Advanced Computation:
Computational Electromagnetics

RCWA Extras

Outline

• One spatial harmonic: $P=Q=1$
• Simulation of 1D Gratings with 3D-RCWA
• Formulation of 2D-RCWA with fast Fourier factorization
• Danger of RCWA and convergence
• RCWA and curved structures
• Strategically truncating the set of spatial harmonics
• RCWA for generalized symmetries
• Modeling hexagonal gratings with rectangular RCWA
• Enhanced transmittance matrix approach
One Spatial Harmonic

\[ P = Q = 1 \]

Anatomy of the Convolution Matrix

- Device \( \varepsilon_r(x, y) \)
- Convolution Matrix \( [\varepsilon_r] \)

- Higher order Fourier coefficients describe periodic variations in \( \varepsilon_r(x, y) \)
- 0-order Fourier coefficient is the average value
One Spatial Harmonic \((P=Q=1)\)

When only one spatial harmonic is used, RCWA reduces to the 1D transfer matrix method, but uses the average value for the material properties in each layer. This is first-order “effective medium theory” and can make use of fast Fourier factorization so it is more sophisticated than the standard TMM.

\[\varepsilon_r \ll \varepsilon_1 < \varepsilon_2 < \varepsilon_3 < \varepsilon_4 < \varepsilon_r\]

Physical Device

Effective Medium Approximation in 1D
Grating Terminology

1D grating
Ruled grating

Requires a 2D simulation

2D grating
Crossed grating

Requires a 3D simulation

3D-RCWA for 1D Gratings

Three-dimensional RCWA simulates all polarizations at the same time.

For 1D Gratings, Maxwell’s equations decouple into the $E$ mode and the $H$ mode.

It is possible to reformulate RCWA specifically for 1D Gratings so that it will only simulate either the $E$ mode or $H$ mode, but not both. This approach will be several times faster due to smaller matrices.

It is my experience that 3D-RCWA is fast enough that few applications warrant formulating a 2D-RCWA code. Exceptions include when fast Fourier factorization is important or for running optimizations that require many thousands of simulations to be iterated.

There are some tricks that can be used when using RCWA to model 1D gratings to maximize the speed and efficiency.
Number of Spatial Harmonics

You must use some number of spatial harmonics in this direction.

\[ P \sim 7 \frac{\Lambda}{\lambda} \]

\[ N_x > 500 \]

There is no contrast in this direction. Only one spatial harmonic is needed.

\[ Q = 1 \]
\[ N_y = 1 \]

E and H Modes with 3D-RCWA

The field for both \( E \) and \( H \) modes are thought of completely in terms of the polarization vector for the electric field. If you have implemented your codes following these lectures, no changes to your code are needed outside of the dashboard.

Source Wave Vector

\[
\vec{k}_{inc} = k_x n_{inc} \begin{bmatrix} \sin \theta \\ 0 \\ \cos \theta \end{bmatrix}
\]

Source Polarization Vector

\[
\vec{P}_E = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \text{E Mode}
\]
\[
\vec{P}_H = \begin{bmatrix} \cos \theta \\ 0 \\ \sin \theta \end{bmatrix} \quad \text{H Mode}
\]

The concept of \( E \) and \( H \) modes only apply when: (1) LHI or diagonally anisotropic materials, (2) device is uniform in \( y \) direction, and (3) propagation is restricted to \( x-z \) plane.
FDFD Vs. RCWA Representations

FDFD represents a device as the side view of a single unit cell.

In contrast, RCWA represents a device as a top view of each layer of a single unit cell.

Formulation of 2D RCWA with FFF (1D Gratings)
Starting Point for Derivation

We will start with the semi-analytical matrix form of Maxwell’s equations in Fourier-space. These are rigorous and valid even for 3D devices.

\[-j\tilde{K}_y u_z - \frac{d}{dz} u_y = [\varepsilon_r] s_x\]
\[\frac{d}{dz} u_x + j\tilde{K}_x u_z = [\varepsilon_r] s_y\]
\[\tilde{K}_x u_y - \tilde{K}_y u_x = j[\varepsilon_r] s_z\]
\[-j\tilde{K}_y s_z - \frac{d}{dz} s_y = [\mu_r] u_x\]
\[\frac{d}{dz} s_x + j\tilde{K}_x s_z = [\mu_r] u_y\]
\[\tilde{K}_x s_y - \tilde{K}_y s_x = j[\mu_r] u_z\]

