Normal vector method for convergence improvement using the RCWA for crossed gratings

Thomas Schuster,[1,*] Johannes Ruoff,[2] Norbert Kerwien,[1] Stephan Rafler,[1] and Wolfgang Osten[1]

[1]*Institut für Technische Optik, Universität Stuttgart, Pfaffenwaldring 9, 70569 Stuttgart, Germany*
[2]*Carl Zeiss SMT AG, 73447 Oberkochen, Germany*
[*]*Corresponding author: schuster@ito.uni-stuttgart.de*

Received April 3, 2007; revised May 21, 2007; accepted May 23, 2007;
posted May 29, 2007 (Doc. ID 81796); published August 21, 2007

The rigorous coupled wave analysis (RCWA) is a widely used method for simulating diffraction from periodic structures. Since its recognized formulation by Moharam *et al.* [J. Opt. Soc. Am. A **12**, 1068 and 1077 (1995)], there still has been a discussion about convergence problems. Those problems are more or less solved for the diffraction from line gratings, but there remain different concurrent proposals about the convergence improvement for crossed gratings. We propose to combine Popov and Nevière's formulation of the differential method [*Light Propagation in Periodic Media* (Dekker, 2003) and J. Opt. Soc. Am. A **18**, 2886 (2001)] with the classical RCWA. With a suitable choice of a normal vector field we obtain a better convergence than for the formulations that are known from the literature.

OCIS codes: 050.0050, 050.1960, 260.0260, 260.1960.

1. INTRODUCTION

In this paper a method for convergence improvement for the two-dimensional (2D) rigorous coupled wave analysis (RCWA) is reported. To that end we start with a brief overview of the history of convergence problems. In Section 2 the topic of convergence problems for one-dimensional (1D) gratings is reviewed. Afterward, the different known formulations for 2D gratings are introduced. Section 4 presents the theoretical formulation of the proposal of this work. It is a combination of Popov and Nevière's [1,2] formulation of the differential method with the RCWA [3,4]. The key point of Popov and Nevière's formulation is the introduction of a normal vector (NV) field. We will see that finding an appropriate NV field is not a trivial task. In Section 5 some instructive examples are treated that reveal the problems of the various methods. Finally, different approaches for setting up a suitable NV field are developed. Such algorithms are required to apply the method to structures with practical relevance.

2. CONVERGENCE PROBLEMS USING THE RCWA

In 1995, Moharam *et al.* published a formulation of the RCWA [3,4], which since that time has been an appreciated and often implemented formulation of this method, e.g., by Totzeck [5], whose implementation is the basis for this work. Numerical problems due to antievanescent waves and matrix inversions were definitely avoided in this formulation. Nevertheless, it was found that the TM polarization shows a worse convergence performance with the number of retained Fourier modes than the TE polarization. Lalanne and Morris [6] found that a replacement of one single matrix $\mathbf{E}$ (the Toeplitz matrix of the dielectric function) in Moharam's formulation leads to a considerable improvement of the convergence for the TM polarization. Lalanne called the new matrix $\mathbf{A}^{-1}$; it is the inverse of the Toeplitz matrix of the reciprocal dielectric function $1/\varepsilon$. In the limit of infinite Fourier series both matrices are identical, which led Lalanne to the idea of choosing any of them to obtain a better convergence. Li [7] gave the theoretical explanation for the different behavior of the two matrices for truncated Fourier series and formulated three factorization rules that have to be obeyed to avoid such problems. With the work of Lalanne and Li, the problem of poor convergence due to a wrong treatment of products in truncated Fourier space appearing in the RCWA applied on line gratings has been solved. We should note, however, that one still can face serious convergence problems for metallic gratings with TM polarization, as has been described in detail by Popov *et al.* [8].

We briefly restate Li's factorization rules here for the considered product $\mathbf{D}=\epsilon_0\varepsilon\mathbf{E}$, where $\mathbf{D}$ is the electric displacement, $\epsilon_0\varepsilon$ the permittivity, and $\mathbf{E}$ the electric field. Of course, the rules hold for any product of two functions. We first introduce Laurent's rule and its notation using Toeplitz matrices:

Let $a(x)$ and $b(x)$ be periodic continuous functions in ordinary space; then the product of their truncated Fourier series can be written as

$$c(x)=a(x)b(x)\leftrightarrow c_j=\sum_{k=-N}^{N}a_{j-k}b_k. \tag{1}$$

Laurent's rule is the equivalent for Fourier series expansion to the convolution theorem for the Fourier transform. The Fourier vector $[c]$ containing all coefficients c_j may be written as a matrix product $[c]=[\![a]\!][b]$, where $[\![a]\!]$ denotes the Toeplitz matrix of the Fourier vector $[a]$ whose entries are defined as $a_{jk}=a_{j-k}$. Now we can finally write down Li's rules [7]:

1. Let $\mathbf{D}(x)=\epsilon_0\varepsilon(x)\mathbf{E}(x)$ and either $\varepsilon(x)$ or $\mathbf{E}(x)$ be continuous at some $x=x_0$. The other quantity may be discontinuous there. Then Laurent's rule still holds and

$$[\mathbf{D}] = \epsilon_0[\![\varepsilon]\!][\mathbf{E}]. \tag{2}$$

2. Let $\mathbf{D}(x)=\epsilon_0\varepsilon(x)\mathbf{E}(x)$ and both $\varepsilon(x)$ and $\mathbf{E}(x)$ be discontinuous at some $x=x_0$ but the product $\varepsilon(x)\mathbf{E}(x)$ be continuous there. Then the inverse rule holds, which is given by

$$[\mathbf{D}] = \epsilon_0[\![1/\varepsilon]\!]^{-1}[\mathbf{E}] \tag{3}$$

3. Let $\mathbf{D}(x)=\epsilon_0\varepsilon(x)\mathbf{E}(x)$ and both $\varepsilon(x)$ and $\mathbf{E}(x)$ be discontinuous at some $x=x_0$ and the product $\varepsilon(x)\mathbf{E}(x)$ be discontinuous there as well. Then the product of the two functions in Fourier space cannot be formed by either Laurent's rule or the inverse rule.

Once those rules are known, it is easy to apply them to line gratings, as the electric field can easily be decomposed into components perpendicular and parallel to material boundaries. The parallel component is continuous, as is known from Maxwell's theory, so Laurent's rule can be applied. The perpendicular component of the electric field is discontinuous, but the product $\varepsilon E_\perp$, which is equal to the perpendicular component of the electric displacement, is continuous. Therefore, the inverse rule has to be applied. Applying Li's rules, the convergence rate for the TM case becomes comparable to the TE case in many cases.

3. DIFFERENT PROPOSALS FOR CONVERGENCE IMPROVEMENT CONSIDERING CROSSED GRATINGS

The problem becomes more complicated when crossed gratings are considered. For line gratings the orientation of the lateral boundary is uniquely determined and constant within the elementary cell. Usually it is chosen to be parallel to the y axis. With this choice the orientation of the electrical field components E_x and E_y relative to the boundary takes one of the extreme cases, parallel or perpendicular. E_x is perpendicular to the boundary, E_y is parallel. For crossed gratings the situation is different. Any orientation of the boundary is allowed and the field components E_x and E_y are in most points of the boundary neither perpendicular nor parallel to the latter. For that reason it is rather complicated to apply Li's factorization rules correctly.

