Fourier Modal Method

Jungho Mun

September 13, 2025

Contents

Chapter 1

Formulations

1.1 Maxwell's equations in two-dimensional periodic systems

The starting point of describing the electromagnetic fields is Maxwell's equations in the frequency domain given as

$$\begin{aligned}\nabla \cdot \mathbf{D}(\mathbf{r},\omega) &= \rho(\mathbf{r},\omega)\\ \nabla \cdot \mathbf{B}(\mathbf{r},\omega) &= 0\\ \nabla \times \mathbf{E}(\mathbf{r},\omega) &= i\omega\mathbf{B}(\mathbf{r},\omega)\\ \nabla \times \mathbf{H}(\mathbf{r},\omega) &= -i\omega\mathbf{D}(\mathbf{r},\omega) + \mathbf{j}(\mathbf{r},\omega),\end{aligned}$$

where $\mathbf{E}$ is the electric field [V/m], $\mathbf{H}$ is the magnetic fields [A/m], $\mathbf{D}$ is the electric displacement field, $\mathbf{B}$ is the magnetic flux density, ρ is the charge density, $\mathbf{j}$ is the current density, and SI units and $e^{-i\omega t}$ convention are used. By assuming anisotropic linear media with the constitutive relations

$$\begin{aligned}\mathbf{D}(\mathbf{r},\omega) &= \varepsilon_0\bar{\bar{\varepsilon}}(\mathbf{r},\omega)\cdot\mathbf{E}(\mathbf{r},\omega)\\ \mathbf{B}(\mathbf{r},\omega) &= \mu_0\bar{\bar{\mu}}(\mathbf{r},\omega)\cdot\mathbf{H}(\mathbf{r},\omega),\end{aligned}$$

and normalizing magnetic field $\tilde{\mathbf{H}}(\mathbf{r},\omega) = i\eta_0\mathbf{H}(\mathbf{r},\omega)$, we reformulate the equations in the source-less condition as

$$\begin{aligned}\nabla \times \mathbf{E}(\mathbf{r},\omega) &= k_0\bar{\bar{\mu}}(\mathbf{r},\omega)\cdot\tilde{\mathbf{H}}(\mathbf{r},\omega)\\ \nabla \times \tilde{\mathbf{H}}(\mathbf{r},\omega) &= k_0\bar{\bar{\varepsilon}}(\mathbf{r},\omega)\cdot\mathbf{E}(\mathbf{r},\omega),\end{aligned} \tag{1.1}$$

where ε_0, μ_0, η_0 are the vacuum permittivity, vacuum permeability, and vacuum wave impedance, respectively.

For simplicity, We assume rectangular lattice with lattice vectors $\mathbf{a}_1 = a_x\hat{x}$ and $\mathbf{a}_2 = a_y\hat{y}$. The Fourier expansion of a function with translational symmetry $f(x + a_x) = f(x)$ is

$$\begin{aligned}\mathcal{F}^{-1}[f_m] \equiv f(x) &= \sum_{m=-\infty}^{\infty} f_m e^{-imb_x x}\\ \mathcal{F}[f] \equiv f_m &= \frac{1}{a_x}\int_0^{a_x} f(x)e^{imb_x x}dx,\end{aligned}$$

where $b_x = 2\pi/a_x$. The Fourier expansion of the material properties are expressed as $\varepsilon(x,y) = \mathcal{F}^{-1}[\varepsilon_{mn}]$ and $\varepsilon_{mn} = \mathcal{F}[\varepsilon]$. Because the fields are quasi-periodic with extra phase factor as $f(x+a) = f(x)e^{i\beta a}$ (i.e., Bloch-Floquet theorem), the Fourier expansion of electromagnetic fields is expressed as

$$\begin{aligned}\mathbf{E}(x,y,z) &= E_0\sum_{mn}\mathbf{e}_{mn}(z)e^{i(k_{xm}x+k_{yn}y)}\\ \tilde{\mathbf{H}}(x,y,z) &= E_0\sum_{mn}\mathbf{h}_{mn}(z)e^{i(k_{xm}x+k_{yn}y)},\end{aligned} \tag{1.2}$$

where the partial wave vector $k_{xm} = k_{x0} - mb_x = \beta_x - mb_x$ and $k_{yn} = k_{y0} - nb_y = \beta_y - nb_y$, and β_x and β_y are the bloch periodicities in x- and y-directions, respectively. The fourier coefficients $\mathbf{e}_{mn}$ and $\mathbf{h}_{mn}$ are termed the space harmonics fields, and E_0 carries the information of units (V/m), such that the space harmonics are unit-less.

By inserting eq. 1.2 into eq. 1.1, we obtain Maxwell's equations in the Fourier space as

$$\begin{aligned}iK_{ny}h_{zmn}(z) - \partial_{\tilde{z}}h_{ymn}(z) &= \sum_{pq}(\varepsilon_{xx})_{m-p,n-q}e_{xpq}(z)\\ \partial_{\tilde{z}}h_{xmn}(z) - iK_{xm}h_{zmn}(z) &= \sum_{pq}(\varepsilon_{yy})_{m-p,n-q}e_{ypq}(z)\\ iK_{xm}h_{ymn}(z) - iK_{yn}h_{xmn}(z) &= \sum_{pq}(\varepsilon_{zz})_{m-p,n-q}e_{zpq}(z)\\ iK_{yn}e_{zmn}(z) - \partial_{\tilde{z}}e_{ymn}(z) &= \sum_{pq}(\mu_{xx})_{m-p,n-q}h_{xpq}(z)\\ \partial_{\tilde{z}}e_{xmn}(z) - iK_{xm}e_{zmn}(z) &= \sum_{pq}(\mu_{yy})_{m-p,n-q}h_{ypq}(z)\\ iK_{xm}e_{ymn}(z) - iK_{yn}e_{xmn}(z) &= \sum_{pq}(\mu_{zz})_{m-p,n-q}h_{zpq}(z)\end{aligned} \tag{1.3}$$

where $K_x = k_x/k_0$ and $K_y = k_y/k_0$ are the normalized wavevectors, and $\partial_{\tilde{z}} = \frac{1}{k_0}\frac{\partial}{\partial z}$.

1.2 Matrix representation

The Fourier representation above involves the Fourier index $n, m = -\infty, \dots, -1, 0, 1, 2, \dots, \infty$, which needs to be truncated, for instance, as $n, m = -n_{\max}, \dots, -1, 0, 1, 2, \dots, n_{\max}$. We may combine all involved indices into a matrix form with size of $N = (2n_{\max}+1)^2$. The Fourier matrix $[\mathrm{K}_y]$ becomes a diagonal matrix of size $N \times N$, the space harmonics field vector $[\mathbf{h}_z]$ has size $N(\times 1)$, and the convolution matrix $[\![\varepsilon_{xx}]\!]$ is a matrix of size $N \times N$. Then, eq. 1.3 is expressed in matrix form as

$$\begin{aligned}
i[\mathrm{K}_y][\mathbf{h}_z] - \partial_{\tilde{z}}[\mathbf{h}_y] &= [\![\varepsilon_{xx}]\!][\mathbf{e}_x] \\
\partial_{\tilde{z}}[\mathbf{h}_x] - i[\mathrm{K}_x][\mathbf{h}_z] &= [\![\varepsilon_{yy}]\!][\mathbf{e}_y] \\
[\mathrm{K}_x][\mathbf{h}_y] - [\mathrm{K}_y][\mathbf{h}_x] &= -i[\![\varepsilon_{zz}]\!][\mathbf{e}_z] \\
i[\mathrm{K}_y][\mathbf{e}_z] - \partial_{\tilde{z}}[\mathbf{e}_y] &= [\![\mu_{xx}]\!][\mathbf{h}_x] \\
\partial_{\tilde{z}}[\mathbf{e}_x] - i[\mathrm{K}_x][\mathbf{e}_z] &= [\![\mu_{yy}]\!][\mathbf{h}_y] \\
[\mathrm{K}_x][\mathbf{e}_y] - [\mathrm{K}_y][\mathbf{e}_x] &= -i[\![\mu_{zz}]\!][\mathbf{h}_z]
\end{aligned} \tag{1.4}$$

where $[\![\cdot]\!]$ denotes the convolution matrix. By noting that $[\mathbf{e}_z] = i[\![\varepsilon_{zz}]\!]^{-1}([\mathrm{K}_x][\mathbf{h}_y] - [\mathrm{K}_y][\mathbf{h}_x])$ and $[\mathbf{h}_z] = i[\![\mu_{zz}]\!]^{-1}([\mathrm{K}_x][\mathbf{e}_y] - [\mathrm{K}_y][\mathbf{e}_x])$, the longitudinal field components can be eliminated as

$$\begin{aligned}
-[\mathrm{K}_y][\![\mu_{zz}]\!]^{-1}([\mathrm{K}_x][\mathbf{e}_y] - [\mathrm{K}_y][\mathbf{e}_x]) - \partial_{\tilde{z}}[\mathbf{h}_y] &= [\![\varepsilon_{xx}]\!][\mathbf{e}_x] \\
\partial_{\tilde{z}}[\mathbf{h}_x] + [\mathrm{K}_x][\![\mu_{zz}]\!]^{-1}([\mathrm{K}_x][\mathbf{e}_y] - [\mathrm{K}_y][\mathbf{e}_x]) &= [\![\varepsilon_{yy}]\!][\mathbf{e}_y] \\
-[\mathrm{K}_y][\![\varepsilon_{zz}]\!]^{-1}([\mathrm{K}_x][\mathbf{h}_y] - [\mathrm{K}_y][\mathbf{h}_x]) - \partial_{\tilde{z}}[\mathbf{e}_y] &= [\![\mu_{xx}]\!][\mathbf{h}_x] \\
\partial_{\tilde{z}}[\mathbf{e}_x] + [\mathrm{K}_x][\![\varepsilon_{zz}]\!]^{-1}([\mathrm{K}_x][\mathbf{h}_y] - [\mathrm{K}_y][\mathbf{h}_x]) &= [\![\mu_{yy}]\!][\mathbf{h}_y]
\end{aligned} \tag{1.5}$$

