DESIGN AND OPTIMIZATION OF NANO-OPTICAL
ELEMENTS BY COUPLING FABRICATION TO
OPTICAL BEHAVIOR

by

RAYMOND C. RUMPF
B.S.E.E. Florida Institute of Technology, 1995
M.S.E.E. Florida Institute of Technology, 1997

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the College of Optics / CREOL&FPCE
at the University of Central Florida
Orlando, Florida

Major Professor: Eric Johnson

Spring Term
2006

ABSTRACT

Photonic crystals and nanophotonics have received a great deal of attention over the last decade, largely due to improved numerical modeling and advances in fabrication technologies. To this day, fabrication and optical behavior remain decoupled during the design phase and numerous assumptions are made about "perfect" geometry. As research moves from theory to real devices, predicting device behavior based on realistic geometry becomes critical. In this dissertation, a set of numerical tools was developed to model micro and nano fabrication processes. They were combined with equally capable tools to model optical performance of the simulated structures. Using these tools, it was predicted and demonstrated that 3D nanostructures may be formed on a standard mask aligner. A space-variant photonic crystal filter was designed and optimized based on a simple fabrication method of etching holes through hetero-structured substrates. It was found that hole taper limited their optical performance and a method was developed to compensate. A method was developed to tune the spectral response of guided-mode resonance filters at the time of fabrication using models of etching and deposition. Autocloning was modeled and shown that it could be used to form extremely high aspect ratio structures to improve performance of form-birefringent devices. Finally, the numerical tools were applied to metallic photonic crystal devices.

To my wife, family, friends, and coworkers

from which I have sacrificed so much time.

ACKNOWLEDGMENTS

Above all, I want to thank my wife Krissy for her love, patience, and unending support through my entire Ph.D. process. Her love and encouragement helped make it possible for me to pursue this degree. Her help in creating and formatting this document is deeply appreciated.

I must thank my advisor and friend Dr. Eric Johnson for his wisdom, guidance, and tolerance through this process. I am extremely proud to have studied under him. I also appreciate the time and contributions of my other committee members: Dr. Jim Moharam, Dr. Stephen Kuebler, Dr. Glenn Boreman, and Dr. Thomas Wu.

I want to thank my family for their support and apologize for all of the time we did not share due to my studies. My love and thoughts were always with them.

My friends and coworkers who supported me and provided guidance during my graduate studies were of great help. Their contributions are greatly appreciated.

Finally, I want to thank my dog Rocky, who sat by my side through most of my research. He always seemed to know when I needed a break and tolerated me when I taught him Maxwell's equations.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS/ABBREVIATIONS

ABC	Absorbing Boundary Condition
BCC	Body-Centered-Cubic
BPM	Beam Propagation Method
CCD	Charge Coupled Device
CP	Circular Polarization
DOE	Diffractive Optical Element
FCC	Face-Centered-Cubic
FDFD	Finite-Difference Frequency-Domain
FDM	Finite-Difference Method
FDTD	Finite-Difference Time-Domain
FMM	Fast Marching Method
GMR	Guided-Mode Resonance
GMRF	Guided-Mode Resonance Filter
IBZ	Irreducible Brillouin Zone
LCP	Left-Hand Circular Polarization
LP	Linear Polarization
LSM	Level Set Method
MOL	Method of Lines
NFNP	Near-Field Nano-Patterning
NOE	Nano-Optical Element

ODOS	Optical Density of States
PC	Photonic Crystal
PML	Perfectly Matched Layer
PWEM	Plane Wave Expansion Method
RCP	Right-Hand Circular Polarization
RCWA	Rigorous Coupled-Wave Analysis
TE	Transverse Electric
TM	Transverse Magnetic

CHAPTER 1
INTRODUCTION

1.1. Overview of Dissertation

Photonic crystals and nanophotonics have received a great deal of attention over the last decade, largely due to improved numerical modeling and advances in fabrication technologies. To this day, fabrication and optical behavior remain decoupled during the design phase and numerous assumptions are made about "perfect" geometry. Little effort has been made to incorporate distorting artifacts of fabrication or to study how these deformations affect device performance.

As research moves from theory to real devices, predicting device behavior based on realistic geometry becomes critical [1-7]. Even small distortions can lead to dramatic changes in optical behavior [8]. In this dissertation, a set of numerical tools was developed to model micro and nano fabrication processes. They were combined with equally capable tools to model optical performance of the simulated structures. Using these tools, nano-optical elements were design and optimized by coupling fabrication to optical behavior. Experimental results achieved in the lab verify the concepts developed in this dissertation.

1.1.1. Outline of Dissertation

Chapter 1 introduces the topic of the thesis and presents the rationale and significance of results obtained in this work. The concept of nano-optical elements (NOE) is presented and the importance of modeling their fabrication is illustrated. As a prelude to following chapters, Maxwell's equations are introduced and used to derive wave equations and important theorems.

The microscopic behavior of dielectric materials is explained using the Lorentz oscillator model from which Drude theory of metals is derived.

Chapter 2 introduces photonic crystals. Lattice symmetries are identified and illustrated and a mathematical framework is presented to quantify them in terms of lattice vectors. The concepts of reciprocal space, Wigner-Seitz cells, and Brillouin zones are introduced and used to describe how photonic band diagrams are constructed and interpreted. Origin of the photonic band gap is explained and dispersive behavior is compared to the microscopic response of a homogeneous dielectric.

Chapters 3-6 describe the main electromagnetic simulation tools used in this dissertation. These are finite-difference frequency-domain (FDFD), finite-difference time-domain (FDTD), plane wave expansion method (PWEM), and rigorous coupled-wave analysis (RCWA). The methods are introduced and their formulations are presented in detail. To aid in their implementation, detailed block diagrams are provided that are consistent with the formulations. Benchmark simulations are given later in this chapter to verify their accuracy.

Chapters 7 and 8 describe numerical tools used to model etching, deposition, and developing processes. The string method was used to model etching, deposition, and autocloning processes, while the fast marching method (FMM) was used to model developing. The methods are introduced and detailed formulations are presented. To aid in their implementation, detailed block diagrams are provided. Benchmark simulations or numerical examples are performed to verify accuracy of the codes or to illustrate the method.

Chapter 9 summarizes the earliest efforts in this dissertation on holographic lithography. A method is presented to synthesize exposure configurations that lead to the desired photonic crystal symmetry. Physics of the process are described and numerical tools from earlier chapters

are combined into a comprehensive model to simulate the entire process. It is predicted that 3D structures can be formed by near-field nano-patterning (NFNP) on a standard mask aligner using an ultraviolet lamp instead of a laser. Experimental results are presented that confirm this prediction.

Chapters 10-12 design and optimize specific NOEs based on realistic device geometry determined by modeling fabrication. These include photonic crystal devices, guided-mode resonance filters, and form-birefringent elements. Numerical simulations are compared to experimental results where possible.

Chapter 13 summarizes the most recent work in this dissertation pertaining to metallic photonic crystals. Due to the complexity of the structures, FDTD is used exclusively to calculate photonic band diagrams as well as transmission and reflection through finite slabs. Optical properties of photonic crystals comprised of metal or dielectric are compared and important differences identified. A comparison of gold and silver is made that accounts for the dispersive nature of the metals. A comparison of photonic crystals with ideal and realistic geometries are compared.

Chapter 14 summarizes the dissertation and reviews key developments achieved in this work. Suggestions for future work are identified.

1.2. Nano-Optical Elements

Conventional micro and diffractive optics consists of macroscopic surface-relief structures usually designed using scalar methods [9]. Basic passive optical components such as lenses [10-12], prisms [13], and mirrors [14] can be miniaturized and arbitrary shapes realized. These devices have found a host of applications in imaging [15, 16], adaptive optics [17],

4

sensors [18, 19], microscopy [20, 21], communications [22-24], micro-fabrication [2, 5, 6, 25], beam shaping [26-30], and more. Material properties, however, remain a fundamental limitation.

Photonic crystals (PC) are structures with sub-wavelength periodic material properties that can produce sharp electromagnetic resonances. These structures provide a means of controlling and engineering the bulk electromagnetic properties of materials. Many novel and useful phenomena have been observed that promise significant impact on future photonic systems. Useful properties include forbidden transmission [31] and unique dispersive properties like negative refractive index [32].
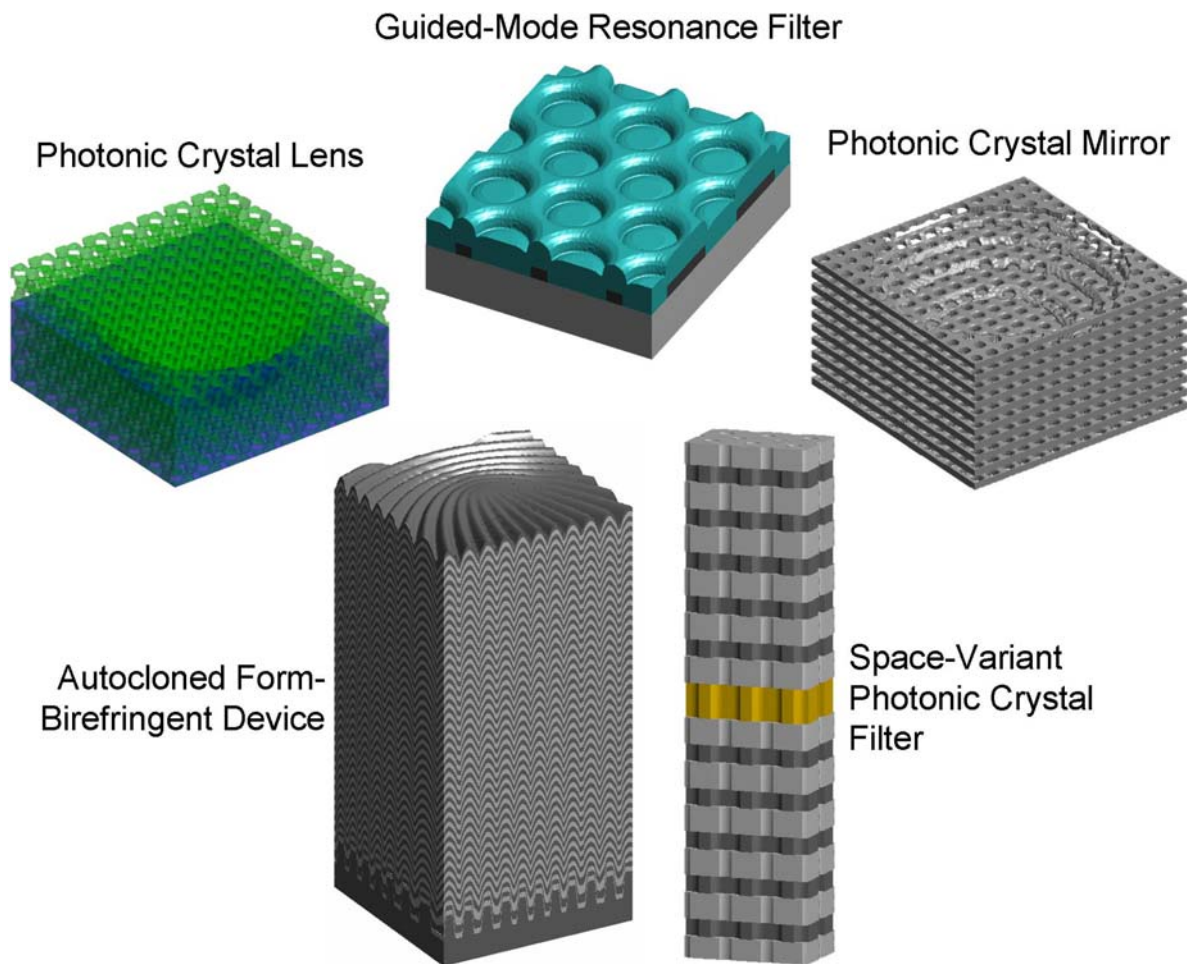
Figure 1-1: Nano-optical elements

5

Nano-optical elements combine functionality of micro and diffractive optical elements with the optical properties of nanophotonic structures like photonic crystals. NOEs can enhance performance of devices by providing a means to circumvent fundamental limitations imposed by conventional materials. Concept diagrams of several nano-optical elements are provided in Figure 1-1. An example simulation is depicted in Figure 1-2 where a lens has been formed inside a photonic crystal. When light within the photonic band gap is applied, it is completely reflected. When light from outside the band gap is applied, it is transmitted and focused by the lens. In this manner, multiple functions have been multiplexed into a single device.
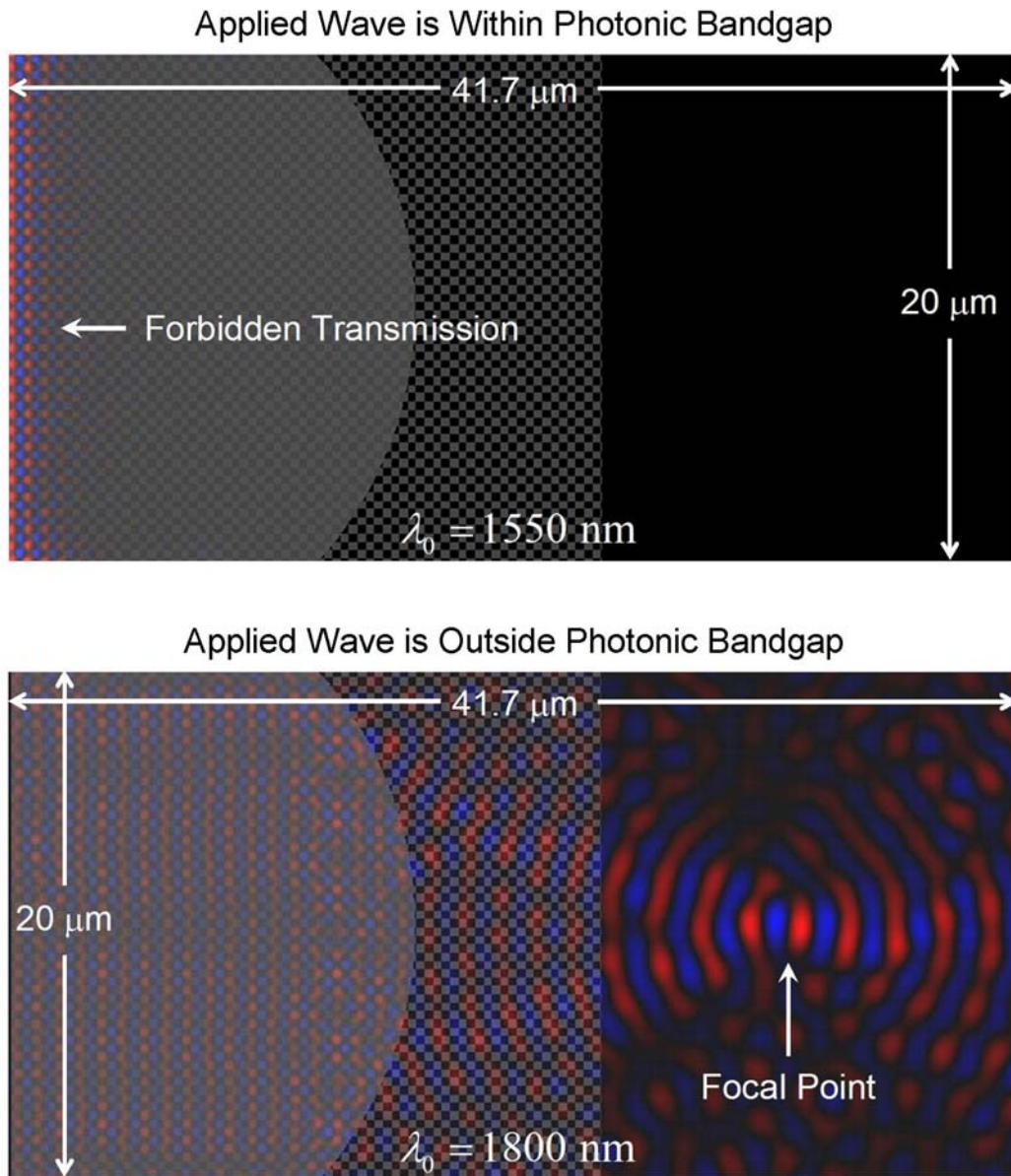
Figure 1-2: Multifunctional nano-optical element

1.3. Significance of Nano-Optical Elements

Photonic crystals and other nanophotonic structures hold great promise for many applications, and the research community is beginning to integrate them into higher level systems like NOEs.

NOEs should find a host of applications multiplexing and demultiplexing signals due to inherent spectral filtering properties of photonic crystals,. This could include wavelength division multiplexing (WDM) in fiber optic systems, color filtering in imaging systems, and more.

Devices based on NOEs could be an enabling technology for system miniaturization. Dimensions of optical integrated circuits may be reduced through more efficient routing of waveguides using tighter bends and improved crossing. Imaging systems can be made more compact by replacing complex multi-element lens systems with single element NOEs. Reducing size and number of components in a system can also lead to improved reliability and lower cost.

Display technology may benefit as well. Due to the porous nature of photonic crystals, they can be backfilled with a number of materials with tunable optical properties. Liquid crystals could tune the index contrast, moving the band gap, and effectively changing color.

Sensor technology can exploit highly sensitive optical resonances to detect many things. Chemical sensors could be formed by backfilling with chemically sensitive materials. Highly sensitive acoustic, seismic, pressure, stress, strain, and vibration sensors could be made by detecting small deformations of the lattice.

Imaging systems are likely to benefit from the ability to engineer optical materials. Properties such as negative refractive index could be designed into imaging optics to reduce aberrations. This issue is particularly severe in ultra-miniaturized systems.

When more is known about NOEs other applications should become apparent. These may include beam steering, bio-photonics, adaptive optics, and more.

1.4. Impact of Fabrication on Optical Behavior

Mechanisms causing nano-optical elements to deviate from their "perfect" geometry fall into two categories. First, random distortions arise from random fluctuations in size, alignment, surface roughness, and other volumetric inhomogeneities. Random distortions are difficult to predict and quantify because simulations are often very large and statistics of the fluctuations must be understood. Second, deterministic distortions produce similar effects, but repeat throughout the element. Deterministic distortions tend to affect optical behavior most severely because the same electromagnetic disturbances are encountered many times throughout the entire structure. Fortunately, these can be predicted by modeling fabrication. Despite this, random distortions remain the most heavily studied. As nano-optical elements move from theory to practice, understanding both mechanisms will be critical.

For random distortions, the most commonly studied photonic crystal is the stacked-bar, or woodpile, structure [33, 34]. Two-dimensional arrays of cylinders [35, 36] and inverse opals [8] have also been studied among others [37, 38]. Width of the photonic band gap seems to be the dominant figure of merit to asses how resilient a photonic crystal is to fluctuations in its geometry. Virtually no work can be found on how distortions affect the dispersive behavior of photonic crystals. Most literature concludes that fluctuations on the order of 10% to 15% of the lattice constant lead to a 40% to 50% reduction in width of the photonic band gap. A notable exception is the inverse opal, where the band gap resides at much higher frequencies between the 8th and 9th bands. For inverse opals, the band gap was found to completely close if fluctuations exceeded just 2%. Two-dimensional photonic crystals were found to be more sensitive to size fluctuations than variations in alignment. As may be expected, transverse magnetic (TM) modes were shown to be more sensitive than transverse electric (TE) modes.

Very little work investigating deterministic distortion can be found in the literature. Deformations arising in an autocloning process were addressed in Ref. [39]. This dissertation has contributed to studies on holographic lithography, NFNP, and more [1-7]. It was shown that optical absorption during holographic exposure leads to photonic crystals where fill factor varies with depth. For negative photoresists like Epon® SU-8 [4], photonic crystals become thinner with depth leading to chirped lattices with broadened spectral response. High contrast exposures were shown to be quite robust to reflections and standing waves during exposure. A face-centered-cubic photonic crystal was able to tolerate 20% reflection from a substrate while maintaining similar optical performance. It was shown that performing multiple exposures can superimpose different lattices forming hybrids with optical properties of both.

Figure 1-3 compares severity of random and deterministic distortion for one specific photonic crystal. The lattice is a 2D array of air cylinders in a dielectric medium with $\varepsilon_r=12$. The leftmost diagram in Figure 1-3(a) shows the "perfect" photonic crystal with uniform holes, its Brillouin zone, and the photonic band structure for both TE and TM modes. Bands are symmetric about Γ because the lattice is symmetric through a rotation of 90°. In Figure 1-3(b) and (c), the lattice is distorted leading to altered band structures. Distorted bands, represented by solid lines, are compared to bands of the "perfect" photonic crystal, indicated by dashed lines. In Figure 1-3(b), deterministic distortion is introduced by increasing hole size by 10% of the lattice constant. Bands of the distorted lattice retain their symmetry, but are offset in frequency with increasing severity at higher frequencies. A band gap has opened for the TE modes between the first and second bands. The TM band gap has shifted by around 30% to higher frequencies and is wider. In Figure 1-3(c), random distortion is introduced by randomly varying hole size by 20% of the lattice constant. It is immediately obvious how less the bands are affected by random

fluctuations. This is partially due to overall average refractive index remaining the same. For purely random variations, bands should retain their symmetry. The small degree of asymmetry observed in this example is attributed to a slight anisotropy of the supercell used in the calculation, so it is a purely numerical artifact. It should be noted that while random distortion is visually most dramatic, it is the more subtle deterministic distortion that is most serious. For this reason, it is important to predict optical behavior of nano-optical elements based on realistic geometry.
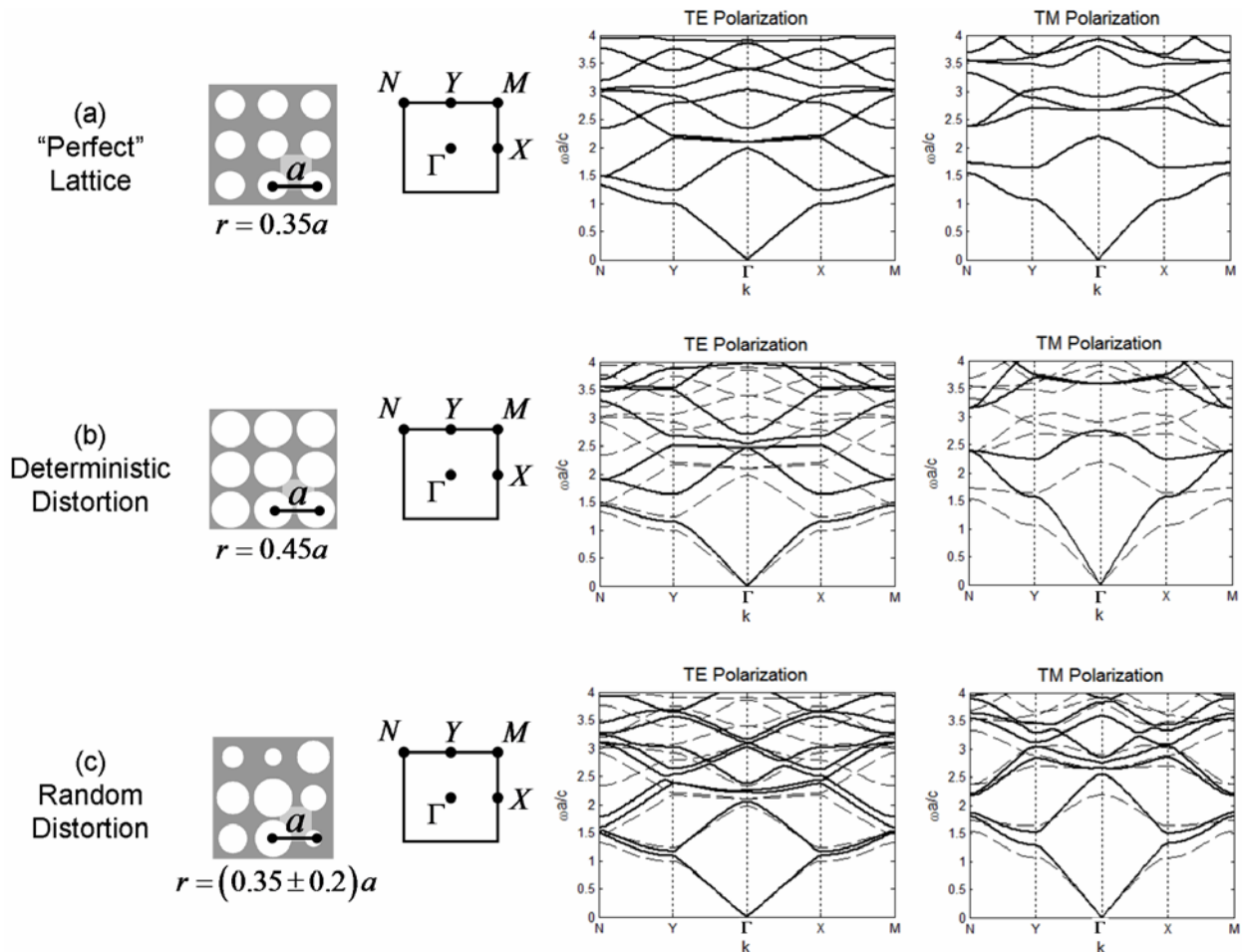
Figure 1-3: Comparison between random and deterministic distortion

To demonstrate the impact of realistic geometry on a real structure, reflection from two face-centered-cubic photonic crystals are compared in Figure 1-4. A depiction of the unit cells are shown at the left side of the diagram. The "perfect" geometry is rendered in blue, while a more realistic geometry resulting from holographic lithography is rendered in red. Reflection from a ten layer slab from both lattices are compared at the right side of the diagram. Width of the band gap is slightly wider for the realistic case and red-shifted to a longer wavelength. Peak reflection of the realistic structure is lower. A kink observed in the spectrum of the realistic photonic crystal suggests there is interesting dispersive behavior that is completely missed when perfect geometry is assumed.
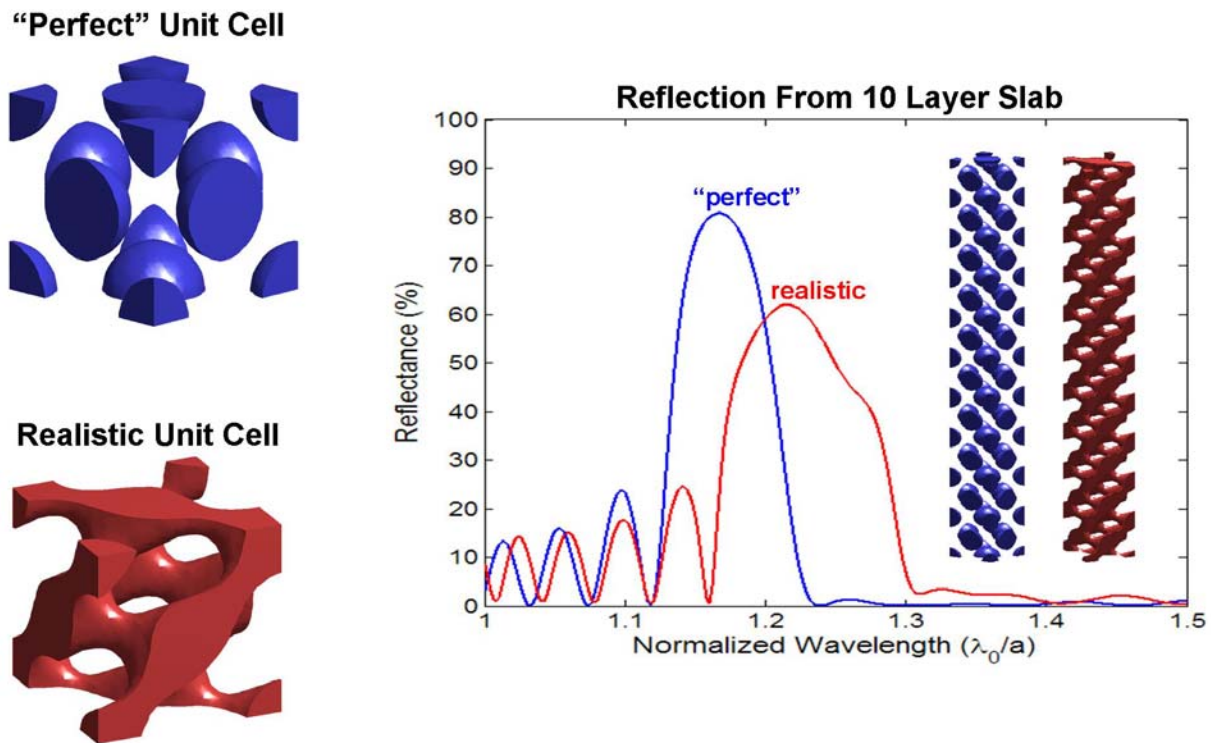
Figure 1-4: Comparison of "perfect" and realistic FCC photonic crystal

1.5. General Considerations for Numerical Modeling

To predict geometry of NOEs, key steps in fabrication must be identified and modeled accurately. A background in suitable numerical methods is helpful. It is best to seek techniques capable of handling large scale problems with high volumetric complexity since realistic devices are rarely comprised of perfect geometric patterns and can be quite large in terms of the optical wavelengths of interest.

For electromagnetic simulations, finite-element methods [40] and finite-difference methods [41-47] are attractive, particularly when metals are incorporated. For 2D simulations, the finite-difference frequency-domain (FDFD) method is accurate, simple to implement, and excellent for modeling complex structures of finite size [45, 46]. The transfer matrix method [42, 48] is popular for 2D and 3D structures, especially when they contain metals. For layered and periodic dielectric devices with moderate to low index contrast, rigorous coupled-wave analysis (RCWA) [49-52] is very fast and efficient. Finite-difference time-domain (FDTD) [43, 44] remains a very powerful tool for modeling finite sized devices, large scale structures, and characterizing devices over very broad frequency range. FDTD has difficulty resolving sharp resonances or any abrupt change in spectral response, but it is exceptionally capable of identifying if resonances are present and locating their position in frequency.
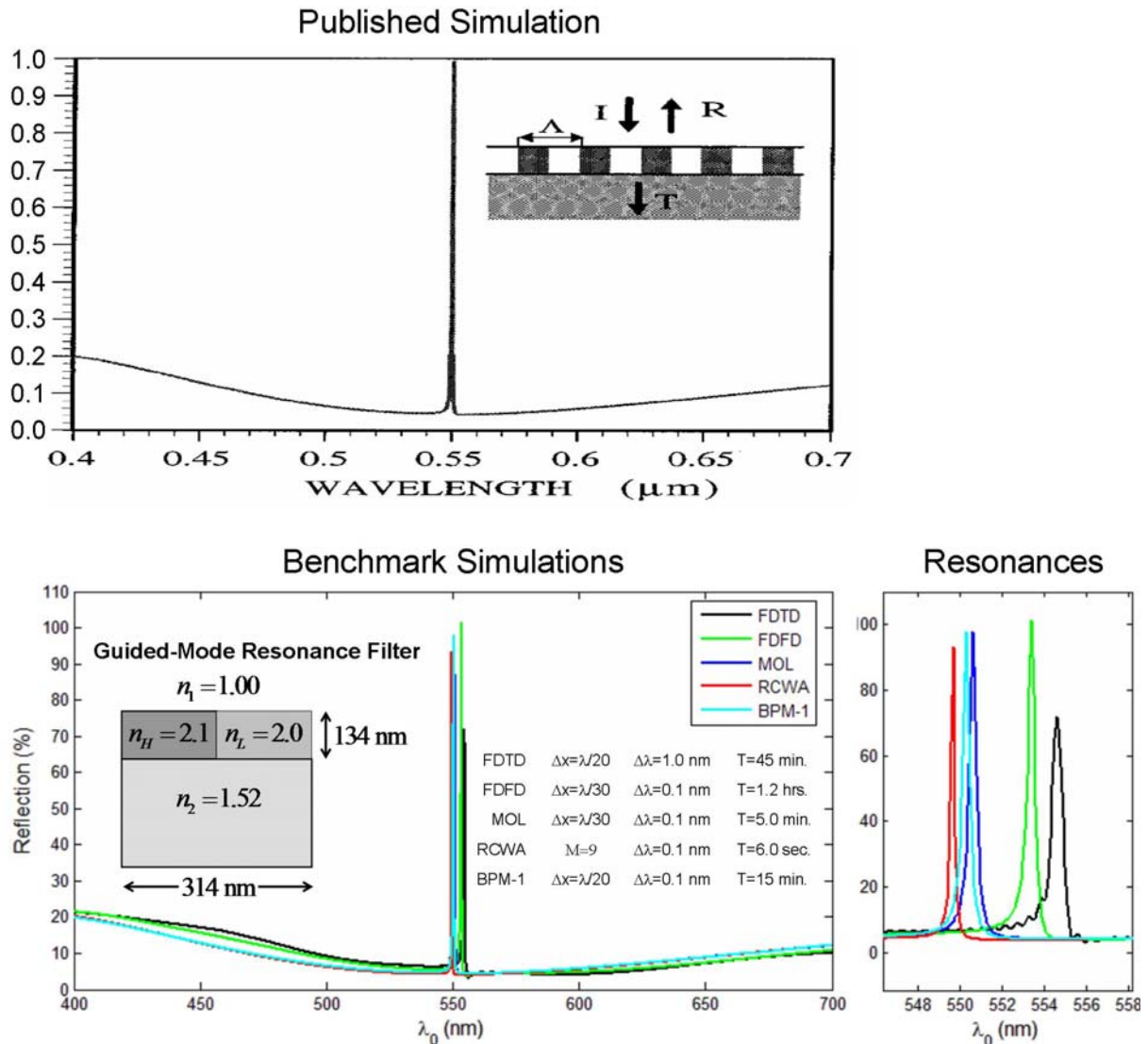
Figure 1-5: Benchmark simulations for EM simulation tools

Figure 1-5 compares theoretical results obtained in Ref. [53] to benchmark simulations for all electromagnetic simulation tools outlined in this dissertation. Numerical resolution, spectral resolution, and simulation time is summarized in this figure. A close-up of the simulated resonances is provided in the bottom right diagram. Numerical dispersion has shifted the resonance peaks to slightly longer wavelengths in the models employing finite-differences.

Etching, developing, and deposition processes can be modeled by tracking progression of a surface. There are four popular approaches found in literature. These are string methods [39, 54, 55], cell volume methods [3, 56-60], level set methods (LSM) [61-63], and fast marching methods (FMM) [63]. String, or marker particle, methods resolve the surface with a number of discrete points. During simulation, position of each point is updated based on position of surrounding points and a local rate function. These methods are very fast and efficient when surface topology is simple, smooth, and continuous. This is the case for much of grayscale lithography and analog devices. When structures are complex with detached surfaces and sharp bends, the method can become unstable and is difficult to implement. For these reasons, the string method is mostly limited to 2D problems.

Cell volume methods represent bulk materials as an array of cube shaped cells. Each cell is assigned a fictitious volume that is adjusted as the surface progresses. Cells with zero volume represent points where the bulk material does not exist. Cells where the volume is filled represent points where the bulk material is present. Cells with partially filled volumes interpolate position of the surface. The main trick implementing these methods is modifying how cell volume is reduced based on the direction the surface is moving through the cell. This is a difficult calculation in the cell volume framework and fine grids are needed to obtain accurate results. For this reason, cell volume methods require large amounts of memory and run slower than string methods, but are able to handle arbitrarily complex surfaces and extend well to three dimensions.

Of all techniques, LSM and FMM seem to offer the greatest compromise between speed, flexibility, and efficiency while being able handle very complex geometries. Level set methods may be the most generalized and most rigorous technique for modeling evolution of surfaces.

They are best tailored for modeling etching and deposition due to the bidirectional nature of the surface and surface dependant growth rates encountered in the process. Fast marching methods are incredibly fast and accurate algorithms best suited to model processes like developing where surfaces progress in only one direction and rate depends only on position. Both LSM and FMM offer much simpler means of calculating geometric properties of the surface such as surface normal and curvature.

Table 1-1: Maxwell's equations

Integral Form	Differential Form	Description	
$\oiint_S \vec{D}(t) \cdot d\vec{S} = \int_V \rho_v(t)\, dv$	$\nabla \cdot \vec{D}(t) = \rho_v(t)$	Gauss' law	Time-Domain
$\oiint_S \vec{B}(t) \cdot d\vec{S} = 0$	$\nabla \cdot \vec{B}(t) = 0$	No magnetic charge	
$\oint_L \vec{E}(t) \cdot d\vec{L} = -\dfrac{\partial}{\partial t} \iint_S \vec{B}(t) \cdot d\vec{S}$	$\nabla \times \vec{E}(t) = -\dfrac{\partial \vec{B}(t)}{\partial t}$	Faraday's law	
$\oint_L \vec{H}(t) \cdot d\vec{L} = -\iint_S \left[\vec{J}(t) + \dfrac{\partial \vec{D}(t)}{\partial t} \right] \cdot d\vec{S}$	$\nabla \times \vec{H}(t) = \vec{J}(t) + \dfrac{\partial \vec{D}(t)}{\partial t}$	Ampere's circuit law	
$\vec{D}(t) = \tilde{\varepsilon}(t) * \vec{E}(t)$	$\underline{B}(t) = \tilde{\mu}(t) * \vec{H}(t)$		
$\oiint_S \vec{D} \cdot d\vec{S} = \int_V \rho_v\, dv$	$\nabla \cdot \vec{D} = \rho_v$	Gauss' law	Frequency-Domain
$\oiint_S \vec{B} \cdot d\vec{S} = 0$	$\nabla \cdot \vec{B} = 0$	No magnetic charge	
$\oint_L \vec{E} \cdot d\vec{L} = -j\omega \iint_S \vec{B} \cdot d\vec{S}$	$\nabla \times \vec{E} = -j\omega \vec{B}$	Faraday's law	
$\oint_L \vec{H} \cdot d\vec{L} = -\iint_S (\vec{J} + j\omega \vec{D}) \cdot d\vec{S}$	$\nabla \times \vec{H} = \vec{J} + j\omega \vec{D}$	Ampere's circuit law	
$\vec{D} = \tilde{\varepsilon} \vec{E}$	$\underline{B} = \tilde{\mu} \vec{H}$		

1.6. Maxwell's Equations and Theorems

Maxwell's equations are the fundamental laws that govern behavior of all electromagnetic energy. They are based on a macroscopic view of how waves interact with matter and may take many forms. To understand optical behavior in nano-optical elements, it is necessary to have an understanding of Maxwell's equations and how they are applied. A summary of Maxwell's equations is provided in Table 1-1.

1.6.1.   Time-Domain Form

The most widely used form of Maxwell's equations are differential equations, as opposed to integral equations. In time-domain form [64], these are

$$\nabla \bullet \vec{D}(t) = \rho_v(t) \tag{1.1}$$

$$\nabla \bullet \vec{B}(t) = 0 \tag{1.2}$$

$$\nabla \times \vec{E}(t) = -\frac{\partial \vec{B}(t)}{\partial t} \tag{1.3}$$

$$\nabla \times \vec{H}(t) = \vec{J}(t) + \frac{\partial \vec{D}(t)}{\partial t} \tag{1.4}$$

The field parameters $\vec{D}$, $\vec{E}$, $\vec{B}$, and $\vec{H}$ represent electric flux density (C/m^2), electric field intensity (V/m), magnetic flux density (W/m^2), and magnetic field intensity (A/m) respectively. Electric charge density ρ_v (C/m^3) and electric current density $\vec{J}$ (A/m^2) represent sources that can induce electromagnetic fields or be induced by the fields.

All materials are comprised of charged particles that are displaced in the presence of an applied electric field. Accelerating charges radiate fields that combine out of phase with the applied wave. The overall effect is to modify propagation so waves behave differently than they would in a vacuum. To incorporate these effects in a simplified manner, they are treated

macroscopically through the constitutive parameters $\overline{\varepsilon}(t)$, $\overline{\mu}(t)$, and sometimes $\overline{\sigma}(t)$. In general, these are time-varying tensor quantities that relate to the fields through the following equations that involve convolutions.

$$\vec{D}(t) = \overline{\varepsilon}(t) * \vec{E}(t) \tag{1.5}$$

$$\vec{B}(t) = \overline{\mu}(t) * \vec{H}(t) \tag{1.6}$$

$$\vec{J}(t) = \overline{\sigma}(t) * \vec{E}(t) \tag{1.7}$$

The dielectric permittivity $\overline{\varepsilon}$ characterizes how well a material can store energy imposed by an electric field. Permeability $\overline{\mu}$ quantifies how efficiently energy is stored due to an applied magnetic field. The conductivity $\overline{\sigma}$ arises from free charges that form electrical currents in the presence of an applied electric field.

1.6.2.   *Frequency-Domain Form*

Transforming Eqs. (1.1)-(1.7) to the frequency-domain simplifies the mathematical framework by reducing convolutions to simple products.

$$\nabla \bullet \vec{D} = \rho_v \tag{1.8}$$

$$\nabla \bullet \vec{B} = 0 \tag{1.9}$$

$$\nabla \times \vec{E} = -j\omega\vec{B} \tag{1.10}$$

$$\nabla \times \vec{H} = \vec{J} + j\omega\vec{D} \tag{1.11}$$

$$\vec{D} = \overline{\varepsilon}\vec{E} \tag{1.12}$$

$$\vec{B} = \overline{\mu}\vec{H} \tag{1.13}$$

$$\vec{J} = \overline{\sigma}\vec{E} \tag{1.14}$$

In this form, it is possible to eliminate the terms $\vec{D}$, $\vec{B}$ and $\vec{J}$ from Maxwell's equations.

$$\nabla \bullet \left(\overline{\varepsilon} \vec{E} \right) = q \tag{1.15}$$

$$\nabla \bullet \left(\overline{\mu} \vec{H} \right) = 0 \tag{1.16}$$

$$\nabla \times \vec{E} = -j\omega \overline{\mu} \vec{H} \tag{1.17}$$

$$\nabla \times \vec{H} = \overline{\sigma} \vec{E} + j\omega \overline{\varepsilon} \vec{E} \tag{1.18}$$

Most often permittivity and permeability are not tensor quantities. In addition, charge ρ_v is usually ignored and conductivity is incorporated into a complex permittivity term. Under these conditions, Maxwell's equations assume their most familiar form.

$$\nabla \bullet \left(\varepsilon \vec{E} \right) = 0 \tag{1.19}$$

$$\nabla \bullet \left(\mu \vec{H} \right) = 0 \tag{1.20}$$

$$\nabla \times \vec{E} = -j\omega\mu \vec{H} \tag{1.21}$$

$$\nabla \times \vec{H} = j\omega\varepsilon \vec{E} \tag{1.22}$$

1.6.3.   Wave Equations

Perhaps the most significant aspect of Maxwell's equations is that they predict propagating waves. The curl equations show that time changing electric fields induce curling, or circulating, magnetic fields in the immediate vicinity. Likewise, time changing magnetic fields induce circulating electric fields. In this manner, oscillating fields continue to induce other oscillating fields and a wave is formed. To better understand the wave phenomenon, it makes sense to combine Maxwell's curl equations since it is the interaction of these two equations that predicts propagation. The combined equation is called a "wave equation" and it is possible to

derive it in terms of just the magnetic field or just the electric field. Wave equations enable fully

rigorous analysis of electromagnetic problems using a single equation while incorporating all

information from both of Maxwell's curl equations.

To begin deriving full vector wave equations, Eqs. (1.21)-(1.22) are written as

$$(1/\mu)\nabla \times \vec{E} = -j\omega\vec{H} \tag{1.23}$$

$$(1/\varepsilon)\nabla \times \vec{H} = j\omega\vec{E} \tag{1.24}$$

Taking the curl of these equations leads to

$$\nabla \times (1/\mu)\nabla \times \vec{E} = -j\omega\left(\nabla \times \vec{H}\right) \tag{1.25}$$

$$\nabla \times (1/\varepsilon)\nabla \times \vec{H} = j\omega\left(\nabla \times \vec{E}\right) \tag{1.26}$$

The curl operations on the right side of these equations can be replaced using the original

curl equations. This leads to the full vector wave equations.

$$\nabla \times \frac{1}{\mu}\nabla \times \vec{E} = \omega^2\varepsilon\vec{E} \tag{1.27}$$

$$\nabla \times \frac{1}{\varepsilon}\nabla \times \vec{H} = \omega^2\mu\vec{H} \tag{1.28}$$

The wave equations can be written in terms of the free space wave number k_0, where

$k_0 = \omega/c$.

$$\nabla \times \frac{1}{\mu_r}\nabla \times \vec{E} = k_0^2\varepsilon_r\vec{E} \tag{1.29}$$

$$\nabla \times \frac{1}{\varepsilon_r}\nabla \times \vec{H} = k_0^2\mu_r\vec{H} \tag{1.30}$$

1.6.4.   Solution in Homogeneous Materials

When materials are homogeneous, the wave equation reduces to the famous Helmholtz equation. For nonmagnetic materials, Eq. (1.29) becomes

$$\nabla^2 \vec{E} - k_0^2 \varepsilon_r \vec{E} = 0 \tag{1.31}$$

Here, the wave equation has a plane wave solution. When permittivity is complex, the wave vector is complex indicating the wave decays exponentially due to loss in the material.

$$\vec{E}(\vec{r}) = \vec{E}_0 \exp\left(-j\vec{k} \bullet \vec{r}\right) \tag{1.32}$$

$$\left|\vec{k}\right|^2 = k_0^2 \varepsilon_r \tag{1.33}$$

1.6.5.   Scaling Properties

A very useful feature of Maxwell's equations is that there is no fundamental length scale. Systems may be scaled to operate at different frequencies simply by scaling dimensions or material properties proportional to the change in frequency. This assumes it is physically possible to scale systems in this manner. In practice, building smaller structures and having arbitrary control over material properties may be difficult.

The first useful scaling property is when dimensions of a system are compressed or stretched by a scaling factor v. The wave equation in the scaled system would be written as

$$(v\nabla) \times \frac{1}{\varepsilon(\vec{r}/v)}(v\nabla) \times \vec{H}\left(\frac{\vec{r}}{v}\right) = \omega^2 \mu\left(\frac{\vec{r}}{v}\right)\vec{H}\left(\frac{\vec{r}}{v}\right) \tag{1.34}$$

Material parameters in the scaled system can be written as $\varepsilon'(\vec{r}) = \varepsilon(\vec{r}/v)$ and $\mu'(\vec{r}) = \mu(\vec{r}/v)$. The wave equation now becomes

$$\nabla \times \frac{1}{\varepsilon'(\vec{r})}\nabla \times \vec{H}\left(\frac{\vec{r}}{v}\right) = \left(\frac{\omega}{v}\right)^2 \mu'(\vec{r})\vec{H}\left(\frac{\vec{r}}{v}\right) \tag{1.35}$$

21

This is simply the original wave equation where the field in the scaled system is $\vec{H}'(\vec{r}) = \vec{H}(\vec{r}/v)$ and frequency is scaled according to $\omega' = \omega/v$. Systems that are scaled versions of each other will behave exactly the same except the field patterns and associated frequencies will be scaled proportionally.

A second useful scaling property is when the material properties are scaled by factors a_ε and a_μ. The wave equation in the scaled system is

$$\nabla \times \frac{1}{(a_\varepsilon \varepsilon)} \nabla \times \vec{H} = \omega^2 (a_\mu \mu) \vec{H} \tag{1.36}$$

Grouping the scaling terms with ω yields

$$\nabla \times \frac{1}{\varepsilon} \nabla \times \vec{H} = \left(\omega \sqrt{a_\mu a_\varepsilon} \right)^2 \mu \vec{H} \tag{1.37}$$

Behavior in the scaled system remains unchanged other than frequency being scaled by factor $\sqrt{a_\mu a_\varepsilon}$. In optics, frequency scales linearly with refractive index.

$$\nabla \times \frac{1}{n^2} \nabla \times \vec{H} = \left(\frac{\omega a_n}{c_0} \right)^2 \vec{H} \tag{1.38}$$

1.6.6. Left-Handed Materials

In the 1960's, Russian physicist Victor Veselago studied the consequences of having both a negative permittivity and negative permeability [65, 66]. This was pure theory at the time because it was not yet known how to realize a negative permeability. He found the Poynting vector and wave vector had opposite sign and reasoned this was only possible if refractive index was negative. This principle is easily shown by writing Maxwell's curl equations with the negative sign associated with μ and ε written explicitly.

$$\nabla \times \vec{E} = -j\omega \left(-\tilde{\mu} \right) \vec{H} \tag{1.39}$$

22

$$\nabla \times \vec{H} = j\omega\left(-\tilde{\varepsilon}\right)\vec{E} \tag{1.40}$$

Associating the signs with the magnetic field leads to

$$\nabla \times \vec{E} = -j\omega\tilde{\mu}\left(-\vec{H}\right) \tag{1.41}$$

$$\nabla \times \left(-\vec{H}\right) = j\omega\tilde{\varepsilon}\vec{E} \tag{1.42}$$

This is simply Maxwell's equations with the sign of the magnetic field reversed. The Poynting vector is then

$$\vec{P} = \vec{E} \times \left(-\vec{H}\right) = -\vec{E} \times \vec{H} \tag{1.43}$$

The right-hand rule in electromagnetics [67] is not valid here because sign of the Poynting vector is reversed. In fact, materials with simultaneously negative permittivity and negative permeability follow an analogous left-hand rule. For this reason, such materials have been titled "left-handed."

At an interface between air and a left-handed material, transmitted energy must keep propagating forward by definition. The Poynting vector is defined to point in the direction of energy flow. Phase matching requires the tangential component of the wave vector to be continuous across the interface. The remaining degree of freedom is the longitudinal component of the wave vector. Signs can only be reconciled if it is made negative. This would be the case if refractive index were negative. For this reason, left-handed materials are assigned a negative refractive index.

1.7. Optical Properties of Bulk Materials

1.7.1. *Lorentz Oscillator Model for Dielectrics*

Motion of an electron bound to a positively charged nucleus is in many ways analogous to the motion of a mass on a spring. Electrostatic force replaces the elastic function of the spring. From mechanics, motion of this sort can be approximated by the Lorentz oscillator model [68].

$$m\frac{\partial^2 \vec{r}}{\partial t^2} + m\Gamma\frac{\partial \vec{r}}{\partial t} + m\omega_0^2 \vec{r} = -q\vec{E} \tag{1.44}$$

In this equation, $\vec{r}$ is position, m is mass of the electron, Γ characterizes loss and is analogous to friction, $m\omega_0^2$ arises from a restoring force analogous to the spring with resonant frequency ω_0, and $q\vec{E}$ is an applied electric force from an electromagnetic wave. This simple model assumes the system is linear. Figure 1-6 illustrates this concept and compares it to the traditional mechanical model of a spring and damper system.

For time-harmonic fields, the equation of motion can be written as the Fourier transform of Eq. (1.44).

$$-m\omega^2 \vec{r}(\omega) - j\omega m\Gamma \vec{r}(\omega) + m\omega_0^2 \vec{r}(\omega) = -q\vec{E}(\omega) \tag{1.45}$$

Solving this for $\vec{r}$ leads to

$$\vec{r}(\omega) = -\frac{q}{m}\frac{\vec{E}(\omega)}{\omega_0^2 - \omega^2 - j\omega\Gamma} \tag{1.46}$$

As $\vec{r}$ deviates from equilibrium, charge is displaced forming a "dipole moment," or offset charge $\vec{\mu}$. This is

$$\vec{\mu}(\omega) = -q\vec{r}(\omega) = \frac{q^2}{m}\frac{\vec{E}(\omega)}{\omega_0^2 - \omega^2 - j\omega\Gamma} \tag{1.47}$$

EQUILIBRIUM STATE

spring
ω_0

damper
Γ

mass
m

atomic nucleus

no applied electric field

no material polarization

electron "cloud"

POLARIZED STATE

$\vec{E}$
applied electric field

$\vec{P}$
material polarization

displaced electron "cloud"

displaced mass
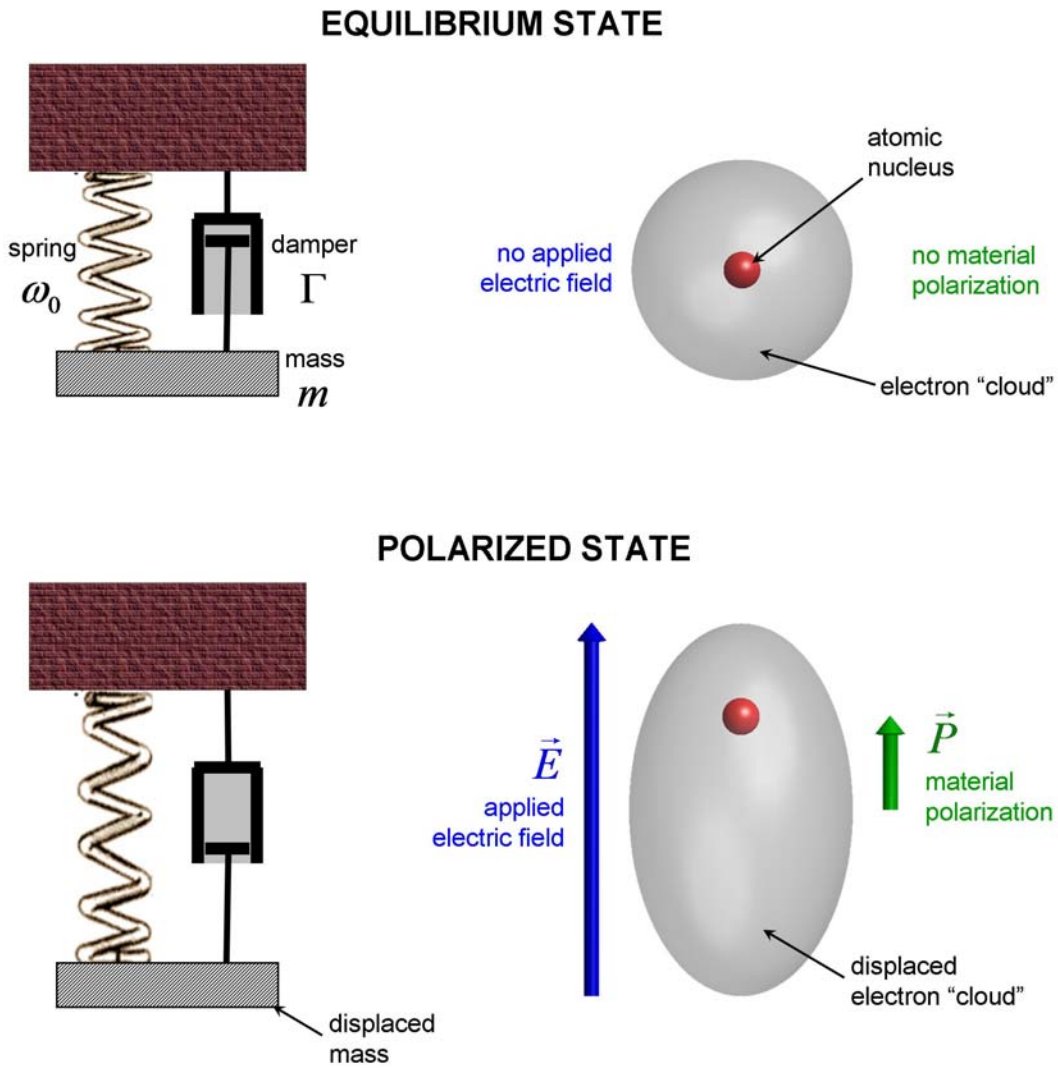
Figure 1-6: Lorentz oscillator model

Material polarization is a macroscopic concept defined as the average dipole moment throughout the volume of the material.

$$\vec{P}(\omega) = N \langle \vec{\mu}(\omega) \rangle = \frac{Nq^2}{m_{\text{eff}}} \frac{\vec{E}(\omega)}{\omega_0^2 - \omega^2 - j\omega\Gamma} \tag{1.48}$$

In this equation, the mass is replaced with the "effective" mass of the electron and parameter N represents density of atoms. The constitutive equation relating $\vec{D}$ and $\vec{E}$ can be written in terms of the material polarization as

$$\vec{D} = \tilde{\varepsilon}\vec{E} = \varepsilon_0\vec{E} + \vec{P} \tag{1.49}$$

From this equation, the dielectric function is related to material polarization through

$$\vec{P} = \varepsilon_0\left(\tilde{\varepsilon}_r - 1\right)\vec{E} \tag{1.50}$$

Comparing Eq. (1.48) to Eq. (1.50) leads to an expression for the dielectric function in terms of the Lorentz parameters.

$$\tilde{\varepsilon}_r = 1 + \frac{\omega_p^2}{\omega_0^2 - \omega^2 - j\omega\Gamma} \tag{1.51}$$

$$\omega_p^2 = \frac{Nq^2}{\varepsilon_0 m_{\text{eff}}} \tag{1.52}$$

Equation (1.51) is referred to as the Lorentz model for dielectrics and predicts electromagnetic behavior of bulk materials with surprising accuracy. In this equation, some parameters have been collected into a new term called the "plasma" frequency ω_p. A plot of a typical dielectric response based on this equation is provided in Figure 1-7. The Lorentz model correctly predicts high absorption near resonance tapering to zero away from resonance. At frequencies just above resonance, the model predicts dielectric constant dips very low and can even be less than one. Far above resonance, it predicts the dielectric function asymptotically approaches 1.0 and the material is essentially transparent. At frequencies below resonance, it predicts the dielectric function asymptotically approaches a value greater than 1 determined by the plasma and resonant frequencies.

$$\tilde{\varepsilon}_r(0) = 1 + \omega_p^2/\omega_0^2 \tag{1.53}$$

26

$$\tilde{\varepsilon}_r(\infty) = 1 \qquad (1.54)$$

Real materials show multiple resonances of many types and are best described by a linear sum of Lorentz oscillators. This can be expressed generally as

$$\tilde{\varepsilon}_r = 1 + \omega_p^2 \sum_i \frac{f_i}{\omega_i^2 - \omega^2 - j\omega\Gamma} \qquad (1.55)$$

The parameter f_i is the "oscillator strength" of each Lorentz oscillator. It characterizes the probability that atoms will take part in absorption or emission of radiation.

Dielectric Function

$$\varepsilon''(\omega_0) = 1 + \frac{\omega_p^2}{\omega_0 \Gamma}$$

$$\varepsilon(\omega) = \varepsilon'(\omega) + i\varepsilon''(\omega)$$

$$\varepsilon'(0) = 1 + \omega_p^2 / \omega_0^2$$

$$\varepsilon'(\omega)$$

$$\varepsilon''(\omega)$$

$\rightarrow \; \leftarrow \Gamma$

Refractive Index

$$\eta(\omega) = n(\omega) + i\kappa(\omega)$$

$$n(0) = \sqrt{1 + \omega_p^2 / \omega_0^2}$$

$$n(\omega)$$

$$\kappa(\omega)$$

$$\omega_0$$
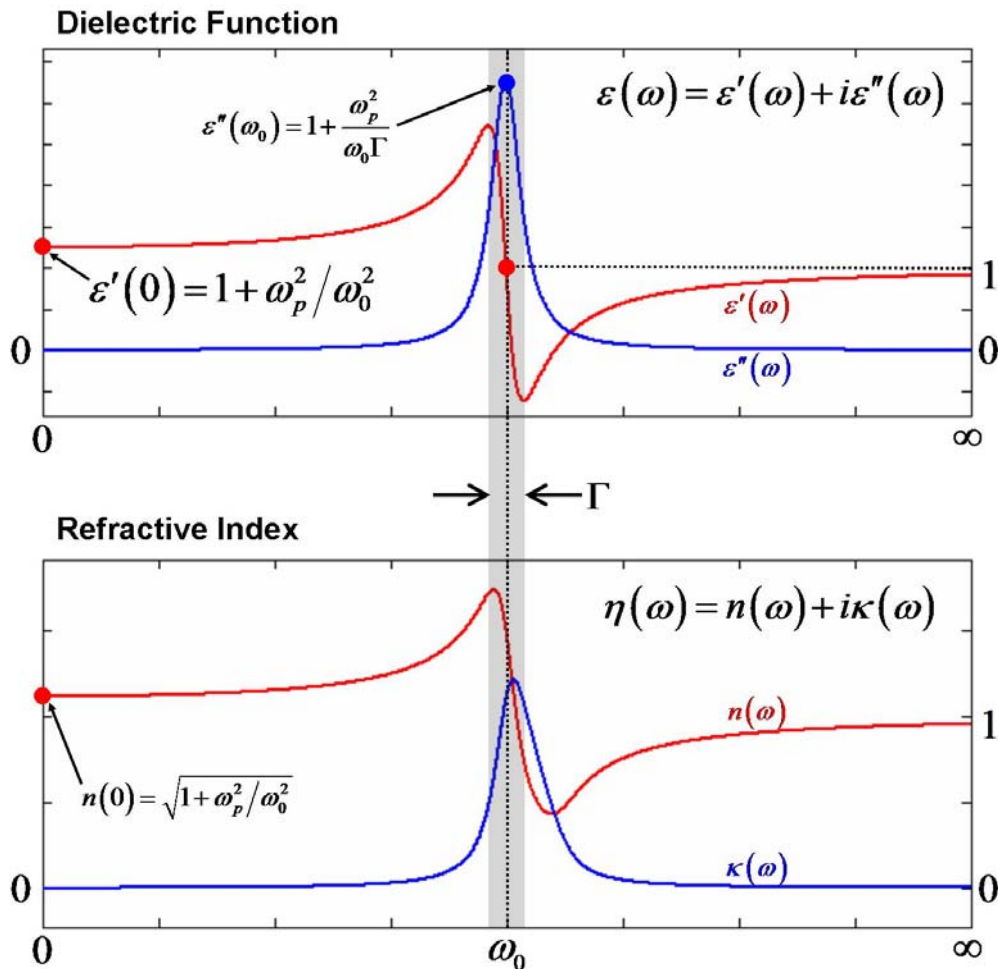
Figure 1-7: Typical dielectric response of Lorentz oscillator

1.7.2. Drude Model for Metals

In metals, most electrons are free and not bound to a nucleus so restoring force is negligible. Based on this reasoning, the Drude model for metals can be derived from the Lorentz model by assuming $\omega_0 \approx 0$ [68]. This is

$$\tilde{\varepsilon}_r = 1 - \frac{\omega_p^2}{\omega^2 + j\omega\Gamma} \tag{1.56}$$

Material properties can be treated as a real-valued permittivity and a conductivity. In terms of the mean collision rate defined as $\tau = 1/\Gamma$, these are

$$\varepsilon_r = 1 - \frac{\omega_p^2 \tau^2}{1 + \omega^2 \tau^2} \tag{1.57}$$

$$\sigma = \frac{\sigma_0}{1 + \omega^2 \tau^2} \qquad \sigma_0 = \varepsilon_0 \omega_p^2 \tau \tag{1.58}$$

The parameter σ_0 is conductivity at $\omega = 0$. From these equations, it can be seen that at very high frequencies, loss vanishes and metals become transparent.

The plasma frequency ω_p for typical metals lies in the ultraviolet region. Its role can be more clearly understood if the dielectric function is separated into real and imaginary parts.

$$\tilde{\varepsilon}_r = \left(1 - \frac{\omega_p^2 \tau^2}{1 + \omega^2 \tau^2}\right) + j\left(\frac{\tau \omega_p^2 / \omega}{1 + \omega^2 \tau^2}\right) \tag{1.59}$$

From this equation it can be seen that at frequencies below the plasma frequency, dielectric constant is mostly imaginary and metals behave like good conductors. Near the plasma frequency, metals are very lossy because both real and imaginary parts are significant. Above the plasma frequency, the imaginary part of the dielectric function is very small

indicating metals in this region are weakly absorbing. In addition, the real part becomes less than one.

1.7.3. *Skin Depth and Loss at Low Frequencies*

Most applications use frequencies well below ultraviolet so behavior of metals in the low frequency limit is of great interest. From radio frequencies to infrared, $\omega \ll (1/\tau) \ll \omega_p$ and Eqs. (1.57)-(1.58) reduce to

$$\varepsilon_r \approx 1 \tag{1.60}$$

$$\sigma = \sigma_0 \tag{1.61}$$

The complex dielectric function in this case is essentially entirely imaginary because $1 \ll \sigma/\omega$.

$$\tilde{\varepsilon}_r \approx j\frac{\sigma_0}{\omega} \tag{1.62}$$

From Eq. (1.33), the complex wave number of a wave propagating through a metal can be written as

$$k = (1+j)\sqrt{\frac{\sigma_0 \omega}{2c^2}} \tag{1.63}$$

The imaginary part of the wave number dictates how quickly the wave will attenuate due to loss. The "skin depth" $\delta(\omega)$ is defined as the distance a wave propagates where amplitude decays to a factor of $1/e$ from its starting amplitude. This corresponds to the inverse of the absorption coefficient $\alpha(\omega)$ and is calculated from Eq. (1.63) to be

$$\delta(\omega) = \frac{1}{\alpha(\omega)} = \sqrt{\frac{2c^2}{\sigma_0 \omega}} \tag{1.64}$$

From this equation it can be concluded that higher frequencies experience greater loss and decay faster. For this reason, metallic structures are most useful at lower frequencies.

Refractive index of a metal can be derived from Eq. (1.62) through the relation $\tilde{n}^2 = \tilde{\varepsilon}_r$. This is

$$\tilde{n} = (1 + j)\sqrt{\frac{\sigma_0}{2\omega}} \tag{1.65}$$

When light is incident from air onto a metal, the reflection coefficient is given by

$$R = \left(\frac{\tilde{n} - 1}{\tilde{n} + 1}\right)^2 \tag{1.66}$$

At very low frequencies, refractive index becomes very large, reflection approaches 100%, and metals behave much like "perfect" electric conductors. Greater reflection loss will occur at higher frequencies.

Table 1-2: Optical properties of selected metals at 1550 nm

Metal	Symbol	Plasma Res. λ (nm)	Refractive Index n	Refractive Index k	Drude ε	Conductivity σ (S/m)	Reflectance R (90°)	Absorption α (1/μm)	Skin Depth δ (nm)
Aluminum	Al	82.78	1.444	15.955	-252.4769	5.60E+16	97.78%	129.3524	7.7308
Chromium	Cr	115.35	4.200	4.950	-6.8625	5.06E+16	67.50%	40.1313	24.9182
Copper	Cu	114.50	0.500	8.000	-63.75	9.73E+15	97.70%	64.8586	15.4181
Gold	Au	137.32	0.180	10.210	-104.2117	4.47E+15	99.30%	82.7758	12.0808
Nickel	Ni	77.89	3.380	6.820	-35.088	5.61E+16	79.40%	55.2920	18.0858
Silver	Ag	137.62	0.400	10.000	-99.84	9.73E+15	98.99%	81.0733	12.3345

Table 1-2 summarizes the optical properties of commonly used metals at a wavelength of 1550 nm. Data was borrowed from Refs. [69, 70].

CHAPTER 2
PHOTONIC CRYSTALS

2.1. Overview

Photonic crystals are a broad class of artificial materials that exploit scattering and resonance in periodic structures to control propagation of electromagnetic energy. They are called "photonic crystals" to compare the behavior of photons to the behavior of electrons in real crystals. Most literature credits the pioneering work of Eli Yablonovitch [71] and Sajeev John [72], however as early as 1919 Guglielmo Marconi patented a class of periodic metallic structures for reflecting and filtering radio waves [73]. The concept of constructing artificial dielectrics from periodic structures arose in the 1950's [74]. An excellent review of the early work in periodic structures can be found in Refs. [74, 75].

The purpose of a photonic crystal is to engineer the bulk electromagnetic properties of a material and to control photons similar to how electrons are controlled in semiconductors. Many novel phenomena have been observed or predicted which promise significant impact on future photonic systems. Photonic band gaps can improve efficiency of optoelectronic devices by suppressing emission of photons within the band gap [71, 76]. Minibands within the band gap may offer even higher degrees of spectral selectivity for spectroscopy and dense wavelength division multiplexing (DWDM) [77]. Photonic band gaps enable waveguiding around tight bends [78-80] which can accommodate denser waveguide routing and may lead to a host of new and novel devices. Highly dispersive photonic crystals [81, 82] can compensate for large amounts of channel dispersion, slow the propagation of light, or perform difficult phase transfer

31

functions. Negative refraction [82, 83] may lead to dramatically improved imaging optics or improve the quality of micro-optic devices such as micro-mirrors and micro-lenses.

One-dimensional and two-dimensional photonic crystals have received considerable attention in literature primarily due to their ease of simulation, analysis, and fabrication compared to their three-dimensional counterparts. While 1D and 2D crystals have many applications, 3D crystals have at least two important advantages. First, they are capable of having a complete photonic band gap where propagation is forbidden at any angle and polarization. Second, they are capable of localizing photons at a point defect in the lattice.

This chapter will introduce the concept of photonic crystals and describe the general symmetries they may have in terms of the famous Bravais lattices. A mathematical framework will be presented starting with a description of lattice vectors, reciprocal lattices, Wigner-Seitz primitive cells, and Brillouin zones. A brief overview of the wave equation as an eigen-value problem will lead to some intuitive conclusions to help explain optical properties of photonic crystals. Construction of photonic band diagrams and how they may be interpreted will be presented. The chapter will conclude with a qualitative description of the physical origin of dispersion and band gaps in photonic crystals. The special case of metallic photonic crystals will be discussed in CHAPTER 13.

2.2. Photonic Crystal Symmetry

Much can be inferred about the behavior of a photonic crystal just from knowing its symmetry. For this reason, it is important to understand how symmetries are classified and how they may be described quantitatively. In solid state physics [84], a systematic and precise method was developed for classifying crystal structures based on lattice "operations." These are ways of physically manipulating lattice symmetry and include translations, inversions,

reflections, rotations, rotation-reflections, and rotation-inversions. Symmetry can be deduced by performing combinations of lattice operations and checking to see if the crystal remains unchanged, or often worded as "taken into itself."

In three dimensions, there are 230 possible symmetries, or "space groups," that a lattice with a basis can have. A basis is a fundamental unit cell that is repeated to construct the overall lattice. More possibilities exist when quasi-crystals [85, 86] and Archimedian tilings [87] are considered. These lack a basis, but offer the highest degrees of symmetry known to exist. Thirty-two "points groups" can be defined when only lattice operations are used that leave one point fixed. This means no translational operations are performed.

When the fundamental objects in a crystal lattice are perfect spheres, the total number of space groups reduces to 14 and are called Bravais lattices. A Bravais lattice is defined as an infinite array of discrete points with an arrangement and orientation that appears exactly the same from whichever of the points the array is viewed [84]. Similarly, the point groups reduce to the famous seven crystal systems: cubic, tetragonal, orthorhombic, monoclinic, triclinic, rhombohedral (trigonal), and hexagonal. The concept of Bravais lattices and the seven crystal systems can be applied to crystals without spherical units, but they can only be used to describe the general manner in which a structure is periodic. They cannot describe the geometry of the physical objects comprising the lattice.

Figure 2-1 summarizes the seven crystal systems and enumerates the Bravais lattices belonging to each. The cubic and hexagonal lattices have been the most intensely studied due to their high degrees of symmetry.
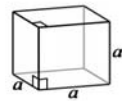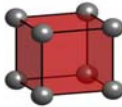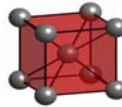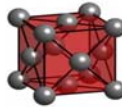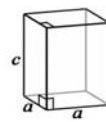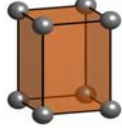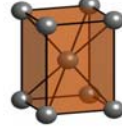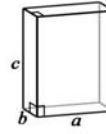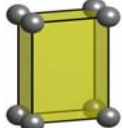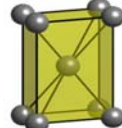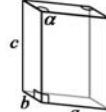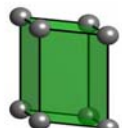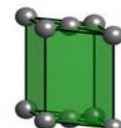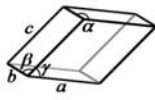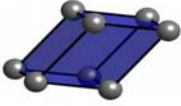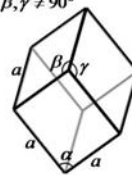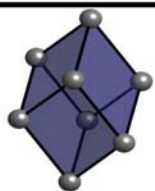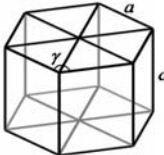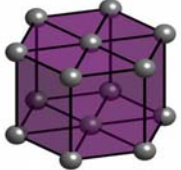
	Unit Cell	simple	body-centered	face-centered	base-centered
Cubic	a, a, a				
Tetragonal	$a \neq c$ c, a, a				
Orthorhombic	$a \neq b \neq c$ c, b, a				
Monoclinic	$\alpha \neq 90°$ $a \neq b \neq c$ α, c, b, a				
Triclinic	$a \neq b \neq c$ $\alpha, \beta, \gamma \neq 90°$ c, α, β, γ, b, a				
Rhombohedral (Trigonal)	$\alpha, \beta, \gamma \neq 90°$ a, β, γ, a, α, a				
Hexagonal	$\gamma = 120°$ a, γ, c				

Figure 2-1: The seven crystal systems and fourteen Bravais lattices

2.3. Mathematical Framework

2.3.1. Lattice Vectors

All Bravais lattices can be described by a set of three lattice vectors. These can be either axis vectors or translation vectors and are illustrated in Figure 2-2. Primitive axis vectors point along the edges of a unit cell in the direction of the lattice axes. The magnitudes of these vectors are the lattice constants. Axis vectors most intuitively describe size and orientation of the crystal unit cell, but do not completely describe the symmetry. For example, simple cubic, body-centered cubic, and face-centered cubic crystals all share the same axis vectors.
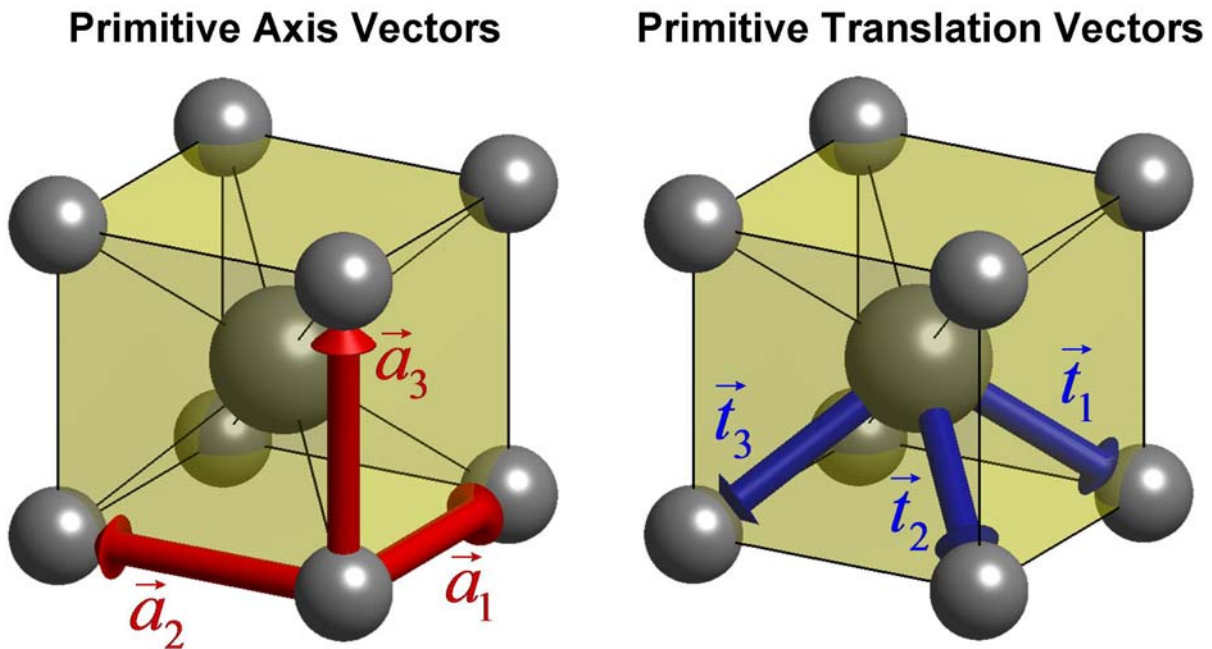
Figure 2-2: Axis and translation vectors

Primitive translation vectors $\left[\vec{t}_1, \vec{t}_2, \vec{t}_3\right]$ connect adjacent lattice points within the unit cell in a manner that completely characterizes size, orientation, and symmetry of the entire crystal.

Choice of these vectors is not unique. If chosen correctly, any two lattice points anywhere in the

whole crystal can be connected through an integer combination of the primitive translation

vectors. The composite vector $\vec{g}$ is also called a translation vector, but it is not necessarily

primitive.

$$\vec{g} = p\vec{t_1} + q\vec{t_2} + r\vec{t_3} \qquad \begin{aligned} p &= ..., -2, -1, 0, +1, +2, ... \\ q &= ..., -2, -1, 0, +1, +2, ... \\ r &= ..., -2, -1, 0, +1, +2, ... \end{aligned} \qquad (2.1)$$

Calculating translation vectors from axis vectors is possible if the symmetry is known.

The following equations summarize the conversion.

$$\text{Simple:} \quad \begin{bmatrix} \vec{t_1} \\ \vec{t_2} \\ \vec{t_3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \vec{a_1} \\ \vec{a_2} \\ \vec{a_3} \end{bmatrix} \qquad (2.2)$$

$$\text{Body-Centered:} \quad \begin{bmatrix} \vec{t_1} \\ \vec{t_2} \\ \vec{t_3} \end{bmatrix} = \begin{bmatrix} -1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & 1/2 \\ 1/2 & 1/2 & -1/2 \end{bmatrix} \begin{bmatrix} \vec{a_1} \\ \vec{a_2} \\ \vec{a_3} \end{bmatrix} \qquad (2.3)$$

$$\text{Face-Centered:} \quad \begin{bmatrix} \vec{t_1} \\ \vec{t_2} \\ \vec{t_3} \end{bmatrix} = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix} \begin{bmatrix} \vec{a_1} \\ \vec{a_2} \\ \vec{a_3} \end{bmatrix} \qquad (2.4)$$

$$\text{Base-Centered:} \quad \begin{bmatrix} \vec{t_1} \\ \vec{t_2} \\ \vec{t_3} \end{bmatrix} = \begin{bmatrix} -1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1/2 \end{bmatrix} \begin{bmatrix} \vec{a_1} \\ \vec{a_2} \\ \vec{a_3} \end{bmatrix} \qquad (2.5)$$

$$\text{Trigonal:} \quad \begin{bmatrix} \vec{t_1} \\ \vec{t_2} \\ \vec{t_3} \end{bmatrix} = \begin{bmatrix} 1 & 1/3 & 1/3 \\ -1/3 & 1/3 & 1/3 \\ -1/3 & -1 & 1/3 \end{bmatrix} \begin{bmatrix} \vec{a_1} \\ \vec{a_2} \\ \vec{a_3} \end{bmatrix} \qquad (2.6)$$

2.3.2. *Reciprocal Lattices*

It is often useful to construct a "reciprocal" lattice because it relates more directly to grating vectors and diffraction. Every Bravais lattice has a reciprocal lattice that can be described by a set of reciprocal lattice vectors $\left[\vec{T}_1, \vec{T}_2, \vec{T}_3 \right]$. The original crystal is called the "direct" lattice. The reciprocal of the reciprocal lattice is the direct lattice again. The reciprocal of a simple cubic lattice is simple cubic. Face-centered-cubic (FCC) and body-centered-cubic (BCC) are reciprocal pairs.
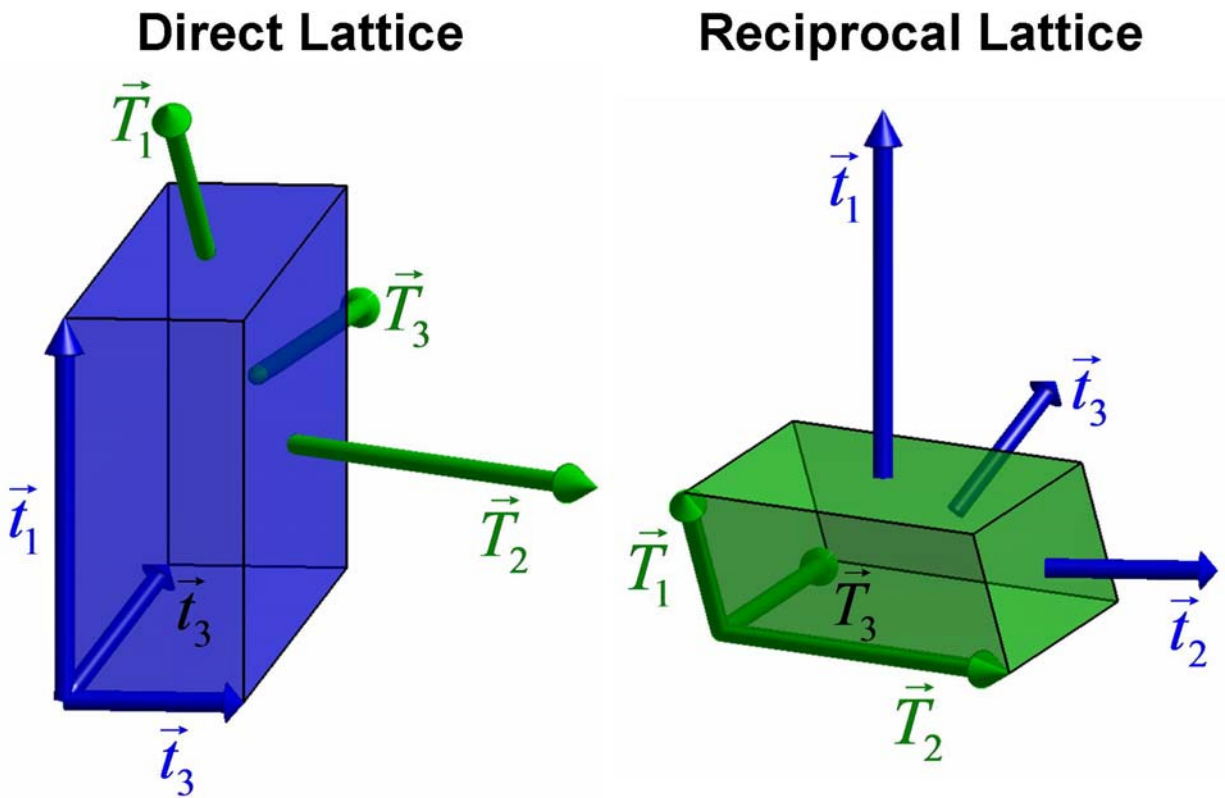
Figure 2-3: Direct and reciprocal lattices

Each pair of direct lattice vectors defines a plane from which a reciprocal lattice vector is defined. The reciprocal lattice vector is normal to this plane and has a magnitude equal to 2π divided by the spacing between adjacent planes. The geometric construction is illustrated in Figure 2-3. In this regard, reciprocal lattice vectors are equivalent to optical grating vectors where crystal planes of the direct lattice correspond to grating planes. This is a fundamental concept in grating diffraction and holography.

