A COMBINATION OF UP- AND DOWN-GOING FLOQUET MODAL FUNCTIONS USED TO DESCRIBE THE FIELD INSIDE GROOVES OF A DEEP GRATING

M. Ohtsu, Y. Okuno, A. Matsushima, and T. Suyama

Graduate School of Science and Technology
Kumamoto University
2-39-1 Kurokami Kumamoto, 860-8555 Japan

Abstract—An effective computational method based on a conventional modal-expansion approach is presented for solving the problem of diffraction by a deep grating. The groove depth can be the same as or a little more than the grating period. The material can be a perfect conductor, a dielectric, or a metal. The method is based on Yasuura's modal expansion, which is known as a least-squares boundary residual method or a modified Rayleigh method. The feature of the present method is that: (1) The semi-infinite region U over the grating surface is divided into an upper half plane U_0 and a groove region U_G by a fictitious boundary (a horizontal line); (2) The latter is further divided into shallow horizontal layers U_1, U_2, $\cdots$, U_Q again by fictitious boundaries; (3) An approximate solution in U_0 is defined in a usual manner, i.e., a finite summation of up-going Floquet modal functions with unknown coefficients, while the solutions in U_q ($q = 1, 2, \cdots, Q$) include not only the up-going but also the down-going modal functions; (4) If the grating is made of a dielectric or a metal, the semi-infinite region L below the surface is partitioned similarly into L_0, L_1, $\cdots$, L_Q, and approximate solutions are defined in each region; (5) A huge-sized least squares problem that appears in finding the modal coefficients is solved by the QR decomposition accompanied by sequential accumulation. The method of solution for a grating made of a perfect conductor is described in the text. The method for dielectric gratings can be found in an appendix. Numerical examples include the results for perfectly conducting and dielectric gratings.

1. INTRODUCTION

Yasuura's method [1–4] is one of the numerical techniques for solving the problem of diffraction by a grating. Although alternative terminology for the method (e.g., a least-squares boundary residual method [5] or a modified Rayleigh method [6]) exists, we employ the name throughout this paper. The method is a standard technique having reliability for a numerical solution to the boundary value problems of the Helmholtz equation. It is an accepted knowledge, however, that the conventional Yasuura's method (CYM) [1–4] with Floquet modes as basis functions does not have a wide range of application [6, 7].

Let us assume that we employ the CYM (see Appendix A) in solving the problem of plane-wave diffraction from a perfectly-conducting sinusoidal grating having a period D and a depth $2H$. The period is comparable to the wavelength as usual. For an E-wave (s polarization) problem where $2H/D = 0.5$, we can employ 45 Floquet modal functions in the semi-infinite region over the grating surface and obtain a solution with 0.1 percent error in both energy conservation and boundary condition. Employment of additional modal functions easily causes numerical trouble in making least-squares boundary matching. Hence, a practical limit in $2H/D$ in the E-wave case is 0.5 so long as we use conventional double-precision arithmetic. Similarly, the limit in an H-wave (p polarization) case seems to be a little less than 0.4.

Yasuura's method has been equipped with a smoothing procedure (SP) [8, 9] in order that the method be capable of solving a wider class of problems. It has been shown that: in the above problem, we can obtain a solution with 0.1 percent error using $17-41$ modal functions (the number depends on the order of the smoothing procedure [8, 9] and on the polarization). Although Yasuura's method with the smoothing procedure solves most of the problems for commonly used gratings, the limit in $2H/D$ has as yet been scarcely dealt with. This is because there still is a limit at about $2H/D = 0.7$ and because the SP is not convenient to employ due to difficult theory and increase of computational complexity.

In the following sections we present an alternative way of extending the range of application: First, we partition the semi-infinite region over the grating surface into a half plane over the grating and a couple of stratified horizontal layers in the groove region. Second, we define approximate solutions in each region in the form of truncated modal expansions with unknown coefficients. The solution in the half plane is a commonly employed Floquet modal expansion, while one in a layer is a sum of two finite expansions: one is up-going and another
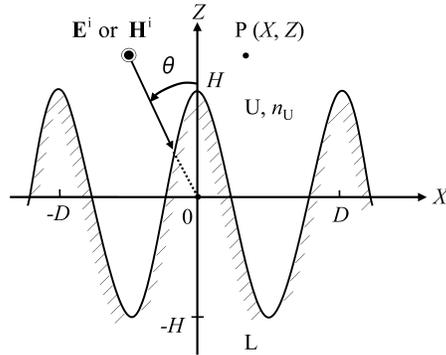
Figure 1. Cross section of a perfectly conducting grating.

is down-going. Then, applying the boundary conditions, we find the modal coefficients that minimize a mean-squares boundary residual.

Although we present, for simplicity, the method assuming an E-wave problem for a perfectly conducting grating, we include the results for an H-wave problem and for a dielectric grating in the numerical examples section. The methods for these types of problems are concisely described in Appendices B and D.

2. STATEMENT OF THE PROBLEM

The cross section of the grating made of a perfectly conducting metal and a coordinate system are shown in Fig. 1. We assume that the grating is periodic in X and uniform in Y. The surface profile of the grating is assumed to be

$$z = f(x) = H \cos \frac{2\pi x}{D} \qquad (1)$$

where H and D are the half depth (amplitude) and the period of the grating. The profile of the grating is a boundary between two regions U $(Z > f(X))$ and L $(Z < f(X))$. The region U is filled with a dielectric medium having a relative refractive index n_U, and the region L is occupied by a perfect conductor. Note that the pair of capital letters (X, Z) represents a point in U; and lower-case letters (x, z) denote a point on the boundary.

We consider the problem to seek the diffracted waves in U for the case of an E-wave incidence. Statement for the case of an H-wave incidence may be found in Appendix B together with the method of

solution. We write the incident electric field as

$$\mathbf{E}^{\mathrm{i}}(\mathbf{P}) = \mathbf{i}_Y \Phi^{\mathrm{i}}(\mathbf{P}) \tag{2}$$

where

$$\Phi^{\mathrm{i}}(\mathbf{P}) = \exp[i(\alpha_0 X - \beta_0^{\mathrm{U}} Z)] \quad (\mathbf{P} \in \mathrm{U}) \tag{3}$$

In (2) and (3) $\mathbf{P} = (X, Z)$ is a position vector of a point P in U, $\mathbf{i}_Y$ is a unit vector in the Y direction, $\alpha_0 = n_{\mathrm{U}} k \sin\theta$, $\beta_0^{\mathrm{U}} = n_{\mathrm{U}} k \cos\theta$, $k = \omega\sqrt{\varepsilon_0\mu_0} = 2\pi/\lambda$ and θ is the angle of incidence shown in Fig. 1. The time dependence $\exp(-i\omega t)$ is suppressed throughout.

We denote by $\Phi^{\mathrm{d}}(\mathbf{P})$ the Y component of the diffracted electric field in U, which satisfies the following conditions:

1. The 2-D Helmholtz equation;

2. A periodicity condition that $\Phi^{\mathrm{d}}(X + D, Z) = e^{i\alpha_0 D}\Phi^{\mathrm{d}}(X, Z)$;

3. A radiation condition in the Z direction that $\Phi^{\mathrm{d}}(X, Z)$ propagates or attenuates in the positive Z direction; and

4. A boundary condition that the tangential component of the total electric field vanishes on the boundary:

$$\left. (\Phi^{\mathrm{i}} + \Phi^{\mathrm{d}}) \right|_{z=f(x)+0} = 0 \tag{4}$$

3. METHOD OF ANALYSIS

As shown in Fig. 2, we first divide the region U over the grating surface into a half plane U_0 over the grating $(Z > H)$ and the groove region U_{G} $(f(X) < Z < H)$ below the half plane U_0. U_0 and U_{G} are sub regions of U having a common border at $Z = H$. Next, we partition the groove region U_{G} into shallow horizonal layers $\mathrm{U}_1, \mathrm{U}_2, \cdots, \mathrm{U}_Q$, where Q is the number of the layers. The boundary between U_q and U_{q+1} is a horizontal line given by (see Fig. 2)

$$Z = z_q = (1 - 2q/Q)H \qquad (q = 1, 2, \cdots, Q - 1) \tag{5}$$

Consequently, we have $Q + 1$ sub-regions in one period $(0 < X < D)$.