Figure 1 depicts two illustrative kinds of structures. For a rectangular cavity or column in the elementary cell

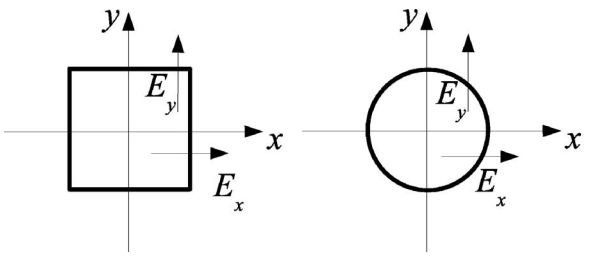


Fig. 1. Relative orientation of Cartesian field components and boundaries for two example structures.

there are only such points along the boundary where the field components are either perpendicular or parallel. But even in this case, where only both extreme cases of the relative orientations appear, it is not trivial to apply Li's factorization rules in Fourier space, where the information about relative orientations in ordinary space is not manifest. However, the RCWA is formulated in Fourier space. For a cylindrical cavity or column, all intermediate states of relative orientations between parallel and perpendicular appear. Looking at this example one has to recognize that no descriptive understanding of how to apply Li's rules in Fourier space is possible. Both example structures show that an obeyance to Li's rule cannot be achieved by a simple adaptation of the equations in Fourier space.

Lalanne [9] proposes an obvious way to make the best out of this dilemma. He just uses a weighted average of the matrices $\mathbf{E}$ and $\mathbf{A}^{-1}$ ($[\![\varepsilon]\!]$ and $[\![1/\varepsilon]\!]^{-1}$ using Li's notation), which is a change that is extremely easy to implement. This method tries to minimize the length of boundary, where Li's rules are violated.

Li [10] presents a reformulation of the theory in much greater generality, assuming nonorthogonal coordinate systems. In his derivation he uses a not necessarily rectangular grid and approximates the boundary by zigzag lines from this grid. By doing so, he constructs two new matrices $[\![\varepsilon]\!]$ and $[\![\varepsilon]\!]$, which are block Toeplitz matrices containing a mixture of Laurent's rule in one direction and the inverse rule in the other direction. Speaking in a very descriptive manner, this procedure is a step-by-step transition to Fourier space that allows it to obey Li's rules. Li's proposal for convergence improvement works very well for quadrangles, but for structures with curved boundaries or boundaries that are oblique with respect to the chosen grid, it suffers from the approximation that the boundary has to be described as a zigzag line. The *shape* of the boundary can be approximated that way with arbitrary accuracy, but its *local orientation* is in most points erroneous. Moreover, the *length* of the zigzag boundary never converges to the length of the smooth boundary, regardless of the level of refinement.

The described problem is similar to the staircase approximation of which the RCWA makes inherent use. Popov *et al.* [11] investigated the problem in detail considering 1D gratings with oblique facets, such as trapezoidal or blaze gratings. They showed that especially for metallic gratings in TM polarization, the convergence is very bad, which they ascribed to the appearance of field enhancements at the edges of the staircase boundary, which do not disappear with successive increase of the number of slices. Considering 1D problems with oblique facets, one cannot completely avoid edges. As Popov *et al.* [11] showed, it is possible to use a correct normal vector field for a decomposition of the electric field, but the material boundaries remain staircaselike unless one uses the differential method. In contrast to that, for 2D gratings with perpendicular sidewalls the shape as well as the orientation of the boundary curve can be modeled with arbitrary accuracy using the RCWA, which is described in the following.

Popov and Nevière [1,2] presented a reformulation of the differential method where Li's rules are obeyed, intro-

ducing an NV field. They call this technique fast Fourier factorization (FFF). In this formulation they do the most obvious thing to comply with Li's factorization rules. They perform a decomposition of the fields in ordinary space into components perpendicular and parallel to the boundary in any point. To this end they introduce an NV at each boundary point and decompose the electric fields into a parallel and perpendicular component. Afterward, they take the full information about the orientation of the boundary over to Fourier space by considering the Fourier series of the NV field. This step, however, requires an extension of the NV onto the whole unit cell, even to those locations where the permittivity is continuous. By this method they manage to fulfill Li's factorization rules in Fourier space.

In Section 4 we present a proposal for the improvement of the convergence of the RCWA applied to crossed gratings that adopts Popov and Nevière's method to fulfill Li's factorization rules. Considering smooth structures such as sinusoidal gratings, it is certainly the best idea to choose the differential method instead of the RCWA. But if one wants to model structures such as cylindrical or elliptical cylindrical holes, the RCWA with the convergence improvement presented here seems to be best suited.

4. THEORETICAL FORMULATION OF THE PROPOSED METHOD

The following considerations are presented following the terminology of Moharam *et al.* [3,4]. Sometimes, for better understanding, we refer to Popov and Nevière or Li, using their notation at first, then transform it to a consistent formulation. We adopt the coordinate system of Moharam *et al.*, too, where the sample surface is lying in the xy plane, the z axis pointing toward the sample substrate.

We propose to combine Popov and Nevière's [1,2] method to fulfill Li's factorization rules with the classical formulation of the RCWA [3,4]. The latter makes use of the time harmonic Maxwell curl equations for nonmagnetic materials. The mentioned convergence problems arise only from the equation that contains the electric displacement $\mathbf{D}$, i.e.,

$$\nabla \times \mathbf{H_g} = i\omega \mathbf{D_g}, \tag{4}$$

where $\mathbf{H_g}$ and $\mathbf{D_g}$ are the magnetic field and electric displacement in the grating region and i is the imaginary unit. This equation can be found in [3] within Eq. (56). There the $\mathbf{D}$ field is replaced using the material equation $\mathbf{D}=\epsilon_0 \varepsilon(x)\mathbf{E}$, which is allowed in this simple manner only in ordinary space, but not in Fourier space. The fields in Eq. (4) are then replaced in [3] by their Fourier series according to

$$\mathbf{E_g} = \sum_j [S_{xj}(z)\mathbf{x} + S_{yj}(z)\mathbf{y} + S_{zj}(z)\mathbf{z}] \times \exp[-i(k_{xj}x + k_{yj}y)],$$

$$\mathbf{H_g} = -i\left(\frac{\epsilon_0}{\mu_0}\right)^{1/2} \sum_j [U_{xj}(z)\mathbf{x} + U_{yj}(z)\mathbf{y} + U_{zj}(z)\mathbf{z}] \times \exp[-i(k_{xj}x + k_{yj}y)]. \tag{5}$$

The discrete k-vector elements k_{xj} and k_{yj} are determined from Floquet's theorem. Performing this replacement Moharam *et al.* [3] set up a matrix equation in Fourier space (we exchange the order of the x and y components):

$$\begin{bmatrix} \frac{\partial \mathbf{U}_x}{\partial z'} \\ \frac{\partial \mathbf{U}_y}{\partial z'} \end{bmatrix} = \begin{bmatrix} -\mathbf{K}_x\mathbf{K}_y & \mathbf{K}_x^2 - \mathbf{E} \\ \mathbf{E} - \mathbf{K}_y^2 & \mathbf{K}_x\mathbf{K}_y \end{bmatrix} \begin{bmatrix} \mathbf{S}_x \\ \mathbf{S}_y \end{bmatrix}, \tag{6}$$

where $z'=k_0 z$ is a normalized z coordinate, $\mathbf{E}$ is the Toeplitz matrix of the relative permittivity (i.e., $[\![\varepsilon]\!]$ using Li's notation), and $\mathbf{K}_x$ and $\mathbf{K}_y$ are diagonal matrices with the entries $\mathbf{K}_{x,jj}=k_{xj}/k_0$ and $\mathbf{K}_{y,jj}=k_{yj}/k_0$. This equation corresponds to the third and fourth lines of Eq. (57) in [3]. The first two lines can be derived in a similar manner using the other Maxwell curl equation. We omit these lines here as they do not contain the electric displacement and thus do not require the application of Li's rules. Note, however, that the mentioned Eq. (57) in [3] is a formulation for 1D gratings in conical mount. The extension to 2D gratings can be easily achieved by replacing the k_y component by the discrete k_{yj} and by replacing simple Toeplitz matrices, by block Toeplitz matrices with Toeplitz blocks (BTTB). The vectors of Fourier coefficients become matrices, which have to be reshaped to vectors using the same sorting scheme as for setting up the BTTB matrices.