Reciprocal space is often called "k-space" because both grating vectors and wave vectors have a magnitude equal to 2π divided by a length. It is often preferred to analyze photonic crystals in k-space because of the elegant relation between wave vectors and lattice vectors. In addition, both spatial frequency and direction of propagation are conveyed simultaneously through the wave vector.

As with the direct lattice, primitive translation vectors of the reciprocal lattice connect lattice points within the same unit cell. Any two lattice points in the entire reciprocal lattice can be connected by a composite reciprocal translation vector that is an integer combination of the primitive reciprocal translation vectors.

$$\vec{G} = P\vec{T}_2 + Q\vec{T}_2 + R\vec{T}_3 \qquad \begin{aligned} P &= ...,-2,-1,0,+1,+2,... \\ Q &= ...,-2,-1,0,+1,+2,... \\ R &= ...,-2,-1,0,+1,+2,... \end{aligned} \qquad (2.7)$$

To convert between reciprocal lattice vectors and direct lattice vectors, the following vector equations can be used.

$$\vec{T}_1 = 2\pi \frac{\vec{t}_2 \times \vec{t}_3}{\vec{t}_1 \bullet (\vec{t}_2 \times \vec{t}_3)} \qquad \vec{T}_2 = 2\pi \frac{\vec{t}_3 \times \vec{t}_1}{\vec{t}_1 \bullet (\vec{t}_2 \times \vec{t}_3)} \qquad \vec{T}_3 = 2\pi \frac{\vec{t}_1 \times \vec{t}_2}{\vec{t}_1 \bullet (\vec{t}_2 \times \vec{t}_3)} \qquad (2.8)$$

$$\vec{t}_1 = 2\pi \frac{\vec{T}_2 \times \vec{T}_3}{\vec{T}_1 \bullet (\vec{T}_2 \times \vec{T}_3)} \qquad \vec{t}_2 = 2\pi \frac{\vec{T}_3 \times \vec{T}_1}{\vec{T}_1 \bullet (\vec{T}_2 \times \vec{T}_3)} \qquad \vec{t}_3 = 2\pi \frac{\vec{T}_1 \times \vec{T}_2}{\vec{T}_1 \bullet (\vec{T}_2 \times \vec{T}_3)} \qquad (2.9)$$

From these equations, it can be shown that the dot product between a direct lattice vector and a reciprocal lattice vector must always be an integer multiple of 2π.

$$\vec{g} \bullet \vec{G} = m(2\pi) \qquad m = ..., -2, -1, 0, +1, +2, ... \qquad (2.10)$$

2.3.3.   Wigner-Seitz Cells and Brillouin Zones

A unit cell is a solid volume that completely fills space without any overlap when stacked. The most well known images of unit cells, like those depicted in Figure 2-1, are called "conventional" unit cells and are constructed to most clearly represent the symmetry of the lattice. A primitive unit cell is usually much smaller than the conventional unit cell and contains only a single lattice point that is not on its surface.

Choice of the unit cell is not unique and often many possibilities exist. The most common choice is the Wigner-Seitz cell [84]. This is defined as the volume of space around any lattice point which is closer to that lattice point than any other point in the entire lattice. The Brillouin zone is defined in the same manner except it is constructed in reciprocal space.
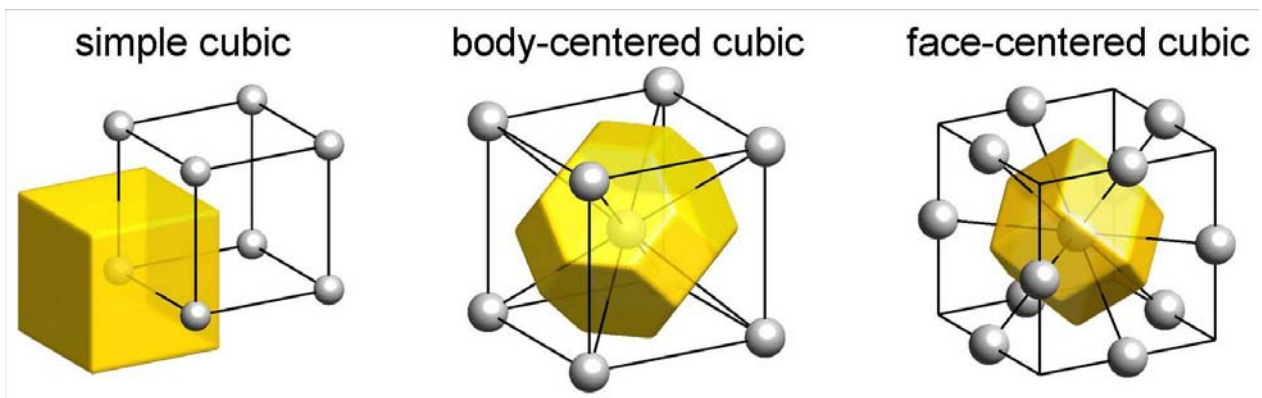
Figure 2-4: Wigner-Seitz cells of cubic crystals

Figure 2-4 shows the Wigner-Seitz cell for all three cubic crystal symmetries. For a simple cubic lattice, the Wigner-Seitz cell is a cube with 6 sides. For a body-centered crystal, it is a "truncated" octahedron with 14 sides. For a face-centered crystal, it is a rhombic dodecahedron and has 12 sides. The Wigner-Seitz cell of the body-centered cubic (BCC) crystal is the most spherical of all Bravais lattices. This means the Brillouin zone of the face-centered-cubic (FCC) crystal is the most spherical so FCC crystals are said have the highest symmetry of the Bravais lattices.

2.3.4.   *Wave Equation as an Eigen-Value Problem*

To study wave propagation in photonic crystals, it is most convenient to start with the wave equation. It can be written in the form of an eigen-value problem. For non-magnetic materials, this is

$$\nabla \times \frac{1}{\varepsilon_r} \nabla \times \vec{H} = k_0^2 \vec{H} \tag{2.11}$$

Eigen-value problems arise so often in physics and engineering that much is known about their properties and many efficient algorithms exist for calculating their solution. They have the following general form.

$$L\{\vec{H}\} = v\vec{H} \tag{2.12}$$

The linear mathematical operation represented by L is called the "eigen-operator," v is called the "eigen-value," and in this context $\vec{H}$ is the "eigen-state." This equation can be thought of as a way to test if $\vec{H}$ is an allowed state of the system. If performing the mathematical operation L on $\vec{H}$ just equals $\vec{H}$ scaled in magnitude by a constant v, then $\vec{H}$ must be an allowed state of the system. The scaling constant v is the eigen-value. It makes intuitive sense that many

possible solutions may exist. It is easy to see the wave equation in Eq. (2.11) is an eigen-value

problem where the eigen-value is k_0^2 and the eigen-operator is $\nabla \times \varepsilon_r^{-1} \nabla \times$.

To assess the similarity between two possible solutions $\vec{H}_1(\vec{r})$ and $\vec{H}_2(\vec{r})$, an "inner

product" may be defined as

$$\vec{H}_1 \otimes \vec{H}_2 = \int_\Omega \vec{H}_1^*(\vec{r}) \bullet \vec{H}_2(\vec{r}) d\vec{r} \tag{2.13}$$

In this equation $\vec{r}$ represents position and Ω represents the region of a single unit cell. If $\vec{H}_1$ and

$\vec{H}_2$ are very similar, the inner product will yield a large number. If they are very different, the

answer will be small. When the inner product exactly equals zero, $\vec{H}_1$ and $\vec{H}_2$ are said to be

"orthogonal."

The inner product may be used to test if a linear operator, L, is "Hermitian" if the

following property is satisfied.

$$L\{\vec{H}_1\} \otimes \vec{H}_2 = \vec{H}_1 \otimes L\{\vec{H}_2\} \tag{2.14}$$

It is easily shown that for lossless materials, the eigen-operator in Eq. (2.11) is Hermitian

[88]. This would not be the case if the wave equation were derived in terms of the electric field.

Hermitian eigen-value problems have three important properties. First, the eigen-values are real.

This means calculated frequencies will be real-valued and have physical meaning. Second, the

eigen-states are orthogonal. This means very different answers will be obtained from the

solution so no wasted calculations will be performed. Third, the eigen-states are "complete" so

the full set of eigen-states will completely describe the system. No information will be missing

from the solution. Once the eigen-states are calculated, all possible field patterns of propagating

waves in the photonic crystal must be a linear sum of the eigen-states. If it is not, the field must be evanescent.

Photonic crystals are periodic so the eigen-value problem will lead to discrete solutions. In addition, modes of the system will obey the Bloch-Floquet theorem [88].

$$\vec{H}(\vec{r}) = e^{j(\vec{k}\bullet\vec{r})}\vec{H}_{\vec{k}} \qquad (2.15)$$

The Bloch-Floquet theorem describes modes as having the general form of a plane wave, but with a periodic envelope $\vec{H}_{\vec{k}}$ that takes on the same symmetry as the photonic crystal. An important point to notice is that a Bloch wave is not a plane wave. A more accurate description is a bundle of coupled plane waves that propagate independently from other waves. While the component plane waves scatter within the lattice, the overall Bloch wave does not.

Substituting Eq. (2.15) into Eq. (2.11) leads to a wave equation in terms of the Bloch wave parameters.

$$\left(\nabla + j\vec{k}\right) \times \frac{1}{\varepsilon_r}\left(\nabla + j\vec{k}\right) \times \vec{H}_{i,\vec{k}} = \left(\frac{\omega_{i,\vec{k}}}{c_0}\right)^2 \vec{H}_{i,\vec{k}} \qquad (2.16)$$

The variational theorem states the lowest energy band satisfying Eq. (2.16) minimizes the following variational problem [89].

$$k^2_{0,i,\vec{k}} \Rightarrow \min \frac{\int\limits_\Omega \left|\left(\nabla + j\vec{k}\right) \times \vec{E}_{i,\vec{k}}\right|^2}{\int\limits_\Omega \varepsilon\left|\vec{E}_{i,\vec{k}}\right|^2} \qquad (2.17)$$

To minimize this function, the denominator must be maximized. This indicates the most intense electric fields will reside within the highest dielectric constant materials. In simpler terms, electric fields of the lowest order mode prefer to reside in the highest index regions.

2.4. Photonic Band Diagrams

Photonic band diagrams are a compact, but incomplete, means of characterizing the optical behavior of a photonic crystal. It is essentially a map of what temporal frequencies are allowed to propagate through the lattice with a given direction and spatial period. They provide a convenient summary of the transmission and dispersive properties of a photonic crystal. While they are only valid for crystals of infinite extent, they are much faster to calculate and behavior of finite crystals is usually very consistent.

The following text will outline at a high level how photonic band diagrams are constructed and how optical behavior may be interpreted from them. Details of the numerical solution to the wave equation will be discussed in later chapters.

2.4.1.   The Irreducible Brillouin Zone

In real space, everything is known about the solution if the field is known everywhere in the unit cell. To reduce the number of calculations, a primitive unit cell such as the Wigner-Seitz cell should be used. Similarly, in reciprocal space everything is known when the solution to the eigen-value problem is obtained at every point within the Brillouin zone.

The problem can be reduced even further through the concept of the "irreducible Brillouin zone" (IBZ). Propagation of light through a simple cubic lattice along any of the axes will behave the same as long as the crystal has 90° rotational symmetry. If this is the case, there is clearly redundancy within the Brillouin zone. A similar argument can be made for any crystal. The smallest region within the Brillouin zone that completely characterizes the photonic crystal is the IBZ. Figure 2-5 depicts the full Brillouin zone of a FCC crystal along with the irreducible Brillouin zone when spherical bases are assumed. The vertices of the IBZ are key points of

symmetry and are labeled with bold letters in this figure. $\vec{T_1}$, $\vec{T_2}$, and $\vec{T_3}$ are the reciprocal lattice vectors.
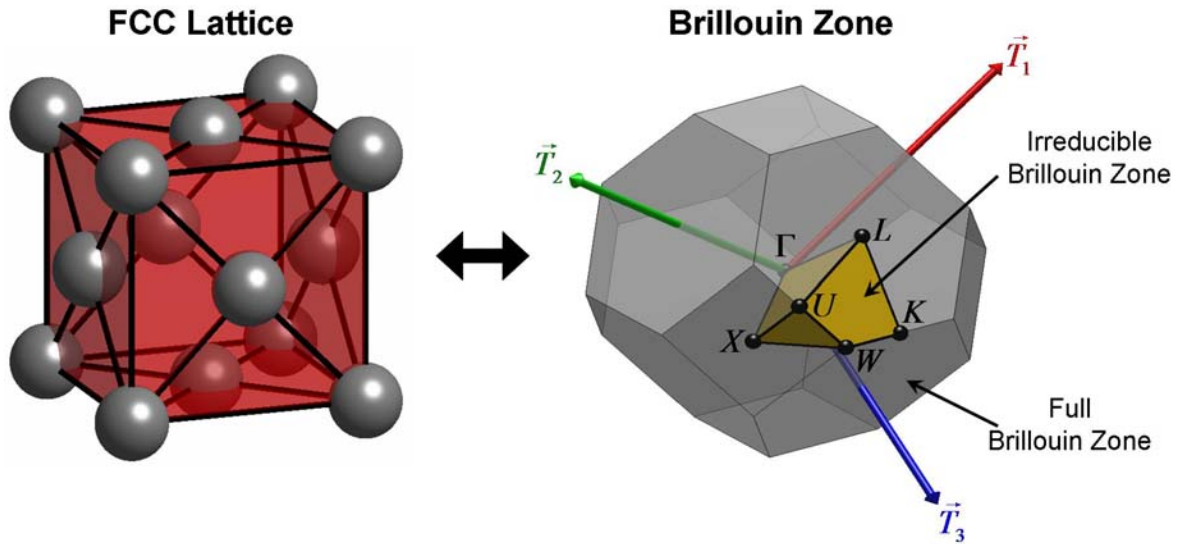
Figure 2-5: Brillouin zone, IBZ, and key points of symmetry for FCC crystals

2.4.2.   Construction of the Band Diagram

A numerical solution of the wave equation requires only the dielectric function ε_r describing materials in the unit cell and a Bloch wave vector $\vec{\beta}$ describing direction and spatial period of a Bloch wave. As illustrated in Figure 2-6, the solution produces a set of eigen-values and eigen-vectors. The eigen-values are interpreted as temporal frequencies that lead to allowed modes described by the Bloch wave vector. The eigen-vectors describe field profiles of the Bloch modes, but only the eigen-values are needed to construct a photonic band diagram. Details of the numerical solution to the wave equation are discussed in later chapters, but the

procedure simply calculates the set of discrete frequencies which lead to Bloch modes described by the given Bloch wave vector.

A photonic band diagram is constructed by lining up the eigen-values computed over a sequence of Bloch wave vectors. The sequence contains values of $\vec{\beta}$ that progress around the perimeter of the IBZ connecting all key points of symmetry. A solution at a sufficient number of points along the perimeter must be obtained to adequately resolve the photonic band structure.
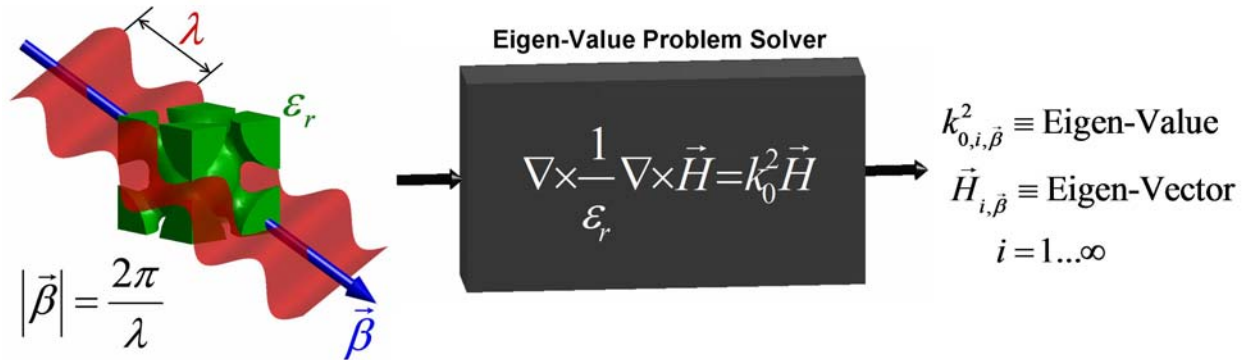
Figure 2-6: Numerical solution of the eigen-value wave equation

For 1D photonic crystals, the IBZ is one dimensional so the complete solution is a two-dimensional set of data. In this case, it is possible to completely represent the photonic band structure on paper. For 2D photonic crystals, the IBZ has two dimensions so the complete solution is a three-dimensional set of data. Bands are represented as sheets, but it is not entirely feasible to represent a three dimensional set of data on paper. The IBZ of 3D photonic crystals has three dimensions so the complete solution is a four-dimensional set of data. This is clearly not possible to completely represent on paper.
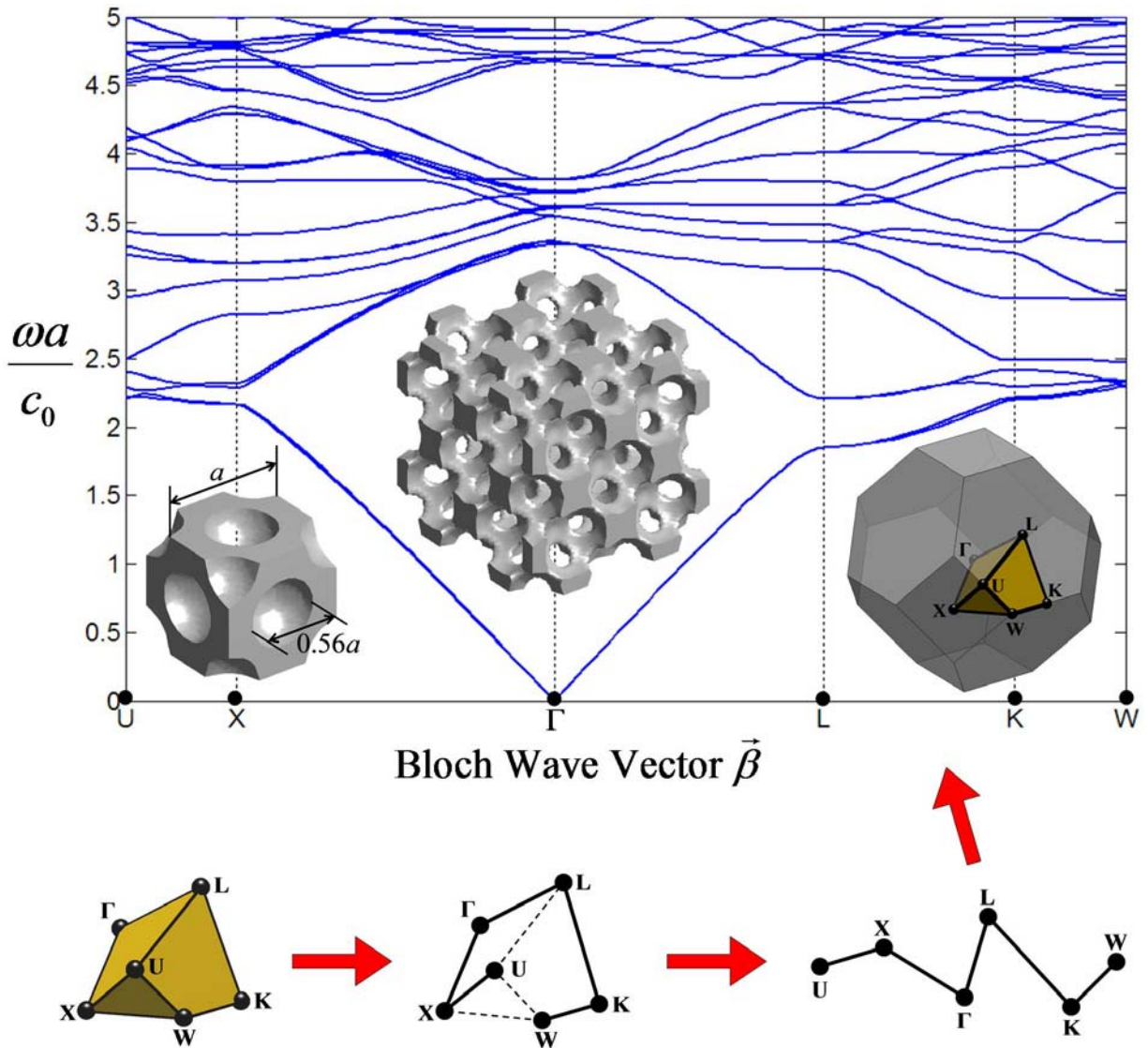
Figure 2-7: Photonic band diagram of inverted FCC photonic crystal

To describe the photonic band structure of 2D and 3D photonic crystals in a simple manner, a method for representing the data in a two-dimensional figure is needed. Recognizing that the band extremes are most likely to occur at the key points of symmetry in the IBZ, a good approach is to solve the eigen-value problem only for points along a line around the perimeter of the IBZ passing through all key points of symmetry. While the line may wonder through two

and three dimensions, the final set of data will only have two dimensions because solutions were obtained along a single line path. It should be mentioned that bypassing most of the IBZ in this manner leads to a small possibility that interesting behavior or band extremes may be missed.

An example is shown in Figure 2-7 where the photonic band structure was calculated using PWEM for a fully three dimensional FCC lattice comprised of air holes in a dielectric medium with n=3.6. The leftmost inset shows the unit cell of the photonic crystal from which the band structure was calculated. The center inset illustrates an extended lattice. The rightmost inset shows the Brillouin zone and IBZ discussed previously. Key points of symmetry are labeled **U**, **X**, **Γ**, **L**, **K**, and **W**. An array of around 200 Bloch wave vectors was generated progressing from **U** to **X** to **Γ** to **L** to **K** and finally to **W**. At each vector, eigen-frequencies were computed and plotted in the figure to form the continuous bands. This crystal does not posses a complete photonic band gap. The path taken around the perimeter of the IBZ and how it is unfolded to produce the horizontal axis is illustrated below the band diagram.

2.4.3.   *Reading Photonic Band Diagrams*

There are at least five optical properties that can be estimated from a photonic band diagram. These are band gaps, transmission spectrum, phase velocity (phase refractive index), group velocity (group refractive index), and dispersion. The following paragraphs outline how to estimate these properties from the photonic band diagram shown in Figure 2-8. In this diagram, bands of the photonic crystal are represented as solid blue lines. Dashed green lines correspond to the "light lines" of a homogeneous material set to the average dielectric constant of the photonic crystal. Transmission through a five layer slab is depicted on the right and shares the same frequency axis as the photonic band diagram for easy comparison.

A photonic band gap exists when no bands are present within some range of frequencies. If a photonic crystal is excited by a frequency in this range, the field will be evanescent and the lattice will appear as a mirror throughout. In Figure 2-8, the band gaps are highlighted in light gray while the gold highlights identify partial band gaps for light at normal incidence. A "complete" photonic band gap exists only if light at any angle and polarization is forbidden. Polarization of the Bloch modes cannot be determined from a band diagram unless it is identified explicitly.
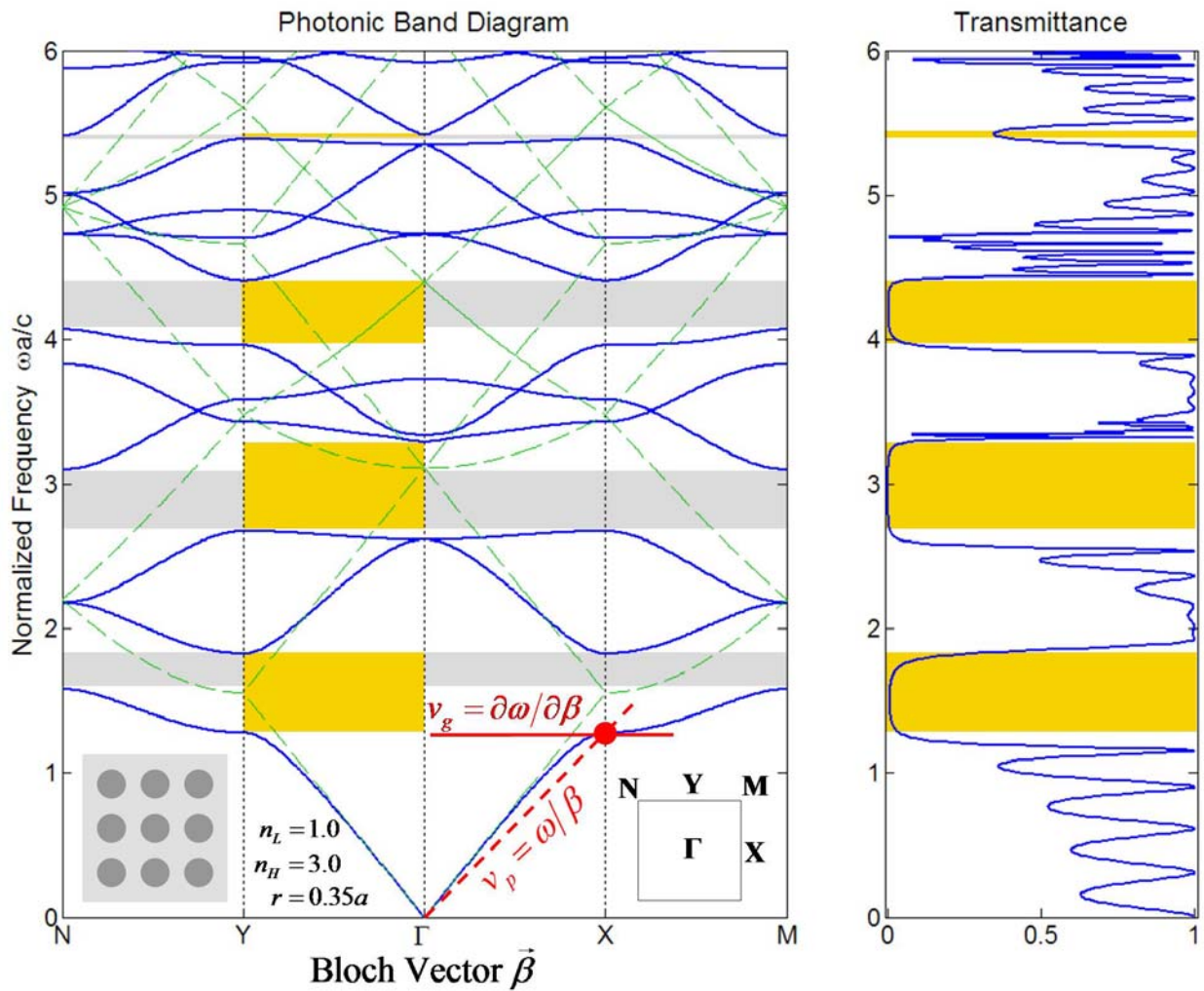
Figure 2-8: Interpreting photonic band diagrams

The transmission spectrum can be estimated through the concept of optical density of states (ODOS). The ODOS is the number of bands that exist for each temporal frequency. Transmission through the photonic crystal is greatly enhanced at frequencies where many states exist. Transmission is very low where there are few states or a band gap is present. While there are more factors leading to the real transmission spectrum, this provides a simple means for estimation. The transmission spectrum often shows many oscillations in regions where many photonic bands exist.

Phase velocity is the speed in which points of equal phase appear to propagate. This is defined as

$$\vec{v}_p = \omega \frac{\vec{\beta}}{\left|\vec{\beta}\right|^2} \tag{2.18}$$

From Eq. (2.18), phase velocity of a Bloch mode depends on position in the diagram. It can also be understood as the slope of the line connecting the point at $\omega = \Gamma = 0$ to the point of interest. It follows that the phase refractive index is

$$n_p = \frac{c}{\left|\vec{v}_p\right|} \tag{2.19}$$

Energy in a photonic crystal propagates at the group velocity which does not have to be in the same direction or the same magnitude as phase velocity. Group velocity is deduced from the slope of the band at the point of interest. For two and three dimensional crystals, the group velocity is calculated from the gradient of ω in k-space.

$$\vec{v}_g = \nabla_k \omega = \frac{\partial \omega}{\partial \beta_x} \hat{x} + \frac{\partial \omega}{\partial \beta_y} \hat{y} + \frac{\partial \omega}{\partial \beta_z} \hat{z} \tag{2.20}$$

The group refractive index is then

$$n_g = \frac{c}{|\vec{v}_g|} \tag{2.21}$$

The group velocity is related to the phase velocity through Rayleigh's equation [87].

$$\vec{v}_g = \vec{v}_p - \lambda_0 \frac{\partial \vec{v}_p}{\partial \lambda} \tag{2.22}$$

It follows that the group and phase refractive indices are related through

$$n_g \approx n_p - \lambda_0 \frac{\partial n_p}{\partial \lambda} \tag{2.23}$$

From Eqs. (2.22)-(2.23), it can be seen that group velocity deviates from phase velocity only when there is significant dispersion $\partial n_p / \partial \lambda$. Dispersion can be identified anywhere the band deviates from the light line.

2.5. Physical Origin of Electromagnetic Behavior in Photonic Crystals

2.5.1.   The Photonic Band Gap

In some photonic crystals, a range of wavelengths exists for which propagation is forbidden. This is called a "photonic band gap" and arises due to coherent scattering of light in the lattice. The physical origin of the band gap can be understood through arguments discussed previously. Three important concepts are:

1. Bloch modes must have the same symmetry as the photonic crystal.

2. Electric fields of the lowest order mode prefer to reside in higher index regions.

3. Modes corresponding to different bands must be orthogonal.

50

Consider the 1D photonic crystal and its band structure shown in Figure 2-9. The lattice is a stack of alternating layers of dielectric that has a photonic band gap. Consider the lowest energy band at $k=\pi/a$, labeled with a blue point. The irradiance of the mode is depicted throughout the lattice. The mode follows rule 1 by taking on the symmetry of lattice. To be consistent with rule 2, the most intense portions of the field are placed in the regions with n_2. Since there is no lower order mode, rule 3 does not yet apply.

Now consider the next higher energy band at $k=\pi/a$ labeled with a red point. This mode also follows rule 1 by taking on the same symmetry as the lattice. Rule 2 does not apply, but the mode is forced to be orthogonal to the lower order mode due to rule 3. For this reason, the intense portions of the field are shifted to the lower index regions corresponding to n_1.
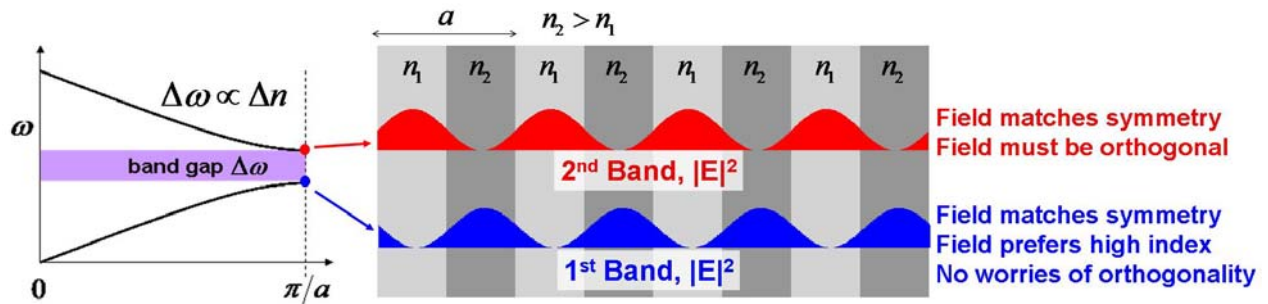
Figure 2-9: Physical origin of the photonic band gap

The lower order mode must have a higher effective refractive index because it resides mainly in the higher index regions. The higher order mode must have a lower effective refractive index because it resides mainly in the lower index regions. Both modes, however, have the same spatial frequency of π/a. The only way to reconcile the two modes having different effective index with the same spatial frequency is if they are distinct in temporal

frequency. Therefore, a distinct gap between the bands must exist if there is any index contrast. Based on this argument, it makes intuitive sense that higher contrast leads to greater disparity between the modes and a larger photonic band gap.

2.5.2.   *Realizing a Complete Photonic Band Gap*

From discussion above, it is possible to identify what lattices support photonic band gaps for transverse electric (TE) modes, transverse magnetic (TM) modes, or both. For 2D photonic crystals where wave propagation is restricted to the crystal plane, TE modes have the electric field polarized normal to the crystal plane. These modes can form isolated regions of high intensity surrounded by connected regions of low intensity. To maximize the disparity discussed above, lattices favoring TE band gaps should be comprised of isolated regions of high dielectric constant material. One such structure is an array of dielectric cylinders in air. The photonic band structure and two Bloch modes are shown in Figure 2-10. At the highest point in the lowest band marked **A**, the field is contained almost entirely within the dielectric rods. For this reason, the lower band is sometimes referred to as the 'dielectric band.' At the upper band marked **B**, the field resides almost entirely outside of the dielectric rods. This band is sometimes referred to as the 'air-band.'
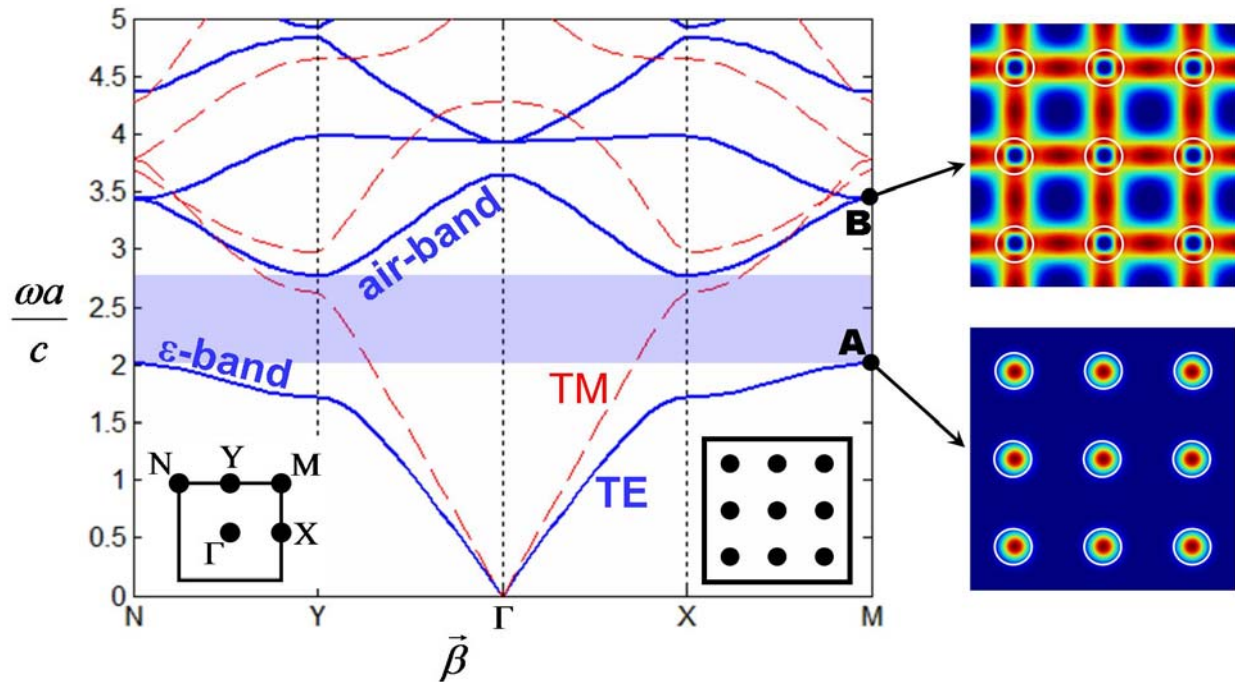
Figure 2-10: Photonic crystal favoring TE band gaps

TM modes have the electric field polarized parallel to the crystal plane. In this case the electric field is a vector quantity and must form circular loops to satisfy Maxwell's equations which dictate electric field lines should be continuous. To maximize the disparity discussed above, lattices favoring TM band gaps should be comprised of connected veins of high dielectric constant material. One possible structure is a series of interconnected dielectric rods. The photonic band structure and two Bloch modes for this lattice are shown in Figure 2-11. At the highest point in the lowest energy band marked **A**, the field forms loops and is contained almost entirely in the dielectric rods. At the upper band marked **B**, the field still forms loops, but resides almost entirely outside of the rods.
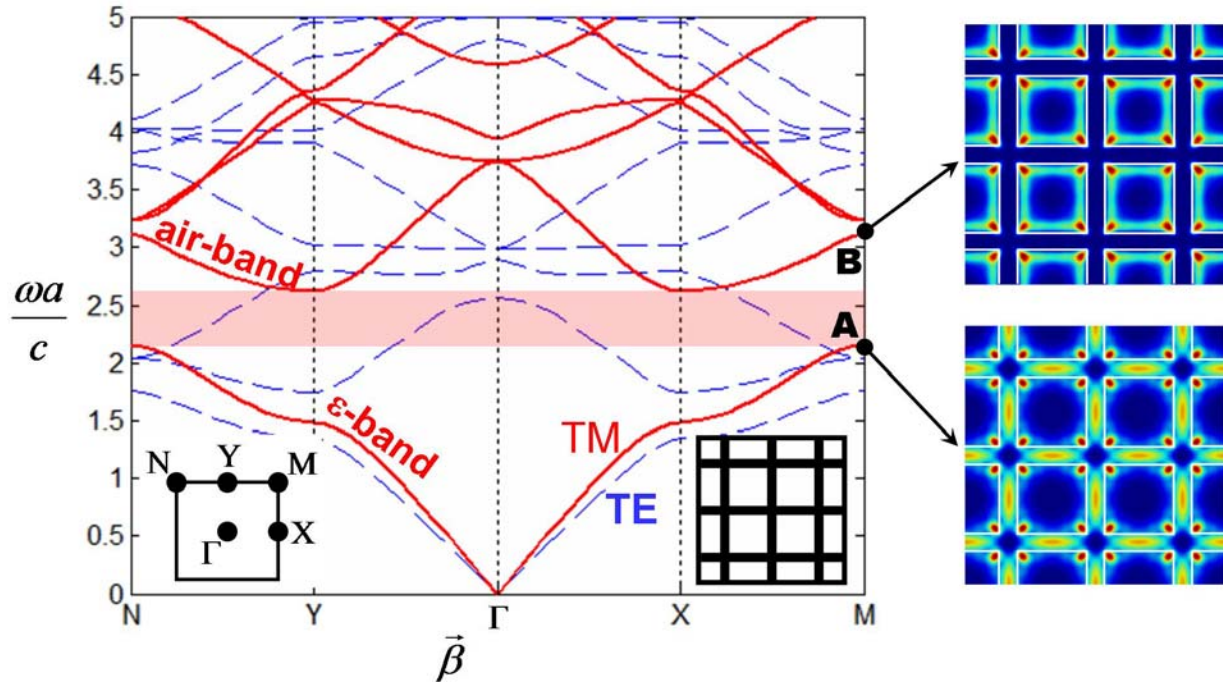
Figure 2-11: Photonic crystal favoring TM band gaps

Another aspect of lattice symmetry is dielectric inversion. This entails swapping high and low dielectric materials in the lattice. Figure 2-12 shows photonic band structures for a square array of dielectric cylinders in air and a square array of air cylinders in dielectric. The photonic band structures are very different, but one key difference can be observed. The array of dielectric cylinders supports TE band gaps, but not TM band gaps. The inverted array supports TM band gaps, but not TE band gaps. This is a logical conclusion since the first lattice is comprised of isolated regions of dielectric and the second is a connected lattice.
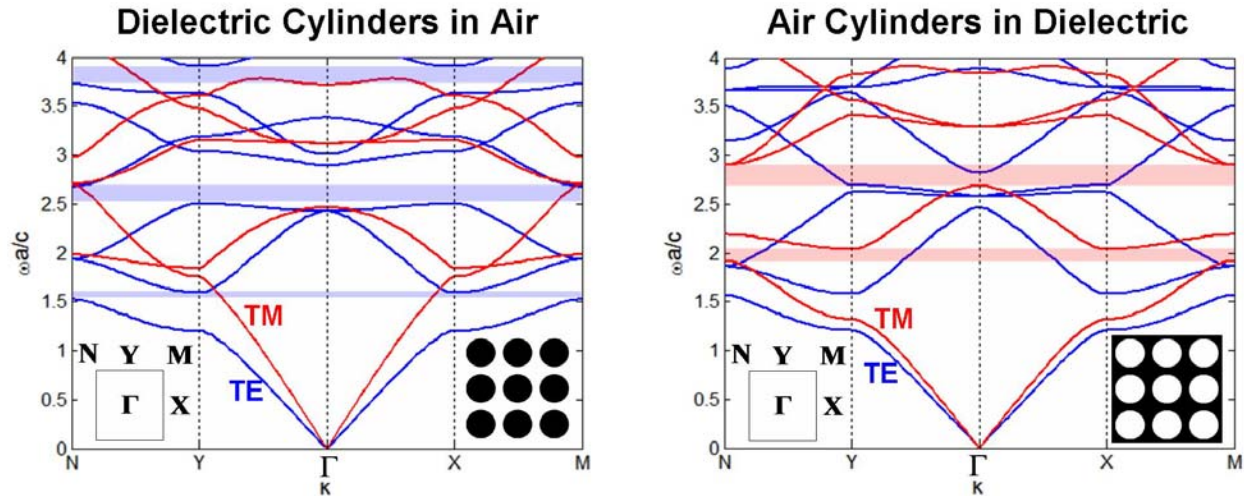
Figure 2-12: Dielectric inversion

The role of lattice symmetry can be understood through Figure 2-13 where the photonic band structure is compared between four cases. The square array at the upper left provides a number of small band gaps. A rectangular array is formed by increasing the lattice constant along one axis. This offsets the gaps seen by light traveling on each axis so they are no longer aligned. For this reason, no complete photonic band gap is observed because the lattice has lost some of its symmetry. When the shape of the dielectric rod is distorted but square symmetry is maintained, a similar effect happens and only a very small band gap is observed. When perfectly circular rods are placed in a hexagonal array, much larger band gaps appear due to the higher degree of symmetry. It can be concluded that band gaps are favored by lattices with high symmetry. That is, lattices with near circular/spherical Brillouin zones.

55
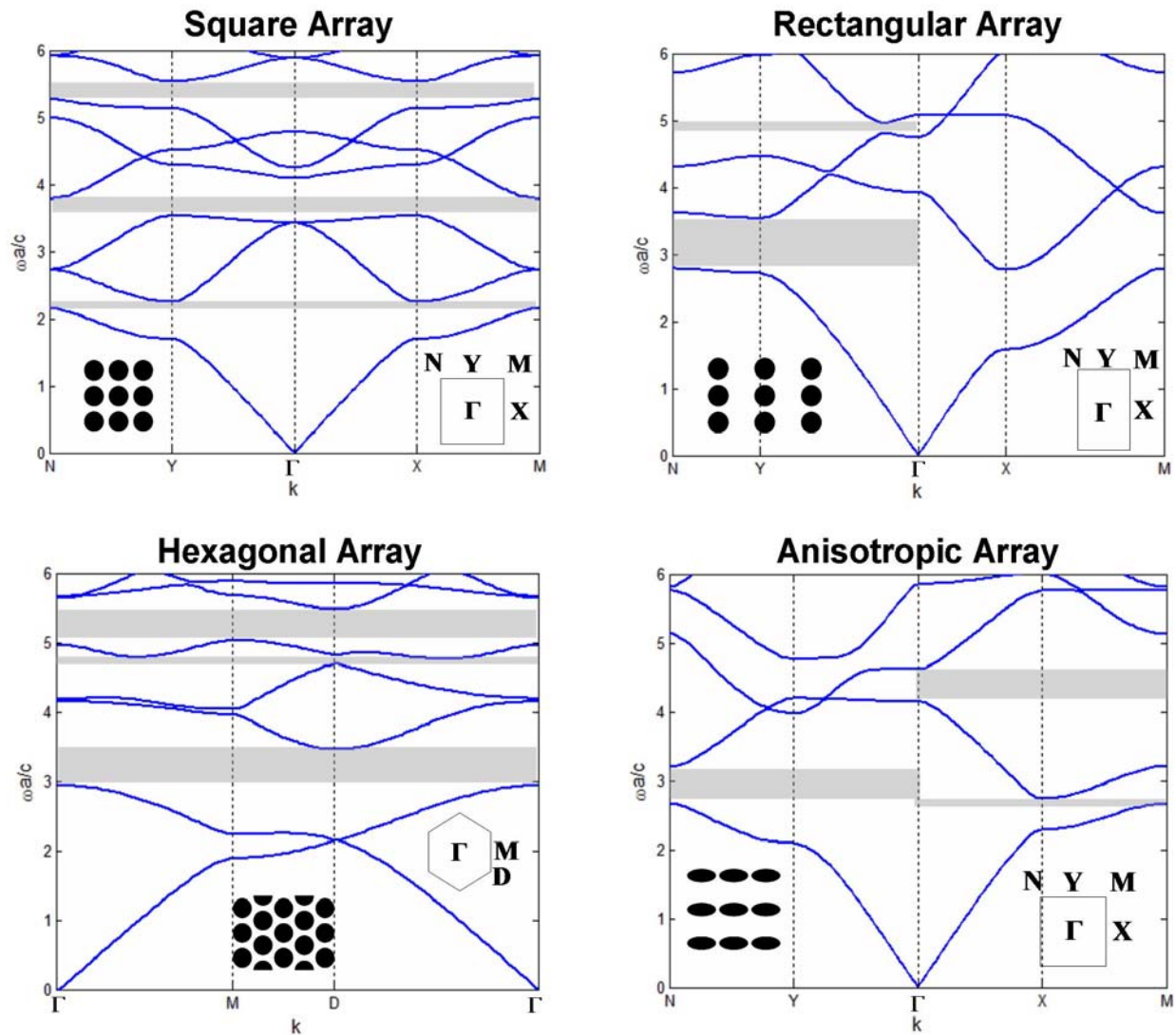
Figure 2-13: Role of lattice symmetry

To realize a complete photonic band gap, high symmetry lattices with isolated "globs" of dielectric connected by thin veins are needed. In two dimensions, a common solution is the hexagonal lattice depicted in Figure 2-14 along with its photonic band structure. It incorporates isolated "globs" of dielectric to promote TE band gaps and connects them with thin dielectric "veins" to support TM band gaps. It is a hexagonal array for high symmetry and has high index contrast to realize a wide band gap. As can be seen from the photonic band structure, this lattice

proves a very large complete photonic band gap. Based on this reasoning, it is not surprising that the inverse type photonic crystals with FCC or diamond symmetry have the widest known 3D photonic band gaps.
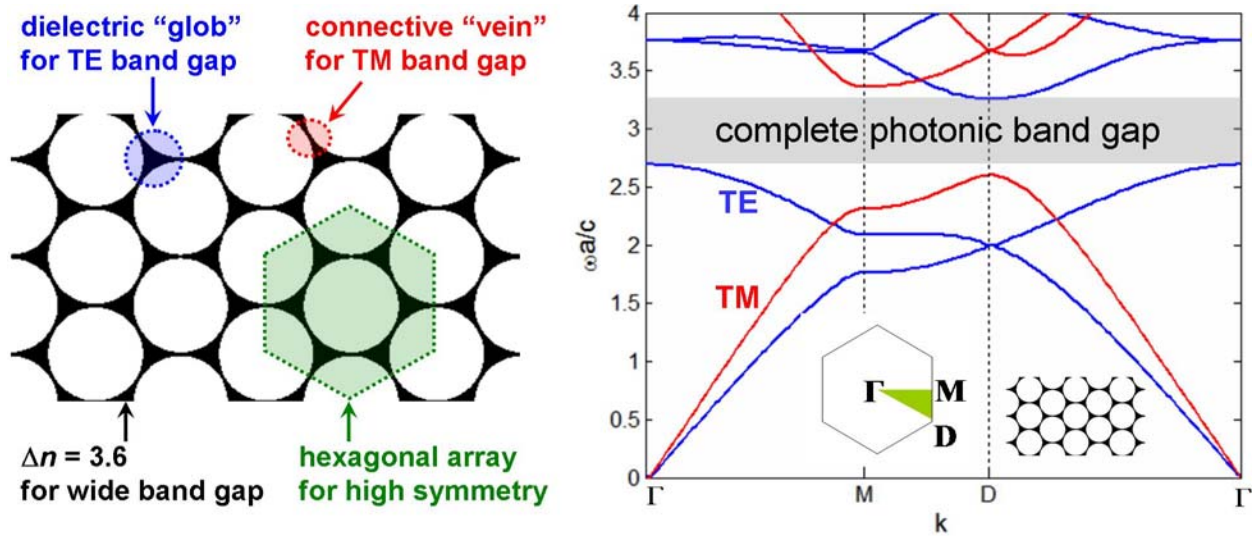
Figure 2-14: Photonic crystal with a complete band gap

2.5.3. *Dispersion*

The strong dispersive properties of a photonic crystal can be reasoned through the illustrations in Figure 2-15. This compares the optical response of an atomic crystal to the optical response of a photonic crystal. The mechanism producing a macroscopic refractive index in an atomic lattice is fundamentally the same mechanism in a photonic crystal producing an effective refractive index.
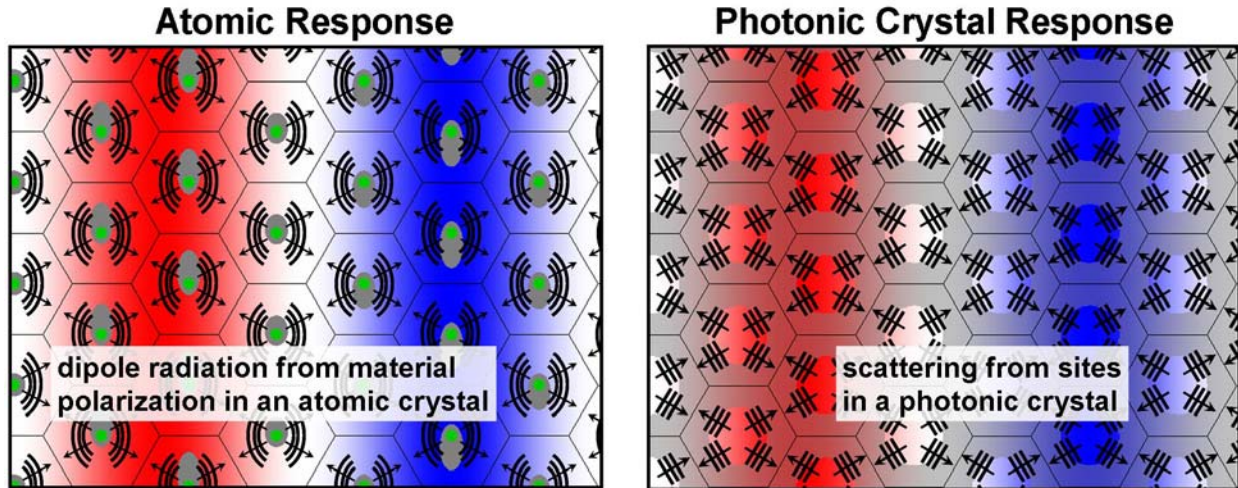
Figure 2-15: Atomic crystal and photonic crystal response to an optical wave

The left diagram in Figure 2-15 depicts an atomic crystal in the presence of an applied optical wave. Charges are displaced proportional to the applied electric field causing the material to become "polarized." Because the field is oscillating, material polarization oscillates at the same frequency. As charges are accelerated during oscillation, they radiate energy at the same frequency as the applied wave. This takes the form of dipole radiation. These combine out of phase with the applied wave producing an apparent change in velocity. An effective dielectric function is used to describe the phenomenon at a macroscopic level. Materials are dispersive because charges are bound to the lattice through a restoring force producing a resonance. This alters the response of the material as a function of wavelength.

The right diagram in Figure 2-15 shows a photonic crystal in the presence of an applied optical wave. Waves scattered from the lattice combine out of phase with the applied wave in a manner analogous to the dipole radiation in an atomic lattice. The overall affect is an apparent change in velocity described in terms of an effective dielectric constant. Photonic crystals are dispersive because the periodic lattice has a resonance causing the response to change with

wavelength. It should be expected that the response of a photonic crystal can be much stronger than standard dielectric materials could provide because index contrast can be very high. This leads to stronger scattered fields in a photonic crystal than there would be dipole radiation in an atomic crystal.

CHAPTER 3
FINITE-DIFFERENCE FREQUENCY-DOMAIN METHOD

3.1. Overview

The finite-difference frequency-domain (FDFD) method [45, 46, 90-95] may be the easiest of all methods to formulate and implement. It is a completely numerical method excellent for field visualization and modeling devices with complex geometry or structures of finite size. It is accurate, stable, and sources of error are well understood. It is fully vectorial and obtains a rigorous solution to Maxwell's equations. Being a frequency-domain method, it is able to resolve sharp resonances and obtain solutions at a single frequency more efficiently than time-domain methods. Being a frequency-domain method, it is able to account for oblique angles of incidence and material dispersion quite easily.

The method transforms Maxwell's equations into a set of linear algebraic equations using the finite-difference method [96]. A source is incorporated and the problem solved through matrix division. The method is capable of modeling 2D devices that are hundreds of wavelengths square. Non-uniform [90] and unstructured grids [93] can also be used to improve accuracy and efficiency.

Despite its power and simplicity, literature on FDFD remains surprisingly sparse. Fortunately, much of the literature on FDTD is directly applicable to FDFD. This chapter will introduce finite-differences and outline a new and much improved formulation of the method in the framework of matrix operators on a Yee grid [97]. This approach is more compact, less tedious to formulate, and requires only first order derivatives. It is stable and accurate for both TE and TM modes. A simple approach to incorporate a source will be introduced that can be

extended to arbitrary sources such as cylindrical waves or dipole sources and inherently accommodates implementation on unstructured grids or using higher-order accurate finite-difference equations. Compensating for numerical dispersion will also be discussed.

3.2. Finite-Difference Approximations

In the finite-difference method (FDM), continuous functions are fit to a grid where function values are known only at discrete points. Intermediate values must be interpolated if they are needed. The method approximates derivatives using weighted sums of function values at neighboring points. For example, suppose function values $f_0, f_1,$ and f_2 are known at three consecutive points on a grid with spacing Δ. The first derivative at point 1 could be approximated by calculating the slope between points 0 and 2.

$$\frac{\partial f_1}{\partial x} \cong \frac{f_2 - f_0}{2\Delta} \tag{3.1}$$

Similarly, the derivative midway between two consecutive points on a grid can be estimated as

$$\frac{\partial f_{1.5}}{\partial x} \cong \frac{f_2 - f_1}{\Delta} \tag{3.2}$$

These equations are second-order accurate finite-difference approximations of first-order derivatives. Finite-difference equations can also be written on non-uniform grids or even unstructured grids. Equation (3.2) is the only finite-difference approximation used in this dissertation and is second-order accurate on a uniform grid. For more details on finite-difference approximations see Ref. [96].

3.3. Representing Optical Structures on Discrete Grids

In the FDM framework, fields and materials are defined only at discrete points. Grid resolution, defined as the spacing between grid points, must be sufficiently small to accurately

61

portray rapidly varying fields and resolve structures with fine dimensions. When grid resolution is made too small, matrices may become too large to solve by conventional means. When grid resolution is made too large, the method is less accurate and can potentially be unstable.

In practice, it is best to use as coarse of a grid as possible while achieving an acceptable level of accuracy. This tends to be on the order of $\lambda_{min}/10$ to $\lambda_{min}/40$, where λ_{min} is the wavelength inside the highest refractive index found in the problem.

3.3.1.   *Representing Fields*

Despite there being four Maxwell's equations, most numerical algorithms are formulated only from the curl equations. The divergence equations in charge free media are a direct consequence of the curl equations. It is necessary in a grid-based system to ensure the divergence conditions are implicitly enforced by correct placement and orientation of field components on the grid. Grids where all fields are collocated at the same point, do not necessarily satisfy the divergence equations. The problem becomes apparent when modeling TM modes or fully three-dimensional systems.

The most popular grid scheme is that described by Kane Yee in 1966 [97] where electric and magnetic fields are interleaved on a Cartesian grid as depicted in Figure 3-1. Each electric field is immediately surrounded by four circulating magnetic fields and each magnetic field is immediately surrounded by four circulating electric fields. This clever arrangement makes numerical representation of Maxwell's curl equations highly efficient and inherently at least second-order accurate on uniform grids [44]. The Yee grid ensures divergence-free behavior [44] so only Maxwell's curl equations need to be considered. Boundary conditions are naturally satisfied if the interface between two materials is parallel to one of the lattice coordinates.
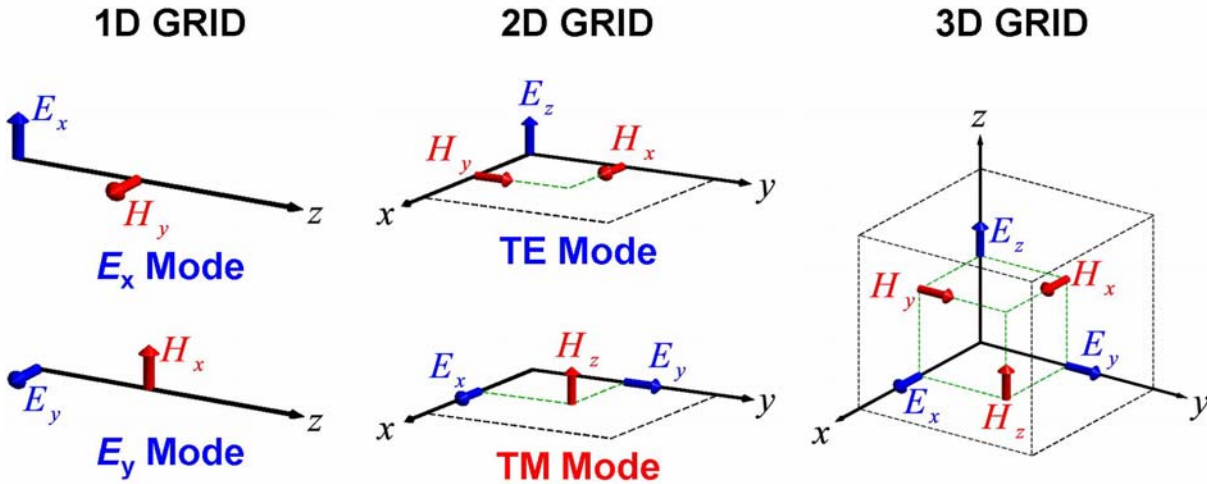
Figure 3-1: Unit cell of Yee grid for different dimensional systems

3.3.2. *Representing Materials*

Each field component is placed at a unique physical point due to the staggered nature of the Yee grid. As a consequence, field components have different phase and may reside within a different material. It is common to assign unique material properties to each field component. Materials represented in this manner are inherently anisotropic. In addition, there exists potential for subcell resolution of very small structures. Several methods exist for assigning material properties and the most common are summarized in Figure 3-2.

A simple and powerful method of assigning materials properties to points on the grid is to construct the problem at twice the desired resolution and overlay this onto the Yee grid. This approach has the advantage of being straight forward to implement and was found to be more accurate then Dey-Mittra averaging [98] when fine structures had to be resolved [1]. This approach is labeled "discrete" in Figure 3-2.
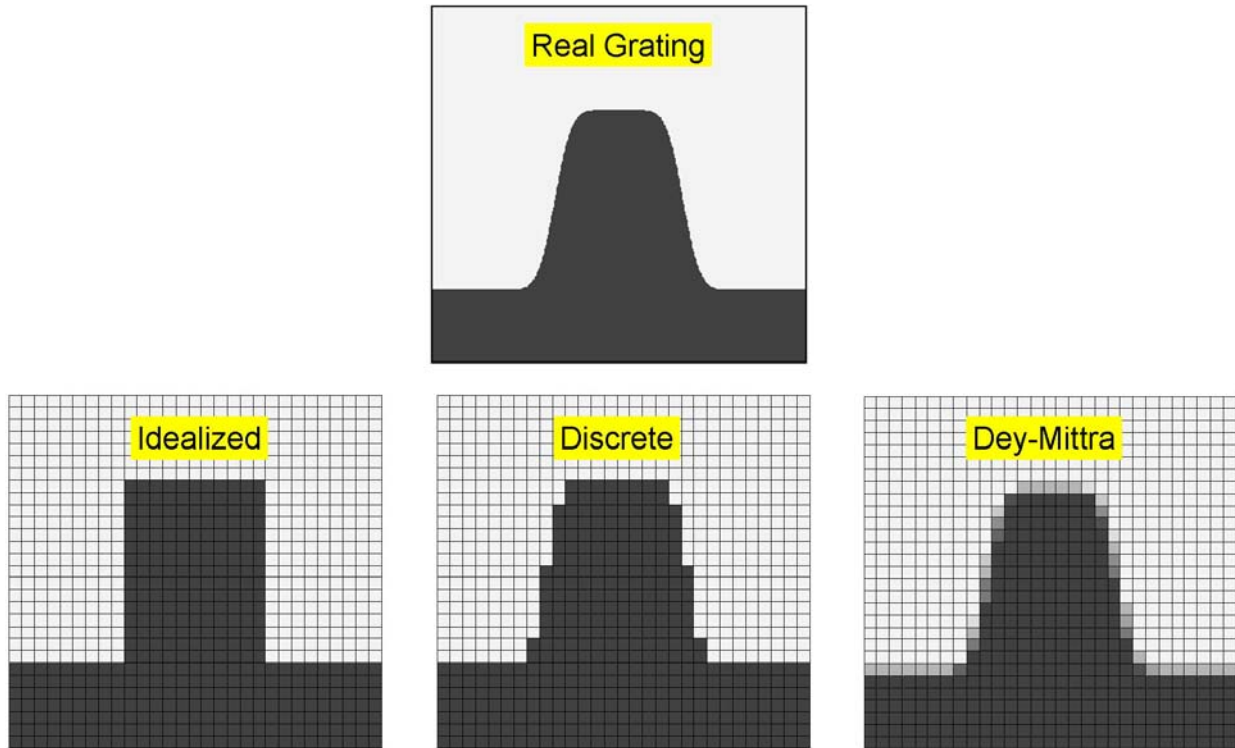
Figure 3-2: Material representation

Cartesian grids do not conform well to curved surfaces. While the "2X" grid method works well for fine features, larger structures are usually more accurately represented using Dey-Mittra averaging. This method simply assigns materials properties to points on the Yee grid based on a local average. To implement this method, it is easiest to construct the problem on a very high resolution grid. The average of local groups of points on the high resolution grid, are assigned to single points on the Yee grid. A typical result is labeled "Dey-Mittra" in Figure 3-2.

3.4. Matrix Representation of Maxwell's Equations

3.4.1.   Maxwell's Equations in Normalized Form

Maxwell's curl equations for source-free media in the frequency-domain are

$$\nabla \times \vec{E} = -j\omega\mu\vec{H} \tag{3.3}$$

$$\nabla \times \vec{H} = j\omega\varepsilon\vec{E} \tag{3.4}$$

The magnitude of the electric and magnetic fields differ by around three orders of magnitude. To avoid numerical rounding errors, either the electric or magnetic field should be normalized. It is convenient to normalize the magnetic field in the following manner.

$$\vec{\tilde{H}} = -j\sqrt{\frac{\varepsilon_0}{\mu_0}}\vec{H} \tag{3.5}$$

Maxwell's curl equations can be written in terms of the normalized magnetic field where $k_0 = \omega/c_0$, ε_r is relative permittivity, and μ_r is relative permeability.

$$\nabla \times \vec{E} = k_0\mu_r\vec{\tilde{H}} \tag{3.6}$$

$$\nabla \times \vec{\tilde{H}} = k_0\varepsilon_r\vec{E} \tag{3.7}$$

3.4.2.   Incorporation of Perfectly Matched Layer Absorbing Boundary

Conditions

To prevent nonphysical reflections from the numerical boundaries of the grid, the perfectly matched layer (PML) boundary condition may be used [92, 99]. It is introduced into Maxwell's curl equations through the tensor $\overline{\overline{s}}$.

$$\nabla \times \vec{E} = k_0\mu_r\overline{\overline{s}}\vec{\tilde{H}} \tag{3.8}$$

$$\nabla \times \vec{\tilde{H}} = k_0\varepsilon_r\overline{\overline{s}}\vec{E} \tag{3.9}$$

The PML tensor can be written as

$$\bar{\bar{s}} = \begin{bmatrix} \dfrac{s_y s_z}{s_x} & 0 & 0 \\[2ex] 0 & \dfrac{s_x s_z}{s_y} & 0 \\[2ex] 0 & 0 & \dfrac{s_x s_y}{s_z} \end{bmatrix} \qquad \begin{aligned} s_x(x) &= s_{x0}(x)\left[1+\dfrac{\eta_0}{jk_0}\sigma_x^*(x)\right] \\[1ex] s_y(y) &= s_{y0}(x)\left[1+\dfrac{\eta_0}{jk_0}\sigma_y^*(y)\right] \\[1ex] s_z(z) &= s_{z0}(x)\left[1+\dfrac{\eta_0}{jk_0}\sigma_z^*(z)\right] \end{aligned} \tag{3.10}$$

where the PML gradient is expressed as

$$s_{z0}(z) = 1 + s_{max}\left(\frac{z}{L}\right)^p \qquad\qquad \sigma_z^*(z) = \sigma_{max}^* \sin^2\left(\frac{\pi z}{2L}\right) \tag{3.11}$$

Typical values are $0 \le s_{max} \le 5$ and $\sigma_{max}^* \approx 1$. This is similar to a conventional PML with the

addition of the s_0 parameter to better attenuate evanescent fields. Using normalized coordinates

($x' = k_0 x$, $y' = k_0 y$, and $z' = k_0 z$), Eqs. (3.8)-(3.9) expand to

$$\frac{\partial E_z}{\partial y'} - \frac{\partial E_y}{\partial z'} = \mu_r \frac{s_y s_z}{s_x} \tilde{H}_x \tag{3.12}$$

$$\frac{\partial E_x}{\partial z'} - \frac{\partial E_z}{\partial x'} = \mu_r \frac{s_x s_z}{s_y} \tilde{H}_y \tag{3.13}$$

$$\frac{\partial E_y}{\partial x'} - \frac{\partial E_x}{\partial y'} = \mu_r \frac{s_x s_y}{s_z} \tilde{H}_z \tag{3.14}$$

$$\frac{\partial \tilde{H}_z}{\partial y'} - \frac{\partial \tilde{H}_y}{\partial z'} = \varepsilon_r \frac{s_y s_z}{s_x} E_x \tag{3.15}$$

$$\frac{\partial \tilde{H}_x}{\partial z'} - \frac{\partial \tilde{H}_z}{\partial x'} = \varepsilon_r \frac{s_x s_z}{s_y} E_y \tag{3.16}$$

$$\frac{\partial \tilde{H}_y}{\partial x'} - \frac{\partial \tilde{H}_x}{\partial y'} = \varepsilon_r \frac{s_x s_y}{s_z} E_z \tag{3.17}$$

3.4.3. Approximating Maxwell's Curl Equations with Finite-Differences

Applying the finite-difference approximation of Eq. (3.2) to Eqs. (3.12)-(3.17) on a Yee grid, leads to a set of six algebraic equations.

$$\frac{E_z^{i,j+1,k} - E_z^{i,j,k}}{\Delta y'} - \frac{E_y^{i,j,k+1} - E_y^{i,j,k}}{\Delta z'} = \mu_x^{i,j,k} \tilde{H}_x^{i,j,k} \tag{3.18}$$

$$\frac{E_x^{i,j,k+1} - E_x^{i,j,k}}{\Delta z'} - \frac{E_z^{i+1,j,k} - E_z^{i,j,k}}{\Delta x'} = \mu_y^{i,j,k} \tilde{H}_y^{i,j,k} \tag{3.19}$$

$$\frac{E_y^{i+1,j,k} - E_y^{i,j,k}}{\Delta x'} - \frac{E_x^{i,j+1,k} - E_x^{i,j,k}}{\Delta y'} = \mu_z^{i,j,k} \tilde{H}_z^{i,j,k} \tag{3.20}$$

$$\frac{\tilde{H}_z^{i,j,k} - \tilde{H}_z^{i,j-1,k}}{\Delta y'} - \frac{\tilde{H}_y^{i,j,k} - \tilde{H}_y^{i,j,k-1}}{\Delta z'} = \varepsilon_x^{i,j,k} E_x^{i,j,k} \tag{3.21}$$

$$\frac{\tilde{H}_x^{i,j,k} - \tilde{H}_x^{i,j,k-1}}{\Delta z'} - \frac{\tilde{H}_z^{i,j,k} - \tilde{H}_z^{i-1,j,k}}{\Delta x'} = \varepsilon_y^{i,j,k} E_y^{i,j,k} \tag{3.22}$$

$$\frac{\tilde{H}_y^{i,j,k} - \tilde{H}_y^{i-1,j,k}}{\Delta x'} - \frac{\tilde{H}_x^{i,j,k} - \tilde{H}_x^{i,j-1,k}}{\Delta y'} = \varepsilon_z^{i,j,k} E_z^{i,j,k} \tag{3.23}$$

The PML parameters have been combined with the permittivity and permeability according to

$$\varepsilon_x = \varepsilon_r \frac{s_y s_z}{s_x} \qquad \varepsilon_y = \varepsilon_r \frac{s_x s_z}{s_y} \qquad \varepsilon_z = \varepsilon_r \frac{s_x s_y}{s_z} \tag{3.24}$$

$$\mu_x = \mu_r \frac{s_y s_z}{s_x} \qquad \mu_y = \mu_r \frac{s_x s_z}{s_y} \qquad \mu_z = \mu_r \frac{s_x s_y}{s_z} \tag{3.25}$$

3.4.4. Numerical Dispersion

As a consequence of representing continuous fields on a discrete grid, nonphysical dispersion is introduced [44]. Phase velocity of simulated waves differs slightly from what would be observed in physical systems. Numerical dispersion is a function of wavelength, direction of propagation, and grid parameters. Simulated waves accumulate extra phase as they

propagate through the numerical grid that leads to nonphysical results such as shifted spectral response, pulse broadening, ringing of pulsed waveforms, imprecise cancellation of scattered waves, anisotropy, and pseudo-refraction. The problem is particularly prevalent when modeling large structures, using periodic boundary conditions, or using a very coarse grid.

The numerical dispersion relation is derived by substituting a plane wave solution into Eqs. (3.18)-(3.25) and determining the condition placed on the wave vector. The result for a homogeneous and isotropic material is

$$\left(\frac{\omega}{v}\right)^2 \varepsilon_r = \left[\frac{2}{\Delta x}\sin\left(\frac{k_x\Delta x}{2}\right)\right]^2 + \left[\frac{2}{\Delta y}\sin\left(\frac{k_y\Delta y}{2}\right)\right]^2 + \left[\frac{2}{\Delta z}\sin\left(\frac{k_z\Delta z}{2}\right)\right]^2 \tag{3.26}$$

Here, velocity is represented by the parameter v indicating it is different than the free space speed of light. It can be seen from this equation that dispersion is purely an artifact of grid resolution. At the resolution parameters approach zero, the dispersion equation reduces to that of a real material.

Using Eq. (3.26), it is possible by simple means to exactly compensate for numerical dispersion in one direction. The correction is implemented by multiplying the dielectric function by a factor f. If f is chosen correctly, the parameter v becomes the true vacuum speed of light.

$$\left(\frac{\omega}{c}\right)^2 (f\varepsilon_r) = \left[\frac{2}{\Delta x}\sin\left(\frac{k_x\Delta x}{2}\right)\right]^2 + \left[\frac{2}{\Delta y}\sin\left(\frac{k_y\Delta y}{2}\right)\right]^2 + \left[\frac{2}{\Delta z}\sin\left(\frac{k_z\Delta z}{2}\right)\right]^2 \tag{3.27}$$

Solving this for the correction factor f leads to

$$f = \frac{1}{k_0^2\varepsilon_r}\left\{\left[\frac{2}{\Delta x}\sin\left(\frac{k_x\Delta x}{2}\right)\right]^2 + \left[\frac{2}{\Delta y}\sin\left(\frac{k_y\Delta y}{2}\right)\right]^2 + \left[\frac{2}{\Delta z}\sin\left(\frac{k_z\Delta z}{2}\right)\right]^2\right\} \tag{3.28}$$

To compensate for numerical dispersion, the grid parameters Δx, Δy, and Δz are calculated first. Given a wave vector and average dielectric constant throughout the problem

68

space, the correction factor is calculated using Eq. (3.28). Immediately after the materials are assigned to points on the grid, they are all multiplied by the factor f. From there, matrices are calculated and solved without numerical dispersion. When there are waves in multiple directions, exact compensation is no longer possible by this method, but an average correction factor will still improve model accuracy.

3.4.5. *Matrix Form of Maxwell's Curl Equations*

When all six equations are written for every cell in the Yee grid, a large set of algebraic equations is generated. These can be written as separate matrix equations.

$$\mathbf{D}_{y'}^{E}\mathbf{E}_z - \mathbf{D}_{z'}^{E}\mathbf{E}_y = \boldsymbol{\mu}_x\tilde{\mathbf{H}}_x \tag{3.29}$$

$$\mathbf{D}_{z'}^{E}\mathbf{E}_x - \mathbf{D}_{x'}^{E}\mathbf{E}_z = \boldsymbol{\mu}_y\tilde{\mathbf{H}}_y \tag{3.30}$$

$$\mathbf{D}_{x'}^{E}\mathbf{E}_y - \mathbf{D}_{y'}^{E}\mathbf{E}_x = \boldsymbol{\mu}_z\tilde{\mathbf{H}}_z \tag{3.31}$$

$$\mathbf{D}_{y'}^{H}\tilde{\mathbf{H}}_z - \mathbf{D}_{z'}^{H}\tilde{\mathbf{H}}_y = \boldsymbol{\varepsilon}_x\mathbf{E}_x \tag{3.32}$$

$$\mathbf{D}_{z'}^{H}\tilde{\mathbf{H}}_x - \mathbf{D}_{x'}^{H}\tilde{\mathbf{H}}_z = \boldsymbol{\varepsilon}_y\mathbf{E}_y \tag{3.33}$$

$$\mathbf{D}_{x'}^{H}\tilde{\mathbf{H}}_y - \mathbf{D}_{y'}^{H}\tilde{\mathbf{H}}_x = \boldsymbol{\varepsilon}_z\mathbf{E}_z \tag{3.34}$$

$\mathbf{E}_x$, $\mathbf{E}_y$, and $\mathbf{E}_z$ are column vectors containing a sequential list of all electric field values in the grid. Likewise, $\tilde{\mathbf{H}}_x$, $\tilde{\mathbf{H}}_x$, and $\tilde{\mathbf{H}}_x$ are column vectors containing all normalized magnetic field values in the grid. $\boldsymbol{\varepsilon}_x$, $\boldsymbol{\varepsilon}_y$, and $\boldsymbol{\varepsilon}_z$ are diagonal matrices containing the dielectric function and PML parameters at each point in the grid. In a similar manner, $\boldsymbol{\mu}_x$, $\boldsymbol{\mu}_y$, and $\boldsymbol{\mu}_z$ contain the relative permeability and PML parameters. $\mathbf{D}_{x'}^{E}$, $\mathbf{D}_{y'}^{E}$, and $\mathbf{D}_{z'}^{E}$ are matrices that perform spatial derivative operations on electric fields. Likewise, $\mathbf{D}_{x'}^{H}$, $\mathbf{D}_{y'}^{H}$, and $\mathbf{D}_{z'}^{H}$ are matrices that perform

spatial derivative operations on magnetic fields. The derivative operations are different for electric and magnetic fields due to the staggered nature of the Yee grid. They are defined as

$$\left(\mathbf{D}_x^E \mathbf{E}\right)_{\text{row } i,j,k} = \frac{E^{i+1,j,k} - E^{i,j,k}}{\Delta x'} \tag{3.35}$$

$$\left(\mathbf{D}_y^E \mathbf{E}\right)_{\text{row } i,j,k} = \frac{E^{i,j+1,k} - E^{i,j,k}}{\Delta y'} \tag{3.36}$$

$$\left(\mathbf{D}_z^E \mathbf{E}\right)_{\text{row } i,j,k} = \frac{E^{i,j,k+1} - E^{i,j,k}}{\Delta z'} \tag{3.37}$$

$$\left(\mathbf{D}_{x'}^H \tilde{\mathbf{H}}\right)_{\text{row } i,j,k} = \frac{\tilde{H}^{i,j,k} - \tilde{H}^{i-1,j,k}}{\Delta x'} \tag{3.38}$$

$$\left(\mathbf{D}_{y'}^H \tilde{\mathbf{H}}\right)_{\text{row } i,j,k} = \frac{\tilde{H}^{i,j,k} - \tilde{H}^{i,j-1,k}}{\Delta y'} \tag{3.39}$$

$$\left(\mathbf{D}_{z'}^H \tilde{\mathbf{H}}\right)_{\text{row } i,j,k} = \frac{\tilde{H}^{i,j,k} - \tilde{H}^{i,j,k-1}}{\Delta z'} \tag{3.40}$$

3.4.6.   *Two-Dimensional and One-Dimensional Systems*

Equations (3.29)-(3.34) are valid for fully three-dimensional systems. In many cases, dimensionality of the problem can be reduced to two dimensions or even just one. For problems that can be accurately described in less than three dimensions, this approach offers a tremendous improvement in speed and efficiency and should be considered very good practice. For two dimensional systems where the materials are uniform in the z-direction and propagation is restricted to the *xy*-plane, it follows that $\mathbf{D}_z^E = \mathbf{D}_z^H = \mathbf{0}$. In this case, Eqs. (3.29)-(3.34) separate into two independent modes.