An approximation for the diffracted electric field $\Phi^{\mathrm{d}}(\mathbf{P})$ in U_0 is defined in terms of a finite Floquet modal expansion. Here, the Floquet modal functions are separated solutions of Helmholtz's equation satisfying the radiation and periodicity condition, and are given by

$$\phi_m^+(X, Z) = \exp[i(\alpha_m X + \beta_m^{\mathrm{U}} Z)] \quad (m = 0, \pm 1, \cdots) \tag{6}$$
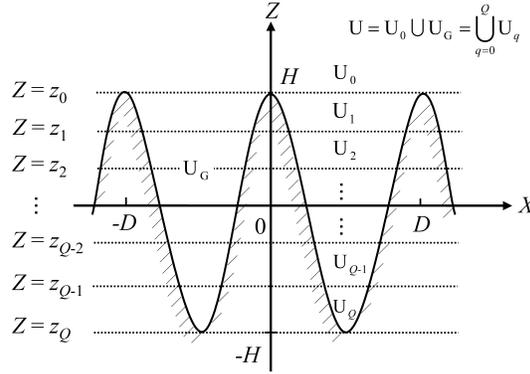
Figure 2. Partition of the semi-infinite region U.

with

$$\alpha_m = \alpha_0 + \frac{2m\pi}{D}, \quad \beta_m^{\mathrm{U}} = (n_{\mathrm{U}}^2 k^2 - \alpha_m^2)^{1/2} \quad (\mathrm{Im}\ \beta_m^{\mathrm{U}} \geq 0) \qquad (7)$$

In (6) the superscript $+$ means that the modal function ϕ_m^+ propagates or attenuates in the positive Z direction (up-going wave). Thus the approximate solution in U_0 takes the form that

$$\Phi_{0N}^{\mathrm{d}}(\mathbf{P}) \;\; = \;\; \sum_{m=-N}^{N} A_m^{(0)} \phi_m^+(X, Z - H) \;\; (\mathbf{P} \in \mathrm{U}_0) \qquad (8)$$

where N is a number of truncation. Note that: although the modal coefficients $A_m^{(0)}$ are functions of the truncation number and should be represented as $A_m^{(0)}(N)$, we omit (N) for simplicity.

In defining an approximate solution in the thin layers U_q ($q = 1, 2, \cdots, Q$), we employ not only the up-going but also down-going Floquet modal functions:

$$\phi_m^-(X, Z) = \exp[i(\alpha_m X - \beta_m^{\mathrm{U}} Z)] \quad (m = 0, \pm 1, \cdots) \qquad (9)$$

where $-$ denote down-going waves. Approximate solutions in the sub-regions U_q, hence, are commonly represented as

$$\Phi_{qN}^{\mathrm{d}}(\mathbf{P}) = \sum_{m=-N}^{N} \left\{ A_m^{(q)} \phi_m^+(X, Z - z_q) + B_m^{(q)} \phi_m^-(X, Z - z_{q-1}) \right\}$$

$$(\mathbf{P} \in \mathrm{U}_q\ ;\ q = 1, 2, \cdots, Q) \qquad (10)$$

Thus we have defined the approximate solutions in U_0 and U_q ($q = 1, 2, \cdots, Q$). The total number of unknown coefficients is $N_T = (2N+1)(2Q+1)$: $(2N+1)$ for U_0 and $2Q(2N+1)$ for U_q's.

The solutions in (8) and (10) satisfy the Helmholtz equation and the periodicity condition. In addition, (8) satisfies the radiation condition. The unknown coefficients, hence, should be determined in order that the solutions meet the boundary condition (4) and an additional set of boundary conditions on the fictitious boundaries:

$$\left.(\Phi^i \delta_{q0} + \Phi_q^d)\right|_{z=z_q+0} = \left.\Phi_{q+1}^d\right|_{z=z_q-0} \qquad (q = 0, 1, ..., Q) \quad (11)$$

$$\left.\frac{\partial(\Phi^i \delta_{q0} + \Phi_q^d)}{\partial\nu}\right|_{z=z_q+0} = \left.\frac{\partial\Phi_{q+1}^d}{\partial\nu}\right|_{z=z_q-0} \qquad (q = 0, 1, ..., Q) \quad (12)$$

where $\partial/\partial\nu$ $(= \partial/\partial Z|_{z=z_q})$ represents the normal derivative on the fictitious boundary and δ_{q0} is Kronecker's delta.

We employ the least-squares method to find the $A_m^{(q)}$ and $B_m^{(q)}$ coefficients. To do this, we define a mean-square error

$$
\begin{aligned}
I_N &= \sum_{q=0}^{Q-1} \left\{ \int_{\Gamma_q} \left| \left[\Phi^i \delta_{q0} + \Phi_{qN}^d - \Phi_{q+1N}^d\right] (x, z(x)) \right|^2 dx \right.\\
&\quad + W^2 \left. \int_{\Gamma_q} \left| \left[\frac{\partial(\Phi^i \delta_{q0} + \Phi_{qN}^d)}{\partial\nu} - \frac{\partial\Phi_{q+1N}^d}{\partial\nu}\right] (x, z(x)) \right|^2 dx \right\} \\
&\quad + \sum_{q=1}^{Q} \left\{ \int_{C_{0q}} \left|\Phi_{qN}^d(x, z(x))\right|^2 dx \right\} \qquad\qquad (13)
\end{aligned}
$$

where W is an intrinsic impedance of vacuum and Γ_q and C_{0q} denote fictitious boundaries and segments of C_0 that are the boundaries of U_q (Fig. 3). The method for solving the least-squares problem on a computer, i.e., the method of discretization and the method of solution for the discretized problem is described in Appendix C.

4. NUMERICAL RESULTS

4.1. Preparation

The mth order reflection efficiency for a propagating order m ($\beta_m^U > 0$) is given by:

$$\rho_m = \frac{\beta_m^U}{\beta_0^U} |A_m^{(0)}|^2 \qquad\qquad (14)$$
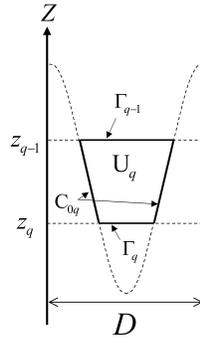
Figure 3. A layer U_q with boundaries Γ_{q-1}, Γ_q and C_{0q}.

For the case of a perfectly conducting grating, the energy error is calculated by:

$$\varepsilon_N = 1 - \sum_m{}' \rho_m \qquad (15)$$

where $\sum'$ stands for the summation over the propagating orders. The energy error ε_N and the norm error on the boundary I_N (see Appendix C for details) will be used in checking the accuracy of solutions.

4.2. Convergence of Solution

We here check the convergence of the results obtained by the present method. Figure 4 shows the variation of the energy error ε_N and the norm error I_N as functions of the number of truncation N at $2H/D = 1.0$ and $Q = 20$ for both E-wave (Fig. 4(a)) and H-wave (Fig. 4(b)) incidence. Other parameters are given in the figure caption. As we observe in these figures, I_N decreases as N increases. An approximate solution with 0.1 percent energy error is obtained at $N = 14$ for an E-wave. In an H-wave case convergence of solutions is not so fast as in the E-wave case. We obtain a solution with 1.0 percent energy error at $N = 23$ in that case of polarization.