Reduction to Two Dimensions

When devices are uniform in the y-direction and no wave propagation occurs in this direction, we have
\[\tilde{K}_y = 0\]

Our matrix equations reduce to
\[-\frac{d}{dz} u_y = [\varepsilon_r] s_x\]
\[\frac{d}{dz} u_x + j\tilde{K}_x u_z = [\varepsilon_r] s_y\]
\[\tilde{K}_x u_y = j[\varepsilon_r] s_z\]
\[-\frac{d}{dz} s_y = [\mu_r] u_x\]
\[\frac{d}{dz} s_x + j\tilde{K}_x s_z = [\mu_r] u_y\]
\[\tilde{K}_x s_y = j[\mu_r] u_z\]
Two Independent Modes

We see that Maxwell’s equations have decoupled into two independent modes.

\[
-\frac{d}{dx} u_x = [\varepsilon] s_x, \\
\frac{d}{dx} u_y + jK_y u_y = [\varepsilon] s_y, \\
K_x u_x = j[\varepsilon] s_x, \\
K_y u_y = j[\varepsilon] s_y.
\]

E Mode

\[
-\frac{d}{dx} s_x = [\mu] u_x, \\
\frac{d}{dx} s_y + jK_y s_y = [\mu] u_y, \\
K_x s_x = j[\mu] u_x, \\
K_y s_y = j[\mu] u_y.
\]

H Mode

Orientation of the Field Components

For 1D gratings, the orientation of the field components relative to the interfaces is fixed. In this case, it is straightforward to incorporate fast Fourier factorization into the formulation.

Within the layer...
\( S_x \) is always perpendicular to interfaces. 
\( S_y \) is always parallel to interfaces. 
\( S_z \) is always parallel to interfaces.

Within the layer...
\( U_x \) is always perpendicular to interfaces. 
\( U_y \) is always parallel to interfaces. 
\( U_z \) is always parallel to interfaces.
Incorporating Fast Fourier Factorization

For 1D gratings, it is very straightforward to incorporate fast Fourier factorization rules. This will improve convergence rates.

**E Mode**

\[
\frac{d}{dz} \mathbf{u}_x + j\mathbf{K}_x \mathbf{u}_x = \left[ \varepsilon_r \right] \mathbf{s}_y \\
-\frac{d}{dz} \mathbf{s}_x = \left[ 1/\mu_r \right] \mathbf{u}_z \\
\mathbf{K}_x \mathbf{s}_y = j\left[ \mu_r \right] \mathbf{u}_z
\]

\( \mathbf{s}_y \) is always parallel to interfaces so \( \left[ \varepsilon_r \right] \) is the standard convolution matrix.

\( \mathbf{u}_z \) is always perpendicular so FFF rules are used to construct \( \left[ 1/\mu_r \right] \).

\( \mathbf{u}_z \) is always parallel to interfaces so \( \left[ \mu_r \right] \) is the standard convolution matrix.

**H Mode**

\[
\frac{d}{dz} \mathbf{s}_x + j\mathbf{K}_x \mathbf{s}_z = \left[ \mu_r \right] \mathbf{u}_y \\
-\frac{d}{dz} \mathbf{u}_y = \left[ 1/\varepsilon_r \right] \mathbf{s}_x \\
\mathbf{K}_x \mathbf{u}_y = j\left[ \varepsilon_r \right] \mathbf{s}_z
\]

\( \mathbf{u}_y \) is always parallel to interfaces so \( \left[ \mu_r \right] \) is the standard convolution matrix.

\( \mathbf{s}_z \) is always perpendicular so FFF rules are used to construct \( \left[ 1/\varepsilon_r \right] \).

\( \mathbf{s}_z \) is always parallel to interfaces so \( \left[ \varepsilon_r \right] \) is the standard convolution matrix.