By reformulating this equation into a block matrix form, the PQ matrices are defined as

$$\begin{aligned}
\partial_{\tilde{z}} \begin{bmatrix} \mathbf{e}_x \\ \mathbf{e}_y \end{bmatrix} &= [\mathrm{P]} \begin{bmatrix} \mathbf{h}_x \\ \mathbf{h}_y \end{bmatrix} \\
\partial_{\tilde{z}} \begin{bmatrix} \mathbf{h}_x \\ \mathbf{h}_y \end{bmatrix} &= [\mathrm{Q}] \begin{bmatrix} \mathbf{e}_x \\ \mathbf{e}_y \end{bmatrix}
\end{aligned} \tag{1.6a}$$

$$[\mathrm{P}] = \begin{bmatrix} [\mathrm{K}_x][\![\varepsilon_{zz}]\!]^{-1}[\mathrm{K}_y] & [\![\mu_{yy}]\!] - [\mathrm{K}_x][\![\varepsilon_{zz}]\!]^{-1}[\mathrm{K}_x] \\ [\mathrm{K}_y][\![\varepsilon_{zz}]\!]^{-1}[\mathrm{K}_y] - [\![\mu_{xx}]\!] & -[\mathrm{K}_y][\![\varepsilon_{zz}]\!]^{-1}[\mathrm{K}_x] \end{bmatrix}$$
$$[\mathrm{Q}] = \begin{bmatrix} [\mathrm{K}_x][\![\mu_{zz}]\!]^{-1}[\mathrm{K}_y] & [\![\varepsilon_{yy}]\!] - [\mathrm{K}_x][\![\mu_{zz}]\!]^{-1}[\mathrm{K}_x] \\ [\mathrm{K}_y][\![\mu_{zz}]\!]^{-1}[\mathrm{K}_y] - [\![\varepsilon_{xx}]\!] & -[\mathrm{K}_y][\![\mu_{zz}]\!]^{-1}[\mathrm{K}_x] \end{bmatrix}. \tag{1.6b}$$

Combining the two equations (eq. 1.6a), we obtain the wave equation $\partial_{\tilde{z}}^2 \begin{bmatrix} \mathbf{e}_x \\ \mathbf{e}_y \end{bmatrix} = [\Omega^2] \begin{bmatrix} \mathbf{e}_x \\ \mathbf{e}_y \end{bmatrix}$, where $[\Omega^2] = [\mathrm{P}][\mathrm{Q}]$. The general solution of the space

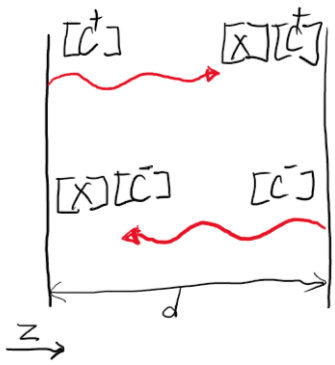

Figure 1: Forward and backward propagating mode coefficients in a single layer with thickness d.

harmonic fields is obtained as

$$\begin{aligned}
\begin{bmatrix} \mathbf{e}_x(z) \\ \mathbf{e}_y(z) \end{bmatrix} &= e^{-[\Omega]\tilde{z}}[\mathbf{e}^+(z=0)] + e^{[\Omega](\tilde{z}-\tilde{d})}[\mathbf{e}^-(z=d)] \\
&= [\mathrm{W}]e^{-[\mathrm{D}]\tilde{z}}[\mathrm{W}]^{-1}[\mathbf{e}^+(z=0)] \\
&\quad + [\mathrm{W}]e^{[\mathrm{D}](\tilde{z}-\tilde{d})}[\mathrm{W}]^{-1}[\mathbf{e}^-(z=d)] \\
&= [\mathrm{W}]e^{-[\mathrm{D}]\tilde{z}}[\mathbf{c}^+] + [\mathrm{W}]e^{[\mathrm{D}](\tilde{z}-\tilde{d})}[\mathbf{c}^-],
\end{aligned} \tag{1.7}$$

where $[\mathrm{W}]$ and $[\mathrm{D}^2]$ are the eigenvector and eigenvalue matrices of $[\Omega^2]$, and $\tilde{d} = k_0 d$ is the normalized layer thickness. Similarly, the magnetic fields are expressed as

$$\begin{bmatrix} \mathbf{h}_x(z) \\ \mathbf{h}_y(z) \end{bmatrix} = -[\mathrm{V}]e^{-[\mathrm{D}]\tilde{z}}[\mathbf{c}^+] + [\mathrm{V}]e^{[\mathrm{D}](\tilde{z}-\tilde{d})}[\mathbf{c}^-], \tag{1.8}$$

where $[\mathrm{V}] = [\mathrm{Q}][\mathrm{W}][\mathrm{D}]^{-1}$. Each mode coefficient consists of x- and y modes as $[\mathbf{c}^+] := \begin{bmatrix} \mathbf{c}_x^+ \\ \mathbf{c}_y^+ \end{bmatrix}$, and the mode coefficients of forward (+) and backward (-) propagating waves can be combined as $[\mathbf{c}] := \begin{bmatrix} \mathbf{c}^+ \\ \mathbf{c}^- \end{bmatrix}$. It should be noted that the reference z coordinate of forward propagation mode $[\mathbf{c}^+]$ is $z = 0$, and that of backward propagating mode $[\mathbf{c}^-]$ is $z = d$, because the modes can be highly evanescent. The reference is chosen, such that the forward propagating mode decays as they propagate forwards, and the backward propagating mode decays as they propagate backwards (fig. 1). The combined tangential field $\Psi(z) = [\mathbf{e}_x, \mathbf{e}_y, \mathbf{h}_x, \mathbf{h}_y]^\top$ of eqs. 1.7 and 1.8 is expressed as

$$\Psi(z) = [\overline{\overline{\mathrm{F}}}] \begin{bmatrix} e^{-[\mathrm{D}]\tilde{z}} & \\ & e^{-[\mathrm{D}](\tilde{d}-\tilde{z})} \end{bmatrix} [\mathbf{c}], \tag{1.9}$$

where the eigenvector matrices $[\mathrm{W}]$ and $[\mathrm{V}]$ provide the mode profiles of electric and magnetic fields,

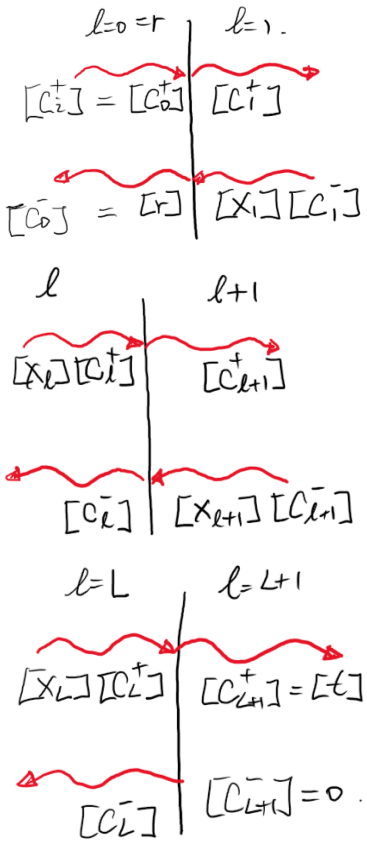

Figure 2: Boundary condition matching

respectively, and have size of $2N \times 2N$, and the mode matrix $[\overline{\overline{\mathrm{F}}}] := \begin{bmatrix} \mathrm{W} & \mathrm{W} \\ -\mathrm{V} & \mathrm{V} \end{bmatrix}$ of size $4N \times 4N$ relates the mode coefficients to the tangential field vectors. The eigenvalue matrix [D] provides the mode propagation index and is a diagonal matrix of size $2N \times 2N$.

