Note on formulation of the enhanced scattering-(transmittance-) matrix approach

Eng Leong Tan

Centre for Wireless Communications, 20 Science Park Road, #02-34/37, TeleTech Park, Singapore Science Park II, 117674 Singapore

Received September 7, 2001; revised manuscript received January 14, 2002; accepted January 14, 2002

The enhanced transmittance matrix approach developed by Moharam *et al.* [J. Opt. Soc. Am. A **12**, 1077 (1995)] is reformulated in a concise and illuminating form in terms of scattering (reflection and transmission) matrices directly. Two equivalent recursive formulations, corresponding to their full- and partial-solution approaches, are presented and extended to allow simultaneous determination of both reflected and transmitted amplitudes. The relationships between these formulations and the *S*-matrix algorithm, together with their relative efficiencies and usefulness, are ascertained and compared by means of compact formulas featuring parallel algebraic structures. It is made evident that given the eigenmode solutions, the enhanced approach is the most direct and efficient way for deducing global scattering matrices. © 2002 Optical Society of America

OCIS codes: 050.2770, 050.1950, 230.4170.

1. INTRODUCTION

It is well known that the standard transmittance (transfer) matrix methods, useful for analysis of multilayered homogeneous and grating structures, may suffer from numerical instabilities in the presence of evanescent growing and decaying fields. One of the celebrated techniques that can circumvent this problem is the scattering-matrix method established by Ko, Sambles, and others.[1,2] This method has been well documented as the *S*-matrix algorithm, along with the *R*-matrix algorithms and many of their implementation variants, by Li.[3] Another numerically stable algorithm has been developed by Moharam *et al.*[4] and is called the enhanced transmittance-matrix approach. This approach anticipates and preempts the potential numerical problems by proper normalizations and substitutions of exponentially growing elements. The resultant formulas have been shown to yield stable and convergent results even for excessively deep gratings with many levels.

The enhanced transmittance-matrix approach has an algebraic structure somewhat different from that of conventional *S*-matrix algorithm, and it was not considered as another variant in Ref. 3. In their paper, Moharam *et al.*[4] presented two seemingly distinct formulations for their full- and partial-solution approaches. The first approach permits the calculation of all the diffraction reflected and transmitted amplitudes, whereas the second approach allows only one or the other amplitude to be calculated and is simpler and more efficient. From their presentation the connection between these two approaches is not obvious. There is also no discussion of the relationships between the algorithms or any indication of their relative numerical efficiencies compared with the *S*-matrix or other algorithms. Furthermore, it will be useful to bring to completion the partial solution obtained by Moharam *et al.*[4] for only the reflected amplitudes. The question then is whether this completed approach remains simpler and more efficient.

This paper will resolve the above doubts and tie the enhanced transmittance-matrix approach in with the scattering-matrix method (*S*-matrix algorithm). In fact, it would seem more instructive to refer to the approach as the enhanced scattering-matrix approach, since all subsequent derivations are to be carried out directly in terms of the more illuminating scattering (reflection and transmission) matrices. In Section 2, two equivalent recursive formulations that correspond to the full- and the partial-solution approaches of Ref. 4 are presented. Through proper choice of field representation, both formulations are delineated in a manner that obviates most manipulations and substitutions of auxiliary notations. Moreover, the partial solution provided by Moharam *et al.*[4] is completed to allow simultaneous determination of both reflected and transmitted amplitudes. In Section 3, the relationships between the enhanced scattering-matrix approach and the classical *S*-matrix algorithm are ascertained by means of compact formulas featuring parallel algebraic structures. The aspects for which this approach is more efficient are indicated, and relative merits and usefulness of the approaches are discussed. Further inference from the formulations also relates them to the variants of the *S*-matrix algorithm implemented with transmittance matrices.