TE Mode

$$\mathbf{D}_{x'}^H \tilde{\mathbf{H}}_y - \mathbf{D}_{y'}^H \tilde{\mathbf{H}}_x = \varepsilon_z \mathbf{E}_z \tag{3.41}$$

$$\mathbf{D}_{y'}^{E}\mathbf{E}_{z} = \mu_{x}\tilde{\mathbf{H}}_{x} \tag{3.42}$$

$$-\mathbf{D}_{x'}^{E}\mathbf{E}_{z} = \mu_{y}\tilde{\mathbf{H}}_{y} \tag{3.43}$$

TM Mode

$$\mathbf{D}_{x'}^{E}\mathbf{E}_{y} - \mathbf{D}_{y'}^{E}\mathbf{E}_{x} = \mu_{z}\tilde{\mathbf{H}}_{z} \tag{3.44}$$

$$\mathbf{D}_{y'}^{H}\tilde{\mathbf{H}}_{z} = \varepsilon_{x}\mathbf{E}_{x} \tag{3.45}$$

$$-\mathbf{D}_{x'}^{H}\tilde{\mathbf{H}}_{z} = \varepsilon_{y}\mathbf{E}_{y} \tag{3.46}$$

One dimensional systems arise when the problem is uniform along two dimensions such

that $\mathbf{D}_{x}^{E} = \mathbf{D}_{y}^{E} = \mathbf{D}_{x}^{H} = \mathbf{D}_{y}^{H} = \mathbf{0}$. In this case, the longitudinal fields are always zero and

Maxwell's equations reduce to two independent modes as suggested by the 1D Yee grids

illustrated in Figure 3-1. These modes are identical in form and may be treated using the same

formulation. These are

E_x Mode

$$-\mathbf{D}_{z'}^{H}\tilde{\mathbf{H}}_{y} = \varepsilon_{x}\mathbf{E}_{x} \tag{3.47}$$

$$\mathbf{D}_{z'}^{E}\mathbf{E}_{x} = \mu_{y}\tilde{\mathbf{H}}_{y} \tag{3.48}$$

E_y Mode

$$\mathbf{D}_{z'}^{H}\tilde{\mathbf{H}}_{x} = \varepsilon_{y}\mathbf{E}_{y} \tag{3.49}$$

$$-\mathbf{D}_{z'}^{E}\mathbf{E}_{y} = \mu_{x}\tilde{\mathbf{H}}_{x} \tag{3.50}$$

Longitudinal Fields

$$\mathbf{E}_{z} = \tilde{\mathbf{H}}_{z} = \mathbf{0} \tag{3.51}$$

3.4.7. Matrix Operators for Spatial Derivatives on a 2D Yee Grid

Maxwell's equations have been cast into matrix form, but little has been said of the square matrices representing derivative operations. In general, any linear operation can be written as a square matrix multiplying a column vector. In addition to derivative operations, these can perform discrete Fourier transforms, convolutions, integrations, multiplications, and more. In any matrix equation, column vectors are usually interpreted as a "state" of the system, while square matrices are interpreted as linear mathematical operations that are performed on the column vectors. More details on finite-difference matrix operators and their application can be found in Refs. [100-103].

From Eqs. (3.35)-(3.40), the derivative operators on a 2D Yee grid are defined as

$$\left(\mathbf{D}_x^E \mathbf{E}\right)_{\text{row } i,j} = \frac{E^{i+1,j} - E^{i,j}}{\Delta x'} \tag{3.52}$$

$$\left(\mathbf{D}_y^E \mathbf{E}\right)_{\text{row } i,j} = \frac{E^{i,j+1} - E^{i,j}}{\Delta y'} \tag{3.53}$$

$$\left(\mathbf{D}_{x'}^H \mathbf{H}\right)_{\text{row } i,j} = \frac{\tilde{H}^{i,j} - \tilde{H}^{i-1,j}}{\Delta x'} \tag{3.54}$$

$$\left(\mathbf{D}_{y'}^H \mathbf{H}\right)_{\text{row } i,j} = \frac{\tilde{H}^{i,j} - \tilde{H}^{i,j-1}}{\Delta y'} \tag{3.55}$$

Figure 3-3 shows a small 4×4 2D Yee grid corresponding to the TE mode. A single electric field component is polarized transverse to the xy-plane. Two magnetic field components are parallel to the xy-plane.

Figure 3-3: Two-dimensional Yee grid for TE mode

The derivative operators on this grid using Dirichlet boundary conditions [103] are

$$
\mathbf{D}_{x'}^{E}=\frac{1}{\Delta x'}
\begin{bmatrix}
-1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1
\end{bmatrix}
$$

$$
\mathbf{D}_{y'}^{E}=\frac{1}{\Delta y'}
\begin{bmatrix}
-1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1
\end{bmatrix}
\tag{3.56}
$$

$$
\mathbf{D}_{x'}^{H}=\frac{1}{\Delta x'}
\begin{bmatrix}
+1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
-1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1
\end{bmatrix}
$$

$$
\mathbf{D}_{y'}^{H}=\frac{1}{\Delta y'}
\begin{bmatrix}
+1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
-1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1
\end{bmatrix}
\tag{3.57}
$$

Each of these matrices contain nonzero elements on just two diagonals leading to very sparse matrices that can be constructed very quickly and stored efficiently. All diagonals are uniform except where quantities are written in red. These exceptions arise from the Dirichlet boundary conditions.

Periodic boundaries can also be implemented in the derivative operators. For an incident plane wave $\vec{A}e^{-j\vec{k}\cdot\vec{r}}$, the derivative operators become

$$\mathbf{D}_{x'}^{E}=\frac{1}{\Delta x'}\begin{bmatrix}
-1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
\textcolor{blue}{\phi_x} & 0 & 0 & -1 & \textcolor{red}{0} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & \textcolor{blue}{\phi_x} & 0 & 0 & 0 & -1 & \textcolor{red}{0} & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \textcolor{blue}{\phi_x} & 0 & 0 & -1 & \textcolor{red}{0} & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \textcolor{blue}{\phi_x} & 0 & 0 & -1 & \\
\end{bmatrix}$$

$$\mathbf{D}_{y'}^{E}=\frac{1}{\Delta y'}\begin{bmatrix}
-1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1\\
\textcolor{blue}{\phi_y} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0\\
0 & \textcolor{blue}{\phi_y} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0\\
0 & 0 & \textcolor{blue}{\phi_y} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0\\
0 & 0 & 0 & \textcolor{blue}{\phi_y} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1\\
\end{bmatrix} \tag{3.58}$$

$$\mathbf{D}_{x'}^{H}=\frac{1}{\Delta x'}\begin{bmatrix}
+1 & 0 & 0 & -\textcolor{blue}{\phi_x^*} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
-1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & \textcolor{red}{0} & +1 & 0 & 0 & -\textcolor{blue}{\phi_x^*} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \textcolor{red}{0} & +1 & 0 & 0 & -\textcolor{blue}{\phi_x^*} & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \textcolor{red}{0} & +1 & 0 & 0 & -\textcolor{blue}{\phi_x^*}\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1\\
\end{bmatrix}$$

$$\mathbf{D}_{y'}^{H}=\frac{1}{\Delta y'}\begin{bmatrix}
+1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\textcolor{blue}{\phi_y^*} & 0 & 0 & 0\\
0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\textcolor{blue}{\phi_y^*} & 0 & 0\\
0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\textcolor{blue}{\phi_y^*} & 0\\
0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\textcolor{blue}{\phi_y^*}\\
-1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & +1\\
\end{bmatrix} \tag{3.59}$$

$$\phi_x = e^{-jk_x N_x \Delta x} \qquad\qquad \phi_x^* = e^{+jk_x N_x \Delta x} \tag{3.60}$$

$$\phi_y = e^{-jk_y N_y \Delta y} \qquad\qquad \phi_y^* = e^{+jk_y N_y \Delta y} \tag{3.61}$$

These matrices include terms written in blue along an additional diagonal. When a field is required from outside of the problem space, periodic boundaries dictate it is possible to use a field from the opposite side of the grid. Phase due to an oblique angle of incidence is compensated through the terms ϕ_x and ϕ_y.

3.5. Formulation

3.5.1. *Matrix Wave Equation*

Solving Eqs. (3.42)-(3.43) for the magnetic field quantities and substituting the expressions into Eq. (3.41) leads to the wave equation for the TE mode in terms of just $\mathbf{E}_z$. A similar treatment of Eqs. (3.44)-(3.46) leads to the wave equation for the TM mode.

TE Mode

$$\left(\mathbf{D}_{x'}^{H} \boldsymbol{\mu}_y^{-1} \mathbf{D}_{x'}^{E} + \mathbf{D}_{y'}^{H} \boldsymbol{\mu}_x^{-1} \mathbf{D}_{y'}^{E} + \boldsymbol{\varepsilon}_z \right) \mathbf{E}_z = \mathbf{0} \tag{3.62}$$

$$\tilde{\mathbf{H}}_x = \boldsymbol{\mu}_x^{-1} \mathbf{D}_{y'}^{E} \mathbf{E}_z \tag{3.63}$$

$$\tilde{\mathbf{H}}_y = -\boldsymbol{\mu}_y^{-1} \mathbf{D}_{x'}^{E} \mathbf{E}_z \tag{3.64}$$

TM Mode

$$\left(\mathbf{D}_{x'}^{E} \boldsymbol{\varepsilon}_y^{-1} \mathbf{D}_{x'}^{H} + \mathbf{D}_{y'}^{E} \boldsymbol{\varepsilon}_x^{-1} \mathbf{D}_{y'}^{H} + \boldsymbol{\mu}_z \right) \tilde{\mathbf{H}}_z = \mathbf{0} \tag{3.65}$$

$$\mathbf{E}_x = \boldsymbol{\varepsilon}_x^{-1} \mathbf{D}_{y'}^{H} \tilde{\mathbf{H}}_y \tag{3.66}$$

$$\mathbf{E}_y = -\boldsymbol{\varepsilon}_y^{-1} \mathbf{D}_{x'}^{H} \tilde{\mathbf{H}}_z \tag{3.67}$$

3.5.2. *Incorporation of a Source*

It is important to note the matrix equations derived above are not yet solvable. They relate field components in a scattering environment, but a source must be incorporated or the solution will be trivial. There are two alternative ways to incorporate a source. These are the total-field/scattered-field (TF/SF) technique [104] and the pure scattered-field formulation [94, 105]. The TF/SF method is preferred here because it leads to a more intuitive picture of the fields and the matrix equation is in a form more suitable to efficient solution.
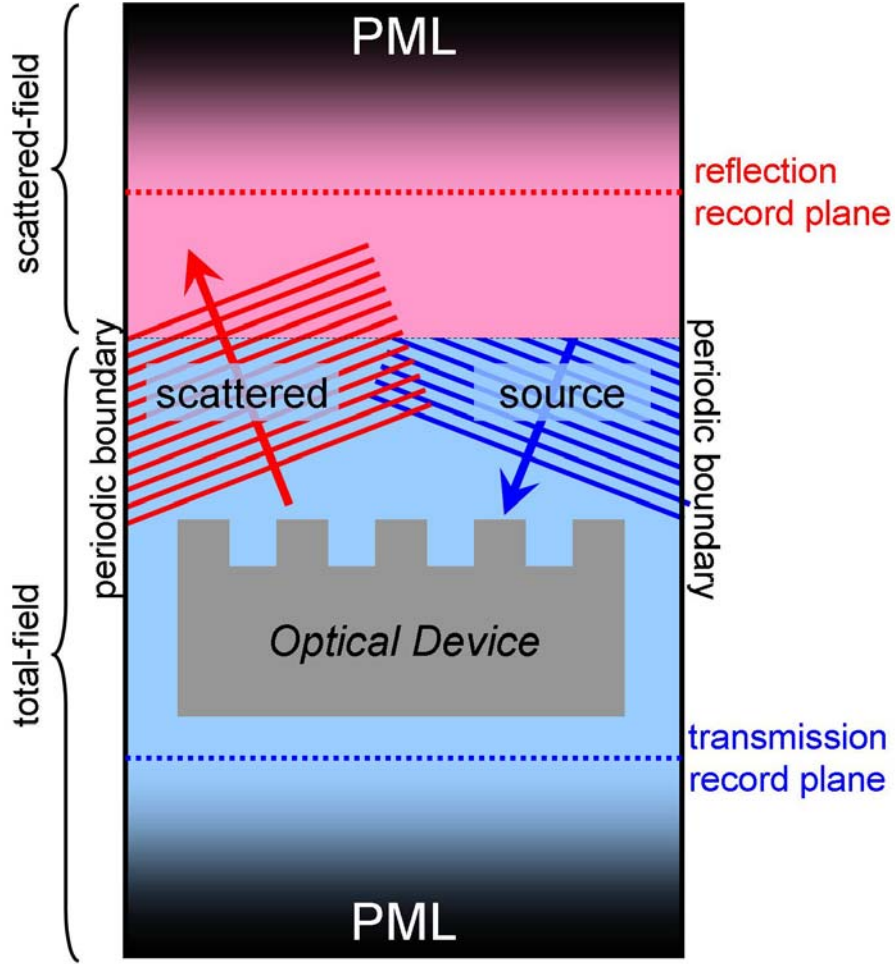
Figure 3-4: Construction of FDFD model

Both TE and TM matrix equations can be written in the following general form.

$$\mathbf{L}\mathbf{f}_z = \mathbf{0} \tag{3.68}$$

where $\mathbf{f}_z$ represents the transverse field and $\mathbf{L}$ is a "field" matrix representing the linear operation dictated by the wave equation. These were found to be

$$\text{2D Systems: } \mathbf{L} = \begin{cases} \boldsymbol{\varepsilon}_z + \mathbf{D}_{x'}^{H}\boldsymbol{\mu}_y^{-1}\mathbf{D}_{x'}^{E} + \mathbf{D}_{y'}^{H}\boldsymbol{\mu}_x^{-1}\mathbf{D}_{y'}^{E} & \text{TE} \\ \boldsymbol{\mu}_z + \mathbf{D}_{x'}^{E}\boldsymbol{\varepsilon}_y^{-1}\mathbf{D}_{x'}^{H} + \mathbf{D}_{y'}^{E}\boldsymbol{\varepsilon}_x^{-1}\mathbf{D}_{y'}^{H} & \text{TM} \end{cases} \tag{3.69}$$

$$\text{1D Systems: } \mathbf{L} = \mathbf{D}_{z'}^{H}\boldsymbol{\mu}^{-1}\mathbf{D}_{z'}^{E} + \boldsymbol{\varepsilon} \tag{3.70}$$

76

The TF/SF formulation divides the problem space into two regions as illustrated in Figure 3-4. In the total-field region, field quantities contain the source and fields reflected by the structure being modeled. The scattered-field region does not contain the source.

When finite-difference equations contain field terms from both sides of the interface, the source must added to scattered-field quantities or subtracted from total-field quantities. These "correction" terms are known quantities and can be moved to the right side of Eq. (3.68) to obtain an equation of the form

$$\mathbf{L}\mathbf{f}_z = \mathbf{S} \tag{3.71}$$

This can be solved through matrix division using any of a variety of elimination or iterative methods [96].

$$\mathbf{f}_z = \mathbf{L}^{-1}\mathbf{S} \tag{3.72}$$

To calculate the source vector $\mathbf{S}$, the source field vector $\mathbf{f}_{src}$ should be calculated in a problem space that is made homogenous consistent with the material properties where the source is to be injected. Next, the scattered-field "masking" matrix $\mathbf{Q}$ is constructed where diagonal elements are set to 1 where they correspond to points in the scattered-field, and 0 elsewhere. Using this matrix, the source field can be isolated to the scattered-field or total-field regions.

$$\mathbf{f}_{SF} = \mathbf{Q}\mathbf{f}_{src} \tag{3.73}$$

$$\mathbf{f}_{TF} = (\mathbf{I} - \mathbf{Q})\mathbf{f}_{src} \tag{3.74}$$

To match the scattered-field to the total-field, the source must be subtracted from total-field terms appearing in scattered-field finite-difference equations. The quantity that must be subtracted is $\mathbf{L}\mathbf{f}_{TF}$, but this should only be subtracted from the scattered-field equations. Using the $\mathbf{Q}$ matrix do this, Eq. (3.68) is modified to match the scattered-field equations to the total-field region.

$$\mathbf{Lf}_z - \mathbf{QLf}_{TF} = \mathbf{0} \tag{3.75}$$

In a similar manner, the source must be added to scattered-field terms appearing in total-field finite-difference equations. The quantity that must be added is $\mathbf{Lf}_{SF}$, but it should only be added to total-field equations. Using the matrix $\mathbf{Q}$ do this, Eq. (3.75) is further modified to match the total-field equations to the scattered-field region.

$$\mathbf{Lf}_z - \mathbf{QLf}_{TF} + (\mathbf{I} - \mathbf{Q})\mathbf{Lf}_{SF} = \mathbf{0} \tag{3.76}$$

The last two terms on the left side of this equation are known quantities and can be moved to the right side of the equation. Comparing this to Eq. (3.71) shows the source vector to be

$$\mathbf{S} = \mathbf{QLf}_{TF} - (\mathbf{I} - \mathbf{Q})\mathbf{Lf}_{SF} \tag{3.77}$$

A simpler expression can be obtained by substituting Eqs. (3.73)-(3.74) into this equation. After some algebraic manipulation, the source vector $\mathbf{S}$ can be calculated according to

$$\mathbf{S} = (\mathbf{QL} - \mathbf{LQ})\mathbf{f}_{src} \tag{3.78}$$

This method of calculating the source vector is very generic and accommodates any type of grid, any order-accurate derivatives, and any shape scattered-field. Equation (3.78) only modifies finite-difference equations containing terms from both sides of the TF/SF boundary. To speed calculation, finite-difference equations can be modified directly without implementing this general purpose approach.

3.5.3. *Calculation of Reflected and Transmitted Fields*

The finite-difference frequency-domain method can be used to visualize fields or calculate power in reflected and transmitted waves. In periodic structures, diffraction efficiency is most often employed. After Eq. (3.72) is solved, the field in the reflection and transmission record planes are extracted from $\mathbf{f}_z$. These are $f_{ref}(x)$ and $f_{trn}(x)$ respectively.

First, the phase gradient due to oblique angle of incidence is removed. In addition, field amplitudes are normalized to the source amplitude. This is accomplished using the following equations.

$$f'_{\text{ref}}(x) = f_{\text{ref}}(x)/f_{\text{src}}(x) \qquad f'_{\text{trn}}(x) = f_{\text{trn}}(x)/f_{\text{src}}(x) \qquad (3.79)$$

Second, the fast Fourier transform (FFT) [106] of these functions is calculated to obtain the angular spectrum of reflected and transmitted waves. These are expressed as.

$$R(m) = \text{FFT}\left[f'_{\text{ref}}(x)\right] \qquad T(m) = \text{FFT}\left[f'_{\text{trn}}(x)\right] \qquad (3.80)$$

Third, wave vector components corresponding to spatial harmonics comprising the angular spectrum are calculated. For an angle of incidence θ, these are

$$k_m^x = k_0 n_{\text{inc}} \sin\theta - \frac{2\pi m}{\Lambda} \qquad (3.81)$$

$$k_m^{y,\text{ref}} = \begin{cases} \sqrt{\left(k_0 n_{\text{ref}}\right)^2 - \left(k_m^x\right)^2} & k_0 n_{\text{ref}} > k_m^x \\ -j\sqrt{\left(k_m^x\right)^2 - \left(k_0 n_{\text{ref}}\right)^2} & k_0 n_{\text{ref}} < k_m^x \end{cases} \qquad (3.82)$$

$$k_m^{y,\text{trn}} = \begin{cases} \sqrt{\left(k_0 n_{\text{trn}}\right)^2 - \left(k_m^x\right)^2} & k_0 n_{\text{trn}} > k_m^x \\ -j\sqrt{\left(k_m^x\right)^2 - \left(k_0 n_{\text{trn}}\right)^2} & k_0 n_{\text{trn}} < k_m^x \end{cases} \qquad (3.83)$$

Fourth, the fraction of power in the spatial harmonics is calculated using Eqs. (3.84)-(3.85). A derivation of these equations can be found in Section 6.2.13

$$P_{\text{ref}}(m) = R^2(m) \cdot \text{Re}\left[\frac{k_m^{y,\text{ref}}}{k_0 n_{\text{inc}} \cos\theta}\right] \qquad (3.84)$$

$$P_{\text{trn}}(m) = T^2(m) \cdot \text{Re}\left[\frac{k_m^{y,\text{trn}}}{k_0 n_{\text{inc}} \cos\theta}\right] \qquad (3.85)$$

Start →

COMPUTE YEE GRID
$N_x, N_y, \Delta x, \Delta y$

COMPUTE 2X GRID
$N'_x, N'_y, \Delta x', \Delta y'$

CONSTRUCT 2X PROBLEM
$\varepsilon'_{i,j}$ and $\mu'_{i,j}$

CALCULATE 2X PML
s'_x, s'_y

OVERLAY PML AND MATERIALS ON YEE GRID
$$\varepsilon^x_{i,j} = \varepsilon'_{2i,2j-1} s^y_{2i,2j-1} / s^x_{2i,2j-1} \qquad \mu^x_{i,j} = \mu'_{2i,2j-1} s^y_{2i,2j-1} / s^x_{2i,2j-1}$$
$$\varepsilon^y_{i,j} = \varepsilon'_{2i,2j-1} s^x_{2i,2j-1} / s^y_{2i,2j-1} \qquad \mu^y_{i,j} = \mu'_{2i,2j-1} s^x_{2i,2j-1} / s^y_{2i,2j-1}$$

COMPUTE DERIVATIVE OPERATORS
$$\mathbf{D}^E_x, \ \mathbf{D}^E_y, \ \mathbf{D}^E_z \qquad \mathbf{D}^H_x, \ \mathbf{D}^H_y, \ \mathbf{D}^H_z$$

COMPUTE FIELD MATRIX
$$\mathbf{L} = \begin{cases} \boldsymbol{\varepsilon}_z + \mathbf{D}^H_x \boldsymbol{\mu}^{-1}_y \mathbf{D}^E_x + \mathbf{D}^H_y \boldsymbol{\mu}^{-1}_x \mathbf{D}^E_y & \text{TE} \\ \boldsymbol{\mu}_z + \mathbf{D}^E_x \boldsymbol{\varepsilon}^{-1}_y \mathbf{D}^H_x + \mathbf{D}^E_y \boldsymbol{\varepsilon}^{-1}_x \mathbf{D}^H_y & \text{TM} \end{cases}$$

COMPUTE SOURCE VECTOR
$$\mathbf{S} = (\mathbf{QL} - \mathbf{LQ})\mathbf{F}_{\text{src}}$$

COMPUTE SOURCE FIELD
$$F_{\text{src}}(x,y) = \exp\left[-jk_0 n_{\text{inc}}(x\sin\theta + y\cos\theta)\right]$$

SOLVE MATRIX DIVISION
$$\mathbf{F}_z = \mathbf{L}^{-1}\mathbf{S}$$

EXTRACT OUTPUT FIELDS
$\mathbf{F}_{\text{ref}}$ and $\mathbf{F}_{\text{trn}}$

NORMALIZE FIELDS
$$\mathbf{F}'_{\text{ref}} = \mathbf{F}'_{\text{ref}} \div \mathbf{F}_{\text{src}} \qquad \mathbf{F}'_{\text{trn}} = \mathbf{F}'_{\text{trn}} \div \mathbf{F}_{\text{src}}$$

COMPUTE ANGULAR SPECTRUM
$$\mathbf{R} = \text{FFT}\left[\mathbf{F}'_{\text{ref}}\right] \qquad \mathbf{T} = \text{FFT}\left[\mathbf{F}'_{\text{trn}}\right]$$

CALCULATE WAVE VECTOR COMPONENTS
$$k^x_m = k_0 n_{\text{inc}} \sin\theta - \frac{2\pi m}{\Lambda} \qquad k^y_m = \begin{cases} \sqrt{(k_0 n)^2 - (k^x_m)^2} & k_0 n > k^x_m \\ -j\sqrt{(k^x_m)^2 - (k_0 n)^2} & k_0 n < k^x_m \end{cases}$$

CALCULATE POWER IN DIFFRACTED ORDERS
$$\mathbf{P}_{\text{ref}} = \mathbf{R}^2 \cdot \text{Re}\left[\frac{\mathbf{K}_{y,\text{ref}}}{k_0 n_{\text{inc}} \cos\theta}\right] \qquad \mathbf{P}_{\text{trn}} = \mathbf{T}^2 \cdot \text{Re}\left[\frac{\mathbf{K}_{y,\text{trn}}}{k_0 n_{\text{inc}} \cos\theta}\right]$$

Stop ←

Figure 3-5: Block diagram of FDFD implementation

3.6. Implementation

Figure 3-5 shows a block diagram of how FDFD can be implemented. Five main steps are highlighted in distinct colors. The first step calculates grid parameters, builds the structure to

be modeled, and assigned material properties to points on the Yee grid. The second step calculates the field matrix. Step three calculates the source vector. Step four solves the matrix division problem to calculate the field throughout the problem space. Step five calculates reflected and transmitted energy.

3.6.1.   Causes of Error

It is important to understand what mechanisms produce modeling errors to better understand limitations of the method and how to improve accuracy. Numerical dispersion has already been discussed and a means to compensate for it was outlined. Some of the most common sources of error are demonstrated in Figure 3-6.

Numerical error arising from the finite-difference approximations of the spatial derivatives can be severe when grid resolution is made too coarse. Experience has shown that grid resolution should be finer than a tenth of a wavelength, or a tenth of the finest feature to be resolved. In Figure 3-6(a), reflection spectra is calculated using various grid resolutions and compared to an exact solution obtained using RCWA. The most obvious artifact of poor grid resolution is shifted spectral response due to numerical dispersion. As grid resolution is made higher, position of the resonance asymptotically approaches the exact solution.

In contrast, rounding errors become more severe when grid resolution is high. As the number of simultaneous equations increase, many computation algorithms such as Gauss elimination will show increasing error as more equations are solved. For these reasons, it is good practice to use as coarse of a grid as possible while still achieving an acceptable level of accuracy. In many cases, iterative methods are able to circumvent this source of error [96].
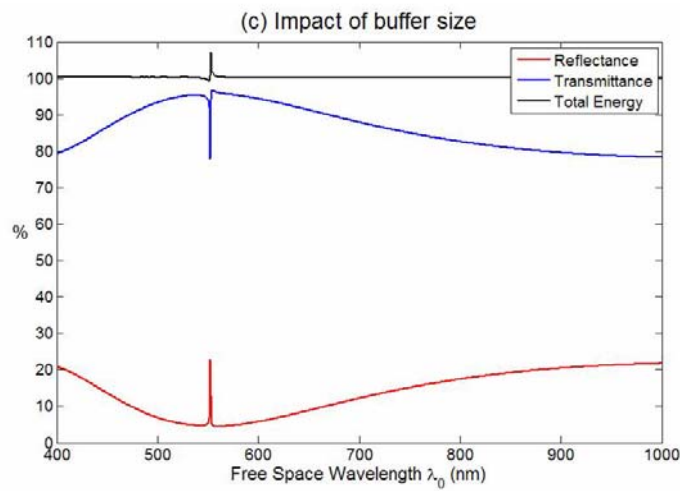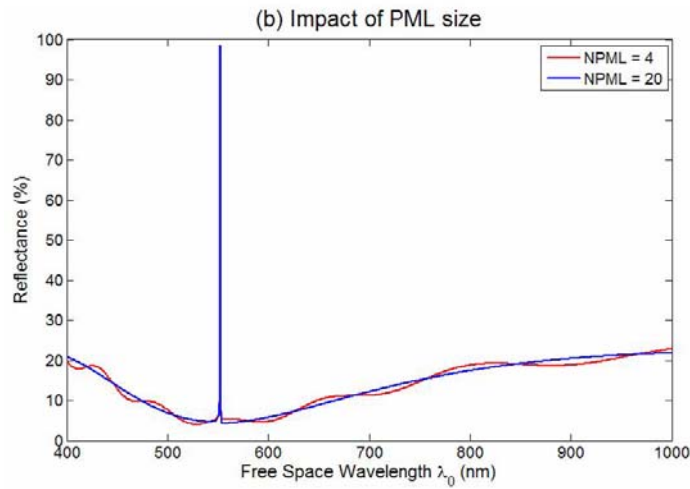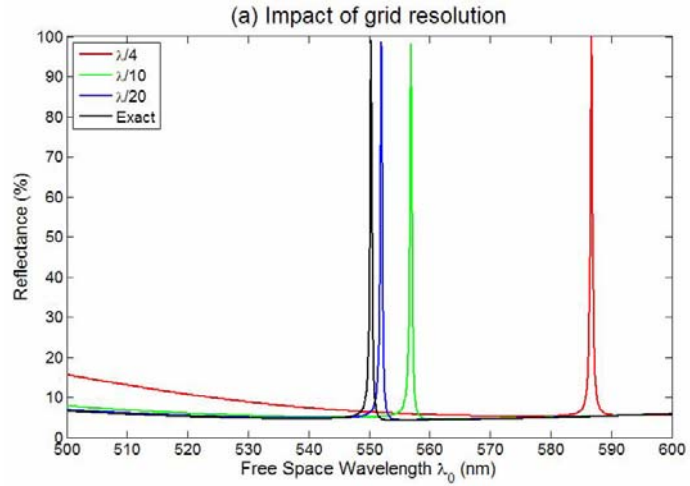
Figure 3-6: Examples of numerical error

The Yee grid is Cartesian in nature which makes it difficult to represent curved surfaces. When such surfaces must be resolved, it is often necessary to use a fine grid along with any of the techniques discussed previously for representing materials on discrete grid. It is good practice to test for "convergence." Simulated parameters are plotted as a function of grid resolution. An exponentially converging profile is usually observed as the structure becomes better represented. An optimum grid resolution can be chosen by selecting the grid resolution where accuracy no longer improves significantly and run-time is sufficiently fast.

Nonphysical reflections from grid boundaries will always be present regardless of what absorbing boundary condition is used. This results in a standing wave in the region where reflected and transmitted power is calculated. Field amplitudes will be determined inaccurately and error will result in a form that produces oscillations in the spectral response. This is demonstrated in Figure 3-6(b) where results are compared using a PML of different dimensions. The PML with only four grid cells produces higher reflections from the boundaries producing the oscillations.

Wood's anomalies of the Rayleigh type are particularly problematic when they exist in the transmitted or reflected regions. Waves propagating at near 90° are not handled well at the longitudinal boundaries and very large error can result that violates conservation of energy. Fortunately, most devices are not designed to operate this way so results in this part of the spectrum can usually be ignored.

An often overlooked source of error is the size of "empty" space placed between the structure being modeled and the absorbing boundaries. Devices with large evanescent fields tend to couple to absorbing boundaries if they are too close. This produces inaccurate calculations of

reflected and transmitted power and affects conservation of energy. Figure 3-6(c) shows the spectral response of the same device where the buffer region ("empty space") between the structure and boundary was made too small. Transmission and reflection are no longer determined accurately and conservation of energy is violated on resonance. This is due to the evanescent field of the guided mode coupling to the PML boundary condition.

To identify numerical error, it is good practice to check for conservation of energy. When loss is not being considered, total reflected power plus total transmitted power should always equal 100%. Nonphysical reflections from the boundary usually lead to an oscillating error in conservation of energy. Wood's anomalies usually produce sharp and isolated spikes or dips in conservation of energy.

CHAPTER 4
FINITE-DIFFERENCE TIME-DOMAIN METHOD

4.1. Overview

The finite-difference time-domain (FDTD) method [43, 44, 97] is a very powerful technique for broadband characterization and field visualization. It accommodates parallel processing and can be implemented without using linear algebra so it is very popular for rigorous simulations of large scale problems and structures with complicated geometry. It is a time-domain method so it naturally incorporates material and device nonlinearities and may be used to model active devices. The method is incredibly versatile and has likely been applied to more problems than any other method. It is accurate and sources of numerical error are well understood. It is fundamentally a scattering algorithm, but it can be used to calculate eigen-modes and photonic band diagrams [107-109]. An enormous library of literature can be found on this method including books devoted entirely to FDTD.

The method works by iterating Maxwell's curl equations to evolve fields over time, essentially making an animated movie of the fields. By exciting the problem with an impulse, the response over an enormous span of frequencies can be obtained in a single simulation. Steady-state fields can also be obtained at multiple frequencies in the same simulation by integration. The method requires an initial investment of computer resources, but scales linearly from there because it is not based on linear algebra. For this reason, FDTD is slow and inefficient for smaller problems, but becomes the more efficient method as problem size grows. FDTD is considered a poor method for simulating highly resonant devices because an enormous number of iterations is necessary to resolve the shape of sharp resonances or other abrupt

85

features in the spectral response. It is a very powerful tool, however, for determining if a structure is resonant and determining location of the resonances.
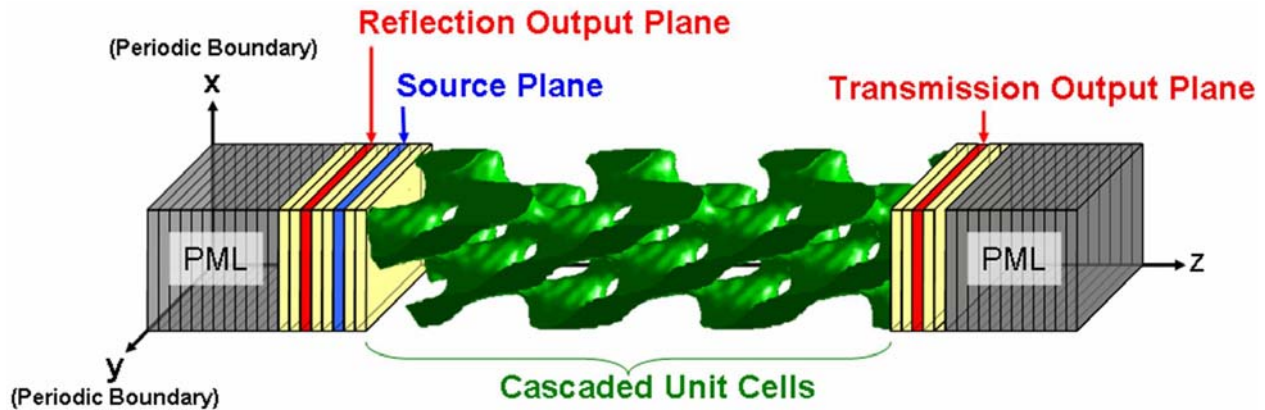
Figure 4-1: FDTD model construction

A serious drawback of FDTD is that it is not always possible to handle oblique angles of incidence. When a structure is excited by a pulsed source, periodic boundaries make this difficult. The angled-update method can be used, but angle of incidence is limited to less than 45° for 2D simulations and less then 35° for 3D simulations [110]. When only a single frequency is of interest, no limitation is posed and a variety of methods can be used [111-114].

This chapter will step the reader through the formulation and implementation of a rigorous three-dimensional code that incorporates loss, gain, permittivity, permeability, material dispersion, and a perfectly matched layer (PML) absorbing boundary condition. To speed simulation, the formulation includes a "background" loss, permittivity, and permeability that operates independent of the dispersion model. While more efficient implementations are possible, this approach is more accommodating in a research environment because it is more modular. Features such as material dispersion, loss, gain, or nonlinearities can be easily added or

86

removed from the code. Dispersion and gain are implemented by calculating material

polarization and polarization current as intermediate parameters. This offers the benefit of

utilizing a physically meaningful parameter that can be visualized instead of using abstract

integration terms.

4.2. Formulation

This section will present the so-called "D-H" formulation of FDTD. This approach

enables the PML to be implemented independent of dielectric materials without splitting the

fields. The fields $\vec{H}$, $\vec{D}$, and $\vec{E}$ are computed at each time step. All material properties,

including dispersion and nonlinearities, are incorporated in the step where $\vec{E}$ is computed from

$\vec{D}$ so features are easily added or removed. The PML operates independently of any dielectric

materials so no interaction has to be considered. Dielectric structures extending into the PML

can be made to appear as if they extend indefinitely.

Like FDFD described last chapter, FDTD will make use of the Yee grid so the divergence

conditions will be implicitly satisfied. In this manner, field update equations can be based only

on Maxwell's curl equations and the constitutive relations. The Yee grid is arranged so that

electric fields circulate around magnetic fields and magnetic fields circulate around electric

fields. This highly efficiency arrangement is illustrated in Figure 4-2 and is a critical aspect of

any three dimensional finite-difference method. Fields from outside the grid are shown as faded

arrows. Yee grid unit cells are enclosed in gray boxes.

Formulation of the update equations starts with Maxwell's equations in the frequency-

domain where material and PML parameters are incorporated more intuitively. These equations

will be converted to the time-domain where derivatives will be approximated by finite-

differences and integrations approximated by summations. Field update equations are derived by

87

solving these equations for the future field values. To speed simulation and simplify the update

equations, all constants are collected into update coefficients so they may be calculated outside

of the main FDTD loop.

Single Unit Cell   **2×2 Grid**
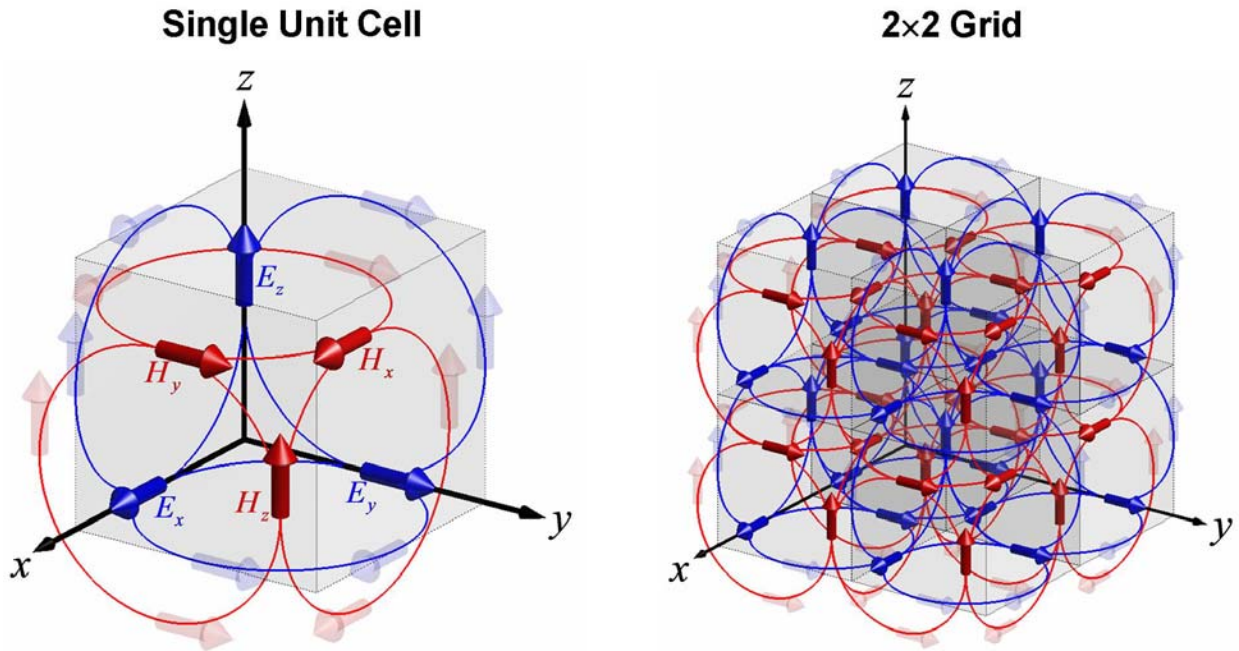
Figure 4-2: Circulating fields in 3D Yee grid

4.2.1. *Maxwell's Equations*

Maxwell's curl equations in the frequency-domain can be written as

$$\nabla \times \vec{E}_{\omega} = -j\omega\mu\vec{H}_{\omega} \tag{4.1}$$

$$\nabla \times \vec{H}_{\omega} = \sigma\vec{E}_{\omega} + j\omega\vec{D}_{\omega} \tag{4.2}$$

where all physical materials properties except an optional "background" loss are implemented through the constitutive relation

$$\vec{D}_\omega = \varepsilon_\omega^* \vec{E}_\omega = \varepsilon_0 \varepsilon_{r,\infty} \vec{E}_\omega + \vec{P}_\omega \tag{4.3}$$

This equation is not consistent with Eq. (1.49) because an additional "background" dielectric constant $\varepsilon_{r,\infty}$ has been incorporated that operates separately from the dispersion model. This improves efficiency of the code when it is not necessary to account for material dispersion.

It is well known for plane waves that the magnitude of $\vec{E}_\omega$ and $\vec{H}_\omega$ are related through material impedance as $|\vec{E}_\omega|/|\vec{H}_\omega| = \eta$. For free space, $\eta \approx 377\ \Omega$ illustrating that the electric and magnetic fields are roughly three orders of magnitude different. To minimize rounding errors, all parameters associated with the electric field are normalized in the following manner.

$$\tilde{\vec{E}}_\omega = \frac{1}{\eta_0} \vec{E}_\omega = \sqrt{\frac{\varepsilon_0}{\mu_0}} \vec{E}_\omega \qquad \tilde{\vec{D}}_\omega = c_0 \vec{D}_\omega = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} \vec{D}_\omega \tag{4.4}$$

$$\tilde{\vec{J}}_\omega = c_0 \vec{J}_\omega = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} \vec{J}_\omega \qquad \tilde{\vec{P}}_\omega = c_0 \vec{P}_\omega = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} \vec{P}_\omega \tag{4.5}$$

Using normalized parameters, Eqs. (4.1)-(4.3) become

$$\frac{c_0}{\mu_r} \nabla \times \tilde{\vec{E}}_\omega = -j\omega \vec{H}_\omega \tag{4.6}$$

$$c_0 \nabla \times \vec{H}_\omega = \frac{\sigma}{\varepsilon_0} \tilde{\vec{E}}_\omega + j\omega \tilde{\vec{D}}_\omega \tag{4.7}$$

$$\tilde{\vec{D}}_\omega = \varepsilon_r^*(\omega) \tilde{\vec{E}}_\omega = \varepsilon_{r,\infty} \tilde{\vec{E}}_\omega + \tilde{\vec{P}}_\omega \tag{4.8}$$

4.2.2. Incorporation of Boundary Conditions

To absorb outgoing waves, a PML is incorporated into Maxwell's curl equations through the tensor $\bar{\bar{S}}$ as follows.

$$\frac{c_0}{\mu_r} \nabla \times \vec{\tilde{E}} = -j\omega\bar{\bar{S}}\vec{H}$$ (4.9)

$$c_0 \nabla \times \vec{H} = \frac{\sigma}{\varepsilon_0}\vec{\tilde{E}}_\omega + j\omega\bar{\bar{S}}\vec{\tilde{D}}$$ (4.10)

The tensor $\bar{\bar{S}}$ is defined as

$$\bar{\bar{S}} = \begin{bmatrix} \dfrac{s_y s_z}{s_x} & 0 & 0 \\ 0 & \dfrac{s_x s_z}{s_y} & 0 \\ 0 & 0 & \dfrac{s_x s_y}{s_z} \end{bmatrix} \qquad \begin{aligned} s_x &= 1 + \frac{\sigma'_x(x)}{j\omega\varepsilon_0} \\ s_y &= 1 + \frac{\sigma'_y(y)}{j\omega\varepsilon_0} \\ s_z &= 1 + \frac{\sigma'_z(z)}{j\omega\varepsilon_0} \end{aligned}$$ (4.11)

The PML works by incorporating fictitious loss terms σ'_x, σ'_y, and σ'_z while simultaneously matching impedance. In this manner, fields decay without being artificially reflected back into the problem space. For best performance, the loss terms should increase gradually into the PML.

4.2.3. Vector Expansion of Maxwell's Equations

Expanding Eqs. (4.9)-(4.10) and Eq. (4.8) into vector components leads to the following nine equations. For simplicity the ω subscripts have been dropped, but the fields are still frequency-domain quantities at this point.

$$j\omega\left(1 + \frac{\sigma^*_x}{j\omega\varepsilon_0}\right)^{-1}\left(1 + \frac{\sigma^*_y}{j\omega\varepsilon_0}\right)\left(1 + \frac{\sigma^*_z}{j\omega\varepsilon_0}\right)H_x = -\frac{c_0}{\mu_r}\left(\frac{\partial \tilde{E}_z}{\partial y} - \frac{\partial \tilde{E}_y}{\partial z}\right)$$ (4.12)

$$j\omega\left(1+\frac{\sigma_x^*}{j\omega\varepsilon_0}\right)\left(1+\frac{\sigma_y^*}{j\omega\varepsilon_0}\right)^{-1}\left(1+\frac{\sigma_z^*}{j\omega\varepsilon_0}\right)H_y = -\frac{c_0}{\mu_r}\left(\frac{\partial\tilde{E}_x}{\partial z}-\frac{\partial\tilde{E}_z}{\partial x}\right)$$ (4.13)

$$j\omega\left(1+\frac{\sigma_x^*}{j\omega\varepsilon_0}\right)\left(1+\frac{\sigma_y^*}{j\omega\varepsilon_0}\right)\left(1+\frac{\sigma_z^*}{j\omega\varepsilon_0}\right)^{-1}H_z = -\frac{c_0}{\mu_r}\left(\frac{\partial\tilde{E}_y}{\partial x}-\frac{\partial\tilde{E}_x}{\partial y}\right)$$ (4.14)

$$j\omega\left(1+\frac{\sigma_x^*}{j\omega\varepsilon_0}\right)^{-1}\left(1+\frac{\sigma_y^*}{j\omega\varepsilon_0}\right)\left(1+\frac{\sigma_z^*}{j\omega\varepsilon_0}\right)\tilde{D}_x = c_0\left(\frac{\partial H_z}{\partial y}-\frac{\partial H_y}{\partial z}\right)-\frac{\sigma_x}{\varepsilon_0}\tilde{E}_x$$ (4.15)

$$j\omega\left(1+\frac{\sigma_x^*}{j\omega\varepsilon_0}\right)\left(1+\frac{\sigma_y^*}{j\omega\varepsilon_0}\right)^{-1}\left(1+\frac{\sigma_z^*}{j\omega\varepsilon_0}\right)\tilde{D}_y = c_0\left(\frac{\partial H_x}{\partial z}-\frac{\partial H_z}{\partial x}\right)-\frac{\sigma_y}{\varepsilon_0}\tilde{E}_y$$ (4.16)

$$j\omega\left(1+\frac{\sigma_x^*}{j\omega\varepsilon_0}\right)\left(1+\frac{\sigma_y^*}{j\omega\varepsilon_0}\right)\left(1+\frac{\sigma_z^*}{j\omega\varepsilon_0}\right)^{-1}\tilde{D}_z = c_0\left(\frac{\partial H_y}{\partial x}-\frac{\partial H_x}{\partial y}\right)-\frac{\sigma_z}{\varepsilon_0}\tilde{E}_z$$ (4.17)

$$\tilde{D}_x = \varepsilon_x^*\tilde{E}_x = \varepsilon_{x,\infty}\tilde{E}_x + \tilde{P}_x$$ (4.18)

$$\tilde{D}_y = \varepsilon_y^*\tilde{E}_y = \varepsilon_{y,\infty}\tilde{E}_y + \tilde{P}_y$$ (4.19)

$$\tilde{D}_z = \varepsilon_z^*\tilde{E}_z = \varepsilon_{z,\infty}\tilde{E}_z + \tilde{P}_z$$ (4.20)

4.2.4. H-Field Update Equations

To derive update equations from Eqs. (4.12)-(4.14), they must first be converted to the time-domain. Proceeding with Eq. (4.12), the inverse term on the left is brought to the right and the remaining terms multiplied. This leads to

$$j\omega H_x + \frac{\sigma_y^*+\sigma_z^*}{\varepsilon_0}H_x + \frac{\sigma_y^*\sigma_z^*}{\varepsilon_0^2}\frac{1}{j\omega}H_x = -\frac{c_0}{\mu_r}C_{Ex} - \frac{c_0\sigma_x^*}{\mu_r\varepsilon_0}\frac{1}{j\omega}C_{Ex}$$ (4.21)

$$C_{Ex} = \frac{\partial\tilde{E}_z}{\partial y} - \frac{\partial\tilde{E}_y}{\partial z}$$ (4.22)

This can be converted to the time-domain using the following properties of the Fourier transform.

$$F\left\{\frac{\partial g(t)}{\partial t}\right\} = j\omega G(\omega) \qquad F\left\{\int_{-\infty}^{t} g(\tau) d\tau\right\} = \frac{1}{j\omega} G(\omega) \qquad (4.23)$$

Equation (4.21) in the time-domain is then

$$\frac{\partial H_x}{\partial t} + \frac{\sigma_y^* + \sigma_z^*}{\varepsilon_0} H_x + \frac{\sigma_y^* \sigma_z^*}{\varepsilon_0^2} \int_{-\infty}^{t} H_x d\tau = -\frac{c_0}{\mu_r} C_{Ex} - \frac{c_0 \sigma_x^*}{\mu_r \varepsilon_0} \int_{-\infty}^{t} C_{Ex} d\tau \qquad (4.24)$$

The derivatives are numerically approximated using finite-differences while the integrations are approximated by summations.

$$\frac{H_x^{t+\Delta t/2} - H_x^{t-\Delta t/2}}{\Delta t} + \frac{\sigma_y^* + \sigma_z^*}{\varepsilon_0} \frac{H_x^{t+\Delta t/2} + H_x^{t-\Delta t/2}}{2} + \frac{\sigma_y^* \sigma_z^*}{\varepsilon_0^2} \sum_{\tau=0}^{t} H_x^\tau \Delta t = -\frac{c_0}{\mu_r} C_{Ex}^t - \frac{c_0 \sigma_x^*}{\mu_r \varepsilon_0} \sum_{\tau=0}^{t} C_{Ex}^t \Delta t \qquad (4.25)$$

The first summation in this equation requires magnetic fields at integer time steps. In the FDTD framework, they are only computed at half time steps. For this reason, the summation should be written as

$$\sum_{\tau=0}^{t} H_x^\tau \Delta t = \frac{H_x^{\tau+\Delta t/2} + H_x^{\tau-\Delta t/2}}{2} \frac{\Delta t}{2} + \sum_{\tau=\Delta t/2}^{t-\Delta t/2} H_x^\tau \Delta t \qquad (4.26)$$

Applying this to Eq. (4.25) and solving for $H_x^{t+\Delta t/2}$ leads to the update equation for this field component.

$$H_x^{t+\Delta t/2} = m_{Hx1} H_x^{t-\Delta t/2} - m_{Hx2} C_{Ex}^t - m_{Hx3} I_{CEx}^t - m_{Hx4} I_{Hx}^{t-\Delta t/2} \qquad (4.27)$$

$$m_{Hx0} = \frac{1}{\Delta t} + \frac{\sigma_y^* + \sigma_z^*}{2\varepsilon_0} + \frac{\sigma_y^* \sigma_z^* \Delta t}{4\varepsilon_0^2} \qquad m_{Hx1} = \frac{1}{m_{Hx0}}\left(\frac{1}{\Delta t} - \frac{\sigma_y^* + \sigma_z^*}{2\varepsilon_0} - \frac{\sigma_y^* \sigma_z^* \Delta t}{4\varepsilon_0^2}\right)$$

$$m_{Hx2} = \frac{c_0}{m_{Hx0} \mu_r} \qquad m_{Hx3} = \frac{c_0 \sigma_x^* \Delta t}{m_{Hx0} \mu_r \varepsilon_0} \qquad m_{Hx4} = \frac{\sigma_y^* \sigma_z^* \Delta t}{m_{Hx0} \varepsilon_0^2} \qquad (4.28)$$

$$I_{CEx}^t = \sum_{\tau=0}^{t} C_{Ex}^t \qquad I_{Hx}^{t-\Delta t/2} = \sum_{\tau=\Delta t/2}^{t-\Delta t/2} H_x^\tau \qquad (4.29)$$

Here, all of the constants have been collected into "update coefficients" m_{Hx0} through m_{Hx4} that can be computed prior to iteration. The update equations for $H_y^{t+\Delta t/2}$ and $H_z^{t+\Delta t/2}$ can be written directly by analogy and are summarized later.

4.2.5. *Material Polarization Update Equations*

To be most general, the dispersive portion of the dielectric function is written as a Lorentz model of arbitrary order M.

$$\varepsilon_r^*(\omega) = \sum_{m=1}^{M} \frac{f_m \Omega_p^2}{\omega_m^2 - \omega^2 + j\omega\Gamma_m} \qquad (4.30)$$

Using this equation, materials can be made to have gain by making the plasma frequency imaginary. From Eq. (4.8), the dispersive portion of material polarization can be written as

$$\vec{P}(\omega) = \varepsilon_r^*(\omega)\vec{E}(\omega) \qquad (4.31)$$

Substituting Eq. (4.30) into Eq. (4.31) leads to

$$\vec{P}(\omega) = \sum_{m=1}^{M} \frac{f_m \Omega_p^2}{\omega_m^2 - \omega^2 + j\omega\Gamma_m}\vec{E}(\omega) \qquad (4.32)$$

This equation can be thought of as a sum of component polarizations; one for each Lorentz oscillator. This interpretation is written as

$$\vec{P}(\omega) = \sum_{m=1}^{M} \vec{P}_m(\omega) \qquad (4.33)$$

$$\vec{P}_m(\omega) = \frac{f_m \Omega_p^2}{\omega_m^2 - \omega^2 + j\omega\Gamma_m}\vec{E}(\omega) \qquad (4.34)$$

By multiplying both sides of Eq. (4.34) by the denominator, it can be written in a form that better accommodates transformation to the time-domain.

$$\omega_m^2 \vec{P}_m(\omega) + (j\omega)^2 \vec{P}_m(\omega) + j\omega\Gamma_m \vec{P}_m(\omega) = f_m \Omega_p^2 \vec{E}(\omega) \qquad (4.35)$$

93

Converting this to the time-domain leads to

$$\frac{\partial^2 \vec{\tilde{P}}_m(t)}{\partial t^2} + \Gamma_m \frac{\partial \vec{\tilde{P}}_m(t)}{\partial t} + \omega_m^2 \vec{\tilde{P}}_m(t) = f_m \Omega_p^2 \vec{\tilde{E}}(t) \tag{4.36}$$

Polarization current associated with the mth Lorentz oscillator is defined as

$$\vec{\tilde{J}}_m(t) = \frac{\partial \vec{\tilde{P}}_m(t)}{\partial t} \tag{4.37}$$

Approximating the time derivative with a finite-difference and solving for $\vec{\tilde{P}}_m^{t+\Delta t}$ leads to the

update equation for material polarization in terms of the polarization current.

$$\vec{\tilde{P}}_m^{t+\Delta t} = \vec{\tilde{P}}_m^t + \Delta t \vec{\tilde{J}}_m^{t+\Delta t/2} \tag{4.38}$$

Expressing Eq. (4.36) in terms of $\vec{\tilde{J}}_m(t)$ yields

$$\frac{\partial \vec{\tilde{J}}_m(t)}{\partial t} + \Gamma_m \vec{\tilde{J}}_m(t) + \omega_m^2 \vec{\tilde{P}}_m(t) = f_m \Omega_p^2 \vec{\tilde{E}}(t) \tag{4.39}$$

Approximating this equation with finite-differences and solving for $\vec{\tilde{J}}_m^{t+\Delta t/2}$ leads to an update

equation for polarization current in terms of material polarization and electric field.

$$\vec{\tilde{J}}_m^{t+\Delta t/2} = \left(\frac{2 - \Gamma_m \Delta t}{2 + \Gamma_m \Delta t} \right) \vec{\tilde{J}}_m^{t-\Delta t/2} - \left(\frac{2 \omega_m^2 \Delta t}{2 + \Gamma_m \Delta t} \right) \vec{\tilde{P}}_m^t + \left(\frac{2 f_m \Omega_p^2 \Delta t}{2 + \Gamma_m \Delta t} \right) \vec{\tilde{E}}^t \tag{4.40}$$

This method of accounting for polarization is sufficiently general to incorporate other

more complex behavior into the simulation. Specifically, minor modifications to the equations

in this section would enable the model to account for gain, loss, or nonlinear behavior. Values

for the Lorentz model parameters were borrowed from Ref. [69] and are summarized in Table

4-1.

Table 4-1: Values for Lorentz model parameters

LORENTZ-DRUDE PARAMETERS (eV)											
Parameter	Ag	Au	Cu	Al	Be	Cr	Ni	Pd	Pt	Ti	W
wp	9.010	9.030	10.830	14.980	18.510	10.750	15.920	9.720	9.590	7.290	13.220
f0	0.845	0.760	0.575	0.523	0.084	0.168	0.096	0.330	0.333	0.148	0.206
G0	0.048	0.053	0.030	0.047	0.035	0.047	0.048	0.008	0.080	0.082	0.064
w0	0	0	0	0	0	0	0	0	0	0	0
f1	0.065	0.024	0.061	0.227	0.031	0.151	0.100	0.649	0.191	0.899	0.054
G1	3.886	0.241	0.378	0.333	1.664	3.175	4.511	2.950	0.517	2.276	0.530
w1	0.816	0.415	0.291	0.162	0.100	0.121	0.174	0.336	0.780	0.777	1.004
f2	0.124	0.010	0.104	0.050	0.140	0.150	0.135	0.121	0.659	0.393	0.166
G2	0.452	0.345	1.056	0.312	3.395	1.305	1.334	0.555	1.838	2.518	1.281
w2	4.481	0.830	2.957	1.544	1.032	0.543	0.582	0.501	1.314	1.545	1.917
f3	0.011	0.071	0.723	0.166	0.530	1.149	0.106	0.638	0.547	0.187	0.706
G3	0.065	0.870	3.213	1.351	4.454	2.676	2.178	4.621	3.668	1.663	3.332
w3	8.185	2.969	5.300	1.808	3.183	1.970	1.597	1.659	3.141	2.509	3.580
f4	0.840	0.601	0.638	0.030	0.130	0.825	0.729	0.453	3.576	0.001	2.590
G4	0.916	2.494	4.305	3.382	1.802	1.335	6.292	3.236	8.517	1.762	5.836
w4	9.083	4.304	11.180	3.473	4.604	8.775	6.089	5.715	9.249	19.430	7.498
f5	5.646	4.384									
G5	2.419	2.214									
w5	20.290	13.320									
Min n:	0.130	0.080	0.100	0.033	1.350	0.310	0.860	0.720	0.730	0.760	0.490

4.2.6. E-Field Update Equations

Substituting Eq. (4.33) into Eq. (4.18) yields

$$\vec{\tilde{D}} = \varepsilon_{x,\infty} \vec{\tilde{E}} + \sum_{m=1}^{M} \vec{\tilde{P}}_m (\omega) \tag{4.41}$$

In the time-domain, this is simply

$$\vec{\tilde{D}}(t) = \varepsilon_{x,\infty} \vec{\tilde{E}}(t) + \sum_{m=1}^{M} \vec{\tilde{P}}_m (t) \tag{4.42}$$

Solving this for $\vec{\tilde{E}}(t)$ and converting to discrete-time leads to the update equation for the electric field in terms of material polarization $\vec{\tilde{P}}$, displacement field $\vec{\tilde{D}}$, and background dielectric constant $\varepsilon_{x,\infty}$.

$$\vec{\tilde{E}}^{t+\Delta t} = \frac{1}{\varepsilon_{x,\infty}} \left(\vec{\tilde{D}}^{t+\Delta t} - \sum_{m=1}^{M} \vec{\tilde{P}}^{t+\Delta t} \right) \tag{4.43}$$

4.2.7.   D-Field Update Equations

Following the same procedure taken in section 4.1.4, Eq. (4.15) is written in a form that accommodates conversion to the time-domain.

$$j\omega \tilde{D}_x + \frac{\sigma_y^* + \sigma_z^*}{\varepsilon_0} \tilde{D}_x + \frac{\sigma_y^* \sigma_z^*}{\varepsilon_0^2} \frac{1}{j\omega} \tilde{D}_x = c_0 C_{Hx} + \frac{c_0 \sigma_x^*}{\varepsilon_0} \frac{1}{j\omega} C_{Hx} - \frac{\sigma_x}{\varepsilon_0} \tilde{E}_x \tag{4.44}$$

$$C_{Hx} = \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} \tag{4.45}$$

In the time-domain, this is

$$\frac{\partial \tilde{D}_x(t)}{\partial t} + \frac{\sigma_y^* + \sigma_z^*}{\varepsilon_0} \tilde{D}_x(t) + \frac{\sigma_y^* \sigma_z^*}{\varepsilon_0^2} \int_{-\infty}^{t} \tilde{D}_x(\tau) d\tau = c_0 C_{Hx}(t) + \frac{c_0 \sigma_x^*}{\varepsilon_0} \int_{-\infty}^{t} C_x^H(\tau) d\tau - \frac{\sigma_x}{\varepsilon_0} \tilde{E}_x(t) \tag{4.46}$$

Approximating the derivatives with finite-differences, approximating integrations with summations, and solving for $\tilde{D}_x^{t+\Delta t}$ leads to the update equation.

$$\tilde{D}_x^{t+\Delta t} = m_{Dx1}' \tilde{D}_x^t + m_{Dx2}' C_{Hx}^{t+\Delta t/2} + m_{Dx3}' I_{CHx}^{t+\Delta t/2} - m_{Dx4}' I_{Dx}^t - m_{Dx5}' \left(\tilde{E}_x^{t+\Delta t} + \tilde{E}_x^t \right) \tag{4.47}$$

$$m_{Dx0}' = \frac{1}{\Delta t} + \frac{\sigma_y^* + \sigma_z^*}{2\varepsilon_0} + \frac{\sigma_y^* \sigma_z^* \Delta t}{4\varepsilon_0^2} \qquad m_{Dx1}' = \frac{1}{m_{Dx0}'} \left(\frac{1}{\Delta t} - \frac{\sigma_y^* + \sigma_z^*}{2\varepsilon_0} - \frac{\sigma_y^* \sigma_z^* \Delta t}{4\varepsilon_0^2} \right)$$

$$m_{Dx2}' = \frac{c_0}{m_{Dx0}'} \qquad m_{Dx3}' = \frac{c_0 \sigma_x^* \Delta t}{m_{Dx0}' \varepsilon_0} \qquad m_{Dx4}' = \frac{\sigma_y^* \sigma_z^* \Delta t}{m_{Dx0}' \varepsilon_0^2} \qquad m_{Dx5}' = \frac{\sigma_x}{2 m_{Dx0}' \varepsilon_0} \tag{4.48}$$

96

$$I_{CHx}^{t+\Delta t/2} = \sum_{\tau=\Delta t/2}^{t+\Delta t/2} C_{Hx}^{\tau} \qquad I_{Dx}^{t} = \sum_{\tau=0}^{t} \tilde{D}_{x}^{\tau} \qquad (4.49)$$

This equation, however, requires apriori knowledge of the E-field. This is mitigated by

substituting Eq. (4.43) into Eq. (4.47) and solving for $\tilde{D}_{x}^{t+\Delta t}$.

$$\tilde{D}_{x}^{t+\Delta t} = m_{Dx1}\tilde{D}_{x}^{t} + m_{Dx2}C_{Hx}^{t+\Delta t/2} + m_{Dx3}I_{CHx}^{t+\Delta t/2} - m_{Dx4}I_{Dx}^{t} - m_{Dx5}\tilde{E}_{x}^{t} + m_{Dx6}\tilde{P}_{x}^{t+\Delta t} \qquad (4.50)$$

$$m_{Dx0} = 1 + \frac{m'_{Dx5}}{\varepsilon_{r,\infty}} \qquad m_{Dx1} = \frac{m'_{Dx1}}{m_{Dx0}} \qquad m_{Dx2} = \frac{m'_{Dx2}}{m_{Dx0}} \qquad m_{Dx3} = \frac{m'_{Dx3}}{m_{Dx0}}$$

$$m_{Dx4} = \frac{m'_{Dx4}}{m_{Dx0}} \qquad m_{Dx5} = \frac{m'_{Dx5}}{m_{Dx0}} \qquad m_{Dx6} = \frac{m'_{Dx5}}{m_{Dx0}\varepsilon_{x,\infty}} \qquad (4.51)$$

The remaining update equations can be written by analogy and will be summarized in a later

section.

4.2.8.   *Incorporation of a Source*

In the total-field/scattered-field (TF/SF) framework, update equations must be modified

at the boundary between the total-field (TF) and scattered-field (SF). Only the tangential field

update equations contain field terms from both sides of the interface so only these equations have

to be modified.

For the D-field, update equations in the TF region contain magnetic field terms from the

SF region. The source must be added to the SF terms to make them compatible with TF

quantities.

$$\tilde{D}_{x}^{t+\Delta t} = m_{Dx1}\tilde{D}_{x}^{t} + m_{Dx2}\left(\frac{H_{i,j,k}^{z} - H_{i,j-1,k}^{z}}{\Delta y} - \frac{H_{i,j,k}^{y} - H_{i,j,k-1}^{y}}{\Delta z} \right) + m_{Dx2}\frac{H_{i,j,k-1}^{y,\text{inc}}}{\Delta z} \qquad (4.52)$$

$$\tilde{D}_{y}^{t+\Delta t} = m_{Dy1}\tilde{D}_{y}^{t} + m_{Dy2}\left(\frac{H_{i,j,k}^{x} - H_{i,j,k-1}^{x}}{\Delta z} - \frac{H_{i,j,k}^{z} - H_{i-1,j,k}^{z}}{\Delta x} \right) - m_{Dy2}\frac{H_{i,j,k-1}^{x,\text{inc}}}{\Delta z} \qquad (4.53)$$

For the H-field, update equations in the SF region contain electric field terms from the TF region. The source must be subtracted from the TF terms to make them compatible with SF quantities.

$$H_x^{t+\Delta t} = m_{Hx1}H_x^t - m_{Hx2}\left(\frac{\tilde{E}_{i,j+1,k}^z - \tilde{E}_{i,j,k}^z}{\Delta y} - \frac{\tilde{E}_{i,j,k+1}^y - \tilde{E}_{i,j,k}^y}{\Delta z}\right) - m_{Hx2}\frac{\tilde{E}_{i,j,k+1}^{y,\text{inc}}}{\Delta z} \qquad (4.54)$$

$$H_y^{t+\Delta t} = m_{Hy1}H_y^t - m_{Hy2}\left(\frac{\tilde{E}_{i,j,k+1}^x - \tilde{E}_{i,j,k}^x}{\Delta z} - \frac{\tilde{E}_{i+1,j,k}^z - \tilde{E}_{i,j,k}^z}{\Delta x}\right) + m_{Hy2}\frac{\tilde{E}_{i,j,k+1}^{x,\text{inc}}}{\Delta z} \qquad (4.55)$$

In these equations, terms relating to the PML have been dropped because the source is incorporated outside the PML. To make the code more modular, the source terms written in red can be implemented separate from the main update equation.

To implement a plane wave source, polarization of the electric and magnetic field must be known. Given a polarization vector $\vec{p}$, the source electric field can be written as

$$\tilde{E}(t) = \vec{p}e^{-jk_z z} \qquad (4.56)$$

Substituting this into Eq. (4.6) and solving for $\vec{H}$ leads to an equation for polarization of the magnetic field.

$$\vec{h} = \frac{\sqrt{\mu_{r,x}\varepsilon_{r,x}}\, p_x}{\mu_{r,y}}\hat{a}_y - \frac{\sqrt{\mu_{r,y}\varepsilon_{r,y}}\, p_y}{\mu_{r,x}}\hat{a}_x \qquad (4.57)$$

4.2.9.   Fourier Transforms

It is often necessary to obtain frequency-domain quantities from time-domain data. One approach is to store the temporal response observed during simulation and Fourier transform the data after completion. For large data sets over many time iterations, this approach is inefficient. A better method is to continually perform an integration that computes the Fourier transform as the code iterates. This section will formulate this technique.

As the code iterates, a field at a particular point in the grid can be viewed as a function $f(t)$. The Fourier transform of this function is defined as

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \tag{4.58}$$

This can be approximated numerically as a summation.

$$F(\omega) \cong \sum_{k=0}^{N} f(k) e^{-j\omega k \Delta t} \Delta t \tag{4.59}$$

Angular frequency ω is related to wavelength λ_0 through $\omega = 2\pi c_0 / \lambda_0$ so Eq. (4.59) can be written as

$$F(\lambda_0) \cong \Delta t \sum_{k=0}^{N} f(t) \left(e^{-j\frac{2\pi c_0}{\lambda_0} \Delta t} \right)^k \tag{4.60}$$

It is most efficient if the complex exponential inside parentheses is computed prior to iteration. One must be computed for each wavelength of interest. In this manner, Eq. (4.60) can be implemented inside the main FDTD loop to compute steady-state quantities at any number of wavelengths.

4.2.10. Numerical Dispersion

FDTD uses the same Yee grid as FDFD so the numerical dispersion equation is quite similar, except that it contains a time resolution parameter Δt.

$$\left[\frac{2}{c\Delta t} \sin\left(\frac{\omega \Delta t}{2} \right) \right]^2 \varepsilon_r = \left[\frac{2}{\Delta x} \sin\left(\frac{k_x \Delta x}{2} \right) \right]^2 + \left[\frac{2}{\Delta y} \sin\left(\frac{k_y \Delta y}{2} \right) \right]^2 + \left[\frac{2}{\Delta z} \sin\left(\frac{k_z \Delta z}{2} \right) \right]^2 \tag{4.61}$$

Since the wave is propagating through a numerical grid, phase velocity is less than it would be in a physical system. Equation (4.61) can be written in a form that includes the numerical phase velocity.

$$\left(\frac{\omega}{v}\right)^2 \varepsilon_r = \left[\frac{2}{c\Delta t}\sin\left(\frac{\omega\Delta t}{2}\right)\right]^2 \varepsilon_r = \left[\frac{2}{\Delta x}\sin\left(\frac{k_x\Delta x}{2}\right)\right]^2 + \left[\frac{2}{\Delta y}\sin\left(\frac{k_y\Delta y}{2}\right)\right]^2 + \left[\frac{2}{\Delta z}\sin\left(\frac{k_z\Delta z}{2}\right)\right]^2 \qquad (4.62)$$

In this equation, the numerical phase velocity is assigned the parameter v to indicate it is different than the free space speed of light c due to numerical dispersion. Multiplying the dielectric function by a correction factor, this equation becomes

$$\left(\frac{\omega}{c}\right)^2 (f\varepsilon_r) = \left[\frac{2}{c\Delta t}\sin\left(\frac{\omega\Delta t}{2}\right)\right]^2 (f\varepsilon_r) = \left[\frac{2}{\Delta x}\sin\left(\frac{k_x\Delta x}{2}\right)\right]^2 + \left[\frac{2}{\Delta y}\sin\left(\frac{k_y\Delta y}{2}\right)\right]^2 + \left[\frac{2}{\Delta z}\sin\left(\frac{k_z\Delta z}{2}\right)\right]^2 \qquad (4.63)$$

Numerical dispersion can be exactly compensated in one direction when the correction factor f and time step Δt are chosen such that

$$f = \frac{1}{k_0^2 \varepsilon_r}\left\{\left[\frac{2}{\Delta x}\sin\left(\frac{k_x\Delta x}{2}\right)\right]^2 + \left[\frac{2}{\Delta y}\sin\left(\frac{k_y\Delta y}{2}\right)\right]^2 + \left[\frac{2}{\Delta z}\sin\left(\frac{k_z\Delta z}{2}\right)\right]^2\right\} \qquad (4.64)$$

$$\frac{k_0 c\Delta t}{2} = \sin\left(\frac{k_0 c\Delta t}{2}\right) \qquad (4.65)$$

4.2.11. Calculation of Photonic Band Diagrams

The FDTD algorithm can be used to calculate the band structure of photonic crystals [107-109]. It is a particularly powerful method for metallic lattices or for any difficult lattice not easily handled by other methods. FDTD is able to account for dispersion and nonlinearities because it is a time-domain method and Maxwell's equations are not handled as an eigen-value problem with frequency as the eigen-value.

For each Bloch wave vector, Maxwell's equations are iterated in the usual manner using the appropriate periodic boundary conditions. The temporal response is recorded during simulation and a Fourier transform calculated after completion. Eigen-frequencies corresponding to photonic bands are identified as sharp peaks in the spectrum. It is sometimes difficult to discriminate between bands when they are in close proximity or cross.

To implement periodic boundary conditions, the method requires field values to be complex. This doubles the amount of required storage compared to conventional FDTD simulations with real-valued fields. Given a Bloch wave vector $\vec{\beta}$, the periodic boundary conditions can be derived from the Bloch-Floquet theorem. Given any lattice vector $\vec{g}$, these are

$$\vec{\tilde{E}}(\vec{r} + \vec{g}) = \exp\left(j\vec{\beta} \bullet \vec{g}\right)\vec{\tilde{E}}(\vec{r}) \tag{4.66}$$

$$\vec{H}(\vec{r} + \vec{g}) = \exp\left(j\vec{\beta} \bullet \vec{g}\right)\vec{H}(\vec{r}) \tag{4.67}$$

This only requires the curl equations to be modified at the boundaries. For cubic symmetries, these are

$$C_{Ex}^{i,N_y,k} = \frac{\phi_y \tilde{E}_{i,1,k}^z - \tilde{E}_{i,j,k}^z}{\Delta y} - \frac{\tilde{E}_{i,j,k+1}^y - \tilde{E}_{i,j,k}^y}{\Delta z} \qquad C_{Ex}^{i,j,N_z} = \frac{\tilde{E}_{i,j+1,k}^z - \tilde{E}_{i,j,k}^z}{\Delta y} - \frac{\phi_z \tilde{E}_{i,j,1}^y - \tilde{E}_{i,j,k}^y}{\Delta z} \tag{4.68}$$

$$C_{Ey}^{i,j,N_z} = \frac{\phi_z \tilde{E}_{i,j,1}^x - \tilde{E}_{i,j,k}^x}{\Delta z} - \frac{\tilde{E}_{i+1,j,k}^z - \tilde{E}_{i,j,k}^z}{\Delta x} \qquad C_{Ey}^{N_x,j,k} = \frac{\tilde{E}_{i,j,k+1}^x - \tilde{E}_{i,j,k}^x}{\Delta z} - \frac{\phi_x \tilde{E}_{1,j,k}^z - \tilde{E}_{i,j,k}^z}{\Delta x} \tag{4.69}$$

$$C_{Ez}^{N_x,j,k} = \frac{\phi_x \tilde{E}_{1,j,k}^y - \tilde{E}_{i,j,k}^y}{\Delta x} - \frac{\tilde{E}_{i,j+1,k}^x - \tilde{E}_{i,j,k}^x}{\Delta y} \qquad C_{Ez}^{i,N_y,k} = \frac{\tilde{E}_{i+1,j,k}^y - \tilde{E}_{i,j,k}^y}{\Delta x} - \frac{\phi_y \tilde{E}_{i,1,k}^x - \tilde{E}_{i,j,k}^x}{\Delta y} \tag{4.70}$$

$$C_{Hx}^{i,1,k} = \frac{H_{i,j,k}^z - \phi_y^* H_{i,N_y,k}^z}{\Delta y} - \frac{H_{i,j,k}^y - H_{i,j,k-1}^y}{\Delta z} \qquad C_{Hx}^{i,j,1} = \frac{H_{i,j,k}^z - H_{i,j-1,k}^z}{\Delta y} - \frac{H_{i,j,k}^y - \phi_z^* H_{i,j,N_z}^y}{\Delta z} \tag{4.71}$$

$$C_{Hy}^{i,j,1} = \frac{H_{i,j,k}^x - \phi_z^* H_{i,j,N_z}^x}{\Delta z} - \frac{H_{i,j,k}^z - H_{i-1,j,k}^z}{\Delta x} \qquad C_{Hy}^{1,j,k} = \frac{H_{i,j,k}^x - H_{i,j,k-1}^x}{\Delta z} - \frac{H_{i,j,k}^z - \phi_x^* H_{N_x,j,k}^z}{\Delta x} \tag{4.72}$$

$$C_{Hz}^{1,j,k} = \frac{H_{i,j,k}^{y} - \phi_x^* H_{N_x,j,k}^{y}}{\Delta x} - \frac{H_{i,j,k}^{x} - H_{i,j-1,k}^{x}}{\Delta y} \qquad C_{Hz}^{i,1,k} = \frac{H_{i,j,k}^{y} - H_{i-1,j,k}^{y}}{\Delta x} - \frac{H_{i,j,k}^{x} - \phi_y^* H_{i,N_y,k}^{x}}{\Delta y} \qquad (4.73)$$

$$\phi_x = e^{-jk_x N_x \Delta x} \qquad\qquad \phi_x^* = e^{+jk_x N_x \Delta x} \qquad\qquad (4.74)$$

$$\phi_y = e^{-jk_y N_y \Delta y} \qquad\qquad \phi_y^* = e^{+jk_y N_y \Delta y} \qquad\qquad (4.75)$$

To ensure energy is present in all possible modes, multiple dipole sources with multiple polarizations must be implemented. Position of the sources must be away from any point of symmetry or Bragg plane to ensure the sources are not placed at the nulls of a Bloch mode. Similarly, time response should be recorded at several locations within the unit cell that are also away from any points of symmetry. An overall Fourier transform is determined by summing the Fourier transforms calculated for each recorded response.

4.3. Implementation

4.3.1. Initialization

A FDTD simulation starts by defining the structure to be modeled and identifying the range of wavelengths for which to compute a response. Based on this information a grid may be defined and material properties assigned to each point on the grid. This includes all parameters implementing the PML.

There are two considerations when choosing grid resolution. First, grid spacing must be small enough to sufficiently resolve the shortest wavelength of interest. This is typically $\lambda/10$ to $\lambda/30$, where λ is the shortest wavelength in the region of highest refractive index. Second, grid spacing must be small enough to sufficiently resolve the finest feature of the structure.

Energy can never travel farther than one grid spacing in a single time step due to the manner in which the fields are updated. For this reason, the time step should be small enough to ensure this would not have to be the case to accurately portray physical fields. This is called the Courant stability condition [44]. For dispersive materials, refractive index can be less than 1.0 and the conventional stability condition must be modified to account for this. This becomes

$$\Delta t \leq \frac{n_{min}}{c_0 \sqrt{\dfrac{1}{\Delta_x^2} + \dfrac{1}{\Delta_y^2} + \dfrac{1}{\Delta_z^2}}} \tag{4.76}$$

In practice, the time step is calculated from the minimum grid spacing Δ_{min} found anywhere in the Yee grid.

$$\Delta t \leq \frac{n_{min} \Delta_{min}}{2 c_0} \tag{4.77}$$

Another necessary part of initialization is setting all fields and integration terms to zero. If steady-state fields are to be calculated, the complex exponential kernels should also be computed prior to executing the main FDTD loop.

4.3.2.   *Compute Update Coefficients*

To improve speed, material related parameters are collected into update coefficients. These should be computed prior to executing the main FDTD loop.