1.3 Boundary condition matching

The boundary conditions of classical electromagnetics state that the tangential components of electric and magnetic fields should be continuous at the interfaces; that is, $\Psi(z_l - \epsilon_+) = \Psi(z_l + \epsilon_+)$, where z_l is the interface between layers l and $l+1$. Then, using eq. 1.9, we obtain the following set of equations:

$$\Psi(z_0) = [\overline{\overline{\mathrm{F}_r}}] \begin{bmatrix} [\mathbf{c}_i] \\ [\mathbf{r}] \end{bmatrix} = [\overline{\overline{\mathrm{F}_1}}] \begin{bmatrix} [\mathbf{c}_1^+] \\ [\mathrm{X}_1][\mathbf{c}_1^-] \end{bmatrix}, \tag{1.10a}$$

$$\Psi(z_l) = [\overline{\overline{\mathrm{F}_l}}] \begin{bmatrix} [\mathrm{X}_l][\mathbf{c}_l^+] \\ [\mathbf{c}_l^-] \end{bmatrix} = [\overline{\overline{\mathrm{F}_{l+1}}}] \begin{bmatrix} [\mathbf{c}_{l+1}^+] \\ [\mathrm{X}_{l+1}][\mathbf{c}_{l+1}^-] \end{bmatrix}, \tag{1.10b}$$

$$\Psi(z_L) = [\overline{\overline{\mathrm{F}_L}}] \begin{bmatrix} [\mathrm{X}_L][\mathbf{c}_L^+] \\ [\mathbf{c}_L^-] \end{bmatrix} = [\mathrm{F}_t^+][\mathbf{t}], \tag{1.10c}$$

where $[\mathrm{X}_l] = e^{-[\mathrm{D}_l]\tilde{d}_l}$. Eq. 1.10a applies to the first interface between the substrate and the first layer, and $[\mathbf{c}_i] := [\mathbf{c}_0^+]$ and $[\mathbf{r}] := [\mathbf{c}_0^-]$ are the mode coefficients of incident and reflected waves, respectively. Eq. 1.10b applies to the intermediate interfaces with $l = 1, 2, \ldots, L$, and L is the total number of layers. Eq. 1.10c applies to the last interface between the last layer and the superstrate, and $[\mathbf{t}] := [\mathbf{c}_{L+1}^+]$ is the mode coefficient of the transmitted wave. The subscripts r and t denote to 0 and $L+1$, respectively, such that $[\overline{\overline{\mathrm{F}_r}}] := [\overline{\overline{\mathrm{F}_0}}]$ is the mode matrix of the substrate (reflection-side), and $[\mathrm{F}_r^-] = \begin{bmatrix} [\mathrm{W}_r] \\ [\mathrm{V}_r] \end{bmatrix}$ is the mode matrix of only the backward propagating waves in the substrate. Likewise, $[\mathrm{F}_t^+] = \begin{bmatrix} [\mathrm{W}_t] \\ -[\mathrm{V}_t] \end{bmatrix}$ is the mode matrix of only the forward propagating waves in the superstrate (transmission-side).

Chapter 2

Algorithms

In this chapter, we introduce several techniques that utilize the given set of equations (eq. 1.10) to solve for the transmission/reflection coefficients or mode coefficients. Specifically, the classical transfer matrix method is discussed, and its improved algorithm in terms of numerical stability is covered. Then, the scattering matrix method is discussed. There are other methods including the R-matrix method, impedance or admittance matrix method[1], the H-matrix method, and the enhanced S-matrix method[2]. These methods are analytically identical, but may perform differently in actual numerical calculations.

2.1 Transfer matrix method

First, we examine the transfer matrix in periodic multilayer systems. By inserting $z = 0$ and $z = d$ to eq. 1.9, we obtain

$$\begin{aligned} \Psi(d) &= [\overline{\overline{\mathrm{F}}}] \begin{bmatrix} [\mathrm{X}] & \\ & [\mathrm{I}] \end{bmatrix} \begin{bmatrix} \mathbf{c}^+ \\ \mathbf{c}^- \end{bmatrix} \\ \Psi(0) &= [\overline{\overline{\mathrm{F}}}] \begin{bmatrix} [\mathrm{I}] & \\ & [\mathrm{X}] \end{bmatrix} \begin{bmatrix} \mathbf{c}^+ \\ \mathbf{c}^- \end{bmatrix} \end{aligned} \tag{2.1}$$

which leads to

$$\begin{aligned} \begin{bmatrix} \mathbf{c}^+ \\ \mathbf{c}^- \end{bmatrix} &= \begin{bmatrix} [\mathrm{X}]^{-1} & \\ & [\mathrm{I}] \end{bmatrix} [\overline{\overline{\mathrm{F}}}]^{-1} \Psi(0) \\ &= \begin{bmatrix} [\mathrm{I}] & \\ & [\mathrm{X}]^{-1} \end{bmatrix} [\overline{\overline{\mathrm{F}}}]^{-1} \Psi(d) \end{aligned} \tag{2.2}$$

from which we see that the transfer matrix of a layer relates the tangential fields at both ends of the layer as

$$\begin{aligned} \Psi(d) &= [\overline{\overline{\mathrm{F}}}] \begin{bmatrix} [\mathrm{X}]^{-1} & \\ & [\mathrm{X}] \end{bmatrix} [\overline{\overline{\mathrm{F}}}]^{-1} \Psi(0) \\ &= [\overline{\overline{\mathrm{T}}}] \Psi(0). \end{aligned} \tag{2.3}$$

We note that the transfer matrix $[\overline{\overline{\mathrm{T}}}]$ relates the tangential fields at both ends of the layer.

To solve for the device responses consisting of multiple layers, a set of self-consistent linear equations is constructed as

$$\begin{aligned} [\overline{\overline{\mathrm{F}_r}}] \begin{bmatrix} [\mathbf{c}_i] \\ [\mathbf{r}] \end{bmatrix} &= \prod_{l=1}^{L} \left\{ [\overline{\overline{\mathrm{F}_l}}] \begin{bmatrix} [\mathrm{X}_l]^{-1} & \\ & [\mathrm{X}_l] \end{bmatrix} [\overline{\overline{\mathrm{F}_l}}]^{-1} \right\} [\mathrm{F}_t^+][\mathbf{t}] \\ &= \left\{ \prod_{l=1}^{L} [\overline{\overline{\mathrm{T}_l}}] \right\} [\mathrm{F}_t^+][\mathbf{t}] \\ &= [\overline{\overline{\mathrm{T}_d}}]\ [\mathrm{F}_t^+][\mathbf{t}], \end{aligned} \tag{2.4}$$

where $[\overline{\overline{\mathrm{T}_l}}]$ is the transfer matrix of the layer l, and $[\overline{\overline{\mathrm{T}_d}}]$ is the transfer matrix of the entire device. The transfer matrix method (TMX) solves this set of equations to obtain the transmission and reflection coefficients as the unknowns.

Despite the simplicity of TMX, it has not been widely used in FMM, because the direct evaluation of eq. 2.4 is numerically instable. This instability comes from the evaluation of the inverse phase matrix $[\mathrm{X}]^{-1}$, which is exponentially diverging for evanescent modes with high fourier number and large thickness. The resulting T-matrix is ill-conditioned due to the phase accumulation at large Fourier order and thickness.

Still, for homogeneous cases (i.e., keeping only the zeroth Fourier order), TMX works well. TMX can calculate multiple reflection problems in multilayer systems, e.g., distributed Bragg reflector and anti-reflection layers, and for a single interface, the problem becomes the Fresnel equation.

2.2 Enhanced transmittance matrix method

The enhanced transmittance matrix (ETM) method[3] achieves numerical stability by avoiding the evaluation of $[\mathrm{X}]^{-1}$. In ETM, eq. 2.4 is expressed as

$$\begin{aligned}
[\overline{\overline{\mathrm{F}_r}}]\begin{bmatrix}[\mathbf{c}_i]\\ [\mathbf{r}]\end{bmatrix} &= \left\{\prod_{l=1}^{L}[\overline{\overline{\mathrm{T}^l}}]\right\}[\mathrm{F}_t^+][\mathbf{t}] \\
&= \left\{\prod_{l=1}^{L}[\overline{\overline{\mathrm{T}^l}}]\right\}[\mathrm{G}_{L+1}][\mathbf{t}_{L+1}] \\
&= \left\{\prod_{l=1}^{L-1}[\overline{\overline{\mathrm{T}^l}}]\right\}[\mathrm{F}_L]\begin{bmatrix}[\mathrm{X}_L]^{-1}[\mathrm{A}_L]\\ [\mathrm{X}_L][\mathrm{B}_L]\end{bmatrix}[\mathbf{t}_{L+1}] \\
&= \left\{\prod_{l=1}^{L-1}[\overline{\overline{\mathrm{T}^l}}]\right\}[\mathrm{F}_L]\begin{bmatrix}[\mathrm{I}]\\ [\mathrm{X}_L][\mathrm{B}_L][\mathrm{A}_L]^{-1}[\mathrm{X}_L]\end{bmatrix}[\mathbf{t}_L] \\
&= \left\{\prod_{l=1}^{L-1}[\overline{\overline{\mathrm{T}^l}}]\right\}[\mathrm{G}_L][\mathbf{t}_L] \\
&= \cdots \\
&= [\mathrm{G}_1][\mathbf{t}_1],
\end{aligned} \tag{2.5}$$