4.3. Comparison with Other Methods

We examine our results for a perfectly conducting grating by comparing them with existing results obtained by an integral-equation method [10, 11]. We then show a couple of results for a dielectric grating and compare them with the results obtained by the Rigorous Coupled Wave Analysis (RCWA) method [12].

(a) E-wave   (b) H-wave
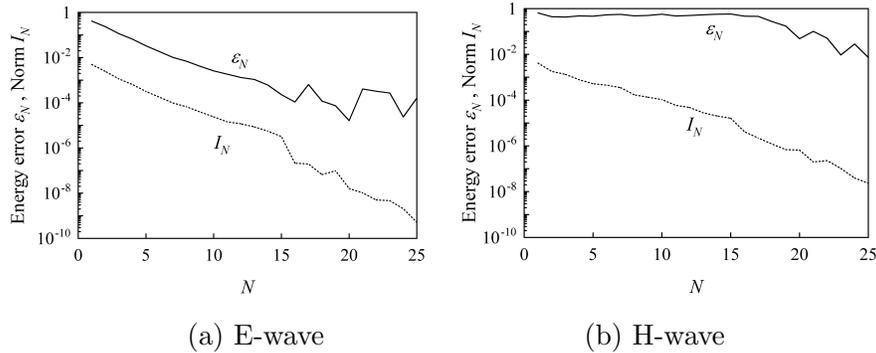
Figure 4. Decrease of the energy error ε_N and the norm I_N as functions of the number of truncation. $Q = 20$, $n_U = 1$, $2H/D = 1.0$, $D = 0.5\lambda$, and $\theta = 30°$.

(a) 0 and -1 order   (b) -2 and -3 order
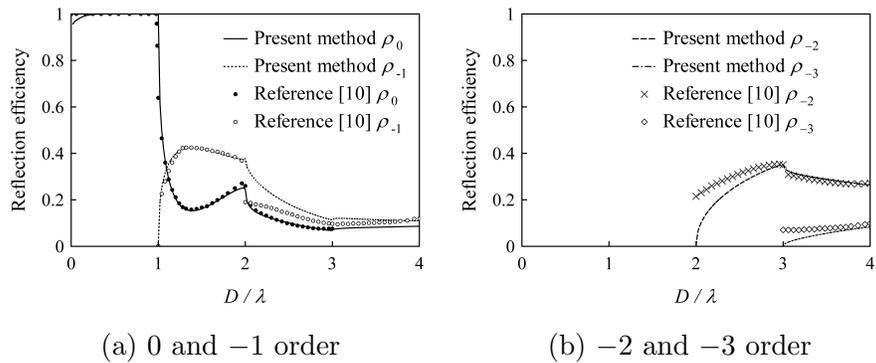
Figure 5. Diffraction efficiency as functions of D/λ (E-wave). $N = 25$, $Q = 25$, $n_U = 1$, $H = 0.25\lambda$, and $\theta = 0°$.

4.3.1. Perfectly Conducting Grating: E-wave Case

Figure 5 shows the comparison of reflection efficiency as a function of D/λ, the grating period normalized by the wavelength. E-wave incidence is assumed and the amplitude of the surface modulation is given by $H/\lambda = 0.25$. The curves and the symbols, respectively, represent the present results and the results in Reference [10]. We observe good agreement in zeroth order efficiency and notice considerable discrepancy in non-zero orders. We believe that the present results are accurate because they have been obtained by Yasuura's method that has proof of convergence and because their error is less than 0.01 percent in the range $D/\lambda > 0.5$ in both energy
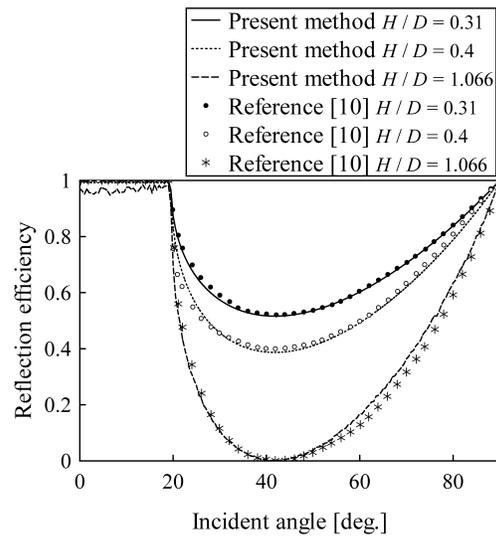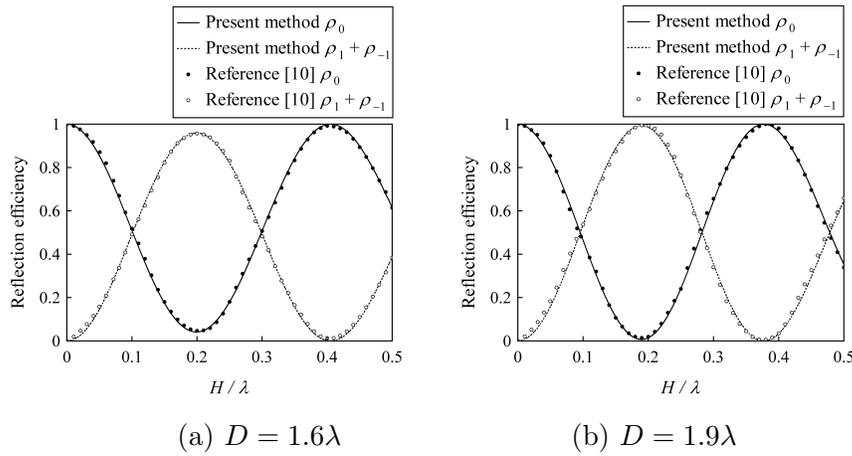
Figure 6. Diffraction efficiency as functions of θ (E-wave). $n_U = 1$ and $D = 0.75\lambda$. $H/D = 0.31$, $(N,Q) = (15,4)$; $H/D = 0.4$, $(15,5)$; and $H/D = 1.066$, $(30,20)$.

conservation and boundary condition. The difference, hence, may come from inadequate computational technology at the time when Reference [10] was published.

Figure 6 shows comparison of reflection efficiency as functions of the incident angle for E-wave incidence and $D = 0.75\lambda$. The curves and the symbols, again, represent the present results and the results in [10]. We find good agreement between the results.

Figure 7 shows comparison of reflection efficiency as functions of H/λ, the surface amplitude normalized by the wavelength. The grating period is a parameter, which is given by $D = 1.6\lambda$ in Fig. 7(a) and $D = 1.9\lambda$ in Fig. 7(b). The curves and the circles denote the present results and the results from [10]. Since the depth-to-period ratio $2H/D$ in these figures is relatively small, i.e., $0 < 2H/D < 0.625$ in Fig. 7(a) and $0 < 2H/D < 0.526$ in Fig. 7(b), we have set $N = 25$ and $Q = 5$ to obtain solutions with 0.1 percent energy error.

In Figs. 8–10 we show the results for relatively deep gratings. The curves and the circles denote the present results and the results from [10]. We choose Q and N given in Table 1 to keep the accuracy of solutions.

