Note on the S-matrix propagation algorithm

Lifeng Li

Department of Precision Instruments, Tsinghua University, Beijing 100084, China

Received July 19, 2002; revised manuscript received November 4, 2002; accepted November 19, 2002

A set of full-matrix recursion formulas for the $W \to S$ variant of the S-matrix algorithm is derived, which includes the recent results of some other authors as a subset. In addition, a special type of symmetry that is often found in the structure of coefficient matrices (W matrices) that appear in boundary-matching conditions is identified and fully exploited for the purpose of increasing computation efficiency. Two tables of floating-point operation (flop) counts for both the new $W \to S$ variant and the old $W \to t \to S$ variant of the S-matrix algorithm are given. Comparisons of flop counts show that in performing S-matrix recursions in the absence of the symmetry, it is more efficient to go directly from W matrices to S matrices. In the presence of the symmetry, however, using t matrices is equally and sometimes more advantageous, provided that the symmetry is utilized.

OCIS codes: 000.3870, 050.1950, 050.2770, 050.7330.

1. INTRODUCTION

After having been applied with great success in numerical modeling of multilayered diffraction gratings for a short time (short relative to the history of electromagnetic grating theory) the S-matrix propagation algorithm is favored by many researchers as a simple and effective means of circumventing the numerical difficulties associated with evanescent waves. It is not only elegant in form, rendering clear physical interpretations, but also efficient in computation cost, since the number of floating-point operations (flops) is in most cases lower than with the naïve and numerically unstable T-matrix algorithm. The S-matrix algorithm has many implementation variants. Some of the most important ones, along with their operation counts, are presented in Ref. 1.

The author's initial intention in writing this paper was to introduce to the readers of this journal, and to extend, the recent work of some Chinese authors on what they call the reflection and transmission coefficient matrix (RTCM) algorithm.[2–5] These references are inaccessible to most readers outside China, because Refs. 2, 4 and 5 are printed in Chinese, and the English language journal that contains Ref. 3 may not be available in many science libraries. Three different algebraic forms of the RTCM have been presented with varying degrees of generality. According to Fu and co-workers,[2–5] RTCM is numerically more efficient than any of the known variants of the S-matrix algorithm. Since reflection and transmission matrices are submatrices of the full S matrix, the relationship between the RTCM and the S-matrix algorithms should be investigated, but it was not touched upon by Fu and co-workers. The scope of the present paper is slightly expanded because during its preparation a paper by Tan[6] appeared in this journal. Tan reformulated, in a remarkably concise form, the enhanced transmission matrix approach of Moharam *et al.*[7] and termed the resultant formulation the enhanced scattering-matrix approach. He also presented a clear discussion of the relationship between his enhanced approach and the S-matrix algorithm and correctly pointed out that the former can be considered a $W \to S$ variant of the latter, which is unfortunately missing from Table 1 of Ref. 1. He concluded that, given eigensolutions, the enhanced scattering-matrix approach was the most direct and efficient way of performing S-matrix recursion.

This paper augments the half-matrix recursion formula of Refs. 5 and 6 to full-matrix recursion, clarifies the relationship between the RTCM and the S-matrix algorithms, and presents some important results as a consequence of a special but common type of structural symmetry of the W matrices (in the notation of Ref. 1). In Section 2, the full-matrix recursion of the $W \to S$ variant is derived for the general form of boundary conditions. In Section 3, special formulas of both the $W \to S$ variant and the $W \to t \to S$ variant are derived as a result of the above-mentioned symmetry of the W matrices. Section 4 identifies the RTCM as a special form of the $W \to S$ variant of the S-matrix algorithm and discusses, with the aid of a table of flop counts, the relative merits of the $W \to S$ and $W \to t \to S$ variants in the presence and absence of the symmetry. Appendix A describes the use of another type of W-matrix symmetry that arises from applying the Fourier modal method to lamellar gratings in conical mountings.

2. DERIVATION OF THE $W \to S$ VARIANT

In this paper we will follow the notation and convention of Ref. 1. Although this notation system is apparently not as compact as that in Ref. 6, it seems to be clearer and easier to follow, at least for those who are already familiar with Ref. 1. Layer p in the multilayered structure has its upper boundary labeled y_p and its lower boundary labeled y_{p-1}, the upward direction being along the positive y axis. The top and bottom semi-infinite layers are labeled $n + 1$ and 0, respectively. We assume that the S-matrix recursion starts upward with layer 0.

