JOURNAL OF MODERN OPTICS, 1993, VOL. 40, NO. 4, 553–573

A modal analysis of lamellar diffraction gratings in conical mountings

LIFENG LI

Optical Sciences Center, University of Arizona,
Tucson, Arizona 85721, USA

(*Received 3 March 1992; accepted 8 June 1992*)

Abstract. A rigorous modal analysis of lamellar gratings, i.e. gratings having rectangular grooves, in conical mountings is presented. It is an extension of the analysis of Botten *et al.* which considered non-conical mountings. A key step in the extension is a decomposition of the electromagnetic field in the grating region into two orthogonal components. A computer program implementing this extended modal analysis is capable of dealing with plane wave diffraction by dielectric and metallic gratings with deep grooves, at arbitrary angles of incidence, and having arbitrary incident polarizations. Some numerical examples are included.

1. Introduction

The modal approach has been applied by many authors to lamellar, non-perfectly conducting gratings in the past [1–10]. Most noticeably, Botten *et al.* [5–7] presented a series of three papers, in 1981, on the modal analysis of dielectric, finitely conducting, and highly conducting lamellar gratings. Their work was later formulated in a more systematic way, and its certain numerical aspects were improved, by Suratteau *et al.* [8] and by Tayeb and Petit [9]. However, these analyses are limited to non-conical mountings. In many applications, lamellar gratings are used in conical mountings [11]. Recently, a modal analysis of lamellar gratings in conical mountings was presented by Peng [12].

The present work differs substantially from that of Peng [12] in mathematical formulation and numerical implementation. In Peng's work, the validity of the decomposition of the electromagnetic field in the corrugated region into two orthogonal components is assumed without proof. The eigenfunctions (the modal fields in the corrugated region) for a grating in a conical mounting are constructed by geometrical means from the TE and TM eigenfunctions for the grating in an equivalent non-conical (also called classical) mounting. The completeness and orthogonality of the eigenfunctions assembled in this manner are not addressed. In this paper, the eigenfunctions along with their completeness and orthogonality are derived rigorously and systematically from the boundary-value problems.

The present work can be considered as an extension of the works of Botten *et al.* and Suratteau *et al.* to conical mountings. A key step in its development is the proof of the field decomposition mentioned above. Once this is done, the task of finding eigenvalues and eigenfunctions for a conical mounting reduces to that of a classical mounting, and the previous results of the above authors, including their powerful and sophisticated numerical methods for finding the eigenvalues, can be used.

The mathematical formulation of the modal analysis is presented in Section 2, and the numerical aspects of the analysis are addressed in Section 3. Section 4 provides some numerical results. The validity of the field decomposition is proved in

Appendix A. For the sake of the normal flow of the paper, some of the results available in [5–9] are re-stated in Sections 2 and 3, but often are formulated differently. In other instances, the reader is referred to the original references.

2. Mathematical formulation

2.1. *Notation*

A lamellar grating in a conical diffraction configuration is depicted in figure 1. The coordinate system is chosen such that the x axis is perpendicular to, and the z axis is parallel to, the grating grooves, and the y axis is the normal of the overall structure. A monochromatic plane wave of vacuum wavelength λ_0 is incident on the lamellar grating at a polar angle θ and an azimuthal angle ϕ. The range of θ is $0 \leqslant \theta < \pi/2$ and that of ϕ is $-\pi < \phi \leqslant \pi$, with the clockwise direction being the positive direction for ϕ. The same conventions will be used for the diffracted waves that are not shown. The incident polarization is in general elliptical.

The geometry of the lamellar grating is shown in figure 2. The grating period is d, and the widths of medium 1 and medium 2 are d_1 and d_2. We shall call the regions of space where $y > h/2$, $y < -h/2$, and $-h/2 < y < h/2$, regions 1, 2, and 0, respectively, where h is the grating groove depth. We shall use the superscript (j), where $j = 1, 2$, to denote quantities associated with regions j and the subscript j to denote quantities associated with the two media in region 0. Thus, the permittivity and permeability of the medium in region 1 are $\epsilon^{(1)}$ and $\mu^{(1)}$, and those in region 2 are $\epsilon^{(2)}$ and $\mu^{(2)}$. The permittivity and permeability of region 0 of periodic functions of x,

$$\left.\begin{aligned} \epsilon(x) &= \epsilon_1, \quad \mu(x) = \mu_1, \quad 0 \leqslant |x| \leqslant d_1/2, \\ \epsilon(x) &= \epsilon_2, \quad \mu(x) = \mu_2, \quad d_1/2 < |x| \leqslant d/2. \end{aligned}\right\} \tag{1}$$

Although for most optical applications, the permeability is a constant and equals that of the vacuum, to reveal the symmetry of the electric and magnetic fields, μ_1 and μ_2 are formally assumed to be different. The Gaussian system of units is used in this paper.

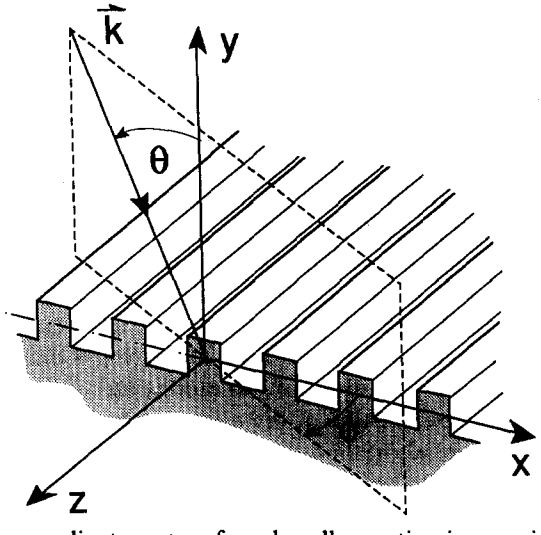

Figure 1. The coordinate system for a lamellar grating in a conical mounting.

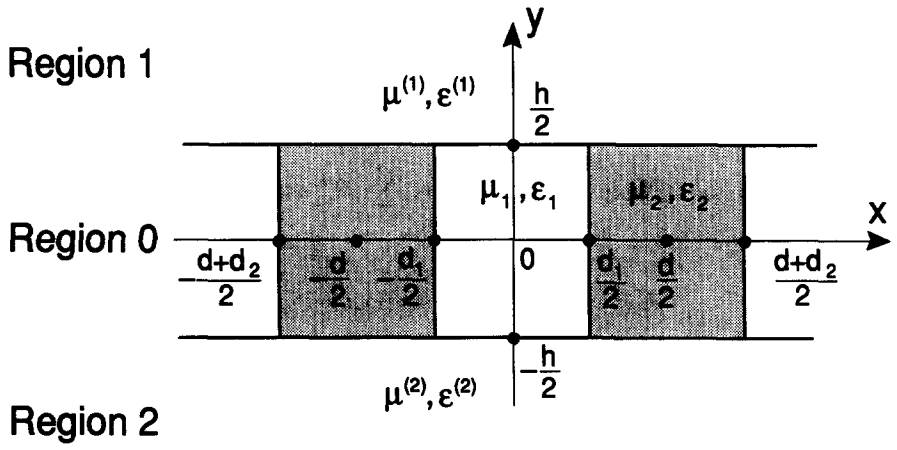

Figure 2. The geometry of a lamellar grating.

Let k_0 be the magnitude of the vacuum wave-vector. The magnitudes of the wave-vector in regions $j=1$, 2 and region 0 are denoted by $k^{(j)}$ and $k(x)$,

$$k^{(j)2} = \epsilon^{(j)}\mu^{(j)}k_0^2, \tag{2}$$

$$k^2(x) = \epsilon(x)\mu(x)k_0^2. \tag{3}$$

The wave-vector of the incident plane wave is

$$\mathbf{k} = k^{(1)}(\hat{\mathbf{x}}\sin\theta\cos\phi - \hat{\mathbf{y}}\cos\theta + \hat{\mathbf{z}}\sin\theta\sin\phi). \tag{4}$$

We denote the z component of the incident wave-vector by k_z,

$$k_z = k^{(1)}\sin\theta\sin\phi. \tag{5}$$

and we define the reduced magnitudes of the wave-vectors by $\tilde{k}^{(j)}$ and $\tilde{k}(x)$,

$$\tilde{k}^{(j)2} = k^{(j)2} - k_z^2, \tag{6}$$

$$\tilde{k}^2(x) = k^2(x) - k_z^2. \tag{7}$$

Since the physical problem is time and z invariant, the electromagnetic field may be assumed to be of the form

$$\mathbf{E}(x, y, z, t) = \mathbf{E}(x, y)\exp(\mathrm{i}k_z z - \mathrm{i}\omega t), \tag{8a}$$

$$\mathbf{H}(x, y, z, t) = \mathbf{H}(x, y)\exp(\mathrm{i}k_z z - \mathrm{i}\omega t). \tag{8b}$$

Throughout this paper, $\mathrm{i}=\sqrt{-1}$. Substituting (8 a) and (8 b) into Maxwell's equations we may express the transverse components of the electromagnetic field in terms of the longitudinal (z) components,

$$\mathbf{E}_t(x, y) = \frac{\mathrm{i}}{\tilde{k}^2(x)}[k_z\nabla_t E_z(x, y) - \mu k_0\hat{\mathbf{z}}\times\nabla_t H_z(x, y)], \tag{9a}$$

$$\mathbf{H}_t(x, y) = \frac{\mathrm{i}}{\tilde{k}^2(x)}[k_z\nabla_t H_z(x, y) + \epsilon k_0\hat{\mathbf{z}}\times\nabla_t E_z(x, y)], \tag{9b}$$

where $\nabla_t = \hat{\mathbf{x}}\partial_x + \hat{\mathbf{y}}\partial_y$. So in solving the conical diffraction problem, it is only necessary to work with the z components of the electric and magnetic fields.