Looking at the right-hand side of Eq. (6) one recognizes that Laurent's rule has been applied for all the crucial products, which are denoted by $\mathbf{ES}_{x,y}$ here. Assuming arbitrary orientation of a boundary, this clearly violates Li's rules.

Popov and Nevière set up a kind of generalized material equation in Fourier space that they need for the reformulation of the differential method. In [1] this equation is denoted by (IX.11), which we now reproduce assuming the z component of the NV $N_z=0$:

$$[D_x] = \{[\![\varepsilon]\!] - \mathbf{\Delta}[\![N_x^2]\!]\}[E_x] - \{\mathbf{\Delta}[\![N_xN_y]\!]\}[E_y],$$

$$[D_y] = \{[\![\varepsilon]\!] - \mathbf{\Delta}[\![N_y^2]\!]\}[E_y] - \{\mathbf{\Delta}[\![N_xN_y]\!]\}[E_x], \tag{7}$$

where $\mathbf{\Delta}=[\![\varepsilon]\!]-[\![1/\varepsilon]\!]^{-1}$ and N_x, N_y are the x and y component, respectively, of the NV of the material boundary. The derivation of Eq. (7) presented in [1] is straightforward. A decomposition of the fields is performed in ordinary space. The parallel and perpendicular components are transformed to Fourier space, applying the first and the second of Li's rules, i.e., Eqs. (2) and (3), respectively.

Now we just have to replace all the products of the form $\mathbf{ES}_{x,y}$ in Eq. (6) using Eq. (7). To obtain a consistent notation we discard Popov and Nevière's bracket notation, denote the Fourier vector of the electric field $[E]$ by $\mathbf{S}$, and denote the Toeplitz matrix of the dielectric constant $[\![\varepsilon]\!]$ by $\mathbf{E}$. Furthermore, we accept the above defined matrix $\mathbf{\Delta}$ and introduce the abbreviation $[\![N_xN_x]\!]=\mathbf{N}_{xx}$ (similar for other combination of indices x and y). Finally, we obtain

$$\begin{bmatrix} \frac{\partial \mathbf{U}_x}{\partial z'} \\ \frac{\partial \mathbf{U}_y}{\partial z'} \end{bmatrix} = \underbrace{\begin{bmatrix} -\mathbf{K}_x\mathbf{K}_y + \mathbf{\Delta}\mathbf{N}_{xy} & \mathbf{K}_x^2 - \mathbf{E} + \mathbf{\Delta}\mathbf{N}_{yy} \\ \mathbf{E} - \mathbf{K}_y^2 - \mathbf{\Delta}\mathbf{N}_{xx} & \mathbf{K}_x\mathbf{K}_y - \mathbf{\Delta}\mathbf{N}_{xy} \end{bmatrix}}_{\mathbf{G}} \begin{bmatrix} \mathbf{S}_x \\ \mathbf{S}_y \end{bmatrix} \tag{8}$$

This equation has to be used instead of Eq. (57) (lines 3 and 4) in [3]. It is equivalent to the last two equations of (IX.28) of [1]. Setting $\boldsymbol{\Delta}=0$, we recover the original formulation of Moharam *et al.*, i.e., Eq. (6). The formulation of Li can be obtained by replacing $\mathbf{E}-\boldsymbol{\Delta}\mathbf{N}_{xx}$ by $[\![\varepsilon]\!]$, $\mathbf{E}-\boldsymbol{\Delta}\mathbf{N}_{yy}$ by $[\![\varepsilon]\!]$, and $\boldsymbol{\Delta}\mathbf{N}_{xy}$ by zero.

We can try to understand the relationship between the known formulations and the presented one. Moharam and co-workers' [3] formulation ignores Li's rules, which can be expressed by the assumption $\boldsymbol{\Delta}=0$ as mentioned in the previous paragraph. Lalanne and Morris's [6] formulation for line gratings in conical mountings can be derived from Eq. (8) using $\mathbf{N}=[1,0]$ or $\mathbf{N}=[0,1]$ if the grating rules are parallel to the y or x axis, respectively. Li's [10] formulation for 2D gratings, however, cannot easily be derived from Eq. (8). It can be understood as a generalization of the correct 1D formulation to 2D gratings, which can only be achieved by introducing zigzag boundary curves whose facets are parallel to the coordinate axes. Thus the coupling terms $\boldsymbol{\Delta}\mathbf{N}_{xy}$ in Eq. (8) become zero and the terms $\mathbf{E}-\boldsymbol{\Delta}\mathbf{N}_{xx}$ and $\mathbf{E}-\boldsymbol{\Delta}\mathbf{N}_{yy}$ are replaced by the results of a step-by-step transition to Fourier space. This transition assumes a correct NV for each of the two Fourier transforms (with respect to x and y) and thus treats the problem correctly. This is, however, possible only by the restriction to zigzag lines, which leads to a cancellation of the coupling terms.

5. SOME INSTRUCTIVE EXAMPLES

The theoretical formulation presented in Section 4 is only a first step for our work. Knowing the work of Popov and Nevière [1,2], it is quite obvious to apply the method of the NV fields not only to the differential method but also to the RCWA of crossed gratings. A second important step consists in setting up a proper NV field. In this section some examples will be treated, where setting up the NV field is either trivial or, at least, possible by descriptive considerations. The purpose of the following examples is to get a better understanding why the method leads to a better convergence than others and from which problems other formulations suffer. In Section 6, strategies for automatic generation of NV fields are developed and investigated. Without such strategies the method is less useful for practical purposes.

A. Tilted Line Grating

The most simple yet the most instructive example is a 1D binary line grating, which is modeled as a 2D grating with the groove direction within the unit cell tilted by a certain angle α as shown in Fig. 2. Since all material boundaries are parallel, the NV field can be chosen to be the constant field:

$$\mathbf{N}=\begin{bmatrix}-\sin\alpha\\ \cos\alpha\end{bmatrix}. \tag{9}$$

Hence its Fourier series has only one single nonvanishing component,

$$[N_x]_k=-\sin\alpha\,\delta_{k0},\quad [N_y]_k=\cos\alpha\,\delta_{k0}, \tag{10}$$

and the matrix $\mathbf{G}$ of Eq. (8) becomes

$$\mathbf{G}=\begin{bmatrix}-\mathbf{K}_x\mathbf{K}_y-\boldsymbol{\Delta}\sin\alpha\cos\alpha & \mathbf{K}_x^2-[\![\varepsilon]\!]\sin^2\alpha-[\![1/\varepsilon]\!]^{-1}\cos^2\alpha\\ [\![\varepsilon]\!]\cos^2\alpha+[\![1/\varepsilon]\!]^{-1}\sin^2\alpha-\mathbf{K}_y^2 & \mathbf{K}_x\mathbf{K}_y+\boldsymbol{\Delta}\sin\alpha\cos\alpha\end{bmatrix}. \tag{11}$$

To conform with the notation of 1D gratings, we define TE as the polarization, where the electric field is parallel to the grooves and TM where it is perpendicular.

It is instructive to look at the case $\alpha=90°$, where the matrix $\mathbf{G}$ reduces to

$$\mathbf{G}=\begin{bmatrix}-\mathbf{K}_x\mathbf{K}_y & \mathbf{K}_x^2-[\![\varepsilon]\!]\\ [\![1/\varepsilon]\!]^{-1}-\mathbf{K}_y^2 & \mathbf{K}_x\mathbf{K}_y\end{bmatrix}, \tag{12}$$

i.e., the correct formulation for 1D gratings by Lalanne and Morris [6]. Moreover, we restrict ourselves to in-plane incidence, where $\mathbf{K}_y=0$. In this case, for TE polarization, it is $\mathbf{S}_x=\mathbf{U}_y=0$, and we obtain from Eqs. (8) and (12)

$$\frac{\partial\mathbf{U}_x}{\partial z'}=(\mathbf{K}_x^2-[\![\varepsilon]\!])\mathbf{S}_y, \tag{13}$$

whereas for TM it is $\mathbf{S}_y=\mathbf{U}_x=0$ and we have

$$\frac{\partial\mathbf{U}_y}{\partial z'}=[\![1/\varepsilon]\!]^{-1}\mathbf{S}_x. \tag{14}$$

Hence, we recover the correct formulations for 1D gratings, as given by Eqs. (IX.30) and (IX.31) from [1]. We should note, however, that for this case the formulation of Li also yields the correct 1D limit. Yet, this is no longer true in the case of the tilted grating.