2. ENHANCED SCATTERING-MATRIX APPROACH

Figure 1 depicts the geometry of a multilayered homogeneous or grating structure stratified in the $\hat{\mathbf{z}}$ direction. There are N layers, including the two semi-infinite regions containing the incident (layer 1) and the transmitted (layer N) waves. The upper and lower bounding interfaces of each layer l ($l = 1, 2, \ldots, N$) are denoted respectively by $Z_l^>$ and $Z_l^<$ such that $Z_l^> \geqslant z$ and $Z_l^< \leqslant z$ for z within the layer. Assume that the electromagnetic

0740-3232/2002/061157-05$15.00

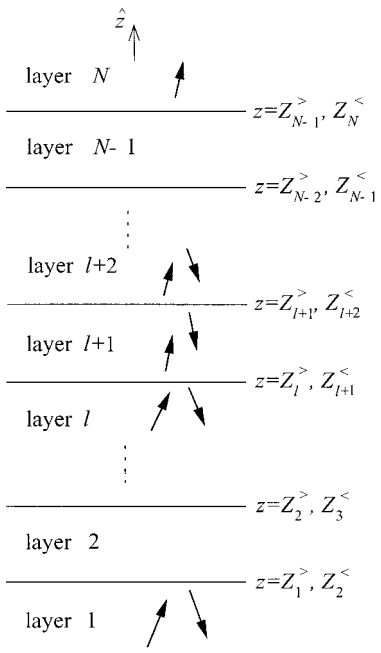

Fig. 1. Geometry of multilayered homogeneous or grating structure. There are N layers including the two semi-infinite regions containing the incident (layer 1) and transmitted (layer N) waves. The upper and lower bounding interfaces of each layer l are denoted by $Z_l^{>}$ and $Z_l^{<}$, respectively.

fields in every layer l have been expanded in terms of a set of eigenmodes that can be written symbolically as

$$\mathbf{f}_l(z) = \mathbf{\Psi}_l \mathbf{P}_l(z) \mathbf{u}_l . \tag{1}$$

Here $\mathbf{f}_l$ is the field vector comprising the tangential electric and magnetic components subjected to interface continuity conditions, $\mathbf{\Psi}_l$ is the eigenmatrix formed by the eigenvectors with their corresponding eigenvalues contained in the exponential terms of $\mathbf{P}_l$, and $\mathbf{u}_l$ is the vector of unknown constants to be specified or determined from the boundary and radiation conditions. The meanings of these notations can be understood further from the context of Ref. 4; e.g., $\mathbf{f}_l$ corresponds to the vector composed of $\mathbf{U}_l$ and $\mathbf{S}_l$, while $\mathbf{\Psi}_l$ corresponds to the matrix of $\mathbf{W}_l$ and $\pm\mathbf{V}_l$, or $\mathbf{I}$ and $\pm j\mathbf{Z}_l$ for the incident and the transmitted layers specifically.

In general, the eigenmodes consist of both upward-bounded and downward-bounded types that remain bounded in the upper and lower regions, respectively. (Association with boundedness, rather than propagation, is more pertinent to the numerical instability of concern). To be specific henceforth, the upward-bounded and the downward-bounded partitions of the matrices in Eq. (1) are arranged and denoted by superscripts $>$ and $<$, respectively, as

$$\mathbf{\Psi}_l = [\mathbf{\Psi}_l^{>} \quad \mathbf{\Psi}_l^{<}], \qquad \mathbf{P}_l(z) = \begin{bmatrix} \mathbf{P}_l^{>}(z) & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_l^{<}(z) \end{bmatrix},$$

$$\mathbf{u}_l = \begin{pmatrix} \mathbf{u}_l^{>} \\ \mathbf{u}_l^{<} \end{pmatrix}. \tag{2}$$

Moreover, $\mathbf{P}_l^{>}(z)$ will not grow infinitely large even when $z \to +\infty$ in the region $z > 0$, and similarly for $\mathbf{P}_l^{<}(z)$ as $z \to -\infty$ in the region $z < 0$; but they will in the respective complementary regions. Such growing of exponentials (hence inaccuracy) is first preempted in Eq. (1) by proper normalization with respect to the lower and upper bounding interfaces:

$$\mathbf{f}_l(z) = \mathbf{\Psi}_l \mathbf{P}_l(z) \mathbf{N}_l \mathbf{c}_l , \qquad \mathbf{N}_l = \begin{bmatrix} \mathbf{P}_l^{>}(-Z_l^{<}) & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_l^{<}(-Z_l^{>}) \end{bmatrix}, \tag{3}$$

where $\mathbf{c}_l$ is just another constant vector to be partitioned like $\mathbf{u}_l$; cf. $\mathbf{c}_l^{\gtrless}$ corresponding to $c_l^{\pm}$ in Ref. 4 Observe that $\mathbf{P}_l(z)\mathbf{N}_l$ taken together will always be bounded for z within layer l. Still, further normalization and introduction of new notation are required for absorption of the inverses of $\mathbf{P}_l^{\gtrless}(z - Z_l^{\lessgtr})$ possibly encountered during subsequent imposition of boundary conditions. This is what has been performed in the enhanced transmittance-matrix approach of Ref. 4 and accounts for its numerical stability, in contrast to the instability in the standard transfer-matrix method.

