shall therefore use Monk's idea<sup>20</sup> and redefine the scattered field inside the PML such that the differential equation contains only the classical curl operator throughout  $\tilde{\Omega}$ .

The modified scattered field is defined by

$$\mathbf{U}^{s'} = \underline{D}(\zeta) \mathbf{U}^s. \quad (54)$$

Note that inside  $\Omega$  we have  $\mathbf{U}^{s'} = \mathbf{U}^s$ . Now we remark that  $\underline{D}(\zeta) \mathbf{k} = \mathbf{k}$ . This follows from the observation that if  $k_i \neq 0$  the configuration is periodic in the  $x_i$  direction and hence  $\zeta_i = 1$ . By using the property that  $\zeta_j$  is a function of  $x_j$  only, one can verify that

$$\nabla_{\mathbf{k}}^\zeta \times \mathbf{U}^s = \frac{1}{\det(\underline{D}(\zeta))} \underline{D}(\zeta) (\nabla_{\mathbf{k}} \times \mathbf{U}^{s'}). \quad (55)$$

Then Eq. (53) can be transformed into

$$k_0^2 \underline{\alpha}' \mathbf{U}^{s'} - \nabla_{\mathbf{k}} \times \underline{\beta}' \nabla_{\mathbf{k}} \times \mathbf{U}^{s'} = \mathbf{F}^s + \nabla_{\mathbf{k}} \times \mathbf{G}^s, \quad \text{in } \tilde{\Omega}, \quad (56)$$

where  $\underline{\alpha}'$  and  $\underline{\beta}'$  are given by

$$\underline{\alpha}' = \det(\underline{D}(\zeta)) \underline{D}(\zeta)^{-1} \underline{\alpha} \underline{D}(\zeta)^{-1}, \quad \text{in } \tilde{\Omega}, \quad (57)$$

$$\underline{\beta}' = \det(\underline{D}(\zeta))^{-1} \underline{D}(\zeta) \underline{\beta} \underline{D}(\zeta), \quad \text{in } \tilde{\Omega}. \quad (58)$$

Note that  $\underline{\alpha}'$  and  $\underline{\beta}'$  are always anisotropic and functions of position in the PML.

We now state two continuity properties of the field  $\mathbf{U}^{s'}$  across  $\partial\Omega$ . These will be needed later. First, expression (52) implies that  $\mathbf{U}^{s'} \in \mathbf{H}(\text{curl}; \tilde{\Omega})$ , and therefore the tangential components of  $\mathbf{U}'$  are continuous across  $\partial\Omega$ :

$$\lim_{\mathbf{r}|\mathbf{r}'} \mathbf{n} \times \mathbf{U}^{s'} = \lim_{\mathbf{r}|\mathbf{r}'} \mathbf{n} \times \mathbf{U}^{s'}, \quad \text{for } \mathbf{r}' \in \partial\Omega, \quad (59)$$

where  $\mathbf{n}$  is the unit normal on  $\partial\Omega$  and  $\lim_{\mathbf{r}|\mathbf{r}'}$  and  $\lim_{\mathbf{r}'|\mathbf{r}}$  are the limits taken from the side of the PML and  $\Omega$ , respectively. Furthermore, Eq. (56) implies that  $\underline{\beta}' \nabla_{\mathbf{k}} \times \mathbf{U}^{s'} + \mathbf{G}^s \in \mathbf{H}(\text{curl}; \tilde{\Omega})$ , and therefore the tangential components of this vector field are also continuous across  $\partial\Omega$ . Because  $\mathbf{G}^s = \mathbf{0}$  in  $\tilde{\Omega} \setminus \Omega$ , we get

$$\lim_{\mathbf{r}|\mathbf{r}'} \mathbf{n} \times [\underline{\beta}' \nabla_{\mathbf{k}} \times \mathbf{U}^{s'}] + \lim_{\mathbf{r}'|\mathbf{r}} \mathbf{G}^s = \lim_{\mathbf{r}'|\mathbf{r}} \mathbf{n} \times [\underline{\beta}' \nabla_{\mathbf{k}} \times \mathbf{U}^{s'} + \mathbf{G}^s], \quad (60)$$

for  $\mathbf{r}' \in \partial\Omega$ .

We still have to impose a boundary condition for  $\mathbf{U}^{s'}$  on the outer boundary  $\partial\tilde{\Omega}$ . One can choose among the Dirichlet condition,

$$\mathbf{n} \times \mathbf{U}^{s'} = \mathbf{0}, \quad \text{on } \partial\tilde{\Omega}; \quad (61)$$

the Neumann condition,

$$\mathbf{n} \times \underline{\beta}' \nabla_{\mathbf{k}} \times \mathbf{U}^{s'} = \mathbf{0}, \quad \text{on } \partial\tilde{\Omega}; \quad (62)$$

and an impedance type of condition. Often Eq. (62) gives the most accurate results.

### C. Reformulation of the Boundary-Value Problem on the Extended Computational Domain

The boundary-value problem [Eqs. (56) and (62)] for  $\mathbf{U}^{s'}$  is still not optimal from the computational point of view when there is a multilayer background. As follows from Table 1,  $\mathbf{G}^s$  is discontinuous across the extension of the interfaces  $x_3 = x_3^{(j)}$  of the multilayer into  $\Omega$  ( $\mathbf{G}^s$  and  $\mathbf{F}^s$  vanish inside the PML; hence there are no discontinuities inside the PML). The mesh generator will always take account of interfaces between different materials inside the ECD so that a finite element will never have a nonzero intersection with more than one material. But it would be impractical to require that the mesh should also take account of the extensions into  $\Omega$  of the interfaces of the multilayer background. Therefore, the discontinuity of  $\mathbf{G}^s$



across these interfaces will, in general, cause numerical inaccuracies. To prevent these, we will use a formulation in which the unknown inside  $\Omega$  is the total field  $\mathbf{U}$  while the unknown in the PML is the modified scattered field  $\mathbf{U}^{s'}$ :

$$k_0^2 \alpha \mathbf{U} - \nabla_{\mathbf{k}} \times (\beta \nabla_{\mathbf{k}} \times \mathbf{U}) = \mathbf{F} + \nabla_{\mathbf{k}} \times \mathbf{G}, \quad \text{in } \Omega, \quad (63)$$

$$k_0^2 \alpha' \mathbf{U}^{s'} - \nabla_{\mathbf{k}} \times (\beta' \nabla_{\mathbf{k}} \times \mathbf{U}^{s'}) = \mathbf{0}, \quad \text{in } \tilde{\Omega} \setminus \Omega, \quad (64)$$

where we have used  $\mathbf{F}^s = \mathbf{G}^s = \mathbf{0}$  outside of  $\Omega$ . As follows from Table 1, interfaces of the multilayer do not cause discontinuities in  $\mathbf{F}$  and  $\mathbf{G}$ .