H-Field Update Coefficients

$$m_{Hx0} = \frac{1}{\Delta t} + \frac{\sigma_y^* + \sigma_z^*}{2\varepsilon_0} + \frac{\sigma_y^* \sigma_z^* \Delta t}{4\varepsilon_0^2} \qquad m_{Hx1} = \frac{1}{m_{Hx0}} \left(\frac{1}{\Delta t} - \frac{\sigma_y^* + \sigma_z^*}{2\varepsilon_0} - \frac{\sigma_y^* \sigma_z^* \Delta t}{4\varepsilon_0^2} \right)$$

$$m_{Hx2} = \frac{c_0}{m_{H0} \mu_r} \qquad m_{Hx3} = \frac{c_0 \sigma_x^* \Delta t}{m_{H0} \mu_r \varepsilon_0} \qquad m_{Hx4} = \frac{\sigma_y^* \sigma_z^* \Delta t}{m_{H0} \varepsilon_0^2}$$

$$m_{Hy0} = \frac{1}{\Delta t} + \frac{\sigma_x^* + \sigma_z^*}{2\varepsilon_0} + \frac{\sigma_x^*\sigma_z^*\Delta t}{4\varepsilon_0^2} \qquad m_{Hy1} = \frac{1}{m_{Hy0}}\left(\frac{1}{\Delta t} - \frac{\sigma_x^* + \sigma_z^*}{2\varepsilon_0} - \frac{\sigma_x^*\sigma_z^*\Delta t}{4\varepsilon_0^2}\right)$$

$$m_{Hy2} = \frac{c_0}{m_{Hy0}\mu_r} \qquad m_{Hy3} = \frac{c_0\sigma_y^*\Delta t}{m_{Hy0}\mu_r\varepsilon_0} \qquad m_{Hy4} = \frac{\sigma_x^*\sigma_z^*\Delta t}{m_{Hy0}\varepsilon_0^2}$$

$$m_{Hz0} = \frac{1}{\Delta t} + \frac{\sigma_x^* + \sigma_y^*}{2\varepsilon_0} + \frac{\sigma_x^*\sigma_y^*\Delta t}{4\varepsilon_0^2} \qquad m_{Hz1} = \frac{1}{m_{Hz0}}\left(\frac{1}{\Delta t} - \frac{\sigma_x^* + \sigma_y^*}{2\varepsilon_0} - \frac{\sigma_x^*\sigma_y^*\Delta t}{4\varepsilon_0^2}\right)$$

$$m_{Hz2} = \frac{c_0}{m_{Hz0}\mu_r} \qquad m_{Hz3} = \frac{c_0\sigma_z^*\Delta t}{m_{Hz0}\mu_r\varepsilon_0} \qquad m_{Hz4} = \frac{\sigma_x^*\sigma_y^*\Delta t}{m_{Hz0}\varepsilon_0^2}$$

D-Field Update Coefficients

$$m_{Dx0}' = \frac{1}{\Delta t} + \frac{\sigma_y^* + \sigma_z^*}{2\varepsilon_0} + \frac{\sigma_y^*\sigma_z^*\Delta t}{4\varepsilon_0^2} \qquad m_{Dx1}' = \frac{1}{m_{Dx0}'}\left(\frac{1}{\Delta t} - \frac{\sigma_y^* + \sigma_z^*}{2\varepsilon_0} - \frac{\sigma_y^*\sigma_z^*\Delta t}{4\varepsilon_0^2}\right)$$

$$m_{Dx2}' = \frac{c_0}{m_{Dx0}'} \qquad m_{Dx3}' = \frac{c_0\sigma_x^*\Delta t}{m_{Dx0}'\varepsilon_0} \qquad m_{Dx4}' = \frac{\sigma_y^*\sigma_z^*\Delta t}{m_{Dx0}'\varepsilon_0^2} \qquad m_{Dx5}' = \frac{\sigma_x}{m_{Dx0}'\varepsilon_0}$$

$$m_{Dx0} = 1 + \frac{m_{Dx5}'}{\varepsilon_{r,\infty}} \qquad m_{Dx1} = \frac{m_{Dx1}'}{m_{Dx0}} \qquad m_{Dx2} = \frac{m_{Dx2}'}{m_{Dx0}} \qquad m_{Dx3} = \frac{m_{Dx3}'}{m_{Dx0}}$$

$$m_{Dx4} = \frac{m_{Dx4}'}{m_{Dx0}} \qquad m_{Dx5} = \frac{m_{Dx5}'}{m_{Dx0}} \qquad m_{Dx6} = \frac{m_{Dx5}'}{m_{Dx0}\varepsilon_{r,\infty}}$$

$$m_{Dy0}' = \frac{1}{\Delta t} + \frac{\sigma_x^* + \sigma_z^*}{2\varepsilon_0} + \frac{\sigma_x^*\sigma_z^*\Delta t}{4\varepsilon_0^2} \qquad m_{Dy1}' = \frac{1}{m_{Dy0}'}\left(\frac{1}{\Delta t} - \frac{\sigma_x^* + \sigma_z^*}{2\varepsilon_0} - \frac{\sigma_x^*\sigma_z^*\Delta t}{4\varepsilon_0^2}\right)$$

$$m_{Dy2}' = \frac{c_0}{m_{Dy0}'} \qquad m_{Dy3}' = \frac{c_0\sigma_y^*\Delta t}{m_{Dy0}'\varepsilon_0} \qquad m_{Dy4}' = \frac{\sigma_x^*\sigma_z^*\Delta t}{m_{Dy0}'\varepsilon_0^2} \qquad m_{Dy5}' = \frac{\sigma_y}{m_{Dy0}'\varepsilon_0}$$

$$m_{Dy0} = 1 + \frac{m_{Dy5}'}{\varepsilon_{r,\infty}} \qquad m_{Dy1} = \frac{m_{Dy1}'}{m_{Dy0}} \qquad m_{Dy2} = \frac{m_{Dy2}'}{m_{Dy0}} \qquad m_{Dy3} = \frac{m_{Dy3}'}{m_{Dy0}}$$

$$m_{Dy4} = \frac{m_{Dy4}'}{m_{Dy0}} \qquad m_{Dx5} = \frac{m_{Dy5}'}{m_{Dy0}} \qquad m_{Dx6} = \frac{m_{Dy5}'}{m_{Dy0}\varepsilon_{r,\infty}}$$

$$m'_{Dz0} = \frac{1}{\Delta t} + \frac{\sigma_x^* + \sigma_y^*}{2\varepsilon_0} + \frac{\sigma_x^* \sigma_y^* \Delta t}{4\varepsilon_0^2} \qquad m'_{Dz1} = \frac{1}{m'_{Dz0}}\left(\frac{1}{\Delta t} - \frac{\sigma_x^* + \sigma_y^*}{2\varepsilon_0} - \frac{\sigma_x^* \sigma_y^* \Delta t}{4\varepsilon_0^2} \right)$$

$$m'_{Dz2} = \frac{c_0}{m'_{Dz0}} \qquad m'_{Dz3} = \frac{c_0 \sigma_z^* \Delta t}{m'_{Dz0}\varepsilon_0} \qquad m'_{Dy4} = \frac{\sigma_x^* \sigma_y^* \Delta t}{m'_{Dz0}\varepsilon_0^2} \qquad m'_{Dz5} = \frac{\sigma_z}{m'_{Dz0}\varepsilon_0}$$

$$m_{Dz0} = 1 + \frac{m'_{Dz5}}{\varepsilon_{r,\infty}} \qquad m_{Dz1} = \frac{m'_{Dz1}}{m_{Dz0}} \qquad m_{Dz2} = \frac{m'_{Dz2}}{m_{Dz0}} \qquad m_{Dz3} = \frac{m'_{Dz3}}{m_{Dz0}}$$

$$m_{Dz4} = \frac{m'_{Dz4}}{m_{Dz0}} \qquad m_{Dz5} = \frac{m'_{Dz5}}{m_{Dz0}} \qquad m_{Dx6} = \frac{m'_{Dz5}}{m_{Dz0}\varepsilon_{r,\infty}}$$

4.3.3.   *Main FDTD Loop*

In this chapter, many equations were derived to update the fields and material polarization. For a stable and robust code, the order in which the quantities are updated is very important. The magnetic field $\vec{H}$ and polarization current $\vec{J}$ are computed at the half time steps. The electric flux density $\vec{\tilde{D}}$, electric field intensity $\vec{\tilde{E}}$, and material polarization $\vec{\tilde{P}}$ are computed at integer time steps. The following text summarizes the complete set of update equations and the order in which all quantities should be computed.

Update H-Field

$$C_{Ex}^{i,j,k} = \frac{\tilde{E}_{i,j+1,k}^z - \tilde{E}_{i,j,k}^z}{\Delta y} - \frac{\tilde{E}_{i,j,k+1}^y - \tilde{E}_{i,j,k}^y}{\Delta z}$$

$$C_{Ey}^{i,j,k} = \frac{\tilde{E}_{i,j,k+1}^x - \tilde{E}_{i,j,k}^x}{\Delta z} - \frac{\tilde{E}_{i+1,j,k}^z - \tilde{E}_{i,j,k}^z}{\Delta x}$$

$$C_{Ez}^{i,j,k} = \frac{\tilde{E}_{i+1,j,k}^y - \tilde{E}_{i,j,k}^y}{\Delta x} - \frac{\tilde{E}_{i,j+1,k}^x - \tilde{E}_{i,j,k}^x}{\Delta y}$$

$$I_{CEx}^t = \sum_{\tau=0}^{t} C_{Ex}^{\tau} \qquad I_{CEy}^t = \sum_{\tau=0}^{t} C_{Ey}^{\tau} \qquad I_{CEz}^t = \sum_{\tau=0}^{t} C_{Ez}^{\tau}$$

$$H_x^{t+\Delta t/2} = m_{Hx1} H_x^{t-\Delta t/2} - m_{Hx2} C_{Ex}^t - m_{Hx3} I_{CEx}^t - m_{Hx4} I_{Hx}^{t-\Delta t/2}$$

$$H_y^{t+\Delta t/2} = m_{Hy1} H_y^{t-\Delta t/2} - m_{Hy2} C_{Ey}^t - m_{Hy3} I_{CEy}^t - m_{Hy4} I_{Hy}^{t-\Delta t/2}$$

$$H_z^{t+\Delta t/2} = m_{Hz1} H_z^{t-\Delta t/2} - m_{Hz2} C_{Ez}^t - m_{Hz3} I_{CEz}^t - m_{Hz4} I_{Hz}^{t-\Delta t/2}$$

Inject H-Field Source

$$H_x^{t+\Delta t} = H_x^{t+\Delta t} - m_{Hx2} \frac{\tilde{E}_{i,j,k+1}^{y,\text{inc}}}{\Delta z}$$

$$H_y^{t+\Delta t} = H_y^{t+\Delta t} + m_{Hy2} \frac{\tilde{E}_{i,j,k+1}^{x,\text{inc}}}{\Delta z}$$

Update Polarization

$$\tilde{J}_{x,m}^{t+\Delta t/2} = \left(\frac{2 - \Gamma_{x,m}\Delta t}{2 + \Gamma_{x,m}\Delta t} \right) \tilde{J}_{x,m}^{t-\Delta t/2} - \left(\frac{2\omega_{x,m}^2 \Delta t}{2 + \Gamma_{x,m}\Delta t} \right) \tilde{P}_{x,m}^t + \left(\frac{2 f_{x,m} \Omega_{x,p}^2 \Delta t}{2 + \Gamma_{x,m}\Delta t} \right) \tilde{E}_{x,m}^t$$

$$\tilde{J}_{y,m}^{t+\Delta t/2} = \left(\frac{2 - \Gamma_{y,m}\Delta t}{2 + \Gamma_{y,m}\Delta t} \right) \tilde{J}_{y,m}^{t-\Delta t/2} - \left(\frac{2\omega_{y,m}^2 \Delta t}{2 + \Gamma_{y,m}\Delta t} \right) \tilde{P}_{y,m}^t + \left(\frac{2 f_{y,m} \Omega_{y,p}^2 \Delta t}{2 + \Gamma_{y,m}\Delta t} \right) \tilde{E}_{y,m}^t$$

$$\tilde{J}_{z,m}^{t+\Delta t/2} = \left(\frac{2 - \Gamma_{z,m}\Delta t}{2 + \Gamma_{z,m}\Delta t} \right) \tilde{J}_{z,m}^{t-\Delta t/2} - \left(\frac{2\omega_{z,m}^2 \Delta t}{2 + \Gamma_{z,m}\Delta t} \right) \tilde{P}_{z,m}^t + \left(\frac{2 f_{z,m} \Omega_{z,p}^2 \Delta t}{2 + \Gamma_{z,m}\Delta t} \right) \tilde{E}_{z,m}^t$$

$$\tilde{P}_{x,m}^{t+\Delta t} = \tilde{P}_{x,m}^t + \Delta t \tilde{J}_{x,m}^{t+\Delta t/2}$$

$$\tilde{P}_{y,m}^{t+\Delta t} = \tilde{P}_{y,m}^t + \Delta t \tilde{J}_{y,m}^{t+\Delta t/2}$$

$$\tilde{P}_{z,m}^{t+\Delta t} = \tilde{P}_{z,m}^{t} + \Delta t \tilde{J}_{z,m}^{t+\Delta t/2}$$

Update D-Field

$$C_{Hx}^{i,j,k} = \frac{H_{i,j,k}^{z} - H_{i,j-1,k}^{z}}{\Delta y} - \frac{H_{i,j,k}^{y} - H_{i,j,k-1}^{y}}{\Delta z}$$

$$C_{Hy}^{i,j,k} = \frac{H_{i,j,k}^{x} - H_{i,j,k-1}^{x}}{\Delta z} - \frac{H_{i,j,k}^{z} - H_{i-1,j,k}^{z}}{\Delta x}$$

$$C_{Hz}^{i,j,k} = \frac{H_{i,j,k}^{y} - H_{i-1,j,k}^{y}}{\Delta x} - \frac{H_{i,j,k}^{x} - H_{i,j-1,k}^{x}}{\Delta y}$$

$$I_{Dx}^{t} = \sum_{\tau=0}^{t} \tilde{D}_{x}^{\tau} \qquad I_{Hy}^{t} = \sum_{\tau=0}^{t} \tilde{D}_{y}^{\tau} \qquad I_{Hz}^{t} = \sum_{\tau=0}^{t} \tilde{D}_{z}^{\tau}$$

$$I_{CHx}^{t+\Delta t/2} = \sum_{\tau=\Delta t/2}^{t+\Delta t/2} C_{Hx}^{\tau} \qquad I_{CHy}^{t+\Delta t/2} = \sum_{\tau=\Delta t/2}^{t+\Delta t/2} C_{Hy}^{\tau} \qquad I_{CHz}^{t+\Delta t/2} = \sum_{\tau=\Delta t/2}^{t+\Delta t/2} C_{Hz}^{\tau}$$

$$\tilde{D}_{x}^{t+\Delta t} = m_{Dx1}\tilde{D}_{x}^{t} + m_{Dx2}C_{Hx}^{t+\Delta t/2} + m_{Dx3}I_{CHx}^{t+\Delta t/2} - m_{Dx4}I_{Dx}^{t} - m_{Dx5}\tilde{E}_{x}^{t} + m_{Dx6}\tilde{P}_{x}^{t+\Delta t}$$

$$\tilde{D}_{y}^{t+\Delta t} = m_{Dy1}\tilde{D}_{y}^{t} + m_{Dy2}C_{Hy}^{t+\Delta t/2} + m_{Dy3}I_{CHy}^{t+\Delta t/2} - m_{Dy4}I_{Dy}^{t} - m_{Dy5}\tilde{E}_{y}^{t} + m_{Dy6}\tilde{P}_{y}^{t+\Delta t}$$

$$\tilde{D}_{z}^{t+\Delta t} = m_{Dz1}\tilde{D}_{z}^{t} + m_{Dz2}C_{Hz}^{t+\Delta t/2} + m_{Dz3}I_{CHz}^{t+\Delta t/2} - m_{Dz4}I_{Dz}^{t} - m_{Dz5}\tilde{E}_{z}^{t} + m_{Dz6}\tilde{P}_{z}^{t+\Delta t}$$

Inject D-Field Source

$$\tilde{D}_{x}^{t+\Delta t} = \tilde{D}_{x}^{t+\Delta t} + m_{Dx2}\frac{H_{i,j,k-1}^{y,\text{inc}}}{\Delta z}$$

$$\tilde{D}_{y}^{t+\Delta t} = \tilde{D}_{y}^{t+\Delta t} - m_{Dy2}\frac{H_{i,j,k-1}^{x,\text{inc}}}{\Delta z}$$

Update E-Field

$$\tilde{E}_x^{t+\Delta t} = \frac{1}{\varepsilon_{x,\infty}} \left(\tilde{D}_x^{t+\Delta t} - \sum_{m=1}^{M} \tilde{P}_{x,m}^{t+\Delta t} \right)$$

$$\tilde{E}_y^{t+\Delta t} = \frac{1}{\varepsilon_{y,\infty}} \left(\tilde{D}_y^{t+\Delta t} - \sum_{m=1}^{M} \tilde{P}_{y,m}^{t+\Delta t} \right)$$

$$\tilde{E}_z^{t+\Delta t} = \frac{1}{\varepsilon_{z,\infty}} \left(\tilde{D}_z^{t+\Delta t} - \sum_{m=1}^{M} \tilde{P}_{z,m}^{t+\Delta t} \right)$$

Record Output / Compute Fourier Transforms

$$F(\lambda_0) \cong \Delta t \sum_{k=0}^{N} f(t) \left(e^{-j\frac{2\pi c_0}{\lambda_0}\Delta t} \right)^k$$

4.3.4.   *Causes of Error*

Causes of error in FDTD are the same as that in FDFD with one addition. In FDTD, the time step contributes to grid dispersion, rounding errors, and inaccuracy of the finite-difference approximation. Errors in FDTD accumulate as simulation progresses so it is important to use as few time steps as possible.

CHAPTER 5

PLANE WAVE EXPANSION METHOD

5.1. Overview

The plane wave expansion method (PWEM) [115-120] may be the most popular method for calculating dispersion diagrams of periodic structures like photonic crystals. PWEM is best suited to analyze periodic structures with low to moderate index contrast where size of the unit cell is less than a wavelength. Using the supercell method or absorbing boundary conditions [119], the method can be used to compute eigen-modes of waveguides, resonators, and structures of finite size. It is stable, fully vectorial, and rigorous in the sense that no approximations are made to Maxwell's equations other than truncating the number of spatial harmonics representing the fields. It is a simple method to implement and enables rigorous analysis of 3D structures on standard desktop computers. PWEM can efficiently analyze all 14 Bravais lattice symmetries and all wave angles through the lattice. It solves the wave equation as an eigen-value problem so it is able to determine and distinguish all photonic bands.

The method exploits periodicity by operating in reciprocal space where structures can often be described with considerably fewer terms. Field patterns are expanded into a finite set of plane waves as illustrated in Figure 5-1. The wave equation can then be written as an eigen-value problem where the eigen-values represent frequency and the eigen-vectors contain amplitude coefficients of the component plane waves. When a photonic band structure is being constructed, only the eigen-values have to be calculated and highly efficient algorithms like conjugate-gradient minimization can be used [121].
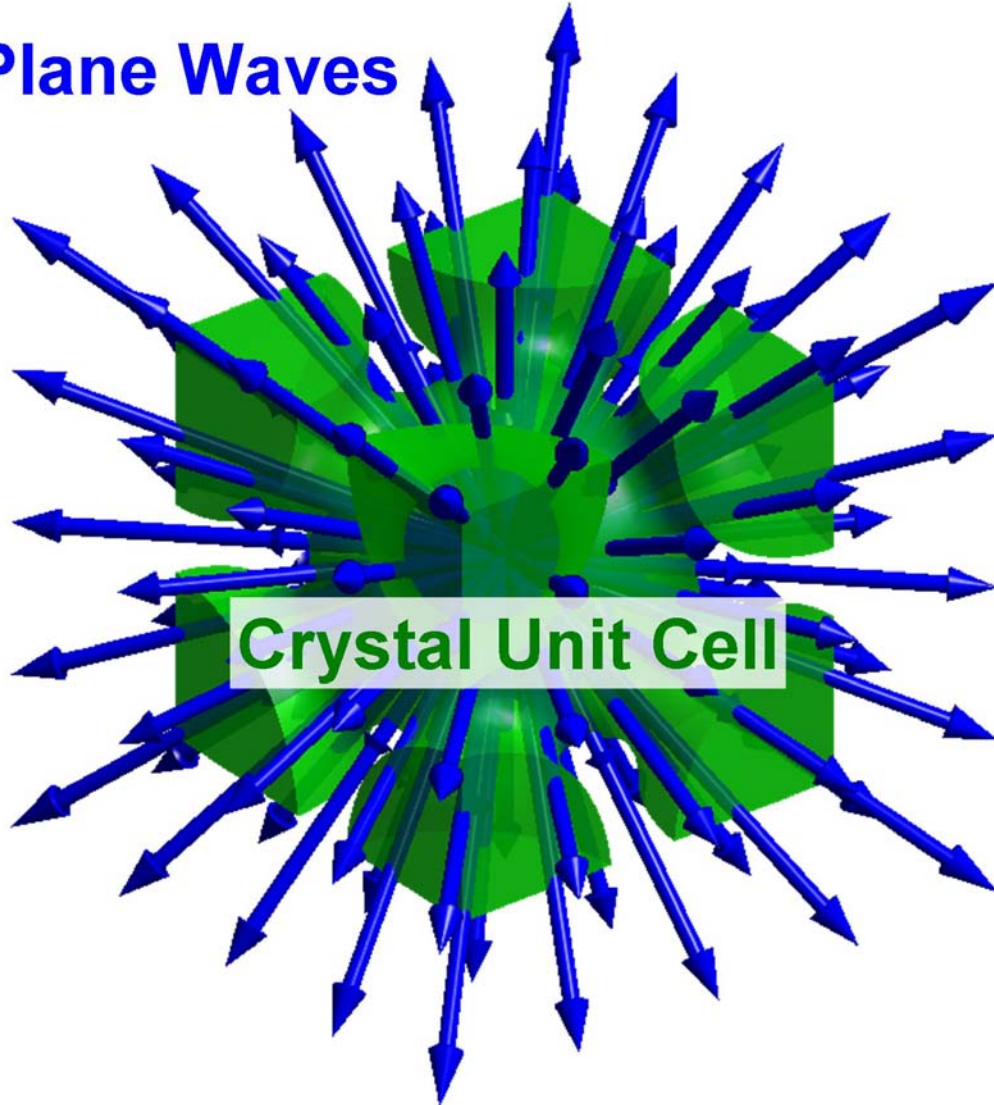
Figure 5-1: Concept of plane wave expansion method

The method is most efficiently applied to smoothly varying structures with low to moderate index contrast. It is very inefficient for modeling metallic structures. For high index contrast and metals, convergence is slow due to Gibb's phenomenon, but can be improved somewhat by using a "smoothed" dielectric tensor [118] and/or preconditioning [118, 121, 122]. Depending how the method is formulated, a singularity may exist at β=0. This is easily

overcome by obtaining an analytical solution at this point in *k*-space. PWEM cannot directly

incorporate material dispersion because frequency is the eigen-value being solved.

5.2. Maxwell's Equations in Reciprocal Space

5.2.1. *Plane Wave Expansion*

Consistent with the name of the method, the fields are expanded into a set of plane

waves, or spatial harmonics, according to the following equations.

$$\vec{E}(\vec{r}) = \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} \hat{e}_{p,q,r} e^{j\vec{k}_{p,q,r} \cdot \vec{r}} \tag{5.1}$$

$$\vec{H}(\vec{r}) = \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} \hat{h}_{p,q,r} e^{j\vec{k}_{p,q,r} \cdot \vec{r}} \tag{5.2}$$

The wave vector corresponding to the *(p,q,r)* plane wave is written in terms of the Bloch wave

vector $\vec{\beta}$ and an integer combination of reciprocal lattice vectors $\vec{T}_1$, $\vec{T}_2$, and $\vec{T}_3$.

$$\vec{k}_{p,q,r} = \vec{\beta} - \vec{G}_{p,q,r} \tag{5.3}$$

$$\vec{G}_{p,q,r} = p\vec{T}_1 + q\vec{T}_2 + r\vec{T}_3 \tag{5.4}$$

The field expressions can be written in a more compact notation as

$$\vec{E}(\vec{r}) = \sum_{\vec{G}} \hat{e}_{\vec{G}} e^{j\vec{k}_{\vec{G}} \cdot \vec{r}} \tag{5.5}$$

$$\vec{H}(\vec{r}) = \sum_{\vec{G}} \hat{h}_{\vec{G}} e^{j\vec{k}_{\vec{G}} \cdot \vec{r}} \tag{5.6}$$

Analysis in three dimensions requires each plane wave to be expressed as the sum of two

orthogonal polarizations. Both polarizations must be orthogonal to direction of the plane wave

as well as to each other. Valid unit polarizations must satisfy the following equations.

$$\hat{p}_{1,p,q,r} \bullet \hat{p}_{2,p,q,r} = 0 \tag{5.7}$$

$$\hat{p}_{1,p,q,r} \times \hat{p}_{2,p,q,r} = \frac{\vec{k}_{p,q,r}}{\left|\vec{k}_{p,q,r}\right|} \tag{5.8}$$

Like the field quantities, permittivity and permeability are expanded in terms of the reciprocal lattice vectors.

$$\varepsilon_r(\vec{r}) = \sum_{\vec{G}} a_{\vec{G}} e^{j\vec{G}\bullet\vec{r}} \qquad a_{\vec{G}} = \frac{1}{V} \iiint_V \varepsilon_r(\vec{r}) e^{-j\vec{G}\bullet\vec{r}} dv \tag{5.9}$$

$$\varepsilon_r^{-1}(\vec{r}) = \sum_{\vec{G}} c_{\vec{G}} e^{j\vec{G}\bullet\vec{r}} \qquad c_{\vec{G}} = \frac{1}{V} \iiint_V \varepsilon_r^{-1}(\vec{r}) e^{-j\vec{G}\bullet\vec{r}} dv \tag{5.10}$$

$$\mu_r(\vec{r}) = \sum_{\vec{G}} b_{\vec{G}} e^{j\vec{G}\bullet\vec{r}} \qquad b_{\vec{G}} = \frac{1}{V} \iiint_V \mu_r(\vec{r}) e^{-j\vec{G}\bullet\vec{r}} dv \tag{5.11}$$

$$\mu_r^{-1}(\vec{r}) = \sum_{\vec{G}} d_{\vec{G}} e^{j\vec{G}\bullet\vec{r}} \qquad d_{\vec{G}} = \frac{1}{V} \iiint_V \mu_r^{-1}(\vec{r}) e^{-j\vec{G}\bullet\vec{r}} dv \tag{5.12}$$

5.2.2.   Substitution into Maxwell's Curl Equations

In source-free media, Maxwell's curl equations in the frequency-domain are

$$\nabla \times \vec{E} = -j\omega\mu_0\mu_r\vec{H} \tag{5.13}$$

$$\nabla \times \vec{H} = j\omega\varepsilon_0\varepsilon_r\vec{E} \tag{5.14}$$

Substituting the expansions defined in section 5.1.1 into the curl equations leads to

$$\nabla \times \sum_{\vec{G}} \hat{e}_{\vec{G}} e^{j\vec{k}_{\vec{G}}\bullet\vec{r}} = -j\omega\mu_0 \left[\sum_{\vec{G}} b_{\vec{G}} e^{j\vec{G}\bullet\vec{r}} \right] \left[\sum_{\vec{G}} \hat{h}_{\vec{G}} e^{j\vec{k}_{\vec{G}}\bullet\vec{r}} \right] \tag{5.15}$$

$$\nabla \times \sum_{\vec{G}} \hat{h}_{\vec{G}} e^{j\vec{k}_{\vec{G}}\bullet\vec{r}} = j\omega\varepsilon_0 \left[\sum_{\vec{G}} a_{\vec{G}} e^{j\vec{G}\bullet\vec{r}} \right] \left[\sum_{\vec{G}} \hat{e}_{\vec{G}} e^{j\vec{k}_{\vec{G}}\bullet\vec{r}} \right] \tag{5.16}$$

After some algebra, these equations can be written as

$$\vec{k}_{\vec{G}} \times \hat{e}_{\vec{G}} = \omega\mu_0 \sum_{\vec{G}'} b_{\vec{G}-\vec{G}'} e^{j(\vec{G}-\vec{G}')\bullet\vec{r}} \hat{h}_{\vec{G}'} \tag{5.17}$$

$$\vec{k}_{\vec{G}} \times \hat{h}_{\vec{G}} = -\omega\varepsilon_0 \sum_{\vec{G}'} a_{\vec{G}-\vec{G}'} e^{j(\vec{G}-\vec{G}')\cdot\vec{r}} \hat{e}_{\vec{G}'} \tag{5.18}$$

The summations on the right side of these equations are discrete convolutions. Spatial

multiplication with permittivity/permeability in real space has become convolution in reciprocal

space. The final form of Maxwell's curl equations in reciprocal space is then

$$\vec{k} \times \hat{e} = \omega\mu_0\mu_r * \hat{h} \tag{5.19}$$

$$\vec{k} \times \hat{h} = -\omega\varepsilon_0\varepsilon_r * \hat{e} \tag{5.20}$$

5.2.3. *Numerical Representation*

To implement PWEM, the set of plane waves must be finite. When indices of the

harmonics are $-P \le p \le P$, $-Q \le q \le Q$, and $-R \le r \le R$, the number of harmonics retained

along $\vec{T_1}$ is 2P+1, along $\vec{T_2}$ is 2Q+1, and along $\vec{T_3}$ is 2R+1. The total number of harmonics is

$N = (2P+1)(2Q+1)(2R+1)$ and Eqs. (5.19)-(5.20) represent two sets of 3N equations that can

be written in matrix form as

$$\vec{\mathbf{K}} \times \vec{\mathbf{E}} = \omega\mu_0\boldsymbol{\mu}_r\vec{\mathbf{H}} \tag{5.21}$$

$$\vec{\mathbf{K}} \times \vec{\mathbf{H}} = -\omega\varepsilon_0\boldsymbol{\varepsilon}_r\vec{\mathbf{E}} \tag{5.22}$$

where

$$\vec{\mathbf{K}} = \vec{\beta} - \vec{\mathbf{G}} \qquad \vec{\mathbf{G}} = \begin{bmatrix} \mathbf{G}_x \\ \mathbf{G}_y \\ \mathbf{G}_z \end{bmatrix} \tag{5.23}$$

$\vec{\mathbf{G}}$ is a column vector of three square diagonal matrices $\mathbf{G}_x$, $\mathbf{G}_y$, and $\mathbf{G}_z$ containing the vector

components of $\vec{\mathbf{G}}$ over the entire expansion. $\boldsymbol{\mu}_r$ and $\boldsymbol{\varepsilon}_r$ are matrix operators representing the

113

convolutions in Eqs. (5.19)-(5.20) as defined in Eqs. (5.17)-(5.18). $\vec{\mathbf{E}}$ and $\vec{\mathbf{H}}$ are column vectors containing amplitude coefficients of the component plane waves.

5.2.4.   Truncation of Plane Wave Expansion

There is greater flexibility in which terms are retained in the expansion than suggested above. A region of *k*-space is defined by indices of the spatial harmonics retained in the expansion. Using the conventional approach this region is orthorhombic, or cubic when *P=Q=R,* as depicted at the top of Figure 5-2. The number of harmonics retained in any direction determines the resolution of structures with contrast in that direction. Conventional formulation of PWEM favors spatial resolution in directions diagonal to the crystal axes. A more efficient approach would favor all directions equally, favor directions with higher index contrast, or be based on other special knowledge of the device being modeled.

A simple and powerful truncation scheme [123, 124] is given by

$$\left|\frac{p}{P}\right|^{2\gamma}+\left|\frac{q}{Q}\right|^{2\gamma}+\left|\frac{r}{R}\right|^{2\gamma}\leq 1 \qquad (5.24)$$

The power parameter γ is any number greater than zero. Figure 5-2 shows the range of truncation schemes that are possible through proper selection of γ. Very small values only retain harmonics along the reciprocal lattice axes. Powers less than 0.5 lead to "pincushion" truncation that heavily favors on-axis harmonics, but includes some off axis. Diamond truncation is achieved for $\gamma=0.5$. Powers greater than 1.0 leads to barrel truncation that contains many more of the off axis harmonics, neglecting harmonics mostly from the extreme corners of the conventional cubic volume. Very large values approach the conventional cubic truncation scheme. It is also possible to assign different power exponents to each term. As a starting point, spherical truncation seems preferable because it leads to equal spatial resolution in all directions.
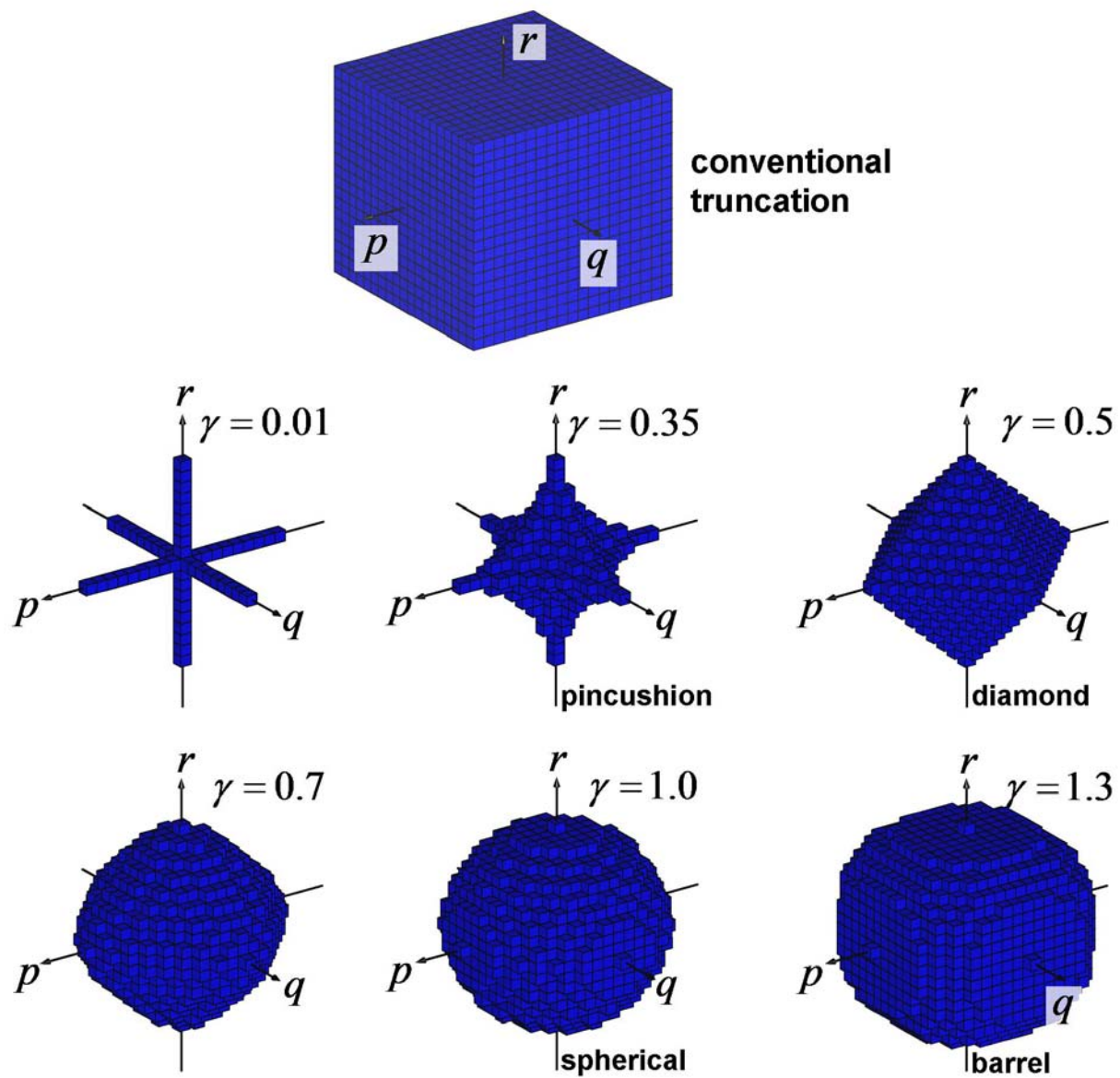
Figure 5-2: Truncation schemes

5.3. Formulation

In this section, a unique formulation of the PWEM is presented that offers many important benefits. It is more compact and intuitive than formulations with summations in their final form. By projecting the fields onto two orthogonal polarization vectors, matrix size is reduced by more than half and solutions can be found for any Bloch wave vector except $\vec{\beta} = 0$. Formulations like that found in Ref. [125] use Maxwell's divergence equations to reduce matrix size, but solutions cannot be found for Bloch wave vectors parallel to one crystal plane. The formulation more directly accommodates tensor material properties for incorporating a PML or analyzing structures with anisotropic materials. The formulation is more straightforward to implement truncation schemes as described previously. Finally, the formulation leads to more compact computer code that is easier to read, write, and troubleshoot.

5.3.1.   *Useful Vector Identities*

In this formulation, spatial vectors, numerical column vectors, and matrices are mixed and require special identities to simplify equations. Useful identities are

$$\vec{\mathbf{A}} \bullet \vec{\mathbf{B}} \neq \vec{\mathbf{B}} \bullet \vec{\mathbf{A}} \quad \text{unless } \mathbf{A}_x, \mathbf{A}_y, \mathbf{A}_z, \mathbf{B}_x, \mathbf{B}_y, \text{ and } \mathbf{B}_z \text{ are all diagonal} \tag{5.25}$$

$$\vec{\mathbf{A}} \times \vec{\mathbf{B}} \neq -\vec{\mathbf{B}} \times \vec{\mathbf{A}} \quad \text{unless } \mathbf{A}_x, \mathbf{A}_y, \mathbf{A}_z, \mathbf{B}_x, \mathbf{B}_y, \text{ and } \mathbf{B}_z \text{ are all diagonal} \tag{5.26}$$

$$\vec{\mathbf{A}} \bullet \left(\vec{\mathbf{B}} \times \vec{\mathbf{C}} \right) = \left(\vec{\mathbf{A}} \times \vec{\mathbf{B}} \right) \bullet \vec{\mathbf{C}} \tag{5.27}$$

where

$$\vec{\mathbf{A}} = \begin{bmatrix} \mathbf{A}_x \\ \mathbf{A}_y \\ \mathbf{A}_z \end{bmatrix} \qquad \vec{\mathbf{B}} = \begin{bmatrix} \mathbf{B}_x \\ \mathbf{B}_y \\ \mathbf{B}_z \end{bmatrix} \qquad \vec{\mathbf{C}} = \begin{bmatrix} \mathbf{C}_x \\ \mathbf{C}_y \\ \mathbf{C}_z \end{bmatrix} \tag{5.28}$$

The matrices $\mathbf{A}_x$, $\mathbf{A}_y$, $\mathbf{A}_z$, $\mathbf{B}_x$, $\mathbf{B}_y$, $\mathbf{B}_z$, $\mathbf{C}_x$, $\mathbf{C}_y$, and $\mathbf{C}_z$ are square matrices with individual vector components along the diagonal. The matrices $\vec{\mathbf{A}}$, $\vec{\mathbf{B}}$, and $\vec{\mathbf{C}}$ are "column vectors" containing the vector component matrices. It is technically incorrect to call these column vectors because they can contain more than one column. They are "column vectors" with square matrix elements.

Three types of multiplication are possible. These are multiplication with a matrix of scalars, a dot product, and a cross product.

$$\vec{\mathbf{A}}\cdot = \begin{bmatrix} \mathbf{A}_x & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_y & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_z \end{bmatrix} \tag{5.29}$$

$$\vec{\mathbf{A}}\bullet = \begin{bmatrix} \mathbf{A}_x & \mathbf{A}_y & \mathbf{A}_z \end{bmatrix} \tag{5.30}$$

$$\vec{\mathbf{A}}\times = \begin{bmatrix} \mathbf{0} & -\mathbf{A}_z & \mathbf{A}_y \\ \mathbf{A}_z & \mathbf{0} & -\mathbf{A}_x \\ -\mathbf{A}_y & \mathbf{A}_x & \mathbf{0} \end{bmatrix} \tag{5.31}$$

It is possible to project vector $\vec{\mathbf{A}}$ onto vector $\vec{\mathbf{B}}$ using the following equation.

$$\left\langle \vec{\mathbf{A}} \right\rangle_{\vec{\mathbf{B}}} = \left| \vec{\mathbf{B}} \right|^{-2} \left(\vec{\mathbf{B}} \bullet \vec{\mathbf{A}} \right) \vec{\mathbf{B}} \tag{5.32}$$

5.3.2. Three-Dimensional Formulation

Solving Eq. (5.22) for the electric field yields

$$\vec{\mathbf{E}} = -\left(1/\omega\varepsilon_0 \right) \varepsilon_r^{-1} \vec{\mathbf{K}} \times \vec{\mathbf{H}} \tag{5.33}$$

Substituting this into Eq. (5.21) leads to the wave equation in reciprocal space.

$$\vec{\mathbf{K}} \times \varepsilon_r^{-1} \vec{\mathbf{K}} \times \vec{\mathbf{H}} = -k_0^2 \mathbf{\mu}_r \vec{\mathbf{H}} \tag{5.34}$$

117

The magnetic field should be written as the sum of two orthogonal polarizations.

$$\vec{H} = \hat{\mathbf{p}}_1 \cdot \mathbf{h}_1 + \hat{\mathbf{p}}_2 \cdot \mathbf{h}_2 \tag{5.35}$$

$\hat{\mathbf{p}}_1$ and $\hat{\mathbf{p}}_2$ are polarization vectors and will be discussed in more detail in Section 5.3.3. The column vectors $\mathbf{h}_1$ and $\mathbf{h}_2$ contain the complex amplitude coefficients of the plane waves. Until a solution is obtained, both are unknown quantities.

From Eqs. (5.7)-(5.8), the polarization vectors must satisfy the following matrix equations.

$$\hat{\mathbf{p}}_1 \times \hat{\mathbf{p}}_2 = \left|\vec{\mathbf{K}}\right|^{-1} \vec{\mathbf{K}} \tag{5.36}$$

$$\vec{\mathbf{K}} \times \hat{\mathbf{p}}_1 = \left|\vec{\mathbf{K}}\right| \hat{\mathbf{p}}_2 \tag{5.37}$$

$$\vec{\mathbf{K}} \times \hat{\mathbf{p}}_2 = -\left|\vec{\mathbf{K}}\right| \hat{\mathbf{p}}_1 \tag{5.38}$$

$$\left|\vec{\mathbf{K}}\right| = \begin{bmatrix} \sqrt{\mathbf{K}_x^2 + \mathbf{K}_y^2 + \mathbf{K}_z^2} & 0 & 0 \\ 0 & \sqrt{\mathbf{K}_x^2 + \mathbf{K}_y^2 + \mathbf{K}_z^2} & 0 \\ 0 & 0 & \sqrt{\mathbf{K}_x^2 + \mathbf{K}_y^2 + \mathbf{K}_z^2} \end{bmatrix} \tag{5.39}$$

Substituting Eq. (5.35) into the wave equation yields

$$\vec{\mathbf{K}} \times \varepsilon_r^{-1} \vec{\mathbf{K}} \times \left(\hat{\mathbf{p}}_1 \cdot \mathbf{h}_1 + \hat{\mathbf{p}}_2 \cdot \mathbf{h}_2\right) = -k_0^2 \boldsymbol{\mu}_r \left(\hat{\mathbf{p}}_1 \cdot \mathbf{h}_1 + \hat{\mathbf{p}}_2 \cdot \mathbf{h}_2\right) \tag{5.40}$$

This matrix equation can be solved to obtain eigen-values and eigen-vectors, but it is possible to reduce matrix size by more than half by projecting both sides of this equation onto common vector components $\hat{\mathbf{p}}_1$ and $\hat{\mathbf{p}}_2$. This can be performed using Eq. (5.32) to obtain

$$\begin{aligned} &\hat{\mathbf{p}}_1 \cdot \left\{\hat{\mathbf{p}}_1 \bullet \left(\vec{\mathbf{K}} \times \varepsilon_r^{-1} \left|\vec{\mathbf{K}}\right| \cdot \hat{\mathbf{p}}_2 \cdot \mathbf{h}_1 - \vec{\mathbf{K}} \times \varepsilon_r^{-1} \left|\vec{\mathbf{K}}\right| \cdot \hat{\mathbf{p}}_1 \cdot \mathbf{h}_2\right)\right\} \\ &+\hat{\mathbf{p}}_2 \cdot \left\{\hat{\mathbf{p}}_2 \bullet \left(\vec{\mathbf{K}} \times \varepsilon_r^{-1} \left|\vec{\mathbf{K}}\right| \cdot \hat{\mathbf{p}}_2 \cdot \mathbf{h}_1 - \vec{\mathbf{K}} \times \varepsilon_r^{-1} \left|\vec{\mathbf{K}}\right| \cdot \hat{\mathbf{p}}_1 \cdot \mathbf{h}_2\right)\right\} \end{aligned} = \begin{aligned} &\hat{\mathbf{p}}_1 \cdot \left\{\hat{\mathbf{p}}_1 \bullet \left(-k_0^2 \boldsymbol{\mu}_r \hat{\mathbf{p}}_1 \cdot \mathbf{h}_1 - k_0^2 \boldsymbol{\mu}_r \hat{\mathbf{p}}_2 \cdot \mathbf{h}_2\right)\right\} \\ &+\hat{\mathbf{p}}_2 \cdot \left\{\hat{\mathbf{p}}_2 \bullet \left(-k_0^2 \boldsymbol{\mu}_r \hat{\mathbf{p}}_1 \cdot \mathbf{h}_1 - k_0^2 \boldsymbol{\mu}_r \hat{\mathbf{p}}_2 \cdot \mathbf{h}_2\right)\right\} \end{aligned} \tag{5.41}$$

This can be written as two equations because $\hat{\mathbf{p}}_1$ and $\hat{\mathbf{p}}_2$ are orthogonal.

$$\hat{\mathbf{p}}_1 \bullet \left(\vec{\mathbf{K}} \times \boldsymbol{\varepsilon}_r^{-1} \left| \vec{\mathbf{K}} \right| \cdot \hat{\mathbf{p}}_2 \cdot \mathbf{h}_1 - \vec{\mathbf{K}} \times \boldsymbol{\varepsilon}_r^{-1} \left| \vec{\mathbf{K}} \right| \cdot \hat{\mathbf{p}}_1 \cdot \mathbf{h}_2 \right) = \hat{\mathbf{p}}_1 \bullet \left(-k_0^2 \boldsymbol{\mu}_r \hat{\mathbf{p}}_1 \cdot \mathbf{h}_1 - k_0^2 \boldsymbol{\mu}_r \hat{\mathbf{p}}_2 \cdot \mathbf{h}_2 \right) \qquad (5.42)$$

$$\hat{\mathbf{p}}_2 \bullet \left(\vec{\mathbf{K}} \times \boldsymbol{\varepsilon}_r^{-1} \left| \vec{\mathbf{K}} \right| \cdot \hat{\mathbf{p}}_2 \cdot \mathbf{h}_1 - \vec{\mathbf{K}} \times \boldsymbol{\varepsilon}_r^{-1} \left| \vec{\mathbf{K}} \right| \cdot \hat{\mathbf{p}}_1 \cdot \mathbf{h}_2 \right) = \hat{\mathbf{p}}_2 \bullet \left(-k_0^2 \boldsymbol{\mu}_r \hat{\mathbf{p}}_1 \cdot \mathbf{h}_1 - k_0^2 \boldsymbol{\mu}_r \hat{\mathbf{p}}_2 \cdot \mathbf{h}_2 \right) \qquad (5.43)$$

Applying identities (5.25)-(5.27) and Eqs. (5.36)-(5.38), this system of equations can be simplified to

$$\left(\left| \vec{\mathbf{K}} \right| \hat{\mathbf{p}}_2 \bullet \boldsymbol{\varepsilon}_r^{-1} \left| \vec{\mathbf{K}} \right| \hat{\mathbf{p}}_2 \right) \mathbf{h}_1 - \left(\left| \vec{\mathbf{K}} \right| \hat{\mathbf{p}}_2 \bullet \boldsymbol{\varepsilon}_r^{-1} \left| \vec{\mathbf{K}} \right| \hat{\mathbf{p}}_1 \right) \mathbf{h}_2 = k_0^2 \left(\hat{\mathbf{p}}_1 \bullet \boldsymbol{\mu}_r \hat{\mathbf{p}}_1 \right) \mathbf{h}_1 + k_0^2 \left(\hat{\mathbf{p}}_1 \bullet \boldsymbol{\mu}_r \hat{\mathbf{p}}_2 \right) \mathbf{h}_2 \qquad (5.44)$$

$$-\left(\left| \vec{\mathbf{K}} \right| \hat{\mathbf{p}}_1 \bullet \boldsymbol{\varepsilon}_r^{-1} \left| \vec{\mathbf{K}} \right| \hat{\mathbf{p}}_2 \right) \mathbf{h}_1 + \left(\left| \vec{\mathbf{K}} \right| \hat{\mathbf{p}}_1 \bullet \boldsymbol{\varepsilon}_r^{-1} \left| \vec{\mathbf{K}} \right| \hat{\mathbf{p}}_1 \right) \mathbf{h}_2 = k_0^2 \left(\hat{\mathbf{p}}_2 \bullet \boldsymbol{\mu}_r \hat{\mathbf{p}}_1 \right) \mathbf{h}_1 + k_0^2 \left(\hat{\mathbf{p}}_2 \bullet \boldsymbol{\mu}_r \hat{\mathbf{p}}_2 \right) \mathbf{h}_2 \qquad (5.45)$$

The eigen-value equation is constructed by writing these equations in matrix form.

$$\begin{bmatrix} \left| \vec{\mathbf{K}} \right| \hat{\mathbf{p}}_2 \bullet \boldsymbol{\varepsilon}_r^{-1} \left| \vec{\mathbf{K}} \right| \hat{\mathbf{p}}_2 & -\left| \vec{\mathbf{K}} \right| \hat{\mathbf{p}}_2 \bullet \boldsymbol{\varepsilon}_r^{-1} \left| \vec{\mathbf{K}} \right| \hat{\mathbf{p}}_1 \\ -\left| \vec{\mathbf{K}} \right| \hat{\mathbf{p}}_1 \bullet \boldsymbol{\varepsilon}_r^{-1} \left| \vec{\mathbf{K}} \right| \hat{\mathbf{p}}_2 & \left| \vec{\mathbf{K}} \right| \hat{\mathbf{p}}_1 \bullet \boldsymbol{\varepsilon}_r^{-1} \left| \vec{\mathbf{K}} \right| \hat{\mathbf{p}}_1 \end{bmatrix} \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{bmatrix} = k_0^2 \begin{bmatrix} \hat{\mathbf{p}}_1 \bullet \boldsymbol{\mu}_r \hat{\mathbf{p}}_1 & \hat{\mathbf{p}}_1 \bullet \boldsymbol{\mu}_r \hat{\mathbf{p}}_2 \\ \hat{\mathbf{p}}_2 \bullet \boldsymbol{\mu}_r \hat{\mathbf{p}}_1 & \hat{\mathbf{p}}_2 \bullet \boldsymbol{\mu}_r \hat{\mathbf{p}}_2 \end{bmatrix} \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{bmatrix} \qquad (5.46)$$

Matrix size has been reduced from $9N^2$ elements in Eq. (5.40) to $4N^2$ elements in Eq. (5.46). For non-magnetic materials, the square matrix on the right side of this equation reduces to the unity matrix and can be neglected.

5.3.3.   *Calculation of Orthogonal Polarization Vectors*

Three dimensional analysis requires the field to be written in terms of two orthogonal polarizations. It is convenient and meaningful to select polarization vectors consistent with parallel and perpendicular polarizations if a predominant direction of propagation can be defined. Assuming the predominant direction to be in the z-direction, unit vectors in the direction of perpendicular and parallel polarizations are

$$\hat{p}_{p,q,r}^{\perp} = \hat{z} \times \frac{\vec{k}_{p,q,r}}{\left| \vec{k}_{p,q,r} \right|} \qquad \hat{p}_{p,q,r}^{\|} = \frac{\vec{k}_{p,q,r}}{\left| \vec{k}_{p,q,r} \right|} \times \hat{p}_{p,q,r}^{\perp} \qquad \text{for } \frac{\vec{k}_{p,q,r}}{\left| \vec{k}_{p,q,r} \right|} \neq \hat{z} \qquad (5.47)$$

119

These equations are not valid when the wave vector is parallel to the *z*-axis. In this case, it is easiest to set $\hat{p}^{\perp}_{p,q,r} = \hat{x}$ and $\hat{p}^{\parallel}_{p,q,r} = \hat{y}$. Collected into matrix form, these are

$$\hat{\mathbf{p}}_1 = \begin{bmatrix} \mathbf{p}_{x,1} \\ \mathbf{p}_{y,1} \\ \mathbf{p}_{z,1} \end{bmatrix} \qquad \hat{\mathbf{p}}_2 = \begin{bmatrix} \mathbf{p}_{x,2} \\ \mathbf{p}_{y,2} \\ \mathbf{p}_{z,2} \end{bmatrix} \tag{5.48}$$

The matrices $\hat{\mathbf{p}}_2$ and $\hat{\mathbf{p}}_2$ are column vectors of square matrices like $\vec{\mathbf{A}}$, $\vec{\mathbf{B}}$, and $\vec{\mathbf{C}}$ in Eq. (5.28).

Matrices $\mathbf{p}_{x,1}$, $\mathbf{p}_{y,1}$, $\mathbf{p}_{z,1}$, $\mathbf{p}_{x,2}$, $\mathbf{p}_{y,2}$, and $\mathbf{p}_{z,2}$ are square matrices with polarization components of the plane waves along their diagonals.

5.3.4.   Two-Dimensional Formulation

Expanding Eqs. (5.21)-(5.22) into vector components leads to a set of six equations.

$$\mathbf{K}_y \mathbf{E}_z - \mathbf{K}_z \mathbf{E}_y = \omega \mu_0 \mathbf{\mu}_x \mathbf{H}_x \tag{5.49}$$

$$\mathbf{K}_z \mathbf{E}_x - \mathbf{K}_x \mathbf{E}_z = \omega \mu_0 \mathbf{\mu}_y \mathbf{H}_y \tag{5.50}$$

$$\mathbf{K}_x \mathbf{E}_y - \mathbf{K}_y \mathbf{E}_x = \omega \mu_0 \mathbf{\mu}_z \mathbf{H}_z \tag{5.51}$$

$$\mathbf{K}_y \mathbf{H}_z - \mathbf{K}_z \mathbf{H}_y = -\omega \varepsilon_0 \mathbf{\varepsilon}_x \mathbf{E}_x \tag{5.52}$$

$$\mathbf{K}_z \mathbf{H}_x - \mathbf{K}_x \mathbf{H}_z = -\omega \varepsilon_0 \mathbf{\varepsilon}_y \mathbf{E}_y \tag{5.53}$$

$$\mathbf{K}_x \mathbf{H}_y - \mathbf{K}_y \mathbf{H}_x = -\omega \varepsilon_0 \mathbf{\varepsilon}_z \mathbf{E}_z \tag{5.54}$$

For two-dimension problems with plane waves restricted to the xy-plane, $\mathbf{K}_z = \mathbf{0}$ and the above equations decouple into two sets of equations corresponding to two distinct modes.

TE Mode

$$\mathbf{K}_x \mathbf{H}_y - \mathbf{K}_y \mathbf{H}_x = -\omega \varepsilon_0 \mathbf{\varepsilon}_z \mathbf{E}_z \tag{5.55}$$

$$K_y E_z = \omega \mu_0 \mu_x H_x \tag{5.56}$$

$$-K_x E_z = \omega \mu_0 \mu_y H_y \tag{5.57}$$

TM Mode

$$K_x E_y - K_y E_x = \omega \mu_0 \mu_z H_z \tag{5.58}$$

$$K_y H_z = -\omega \varepsilon_0 \varepsilon_x E_x \tag{5.59}$$

$$-K_x H_z = -\omega \varepsilon_0 \varepsilon_y E_y \tag{5.60}$$

The eigen-value equation for the TE mode is derived by solving Eqs. (5.56)-(5.57) for H_x and H_y and substituting these into Eq. (5.55). A similar treatment of Eqs. (5.58)-(5.60) leads to the TM mode eigen-value equation.

TE Mode

$$\left(K_x \mu_y^{-1} K_x + K_y \mu_x^{-1} K_y \right) E_z = k_0^2 \varepsilon_z E_z \tag{5.61}$$

$$H_x = \left(1/\omega \mu_0 \right) \mu_x^{-1} K_y E_z \tag{5.62}$$

$$H_y = -\left(1/\omega \mu_0 \right) \mu_y^{-1} K_x E_z \tag{5.63}$$

TM Mode

$$\left(K_x \varepsilon_y^{-1} K_x + K_y \varepsilon_x^{-1} K_y \right) H_z = k_0^2 \mu_z H_z \tag{5.64}$$

$$E_x = -\left(1/\omega \varepsilon_0 \right) \varepsilon_x^{-1} K_y H_z \tag{5.65}$$

$$E_y = \left(1/\omega \varepsilon_0 \right) \varepsilon_y^{-1} K_x H_z \tag{5.66}$$

5.3.5. *One-Dimensional Formulation*

Equations (5.49)-(5.54) can be reduced further for one-dimension systems where

$\mathbf{K}_x = \mathbf{K}_y = \mathbf{0}$. For this case $\mathbf{E}_z = \mathbf{H}_z = \mathbf{0}$ and both remaining polarizations are described by the

same eigen-value equation.

$$\left(\mathbf{K}_z \boldsymbol{\varepsilon}^{-1} \mathbf{K}_z \right) \mathbf{H} = k_0^2 \boldsymbol{\mu} \mathbf{H} \tag{5.67}$$

5.4. Implementation

To clarify how equations derived in this chapter are to be used, Figure 5-3 shows block

diagrams of how PWEM can be implemented for 3D, 2D, and 1D problems.
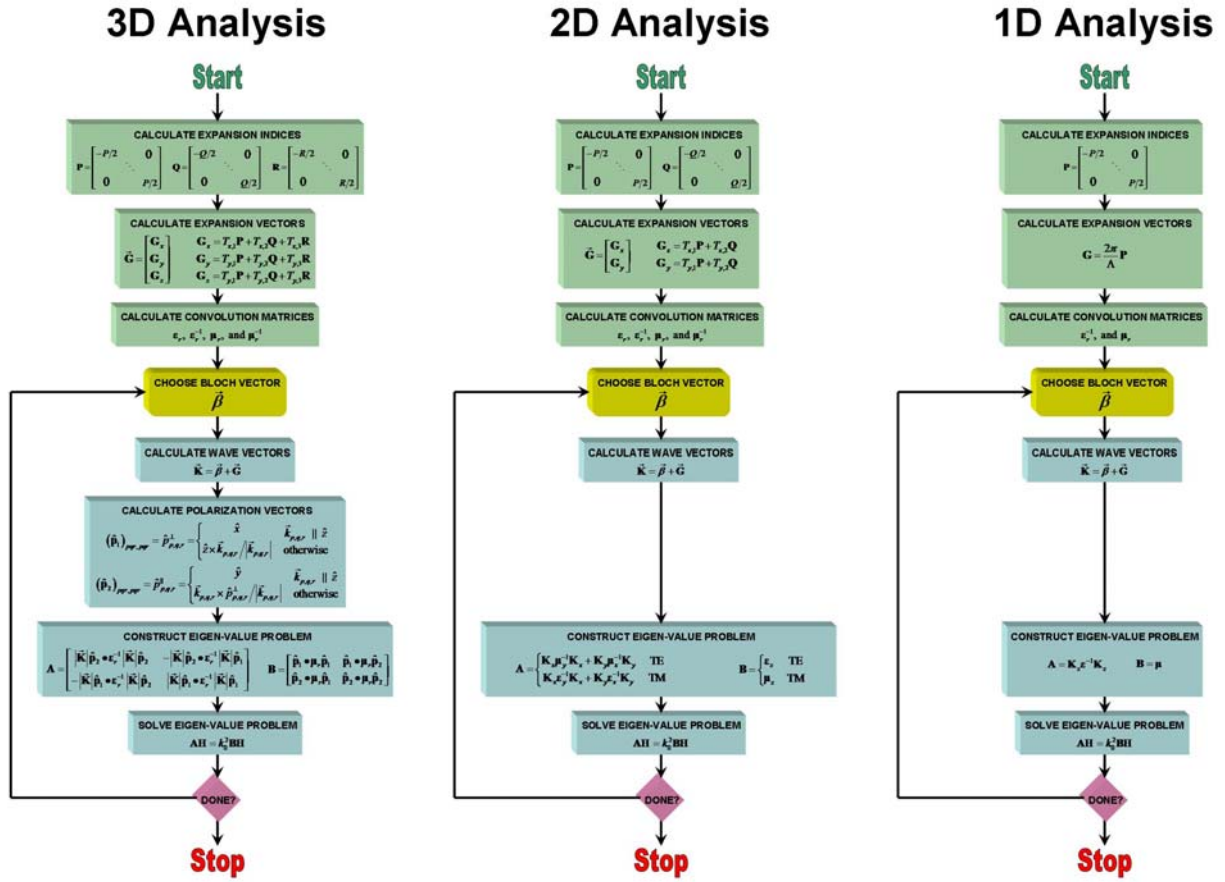
Figure 5-3: Block diagram of PWEM implementations

5.5. Benchmark Simulation

To verify the accuracy of the fully three-dimensional PWEM algorithm developed in this dissertation, the photonic band structure of a simple cubic lattice comprised of dielectric rods (n=1.53) was calculated using PWEM and FDTD. These are compared in Figure 5-4.
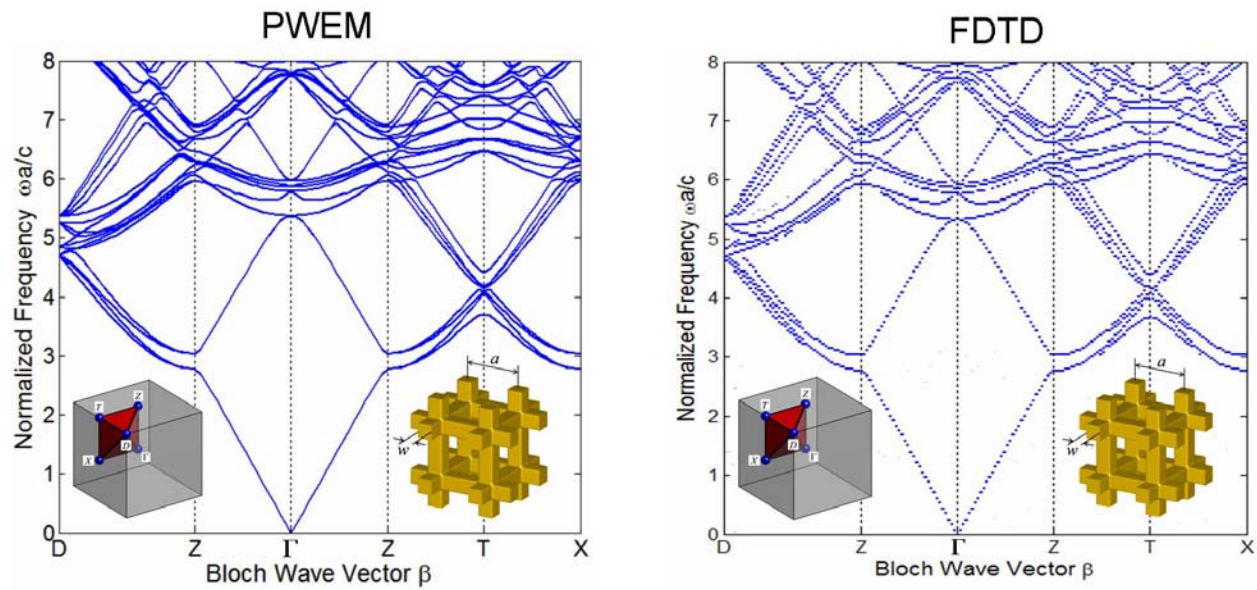
Figure 5-4: Benchmark simulation compared to FDTD

These simulations took roughly the same amount of time to calculate. PWEM offers the advantage providing a clear distinction between bands. As refractive index contrast is raised or if metals are incorporated, FDTD becomes the more efficient method.

CHAPTER 6
RIGOROUS COUPLED-WAVE ANALYSIS

6.1. Overview

Rigorous coupled-wave analysis (RCWA) [50-52, 124-133] seems to be the most popular method to model scattering from periodic structures with low to moderate index contrast. It offers incredibly efficient analysis of layered dielectric structures that are periodic on the scale of a wavelength. It is unconditionally stable, full vector, and rigorous in the sense that no approximations are made to Maxwell's equations. Being a frequency-domain method, it easily incorporates material dispersion. It is a semi-analytical method where the wave equation is solved analytically in the longitudinal direction. Analytical treatment of the wave equation in the direction of propagation enables RCWA to model structures of arbitrary length. Numerical treatment of the transverse problem limits the width of 3D structures to be on the order of a wavelength. For 2D structures, multiple wavelengths are feasible. It is able to handle incident waves at oblique angles. A good amount of literature can be found on the method and considerable effort has been devoted to improving its speed and efficiency.

The concept and geometrical construction of RCWA is illustrated in Figure 6-1. To implement the method, structures are divided into layers that are uniform in the longitudinal direction. The transverse problem is solved in reciprocal space by expressing the field as a sum of spatial harmonics. This transforms the wave equation into a set of ordinary differential equations that is solved as an eigen-value problem. The eigen-vectors characterize the configurations of spatial harmonics that can exist in each layer. The eigen-values describe longitudinal behavior in terms of a complex propagation constant that incorporates loss, gain,

and coupling between modes. An overall solution is obtained through the boundary conditions by matching tangential fields at the segment interfaces.
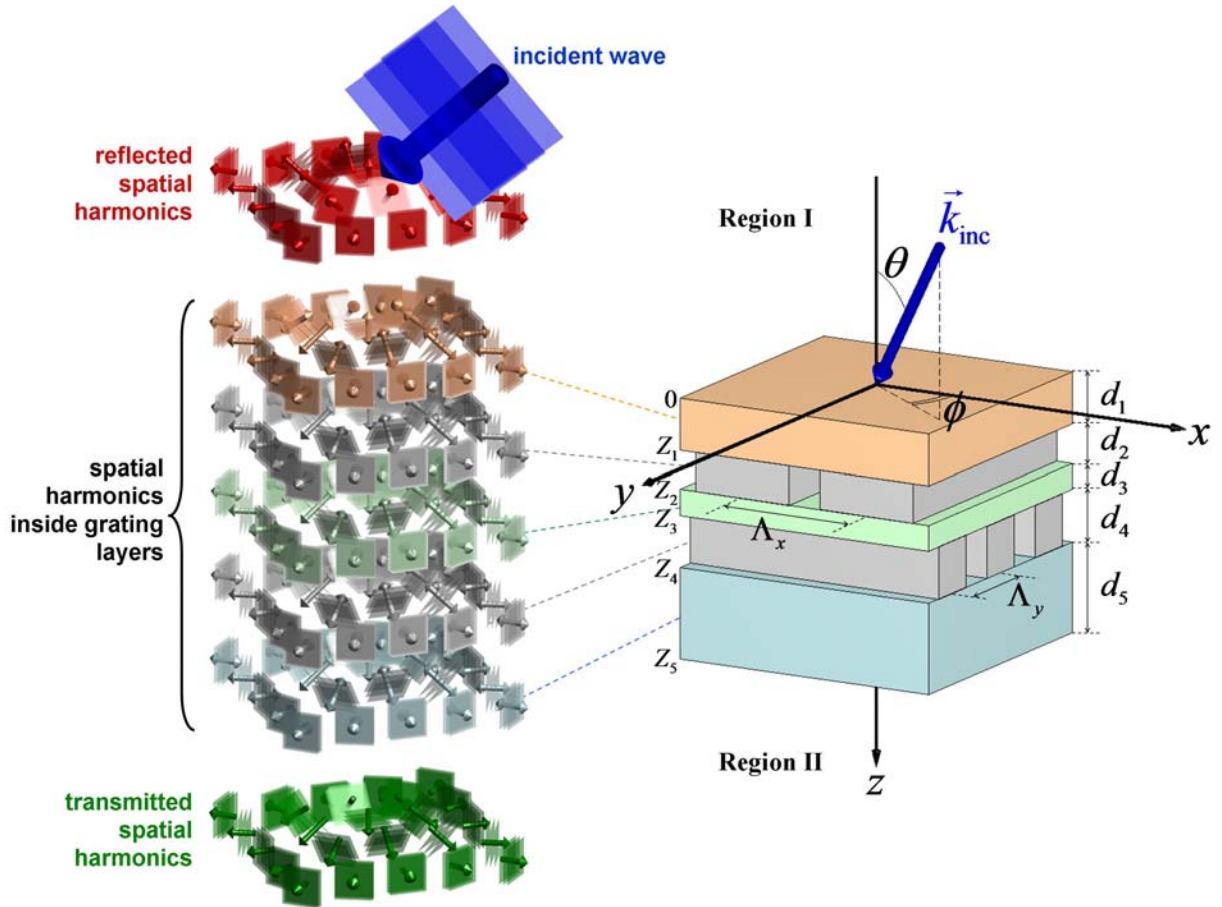

Figure 6-1: RCWA concept and geometry

RCWA is best suited to analyze layered problems with minimal volumetric complexity due to the eigen-mode computations in each unique layer. Arbitrarily shaped devices must be handled by resolving them with many thin segments leading to a "stair-case" approximation of their geometry. This can become very computationally intensive for large elements because a complete set of eigen-modes must be computed in each segment.

For structures that are described well by Fourier series with few terms, RCWA is incredibly fast and efficient. This includes virtually all periodic dielectric structures of low to moderate index contrast that do not have very fine features. When problems have very sharp index contrast or metals, many Fourier terms are required to accurately describe the structure and fields, so the method quickly becomes bogged down computing the eigen-modes. This is referred to as "slow convergence." Modeling TM polarized waves is particularly problematic [125, 129, 134-136] for reasons discussed in Ref. [137]. It is argued in Refs. [125, 138] that for TM polarization the staircase approximation of arbitrarily shaped gratings never converges to the correct answer regardless of the number of layers used. Application of the method for TM polarization is restricted to truly lamellar profiles for this reason. The problem does not exist for TE polarization and arbitrarily shaped structures are approximated well by the staircase approximation.

While the formulation of RCWA may seem more tedious and complicated than other methods, its implementation is surprisingly simple and compact.

6.2. Formulation

This section will derive all necessary equations to implement RCWA using the enhanced transmittance matrix approach for the multilayer geometry depicted in Figure 6-1 [51, 52]. The enhanced transmittance matrix approach was selected over a scattering matrix formulation for its greater speed. A single plane wave is incident at an oblique angle on a grating stack from region I. Each layer in the stack is uniform in the longitudinal direction with thickness d_i for the ith layer. Reflected waves are calculated in region I and transmitted waves are calculated in region II. The following formulation will also derive equations for calculating fields in the intermediate layers when that is needed.

6.2.1.   Maxwell's Equations

Maxwell's curl equations for source-free materials are

$$\nabla \times \vec{E} = -j\omega\mu_0\mu_r\vec{H} \tag{6.1}$$

$$\nabla \times \vec{H} = j\omega\varepsilon_0\varepsilon_r\vec{E} \tag{6.2}$$

It is convenient to normalize the magnetic field in the following manner.

$$\tilde{\vec{H}} = -j\sqrt{\frac{\mu_0}{\varepsilon_0}}\vec{H} \tag{6.3}$$

In terms of the normalized magnetic field, Maxwell's equations become

$$\nabla \times \vec{E} = k_0\mu_r\tilde{\vec{H}} \tag{6.4}$$

$$\nabla \times \tilde{\vec{H}} = k_0\varepsilon_r\vec{E} \tag{6.5}$$

Expanding Eqs. (6.4)-(6.5) into vector components leads to six coupled equations.

$$\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = k_0\mu_r\tilde{H}_x \tag{6.6}$$

$$\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = k_0\mu_r\tilde{H}_y \tag{6.7}$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = k_0\mu_r\tilde{H}_z \tag{6.8}$$

$$\frac{\partial \tilde{H}_z}{\partial y} - \frac{\partial \tilde{H}_y}{\partial z} = k_0\varepsilon_r E_x \tag{6.9}$$

$$\frac{\partial \tilde{H}_x}{\partial z} - \frac{\partial \tilde{H}_z}{\partial x} = k_0\varepsilon_r E_y \tag{6.10}$$

$$\frac{\partial \tilde{H}_y}{\partial x} - \frac{\partial \tilde{H}_x}{\partial y} = k_0\varepsilon_r E_z \tag{6.11}$$

6.2.2. Fourier Expansion of Terms

Assuming the problem is periodic, fields can be expanded into a 2D Fourier series where the terms in the series correspond to spatial harmonics. For the ith grating layer, the expansions can be written as

$$\vec{E}_i(x, y, z) = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \vec{S}_{i,m,n}(z) \exp\left[-j\left(k_{m,n}^x x + k_{m,n}^y y\right)\right] \tag{6.12}$$

$$\vec{H}_i(x, y, z) = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \vec{U}_{i,m,n}(z) \exp\left[-j\left(k_{m,n}^x x + k_{m,n}^y y\right)\right] \tag{6.13}$$

Wave vector components of the spatial harmonics are defined in terms of reciprocal lattice vectors $\vec{G}_1$ and $\vec{G}_2$ of the grating unit cell. These are

$$k_{m,n}^x = k_{\text{inc}}^x - \left(m\vec{G}_1 + n\vec{G}_2\right) \bullet \hat{x} \tag{6.14}$$

$$k_{m,n}^y = k_{\text{inc}}^y - \left(m\vec{G}_1 + n\vec{G}_2\right) \bullet \hat{y} \tag{6.15}$$

$$k_{i,m,n}^z = \begin{cases} \sqrt{k_0^2 \varepsilon_i - \left(k_{m,n}^x\right)^2 - \left(k_{m,n}^y\right)^2} & k_0^2 \varepsilon_i > \left(k_{m,n}^x\right)^2 + \left(k_{m,n}^y\right)^2 \\ -j\sqrt{\left(k_{m,n}^x\right)^2 + \left(k_{m,n}^y\right)^2 - k_0^2 \varepsilon_i} & k_0^2 \varepsilon_i < \left(k_{m,n}^x\right)^2 + \left(k_{m,n}^y\right)^2 \end{cases} \tag{6.16}$$

In a similar manner, the permittivity and permeability functions can be expanded into Fourier series. To improve convergence, Fourier expansions are computed for the inverse of the material properties as well.

$$\varepsilon_i(x, y) = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} a_{i,m,n} \exp\left(j\vec{G}_{m,n} \bullet \vec{r}\right) \tag{6.17}$$

$$\mu_i(x, y) = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} c_{i,m,n} \exp\left(j\vec{G}_{m,n} \bullet \vec{r}\right) \tag{6.18}$$

$$\varepsilon_i^{-1}(x, y) = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} b_{i,m,n} \exp\left(j\vec{G}_{m,n} \bullet \vec{r}\right) \tag{6.19}$$

$$\mu_i^{-1}(x, y) = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} d_{i,m,n} \exp\left(j\vec{G}_{m,n} \bullet \vec{r}\right) \tag{6.20}$$

where the Fourier series coefficients are calculated as

$$a_{i,m,n} = \frac{1}{A} \iint_A \varepsilon_i(x, y) \exp\left(-j\vec{G}_{m,n} \bullet \vec{r}\right) \tag{6.21}$$

$$b_{i,m,n} = \frac{1}{A} \iint_A \mu_i(x, y) \exp\left(-j\vec{G}_{m,n} \bullet \vec{r}\right) \tag{6.22}$$

$$c_{i,m,n} = \frac{1}{A} \iint_A \varepsilon_i^{-1}(x, y) \exp\left(-j\vec{G}_{m,n} \bullet \vec{r}\right) \tag{6.23}$$

$$d_{i,m,n} = \frac{1}{A} \iint_A \mu_i^{-1}(x, y) \exp\left(-j\vec{G}_{m,n} \bullet \vec{r}\right) \tag{6.24}$$

6.2.3. *Semi-Analytical Maxwell's Equations in Reciprocal Space*

Maxwell's equations can be transformed into reciprocal space by substituting the above expansions into Eqs. (6.6)-(6.11). For brevity, the details of the substitution will only be outlined for Eq. (6.6). The others may be written by analogy. Upon substitution, Eq. (6.6) becomes

$$\frac{\partial}{\partial y} \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} S_{i,m,n}^z(z)\exp\left[-j\left(k_{m,n}^x x + k_{m,n}^y y\right)\right] - \frac{\partial}{\partial z} \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} S_{i,m,n}^y(z)\exp\left[-j\left(k_{m,n}^x x + k_{m,n}^y y\right)\right]$$
$$= k_0 \left\{ \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} c_{i,m,n} \exp\left[j\left(mG_{x1} + nG_{x2}\right)\right] \right\} \left\{ \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} U_{i,m,n}^x(z)\exp\left[-j\left(k_{m,n}^x x + k_{m,n}^y y\right)\right] \right\} \tag{6.25}$$

After some algebra, this can be written as

$$\sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} -jk_n^y S_{i,m,n}^z(z)\exp\left[-j\left(k_{m,n}^x x + k_{m,n}^y y\right)\right] - \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \frac{dS_{i,m,n}^y(z)}{dz}\exp\left[-j\left(k_{m,n}^x x + k_{m,n}^y y\right)\right]$$
$$= k_0 \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \left\{ \exp\left[-j\left(k_{m,n}^x x + k_{m,n}^y y\right)\right] \sum_{q=-\infty}^{+\infty} \sum_{r=-\infty}^{+\infty} c_{i,m-q,n-r} U_{i,q,r}^x(z) \right\} \tag{6.26}$$

The partial derivative is written as an ordinary derivative because $S_{i,m,n}^y(z)$ is only a function of z. The entire equation can be brought inside the double summation over m and n.

This defines an infinite set of equations corresponding to an infinite set of spatial harmonics. For the $(m,n)^{th}$ spatial harmonic, the equation is

$$-jk_{m,n}^{y}S_{i,m,n}^{z}(z)-\frac{dS_{i,m,n}^{y}(z)}{dz}=k_{0}\sum_{q=-\infty}^{+\infty}\sum_{r=-\infty}^{+\infty}c_{i,m-q,n-r}U_{i,q,r}^{x}(z) \tag{6.27}$$

The double summation on the right of this equation defines a convolution of μ_i with $U_i^x(z)$. Applying this same procedure to Eqs. (6.7)-(6.11) leads to the complete set of Maxwell's curl equations in reciprocal space.

$$-jk_{m,n}^{y}S_{i,m,n}^{z}(z)-\frac{dS_{i,m,n}^{y}(z)}{dz}=k_{0}\mu_{i}^{x}*U_{i}^{x}(z) \tag{6.28}$$

$$\frac{dS_{i,m,n}^{x}(z)}{dz}+jk_{m,n}^{x}S_{i,m,n}^{z}(z)=k_{0}\mu_{i}^{y}*U_{i}^{y}(z) \tag{6.29}$$

$$-jk_{m,n}^{x}S_{i,m,n}^{y}(z)+jk_{m,n}^{y}S_{i,m,n}^{x}(z)=k_{0}\mu_{i}^{z}*U_{i}^{z}(z) \tag{6.30}$$

$$-jk_{m,n}^{y}U_{i,m,n}^{z}(z)-\frac{dU_{i,m,n}^{y}(z)}{dz}=k_{0}\varepsilon_{i}^{x}*S_{i}^{x}(z) \tag{6.31}$$

$$\frac{dU_{i,m,n}^{x}(z)}{dz}+jk_{m,n}^{x}U_{i,m,n}^{z}(z)=k_{0}\varepsilon_{i}^{y}*S_{i}^{y}(z) \tag{6.32}$$

$$-jk_{m,n}^{x}U_{i,m,n}^{y}(z)+jk_{m,n}^{y}U_{i,m,n}^{x}(z)=k_{0}\varepsilon_{i}^{z}*S_{i}^{z}(z) \tag{6.33}$$

6.2.4.   Numerical Representation

To be solved on a computer, the infinite set of equations above must be truncated to a finite set. For $2M+1$ spatial harmonics retained along $\vec{G}_1$ and $2N+1$ spatial harmonics retained along $\vec{G}_2$, the indices of the harmonics are restricted to

$$-M\leq m\leq M \qquad\qquad -N\leq n\leq N \tag{6.34}$$

131

The finite set of equations can be written in matrix form as

$$-jk_0 \mathbf{K}_y \mathbf{S}_i^z(z) - \frac{d}{dz} \mathbf{S}_i^y(z) = k_0 \boldsymbol{\mu}_{i,x} \mathbf{U}_i^x(z) \tag{6.35}$$

$$\frac{d}{dz} \mathbf{S}_i^x(z) + jk_0 \mathbf{K}_x \mathbf{S}_i^z(z) = k_0 \boldsymbol{\mu}_{i,y} \mathbf{U}_i^y(z) \tag{6.36}$$

$$-\mathbf{K}_x \mathbf{S}_i^y(z) + \mathbf{K}_y \mathbf{S}_i^x(z) = -j\boldsymbol{\mu}_{i,z} \mathbf{U}_i^z(z) \tag{6.37}$$

$$-jk_0 \mathbf{K}_y \mathbf{U}_i^z(z) - \frac{d}{dz} \mathbf{U}_i^y(z) = k_0 \boldsymbol{\varepsilon}_{i,x} \mathbf{S}_i^x(z) \tag{6.38}$$

$$\frac{d}{dz} \mathbf{U}_i^x(z) + jk_0 \mathbf{K}_x \mathbf{U}_i^z(z) = k_0 \boldsymbol{\varepsilon}_{i,y} \mathbf{S}_i^y(z) \tag{6.39}$$

$$-\mathbf{K}_x \mathbf{U}_i^y(z) + \mathbf{K}_y \mathbf{U}_i^x(z) = -j\boldsymbol{\varepsilon}_{i,z} \mathbf{S}_i^z(z) \tag{6.40}$$

$\mathbf{K}_x$ and $\mathbf{K}_y$ are diagonal matrices containing the wave vector components defined in Eqs. (6.14) and (6.15) respectively, but are normalized by dividing them by k_0. The quantities $\mathbf{S}_i^x(z)$, $\mathbf{S}_i^y(z)$, and $\mathbf{S}_i^z(z)$ are column vectors containing complex amplitude coefficients of the spatial harmonics for the electric field. Likewise $\mathbf{U}_i^x(z)$, $\mathbf{U}_i^y(z)$, and $\mathbf{U}_i^z(z)$ correspond to magnetic field coefficients. The square matrices $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\mu}_i$ are matrix operators for the convolution operations defined in Eqs. (6.28)-(6.33). These should not be confused with diagonal matrices of material properties in real space. Construction of these matrices will be discussed later.