In a more concise manner, the field vector of Eq. (1) can be rewritten as

$$\mathbf{f}_l(z) = \mathbf{\Psi}_l \mathbf{w}_l(z). \tag{4}$$

Here $\mathbf{P}_l$ has been lumped together with $\mathbf{u}_l$ so that each unknown constant is now weighted with an exponential term. The weight vector $\mathbf{w}_l$ is then no longer constant but a function of z within each layer. Furthermore, the usual reflection and transmission matrices are to be defined with respect to the partitions of $\mathbf{w}_l$; i.e., $\mathbf{w}_l^{\gtrless}$ instead of $\mathbf{u}_l^{\gtrless}$ or $\mathbf{c}_l^{\gtrless}$ above. With such choice of variables, the determination of these matrices can be carried out succinctly for each individual layer as well as for the whole structure. Referring to the layers l, $l+1$, and $l+2$ in Fig. 1 (serving as the building blocks for the entire geometry), let us consider the upward-bounded waves $\mathbf{w}_l^{>}$ incident from layer l to $l+1$. At the interface between these layers, the reflected downward-bounded wave in layer l and the transmitted upward-bounded wave in layer $l+1$ can be described conveniently through the reflection and transmission matrices in

$$\mathbf{w}_l^{<}(Z_l^{>}) = \mathbf{r}_{l,l+1}^{g} \mathbf{w}_l^{>}(Z_l^{>}), \tag{5}$$

$$\mathbf{w}_{l+1}^{>}(Z_{l+1}^{<}) = \mathbf{t}_{l,l+1}^{g} \mathbf{w}_l^{>}(Z_l^{>}). \tag{6}$$

These matrices are superscripted with g to signify the global incorporation of multiple-reflection effects due to the presence of layer $l+2$ and beyond. In the similar manner, all reflections from layer $l+2$ and more are to be lumped into the global reflection matrix at $z = Z_{l+1}^{>}$:

$$\mathbf{w}_{l+1}^{<}(Z_{l+1}^{>}) = \mathbf{r}_{l+1,l+2}^{g} \mathbf{w}_{l+1}^{>}(Z_{l+1}^{>}). \tag{7}$$

This relation can in turn be transferred to another interface at $z = Z_{l+1}^{<}$ via

$$\mathbf{w}_{l+1}^{<}(Z_{l+1}^{<}) = \mathbf{R}_{l+1,l+2}^{g} \mathbf{w}_{l+1}^{>}(Z_{l+1}^{<}), \tag{8}$$

$$\begin{aligned} \mathbf{R}_{l+1,l+2}^{g} &= \mathbf{P}_{l+1}^{<}(Z_{l+1}^{<} - Z_{l+1}^{>}) \mathbf{r}_{l+1,l+2}^{g} \\ &\quad \times \mathbf{P}_{l+1}^{>}(Z_{l+1}^{>} - Z_{l+1}^{<}) \\ &= \mathbf{P}_{l+1}^{<}(-d_{l+1}) \mathbf{r}_{l+1,l+2}^{g} \mathbf{P}_{l+1}^{>}(d_{l+1}), \end{aligned} \tag{9}$$

where d_{l+1} is the thickness of layer $l+1$. Tracing from the right-hand sides of Eqs. (8) and (9), $\mathbf{R}^g_{l+1,l+2}$ is seen to describe the round-trip reflection that includes the phase traversal in layer $l+1$ from $Z^<_{l+1}$ to $Z^>_{l+1}$, followed by global reflection back to $Z^<_{l+1}$. Owing to the boundedness of both $\mathbf{P}^>_{l+1}(d_{l+1})$ and $\mathbf{P}^<_{l+1}(-d_{l+1})$, $\mathbf{R}^g_{l+1,l+2}$ will stay bounded as well.