---

**Eliminate Longitudinal Components**

We solve the third equation for \( \mathbf{u}_z \) and the sixth equation for \( \mathbf{s}_z \) to obtain

\[
\mathbf{u}_z = -j\left[ \mu_r \right]^{-1} \mathbf{K}_x \mathbf{s}_y \\
\mathbf{s}_z = -j\left[ \varepsilon_r \right]^{-1} \mathbf{K}_x \mathbf{u}_y
\]

We substitute these expressions into the remaining Maxwell’s equations.

**E Mode**

\[
\frac{d}{dz} \mathbf{u}_x = \left[ \varepsilon_r \right] \mathbf{s}_y - \mathbf{K}_x \left[ \mu_r \right]^{-1} \mathbf{K}_x \mathbf{s}_y \\
\frac{d}{dz} \mathbf{s}_y = -\left[ 1/\mu_r \right]^{-1} \mathbf{u}_x
\]

**H Mode**

\[
\frac{d}{dz} \mathbf{s}_x = \left[ \mu_r \right] \mathbf{u}_y - \mathbf{K}_x \left[ \varepsilon_r \right]^{-1} \mathbf{K}_x \mathbf{u}_y \\
\frac{d}{dz} \mathbf{u}_y = -\left[ 1/\varepsilon_r \right]^{-1} \mathbf{s}_x
\]
Standard P and Q Form

**E Mode**
\[
\frac{d}{d\tilde{z}} s_y = Pu_x \\
\frac{d}{d\tilde{z}} u_x = Qs_y
\]

**H Mode**
\[
\frac{d}{d\tilde{z}} u_y = Ps_x \\
\frac{d}{d\tilde{z}} s_x = Qu_y
\]

\[
P = -\left[\frac{1}{\mu_r}\right]^{-1} \\
Q = \left[\varepsilon_r\right] - \tilde{K}_x \left[\mu_r\right]^{-1} \tilde{K}_x
\]

We see that FFF is incorporated solely into the matrix \( P \).

---

Matrix Wave Equations

Now that we have our equations in standard \( P \) and \( Q \) form, we derive the wave equations in the same manner as before.

**E Mode**
\[
\frac{d^2}{d\tilde{z}^2} s_y - \Omega^2 s_y = 0
\]
\[
\Omega^2 = PQ
\]

**H Mode**
\[
\frac{d^2}{d\tilde{z}^2} u_y - \Omega^2 u_y = 0
\]
\[
\Omega^2 = PQ
\]

...and now you know the rest of the story.
Convergence Study for 1D Gratings

\[ \mu_r = 1.0, \varepsilon_r = 3.0 \]

Ordinary FFF

\[ \mu_r = 1.0, \varepsilon_r = 40.0 \]

Ordinary FFF

H-Mode

Convergence Study for 1D Curved Structures

\[ \mu_r = 1.0, \varepsilon_r = 3.0 \]

Ordinary FFF

H-Mode

\[ \mu_r = 1.0, \varepsilon_r = 40.0 \]

Ordinary FFF
Danger of RCWA and Convergence

In real-space, poor grid resolution led to fluctuations in conservation of power and other very recognizable signs that things are wrong.

The danger of RCWA is that results can “look” correct even with very few spatial harmonics.

Conservation of power will always be obeyed in RCWA even using just one spatial harmonic.

It must become habit to look for convergence, as there are few other signs that more spatial harmonics are needed.
Typical Convergence Plot

Device from last lecture
@ $f_0 = 47.6$ GHz

Convergence
250 harmonics
$\approx 15 \times 15$ harmonics

RCWA and Curved Structures
EBG Material

Divide into Thin Layers

Layers 1 to 20  Staircase Approximation
Strategically Truncating the Set of Spatial Harmonics

Notes on Truncating the Set of Spatial Harmonics

- The choice of which spatial harmonics to include in the expansion is arbitrary.
- Improper choice can lead to slow convergence and inaccurate results.
- We chose directions consistent with the physics of diffraction and a rectangular Fourier-space grid for simplicity.
- The number of harmonics retained in a particular direction determines the spatial resolution of structures with contrast in that direction.
- It seems optimal to keep the number of spatial harmonics uniform in all directions.
Fourier-Space Grid Notation

The components of the wave vector expansion look like:

\[
\tilde{k}_x (m) \quad \tilde{k}_y (n)
\]

