Efficient finite-element, Green's function approach for critical-dimension metrology of three-dimensional gratings on multilayer films

Yia-Chung Chang

Department of Physics, University of Illinois at Urbana–Champaign, 1110 W. Green St., Urbana, Illinois 61801

Guangwei Li, Hanyou Chu, and Jon Opsal

Thermawave Inc., 1250 Reliance Way, Fremont, California 94539

Received May 17, 2005; revised July 30, 2005; accepted August 10, 2005

We present an efficient method for calculating the reflectivity of three-dimensional gratings on multilayer films based on a finite-element, Green's function approach. Our method scales as $N \log N$, where N is the number of plane waves used in the expansion. Therefore, it is much more efficient than the commonly adopted rigorous-coupled-wave analysis (RCWA), which scales as N^3. We demonstrate the effectiveness of this method by applying it to a two-dimensional periodic array of contact holes on a multilayer film. We find that our Green's function approach is about one order of magnitude faster than the RCWA approach when applied to typical contact holes considered in industry. For most cases, this method is efficient enough for application as a real-time, critical-dimension metrology tool. © 2006 Optical Society of America

OCIS codes: 120.2130, 050.1960, 120.3940.

1. INTRODUCTION

Recently, optical scatterometry has been recognized as a very important tool for critical-dimension (CD) metrology in the integrated circuit industry. The most popular scatterometry data analysis approach is the so-called rigorous-coupled-wave analysis (RCWA) (or simply Fourier expansion method) popularized by Moharam *et al.*[1,2] and Li.[3]

Initially, RCWA was plagued with convergence problems associated with the Fourier expansion in the in-plane direction and exponential instabilities in the vertical direction, though the latter problem is not limited just to RCWA. The exponential instability can be overcome with the so-called enhanced-transmittance method.[2] The poor convergence problem related to Fourier expansion in the in-plane direction has been overcome by various authors,[4–10] making RCWA a truly viable approach for grating calculations.

However, the RCWA approach becomes prohibitively inefficient when applied to three-dimensional (3D) gratings, because it involves a matrix diagonalization procedure, which scales as N^3, where N is the number of coupled waves used. For 3D objects, expansion is needed in each of the two in-plane directions. Thus, N is typically larger than 100, which makes the realization of real-time CD metrology impractical. In this paper, we describe the development of an efficient numerical method based on the Green's function (GF) approach that can outperform the RCWA method substantially when applied to 3D gratings.

Here, we consider a two-dimensional (2D) periodic array (in the x–y plane) of 3D gratings on a multilayer film as illustrated in Fig. 1. Our goal is to find an efficient numerical method to calculate the principal-order reflection coefficients for both TE and TM modes as functions of the wavelength of incident radiation. In the current approach, we obtain the scattered-wave function by solving the Lippmann–Schwinger (LS) equation,[11]

$$\Psi(\mathbf{r}) = \Psi_0(\mathbf{r}) + \int G(\mathbf{r},\mathbf{r}')V(\mathbf{r}')\Psi(\mathbf{r}')\mathrm{d}\mathbf{r}', \tag{1}$$

where $G(\mathbf{r},\mathbf{r}')$ is the 3×3 tensor GF of the unperturbed system, which is just the multilayer film in the absence of the grating structure. $V(\mathbf{r})=[1-\epsilon(\mathbf{r})]k_0^2$ defines the perturbation, which describes the difference due to the creation of a 3D grating on multilayer film. $k_0=2\pi/\lambda_0$, where λ_0 is the wavelength of the light in free space. Once the scattered-wave function (a three-component vector) is obtained, we calculate the principal-order reflection coefficient r by solving the projected LS equation

$$r = r_0 + \int \mathrm{d}\mathbf{r} e^{ik_{0x}x+ik_{0y}y}\delta(z)\int G(\mathbf{r},\mathbf{r}')V(\mathbf{r}')\Psi(\mathbf{r}')\mathrm{d}\mathbf{r}', \tag{2}$$

where r_0 is the reflection coefficient for the zeroth-order structure and k_{0x},k_{0y} are the two in-plane components of the incident wave vector. Paulus and Martin[12] adopted the tensor GF approach to calculate the scattering of the EM field from isolated objects embedded in stratified media. In their approach, the GF is evaluated in real space via a $\mathbf{k}$-space integration, then the LS equation given in Eq. (1) is solved. Here, we are interested in the scattering from a 3D grating (a 2D periodic array of 3D objects) on a multilayer film, and we find that it is more efficient to expand the tensor GF in terms of a finite set of plane waves

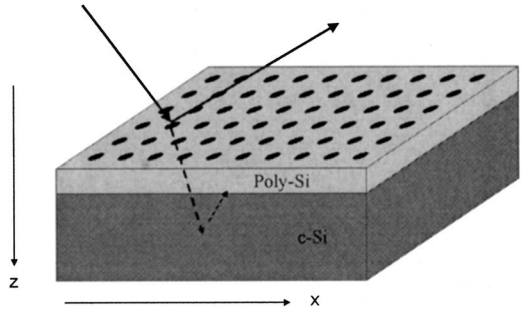

Fig. 1. Schematic diagram for ellipsometric measurement of a 2D periodic array of contact holes.

(in x and y) and perform the multiplications in reciprocal space.

2. FORMULATION OF THE TENSOR GREEN'S FUNCTION IN RECIPROCAL SPACE

Let the primitive lattice vectors be $\mathbf{a}_1$ and $\mathbf{a}_2$. The corresponding primitive reciprocal lattice vectors are denoted $\mathbf{b}_1$ and $\mathbf{b}_2$ with $\mathbf{b}_i \cdot \mathbf{a}_j = 2\pi\delta_{ij}$. We expand the wave function and GF in terms of plane wave $e^{i\mathbf{k}_n \cdot \vec{\rho}}$, where $\mathbf{k}_n = \mathbf{k}_0 + \mathbf{g}_n$; $n = 1, \cdots, N$, with $\mathbf{k}_0$ being the incident in-plane wave vector and $\mathbf{g}_n$ the reciprocal lattice vectors. We denote the x and y components of $\mathbf{k}_n$ as k_{xn} and k_{yn}.