6.2.5. Truncation of Spatial Harmonics

The choice of which spatial harmonics are retained is arbitrary, but improper choice can lead to inaccurate results. The number of spatial harmonics retained in a particular direction determines the spatial resolution of structures with index contrast in that direction. A region of

132

k-space is defined by the indices of the spatial harmonics retained in the expansion. Using the conventional approach, this region is rectangular as depicted at the top of Figure 6-2. This approach seems inefficient because it favors spatial resolution along the diagonals.

A simple and powerful truncation scheme [123, 124] is given by

$$\left|\frac{m}{M}\right|^{2\gamma} + \left|\frac{n}{N}\right|^{2\gamma} \leq 1 \tag{6.41}$$

The power parameter γ is any number greater than zero. Figure 6-2 shows the range of truncation schemes that are possible through choice of γ. Very small values only retain harmonics along the Fourier axes. Powers less than 0.5 lead to "pincushion" truncation heavily favoring on-axis harmonics, but does include some off axis harmonics. Diamond truncation is achieved for $\gamma=0.5$ and is like the conventional approach, but favors resolution along the Fourier axes. Powers greater than 1 lead to barrel truncation. This contains more off-axis harmonics and only neglects harmonics from the extreme corners of the conventional rectangle. Very large values approach the conventional rectangular truncation scheme. When nothing is known about which harmonics to favor, circular truncation ($\gamma=1$) is a good choice because it provides equal spatial resolution in all directions.
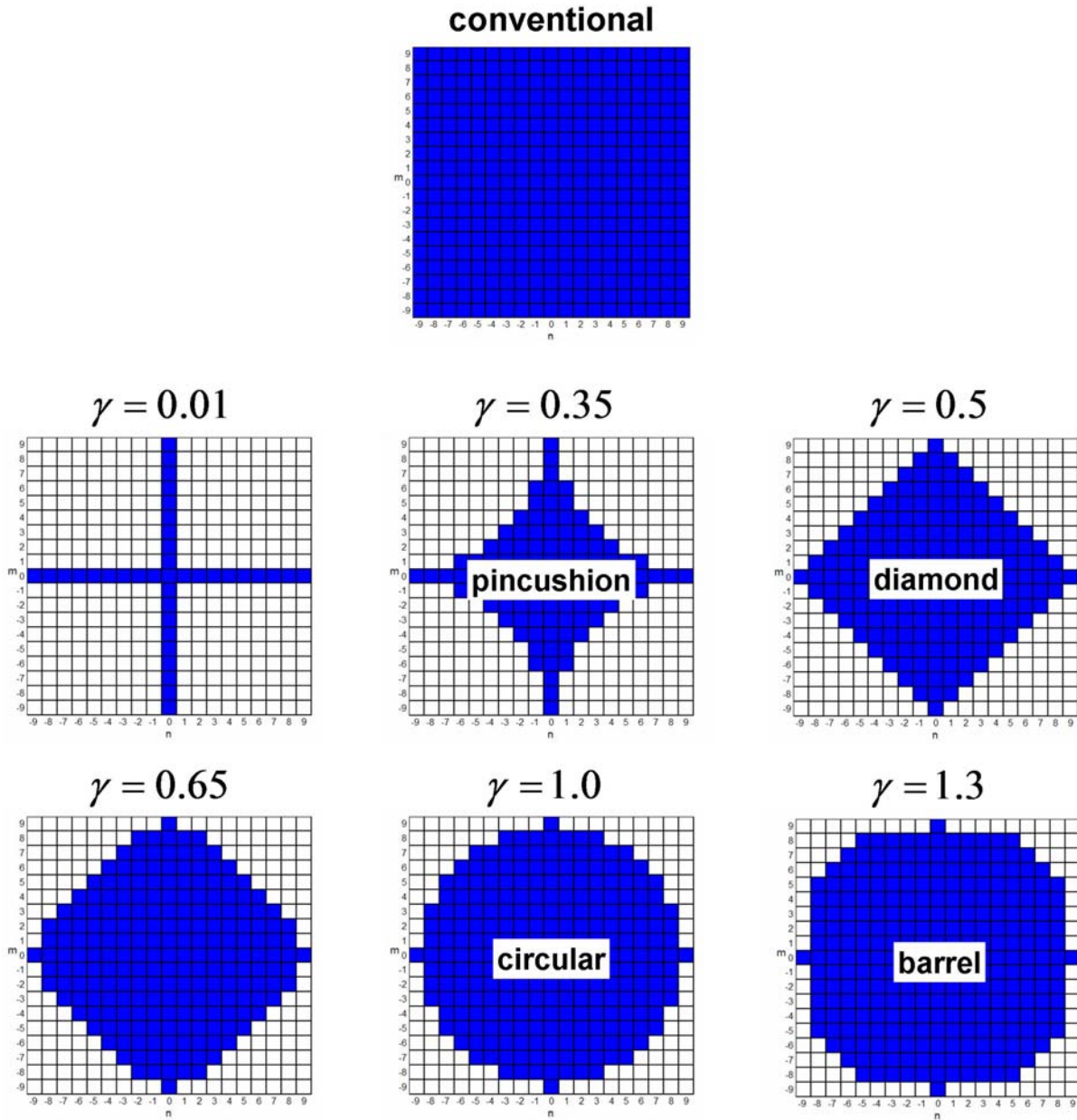
Figure 6-2: Truncation schemes

6.2.6. Construction of Convolution Matrices using FFT

Constructing the convolution matrices is often confusing to those new to RCWA. Figure 6-3 illustrates the three main steps for constructing these matrices without truncation.
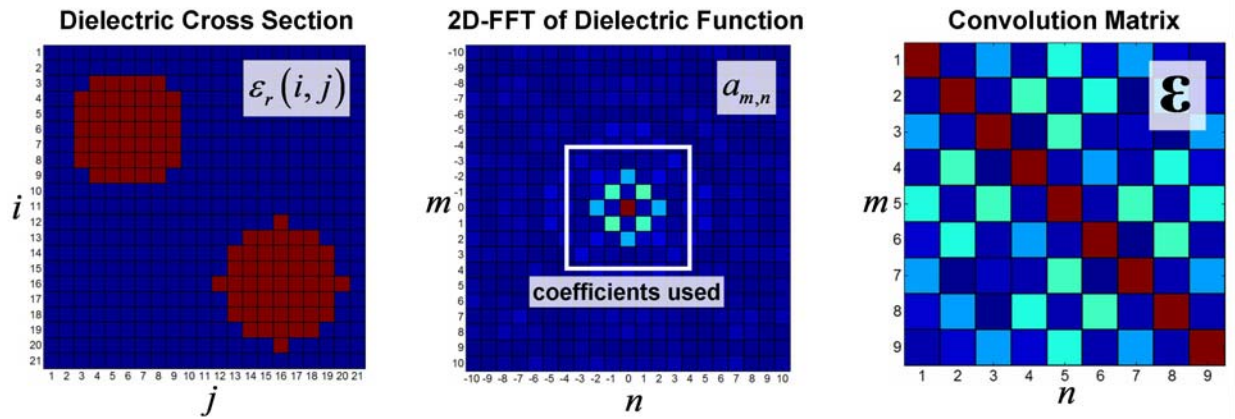
Figure 6-3: Steps for calculating RCWA convolution matrices

First, a 2D dielectric function is constructed that describes the cross section of the grating. This can be analytical or numerical, but analytical is preferred. Second, Fourier series coefficients are calculated for the dielectric function. When the dielectric function is numerical and analytical expressions are not available, the Fourier coefficients can be obtained from the lowest order terms calculated from a 2D fast Fourier transform (FFT). It was observed that many points were needed in the calculation to obtain accurate values for the Fourier coefficients. Grids were often set to 1024×1024 grid points or higher. Most FFT algorithms need to be normalized by dividing by the total number of points in the grid, $N_i N_j$.

$$a(m,n) = \frac{1}{N_i N_j} \mathrm{FFT}\big[\varepsilon(i,j)\big] \tag{6.42}$$

Third, the convolution matrix is constructed by reordering the lowest order Fourier series coefficients. Given Fourier series coefficients calculated in Eq. (6.42), the convolution matrix retaining nine spatial harmonics (i.e. $M=N=1$) would be constructed from the definition given in Eq. (6.27) as

$$
\varepsilon = \begin{bmatrix}
a_{0,0} & a_{-1,0} & a_{-2,0} & a_{0,-1} & a_{-1,-1} & a_{-2,-1} & a_{0,-2} & a_{-1,-2} & a_{-2,-2} \\
a_{1,0} & a_{0,0} & a_{-1,0} & a_{1,-1} & a_{0,-1} & a_{-1,-1} & a_{1,-2} & a_{0,-2} & a_{-1,-2} \\
a_{2,0} & a_{1,0} & a_{0,0} & a_{2,-1} & a_{1,-1} & a_{0,-1} & a_{2,-2} & a_{1,-2} & a_{0,-2} \\
a_{0,1} & a_{-1,1} & a_{-2,1} & a_{0,0} & a_{-1,0} & a_{-2,0} & a_{0,-1} & a_{-1,-1} & a_{-2,-1} \\
a_{1,1} & a_{0,1} & a_{-1,1} & a_{1,0} & a_{0,0} & a_{-1,0} & a_{1,-1} & a_{0,-1} & a_{-1,-1} \\
a_{2,1} & a_{1,1} & a_{0,1} & a_{2,0} & a_{1,0} & a_{0,0} & a_{2,-1} & a_{1,-1} & a_{0,-1} \\
a_{0,2} & a_{-1,2} & a_{-2,2} & a_{0,1} & a_{-1,1} & a_{-2,1} & a_{0,0} & a_{-1,0} & a_{-2,0} \\
a_{1,2} & a_{0,2} & a_{-1,2} & a_{1,1} & a_{0,1} & a_{-1,1} & a_{1,0} & a_{0,0} & a_{-1,0} \\
a_{2,2} & a_{1,2} & a_{0,2} & a_{2,1} & a_{1,1} & a_{0,1} & a_{2,0} & a_{1,0} & a_{0,0}
\end{bmatrix}
\tag{6.43}
$$

6.2.7. *Elimination of Longitudinal Fields*

Solving Eq. (6.37) and Eq. (6.40) for the longitudinal fields yields

$$
\mathbf{U}_i^z(z) = j\boldsymbol{\mu}_{i,z}^{-1}\mathbf{K}_y\mathbf{S}_i^x(z) - j\boldsymbol{\mu}_{i,z}^{-1}\mathbf{K}_x\mathbf{S}_i^y(z)
\tag{6.44}
$$

$$
\mathbf{S}_i^z(z) = j\boldsymbol{\varepsilon}_{i,z}^{-1}\mathbf{K}_y\mathbf{U}_i^x(z) - j\boldsymbol{\varepsilon}_{i,z}^{-1}\mathbf{K}_x\mathbf{U}_i^y(z)
\tag{6.45}
$$

These equations require inverse convolution matrices. For best performance, these should be constructed separately using Eqs. (6.19)-(6.20) to form unique convolution matrices. Substituting Eq. (6.45) into Eqs. (6.35)-(6.36) and Eq. (6.44) into Eqs. (6.38)-(6.39) reduces the problem to

$$
\frac{1}{k_0}\frac{d}{dz}\mathbf{S}_i^y(z) = \left[\mathbf{K}_y\boldsymbol{\varepsilon}_{i,z}^{-1}\mathbf{K}_y - \boldsymbol{\mu}_{i,x}\right]\mathbf{U}_i^x(z) - \left[\mathbf{K}_y\boldsymbol{\varepsilon}_{i,z}^{-1}\mathbf{K}_x\right]\mathbf{U}_i^y(z)
\tag{6.46}
$$

$$
\frac{1}{k_0}\frac{d}{dz}\mathbf{S}_i^x(z) = \left[\mathbf{K}_x\boldsymbol{\varepsilon}_{i,z}^{-1}\mathbf{K}_y\right]\mathbf{U}_i^x(z) + \left[\boldsymbol{\mu}_{i,y} - \mathbf{K}_x\boldsymbol{\varepsilon}_{i,z}^{-1}\mathbf{K}_x\right]\mathbf{U}_i^y(z)
\tag{6.47}
$$

$$\frac{1}{k_0}\frac{d}{dz}\mathbf{U}_i^y(z) = \left[\mathbf{K}_y\boldsymbol{\mu}_{i,z}^{-1}\mathbf{K}_y - \boldsymbol{\varepsilon}_{i,x}\right]\mathbf{S}_i^x(z) - \left[\mathbf{K}_y\boldsymbol{\mu}_{i,z}^{-1}\mathbf{K}_x\right]\mathbf{S}_i^y(z) \tag{6.48}$$

$$\frac{1}{k_0}\frac{d}{dz}\mathbf{U}_i^x(z) = \left[\mathbf{K}_x\boldsymbol{\mu}_{i,z}^{-1}\mathbf{K}_y\right]\mathbf{S}_i^x(z) + \left[\boldsymbol{\varepsilon}_{i,y} - \mathbf{K}_x\boldsymbol{\mu}_{i,z}^{-1}\mathbf{K}_x\right]\mathbf{S}_i^y(z) \tag{6.49}$$

6.2.8.   Formulation of Semi-Analytical Wave Equation

After normalizing the remaining spatial parameter according to $z' = k_0 z$, Eqs. (6.46)-

(6.49) can be written as two coupled matrix equations.

$$\frac{d}{dz'}\begin{bmatrix}\mathbf{S}_i^x(z')\\\mathbf{S}_i^y(z')\end{bmatrix} = \mathbf{P}_i\begin{bmatrix}\mathbf{U}_i^x(z')\\\mathbf{U}_i^y(z')\end{bmatrix} \tag{6.50}$$

$$\frac{d}{dz'}\begin{bmatrix}\mathbf{U}_i^x(z')\\\mathbf{U}_i^y(z')\end{bmatrix} = \mathbf{Q}_i\begin{bmatrix}\mathbf{S}_i^x(z')\\\mathbf{S}_i^y(z')\end{bmatrix} \tag{6.51}$$

where

$$\mathbf{P}_i = \begin{bmatrix}\mathbf{K}_x\boldsymbol{\varepsilon}_{i,z}^{-1}\mathbf{K}_y & \boldsymbol{\mu}_{i,y} - \mathbf{K}_x\boldsymbol{\varepsilon}_{i,z}^{-1}\mathbf{K}_x\\\mathbf{K}_y\boldsymbol{\varepsilon}_{i,z}^{-1}\mathbf{K}_y - \boldsymbol{\mu}_{i,x} & -\mathbf{K}_y\boldsymbol{\varepsilon}_{i,z}^{-1}\mathbf{K}_x\end{bmatrix} \tag{6.52}$$

$$\mathbf{Q}_i = \begin{bmatrix}\mathbf{K}_x\boldsymbol{\mu}_{i,z}^{-1}\mathbf{K}_y & \boldsymbol{\varepsilon}_{i,y} - \mathbf{K}_x\boldsymbol{\mu}_{i,z}^{-1}\mathbf{K}_x\\\mathbf{K}_y\boldsymbol{\mu}_{i,z}^{-1}\mathbf{K}_y - \boldsymbol{\varepsilon}_{i,x} & -\mathbf{K}_y\boldsymbol{\mu}_{i,z}^{-1}\mathbf{K}_x\end{bmatrix} \tag{6.53}$$

The semi-analytical wave equation is derived by differentiating Eq. (6.50) with respect to z' and

eliminating the magnetic field using Eq. (6.51).

$$\frac{d^2}{dz'^2}\begin{bmatrix}\mathbf{S}_i^x(z')\\\mathbf{S}_i^y(z')\end{bmatrix} - \boldsymbol{\Omega}_i^2\begin{bmatrix}\mathbf{S}_i^x(z')\\\mathbf{S}_i^y(z')\end{bmatrix} = 0 \tag{6.54}$$

where

$$\boldsymbol{\Omega}_i^2 = \mathbf{P}_i\mathbf{Q}_i \tag{6.55}$$

6.2.9. Semi-Analytical Solution

The set of differential equations represented in Eq. (6.54) can each be solved analytically. The solutions require calculation of all eigen-vectors and eigen-values of the matrix $\mathbf{\Omega}_i^2$.

$$\mathbf{\Omega}_i^2 \quad \rightarrow \quad \begin{aligned} \mathbf{W}_i &\equiv \text{Eigen-Vectors} \\ \lambda_i^2 &\equiv \text{Eigen-Values} \end{aligned} \tag{6.56}$$

In matrix form, the solutions to Eq. (6.54) can be written as

$$\begin{bmatrix} \mathbf{S}_i^x(z_i') \\ \mathbf{S}_i^y(z_i') \end{bmatrix} = e^{-\mathbf{\Omega}_i z_i'}\mathbf{A}_i^+ + e^{\mathbf{\Omega}_i z_i'}\mathbf{A}_i^- \tag{6.57}$$

$\mathbf{A}_i^+$ and $\mathbf{A}_i^-$ are proportionality constants corresponding to forward and backward propagating waves respectively. The parameter z_i' is the local longitudinal coordinate in the ith grating layer. Assuming the matrix $\mathbf{\Omega}_i^2$ is full rank, the exponentials can be written in terms of its eigen-vectors and positive square-roots of the eigen-values.

$$e^{\mathbf{\Omega}_i z_i'} = \mathbf{W}_i e^{\lambda_i z_i'}\mathbf{W}_i^{-1} = \mathbf{W}_i e^{k_0\lambda_i(z-d_i)}\mathbf{W}_i^{-1} \tag{6.58}$$

$$e^{-\mathbf{\Omega}_i z_i'} = \mathbf{W}_i e^{-\lambda_i z_i'}\mathbf{W}_i^{-1} = \mathbf{W}_i e^{-k_0\lambda_i(z-d_{i-1})}\mathbf{W}_i^{-1} \tag{6.59}$$

$$e^{-\lambda_i z_i'} = \begin{bmatrix} \lambda_{i,1} & & 0 \\ & \ddots & \\ 0 & & \lambda_{i,L} \end{bmatrix} \tag{6.60}$$

These are substituted into Eq. (6.57) and $\mathbf{W}^{-1}$ is absorbed into the proportionality constants now written as $\mathbf{C}_i^+$ and $\mathbf{C}_i^-$.

$$\begin{bmatrix} \mathbf{S}_i^x(z) \\ \mathbf{S}_i^y(z) \end{bmatrix} = \mathbf{W}_i e^{-k_0\lambda_i(z-Z_{i-1})}\mathbf{C}_i^+ + \mathbf{W}_i e^{k_0\lambda_i(z-Z_i)}\mathbf{C}_i^- \tag{6.61}$$

The solution for the magnetic fields can be written in a similar form.

$$\begin{bmatrix} \mathbf{U}_i^x(z) \\ \mathbf{U}_i^y(z) \end{bmatrix} = -\mathbf{V}_i e^{-k_0\lambda_i(z-Z_{i-1})}\mathbf{C}_i^+ + \mathbf{V}_i e^{k_0\lambda_i(z-Z_i)}\mathbf{C}_i^-$$ (6.62)

To compute $\mathbf{V}_i$, Eq. (6.62) is differentiated with respect to z'.

$$\frac{d}{dz'}\begin{bmatrix} \mathbf{U}_i^x(z) \\ \mathbf{U}_i^y(z) \end{bmatrix} = \mathbf{V}_i\lambda_i e^{-k_0\lambda_i(z-Z_{i-1})}\mathbf{C}_i^+ + \mathbf{V}_i\lambda_i e^{k_0\lambda_i(z-Z_i)}\mathbf{C}_i^-$$ (6.63)

Comparing this to Eq. (6.51) and Eq. (6.61) shows that

$$\mathbf{V}_i = \mathbf{Q}_i\mathbf{W}_i\lambda_i^{-1}$$ (6.64)

Combining Eq. (6.61) with Eq. (6.62) leads to the analytical solution of the fields. To compute the proportionality constants, boundary conditions must be solved by matching tangential fields at each interface.

$$\begin{bmatrix} \mathbf{S}_i^x(z) \\ \mathbf{S}_i^y(z) \\ \mathbf{U}_i^x(z) \\ \mathbf{U}_i^y(z) \end{bmatrix} = \begin{bmatrix} \mathbf{W}_i & \mathbf{W}_i \\ -\mathbf{V}_i & \mathbf{V}_i \end{bmatrix}\begin{bmatrix} e^{-k_0\lambda_i(z-Z_{i-1})} & \mathbf{00} \\ \mathbf{00} & e^{k_0\lambda_i(z-Z_i)} \end{bmatrix}\begin{bmatrix} \mathbf{C}_i^+ \\ \mathbf{C}_i^- \end{bmatrix}$$ (6.65)

6.2.10.   Field Expressions Outside Grating Stack

Before the boundary conditions may be solved, fields outside the grating stack must be defined. It will be assumed a single plane wave is incident on the grating stack from the reflected side.

$$\vec{E}_{inc}(x, y, z) = \vec{P}\exp\left(-j\vec{k}_{inc}\bullet\vec{r}\right)$$ (6.66)

The total electric field on the reflected side is the incident wave plus the reflected field.

$$\vec{E}_I(x, y, z) = \vec{E}_{inc}(x, y, z) + \sum_{m=-\infty}^{+\infty}\sum_{n=-\infty}^{+\infty}\vec{R}_{m,n}\exp\left[-j\left(k_{m,n}^x x + k_{m,n}^y y - k_{I,m,n}^z z\right)\right]$$ (6.67)

The total electric field on the transmitted side is simply

$$\vec{E}_{II}(x,y,z) = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \vec{T}_{m,n} \exp\left\{-j\left[k_{m,n}^{x}x + k_{m,n}^{y}y + k_{I,m,n}^{z}(z-d_{L})\right]\right\} \qquad (6.68)$$

In matrix form, the above equations can be written as

$$\mathbf{S}_{inc}^{x}(z') = p_{x}\exp\left(-jk_{inc}^{z}z'\right)\boldsymbol{\delta}_{m,n,0} \qquad (6.69)$$

$$\mathbf{S}_{inc}^{y}(z') = p_{y}\exp\left(-jk_{inc}^{z}z'\right)\boldsymbol{\delta}_{m,n,0} \qquad (6.70)$$

$$\mathbf{S}_{inc}^{z}(z') = p_{z}\exp\left(-jk_{inc}^{z}z'\right)\boldsymbol{\delta}_{m,n,0} \qquad (6.71)$$

$$\mathbf{S}_{I}^{x}(z') = p_{x}\exp\left(-jk_{inc}^{z}z'\right)\boldsymbol{\delta}_{m,n,0} + \exp\left(jK_{I}^{z}z'\right)\mathbf{R}_{x} \qquad (6.72)$$

$$\mathbf{S}_{I}^{y}(z') = p_{y}\exp\left(-jk_{inc}^{z}z'\right)\boldsymbol{\delta}_{m,n,0} + \exp\left(jK_{I}^{z}z'\right)\mathbf{R}_{y} \qquad (6.73)$$

$$\mathbf{S}_{I}^{z}(z') = p_{z}\exp\left(-jk_{inc}^{z}z'\right)\boldsymbol{\delta}_{m,n,0} + \exp\left(jK_{I}^{z}z'\right)\mathbf{R}_{z} \qquad (6.74)$$

$$\mathbf{S}_{II}^{x}(z) = \exp\left[-jk_{0}\mathbf{K}_{I}^{z}(z-Z_{L})\right]\mathbf{T}_{x} \qquad (6.75)$$

$$\mathbf{S}_{II}^{y}(z) = \exp\left[-jk_{0}\mathbf{K}_{I}^{z}(z-Z_{L})\right]\mathbf{T}_{y} \qquad (6.76)$$

$$\mathbf{S}_{II}^{z}(z) = \exp\left[-jk_{0}\mathbf{K}_{I}^{z}(z-Z_{L})\right]\mathbf{T}_{z} \qquad (6.77)$$

The column vector $\boldsymbol{\delta}_{m,n,0}$ contains all zeros except the element at $m=n=0$ which is set to 1. This corresponds to the zero-order being the incident and refracted waves. Outside the first ($z=0$) and last ($z=Z_{L}$) interfaces, the total electric field is

$$\begin{bmatrix} \mathbf{S}_{I}^{x}(0) \\ \mathbf{S}_{I}^{y}(0) \end{bmatrix} = \begin{bmatrix} p_{x}\boldsymbol{\delta}_{m,n,0} \\ p_{y}\boldsymbol{\delta}_{m,n,0} \end{bmatrix} + \begin{bmatrix} \mathbf{R}_{x} \\ \mathbf{R}_{y} \end{bmatrix} \qquad (6.78)$$

$$\begin{bmatrix} \mathbf{S}_{II}^{x}(Z_{L}) \\ \mathbf{S}_{II}^{y}(Z_{L}) \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{x} \\ \mathbf{T}_{y} \end{bmatrix} \qquad (6.79)$$

The incident magnetic field can be computed from the incident electric field by substituting Eq. (6.66) into Eqs. (6.6)-(6.7).

$$\tilde{H}_{inc}^{x} = \frac{j}{\mu_I}\left(\frac{k_{inc}^{z}}{k_0}p_y - \frac{k_{inc}^{y}}{k_0}p_z\right)e^{-j\vec{k}_{inc}\cdot\vec{r}} \tag{6.80}$$

$$\tilde{H}_{y} = \frac{j}{\mu_I}\left(\frac{k_{inc}^{x}}{k_0}p_z - \frac{k_{inc}^{z}}{k_0}p_x\right)e^{-j\vec{k}_{inc}\cdot\vec{r}} \tag{6.81}$$

In matrix form, the above equations at $z = 0$ become

$$\begin{bmatrix} \mathbf{U}_{inc}^{x}(0) \\ \mathbf{U}_{inc}^{y}(0) \end{bmatrix} = \frac{j}{\mu_I}\begin{bmatrix} \left(k_{inc}^{z}p_y - k_{inc}^{y}p_z\right)\boldsymbol{\delta}_{m,n,0} \\ \left(k_{inc}^{x}p_z - k_{inc}^{z}p_x\right)\boldsymbol{\delta}_{m,n,0} \end{bmatrix} \tag{6.82}$$

Next, equations describing the reflected magnetic field must be derived. Since this region is homogeneous, Eq. (6.50) reduces to

$$\frac{\partial}{\partial z'}\begin{bmatrix} \mathbf{S}_{I}^{x}(z') \\ \mathbf{S}_{I}^{y}(z') \end{bmatrix} = \frac{1}{\varepsilon_I}\begin{bmatrix} \mathbf{K}_x\mathbf{K}_y & \mu_I\varepsilon_I - \mathbf{K}_x^{2} \\ \mathbf{K}_y^{2} - \mu_I\varepsilon_I & -\mathbf{K}_x\mathbf{K}_y \end{bmatrix}\begin{bmatrix} \mathbf{U}_{I}^{x}(z') \\ \mathbf{U}_{I}^{y}(z') \end{bmatrix} \tag{6.83}$$

Substituting the reflected component of Eqs. (6.72)-(6.73) into this equation yields

$$\begin{bmatrix} j\mathbf{K}_{I}^{z}\exp\left(j\mathbf{K}_{I}^{z}z'\right)\mathbf{R}_x \\ j\mathbf{K}_{I}^{z}\exp\left(j\mathbf{K}_{I}^{z}z'\right)\mathbf{R}_y \end{bmatrix} = \frac{1}{\varepsilon_I}\begin{bmatrix} \mathbf{K}_x\mathbf{K}_y & \mu_I\varepsilon_I - \mathbf{K}_x^{2} \\ \mathbf{K}_y^{2} - \mu_I\varepsilon_I & -\mathbf{K}_x\mathbf{K}_y \end{bmatrix}\begin{bmatrix} \mathbf{U}_{I,ref}^{x}(z') \\ \mathbf{U}_{I,ref}^{y}(z') \end{bmatrix} \tag{6.84}$$

Evaluating this at $z' = 0$ and solving for the magnetic field coefficients leads to

$$\begin{bmatrix} \mathbf{U}_{I,ref}^{x}(0) \\ \mathbf{U}_{I,ref}^{y}(0) \end{bmatrix} = \frac{j}{\mu_I}\begin{bmatrix} -\dfrac{\mathbf{K}_x\mathbf{K}_y}{\mathbf{K}_I^{z}} & -\dfrac{\mathbf{K}_y^{2} + \mathbf{K}_{I,z}^{2}}{\mathbf{K}_I^{z}} \\ \dfrac{\mathbf{K}_x^{2} + \mathbf{K}_{I,z}^{2}}{\mathbf{K}_I^{z}} & \dfrac{\mathbf{K}_x\mathbf{K}_y}{\mathbf{K}_I^{z}} \end{bmatrix}\begin{bmatrix} \mathbf{R}_x \\ \mathbf{R}_y \end{bmatrix} \tag{6.85}$$

The total magnetic field in region I at the first interface is the sum of Eq. (6.82) and (6.85).

$$
\begin{bmatrix} \mathbf{U}_I^x(0) \\ \mathbf{U}_I^y(0) \end{bmatrix} = \frac{j}{\mu_I} \begin{bmatrix} \left(k_{\text{inc}}^z P_y - k_{\text{inc}}^y P_z \right) \boldsymbol{\delta}_{m,n,0} \\ \left(k_{\text{inc}}^x P_z - k_{\text{inc}}^z P_x \right) \boldsymbol{\delta}_{m,n,0} \end{bmatrix} + \frac{j}{\mu_I} \begin{bmatrix} -\dfrac{\mathbf{K_x K_y}}{\mathbf{K}_I^z} & -\dfrac{\mathbf{K}_y^2 + \mathbf{K}_{I,z}^2}{\mathbf{K}_I^z} \\ \dfrac{\mathbf{K}_x^2 + \mathbf{K}_{I,z}^2}{\mathbf{K}_I^z} & \dfrac{\mathbf{K_x K_y}}{\mathbf{K}_I^z} \end{bmatrix} \begin{bmatrix} \mathbf{R}_x \\ \mathbf{R}_y \end{bmatrix}
\tag{6.86}
$$

The magnetic field in region II is found by substituting Eqs. (6.75)-(6.77) into Eq. (6.83) written in terms of region II properties.

$$
\begin{bmatrix} \mathbf{K}_{II}^z \exp\left[jk_0 \mathbf{K}_{II}^z (z - Z_L) \right] \mathbf{T}_x \\ \mathbf{K}_{II}^z \exp\left[jk_0 \mathbf{K}_{II}^z (z - Z_L) \right] \mathbf{T}_y \end{bmatrix} = \frac{j}{\varepsilon_{II}} \begin{bmatrix} \mathbf{K_x K_y} & \mu_{II}\varepsilon_{II} - \mathbf{K}_x^2 \\ \mathbf{K}_y^2 - \mu_{II}\varepsilon_{II} & -\mathbf{K_x K_y} \end{bmatrix} \begin{bmatrix} \mathbf{U}_{II}^x(z') \\ \mathbf{U}_{II}^y(z') \end{bmatrix}
\tag{6.87}
$$

Evaluating this at $z = Z_L$ yields

$$
\begin{bmatrix} \mathbf{U}_{II}^x(Z_L) \\ \mathbf{U}_{II}^y(Z_L) \end{bmatrix} = \frac{j}{\mu_{II}} \begin{bmatrix} \dfrac{\mathbf{K_x K_y}}{\mathbf{K}_{II}^z} & \dfrac{\mathbf{K}_y^2 + \mathbf{K}_{II,z}^2}{\mathbf{K}_{II}^z} \\ -\dfrac{\mathbf{K}_x^2 + \mathbf{K}_{II,z}^2}{\mathbf{K}_{II}^z} & -\dfrac{\mathbf{K_x K_y}}{\mathbf{K}_{II}^z} \end{bmatrix} \begin{bmatrix} \mathbf{T}_x \\ \mathbf{T}_y \end{bmatrix}
\tag{6.88}
$$

6.2.11. Boundary Conditions

The fields inside the first grating layer at $z = 0$ are computed from Eq. (6.65).

$$
\begin{bmatrix} \mathbf{S}_1^x(0) \\ \mathbf{S}_1^y(0) \\ \mathbf{U}_1^x(0) \\ \mathbf{U}_1^y(0) \end{bmatrix} = \begin{bmatrix} \mathbf{W}_1 & \mathbf{W}_1 \\ -\mathbf{V}_1 & \mathbf{V}_1 \end{bmatrix} \begin{bmatrix} \mathbf{II} & \mathbf{00} \\ \mathbf{00} & e^{-k_0\lambda_1 d_1} \end{bmatrix} \begin{bmatrix} \mathbf{C}_1^+ \\ \mathbf{C}_1^- \end{bmatrix}
\tag{6.89}
$$

Boundary conditions at the first interface require Eq. (6.78) and Eq. (6.86) to equal Eq. (6.89).

$$
\begin{bmatrix} p_x \boldsymbol{\delta}_{m,n,0} \\ p_y \boldsymbol{\delta}_{m,n,0} \\ \dfrac{j}{\mu_I}\left(k_{\text{inc}}^z P_y - k_{\text{inc}}^y P_z \right)\boldsymbol{\delta}_{m,n,0} \\ \dfrac{j}{\mu_I}\left(k_{\text{inc}}^x P_z - k_{\text{inc}}^z P_x \right)\boldsymbol{\delta}_{m,n,0} \end{bmatrix} + \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \\ -\dfrac{j}{\mu_I}\dfrac{\mathbf{K_x K_y}}{\mathbf{K}_I^z} & -\dfrac{j}{\mu_I}\dfrac{\mathbf{K}_y^2 + \mathbf{K}_{I,z}^2}{\mathbf{K}_I^z} \\ \dfrac{j}{\mu_I}\dfrac{\mathbf{K}_x^2 + \mathbf{K}_{I,z}^2}{\mathbf{K}_I^z} & \dfrac{j}{\mu_I}\dfrac{\mathbf{K_x K_y}}{\mathbf{K}_I^z} \end{bmatrix} \begin{bmatrix} \mathbf{R}_x \\ \mathbf{R}_y \end{bmatrix} = \begin{bmatrix} \mathbf{W}_1 & \mathbf{W}_1 \\ -\mathbf{V}_1 & \mathbf{V}_1 \end{bmatrix} \begin{bmatrix} \mathbf{II} & \mathbf{00} \\ \mathbf{00} & e^{-k_0\lambda_1 d_1} \end{bmatrix} \begin{bmatrix} \mathbf{C}_1^+ \\ \mathbf{C}_1^- \end{bmatrix}
\tag{6.90}
$$

Fields inside the first grating layer at the second interface are computed from Eq. (6.65).

$$
\begin{bmatrix} \mathbf{S}_1^x(Z_1) \\ \mathbf{S}_1^y(Z_1) \\ \mathbf{U}_1^x(Z_1) \\ \mathbf{U}_1^y(Z_1) \end{bmatrix} = \begin{bmatrix} \mathbf{W}_1 & \mathbf{W}_1 \\ -\mathbf{V}_1 & \mathbf{V}_1 \end{bmatrix} \begin{bmatrix} e^{-k_0\lambda_1 d_1} & \mathbf{00} \\ \mathbf{00} & \mathbf{II} \end{bmatrix} \begin{bmatrix} \mathbf{C}_1^+ \\ \mathbf{C}_1^- \end{bmatrix}
\tag{6.91}
$$

Similarly, fields inside the second grating layer at the second interface are

$$
\begin{bmatrix} \mathbf{S}_2^x(Z_1) \\ \mathbf{S}_2^y(Z_1) \\ \mathbf{U}_2^x(Z_1) \\ \mathbf{U}_2^y(Z_1) \end{bmatrix} = \begin{bmatrix} \mathbf{W}_2 & \mathbf{W}_2 \\ -\mathbf{V}_2 & \mathbf{V}_2 \end{bmatrix} \begin{bmatrix} \mathbf{II} & \mathbf{00} \\ \mathbf{00} & e^{-k_0\lambda_2 d_2} \end{bmatrix} \begin{bmatrix} \mathbf{C}_2^+ \\ \mathbf{C}_2^- \end{bmatrix}
\tag{6.92}
$$

Boundary conditions require Eq. (6.91) to equal (6.92).

$$
\begin{bmatrix} \mathbf{W}_1 & \mathbf{W}_1 \\ -\mathbf{V}_1 & \mathbf{V}_1 \end{bmatrix} \begin{bmatrix} e^{-k_0\lambda_1 d_1} & \mathbf{00} \\ \mathbf{00} & \mathbf{II} \end{bmatrix} \begin{bmatrix} \mathbf{C}_1^+ \\ \mathbf{C}_1^- \end{bmatrix} = \begin{bmatrix} \mathbf{W}_2 & \mathbf{W}_2 \\ -\mathbf{V}_2 & \mathbf{V}_2 \end{bmatrix} \begin{bmatrix} \mathbf{II} & \mathbf{00} \\ \mathbf{00} & e^{-k_0\lambda_2 d_2} \end{bmatrix} \begin{bmatrix} \mathbf{C}_2^+ \\ \mathbf{C}_2^- \end{bmatrix}
\tag{6.93}
$$

This equation can be generalized to match fields at any intermediate interface.

$$
\begin{bmatrix} \mathbf{W}_i & \mathbf{W}_i \\ -\mathbf{V}_i & \mathbf{V}_i \end{bmatrix} \begin{bmatrix} e^{-k_0\lambda_i d_i} & \mathbf{00} \\ \mathbf{00} & \mathbf{II} \end{bmatrix} \begin{bmatrix} \mathbf{C}_i^+ \\ \mathbf{C}_i^- \end{bmatrix} = \begin{bmatrix} \mathbf{W}_{i+1} & \mathbf{W}_{i+1} \\ -\mathbf{V}_{i+1} & \mathbf{V}_{i+1} \end{bmatrix} \begin{bmatrix} \mathbf{II} & \mathbf{00} \\ \mathbf{00} & e^{-k_0\lambda_{i+1} d_{i+1}} \end{bmatrix} \begin{bmatrix} \mathbf{C}_{i+1}^+ \\ \mathbf{C}_{i+1}^- \end{bmatrix}
\tag{6.94}
$$

Fields inside the last grating layer at the last interface are computed from Eq. (6.65).

$$
\begin{bmatrix} \mathbf{S}_L^x(Z_L) \\ \mathbf{S}_L^y(Z_L) \\ \mathbf{U}_L^x(Z_L) \\ \mathbf{U}_L^y(Z_L) \end{bmatrix} = \begin{bmatrix} \mathbf{W}_L & \mathbf{W}_L \\ -\mathbf{V}_L & \mathbf{V}_L \end{bmatrix} \begin{bmatrix} e^{-k_0\lambda_L d_L} & \mathbf{00} \\ \mathbf{00} & \mathbf{II} \end{bmatrix} \begin{bmatrix} \mathbf{C}_L^+ \\ \mathbf{C}_L^- \end{bmatrix}
\tag{6.95}
$$

Boundary conditions at the last interface require Eq. (6.95) to equal Eq. (6.79) combined with

Eq. (6.88).

$$\begin{bmatrix} \mathbf{W}_L & \mathbf{W}_L \\ -\mathbf{V}_L & \mathbf{V}_L \end{bmatrix} \begin{bmatrix} e^{-k_0 \lambda_L d_L} & \mathbf{00} \\ \mathbf{00} & \mathbf{II} \end{bmatrix} \begin{bmatrix} \mathbf{C}_L^+ \\ \mathbf{C}_L^- \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \\ \dfrac{j}{\mu_{II}} \dfrac{\mathbf{K_x K_y}}{\mathbf{K}_{II}^z} & \dfrac{j}{\mu_{II}} \dfrac{\mathbf{K}_y^2 + \mathbf{K}_{II,z}^2}{\mathbf{K}_{II}^z} \\ -\dfrac{j}{\mu_{II}} \dfrac{\mathbf{K}_x^2 + \mathbf{K}_{II,z}^2}{\mathbf{K}_{II}^z} & -\dfrac{j}{\mu_{II}} \dfrac{\mathbf{K_x K_y}}{\mathbf{K}_{II}^z} \end{bmatrix} \begin{bmatrix} \mathbf{T}_x \\ \mathbf{T}_y \end{bmatrix} \qquad (6.96)$$

    The boundary conditions can be written in a more compact form by defining the

following matrices. **A** relates to the fields on the reflected side, while **B** relates to the fields on

the transmitted side. **S** describes the source.

$$\mathbf{A} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \\ -\dfrac{j}{\mu_I} \dfrac{\mathbf{K_x K_y}}{\mathbf{K}_I^z} & -\dfrac{j}{\mu_I} \dfrac{\mathbf{K}_y^2 + \mathbf{K}_{I,z}^2}{\mathbf{K}_I^z} \\ \dfrac{j}{\mu_I} \dfrac{\mathbf{K}_x^2 + \mathbf{K}_{I,z}^2}{\mathbf{K}_I^z} & \dfrac{j}{\mu_I} \dfrac{\mathbf{K_x K_y}}{\mathbf{K}_I^z} \end{bmatrix} \qquad (6.97)$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \\ \dfrac{j}{\mu_{II}} \dfrac{\mathbf{K_x K_y}}{\mathbf{K}_{II}^z} & \dfrac{j}{\mu_{II}} \dfrac{\mathbf{K}_y^2 + \mathbf{K}_{II,z}^2}{\mathbf{K}_{II}^z} \\ -\dfrac{j}{\mu_{II}} \dfrac{\mathbf{K}_x^2 + \mathbf{K}_{II,z}^2}{\mathbf{K}_{II}^z} & -\dfrac{j}{\mu_{II}} \dfrac{\mathbf{K_x K_y}}{\mathbf{K}_{II}^z} \end{bmatrix} \qquad (6.98)$$

$$\mathbf{S} = \begin{bmatrix} p_x \delta_{m,n,0} \\ p_y \delta_{m,n,0} \\ \dfrac{j}{\mu_I} \left(k_{\text{inc}}^z p_y - k_{\text{inc}}^y p_z \right) \delta_{m,n,0} \\ \dfrac{j}{\mu_I} \left(k_{\text{inc}}^x p_z - k_{\text{inc}}^z p_x \right) \delta_{m,n,0} \end{bmatrix} \qquad (6.99)$$

For each grating layer, the following matrices can be defined.

$$\mathbf{F}_i = \begin{bmatrix} \mathbf{W}_i & \mathbf{W}_i \\ -\mathbf{V}_i & \mathbf{V}_i \end{bmatrix} \qquad \mathbf{X}_i = e^{-k_0 \lambda_i d_i} \tag{6.100}$$

Using the above definitions, the boundary conditions are as follows:

First Interface

$$\mathbf{S} + \mathbf{A}\mathbf{R} = \mathbf{F}_1 \begin{bmatrix} \mathbf{II} & \mathbf{00} \\ \mathbf{00} & \mathbf{X}_1 \end{bmatrix} \mathbf{C}_1 \tag{6.101}$$

Intermediate Interfaces

$$\mathbf{F}_i \begin{bmatrix} \mathbf{X}_i & \mathbf{00} \\ \mathbf{00} & \mathbf{II} \end{bmatrix} \mathbf{C}_i = \mathbf{F}_{i+1} \begin{bmatrix} \mathbf{II} & \mathbf{00} \\ \mathbf{00} & \mathbf{X}_{i+1} \end{bmatrix} \mathbf{C}_{i+1} \tag{6.102}$$

Last Interface

$$\mathbf{F}_L \begin{bmatrix} \mathbf{X}_L & \mathbf{00} \\ \mathbf{00} & \mathbf{II} \end{bmatrix} \mathbf{C}_L = \mathbf{B}\mathbf{T} \tag{6.103}$$

6.2.12. *Enhanced Transmittance Matrix Method*

To solve the boundary conditions, proportionality constants may be circumvented using the enhanced transmittance matrix method [52]. This method avoids computing the inverse of $\mathbf{X}_i$ which could lead to poorly conditioned matrices for large positive eigen-values.

The algorithm for enhanced transmittance matrices usually starts at the last interface and works the boundary conditions backward to the first interface. First, Eq. (6.103) is solved for $\mathbf{C}_L$.

$$\mathbf{C}_L = \begin{bmatrix} \mathbf{X}_L^{-1} & \mathbf{00} \\ \mathbf{00} & \mathbf{II} \end{bmatrix} \mathbf{F}_L^{-1} \mathbf{B}\mathbf{T} \tag{6.104}$$

This can be written in the following form.

$$\mathbf{C}_L = \begin{bmatrix} \mathbf{X}_L^{-1} & \mathbf{00} \\ \mathbf{00} & \mathbf{II} \end{bmatrix} \begin{bmatrix} \mathbf{a}_L \\ \mathbf{b}_L \end{bmatrix} \mathbf{T} \tag{6.105}$$

$$\begin{bmatrix} \mathbf{a}_L \\ \mathbf{b}_L \end{bmatrix} = \mathbf{F}_L^{-1} \mathbf{B} \tag{6.106}$$

To eliminate the potentially ill conditioned matrix $\mathbf{X}_L^{-1}$, an intermediate parameter $\mathbf{T}_L$ is defined

such that

$$\mathbf{T} = \mathbf{a}_L^{-1} \mathbf{X}_L \mathbf{T}_L \tag{6.107}$$

Substituting this into Eq. (6.105) eliminates $\mathbf{X}_L^{-1}$.

$$\mathbf{C}_L = \begin{bmatrix} \mathbf{II} \\ \mathbf{b}_L \mathbf{a}_L^{-1} \mathbf{X}_L \end{bmatrix} \mathbf{T}_L \tag{6.108}$$

Working backward through the grating stack, the expression for $\mathbf{C}_L$ is substituted into

Eq. (6.102) written for the second to last interface.

$$\mathbf{C}_{L-1} = \begin{bmatrix} \mathbf{X}_{L-1}^{-1} & \mathbf{00} \\ \mathbf{00} & \mathbf{II} \end{bmatrix} \mathbf{F}_{L-1}^{-1} \mathbf{F}_L \begin{bmatrix} \mathbf{II} & \mathbf{00} \\ \mathbf{00} & \mathbf{X}_L \end{bmatrix} \begin{bmatrix} \mathbf{II} \\ \mathbf{b}_L \mathbf{a}_L^{-1} \mathbf{X}_L \end{bmatrix} \mathbf{T}_L \tag{6.109}$$

This is written to be in the same form as Eqs. (6.105)-(6.106).

$$\mathbf{C}_{L-1} = \begin{bmatrix} \mathbf{X}_{L-1}^{-1} & \mathbf{00} \\ \mathbf{00} & \mathbf{II} \end{bmatrix} \begin{bmatrix} \mathbf{a}_{L-1} \\ \mathbf{b}_{L-1} \end{bmatrix} \mathbf{T}_L \tag{6.110}$$

$$\begin{bmatrix} \mathbf{a}_{L-1} \\ \mathbf{b}_{L-1} \end{bmatrix} = \mathbf{F}_{L-1}^{-1} \mathbf{F}_L \begin{bmatrix} \mathbf{II} & \mathbf{00} \\ \mathbf{00} & \mathbf{X}_L \end{bmatrix} \begin{bmatrix} \mathbf{II} \\ \mathbf{b}_L \mathbf{a}_L^{-1} \mathbf{X}_L \end{bmatrix} \tag{6.111}$$

Like before, the potentially ill conditioned matrix $\mathbf{X}_{L-1}^{-1}$ is eliminated by defining $\mathbf{T}_{L-1}$ such that

$$\mathbf{T}_L = \mathbf{a}_{L-1}^{-1} \mathbf{X}_{L-1} \mathbf{T}_{L-1} \tag{6.112}$$

146

Substituting this into Eq. (6.110) eliminates $\mathbf{X}_{L-1}^{-1}$.

$$\mathbf{C}_{L-1} = \begin{bmatrix} \mathbf{X}_{L-1}^{-1}\mathbf{a}_{L-1} \\ \mathbf{b}_{L-1} \end{bmatrix} \mathbf{a}_{L-1}^{-1}\mathbf{X}_{L-1}\mathbf{T}_{L-1} = \begin{bmatrix} \mathbf{II} \\ \mathbf{b}_{L-1}\mathbf{a}_{L-1}^{-1}\mathbf{X}_{L-1} \end{bmatrix} \mathbf{T}_{L-1} \tag{6.113}$$

This equation is again written in the form of Eq. (6.105) and the process repeats through the entire grating stack. When the last interface is reached, the final equation will be

$$\mathbf{S} + \mathbf{AR} = \mathbf{F}_1 \begin{bmatrix} \mathbf{II} & \mathbf{00} \\ \mathbf{00} & \mathbf{X}_1 \end{bmatrix} \begin{bmatrix} \mathbf{II} \\ \mathbf{b}_1\mathbf{a}_1^{-1}\mathbf{X}_1 \end{bmatrix} \mathbf{T}_1 \tag{6.114}$$

This can be solved for $\mathbf{R}$ and $\mathbf{T}_1$ in the following manner.

$$\begin{bmatrix} \mathbf{R} \\ \mathbf{T}_1 \end{bmatrix} = \begin{bmatrix} -\mathbf{A} & \mathbf{B}' \end{bmatrix}^{-1} \mathbf{S} \qquad\qquad \mathbf{B}' = \mathbf{F}_1 \begin{bmatrix} \mathbf{II} \\ \mathbf{X}_1\mathbf{b}_1\mathbf{a}_1^{-1}\mathbf{X}_1 \end{bmatrix} \tag{6.115}$$

From here, all $\mathbf{T}_i$ vectors can be computed from Eq. (6.112) by working forward through the grating stack. Overall transmission through the grating stack is then

$$\mathbf{T} = \mathbf{a}_L^{-1}\mathbf{X}_L\mathbf{a}_{L-1}^{-1}\mathbf{X}_{L-1}\cdots\mathbf{a}_2^{-1}\mathbf{X}_2\mathbf{a}_1^{-1}\mathbf{X}_1\mathbf{T}_1 \tag{6.116}$$

6.2.13. *Computation of Diffraction Efficiencies*

RMS power is defined as

$$\vec{P} = \tfrac{1}{2}\text{Re}\left[\vec{E} \times \vec{H} \right] \tag{6.117}$$

For plane waves in isotropic media, $\vec{E}$ and $\vec{H}$ are perpendicular and power flows in the direction of the wave vector $\vec{k}$. Therefore, $\vec{P}/|\vec{P}| = \vec{k}/|\vec{k}|$ so Eq. (6.117) can be written as

$$\vec{P} = \frac{1}{2}\text{Re}\left[\frac{\vec{k}}{|\vec{k}|}|\vec{E}||\vec{H}| \right] \tag{6.118}$$

Field magnitudes are related through the material impedance as $\left|\vec{E}\right| = \eta\left|\vec{H}\right|$. This is used to eliminate $\vec{H}$ in Eq. (6.118).

$$\vec{P} = \frac{1}{2}\mathrm{Re}\left[\frac{\vec{k}}{\left|\vec{k}\right|}\frac{\left|\vec{E}\right|^2}{\eta}\right] \tag{6.119}$$

It is only the power flow in the z-direction that is of interest because this is what transports energy into and out of the grating stack.

$$P_z = \frac{1}{2}\mathrm{Re}\left[\frac{k_z}{\left|\vec{k}\right|}\frac{\left|\vec{E}\right|^2}{\eta}\right] \tag{6.120}$$

For the $(m,n)^{\text{th}}$ harmonic, this becomes

$$P_{m,n}^z = \frac{1}{2}\mathrm{Re}\left[\frac{k_{m,n}^z}{\left|\vec{k}_{m,n}\right|}\frac{\left|\vec{S}_{m,n}\right|^2}{\eta}\right] \tag{6.121}$$

Diffraction efficiency of the $(m,n)^{\text{th}}$ harmonic is defined as $\mathrm{DE}_{m,n} = P_{m,n}^z \big/ P_{\text{inc}}^z$ which is

$$\mathrm{DE}_{m,n} = \frac{\left|\vec{S}_{m,n}\right|^2}{\left|\vec{S}_{\text{inc}}\right|^2}\mathrm{Re}\left[\frac{k_{m,n}^z}{k_{\text{inc}}^z}\frac{\mu_{r,\text{inc}}}{\mu_r}\right] \tag{6.122}$$

Typically the incident wave is made to have unit amplitude so this equation reduces to

$$\mathrm{DE}_{m,n} = \left|\vec{S}_{m,n}\right|^2\mathrm{Re}\left[\frac{k_{m,n}^z}{k_{\text{inc}}^z}\frac{\mu_{r,\text{inc}}}{\mu_r}\right] \tag{6.123}$$

The reflection and transmission coefficients of the longitudinal fields must be known to compute diffraction efficiency. Since wave polarization is orthogonal to the direction of propagation, the following equations must be valid.

$$\vec{k}_I \bullet \vec{R} = 0 \tag{6.124}$$

$$\vec{k}_{II} \bullet \vec{T} = 0 \tag{6.125}$$

Expanding these equations into vector components leads to

$$k_I^x R_x + k_I^y R_y + k_I^z R_z = 0 \tag{6.126}$$

$$k_{II}^x T_x + k_{II}^y T_y + k_{II}^z T_z = 0 \tag{6.127}$$

Solving these equations for R_z and T_z yields

$$R_z = -\frac{k_I^x}{k_I^z} R_x - \frac{k_I^y}{k_I^z} R_y \tag{6.128}$$

$$T_z = -\frac{k_{II}^x}{k_{II}^z} T_x - \frac{k_{II}^y}{k_{II}^z} T_y \tag{6.129}$$

Using the adopted matrix notation, these equations are

$$\mathbf{R}_z = -\mathbf{K}_{z,I}^{-1} \left(\mathbf{K}_x \mathbf{R}_x + \mathbf{K}_y \mathbf{R}_y \right) \tag{6.130}$$

$$\mathbf{T}_z = -\mathbf{K}_{z,II}^{-1} \left(\mathbf{K}_x \mathbf{T}_x + \mathbf{K}_y \mathbf{T}_y \right) \tag{6.131}$$

Diffraction efficiency is then

$$\mathbf{D_R} = \left| \vec{\mathbf{R}} \right|^2 \operatorname{Re} \left[\frac{\mu_{r,\text{inc}}}{\mu_I} \frac{\mathbf{K}_{z,I}}{k_{\text{inc}}^z} \right] \tag{6.132}$$

$$\mathbf{D_T} = \left| \vec{\mathbf{T}} \right|^2 \operatorname{Re} \left[\frac{\mu_{r,\text{inc}}}{\mu_{II}} \frac{\mathbf{K}_{z,II}}{k_{\text{inc}}^z} \right] \tag{6.133}$$

where

$$\left| \vec{\mathbf{R}} \right|^2 = \left| \mathbf{R}_x \right|^2 + \left| \mathbf{R}_y \right|^2 + \left| \mathbf{R}_z \right|^2 \tag{6.134}$$

$$\left| \vec{\mathbf{T}} \right|^2 = \left| \mathbf{T}_x \right|^2 + \left| \mathbf{T}_y \right|^2 + \left| \mathbf{T}_z \right|^2 \tag{6.135}$$

6.2.14. Calculation of Intermediate Fields

Fields inside an intermediate grating layer can be computed from Eq. (6.65) if the proportionality constants are known. These are not computed using the conventional enhanced transmittance matrix method described above. They may be computed while working forward through the grating layers using Eq. (6.108). For the ith grating layer, this is

$$\begin{bmatrix} \mathbf{C}_i^+ \\ \mathbf{C}_i^- \end{bmatrix} = \begin{bmatrix} \mathbf{II} \\ \mathbf{b}_i \mathbf{a}_i^{-1} \mathbf{X}_i \end{bmatrix} \mathbf{T}_i \tag{6.136}$$

Once these have been computed, the tangential field coefficients in the ith grating layer are computed using Eq. (6.65). This can be written as

$$\begin{bmatrix} \mathbf{S}_i^x(z) \\ \mathbf{S}_i^y(z) \\ \mathbf{U}_i^x(z) \\ \mathbf{U}_i^y(z) \end{bmatrix} = \mathbf{F}_i \mathbf{X}_i(z) \mathbf{C}_i \tag{6.137}$$

where

$$\mathbf{X}_i(z) = \begin{bmatrix} e^{-k_0 \lambda_i (z - Z_{i-1})} & \mathbf{00} \\ \mathbf{00} & e^{k_0 \lambda_i (z - Z_i)} \end{bmatrix} \tag{6.138}$$

Next, the longitudinal field coefficients can be calculated using Eqs. (6.44)-(6.45). These were

$$\mathbf{S}_i^z(z) = j\varepsilon_{i,z}^{-1} \left[\mathbf{K}_y \mathbf{U}_i^x(z) - \mathbf{K}_x \mathbf{U}_i^y(z) \right] \tag{6.139}$$

$$\mathbf{U}_i^z(z) = j\mu_{i,z}^{-1} \left[\mathbf{K}_y \mathbf{S}_i^x(z) - \mathbf{K}_x \mathbf{S}_i^y(z) \right] \tag{6.140}$$

The total field is reconstructed by substituting the amplitude coefficients of the spatial harmonics into the field expansions of Eqs. (6.12)-(6.13). The inverse-FFT function may also be used.

150

6.3. Implementation

At first, it can be unclear how to implement RCWA due to the large number of equations involved in the formulation. Figure 6-4 illustrates how RCWA can be implemented using the enhanced transmittance matrix approach. The block diagram also shows how intermediate fields may be calculated by recording intermediate parameters and reconstructing the field patterns at the end.
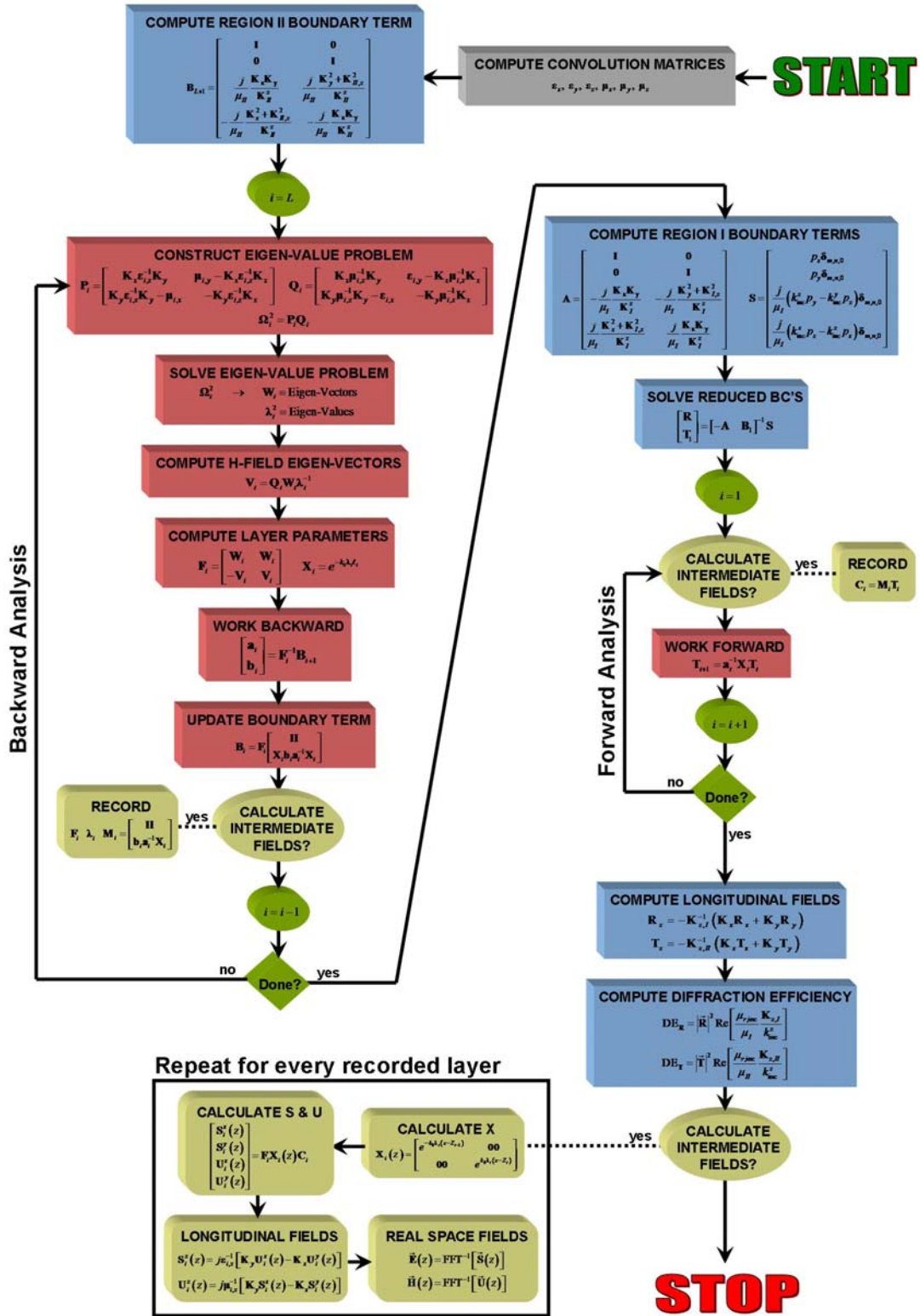
Figure 6-4: Block diagram of RCWA using enhanced transmittance matrix approach

152

CHAPTER 7
STRING METHOD

7.1. Overview

The string method [39, 54, 55] is an incredibly fast and efficient method for simulating the evolution of surfaces that are mostly smooth and continuous. It is simple to implement and can incorporate multiple physical processes. It has been used to model a variety of fabrication processes including etching [55], deposition [55], and autocloning [39].

The main drawback of the method is that it is inherently unstable [139] and therefore restricted to smooth and continuous surfaces [62]. Proven techniques exist to improve stability, but they all modify the equation of motion in unobvious and unrealistic ways [63]. Due to the manner in which the method is stabilized, it can rarely be extended to model three-dimensional problems.

7.2. Algorithm

The string method is surprisingly easy to implement and a block diagram is provided in Figure 7-1. Prior to entering the main loop, the initial surface is described by a series of marker particles placed along the edges. Each particle is described by a position vector $\vec{x}_i$. When N marker particles are used, the complete list of position vectors can be written as

$$\vec{\mathbf{X}} = \begin{bmatrix} \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_N \end{bmatrix} \tag{7.1}$$

The first step in the main loop is to calculate how fast, and in what direction, each particle on the surface must move to best model the physics of the process. This is called the rate function and is essentially a velocity associated with each marker point.

$$\vec{\mathbf{R}} = \begin{bmatrix} \vec{r}_1 & \vec{r}_2 & \cdots & \vec{r}_N \end{bmatrix} \tag{7.2}$$

153

To update position vectors based on the rate function, a suitable time step Δt must be chosen that is small enough to prevent particles from crossing paths. After this is calculated, position is updated according to

$$\vec{\mathbf{X}}_{t+\Delta t} = \vec{\mathbf{X}}_t + \Delta t \vec{\mathbf{R}}_t \qquad (7.3)$$
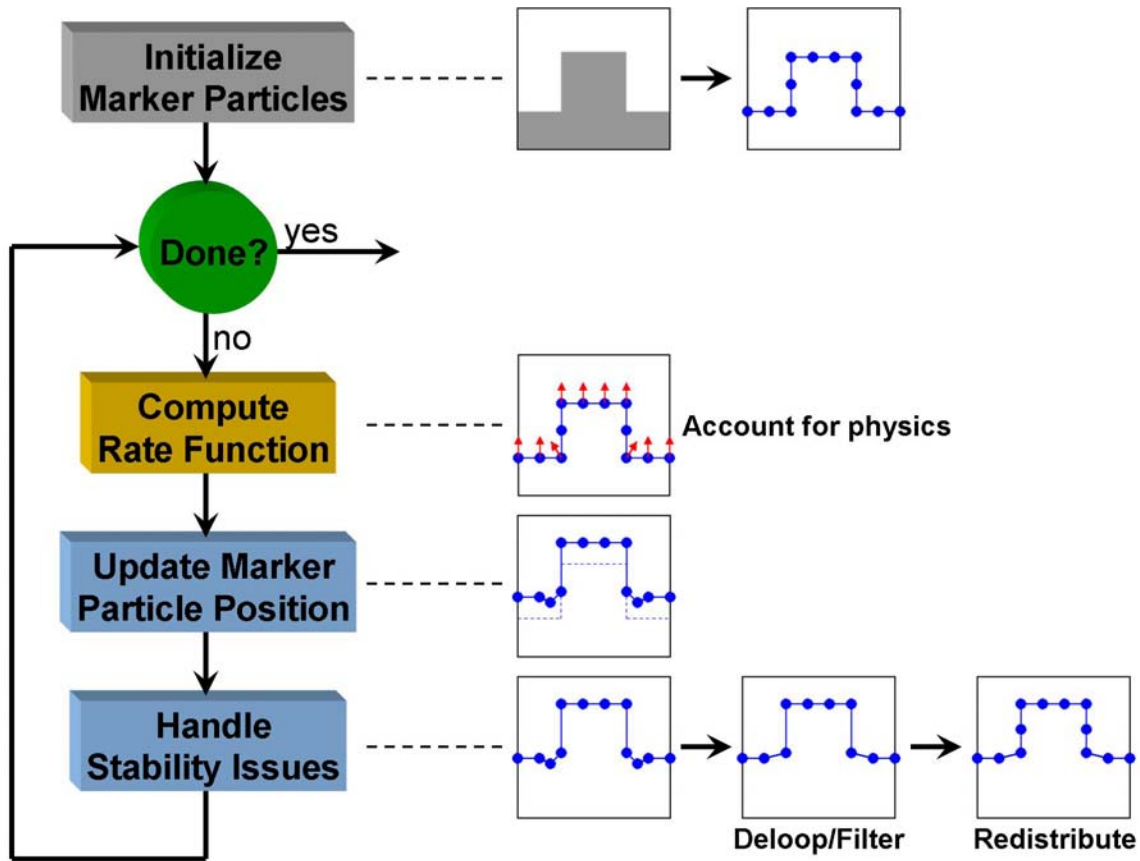
Figure 7-1: Block diagram of string method

As mentioned previously, the string method is inherently unstable for two main reasons. First, as particles converge they may reduce accuracy of the rate function or form unstable loops in the string. To mitigate this, a "smoothing" function can be applied to the rate function and marker points can be removed as they converge. Second, as marker points diverge they can

154

become too sparse to adequately resolve the surface. In this case, additional marker particles must be added to the list. Care must be taken to correctly interpolate where they are to be placed. Performing a filtering function on the position vectors after the update can also help maintain stability.
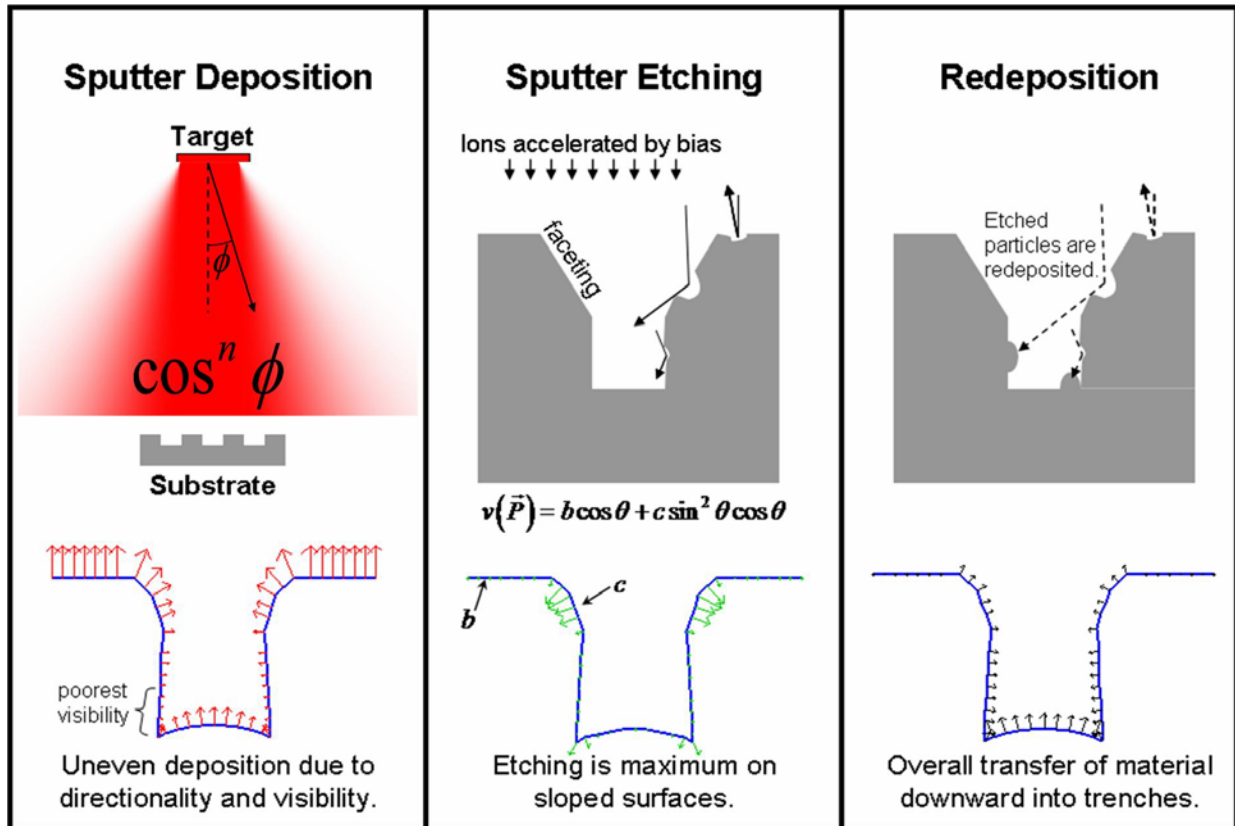
Figure 7-2: Mechanisms leading to autocloning

7.3. Calculation of Rate Function for Autocloning

Autocloning [39, 140-144] is a process where alternating layers of dielectric are deposited onto a corrugated surface. A self-shaping effect causes surface topology to quickly converge to a profile that is maintained over a virtually unlimited number of layers. Three

155

mechanisms are responsible for this phenomenon [39]. These are sputter deposition, sputter etching, and redeposition of etched particles. These mechanisms are illustrated in Figure 7-2 along with "snapshots" from the string method. Arrows indicate velocity of the specific mechanism being illustrated. The overall rate function is the sum of all component velocities.

7.3.1. *Sputter Deposition*

During sputter deposition, neutral particles are deposited from a target to the substrate with an angular dependence of $\cos^n \phi$. Parameter ϕ denotes angle from the vertical axis, and n characterizes the diffusion profile. Isotropic deposition corresponds to $n = 0$ while deposition in a direction strictly normal to the substrate corresponds to $n = \infty$. While most particles diffuse vertically, diffusion at oblique angles leads to deposition on sidewalls.

The deposition rate at any point $\vec{x}_i$ along the surface can be computed as

$$R_{\text{dep}}\left(\vec{x}_i\right) = d \frac{\displaystyle\int_{-\pi/2}^{\pi/2} V_{\text{T}}\left(\vec{x}_i,\phi\right)\cos^n\left(\phi\right)\cos\left(\phi-\theta\right)d\phi}{\displaystyle\int_{-\pi/2}^{\pi/2} \cos^{n+1}\left(\phi\right)d\phi} \tag{7.4}$$

Parameter d describes the magnitude of flux of incident particles, θ is the angle of the surface normal from the vertical axis, and $V_{\text{T}}\left(\vec{x}_i,\phi\right)$ is visibility of the target at angle ϕ from point P. Visibility equals 1 if the target is visible and 0 if it is not. Another way of writing Eq. (7.4) incorporates visibility by limiting the range of angles to those that are visible (i.e. $\phi_1 \le \phi \le \phi_2$).

$$R_{\text{dep}}\left(\vec{x}_i\right) = d \frac{\displaystyle\int_{\phi_1}^{\phi_2} \cos^n\left(\phi\right)\cos\left(\phi-\theta\right)d\phi}{\displaystyle\int_{-\pi/2}^{\pi/2} \cos^{n+1}\left(\phi\right)d\phi} \tag{7.5}$$

156

The angular dependence on deposition leads to horizontal surfaces growing faster that vertical surfaces. Visibility leads to a shadowing effect that produces "kinks" near the bottom of vertical walls.

7.3.2. *Sputter Etching*

Charged rare-gas ions accelerated by the field during bias sputtering leads to etching of the substrate. The etch rate (i.e. negative deposition rate) at point $\vec{x}_i$ along the surface can be computed as

$$R_{\text{etch}}\left(\vec{x}_i\right) = b\cos\theta + c\sin^2\theta\cos\theta \qquad\qquad b,c < 0 \qquad\qquad (7.6)$$

Parameter b denotes sputter etch rate of the horizontal surface. Parameter c characterizes angular selectivity of the mechanism. The function $\sin^2\theta\cos\theta$ has a maximum value around $\theta = 55°$. The peak etch rate at this angle is emphasized with increasing $|c|$. It is this mechanism that leads to regions of the surface profile converging to constant slope near $55°$.

7.3.3. *Redeposition*

Materials that are etched during the sputtering process can be deposited back onto the surface. The path particles take between being etched and redeposited is quite complicated and difficult to model, particularly for deep trenches. The tendency is for particles to move from upper parts of trenches to lower parts. For autocloning, it is reasonable to assume particles come uniformly from surrounding surfaces. This is especially true after the surface profile has converged to V-shaped grooves and θ is constant nearly everywhere.

157

Redeposition rate can be expressed as an integration over the full range of angles from which particles can diffuse. This is

$$R_{\text{redep}}\left(\vec{x}_i\right) = \frac{e}{2} \int_{-\pi/2}^{+\pi/2} V_S\left(\vec{x}_i, \phi\right) \cos\left(\phi - \theta\right) d\phi \tag{7.7}$$

Parameter e characterizes the rate of redeposition. $V_S\left(\vec{x}_i, \phi\right)$ indicates what portion of the surface is visible at angle ϕ for point P. Points at deeper parts of a groove will receive greater redeposition because greater surface area is visible from which to receive particles. It is this mechanism in autocloning that enables lower portions of the V-grooves to keep pace with the top portions despite shadowing effects inhibiting sputter deposition in the lower portions.

7.3.4. Assembling the Rate Function

After rates of the different mechanisms have been calculated, they are added to calculate the overall rate vector. The direction of the rate vector at each point is in the direction of the surface normal $\hat{n}_i$. The overall rate function at point $\vec{x}_i$ is then

$$\vec{R}\left(\vec{x}_i\right) = \left[R_{\text{dep}}\left(\vec{x}_i\right) + R_{\text{etch}}\left(\vec{x}_i\right) + R_{\text{redep}}\left(\vec{x}_i\right)\right]\hat{n}_i \tag{7.8}$$

The surface normals can be calculated as

$$\hat{n}_i = \frac{-\delta_{z,i}\hat{a}_x + \delta_{x,i}\hat{a}_z}{\sqrt{\delta_{x,i}^2 + \delta_{z,i}^2}} \tag{7.9}$$

$$\delta_{x,i} = x_{i+1} - x_{i-1} \qquad\qquad \delta_{z,i} = z_{i+1} - z_{i-1} \tag{7.10}$$

7.4. Benchmark Simulation

In Ref. [39], an autocloning process was developed using amorphous silicon (a-Si) and silicon dioxide (SiO$_2$). Simulation parameters for the string method were calibrated from experimental results are summarized in Table 7-1.

Table 7-1: String method parameters for autocloning

SIMULATION PARAMETERS		
Parameter	a-Si	SiO$_2$
d	1	1
n	0.4	1
b	0	-0.1
c	0	-0.4
e	0.15	0.5

PROCESS CONDITIONS		
Parameter	a-Si	SiO$_2$
Target	poly-Si	fused silica
Sputtering Gas	Ar+H$_2$	Ar+H$_2$
Gas Pressure	3.5 mTorr	2.0 mTorr
RF Power	400 W	400 W
Bias	0	50 W

To verify accuracy of the string method used in this dissertation, numerical results presented in Ref. [39] were duplicated. A comparison is shown in Figure 7-3. Parameters describing the trapezoidal grating profile and apparent alternating layer thicknesses were not provided so they were estimated. Grating period was set to 1 μm, grating depth to 0.4 μm, duty cycle of the trapezoid base was set to 0.5, and duty cycle of the top was set to 0.4. Layer thicknesses were each set to 0.2 μm as prescribed in Ref. [39], but the SiO$_2$ layers appear to be somewhat thinner. This discrepancy was neglected for the benchmark simulation and equal film thicknesses were used.
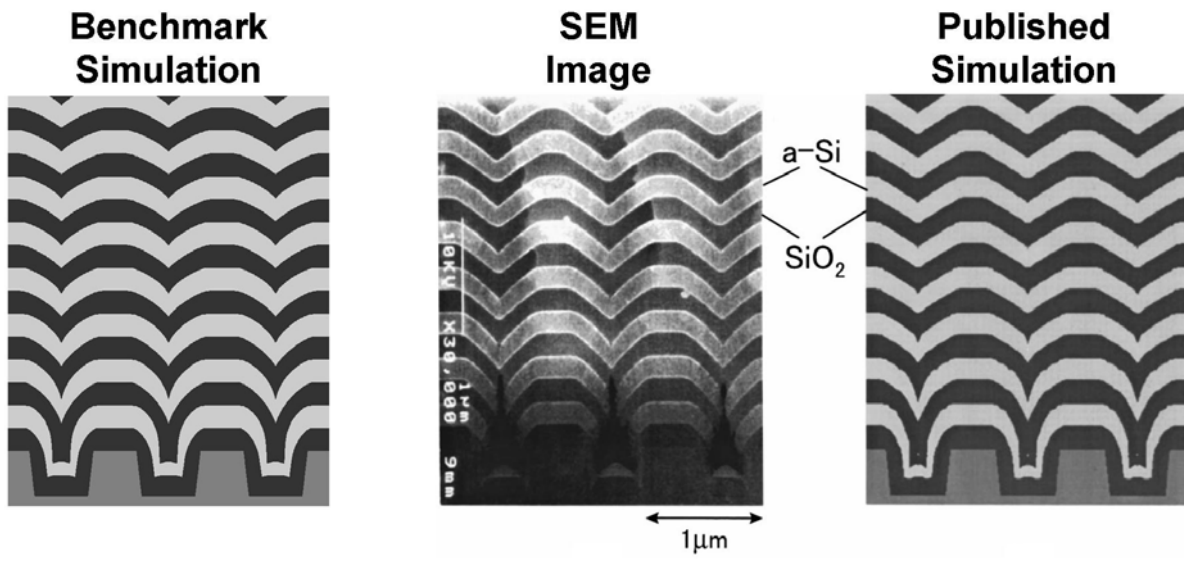
Figure 7-3: Benchmark simulation for string method

CHAPTER 8

LEVEL SET METHOD AND FAST MARCHING METHOD

8.1. Overview

Level set methods (LSM) [61-63, 145-147] are very efficient algorithms for modeling the progression of a surface or interface. It is a rigorous and stable method that is simple to implement. Multiple physical processes operating on the interface may be incorporated at the same time. When surfaces progress in just one direction and rate of progression depends only on position, fast marching methods (FMM) [63, 148, 149] offer even greater speed and efficiency. The methods have been applied to solve a host of applications including micro-fabrication [63, 146-149], computational geometry [63], grid generation [63, 150], image processing and noise removal [151-154], computer vision [63], combustion [139], fluid mechanics [61], path planning [155, 156], and more.