As usual, the fields above and below the corrugated region may be written in Rayleigh expansions [13]:

$$E_z(x,y)=I_z^{(e)}\exp[i\alpha_0 x-i\beta_0^{(1)}y]+\sum_{n=-\infty}^{+\infty}R_n^{(e)}\exp[i\alpha_n x+i\beta_n^{(1)}y], \tag{10a}$$

$$H_z(x,y)=I_z^{(h)}\exp[i\alpha_0 x-i\beta_0^{(1)}y]+\sum_{n=-\infty}^{+\infty}R_n^{(h)}\exp[i\alpha_n x+i\beta_n^{(1)}y]. \tag{10b}$$

for $y>h/2$, and

$$E_z(x,y)=\sum_{n=-\infty}^{+\infty}T_n^{(e)}\exp[i\alpha_n x-i\beta_n^{(2)}y], \tag{11a}$$

$$H_z(x,y)=\sum_{n=-\infty}^{+\infty}T_n^{(h)}\exp[i\alpha_n x-i\beta_n^{(2)}y], \tag{11b}$$

for $y<h/2$, where

$$\alpha_n=\alpha_0+2n\pi/d, \qquad \alpha_0=k^{(1)}\sin\theta\cos\phi, \tag{12}$$

$$\beta_n^{(j)2}=\tilde{k}^{(j)2}-\alpha_n^2, \qquad \mathrm{Re}[\beta_n^{(j)}]+\mathrm{Im}[\beta_n^{(j)}]>0. \tag{13}$$

In (10) and (11), $I_z^{(e)}$, $I_z^{(h)}$, $R_n^{(e)}$, $R_n^{(h)}$, $T_n^{(e)}$, and $T_n^{(h)}$ are, respectively, the complex amplitudes of the z components of the incident and diffracted electric and magnetic fields in regions 1 and 2.

2.2. *Field decomposition*

Suppose h tends to infinity in figure 2, yielding a medium that is periodic in x and infinite in y and z. If the general expression for the electromagnetic field in this infinite medium is known, by imposing the interface conditions at $y=\pm h/2$ between this field and those given in (10) and (11), we can determine all the unknown field amplitudes.

For a z- and y-invariant medium, the following system of equations for the fields E_z and H_z can be derived from the original Maxwell's equations

$$\begin{pmatrix} \dfrac{\tilde{k}^2}{\epsilon}\dfrac{\partial}{\partial x}\left(\dfrac{\epsilon}{\tilde{k}^2}\dfrac{\partial}{\partial x}\right)+\dfrac{\partial^2}{\partial y^2}+\tilde{k}^2 & +\dfrac{k_z}{k_0}\dfrac{\tilde{k}^2}{\epsilon}\left(\dfrac{\mathrm{d}}{\mathrm{d}x}\dfrac{1}{\tilde{k}^2}\right)\dfrac{\partial}{\partial y} \\ -\dfrac{k_z}{k_0}\dfrac{\tilde{k}^2}{\mu}\left(\dfrac{\mathrm{d}}{\mathrm{d}x}\dfrac{1}{\tilde{k}^2}\right)\dfrac{\partial}{\partial y} & \dfrac{\tilde{k}^2}{\mu}\dfrac{\partial}{\partial x}\left(\dfrac{\mu}{\tilde{k}^2}\dfrac{\partial}{\partial x}\right)+\dfrac{\partial^2}{\partial y^2}+\tilde{k}^2 \end{pmatrix}\begin{pmatrix} E_z \\ H_z \end{pmatrix}=0. \tag{14}$$

To determine the modal representation of the electromagnetic field we need to solve (14) subject to the pseudo-periodic conditions [13].

$$\left.\begin{aligned} &E_z(d/2,y)=\exp(i\alpha_0 d)\,E_z(-d/2,y), \qquad H_z(d/2,y)=\exp(i\alpha_0 d)\,H_z(-d/2,y),\\ &\frac{\partial E_z}{\partial x}(d/2,y)=\exp(i\alpha_0 d)\frac{\partial E_z}{\partial x}(-d/2,y), \qquad \frac{\partial H_z}{\partial x}(d/2,y)=\exp(i\alpha_0 d)\frac{\partial H_z}{\partial x}(-d/2,y). \end{aligned}\right\} \tag{15}$$

In addition, we also need the interface conditions for the fields and their derivatives at the medium discontinuities. However, the burden of mentioning the interface conditions can be relieved if we understand Maxwell's equations in the sense of distribution [13].

Equation (14) is a coupled system of equations for E_z and H_z, whose direct solution seems to be difficult. However, by the folllowing field decomposition, its solution is simplified. In a z- and y-invariant medium, an electromagnetic field is said to be $E_\perp$ ($H_\perp$) if the x-component of its electric (magnetic) field vanishes. Let the

superscript (e) denote the $E_\perp$ field, and (h) the $H_\perp$ field. Then from (9)

$$k_x \frac{\partial}{\partial x} E_z^{(e)} + \mu k_0 \frac{\partial}{\partial y} H_z^{(e)} = 0, \tag{16a}$$

$$k_x \frac{\partial}{\partial x} H_z^{(h)} - \epsilon k_0 \frac{\partial}{\partial y} E_z^{(h)} = 0. \tag{16b}$$

In Appendix A, we prove that any field $(E_z, H_z)^T$, where superscript T signifies matrix transpose, satisfying (14) and (15) can be uniquely expressed as a sum of an $E_\perp$ field and an $H_\perp$ field, and these two component fields satisfy (14) and (15) independently. Thus the task of solving the vector-valued boundary-value problem (14) plus (15) is reduced to two independent scalar ones to be derived below.

Using (16 *a*) to eliminate H_x from the first equation of (14), and using (16 *b*) to eliminate E_x in the second of (14), we obtain

$$\mu \frac{\partial}{\partial x}\left(\frac{1}{\mu}\frac{\partial}{\partial x} E_z^{(e)}\right) + \frac{\partial^2}{\partial y^2} E_z^{(e)} + \tilde{k}^2 E_z^{(e)} = 0, \tag{17a}$$

$$\epsilon \frac{\partial}{\partial x}\left(\frac{1}{\epsilon}\frac{\partial}{\partial x} H_z^{(h)}\right) + \frac{\partial^2}{\partial y^2} H_z^{(h)} + \tilde{k}^2 H_z^{(h)} = 0. \tag{17b}$$

In these two equations, the electric field and the magnetic field are no longer coupled. Actually, these two equations are identical to the equations for the TE and TM polarized fields in classical mountings [5–8], provided that $\tilde{k}^2$ is replaced by k^2.

The symmetry exhibited by (17 *a*) and (17 *b*) with respect to $E_z^{(e)}$ and $H_z^{(h)}$ and with respect to ϵ and μ suggests that they can be rewritten as

$$\sigma^{(s)} \frac{\partial}{\partial x}\left(\frac{1}{\sigma^{(s)}}\frac{\partial}{\partial x} F^{(s)}\right) + \frac{\partial^2}{\partial y^2} F^{(s)} + \tilde{k}^2 F^{(s)} = 0, \tag{18}$$

where, and henceforth, $s = e$, h, and

$$F^{(e)} = E_z^{(e)}, \qquad F^{(h)} = H_z^{(h)}, \qquad \sigma^{(e)} = \mu(x), \qquad \sigma^{(h)} = \epsilon(x). \tag{19}$$

Let $G^{(s)}$ denote quantities complementary to $F^{(s)}$ such that

$$G^{(e)} = H_z^{(e)}, \qquad G^{(h)} = E_z^{(h)}. \tag{20}$$

Then, (16 *a*) and (16 *b*) become

$$\frac{\partial}{\partial y} G^{(s)} = \frac{\delta^{(s)} k_x}{\sigma^{(s)} k_0} \frac{\partial}{\partial x} F^{(s)}, \tag{21}$$

where

$$\delta^{(e)} = -1, \qquad \delta^{(h)} = 1. \tag{22}$$

The new notations introduced in (19), (20), and (22) allow the $E_\perp$ and $H_\perp$ fields to be treated identically in the rest of this paper.

Let a trial solution of (18) be

$$F^{(s)}(x, y) = u^{(s)}(x) w^{(s)}(y). \tag{23}$$

Then the standard procedure of separation of variables leads to

$$\sigma^{(s)} \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{\sigma^{(s)}}\frac{\mathrm{d}}{\mathrm{d}x} u^{(s)}\right) + (\tilde{k}^2 - \lambda^{(s)2}) u^{(s)} = 0, \tag{24}$$

where $\lambda^{(s)}$ is a constant. The differential equation (24) and the conditions (15) together pose a boundary-value problem which is considered in the next Section.

2.3. *The boundary-value problem*

In this Section and the next Section, for simplicity, we omit the superscript (s) in the relevant quantities. Let L be a differential operator defined by

$$L=\sigma\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{\sigma}\frac{\mathrm{d}}{\mathrm{d}x}\right)+\bar{k}^2. \tag{25}$$

Then the boundary-value problem for determining the eigenvalues and the eigenfunctions of the modal fields is given by

$$\left.\begin{aligned} Lu&=\rho u,\\ u(d/2)&=\exp(\mathrm{i}\alpha_0 d)\,u(-d/2),\\ u'(d/2)&=\exp(\mathrm{i}\alpha_0 d)\,u'(-d/2),\end{aligned}\right\} \tag{26}$$

where a prime indicates the differentiation with respect to x, and

$$\rho=\lambda^2. \tag{27}$$

Let us now define an inner product $(,)$ for any two integrable, complex-valued functions $u(x)$ and $v(x)$ by

$$(u,v)=\int_{-d/2}^{d/2}\frac{1}{\sigma(x)}u(x)\bar{v}(x)\,\mathrm{d}x, \tag{28}$$

where a bar indicates the complex conjugate. For a lossless dielectric grating ($\epsilon(x)>0, \mu(x)>0$), it is easily seen that L is self-adjoint, i.e.

$$(Lu,v)=(u,Lv). \tag{29}$$

From the theory of ordinary differential equations [14], we know that the eigenvalues determined by the boundary-value problem (26) are real and they form a denumerable sequence. Furthermore, the eigenfunctions form a complete, orthonormal basis in the sense that any continuous and piecewise differentiable function $f(x)$ satisfying the boundary conditions in (26) can be expanded in the eigenfunctions.