where $[\mathbf{t}] = [\mathbf{t}_{L+1}]$ and $[\mathrm{G}_{L+1}] = [\mathrm{F}_t^+]$, and the derivation utilizes the following expressions:

$$[\mathrm{G}_l] = [\overline{\overline{\mathrm{F}_l}}]\begin{bmatrix}[\mathrm{I}]\\ [\mathrm{X}_l][\mathrm{B}_l][\mathrm{A}_l]^{-1}[\mathrm{X}_l]\end{bmatrix} \tag{2.6a}$$

$$\begin{bmatrix}[\mathrm{A}_l]\\ [\mathrm{B}_l]\end{bmatrix} = [\overline{\overline{\mathrm{F}_l}}]^{-1}[\mathrm{G}_{l+1}] \tag{2.6b}$$

$$[\mathbf{t}_{l+1}] = [\mathrm{A}_l]^{-1}[\mathrm{X}_l][\mathbf{t}_l] \tag{2.6c}$$

$$[\mathbf{c}_l] = \begin{bmatrix}[\mathrm{I}]\\ [\mathrm{X}_l][\mathrm{B}_l][\mathrm{A}_l]^{-1}[\mathrm{X}_l]\end{bmatrix}[\mathbf{t}_l] \tag{2.6d}$$

The ETM routine is recursive and performed backwards, where $[\mathbf{r}]$ and $[\mathbf{t}_1]$ are obtained first, and the other mode coefficients $[\mathbf{t}_l]$ are obtained iteratively using eq. 2.6c and eq. 2.6d.

Summary. ETM is stable while fast. ETM is memory-inefficient, because the intermediate matrices [A] and [B] are saved for future calculations of mode coefficients. Also, ETM can be inefficient when the mode coefficient in certain layer is required, because the mode coefficients in all layers are calculated, which are scanned iteratively.

2.3 Scattering matrix method

Now, we examine the scattering (S-) matrix method (SMX), which has been most widely used in FMM.

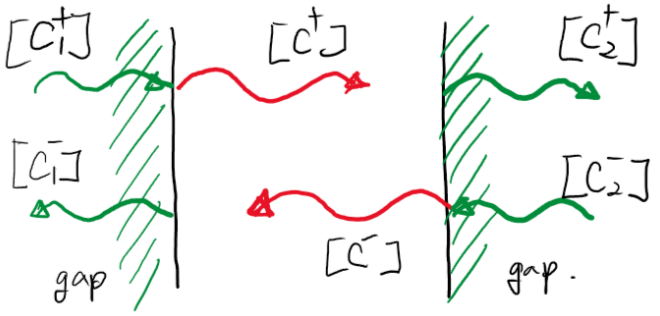

Figure 3: Modes of S-matrix of a layer between gap layers.

An S-matrix of a layer is defined as (fig. 3)

$$\begin{bmatrix}\mathbf{c}_1^-\\ \mathbf{c}_2^+\end{bmatrix} = \begin{bmatrix}[\mathrm{S}_{11}] & [\mathrm{S}_{12}]\\ [\mathrm{S}_{21}] & [\mathrm{S}_{22}]\end{bmatrix}\begin{bmatrix}\mathbf{c}_1^+\\ \mathbf{c}_2^-\end{bmatrix} = [\bar{\mathrm{S}}]\begin{bmatrix}\mathbf{c}_1^+\\ \mathbf{c}_2^-\end{bmatrix}, \tag{2.7}$$

where the mode coefficients $[\mathbf{c}_{1/2}^{\pm}]$ are defined in the external region of the layer, which is contrary to the mode coefficients discussed before ($[\mathbf{c}^{\pm}]$). The subscripts 1 and 2 denote the left and right regions, respectively. The S-matrix elements $[\mathrm{S}_{ij}]$ relate the mode coefficients, such that $[\mathrm{S}_{11}]$ is the forward reflectance, $[\mathrm{S}_{21}]$ is the forward transmittance, $[\mathrm{S}_{12}]$ is the backward transmittance, and $[\mathrm{S}_{22}]$ is the backward reflectance. Note that an S-matrix depends on the external layers 1 and 2. To improve the memory and calculation efficiencies, an artificial gap medium of zero thickness is assumed to exist between the layers. Then, the mode coefficients are defined at the gap medium. The layer becomes symmetric, such that $[\mathrm{S}_{11}] = [\mathrm{S}_{22}]$ and $[\mathrm{S}_{12}] = [\mathrm{S}_{21}]$, and only half of the matrices need to be calculated and saved.

Using the eigen-matrices, the S-matrix elements of a layer are calculated as

$$\begin{aligned}
[\mathrm{S}_{11}] &= [\mathrm{C}]^{-1}\left([\mathrm{X}][\mathrm{B}][\mathrm{A}]^{-1}[\mathrm{X}][\mathrm{A}] - [\mathrm{B}]\right) \\
[\mathrm{S}_{12}] &= [\mathrm{C}]^{-1}[\mathrm{X}]\left([\mathrm{A}] - [\mathrm{B}][\mathrm{A}]^{-1}[\mathrm{B}]\right) \\
[\mathrm{S}_{21}] &= [\mathrm{S}_{12}] \\
[\mathrm{S}_{22}] &= [\mathrm{S}_{11}],
\end{aligned} \tag{2.8}$$

where $[\mathrm{A}] = [\mathrm{W}]^{-1}[\mathrm{W}_g] + [\mathrm{V}]^{-1}[\mathrm{V}_g]$, $[\mathrm{B}] = [\mathrm{W}]^{-1}[\mathrm{W}_g] - [\mathrm{V}]^{-1}[\mathrm{V}_g]$, $[\mathrm{C}] = [\mathrm{A}] - [\mathrm{X}][\mathrm{B}][\mathrm{A}]^{-1}[\mathrm{X}][\mathrm{B}]$, and $[\mathrm{X}] = e^{-[\mathrm{D}]\tilde{d}}$. The S-matrix of reflection-side layer is given as

$$\begin{aligned}
[\mathrm{S}_{11}] &= -[\mathrm{A}]^{-1}[\mathrm{B}] \\
[\mathrm{S}_{12}] &= 2[\mathrm{A}]^{-1} \\
[\mathrm{S}_{21}] &= 0.5([\mathrm{A}] - [\mathrm{B}][\mathrm{A}]^{-1}[\mathrm{B}]) \\
[\mathrm{S}_{22}] &= [\mathrm{B}][\mathrm{A}]^{-1}
\end{aligned} \tag{2.9}$$

where $[\mathrm{A}] = [\mathrm{W}_g]^{-1}[\mathrm{W}_r] + [\mathrm{V}_g]^{-1}[\mathrm{V}_r]$, and $[\mathrm{B}] = [\mathrm{W}_g]^{-1}[\mathrm{W}_r] - [\mathrm{V}_g]^{-1}[\mathrm{V}_r]$. The S-matrix of transmission side layer is given as

$$\begin{aligned}
[\mathrm{S}_{11}] &= [\mathrm{B}][\mathrm{A}]^{-1} \\
[\mathrm{S}_{12}] &= 0.5([\mathrm{A}] - [\mathrm{B}][\mathrm{A}]^{-1}[\mathrm{B}]) \\
[\mathrm{S}_{21}] &= 2[\mathrm{A}]^{-1} \\
[\mathrm{S}_{22}] &= -[\mathrm{A}]^{-1}[\mathrm{B}]
\end{aligned} \tag{2.10}$$

where $[\mathrm{A}] = [\mathrm{W}_g]^{-1}[\mathrm{W}_t] + [\mathrm{V}_g]^{-1}[\mathrm{V}_t]$, and $[\mathrm{B}] = [\mathrm{W}_g]^{-1}[\mathrm{W}_t] - [\mathrm{V}_g]^{-1}[\mathrm{V}_t]$.