The Maxwell's equations for electric fields are

$$\nabla \times \nabla \times \mathbf{E} = k^2 \mathbf{E},$$

$$\nabla \cdot \epsilon \mathbf{E} = 0,$$

where $k^2 = \epsilon k_0^2$ and $k_0 = \omega/c$ is the wave number. The electric field is related to the magnetic field via

$$\mathbf{H}(\mathbf{r}) \sim \nabla \times \mathbf{E}(\mathbf{r}).$$

We denote the dyadic Green's functions for the uniform multilayer material as G, which satisfies

$$\nabla \times \nabla \times G - k^2 G = \nabla\nabla \cdot G - \nabla^2 G - k^2 G = \delta(\mathbf{r}' - \mathbf{r}), \quad (3)$$

and the boundary conditions are that G and $\nabla \times G$ are continuous across a z boundary. ϵ_a is the dielectric constant of the host medium.

Because of the periodicity of the system, we can write G in terms of a plane-wave basis (for $0 < z' < d$) in the form

$$G(\mathbf{r}' - \mathbf{r}) = \sum_n e^{i\mathbf{k}_n \cdot (\vec{\rho} - \vec{\rho}')} \mathcal{G}_n(z, z'). \quad (4)$$

Substituting this into Eq. (3) yields

$$[(i\mathbf{k}_n + \hat{z}\partial_z)(i\mathbf{k}_n + \hat{z}\partial_z) \cdot \mathcal{G}_n + (k_n^2 - \partial_z^2)\mathcal{G}_n] - k^2 \mathcal{G}_n = I\delta(z - z'),$$

where I is a 3×3 identity tensor. Taking the dot product of $(i\mathbf{k}_n + \hat{z}\partial_z)$ with the above gives

$$-k^2 (i\mathbf{k}_n + \hat{z}\partial_z) \cdot \mathcal{G}_n = (i\mathbf{k}_n + \hat{z}\partial_z)\delta(z - z'), \quad (5)$$

and we have

$$(q_n^2 - \partial_z^2)\mathcal{G}_n = \left[I + \frac{(i\mathbf{k}_n + \hat{z}\partial_z)(i\mathbf{k}_n + \hat{z}\partial_z)}{k^2} \right] \delta(z - z'), \quad (6)$$

where $q_n^2 \equiv k_n^2 - k^2$. It is convenient to choose a different coordinate frame (x', y') for each index n, such that $\mathbf{k}_n$ is parallel to the x' axis. We have

$$\hat{x} = \cos\phi_n \hat{x}' - \sin\phi_n \hat{y}',$$

$$\hat{y} = \sin\phi_n \hat{x}' + \cos\phi_n \hat{y}',$$

where $\phi_n = \tan^{-1}(k_{yn}/k_{xn})$. In this frame, $\mathcal{G}$ takes the simplified form

$$[\mathcal{G}] = \begin{bmatrix} \mathcal{G}_{xx} & 0 & \mathcal{G}_{xz} \\ 0 & \mathcal{G}_{yy} & 0 \\ \mathcal{G}_{zx} & 0 & \mathcal{G}_{zz} \end{bmatrix}$$

with the following boundary conditions for the three independent components:

$$\mathcal{G}_{xx}, \quad \mathcal{G}_{xz}, \quad \mathcal{G}_{yy}$$

are continuous, and

$$\frac{\epsilon_a(z)}{q_n^2}\partial_z \mathcal{G}_{xx}, \quad \frac{\epsilon_a(z)}{q_n^2}\partial_z \mathcal{G}_{xz}, \quad \partial_z \mathcal{G}_{yy},$$

are continuous across a z boundary.

The $y'y'$ component of $\mathcal{G}$ is obtained by writing

$$\mathcal{G}_{yy} = (e^{-q_n z} + e^{q_n z} R_n) f_n(z'), \quad \text{for } z > z',$$

$$\mathcal{G}_{yy} = (e^{q_n z} + e^{-q_n z} \bar{r}_n) g_n(z'), \quad \text{for } z < z',$$

and requiring the discontinuity of $\partial_z \mathcal{G}_{yy}$ at $z = z'$ to be -1. We have (in column vector form)

$$\begin{pmatrix} f_n \\ g_n \end{pmatrix} = \frac{1}{2q_n(z')} \begin{pmatrix} e^{Qz'} + \bar{r}u(Re^{Qz'} + e^{-Qz'}) \\ u(Re^{Qz'} + e^{-Qz'}) \end{pmatrix}, \quad (7)$$

where $u \equiv (1 - R\bar{r})^{-1}$ and $Q = q_n$. R and $\bar{r}$ are the forward and backward reflection coefficients, which can be obtained by the transfer-matrix method with the boundary conditions appropriate for E_y. We obtain the recursion relations

$$R_l = e^{-Q_l d_l}(M_l^- + M_l^+ R_{l+1}) T_l e^{-Q_l d_l}, \quad (8)$$

$$T_l \equiv (M_l^+ + M_l^- R_{l+1})^{-1}, \quad (9)$$

where l labels the layers, and $M_l^{\pm} \equiv 1 \pm J_l^{-1} J_{l+1}$ and $J_l = Q_l$. Similarly, $\bar{\mathbf{r}}_l$ are given by

$$\bar{\mathbf{r}}_{l+1} = (\mathbf{m}^- + \mathbf{m}^+ e^{-Q_l d_l} \bar{\mathbf{r}}_l e^{-Q_l d_l}) \mathbf{t}_l, \quad (10)$$

$$\mathbf{t}_l \equiv (\mathbf{m}^+ + \mathbf{m}^- e^{-Q_l d_l} \bar{\mathbf{r}}_l e^{-Q_l d_l})^{-1}, \quad (11)$$

where $\mathbf{m}_l^{\pm} \equiv 1 \pm J_{l+1}^{-1} J_l$.