(a) $D = 1.6\lambda$ (b) $D = 1.9\lambda$

Figure 7. Diffraction efficiency as functions of H/λ (E-wave). $N = 25$, $Q = 5$, $n_{\mathrm{U}} = 1$, and $\theta = 0°$.
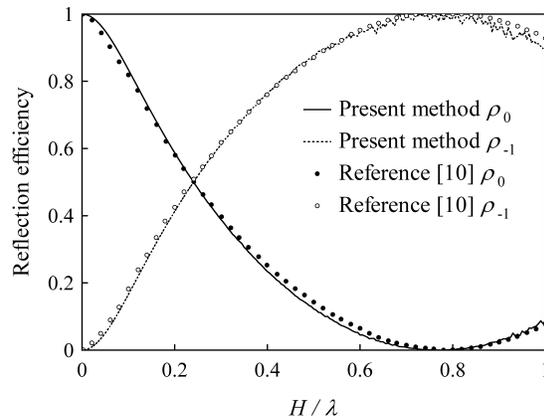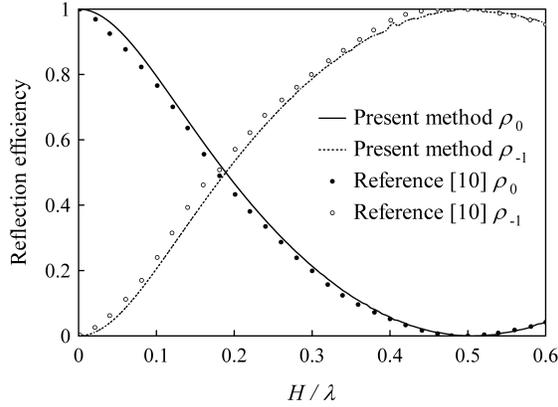

Figure 8. Diffraction efficiency as functions of H/λ (E-wave). $n_{\mathrm{U}} = 1$, $D = 0.75\lambda$, $\theta = 41.8°$, and $N = 20$. $Q = 10$ ($H/\lambda < 0.375$); $Q = 20$ ($0.375 \leq H/\lambda$).
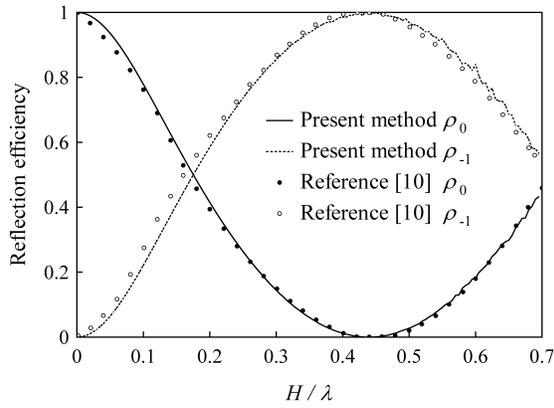
4.3.2. Perfectly Conducting Grating: H-wave Case

We solved the problem of an H-wave incidence by the present method. The statement of problem and the method of analysis are concisely described in Appendix B.

Figure 11 shows reflection efficiency as a function of the incident

Figure 9. Diffraction efficiency as functions of H/λ (E-wave). $n_{\mathrm{U}} = 1$, $D = 0.8\lambda$, $\theta = 36.0°$, $N = 20$. $Q = 10$ $(H/\lambda < 0.425)$; $Q = 20$ $(0.425 \leq H/\lambda)$.
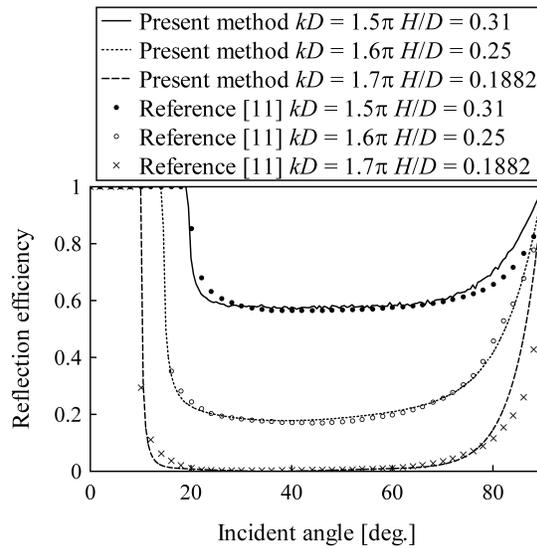

Figure 10. Diffraction efficiency as functions of H/λ (E-wave). $n_{\mathrm{U}} = 1$, $D = 0.9\lambda$, $\theta = 33.72°$, and $N = 20$. $Q = 10$ $(H/\lambda < 0.45)$; $Q = 20$ $(0.45 \leq H/\lambda)$.

angle. The curves and the symbols represent the present results and the results in [11]. We find that the results agree with each other except for the grazing limit.

Table 1. Parameters in Figs. 8–10.

| Fig. | N | Q | | D/λ | θ | Range of | Energy error |
		$2H/D < 1.0$	≥ 1.0		[deg]	$2H/D$	ε_N [%]
8*	20	10	20	0.75	41.8	< 2.666	$\varepsilon_N < 3$
9	20	10	20	0.85	36.0	< 1.412	$\varepsilon_N < 1$
10	20	10	20	0.9	33.72	< 1.556	$\varepsilon_N < 1$

*We have set $Q = 40$ and $N = 25$ in the range $1.6 \leq 2H/D$.

Figure 11. Diffraction efficiency as functions of θ (H-wave). $n_U = 1$, $N = 15$, and $Q = 4$. $D = 0.75\lambda$ ($H/D = 0.31$); $D = 0.8\lambda$ ($H/D = 0.25$); and $D = 0.85\lambda$ ($H/D = 0.1882$).

4.3.3. A Dielectric Grating

We solved the problem of a dielectric grating by the present method and compared the results with those obtained by the Rigorous Coupled Wave Analysis (RCWA) method [12]. Interested readers can find the details of our analysis in Appendix D.

Figure 12 shows the reflection and transmission efficiency calculated by the present method (solid curves) and by the RCWA
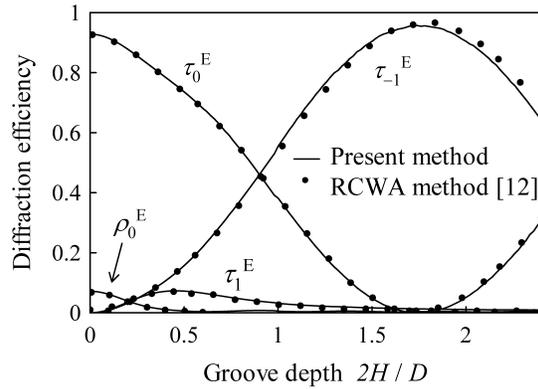
Figure 12. Diffraction efficiency of a dielectric grating as functions of the groove depth $2H$ (E-wave). $D = \lambda$, $n_{\mathrm{U}} = 1$, $n_{\mathrm{L}} = 1.58$, and $\theta = 30°$.

(dots). Setting $Q = 19$ and $N = 9$, we obtained the solutions with 0.3 percent energy error in the range $2H/D \leq 2.4$. Our results agree well with those given in [12].

4.4. Discussions

The basic idea of the present method includes two main points:
(i) Partition of the groove region: the groove region is divided into stratified shallow layers and the Floquet modal functions are defined in each layer.
(ii) Employment of up- and down-going modal functions: an approximate solution in a layer consists not only of up-going but also of down-going modal functions.

The partition can be understood as normalization of modal functions. This is needed when we solve the problem of a deep grating to reduce the strongly oscillating nature of higher evanescent orders (they strongly oscillate in X and rapidly change their magnitude in Z).