Additional equations that link the fields  $\mathbf{U}$  and  $\mathbf{U}^{s'}$  on  $\partial\Omega$  are, of course, now needed. These are derived from Eqs. (59) and (60) and from  $\mathbf{U}^{s'} = \mathbf{U} - \mathbf{U}^0$  in  $\Omega$ , where  $\mathbf{U}^0$  is the solution when there is only the multilayer. We obtain

$$\lim_{\mathbf{r} \uparrow \mathbf{r}'} \mathbf{n} \times \mathbf{U}(\mathbf{r}) - \lim_{\mathbf{r} \downarrow \mathbf{r}'} \mathbf{n} \times \mathbf{U}^{s'}(\mathbf{r}) = \lim_{\mathbf{r} \uparrow \mathbf{r}'} \mathbf{n} \times \mathbf{U}^0(\mathbf{r}), \quad \text{for all } \mathbf{r}' \in \partial\Omega, \quad (65)$$

$$\begin{aligned} \lim_{\mathbf{r} \uparrow \mathbf{r}'} \mathbf{n} \times \beta \nabla_{\mathbf{k}} \times \mathbf{U} - \lim_{\mathbf{r} \downarrow \mathbf{r}'} \mathbf{n} \times [\beta' \nabla_{\mathbf{k}} \times \mathbf{U}^{s'}] \\ = \lim_{\mathbf{r} \uparrow \mathbf{r}'} \mathbf{n} \times [\beta \nabla_{\mathbf{k}} \times \mathbf{U}^0 + \mathbf{G}^s]. \end{aligned} \quad (66)$$

Summarizing, the boundary-value problem that is solved numerically consists of Eqs. (64) and (62) for  $\mathbf{U}^{s'}$  in  $\tilde{\Omega} \setminus \Omega$ , Eq. (63) for  $\mathbf{U}$  in  $\Omega$ , and the continuity equations (65) and (66). The fields  $\mathbf{U}$ ,  $\mathbf{U}^s$  are sought in the spaces  $\mathbf{H}(\text{curl}; \Omega)$  and  $\mathbf{H}(\text{curl}; \tilde{\Omega} \setminus \Omega)$ , respectively. Because the torus  $T^{n_p}$  is used in the definition of  $\Omega$  and  $\tilde{\Omega}$ , periodic boundary conditions do not have to be imposed explicitly.

#### 4. DISCRETIZATION OF THE PROBLEM ON THE EXTENDED COMPUTATIONAL DOMAIN

We consider the discretization of the boundary-value problem derived in Section 3.

##### A. Variational Formulation

Since the unknown fields  $\mathbf{U}$  and  $\mathbf{U}^{s'}$  are sought in the spaces  $\mathbf{H}(\text{curl}; \Omega)$  and  $\mathbf{H}(\text{curl}; \tilde{\Omega} \setminus \Omega)$ , respectively,  $\mathbf{U}$  and  $\mathbf{U}^{s'}$  are periodic on the lattice:

$$\mathbf{a}_j \times \mathbf{U}(\mathbf{r} + \ell \mathbf{a}_j) = \mathbf{a}_j \times \mathbf{U}(\mathbf{r}), \quad (67)$$

$$\mathbf{a}_j \times \mathbf{U}^{s'}(\mathbf{r} + \ell \mathbf{a}_j) = \mathbf{a}_j \times \mathbf{U}^{s'}(\mathbf{r}), \quad (68)$$

for all  $\mathbf{r} \in \Omega$ , all integers  $\ell$ , and for  $1 \leq j \leq n_p$ .

Now let  $\mathbf{V}$  be a vector field in the space  $\mathbf{H}(\text{curl}; \tilde{\Omega})$ . Take the scalar product of  $\mathbf{V}$  with Eq. (63), integrate over  $\Omega$ , and apply a partial integration. Next, take the scalar product of  $\mathbf{V}$  with Eq. (64) and partially integrate over the PML  $\tilde{\Omega} \setminus \Omega$ . By adding the resulting expressions, we get

$$\begin{aligned} \int \int \int_{\Omega} [k_0^2 \alpha \mathbf{U} \cdot \mathbf{V} - \beta \nabla_{\mathbf{k}} \times \mathbf{U} \cdot \nabla_{-\mathbf{k}} \times \mathbf{V}] dx_1 dx_2 dx_3 \\ + \int \int \int_{\tilde{\Omega} \setminus \Omega} [k_0^2 \alpha' \mathbf{U}^{s'} \cdot \mathbf{V} - \beta' \nabla_{\mathbf{k}} \times \mathbf{U}^{s'} \cdot \nabla_{-\mathbf{k}} \times \mathbf{V}] \\ \times dx_1 dx_2 dx_3 \\ = \int \int \int_{\Omega} \mathbf{F} \cdot \mathbf{V} dx_1 dx_2 dx_3 \\ + \int \int \int_{\Omega} \mathbf{G} \cdot \nabla_{-\mathbf{k}} \times \mathbf{V} dx_1 dx_2 dx_3 \\ + \int \int \int_{\partial\Omega} \lim_{\mathbf{r} \uparrow \mathbf{r}'} \mathbf{n} \times [\beta \nabla_{\mathbf{k}} \times \mathbf{U}^0 + \mathbf{G} - \mathbf{G}^s] \cdot \mathbf{V} dS. \end{aligned} \quad (69)$$

The integral over  $\partial\Omega$  is only over the nonperiodic parts of the boundary because the contributions of the periodic parts cancel, since  $\mathbf{U}^s$  and  $\mathbf{r} \mapsto \mathcal{G}(\mathbf{r}, \mathbf{r}^0)$  are periodic and the outward-pointing normal  $\mathbf{n}$  is opposite on opposite parts of the periodic boundaries.

Hence the variational problem is to find  $\mathbf{U} \in \mathbf{H}(\text{curl}; \Omega)$  and  $\mathbf{U}^{s'} \in \mathbf{H}(\text{curl}; \tilde{\Omega} \setminus \Omega)$  such that Eq. (69) holds for all  $\mathbf{V} \in \mathbf{H}(\text{curl}; \tilde{\Omega})$  and such that condition (65) is satisfied.

*Remark.* The periodicity conditions for  $\beta \nabla_{\mathbf{k}} \times \mathbf{U}$  and  $\beta' \nabla_{\mathbf{k}} \times \mathbf{U}^{s'}$  analogous to Eqs. (67) and (68) are not imposed explicitly but are taken into account in a weak sense by variational formulation (69).

##### B. Curl-Conforming Elements

Curl-conforming elements yield approximations in  $\mathbf{H}(\text{curl}; \tilde{\Omega})$ . These approximations are such that the components of the fields that are tangential to the faces and edges of the mesh are continuous. The elements of order  $k$  are on every tetrahedron or hexahedron polynomial of order  $k$ . Corresponding to a curl-conforming element of given order  $k$ , there exists a divergence-conforming element of order  $k$  such that a discrete version of the de Rham diagram commutes. This property is very important because it implies the convergence of the FEM<sup>20</sup> and precludes the occurrence of spurious numerical solutions.