It seems most obvious to consider a moving boundary from a geometry perspective where shapes of structures or positions of interfaces change over time. Therefore, it is tempting to formulate codes to track the motion as a function of time leading to algorithms such as string or cell-volume methods. Level set methods and fast marching methods approach the problem from a completely different perspective in terms of partial differential equations to compute time-of-arrival functions and distance level sets. This enables robust schemes based on well understood methods to be formulated that are more rigorous and easier to adapt to different conditions.

8.2. Mathematical Treatment of Interface Propagation

Forward movement of a surface is usually defined to be in the direction of the outward surface normal. The velocity at each point on the surface is controlled by a rate function R. A negative rate function leads to backward motion of the surface. The rate function may depend on properties of the surface such as curvature, orientation of the surface normal, or visibility to a source. It may also depend on properties not directly related to the surface such as time or position. Perhaps the biggest challenge in formulating a level set method or fast marching method is calculating a rate function that accurately describes what is being modeled. To simulate the developing process, the rate function is simply set to the dissolution rate of the material being developed.

In the most general setting, a surface may progress backward or forward. Some portions of a front may move forward while other portions move backward. Geometry and motion of the surface can be described by a level set function $\psi(\vec{r}, t)$ that is a function of position $\vec{r}$ and time t. The shape of a surface at time t_0 is determined by constructing the isometric surface defined by $\psi(\vec{r}, t_0) = 0$. The level set equation is an initial value problem in the form of a partial differential equation that must be solved for the level set function. This is

$$\frac{\partial \psi}{\partial t} + |\nabla \psi| R = 0 \qquad \text{given } \psi(\vec{r}, 0) \qquad (8.1)$$

When the rate function only depends on position and/or $\partial \psi / \partial t$, Eq. (8.1) reduces to the well known "Hamilton-Jacobi" equation [63].

Figure 8-1 shows a block diagram of how LSM is implemented along with example level set data during simulation of a deposition process. Computing the rate function is perhaps the biggest and most significant step. Physics of the process has to be considered and a rate function

calculated in a manner that best represents motion of the surface. The surface is always defined as the isometric surface at $\psi = 0$.
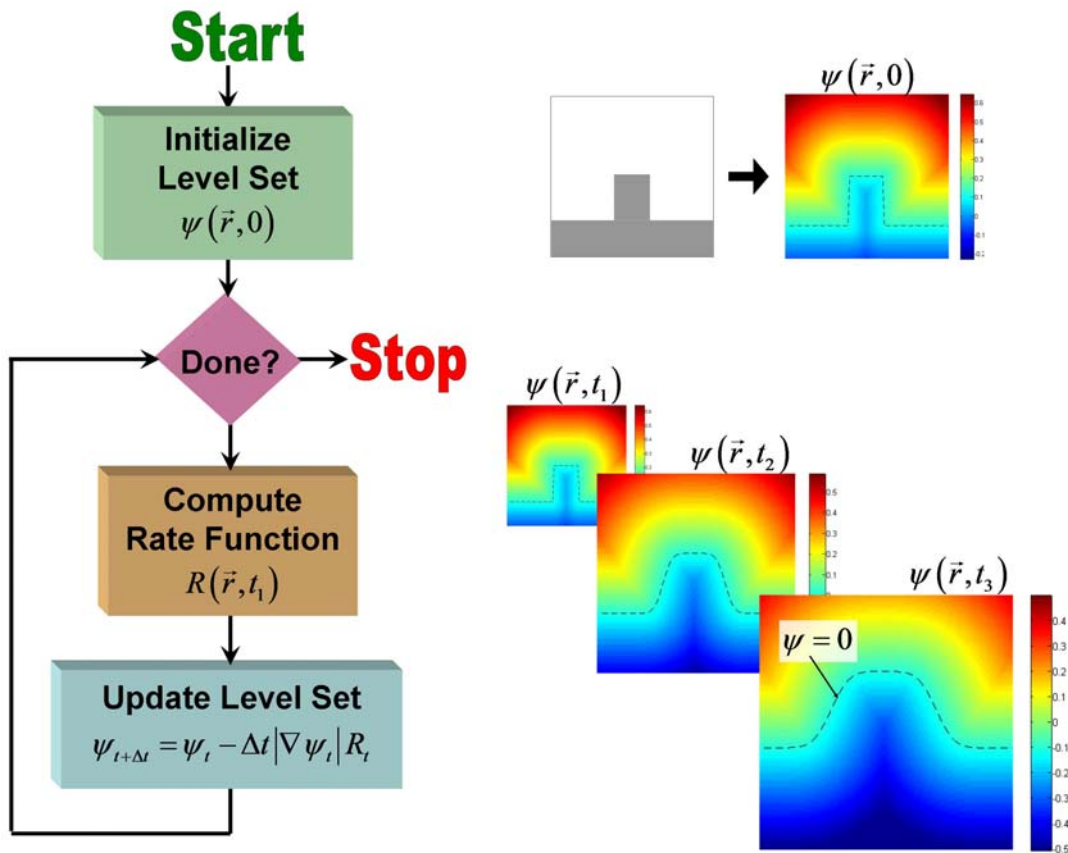
Figure 8-1: Block diagram of level set method

The level set formulation can be dramatically simplified if backward motion does not exist or can be neglected. This requires the rate function to always be positive. In this case, it is easiest to describe the problem in terms of a time-of-arrival function $T(\vec{r})$. This is only a function of position and quantifies the time that the surface arrives at position $\vec{r}$. The shape of

the surface at time t_0 is determined by constructing an isometric surface defined by $T(\vec{r}) = t_0$.

The governing equation is a boundary value problem of the form

$$|\nabla T| R = 1 \qquad \text{initial surface } T(\vec{r}) = 0 \tag{8.2}$$

When the rate function only depends on position, Eq. (8.2) is the well known "Eikonal" equation and fast marching methods may be used to solve it. The remainder of this chapter is devoted to the formulation and implementation of the FMM.

8.3. Formulation of Fast Marching Method

8.3.1. *Upwind Finite-Difference Scheme*

Spatial derivatives in Eq. (8.2) are written explicitly by squaring the equation and moving the rate function to the right-hand side of the equation.

$$\left(\frac{\partial T}{\partial x}\right)^2 + \left(\frac{\partial T}{\partial y}\right)^2 + \left(\frac{\partial T}{\partial z}\right)^2 = \frac{1}{R^2} \tag{8.3}$$

Derivatives in Eq. (8.3) can be approximated using finite-differences by fitting the time function T to a discrete grid. The main assumption leading to this equation was that the surface progressed in a single direction (i.e. $R \geq 0$). In this framework, time values are only known on the "upwind" side of the front. Motion is defined to be in the downwind direction. Time values at grid points along the interface can only be updated based on time values that are upwind. This must be taken into account when approximating the derivatives.

The term $\partial T/\partial x$ can be approximated by forward or backward finite-differences.

$$\text{Backward Finite-Difference: } D_{x-}T_{i,j,k} = \frac{T_{i,j,k} - T_{i-1,j,k}}{\Delta x} \tag{8.4}$$

$$\text{Forward Finite-Difference: } D_{x+}T_{i,j,k} = \frac{T_{i+1,j,k} - T_{i,j,k}}{\Delta x} \tag{8.5}$$

164

In the most general setting, arrival of the interface may come from any direction so the derivative at any grid point is dictated by the side from which the front is approaching most quickly. Therefore, the derivative should be set to the maximum computed from Eqs. (8.4)-(8.5). When a time value at a grid point is less than any of its neighboring points, all finite-differences will be negative and should be neglected. To handle this condition numerically, the upwind finite-difference can be computed as

$$\frac{\partial T_{i,j,k}}{\partial x} \cong \max\left[+D_{x_-}T_{i,j,k}, \ -D_{x_+}T_{i,j,k}, \ 0\right] \tag{8.6}$$

Applying the same procedure to the other spatial derivatives $\partial T/\partial y$ and $\partial T/\partial z$, Eq. (8.3) can be written completely in terms of upwind finite-differences.

$$\left(\max\left[D_{x_-}T, \ -D_{x_+}T, \ 0\right]\right)^2 + \left(\max\left[D_{y_-}T, \ -D_{y_+}T, \ 0\right]\right)^2 + \left(\max\left[D_{z_-}T, \ -D_{z_+}T, \ 0\right]\right)^2 = \frac{1}{R^2} \tag{8.7}$$

8.3.2. *Update Equation for Time*

Equation (8.7) can be viewed as a quadratic equation to be solved for $T_{i,j,k}$. To do this, $T_{i,j,k}$ must be pulled out of the 'max' functions. This can be accomplished by rewriting the upwind finite-difference equations as

$$\frac{\partial T_{i,j,k}}{\partial x} \cong \frac{T_{i,j,k}^{\text{new}} - \min\left(T_{i-1,j,k}, T_{i+1,j,k}, T_{i,j,k}^{\text{old}}\right)}{\Delta x} \tag{8.8}$$

Using this form of finite-difference, Eq. (8.7) can be written as

$$\left(\frac{T_{i,j,k}^{\text{new}} - m_x}{\Delta x}\right)^2 + \left(\frac{T_{i,j,k}^{\text{new}} - m_y}{\Delta y}\right)^2 + \left(\frac{T_{i,j,k}^{\text{new}} - m_z}{\Delta z}\right)^2 = \frac{1}{R_{i,j,k}^2} \tag{8.9}$$

$$m_x = \min\left(T_{i-1,j,k}, T_{i+1,j,k}, T_{i,j,k}^{\text{old}}\right) \tag{8.10}$$

$$m_y = \min\left(T_{i,j-1,k}, T_{i,j+1,k}, T_{i,j,k}^{\text{old}}\right) \tag{8.11}$$

$$m_z = \min\left(T_{i,j,k-1}, T_{i,j,k+1}, T_{i,j,k}^{\text{old}}\right) \tag{8.12}$$

After some algebra, this can be expressed in the form of a standard quadratic equation.

$$a\left(T_{i,k,k}^{\text{new}}\right)^2 + b\left(T_{i,k,k}^{\text{new}}\right) + c = 0 \tag{8.13}$$

$$a = \frac{1}{\Delta x^2} + \frac{1}{\Delta y^2} + \frac{1}{\Delta z^2} \tag{8.14}$$

$$b = -2\left(\frac{m_x}{\Delta x^2} + \frac{m_y}{\Delta y^2} + \frac{m_z}{\Delta z^2}\right) \tag{8.15}$$

$$c = \frac{m_x^2}{\Delta x^2} + \frac{m_y^2}{\Delta y^2} + \frac{m_z^2}{\Delta z^2} - \frac{1}{R^2} \tag{8.16}$$

In general, this has two solutions. To ensure time values are always positive, the following solution is selected.

$$T_{i,j,k}^{\text{new}} = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \tag{8.17}$$

8.4. Implementation

A block diagram of how FMM is implemented in provided in Figure 8-2. The algorithm is initialized by setting all time values upwind of the initial interface to zero and all downwind time values to infinity. The narrow band of grid points immediately associated with the surface are initialized to be consistent with Eq. (8.2) and are stored in a separate list. It is only these points that need to be processed during simulation so a "narrow band" approach is most efficient. As the surface evolves, grid points are removed from, and added to, the narrow band.
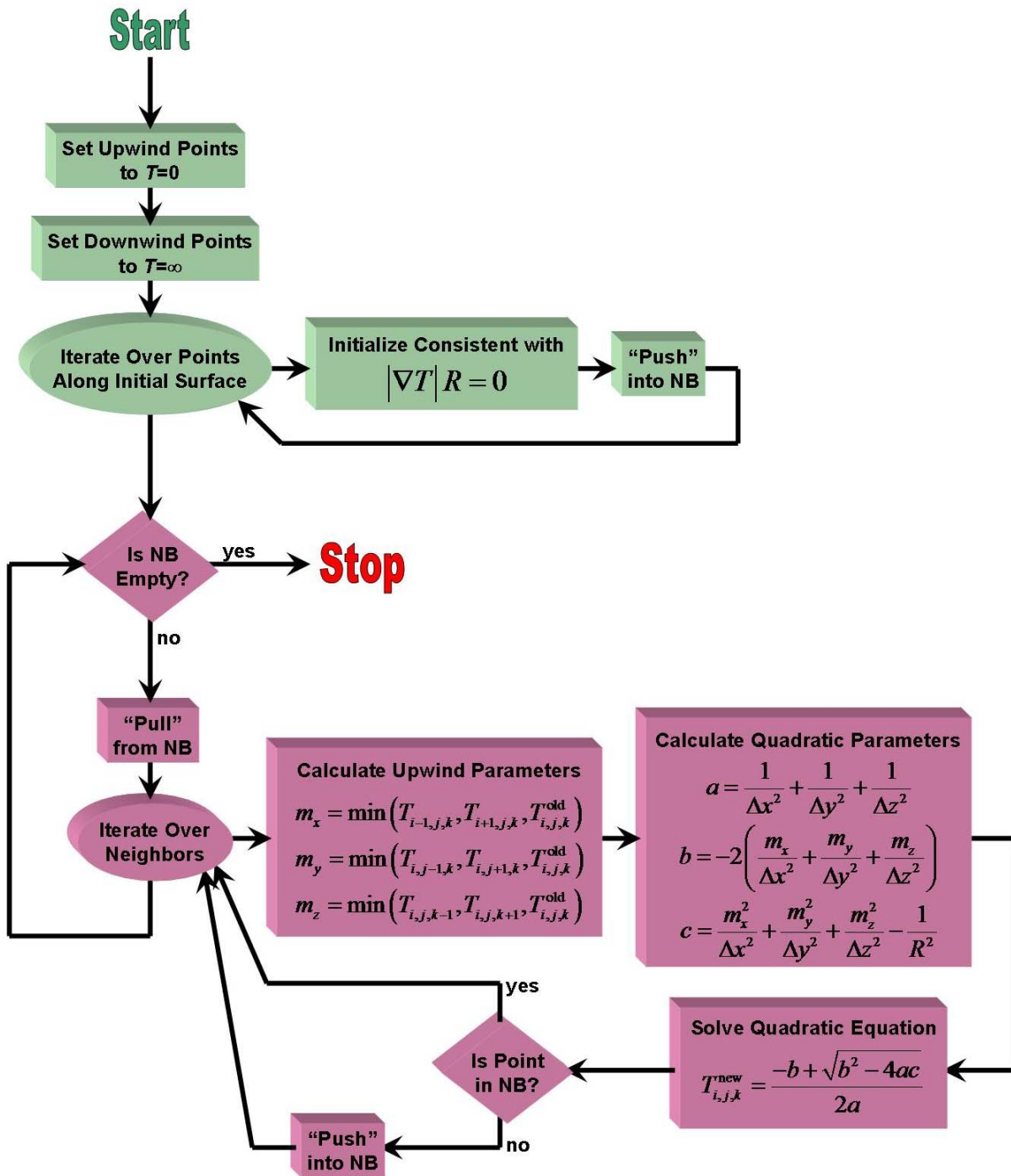
Figure 8-2: Block diagram of fast marching method

Iteration begins at the grid point stored in the narrow band with the smallest time value. This value must be correct since time values may only increase in the upwind scheme. For this reason, it is removed from the narrow band and made permanent. Time values at all neighboring grid points are updated according to Eq. (8.17). If any of these neighbors are not part of the narrow band, they are added to the list. Iteration continues at the grid point with the smallest time value in the narrow band. This process repeats until all grid points have been added to and finally removed from the narrow band. For each iteration, one grid point is always removed from the narrow band so the total number of iterations equals the total number of unknown grid points. Figure 8-2 illustrates FMM data at various stages of the model.
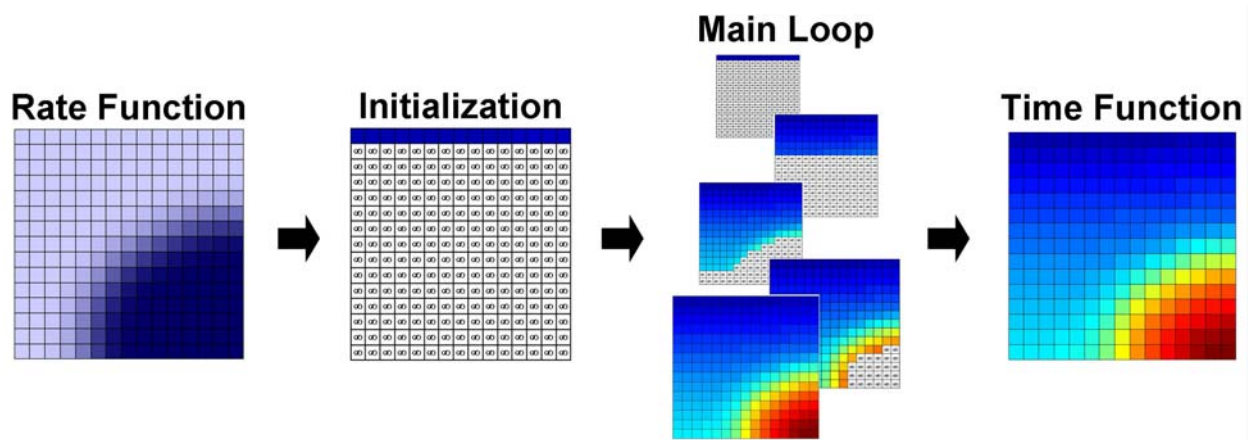
Figure 8-3: FMM data at each step

8.4.1.   *Min-Heap Data Structure*

The most time consuming task by far is finding the grid point in the narrow band with the smallest associated time value. The key to a fast and efficient FMM is to quickly find this point. A good approach is to store the narrow band grid points in a min-heap data structure [62, 63,

168

146, 149]. This is a form a binary tree that can be stored in a linear array without the need of data pointers. It is ordered such that the children of any element always have an equal or greater time value associated with them. In this data structure, the top element will always have the smallest time value. Data is stored in a linear array by locating the children of the element at position n at positions $2n$ and $2n+1$. This is depicted in Figure 8-4. Finding the grid point with the smallest time value simply involves selecting the top element of the tree, which is the first element of the array.
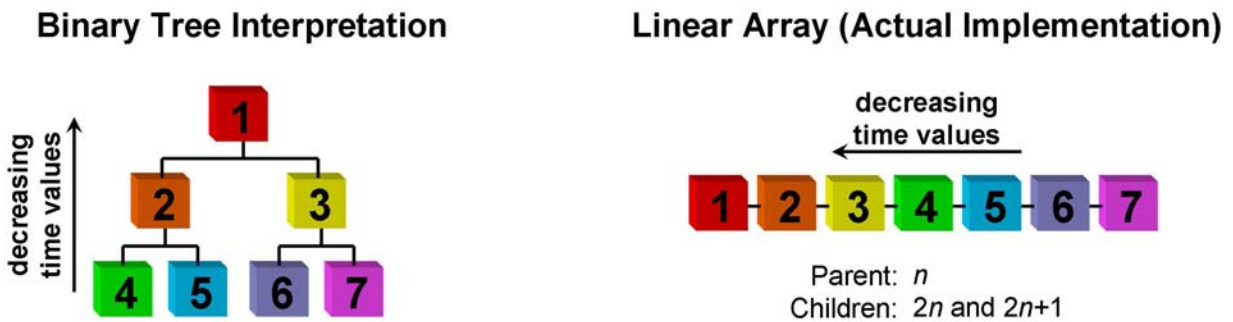
Figure 8-4: Min-heap data structure

Interfacing with the min-heap data structure involves "push" and "pull" operations. The "push" operation adds an element to the tree and adjusts element positions to maintain proper order. The "pull" operation removes the top element and adjusts the remaining elements to fill the top position and maintain proper order. A block diagram of both operations is provided in Figure 8-4.

The push operation is implemented by adding the new element to the bottom of the tree and performing a "bubble-up" procedure. The bottom element is successively swapped with its

parent if the parent has a larger time value. In this manner, the element moves up the tree until its parent has a smaller or equal time value, or it reaches the top of the tree.

The pull operation is implemented by moving the last element in the tree to the empty position at the top and performing a "bubble-down" procedure. The top element is successively swapped with children of smaller time-value. Preference is given to the child with the smallest time value. In this manner, the element moves down the tree until its time value is less than or equal to both its children, or reaches the end of the tree.
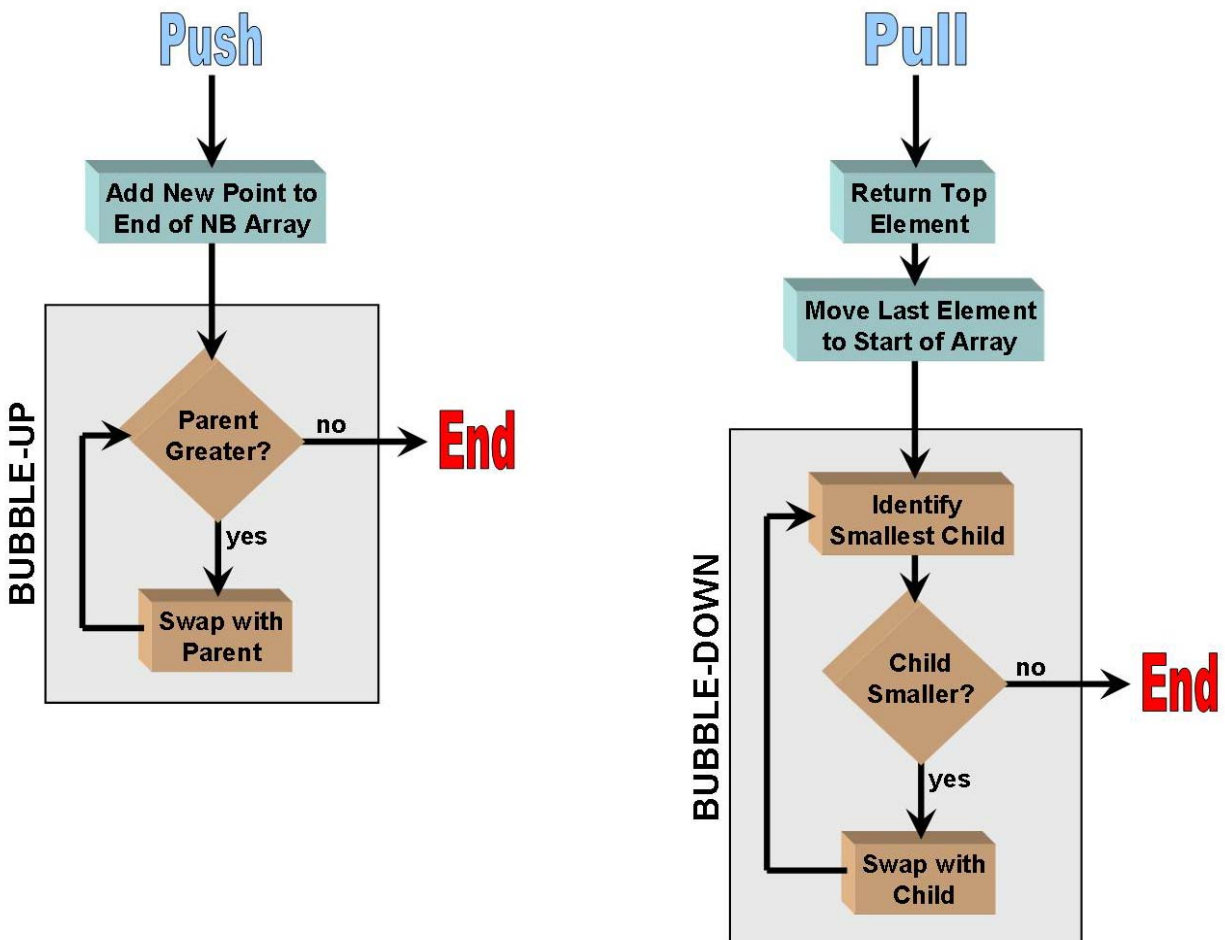
Figure 8-5: Min-heap operations

CHAPTER 9
NEAR-FIELD NANO-PATTERNING

9.1. Overview

Various techniques have been employed to build photonic crystals, but a particularly promising method is holographic lithography, also called interference lithography. It has been shown that all 14 Bravais lattices can be formed by interfering just four noncoplanar beams inside a photoresist [157]. The method is flexible and easily integrated with other optoelectronic devices. Near perfect periodicity can be achieved throughout a large volume in just a single process step because no alignments are needed. Photonic crystals, and photonic crystal templates, can be fabricated in low cost photoresist materials such as Epon® SU-8 [4, 158-160] or Futurrex® NR5-8000 [161, 162]. SU-8 has the added advantage of being environmentally robust and having many other applications in micro-photonics [4] and microelectromechanical systems (MEMS) [163-165]. Low absorption of SU-8 makes it ideal for uniform exposures through very thick films of resist.

Holographic lithography has two major drawbacks. First, it is highly sensitive to mechanical vibration. Even sub-micrometer translations can blur the exposed image such that no crystal can form. Second, a high degree of coherence is needed to insure the beams will produce sufficient interference. This dictates lasers be used as the illumination source. These drawbacks lead to sophisticated and expensive laboratory setups to implement the method.

Near-field nano-patterning (NFNP) is a form of holographic lithography that greatly simplifies the process and mitigates the major drawbacks [2, 5, 6, 25]. A high-resolution phase mask is placed in close proximity or in direct contact with photoresist. The phase mask diffracts incident light, forming a near-field irradiance pattern that exposes the photoresist. This method

is more immune to mechanical vibration because masks can be rigidly fixed in position with the resist. Standard ultraviolet (UV) lamps can be used as sources because required coherence length is only tens of microns. Figure 9-1 illustrates the process steps to implement the method.
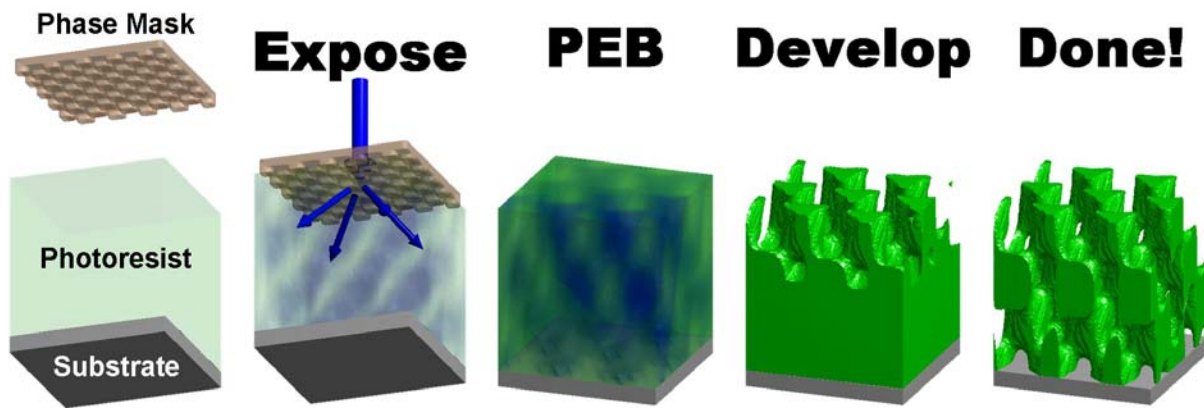
Figure 9-1: Near-field nano-patterning process

NFNP has some drawbacks. The method is less flexible and imposes some limits on what can be fabricated. It is difficult to simultaneously control diffraction efficiency and polarization from a grating so polarization is often not controlled when designing the mask. Masks require small features to diffract ultraviolet light. This requires advanced equipment to fabricate, but masks may be reused many times.

Simulation tools and design methods for photonic crystals are still evolving. Considerable analysis has been performed on ideal 3D lattices [166], but virtually no work has been devoted to modeling the fabrication of these structures. Throughout the literature, aerial images formed by holographic lithography are predicted using scalar models of exposure and simple intensity-threshold methods [167-169] are used to predict lattice geometry from the aerial image. For negative resists such as SU-8, thresholding methods assume all portions of resist

172

exposed with an irradiance above some threshold will remain after developing to form the lattice. While very fast and simple to implement, this technique ignores virtually all of the physics of lithography and all phenomena that produce distortions. Intensity-threshold methods are accurate only for very large features.

A comprehensive model of the fabrication and behavior of photonic crystals formed by near-field nano-patterning was developed in this dissertation. It includes exposure, post exposure baking, developing, and simulation of transmission and reflection spectra of the predicted crystal. This is the first known effort to incorporate physics of lithography into models of photonic crystals.

This chapter will begin by discussing theory of holographic lithography and outline how to design four-beam exposures to form any Bravais lattice symmetry. The procedure outlined in this chapter incorporates more degrees of freedom in the design process than identified elsewhere in literature. This is an enabling concept for designing masks intended for NFNP because the process imposes severe limits on beam configurations. Second, the method for designing the phase mask will be presented. Third, numerical and experimental results will be presented to show that holographic lithography inherently produces photonic crystals that vary in fill factor with depth. It will also be shown that illumination from standard UV lamps can be used to form 3D structures using NFNP. Impact of partial coherence, unpolarized light, unfiltered light, and light with an angular spectrum are demonstrated through simulation. Fourth, angular spectrum of source illumination is investigated and shown to blur the aerial image and increase in severity with depth. Angular spectrum is perhaps the most limiting parameter in NFNP. Finally, experimental results of fully three-dimensional photonic crystals formed in SU-8 in a standard mask aligner are presented to validate theoretical predictions.

9.2. Multiple Beam Interference

In holographic lithography, multiple noncoplanar beams are combined and in the volume of space where they all overlap an overall irradiance pattern, or aerial image, forms. The image formed by the interference of N beams is given by

$$I(\vec{r}) = \frac{n}{2\eta_0} \left| \sum_{p=1}^{N} \vec{E}_p e^{-j\vec{k}_p \bullet \vec{r}} \right|^2 \tag{9.1}$$

where $\vec{E}_p$ is the complex amplitude of the pth beam, $\vec{k}_p$ is the wave vector of pth beam, $\vec{r}$ is position, n is refractive index of the medium, and η_0 is the impedance of free space. After some algebraic manipulation, this can be written as the sum of $(N^2-N)/2$ sinusoidal gratings.

$$I(\vec{r}) = I_0 \left\{ 1 + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} V_{ij} \cos\left(\vec{K}_{ij} \bullet \vec{r} + \phi_{ij} \right) \right\} \tag{9.2}$$

$$I_0 = \frac{1}{2\eta} \sum_{p=1}^{N} \left| \vec{E}_p \right|^2 \tag{9.3}$$

$$\vec{K}_{ij} = \vec{k}_i - \vec{k}_j \tag{9.4}$$

$$V_{ij} = \frac{n}{2\eta_0 I_0} \left| \vec{E}_i \bullet \vec{E}_j^* \right| \tag{9.5}$$

$$\phi_{ij} = \text{Phase}\left[\vec{E}_i \bullet \vec{E}_j^* \right] \tag{9.6}$$

Parameter I_0 characterizes overall irradiance. V_{ij} is visibility of interference between the ith and j^{th} beams. $\vec{K}_{ij}$ is the grating vector of an imposed sinusoidal grating formed by the interference between the ith and j^{th} beams. ϕ_{ij} is a translation term resulting mostly from beam phase. The asterisk superscript indicates a complex conjugate. It can be shown that overall visibility is the sum of all visibility terms.

174

$$V_{\text{overall}} = \sum_{i,j} V_{i,j} \qquad (9.7)$$

9.2.1. Four Beam Interference

When four beams are used to produce interference, Eq. (9.2) reduces to the sum of six sinusoidal gratings. Any set of three grating vectors can completely describe the full set of six as long as they contain information from all four wave vectors. Out of 120 possible combinations, only five are valid and unique. These are

$$\left[\vec{K}_{12}, \vec{K}_{13}, \vec{K}_{14} \right] \qquad \left[\vec{K}_{12}, \vec{K}_{13}, \vec{K}_{24} \right] \qquad \left[\vec{K}_{12}, \vec{K}_{13}, \vec{K}_{34} \right]$$
$$\left[\vec{K}_{13}, \vec{K}_{14}, \vec{K}_{24} \right] \qquad \left[\vec{K}_{14}, \vec{K}_{23}, \vec{K}_{24} \right] \qquad (9.8)$$

Expanding on discussion in Chap. 2, any one set of the vector triplets above can define the primitive translation vectors of the reciprocal lattice. Lattice vectors of the direct lattice can be obtained by reciprocation.

9.3. Beam Synthesis

Before holographic lithography can be performed, a beam synthesis procedure must be formulated to be able to determine what four beams will correctly form the desired photonic crystal. Parameters that should be obtained include beam orientation, direction, wavelength, polarization, and phase. For every crystal, a large number of potential exposure configurations are possible. These must be sorted until a configuration is found with feasible beam angles at a realizable wavelength and provides the desired lattice constant. For example, it is often not feasible to expose a photoresist from the top and bottom simultaneously, so a beam configuration must be found where all beams are incident from the top.

The following sections outline the beam synthesis procedure. As an example, the procedures will be demonstrated by deriving the famous "umbrella" configuration known to form a FCC photonic crystal.

9.3.1. *Calculation of Lattice Vectors*

Beam synthesis begins by describing the desired crystal using primitive translation vectors of the reciprocal lattice $\left[\vec{T}_1, \vec{T}_2, \vec{T}_3 \right]$. It is usually most straight forward to begin by defining the primitive axis vectors of the direct lattice to control size and orientation of the unit cell more intuitively. A cubic photonic crystal with lattice constant Λ will have the following primitive axis vectors.

$$
\begin{bmatrix} \vec{a}_1 \\ \vec{a}_2 \\ \vec{a}_3 \end{bmatrix} = \begin{bmatrix} \Lambda & 0 & 0 \\ 0 & \Lambda & 0 \\ 0 & 0 & \Lambda \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{bmatrix}
\tag{9.9}
$$

The primitive translation vectors can be calculated from the axis vectors using Eqs. (2.2)-(2.6). For a FCC photonic crystal, this is

$$
\begin{bmatrix} \vec{t}_1 \\ \vec{t}_2 \\ \vec{t}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix} \begin{bmatrix} \vec{a}_1 \\ \vec{a}_2 \\ \vec{a}_3 \end{bmatrix}
\tag{9.10}
$$

The primitive translation vectors of the reciprocal lattice are calculated using Eq. (2.8). For the FCC crystal, these are

$$
\vec{T}_1 = 2\pi \frac{\vec{t}_2 \times \vec{t}_3}{\vec{t}_1 \bullet \left(\vec{t}_2 \times \vec{t}_3 \right)} = \frac{2\pi}{\Lambda} \left(-\hat{x} + \hat{y} + \hat{z} \right)
\tag{9.11}
$$

$$
\vec{T}_2 = 2\pi \frac{\vec{t}_3 \times \vec{t}_1}{\vec{t}_1 \bullet \left(\vec{t}_2 \times \vec{t}_3 \right)} = \frac{2\pi}{\Lambda} \left(\hat{x} - \hat{y} + \hat{z} \right)
\tag{9.12}
$$

$$\vec{T_3} = 2\pi \frac{\vec{t_1} \times \vec{t_2}}{\vec{t_1} \bullet \left(\vec{t_2} \times \vec{t_3}\right)} = \frac{2\pi}{\Lambda}\left(\hat{x} + \hat{y} - \hat{z}\right) \tag{9.13}$$

9.3.2. Calculation of Grating Vectors

From Eq. (2.10), it can be shown that the desired crystal will form as long as the induced grating vectors $\left[\vec{K}_{12}, \vec{K}_{13}, \vec{K}_{14}\right]$ are an integer combination of the primitive translation vectors of the reciprocal lattice. This realization leads to a much greater number of potential configurations to explore. The drawback is an altered geometry, but the basic symmetry will be preserved.

$$\begin{bmatrix} \vec{K}_A \\ \vec{K}_B \\ \vec{K}_C \end{bmatrix} = \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix} \begin{bmatrix} \vec{T_1} \\ \vec{T_2} \\ \vec{T_3} \end{bmatrix} \tag{9.14}$$
$$\underset{\mathbf{Q}}{\downarrow}$$

Each lattice vector must be incorporated at least once into the grating vectors, so no columns in $\mathbf{Q}$ can be all zeros. Each grating vector must be unique so no two rows in $\mathbf{Q}$ can be the same.

To arrive at the famous umbrella configuration, the following choice is made.

$$\begin{bmatrix} \vec{K}_A \\ \vec{K}_B \\ \vec{K}_C \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \vec{T_1} \\ \vec{T_2} \\ \vec{T_3} \end{bmatrix} \tag{9.15}$$

9.3.3. Calculation of Wave Vectors

Before computing wave vectors, the manner in which the grating vectors relate to the wave vectors must be chosen. This entails selection of a vector triplet from Eq. (9.8). Having reduced all possible combinations down to just five provides a more efficient search of possible beam configurations. The umbrella configuration proceeds with $\left[\vec{K}_{12}, \vec{K}_{13}, \vec{K}_{14}\right]$ which is also

the most common choice. Based on this selection, the wave vectors are related to the grating vectors through the following equations.

$$\vec{k}_1 - \vec{k}_2 = \vec{K}_A \tag{9.16}$$

$$\vec{k}_1 - \vec{k}_3 = \vec{K}_B \tag{9.17}$$

$$\vec{k}_1 - \vec{k}_4 = \vec{K}_C \tag{9.18}$$

This set of equations cannot be solved because there are more unknown vector quantities than there are known. Additional information is needed and this comes from the condition that all four wave vectors must have the same magnitude.

$$\left|\vec{k}_1\right| = \left|\vec{k}_2\right| = \left|\vec{k}_3\right| = \left|\vec{k}_4\right| \tag{9.19}$$

The system of equations defined by Eqs. (9.16)-(9.19) is not linear and cannot be solved algebraically. If one wave vector is known, however, the remaining three can be computed. The problem reduces to searching k-space to find magnitude and direction of the first wave vector that leads to all four having the same magnitude. A simple error-reduction algorithm can perform this operation.

Continuing with the example, if a guess is made for $\vec{k}_1$ the other three wave vectors can be calculated using Eqs. (9.16)-(9.18) as

$$\vec{k}_2 = \vec{k}_1 - \vec{K}_A \tag{9.20}$$

$$\vec{k}_3 = \vec{k}_1 - \vec{K}_B \tag{9.21}$$

$$\vec{k}_4 = \vec{k}_1 - \vec{K}_C \tag{9.22}$$

An error function resulting from the guess can be calculated as

$$\varepsilon\left(\vec{k}_1\right) = \left(\left|\vec{k}_1\right| - \left|\vec{k}_2\right|\right)^2 + \left(\left|\vec{k}_1\right| - \left|\vec{k}_3\right|\right)^2 + \left(\left|\vec{k}_1\right| - \left|\vec{k}_4\right|\right)^2 \tag{9.23}$$

178

The full set of wave vectors is determined by searching *k*-space for the minimum error. When this is performed, the calculated wave vectors turn out to be

$$\vec{k}_1 = \frac{\pi}{\Lambda}(3\hat{x} + 3\hat{y} + 3\hat{z}) \tag{9.24}$$

$$\vec{k}_2 = \frac{\pi}{\Lambda}(5\hat{x} + \hat{y} + \hat{z}) \tag{9.25}$$

$$\vec{k}_3 = \frac{\pi}{\Lambda}(\hat{x} + 5\hat{y} + \hat{z}) \tag{9.26}$$

$$\vec{k}_4 = \frac{\pi}{\Lambda}(\hat{x} + \hat{y} + 5\hat{z}) \tag{9.27}$$

Inspection of these equations shows $\vec{k}_1$ to be the central beam with $\vec{k}_2$, $\vec{k}_3$, and $\vec{k}_4$ distributed evenly around it. Angle between the central beam any of the radial beams is 38.94°.

9.3.4.   *Calculation of Exposure Wavelength or Lattice Period*

After the four wave vectors are found, exposure wavelength can be calculated through the following equation.

$$\left|\vec{k}\right| = \frac{2\pi n}{\lambda_0} \tag{9.28}$$

$\vec{k}$ can be any of the four calculated wave vectors, *n* is refractive index of the medium, and λ_0 is the free space wavelength.

From Eqs. (9.24)-(9.27), $\left|\vec{k}\right| = 3\pi\sqrt{3}/\Lambda$ which leads to an equation relating exposure wavelength λ_0 to the lattice constant Λ.

$$\frac{\lambda_0}{n\Lambda} = \frac{2}{3\sqrt{3}} \approx 0.3849 \tag{9.29}$$

179

Typically the exposure wavelength is fixed and this equation is used to calculate the lattice constant that would result from this specific exposure configuration. If it is critical to realize a particular lattice constant, this equation can be used to determine what exposure wavelength is necessary. Remember, however, that different exposure configurations will lead to different relations between lattice constant and wavelength so more possibilities exist. It should be mentioned that most photoresists are typically only sensitive to a small range of wavelengths.

9.3.5. Beam Reorientation

Many exposure configurations derived from the above procedure appear to be unfeasible at first glance. It many cases it is possible to reorient the vectors so the beams are realizable. While this enables more exposure configurations to work, the resulting lattices will be reoriented in the same manner as the beams. Figure 9-2 illustrates the reorientation concept.

The left diagram in Figure 9-2 shows the vectors as calculated so far. The umbrella configuration here is obviously tilted leading to an exposure that is very difficult to realize. To obtain the easiest possible exposure configuration, the vectors are rotated about the z-axis by 45°, then around the x-axis by 54.74°, so the central beam is perfectly vertical. This leads to the system of vectors shown in the right diagram of Figure 9-2. Beams in the new orientation are clearly easier to realize in the lab. Prisms may be used to achieve beam angles greater than the critical angle [159].

Axis Vectors: $\vec{a}_1, \vec{a}_2, \vec{a}_3$
Translation Vectors: $\vec{t}_1, \vec{t}_2, \vec{t}_3$
Grating Vectors: $\vec{T}_1, \vec{T}_2, \vec{T}_3$
Wave Vectors: $\vec{k}_1, \vec{k}_2, \vec{k}_3, \vec{k}_4$

Vectors as calculated

Vectors in rotated system
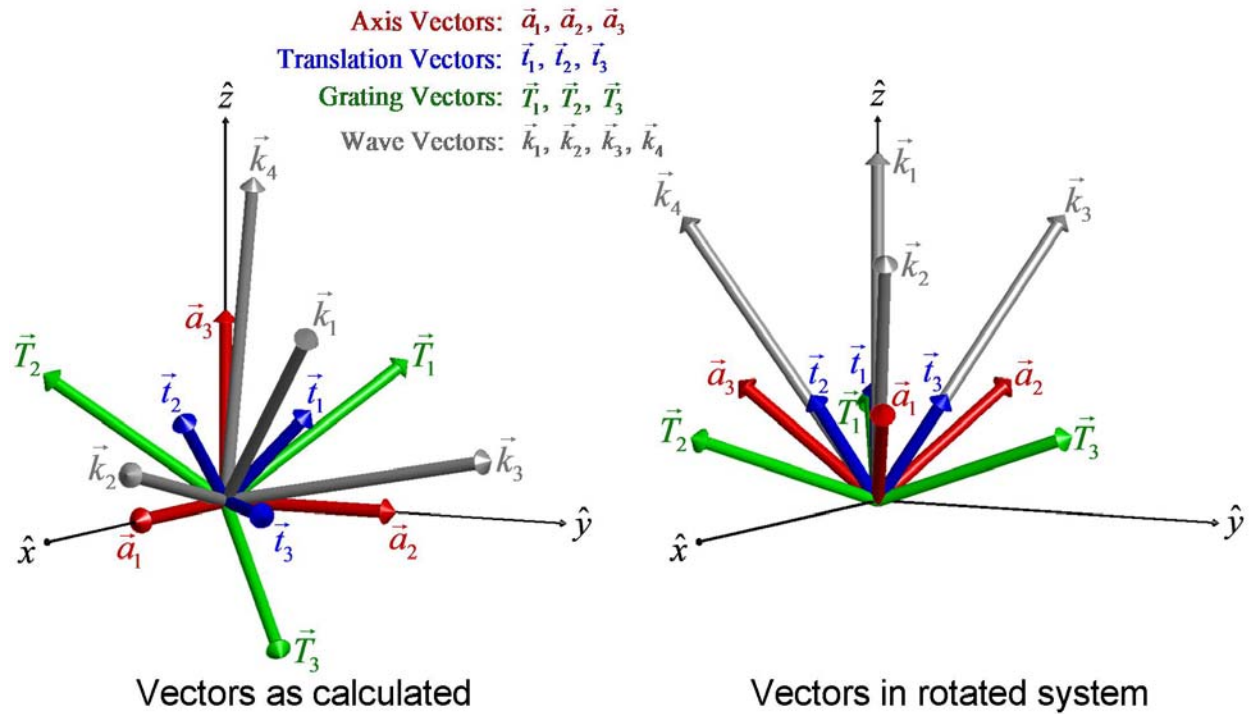
Figure 9-2: Beam reorientation

9.3.6. *Beam Polarization and Phase*

The last step in beam synthesis is determining optimum beam polarization and phase. Polarization can have dramatic impact on image contrast, lattice connectivity, and shape of objects in the lattice [170, 171]. All are very important to reliably form quality photonic crystals. This is apparent by inspecting the dot product in the visibility term defined in Eq. (9.5). If beams are orthogonally polarized, they will not interfere. Foundations of numerical methods to optimize polarization have been addressed in the literature [171]. In most cases, polarization is chosen to maximize interference the between beam pairs that produce the three chosen grating vectors and to minimize interference between all other beam pairs.

Beam phase has a similar impact as polarization. It has been shown to affect image contrast as well as shape of the interference pattern [172]. Phase affects interference through the visibility term and translation term defined in Eq. (9.5) and Eq. (9.6) respectively. In most laboratory setups, relative phase between the beams is difficult to control and is usually neglected unless stability is addressed. Most often, the biggest impact of beam phase is a simple translation of lattice points so it is often ignored.
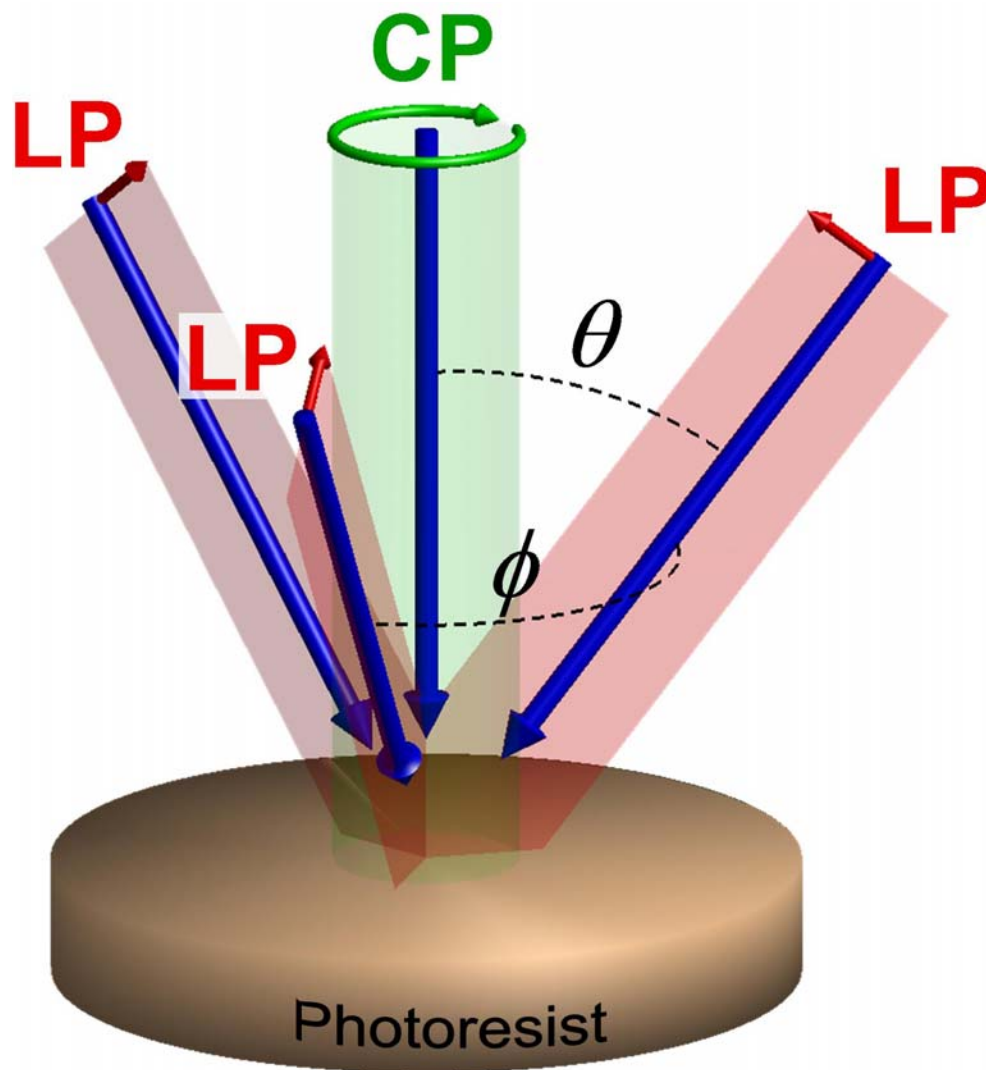
Figure 9-3: Famous "umbrella" configuration for FCC symmetry

For the umbrella configuration, interference between the side beams can be minimized if they are linearly polarized (LP) in their planes of incidence because they are oriented at nearly orthogonal angles. To maximize interference of the side beams with the central beam, the central beam is chosen to be circularly polarized (CP). This fully optimized configuration is illustrated in Figure 9-3 where θ=38.94° and ϕ=120°.

9.4. Mask Design

When an optical wave propagates through a periodic medium like a diffraction grating, the field takes on the same periodicity as the medium. Fourier theory dictates that all periodic functions be comprised of discrete harmonics. This means electromagnetic fields in periodic media can be viewed as a discrete set of plane waves propagating at different angles. Direction of diffracted waves is only a function of size and symmetry of the grating and can be calculated using the famous grating equation [9]. Diffraction occurs in planes defined by the incident wave vector and grating vector. The primary diffraction planes for three common grating symmetries are shown in Figure 9-4.
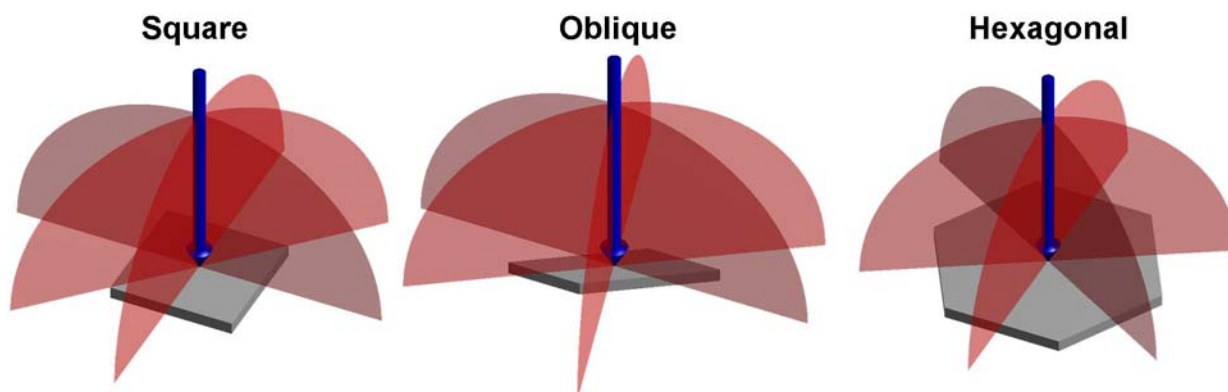
Figure 9-4: Diffraction planes

Grating symmetry is chosen so position of the grating planes are parallel to the planes containing the desired diffracted orders. The specific angle of the diffracted wave within the diffraction plane is calculated using the grating equation.

$$n_g \sin \theta_i = n_{\text{inc}} \sin \theta_{\text{inc}} - i \frac{\lambda_0}{\Lambda} \tag{9.30}$$

Refractive index n_g is the average refractive index in the grating region, n_{inc} is the refractive index of the incident medium, θ_{inc} is angle of incidence, angle of the ith diffracted order is θ_I, period of the grating is Λ, and the free space wavelength is λ_0. From Eq. (9.30), grating period may be chosen to provide the desired beam angles within the diffraction planes.

From discussion above, it can be concluded that grating symmetry determines the radial orientation of the diffracted beams around a central beam. Grating period determines the elevation angle of the diffracted beams within the grating planes. It is still necessary to design the pattern within the grating unit cell and choose a grating relief depth. This is done in a manner to balance energy in the diffracted modes so image contrast is maximized. While many iterative algorithms exist to perform this design, most use scalar approximations that are not accurate when grating dimensions are near a wavelength. At this scale, fabrication becomes a serious limitation and grating design often assumes patterns that accommodate easier fabrication.

Rigorous optimization would entail iterating design parameters until image contrast is maximized or the desired three dimension pattern forms after full comprehensive modeling. In this dissertation it was found that balancing diffraction efficiencies of the diffracted beams was a much faster approach and typically produced near optimal designs.
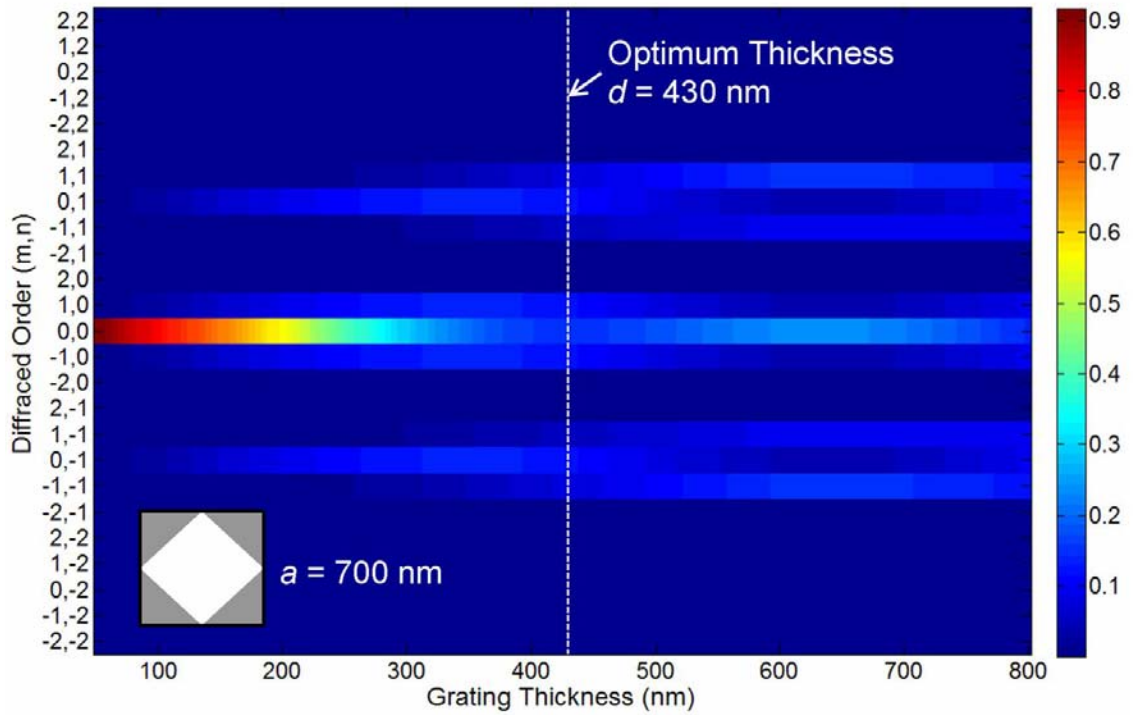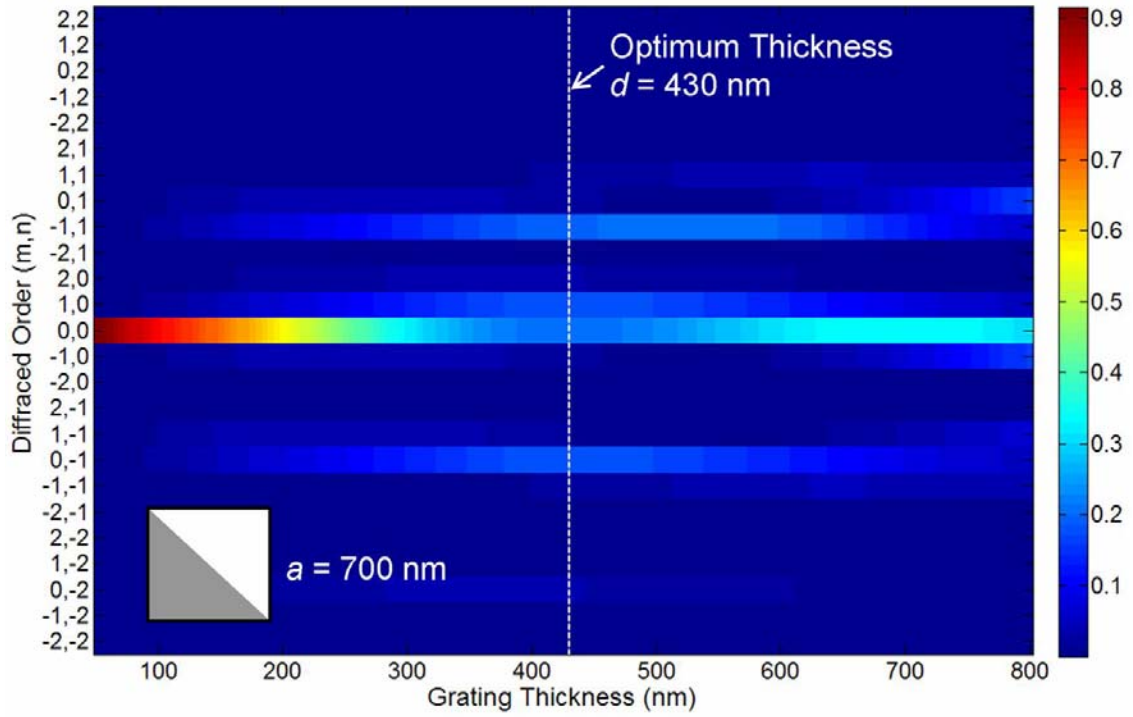
Figure 9-5: Grating optimization

Two grating patterns with square symmetry were selected. Hexagonal symmetry was not considered. Diffraction efficiency for all diffracted harmonics was calculated as a function of grating thickness. This data is provided in Figure 9-5 along with insets showing a single period of the chosen gratings. For both gratings, optimum grating thickness was around 430 nm where diffraction efficiency was approximately balanced. The top diagram contains the best grating design because there are fewer significant harmonics, better balance of diffraction efficiency, and the optimum is maintained over broader range of wavelengths.

9.5. Physics of Near-Field Nano-Patterning

Before a numerical method can be chosen to model NFNP, physics of the process must be understood. Processes considered here were exposure, post exposure baking, and developing because these are the main steps that determine final geometry.

9.5.1.   *Physics of Exposure*

For SU-8, exposure to UV illumination causes a strong acid to form where light is absorbed. Several phenomena combine in the exposure process to determine the overall aerial image. Optical absorption causes irradiance to decay with depth into the photoresist. This is somewhat more complicated if strong reflections and standing waves are present or exposure involves beams incident from both top and bottom. Polarization of the incident and diffracted light is very important. If diffracted beams are orthogonally polarized, they not will produce interference leading to low image contrast and altering geometry of formed structures. Source coherence must also be considered. It has been shown partial coherence produces an overall "smoothing" affect, and structures can still be formed using sources with linewidths exceeding 20 nm [2]. If an air gap is present between the grating and photoresist, angles of the diffracted

beams will be limited by the critical angle of the photoresist and spatial frequency limit of the grating. Three-dimensional structures may still be formed if an air gap is present, but the limited angles will lead to structures elongated in the vertical direction. Perhaps the most critical aspect of NFNP is the angular spectrum of the incident light. Off-axis rays produce tilted aerial images. This causes the overall aerial image to blur with depth, severely reducing image contrast deeper in the photoresist. The angular content of the light must be kept less than a fraction of one degree [2].

During exposure, most photoresists bleach or become transparent as light is absorbed. The phenomena is typically detrimental to holographic lithography, but has been exploited to form digital and analog structures in some positive photoresists [173]. It is most easily quantified using Dill's parameters [174-176] in which optical properties of photoresist are described in terms of refractive index n and absorption coefficient α. During exposure, the "degree of exposure" M decreases according to

$$\frac{d}{dt}M\left(\vec{r},t\right) = -C \cdot M\left(\vec{r},t\right) \cdot I\left(\vec{r},t\right) \tag{9.31}$$

where $I\left(\vec{r},t\right)$ is optical irradiance and C is a parameter characterizing sensitivity of the photoresist. Parameter M is related to the concentration of photoactive compound, but is interpreted here as degree of exposure. Unexposed photoresist corresponds to $M=1$ while fully exposed photoresist corresponds to $M=0$. The absorption coefficient at any instant during exposure is given by

$$\alpha\left(\vec{r},t\right) = A \cdot M\left(\vec{r},t\right) + B \tag{9.32}$$

The parameter A is an exposure dependant absorption coefficient, while B is an exposure independent absorption coefficient. The absorption coefficient of unexposed photoresist is $A+B$,

while the absorption coefficient of fully exposed photoresist is just *B*. The Dill parameters

[*A*,*B*,*C*] are usually provided in the data sheet of the photoresist. Bleaching is negligible for

SU-8 exposed with i-line radiation (i.e. 365 nm) [158]. This is a fortunate circumstance as it

greatly simplifies simulating exposure.

9.5.2.   *Physics of Post Exposure Bake*

During the post exposure bake (PEB), thermal energy catalyzes cross linking in SU-8.

The degree of cross linking determines solubility and ultimately dissolution rate of the resist into

a developer solution. A number of simultaneous chemical reactions occur during the PEB that

are quantified by a system of coupled partial differential equations [177]. These equations

describe the concentration of active compounds as they are produced by reactions during the

bake. At the same time, mobile acids diffuse into surrounding resist. The combined effects of

reaction kinetics and diffusion blur the aerial image to form the latent image.

The general equation describing diffusion of mobile acids is [177]

$$\frac{\partial M}{\partial t} = \nabla \left(D \cdot \nabla M \right) \tag{9.33}$$

where *M* represents chemical concentration and *D* specifies the diffusion coefficient that can

vary spatially or depend on concentration, temperature, and more. Chemically amplified resists

can exhibit different properties on the surfaces because acids are not able to diffuse outside of the

resist and tend to concentrate there. For holographic exposure where dose tends to be low, this

effect was found to be negligible and Eq. (9.33) reduces to

$$\frac{\partial M}{\partial t} = D \left(\nabla^2 M \right) \tag{9.34}$$

Three types of diffusion were identified by Erdmann, *et al* [177], that differed in how the parameter *D* was defined. All of these were found to produce diffusion profiles that were close to Gaussian and could be sufficiently described by the following function.

$$g\left(\vec{r}\right) = \frac{1}{\sqrt{2\pi\rho_{\text{eff}}^2}}\exp\left(-\frac{|\vec{r}|^2}{2\rho_{\text{eff}}^2}\right) \tag{9.35}$$

Vector $\vec{r}$ is offset from the original of diffusion and ρ_{eff} is the effective diffusion length.

Acid diffusion and reaction kinetics combine to blur the aerial image forming the latent image. It is by this mechanism that standing waves and other quickly varying artifacts are usually erased. Image contrast is reduced, fine features in the aerial image are erased, and an overall "smoothing" affect takes place. Acid diffusion, however, contributes to lattice connectivity by enabling isolated pockets of acid to diffuse into each other to form an attached and more mechanically robust structure.

Another important phenomenon to consider is shrinkage. During cross-linking, many resists shrink in size. SU-8 is reported to shrink by approximately 7.5% during the PEB [159]. If this is not considered, the minimum affect on optical behavior will be a 7.5% shift in spectral response. It is reasonable to assume shrinkage occurs mostly in the vertical direction since the photonic crystal is usually a thin slab well adhered to a rigid substrate.

9.5.3. Physics of Developing

The developing process dissolves soluble resist into a developer solution leaving behind the insoluble portions. It is the partially soluble photoresist that must be treated correctly to accurately predict geometry. Strength of a developer solution is locally weakened as dissolved photoresist dilutes the chemistry. For this reason, sufficient agitation may be necessary to remove insoluble photoresist residing deep within a complex 3D structure.
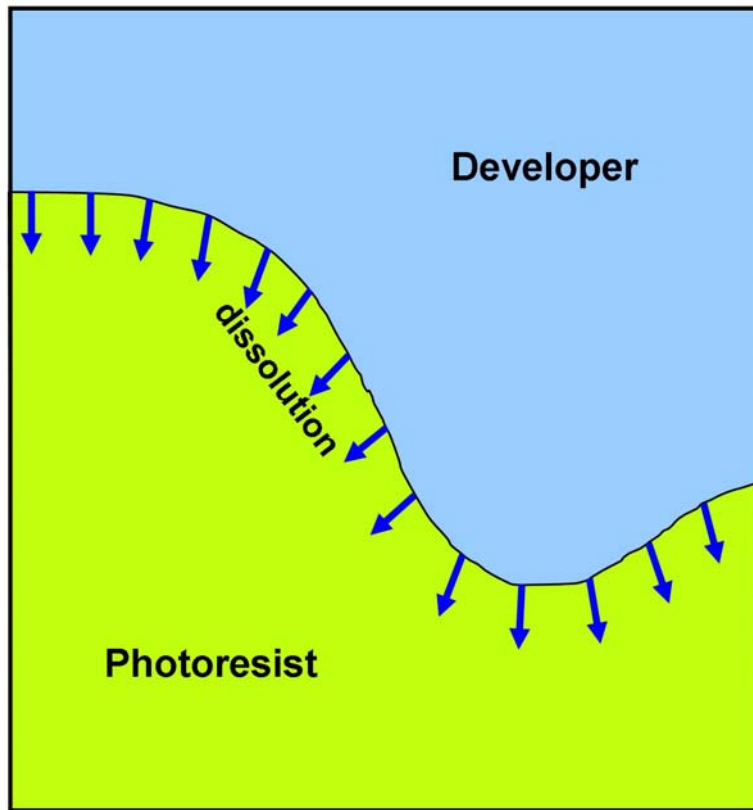
Figure 9-6: Dissolution of photoresist

Quantitatively, the dissolution process is essentially isotropic and described by the dissolution equation, $d=RT$, and is illustrated in Figure 9-6. This relates the distance d that the resist-developer interface will progress if allowed to develop for time T where the local dissolution rate of the resist is R (μm/sec). In addition, dissolution is always in the direction of the surface normal $\hat{n}$ where it is defined pointing toward the resist. The vector form of the dissolution equation is

$$\vec{d} = RT\hat{n} \qquad (9.36)$$

9.6. Comprehensive Modeling

In the comprehensive modeling framework, separate models were constructed for exposure, post exposure baking (PEB), and developing that each successively operated on a common materials mesh. The exposure model computed irradiance throughout the photoresist, or aerial image, and incorporated diffraction through the phase mask, unpolarized light, spectrum of source, optical absorption, scattering, and standing wave effects. The PEB model simulated blurring effects of acid diffusion to determine the latent image and accounted for shrinkage due to cross linking. Dissolution rate throughout the photoresist was computed from the latent image using the enhanced notch model [178]. This data was fed into the final model to simulate the developing process.

By combining these models, geometry of formed structures can be predicted more rigorously. Optical performance can be studied more accurately with a higher degree of fidelity. This tool is seen as an enabling step for designing devices that can exploit the physics and compensate for limiting factors inherent in the fabrication process.

9.6.1. Modeling Exposure

To account for as much physics involved in the exposure process as possible, it is important to use a rigorous method like FDTD [1, 179-181] or RCWA [2, 5]. For near-field nano-patterning, RCWA seems most appropriate because the problem is periodic, comprised of multiple thick layers, and the index contrast tends to be low. Finite-difference time-domain is better suited to problems with larger transverse dimensions, high volumetric complexity, or problems with very high index contrast or metals. RCWA makes no assumptions or approximations to Maxwell's equations so it can account for essentially all phenomena in the

exposure process. Partial coherence, bleaching, unpolarized light, and light with an angular

spectrum can be handled using multiple simulations of RCWA [5].

The aerial image produced by incident light with an angular spectrum was estimated by

integrating the aerial image over the entire angular spectrum of the incident light. Aerial images

formed using unpolarized, or randomly polarized, light can be estimated by adding aerial images

produced by perpendicular and parallel polarizations. After the total aerial image is determined,

irradiance I is computed with the following equation.

$$I(\vec{r}) = \frac{n}{2\eta_0} \left| \vec{E}(\vec{r}) \right|^2 \tag{9.37}$$

$\vec{E}$ is electric field, n is refractive index, and η_0 is the impedance of free space. If the source is

partially coherent or contains multiple spectral lines, overall irradiance can be estimated by

integrating irradiance over the full spectrum of the source.

When bleaching can be ignored, the distribution of absorbed energy $\xi_A(\vec{r})$ after

exposure, is computed from the absorption coefficient α, total exposure time T, and the aerial

image $I(\vec{r})$.

$$\xi_A(\vec{r}) = \alpha \cdot T \cdot I(\vec{r}) \tag{9.38}$$

If bleaching has to be incorporated, the process must be iterated in small enough time

steps to resolve the bleaching process. The time derivative in Eq. (9.31) is approximated with a

finite-difference leading to the following equation to update M.

$$M(\vec{r}, t + \Delta t) = M(\vec{r}, t)\left[1 - C \cdot \Delta t \cdot I(\vec{r}, t)\right] \tag{9.39}$$

Figure 9-7 shows a block diagram of how the bleaching process is modeled using Dill's

parameters.
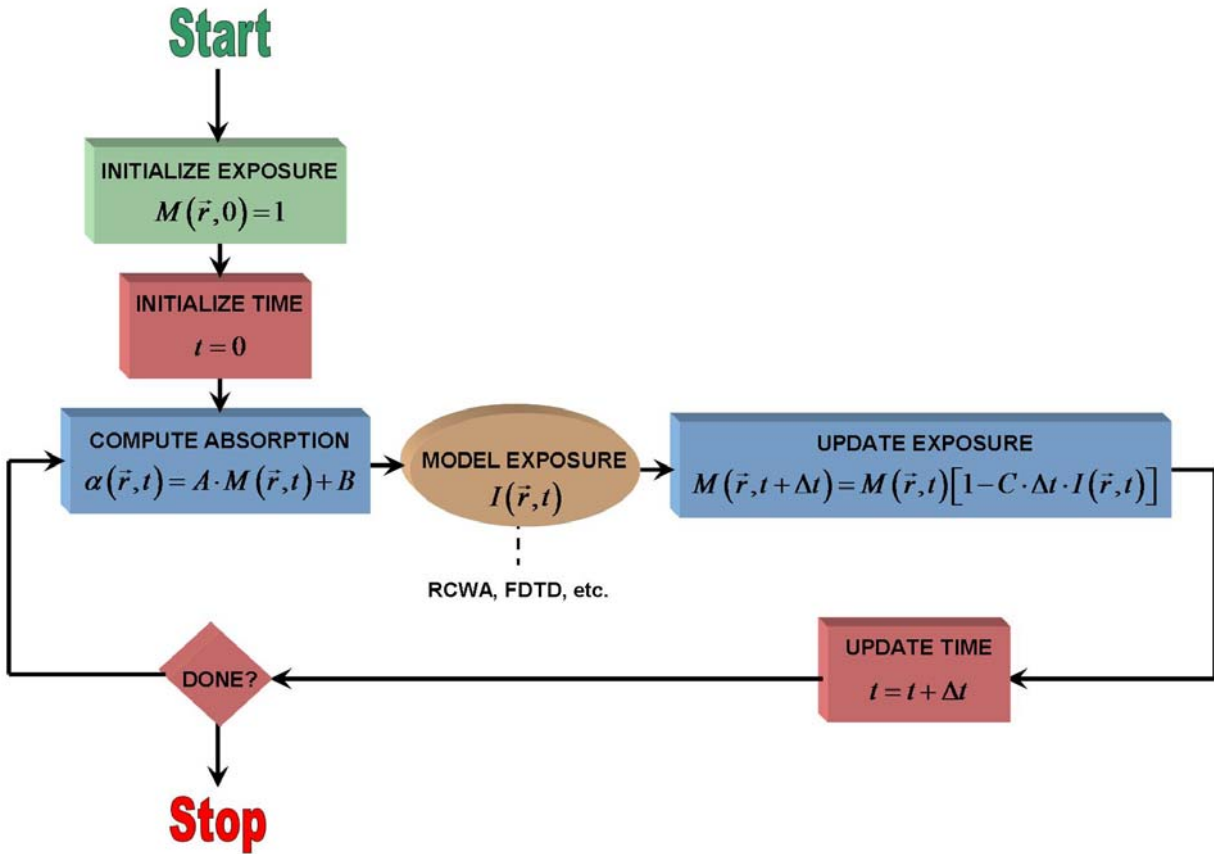
Figure 9-7: Block diagram of modeling the bleaching process

9.6.2. *Modeling Post Exposure Bake*

Three aspects to consider in the PEB are blurring due to acid diffusion, shrinkage due to cross linking, and computation of dissolution rate. Acid diffusion can be modeled by convolving [182] the absorbed energy function $\xi_A(\vec{r})$ with the Gaussian point spread function given in Eq. (9.35).

$$\xi_L(\vec{r}) = \xi_A(\vec{r}) * g(\vec{r}) \tag{9.40}$$

To model shrinkage, resolution parameters [$\Delta x, \Delta y, \Delta z$] of the Cartesian grid can be modified according to a shrinkage tensor. As previously discussed, it may be best to assume $s_{xx}=s_{yy}=0$ since films are usually thin and well adhered to a rigid substrate.

$$\begin{bmatrix} \Delta x' \\ \Delta y' \\ \Delta z' \end{bmatrix} = \begin{bmatrix} s_{xx} & 0 & 0 \\ 0 & s_{yy} & 0 \\ 0 & 0 & s_{zz} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \tag{9.41}$$

To compute dissolution rate R of a high contrast photoresist like SU-8, an empirical equation called the enhanced notch model [178] was used. The model was modified to compute R directly from a normalized absorbed energy function $\bar{\xi}(\vec{r})$. This formulation is

$$R(\bar{\xi}) = R_{max}(1-\bar{\xi})^N \left[\frac{(a_n+1)(1-\bar{\xi})^{N_{notch}}}{a_n+(1-\bar{\xi})^{N_{notch}}} \right] + R_{min} \left[\frac{R_{min}^{\bar{\xi}-1}}{R_{max}^{\bar{\xi}-1}} \right] \left[1 - \frac{(a_n+1)(1-\bar{\xi})^{N_{notch}}}{a_n+(1-\bar{\xi})^{N_{notch}}} \right] \tag{9.42}$$

$$a_n = \frac{N_{notch}+1}{N_{notch}-1}(1-\bar{\xi}_{th})^{N_{notch}} \tag{9.43}$$

$\bar{\xi}$ is a normalized energy parameter in the range $0 \le \bar{\xi} \le 1$. R_{max} is the fastest dissolution rate that occurs where no energy is absorbed (i.e., $\bar{\xi}=0$). R_{min} is the slowest dissolution rate in saturated regions (i.e., $\bar{\xi}=1$). The parameter $\bar{\xi}_{th}$ defines the energy threshold where dissolution rate falls rapidly. Slope of the dissolution rate curve below the notch region is controlled through parameter N. Higher values yield steeper slopes. In a similar manner, N_{notch} controls slope in the central region, or notch region, where dissolution rate falls rapidly. These parameters are discussed in more detail in Ref. [1].