The employment of the down-going waves increases the degree of freedom in the least-squares boundary matching and, consequently, extends the range of application. In order to see how the down-going waves work, we show the comparison of energy error decrease in Fig. 13. The curves (1), (2), and (3) represent the results obtained by the CYM, the results by a method with the partition (normalization) alone, and the results by the present method. When the grating is

(a) $2H/D = 0.2$ (b) $2H/D = 0.6$

(c) $2H/D = 1.0$ (d) $2H/D = 2.0$
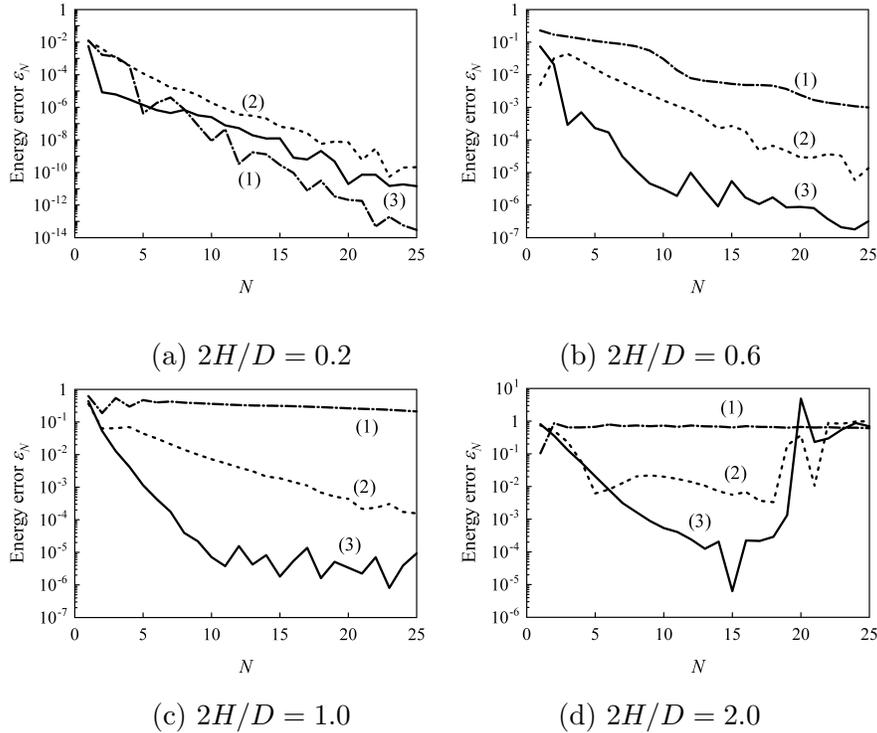
Figure 13. Comparison of the solutions obtained by the CYM (1), a method with normalization alone (2), and by the present method (3). $D = 0.6\lambda$, $n_{\mathrm{U}} = 1.0$, $n_{\mathrm{L}} = 1.58$, $\theta = 30°$, and $Q = 20$ (for (2) and (3)).

shallow ($2H/D = 0.2$), the CYM gives the best solution. If $2H/D$ is increased beyond 0.6, however, the present method is found to be the best: the CYM fails to give a solution for $2H/D = 1.0$; and the solution by a method with the partition alone converges not so rapidly as the solution by the present method. In particular, when $2H/D = 2.0$, the present method with $Q = 20$ and $N = 13$ gives a solution with 0.01 percent energy error while the error of the solution by the method with the partition alone cannot be less than 0.5 percent. We, hence, can conclude that the partition (normalization) alone is not so effective as the combined ideas.

It is useful to remember the argument by Lippmann [13] that the Rayleigh expansion (an infinite series in terms of up-going Floquet modes in U) seems unrealistic because not only up-going but also down-going waves are expected to exist in U_{G}. This was the first doubt about the Rayleigh expansion and was the starting point of the controversy

on the validity of the Rayleigh hypothesis. It was shown, however, that the expansion can represent the field in U provided that the infinite series converges inside U_G and on the grating surface. Hence, to find the range of validity of the infinite series was considered to be the main subject of the controversy. We, however, do not go further in this topic because the aim of this article is not to give a review of the Rayleigh hypothesis. Readers can find a list of interesting references in [14].

It is an accepted knowledge that the set of up-going Floquet modes alone is necessary to describe the field in U_G. From a physical point of view, however, it is natural to assume a discontinuity at the border between U_0 and U_G and, hence, the existence of down-going waves in U_G. In particular, when the grooves are deep, this really seems to be the case. This is the reason why employment of down-going Floquet modes, although it is not necessary from a mathematical point of view, has increased the effectiveness of Yasuura's method of modal expansion.

5. CONCLUSIONS

We have shown that Yasuura's method, when combined with the partition of the groove region, can solve the problem of diffraction from a deep grating with a depth-to-period ratio beyond unify. Although there are a couple of methods [15, 16] that are capable of solving the problems of extremely deep gratings, the present results make sense showing a limit of a conventional modal-expansion approach with the Floquet modes as basis functions.

We are working with a doubly periodic grating (the grating surface is periodic in both X- and Y-direction) [17] and with a stratified grating (gratings made of different materials are stratified in Z) [18]. When the amplitude of the surface modulation becomes large in these problems, the present method works to obtain reliable solutions.

APPENDIX A. CYM FOR AN E-WAVE PROBLEM

For readers' convenience, we here state briefly the method of solution by the conventional Yasuura's method (CYM) [1–4] applied to a problem of diffraction by a perfectly conducting grating.

We define an approximate solution

$$\Phi_N^{\mathrm{d}}(\mathbf{P}) = \sum_{m=-N}^{N} A_m \phi_m^+(X, Z) \quad (\mathbf{P} \in \mathrm{U}) \tag{A1}$$

where $\phi_m^+(X, Z)$ is given by (6).

We determine the A_m coefficients in order that the solution meets the boundary condition (4) in the sense of least-squares. That is, we define a mean-square error

$$I'_N = \int_{C_0} \left| \Phi_N^{\mathrm{d}}(x, z(x)) + \Phi^{\mathrm{i}}(x, z(x)) \right|^2 dx \qquad \text{(A2)}$$

and find the A_m's that minimize (A2).

APPENDIX B. H-WAVE CASE

In an H-wave problem we assume that the incident wave is given by

$$\mathbf{H}^{\mathrm{i}}(\mathbf{P}) = \mathbf{i}_Y \Phi^{\mathrm{i}}(\mathbf{P}) \qquad \text{(B1)}$$

instead of (2). The solution of this problem $\Phi^{\mathrm{d}}(\mathbf{P})$, the Y component of the diffracted magnetic field, satisfies the requirements 1., 2., 3., and the boundary condition

$$\left. \frac{\partial(\Phi^{\mathrm{i}} + \Phi^{\mathrm{d}})}{\partial \nu} \right|_{z=f(x)+0} = 0 \qquad \text{(B2)}$$

The method of solution for this problem is almost the same as that described in Section 3 except that the last term in the right-hand side of (13) should be replaced by

$$\sum_{q=1}^{Q} \left\{ W^2 \int_{C_{0q}} \left| \frac{\partial \Phi_{qN}^{\mathrm{d}}}{\partial \nu}(x, z(x)) \right|^2 dx \right\} \qquad \text{(B3)}$$

If we solve this problem by the CYM (Appendix A), the quadratic form

$$I''_N = \int_{C_0} \left| \frac{\partial \Phi_N^{\mathrm{d}}}{\partial \nu}(x, z(x)) + \frac{\partial \Phi^{\mathrm{i}}}{\partial \nu}(x, z(x)) \right|^2 dx \qquad \text{(B4)}$$

should be minimized instead of (A2).

APPENDIX C. DISCRETIZATION AND A METHOD OF SOLUTION FOR THE DISCRETIZED PROBLEM

We show how to find the modal coefficients that minimize the mean-square error (13) on a computer. Because (13) defines a least-squares problem in a function space, we first state how to discretize the problem.
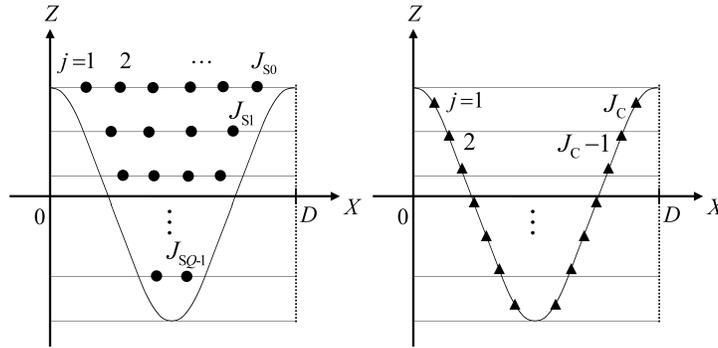
(a) Sampling points on the fictitious boundaries

(b) Sampling points on the grating surface

Figure C1. Location of the sampling points on the boundaries.