The most important curl-conforming elements are Nédélec elements of the first type on tetrahedra,<sup>17</sup> Mur-Nédélec elements of the second type on tetrahedra,<sup>18,19,30</sup> and Nédélec elements of the first type on hexahedra.<sup>21</sup> Nédélec elements of the second type on hexahedra also exist, but these do not satisfy a discrete de Rham diagram. The elements of the second type on tetrahedra give more accurate results than those of the first type of the same order. In fact, if  $h$  is a typical edge length of the mesh, the elements of the first type of order  $k$  yield approximations  $\mathbf{E}_h, \mathbf{H}_h$  of the actual electromagnetic field  $\mathbf{E}, \mathbf{H}$  with error of order  $h^k$  measured in the  $L^2$  norm:

$$\|\mathbf{E} - \mathbf{E}_h\|_{L^2} + \|\mathbf{H} - \mathbf{H}_h\|_{L^2} \leq Ch^k, \quad (70)$$

where  $C$  is a constant that depends on the configuration. In contrast, elements of the second type give an error of order  $h^{k+1}$  for the field that is solved numerically while

the error for the curl of this field is  $h^k$ . Elements of the second type involve more unknowns than those of type 1. For accuracy it is desirable to locally adapt the mesh and/or the order of the elements, e.g., near an interface with an obtuse angle where the field is infinite. Using several orders in one mesh (hp FEM, where  $h$  denotes diameter and  $p$  polynomial degree) is possible by using a hierarchical basis for the elements.<sup>23–25</sup>

We write the finite-element approximations of the total field  $\mathbf{U}$  on the CD ( $\Omega$ ) and the modified scattered field  $\mathbf{U}^{s'}$  on the PML as

$$\mathbf{U} = \sum_m U_m \phi_m, \quad \text{in } \Omega, \quad (71)$$

$$\mathbf{U}^{s'} = \sum_m U_m^{s'} \phi_m, \quad \text{in } \tilde{\Omega} \setminus \Omega, \quad (72)$$

where  $U_m$  and  $U_m^{s'}$  are scalar degrees of freedom (the unknowns). The basis functions must be in the space  $\mathbf{H}(\text{curl}; \tilde{\Omega})$ , which means in particular that they have to be periodic with respect to the variables  $x_1, \dots, x_{n_p}$ . By substituting these expressions into variational formulation (69) and by letting  $\mathbf{V}$  run through the set  $\{\phi_m\}$ , we obtain a large linear system of equations for the unknowns  $\{U_m\}$  and  $\{U_m^{s'}\}$ :

$$A_{nm} \begin{pmatrix} U_m \\ U_m^{s'} \end{pmatrix} = \begin{pmatrix} F_n \\ F_n^s \end{pmatrix}, \quad (73)$$

where  $A$  is the large system matrix and the  $F_n$  and  $F_n^s$  are the coefficients of the right-hand side.

### C. Iterative Solution Method

Although in many applications the CD is comparable to or smaller than the wavelength, the number of unknowns is often several hundred thousands. Because a fine grid is needed to model the skin effect in a metal and a fine mesh in a particular region tends to make the mesh in adjacent regions also finer, the number of unknowns is particularly high in the presence of metals. It is obvious that an iterative solver must be used. However, because the system is indefinite, preconditioning is necessary to achieve convergence of the iteration. The construction of a robust preconditioner that is adequate for all types of problems is highly nontrivial, especially on an unstructured mesh. A preconditioner based on a multigrid<sup>31–34</sup> may show satisfactory performance, but it is difficult to implement. The same holds for the Schatz method.<sup>20</sup> We therefore used Saad's preconditioner ILUTP<sup>35</sup> with BICGSTAB as the iterative solver. In ILUTP the LU factorization is constructed of a modified system matrix obtained by replacing by zero all elements of the system matrix of which the ratio of their absolute value and the maximum absolute value occurring in their row is below a specified threshold. Furthermore, all entries of the ILUTP matrix that are in the complement of a band of certain width specified by the user are set equal to zero. Furthermore, when nested dissection (ND) and approximate minimum fill-in (AMF) reordering procedures are applied, the required fill-in and hence memory and computation times that are needed have been drastically reduced. The iterative solution

method is rather robust but requires a considerable amount of storage. More details about the optimized iterative scheme will appear in Ref. 36.

## 5. PROPAGATION OF THE SCATTERED FIELD INTO THE EXTERIOR OF THE COMPUTATIONAL DOMAIN

Often the scattered field is required outside the CD. We briefly describe a rigorous method to compute the scattered field in an arbitrary point in the exterior of  $\Omega$ .

Consider again a quasi-periodic boundary-value problem for some  $\mathbf{k}$  in the Brillouin zone, and let  $\mathbf{U}_{\mathbf{k}}$  be the  $\mathbf{k}$ -quasi-periodic total field inside  $\Omega$  given by Eq. (19). The field  $\mathbf{U}$  is the corresponding periodic field, i.e.,  $\mathbf{U}_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k} \cdot \mathbf{r}} \mathbf{U}(\mathbf{r})$ . Suppose that  $\mathbf{r}_0$  is a point in the exterior of  $\Omega$  for which there is a lattice vector  $\ell = \sum_{i=1}^{n_p} \ell_i \mathbf{a}_i$  such that  $\mathbf{r}_0 + \ell$  is in  $\Omega$ . Then the total field in  $\mathbf{r}$  can immediately be obtained from the quasi-periodicity:

$$\mathbf{U}_{\mathbf{k}}(\mathbf{r}_0) = e^{i\mathbf{k} \cdot \ell} \mathbf{U}_{\mathbf{k}}(\mathbf{r}_0 + \ell). \quad (74)$$

The scattered field then obviously follows by subtracting the incident field  $\mathbf{U}_{\mathbf{k}}^i$  from the extended total field.

Suppose next that  $\mathbf{r}_0 = (x_1^0, x_2^0, x_3^0)^T$  is a point in the exterior of  $\Omega$  for which no such lattice vector  $\ell$  exists. This happens when, for some  $i = n_p + 1, \dots, 3$ , we have  $x_i^0 < x_i^s$  or  $x_i^0 > x_i^l$ . The scattered field in  $\mathbf{r}_0$  will then be obtained from a formula of the Stratton–Chu type that contains the Green's tensor of the vector Helmholtz equation.

In the Stratton–Chu formula for the scattered field of the  $\mathbf{k}$ -quasi-periodic boundary-value problem, the Green's tensor occurs, which is  $(-\mathbf{k})$ -quasi-periodic. Let  $\underline{\mathcal{G}}_{-\mathbf{k}}(\mathbf{r}, \mathbf{r}_0)$  be the Green's tensor when the dipole is in  $\mathbf{r}_0$  in the fundamental cell. Write

$$\underline{\mathcal{G}}_{-\mathbf{k}}(\mathbf{r}, \mathbf{r}_0) = \exp(-i\mathbf{k} \cdot \mathbf{r}) \underline{\mathcal{G}}(\mathbf{r}, \mathbf{r}_0), \quad (75)$$

where  $\mathbf{r} \mapsto \underline{\mathcal{G}}(\mathbf{r}, \mathbf{r}_0)$  is the corresponding periodic Green's tensor. If  $\mathbf{p}$  is a constant complex vector,  $\mathbf{r} \mapsto \underline{\mathcal{G}}(\mathbf{r}, \mathbf{r}_0) \mathbf{p}$  is the periodic field that satisfies

$$k_0^2 \tilde{\alpha} \underline{\mathcal{G}} \mathbf{p} - \nabla_{-\mathbf{k}} \times \tilde{\beta} \nabla_{-\mathbf{k}} \times \underline{\mathcal{G}} \mathbf{p} = \mathbf{p} \delta(\mathbf{r} - \mathbf{r}_0), \quad (76)$$

with Sommerfeld's radiation condition for  $|x_3| \mapsto \infty$ . Recall that  $\tilde{\alpha}$  and  $\tilde{\beta}$  describe the material properties of the multilayer background. For multilayers consisting of several layers, a closed formula for the Green's tensor cannot be written down. Instead, its Fourier transform with respect to  $x_1$  and  $x_2$  can be determined by a standard multilayer calculation.