Having set forth the definitions of various scattering matrices in terms of $\mathbf{w}^{\gtrless}_l$, we impose the field-continuity condition at the interface between layer l and $l+1$, obtaining

$$\boldsymbol{\Psi}^>_l + \boldsymbol{\Psi}^<_l \mathbf{r}^g_{l,l+1} = [\boldsymbol{\Psi}^>_{l+1} + \boldsymbol{\Psi}^<_{l+1}\mathbf{R}^g_{l+1,l+2}]\mathbf{t}^g_{l,l+1}. \quad (10)$$

From this point, there are a couple of ways to proceed for deducing $\mathbf{r}^g_{l,l+1}$ and $\mathbf{t}^g_{l,l+1}$, assuming that $\mathbf{R}^g_{l+1,l+2}$ is known. These ways may have subtle differences in their implications and convenience of implementation. In fact, two of them have already been implied in the full- and partial-solution approaches of Ref. 4, although they are not as clear-cut as those to be illustrated here. The first approach to solving for the unknowns is through the intermediate solution of

$$\begin{bmatrix} (\mathbf{t}^g_{l,l+1})^{-1} \\ \mathbf{r}^g_{l,l+1}(\mathbf{t}^g_{l,l+1})^{-1} \end{bmatrix} = \boldsymbol{\Psi}_l^{-1}[\boldsymbol{\Psi}^>_{l+1} + \boldsymbol{\Psi}^<_{l+1}\mathbf{R}^g_{l+1,l+2}]. \quad (11)$$

With the matrix result computed from the right-hand side of Eq. (11), one takes the inverse of its top partition to obtain $\mathbf{t}^g_{l,l+1}$ and then multiplies it by the bottom partition to arrive at $\mathbf{r}^g_{l,l+1}$. The second approach is simply to rearrange Eq. (10) in the form

$$\begin{bmatrix} \mathbf{r}^g_{l,l+1} \\ \mathbf{t}^g_{l,l+1} \end{bmatrix} = [-\boldsymbol{\Psi}^<_l \quad [\boldsymbol{\Psi}^>_{l+1} + \boldsymbol{\Psi}^<_{l+1}\mathbf{R}^g_{l+1,l+2}]]^{-1}[\boldsymbol{\Psi}^>_l], \quad (12)$$

from which we have the solutions $\mathbf{r}^g_{l,l+1}$ and $\mathbf{t}^g_{l,l+1}$ available immediately. Equations (11) and (12) constitute two equivalent recursive schemes that enable one to determine the global reflection and transmission matrices for all layers. Both schemes can be initiated from the top side with $\mathbf{r}^g_{N,N+1} = \mathbf{0}$ for semi-infinite layer N. Meanwhile, the intervening global transmission matrices can be used to generalize relation (6) by connecting the upward-bounded wave at $z = Z^<_N$ directly to that at $z = Z^>_1$:

$$\mathbf{t}^g_{1,N} = \mathbf{t}^g_{N-1,N}\mathbf{P}^>_{N-1}(d_{N-1})\cdots\mathbf{t}^g_{2,3}\mathbf{P}^>_2(d_2)\mathbf{t}^g_{1,2}. \quad (13)$$

Tracing from the right-hand of Eq. (13) reveals that such a cascading scheme complies with the structure geometry of Fig. 1 upward from layer 1 to N; in comparison, $\mathbf{t}^g_{l,l+1}$ mimics an interface while $\mathbf{P}^>_l(d_l)$ mimics a layer.

Equations (11) and (12) can be found to parallel the equations of Sections 6 and 7, respectively, in Ref. 4. While both formulations are connected simply via Eq. (10), this is not immediately apparent from that paper after the procedure has been obscured by many manipulations and substitutions of auxiliary notation. For their partial-solution approach, Moharam *et al.*[4] have reiterated that either the reflected or the transmitted amplitudes only (but not both) are obtained. They also have not elaborated on the (seemingly additional) formulation for transmission, which is very similar (to that for reflection) but with a matching process started at the input side. From Eq. (12), it is evident that reflection and transmission matrices can be determined simultaneously within the realm of the same formulation if desired, and $\mathbf{t}^g_{l,l+1}$ has supplemented their partial solution that was provided merely for the reflected amplitudes. Furthermore, both approaches of Ref. 4 still call for one final step each for deducing the total reflected and transmitted amplitudes; cf. Eqs. (31) and (37) of Ref. 4. These steps are reminiscent of the form of Eq. (10) [multiplied by $\mathbf{w}^>_1(Z^>_1)$ for the incident vector], whose solutions are given explicitly by Eq. (11) or Eq. (12) much as in the case for the intermediate steps. From here, one can appreciate the unification and directness feasible in the manipulation through the proper choice of field representation in Eq. (4). Note that the adoption of Eq. (3) does not lead straightaway to the usual reflection and transmission matrices but rather, e.g., to $\mathbf{c}^<_l = [\mathbf{r}^g_{l,l+1}\mathbf{P}^>_l(d_l)]\mathbf{c}^>_l$; cf. Eq. (34) of Ref. 4.