Before doing this, we modify the integrand term of (13) slightly: We remove a common factor $e^{i\alpha_0 x}$ from each function in the absolute symbol. Resultant functions are periodic in X with a period D. Noting that the rectangular rule gives the same result as the trapezoidal rule for a definite integral of a periodic function over one period, we apply the rectangular rule to the modified (13) to have a discretized least-surface problem.

Determination of the number of sampling points. We locate sampling points on the fictitious boundaries Γ_q $(q = 0, 1, \cdots, Q - 1)$ and on the grating surface C_{0q} $(q = 1, 2, \cdots, Q)$ as shown in Fig. C1. Let the total number of the sampling points be

$$J = J_{\mathrm{S}} + J_{\mathrm{C}} \tag{C1}$$

where J_{S} and J_{C} stand for the numbers on the fictitious boundaries and on the grating surface. We determine the numbers by the following rules:

(R1) The number of lows of a Jacobian in the discretized least-squares problem (the number of linear equations in an over-determined set of equations) is twice as many as the number of columns (the number of unknowns N_{T});

(R2) The distance between two neighboring sampling points is fixed approximately.

Let an approximation of the surface length be

$$\ell' = D\sqrt{1 + \left(\frac{4H}{D}\right)^2} \tag{C2}$$

Then, according to (R2), we have a relation between J_C and J_{S0}, the number on the top fictitious boundary Γ_0:

$$\frac{J_C}{J_{S0}} = \frac{\ell'}{D} = \sqrt{1 + \left(\frac{4H}{D}\right)^2} \tag{C3}$$

If we adopt an additional approximation

$$J_S = (Q+1)J_{S0}/2 \tag{C4}$$

and if we combine (C3) and (C4) with (R1), we have

$$J_S \approx \frac{1}{1 + \frac{1}{Q+1}\sqrt{1 + \left(\frac{4H}{D}\right)^2}} \cdot (2Q+1)(2N+1), \quad J_C \approx \frac{2}{Q+1}\sqrt{1 + \left(\frac{4H}{D}\right)^2} \cdot J_S \tag{C5}$$

because the number of unknowns is $(2Q+1)(2N+1)$ and the number of equations at a point on the fictitious boundary is 2 while the number at a surface point is 1.

Location of the sampling points. Having determined the numbers of sampling points on each boundary, we can locate the points as follows: The sampling points on fictitious boundaries Γ_q are given by

$$(x, z) = (x_j^S, z_q) \quad (j = 1, 2, \cdots, J_{Sq} ; \quad q = 0, 1, \cdots, Q - 1) \tag{C6}$$

$$x_j^S = \frac{2j-1}{2J_{S0}}D \quad (j = 1, 2, \cdots, J_{S0} ; \quad z_q > f(x_j^S)) \tag{C7}$$

And the points on the grating surface are given by

$$(x, z) = (x_j^C, f(x_j^C)) \quad (j = 1, 2, \cdots, J_C) \tag{C8}$$

$$x_j^C = \frac{2j-1}{2J_C}D \quad (j = 1, 2, \cdots, J_C) \tag{C9}$$

Discritized form of (13). Application of the rectangular rule to (13) gives

$$I_N \simeq I_{NJ} \equiv \frac{1}{J}||\Phi\mathbf{A} - \mathbf{b}||_{M_T}^2 \tag{C10}$$

where $|| \cdot ||_{M_T}$ denotes the M_T-dimensional Euclidean norm with $M_T = 2J_S + J_C$ and Φ is a Jacobian having a block-diagonal structure

$$\Phi = \begin{bmatrix} \Phi_{11} & \Phi_{12} & 0 & \cdots & 0 \\ 0 & \Phi_{22} & \Phi_{23} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \Phi_{QQ} & \Phi_{QQ+1} \end{bmatrix} \quad (M_T \times N_T) \tag{C11}$$

$\Phi_{q\ell}(\ell = q, q+1; \ q = 1, 2, \cdots, Q)$ are $m_q \times n_\ell$ partial matrices where $m_q = 2J_{Sq-1} + J_C/Q$ $(q = 1, 2, \cdots, Q)$, $n_1 = 2N+1$, and $n_q = 2(2N+1)$ $(q = 2, 3, \cdots, Q+1)$. The elements of $\Phi_{q\ell}$ are the modified boundary values of the modal functions at the sampling points. We assume that Φ is full rank $(\mathrm{rank}(\Phi) = N_T)$, i.e., the N_T column vectors of Φ are linearly independent.

The vector $\mathbf{A}$ is a solution vector defined by

$$\mathbf{A} = \left[\begin{array}{cccc} \mathbf{A}_1{}^T & \mathbf{A}_2{}^T & \cdots & \mathbf{A}_Q{}^T & \mathbf{A}_{Q+1}{}^T \end{array} \right]^T \qquad (N_T \times 1) \quad \text{(C12)}$$

with $\mathbf{A}_1 = \left[A_{-N}^{(0)}, A_{-N+1}^{(0)}, \cdots, A_N^{(0)} \right]^T$ $(n_1 \times 1)$, $\mathbf{A}_q = \left[A_{-N}^{(q-1)}, A_{-N+1}^{(q-1)}, \right.$ $\left. \cdots, A_N^{(q-1)}, B_{-N}^{(q-1)}, B_{-N+1}^{(q-1)}, \cdots, B_N^{(q-1)} \right]^T$ $(n_q \times 1)$ $(q = 2, 3, \cdots, Q+1)$. The vector

$$\mathbf{b} = \left[\begin{array}{cccc} \mathbf{b}_1{}^T & \mathbf{b}_2{}^T & \cdots & \mathbf{b}_Q{}^T \end{array} \right]^T \qquad (M_T \times 1) \qquad \text{(C13)}$$

is a given right-hand side whose first entry ($\mathbf{b}_1$) consists of the sampled values of the incident wave. Other $\mathbf{b}_q$'s are m_q-dimensional zero vectors.

The method of solution. Here, we describe the method of solution for the discretized least-squares problem. The problem becomes large when Q (the number of divisions) or N (the number of truncation) increases. In solving such a large problem, we employ the sequential accumulation (SA) [19] to reduce the computational complexity. The SA-scheme to obtain a QR decomposition of the Jacobian (C11) is as follows:

A QR method with SA.
Step 1. We take partial matrices Φ_{11} and Φ_{12} from Φ and $\mathbf{b}_1$, the corresponding right-hand side, to define

$$\Phi^{(1)} = [\ \Phi_{11} \ \ \Phi_{12} \ ] \ (m_1 \times (n_1 + n_2)), \ \ \mathbf{b}^{(1)} = [\ \mathbf{b}_1 \ ] \ \ (m_1 \times 1) \ \text{(C14)}$$

We then decompose $\Phi^{(1)}$ to obtain

$$\Phi^{(1)} = Q^{(1)} \widetilde{R}^{(1)} = Q^{(1)} \left[\begin{array}{cc} R_{11}^{(1)} & R_{12}^{(1)} \\ 0 & R_{22}^{(1)} \\ \hline \multicolumn{2}{c}{0} \end{array} \right] \qquad \text{(C15)}$$

Hence, if we operate $Q^{(1)*} \ (= \overline{Q^{(1)T}})$ to a combination $[\Phi^{(1)} \ \mathbf{b}^{(1)}]$ from the left, we have

$$Q^{(1)*} \left[\begin{array}{cc} \Phi^{(1)} & \mathbf{b}^{(1)} \end{array} \right] = \left[\begin{array}{cc|c} R_{11}^{(1)} & R_{12}^{(1)} & \mathbf{g}_1^{(1)} \\ 0 & R_{22}^{(1)} & \mathbf{g}_2^{(1)} \\ \hline \multicolumn{2}{c|}{0} & \mathbf{e}^{(1)} \end{array} \right] \qquad \text{(C16)}$$