To see that the matrix $\mathbf{G}$ correctly describes the tilted line grating, we decompose the electric field vector into components $\mathbf{S}_T$ and $\mathbf{S}_N$, which are tangential and normal to the grooves, respectively. The relationship to the Cartesian components is simply given by a rotation:

$$\begin{bmatrix}\mathbf{S}_T\\ \mathbf{S}_N\end{bmatrix}=\begin{bmatrix}\cos\alpha & \sin\alpha\\ -\sin\alpha & \cos\alpha\end{bmatrix}\begin{bmatrix}\mathbf{S}_x\\ \mathbf{S}_y\end{bmatrix}. \tag{15}$$

Similarly, we can construct $\mathbf{U}_T$, $\mathbf{U}_N$, $\mathbf{K}_T$, and $\mathbf{K}_N$. In these quantitites, the equations have to be structurally the same as for a grating in Cartesian coordinates with $\alpha=90°$. The same also holds for $\alpha=0°$, where $\mathbf{G}$ reads

$$\mathbf{G}=\begin{bmatrix}-\mathbf{K}_x\mathbf{K}_y & \mathbf{K}_x^2-[\![1/\varepsilon]\!]^{-1}\\ [\![\varepsilon]\!]-\mathbf{K}_y^2 & \mathbf{K}_x\mathbf{K}_y\end{bmatrix}. \tag{16}$$

In this case, the grooves are parallel to the x axis, and $\mathbf{S}_y$ is the normal component and $\mathbf{S}_x$ the parallel component. Hence, in the new quantitities the matrix $\mathbf{G}$ should take

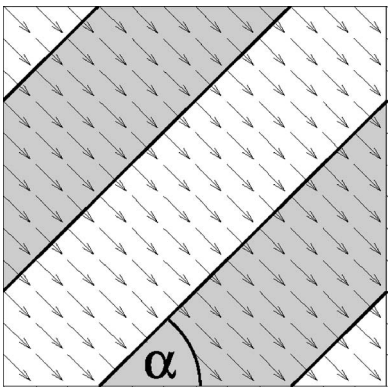


Fig. 2. Tilted line grating modeled as a 2D structure.

the same form as in Eq. (16), with $\mathbf{K}_x$ replaced by $\mathbf{K}_T$ and $\mathbf{K}_y$ by $\mathbf{K}_N$, and Eq. (8) now becomes a matrix equation for $\mathbf{U}_T$, $\mathbf{U}_N$, $\mathbf{S}_T$, and $\mathbf{S}_N$:

$$\begin{bmatrix} \frac{\partial \mathbf{U}_T}{\partial z'} \\ \frac{\partial \mathbf{U}_N}{\partial z'} \end{bmatrix} = \begin{bmatrix} -\mathbf{K}_T\mathbf{K}_N & \mathbf{K}_T^2 - [\![1/\varepsilon]\!]^{-1} \\ [\![\varepsilon]\!] - \mathbf{K}_N^2 & \mathbf{K}_T\mathbf{K}_N \end{bmatrix} \begin{bmatrix} \mathbf{S}_T \\ \mathbf{S}_N \end{bmatrix}. \quad (17)$$

To switch back to Cartesian coordinates, we use relation Eq. (15) and obtain

$$\begin{bmatrix} \frac{\partial \mathbf{U}_x}{\partial z'} \\ \frac{\partial \mathbf{U}_y}{\partial z'} \end{bmatrix} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} -\mathbf{K}_T\mathbf{K}_N & \mathbf{K}_T^2 - [\![1/\varepsilon]\!]^{-1} \\ [\![\varepsilon]\!] - \mathbf{K}_N^2 & \mathbf{K}_T\mathbf{K}_N \end{bmatrix} \times \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} \mathbf{S}_x \\ \mathbf{S}_y \end{bmatrix}. \quad (18)$$

Performing the matrix multiplications and replacing $\mathbf{K}_N$ and $\mathbf{K}_T$ by their respective Cartesian components, we recover the matrix $\mathbf{G}$ as given by Eq. (11). Hence, the matrix $\mathbf{G}$ in Eq. (11) can be obtained via a similarity transformation from the conical 1D equations with appropriate application of the inverse rule and therefore correctly describes the tilted line grating in 2D Cartesian coordinates.

In Fig. 3 we show the convergence curves for the transmitted zeroth order for a tilted grating with tilt angle $\alpha = 45°$. The refractive index is $n = 1.5$, and we assume normal incidence from the substrate. The grating period is 2λ, the width of the grooves is λ, and the grating depth is $\lambda/(2(n-1))$. In each graph we plot the diffraction efficiency as a function of the truncation order M using the three considered formulations: Moharam's original formulation, Li's formulation, and the formulation using the NV field. As usual the Fourier series run from $-M \cdots M$, which yields $2M+1$ Fourier coefficients for each of the two directions of periodic continuation or $(2M+1)^2$ coefficients in total. Moreover, we have included the exact results from 1D computations with a sufficiently high number of Fourier modes, such that convergence was ensured.

The results are quite striking. For TM polarization Moharam's formulation yields slow convergence, which is due to the well known fact that it does not make use of the inverse Laurent rule. For TE polarization Moharam's formulation converges fast, as in this case no inverse rule is needed. This is in complete analogy to the nontilted case. Li's formulation, in contrast, leads to rapid convergence for the TM case but to a poor one for the TE case! It seems clear that Moharam's formulation outperforms Li's one for the TE case, as Moharam's formulation uses Laurent's rule, which is correct, whereas Li's formulation mixes up Laurent's rule with the inverse rule without need. It is not clear, however, why this mixing of the two rules does not worsen the performance of Li's method for the TM case. The NV formulation obeys Li's rules for both cases and thus always shows the best convergence behavior.

Summarizing, we have the following findings: Moharam's formulation leads to poor convergence for the TM case, whereas Li's formulation does so for the TE case. In contrast, the NV formulation yields rapid convergence for both polarizations.

B. Checkerboard Grating

As a second example we consider the checkerboard grating corresponding to example 1 of Li [10]. For convenience we repeat the grating data. The length of a square is 1.25λ, the grating material is characterized by $n = 1.5$, and the thickness of the squares is $h = \lambda$. Incidence is in normal direction from the substrate, and the polarization is parallel to one side of the squares.

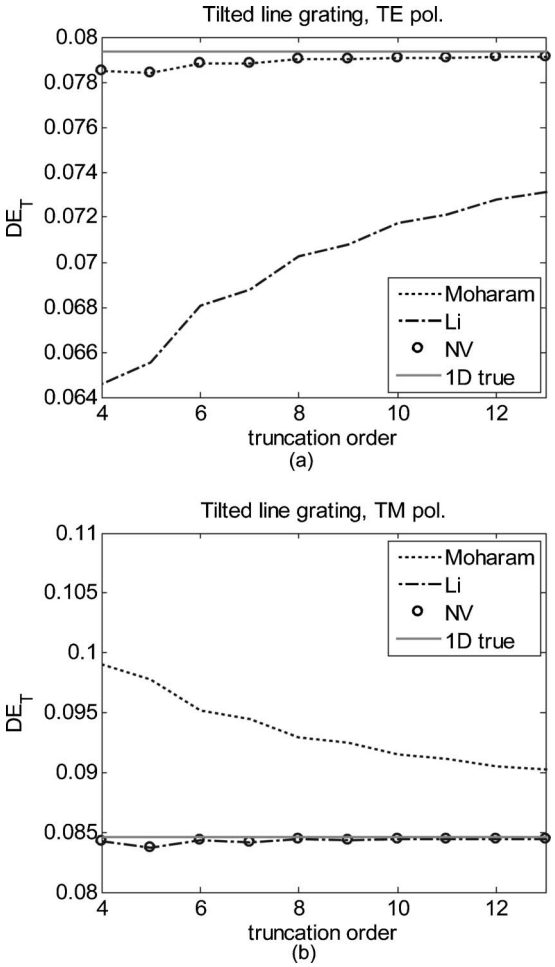


Fig. 3. Convergence curves for tilted line grating.