Our starting point for the derivation is the boundary-matching equation

$$\begin{bmatrix} W_{11}^{(p+1)} & W_{12}^{(p+1)} \\ W_{21}^{(p+1)} & W_{22}^{(p+1)} \end{bmatrix} \begin{pmatrix} \mathbf{u}^{(p+1)} \\ \mathbf{d}^{(p+1)} \end{pmatrix} = \begin{bmatrix} W_{11}^{(p)} & W_{12}^{(p)} \\ W_{21}^{(p)} & W_{22}^{(p)} \end{bmatrix} \begin{bmatrix} \phi_+^{(p)} & 0 \\ 0 & \phi_-^{(p)} \end{bmatrix} \begin{pmatrix} \mathbf{u}^{(p)} \\ \mathbf{d}^{(p)} \end{pmatrix}, \quad (1)$$

where $\mathbf{u}^{(p)}$ and $\mathbf{d}^{(p)}$ are column vectors whose components are amplitudes of the upward- and downward-propagating or decaying modal fields, respectively. The amplitude vectors are always evaluated at lower boundaries, hence in Eq. (1) the appearance of the diagonal matrices $\phi_\pm^{(p)}$, whose elements are $\exp(i\lambda_m^{(p)\pm}h_p)$, where $\lambda_m^{(p)\pm}$ are the eigenvalues of the modal fields and h_p is the thickness of layer p. The 2×2 block matrices $W^{(p+1)}$ and $W^{(p)}$ in Eq. (1) are either formed directly by juxtaposition of solution vectors of an eigenvalue problem or assembled by using such eigenvectors as building blocks. All modal-solution-based grating methods arrive at the stage symbolized by Eq. (1) somewhere in their mathematical developments, although the equations may not be explicitly written in that form. Given Eq. (1) and the global S matrix $S^{(p-1)}$, where

$$\begin{pmatrix} \mathbf{u}^{(p)} \\ \mathbf{d}^{(0)} \end{pmatrix} = S^{(p-1)} \begin{pmatrix} \mathbf{u}^{(0)} \\ \mathbf{d}^{(p)} \end{pmatrix} = \begin{bmatrix} T_{\mathrm{uu}}^{(p-1)} & R_{\mathrm{ud}}^{(p-1)} \\ R_{\mathrm{du}}^{(p-1)} & T_{\mathrm{dd}}^{(p-1)} \end{bmatrix} \begin{pmatrix} \mathbf{u}^{(0)} \\ \mathbf{d}^{(p)} \end{pmatrix}, \quad (2)$$

our aim is to find an expression of $S^{(p)}$ in terms of the submatrices of $S^{(p-1)}$, $W^{(p+1)}$, and $W^{(p)}$ without using the intermediate t or s matrix.

Expanding the above two equations gives

$$W_{11}^{(p+1)}\mathbf{u}^{(p+1)} + W_{12}^{(p+1)}\mathbf{d}^{(p+1)} = W_{11}^{(p)}\phi_+^{(p)}\mathbf{u}^{(p)} + W_{12}^{(p)}\phi_-^{(p)}\mathbf{d}^{(p)},$$

$$W_{21}^{(p+1)}\mathbf{u}^{(p+1)} + W_{22}^{(p+1)}\mathbf{d}^{(p+1)} = W_{21}^{(p)}\phi_+^{(p)}\mathbf{u}^{(p)} + W_{22}^{(p)}\phi_-^{(p)}\mathbf{d}^{(p)}, \quad (3)$$

and

$$\mathbf{u}^{(p)} = T_{\mathrm{uu}}^{(p-1)}\mathbf{u}^{(0)} + R_{\mathrm{ud}}^{(p-1)}\mathbf{d}^{(p)}, \quad (4a)$$

$$\mathbf{d}^{(0)} = R_{\mathrm{du}}^{(p-1)}\mathbf{u}^{(0)} + T_{\mathrm{dd}}^{(p-1)}\mathbf{d}^{(p)}. \quad (4b)$$

Substituting Eq. (4a) into Eqs. (3) to eliminate $\mathbf{u}^{(p)}$, solving the two resultant equations for $\mathbf{u}^{(p+1)}$ and $\phi_-^{(p)}\mathbf{d}^{(p)}$ in terms of $\mathbf{u}^{(0)}$ and $\mathbf{d}^{(p+1)}$, and finally substituting the expression of $\mathbf{d}^{(p)}$ into Eq. (4b), we obtain the desired expressions of the submatrices of $S^{(p)}$,

$$R_{\mathrm{ud}}^{(p)} = (Z^{-1}X_2)_1, \quad (5a)$$

$$T_{\mathrm{dd}}^{(p)} = \tilde{T}_{\mathrm{dd}}^{(p-1)}(Z^{-1}X_2)_2, \quad (5b)$$