The global S-matrix $[\bar{\bar{\mathrm{S}}}]$ relates the incident, reflected, and transmitted waves as

$$\begin{aligned}
\begin{bmatrix} \mathbf{c}_r^- \\ \mathbf{c}_t^+ \end{bmatrix} &= [\bar{\bar{\mathrm{S}}}] \begin{bmatrix} \mathbf{c}_0^+ \\ 0 \end{bmatrix} \\
&= \left\{ [\overline{\overline{\mathrm{S}^0}}] \otimes [\overline{\overline{\mathrm{S}^1}}] \otimes \cdots \otimes [\overline{\overline{\mathrm{S}^L}}] \otimes [\overline{\overline{\mathrm{S}^t}}] \right\} \begin{bmatrix} \mathbf{c}_0^+ \\ 0 \end{bmatrix} \\
&= \left\{ [\overline{\overline{\mathrm{S}^0}}] \otimes [\overline{\overline{\mathrm{S}^d}}] \otimes [\overline{\overline{\mathrm{S}^t}}] \right\} \begin{bmatrix} \mathbf{c}_i^+ \\ 0 \end{bmatrix},
\end{aligned} \tag{2.11}$$

where the redheffer star product $[\overline{\overline{\mathrm{S}^{\mathrm{AB}}}}] = [\overline{\overline{\mathrm{S}^{\mathrm{A}}}}] \otimes [\overline{\overline{\mathrm{S}^{\mathrm{B}}}}]$ is defined as

$$\begin{aligned}
[\mathrm{S}_{11}^{\mathrm{AB}}] &= [\mathrm{S}_{11}^{\mathrm{A}}] + [\mathrm{S}_{12}^{\mathrm{A}}] \left([\mathrm{I}] - [\mathrm{S}_{11}^{\mathrm{B}}][\mathrm{S}_{22}^{\mathrm{A}}] \right)^{-1} [\mathrm{S}_{11}^{\mathrm{B}}][\mathrm{S}_{21}^{\mathrm{A}}] \\
[\mathrm{S}_{12}^{\mathrm{AB}}] &= [\mathrm{S}_{12}^{\mathrm{A}}] \left([\mathrm{I}] - [\mathrm{S}_{11}^{\mathrm{B}}][\mathrm{S}_{22}^{\mathrm{A}}] \right)^{-1} [\mathrm{S}_{12}^{\mathrm{B}}] \\
[\mathrm{S}_{21}^{\mathrm{AB}}] &= [\mathrm{S}_{21}^{\mathrm{B}}] \left([\mathrm{I}] - [\mathrm{S}_{22}^{\mathrm{A}}][\mathrm{S}_{11}^{\mathrm{B}}] \right)^{-1} [\mathrm{S}_{21}^{\mathrm{A}}] \\
[\mathrm{S}_{22}^{\mathrm{AB}}] &= [\mathrm{S}_{22}^{\mathrm{B}}] + [\mathrm{S}_{21}^{\mathrm{B}}] \left([\mathrm{I}] - [\mathrm{S}_{22}^{\mathrm{A}}][\mathrm{S}_{11}^{\mathrm{B}}] \right)^{-1} [\mathrm{S}_{22}^{\mathrm{A}}][\mathrm{S}_{12}^{\mathrm{B}}]
\end{aligned} \tag{2.12}$$

The transmission and reflection coefficients are obtained as $[\mathbf{t}] = [\mathbf{c}_{L+1}^+] = [\mathrm{S}_{21}][\mathbf{c}_0^+]$ and $[\mathbf{r}] = [\mathbf{c}_0^-] = [\mathrm{S}_{11}][\mathbf{c}_0^+]$, and $[\mathbf{c}_0^+] = [\mathrm{W}_0]^{-1}[\mathbf{s}]$ is the incident mode coefficient.

Summary. SMX provides a stable formulation, but is slower than other formulations due to more involved matrix operations. Similar to TMX, several layers can be clustered and reused in different calculations, which may significantly improve the calculation efficiency.

Chapter 3

Source excitations

3.1 Planewaves

Because FMM is formulated upon the planewave basis, planewaves are easily implemented, so most FMM calculations are performed at planewave incidences. A planewave $\mathbf{E} = \mathbf{E}_0 e^{i\mathbf{k}\cdot\mathbf{r}}$ can be defined by its wavevector $\mathbf{k}$ and its electric field vector $\mathbf{E}_0$. Conveniently, the incident wavevector can be defined from the inclination and azimuthal angles as

$$\begin{aligned} k_x &= k_0 \sin\theta\cos\phi \\ k_y &= k_0 \sin\theta\sin\phi \\ k_z &= k_0 \cos\theta, \end{aligned} \tag{3.1}$$

where, in fact, the incident wavevector is determined from the bloch periodicity as $k_x = \beta_x$ and $k_y = \beta_y$. The electric field vector needs only be perpendicular to the wavevector. Convieniently, we may construct a linearly-polarized planewave as

$$\hat{e} = A_p\hat{n}_p + A_s\hat{n}_s = \cos\psi\hat{n}_p + \sin\psi\hat{n}_s, \tag{3.2}$$

where the parameter ψ to control the polarization states of linearly-polarized light, and the p- and s-polarization unit vectors are

$$\begin{aligned} \hat{n}_p &= \begin{bmatrix} \cos\theta\cos\phi \\ \cos\theta\sin\phi \\ -\sin\theta \end{bmatrix} \\ \hat{n}_s &= \begin{bmatrix} -\sin\phi \\ \cos\phi \\ 0 \end{bmatrix} \end{aligned} \tag{3.3}$$

Note, we can use the parameter η to control the ellipticity for general elliptical polarization states (++TBD).

Using the polarization vector, the source vector is constructed as

$$[\mathbf{s}] = \begin{bmatrix} e_x\delta_{m0}\delta_{n0} \\ e_y\delta_{m0}\delta_{n0} \\ h_x\delta_{m0}\delta_{n0} \\ h_y\delta_{m0}\delta_{n0} \end{bmatrix}, \tag{3.4}$$

where (m, n) is the fourier index in x- and y-direction, and typically, only $(0, 0)$ mode is considered.

Power calculation

For planewave excitations, the transmission and reflection at the far fields can be efficiently calculated. Assuming homogeneous substrate and superstrate, the eigen-matrix [W] is the identity matrix, so the tangential electric fields directly relate to the mode coefficients $[\mathbf{t}_{x/y}]$ and $[\mathbf{r}_{x/y}]$. The longitudinal components are obtained as

$$\begin{aligned} [\mathbf{t}_z] &= -[\mathrm{K}_z]^{-1}([\mathrm{K}_x][\mathbf{t}_x] + [\mathrm{K}_y][\mathbf{t}_y]) \\ [\mathbf{r}_z] &= -[\mathrm{K}_z]^{-1}([\mathrm{K}_x][\mathbf{r}_x] + [\mathrm{K}_y][\mathbf{r}_y]). \end{aligned} \tag{3.5}$$

From the xyz-basis, the p/s-basis mode coefficients are obtained as

$$\begin{aligned} [\mathbf{t}_p] &= [\cos\phi]([\mathrm{K}_{2z}][\mathbf{t}_x] - [\mathrm{K}_x][\mathbf{t}_z]) \\ &\quad - [\sin\phi]([\mathrm{K}_y][\mathbf{t}_z] - [\mathrm{K}_{2z}][\mathbf{t}_y]) \\ [\mathbf{t}_s] &= [\cos\phi][\mathbf{t}_y] - [\sin\phi][\mathbf{t}_x] \\ [\mathbf{r}_p] &= [\cos\phi]([\mathrm{K}_{1z}][\mathbf{r}_x] - [\mathrm{K}_x][\mathbf{r}_z]) \\ &\quad - [\sin\phi]([\mathrm{K}_y][\mathbf{r}_z] - [\mathrm{K}_{1z}][\mathbf{r}_y]) \\ [\mathbf{r}_s] &= [\cos\phi][\mathbf{r}_y] - [\sin\phi][\mathbf{r}_x], \end{aligned} \tag{3.6}$$

where $[\tan\phi] = [\mathrm{K}_y][\mathrm{K}_x]^{-1}$. For completeness, the helicity-basis mode coefficients are obtained as

$$\begin{aligned} [\mathbf{t}_+] &= ([\mathbf{t}_p] - i[\mathbf{t}_s])/\sqrt{2} \\ [\mathbf{t}_-] &= ([\mathbf{t}_p] + i[\mathbf{t}_s])/\sqrt{2} \\ [\mathbf{r}_+] &= ([\mathbf{r}_p] - i[\mathbf{r}_s])/\sqrt{2} \\ [\mathbf{r}_-] &= ([\mathbf{r}_p] + i[\mathbf{r}_s])/\sqrt{2} \end{aligned} \tag{3.7}$$

Finally, the transmittance and reflectance are obtained as

$$\begin{aligned}[\mathrm{T}_q] &= \mathrm{Re}([\mathrm{K}_{2z}])|[\mathbf{t}_q]|^2/I_z \\ [\mathrm{R}_q] &= \mathrm{Re}([\mathrm{K}_{1z}])|[\mathbf{r}_q]|^2/I_z,\end{aligned} \tag{3.8}$$

where $I_z = n_1 \cos\theta$ is the incident power flux. Note that it applies to $E_0 = 1$ and nonmagnetic system with $\mu = 1$ or $\eta = 1/n$.