Let  $\mathbf{U}^s$  be the periodic scattered field for the given vector  $\mathbf{k}$  in the Brillouin zone. Let  $\mathbf{r}^0$  be a point in the exterior of  $\Omega$ , and let  $\underline{\mathcal{G}}(\mathbf{r}, \mathbf{r}_0)$  be the periodic Green's tensor for the vector Helmholtz equation for wave vector  $-\mathbf{k}$ . By applying a partial integration, we obtain

$$\begin{aligned}
& \mathbf{U}^s(\mathbf{r}_0) \cdot \mathbf{p} \\
&= \int \int \int_{\mathbb{R}^3 \setminus \Omega} \mathbf{U}^s \cdot [k_0^2 \tilde{\alpha} \underline{\mathbf{G}} \mathbf{p} - \nabla_{-\mathbf{k}} \times \tilde{\beta} \nabla_{-\mathbf{k}} \times \underline{\mathbf{G}} \mathbf{p}] \\
&= - \int \int_{\partial \Omega} \lim_{\mathbf{r} | \mathbf{r}'} \mathbf{n} \times \mathbf{U}^s(\mathbf{r}') \cdot \tilde{\beta} \nabla_{-\mathbf{k}} \times \underline{\mathbf{G}}(\mathbf{r}, \mathbf{r}') \mathbf{p} dS(\mathbf{r}) \\
&\quad - \int \int_{\partial \Omega} \lim_{\mathbf{r} | \mathbf{r}'} \mathbf{n} \times [\tilde{\beta} \nabla_{\mathbf{k}} \times \mathbf{U}^s(\mathbf{r}')] \cdot \underline{\mathbf{G}}(\mathbf{r}, \mathbf{r}') \mathbf{p} dS(\mathbf{r}),
\end{aligned} \tag{77}$$

where  $\mathbf{n}$  is the unit normal that points out of  $\Omega$  and the limits are taken from the exterior of  $\Omega$ . In the derivation of this result, the fact that both the scattered field and the Green's tensor satisfy Sommerfeld's radiation condition is used to infer that the contribution of the integral over the surface of a large sphere vanishes in the limit of infinite radius.

The integral over the boundary  $\partial \Omega$  consists of only those parts that are nonperiodic because the surface integrals over the periodic parts cancel due to the periodicity.

The limits from the exterior of  $\Omega$  of the tangential components  $\mathbf{n} \times \mathbf{U}^s$  and  $\mathbf{n} \times [\tilde{\beta} \nabla_{\mathbf{k}} \times \mathbf{U}^s]$  that occur in the right-hand side of Eq. (77) can be expressed in terms of limits from within  $\Omega$  of the tangential components of the numerically computed total field  $\mathbf{U}$ , the field  $\mathbf{U}^0$ , and their curls. Since both the tangential components of  $\mathbf{U}$  and of  $\beta \nabla \times \mathbf{U}$  are needed, it is advantageous to use Nédélec's first type of elements because then, as has been mentioned before, the errors in the computed  $\mathbf{U}$  and  $\nabla \times \mathbf{U}$  are of the same order in the mesh size.

In many applications it is important to determine the scattered field at large distances. According to the Fraunhofer formula, the scattered field for  $x_3^0 \rightarrow \pm \infty$  is proportional to the Fourier transform with respect to  $x_1^0$  and  $x_2^0$  of the scattered near field. By taking the Fourier transform of Eq. (77), the Fourier transform of  $(x_1^0, x_2^0) \mapsto \mathbf{U}^s(x_1^0, x_2^0, x_3^0)$  can be expressed in terms of surface integrals over  $\partial \Omega$  of the Fourier transform of the Green's tensor. Since the latter is easy to compute in a multilayer, this expression is computationally much more efficient than first computing the scattered near field in some plane  $x_3^0 = \text{constant}$  using Eq. (77) and then computing the Fourier transform afterward. In directions in which the structure is periodic, the Fourier transform is, of course, discrete.

## 6. READING DATA ON AN OPTICAL DISC USING A SOLID IMMERSION LENS

In an optical recording system, the size of the carrier of a bit is of the order of the wavelength. The bit density can be increased by reducing the wavelength or by increasing the numerical aperture of the focusing lens. The effective numerical aperture can also be increased by using a near-field optical system in which a solid immersion lens (SIL) is kept at close distance (20–50 nm) to the disc by an actuator. Because the width of the air gap between the SIL and the disc is small, the waves of high angle of incidence inside the SIL can, although they are evanescent in air, tunnel through the air gap into the disc and thus can con-

tribute to a smaller spot. It is estimated that in this way a storage capacity of up to 300 GB on a compact-disc-size disc may be possible.

In Ref. 37 two different systems are described. Both use the blue wavelength of 405 nm. In the first system a superhemispherical SIL is used with a first-surface read-only (ROM) disc made of silicon. This system gives the highest effective numerical aperture  $NA_{\text{eff}} = n_{\text{SIL}} NA_0$  of 1.9, where  $NA_0$  is the numerical aperture of the objective lens and  $n_{\text{SIL}}$  is the refractive index of the SIL. In the second system a cover layer on the disc is used to protect the SIL and the disc from impact. Because the refractive index of the cover layer is approximately 1.5, the effective numerical aperture is reduced to 1.5. In this paper, we consider only the near-field optical system for readout of first-surface ROM discs, i.e., without the cover layer. The SIL is made of LaSF35 glass with refractive index  $n_{\text{SIL}} = 2.086$  at 405 nm and has a radius of 0.5 mm. Figure 4 shows the optical system with the focusing lens and the spherical SIL. Furthermore, a top view and a cross section are shown of a region of the disc that contains a groove with a row of equal pits in it and two neighboring grooves without pits.

To study the signal modulation and polarization dependence of the readout system, we assume that the Gaussian beam that is focused by the lens is TE polarized (electric field perpendicular to the grooves) or TM polarized (electric field parallel to the grooves). The numerical aperture of the lens is  $NA_0 = 0.91$ , so that the effective numerical aperture of the combined system of the focusing lens and SIL is  $NA_{\text{eff}} = n_{\text{SIL}} * NA_0 = 1.9$ , as stated above. Vector diffraction theory<sup>38</sup> must be applied to take the rotation of the electric field upon transmission through the lens into account. For simplicity, we use a hemispherical SIL instead of a superhemispherical SIL to achieve the  $NA = 1.9$ . The Gaussian beam is focused on the lower surface of the SIL. The field inside the SIL consists of a superposition of plane waves of which the spatial frequencies  $k_x, k_y$  fill a circular disc in reciprocal space of radius

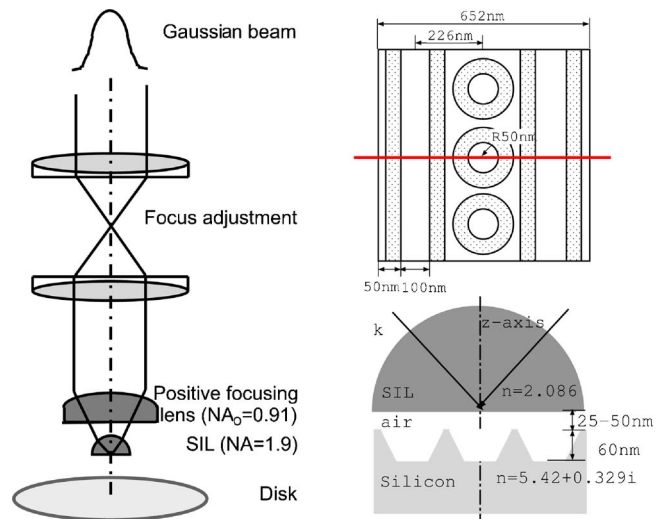


Fig. 4. (Color online) Optical system with a focusing lens and SIL (left) and part of the disc (right) seen from the top and in cross section (the SIL as shown is much too small because its radius is 0.5 mm.)