In order to embrace the most general cases, however, we assume that region 0 is composed of two media of complex permittivities and permeabilities, i.e. the functions $\epsilon(x)$ and $\mu(x)$ are in general complex valued. In addition, we assume α_0, and possibly k_z, to be complex. (This is a minor generalization of the works of [5–9].) This permits us to apply the current model to the homogeneous problem of grating diffraction [15]. In either case, the operator L is no longer self-adjoint; therefore, the eigenvalues of (26) are no longer necessarily real and the eigenfunctions are no longer orthogonal and complete. To be able to use the modal field expansions for the total field, it is necessary to consider the adjoint of (26), which is defined by

$$\left.\begin{aligned} L^+u^+&=\rho^+u^+,\\ u^+(d/2)&=\exp(\mathrm{i}\bar{\alpha}_0 d)\,u^+(-d/2),\\ u^{+\prime}(d/2)&=\exp(\mathrm{i}\bar{\alpha}_0 d)\,u^{+\prime}(-d/2),\end{aligned}\right\} \tag{30}$$

where the superscript $+$ indicates the adjoint and L^+ is the differential operator adjoint to L. It is easily seen that

$$(Lu,v^+)=(u,L^+v^+), \tag{31}$$

if

$$L^+ = \bar{L} = \bar{\sigma}\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{\bar{\sigma}}\frac{\mathrm{d}}{\mathrm{d}x}\right) + \overline{k^2}. \tag{32}$$

From the theory of non-self-adjoint boundary-value problems [14, 16], we know that under certain conditions, which (26) and (30) satisfy, two mutually adjoint boundary-value problems have the following properties.

(*a*) Both boundary-value problems possess an infinite number of eigenvalues and the eigenvalues can be ordered such that

$$\rho_m^+ = \bar{\rho}_m, \qquad m = 0, 1, 2, \ldots . \tag{33}$$

(*b*) The eigenfunctions $\{u_m\}$ and $\{u_n^+\}$, are bi-orthonormal, i.e.

$$(u_m, u_n^+) = \delta_{mn}. \tag{34}$$

(*c*) Any continuous and piecewise differentiable function $f(x)$ satisfying the boundary conditions in (26) has a uniformly convergent formal expansion

$$f(x) = \sum_{m=0}^{\infty} (f, u_m^+) u_m(x). \tag{35}$$

Hence, even for lossy dielectric or metallic gratings, it is still mathematically justified to represent the total electromagnetic field in region 0 by a superposition of modal fields, as has been done by Botten *et al.*

Incidently, the Rayleigh expansions (10 *a*, *b*) and (11 *a*, *b*) can be viewed as expansions in basis functions

$$e_n(x) = \exp(\mathrm{i}\alpha_n x). \tag{36}$$

It is easy to verify that $e_n(x)$ are eigenfunctions of (26) with L replaced by $\mathrm{d}^2/\mathrm{d}x^2$. The adjoint of this new boundary-value problem, with respect to a new inner product $\langle , \rangle$ defined by

$$\langle u, v \rangle = \int_{-d/2}^{d/2} u(x)\bar{v}(x)\,\mathrm{d}x, \tag{37}$$

is (30) with L^+ again replaced by d^2/dx^2, and

$$e_n^+(x) = \frac{1}{d}\exp(\mathrm{i}\bar{\alpha}_n x) \tag{38}$$

are the adjoint eigenfunctions.

2.4. *Eigenvalues and eigenfunctions*

The explicit forms of the characteristic equation for determining the eigenvalues and eigenfunctions can be most conveniently derived by taking advantage of the simplicity and symmetry of $\epsilon(x)$ and $\mu(x)$ given in (1). It is easy to verify that the following two functions are two linearly independent solutions of (24)

$$\varphi_e = \begin{cases} \cos\gamma_1 x, & 0 \leqslant |x| \leqslant \dfrac{d_1}{2}, \\ \cos\dfrac{\gamma_1 d_1}{2}\cos\gamma_2\left(|x| - \dfrac{d_1}{2}\right) - \dfrac{\sigma_2\gamma_1}{\sigma_1\gamma_2}\sin\dfrac{\gamma_1 d_1}{2}\sin\gamma_2\left(|x| - \dfrac{d_1}{2}\right), & \dfrac{d_1}{2} \leqslant |x| \leqslant \dfrac{d}{2}, \end{cases} \tag{39}$$

$$\varphi_o = \begin{cases} \dfrac{1}{\gamma_1}\sin\gamma_1 x, & 0\leqslant|x|\leqslant\dfrac{d_1}{2}, \\ \dfrac{1}{\gamma_1}\operatorname{sgn}(x)\left[\sin\dfrac{\gamma_1 d_1}{2}\cos\gamma_2\left(|x|-\dfrac{d_1}{2}\right)+\dfrac{\sigma_2\gamma_1}{\sigma_1\gamma_2}\cos\dfrac{\gamma_1 d_1}{2}\sin\gamma_2\left(|x|-\dfrac{d_1}{2}\right)\right], & \dfrac{d_1}{2}\leqslant|x|\leqslant\dfrac{d}{2} \end{cases} \tag{40}$$

where, for $j=1,2$,

$$\gamma_j^2=\tilde{k}_j^2-\rho. \tag{41}$$

Clearly, φ_e is an even function and φ_o is an odd function. The general solution of (24) is therefore given by

$$u(x)=A\varphi_e(x)+B\varphi_o(x), \tag{42}$$

where A and B are arbitrary constants. Imposition of the boundary conditions in (26) on (42) gives the characteristic equation

$$\Delta(\rho)=\begin{vmatrix}(1-\kappa)\varphi_e(d/2) & (1+\kappa)\varphi_0(d/2)\\ (1+\kappa)\varphi_e'(d/2) & (1-\kappa)\varphi_0'(d/2)\end{vmatrix}=0, \tag{43}$$

where

$$\kappa=\exp(\mathrm{i}\alpha_0 d). \tag{44}$$

It is obvious that when $\kappa=\pm1$, i.e. in Littrow mountings for which the normal incidence is a special case, $\Delta(\rho)$ is a product of two factors. For general angles of incidence, factorization of $\Delta(\rho)$ is impossible, and the explicit form of the characteristic equation is

$$\cos\gamma_1 d_1\cos\gamma_2 d_2-\frac{1}{2}\left(\frac{\sigma_2\gamma_1}{\sigma_1\gamma_2}+\frac{\sigma_1\gamma_2}{\sigma_2\gamma_1}\right)\sin\gamma_1 d_1\sin\gamma_2 d_2-\cos\alpha_0 d=0. \tag{45}$$

This is a transcendental equation for ρ, whose solutions are, in general, complex numbers.

The eigenfunctions of (26), expressed in terms of φ_e and φ_o are given by

$$u(x)=\begin{cases} \mathrm{i}C\varphi_o(x), & \text{if } \kappa=+1,\ \varphi_o(d/2)=0, \\ \mathrm{i}C\varphi_e(x), & \text{if } k=-1,\ \varphi_e(d/2)=0, \\ C[(1+\kappa)\varphi_o(d/2)\varphi_e(x)-(1-\kappa)\varphi_e(d/2)\varphi_o(x)], & \text{otherwise} \end{cases} \tag{46}$$

where C is the normalization constant. The eigenfunctions of the adjoint problem (30) can be simply obtained by replacing the relevant quantities in (39–46) by their adjoint counterparts. It can be shown that for each eigenvalue, there is in general only one eigenfunction and accidental degeneracy of an eigenvalue can only occur in Littrow mountings. Since the normalization constants C for $u(x)$ and C^+ for $u^+(x)$ are not individually fixed (34), we can demand $\bar{C}^+=C$. Then it can be shown that

$$\overline{u^+}(x)=\kappa^{-1}u(-x). \tag{47}$$

This direct relationship between the two mutually adjoint eigenfunctions is very useful in the numerical implementation of the theory.

2.5. *Modal field representation*

Since the basis functions $\{u_m^{(s)}\}$ are complete and bi-orthogonal in the sense stated in 2.3., the general solution of (18) for $F^{(s)}$ can be written as

$$F^{(s)} = \sum_{m=0}^{\infty} \omega_m^{(s)}(y) u_m^{(s)}(x), \tag{48}$$

where

$$\omega_m^{(s)}(y) = a_m^{(s)} \cos \lambda_m^{(s)} y + b_m^{(s)} \sin \lambda_m^{(s)} y, \tag{49}$$

and $a_m^{(s)}$ and $b_m^{(s)}$ are modal field amplitudes to be determined later. Substituting (48) into (21) and integrating with respect to y, we obtain an expression for $G^{(s)}$,

$$G^{(s)} = \sum_{m=0}^{\infty} \chi_m^{(s)}(y) w_m^{(s)}(x), \tag{50}$$

where

$$\chi_m^{(s)}(y) = \frac{k_z^2 + \lambda_m^{(s)2}}{\lambda_m^{(s)}} [-b_m^{(s)} \cos \lambda_m^{(s)} + a_m^{(s)} \sin \lambda_m^{(s)} y], \tag{51}$$

and

$$w_m^{(s)}(x) = \frac{1}{k_z^2 + \lambda_m^{(s)2}} \frac{k_z}{k_0} \frac{\delta^{(s)}}{\sigma^{(s)}} \frac{\mathrm{d}}{\mathrm{d}x} u_m^{(s)}(x). \tag{52}$$

(The integration constant that would appear in (50) can be shown to be zero.) Thus, by (48), (50), and (A 1) we complete the derivation of the following modal field representation of the total electromagnetic field in region 0:

$$\begin{pmatrix} E_z(x,y) \\ H_z(x,y) \end{pmatrix} = \sum_{m=0}^{\infty} \begin{pmatrix} \omega_m^{(e)}(y) u_m^{(e)}(x) \\ \chi_m^{(e)}(y) w_m^{(e)}(x) \end{pmatrix} + \sum_{m=0}^{\infty} \begin{pmatrix} \chi_m^{(h)}(y) w_m^{(h)}(x) \\ \omega_m^{(h)}(y) u_m^{(h)}(x) \end{pmatrix}. \tag{53}$$

Before closing this Section, we give the orthogonality relation between the vector-valued $E_\perp$ solution and $H_\perp$ solution. For this purpose, we define an inner product $[\,,\,]$ of two vector-valued functions φ and ψ such that

$$[\varphi, \psi] = \int_{-d/2}^{d/2} \varphi^{\mathrm{T}} \begin{pmatrix} -1/\mu & 0 \\ 0 & 1/\epsilon \end{pmatrix} \bar{\psi} \, \mathrm{d}x. \tag{54}$$