3.2 Aperiodic sources in periodic systems

Brillouin zone integration

Although FMM has commonly deals with planewave excitations, FMM can implement nonplanar excitations. In this section, we analyze the nonplanar excitations and their spatial harmonics [**s**], which appear in

$$\begin{aligned}\begin{bmatrix}[\mathrm{W}_r] & [\mathrm{W}_r] \\ -[\mathrm{V}_r] & [\mathrm{V}_r]\end{bmatrix}\begin{bmatrix}\mathbf{c}_0^+ \\ \mathbf{c}_0^-\end{bmatrix} &= [\overline{\overline{\mathrm{F}_r}}][\mathbf{c}_0] \\ &= [\overline{\overline{\mathrm{F}_i^+}}][\mathbf{c}_i^+] + [\overline{\overline{\mathrm{F}_r^-}}][\mathbf{c}_r^-] \\ &= [\mathbf{s}] + [\overline{\overline{\mathrm{F}_r^-}}][\mathbf{r}],\end{aligned} \tag{3.9}$$

According to the angular spectrum representation of electromagnetic fields, the electric fields in free space can be expressed as

$$\mathbf{E}(x,y,z) = \int \mathbf{e}(k_x,k_y)e^{ik_z z}e^{i\mathbf{k}_\parallel\cdot\mathbf{r}_\parallel}d^2\mathbf{k}_\parallel, \tag{3.10}$$

where $\mathbf{k}_\parallel = (k_x, k_y)$, $\mathbf{r}_\parallel = (x, y)$, and $d^2\mathbf{k}_\parallel = dk_x dk_y$. The angular spectrum is obtained as

$$\mathbf{e}(k_x,k_y) = \frac{1}{4\pi^2 E_0}\int \mathbf{E}(x,y,z=z_0)e^{-i\mathbf{k}_\parallel\cdot\mathbf{r}_\parallel}d^2\mathbf{r}_\parallel, \tag{3.11}$$

where $\mathbf{E}(x, y, z_0)$ is the target field at the reference plane at $z = z_0$.

contrary to FMM calculations with discrete fourier space grids, the angular spectrum is given at continuous fourier space.

Brillouin zone integration

TODO: schematics

3.3 Localized optical beams

Simple Gaussian beams

the electric fields of a Gaussian beam at the reference plane is given as

$$\mathbf{E}(x,y,z_0) = \mathbf{E}_0 e^{-w_0^2 r^2} \tag{3.12}$$

where w_0 is the Gaussian beam waist. The angular spectrum is given as

$$\mathbf{e}(k_x,k_y) = \frac{w_0^2}{4\pi}e^{-\frac{w_0^2}{4}(k_x^2+k_y^2)} \tag{3.13}$$

Airy disks

3.4 Vector optical beams

3.5 Dipole radiations

Chapter 4

Improved implementations

4.1 Homogeneous layers

calculations involved with homogeneous layers can be significantly accelerated

analytic calculation of eigen-matrices

the FMM calculation

the eigen-decomposition

the eigen-matrices [W], [V], and [D] can be calculated analytically

Also, the eigen-matrices

4.2 Fourier space grid truncation

In the classical formulation, the Fourier space is defined in a rectangular grids (Sec. 1.2).

In the square grids, the size of associated eigen-matrices is $2N \times 2N$, where $N = (2n_{\max} + 1)^2$, where $n_{\max}$ is the maximum Fourier order in x and y directions.

By truncating the Fourier space grids $i^{2p} + j^{2p} \leq n_{\max}^2$

that is, circular Fourier grids for $p = 1$

for $p = 1$ case, the size of Fourier space is reduced by a factor of $\pi/4 \approx 0.785$ The size of eigen-matrices is reduced by a factor of ≈ 0.617

the reduced involved matrix size significantly reduces the calculation time taken for eigen-decompositior

In addition, the time complexity of matrix multiplication of square matrices of size $N \times N$ is $O(N^3)$

TODO: schematics

4.3 Fast Fourier factorization

It has been pointed out that TM mode converges much slower than TE mode in 1D gratings, and this issue was treated using the Li's inverse rule[4–6].

To understand this problem, the Fourier transform of a function of the form $f = g \cdot h$ should be revisited. Conventionally, the Fourier transform is given as $[f] = [\![g]\!][h]$, where $[\cdot]$ and $[\![\cdot]\!]$ are the Fourier coefficient matrix and its convolution. For truncated Fourier series, the convergence is slow, when f is continuous, but g and h are discontinuous. This issue occurs for the FMM formulations. The boundary condition states that the tangential part of electric fields ($\mathbf{E}$) and the normal part of electric displacement fields ($\mathbf{D}$) are continuous across boundaries. For isotropic materials, $\mathbf{D}(\mathbf{r}) = \varepsilon(\mathbf{r})\mathbf{E}(\mathbf{r})$ and the permittivity distribution $\varepsilon(\mathbf{r})$ is discontinous across the material boundaries.

According to the fast Fourier factorization (FFF) rule, the Fourier transform of $f = g \cdot h$ converges faster in the form of $[f] = [\![g^{-1}]\!]^{-1}[h]$. To make use of this rule, we decompose the electric fields as the normal and tangential components as $\mathbf{E} = \mathbf{E}_n + \mathbf{E}_t$, and they can be expressed as $\mathbf{E}_n = \hat{n}(\hat{n}\cdot\mathbf{E}) = \hat{n}\hat{n}\cdot\mathbf{E}$ and $\mathbf{E}_t = (1 - \hat{n}\hat{n}) \cdot \mathbf{E}$, where $\hat{n}(\mathbf{r})$ is the normal vector field (NVF).

TODO: derivation!

$[\mathbf{d}_n] = [\![\frac{1}{\varepsilon}]\!][\mathbf{e}_n]$

Finally, we obtain the modified Q-matrix as

$$[\mathrm{Q}] = \begin{bmatrix} [\mathrm{K}_x][\mathrm{K}_y] - \Delta_{xy} & [\![\varepsilon]\!] - [\mathrm{K}_x][\mathrm{K}_x] - \Delta_{yy} \\ [\mathrm{K}_y][\mathrm{K}_y] - [\![\varepsilon]\!] + \Delta_{xx} & -[\mathrm{K}_y][\mathrm{K}_x] + \Delta_{xy} \end{bmatrix}, \tag{4.1}$$

Q-matrix with FFF

$$[\mathrm{Q}] = \begin{bmatrix} [\mathrm{K}_x][\mathrm{K}_y] - \Delta_{xy} & [\![\varepsilon]\!] - [\mathrm{K}_x][\mathrm{K}_x] - \Delta_{yy} \\ [\mathrm{K}_y][\mathrm{K}_y] - [\![\varepsilon]\!] + \Delta_{xx} & -[\mathrm{K}_y][\mathrm{K}_x] + \Delta_{xy} \end{bmatrix}, \tag{4.2}$$

where $\Delta_{ij} = [\![\Delta\varepsilon]\!][\![n_i n_j]\!]$, and $[\![\Delta\varepsilon]\!] = [\![\varepsilon]\!] - [\![\frac{1}{\varepsilon}]\!]^{-1}$

$\Delta_{ij} = \frac{1}{2}\left(\llbracket\varepsilon\rrbracket\llbracket n_i n_j\rrbracket + \llbracket n_i n_j\rrbracket\llbracket\varepsilon\rrbracket\right)$

In general, the choice of NVF is not unique, and there are several ways to generate NVF[7–9].

Our method?

[10]

4.4 Symmetry: 1D lattice

1D grating, ++ By assuming the incident wavevector lies in the xz-plane, i.e., $k_y = 0$, the PQ matrices (eq. 1.6a) are further simplified. For s-polarized mode (TE), the PQ matrices are simplified as:

$$\begin{aligned}[\mathrm{P}] &= -\llbracket\mu_{xx}\rrbracket \\ [\mathrm{Q}] &= \llbracket\varepsilon_{yy}\rrbracket - [\mathrm{K}_x]\llbracket\mu_{zz}\rrbracket[\mathrm{K}_x]\end{aligned} \tag{4.3}$$

For p-polarized mode (TM),

$$\begin{aligned}[\mathrm{P}] &= \llbracket\mu_{yy}\rrbracket - [\mathrm{K}_x]\llbracket\varepsilon_{zz}\rrbracket^{-1}[\mathrm{K}_x] \\ [\mathrm{Q}] &= -\llbracket\varepsilon_{xx}\rrbracket\end{aligned} \tag{4.4}$$

For completeness, the PQ matrices for p-polarized mode with FFF are given as

$$\begin{aligned}[\mathrm{P}] &= \llbracket\mu_{yy}\rrbracket - [\mathrm{K}_x]\llbracket\varepsilon_{zz}\rrbracket^{-1}[\mathrm{K}_x] \\ [\mathrm{Q}] &= -\llbracket\tfrac{1}{\varepsilon_{xx}}\rrbracket^{-1}\end{aligned} \tag{4.5}$$

which has been investigated for the Li's inverse rule[4;5].

4.5 Symmetry: 2D lattice

2D FMM

[11–14]

[15]

the NVF may need to be generated to adhere to the symmetries[15]

Bibliography

[1] L. Li, "Formulation and comparison of two recursive matrix algorithms for modeling layered diffraction gratings," *Journal of the Optical Society of America A*, vol. 13, no. 5, pp. 1024–1035, 1996.

[2] E. L. Tan, "Note on formulation of the enhanced scattering-(transmittance-) matrix approach," *Journal of the Optical Society of America A*, vol. 19, no. 6, pp. 1157–1161, 2002.

[3] M. Moharam, E. B. Grann, D. A. Pommet, and T. Gaylord, "Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings," *Journal of the Optical Society of America A*, vol. 12, no. 5, pp. 1068–1076, 1995.