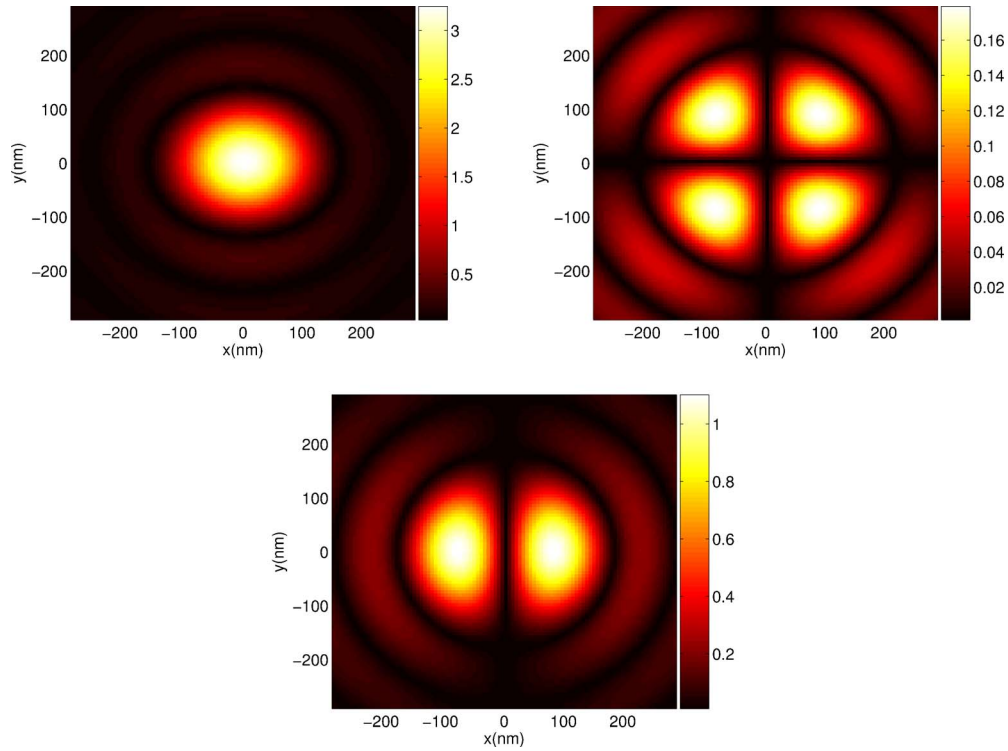


Fig. 5. (Color online) Amplitudes of the  $E_x^i$  (top left),  $E_y^i$  (top right), and  $E_z^i$  components (bottom) of the incident field in the focal plane of the objective lens. The effective numerical aperture is 1.9.

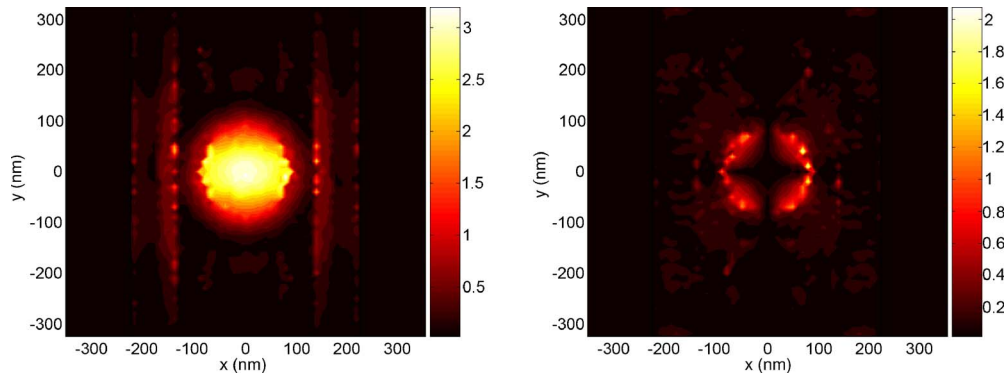


Fig. 6. (Color online) Amplitudes of the  $E_x$  (left) and the  $E_y$  components (right) of the total electric field in a plane inside the air gap and parallel to the surface of the disc at a distance of 5 nm from it. The Gaussian beam is TE polarized, and the incident spot is focused on the center of a pit.

$kn_{\text{SIL}}NA_0$ , with  $k$  as the wavenumber in vacuum. Hence the FWHM of the intensity of the spot is approximately  $\lambda/NA \approx 200$  nm. The absolute values of the incident-field components  $E_x^i$ ,  $E_y^i$ , and  $E_z^i$  in the focal plane, i.e., in the lower surface of the disc (on the side of the SIL), are shown in Fig. 5. This incident field is determined as if the entire image space of the focusing lens consists of the glass of which the SIL is made. Since the plane waves with wave vector  $\mathbf{k}=(k_x, k_y, k_z)$  such that  $k_x^2 + k_y^2 > k^2$  are evanescent in air, the size of the spot in the disc will be somewhat larger than 200 nm.

We determined the modulation of the detected signal for both polarizations by calculating the reflected field imaged by the lens on the detector. Except for the rotation of the polarization upon transmission through the lens, the

field in the detector plane is basically identical to the Fraunhofer pattern of the reflected near field.

The pits considered have minimum length of 100 nm at the bottom and 184 nm at the top (this length is often called 2T, where T is the nominal clock period). As indicated in Fig. 4 the pits are cylindrical with depth of 60 nm, sidewall angle of  $55^\circ$ , and radius at the bottom of 50 nm. The radius of the pit at the surface of the disc is approximately 100 nm. The distance between the grooves (i.e., the track pitch) is 226 nm, and the width of the grooves is 100 nm. The complex refractive index of the silicon disc is  $n_{\text{Si}}=5.42+0.329i$ . The gap width is 20 nm.

In Fig. 6 the amplitudes of the  $x$  and  $y$  components of the total electric field are shown in the plane parallel to the disc at a distance of 5 nm above the disc. The incident



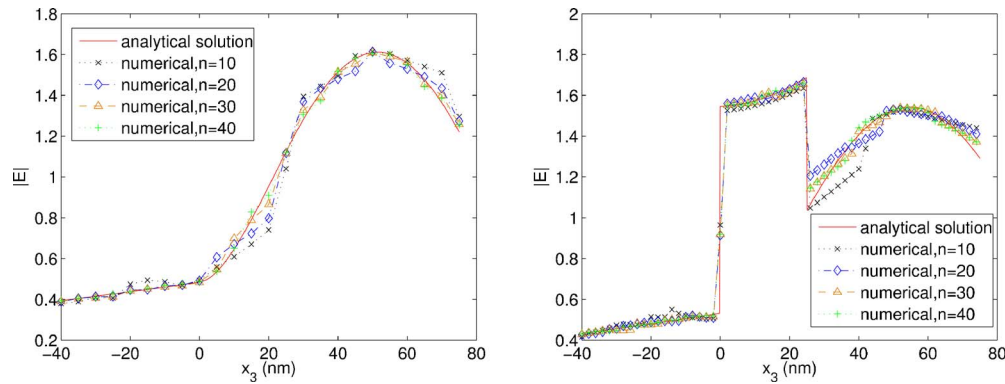


Fig. 7. (Color online) Modulus of the electric field as a function of  $x_3$  in a SIL-air-Si three-layer system, calculated analytically and numerically on different meshes with the indicated number of points per wavelength. The incident plane wave is oblique, and the angle between the wave vector and the  $x_3$  axis is 30 deg. Both  $S$  polarization (left)  $P$  polarization (right) are shown.

spot is TE polarized, and therefore the maximum amplitude of the  $E_x$  component is considerably larger than that of the  $E_y$  component. The spot is focused at the center of a pit. It is seen that the numerically obtained spot has small size, in agreement with the estimation given above.