Let $\varphi_m^{(e)} = (u_m^{(e)}, w_m^{(e)})^{\mathrm{T}}$, and $\varphi_n^{+(h)} = (w_n^{+(h)}, u_n^{+(h)})^{\mathrm{T}}$. Then it can be shown that

$$[\varphi_m^{(e)}, \varphi_n^{+(h)}] = 0. \tag{55}$$

2.6 *Matching interface conditions*

Having obtained the expressions of the electromagnetic fields in all regions of the space, we are now ready to form the final system of linear equations for determining the unknown field amplitudes by applying the interface conditions at $y = \pm h/2$. The interface conditions are the following:

$$E_{z+} = E_{z-}, \qquad H_{z+} = H_{z-}, \tag{56a}$$

$$E_{x+} = E_{x-}, \qquad H_{x+} = H_{x-}, \tag{56b}$$

where the subscripts $\pm$ indicate limits from above and below the interface respectively, and from (9 *a*) and (9 *b*),

$$E_x=\frac{\mathrm{i}}{\tilde{k}^2}\left(k_z\frac{\partial}{\partial x}E_z+\mu k_0\frac{\partial}{\partial y}H_z\right), \tag{57 a}$$

$$H_x=\frac{\mathrm{i}}{\tilde{k}^2}\left(k_z\frac{\partial}{\partial x}H_z-\epsilon k_0\frac{\partial}{\partial y}E_z\right). \tag{57 b}$$

Substituting (10 *a*, *b*), (11 *a*, *b*), and (53) into (56 *a*) and (56 *b*) and carrying out some tedious algebra, we have, for the continuity of E_z and H_z at $y=+h/2$,

$$\sum_{m=0}^{\infty}[(\tilde{a}_m^{(e)}+\tilde{b}_m^{(e)})u_m^{(e)}(x)+(A_m^{(h)}\tilde{a}_m^{(h)}+B_m^{(h)}\tilde{b}_m^{(h)})w_m^{(h)}(x)]=\tilde{I}_z^{(e)}e_0(x)+\sum_{n=-\infty}^{\infty}\tilde{R}_n^{(e)}e_n(x), \tag{58 a}$$

and

$$\sum_{m=0}^{\infty}[(\tilde{a}_m^{(h)}+\tilde{b}_m^{(h)})u_m^{(h)}(x)+(A_m^{(e)}\tilde{a}_m^{(e)}+B_m^{(e)}\tilde{b}_m^{(e)})w_m^{(e)}(x)]=\tilde{I}_z^{(h)}e_0(x)+\sum_{n=-\infty}^{\infty}\tilde{R}_n^{(h)}e_n(x). \tag{58 b}$$

For the continuity of E_z and H_z at $y=-h/2$, we have

$$\sum_{m=0}^{\infty}[(\tilde{a}_m^{(e)}-\tilde{b}_m^{(e)})u_m^{(e)}(x)+(-A_m^{(h)}\tilde{a}_m^{(h)}+B_m^{(h)}\tilde{b}_m^{(h)})w_m^{(h)}(x)]=\sum_{n=-\infty}^{\infty}\tilde{T}_n^{(e)}e_n(x), \tag{58 c}$$

and

$$\sum_{m=0}^{\infty}[(\tilde{a}_m^{(h)}-\tilde{b}_m^{(h)})u_m^{(h)}(x)+(-A_m^{(e)}\tilde{a}_m^{(e)}+B_m^{(e)})w_m^{(e)}(x)]=\sum_{n=-\infty}^{\infty}\tilde{T}_n^{(h)}e_n(x). \tag{58 d}$$

For the continuity of E_x and H_x at $y=+h/2$, we have

$$\sum_{m=0}^{\infty}(A_m^{(e)}\tilde{a}_m^{(e)}+B_m^{(e)}\tilde{b}_m^{(e)})\frac{1}{\mu}u_m^{(e)}(x)$$

$$=(\tau_1^{(1)}\beta_0^{(1)}\tilde{I}_z^{(e)}+\tau_3^{(1)}\alpha_0\tilde{I}_z^{(h)})e_0(x)-\sum_{n=-\infty}^{\infty}(\tau_1^{(1)}\beta_n^{(1)}\tilde{R}_n^{(e)}-\tau_3^{(1)}\alpha_n\tilde{R}_n^{(h)})e_n(x), \tag{59 a}$$

and

$$\sum_{m=0}^{\infty}(A_m^{(h)}\tilde{a}_m^{(h)}+B_m^{(h)}\tilde{b}_m^{(h)})\frac{1}{\epsilon}u_m^{(h)}(x)$$

$$=(\tau_2^{(1)}\beta_0^{(1)}\tilde{I}_z^{(h)}-\tau_3^{(1)}\alpha_0\tilde{I}_z^{(e)})e_0(x)-\sum_{n=-\infty}^{\infty}[\tau_2^{(1)}\beta_n^{(1)}\tilde{R}_n^{(h)}+\tau_3^{(1)}\alpha_n\tilde{R}_n^{(e)}]e_n(x). \tag{59 b}$$

Finally, for the continuity of E_x and H_x at $y=-h/2$, we have

$$\sum_{m=0}^{\infty}(-A_m^{(e)}\tilde{a}_m^{(e)}+B_m^{(e)}\tilde{b}_m^{(e)})\frac{1}{\mu}u_m^{(e)}(x)=\sum_{n=-\infty}^{\infty}(\tau_1^{(2)}\beta_n^{(2)}\tilde{T}_n^{(e)}+\tau_3^{(2)}\alpha_n\tilde{T}_n^{(h)})e_n(x), \tag{59 c}$$

and

$$\sum_{m=0}^{\infty}(-A_m^{(h)}\tilde{a}_m^{(h)}+B_m^{(h)}\tilde{b}_m^{(h)})\frac{1}{\epsilon}u_m^{(h)}(x)=\sum_{n=-\infty}^{\infty}(\tau_2^{(2)}\beta_n^{(2)}\tilde{T}_n^{(h)}-\tau_3^{(2)}\alpha_n\tilde{T}_n^{(e)})e_n(x). \tag{59 d}$$

In the above equations the unknowns are

$$\tilde{a}_m^{(s)}=a_m^{(s)}\cos(\lambda_m^{(s)}h/2), \qquad \tilde{b}_m^{(s)}=b_m^{(s)}\sin(\lambda_m^{(s)}h/2), \tag{60}$$

$$\tilde{R}_n^{(s)} = R_n^{(s)} \exp[i\beta_n^{(1)} h/2], \qquad \tilde{T}_n^{(s)} = T_n^{(s)} \exp[i\beta_n^{(2)} h/2], \tag{61}$$

and the rest of the new notations are defined in Appendix B. The numerical solution of equations (58) and (59) will be considered in Section 3.2.

2.7. *Diffraction efficiency and polarization*

Once $\tilde{R}_n^{(s)}$ and $\tilde{T}_n^{(s)}$ are solved from (58) and (59), $R_n^{(s)}$ and $T_n^{(s)}$ are given by (61). By virtue of (9 *a*) and (9 *b*), all quantities of practical interest can be readily expressed in terms of $R_n^{(s)}$ and $T_n^{(s)}$. Suppose the media in regions 1 and 2 are lossless, and the incident plane wave is normalized such that

$$\frac{\beta_0^{(1)}}{k^{(1)2}} (\epsilon^{(1)} |I_z^{(e)}|^2 + \mu^{(1)} |I_z^{(h)}|^2) = 1. \tag{62}$$

Then, the diffraction efficiencies for the reflected and the transmitted propagating waves of order n are given by

$$\eta_n^{(1)} = \frac{\beta_n^{(1)}}{k^{(1)2}} (\epsilon^{(1)} |R_n^{(e)}|^2 + \mu^{(1)} |R_n^{(h)}|^2), \tag{63 a}$$

and

$$\eta_n^{(2)} = \frac{\beta_n^{(2)}}{k^{(2)2}} (\epsilon^{(2)} |T_n^{(e)}|^2 + \mu^{(2)} |T_n^{(h)}|^2). \tag{63 b}$$

If the media in region 0 are also lossless, the energy balance theorem holds:

$$\sum_n \eta_n^{(1)} + \sum_{n'} \eta_{n'}^{(2)} = 1, \tag{64}$$

where n and n' run through all propagating orders in regions 1 and 2, respectively.

In many applications involving conical mountings, it is very important to be able to predict the states of polarization of the diffracted orders. Let us associate with a propagating order having a wave-vector

$$\mathbf{k}_n^{(j)} = \hat{\mathbf{x}} \alpha_n \pm \hat{\mathbf{y}} \beta_n^{(j)} + \hat{\mathbf{z}} k_z, \tag{65}$$

two unit vectors $\hat{\mathbf{s}}_n^{(j)}$ and $\hat{\mathbf{p}}_n^{(j)}$ such that

$$\hat{\mathbf{s}}_n^{(j)} = \frac{\mathbf{k}_n^{(j)} \times \hat{\mathbf{y}}}{|\mathbf{k}_n^{(j)} \times \hat{\mathbf{y}}|}, \qquad \hat{\mathbf{p}}_n^{(j)} = \hat{\mathbf{s}}_n^{(j)} \times \frac{\mathbf{k}_n^{(j)}}{k^{(j)}}. \tag{66}$$

In (65) the plus sign is for $j = 1$, and the minus sign is for $j = 2$. The polarization of this diffracted order can be described with the following two angular parameters [17]:

$$\left.\begin{aligned} \alpha_n^{(j)} &= \arctan\left(\frac{|E_{ns}^{(j)}|}{|E_{np}^{(j)}|}\right), & 0 \leqslant \alpha_n^{(j)} \leqslant \frac{\pi}{2}, \\ \delta_n^{(j)} &= -\arg\left(\frac{E_{ns}^{(j)}}{E_{np}^{(j)}}\right), & -\pi < \delta_n^{(j)} \leqslant \pi, \end{aligned}\right\} \tag{67}$$

where $E_{ns}^{(j)}$ and $E_{np}^{(j)}$ are the s and p components of the electric field. (The notation $\alpha_n^{(j)}$ should not be confused with α_n defined in (12).) It is a simple exercise to show that for a diffracted wave in medium 1,

$$\frac{E_{ns}^{(1)}}{E_{np}^{(1)}}=\left(\frac{\mu^{(1)}}{\epsilon^{(1)}}\right)^{1/2}\frac{\alpha_n\epsilon^{(1)}k_0R_n^{(e)}+\beta_n^{(1)}k_zR_n^{(h)}}{\alpha_n\mu^{(1)}k_0R_n^{(h)}-\beta_n^{(1)}k_zR_n^{(e)}}. \tag{68}$$

The expression for a diffracted wave in medium 2 may be obtained from (68) by replacing $R_n^{(s)}$ by $T_n^{(s)}$, superscript (1) by (2), and $\beta_n^{(1)}$ by $-\beta_n^{(2)}$.