[4] L. Li and C. W. Haggans, "Convergence of the coupled-wave method for metallic lamellar diffraction gratings," *Journal of the Optical Society of America A*, vol. 10, no. 6, pp. 1184–1189, 1993.

[5] P. Lalanne and G. M. Morris, "Highly improved convergence of the coupled-wave method for tm polarization," *Journal of the Optical Society of America A*, vol. 13, no. 4, pp. 779–784, 1996.

[6] G. Granet and B. Guizal, "Efficient implementation of the coupled-wave method for metallic lamellar gratings in tm polarization," *Journal of the Optical Society of America A*, vol. 13, no. 5, pp. 1019–1023, 1996.

[7] T. Schuster, J. Ruoff, N. Kerwien, S. Rafler, and W. Osten, "Normal vector method for convergence improvement using the rcwa for crossed gratings," *Journal of the Optical Society of America A*, vol. 24, no. 9, pp. 2880–2890, 2007.

[8] P. Götz, T. Schuster, K. Frenner, S. Rafler, and W. Osten, "Normal vector method for the rcwa with automated vector field generation," *Optics express*, vol. 16, no. 22, pp. 17295–17301, 2008.

[9] R. Antos, "Fourier factorization with complex polarization bases in modeling optics of discontinuous biperiodic structures," *Optics express*, vol. 17, no. 9, pp. 7269–7274, 2009.

[10] M. Weismann, D. F. Gallagher, and N. C. Panoiu, "Accurate near-field calculation in the rigorous coupled-wave analysis method," *Journal of Optics*, vol. 17, no. 12, p. 125612, 2015.

[11] C. Zhou and L. Li, "Formulation of the fourier modal method for symmetric crossed gratings in symmetric mountings," *Journal of Optics A: Pure and Applied Optics*, vol. 6, no. 1, p. 43, 2003.

[12] B. Bai and L. Li, "Group-theoretic approach to enhancing the fourier modal method for crossed gratings with one or two reflection symmetries," *Journal of Optics A: Pure and Applied Optics*, vol. 7, no. 7, p. 271, 2005.

[13] B. Bai and L. Li, "Group-theoretic approach to the enhancement of the fourier modal method for crossed gratings: C2 symmetry case," *Journal of the Optical Society of America A*, vol. 22, no. 4, pp. 654–661, 2005.

[14] B. Bai and L. Li, "Group-theoretic approach to enhancing the fourier modal method for crossed gratings with square symmetry," *Journal of the Optical Society of America A*, vol. 23, no. 3, pp. 572–580, 2006.

[15] J. Bischoff, "Formulation of the normal vector rcwa for symmetric crossed gratings in symmetric mountings," *Journal of the Optical Society of America A*, vol. 27, no. 5, pp. 1024–1031, 2010.

[16] V. Liu and S. Fan, "S4: A free electromagnetic solver for layered periodic structures," *Computer Physics Communications*, vol. 183, no. 10, pp. 2233–2244, 2012.

[17] J. P. Hugonin and P. Lalanne, "Reticolo software for grating analysis," *arXiv preprint arXiv:2101.00901*, 2021.

[18] G. Yoon and J. Rho, "Maxim: Metasurfaces-oriented electromagnetic wave simulation software with intuitive graphical user interfaces," *Computer Physics Communications*, vol. 264, p. 107846, 2021.

[19] S. Colburn and A. Majumdar, "Inverse design and flexible parameterization of meta-optics using algorithmic differentiation," *Communications Physics*, vol. 4, no. 1, p. 65, 2021.

[20] C. Kim and B. Lee, "Torcwa: Gpu-accelerated fourier modal method and gradient-based optimization for metasurface design," *Computer Physics Communications*, vol. 282, p. 108552, 2023.

[21] Y. Kim, A. W. Jung, S. Kim, K. Octavian, D. Heo, C. Park, J. Shin, S. Nam, C. Park, J. Park, *et al.*, "Meent: Differentiable electromagnetic simulator for machine learning," *arXiv preprint arXiv:2406.12904*, 2024.

[22] M. F. Schubert and A. M. Hammond, "Fourier modal method for inverse design of metasurface-enhanced micro-leds," 2023.

[23] J. Bischoff and K. Hehl, "Perturbation approach applied to modal diffraction methods," *Journal of the Optical Society of America A*, vol. 28, no. 5, pp. 859–867, 2011.

[24] J. Xu and M. D. Charlton, "Gpu libraries speed performance analysis for rcwa simulation matrix operations," in *Physics and Simulation of Optoelectronic Devices XXXI*, vol. 12415, pp. 151–158, SPIE, 2023.

Appendix A

List of FMM software

Table A.1: List of FMM software. FFF: fast fourier factorization; FSGT: fourier space grid truncation; LSE: localized source excitation; AD: automatic differentiation. *a*: proprietary software.

	algorithm	FFF	FSGT	LSE	AD	interface	code	ref
S4	SMX	✓				C	Github	[16]
RETICOLO	ETM?					MATLAB	Link	[17]
MAXIM	SMX?					GUI	Link	[18]
RCWA-TF	SMX				Tensorflow	Python	Github	[19]
TORCWA	SMX				JAX	Python	Github	[20]
MEENT	ETM				JAX	Python	Github	[21]
FMMAX	SMX	✓		✓	JAX	Python	Github	[22]
Ansys Lumerical[a]						GUI		Link
Synopsys[a]		✓				GUI		Link
OmniSim[a]		✓				GUI		Link

Appendix B

List of the expressions of PQ matrices

B.1 Inhomogeneous layers

Isotropic, nonmagnetic layers

$$
\begin{aligned}
[\mathrm{P}] &= \begin{bmatrix} [\mathrm{K}_x]\llbracket\varepsilon\rrbracket^{-1}[\mathrm{K}_y] & [\mathrm{I}] - [\mathrm{K}_x]\llbracket\varepsilon\rrbracket^{-1}[\mathrm{K}_x] \\ [\mathrm{K}_y]\llbracket\varepsilon\rrbracket^{-1}[\mathrm{K}_y] - [\mathrm{I}] & -[\mathrm{K}_y]\llbracket\varepsilon\rrbracket^{-1}[\mathrm{K}_x] \end{bmatrix} \\
[\mathrm{Q}] &= \begin{bmatrix} [\mathrm{K}_x][\mathrm{K}_y] & \llbracket\varepsilon\rrbracket - [\mathrm{K}_x][\mathrm{K}_x] \\ [\mathrm{K}_y][\mathrm{K}_y] - \llbracket\varepsilon\rrbracket & -[\mathrm{K}_y][\mathrm{K}_x] \end{bmatrix}
\end{aligned}
\tag{B.1}
$$

Isotropic, magnetic layers

$$
\begin{aligned}
[\mathrm{P}] &= \begin{bmatrix} [\mathrm{K}_x]\llbracket\varepsilon\rrbracket^{-1}[\mathrm{K}_y] & \llbracket\mu\rrbracket - [\mathrm{K}_x]\llbracket\varepsilon\rrbracket^{-1}[\mathrm{K}_x] \\ [\mathrm{K}_y]\llbracket\varepsilon\rrbracket^{-1}[\mathrm{K}_y] - \llbracket\mu\rrbracket & -[\mathrm{K}_y]\llbracket\varepsilon\rrbracket^{-1}[\mathrm{K}_x] \end{bmatrix} \\
[\mathrm{Q}] &= \begin{bmatrix} [\mathrm{K}_x]\llbracket\mu\rrbracket^{-1}[\mathrm{K}_y] & \llbracket\varepsilon\rrbracket - [\mathrm{K}_x]\llbracket\mu\rrbracket^{-1}[\mathrm{K}_x] \\ [\mathrm{K}_y]\llbracket\mu\rrbracket^{-1}[\mathrm{K}_y] - \llbracket\varepsilon\rrbracket & -[\mathrm{K}_y]\llbracket\mu\rrbracket^{-1}[\mathrm{K}_x] \end{bmatrix}
\end{aligned}
\tag{B.2}
$$

Anisotropic, nonmagnetic layers

$$
\begin{aligned}
[\mathrm{P}] &= \begin{bmatrix} [\mathrm{K}_x]\llbracket\varepsilon_{zz}\rrbracket^{-1}[\mathrm{K}_y] & [\mathrm{I}] - [\mathrm{K}_x]\llbracket\varepsilon_{zz}\rrbracket^{-1}[\mathrm{K}_x] \\ [\mathrm{K}_y]\llbracket\varepsilon_{zz}\rrbracket^{-1}[\mathrm{K}_y] - [\mathrm{I}] & -[\mathrm{K}_y]\llbracket\varepsilon_{zz}\rrbracket^{-1}[\mathrm{K}_x] \end{bmatrix} \\
[\mathrm{Q}] &= \begin{bmatrix} [\mathrm{K}_x][\mathrm{K}_y] & \llbracket\varepsilon_{yy}\rrbracket - [\mathrm{K}_x][\mathrm{K}_x] \\ [\mathrm{K}_y][\mathrm{K}_y] - \llbracket\varepsilon_{xx}\rrbracket & -[\mathrm{K}_y][\mathrm{K}_x] \end{bmatrix}
\end{aligned}
\tag{B.3}
$$