$$T_{\mathrm{uu}}^{(p)} = (Z^{-1}X_1)_1, \quad (5c)$$

$$R_{\mathrm{du}}^{(p)} = R_{\mathrm{du}}^{(p-1)} + \tilde{T}_{\mathrm{dd}}^{(p-1)}(Z^{-1}X_1)_2, \quad (5d)$$

where

$$Z = \begin{bmatrix} W_{11}^{(p+1)} & -W_{11}^{(p)}\tilde{R}_{\mathrm{ud}}^{(p-1)} - W_{12}^{(p)} \\ W_{21}^{(p+1)} & -W_{21}^{(p)}\tilde{R}_{\mathrm{ud}}^{(p-1)} - W_{22}^{(p)} \end{bmatrix}, \quad (6)$$

$$X = \begin{bmatrix} W_{11}^{(p)}\tilde{T}_{\mathrm{uu}}^{(p-1)} & -W_{12}^{(p+1)} \\ W_{21}^{(p)}\tilde{T}_{\mathrm{uu}}^{(p-1)} & -W_{22}^{(p+1)} \end{bmatrix} = [X_1, X_2], \quad (7)$$

$$\tilde{R}_{\mathrm{ud}}^{(p-1)} = \phi_+^{(p)}R_{\mathrm{ud}}^{(p-1)}\phi_-^{(p)-1}, \quad (8a)$$

$$\tilde{T}_{\mathrm{dd}}^{(p-1)} = T_{\mathrm{dd}}^{(p-1)}\phi_-^{(p)-1}, \quad (8b)$$

$$\tilde{T}_{\mathrm{uu}}^{(p-1)} = \phi_+^{(p)}T_{\mathrm{uu}}^{(p-1)}. \quad (8c)$$

In Eqs. (5) the subscripts 1 and 2 attached to the parentheses refer to the upper and lower blocks of the matrix product, and in Eq. (7) the subscripts 1 and 2 of X refer to the left and right blocks of matrix X. Equations (5)–(8) give the complete set of recursion formulas for the $W \rightarrow S$ variant of the S-matrix algorithm. The reader can be easily self-convinced that this variant is unconditionally stable in the sense defined in Ref. 1.

3. CONSEQUENCES OF SYMMETRY IN *W*-MATRICES

The possibility of using symmetry properties of the W matrices to reduce computation cost was realized by Tan. In Ref. 6 he commented that if the inverse matrix Ψ^{-1} (which corresponds to W^{-1} in this paper) could be computed simply, then the variant of his enhanced scattering-matrix approach represented by his Eq. (11) would be more efficient than that represented by his Eq. (12). He briefly mentioned two examples where such simplification was possible. The first example concerns the problem of light propagation in a stratified homogeneous anisotropic medium analyzed by using a 4×4 matrix formalism, and the second concerns the C method, where a certain fine symmetry of the eigenvalue problem gives W^{-1} directly from W virtually without any numerical computation.[8] As Tan remarked, taking advantage of symmetry that leads to orthogonalizations of eigenvectors can be problematic in the presence of a multiplicity of eigenvalues.

There is one type of symmetry in the W matrix that is much coarser than orthogonality, more common in practice, and easier to use. It significantly reduces the computation cost for the ensuing matrix manipulation, although it still leaves the operation counts at the $O(N^3)$ level. This symmetry is given by

$$W = \begin{bmatrix} W_1 & W_1 \\ W_2 & -W_2 \end{bmatrix}. \quad (9)$$

It is evident that any row in the upper block of this W matrix is orthogonal to any row in the lower block, but the rows within the upper or lower blocks, or the columns of W in general, are not orthogonal to each other. For the sake of conciseness, henceforth the term W_{+-} symmetry will be used to name the symmetry of a W matrix that has the structure as shown in Eq. (9). Many readers surely have seen this type of symmetry somewhere before. For example, it is present throughout Eqs. (24–30) of Ref. 7 and in Eq. (39) of Ref. 9. In the former example, one should factor the diagonal matrix X_l or X_L out of the corresponding matrices. It would not be surprising if some researchers had used the W_{+-} symmetry to increase computation efficiency. However, it appears that there has not been a relevant exposition in the literature as far as the S-matrix algorithm for grating modeling is concerned.