3. Numerical aspects

3.1. *Solution of the characteristic equation*

An efficient, reliable, and accurate numerical method for solving the characteristic equation (45) is of vital importance to the performance of a computer program implementing the modal analysis. Two very different numerical methods have been employed. The method of Botten *et al.* [7, 18] is a general one capable of finding all zeros of an analytic function in a prescribed region of the complex plane. That of Suratteau *et al.* [8] and Tayeb and Petit [9] is a problem-specific method that takes advantage of the fact that (45) can be factored in Littrow mountings. Both of these methods systematically find all eigenvalues of (45) in a prescribed region of the complex plane, and both of them perform well even for highly conducting gratings. As reported by the authors of the second method, their method is as effective as the first method but requires significantly shorter computation time. For this reason, we have decided to use the second method in our numerical implementation of this work. For details of these numerical methods, the interested reader is referred to the references cited above.

3.2. *Solution of the field amplitudes*

Before embarking on the numerical solution of (58) and (59), we first analyse the composition of these equations. In (58) and (59) there are eight equations and eight sets of unknowns. The right-hand sides of these equations are expanded in basis functions $\{e_n\}$, which are bi-orthogonal to their adjoint $\{e_n^+\}$. The left-hand sides are expanded in four different sets of functions $\{u_m^{(e)}\}$, $\{u_m^{(h)}\}$, $\{w_m^{(e)}\}$, and $\{w_m^{(h)}\}$. Of these four, as scalar-valued functions, $\{u_m^{(e)}\}$ and $\{u_m^{(h)}\}$ are bi-orthogonal to their respective adjoints, but not to each other. The functions $\{w_m^{(e)}\}$ and $\{w_m^{(h)}\}$ are proportional to the derivatives of $\{u_m^{(e)}\}$ and $\{u_m^{(h)}\}$ and they are not orthogonal to any other functions.

Equations (58) and (59) constitute a system of equations in known function expansions with unknown expansion coefficients. Such a system can be solved by the method of moments [19], which consists of three steps. First, a projection basis, i.e. a set of linearly independent testing functions, is chosen. Then, both sides of the series expansion equations are projected on to this basis by forming appropriate inner products with the testing functions. This step eliminates the x dependence of the equations and produces an algebraic linear system of equations of infinite dimension. Finally, the linear system is truncated to a finite order and its solution is obtained by the standard numerical techniques.

Since equations (58) and (59) are already expanded in terms of bi-orthonormal basis functions, it is advantageous to choose the adjoints of these basis functions as the testing functions so that the subsequent numerical solution can be simplified. For each of the eight equations we have two convenient projection bases, $\{u_m^{+(s)}\}$ and $\{e_n^+\}$. Therefore, there can be many different combinations of choices of projection bases

for the overall system. Following Suratteau *et al.*, a projection method in which the interface conditions for the z components are projected on to one basis and those for the x components are projected on to the other basis is called a hybrid method. A projection method in which all interface conditions are projected on to one basis is called a homogeneous method. For the case of non-conical mountings, Suratteau *et al.* have proved that the numerical solutions resulting from the hybrid methods satisfy the energy balance and reciprocity criteria automatically (independent of the truncation orders) while those resulting from the homogeneous methods do not. It can be shown that the above statement is also true in the case of conical mountings.

In this paper, we adopt the homogeneous method utilizing the projection basis $\{u_m^{+(s)}\}$, so that the energy balance and reciprocity criteria are not automatically satisfied. Multiplying equations (58) by $\bar{u}_m^{+(s)}(x)/\sigma^{(s)}(x)$, and equations (59) by $\bar{u}_m^{+(s)}(x)$, then integrating over a grating period, and making use of the bi-orthogonality relation (34), we have the following matrix equations:

$$WY = UX + UI, \tag{69}$$

$$DY = QX + PI, \tag{70}$$

where

$$X = \begin{pmatrix} \bar{R}^{(e)} \\ \bar{T}^{(e)} \\ \bar{R}^{(h)} \\ \bar{T}^{(h)} \end{pmatrix}, \qquad Y = \begin{pmatrix} \bar{a}^{(e)} \\ \bar{b}^{(e)} \\ \bar{a}^{(h)} \\ \bar{b}^{(h)} \end{pmatrix}, \qquad I = \begin{pmatrix} \bar{I}_z^{(e)}\delta_{0n} \\ 0 \\ \bar{I}_z^{(h)}\delta_{0n} \\ 0 \end{pmatrix}, \tag{71}$$

δ_{0n} is the Kronecker delta, and the rest of the matrices are defined in Appendix B. In (71), each element of the column vectors is itself a column vector, and the elements of X and Y are related to the unknown field amplitudes listed in (60) and (61). The matrix D in (70) is diagonal, so vector $\mathbf{Y}$ can be expressed in terms of $\mathbf{X}$ without numerical matrix inversion. Substituting the expression of $\mathbf{Y}$ into (69), we have

$$(\mathbf{W}\mathbf{D}^{-1}\mathbf{Q} - \mathbf{U})\mathbf{X} = (\mathbf{U} - \mathbf{W}\mathbf{D}^{-1}\mathbf{P})\mathbf{I}. \tag{72}$$

This is the final linear system of equations from which we numerically determine the field amplitudes.

In order to solve the linear system (72) on a computer, we unavoidably have to truncate the matrices. We designate N as the total number of terms retained in Rayleigh expansions (we truncate the Rayleigh expansions symmetrically with respect to the zero diffraction order) and M as that retained in the modal expansions. The integers N and M are called the truncation orders. It is easily seen that for the solution of (72) to be well (neither under- nor over-) specified, the two truncation orders must be the same. Thus, the matrices $\mathbf{W}$, $\mathbf{U}$, $\mathbf{P}$, $\mathbf{Q}$, and $\mathbf{D}$ in (72) are $4N$ by $4N$ square matrices.

4. Numerical examples

In this Section we present some numerical results. The computer program is written in Fortran 77 and double precision is used for real and complex arithmetic. For the special case of non-conical mountings, the program has been checked using published data with good agreement. For the general case, it meets the energy balance and reciprocity criteria with reasonable accuracy. Table 1 tabulates TE and

Table 1. Numerical comparison with the results of Botten *et al.* for a metallic grating in a non-conical mounting. Parameters: $d=1{\cdot}0\,\mu m$, $d_1=0{\cdot}4001\,\mu m$, $h=0{\cdot}1\,\mu m$, $\epsilon^{(1)}=\epsilon^{(2)}=\epsilon_1=1{\cdot}0$, $\epsilon_2=(1{\cdot}5+i1{\cdot}0)^2$, $\lambda_0=0{\cdot}8\,\mu m$, $\theta=11{\cdot}5°$, $\phi=0°$. Truncation orders: Botten *et al.*, $M=20$, $N=51$; this paper, $M=N=31$.

Diffraction Order	Efficiency, TE		Efficiency, TM	
	Botten *et al.*	This Paper	Botten *et al.*	This Paper
R_{-1}	$2{\cdot}8529\times10^{-2}$	$2{\cdot}8526\times10^{-2}$	$1{\cdot}6772\times10^{-2}$	$1{\cdot}6756\times10^{-2}$
R_0	$6{\cdot}2128\times10^{-2}$	$6{\cdot}2124\times10^{-2}$	$9{\cdot}5755\times10^{-2}$	$9{\cdot}5799\times10^{-2}$
R_1	$4{\cdot}6011\times10^{-3}$	$4{\cdot}6011\times10^{-3}$	$1{\cdot}2779\times10^{-3}$	$1{\cdot}2757\times10^{-3}$
T_{-1}	$3{\cdot}8574\times10^{-2}$	$3{\cdot}8573\times10^{-2}$	$2{\cdot}1372\times10^{-2}$	$2{\cdot}1430\times10^{-2}$
T_0	$4{\cdot}6913\times10^{-1}$	$4{\cdot}6913\times10^{-1}$	$3{\cdot}9939\times10^{-1}$	$3{\cdot}9985\times10^{-1}$
T_1	$5{\cdot}4894\times10^{-3}$	$5{\cdot}4895\times10^{-3}$	$3{\cdot}6226\times10^{-3}$	$3{\cdot}6378\times10^{-3}$

Table 2. Diffraction efficiencies (η), polarization angles (α, δ), and diffraction angles (θ, ϕ) of a dielectric grating in a conical mounting. All angular values are in degrees. Parameters: $d=1{\cdot}0\,\mu m$, $d_1=0{\cdot}5\,\mu m$, $h=0{\cdot}5\,\mu m$, $\epsilon^{(1)}=\epsilon_1=1{\cdot}0$, $\epsilon^{(2)}=\epsilon_2=2{\cdot}25$, $\lambda_0=0{\cdot}5\,\mu m$. Incident polarization: $\alpha=45°$, $\delta=90°$. Incident angle: $\theta=\phi=45°$. Truncation orders: $M=N=31$.