<u>Step q $(q = 2, 3, \cdots, Q)$.</u> We take partial matrices $\mathrm{R}_{qq}^{(q-1)}$ from $\widetilde{\mathrm{R}}^{(q-1)}$, Φ_{qq} and Φ_{qq+1} from Φ, $\mathbf{g}_q^{(q-1)}$, and $\mathbf{b}_q$ to define

$$
\Phi^{(q)} = \begin{bmatrix} \mathrm{R}_{qq}^{(q-1)} & 0 \\ \Phi_{qq} & \Phi_{qq+1} \end{bmatrix} \ ((n_q + m_q) \times (n_q + n_{q+1})),
$$

$$
\mathbf{b}^{(q)} = \begin{bmatrix} \mathbf{g}_q^{(q-1)} \\ \mathbf{b}_q \end{bmatrix} \ ((n_q + m_q) \times 1) \tag{C17}
$$

We obtain $\mathrm{Q}^{(q)}\widetilde{\mathrm{R}}^{(q)}$ by decomposing $\Phi^{(q)}$. Then, operating $\mathrm{Q}^{(q)*}$ to $[\Phi^{(q)} \ \mathbf{b}^{(q)}]$, we have

$$
\Phi^{(q)} = \mathrm{Q}^{(q)}\widetilde{\mathrm{R}}^{(q)} = \mathrm{Q}^{(q)} \begin{bmatrix} \mathrm{R}_{qq}^{(q)} & \mathrm{R}_{qq+1}^{(q)} \\ 0 & \mathrm{R}_{q+1q+1}^{(q)} \\ \hline & 0 \end{bmatrix} \tag{C18}
$$

$$
\mathrm{Q}^{(q)*} \begin{bmatrix} \Phi^{(q)} & \mathbf{b}^{(q)} \end{bmatrix} = \begin{bmatrix} \mathrm{R}_{qq}^{(q)} & \mathrm{R}_{qq+1}^{(q)} & \mathbf{g}_q^{(q)} \\ 0 & \mathrm{R}_{q+1q+1}^{(q)} & \mathbf{g}_{q+1}^{(q)} \\ \hline 0 & & \mathbf{e}^{(q)} \end{bmatrix} \tag{C19}
$$

<u>Step $Q + 1$.</u> If $\mathrm{R}_{qq}^{(q)}$, $\mathrm{R}_{qq+1}^{(q)}$ and $\mathbf{g}_q^{(q)}$ are obtained, $\mathbf{A}_q$'s are found by solving simultaneous linear equations, which have upper triangle coefficient matrices, by backward substitution:

$$
\begin{cases} \mathrm{R}_{Q+1Q+1}^{(Q)}\mathbf{A}_{Q+1} = \mathbf{g}_{Q+1}^{(Q)} \\ \mathrm{R}_{qq}^{(q)}\mathbf{A}_q = \mathbf{g}_q^{(q)} - \mathrm{R}_{qq+1}^{(q)}\mathbf{A}_{q+1} \ \ (q = Q, Q-1, \cdots, 1) \end{cases} \tag{C20}
$$

The above is an algorithm of QR method with SA. We can prove that the solution vector $\mathbf{A}$ composed of $\mathbf{A}_q$'s obtained from (C20) agrees with the solution vector found by the conventional QR method. The minimized error is given by

$$
I_{NJ\mathrm{min}} = ||\mathbf{e}^{(1)}||_{m_1 - (n_1 + n_2)}^2 + \sum_{q=2}^{Q} ||\mathbf{e}^{(q)}||_{m_q - n_{q+1}}^2 \tag{C21}
$$

APPENDIX D. A DIELECTRIC GRATING

When the grating shown in Fig. 1 is made of a dielectric or a metal, we have a transmitted field $\Phi^{\mathrm{L}}(\mathbf{P})$ in L in addition to the incident and reflected field, $\Phi^{\mathrm{i}}(\mathbf{P})$ and $\Phi^{\mathrm{U}}(\mathbf{P})$, in U. $\Phi^{\mathrm{U}}(\mathbf{P})$ and $\Phi^{\mathrm{L}}(\mathbf{P})$ satisfy
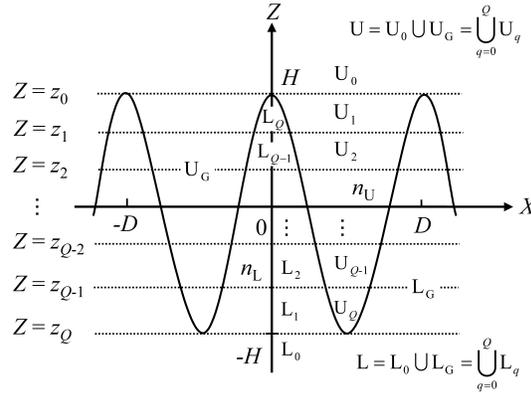
Figure D1. Partition of the upper and the lower regions U and L.

the 2-D Helmholtz equations, the periodicity condition, the radiation condition, and the boundary condition

$$\left.\Phi_q^{\mathrm{U}}\right|_{z=f(x)+0} = \left.\Phi_{Q-q+1}^{\mathrm{L}}\right|_{z=f(x)-0} \quad (q=1,2,\cdots,Q) \quad \text{(D1)}$$

$$\left.\frac{\partial \Phi_q^{\mathrm{U}}}{\partial \nu}\right|_{z=f(x)+0} = \left.\frac{\partial \Phi_{Q-q+1}^{\mathrm{L}}}{\partial \nu}\right|_{z=f(x)-0} \quad (q=1,2,\cdots,Q) \quad \text{(D2)}$$

for an E-wave incidence or

$$\left.\Phi_q^{\mathrm{U}}\right|_{z=f(x)+0} = \left.\Phi_{Q-q+1}^{\mathrm{L}}\right|_{z=f(x)-0} \quad (q=1,2,\cdots,Q) \ \text{(D3)}$$

$$\frac{1}{n_{\mathrm{U}}^2}\left.\frac{\partial \Phi_q^{\mathrm{U}}}{\partial \nu}\right|_{z=f(x)+0} = \frac{1}{n_{\mathrm{L}}^2}\left.\frac{\partial \Phi_{Q-q+1}^{\mathrm{L}}}{\partial \nu}\right|_{z=f(x)-0} \quad (q=1,2,\cdots,Q) \ \text{(D4)}$$

for an H-wave incidence. Here, n_{U} and n_{L} denote refractive indices of the material in U and L. We state the method of solution assuming an E-wave incidence.

First, we divide the semi-infinite region U into the half plane U_0 and the thin layers U_q $(q=1,2,\cdots,Q)$. We have done this in Section 3. Next, we divide the lower semi-infinite region L in a similar way to have L_0 and L_q $(q=1,2,\cdots,Q)$ shown in Fig. D1. The fictitious boundary between L_q and L_{q+1} is a horizontal line

$$Z = z_{Q-q} = (-1 + 2q/Q)H \qquad (q=0,1,\cdots,Q-1) \qquad \text{(D5)}$$

Approximate solutions in U_q $(q=0,1,\cdots,Q)$ take the same forms

as in Section 3. The solutions in L_q $(q = 0, 1, \cdots, Q)$ are defined by

$$\Phi_{0N}^{L}(\mathbf{P}) \quad = \quad \sum_{m=-N}^{N} D_m^{(0)} \phi_m^{L-}(X, Z + H) \quad (\mathbf{P} \in L_0) \qquad (D6)$$

and

$$\Phi_{qN}^{L}(\mathbf{P}) = \sum_{m=-N}^{N} \left\{ C_m^{(q)} \phi_m^{L+}(X, Z - z_{Q-q+1}) + D_m^{(q)} \phi_m^{L-}(X, Z - z_{Q-q}) \right\}$$

$$(\mathbf{P} \in L_q \; ; \; q = 1, 2, \cdots, Q) \qquad (D7)$$

Here, the modal functions in L_q should have the form

$$\phi_m^{L\pm}(X, Z) = \exp[i(\alpha_m X \pm \beta_m^L Z)] \qquad (D8)$$

because they are the solutions of Helmholtz's equation in L.