All simulations were done with Nédélec elements of type 1 and order 1 on a tetrahedral mesh consisting of 10 points per wavelength (measured inside the material). Since the refractive index of the disc is much higher than that of the SIL and air, the mesh inside the disc is required to be much finer than in the air and the SIL. By continuity of the mesh, the grid spacing inside the air and the SIL is actually finer than 10 points per wavelength. The number of (complex) unknowns is  $10^6$ .

To investigate the accuracy of the presented results, we compared the numerical and analytical solutions for a multilayer with the same composition as the optical disc. The only difference is that the interfaces are flat; i.e., the pits and grooves are absent because otherwise an analytic solution does not exist. Instead of a focused spot we computed the field due to a single oblique incident linearly polarized plane wave. The angle between the wave vector and the  $x_3$  axis was 30°. In Fig. 7, the modulus of the electric field [i.e.,  $(|E_1|^2 + |E_2|^2 + |E_3|^2)^{1/2}$ ] is shown as a function of  $x_3$  of the analytical solution and the numerical solutions with the number of mesh points per wavelength varying from 10 to 40. The results were obtained on a relatively small CD of 80 nm × 80 nm × 120 nm because otherwise the computations for 40 points per wavelength would require too much memory. On the left the field is  $S$  polarized (hence the electric field is continuous); on the right it is  $P$  polarized (the normal component of the electric field jumps at the interfaces). The mean error in the modulus of the electric field is shown in Fig. 8. It is seen that the convergence is linear with the number of mesh points, as expected for lowest-order elements. The residual error for 40 points per wavelength was mainly caused by the fact that the quality of the mesh deteriorates when a large number of mesh points are required. The PML seemed to be accurate enough. For the configuration with the grooves and pits present having vertical sidewalls only, we compared our numerical results with a code based on the FDTD. Based on these tests, we estimate the mean error in the numerical solutions for the optical disc to be less than 5% when the number of points

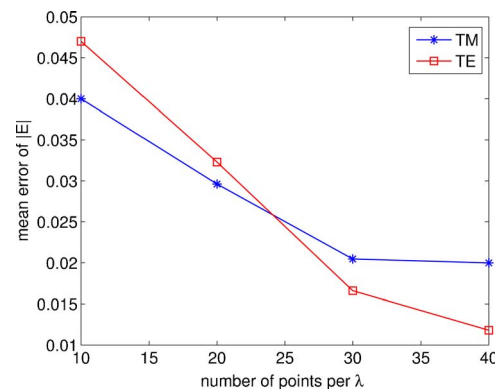


Fig. 8. (Color online) Mean error in the modulus of the electric field shown in Fig. 7 is plotted as a function of the number of points per wavelength. Both polarizations are considered.

per wavelength is 10 (conservative estimate). Using more points per wavelength was not possible on the machine available. Solving the system required approximately 20 h on a Linux HP-DL585 machine with 48 GB memory and a 2.4 GHz processor. Approximately 90% of the CPU is spent in computing the ILUTP.

In Fig. 9 cross sections of  $|E_x|$  in the  $(x, z)$  and  $(y, z)$  planes through the center of the pit corresponding to the case of Fig. 6 are shown. It is seen that the pattern of  $|E_x|$  is more narrow in the  $(y, z)$  plane than in the  $(x, z)$  plane. Furthermore, the amplitude of a field component is maximum close to those parts of the walls of the pits where the field component is perpendicular to these walls. Figure 10 presents the amplitudes of the  $E_x$  (left) and the  $E_y$  component (right) of the total electric field in a plane inside the air gap and parallel to the surface of the disc at a distance of 5 nm above the surface. The incident spot is TM polarized, and therefore the maximum amplitude of the  $E_y$  component is considerably larger than that of the  $E_x$  component.

When the spot is focused in the middle between two pits, the differences between near fields for the two orthogonally polarized spots become larger. It is seen in Figs. 11 and 12 that the total field is smaller than when the spot is focused at the center of a pit. Furthermore, for TM polarization there are two maxima at the fringe of the neighboring pits with a minimum in between. However,

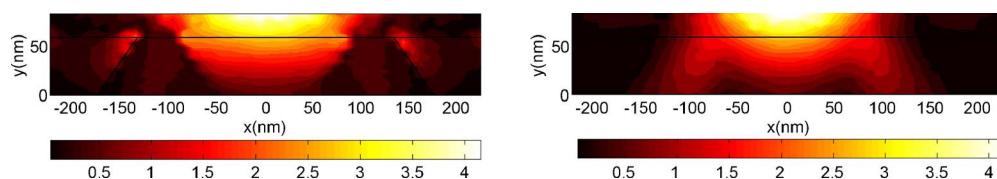


Fig. 9. (Color online) Cross section of the amplitude of the  $E_x$  component of the total field for the case of Fig. 6 in the  $(x, z)$  plane (left) and the  $(y, z)$  plane (right). Both planes are through the center of the spot.

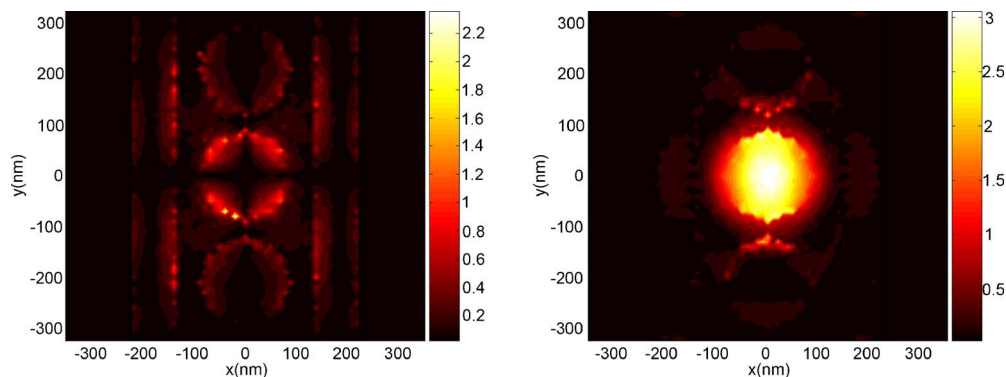


Fig. 10. (Color online) Amplitudes of the  $E_x$  (left) and the  $E_y$  components (right) of the total electric field in a plane inside the air gap and parallel to the surface of the disc at a distance of 5 nm from it. The Gaussian beam is TM polarized, and the incident spot is focused on the center of a pit.