Let us consider the consequences of the W_{+-} symmetry for the recursion formulas derived in the previous section and for the calculation of the t matrix that is defined as $t^{(p)} = W^{(p+1)-1}W^{(p)}$ in Ref. 1. Given the W_{+-} symmetry, matrix Z defined in Eq. (6) can be rewritten as

$$Z = \begin{bmatrix} W_1^{(p+1)} & 0 \\ 0 & W_2^{(p+1)} \end{bmatrix} \begin{bmatrix} 1 & -F^{(p)} \\ 1 & G^{(p)} \end{bmatrix}, \tag{10}$$

where

$$F^{(p)} = Q_1^{(p)}(1 + \tilde{R}_{\mathrm{ud}}^{(p-1)}),$$

$$G^{(p)} = Q_2^{(p)}(1 - \tilde{R}_{\mathrm{ud}}^{(p-1)}),$$

$$Q_l^{(p)} = W_l^{(p+1)-1}W_l^{(p)}, \quad l = 1, 2. \tag{11}$$

Noticing that the second factor on the right-hand side of Eq. (10) can be inverted easily by using the formula

$$\begin{bmatrix} 1 & A \\ 1 & B \end{bmatrix}^{-1} = \begin{bmatrix} -B & A \\ 1 & -1 \end{bmatrix}(A - B)^{-1} \tag{12}$$

for any square matrices A and B of the same size, provided that $A - B$ is nonsingular, from Eqs. (5) and (7) we immediately have

$$R_{\mathrm{ud}}^{(p)} = 1 - 2G^{(p)}\tau^{(p)}, \tag{13a}$$

$$T_{\mathrm{dd}}^{(p)} = 2\tilde{T}_{\mathrm{dd}}^{(p-1)}\tau^{(p)}, \tag{13b}$$

$$T_{\mathrm{uu}}^{(p)} = (F^{(p)}\tau^{(p)}Q_2^{(p)} + G^{(p)}\tau^{(p)}Q_1^{(p)})\tilde{T}_{\mathrm{uu}}^{(p-1)}, \tag{13c}$$

$$R_{\mathrm{du}}^{(p)} = R_{\mathrm{du}}^{(p-1)} + \tilde{T}_{\mathrm{dd}}^{(p-1)}\tau^{(p)}(Q_2^{(p)} - Q_1^{(p)})\tilde{T}_{\mathrm{uu}}^{(p-1)}, \tag{13d}$$

with

$$\tau^{(p)} = (F^{(p)} + G^{(p)})^{-1}. \tag{14}$$

What the W_{+-} symmetry brings to the calculation of the t matrix is an extremely simple result:

$$t^{(p)} = \frac{1}{2}\begin{bmatrix} Q_1^{(p)} + Q_2^{(p)} & Q_1^{(p)} - Q_2^{(p)} \\ Q_1^{(p)} - Q_2^{(p)} & Q_1^{(p)} + Q_2^{(p)} \end{bmatrix}. \tag{15}$$

4. COMPARISON AND DISCUSSION

A. Comparison of Operation Counts

Having derived the recursion formulas for the $W \to S$ variant, it is time for us to estimate its computation cost and, for comparison, that of the $W \to t \to S$ variant before discussing their relative merits. The reader is reminded that if $\mathbf{u}^{(0)} \neq 0$ and $\mathbf{d}^{(n+1)} \neq 0$, recursions of all four submatrices of $S^{(p)}$ must be performed; i.e., a full-matrix recursion is needed. If $\mathbf{u}^{(0)} = 0$, $\mathbf{d}^{(n+1)} \neq 0$, and both transmission and reflection coefficients of the whole system are of interest, only submatrices $R_{\mathrm{ud}}^{(p)}$ and $T_{\mathrm{dd}}^{(p)}$ need to be updated; i.e., only a half-matrix recursion is necessary; and if $R_{\mathrm{ud}}^{(n+1)}$ is the only quantity of interest, a quarter-matrix recursion is sufficient. Although only the last two cases are commonly encountered in practice, certain realistic situations require a full-matrix recursion. For instance, the S-matrix propagation can be performed for a layered structure by using the associativity of the S-matrix product, to avoid unnecessarily repeating some portion of the computation (see Subsection 5.B of Ref. 1). Suppose that $S = S_A * S_B * S_C$, where S_A, S_B, and S_C are S-matrices characterizing the scattering properties of layer groups 0 through i, $i + 1$ through j, and $j + 1$ through n, respectively. Then all four submatrices of S_B are required, even if the incident plane wave exists only in medium 0 or medium n.

In the tabulated results to be given below, only operations with number of flops proportional to N^3 are counted, where N is the dimension of the block matrices $W_{12}^{(p)}$, $R_{\mathrm{ud}}^{(p)}$, and $F^{(p)}$, etc. The method of counting is the same as for obtaining Table 1 in Ref. 1. That is, suppose that A, B, and C are $N \times N$ general complex matrices; then $AB + C$, the LU factorization of A, and $A^{-1}B + C$ (or $BA^{-1} + C$) take N^3, $N^3/3$, and $4N^3/3$ flops, respectively.[10] It is worth emphasizing that the flop counts are with respect to complex numbers, and computation of stand-alone inverse matrices should be avoided whenever possible. The actual counting process is straightforward and will not be carried out here. Table 1 gives the flop counts for the $W \to S$ and $W \to t \to S$ variants of the S-matrix algorithm in the presence and absence of the W_{+-} symmetry, for full-, half-, and quarter-matrix recursions. The $W \to t \to S$ variant is given by Eqs. (19a) and (19a′) of Ref. 1. This table complements Table 1 of Ref. 1.