The parameters were estimated experimentally to be on the order of $R_{max} = 7.5 \frac{\mu m}{sec}$, $R_{min} = 5 \times 10^{-10} \frac{\mu m}{sec}$, $\bar{\xi}_{th} = 0.51$, N=0.8, and N_{notch}=100, but these were observed to vary somewhat for different types of SU-8. Figure 9-8 shows the dissolution rate curve using these parameters.
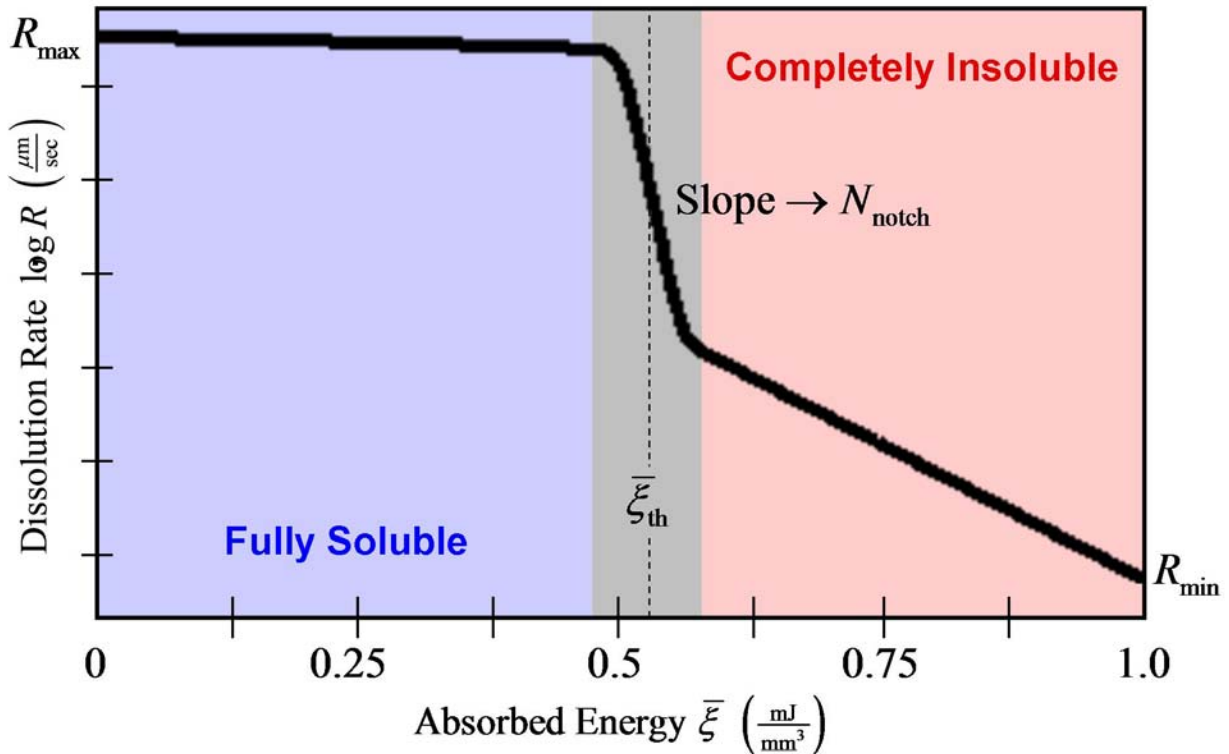
Figure 9-8: Dissolution rate curve of SU-8

9.6.3. Modeling Resist Development

The fast marching method (FMM) was used to model the developing process because it is very fast, efficient, and is able to handle arbitrarily complex structures. Figure 9-9(a) shows an example dissolution rate profile consisting of a sphere of partially soluble resist suspended in a volume of fully soluble resist. In this diagram, black corresponds to the least soluble resist, while the white indicates fully soluble. Figure 9-9(b) shows the time function computed from this data using FMM as described above. Black indicates lowest time values, while white indicates highest time values. It can be observed from this data that the sphere would be last to dissolve because time values in that region are greatest. A slight gradient can also be observed

from top to bottom indicating the short amount of time required to remove the fully soluble resist.
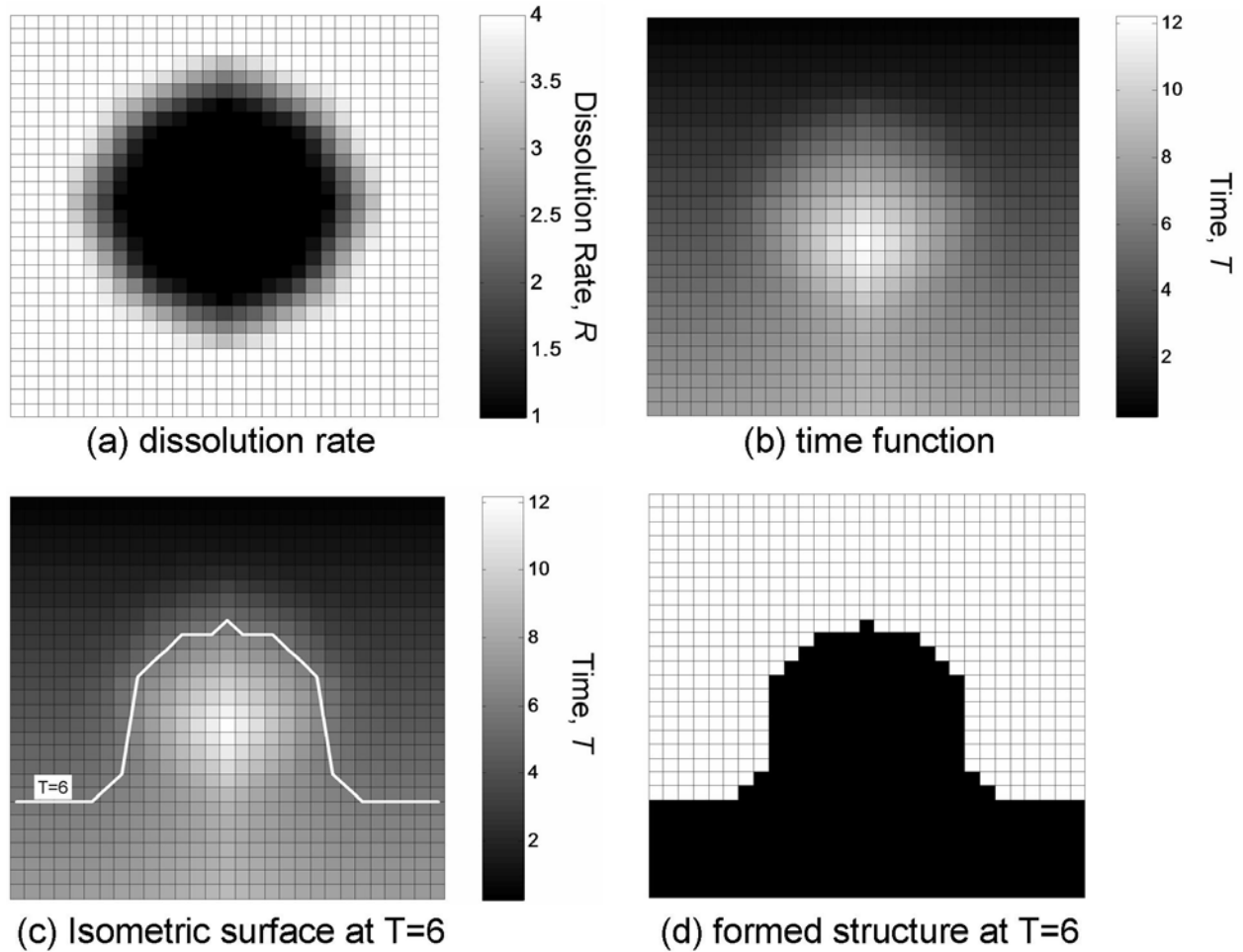
Figure 9-9: Simulating developing with FMM

When simulation completes, the resist profile at any time during developing may be determined by constructing an isometric surface from the time function. For example, resist profile can be determined after developing for 6 seconds. First, the isometric surface at T=6 is constructed as depicted in Figure 9-9(c). From here, all points with greater or equal time values

are not yet dissolved and will remain intact after developing. The resist profile constructed from the isometric surface is shown in Figure 9-9(d).

9.7. Numerical and Experimental Results

9.7.1.   Lattice Chirp

In this dissertation, it was predicted for the first time theoretically that holographic lithography will always form photonic crystals with chirped lattices. Predicting lattice geometry based on optical absorption, image contrast, and solubility is summarized in Figure 9-10. In this diagram, exposure dose of a holographic exposure as a function of depth is superimposed onto a typical solubility curve of a high-contrast negative-tone photoresist. The horizontal axis represents absorbed energy while the vertical axis represents dissolution rate for the dissolution rate curve, and depth for the exposure dose curve. The photonic crystal shown to the right was fabricated in Futurrex® NR5-8000 by Divlianski et al [162] and exhibits a strong chirp due to optical absorption.

From the dissolution rate curve, it is clear that solubility remains relatively unchanged until a critical amount of energy is absorbed. Portions of photoresist exposed with a dose in this range will thoroughly dissolve during developing. As more energy is absorbed, the photoresist enters the notch region where solubility falls rapidly. Analog structures are possible with doses within this region [173]. Beyond the notch region, photoresist is saturated and remains completely intact after developing.

The line representing exposure dose, shows absorbed energy is modulated with depth due to beam interference. Optical absorption produces the tapered envelope evident on this curve that reduces image contrast as well as total dose with depth. To reliably form a photonic crystal, the image contrast must completely span the entire notch region. If this is not the case the crystal

will either completely dissolve or will remain insoluble. When image contrast is shifted to the

left of the notch region, more resist dissolves leading to a photonic crystal with lower fill factor.

When image contrast is shifted to the right, less resist will dissolve producing a photonic crystal

with higher fill factor. Optical absorption tapers the alignment of image contrast with the notch

region producing a photonic crystal where fill factor varies with depth. High absorption causes

the most severe chirp.
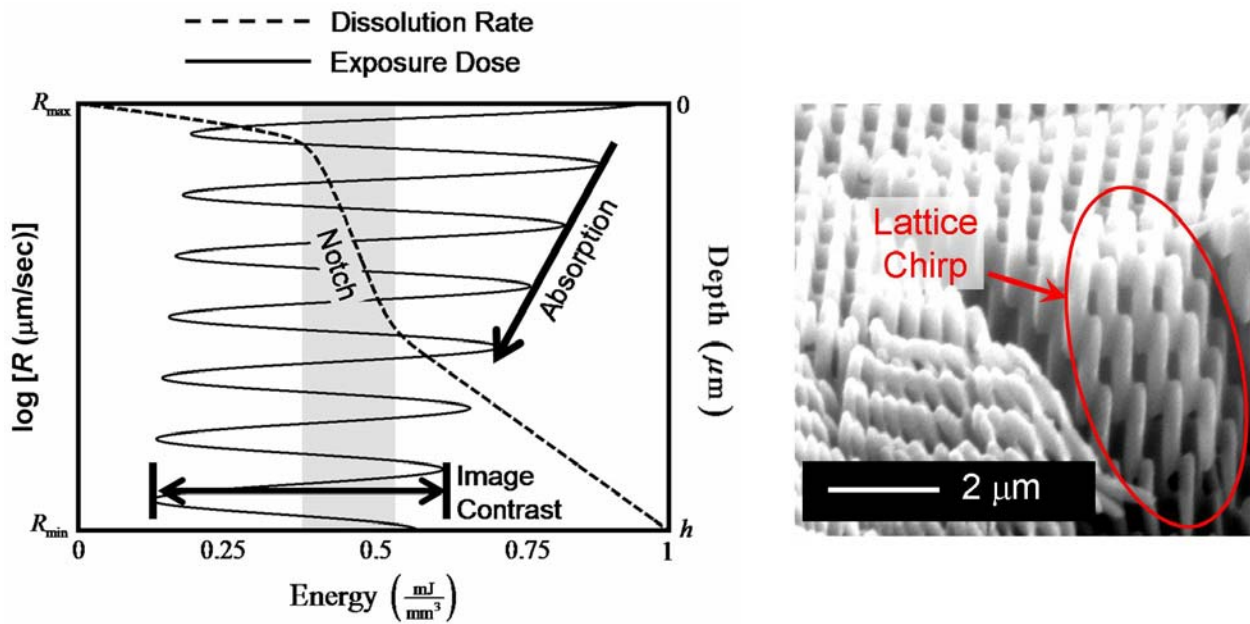
Figure 9-10: Predicting lattice chirp

9.7.2.   *Comparison with Intensity-Threshold Scheme*

A photonic crystal may be formed by exposing SU-8 ($n=1.67$) through a phase grating

like that in Figure 9-11. This grating was designed for operation at 365 nm and can be formed in

PMMA using electron beam lithography. Grating period determined the beam angles. Grating

pattern and thickness determined the relative power in each beam (diffraction efficiency). To achieve good image contrast, these were adjusted until diffraction efficiency of the dominant modes was approximately balanced. More rigorous optimization would consider image contrast, polarization, or quality of the formed structure itself. The optimized film thickness was found to be 525 nm for near-normal incidence.
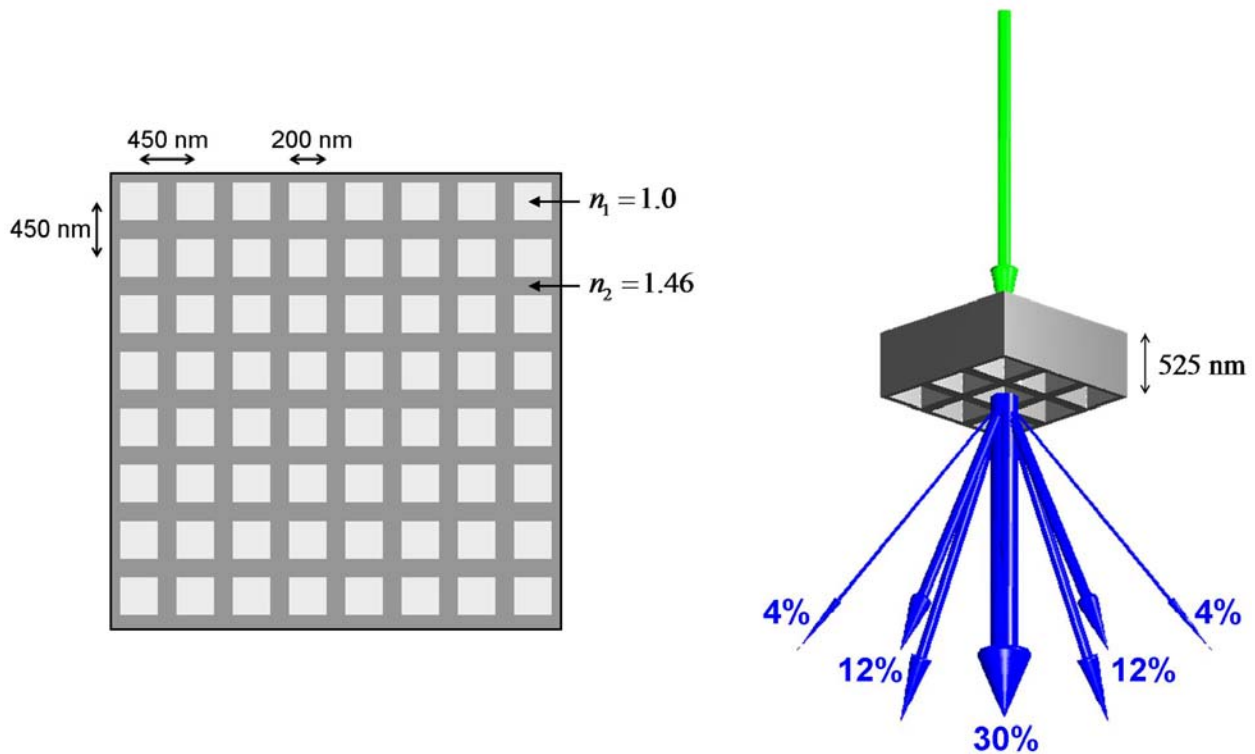
Figure 9-11: Example phase grating and diffracted orders

Diffraction through the mask is illustrated in the right diagram of Figure 9-2 where arrow thickness indicates relative power in the mode. This grating produces five dominant modes, but others exist and must be included for realistic simulation. Reflected modes are not shown in the figure, but were accounted for in simulation.
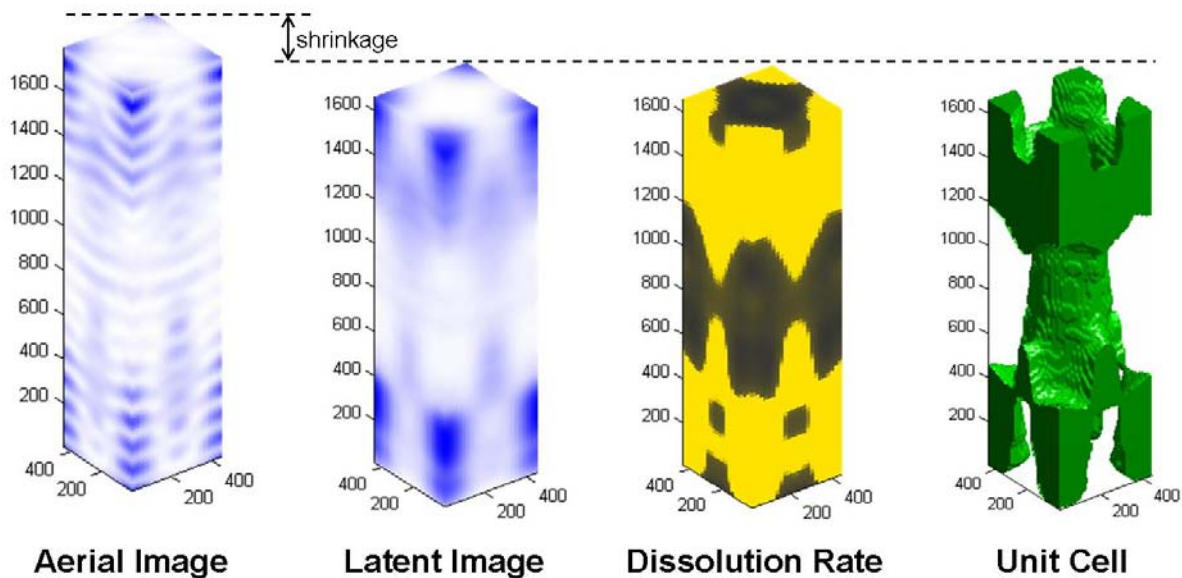
Figure 9-12: Output of comprehensive model at various stages of simulation

Figure 9-12 shows the output of each stage of the comprehensive model. A standing

wave is apparent in the aerial image suggesting reflections from the fused silica substrate

(*n*=1.52). Acid diffusion during the bake tends to counteract this effect by blurring the aerial

image. Vertical dimensions have been reduced by 7.5% due to shrinkage during cross linking

[159]. The third image in Figure 9-12 shows dissolution rate throughout the photoresist

computed from the latent image. Clearly, SU-8 is a very high contrast photoresist leading to few

partially soluble regions. Symmetry appears to be body-centered-tetragonal. Transmission and

reflection spectra from 10 layers of this photonic crystal are shown in Figure 9-13.
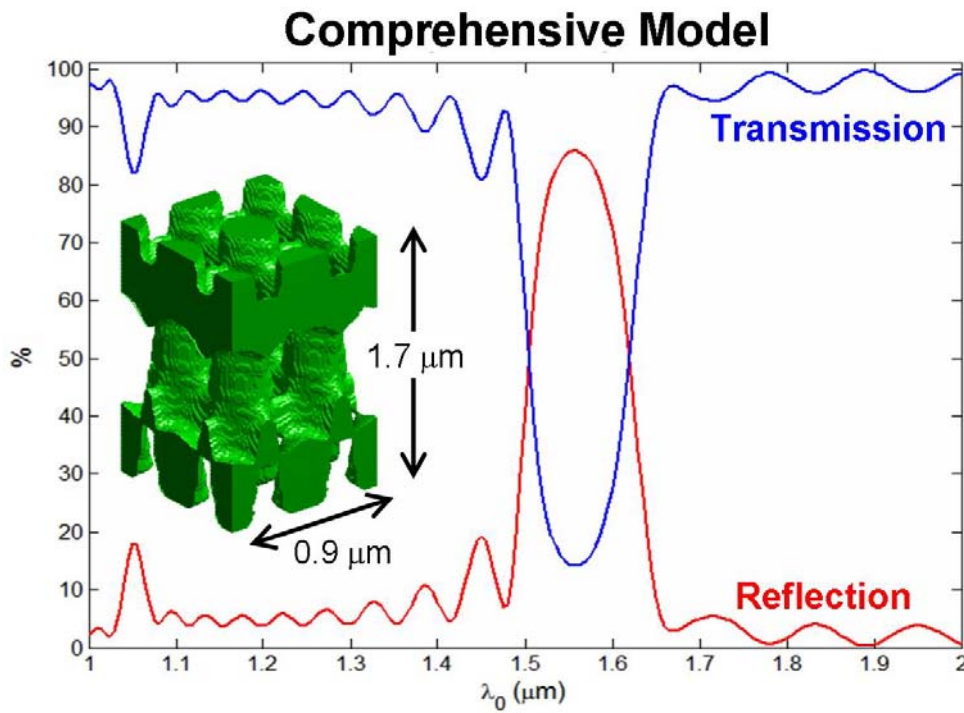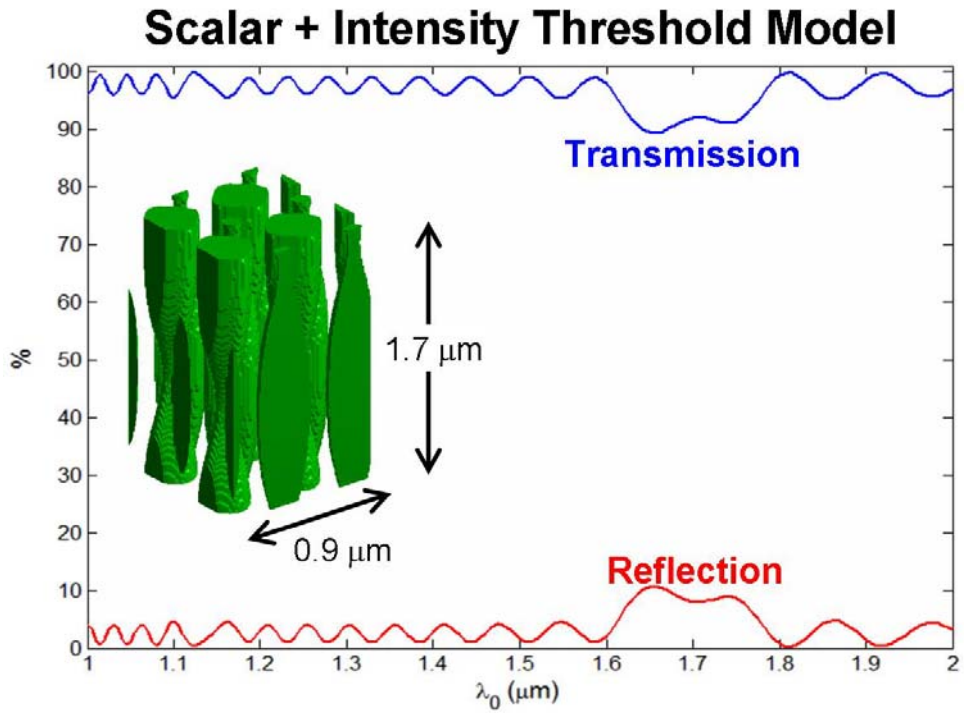
Figure 9-13: Transmission and reflection spectra through 10 layers of photonic crystal

To compare comprehensive modeling to a simple intensity-threshold scheme, irradiance was computed using a scalar sum of the five dominant modes produced by the grating. Intensity-thresholding assumes all portions of photoresist exposed with an irradiance above some threshold will become insoluble and remain intact after developing. A threshold was chosen that produced a photonic crystal with the same fill factor (42%) as comprehensive modeling. Otherwise, numerical results would have been even more disparate. Both diagrams in Figure 9-13 right diagram show a portion of the photonic crystal along with transmission and reflection spectra from a 10 layer slab.

The lattice predicted by intensity-thresholding is not connected at the chosen threshold. Both source polarization and acid diffusion contribute significantly to lattice connectivity. Since these factors are ignored, intensity threshold schemes tend to predict higher fill factor photonic crystals because thresholds must be extended until lattices are connected. This tends to push spectral behavior to longer wavelengths due to higher fill factor. Optical properties can be greatly affected even by subtle error in lattice geometry. For example, the photonic crystal predicted by comprehensive modeling shows a strong band gap in the vicinity of 1550 nm, while the other does not. This indicates more rigorous modeling of fabrication is necessary to accurately predict shape of structures and optical behavior.

9.7.3.   *Exposure Using Standard UV Lamps*

A very attractive feature of NFNP is that 3D structures can be formed using illumination from standard ultra-violet lamps commonly used in photolithography. These sources typically provide partially coherent light that is "unpolarized." Due to lamps being of finite size, illumination is not purely plane-wave and an angular spectrum must also be considered.

202

$$LP_x + \cdots + LP_y + \cdots + CP \approx \text{Unpolarized}$$
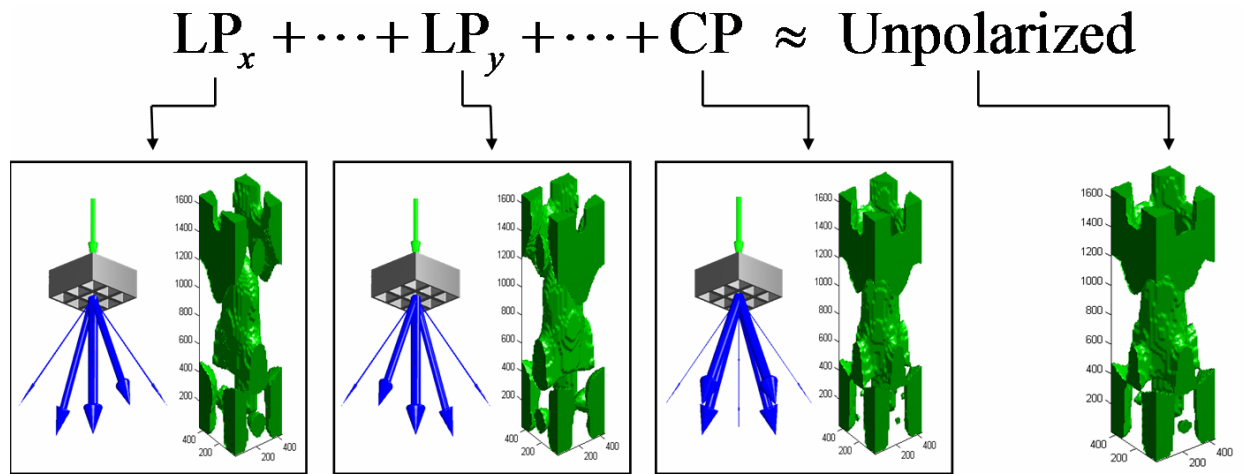
Figure 9-14: Approximation of "unpolarized" light source

Through simulation it was shown that photonic crystals can be formed in the configuration described previously using unpolarized light. This greatly simplifies hardware used for exposure. Figure 9-14 shows how unpolarized light improves lattice connectivity by effectively summing exposures of different polarizations. For this grating, any linear polarization alone does not produce a connected lattice, while unpolarized light does.
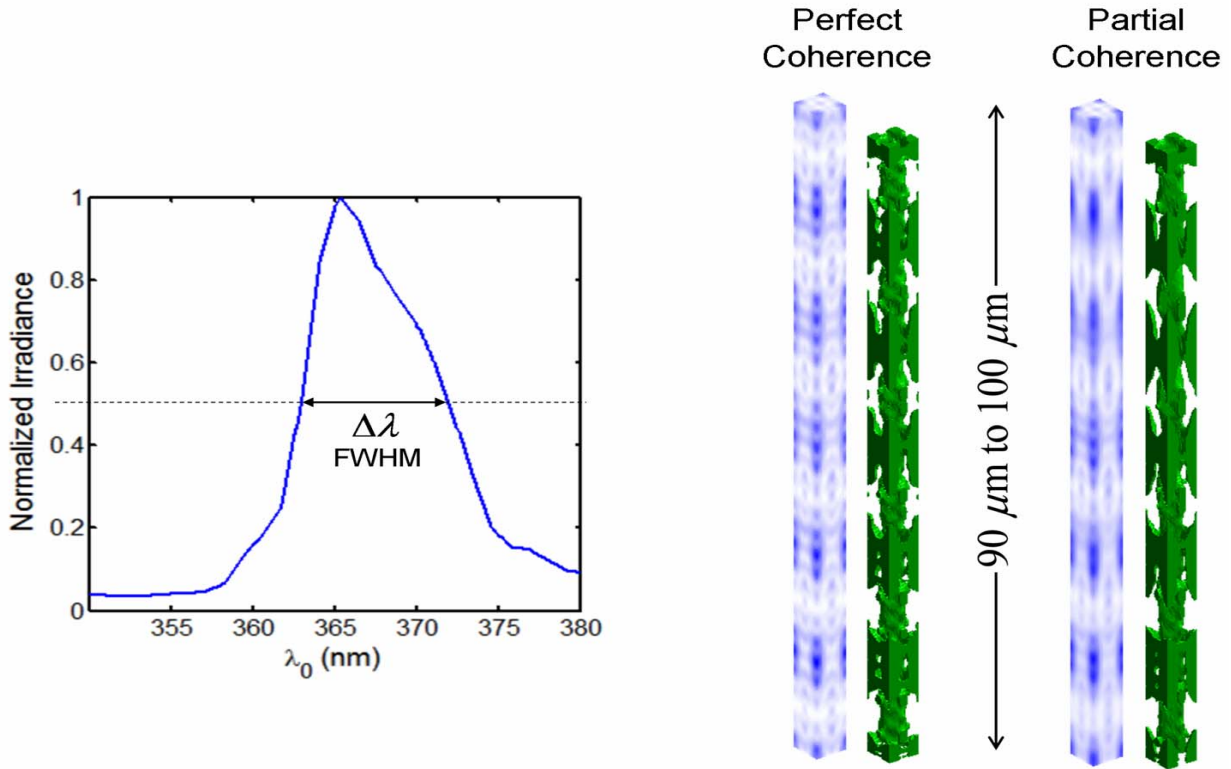
Figure 9-15: Near-field nano-patterning using partially coherent light

Due to dramatically relaxed requirements on beam coherence, 3D structures may be formed using partially coherent light. Figure 9-15 left diagram shows the typical i-line spectrum of a high pressure mercury vapor lamp. It was assumed a narrowband i-line filter was used to remove additional lines from the spectrum. Figure 9-15 right diagram compares aerial images and photonic crystals produced by perfectly coherent and partially coherent light sources. To show useful interference exists far from the grating, the bottom 10 μm of a 100 μm film is depicted in this figure. Partial coherence produces a smoothing effect, but still forms an excellent photonic crystal similar in appearance to the perfectly coherent case. Partial coherence can also contribute to lattice connectivity.
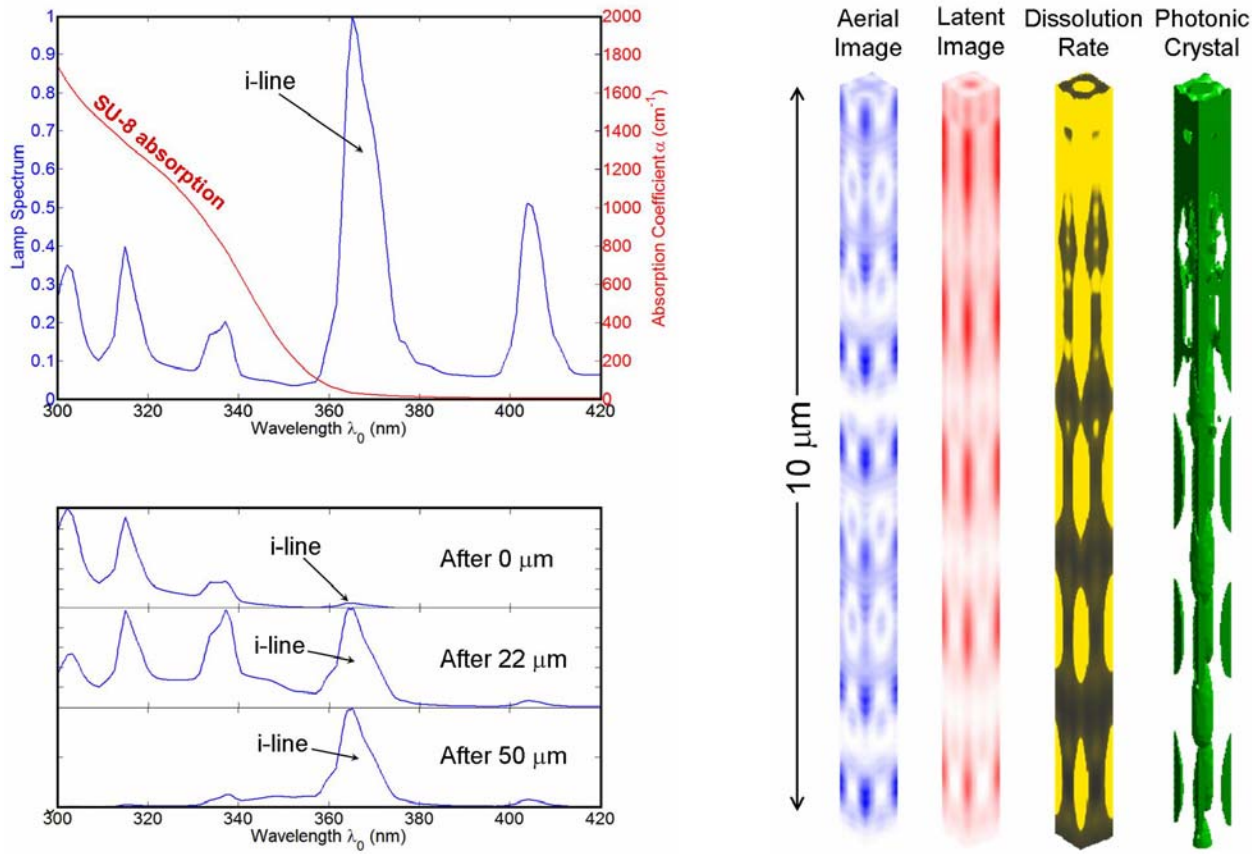
Figure 9-16: NFNP using unfiltered light from typical mercury-vapor lamp

Simulations were performed to explore the effect of using unfiltered light. Figure 9-16 overlays the absorption spectrum of SU-8 against the output spectrum of a typical UV lamp. Despite the i-line being most intense, spectral lines at shorter wavelengths dominate exposure when absorption is considered. This may be predicted using Eq. (9.38). To quantify this effect, "weighted" spectra were calculated by multiplying source irradiance by the absorption coefficient. These are shown in the lower left diagrams in Figure 9-16. The uppermost plot clearly shows spectral lines at shorter wavelengths dominate exposure near the surface because they contribute most to absorbed energy there. Irradiance at the shorter wavelengths, however, decays quickly with depth changing the balance of which lines are dominant. At a depth of 22

μm, the spectral lines contribute about equally to absorbed energy. At 50 μm and greater, the i-line dominates exposure.

The rightmost images in Figure 9-16 show the output after each stage of comprehensive simulation for unfiltered light. Bleaching was ignored at the shorter wavelengths. The aerial image is a measure of irradiance where the i-line is dominant. The latent image is a measure of absorbed energy where which line dominates changes with depth. The latent image is shown to differ greatly from the aerial image, especially near the surface of the film where shorter wavelengths dominate exposure. A sharp contrast in solubility can be observed near the surface of the film. The rightmost image in Figure 9-16 was constructed from isometric surfaces of dissolution rate. The developing model did not show any structure because the top portion of the film was completely insoluble.

The result of using unfiltered light is a region of very high absorbed energy near the surface of the photoresist film. This can lead to the well known "T-topping" or barb-shaped structures that overhang vertical sidewalls. In practice, it would be very difficult to control exposure to form a photonic crystal with unfiltered light. When dose is chosen to form a photonic crystal through most of the volume, the surface of the film would become insoluble and not allow developer to pass through. No photonic crystal would form if dose were lowered to form openings at the surface. With this in mind, it should be possible to form interesting and useful structures by tailoring the relative irradiance of the spectral lines with filters or using photoresist with different absorption properties.

9.7.4.   Angular Spectrum

A critical process parameter to NFNP is the angular spectrum of source illumination. Perfect plane-wave illumination during exposure is ideal, but realistic optical systems suffer

206

from diffraction, scattering, and the penumbra effect where finite size sources produce rays at different angles. Off-axis rays produce tilted aerial images that combine to blur the overall aerial image. Since tilt origin is at the grating, blurring effects increase in severity with distance away from the grating.
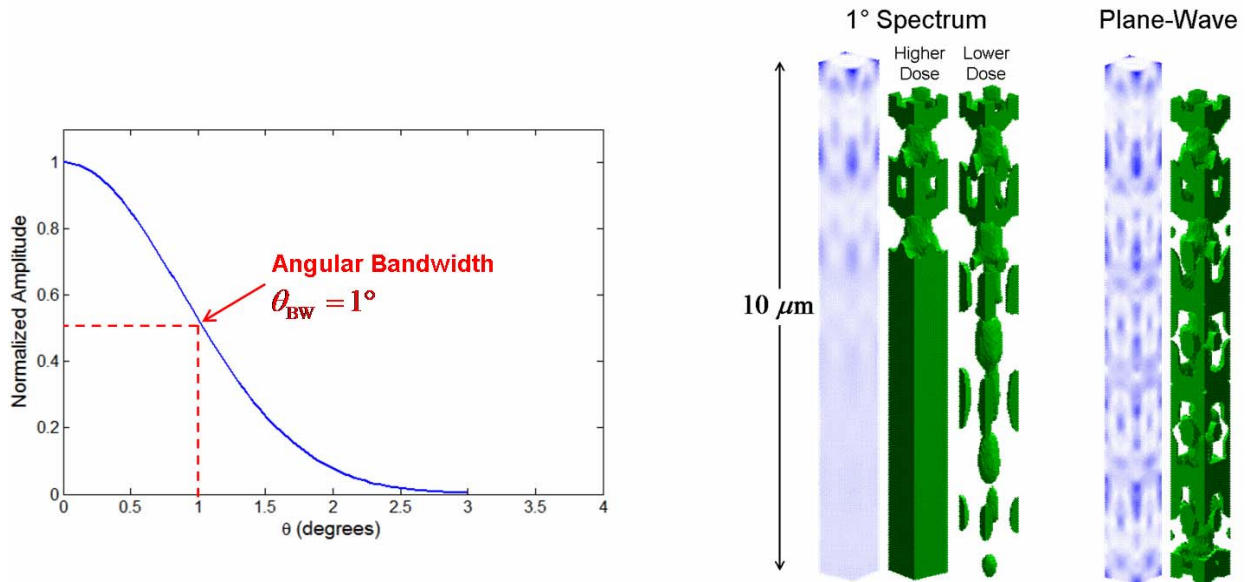
Figure 9-17: Impact of angular spectrum

As an example, angular spectrum was incorporated in the exposure model and results summarized in Figure 9-17. The curve in Figure 9-17 describes the assumed angular dependence of irradiance for an angular bandwidth around 1°. Aerial images and resulting photonic crystals under one grating period are also shown in Figure 9-17. Angular divergence has clearly blurred and "washed out" interference fringes in the lower part of the aerial image. Within a narrow range of dose, two outcomes were possible. For slightly higher doses, the bottom portion was not soluble and a photonic crystal only formed near the surface. For slightly lower doses, the

bottom portion remained soluble and released the photonic crystal from the substrate. When a 5 μm gap was inserted between the grating and photoresist, no crystal formed at all. Under plane-wave illumination, the photonic crystal was virtually uniform with depth. Perhaps this could serve as a mechanism for controlling lattice chirp or compensating for chirp caused by optical absorption during exposure.
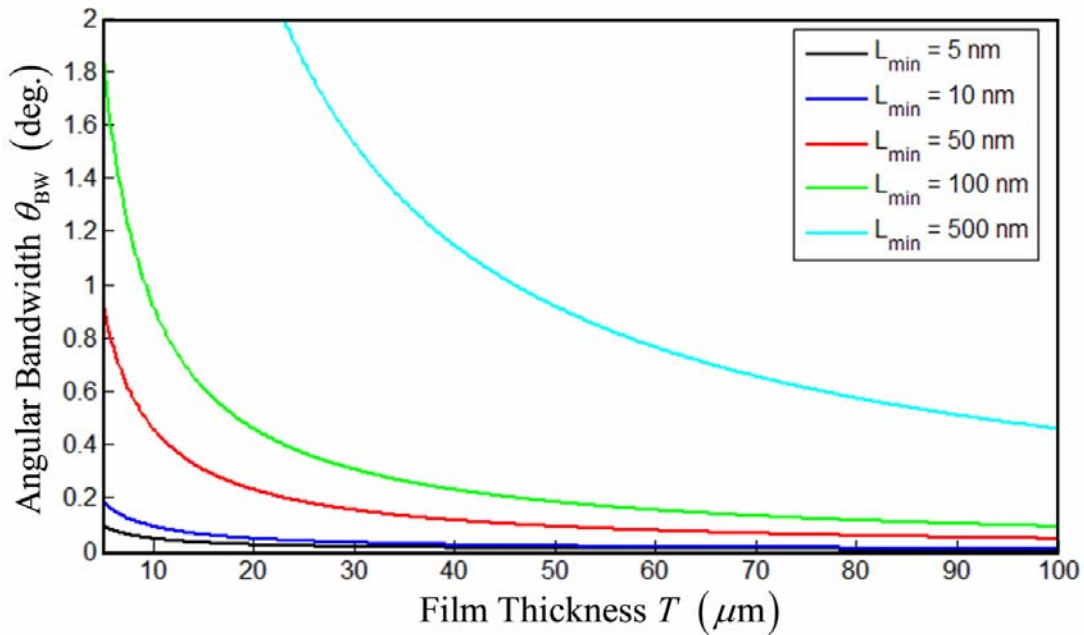
Figure 9-18: Parametric curve for limiting angular spectrum

Bandwidth of angular divergence should be made small as possible. This is usually a fraction of a degree. A rule-of-thumb can be derived from the geometry of the problem. Given minimum feature size of the resulting crystal L_{min}, refractive index of the photoresist n, and film thickness T, angular bandwidth should be limited to

$$\theta_{BW} < \frac{nL_{min}}{T} \tag{9.44}$$

To illustrate this point, parametric curves were calculated from Eq. (9.44) and are shown in Figure 9-18. The curves show maximum angular bandwidth versus film thickness for five minimum feature sizes. A quality exposure should limit angular bandwidth of the source to angles below the curve. Thicker films and smaller feature size place stricter limits on angular bandwidth.

Soft Bake (Oven)

```
3°C/min to 65°C. 25 min hold.
3°C/min to 95°C. 30 min hold.
3°C/min to room temp.
```

Exposure

```
Photoresist: SU-8 2007 (7 μm film)
Wavelength: 365 nm
       Dose:  40 mJ/cm²
```

Post Exposure Bake (Hot Plate)

```
65°C for  2 min.
95°C for 90 sec.
65°C for  3 min.
```

Develop

```
     Time: 4 minutes
Developer: PGMEA
```

Figure 9-19: Fabrication of 3D photonic crystals in SU-8 2007

9.7.5.   Photonic Crystals Formed by NFNP

Through comprehensive modeling of the process, it was predicted that it should be possible to form three-dimensional photonic crystals on a standard mask aligner if a filter is installed to remove spectral energy below the i-line. Experimental results validated this prediction when photonic crystals were fabricated for the first time using a Quintel Ultraline 7000 with an i-line filter. An image of the tool and the process recipe is provided in Figure 9-19.

It was found that an extended soft bake before exposure was necessary to prevent cracking. Exposure dose was typically around half of what was recommended in the data sheet. When a photonic crystal was formed, a brilliant iridescence could be observed after developing.

Figure 9-20 shows four images of photonic crystals formed in SU-8 along with the lattices predicted by comprehensive modeling. Images of the photonic crystals were captured using an atomic force microscope [5, 6].
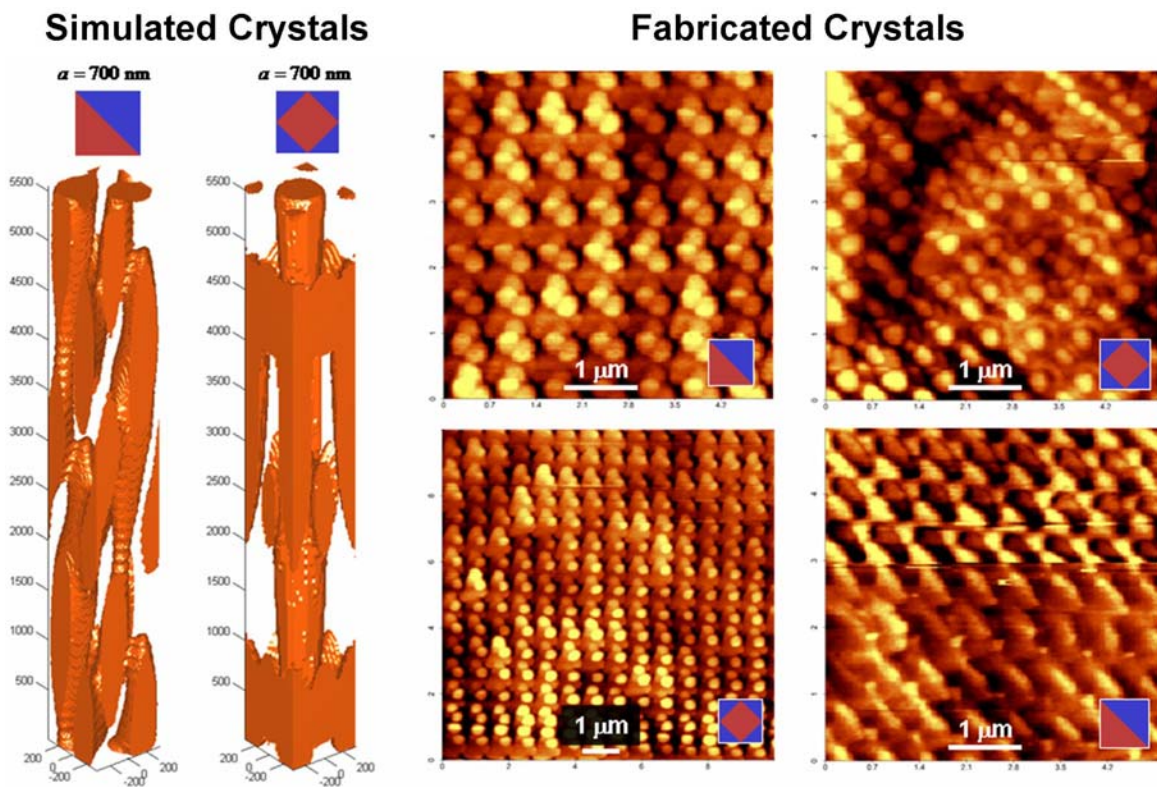
Figure 9-20: Three dimensional photonic crystals formed by NFNP

CHAPTER 10

DESIGN AND OPTIMIZATION OF SPACE-VARIANT PHOTONIC CRYSTAL FILTERS

10.1. Overview

There is a need for easily fabricated devices that can be placed over a detector array to perform filtering functions uniquely tuned for each pixel. This concept is illustrated in Figure 10-1 where a color imaging array is formed by fabricating a space-variant photonic crystal on top of a standard charge-coupled-detector (CCD) array. While the CCD itself is colorblind, fabricating a color filter over the array enables adjacent pixels to detect different colors. The passband of the photonic crystal is tuned across the device aperture by adjusting the diameter of holes etched through the volume. Devices incorporating polarization are possible, but were not considered in this dissertation.

Realizing this concept with conventional multilayer films would require film thickness to be varied across the array. This could only be achieved using a complicated fabrication process with many steps and difficult alignments. A photonic crystal approach has been conceived that greatly simplifies fabrication. Filter response at each pixel may be custom tuned by adjusting hole size in the vicinity of each pixel. This only requires adjustment of transverse dimensions and can be realized with a single mask process. An additional advantage of a photonic crystal is the potential for better performance at oblique angles of incidence due to modulation of the refractive index in three dimensions. Achieving angular tolerance in a multilayer film requires adding additional layers. Thicker films introduce more image distortion, aberrations, and defocusing of the image. Photonic crystals promise easier fabrication, thinner devices, and better performance at oblique angles.
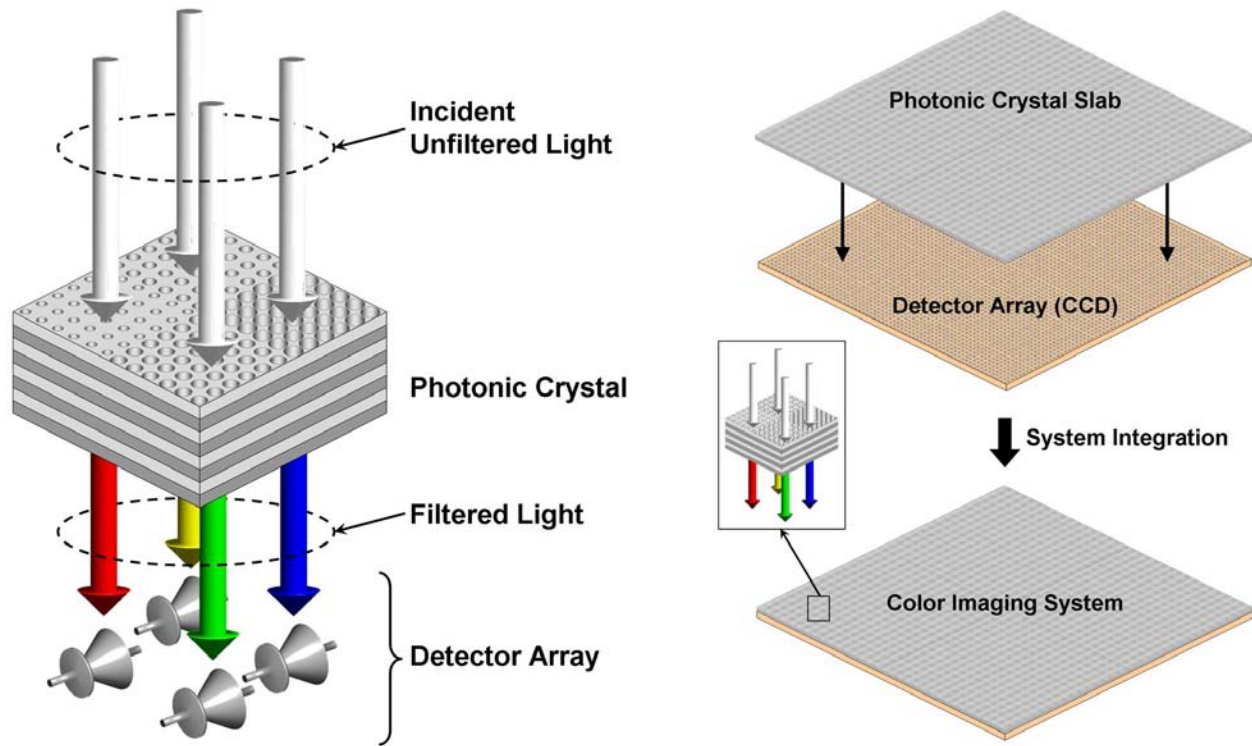
Figure 10-1: System concept for color imaging system

This chapter investigates design and optimization of space-variant photonic crystal filters based on a very simple fabrication approach. It will be shown the main distorting artifact of fabrication is a hole radius that is not uniform depth. Devices will be designed based on two separate material systems and methods of compensating for hole taper will be addressed.

10.2. Fabrication of Photonic Crystal Devices

The following sections describe simple methods for fabricating three-dimensional photonic crystals in either SiO_2/SiN or $GaAs/Al_xGa_{1-x}As$ material systems. They are both essentially any array of holes etched through hetero-structured substrates. Spectral response of both photonic crystals may be tuned across the device aperture by varying hole size.

10.2.1. GaAs/Al$_x$Ga$_{1-x}$As Fabrication Process

Hetero-structured GaAs/Al$_x$Ga$_{1-x}$As wafers grown using a metalorganic chemical vapor deposition (MOCVD) process were obtained from *Bandwidth Semiconductor, LLC* [183]. Each layer was 250 nm thick with a total of 20 layers (10 periods) grown on an n-type GaAs substrate. The fractional aluminum concentration was *x*=0.7. To prepare the wafers for device fabrication, they were cleaned in a 10% solution of ammonium hydroxide for one minute, then dipped into buffered oxide etchant to remove any oxides along with the native oxide layer. The entire fabrication process is illustrated in Figure 10-2.
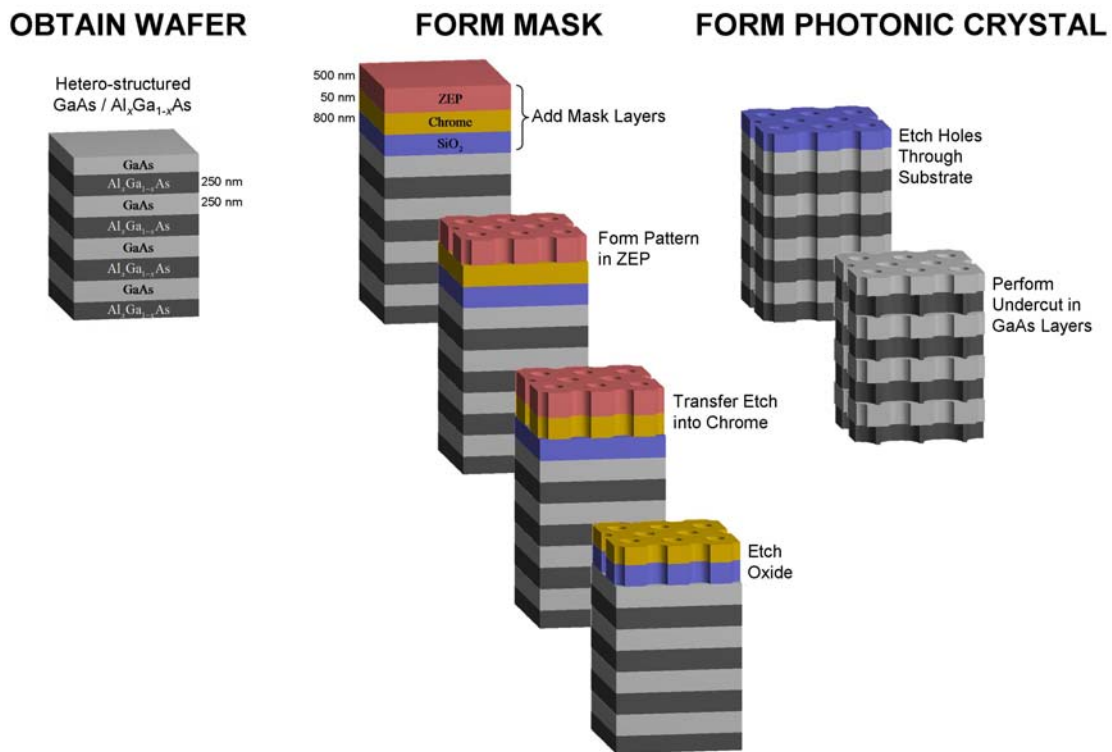
Figure 10-2: Fabrication of GaAs/Al$_x$Ga$_{1-x}$As photonic crystals

Fabrication began by forming a two-layer binary mask onto the surface of the samples. An 800 nm layer of SiO_2 was grown onto the sample using plasma enhanced chemical vapor deposition (PECVD). This was followed by thermally evaporating 50 nm of chrome onto the surface. A 500 nm film of positive e-beam resist ZEP [184] was spun onto the surface and patterned with an array of holes using a Leica EBPG 5000+ writing tool [185]. The resist was developed in ZEP-RD [184] for 90 seconds. The pattern formed in ZEP was transfer etched into the chromium layer using chromium etchant type TFD [186]. The pattern was subsequently transfer etched into the oxide layer using a CHF_3/Ar inductively coupled plasma (ICP). This step also removed the remaining ZEP. Structures obtained in the mask layer are shown in Figure 10-3.
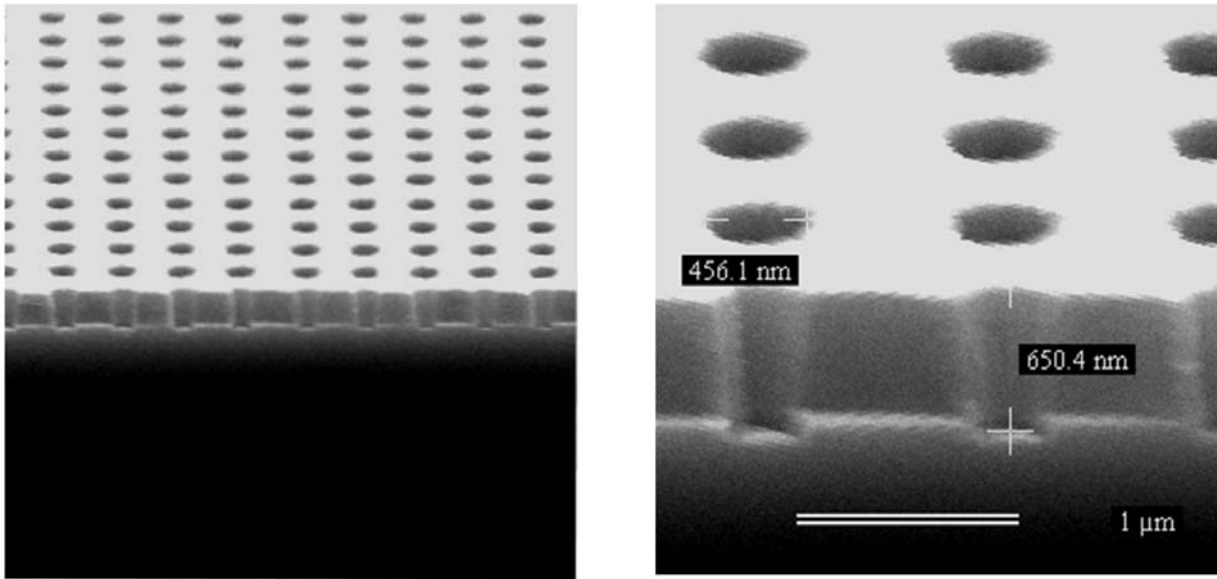
Figure 10-3: SEM images of Cr/SiO$_2$ mask

Substrate temperature was held at 20°C while vertical holes were etched through the GaAs and $Al_xGa_{1-x}As$ layers using a BCl_3/Ar plasma. The chrome layer was also removed

during the etch due to the presence of chlorine. SEM images of the resulting structures are shown Figure 10-4. A slight difference in the etch rate of GaAs and $Al_xGa_{1-x}As$ can be observed in the first few layers producing a small undercut.
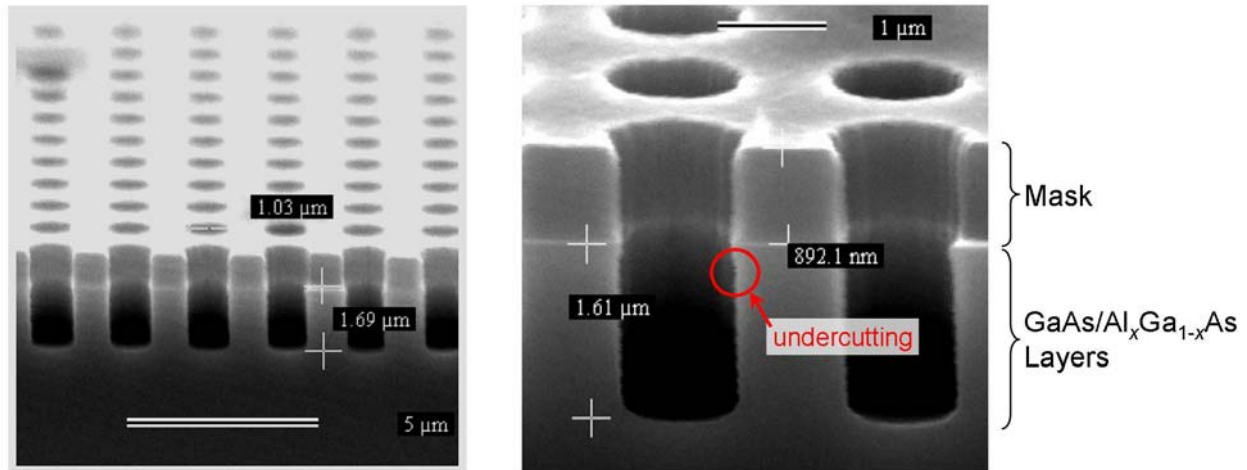
Figure 10-4: Vertical holes formed in GaAs/Al$_x$Ga$_{1-x}$As layers

While still in the vacuum chamber, SF_6 was added to the plasma chemistry to inhibit etching of the $Al_xGa_{1-x}As$ layers. Passivation of the $Al_xGa_{1-x}As$ layers enabled the GaAs layers to etch about six times faster than the $Al_xGa_{1-x}As$ layers. Etch selectivity could be controlled through the flow rate of the SF_6. Some dependence of the undercutting was observed from ICP power and bias voltage. Most of the oxide layer is removed during this step due to the presence of SF_6, but a subsequent etch in CHF_3/O_2 was performed to remove any residual oxide. A SEM image of the undercutting is shown in Figure 10-5.
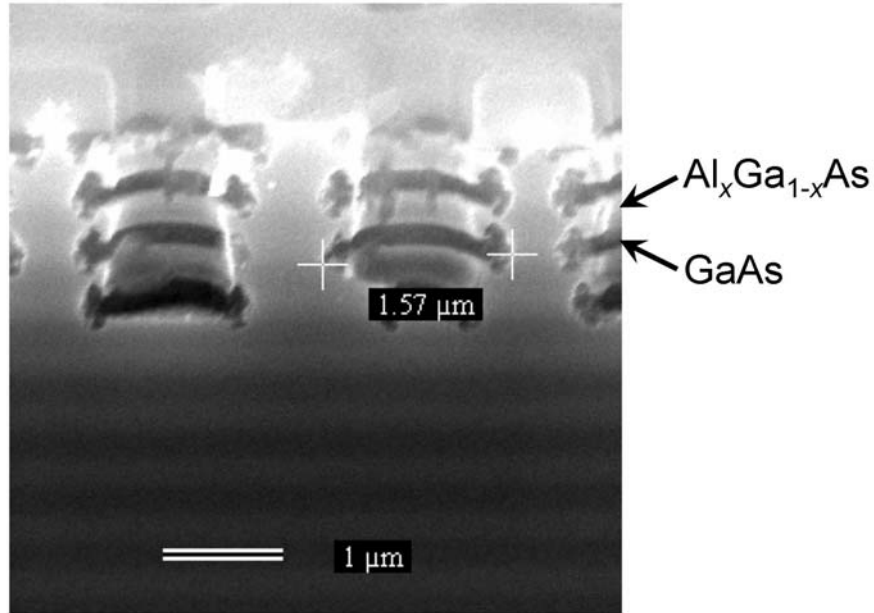
Figure 10-5: SEM image of undercut photonic crystal

10.2.2.   SiO2/SiN Fabrication Process

A similar process was used to fabricate photonic crystals in SiO_2/SiN except that the final

undercut step was not performed. Index contrast between SiO_2 and SiN was sufficient to

produce a photonic band gap for light that was normally incident.

10.3. Device Design

To determine a good starting point for device geometry, properties of the ideal bulk

photonic crystals were optimized before incorporating defects or considering fabrication.

Properties included lattice symmetry, hole size, duty cycle of the layers, degree of undercutting,

and minimum number of dielectric layers for sufficient suppression in the band gap. Next, the

defect was incorporated and device parameters optimized for maximum tuning range while

maintaining acceptable background suppression. Finally, fabrication was considered and taper in

hole radius was found to be the critical limiting factor. Two methods for compensation will be discussed.

10.3.1.   *Square Verses Hexagonal Arrays*

To produce strong suppression in the band gap and offer better performance at oblique angles of incidence, lattices should have a high degree of symmetry so waves see essentially the same perturbations regardless of their direction or polarization. Cubic and hexagonal lattices are well known for high degrees of symmetry. The basic geometry and parameter definitions adopted in this chapter are shown in Figure 10-6.
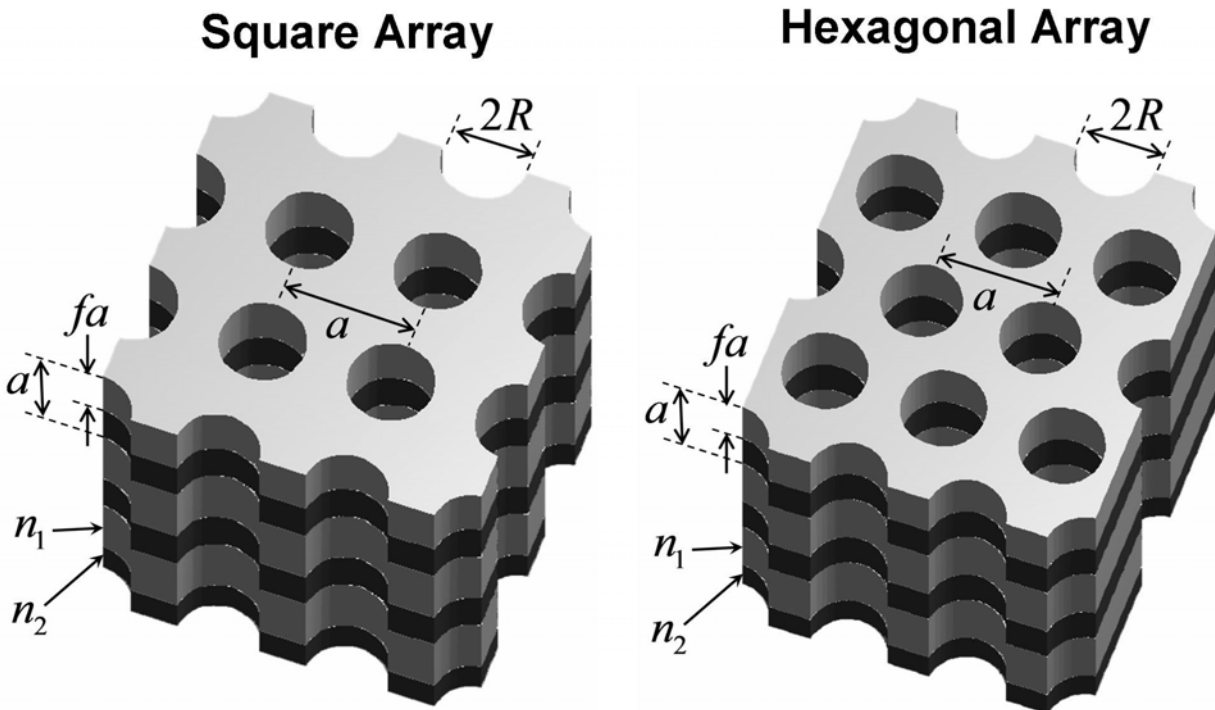
Figure 10-6: Parameter definitions for square and hexagonal arrays
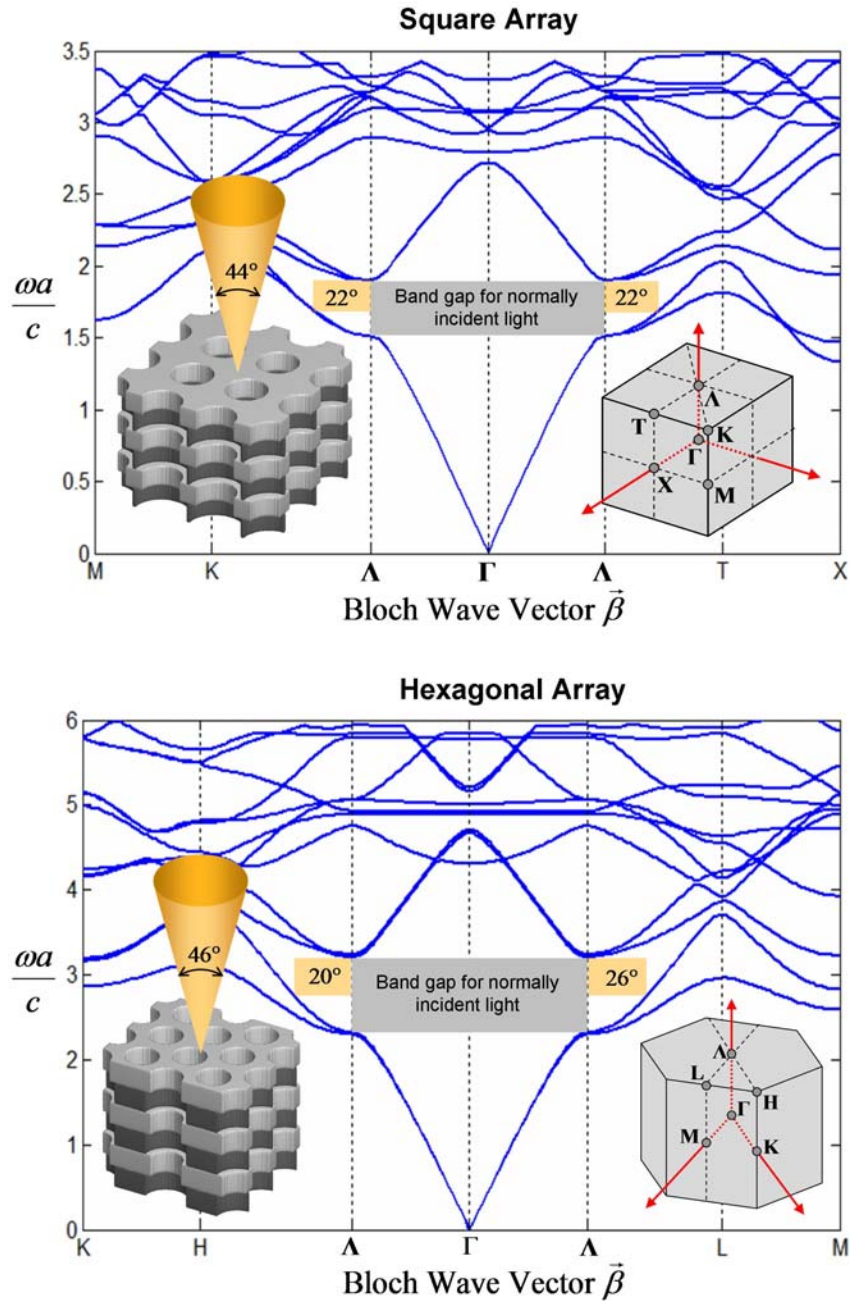
Figure 10-7: Photonic band diagrams for square and hexagonal arrays

To assess the difference between square and hexagonal arrays, their photonic band structures were calculated and are shown in Figure 10-7. Both symmetries provide similar performance in terms of a band gap and tolerance to oblique angles of incidence, but dimensions

of the hexagonal lattice may be nearly twice as large. Neither lattice provides a complete photonic band gap, but both show a partial gap for light propagating near parallel to the vertical axis. Width of the band gap is reduced if oblique angles are considered. As depicted in Figure 10-7, the upper half of the band gap is maintained over a cone of angles roughly 45° wide. From this observation, it was reasoned the structures were behaving much like a multilayer films where hole radius was used to control the effective refractive index of each layer.

10.3.2.   *Optimization of the Partial Photonic Band Gap*

To optimize layer duty cycle and hole radius, the data depicted in Figure 10-8 was generated for both SiO_2/SiN and GaAs/Al_xGa_{1-x}As material systems. Optimum layer duty cycle for the SiO_2/SiN based photonic crystal was found to be 0.58, but decreasing slightly for larger hole radius. For the GaAs/Al_xGa_{1-x}As based photonic crystal, optimum duty cycle was found to be 0.54 decreasing only slightly with increasing hole radius. To maintain a sufficient photonic band gap using SiO_2/SiN, hole radius should be made less than $0.4a$ and duty cycle should be kept between 0.3 and 0.8. Using GaAs/Al_xGa_{1-x}As without an undercut yields a similar conclusion, but width of the band gap is very small. Refractive indices were assumed to be SiO_2 (n=1.4458), SiN (n=1.9767), GaAs (n=3.5), and Al_xGa_{1-x}As (n=3.0). Material dispersion was not considered here because the devices were only required to operate over a short range of wavelengths around 1550 nm.
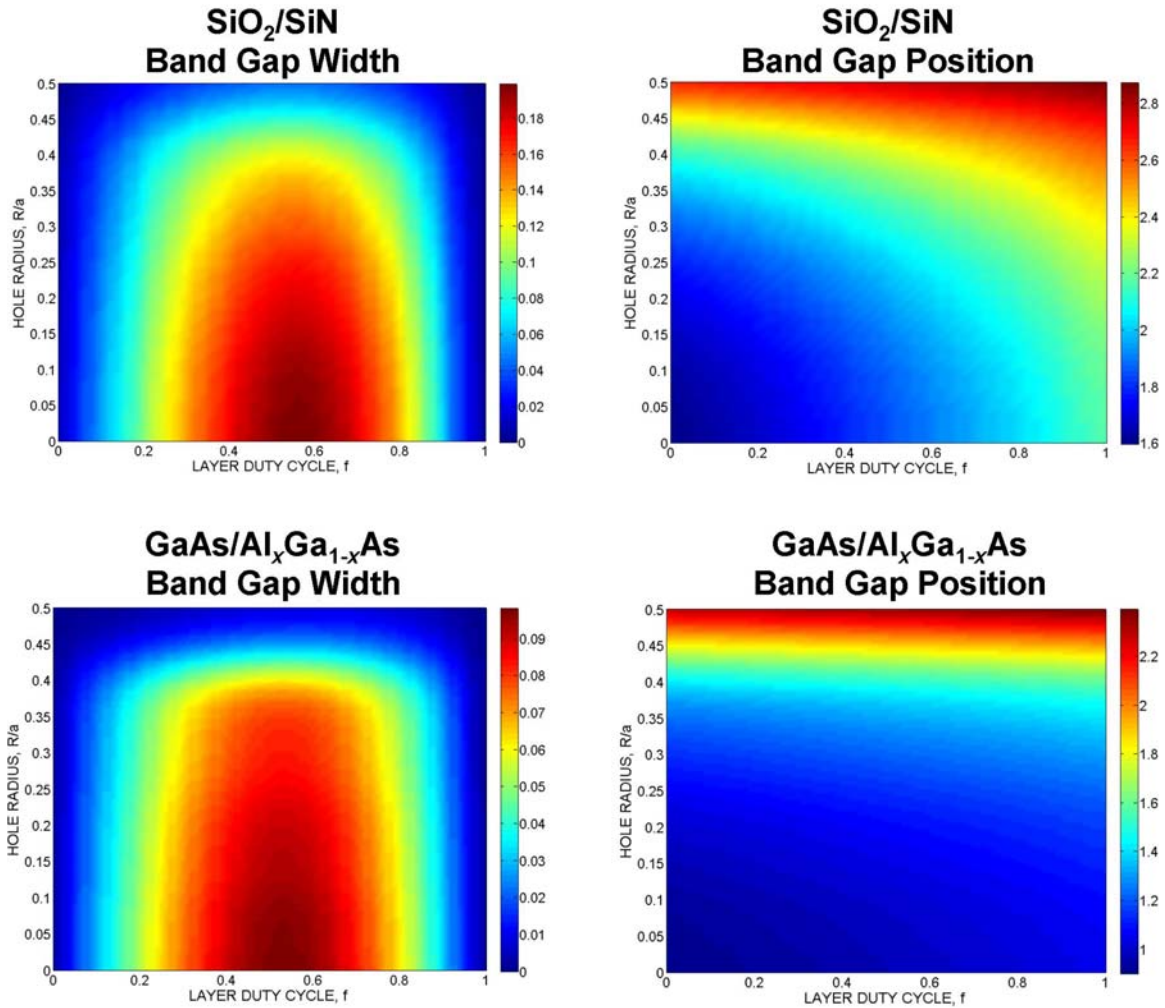
Figure 10-8: Optimization of hole radius and layer duty cycle

For the GaAs/Al$_x$Ga$_{1-x}$As based photonic crystal, GaAs layers may be preferentially etched over the Al$_x$Ga$_{1-x}$As layers to form an undercut. This raises index contrast, potentially leading to a wider photonic band gap. The data provided in Figure 10-9 was calculated to assess optimum hole radius for each layer. Hole radius in the GaAs layers must be greater than in the Al$_x$Ga$_{1-x}$As layers because it is the GaAs layers that are selectively etched. A dashed line indicates zero undercut. Below this line is not a realizable geometry using the prescribed process. It can be seen that for small degrees of undercut, width of the band gap actually

decreases. It is not until the undercut exceeds some breakeven value that undercutting widens the band gap. The ideal case is very small holes in the $Al_xGa_{1-x}As$ layers with very large holes in the GaAs layers. Given sufficient undercutting, it appears more critical to have large hole radius in the GaAs layers than it is to have small hole radius in the $Al_xGa_{1-x}As$ layers.
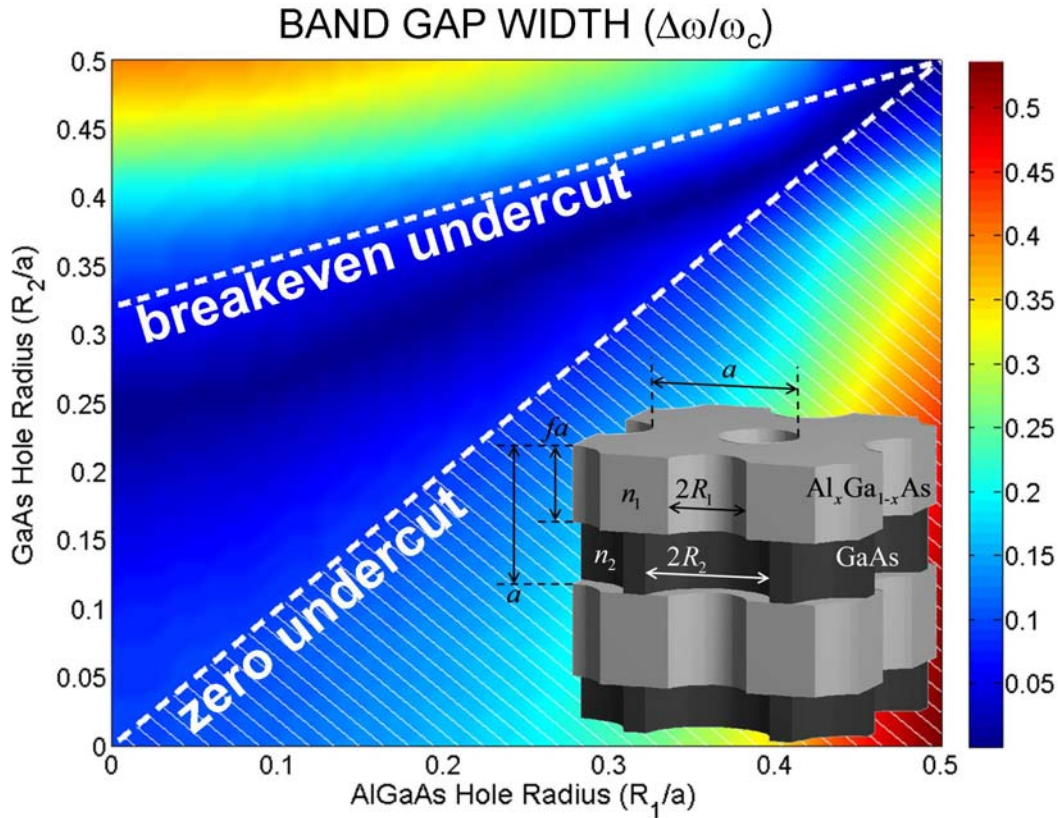
Figure 10-9: Optimization of GaAs undercut

Analysis to this point assumed an infinitely periodic photonic crystal. In practice, photonic crystals are of finite size. To determine how many layers are needed to provide sufficient suppression within the band gap, transmission through a photonic crystal slab with increasing number of lattice periods was calculated for both candidate lattices and summarized in

Figure 10-10. It can be concluded that both material systems require approximately 10 periods (i.e. 20 film layers) to achieve good suppression within the band gap. Based on previous analysis, hole radii in the SiO_2/SiN structure was chosen to be $r_1=r_2=0.3a$ and hole radii in the GaAs/Al_xGa_{1-x}As were chosen to be $r_1=0.2a$ and $r_2=0.4a$. For both material systems, duty cycle of the layer thicknesses was chosen to be $f=0.55$. In this diagram, two film layers comprise one longitudinal period of the photonic crystal.
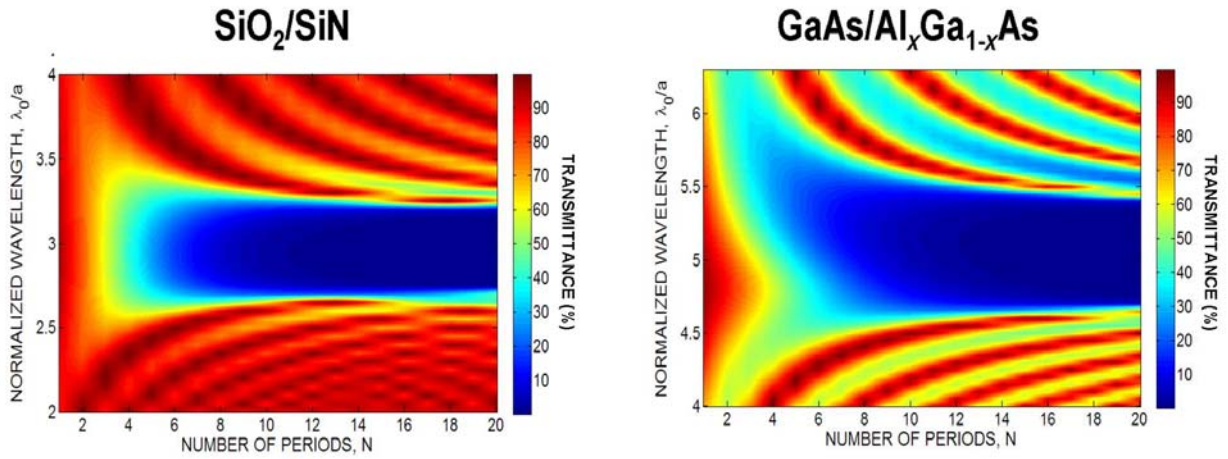
Figure 10-10: Optimization of number of layers (2 layers = 1 period)

These numerical results can be compared to an analytical equation for reflection from a multilayer stack of quarter-wave films [187]. For a stack comprised of $2N$ layers with refractive indices n_L and n_H, reflection is

$$R(N) = \left[\frac{\left(n_H/n_L\right)^{2N} - 1}{\left(n_H/n_L\right)^{2N} + 1} \right]^2 \tag{10.1}$$

Using this equation, a 20 layer stack (i.e. 10 lattice periods) with $n_L=1.4458$ and $n_H=1.9767$ provides over 99% reflection at the center of the band gap.