Order	η	α	δ	θ	ϕ
R_{-2}	$1{\cdot}6137\times10^{-3}$	64·3182	−30·2984	45·0000	135·0000
R_{-1}	$3{\cdot}8070\times10^{-3}$	65·9715	−157·1961	30·0000	90·0000
R_0	$1{\cdot}8548\times10^{-2}$	70·4908	−148·4611	45·0000	45·0000
T_{-3}	$3{\cdot}3631\times10^{-2}$	51·0569	32·2795	48·1897	153·4349
T_{-2}	$1{\cdot}0343\times10^{-1}$	56·2437	110·2136	28·1255	135·0000
T_{-1}	$3{\cdot}1868\times10^{-1}$	46·5484	99·0295	19·4712	90·0000
T_0	$1{\cdot}4186\times10^{-1}$	34·2601	68·3735	28·1255	45·0000
T_1	$3{\cdot}7827\times10^{-1}$	46·3291	86·8095	48·1897	26·5651

TM diffraction efficiencies of a metallic grating in a non-conical mounting. The data of Botten *et al.* are taken from table 1 of [6]. Note that Botten *et al.* adopted a hybrid projection method that allows unequal truncation orders N and M. Clearly, the agreement is very good, especially for the TE polarization.

To date, there are no numerical data, especially data of the polarization parameters, for non-perfectly conducting lamellar gratings in conical mountings available in the literature; therefore, we present some original data in the rest of this section. Listed in table 2 are diffraction efficiencies (η), polarization angles (α, δ), and diffraction angles (θ, ϕ) of a dielectric grating in a conical mounting (the parameters are listed in the table caption). Note that the incident plane wave is right-hand, circularly polarized. Listed in table 3 are diffraction efficiencies (η), polarization angles (α, δ), and diffraction angles (θ, ϕ) of a highly conducting grating in a conical mounting. The incident plane wave is linearly polarized with equal s and p component amplitudes. In both cases, the truncation orders are chosen to ensure that the accuracy of the data is better than one per cent (see discussion below).

In figure 3 (a, b, c) we show the change of diffraction efficiencies η and polarization angles α and δ of a small period dielectric grating as the incident azimuthal angle ϕ sweeps through the first quadrant. The incident plane wave is

Table 3. Diffraction efficiencies (η), polarization angles (α, δ), and diffraction angles (θ, ϕ) of a metallic grating in a conical mounting. All angular values are in degrees. Parameters: $d=1{\cdot}0\,\mu m$, $d_1=0{\cdot}5\,\mu m$, $h=1{\cdot}0\,\mu m$, $\epsilon^{(1)}=\epsilon_1=1{\cdot}0$, $\epsilon^{(2)}=\epsilon_2=(0{\cdot}1+i5{\cdot}0)^2$, $\lambda_0=0{\cdot}5\,\mu m$. Incident polarization: $\alpha=45°$, $\delta=0°$. Incident angle: $\theta=30°$, $\phi=45°$. Truncation orders: $M=N=51$.

Order	η	α	δ	θ	ϕ
R_{-2}	$7{\cdot}3099\times10^{-2}$	62·4788	52·7402	47·4606	151·3249
R_{-1}	$1{\cdot}3511\times10^{-1}$	15·3476	−12·0484	22·5000	112·5000
R_0	$4{\cdot}2986\times10^{-1}$	41·2528	171·2140	30·0000	45·0000
R_1	$3{\cdot}0238\times10^{-1}$	75·2325	168·7752	67·5000	22·5000

always p polarized and it strikes the grating from the optically denser medium at a polar angle greater than the critical angle (for total internal reflection). This configuration is reminiscent of what occurs in a planar waveguide grating coupler for a TM polarized guided-wave [20]. The reflected and the transmitted negative first orders pass off at about $\phi=37°$ and $\phi=60°$, respectively. It is evident that as soon as ϕ is non-zero, the diffraction orders become elliptically polarized. The two first orders are nearly circularly polarized at $\phi\sim20°$. Also, the senses of polarization of the two first orders remain right-handed throughout the angular range of their existence.

Next, we illustrate the excellent convergence rate of the modal method. For this purpose, we define a measure of error Δ_N as follows

$$\Delta_N=\log_{10}\left|\frac{f_N-f^*}{f^*}\right|, \tag{73}$$

where f_N stands for any one of the physical quantities, such as a diffraction efficiency or a polarization angle, computed with truncation order N, and $f^*=f_{N^*}$, where $N^*>N$ is an integer. If f stands for the sum of the diffraction efficiencies for a lossless grating, $f^*=1{\cdot}0$. Roughly speaking, the negative of Δ_N gives the number of correct significant digits in the numerical results.

Figure 4 shows the convergence of the sum of the diffraction efficiencies and the efficiencies of the negative first orders in reflection and in transmission for a dielectric grating. The physical parameters are the same as those of table 2. The truncation order N varies from 11 to 61 in increment of 2, and $N^*=63$. It is evident from the figure that better than 1% accuracy is achieved as soon as the truncation order is greater than 11. The convergence is not monotonic. The large oscillation in the convergence sequence of the reflected order is probably due to the smallness of the diffraction efficiency (see row 2, column 2 of table 2). If we make a low-order polynomial fit of each set of the data in the figure, the three resulting curves will have more or less similar shapes and close locations. This implies that the energy balance criterion can be used as a good accuracy indicator, thanks to our choice of the homogeneous projection method.

Figure 5 shows the convergence of the diffraction efficiency and diffraction angles α and δ of the negative first order of a metallic grating. The physical parameters are the same as those of table 3. The truncation order N varies from 11 to 69 in increments of 2, and $N^*=71$. Since now the grating is metallic, the convergence in this case is, as expected, slower than that of figure 4. However, better than 1% accuracy can still be achieved with a truncation order of 40 or greater.

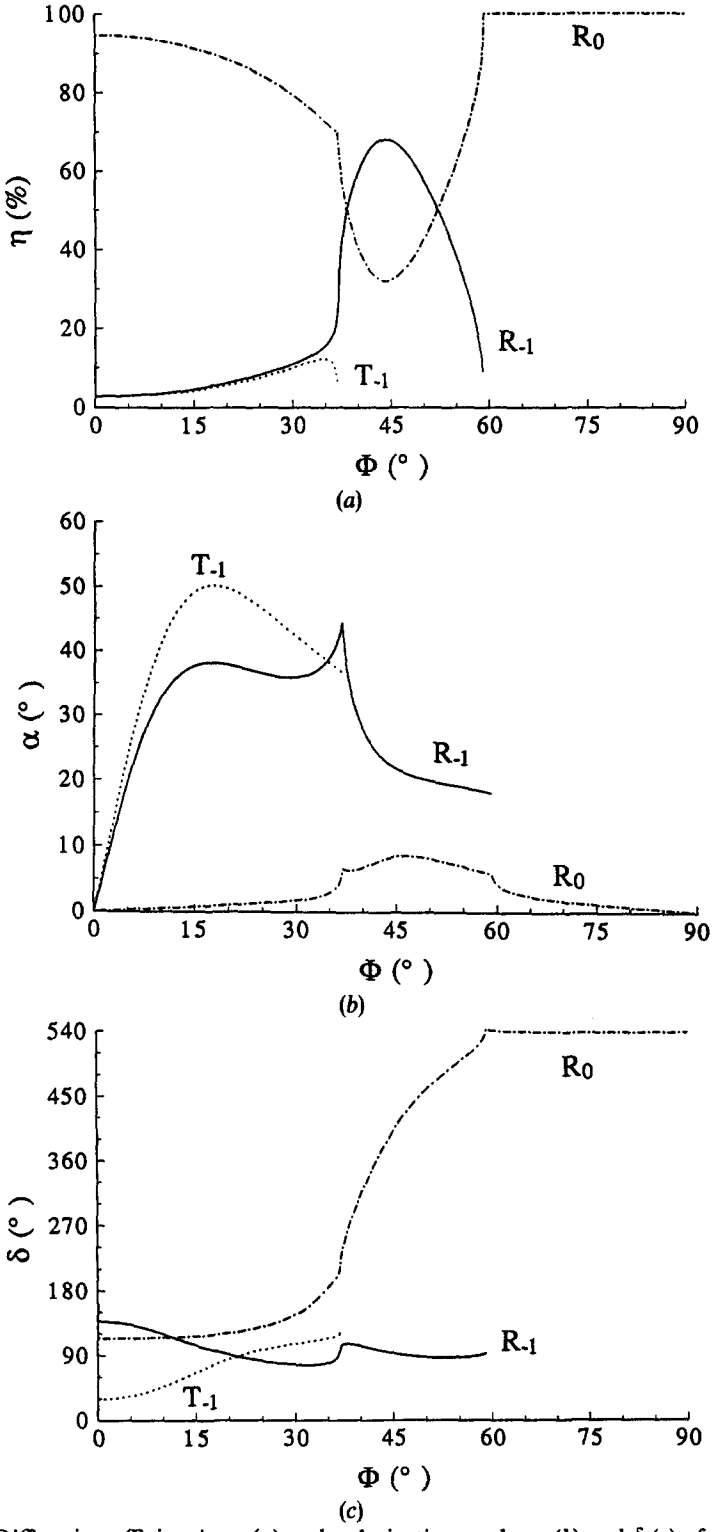

Figure 3. Diffraction efficiencies η (*a*) and polarization angles α (*b*) and δ (*c*) of a dielectric grating versus the incident azimuthal angle ϕ. All angles are measured in degrees. Parameters: $d=0{\cdot}3\,\mu m$, $d_1=0{\cdot}15\,\mu m$, $h=0{\cdot}15\,\mu m$, $\epsilon^{(1)}=\epsilon_1=2{\cdot}25$, $\epsilon^{(2)}=\epsilon_2=1{\cdot}0$, $\lambda_0=0{\cdot}5\,\mu m$. Incident polarization: $\alpha=0°$. Incident angle: $\theta=60°$. Truncation orders: $M=N=45$.