B. Discussion

It is easy to verify that Eqs. (5a) and (5b) are equivalent to Eq. (12) of Ref. 6. From Table 1 we see that when there is no W_{+-} symmetry the conclusion of Ref. 6 is justified. The computation efficiency of the $W \to S$ variant is indeed significantly higher than that of the $W \to t \to S$ variant (and any other variants in Table 1 of Ref. 1). Table 1 also shows that when the entire S-matrix is needed, using the complete set of formulas in Eqs. (5) is more economical than using two half-matrix recursion sequences in opposite directions, even without considering the additional bookkeeping effort for the second recursion sequence.

Equations (13a) and (13b), along with the auxiliary notations in Eqs. (11) and (14), give the RTCM of Refs. 2 and 4. Therefore, although those authors might not have realized it, this version of the RTCM is simply the $W \to S$ variant of the S-matrix algorithm under the special condition of the W_{+-} symmetry. In their more recent work on anisotropic gratings[5] they proposed a more general ap-

Table 1. Operation Counts (in N^3 Flops) per Recursion for Two Variants of the S-matrix Algorithm (Nonconical Mountings)

Algorithm	W_{+-} Symmetry	Flop Count: Full	Half	Quarter
$W \to S$	Without	16 2/3	9 2/3	8 2/3
	With	13	7	6
$W \to t \to S$	Without	19	15	14
	With	11	7	6

proach, also under the name of RTCM, that does not depend on the W_{+-} symmetry. Equation (15) of Ref. 5 can be identified with Eq. (12) of Ref. 6 or with Eqs. (5a) and (5b) of the present paper.

From Table 1 we can conclude that when the W_{+-} symmetry is not available, the $W \to S$ variant is always more efficient than the $W \to t \to S$ variant. When the symmetry is available and utilized, the $W \to S$ variant has exactly the same computation cost as the $W \to t \to S$ variant in the usual situation in which only one or two of the four sub-S-matrices are needed. If all four are needed, then $W \to t \to S$ is actually more efficient than $W \to S$.

It is the simple expression of the t matrix in Eq. (15), made possible by the W_{+-} symmetry, that gives the $W \to t \to S$ variant a slight advantage. This special type of symmetry, although common, is not always available. Let us consider the Fourier modal method for isotropic lamellar gratings in nonconical mountings. The W_{+-} symmetry exists when the sides of the lamellae are parallel to the Oyz plane, assuming that the Oxz plane of a rectangular Cartesian coordinate system is parallel to the grating plane. As soon as the lamellae take a slant angle or either of the two materials in the periodic region becomes anisotropic with an arbitrarily oriented permittivity tensor, the W_{+-} symmetry disappears.

In some cases one can find other types of structural symmetry in the W matrices. In conical mounting these types of symmetry are likely to be found when the W matrices are written in 4×4 block form. Exploiting such symmetry in the implementation of the $W \to S$ and $W \to t \to S$ variants would definitely lead to greatly increased computation efficiency (see Appendix A). In fact, the solution method described in Ref. 3 is the half-matrix recursion version of the $W \to S$ variant of the S-matrix algorithm in disguise, when a certain symmetry of the W matrix in 4×4 block form is taken into account.

The key to success of the work of Fu and co-workers[2–5] and that of Tan[6] is the recognition that, as the latter author pointed out, to perform S-matrix propagation it is not necessary to go through local t or s matrices. With this hindsight, not including $W \to S$ in the discussion of implementation variants of the S-matrix algorithm in Ref. 1 was an obvious oversight. Auxiliary notation is often helpful for pedagogical purposes. For example, the two variants given in Ref. 1 that go through the $\tilde{s}$ matrix are preferred from a physical point of view, because the recursion formulas contain inverse matrices that can readily be expanded into geometric series, giving rise to interpretation of multiple reflections and diffractions. However, from a practical point of view, auxiliary notation should not be introduced at the expense of computation efficiency, as the t matrix has done in the absence of the W_{+-} symmetry.

Naming the variants of the S-matrix algorithm the RTCM method, the enhanced scattering-matrix approach, or something else is a matter of individual preference. The personal opinion of this author is that they all may be simply referred to as the S-matrix algorithm unless the situation warrants a distinction. An advantage of taking this view point is that all situations can be given a unified and systematic treatment.