There are three possible unit cells in Li's paper, from which one uses an oblique-angled coordinate system, which will not be considered in this paper. Instead, we restrict ourselves to unit cell A [see Fig. 4(a)] and unit cell B [see Fig. 4(b) and 4(c)]. In unit cell A all material boundaries are parallel to the coordinate axes, whereas unit cell B is tilted by 45° with respect to unit cell A, which means that all material boundaries are tilted with respect to the coordinate system defined by the borders of unit cell A. In addition, it is smaller than unit cell A, which is expected to yield better convergence.

For the line grating the quest of finding an NV field was trivial, since the constant field did the job perfectly; however, for the checkerboard geometry, an appropriate

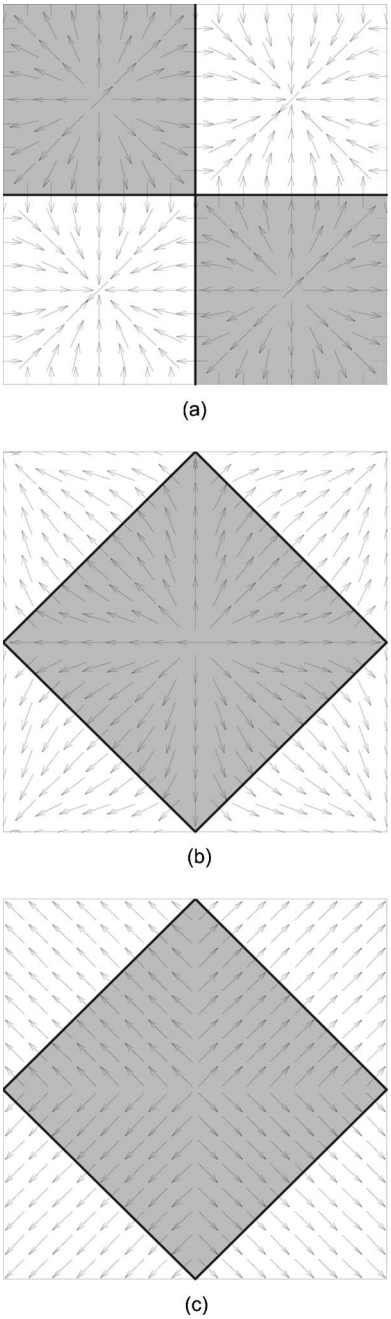


Fig. 4. Different unit cells for a checkerboard grating and different ways to set up the NV field.

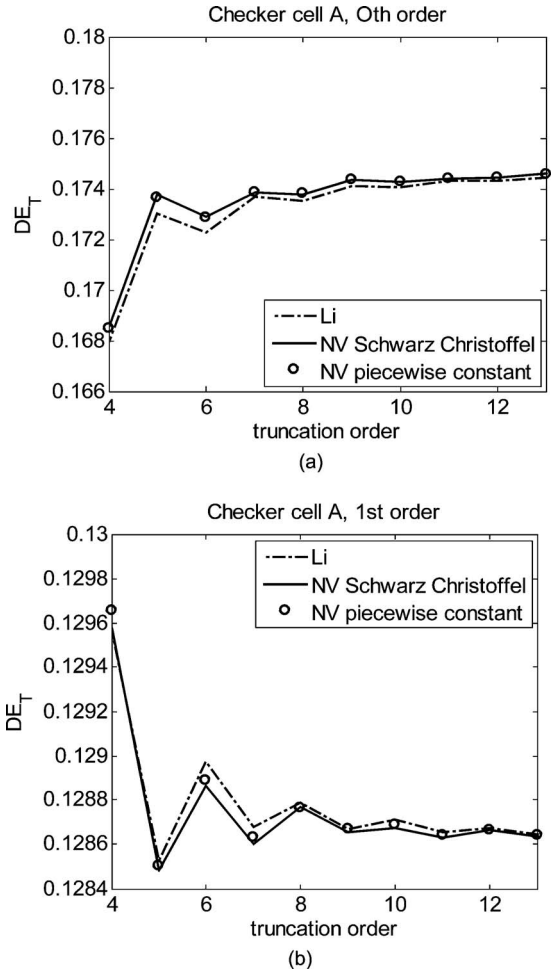


Fig. 5. Convergence curves for checkerboard cell A.

continuation of the NV field is no longer obvious. The most simple way to set up a suitable NV field is to introduce lines of discontinuties along the diagonals of the squares [see Fig. 4(c)]. This field describes the NV correctly at the boundaries and is pretty smooth in the close neighborhood of the latter, which seems to be an advantage. The lines of discontinuties, however, seem to be a disadvantage, as many Fourier coefficients are needed to represent them correctly in Fourier space. So we cannot be sure whether this NV field is ideal or not.

We try a different way to set up an NV field, which will be described in Subsection 6.A. This method makes use of the Schwarz–Christoffel transformation. The resulting NV field is depicted in Figs. 4(a) and 4(b) for the two possible unit cells. Let us state that the NV fields avoid lines of discontinuities but accept a few point singularities, some of them even lying on the material boundaries.

Figures 5 and 6 show the convergence of the (0,0) [Figs. 5(a) and 6(a)] and $(0,-1)$ [Figs. 5(b) and 6(b)] transmitted orders for units cells A and B, respectively, using Li's formulation and the NV formulation. For unit cell A both formulations show a similar convergence behavior. In contrast, for unit cell B the NV formulation using any of the described NV fields converges much more rapidly than Li's formulation. The results for the two different NV fields are almost identical. Moreover, we observe that the convergence of the NV formulation for unit cell B is faster

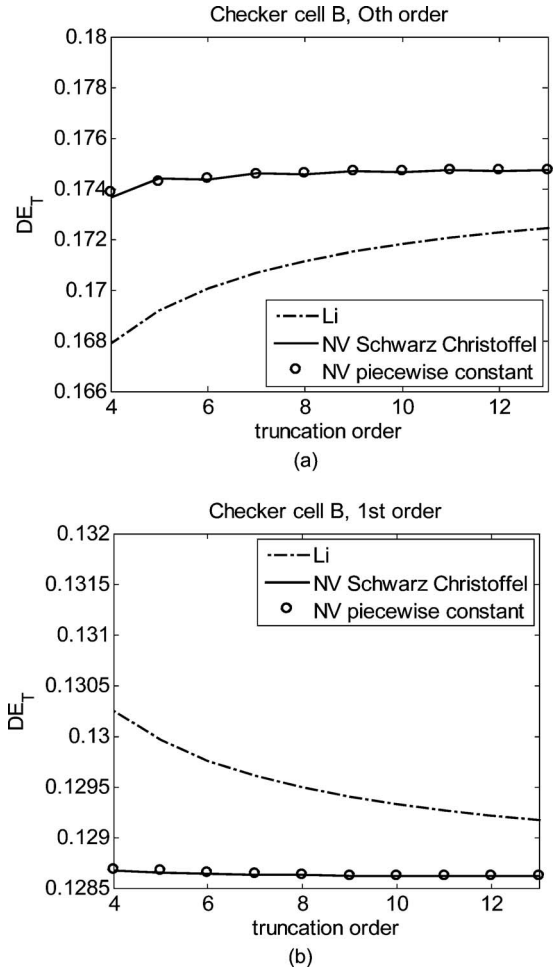


Fig. 6. Convergence curves for checkerboard cell B.