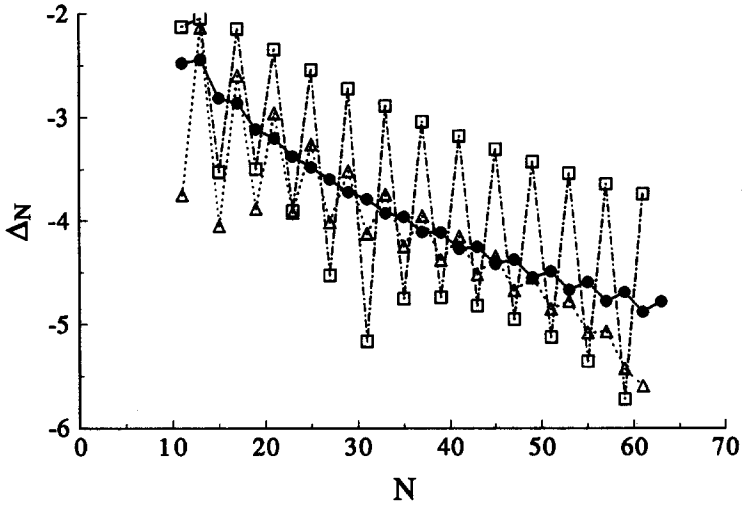

Figure 4. Convergence for a dielectric grating. ●, Sum of all diffraction efficiencies; □, diffraction efficiency of −1 order in reflection; △, diffraction efficiency of −1 order in transmission. The parameters are the same as for table 2.

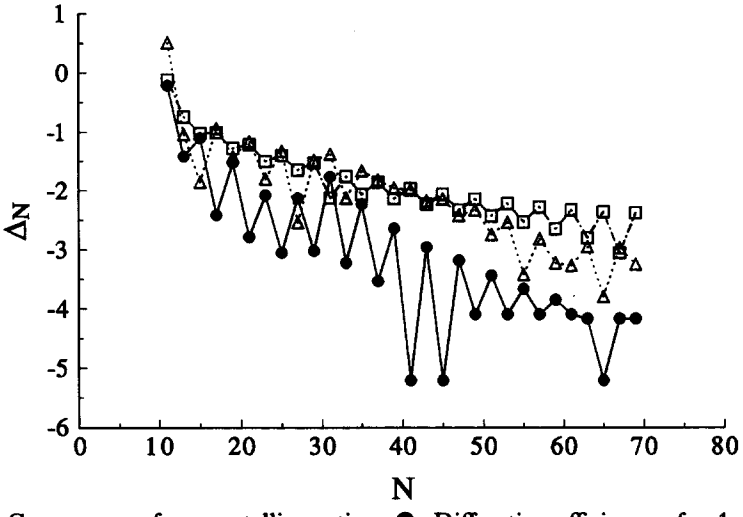

Figure 5. Convergence for a metallic grating. ●, Diffraction efficiency of −1 order; □, polarization angle α; △, polarization angle δ. The parameters are the same as for table 3.

As the authors of [5–8] have shown, one of the unique features of the modal method is that it works very well even for deep, highly conducting gratings. This feature is also true for the extension of the method to conical mounts as developed in this paper. This is understandable, since the fundamental analytical and numerical issues for conical and non-conical mountings are essentially the same. For the sake of saving space, however, we will not provide any numerical evidence here.

5. Summary

In this paper, we have extended the rigorous model method of Botten *et al.* to the case of conical mountings. A crucial step in accomplishing the extension is the field decomposition discussed in Section 2.2. and Appendix A. The field decomposition

reduces the vector-valued boundary-value problem given by (14) and (15) to a scalar one given by (26), thus tremendously simplifying the subsequent analysis and allowing the previous works of Botten *et al.* and Suratteau *et al.* to be used here. The completeness and orthogonality of the modal fields in the corrugated region are carefully established. The computer program implementing the extended model method can treat a plane wave of arbitrary angle of incidence and polarization. It converges very well for highly conducting grating materials and very deep grating grooves. We have included some original data of both diffraction efficiencies and diffraction polarizations for conical diffraction configurations.

The mathematical formulation presented in this paper has been kept general. In fact, the explicit x dependence of $\epsilon(x)$ and $\mu(x)$ given in (1) is not used, except for the derivation of the characteristic equation and the eigenfunctions in Section 2.4. Therefore, the formulation is valid for any periodic $\epsilon(x)$ and $\mu(x)$. However, it is precisely the simple x dependence of $\epsilon(x)$ and $\mu(x)$ in (1) that makes the easy derivations in Section 2.4. possible. For any other permittivity and permeability variations, the solutions for the eigenvalues and eigenfunctions become very complicated, and the modal analysis quickly loses its advantages over other grating methods.

The present analysis can be easily extended to treat gratings of arbitrary groove shapes and waveguide gratings with a number of uniform layers above and below the corrugated region. These extensions will be the subject of a future paper.

Appendices

Appendix A

Theorem

Any solution of (14) and (15) can be decomposed such that

$$\begin{pmatrix} E_z \\ H_z \end{pmatrix} = \begin{pmatrix} E_z^{(e)} \\ H_z^{(e)} \end{pmatrix} + \begin{pmatrix} E_z^{(h)} \\ H_z^{(h)} \end{pmatrix}, \tag{A 1}$$

where $(E_z^{(e)}, H_z^{(e)})^{\mathrm{T}}$ and $(E_z^{(h)}, H_z^{(h)})^{\mathrm{T}}$ are $E_\perp$ and $H_\perp$, respectively, and they satisfy (14) and (15) independently. The decomposition (A 1) is, in general, unique.

Proof

Since there is no explicit y dependence in (14), any solution of (14) is necessarily of the following form

$$\begin{pmatrix} E_z \\ H_z \end{pmatrix} = \sum_\lambda \begin{pmatrix} E_{z\lambda}(x, y) \\ E_{z\lambda}(x, y) \end{pmatrix}, \tag{A 2}$$

where

$$\begin{pmatrix} E_{z\lambda}(x, y) \\ H_{z\lambda}(x, y) \end{pmatrix} = \exp(\mathrm{i}\lambda y) \begin{pmatrix} \xi_\lambda(x) \\ \eta_\lambda(x) \end{pmatrix}. \tag{A 3}$$

Substituting (A 3) into (14), and eliminating the y-dependence, we have

$$\begin{pmatrix} \dfrac{\tilde{k}^2}{\epsilon}\dfrac{\mathrm{d}}{\mathrm{d}x}\left(\dfrac{\epsilon}{\tilde{k}^2}\dfrac{\mathrm{d}}{\mathrm{d}x}\right) + \tilde{k}^2 - \lambda^2 & +\mathrm{i}\lambda\dfrac{k_z}{k_0}\dfrac{\tilde{k}^2}{\epsilon}\left(\dfrac{\mathrm{d}}{\mathrm{d}x}\dfrac{1}{\tilde{k}^2}\right) \\ -\mathrm{i}\lambda\dfrac{k_z}{k_0}\dfrac{\tilde{k}^2}{\mu}\left(\dfrac{\mathrm{d}}{\mathrm{d}x}\dfrac{1}{\tilde{k}^2}\right) & \dfrac{\tilde{k}^2}{\mu}\dfrac{\mathrm{d}}{\mathrm{d}x}\left(\dfrac{\mu}{\tilde{k}^2}\dfrac{\mathrm{d}}{\mathrm{d}x}\right) + \tilde{k}^2 - \lambda^2 \end{pmatrix} \begin{pmatrix} \xi_\lambda(x) \\ \eta_\lambda(x) \end{pmatrix} = 0. \tag{A 4}$$

Equation (A 4) is a linear, two-dimensional vector-valued, second-order ordinary differential equation. It, therefore, has four linearly independent solutions. Suppose $(E_{z\lambda}, H_{z\lambda})^{\mathrm{T}}$ is $E_\perp$. Then, from (16 *a*)

$$\eta_\lambda(x) = -\frac{1}{\mathrm{i}\lambda}\frac{1}{\mu}\frac{k_z}{k_0}\frac{\mathrm{d}}{\mathrm{d}x}\xi_\lambda(x). \tag{A 5}$$

Substitution of (A 5) into (A 4) results in two second-order equations for $\xi_\lambda(x)$. These two equations are not independent; one can be derived from the other. Thus, we have two linearly independent solutions of (A 4) that satisfy the $E_\perp$ condition (16 *a*). Similarly, if we demand $(E_{z\lambda}, H_{z\lambda})^{\mathrm{T}}$ to be $H_\perp$, then

$$\xi_\lambda(x) = \frac{1}{\mathrm{i}\lambda}\frac{1}{\epsilon}\frac{k_z}{k_0}\frac{\mathrm{d}}{\mathrm{d}x}\eta_\lambda(x), \tag{A 6}$$

and we have two linearly independent solutions of (A 4) that satisfy the $H_\perp$ condition (16 *b*). Clearly, the solutions of $E_\perp$ type and $H_\perp$ type are linearly independent. Therefore, any solution of (A 4) has the following form

$$\begin{pmatrix}\xi_\lambda(x)\\ \eta_\lambda(x)\end{pmatrix} = c_{\lambda 1}^{(e)}\begin{pmatrix}u_{\lambda 1}^{(e)}\\ v_{\lambda 1}^{(e)}\end{pmatrix} + c_{\lambda 2}^{(e)}\begin{pmatrix}u_{\lambda 2}^{(e)}\\ v_{\lambda 2}^{(e)}\end{pmatrix} + c_{\lambda 1}^{(h)}\begin{pmatrix}v_{\lambda 1}^{(h)}\\ u_{\lambda 1}^{(h)}\end{pmatrix} + c_{\lambda 2}^{(h)}\begin{pmatrix}v_{\lambda 2}^{(h)}\\ u_{\lambda 2}^{(h)}\end{pmatrix}, \tag{A 7}$$

where $c_{\lambda l}^{(s)}$, $l=1,2, s=e, h$ are constants, $(u_{\lambda l}^{(e)}, v_{\lambda l}^{(e)})^{\mathrm{T}}$ and $(v_{\lambda l}^{(h)}, u_{\lambda l}^{(h)})^{\mathrm{T}}$ are $E_\perp$ and $H_\perp$ solutions of (A 4) respectively. So, the decomposition (A 1) is always possible.