5. CONCLUSION

A set of one-way, full-matrix recursion formulas for the $W \to S$ variant of the S-matrix algorithm has been derived, which includes the key equation of the enhanced scattering-matrix approach recently developed by Tan[6] and the RTCM algorithm of Fu and co-workers[2–5] as a subset. A special type of symmetry, referred to as the W_{+-} symmetry in this paper, is identified. Because of the common occurrence of this symmetry, its full exploitation for the purpose of saving computation time has been conducted and proves worthwhile. Comparisons of flop counts have shown that in performing S-matrix recursions in the absence of the W_{+-} symmetry and other types of symmetry, it is more efficient to go directly from W matrices to S matrices, bypassing t (or s) matrices. In the presence of the W_{+-} symmetry, however, using t matrices is equally or more advantageous, provided that the symmetry is fully utilized.

APPENDIX A

This appendix presents the W matrix in 4×4 block matrix form, resulting from applying the Fourier modal method to lamellar gratings in conical mountings, and demonstrates how the symmetry of this W matrix can be exploited to numerical advantage. All quantities contained in the W matrix are defined here without derivations, which are beyond the scope of this short paper.

The underlying Cartesian coordinate system is such that the y axis is perpendicular to the grating plane and the z axis is parallel to the grating grooves. We assume that the four rows and four columns of the W matrix are arranged to match the column vectors of the Fourier coefficients of the four tangential electromagnetic field components (E_z, H_z, H_x, E_x) and the column vectors of the unknown modal field amplitudes ($\mathbf{u}^{(\mathrm{e})}$, $\mathbf{u}^{(\mathrm{h})}$, $\mathbf{d}^{(\mathrm{e})}$, $\mathbf{d}^{(\mathrm{h})}$), respectively. The meaning of superscripts (e) and (h) will become clear after Eq. (A3). The W matrix is given by

$$W = \begin{bmatrix} A & J & A & -J \\ B & D & -B & D \\ C & 0 & -C & 0 \\ 0 & K & 0 & -K \end{bmatrix}, \tag{A1}$$

with

$$\begin{aligned} A_{mq} &= E_{zmq}, \\ B_{mq} &= -\frac{k_z \alpha_m}{\mu k_0 \lambda_q^{(\mathrm{e})}} E_{zmq}, \\ C_{mq} &= \frac{k_z^2 + \lambda_q^{(\mathrm{e})2}}{\mu k_0 \lambda_q^{(\mathrm{e})}} E_{zmq}, \\ D_{mq} &= H_{zmq}, \\ J_{mq} &= \frac{k_z}{k_0 \lambda_q^{(\mathrm{h})}} \sum_n [\![\epsilon]\!]_{mn}^{-1} \alpha_n H_{znq}, \\ K_{mq} &= -\frac{k_z^2 + \lambda_q^{(\mathrm{h})2}}{k_0 \lambda_q^{(\mathrm{h})}} \sum_n \left[\!\!\left[\frac{1}{\epsilon}\right]\!\!\right]_{mn} H_{znq}. \end{aligned} \tag{A2}$$

In the above equation E_{zmq} and H_{zmq} are the mth Fourier coefficients of the qth eigenvectors of the following two matrix eigenvalue problems,

$$(\mu k_0^2[\![\epsilon]\!] - \boldsymbol{\alpha}^2 - k_z^2)E_z = \lambda^{(e)2}E_z, \quad \text{(A3a)}$$

$$\left[\!\left[\frac{1}{\epsilon}\right]\!\right]^{-1}(\mu k_0^2 - \boldsymbol{\alpha}[\![\epsilon]\!]^{-1}\boldsymbol{\alpha})H_z - k_z^2 H_z = \lambda^{(h)2}H_z, \quad \text{(A3b)}$$

respectively, and $\lambda_q^{(e)}$ and $\lambda_q^{(h)}$ are the corresponding eigenvalues. These two equations can be derived from Eq. (24) of Ref. 11 by taking proper care of Fourier factorization.[12] The rest of the symbols in Eqs. (A2) and (A3) take their usual meanings. The double square brackets denote Toeplitz matrices generated by the Fourier coefficients of the associated functions, superscript -1 denotes matrix inversion, ϵ is the periodic permittivity, $\boldsymbol{\alpha}$ is a diagonal matrix with elements α_m, $m = \ldots, -1, 0, +1, \ldots$, α_m and k_z are defined in Ref. 11, k_0 is the magnitude of the wave vector in vacuum, and $\mu = 1$. The symmetry of the W matrix in Eq. (A1) will be referred to as 4×4 W_{+-} symmetry below.