than for unit cell A, as is expected from its smaller size. However, this is not the case for Li's formulation, where the convergence for unit cell B is even worse than for unit cell A. This is at odds with the previously stated expectation and Li [10] already took notice of this difference and attributed it to the difference in the Fourier transforms. Whereas for unit cell A one can use analytic expressions, this is not the case for unit cell B, where one has to discretize the dielectric constant and perform the fast Fourier transform (FFT) numerically. However, we do not agree with this explanation, for we use the numerical FFTs for all three formulations and for both unit cells. Instead, we argue that the slower convergence for unit cell B rather comes from the fact that the tilted material boundaries are approximated by zigzag lines in the same manner as for the tilted line grating. In this case the formulation of Li leads to slow convergence whenever the polarization is parallel to this boundary. Since, for unit cell B, only half of the boundaries are either parallel or orthogonal to the incident polarization, the convergence is not as bad as in the case for the tilted line grating, where all the boundaries are parallel to the incident polarization in the TE case. As mentioned before, in unit cell A all the material boundaries are parallel to the coordinates, and here Li's formulation is perfectly appropriate.

6. PRACTICAL CONSIDERATIONS ABOUT SETTING UP THE NORMAL VECTOR

As demonstrated in Section 5, a considerable improvement of convergence with respect to the formulations known from the literature can be achieved. Considering the example of the checkerboard grating, we had to recognize that for most structures of practical interest it is not possible to set up an NV field without any discontinuites or singularities. One has the choice between accepting lines of discontinuities, which ought to be located the furthest away from the material boundaries, and point singularities, which are lying on the boundaries. There is no clearly defined rule which of the two choices is the better one.

In the following, two approaches for setting up NV fields for certain classes of structures are introduced. Both avoid lines of discontinuities and accept point singularities, even on the material boundaries, but both lead to better convergence behavior than the previous formulations of the RCWA. They are compared to the competing approach of accepting lines of discontinuities far away from the material boundaries. We cannot find a clear answer regarding which approach is the better one. The following algorithms, however, can be applied to a certain class of more or less arbitrary structures and thus are suitable for practical use. It is not obvious how the competing approach with lines of discontinuities can be realized in comparable algorithms.

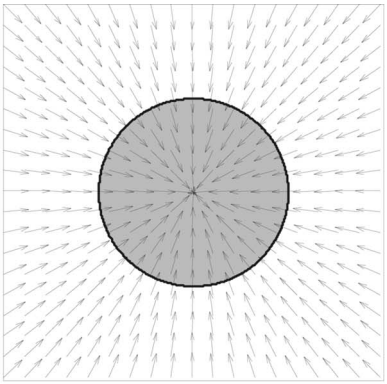
A. Use of the Schwarz–Christoffel Transformation for Polygonial Structures

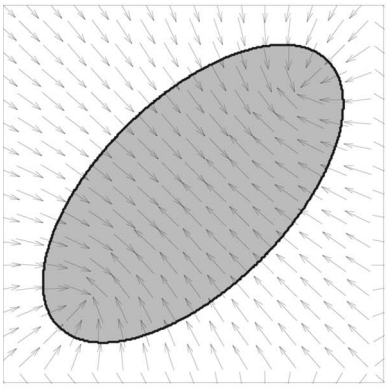
One possible way, which is particularly suited for polygonial geometries, is based on the use of the Schwarz–Christoffel transformation and will be outlined in the following. If the material boundary were a circle, as shown in Fig. 7(a), then a possible NV field would be a radial vector field, which would have just one singular point at the center but would be smooth at all other points. If we consider a polar coordinate system with its origin at the center of the circle, then the material boundary lies on $r=$const. coordinate lines, whereas the NV field is always tangential to the $\phi=$const. coordinate lines. If we now deform the material boundary, say, to an ellipse, the new coordinate field could be found by choosing an appropriate conformal coordinate transform, which transforms the circle into the ellipse. Since conformal transformations conserve angles, any orthogonal coordinate system will be transformed into another orthogonal system, and therefore the tangent vectors to the transformed $\phi'=$const. lines represent a valid NV field for the new geometry.

In the case of a polygonial material boundary, the NV field for the interior can be found by the Schwarz–Christoffel transformation, which transforms the unit disk into an area with an arbitrary polygonial boundary. Originally, the Schwarz–Christoffel transformation is a mapping of the upper complex half-plane into the interior of a polygon, but there exist other formulations with different preimages. Depending on the grating structure, one formulation might be advantageous over the other. For the checkerboard structure, we use the unit circle. To

this end we identify the y axis of the unit cell with the imaginary axis of the complex plane.

The coordinate lines after mapping the unit circle into the interior square of unit cell B are shown in Fig. 8. The NV field is represented by the vector tangent to the transformed coordinate lines $\phi' = \text{const}$. Thus, the NV field obtained by this procedure can then easily be extended to the remaining unit cell by appropriately shifting it into the four corners. The complete NV field has already been considered in Subsection 5.B and is depicted in Fig. 4(b) [and in Fig. 4(a) as well]. Note that the field is continuous everywhere except at nine points, namely, the center and the four corners of the unit cell and the four corners of the tilted square. The former lie in the homogeneous regions





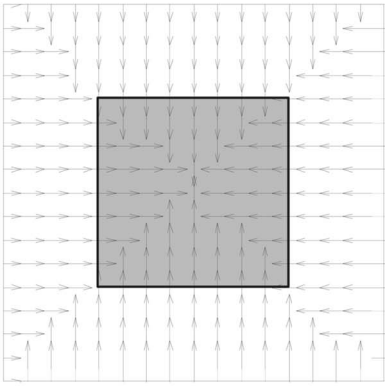


Fig. 7. Most obvious NV fields for simple geometric objects in the unit cell.

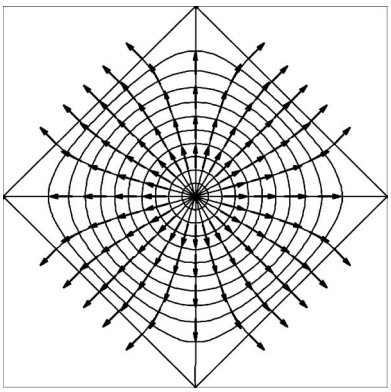


Fig. 8. Coordinate lines obtained by Schwarz–Christoffel transformation.

and represent source or drain points, whereas the latter are saddle points. To generate these figures we used a freely available MATLAB implementation of the Schwarz–Christoffel transformation [12].

It should be mentioned that although it seems to be a very convenient way to use the Schwarz–Christoffel transformation to obtain smooth NV fields for a given polygonial grating geometry, the modeling still can become quite awkward for more complicated grating structures. Moreover, for a finely discretized unit cell, the setting up of the NV field at each point can become a very time-consuming process, which significantly stretches the computation time of the complete solution of the diffraction problem. One advantage, however, is that for a sequence of computations for a range of wavelengths, incidence angles, or height variations, the NV field has to be computed only once, since it only depends on the grating geometry in the xy plane.

B. Use of an Electrostatic Model for Hole or Pillar Arrays

As the Schwarz–Christoffel transformation can only be applied to grating geometries with polygonial material boundaries, for curved smooth boundaries, a different way has to be found to set up the NV field. Particularly simple but widely used examples are gratings consisting of a single convex cavity (or column) within the elementary cell, i.e., they possess a simply connected region of one material surrounded by an ambient medium. Examples are arrays of circular or elliptical pillars or holes with arbitrary orientation.