We visualized it this way:

A simpler view of our 2D Fourier-space grid is

Simple Grid Truncation Scheme

We will retain all spatial harmonics with indices that satisfy the following equation:

\[
\left| \frac{m}{M} \right|^{2\gamma} + \left| \frac{n}{N} \right|^{2\gamma} \leq 1
\]

\[
\gamma = 0.01 \quad \gamma = 0.35 \quad \gamma = 0.5
\]

Conventional 

\[
|n| \leq N \text{ and } |m| \leq M
\]

\[
\gamma = 0.65 \quad \gamma = 1.0 \quad \gamma = 1.3
\]

pincushion 

diamond 

circular 

barrel

**Implementation**

Step 1 – Build Unit Cell on High-Resolution Grid

Step 2 – Calculate Fourier Coefficients Using FFT Technique
Implementation

Step 3 – Assemble Standard Convolution Matrix From Fourier Coefficients

\[ \mathbf{E}_r = \mathbf{F} \mathbf{X} \mathbf{F}^* \]

Implementation

Step 4 – Build Truncation Map

% Construct Truncation Map

\[ \text{TMAP} = \text{abs}(m/M)^{2p} + \text{abs}(n/N)^{2p}; \]

\[ \text{TMAP} = (\text{TMAP} <= 1); \]
Implementation
Step 5 – Extract Indices of Spatial Harmonics to Retain

% Construct Truncation Map
TMAP = abs(m/M).^(2*p) ...
  + abs(n/N).^(2*p);
TMAP = (TMAP <= 1);

% Extract Array Indices
ind = find(TMAP(:));

Implementation
Step 6 – Truncate Convolution Matrix

% Truncate Convolution Matrix
ERCT = ERC(ind,ind);
Implementation
Step 7 – Perform RCWA

The rest of the RCWA algorithm remains virtually unchanged.

You will need to consider your truncation again:

1. When you calculate the source.
2. If you calculate the fields from the eigen-modes.

RCWA for Generalized Symmetries
Revised Fourier Transforms

The materials...

\[ \varepsilon_r (\vec{r}) = \sum_{k_{m,n}} \sum_{m,n} a_{m,n} e^{i(2\pi k_{m,n} \cdot \vec{r})} \]

\[ \mu_r (\vec{r}) = \sum_{k_{m,n}} \sum_{m,n} b_{m,n} e^{i(2\pi k_{m,n} \cdot \vec{r})} \]

\[ a_{m,n} = \frac{1}{A_{\text{cell}}} \int_{\text{cell}} \varepsilon_r (\vec{r}) e^{-i(2\pi k_{m,n} \cdot \vec{r})} dA \]

\[ b_{m,n} = \frac{1}{A_{\text{cell}}} \int_{\text{cell}} \mu_r (\vec{r}) e^{-i(2\pi k_{m,n} \cdot \vec{r})} dA \]

The fields...

\[ E_x (\vec{r},z) = \sum_{k_{m,n}} \sum_{m,n} S_x (m,n,z) e^{-i(2\pi k_{m,n} \cdot \vec{r})} \]

\[ H_x (\vec{r},z) = \sum_{k_{m,n}} \sum_{m,n} H_x (m,n,z) e^{-i(2\pi k_{m,n} \cdot \vec{r})} \]

Wave vector expansion...

\[ k_i (m,n) = k_{\mu} - m \xi_1 - n \eta_1 \]

\[ k_i (m,n) = \sqrt{k_r^2 \mu^2 - k_i (m,n)^2} \]

\[ k_{\mu} = \frac{2 \pi}{\lambda} \]

\[ \vec{\xi}_1 = \text{reciprocal lattice vectors of the unit cell} \]

Modeling Hexagonal Gratings with Rectangular RCWA
Geometry of a Hexagon

Grating Vectors of Hexagonal Structures

\[ \vec{i}_1 = \frac{a}{2} \hat{x} - \frac{a \sqrt{3}}{2} \hat{y}, \quad \vec{i}_2 = \frac{a}{2} \hat{x} + \frac{a \sqrt{3}}{2} \hat{y}, \quad \vec{i}_3 = c \hat{z} \]

\[ \vec{T}_1 = \frac{2 \pi}{a} \hat{x} - \frac{2 \pi}{a \sqrt{3}} \hat{y}, \quad \vec{T}_2 = \frac{2 \pi}{a} \hat{x} + \frac{2 \pi}{a \sqrt{3}} \hat{y}, \quad \vec{T}_3 = \left(\frac{2 \pi}{c}\right) \hat{z} \]
Rectangular Unit Cell in Hexagonal Array

We must identify a rectangular unit cell that reconstructs a hexagonal array.