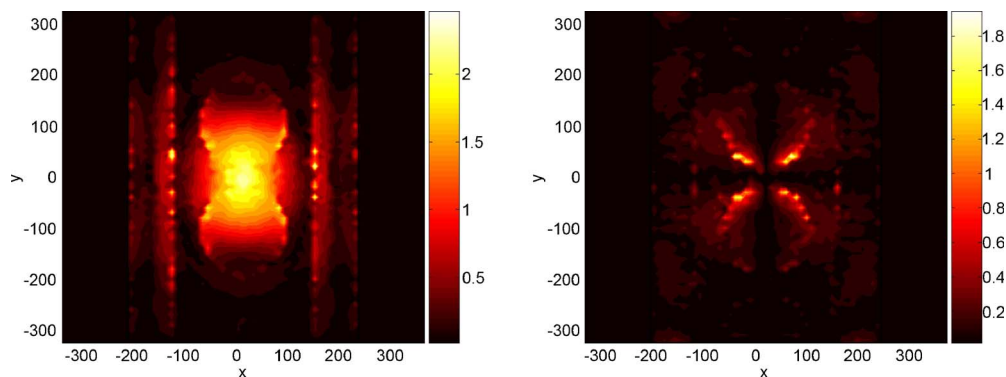


Fig. 11. (Color online) Amplitudes of the  $E_x$  (left) and the  $E_y$  components (right) of the total electric field in a plane inside the air gap and parallel to the surface of the disc at a distance of 5 nm. The Gaussian beam is TE polarized, and the incident spot is focused in the middle between two pits.

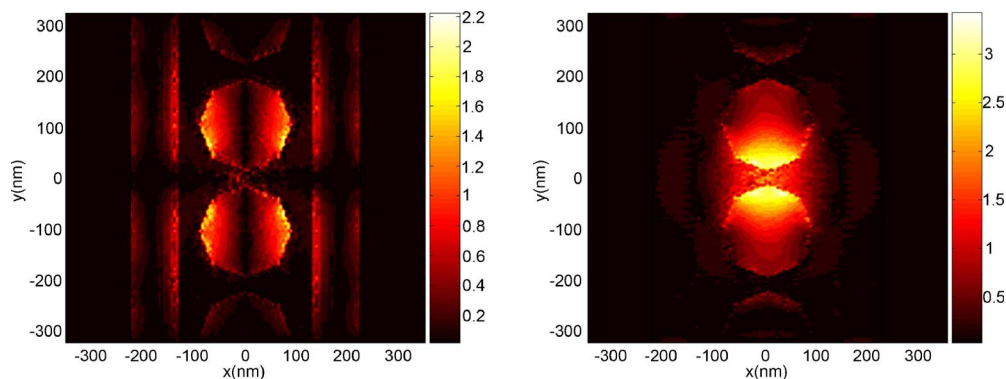


Fig. 12. (Color online) Amplitudes of the  $E_x$  (left) and the  $E_y$  components (right) of the total electric field in a plane inside the air gap and parallel to the surface of the disc at a distance of 5 nm. The Gaussian beam is TM polarized, and the incident spot is focused in the middle between two pits.

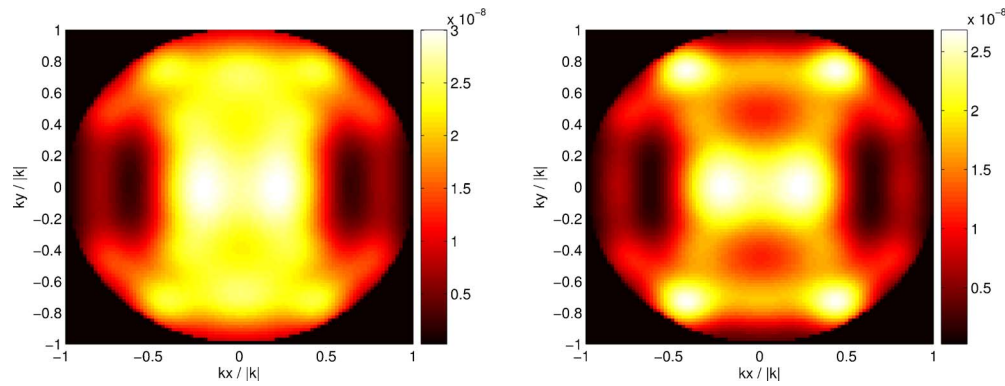


Fig. 13. (Color online) Intensity patterns of the reflected field (i.e.,  $E_x^2 + E_y^2$  of the reflected far field) when the incident Gaussian beam is TE polarized and the spot is focused in the center of a pit (left) and in the middle between two pits (right). The normalized detected intensities are 0.4795 (left) and 0.3724 (right), respectively.

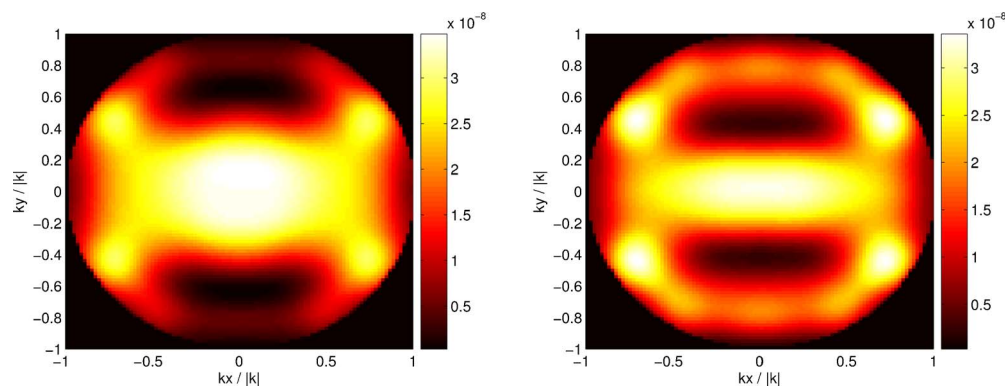


Fig. 14. (Color online) Intensity patterns of the reflected field when the incident Gaussian beam is TM polarized and the spot is focused in the center of a pit (left) or in the middle between two pits (right). The normalized detected intensities are 0.4862 (left) and 0.4797 (right).

in the TE case, the maximum of the  $E_x$  components occurs at the center of the spot between the two pits.

In Figs. 13 and 14, the intensity patterns of the reflected field at the detector are shown when the incident Gaussian beam is TE and TM polarized, respectively, and the spot is focused either at the center of a pit or halfway between two pits. It is seen that the patterns for TE and TM polarizations differ substantially. Furthermore, the total detected reflected intensity is considerably higher when the spot is focused in the center of a pit than when it is focused between the pits. When the pits are longer, for pit lengths of  $3T$  and  $4T$ , for example, the computations predict similar effects. These computational results are confirmed by experiments. They are in contrast to what is measured for conventional compact disc and DVD systems for which the reflected intensity is always observed to be largest when the spot is focused between the pits. This difference between conventional and near-field recording may be due to the contribution to the reflected intensity of scattered waves that are evanescent in air but propagate in the SIL. These scattered waves contribute to the detected signal in the case of near-field recording but cannot contribute in conventional recording. The amplitudes of these scattered evanescent waves may be expected to be larger when the spot is focused on the pit than when the spot is focused on the land between the pits.

## 7. CONCLUSION

A finite-element model for the rigorous modeling of electromagnetic scattering in micro-optics is described. The model applies to nonperiodic structures and to structures that are periodic in one, two, or three directions. The materials may be inhomogeneous and/or anisotropic dielectrics, metals, or magnetic materials. A computational domain was defined that enclosed all nontrivial scatterers. This domain is extended in the directions of nonperiodicity by a PML to prevent reflections. On the extended computational domain a variational formulation was derived, and Nédélec's edge elements were used to discretize the system. The system was solved iteratively by BICGSTAB with ILUTP as the preconditioner. Due to a ND and AMF renumbering of the unknowns, the ILUTP was made quite efficient. An application of the model to our proposal for high-density optical recording using a solid immersion lens has been discussed. It was found that, in contrast to conventional recording, the detected reflected intensity is largest when the spot is focused instead of between the pits.