Anisotropic, magnetic layers

$$
\begin{aligned}
[\mathrm{P}] &= \begin{bmatrix} [\mathrm{K}_x]\llbracket\varepsilon_{zz}\rrbracket^{-1}[\mathrm{K}_y] & \llbracket\mu_{yy}\rrbracket - [\mathrm{K}_x]\llbracket\varepsilon_{zz}\rrbracket^{-1}[\mathrm{K}_x] \\ [\mathrm{K}_y]\llbracket\varepsilon_{zz}\rrbracket^{-1}[\mathrm{K}_y] - \llbracket\mu_{xx}\rrbracket & -[\mathrm{K}_y]\llbracket\varepsilon_{zz}\rrbracket^{-1}[\mathrm{K}_x] \end{bmatrix} \\
[\mathrm{Q}] &= \begin{bmatrix} [\mathrm{K}_x]\llbracket\mu_{zz}\rrbracket^{-1}[\mathrm{K}_y] & \llbracket\varepsilon_{yy}\rrbracket - [\mathrm{K}_x]\llbracket\mu_{zz}\rrbracket^{-1}[\mathrm{K}_x] \\ [\mathrm{K}_y]\llbracket\mu_{zz}\rrbracket^{-1}[\mathrm{K}_y] - \llbracket\varepsilon_{xx}\rrbracket & -[\mathrm{K}_y]\llbracket\mu_{zz}\rrbracket^{-1}[\mathrm{K}_x] \end{bmatrix}
\end{aligned}
\tag{B.4}
$$

B.2 Homogeneous layers

Isotropic, homogeneous layers

$$\begin{aligned}[\mathrm{P}] &= \frac{1}{\varepsilon}\begin{bmatrix}[\mathrm{K}_x][\mathrm{K}_y] & \varepsilon\mu - [\mathrm{K}_x]^2\\ [\mathrm{K}_y]^2 - \varepsilon\mu & -[\mathrm{K}_y][\mathrm{K}_x]\end{bmatrix}\\ [\mathrm{Q}] &= \frac{1}{\mu}\begin{bmatrix}[\mathrm{K}_x][\mathrm{K}_y] & \varepsilon\mu - [\mathrm{K}_x]^2\\ [\mathrm{K}_y]^2 - \varepsilon\mu & -[\mathrm{K}_y][\mathrm{K}_x]\end{bmatrix} = \frac{\varepsilon}{\mu}[\mathrm{P}]\end{aligned} \tag{B.5}$$

For an isotropic homogeneous layer, the eigen-matrices are obtained as

$$\begin{aligned}[\mathrm{W}] &= \begin{bmatrix}[\mathrm{I}] & \\ & [\mathrm{I}]\end{bmatrix}\\ [\mathrm{D}] &= \begin{bmatrix}-i[\mathrm{K}_z] & \\ & -i[\mathrm{K}_z]\end{bmatrix}\\ [\mathrm{V}] &= [\mathrm{Q}][\mathrm{D}]^{-1}\end{aligned} \tag{B.6}$$

where $[\mathrm{K}_z] = \sqrt{\varepsilon\mu - [\mathrm{K}_x]^2 - [\mathrm{K}_y]^2}$

The PQ matrices of an axially anisotropic homogeneous layer is given as

$$\begin{aligned}[\mathrm{P}] &= \frac{1}{\varepsilon_{zz}}\begin{bmatrix}[\mathrm{K}_x][\mathrm{K}_y] & \varepsilon_{zz}\mu_{yy} - [\mathrm{K}_x][\mathrm{K}_x]\\ [\mathrm{K}_y][\mathrm{K}_y] - \varepsilon_{zz}\mu_{xx} & -[\mathrm{K}_y][\mathrm{K}_x]\end{bmatrix}\\ [\mathrm{Q}] &= \frac{1}{\mu_{zz}}\begin{bmatrix}[\mathrm{K}_x][\mathrm{K}_y] & \varepsilon_{yy}\mu_{zz} - [\mathrm{K}_x][\mathrm{K}_x]\\ [\mathrm{K}_y][\mathrm{K}_y] - \varepsilon_{xx}\mu_{zz} & [\mathrm{K}_y][\mathrm{K}_x]\end{bmatrix}\end{aligned} \tag{B.7}$$

for uniaxially anisotropic homogeneous layer with $\varepsilon_{xx} = \varepsilon_{yy} = \varepsilon_{tt}$, the eigen-matrices are obtained analytically as

$$\begin{aligned}[\mathrm{W}] &= \begin{bmatrix}[\mathrm{I}] & -[\mathrm{K}_y][\mathrm{K}_x]^{-1}\\ [\mathrm{K}_x]^{-1}[\mathrm{K}_y] & [\mathrm{I}]\end{bmatrix}\\ [\mathrm{D}] &= \end{aligned} \tag{B.8}$$

-¿ needs to be confirmed!

B.3 One-dimensional lattice

TE

$$\begin{aligned}[\mathrm{P}] &= -[\mathrm{I}]\\ [\mathrm{Q}] &= [\![\varepsilon]\!] - [\mathrm{K}_x]^2\end{aligned} \tag{B.9}$$

TM

$$\begin{aligned}[\mathrm{P}] &= [\mathrm{I}] - [\mathrm{K}_x][\![\varepsilon]\!]^{-1}[\mathrm{K}_x]\\ [\mathrm{Q}] &= -[\![\varepsilon]\!]\end{aligned} \tag{B.10}$$

Appendix C

Linear algebra operations

numerical evaluation

block matrix inversion

$$\begin{aligned}[\mathrm{T}] &= [\mathrm{F}]\begin{bmatrix}[\mathrm{X}]^{-1} & \\ & [\mathrm{X}]\end{bmatrix}[\mathrm{F}]^{-1} \\ &= \frac{1}{2}\begin{bmatrix}[\mathrm{W}]([\mathrm{X}]+[\mathrm{X}]^{-1})[\mathrm{W}]^{-1} & [\mathrm{W}]([\mathrm{X}]-[\mathrm{X}]^{-1})[\mathrm{V}]^{-1} \\ [\mathrm{V}]([\mathrm{X}]-[\mathrm{X}]^{-1})[\mathrm{W}]^{-1} & [\mathrm{V}]([\mathrm{X}]+[\mathrm{X}]^{-1})[\mathrm{V}]^{-1}\end{bmatrix}\end{aligned} \tag{C.1}$$

the inverse of mode matrix is

$$\begin{aligned}[\mathrm{F}]^{-1} &= \begin{bmatrix}[\mathrm{W}] & [\mathrm{W}] \\ -[\mathrm{V}] & [\mathrm{V}]\end{bmatrix}^{-1} \\ &= \frac{1}{2}\begin{bmatrix}[\mathrm{W}]^{-1} & -[\mathrm{V}]^{-1} \\ [\mathrm{W}]^{-1} & [\mathrm{V}]^{-1}\end{bmatrix}\end{aligned} \tag{C.2}$$

in ETM,

$$\begin{bmatrix}-[\mathrm{I}] & [\mathrm{G}_a] \\ -[\mathrm{V}_0] & [\mathrm{G}_b]\end{bmatrix}\begin{bmatrix}\mathbf{r} \\ \mathbf{t}\end{bmatrix} = [\mathbf{s}] \tag{C.3}$$

in TMM,

$$\begin{bmatrix}-[\mathrm{I}] & [\mathrm{G}_a] \\ -[\mathrm{V}_0] & [\mathrm{G}_b]\end{bmatrix}^{-1} = \begin{bmatrix}-[\mathrm{I}]-[\mathrm{G}_a][\Delta][\mathrm{V}_0] & [\mathrm{G}_a][\Delta] \\ -[\Delta][\mathrm{V}_0] & [\Delta]\end{bmatrix}, \tag{C.4}$$

where $[\Delta] = ([\mathrm{G}_b]-[\mathrm{V}_0][\mathrm{G}_a])^{-1}$.

No layer case

$$\begin{bmatrix}-[\mathrm{I}] & [\mathrm{I}] \\ -[\mathrm{V}_0] & [\mathrm{V}_{L+1}]\end{bmatrix}^{-1} = \begin{bmatrix}-[\Delta][\mathrm{V}_{L+1}] & [\Delta] \\ -[\Delta][\mathrm{V}_0] & [\Delta]\end{bmatrix}, \tag{C.5}$$

where $[\Delta] = ([\mathrm{V}_{L+1}]-[\mathrm{V}_0])^{-1}$.

transfer matrix in single interface -¿ fresnel equation

For iterative optimization problems, the eigen-decomposition process can be accelerated by using perturbation approach[23].

FMM calculations may be accelerated using GPUs, although the eigen-decomposition process is not accelerated by much[24].

Appendix D

Reverse mode gradient

Redheffer star product
ETM subroutine

Test

test

++ nonrectangular lattice
++ chiral system?
++ internal excitation