The W matrix in Eq. (A1), thanks to its symmetrical form, can be easily inverted in 4×4 form:

$$W^{-1} = \frac{1}{2}\begin{bmatrix} A^{-1} & 0 & C^{-1} & -A^{-1}JK^{-1} \\ 0 & D^{-1} & -D^{-1}BC^{-1} & K^{-1} \\ A^{-1} & 0 & -C^{-1} & -A^{-1}JK^{-1} \\ 0 & D^{-1} & -D^{-1}BC^{-1} & -K^{-1} \end{bmatrix}. \quad \text{(A4)}$$

Letting A_1, B_1, A_2, B_2, etc., be the building blocks of the W matrices for two adjacent media, the t matrix is given by

$$t = \begin{bmatrix} z_1 + z_2 & z_6 & z_1 - z_2 & -z_6 \\ z_5 & z_3 + z_4 & -z_5 & z_3 - z_4 \\ z_1 - z_2 & z_6 & z_1 + z_2 & -z_6 \\ z_5 & z_3 - z_4 & -z_5 & z_3 + z_4 \end{bmatrix}, \quad \text{(A5)}$$

where

$$z_1 = A_2^{-1}A_1/2, \qquad z_2 = C_2^{-1}C_1/2,$$
$$z_3 = D_2^{-1}D_1/2, \qquad z_4 = K_2^{-1}K_1/2,$$
$$z_5 = D_2^{-1}(B_1 - B_2C_2^{-1}C_1)/2,$$
$$z_6 = A_2^{-1}(J_1 - J_2K_2^{-1}K_1)/2. \quad \text{(A6)}$$

Assuming that the dimension of each block of the W matrix in Eq. (A1) is $N \times N$, it takes only $9\,1/3\,N^3$ flops to construct the t matrix, compared with $85\,1/3\,N^3$ flops without taking the symmetry into account. Since matrices A_i and C_i, $i = 1, 2$, differ only by a diagonal-matrix factor, $C_2^{-1}C_1$ can be obtained from $A_2^{-1}A_1$ by $O(N^2)$ multiplications. Thus the computation cost is further cut down to $8N^3$ flops. Note that the structure of the t matrix in Eq. (A5) has a certain symmetry, which suggests that further time saving is possible in implementing the $W \to t \to S$ variant.

With the 4×4 W_{+-} symmetry, the $W \to S$ variant of the S matrix algorithm can be given by the following set of formulas,

$$R_{\text{ud}}^{(p)} = 1 - 2\hat{G}^{(p)}\hat{\tau}^{(p)}U^{(p)}, \quad \text{(A7a)}$$

$$T_{\text{dd}}^{(p)} = 2\tilde{T}_{\text{dd}}^{(p-1)}\hat{\tau}^{(p)}U^{(p)}, \quad \text{(A7b)}$$

$$T_{\text{uu}}^{(p)} = (\hat{F}^{(p)}\hat{\tau}^{(p)}\hat{Q}_2^{(p)} + \hat{G}^{(p)}\hat{\tau}^{(p)}\hat{Q}_1^{(p)})\tilde{T}_{\text{uu}}^{(p-1)}, \quad \text{(A7c)}$$

$$R_{\text{du}}^{(p)} = R_{\text{du}}^{(p-1)} + \tilde{T}_{\text{dd}}^{(p-1)}\hat{\tau}^{(p)}(\hat{Q}_2^{(p)} - \hat{Q}_1^{(p)})\tilde{T}_{\text{uu}}^{(p-1)}, \quad \text{(A7d)}$$

where

$$\hat{F}^{(p)} = W_{1+}^{(p+1)-1}(W_{1+}^{(p)}\tilde{R}_{\text{ud}}^{(p-1)} + W_{1-}^{(p)}),$$
$$\hat{G}^{(p)} = \hat{Q}_2^{(p)}(1 - \tilde{R}_{\text{ud}}^{(p-1)}),$$
$$\hat{\tau}^{(p)} = (\hat{F}^{(p)} + \hat{G}^{(p)})^{-1},$$
$$\hat{Q}_1^{(p)} = W_{1+}^{(p+1)-1}W_{1+}^{(p)},$$
$$\hat{Q}_2^{(p)} = \hat{W}_2^{(p+1)-1}\hat{W}_2^{(p)},$$
$$U^{(p)} = W_{1+}^{(p+1)-1}\hat{W}_1^{(p+1)}, \quad \text{(A8)}$$

and from Eq. (A1),

$$\hat{W}_1 = \begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix}, \qquad \hat{W}_2 = \begin{bmatrix} C & 0 \\ 0 & K \end{bmatrix},$$
$$W_{1\pm} = \begin{bmatrix} A & \pm J \\ \pm B & D \end{bmatrix}. \quad \text{(A9)}$$