For z and z' in the same medium, we have

$$\mathcal{G}_{yy} = \frac{1}{2q_n} e^{-q_n |z - z'|} + e^{q_n z} R_n f_n(z') + e^{-q_n z} \bar{r}_n g_n(z').$$

Similarly, $\mathcal{G}_{xx}$ is obtained by writing

$$\mathcal{G}_{xx} = (e^{-q_n z} + e^{q_n z}\tilde{R}_n)\tilde{f}_n(z'), \quad \text{for } z > z',$$

$$\mathcal{G}_{xx} = (e^{q_n z} + e^{-q_n z}\tilde{r}_n)\tilde{g}_n(z'), \quad \text{for } z < z',$$

and requiring the discontinuity of $\partial_z \mathcal{G}_{xx}$ at $z=z'$ to be $-q_n^2/k^2$. We then have (in column vector form)

$$\begin{pmatrix} \tilde{f}_n \\ \tilde{g}_n \end{pmatrix} = \frac{-q_n}{2k^2} \begin{pmatrix} e^{Qz'} + \tilde{r}\tilde{u}(\tilde{R}e^{Qz'} + e^{-Qz'}) \\ \tilde{u}(\tilde{R}e^{Qz'} + e^{-Qz'}) \end{pmatrix}, \tag{12}$$

where $\tilde{u} \equiv (1-\tilde{R}\tilde{r})^{-1}$. $\tilde{R}_n$ and $\tilde{r}_n$ are the forward and backward reflection coefficients, which can be obtained by the transfer-matrix method with the boundary conditions appropriate for E_x. The recursion relations for $\tilde{R}_n$ and $\tilde{r}_n$ are also given by Eqs. (8) and (10), except that J_l is replaced by $\epsilon_l Q_l^{-1}$ because of the different boundary conditions applied here.

For z and z' in the same medium, we have

$$\mathcal{G}_{xx} = \frac{-q_n}{2k^2} e^{-q_n|z-z'|} + e^{q_n z}\tilde{R}_n\tilde{f}_n(z') + e^{-q_n z}\tilde{r}_n\tilde{g}_n(z').$$

$\mathcal{G}_{xz}$, is obtained by writing

$$\mathcal{G}_{xx} = (e^{-q_n z} + e^{q_n z}\tilde{R}_n)\bar{f}_n(z'), \quad \text{for } z > z',$$

$$\mathcal{G}_{xx} = (e^{q_n z} + e^{-q_n z}\tilde{r}_n)\bar{g}_n(z'), \quad \text{for } z < z',$$

and requiring the discontinuity of $\mathcal{G}_{xz}$ at $z=z'$ to be $-ik_n/k^2$ [because of Eq. (5)] and $\int \mathcal{G}_{xz} dz$ to be continuous. We thus obtain

$$\begin{pmatrix} \bar{f}_n \\ \bar{g}_n \end{pmatrix} = \frac{-ik_n}{2k^2} \begin{pmatrix} e^{Qz'} + \tilde{r}\tilde{u}(\tilde{R}e^{Qz'} - e^{-Qz'}) \\ \tilde{u}(\tilde{R}e^{Qz'} - e^{-Qz'}) \end{pmatrix}. \tag{13}$$

For z and z' in the same medium, we have

$$\mathcal{G}_{xz} = \frac{-ik_n}{2k^2}\mathrm{sgn}(z-z')e^{-q_n|z-z'|} + e^{q_n z}\tilde{R}_n\bar{f}_n(z')\bar{g}_n(z') + e^{-q_n z}\tilde{r}_n$$

$$= \frac{ik_n}{q_n^2}\partial_z'\mathcal{G}_{xx}. \tag{14}$$

For z and z' located in different layers (labeled by l and l'), we first find the amplitudes $f_{l'}$ and $g_{l'}$ for the case with z in layer l'. We then use the following recursion relations (which are derived by solving the transfer-matrix equations[1]) to find the amplitudes f_l and g_l for z in layer l:

$$f_{l+1} = (M_l^+ + M_l^- R_{l+1})^{-1} e^{-Q_l d_l} f_l \equiv T_l e^{-Q_l d_l} f_l$$

for $z > z'$ and

$$g_l = e^{-Q_l d_l}(\mathbf{m}^+ + \mathbf{m}^- e^{-Q_l d_l}\bar{\mathbf{r}}_l e^{-Q_l d_l})^{-1} g_{l+1} \equiv e^{-Q_l d_l}\mathbf{t}_l g_{l+1}$$

for $z < z'$.