3. COMPARISON AND DISCUSSION

The above formulations have been described in common terminologies of scattering matrices. Their relationships to the classical algebraic structure of the S-matrix algorithm will now be ascertained. For conciseness and clarity, the S-matrix algorithm will be presented in a form that preserves the inversion of matrices with order $(2n \times 2n)$, like Eq. (11) or Eq. (12), rather than in the conventional form involving $(n \times n)$ matrix inversions only. This will lead us to simple estimation and comparison of the required algebraic operations (hence the relative numerical efficiencies), without being bogged down by the detailed partitioned-matrix inversions and manipulations. First, let us specialize the reflection and transmission matrices of Eq. (12) for the tentative case of layer $l+1$ being semi-infinite:

$$\begin{bmatrix} \mathbf{r}^{\mathrm{l}}_{l,l+1} \\ \mathbf{t}^{\mathrm{l}}_{l,l+1} \end{bmatrix} = [-\boldsymbol{\Psi}^<_l \quad \boldsymbol{\Psi}^>_{l+1}]^{-1}[\boldsymbol{\Psi}^>_l]. \quad (14)$$

The superscript l signifies that the interaction is local to the interface of layers l and $l+1$, and there is no reflected wave contribution in the latter; i.e., $\mathbf{r}^g_{l+1,l+2} = \mathbf{0}$. In an analogous manner, one can deduce the downward (from $l+1$ to l) local matrices by interchanging l and $l+1$ together with $>$ and $<$ in Eq. (14). To retain its matrix inverse, we write instead

$$\begin{bmatrix} \mathbf{t}^{\mathrm{l}}_{l+1,l} \\ \mathbf{r}^{\mathrm{l}}_{l+1,l} \end{bmatrix} = [-\boldsymbol{\Psi}^<_l \quad \boldsymbol{\Psi}^>_{l+1}]^{-1}[-\boldsymbol{\Psi}^<_{l+1}]. \quad (15)$$

Equations (14) and (15) correspond to Eq. (22a) of Ref. 3 written separately for the interface s matrices.

After some algebraic manipulations, the global scattering matrices of Eq. (12) can be expressed in terms of the local ones above as

$$\begin{bmatrix} \mathbf{r}^g_{l,l+1} \\ \mathbf{t}^g_{l,l+1} \end{bmatrix} = \left[\begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \quad \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix} - \begin{bmatrix} \mathbf{t}^{\mathrm{l}}_{l+1,l} \\ \mathbf{r}^{\mathrm{l}}_{l+1,l} \end{bmatrix}[\mathbf{R}^g_{l+1,l+2}]\right]^{-1}\begin{bmatrix} \mathbf{r}^{\mathrm{l}}_{l,l+1} \\ \mathbf{t}^{\mathrm{l}}_{l,l+1} \end{bmatrix}. \quad (16)$$

While this form looks quite expensive with the involvement of a large $(2n \times 2n)$ matrix inversion, it is not ex-

actly so if we exploit the block-triangular structure of the matrix. Indeed, Eq. (16) is tantamount to the two familiar recurrent relations,[3,5]

$$\mathbf{r}^g_{l,l+1} = \mathbf{r}^l_{l,l+1} + \mathbf{t}^l_{l+1,l}\mathbf{R}^g_{l+1,l+2}\mathbf{t}^g_{l,l+1}, \tag{17}$$

$$\mathbf{t}^g_{l,l+1} = [\mathbf{I} - \mathbf{r}^l_{l+1,l}\mathbf{R}^g_{l+1,l+2}]^{-1}\mathbf{t}^l_{l,l+1}, \tag{18}$$

which involve inversion and manipulation of $(n \times n)$ matrices only. Keeping the form of Eq. (16) instead of the somewhat elaborated Eqs. (17) and (18) allows us to identify easily and to gather accordingly the ingredients required for global matrices, i.e., Eqs. (14) and (15) (agreeing with $\mathbf{t}^l_{l+1,l}$ on top of $\mathbf{r}^l_{l+1,l}$ and not otherwise). At this point, it is worth noting which aspects of the enhanced scattering-matrix approach make it more efficient than the S-matrix algorithm. Comparing Eq. (12) and the set of Eqs. (14)–(16), we see that the former has bypassed all intermediate steps and storage space for dealing with local scattering matrices. In other words, one can consider Eq. (12) to be one algorithm of $W \rightarrow S$ as compared with others in Table 1 of Ref. 3. Therefore, if those local matrices are not of primary concern, the enhanced scattering matrix approach, Eq. (12), is the most efficient in that it yields the global scattering matrices in the most direct way when supplied with $\mathbf{\Psi}$ matrices.