The hats above various symbols distinguish these symbols from their counterparts without hats previously defined in Section 3. When $U^{(p)}$ becomes an identity matrix, Eqs. (A7) are formally identical to Eqs. (13). If only $R_{\text{ud}}^{(p)}$ and $T_{\text{dd}}^{(p)}$ are needed, the matrix product $\hat{\tau}^{(p)}U^{(p)}$ in Eqs. (A7a) and (A7b) can be written as

$$\hat{\tau}^{(p)}U^{(p)} = [W_{1-}^{(p)} + W_{1+}^{(p+1)}\hat{Q}_2^{(p)} + (W_{1+}^{(p)} - W_{1+}^{(p+1)}\hat{Q}_2^{(p)})\tilde{R}_{\text{ud}}^{(p-1)}]^{-1}\hat{W}_1^{(p+1)}, \quad \text{(A10)}$$

which requires much less computation effort.

Table 2 gives the flop counts for the $W \to S$ and $W \to t \to S$ variants of the S-matrix algorithm in the presence and absence of the 4×4 W_{+-} symmetry, for full-, half-, and quarter-matrix recursions. In deriving the numbers in the second and fourth rows of the table, Eq. (A10) and the symmetry of the t matrix in Eq. (A5), respectively, are used. Once again, whenever the W_{+-} symmetry is present and properly utilized, the $W \to t \to S$ variant is more efficient than the $W \to S$ variant.

Table 2. Operation Counts (in N^3 Flops) per Recursion for Two Variants of the S-matrix Algorithm (Conical Mountings)

Algorithm	4×4 W_{+-} Symmetry	Flop Count		
		Full	Half	Quarter
$W \to S$	Without	133 1/3	77 1/3	69 1/3
	With	124	45 1/3	33 1/3
$W \to t \to S$	Without	152	120	112
	With	66 2/3	38 2/3	30 2/3

ACKNOWLEDGMENTS

This work is supported by the National Science Foundation of China under project 60125514.

The author can be reached by email at lifengli @tsinghua.edu.cn.

REFERENCES AND NOTES

1. L. Li, "Formulation and comparison of two recursive matrix algorithms for modeling layered diffraction gratings," J. Opt. Soc. Am. A **13**, 1024–1035 (1996).
2. K. Fu, Z. Wang, D. Zhang, J. Wen, and J. Tang, "A vector analytical method of phase diffraction grating," Acta Opt. Sin. **17**, 1652–1659 (1997) (in Chinese).
3. K. Fu, Z. Wang, D. Zhang, J. Zhang, and Q. Zhang, "A modal theory and recursion RTCM algorithm for gratings of deep grooves and arbitrary profile," Sci. China, Ser. A **42**, 636–645 (1999).
4. K. Fu, Z. Wang, J. Zhang, and Q. Zhang, "Fast processing of Fourier modal method for perpendicularly crossed surface-relief binary-period gratings," Acta Opt. Sin. **21**, 236–241 (2001) (in Chinese).
5. X. Tang, K. Fu, Z. Wang, and X. Liu, "Analysis of rigorous modal theory for arbitrary dielectric gratings made with anisotropic materials," Acta Opt. Sin. **22**, 774–779 (2002) (in Chinese).
6. E. L. Tan, "Note on formulation of the enhanced scattering-(transmittance-) matrix approach," J. Opt. Soc. Am. A **19**, 1157–1161 (2002).
7. M. G. Moharam, D. A. Pommet, E. B. Grann, and T. K. Gaylord, "Stable implementation of the rigorous coupled-wave analysis for surface-relief gratings: enhanced transmittance matrix approach," J. Opt. Soc. Am. A **12**, 1077–1086 (1995).
8. L. Li, "Multilayer-coated diffraction gratings: differential method of Chandezon *et al.* revisited," J. Opt. Soc. Am. A **11**, 2816–2828 (1994); errata: J. Opt. Soc. Am. A **13**, 543 (1996).
9. L. Li, "New formulation of the Fourier modal method for crossed surface-relief gratings," J. Opt. Soc. Am. A **14**, 2758–2767 (1997).
10. In this paper the flop counts of matrix operations are based on information provided in G. H. Golub and C. F. Van Loan, *Matrix Computations* (John Hopkins University Press, Baltimore, Md., 1983, 1989, and 1996). To be consistent with Table 1 of Ref. 1, the meaning of a flop follows the original definition given by the authors in the first edition of their book. See the footnote on page 18 of the third edition.
11. L. Li, "A modal analysis of lamellar diffraction gratings in conical mountings," J. Mod. Opt. **40**, 553–573 (1993).
12. L. Li, "Use of Fourier series in the analysis of discontinuous periodic structures," J. Opt. Soc. Am. A **13**, 1870–1876 (1996).