This implies that we will need more spatial harmonics along the \( y \) direction than \( x \).

\[ Q = \text{round}(P \cdot \frac{Sy}{Sx}) \]

Enhanced Transmittance Matrix Approach

Motivation

The enhanced transmittance matrix (ETM) method involves less matrix manipulations so it is much faster than using scattering matrices. Maybe \( \sim 10\times \). It also provides easier computation of internal fields.

ETM works by first stepping backward through each layer (backward analysis) and then stepping forward (forward analysis). Intermediate parameters must be stored for each layer during the backward analysis that are recalled during the forward analysis. This leads to severe memory limitations when many layers are used.

Conclusion \( \Rightarrow \) Unless the device is composed of prohibitively large number of layers or if other features of scattering matrices are not needed, use ETM.

The Problem

The source of the instability is the following matrix.

\[
X^+ = e^{-\alpha L}
\]

\[
X^- = e^{\alpha L} \quad \text{Growing exponentials!!!}
\]

The enhanced transmittance matrix (ETM) method was the first technique applied to RCWA that fixed the instability.

ETM is much faster than scattering matrices, but is much less memory efficient because it requires parameters to be stored for all layers at the same time. Scattering matrices can proceed one layer at a time, forgetting everything about previous layers.
Boundary Conditions

First interface:
\[ s + A r = F_1 \begin{bmatrix} I & 0 \\ 0 & X_1 \end{bmatrix} c_1 \]

Intermediate interfaces:
\[ F_j \begin{bmatrix} X_j & 0 \\ 0 & I \end{bmatrix} c_j = F_{j+1} \begin{bmatrix} I & 0 \\ 0 & X_{j+1} \end{bmatrix} c_{j+1} \]

Last interface:
\[ F_N \begin{bmatrix} X_N & 0 \\ 0 & I \end{bmatrix} c_N = B t \]

Work Backward Through Layers (1 of 4)

The goal is to solve for \( r \) and \( t \) without using \( X^{-1} \).

We start by solving the equation at the last interface for \( c_N \).

\[ F_N \begin{bmatrix} X_N & 0 \\ 0 & I \end{bmatrix} c_N = B t \quad \rightarrow \quad c_N = \begin{bmatrix} X_N^{-1} & 0 \\ 0 & I \end{bmatrix} F_N^{-1} B t \]

We write this as
\[ c_N = \begin{bmatrix} X_N^{-1} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} a_N \\ b_N \end{bmatrix} t_N \quad \begin{bmatrix} a_N \\ b_N \end{bmatrix} = F_N^{-1} B \]
Work Backward Through Layers (2 of 4)

To eliminate the potentially ill conditioned matrix $X^{-1}$, we introduce an intermediate transmittance matrix parameter $t_N$ defined as

\[ t = a_N^{-1}X_N t_N \]

\[ \text{t and } t_N \text{ remain unknown.} \]

Our equation for $c_N$ becomes

\[
\begin{bmatrix}
X_N^{-1} & 0 \\
0 & 1
\end{bmatrix}
\begin{bmatrix}
a_N \\
b_N
\end{bmatrix}
\rightarrow
\begin{bmatrix}
1 \\
b_N a_N^{-1} X_N
\end{bmatrix}
\begin{bmatrix}
t_N \\
t_N
\end{bmatrix}
\]

Work Backward Through Layers (3 of 4)

The boundary condition equation at the second-to-last interface is

\[ F_{N-1} \begin{bmatrix}
X_{N-1}^{-1} & 0 \\
0 & 1
\end{bmatrix}
c_{N-1} = F_N \begin{bmatrix}
1 & 0 \\
0 & X_N
\end{bmatrix}
c_N \]