It should be emphasized that the foregoing discussions have been confined to merely a subset of the S-matrix algorithm. The enhancement in efficiency as described above is significant when only upward S submatrices are required, as in Ref. 4. For the more general case in which the entire S matrix is needed, the saving in computation time may not be as pronounced. To elaborate, we realize that the derivation of global scattering matrices for incident downward-bounded waves can be achieved by another recursion, such as (starting from the bottom side)

$$\begin{bmatrix} \mathbf{t}^g_{l+1,l} \\ \mathbf{r}^g_{l+1,l} \end{bmatrix} = [\,-[\mathbf{\Psi}^<_l + \mathbf{\Psi}^>_l\mathbf{R}^g_{l,l-1}] \quad \mathbf{\Psi}^>_{l+1}]^{-1}[-\mathbf{\Psi}^<_{l+1}]. \tag{19}$$

This process costs roughly as much as Eq. (12) for involvement of another large matrix inversion. On the other hand, by making full utilization of the local scattering matrices already available in Eqs. (14) and (15), we can construct the downward global matrices readily from

$$\begin{bmatrix} \mathbf{t}^g_{l+1,l} \\ \mathbf{r}^g_{l+1,l} \end{bmatrix} = \left[\begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{r}^l_{l,l+1} \\ \mathbf{t}^l_{l,l+1} \end{bmatrix}[\mathbf{R}^g_{l,l-1}] \quad \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix}\right]^{-1}\begin{bmatrix} \mathbf{t}^l_{l+1,l} \\ \mathbf{r}^l_{l+1,l} \end{bmatrix}. \tag{20}$$

Again, this equation with its block-triangular structure amounts to inversion and manipulation of $(n \times n)$ matrices. Now with all the pertinent formulas available, one can compare that two inversions of full $(2n \times 2n)$ matrices are required in Eqs. (12) and (19), whereas one inversion of a full $(2n \times 2n)$ matrix is needed in Eqs. (14) and (15), followed by two inversions of block-triangular $(2n \times 2n)$ matrices in Eqs. (16) and (20) [equivalently, four elaborated manipulations of $(n \times n)$ matrices]. More detailed comparisons should also account for the possibility of exploiting the symmetrical properties that may be present in the S matrices for certain structures.

Let us move next to consider Eq. (11), which represents another variant of the enhanced scattering-matrix approach. At first sight, this equation appears to be always less efficient than Eq. (12) for deducing the global scattering matrices, since it demands the extra step of inversion and multiplication of $\mathbf{t}^g_{l,l+1}$. However, in cases where one is able to devise some simple means for computing $\mathbf{\Psi}_l^{-1}$, this extra step may be well compensated for, and the overall efficiency of Eq. (11) may actually become higher than that of Eq. (12), whose inverse is usually more expensive. For lossless homogeneous layers, there have been investigations on using the symmetry and orthogonality properties to facilitate the calculation of $\mathbf{\Psi}_l^{-1}$, e.g., Ref. 6. I shall not pursue further here the exploitation of these properties for diffraction gratings but merely remark that in general, most orthogonalizations become exhaustive in the presence of numerically degenerate eigenvalues.[7] If we denote the downward (from $l+1$ to l) transfer of weight vectors across the interface by

$$\mathbf{w}_l(Z^>_l) = \mathbf{T}^{\mathbf{w}}_{l+1,l}\mathbf{w}_{l+1}(Z^<_{l+1}), \tag{21}$$

$$\mathbf{T}^{\mathbf{w}}_{l+1,l} = [\mathbf{T}^{\mathbf{w}>}_{l+1,l} \quad \mathbf{T}^{\mathbf{w}<}_{l+1,l}] = \mathbf{\Psi}_l^{-1}\mathbf{\Psi}_{l+1}, \tag{22}$$