Using Eqs. (5) and (6), we can relate the remaining two components to $\mathcal{G}_{xx}$ and $\mathcal{G}_{xz}$ via

$$\mathcal{G}_{zx} = \frac{-ik_n}{q_n^2}\partial_z \mathcal{G}_{xx}, \tag{15}$$

$$\mathcal{G}_{zz} = \frac{1}{q_n^2}[-ik_n\partial_z\mathcal{G}_{xz} + \delta(z-z')] = \frac{k_n^2}{q_n^4}\partial_z\partial_z'\mathcal{G}_{xx} + \frac{1}{q_n^2}\delta(z-z'). \tag{16}$$

For z and z' in the same medium, we have

$$\mathcal{G}_{zx} = \frac{-ik_n}{2k^2}\mathrm{sgn}(z-z')e^{-q_n|z-z'|} + \frac{ik_n}{q_n}[-e^{q_n z}\tilde{R}_n\tilde{f}_n(z') + e^{-q_n z}\tilde{r}_n\tilde{g}_n(z')], \tag{17}$$

$$\mathcal{G}_{zz} = \frac{k_n^2}{2q_n k^2}e^{-q_n|z-z'|} + \frac{ik_n}{q_n}[-e^{q_n z}\tilde{R}_n\bar{f}_n(z') + e^{-q_n z}\tilde{r}_n\bar{g}_n(z')] - \frac{1}{k^2}\delta(z-z'). \tag{18}$$

Note that $G(\mathbf{r},\mathbf{r}')$ is symmetric, which implies $\mathcal{G}(k_n;z,z') = \mathcal{G}^T(-k_n;z',z)$. Here $\mathcal{G}(k_n;z,z') = \mathcal{G}_n(z,z')$. Rotating back to the original coordinates, we have

$$\mathcal{G}_n = \begin{bmatrix} \mathcal{G}_{11} & \mathcal{G}_{12} & \mathcal{G}_{13} \\ \mathcal{G}_{21} & \mathcal{G}_{22} & \mathcal{G}_{23} \\ \mathcal{G}_{31} & \mathcal{G}_{32} & \mathcal{G}_{33} \end{bmatrix} = \begin{bmatrix} \frac{k_{xn}^2}{k_n^2}\mathcal{G}_{xx} + \frac{k_{yn}^2}{k_n^2}\mathcal{G}_{yy} & k_{xn}k_{yn}(\mathcal{G}_{xx} - \mathcal{G}_{yy}) & \frac{k_{xn}}{k_n}\mathcal{G}_{xz} \\ \frac{k_{xn}k_{yn}}{k_n^2}(\mathcal{G}_{xx} - \mathcal{G}_{yy}) & \frac{k_{yn}^2}{k_n^2}\mathcal{G}_{xx} + \frac{k_{xn}^2}{k_n^2}\mathcal{G}_{yy} & \frac{k_{yn}}{k_n}\mathcal{G}_{xz} \\ \frac{k_{xn}}{k_n}\mathcal{G}_{zx} & \frac{k_{yn}}{k_n}\mathcal{G}_{zx} & \mathcal{G}_{zz} \end{bmatrix}.$$

Note that in the limit $k_n \to 0$, we have $\mathcal{G}_{xx} = \mathcal{G}_{yy}$, $\mathcal{G}_{xz} = \mathcal{G}_{zx} = 0$, and $\mathcal{G}_{zz} = -1/k^2\delta(z-z')$.

It can be shown that the LS equation holds:

$$\Psi(z) - \Psi^0(z) = \int \mathrm{d}z' \mathcal{G}(z,z')V(z')\Psi(z'), \tag{19}$$

where $V \equiv [\epsilon(z) - \epsilon_a]k_0^2$. Note that $\epsilon(z)$ is anisotropic in the laminated film in the grating region. We have

$$\epsilon = \begin{bmatrix} \alpha P^{-1} + (1-\alpha)\epsilon_3 & 0 & 0 \\ 0 & \beta P^{-1} + (1-\beta)\epsilon_3 & 0 \\ 0 & 0 & \epsilon_3 \end{bmatrix},$$

where ϵ_3 is the zz component of the dielectric tensor, and P as a matrix in the coupled-wave basis is the inverse of the ϵ_3 matrix. The use of P^{-1} in place of ϵ_3 can substantially improve the convergence rate in N as pointed out by Li[4,5] and others.[6–10] α and β are mixing parameters between 0 and 1. For 2D gratings perpendicular to the x axis, we have $\alpha=1$ and $\beta=0$. For 3D gratings, α and β depend on the shape of the grating object.[10] The three-component LS equation can be reduced to a two-component LS equation by utilizing the transverse

condition $\nabla\cdot\epsilon\mathbf{E}=0$ to eliminate one component (E_x). We have

$$\tilde{E}_x \equiv P^{-1}E_x = K_x^{-1}[i\partial_z(\epsilon E_z) - K_y(\epsilon E_y)].$$

Substituting this into the LS equation (for the y and z components) yields

$$E_y(z) - E_y^0(z) = \int dz'(\mathcal{G}_{21}\tilde{V}\tilde{E}_x + \mathcal{G}_{22}VE_y + \mathcal{G}_{23}VE_z), \tag{20}$$

$$E_z(z) - E_z^0(z) = \int dz'(\mathcal{G}_{31}\tilde{V}\tilde{E}_x + \mathcal{G}_{32}VE_y + \mathcal{G}_{33}VE_z), \tag{21}$$

where we have defined

$$\tilde{V} \equiv (1 - P\epsilon_a)k_0^2.$$

Note that both V and $\tilde{V}$ can be factored out as the product of a complex number and a real Toeplitz matrix, and the product $\epsilon E_i = VE_i + \epsilon_a E_i, i = x,y$, can be calculated quickly once VE_i are obtained. In the special case with $K_y = 0, E_z$ is proportional to $K_x\epsilon^{-1}H_y$, and the the above LS equations reduce to two decoupled integral equations, one for the TE mode (E_y) and the other for the TM mode (H_y). The resulting decoupled LS equations take the form

$$E_y(z) - E_y^0(z) = \int dz'\mathcal{G}_{22}(z,z')VE_y, \tag{22}$$

$$H_y - \tilde{E}_z^0(z) = \int dz'[\partial_{z'}\bar{\mathcal{G}}(z,z')\tilde{V}\partial_{z'}H_y + \bar{\mathcal{G}}(z,z')K_xV\epsilon^{-1}K_xH_y], \tag{23}$$

where

$$\bar{\mathcal{G}}(z,z') = \frac{1}{2q_n}e^{-q_n|z-z'|} - \frac{k^2}{ik_nq_n}[-e^{q_nz}\tilde{R}_n\bar{f}_n(z') + e^{-q_nz}\tilde{r}_n\bar{g}_n(z')].$$

Note that the reflection amplitudes $\tilde{R}$ and $\tilde{r}$ are related to the corresponing amplitudes for the TM mode (H_y component) by $\tilde{R} = -R_{TM}$ and $\tilde{r} = -\bar{r}_{TM}$.