Considering such structures, we again have two possibilities, either accept lines of discontinuities away from the boundaries or accept point singularities on the boundaries. The first possibility is depicted in Fig. 7 for three structures: a circle (a), a tilted ellipse (b), and a square (c). For such simple structures the choice of an NV field is quite obvious. For the circle, a field that points in a radial direction throughout the whole elementary cell is chosen. For the ellipse, finding the NV field is a little bit more complicated, but still straightforward. One can look up in standard mathematics textbooks, e.g., [13], the elliptical cylindrical coordinate system, where the lines of constant coordinates form families of confocal ellipses and confocal hyperbolas, respectively. The field lines of the NV field shown in Fig. 7(b) form this family of hyperbolas. For the

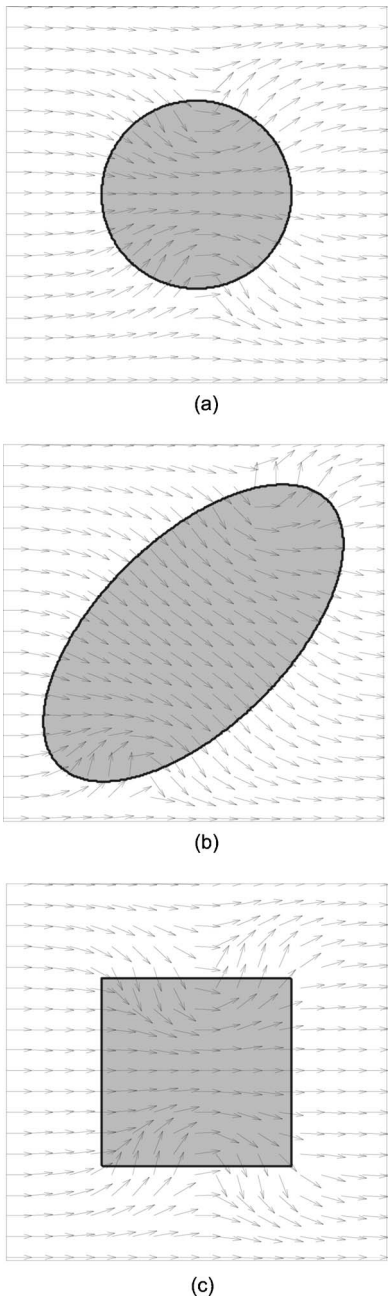


Fig. 9. Normal vector fields set up using the electrostatic model algorithm for three example structures.

square, the NV field can be set up similarly to the checkerboard grating by introducing lines of discontinuities along the diagonals of the square and choosing a constant field in the resulting areas, see Fig. 7(c).

For computing diffraction from more or less arbitrary structures it is desirable to find an algorithm that computes the NV field automatically. We developed such an algorithm, which in contrast to the NV fields described previously avoids line discontinuities but accepts singularities on the material boundaries. The algorithm is based on two electrostatic models applied to the convex cavity. Each of the two models is connected to one of two steps of the computation of the NV field.

In the first step the cavity in the elementary cell is regarded as a perfectly conducting object in an infinitely expanded plate capacitor. We choose the positive plate on the left side, the negative one on the right side. Poisson's equation is solved for that setting. The resulting electrical field is normalized and accepted as the sought-after NV field outside the cavity. At two points, this NV field changes its orientation from inward to outward. Those two points are looked for as they are needed for the second step of the computation. In some cases they are given by symmetry considerations. For the second step, the boundary curve is cut into two pieces at those two points, and now the cavity itself is treated as a plate capacitor with curved plates. Again, Poisson's equation is solved, and the resulting electric field is normalized and accepted as the NV field, this time inside the cavity.

We applied the described algorithm to the same three structures as considered previously in this subsection, a circle, a tilted ellipse, and a square. We used the finite element method for solving Poisson's equation. The resulting NV fields are depicted in Fig. 9. As can easily be seen, there are always two singular points of the NV field lying on the boundary. It seems to be a fundamental limitation of the problem that one cannot get rid of these singularities unless one accepts complete lines of discontinuities.

We simulated diffraction of light with normal incidence from arrays of the described structures using the same three formulations of the RCWA. The simulation data are

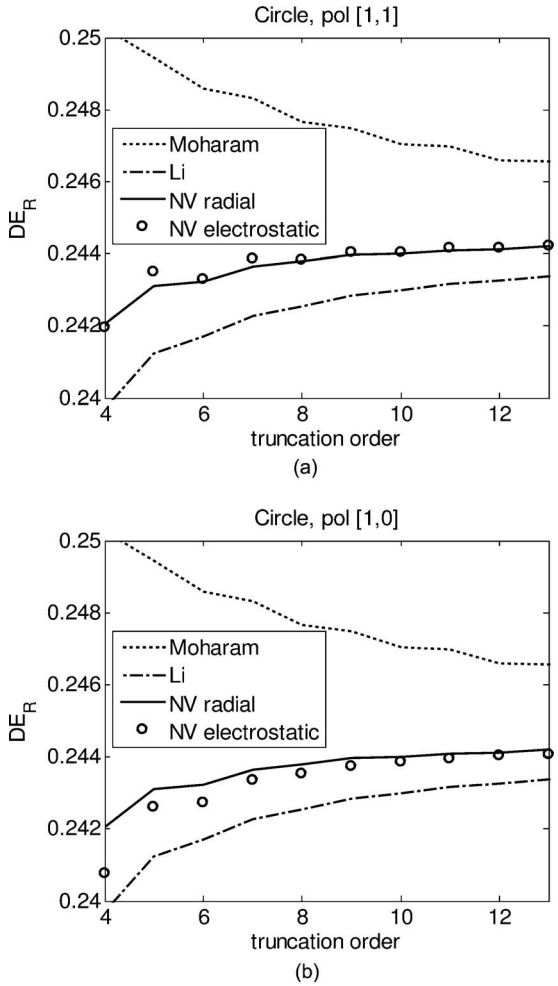


Fig. 10. Convergence curves for an array of circles.

again closely related to Li's [10] paper about convergence improvement using a quadrangular grid. The structures were supposed to be aerial cavities in a metallic layer with a refractive index of $1.75+1.5i$ and a thickness of 50 nm. The substrate was chosen to be glass with a refractive index of 1.5. The wavelength was 500 nm, the period in the x and y directions 1000 nm each. The side length of the square and the diameter of the circle were 500 nm, the semiaxes of the ellipse 1000 and 500 nm.

Figure 10 shows the diffraction efficiencies of the reflected zeroth order for the array of circles as a function of the truncation order M. In Fig. 10(a) diagonal polarization $[E_x,E_y]=[1,1]$ of the incident light is assumed, whereas the polarization in Fig. 10(b) is x polarization $[E_x,E_y]=[1,0]$. As can be seen, the NV method converges the fastest for both polarizations. The two choices of the NV field do not show considerable differences in the resulting convergence curves; the radial field is slightly better.

In Fig. 11 the diffraction efficiencies of the array of oblique ellipses can be seen. This time the two mutually orthogonal diagonal polarizations $[E_x,E_y]=[1,1]$ and $[E_x,E_y]=[1,-1]$ are compared to each other. This comparison is reasonable for only the ellipse array, as for the two other examples both polarizations lead to identical results due to the symmetry. The oblique ellipse somewhat resembles an oblique groove. Enlarging the semimajor axis toward infinity transforms the ellipse array into a tilted line grating. Thus it is reasonable to investigate the diagonal polarizations, which can be considered as "mainly TE" and "mainly TM." What do we mean by these names? For the mainly TE polarization the tangential component of the electric field is larger than the normal component for a maximum fraction of the boundary length; for the mainly TM polarization the opposite is true. The mainly TE polarization is a true TE polarization for the limit of an infinitely long semimajor axis; the same is true for the mainly TM polarization. Thus it is no surprise that the convergence curves show a similar behavior as those of the tilted line grating. For the mainly TE polarization the old formulation of Moharam *et al.* converges better than Li's formulation. For the mainly TM polarization Li's formulation works better than the old one. For both polarizations, the NV method shows the best convergence properties of the three. For the mainly TE polarizations the two different NV fields lead to a considerable difference in the resulting curves, but it is hard to decide which one is better. For the mainly TM polarization, however, both NV fields lead to virtually identical results.