## ACKNOWLEDGMENTS

The authors acknowledge the important contribution of Alexei Kononov of Delft University of Technology to the



optimization of the preconditioner. They furthermore thank their colleague Ferry Zijp of the near-field recording team for fruitful discussions. X. Wei was sponsored by the Dutch Ministry of Economic Affairs under project BTS01044.

Corresponding author X. Wei can be reached by e-mail at wei@natlab.research.philips.com.

## REFERENCES

1. K. Knop, "Rigorous diffraction theory for transmission phase gratings with deep rectangular grooves," *J. Opt. Soc. Am.* **68**, 1206–1210 (1978).
2. M. G. Moharam and T. K. Gaylord, "Diffraction analysis of dielectric surface-relief gratings," *J. Opt. Soc. Am.* **72**, 1385–1392 (1982).
3. M. G. Moharam and T. K. Gaylord, "Three-dimensional vector coupled-wave analysis of planar-grating diffraction," *J. Opt. Soc. Am.* **73**, 1105–1112 (1983).
4. L. Li, "Use of Fourier series in the analysis of discontinuous periodic structures," *J. Opt. Soc. Am. A* **13**, 1870–1876 (1996).
5. V. Rokhlin, "Rapid solution of integral equations of scattering theory in two dimensions," *J. Comput. Phys.* **36**, 414–439 (1990).
6. N. Enghetta, W. Murphy, V. Rokhlin, and M. Vassiliou, "The fast multipole method (FMM) for electromagnetic scattering problems," *IEEE Trans. Antennas Propag.* **40**, 634–642 (1992).
7. J. Song, C. Lu, and W. Chew, "Multilevel fast multipole algorithm for electromagnetic scattering by large complex objects," *IEEE Trans. Antennas Propag.* **45**, 1488–1493 (1997).
8. W. Chew, C. Lu, and Y. Wang, "Efficient computation of three-dimensional scattering of vector electromagnetic waves," *J. Opt. Soc. Am. A* **11**, 1528–1537 (1994).
9. K. Yee, "Numerical solutions of initial boundary value problems involving Maxwell's equations in isotropic media," *IEEE Trans. Antennas Propag.* **AP-14**, 302–307 (1966).
10. K. Shlager and J. Schneider, "A selective survey of the finite-difference time-domain literature," *IEEE Antennas Propag. Mag.* **37**, 39–56 (1995).
11. A. Taflov, *Advances in Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 1st ed. (Artech House, 1998).
12. J. P. Berenger, "A perfectly matched layer for the absorption of electromagnetic waves," *J. Comput. Phys.* **114**, 185–200 (1994).
13. J. Judkins and R. Ziolkowski, "Finite-difference time-domain modeling of nonperfectly conducting metallic thin-film gratings," *J. Opt. Soc. Am. A* **12**, 1974–1983 (1995).
14. W.-C. Liu and D. P. Tsai, "Optical tunneling effect of surface plasmon polaritons and localized surface plasmon resonance," *Phys. Rev. B* **65**, 155423 (2002).
15. G. Bao, "Finite element approximation of time harmonic waves in periodic structures," *SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal.* **32**, 1155–1169 (1995).
16. G. Bao and H. Yang, "A least-squares finite element analysis for diffraction problems," *SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal.* **37**, 665–682 (2000).
17. J. Nédélec, "Mixed finite elements in  $\mathbb{R}^3$ ," *Numer. Math.* **35**, 315–341 (1980).
18. J. Nédélec, "A new family of mixed finite elements in  $\mathbb{R}^3$ ," *Numer. Math.* **50**, 57–81 (1986).
19. G. Mur and A. de Hoop, "A finite-element method for computing three-dimensional electromagnetic fields in inhomogeneous media," *IEEE Trans. Magn.* **MAG-21**, 2188–2191 (1985).
20. P. Monk, *Finite Element Methods for Maxwell's Equations*, 1st ed. (Oxford U. Press, 2003).
21. G. C. Cohen, *High-Order Numerical Methods for Transient Wave Equations*, 1st ed. (Springer, 2001).
22. P. Monk and L. Demkowicz, "Discrete compactness and the approximation of Maxwell's equations in  $\mathbb{R}^3$ ," *Math. Comput.* **70**, 507–523 (2001).
23. W. Rachowicz and L. Demkowicz, "An hp-adaptive finite element method for electromagnetics. Part II: A 3D implementation," *Int. J. Numer. Methods Eng.* **53**, 147–180 (2002).
24. C. Geuzaine, B. Meys, P. Dular, and W. Legros, "Convergence of high order curl-conforming finite elements," *IEEE Trans. Magn.* **35**, 1442–1445 (1999).
25. C. Geuzaine, "High order hybrid finite element schemes for Maxwell's equations taking thin structures and global quantities into account," Ph.D. thesis (Université de Liège, Liège, Belgium, 2001).
26. T. Eibert and V. Hansen, "3-D FEM/BEM-hybrid approach based on a general formulation of Huygen's principle for planar layered media," *IEEE Trans. Microwave Theory Tech.* **45**, 1105–1112 (1997).
27. W. C. Chew and W. H. Weedon, "A 3D perfectly matched medium from modified Maxwell's equations with stretched coordinates," *Microwave Opt. Technol. Lett.* **7**, 599–604 (1994).
28. M. Benzi, "Preconditioning techniques for large linear systems: a survey," *J. Comput. Phys.* **182**, 418–477 (2002).
29. Y. A. Erlangga, C. Vuik, and C. W. Oosterlee, "On a class of preconditioners for the study of the Helmholtz equation," *Appl. Numer. Math.* **50**, 405–425 (2004).
30. G. Mur and A. de Hoop, "The finite-element modeling of three-dimensional time-domain electromagnetic fields in strongly inhomogeneous media," *IEEE Trans. Magn.* **28**, 1130–1133 (1992).
31. J. Gozani, A. Nachshon, and E. Trukel, "Conjugate gradient coupled with multigrid for an indefinite problem," in *Advances in Computer Methods for Partial Differential Equations V* (Springfield, 1984), pp. 425–427.
32. H. Elman, O. Ernst, and D. O'Leary, "A multigrid method enhanced by Krylov subspace iteration for discrete Helmholtz equations," *SIAM J. Sci. Comput. (USA)* **23**, 1291–1315 (2002).
33. S. Kim, "Multigrid simulation for high-frequency solutions of the Helmholtz problem in heterogeneous media," *SIAM J. Sci. Comput. (USA)* **24**, 684–701 (2003).
34. Y. A. Erlangga, C. W. Oosterlee, and C. Vuik, "A novel multigrid-based preconditioner for the heterogeneous Helmholtz equation," *SIAM J. Sci. Comput. (USA)* **27**, 1471–1492 (2006).
35. Y. Saad, *Iterative Methods for Sparse Linear Systems*, 2nd ed. (Society for Industrial and Applied Mathematics, 2003).
36. A. Kononov, X. Wei, and H. Urbach, "An efficient preconditioner for the finite element method applied to the time-harmonic Maxwell equations," *J. Comput. Phys.* (to be published).
37. F. Zijp, M. van der Mark, J. Lee, and C. Verschuren, "Near-field read-out of a 50-GB first-surface disk with NA = 1.9 and a proposal for a cover-layer-incident, dual-layer near-field system," in *Proc. SPIE* **5380**, 209–223 (2004).
38. B. Richards and E. Wolf, "Electromagnetic diffraction in optical systems. II. Structure of the image field in an aplanatic system," *Proc. R. Soc. London, Ser. A* **253**, 358–379 (1959).