3. NUMERICAL IMPLEMENTATION

The above derivation shows that the GF defined in the coupled-wave basis at any values of (z,z') can be evaluated analytically: We are now ready to solve the LS Eq. (19) numerically. We divide the integration domain ($0 < z' < d$) into M segments. We denote the width and midpoint of the jth segment by Δz_j and z_j; $j = 1, \cdots M$. The wave function is expanded in terms of basis functions $B_m(z - z_j)$ of order $m = 0$ (piecewise constant) or $m = 1$ (piecewise linear). We define

$$\Psi(z') = \sum_j \Psi(z_j)B_0(z' - z_j) \quad \text{for } m = 0,$$

$$\Psi(z') = \sum_{j=0}^{M} \Psi(z_j^+)B_1(z' - z_j^+) \quad \text{for } m = 1,$$

where $z_j^\pm \equiv z_j \pm \Delta z_j$. The basis functions are defined by

$$B_0(z' - z_j) = \begin{cases} 1 & \text{for } z_j^- < z' < z_+^j \\ 0 & \text{otherwise} \end{cases},$$

$$B_1(z' - z_j^+) = \begin{cases} (z' - z_j^-)/\Delta z_j & \text{for } z_j^- < z' < z_j^+ \\ (z_{j+1}^+ - z')/\Delta z_{j+1} & \text{for } z_j^+ < z' < z_{j+1}^+ \\ 0 & \text{otherwise} \end{cases}.$$

At end segments, we have $B_1(z' - z_0^+) = (z_1^+ - z')/\Delta z_1$ and $B_1(z' - z_M^+) = (z' - z_M^-)/\Delta z_M$. In each iteration, the wave function Ψ on the boundary mesh point z_j is updated by carrying out the integral over z' within each segment analytically, thus converting the LS equation into a simple sequence of matrix-vector multiplications.

Note that in all cases $G(z,z')$ contains terms that are proportional to the analytic functions $e^{-Qz'}$ or $e^{Qz'}$ and the perturbation V is assumed to be constant in each segment. The LS equation contains the following integrals, which are carried out analytically within each segment. For segment j, we have (for $m = 0$)

$$\int_{z_j^-}^{z_j^+} dz' e^{-Qz'}\Psi(z') = e^{-Qz_j^-}W_j\Psi(z_j)$$

and

$$\int_{z_j^-}^{z_j^+} dz' e^{Qz'}\Psi(z') = e^{Qz_j^+}W_j\Psi(z_j),$$

where $z_j^\pm \equiv z_j \pm \Delta z_j/2$ and

$$W_j \equiv (1 - e^{-Q\Delta z_j})Q^{-1}.$$

For the $m = 1$ case, the integrals can be worked out in similar fashion.

The above procedure is essential for making the LS equations numerically stable, since it brings down a factor Q^{-1} that decays as $1/n$ as the order of the coupled-wave basis increases. Without this extra factor, the calculation is slowly convergent (decaying as $1/n$), and a direct numerical evaluation of the LS equation on a finite mesh will be unstable. We can further improve the accuracy by performing an integration of the resulting function of Eq. (19) multiplied by the basis function $B_m(z - z_j)$. This brings down another factor Q^{-1}. For $m = 0$ expansion, special treatment is needed if z and z' fall in the same segment j. We have

$$\int_{z_j^-}^{z_j^+} \mathrm{d}z \int_{z_j^-}^{z_j^+} \mathrm{d}z' e^{-Q|z-z'|} = W_j^0, \tag{24}$$

where

$$W_j^0 = 2[\Delta z_j - (1 - e^{-Q\Delta z_j})Q^{-1}]Q^{-1} = 2W_j^+ \Delta z_j,$$

and

$$\int_{z_j^-}^{z_j^+} \mathrm{d}z \int_{z_j^-}^{z_j^+} \mathrm{d}z' \ \mathrm{sgn}(z - z') e^{-Q|z-z'|} = 0.$$

After performing the double-site integration, we obtain the following matrix equations:

$$\hat{A}X \equiv OX - \bar{G}VX = X^0, \tag{25}$$

where X^0 and X are $3NM$-dimensional column vectors with entries $X_n^0(j) = \int \mathrm{d}\mathbf{r} e^{-i\mathbf{k}_n \cdot \vec{\rho}} B_m(j,z) \Psi^0(\mathbf{r})$ and $X_n(j) = \int \mathrm{d}\mathbf{r} e^{-i\mathbf{k}_n \cdot \vec{\rho}} B_m(j,z) \Psi(\mathbf{r})$. O, V, and $\bar{G}$ are $3NM \times 3NM$ matrices with matrix elements given by

$$O_{n,n'}(j,j') = \delta_{n,n'} \int \mathrm{d}z B_m(z,j) B_m(z,j'), \tag{26}$$

$$V_{n,n'}(j,j') = I\delta_{j,j'} \int \mathrm{d}\mathbf{r} e^{i(\mathbf{k}_{n'} - \mathbf{k}_n)\cdot\vec{\rho}} B_m(z,j) B_m(z,j) V(z), \tag{27}$$

$$\bar{G}_{n,n'}(j,j') = \delta_{n,n'} \int \mathrm{d}z B_m(z,j) \mathcal{G}_n(z,z') B_m(z',j') dz'. \tag{28}$$

All these matrix elements can be evaluated analytically. The above linear matrix equations are solved iteratively by a quasi-minimal-residue (QMR) method. The initial guess of the wave function $\Psi(z')$ is taken to be $\Psi^0(z')$, which is evaluated on the mesh points z_j $(j = 1, \ldots, M)$.[13] We then perform the following numerical procedures:

1. The matrix-vector operations $VX \equiv Y$; this procedure takes N^2M steps, and it is the bottleneck of the whole algorithm. However, because V is a Toeplitz matrix, we can use a Toeplitz matrix vector multiplication subroutine that scales as $N \log N$. This will be advantageous when N becomes large.