Eq. (11) can be written in a form featuring implications of the (weight) transmittance matrix:

$$\begin{bmatrix} (\mathbf{t}^g_{l,l+1})^{-1} \\ \mathbf{r}^g_{l,l+1}(\mathbf{t}^g_{l,l+1})^{-1} \end{bmatrix} = \mathbf{T}^{\mathbf{w}>}_{l+1,l} + \mathbf{T}^{\mathbf{w}<}_{l+1,l}\mathbf{R}^g_{l+1,l+2}. \tag{23}$$

Another closely akin variant that also retains the eigenmatrix inverse as a whole (in the transmittance matrix) merits presentation for comparison:

$$\begin{bmatrix} \mathbf{r}^g_{l,l+1} \\ \mathbf{t}^g_{l,l+1} \end{bmatrix} = \left[[-\mathbf{T}^{\mathbf{w}<}_{l,l+1}] \quad \begin{bmatrix} \mathbf{I} \\ \mathbf{R}^g_{l+1,l+2} \end{bmatrix}\right]^{-1}[\mathbf{T}^{\mathbf{w}>}_{l,l+1}]. \tag{24}$$

Here we have used instead the upward (from l to $l+1$) transfer of weight vectors denoted by [cf. Eq. (7) of Ref. 3]

$$\mathbf{T}^{\mathbf{w}}_{l,l+1} = [\mathbf{T}^{\mathbf{w}>}_{l,l+1} \quad \mathbf{T}^{\mathbf{w}<}_{l,l+1}] = \mathbf{\Psi}_{l+1}^{-1}\mathbf{\Psi}_l. \tag{25}$$

Even though there is some freedom of choice for implementing upward or downward transference, their computation overhead may differ as dictated by Eqs. (23) and (24), assuming the same operation count for Eqs. (22) and (25). Both formulations are useful especially when supplied with respective $\mathbf{T}^{\mathbf{w}}$ matrices directly, and they complement each other as the variants of Ref. 3 in conjunction with the unconditionally stable algorithm of $t \rightarrow S$.

4. CONCLUSION

The enhanced transmittance-matrix approach developed by Moharam *et al.*[4] has been reformulated in a concise and illuminating form in terms of scattering matrices directly. Two equivalent recursive formulations have been presented and extended to allow simultaneous determination of both reflected and transmitted amplitudes. The relationships between these formulations and the S-matrix algorithm have been ascertained by means of compact formulas featuring parallel algebraic structures. It has been made evident that given the eigenmode solutions, the enhanced scattering-matrix approach is the most direct and efficient way for deducing the global scat-

tering matrices, without the need to deal with the local matrices as in the conventional S-matrix algorithm.

E. L. Tan can be reached by e-mail at cwctanel @nus.edu.sg.

REFERENCES AND NOTES

1. D. Y. K. Ko and J. R. Sambles, "Scattering matrix method for propagation of radiation in stratified media: attenuated total reflection studies of liquid crystals," J. Opt. Soc. Am. A **5**, 1863–1866 (1988).
2. N. P. K. Cotter, T. W. Preist, and J. R. Sambles, "Scattering-matrix approach to multilayer diffraction," J. Opt. Soc. Am. A **12**, 1097–1103 (1995).
3. L. Li, "Formulation and comparison of two recursive matrix algorithms for modeling layered diffraction gratings," J. Opt. Soc. Am. A **13**, 1024–1035 (1996).
4. M. G. Moharam, D. A. Pommet, E. B. Grann, and T. K. Gaylord, "Stable implementation of the rigorous coupled-wave analysis for surface-relief gratings: enhanced transmittance matrix approach," J. Opt. Soc. Am. A **12**, 1077–1086 (1995).
5. Following the common indexing convention for S submatrices, $\mathbf{r}^g_{l,l+1}$ would have been denoted by $\mathbf{r}^g_{l,l+2}$. Since this matrix as well as $\mathbf{R}^g_{l+1,l+2}$ lumps all global reflections including those from far beyond layer $l+2$, the layer indices are simply taken to be two consecutive ones as indicative of the incident boundedness direction.
6. K. Eidner, "Light propagation in stratified anisotropic media: orthogonality and symmetry properties of the 4×4 matrix formalisms," J. Opt. Soc. Am. A **6**, 1657–1660 (1989).
7. L. Li, "Multilayer-coated diffraction gratings: differential method of Chandezon *et al.* revisited," J. Opt. Soc. Am. A **11**, 2816–2828 (1994).