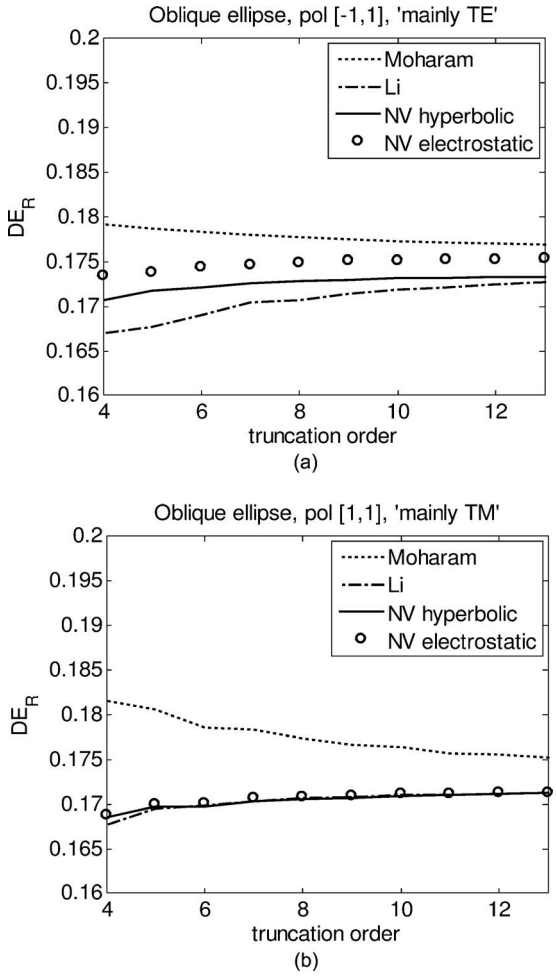


Fig. 11. Convergence curves for an array of oblique ellipses.

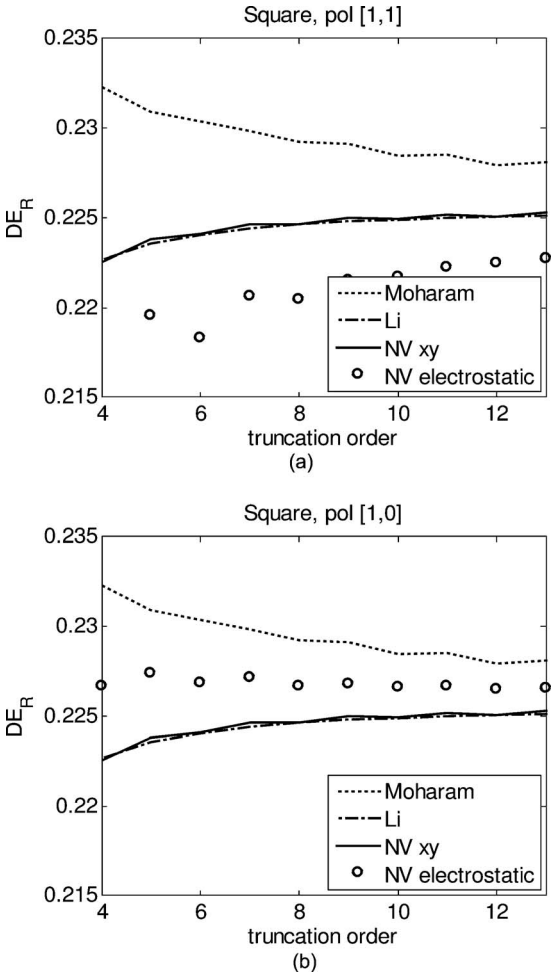


Fig. 12. Convergence curves for an array of squares.

In Fig. 12 the convergence behavior of the methods is investigated for an array of squares. Again, diagonal and x polarization are compared as in the case of the circle ar-

ray. It is no surprise that Li's method works best for this special structure, as it is perfectly adapted to its symmetry. As can be seen in Fig. 12(a), the NV method with the NV field from the electrostatic model algorithm is rather poor, at least for this particular polarization. Obviously, the point singularities are more harmful in this case than in others. The piecewise constant NV field, which is denoted by *xy*, however, leads to almost identical results as Li's method. This is also true for *x*-polarization, which is shown in Fig. 12(b). Here, in contrast, the NV field from the electrostatic model algorithm leads to the best convergence.

7. CONCLUSION

We combined Popov and Nevière's [1,2] formulation of the differential method with the recognized formulation of the RCWA [3,4] for crossed gratings. Thus we achieved similar or better convergence than proposals for convergence improvement, which are known from the literature. This method, however, requires introducing an NV field, which has to be defined over the whole grating unit cell, although it is strictly speaking used only at the material boundaries, where it serves to decompose the electric and magnetic transverse fields into tangential and normal components with respect to the boundaries in order to correctly apply Li's factorization rules for truncated Fourier series.

Unfortunately, there is no unique way to set up the NV field for any refractive index distribution in the elementary cell, and with improper choices the improvement in the convergence rate may be spoiled. We proposed some general approaches to how the NV ought to be set up for certain classes of structures. We will continue our efforts to find more general algorithms for setting up NV fields. Concave cavities or multiply connected regions such as ring systems could be the next targets of interest.

Although we demonstrated a better convergence for some structures than the previous formulations, we have to state that our method is not yet applicable to arbitrary structures. There is still a need to generalize the approaches for setting up the NV fields for arbitrary structures.

REFERENCES

1. E. Popov and M. Nevière, *Light Propagation in Periodic Media* (Dekker, 2003).
2. E. Popov and M. Nevière, "Maxwell equations in Fourier space: fast-converging formulation for diffraction by arbitrary shaped, periodic, anisotropic media," J. Opt. Soc. Am. A **18**, 2886–2894 (2001).
3. M. G. Moharam, E. B. Grann, D. A. Pommet, and T. K. Gaylord, "Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings," J. Opt. Soc. Am. A **12**, 1068–1076 (1995).
4. M. G. Moharam, D. A. Pommet, E. B. Grann, and T. K. Gaylord, "Stable implementation of the rigorous coupled-wave analysis for surface-relief gratings: enhanced transmittance matrix approach," J. Opt. Soc. Am. A **12**, 1077–1086 (1995).
5. M. Totzeck, "Numerical simulation of high-NA quantitative polarization microscopy and corresponding near-fields," Optik **112**, 399–406 (2001).
6. P. Lalanne and G. M. Morris, "Highly improved convergence of the coupled-wave method for TM polarization," J. Opt. Soc. Am. A **13**, 779–784 (1996).
7. L. Li, "Use of Fourier series in the analysis of discontinuous periodic structures," J. Opt. Soc. Am. A **13**, 1870–1876 (1996).
8. E. Popov, B. Chernov, M. Nevière, and N. Bonod, "Differential theory: application to highly conducting gratings," J. Opt. Soc. Am. A **21**, 199–206 (2004).
9. P. Lalanne, "Improved formulation of the coupled-wave method for two-dimensional gratings," J. Opt. Soc. Am. A **14**, 1592–1598 (1997).
10. L. Li, "New formulation of the Fourier modal method for crossed surface-relief gratings," J. Opt. Soc. Am. A **14**, 2758–2767 (1997).
11. E. Popov, M. Neviére, B. Gralak, and G. Tayeb, "Staircase approximation validity for arbitrary-shaped gratings," J. Opt. Soc. Am. A **19**, 33–42 (2002).
12. T. Driscoll, "Schwarz–Christoffel toolbox for MATLAB," http://www.math.udel.edu/~driscoll/software.
13. I. Sneddon, *Encylopaedic Dictionary of Mathematics for Engineers and Applied Scientists* (Wheatons, 1976).