2. The matrix-vector multiplications $\bar{G}Y$; here we utilize the fact that $\bar{G}$ is semiseparable, and the matrix-vector operation can be carried out with just $N \times M$ steps.

3. Evaluate $O^{-1}\hat{A}X \equiv X - O^{-1}B$.

Note that the eigenvalues of $O^{-1}\hat{A}$ are all close to 1 if the magnitude of GV is small compared with 1. This feature makes the QMR method a fast-converging one. The iteration procedure involves repeated multiplication of the form $O^{-1}\hat{A}X$. The number of iterations required typically ranges from 5 to 100, depending on the wavelength and the contrast of dielectric constants. Note that the efficiency of step 2 (the most time-consuming part) in our procedure can be improved by first transforming the vector X from k-space representation to real-space representation via a projection into finite-element basis functions in x and y or via a fast Fourier transform (FFT) with uniform 2D grids. We then perform the multiplication VX in real space, and finally transform the vector VX back into k space. The CPU time scales as $MN^{3/2}$ (for direct transform) or $MN \log N$ (for FFT), because the transformation can be done independently for the x and y directions. This scheme is more advantageous for a medium size of N. For large N, it is comparable to the Toeplitz matrix vector multiplication.

4. APPLICATION FOR THREE-DIMENSIONAL CYLINDRICAL OBJECTS

For gratings made of cylindrical pillars, the transformation between k space and real space can be further simplified. Each segment (labeled j) in the integration domain $(0 < z' < d)$ now contains a disc of radius a_j and thickness Δz_j. The wave function in each segment is then expanded in terms of basis functions defined as $\rho^m e^{i\nu\phi}/\sqrt{2\pi}$. We have

$$\Psi_j(\rho,\phi) = \sum_{m,\nu} C_j(m,\nu) \rho^{m-1} e^{i\nu\phi} S_m^{-1},$$

where $S_m = (\pi/m) a_j^m$ is the normalization constant. Note that in all cases $\mathcal{G}(z',z)$ contains terms that are proportional to the analytic functions $e^{i \sin(\phi+\phi_n) k_n \rho - q_n z}$ or $e^{i \sin(\phi+\phi_n) k_n \rho + q_n z}$. The LS equation contains the following integrals, which are carried out analytically within each segment. For a segment with $z_j \neq z_{j'}$, we have

$$S_m^{-1} \int_{z_j^-}^{z_j^+} \mathrm{d}z \int_0^{a_j} \int_0^{2\pi} \rho^m \mathrm{d}\phi e^{i \sin(\phi+\phi_n) k_n \rho} e^{i\nu\phi} e^{\pm q_n z} \mathrm{d}\rho$$

$$= 2\sqrt{\pi m} e^{i\nu\phi_n} W_j(m,\nu) e^{\pm q_n z_j^{\pm}} (1 - e^{-q_n \Delta z_j}) q_n^{-1}, \tag{29}$$

where

$$\phi_n = \tan^{-1}(k_{nx}/k_{ny}),$$

$$z_j^{\pm} \equiv z_j \pm \Delta z_j/2,$$

$$W_j(m,\nu) = \int_0^{a_j} \mathrm{d}\rho (\rho/a_j)^m J_\nu(k_n \rho).$$

Similarly, integrating over ρ', ϕ' in mesh j' gives $W_{j'}(m',\mu)$.

For diagonal elements with $j = j'$, we have

$$e^{i(\nu-\mu)\phi_n} \int_0^{a_j} \mathrm{d}\rho (\rho/a_j)^m \int_0^{a_j} \mathrm{d}\rho' (\rho'/R)^{m'} J_\mu(k_n \rho') J_\nu(k_n \rho) \times e^{-q_n|z-z'|} = e^{i(\nu-\mu)\phi_n} W_j(m',\mu) W_j(m,\nu) W_n^0(j),$$

$$W_n^0(j) = 2[\Delta s - (1 - e^{-q_n \Delta s})/q_n]/q_n.$$

To evaluate $W_j(m,\nu)$, we use the following formulas:

$$W(\nu+1,\nu) = k^{-1} J_{\nu+1}(k a_j),$$

$$W(\nu,\nu) = 2^{\nu-1} (k a_j)^{-\nu} a_j \sqrt{\pi} \Gamma(\nu + 1/2) [J_\nu(k a_j) \mathbf{H}_{\nu-1}(k a_j) - J_{\nu-1}(k a_j) \mathbf{H}_\nu(k a_j)],$$

where $\mathbf{H}_\nu(x)$ are the Struve functions,[14] which are ob-

tained via the series expansion and the recursion relation

$$\mathbf{H}_{\nu-1}(x)+\mathbf{H}_{\nu+1}(x)=\frac{2\nu}{x}\mathbf{H}_{\nu}(x)+\frac{(x/2)^{\nu}}{\sqrt{\pi}\Gamma(\nu+3/2)}.$$