The approximate solutions should meet an additional set of boundary conditions on the fictitious boundaries:

$$\left. (\Phi^i \delta_{q0} + \Phi_q^U) \right|_{z=z_q+0} \quad = \quad \left. \Phi_{q+1}^U \right|_{z=z_q-0} \quad (q = 0, 1, ..., Q) \qquad (D9)$$

$$\left. \frac{\partial (\Phi^i \delta_{q0} + \Phi_q^U)}{\partial \nu} \right|_{z=z_q+0} \quad = \quad \left. \frac{\partial \Phi_{q+1}^U}{\partial \nu} \right|_{z=z_q-0} \quad (q = 0, 1, ..., Q) \qquad (D10)$$

$$\left. \Phi_{q+1}^L \right|_{z=z_{Q-q}+0} \quad = \quad \left. \Phi_q^L \right|_{z=z_{Q-q}-0} \quad (q = 0, 1, ..., Q) \qquad (D11)$$

$$\left. \frac{\partial \Phi_{q+1}^L}{\partial \nu} \right|_{z=z_{Q-q}+0} \quad = \quad \left. \frac{\partial \Phi_q^L}{\partial \nu} \right|_{z=z_{Q-q}-0} \quad (q = 0, 1, ..., Q) \qquad (D12)$$

Here, δ_{q0} is Kronecker's delta.

We find the coefficients that minimize the quadratic form

$$
\begin{aligned}
I_N \quad = \quad & \sum_{q=0}^{Q-1} \left\{ \int_{\Gamma_q} \left| \left[\Phi^i \delta_{q0} + \Phi_{qN}^U - \Phi_{q+1N}^U \right] (x, z(x)) \right|^2 dx \right. \\
& + \quad W^2 \int_{\Gamma_q} \left. \left| \left[\frac{\partial (\Phi^i \delta_{q0} + \Phi_{qN}^U)}{\partial \nu} - \frac{\partial \Phi_{q+1N}^U}{\partial \nu} \right] (x, z(x)) \right|^2 dx \right\} \\
& + \quad \sum_{q=0}^{Q-1} \left\{ \int_{\Gamma_{Q-q}} \left| \left[\Phi_{q+1N}^L - \Phi_{qN}^L \right] (x, z(x)) \right|^2 dx \right. \\
& + \quad W^2 \int_{\Gamma_{Q-q}} \left. \left| \left[\frac{\partial \Phi_{q+1N}^L}{\partial \nu} - \frac{\partial \Phi_{qN}^L}{\partial \nu} \right] (x, z(x)) \right|^2 dx \right\}
\end{aligned}
$$

$$+ \sum_{q=1}^{Q} \left\{ \int_{C_{0q}} \left| \left[\Phi_{qN}^{U} - \Phi_{Q-q+1N}^{L} \right](x, z(x)) \right|^2 dx \right.$$

$$\left. + W^2 \int_{C_{0q}} \left| \left[\frac{\partial(\Phi_{qN}^{U} - \Phi_{Q-q+1N}^{L})}{\partial \nu} \right](x, z(x)) \right|^2 dx \right\} \qquad \text{(D13)}$$

REFERENCES

1. Yasuura, K. and T. Itakura, "Approximation method for wave functions," Kyushu Univ. Tech. Rep., Vol. 38, No. 1, 72–77, 1965 (in Japanese).

2. Yasuura, K. and T. Itakura, "Complete set of wave functions," Kyushu Univ. Tech. Rep., Vol. 38, No. 4, 378–385, 1966 (in Japanese).

3. Yasuura, K. and T. Itakura, "Approximate algorism by complete set of wave functions," Kyushu Univ. Tech. Rep., Vol. 39, No. 1, 51–56, 1966 (in Japanese).

4. Yasuura, K., "A view of numerical methods in diffraction problems," *Progress in Radio Science*, W. V. Tilson and M. Sauzade (eds.), 1966–1969, 257–270, URSI, Brussels, 1971.

5. Davies, J. B., "A least-squares boundary residual method for the numerical solution of scattering problems," *IEEE Trans., Microwave Theory Tech.*, Vol. MTT-21, No. 2, 99–104, 1973.

6. Van den Berg, P. M., "Reflection by a grating: Rayleigh methods," *J. Opt. Soc. Am.*, Vol. 71, No. 10, 1224–1229, 1981.

7. Hugonin, J. P., R. Petit, and M. Cadilhac, "Plane-wave expansions used to describe the field diffracted by a grating," *J. Opt. Soc. Am.*, Vol. 71, No. 5, 593–598, 1981.

8. Ikuno, H. and K. Yasuura, "Numerical calculation of the scattered field from a periodic deformed cylinder using the smoothing process on the mode-matching method," *Radio Sci.*, Vol. 13, 937–946, 1978.

9. Okuno, Y. and K. Yasuura, "Numerical algorithm based on the mode-matching method with a singular-smoothing procedure for analysing edge-type scattering problems," *IEEE Trans. Antennas Propagat.*, Vol. AP-30, 580–587, 1982.

10. Zaki, K. A. and A. R. Neureuther, "Scattering from a perfectly conducting surface with a sinusoidal height profile: TE polarization," *IEEE Trans. Antennas Propagat.*, Vol. AP-19, No. 2, 208–214, 1971.

11. Zaki, K. A. and A. R. Neureuther, "Scattering from a perfectly conducting surface with a sinusoidal height profile: TM polarization," *IEEE Trans. Antennas Propagat.*, Vol. AP-19, No. 6, 747–751, 1971.

12. Moharam, M. G. and T. K. Gaylord, "Diffraction analysis of dielectric surface-relief gratings," *J. Opt. Soc. Am.*, Vol. 72, No. 10, 1385–1392, 1982.

13. Lippmann, B. A., "Note on the theory of gratings," *J. Opt. Soc. Am.*, Vol. 43, No. 5, 408, 1953.

14. Millar, R. F., "The Rayleigh hypothesis and a related least-squares solution to scattering problems for periodic surfaces and other scatterers," *Radio Sci.*, Vol. 8, No. 9, 785–796, 1973.

15. Chandezon, J., M. T. Dupuis, G. Cornet, and D. Maystre, "Multicoated gratings: a diffraction formalism applicable in the entire optical region," *J. Opt. Soc. Am.*, Vol. 72, No. 7, 839–846, 1982.

16. Moharam, M. G. and T. K. Gaylord, "Rigorous coupled-wave analysis of planar grating diffraction," *J. Opt. Soc. Am.*, Vol. 71, No. 7, 811–818, 1981.

17. Matsuda, T., D. Zhou, and Y. Okuno, "Numerical analysis of plasmon-resonance absorption in bisinusoidal metal gratings," *J. Opt. Soc. Am. A*, Vol. 19, No. 4, 695–699, 2002.

18. Matsuda, T. and Y. Okuno, "A numerical analysis of plane-wave diffraction from a multilayer-overcoated grating," IEICE, Vol. J76-C-I, No. 6, 206–214, 1993 (in Japanese).

19. Lawson, C. L. and R. J. Hanson, *Solving Least-Squares Problems*, Prentice-Hall, Englewood Cliffs, NJ, 1974.