Next, we impose the pseudo-periodic conditions (15) on the general solution (A 7). This leads to the following characteristic equation

$$\begin{vmatrix} U_{\lambda 1}^{(e)} & U_{\lambda 2}^{(e)} & V_{\lambda 1}^{(h)} & V_{\lambda 2}^{(h)} \\ U_{\lambda 1}^{(e)'} & U_{\lambda 2}^{(e)'} & V_{\lambda 1}^{(h)'} & V_{\lambda 2}^{(h)'} \\ V_{\lambda 1}^{(e)} & V_{\lambda 2}^{(e)} & U_{\lambda 1}^{(h)} & U_{\lambda 2}^{(h)} \\ V_{\lambda 1}^{(e)'} & V_{\lambda 2}^{(e)'} & U_{\lambda 1}^{(h)'} & U_{\lambda 2}^{(h)'} \end{vmatrix} = 0, \tag{A 8}$$

where

$$\begin{aligned} U_{\lambda l}^{(s)} &= u_{\lambda l}^{(s)}(d/2) - \exp(\mathrm{i}\alpha_0 d)\, u_{\lambda l}^{(s)}(-d/2), \\ V_{\lambda l}^{(s)} &= v_{\lambda l}^{(s)}(d/2) - \exp(\mathrm{i}\alpha_0 d)\, v_{\lambda l}^{(s)}(-d/2), \\ U_{\lambda l}^{(s)'} &= u_{\lambda l}^{(s)'}(d/2) - \exp(\mathrm{i}\alpha_0 d)\, u_{\lambda l}^{(s)'}(-d/2), \\ V_{\lambda l}^{(s)'} &= v_{\lambda l}^{(s)'}(d/2) - \exp(\mathrm{i}\alpha_0 d)\, v_{\lambda l}^{(s)'}(-d/2). \end{aligned} \tag{A 9}$$

By elementary row manipulation and making use of (A 5) and (A 6) for the $E_\perp$ and $H_\perp$ solutions, it can be shown that the two off-diagonal two-by-two-matrices in (A 8) can be made zero. Therefore, the $E_\perp$ and $H_\perp$ solutions of (A 4), and hence those of (14), satisfy the pseudo-periodic conditions (15) independently.

Suppose the decomposition (A 1) is not unique. Then we may have a decomposition of the zero field into two non-zero orthogonal modal fields. Furthermore, each of these fields satisfies both the $E_\perp$ and the $H_\perp$ conditions simultaneously. However, this leads to $k_z^2 + \lambda^2 = 0$, a condition which is, in general, not true. This completes the proof of the theorem.

Appendix B

As in the main text of the paper, $j=1, 2$, and $s=e, h$. In (58) and (59)

$$\tilde{I}_z^{(s)} = I_z^{(s)} \exp\left[-\mathrm{i}\beta_0^{(1)} h/2\right]. \tag{B 1}$$

$$A_m^{(s)} = \frac{k_z^2 + \lambda_m^{(s)2}}{\lambda_m^{(s)}} \tan\frac{\lambda_m^{(s)} h}{2}, \qquad B_m^{(s)} = -\frac{k_z^2 + \lambda_m^{(s)2}}{\lambda_m^{(s)}} \operatorname{ctan}\frac{\lambda_m^{(s)} h}{2}, \tag{B 2}$$

$$\tau_1^{(j)} = \mathrm{i}k_0^2 \frac{\epsilon^{(j)}}{\tilde{k}^{(j)2}}, \qquad \tau_2^{(j)} = \mathrm{i}k_0^2 \frac{\mu^{(j)}}{\tilde{k}^{(j)2}}, \qquad \tau_3^{(j)} = \mathrm{i}k_0 \frac{k_z}{\tilde{k}^{(j)2}}. \tag{B 3}$$

For the sake of clarity, the matrices in (69) and (70) are expressed in block forms, followed by the definitions of each sub-matrix.

$$\mathbf{W} = \begin{pmatrix} 1 & 1 & W^{(e)}A^{(h)} & W^{(e)}B^{(h)} \\ 1 & -1 & -W^{(e)}A^{(h)} & W^{(e)}B^{(h)} \\ W^{(h)}A^{(e)} & W^{(h)}B^{(e)} & 1 & 1 \\ -W^{(h)}A^{(e)} & W^{(h)}B^{(e)} & 1 & -1 \end{pmatrix}. \tag{B 4}$$

$$\mathbf{U} = \begin{pmatrix} U^{(e)} & 0 & 0 & 0 \\ 0 & U^{(e)} & 0 & 0 \\ 0 & 0 & U^{(h)} & 0 \\ 0 & 0 & 0 & U^{(h)} \end{pmatrix}. \tag{B 5}$$

$$\mathbf{D} = 2\begin{pmatrix} A^{(e)} & 0 & 0 & 0 \\ 0 & B^{(e)} & 0 & 0 \\ 0 & 0 & A^{(h)} & 0 \\ 0 & 0 & 0 & B^{(h)} \end{pmatrix}. \tag{B 6}$$

$$\mathbf{Q} = \begin{pmatrix} \tilde{U}^{(e)} & 0 & 0 & 0 \\ 0 & U^{(e)} & 0 & 0 \\ 0 & 0 & U^{(h)} & 0 \\ 0 & 0 & 0 & \tilde{U}^{(h)} \end{pmatrix} \begin{pmatrix} -\tau_1^{(1)}\beta^{(1)} & -\tau_1^{(2)}\beta^{(2)} & \tau_3^{(1)}\alpha & -\tau_3^{(2)}\alpha \\ -\tau_1^{(1)}\beta^{(1)} & \tau_1^{(2)}\beta^{(2)} & \tau_3^{(1)}\alpha & \tau_3^{(2)}\alpha \\ -\tau_3^{(1)}\alpha & \tau_3^{(2)}\alpha & -\tau_2^{(1)}\beta^{(1)} & -\tau_2^{(2)}\beta^{(2)} \\ -\tau_3^{(1)}\alpha & -\tau_3^{(2)}\alpha & -\tau_2^{(1)}\beta^{(1)} & \tau_2^{(2)}\beta^{(2)} \end{pmatrix}. \tag{B 7}$$

$$A_{mn}^{(s)} = A_m^{(s)}\delta_{mn}, \qquad B_{mn}^{(s)} = B_m^{(s)}\delta_{mn}. \tag{B 8}$$

$$\alpha_{mn} = \alpha_n \delta_{mn}, \qquad \beta_{mn}^{(j)} = \beta_n^{(j)}\delta_{mn}. \tag{B 9}$$

$$U_{mn}^{(s)} = (e_n, u_m^{+(s)})_s, \qquad \tilde{U}_{mn}^{(s)} = \langle e_n, u_m^{+(s)} \rangle. \tag{B 10}$$

$$W_{mn}^{(e)} = (w_n^{(h)}, u_m^{+(e)})_e, \qquad W_{mn}^{(h)} = (w_n^{(e)}, u_m^{+(h)})_h. \tag{B 11}$$

To get the expression for matrix P, we only need to reverse the sign of $\beta^{(j)}$ in (B 7). In (B 10) and (B 11) the inner products are those defined in (28) and (37), and the subscript s indicates that the weight function $\sigma(s)$ should be used in the integral. By using (55) and (47), it can be shown that the two sub-matrices in (B 11) are related such that

$$W_{mn}^{(e)} + W_{nm}^{(h)} = 0. \tag{B 12}$$

Hence, there are five sub-matrices involving the eigenfunctions to be calculated. It can be shown that these sub-matrix elements can be expressed in terms of the left and

right limits of $u_m^{(s)}$ and its derivative at $\pm d_1/2$, if the periodic medium is characterized by (1).

Since $w_m^{(s)}$ and $\tau_3^{(j)}$ are proportional to k_z, the matrices defined in (B 4)–(B 7) become block-diagonal when $k_z=0$. Of course, this means that in non-conical mountings, E_z and H_z are de-coupled.

Acknowledgment

This work is supported by the Optical Data Storage Center at the University of Arizona.

References

[1] COLLIN, R. E., 1956, *Can. J. Phys.*, **34**, 398.
[2] BURCKHARDT, C. B., 1966, *J. opt. Soc. Am.*, **56**, 1502.
[3] KNOP, K., 1978, *J. opt. Soc. Am.*, **68**, 1206.
[4] PENG, S. T., TAMIR, T., and BERTONI, H. L., 1975, *IEEE Trans. microw. Theory Tech.*, **23**, 123.
[5] BOTTEN, L. C., CRAIG, M. S., MCPHEDRAN, R. C., ADAMS, J. L., and ANDREWARTHA, J. R., 1981, *Optica Acta*, **28**, 413.
[6] BOTTEN, L. C., CRAIG, M. S., MCPHEDRAN, R. C., ADAMS, J. L., and ANDREWARTHA, J. R., 1981, *Optica Acta*, **28**, 1087.
[7] BOTTEN, L. C., CRAIG, M. S., and MCPHEDRAN, R. C., 1981, *Optica Acta*, **28**, 1103.
[8] SURATTEAU, J. Y., CADILHAC, M., and PETIT, R., 1983, *J. Optics (Paris)*, **14**, 273.
[9] TAYEB, G., and PETIT, R., 1984, *Optica Acta*, **31**, 1361.
[10] SHENG, P., STEPLEMAN, R. S., and SANDA, P. N., 1982, *Phys. Rev.* B, **26**, 2907.
[11] HAGGANS, C. W., LI, L., FUJITA, T., and KOSTUK, R. K., 1993, *J. mod. Optics*, **40**, 675.
[12] PENG, S. T., 1989, *J. opt. Soc. Am.* A, **6**, 1869.
[13] PETIT, R., 1980, *Electromagnetic Theory of Gratings*, edited by R. Petit (Berlin: Springer-Verlag).
[14] COLE, R. H., 1968, *Theory of Ordinary Differential Equations* (New York: Appleton-Century-Crofts).
[15] NEVIERE, M., 1980, *Electromagnetic Theory of Gratings*, edited by R. Petit (Berlin: Springer-Verlag).
[16] NAIMARK, M. A., 1967, *Linear Differential Operators, Part I* (New York: Frederick Ungar).
[17] See, for example, BORN, M., and WOLF, E., 1989, *Principles of Optics*, sixth edition (Oxford: Pergamon).
[18] BOTTEN, L. C., CRAIG, M. S., and MCPHEDRAN, R. C., 1983, *Computer Phys. Commun.*, **29**, 245.
[19] HARRINGTON, R. F., 1968, *Field Computation by Moment Methods* (New York: Macmillan).
[20] LI, L., GONG, Q., LAWRENCE, G. N., and BURKE, J. J., 1992, *Appl. Optics*, **31**, 4190.