Using the recursion relation

$$J_{\nu+1}(x)+J_{\nu-1}(x)=(2\nu/x)J_{\nu}(x),$$

we obtain

$$W(m+1,\nu-1)=\frac{2\nu}{ka_j}W(m,\nu)-W(m+1,\nu+1).$$

This relation allows us to obtain all $W(m,\nu)$ with $m>\nu$, starting from $W(\nu+1,\nu)$ and $W(\nu+1,\nu+1)$. Next, we use the recursion relation

$$J_{\nu+1}(x)=J_{\nu-1}(x)+2J'_{\nu}(x)$$

to obtain (with integration by parts)

$$W(m,\nu+1)=W(m,\nu-1)+2[a_jJ_{\nu}(ka_j)-mW(m-1,\nu)]/(ka_j).$$

This allows us to obtain all other $W(m,\nu)$, starting from $W(0,\nu)$ and $J_{\nu}(ka_j)$, while $W(0,\nu)=\int_0^{a_j}J_{\nu}(kx)\mathrm{d}x$.

5. RESULTS AND DISCUSSION

To test the efficiency and accuracy of our method, we compare the performance of our algorithm against a calculation based on RCWA. The reflection coefficients obtained from our k-space version of the Greens function approach agree with the corresponding RCWA results within an error less than 10^{-3} for all cases tested. Figure 2 shows the calculated results for the reflection amplitude of a selected grating structure that contains a 2D periodic array of contact holes in a layer of polycrystalline silicon. The grating has a periodicity of 1000 nm in both the x and y directions. The contact holes have a diameter of 307.5 nm and depth of 409 nm. The grating sits on a crystalline silicon substrate. The average contrast of the dielectric constant is ≈10. Fifteen plane waves in each of the x and y directions are used in the calculation, which corresponds

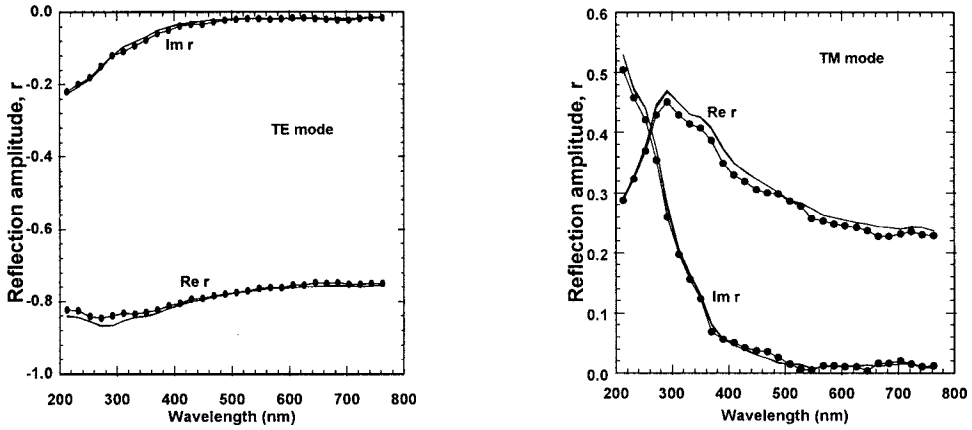

Fig. 2. Complex reflection amplitude r versus wavelength of the incident light calculated by the Green's function method for TE and TM modes. The real and imaginary parts are marked by Re r and Im r, respectively. Solid circles, k-space version; solid curves: r-space version.

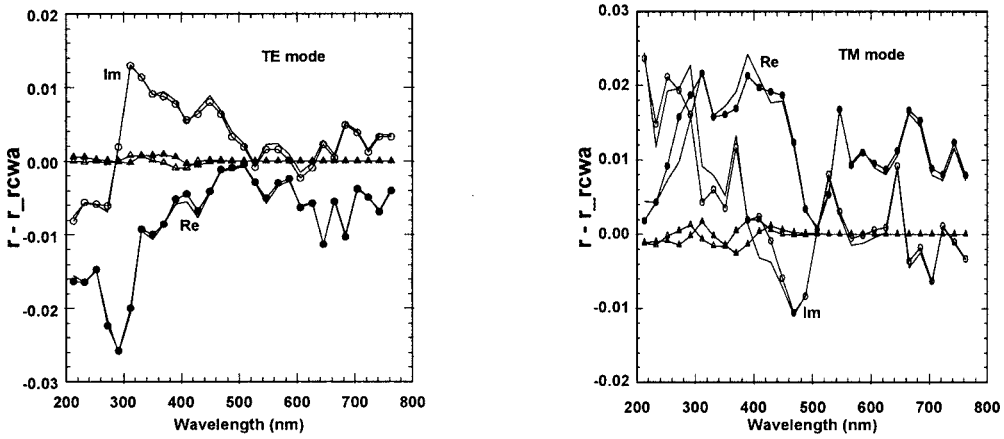

Fig. 3. Difference between reflection amplitude calculated by the GF approach (r) and RCWA (r_{rcwa}) versus wavelength of the incident light for TE and TM modes: solid (open) triangles, real (imaginary) part for k-space version; solid (open) circles, real (imaginary) part for r-space version with uniform-grid sampling. Solid curves, r-space version with high-order basis expansion.