Substituting our expression for $c_N$ into the equation yields

\[
F_{N-1} \begin{bmatrix}
X_{N-1}^{-1} & 0 \\
0 & 1
\end{bmatrix}
c_{N-1} = F_N \begin{bmatrix}
1 & 0 \\
0 & X_N
\end{bmatrix}
\begin{bmatrix}
1 \\
b_N a_N^{-1} X_N
\end{bmatrix}
t_N
\]

Solving this for $c_{N-1}$ leads to

\[
c_{N-1} = \begin{bmatrix}
X_{N-1}^{-1} & 0 \\
0 & 1
\end{bmatrix} F_{N-1}^{-1} F_N \begin{bmatrix}
1 & 0 \\
0 & X_N
\end{bmatrix}
\begin{bmatrix}
1 \\
b_N a_N^{-1} X_N
\end{bmatrix}
t_N
\]

We have now worked backward by one layer while avoiding $X_N^{-1}$. 
Work Backward Through Layers (4 of 4)

This process continues through all the layers.

\[
\begin{align*}
c_{N-1} &= \left[ \begin{array}{c} X_{N-1}^{-1} \\ 0 \\ 1 \\ b_{N-1}^{-1}a_{N-1}^{-1}X_{N-1} \end{array} \right] F_{N-1}^{-1} F_N \left[ \begin{array}{c} 1 \\ 0 \\ X_N \\ b_N a_N^{-1}X_N \end{array} \right] t_N \\
c_{N-1} &= \left[ \begin{array}{c} X_{N-1}^{-1} \\ 0 \\ 1 \\ b_{N-1}^{-1}a_{N-1}^{-1}X_{N-1} \end{array} \right] t_{N-1} \\
t_N &= a_{N-1}^{-1}X_{N-1} t_{N-1}
\end{align*}
\]

\[\text{Note: we must store } a \text{ and } X \text{ for each layer.}
\text{This leads to poor memory efficiency.}\]

Solve for Reflected and Transmitted Fields

After working through all interfaces, we are left with

\[
s + Ar = F_1 \left[ \begin{array}{c} 1 \\ 0 \\ X_1 \\ b_1 a_1^{-1}X_1 \end{array} \right] t_1
\]

We solve this matrix equation for \( r \) and \( t_1 \).

\[
\begin{bmatrix} r \\ t_1 \end{bmatrix} = \left[ -A \quad B' \right]^{-1} S \quad B' = F_1 \left[ \begin{array}{c} 1 \\ 0 \\ X_1 \\ b_1 a_1^{-1}X_1 \end{array} \right] t_1
\]

Now that we know \( t_1 \), we can work forward through the layers to solve for \( t \).

\[
t_2 = a_1^{-1}X_1 t_1 \\
t_3 = a_2^{-1}X_2 t_2 \\
\vdots \\
t_N = a_{N-1}^{-1}X_{N-1} t_{N-1} \\
t = a_N^{-1}X_N t_N
\]
Calculating the Diffraction Efficiencies

First, we calculate the longitudinal field components.

\[ r_z = -\hat{K}_{z,\text{ref}}^{-1} \left( \hat{K}_z r_x + \hat{K}_y r_y \right) \]

\[ t_z = -\hat{K}_{z,\text{ref}}^{-1} \left( \hat{K}_t t_x + \hat{K}_y t_y \right) \]

Second, we calculate the diffraction efficiencies.

\[ R = \frac{\text{Re} \left[ -\hat{K}_{z,\text{ref}} / \mu_{\text{inc}} \right]}{\text{Re} \left[ k_{z,\text{inc}} / \mu_{\text{inc}} \right]} |r|^2 \]

\[ |r|^2 = |r_x|^2 + |r_y|^2 + |r_z|^2 \]

\[ T = \frac{\text{Re} \left[ \hat{K}_{z,\text{trn}} / \mu_{\text{trn}} \right]}{\text{Re} \left[ k_{z,\text{inc}} / \mu_{\text{inc}} \right]} |t|^2 \]

\[ |t|^2 = |t_x|^2 + |t_y|^2 + |t_z|^2 \]

Block Diagram of ETM