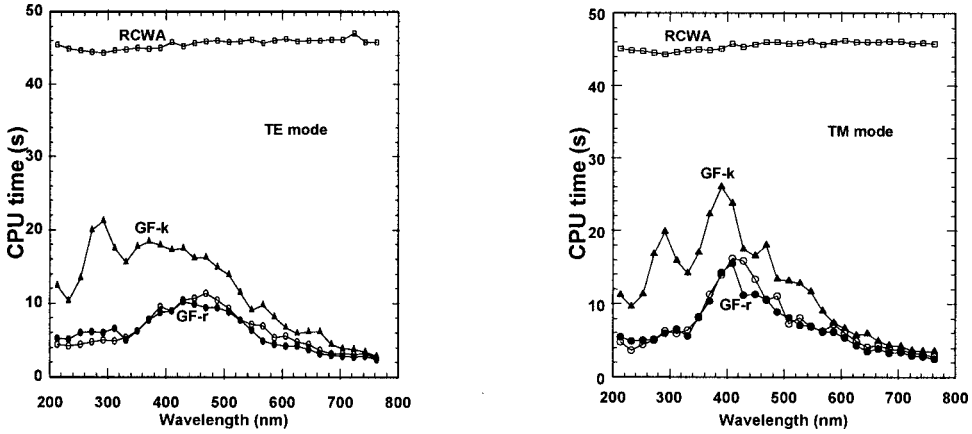

Fig. 4. CPU time required for three versions of the GF approach and RCWA: solid triangles, k-space version; solid (open) circles, r-space version with uniform-grid sampling (high-order basis expansion). Open squares, RCWA.

to $N=225$. The number of segments M used in the GF calculation depends on the wavelength λ and depth of the structure d. We find that the choice $M=15d/\lambda$ typically gives convergent results. Further increase of M would lead to a change of less than 10^{-4} on average. The computation speed is linearly proportional to the value of M. In Fig. 2, the results obtained from the two different real-space approaches (one with uniform-grid sampling and the other with high-order basis function expansion as described in Section 4) are almost indistinguishable. However, there are minor but noticeable differences between the real-space approach (GFr) and the k-space (GFk) approach.

Figure 3 shows the difference between the reflection amplitudes calculated by the GF approach and those from RCWA. The difference between the GFk approach and RCWA is almost negligible, while that between the GFr and RCWA is typically $\approx 10^{-2}$. Although the GFr approach is not as accurate as the GFk approach, it is computationally more efficient.

To demonstrate the superiority of our GF approach in terms of speed, we show in Fig. 4 the CPU time comparison between GF and RCWA approaches. We see that the GF approach is one order of magnitude faster than the RCWA approach when the contact hole considered has a fixed diameter (as shown for this example). For contact holes with varying diameters along the z direction, it requires the use of multiple slices in the RCWA approach. This will further increase the CPU time needed for the RCWA approach by a factor equal to the number of slices. On the other hand, the CPU time for the GF approach remains essentially unchanged, since multiple slices are already used in the GF approach even when the diameter is fixed. For the higher contrast case, the number of coupled waves N needed to get convergent results will increase, and the advantage of the GF approach becomes even more evident, since the CPU time for RCWA approach scales as N^3, while it scales as $N \log N$ for the GF approach.

6. SUMMARY

In summary, we have developed a finite-element GF method for calculating the reflectivity of 3D gratings on multilayer films. Our method scales as $N \log N$, where N is the number of plane waves used in the expansion. We have demonstrated that it is much more efficient than the commonly adopted RCWA method, which scales as N^3. We find that our GF approach is about one order of magnitude faster than the RCWA approach when applied to periodic arrays of contact holes such as those considered in industry. When variation of the hole diameter at different depths is considered, the advantage of this method becomes even more apparent. For most cases (with medium contrast), this method is efficient enough for application as a real-time CD metrology tool. The k-space GF method has also been tested on different geometries (rectangles and triangles) and for different contrasts (up to a contrast ratio of 1:100). Convergent results are always obtainable. However, when the perturbation is too large (e.g., a very large hole in a high-contrast medium), the number of iterations sometimes becomes very large, which makes this method not as attractive for high-contrast applications.

Corresponding author Y.-C. Chang's e-mail address is ychang1@yahoo.com.

REFERENCES AND NOTES

1. 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).
2. M. G. Moharam, D. A. Pommet, E. B. Grann, and T. K. Gaylord, "Stable implementation of the rigorous coupled-wave analysis of surface-relief gratings: enhanced transmittance matrix approach," J. Opt. Soc. Am. A **12**, 1077–1085 (1995).
3. L. Li, "Formulation and comparison of two recursive matrix algorithms for modeling layered diffraction gratings," J. Opt. Soc. Am. A **13**, 1024–1035 (1996).
4. L. Li, "Use of Fourier series in the analysis of discontinuous structures," J. Opt. Soc. Am. A **13**, 1870–1876 (1996).
5. L. Li, "Note on the S-matrix propagation algorithm," J. Opt. Soc. Am. A **20**, 655–660 (2003).
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. G. Granet and B. Guizal, "Really efficient implementation of the coupled-wave method for metallic lamellar gratings in TM polarization," J. Opt. Soc. Am. A **13**, 1019–1023 (1996).
8. P. Lalanne, "Improved formulation of the coupled-wave

method for two-dimensional gratings," J. Opt. Soc. Am. A **14**, 1592–1598 (1997).
9. G. Granet, "Reformulation of the lamellar grating problem through the concept of adaptive spatial resolution," J. Opt. Soc. Am. A **16**, 2510–2516 (1999).
10. E. Popov and M. Neviére, "Grating theory: new equations in Fourier space leading to fast converging results for TM polarization," J. Opt. Soc. Am. A **17**, 1773–1784 (2000).
11. See for example, Fayyazuddin and Riazuddin, *Quantum Mechanics*, (World Scientific, 1990), p. 368.
12. M. Paulus and O. J. F. Martin, "Green's tensor technique for scattering in two-dimensional stratified media," Phys. Rev. E **63**, 066615 (2001).
13. Note that the choice of the initial guess is rather arbitrary. It has been noted that there is no difference in terms of convergence speed whether the estimated initial guess jor the random vector generated by the QMR routine is used.
14. *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*, M. Abramowitz and I. A. Stegun, eds. (National Bureau of Standards, Applied Mathematics Series 55, 1972), p. 496.