10.3.3. Incorporation of a Defect

To construct a device that functions in the manner depicted in Figure 10-1, a defect must be incorporated to produce a narrow transmission spike within the photonic band gap. Figure 10-11and Figure 10-12 illustrate how defect thickness affects the position and number of transmission spikes observed. Geometry of the bulk photonic crystal structure outside of the defect is as described above with 10 lattice periods on either side of the defect.
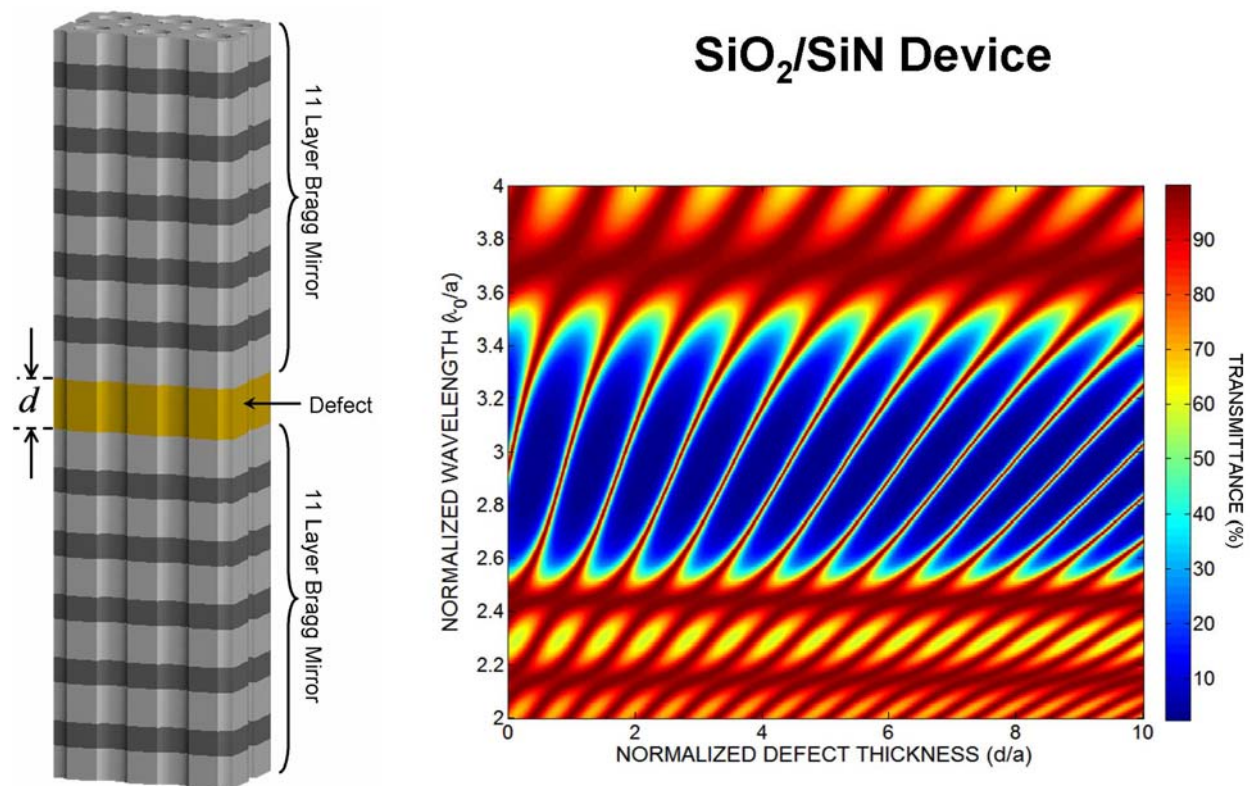

Figure 10-11: Incorporation of defect into SiO$_2$/SiN photonic crystal

Similar conclusions can be made for both material systems. When the defect is made thicker, the transmission peak moves to a longer wavelength. When the transmission peak reaches the edge of the band gap, a new peak emerges from the other edge. For thick defect

223

layers, position of the transmission peaks translate slower leading to multiple peaks present in the band gap. For most applications, a single transmission peak placed at the center of the band gap is desired. To achieve this, thickness of the defect layer should be around $1.17a$ in the GaAs/Al$_x$Ga$_{1-x}$As device and $1.0a$ in the SiO$_2$/SiN device.
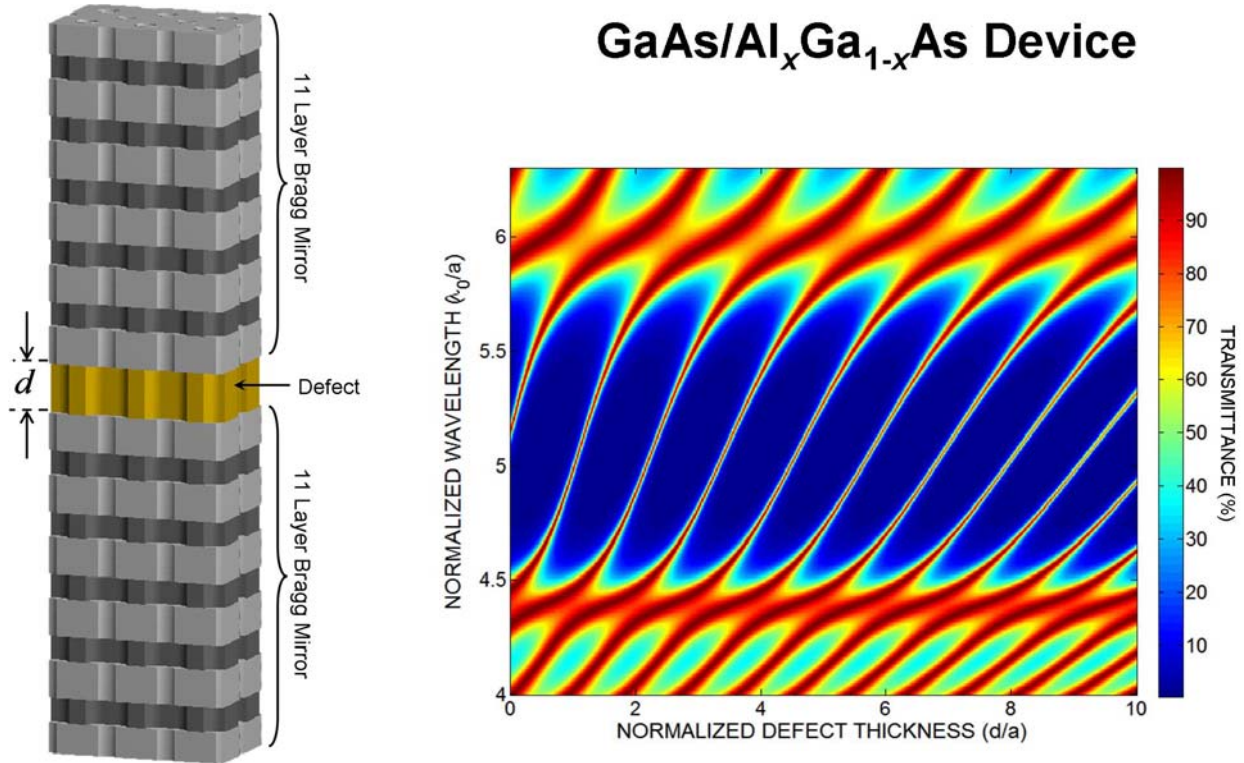
Figure 10-12: Incorporation of defect into GaAs/Al$_x$Ga$_{1-x}$As photonic crystal

While position of the transmission spike can be controlled by thickness of the defect layer, hole radius is the mechanism that should be used to tune its position across the device aperture. Position of the band gap and the transmission spike do not inherently shift evenly due to the distribution of fields at the transmission resonance. To compensate for uneven shift, thickness of the defect should be further optimized to achieve maximum tuning range. For this

reason, defect thickness of the GaAs/Al$_x$Ga$_{1-x}$As device was set to $d=1.3a$ instead of $d=1.17a$ as prescribed previously.

Tuning curves for adjusting hole radius in both material systems are depicted in Figure 10-13. The SiO$_2$/SiN material system clearly has a smaller tuning range, but lineshape is more uniform. Lineshape varies more severely in the GaAs/Al$_x$Ga$_{1-x}$As material system.
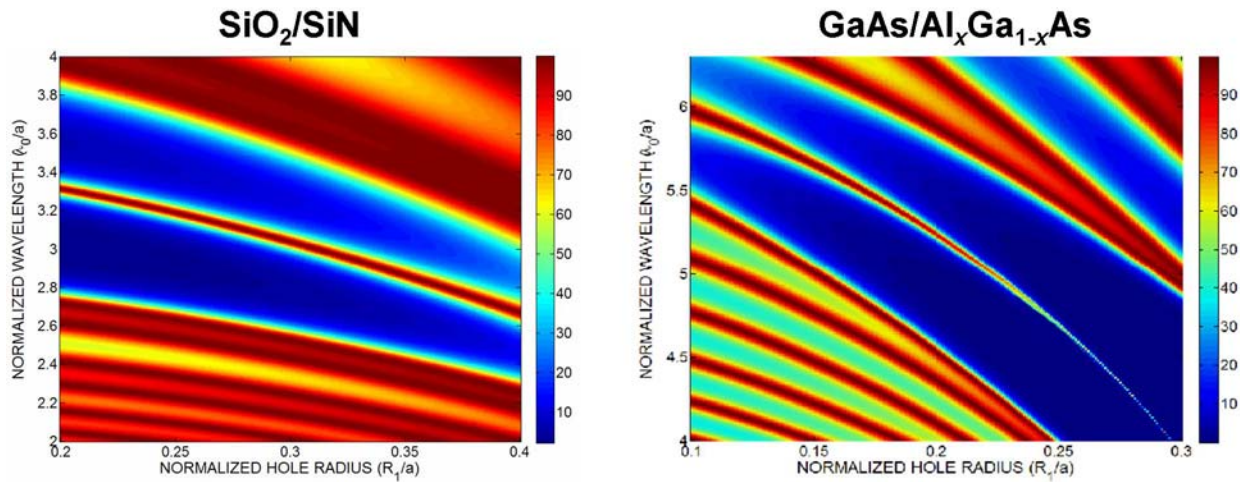
Figure 10-13: Tuning transmission spike using hole radius

10.3.4.   Hole Taper

Due to physics inherent in the etching processing, hole radius can vary with depth forming essentially a chirped photonic crystal. This has the effect of "detuning" the Bragg mirrors above and below the defect reducing efficiency of the device and affecting the band gap. The effect of a tapered hole radius is summarized in Figure 10-14 where transmittance through a 20 layer slab with a defect was calculated for both material systems.
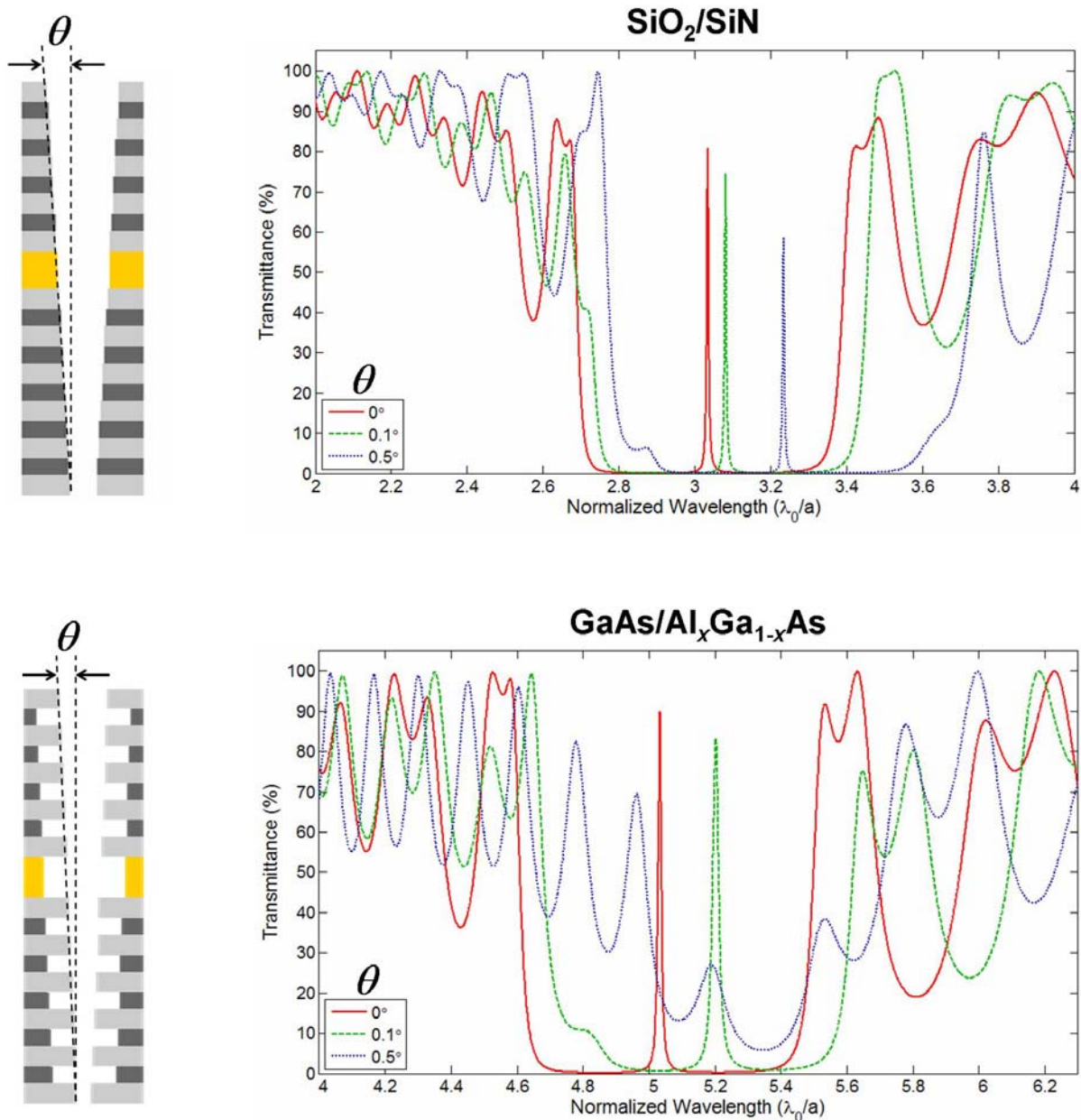
Figure 10-14: Impact of hole taper on transmittance

As taper angle θ increases, the band gap in both material systems shifts to longer wavelengths because the photonic crystal contains more dielectric. As may be expected, the GaAs/Al$_x$Ga$_{1-x}$As system is more sensitive due to higher refractive index and stronger resonance.

In fact, when taper angle approaches just 0.5° in this system, the spectral response becomes distorted well beyond being a useful device. In both systems, position of the transmission notch shifts to longer wavelengths and peak transmission is reduced. Shift in position can be attributed to higher fill factor caused by reducing hole radius. Decrease in peak efficiency is due to reflection bands of the top and bottom Bragg mirrors being misaligned.
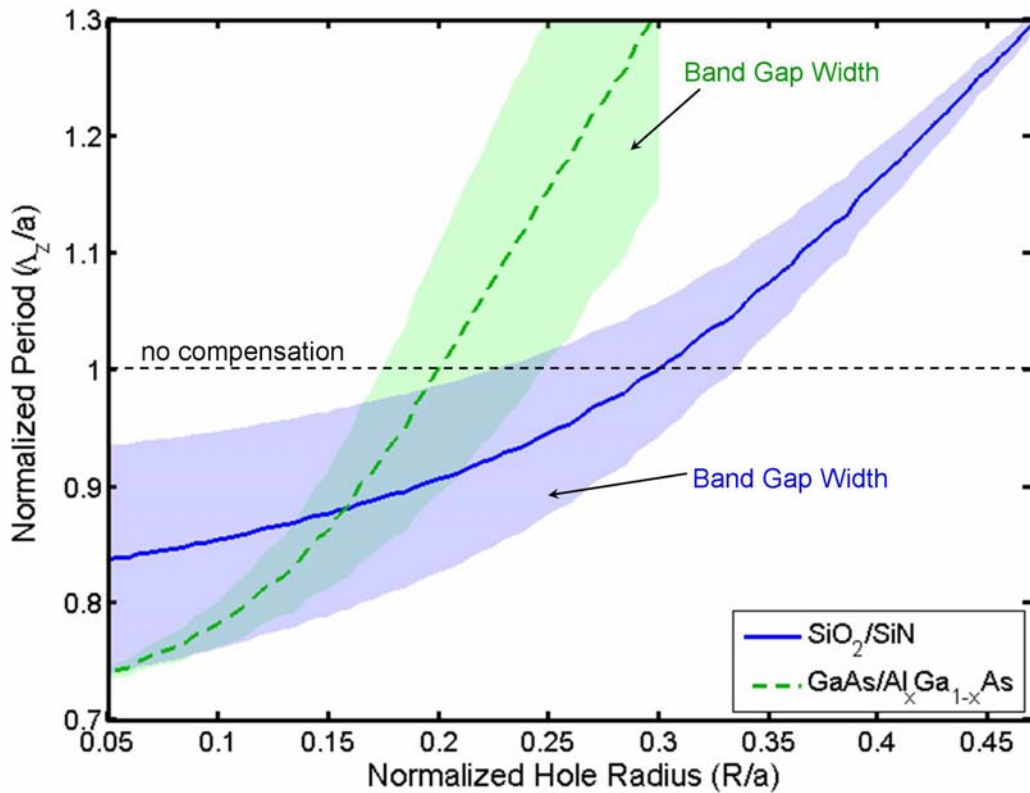
Figure 10-15: Taper compensation by longitudinal period

If hole taper cannot be prevented, there are two potential solutions. First, initial hole size can be increased to reposition the band gap and transmission spike. Second, film thickness can be modified in each layer such that position of the band gap is maintained throughout the crystal. To investigate the second approach, longitudinal period was adjusted for varying hole radius to

maintain position of the band gap in both material systems. These results are shown in Figure 10-15. Shaded regions surrounding the lines show how band gap width is affected by the compensation. Wider colored bands indicate wider band gap. The horizontal dashed line represents the case where hole radius is ideal and no compensation is used. Width of the band gap in the SiO_2/SiN material system decreases with decreasing hole radius, while for $GaAs/Al_xGa_{1-x}As$ it increases with decreasing hole radius.
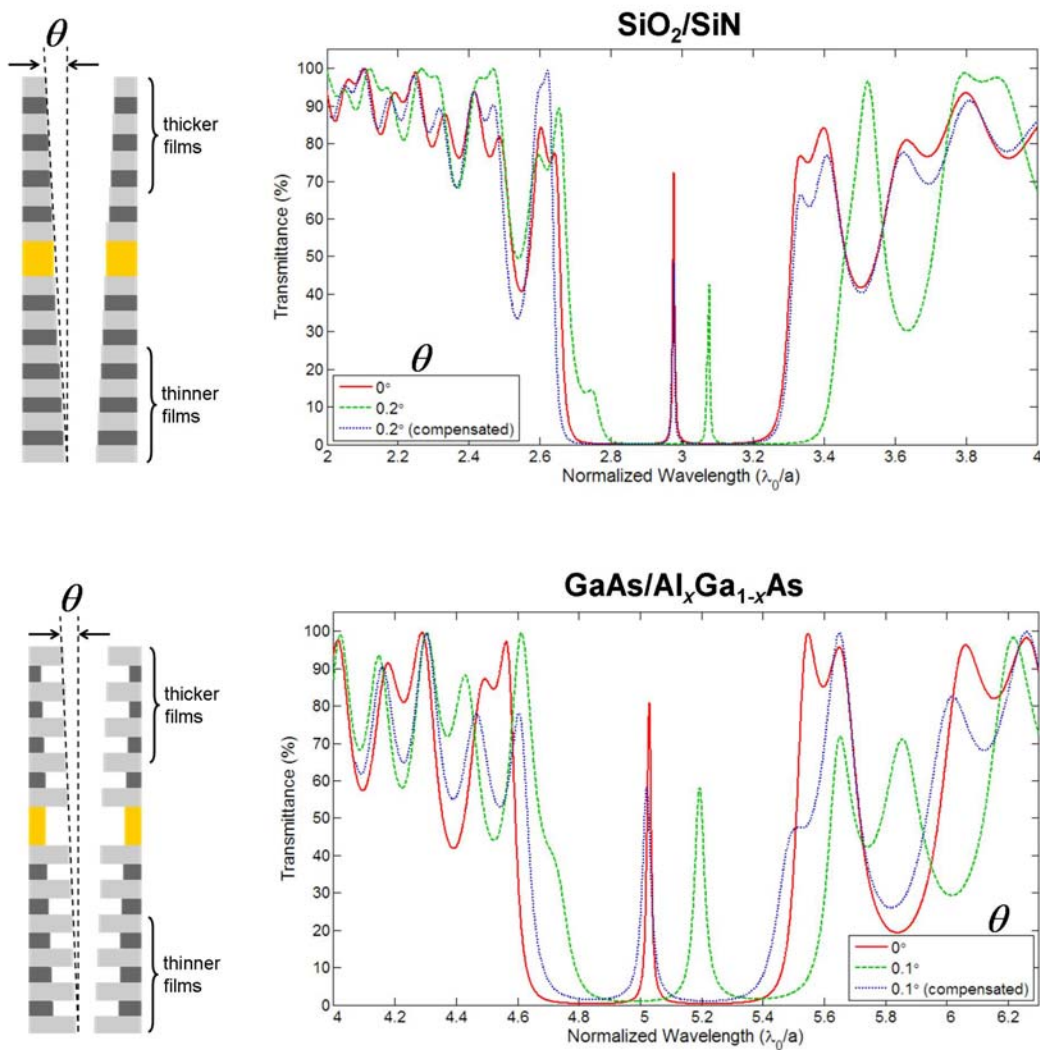
Figure 10-16: Transmittance through a compensated tapered device

To illustrate the compensation technique, data calculated for Figure 10-15 was imported into RCWA to calculate the transmission spectrum through compensated devices. Results are depicted in Figure 10-16. Three responses are shown in each diagram. First, the ideal spectrum with no hole taper is depicted by the solid red line. Second, transmission spectrum where a taper is introduced is represented by the dashed green line. Third, transmission spectrum where taper has been compensated by adjusting film thickness is depicted by the dashed blue line. For both material systems, position of the band gap and position of the transmission notch fall in very close proximity to the ideal case. This is only possible by adjusting film thicknesses and defect thickness appropriately. Compensation is more accurate in the SiO_2/SiN device due to lower sensitivity to structural deformations. Differences in layer thicknesses are conveyed in this figure, but are barely perceivable because the diagrams were drawn to scale.

10.3.5.   *Device Performance at Oblique Incidence*

It is important to understand how devices perform at oblique angles of incidence. Inspired by photonic band diagrams, Figure 10-17 was constructed to show transmittance as a function of wavelength and angle of incidence for both material systems in a real device. The points marked **X** along the horizontal axes correspond to light incident near parallel to the *x*-axis. The points marked **Z** correspond to light at normal incidence. The points marked **D** indicate light that is incident at near 90° aligned along the diagonal of the hexagonal unit cell as depicted in Figure 10-18. Intermediate points are marked by their angle of incidence.
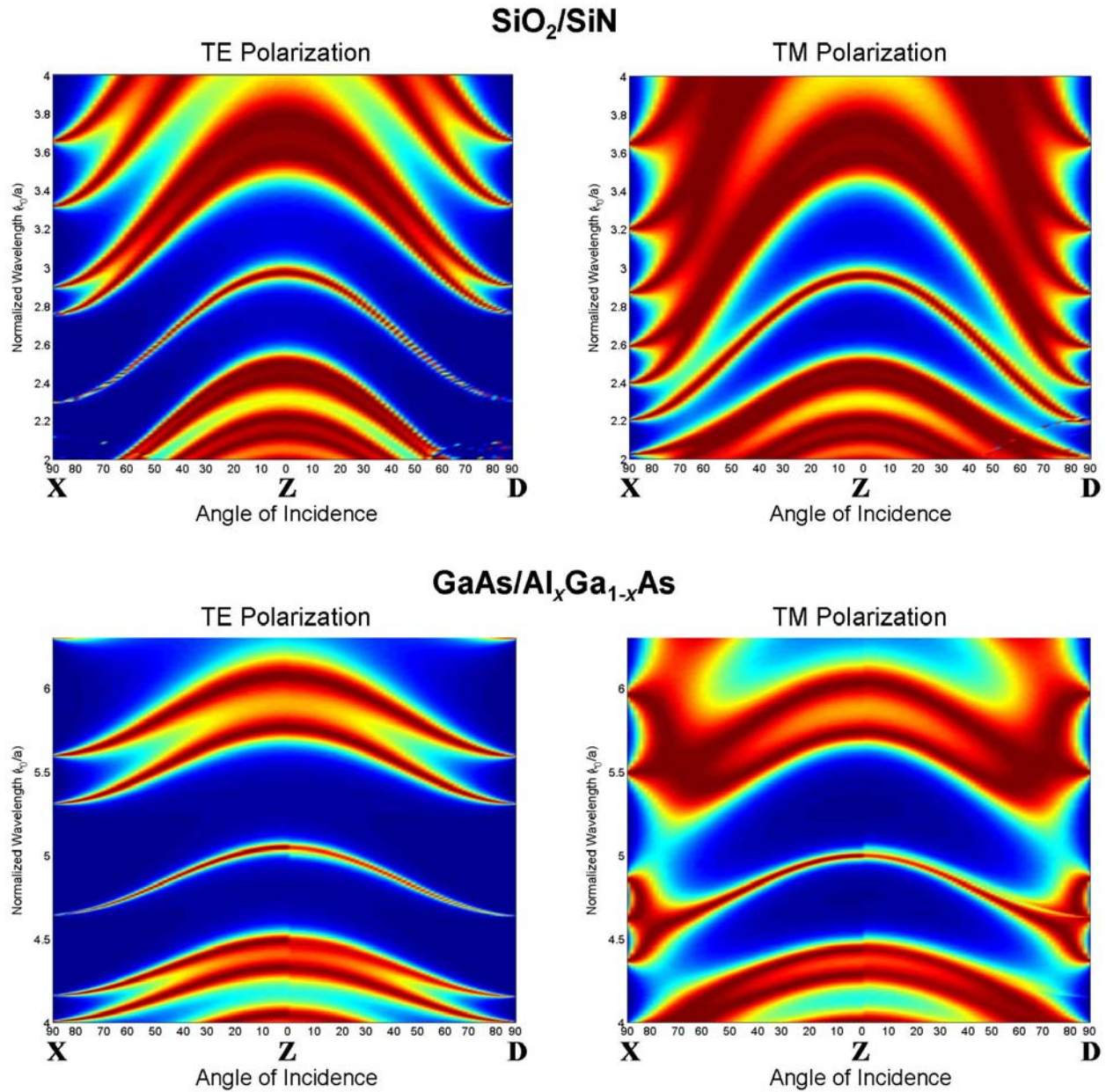
Figure 10-17: Transmittance at oblique angles of incidence

Both devices are more tolerant to oblique angles for TE polarized light than TM, but it is the GaAs/Al$_x$Ga$_{1-x}$As device that is most tolerant. At the same time, it is the GaAs/Al$_x$Ga$_{1-x}$As that most sensitive in TM polarization. It can be concluded from this data, that both devices offer good performance over a cone of angles extending ±5°. This performance parameter is

230

considerably less than predicted in discussion regarding Figure 10-7. The difference is due to the real device being a finite slab instead of an infinitely periodic lattice.
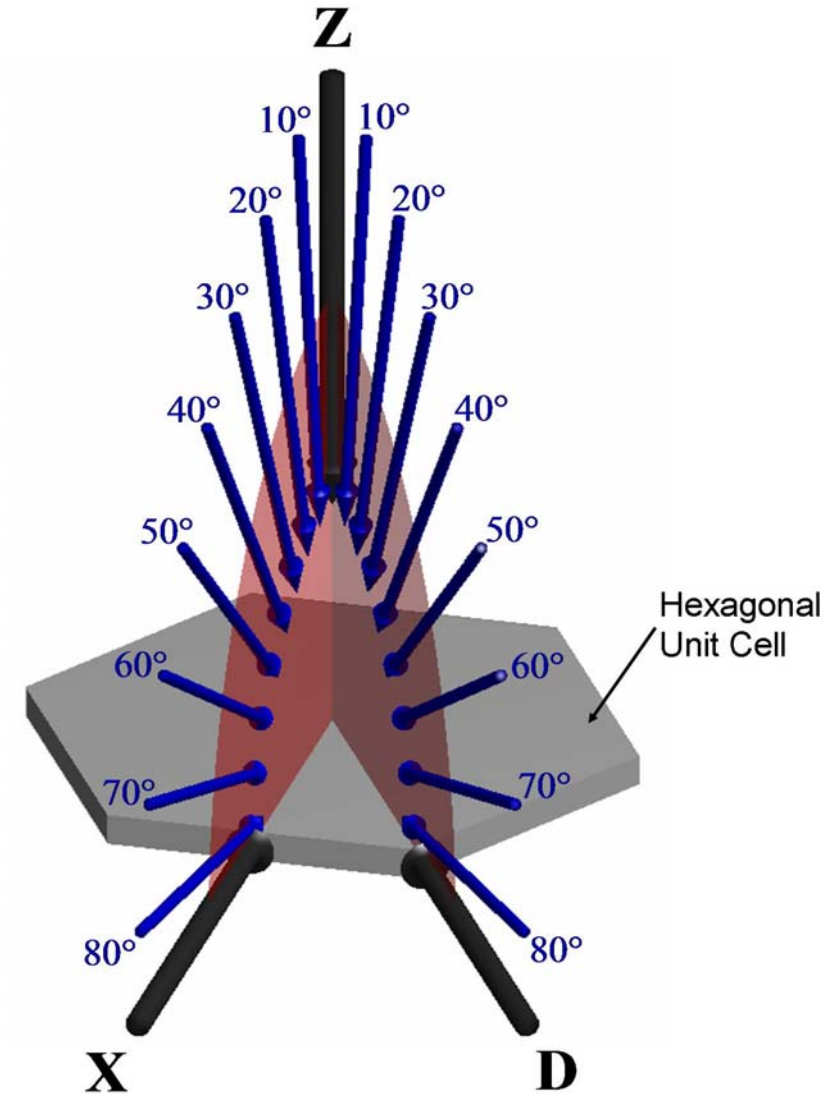
Figure 10-18: Angle of incidence convention

10.4. Conclusion

This chapter designed and optimized space-variant filters based on an easily fabricated photonic crystal structures. Two material systems were considered. These were SiO_2/SiN and $GaAs/Al_xGa_{1-x}As$. The crystals were comprised of an array of holes etched into hetero-structured substrates. A defect layer was incorporated to produce a transmission notch in the center of the transmission stop band. Position of the notch was tuned by adjusting hole radius across the device aperture to form a space-variant filter. It was found that a hexagonal array enabled dimensions to be larger than a square array so it was the preferred geometry. Otherwise, it offered similar performance to a square array.

The material system using SiO_2/SiN was a simpler device to fabricate because it did not require undercutting to produce a sufficient band gap. This system was most limited in tuning range, but spectral response was more robust to the tuning mechanism. The $GaAs/Al_xGa_{1-x}As$ material system required the GaAs layers to be undercut to produce a sufficient band gap. It was found that a minimum amount of undercut was necessary for width of the band gap to be improved. While more sensitive to structural dimensions, this system provided the greatest tuning range. Both material systems provided fair tolerance to oblique angles of incidence within a cone of angles extending $\pm5°$.

CHAPTER 11

DESIGN AND OPTIMIZATION OF GUIDED-MODE RESONANCE FILTERS WITH ACCURATELY POSITIONED RESONANCE

11.1. Overview

In 1902, R. W. Wood discovered "anomalies," or abrupt changes, in diffraction efficiency from gratings when a parameter was varied over a narrow range [188]. A first type of anomaly was explained by Lord Rayleigh in 1907 as a sudden redistribution of energy when additional diffracted orders appear or disappear [189]. In1965, Hessel presented a new theory describing a second anomaly as a resonance type due to a guided wave phenomenon [190]. Mashev used the resonance anomaly to demonstrate a narrow-band reflection filter [191]. Since that time, guided-mode resonance (GMR) devices have found applications in astronomy, WDM filters [192], dichroic reflectors in lasers [193, 194], polarizers [194], sensors [195], security marks and optical tags [196], pulse shaping and dispersion compensation [197, 198], frequency selective surfaces [199], and more.

GMR filters are simple structures comprised of just a grating and a waveguide [200, 201]. When precise phase-matching conditions are met, externally propagating waves can be coupled into guided modes by the grating. The guided modes are "leaky" modes due to index modulation in the waveguide. This causes the guided modes to slowly leak from the waveguide and recombine with the incident wave to establish a resonance. This highly sensitive phenomenon enables filters to be constructed with extremely narrow passbands or stopbands. In fact, linewidth can usually be made arbitrarily narrow just by reducing contrast of the grating.

233

GMR filters are highly compact, easy to fabricate, and typically perform better than conventional thin film stacks for sub-nanometer filters. Filter response can be symmetric with virtually no ripple outside of the passband for both transmission and reflection type filters. Efficiency can approach 100% and they may be tuned over a very large range of wavelengths. Filters with multiple resonances have also been realized. Designs are most often obtained from direct parametric searches or by genetic optimization [202]. While rigorous methods must be used to analyze GMR filters, experimental results are typically in excellent agreement with theoretical predictions.

It appears difficult to achieve strong suppression outside of the passband without resorting to devices with multiple layers. This seems particularly true for transmission type filters, although promising results have been obtained using genetic optimization. GMR filters must be very wide relative to the wavelength to operate efficiently [203]. This ensures enough distance for leaky modes to leak out of the waveguide and contribute to the resonance phenomenon.

GMR filters tend to be very sensitive to polarization, angle of incidence, refractive index, and structural deviations due to the high sensitivity of the phase-matching condition. Crossed gratings and even some ruled grating configurations have been shown to operate independent of polarization [204-208]. Doubly periodic structures have been used to improve tolerance to oblique angles of incidence [209].

This paper outlines an approach for mitigating the high sensitivity to structural deviations and imprecise control of refractive index so that a resonance can be placed accurately at the time of fabrication. The method was inspired by how thin film resistors are trimmed to achieve an accurate value of resistance.

11.2. Theory of Guided-Mode Resonance

There are two interacting mechanisms that must be understood to fully explain the operation of a GMR. The first mechanism is diffraction of an incident wave by a grating and is illustrated in Figure 11-1.
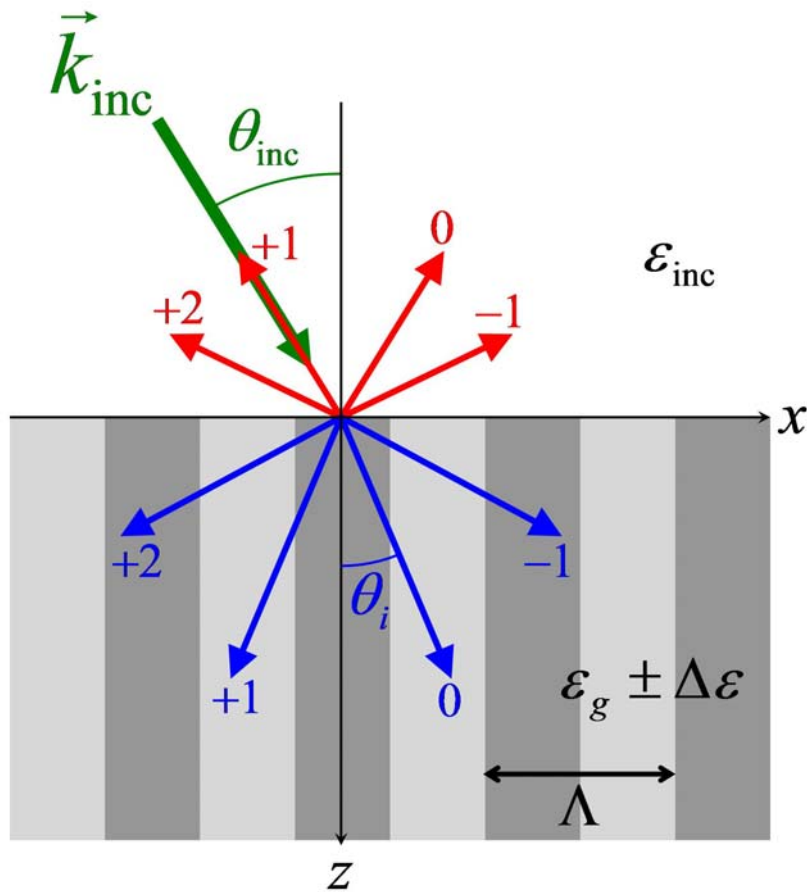

Figure 11-1: Diffraction from a grating

Inside the grating region, amplitude of the propagating field takes on the same variations as the grating. When it is periodic, the electric field can be expressed as a Fourier series where each term corresponds to a different diffracted order. The amplitude and phase of each diffracted

mode must be calculated using Maxwell's equations, but the direction can be determined by the famous grating equation [9].

$$\sqrt{\varepsilon_g} \sin \theta_i = \sqrt{\varepsilon_1} \sin \theta_{inc} - i \frac{\lambda_0}{\Lambda} \qquad (11.1)$$

The second mechanism is wave guiding where light is made to propagate along a confined path by total internal reflection (TIR) [210]. Wave guiding can only occur when the effective index of the guided mode is greater than the surrounding media and less than the refractive index of the core itself. From this, a quantitative condition for wave guiding can be written as

$$\max \left[\sqrt{\varepsilon_1}, \sqrt{\varepsilon_2} \right] \leq \left| \frac{\beta_i}{k_0} \right| < \sqrt{\varepsilon_g} \qquad (11.2)$$

where β_i is the propagation constant and $\left| \beta_i / k_0 \right|$ is the effective index of the ith-order guided mode. From a simple ray picture of a slab waveguide [211], a guided mode can be thought of as a light ray propagating at an angle θ_i such that

$$\frac{\beta_i}{k_0} = \sqrt{\varepsilon_g} \sin \theta_i \qquad (11.3)$$

A guided-mode resonance occurs when a diffracted order exists at the same angle as a guided mode. Guided modes can be related to the incident wave by substituting Eq. (11.3) into the grating equation. Combining this with Eq. (11.2) leads to an inequality that can be used to estimate regions where guided-mode resonances may occur. This is

$$\max \left[\sqrt{\varepsilon_1}, \sqrt{\varepsilon_2} \right] \leq \left| \sqrt{\varepsilon_1} \sin \theta_{inc} - i \frac{\lambda_0}{\Lambda} \right| < \sqrt{\varepsilon_g} \qquad (11.4)$$

236

Figure 11-2 identifies these regions graphically as a function of angle of incidence and grating period. It was generated using Eq. (11.4) for a typical GMR filter with ε_1=1.0, ε_2=2.25, and ε_g=4.0. The parameter ε_g is the average dielectric constant in the grating region estimated by

$$\varepsilon_g \cong \left(\varepsilon_H + \varepsilon_L\right)/2 \tag{11.5}$$
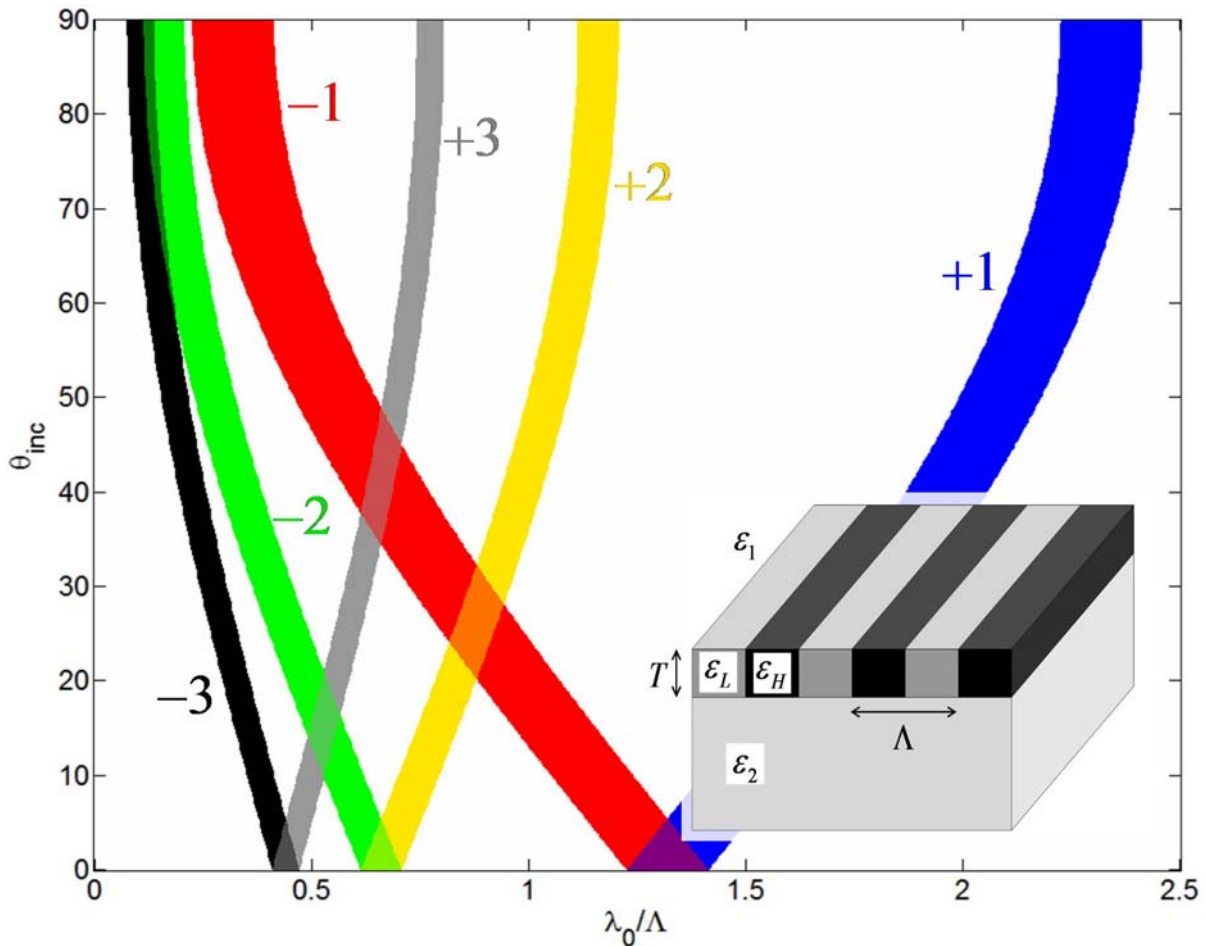
Figure 11-2: Regions of resonance for a guided mode resonance filter

Several important aspects of GMR filters can be observed in Figure 11-2. As angle of incidence moves away from the normal, resonances split. At normal incidence, symmetry

dictates the two resonances occur at the same wavelength, but at oblique angles different phase matching conditions are required to couple external waves into guided modes. As wavelength is decreased, or period increased, the number of resonances increases exponentially. Most designs avoid multiple resonances, but sometimes they can be exploited. For example, Ref. [212] used a doubly resonant structure to form a device with a sharp peak in transmission instead of reflection.

11.3. Typical GMR Devices

Guided-mode resonance filters can be constructed in many ways, but each must simultaneously possess diffractive and guiding properties. Eight representative designs are shown in Figure 11-3. The device in part (a) is a very typical reflection filter that uses a single modulated layer as the grating and the waveguide [53]. It is a ruled grating so TE and TM modes show a different resonance due to birefringence of the grating. In this case, lineshape could be made symmetric if refractive index on both sides of the grating layer were made the same.

A transmission filter is shown in part (b) [213]. While potentially difficult to fabricate, this device demonstrates that very compact filters can be realized from just a single dielectric layer with air serving as the cladding. This device could also function as a polarization filter because the two polarizations behave so differently.

A common method of incorporating multiple resonances, constructing a transmission filter, realizing polarization independence, or improving angular tolerance is to use doubly periodic gratings as shown in part (c) [214]. This device started as the structure in part (a), but a second grating layer was added. This diminished the resonances so a higher index contrast was used in this example.
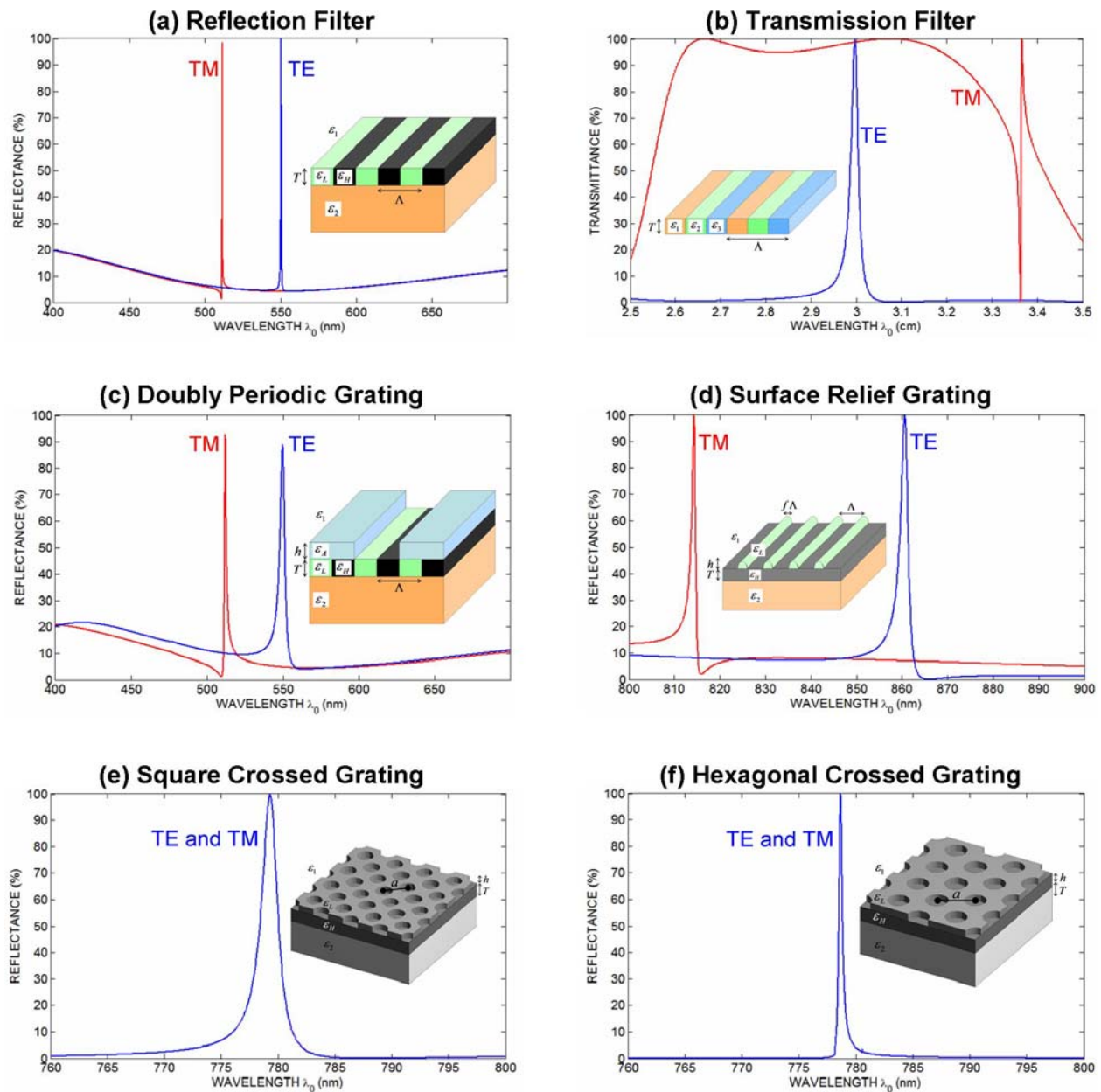
238

Figure 11-3: Typical GMR filters

In practice, many GMR filters take on the topology shown in part (d) where a surface

relief grating is fabricated over a dielectric stack containing a guiding layer [215]. They are easy

239

to fabricate and can perform well. For ruled gratings, birefringence causes the TE and TM resonance peaks to be distinct.

To form a structure that operates independent of polarization, crossed gratings may be used that have 90° rotational symmetry as shown in parts (e) and (f). These are equivalent structures aside from one having square symmetry and the other hexagonal symmetry with a larger period. Dimensions of a hexagonal device can often be larger leading to structures that are potentially easier to fabricate. In this example, resonance of the hexagonal array is narrower than that produced by the square array.

11.4. Device Design

To adjust position of a resonance at the time of fabrication, an additive or subtractive process performed on the surface of the device seems easiest to implement. Refractive index and structural dimensions are much less feasible to adjust after a device has been built. For this reason, it becomes necessary to devise a resonant structure that is sensitive to this approach. A most sensitive device should have the core exposed since position of the resonance is strongly affected by core thickness. In this manner, GMR filters can be adjusted over the greatest possible tuning range at the time of fabrication.

A GMR filter that is easy to build and satisfies requirements discussed above is depicted in Figure 11-4. A surface relief grating is formed directly onto the surface of the guiding layer. By etching or depositing small amounts of material, position of the resonance can be tuned by adjusting thickness of the core region through parameter T. It will be shown that grating depth and grating profile change during the tuning process depending on whether an additive or subtractive process is employed to tune the device.
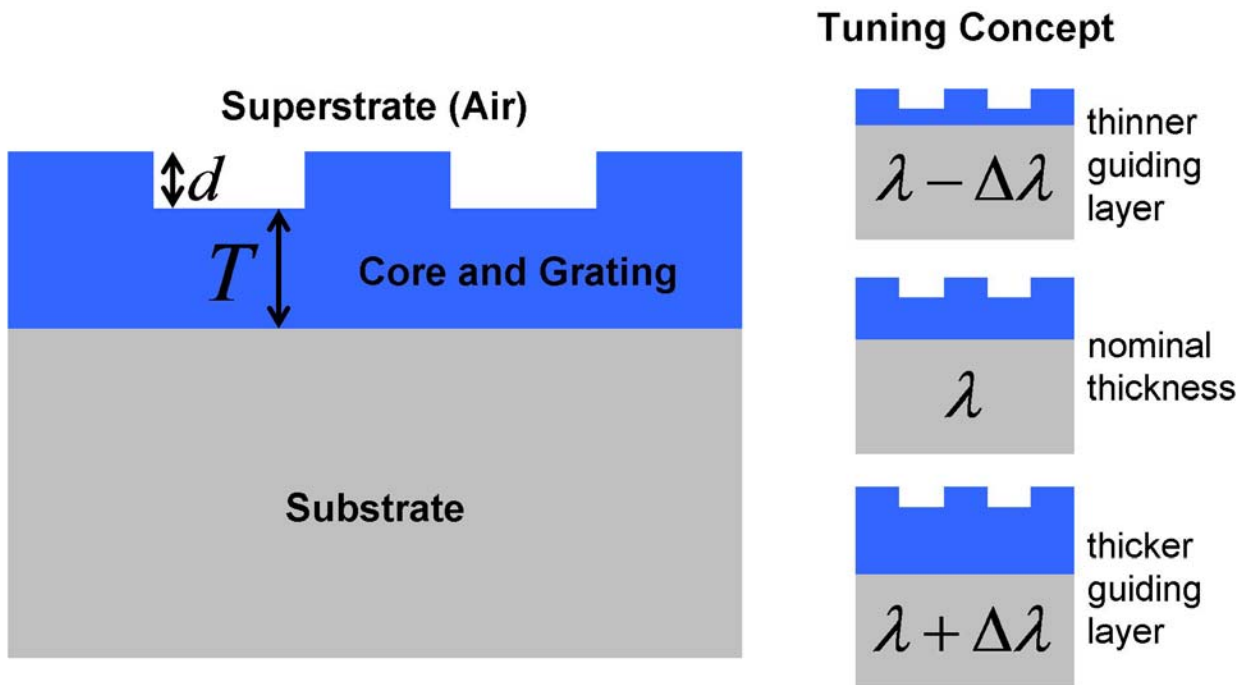
Figure 11-4: Tunable GMR and tuning concept

11.4.1. *Ideal Device*

To design a device that can operate independent of polarization, a crossed grating GMR consistent with that in Figure 11-4 can be used. This device, along with its reflection spectrum, is depicted in Figure 11-5. The device has hexagonal symmetry so grating features can be largest. It is an array of circular holes formed into a high index guide layer. The device has parameters ε_1=1.0, ε_2=2.31, ε_g=3.91, a=1.15 μm, f=0.7, d=230 nm, and T=345 nm. Two resonances can be observed that are spaced around 100 nm apart. This is one factor that limits the overall tuning range to a 100 nm operational window highlighted in the figure.
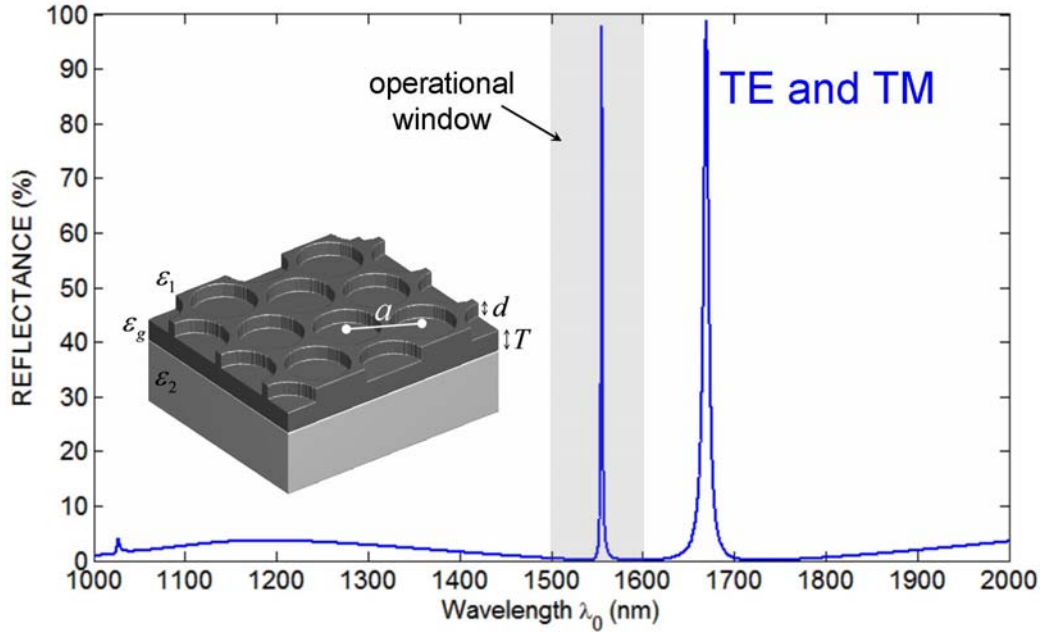
Figure 11-5: Reflectance of "perfect" hexagonal GMR filter

Hexagonal symmetry was chosen for its narrower resonance and larger grating dimensions. Tuning of this device is accomplished by adjusting thickness of core region by deposition or etching processes. The effect on reflection spectrum within the operational window as parameter T is adjusted is depicted in Figure 11-6.

At the left side of Figure 11-6, position and width of the resonance shows a nonlinear response to variations in film thickness. From this observation, designs intended only to adjust position of resonance should use a thicker core layer. If the core layer is too thick, the guide can become multimode and introduce additional resonance peaks into the reflection spectrum. It can also be observed from Figure 11-6 that as resonance is adjusted, suppression at the edges of the operational window is reduced. Nominal thickness of the GMR filter was chosen to minimize this effect over the greatest possible tuning range.
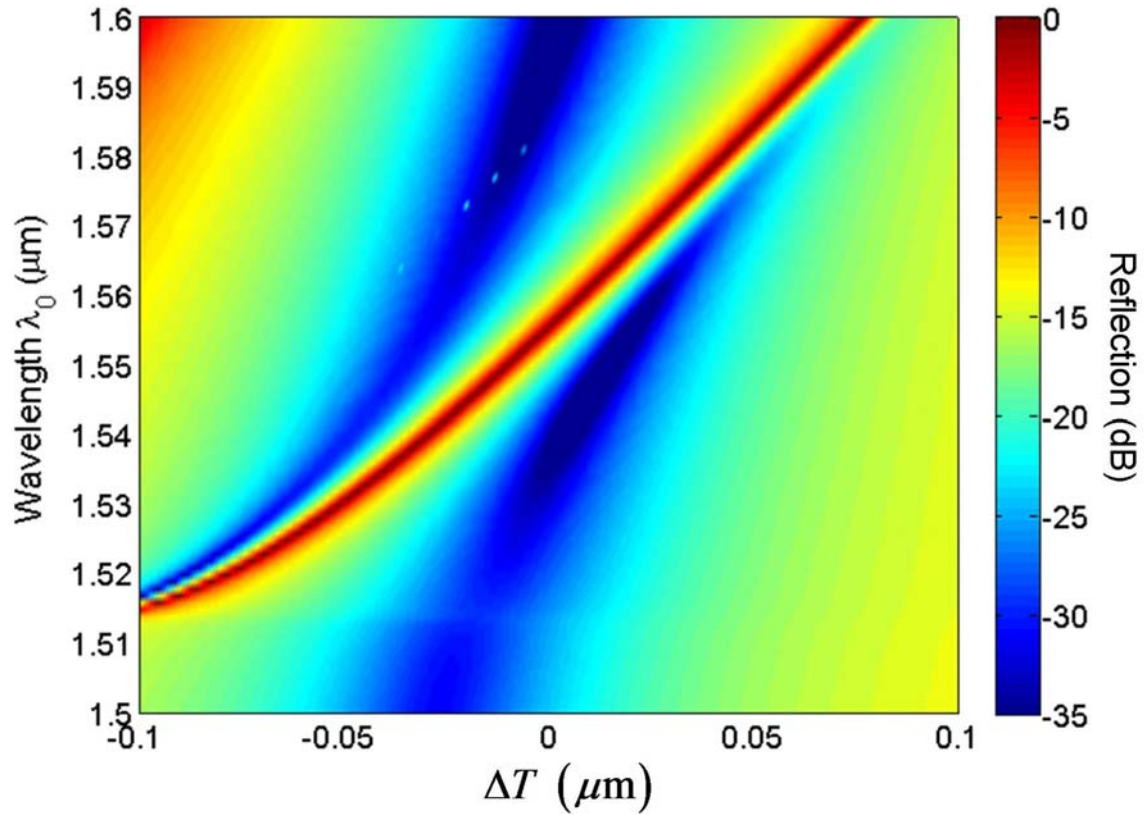
242

Figure 11-6: Tuning of "perfect" GMR filter

The duty cycle and relief depth were chosen to simultaneously optimize suppression outside of the passband and peak reflection on resonance. Duty cycle is defined as the fractional area occupied by the air holes. Hole radius r as a function of duty cycle f and lattice constant a for square and hexagonal symmetry were computed using the following equations.

$$r_{sqr}^2 = a^2 \frac{f}{\pi} \qquad r_{hex}^2 = a^2 \frac{\sqrt{3}f}{2\pi} \qquad (11.6)$$

11.4.2. Tuning by Deposition

Material may be deposited onto an initial grating formed on the surface of a substrate. By measuring resonance position at various stages during deposition, peak resonance can be placed accurately in frequency. This process is summarized in Figure 11-7.
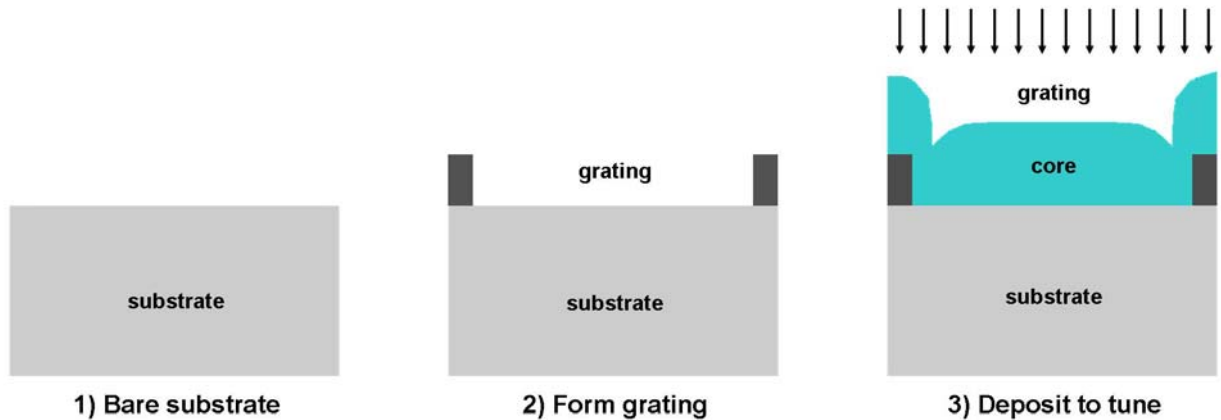

Figure 11-7: GMR filter tune-by-deposition fabrication process

Due to physics inherent in the deposition process, the rectangular profile of the initial grating is not maintained. Visibility from the target to the substrate prevents deposition from being uniform causing trenches to form at the base of vertical sidewalls [39]. One artifact of deposition is decreased duty cycle. Off-axis diffusion leads to deposition on vertical sidewalls. This mechanism works to reduce hole radius as material is deposited. The consequence is much poorer suppression outside of the passband and a shift in position of resonances. To mitigate this effect, duty cycle of the initial grating should be exaggerated such that after deposition, duty cycle is correct. This mechanism places an upper limit on what duty cycle may be realized based on directionality of deposition and how thin of a grating can be made.

A second artifact of deposition is increased relief depth. Portions of the grating at the bottom of the grooves are less visible to the target and receive less deposition. Portions at the top of the grooves are fully visible and receive maximum deposition. This leads to a higher growth rate at the top of the grooves than the bottom, which increases relief depth during deposition. This effect leads to poorer suppression off resonance and a shift in position of the resonance peaks. To mitigate this effect, relief depth of the initial grating should be reduced such that the final relief depth is correct. For this design, the initial grating had a=1.15 μm, f=0.95, d_1=207 nm, ε_1=1.0, ε_2=2.31, and ε_g=3.91.
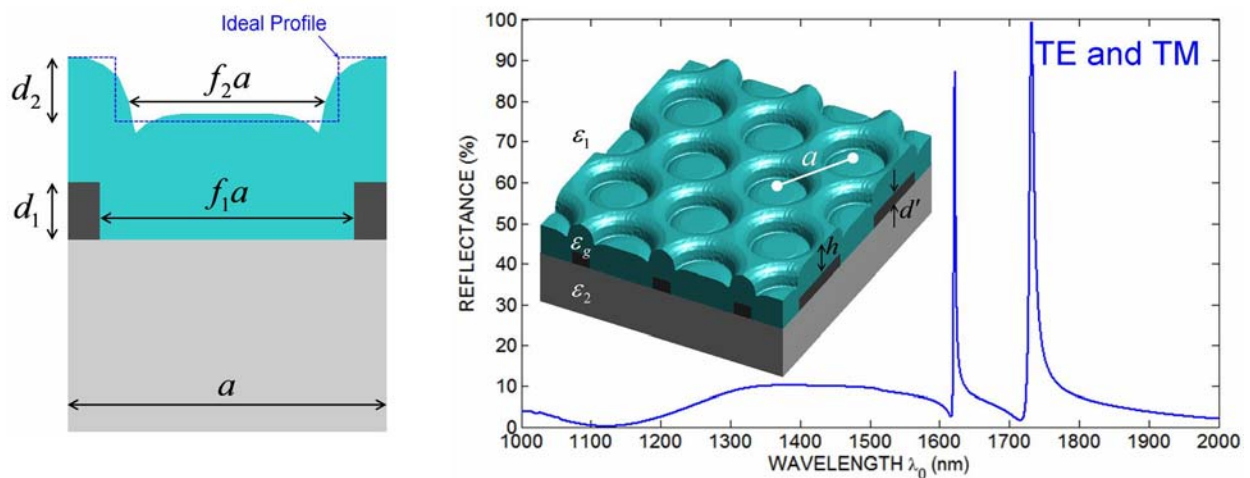

Figure 11-8: Replication of "perfect" GMR by deposition

A third artifact of deposition is a GMR that is somewhat doubly-periodic. Portions of the surface at the bottom of the grating grooves nearest to the sidewalls receive less deposition due to a shadowing effect. This causes a "hump" to appear in the middle of the grooves resembling the hump at the top of the grooves. If conditions are right, multiple resonances may appear in the

245

reflection spectrum. It is entirely possible this mechanism could be exploited in some manner, but it was not considered here.

Figure 11-8 shows a cross section and reflection spectrum of hexagonal GMR filter optimized to replicate the ideal grating profile as much as possible. The resonances have shifted significantly and background suppression is poor. Performance can be enhanced if the notion of having to replicate the ideal grating is discarded. Figure 11-9 shows the same grating, but the deposited layer is thinner, relief depth is greater, and duty cycle is greater than in Figure 11-8. In this case, background suppression is greatly improved and resonances are in the desired position.
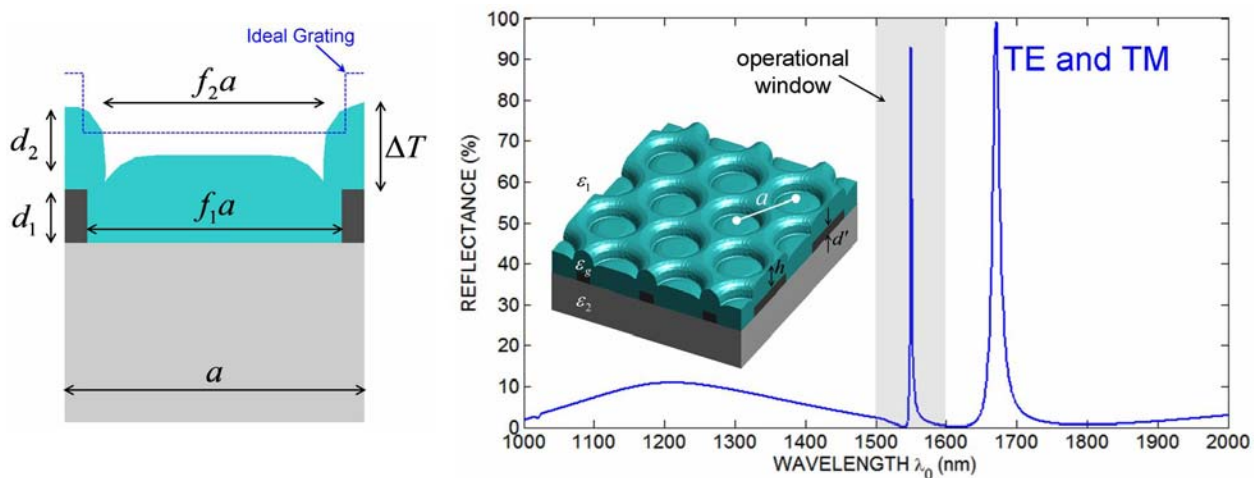

Figure 11-9: Optimized GMR filter using deposition process

Figure 11-10 shows reflectance as material is deposited onto the substrate. If lesser tuning range is tolerable, device performance can be enhanced using a larger lattice constant and adjusting film thickness.

The main drawback of the additive approach is that it is difficult to realize a large grating

duty cycle which is advantageous here. For this reason, it is best to use a highly directional

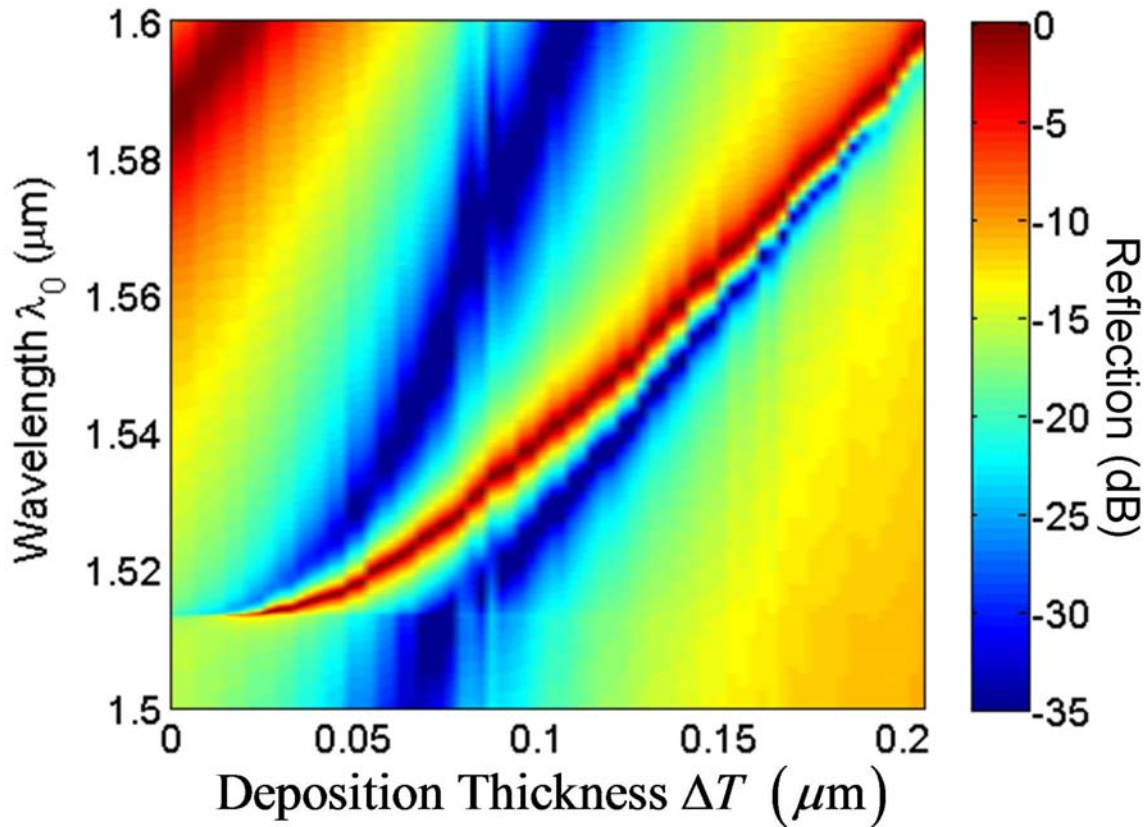deposition or use the subtractive process described next.

Figure 11-10: Tuning curve for deposition process

11.4.3.   Tuning by Etching

As an alternative to deposition, devices may be tuned using an etching process. This

offers several advantages. First, it is an easier process and less expensive to implement. Second,

it does not form a structure that is doubly-periodic. Third, it produces a grating profile with anti-

reflection properties due to smoother transition from air into the dielectric. As a result, the

device offers improved suppression away from resonance. Figure 11-11 summarizes the tune-by-etching fabrication process.
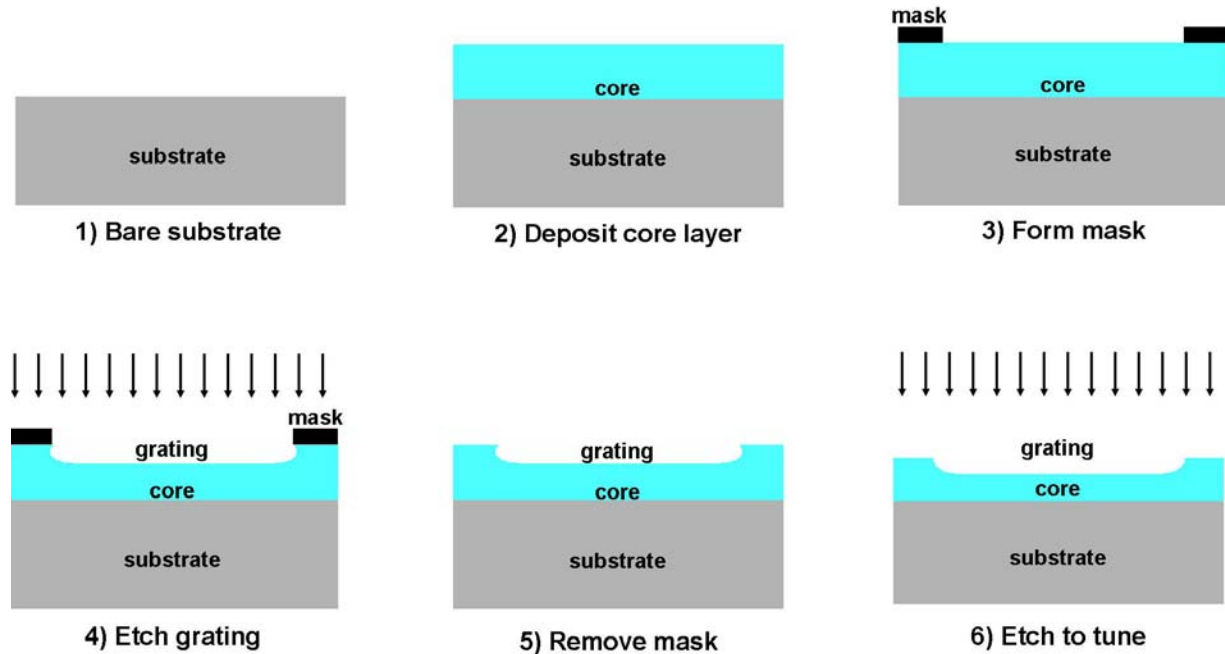

Figure 11-11: GMR filter tune-by-etching fabrication process

In contrast to the tune-by-deposition process, tune-by-etching leads to an increased duty cycle and decreased relief depth. This can be compensated by starting with smaller holes and greater relief depth and allowing the tuning process to make the corrections. This mechanism places a lower limit on what duty cycle can be achieved. In addition, overall relief depth is diminished due to redeposition and shadowing effects. This too can be compensated by exaggerating the initial dimensions and allowing the etching process to reduce relief depth.

The reflection spectrum of an optimized device along with a device cross section is shown in Figure 11-12. The initial grating in this case had a=1.16 mm, f=0.9, d_1=500 nm, and d_2=200 nm.
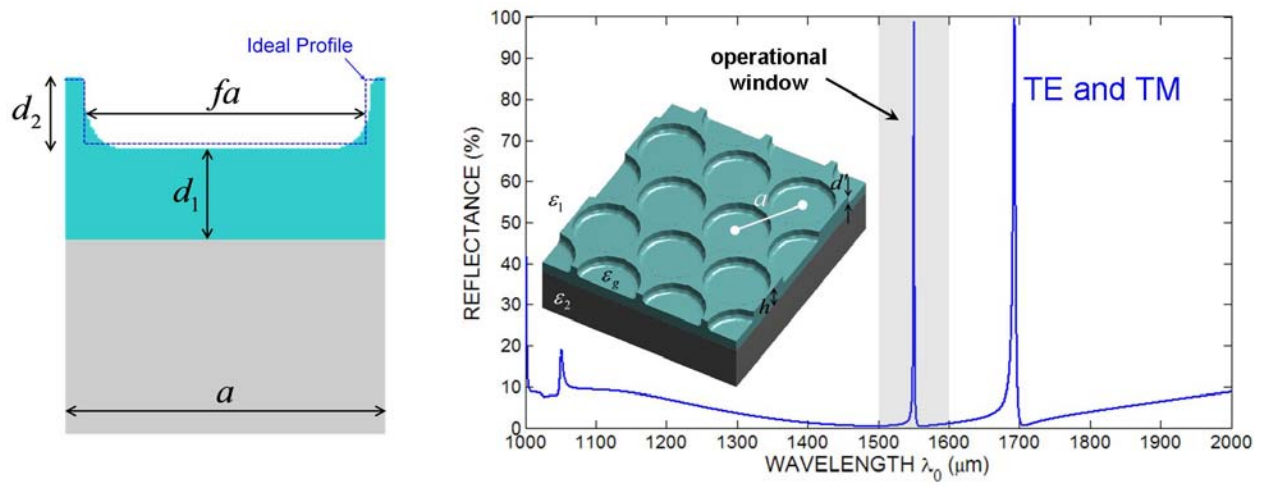
Figure 11-12: Optimized GMR filter using etching process

Figure 11-13 summarizes the spectral response of the GMR filter as the profile is etched.

A tuning range of over 50 nm is observed and shape of the resonance is near constant.

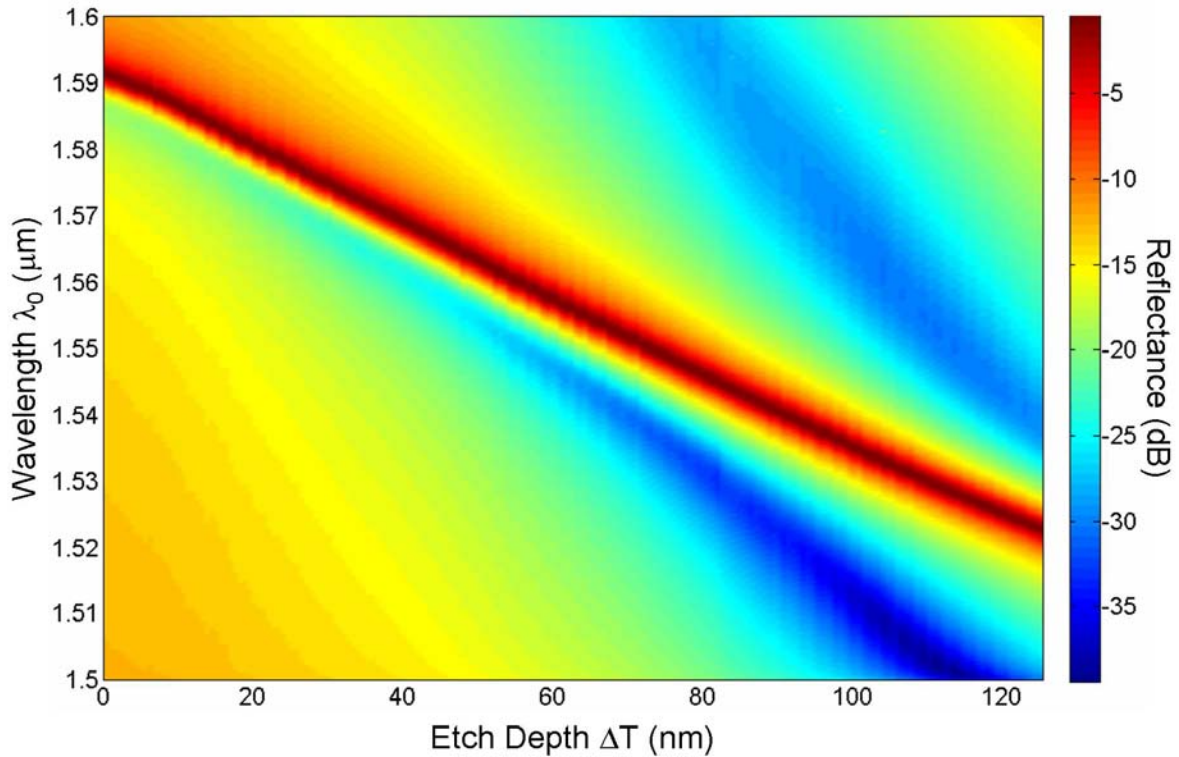Suppression off resonance is near 30 dB.

Figure 11-13: Tuning curve for etching process

11.5. Conclusion

This chapter focused on the design and optimization of guided-mode resonance filters where deposition or etching processes were used to form a device with an accurately placed resonance. Based on analysis in this dissertation, the etching process is preferred because it leads to better device performance and is easier to implement in the lab. To improve suppression outside of the main resonance, multilayer devices will likely bee needed. This would require methods for tuning intermediate layers at intermediate wavelengths.

CHAPTER 12

DESIGN AND OPTIMIZATION OF AUTOCLONED FORM-BIREFRINGENT STRUCTURES

12.1. Overview

Diffractive optical elements (DOE) are proven devices that have found a great number of applications in modern optics [9]. They are typically designed using scalar diffraction theory and can be used to realize complex wavefronts [216]. They are easy to fabricate and operate independent of polarization because feature size is typically very large relative to wavelength.

Polarization is a degree of freedom not often exploited in DOEs, but can be used to construct many types of polarization sensitive devices. A form-birefringent device functions like a DOE, but manipulates polarization by engineering the orientation and degree of birefringence throughout an element comprised of subwavelength features. Xu *et al* were among the first to explore this class of optical elements [217, 218], although their use can be dated back to Heinrich Hertz who used metal-stripe gratings as polarizers to test properties of radio waves in 1888 [219, 220]. Later, others achieved simultaneous control of phase and polarization [221] to address a greater number of applications. Lenses can be formed where sign of the lens is reversed when handedness of polarization is changed [222]. Field distributions like radial polarization or azimuthal polarization can be produced that posses singularities called disclinations. The ability to control polarization across a column of light enables polarization coding of information, optical encryption, tight focusing [223], polarimetric imaging [224, 225], optical tweezers, polarization converters [226-230], image processing, imitation of complex grating profiles [231-233], polarization beam splitters [234-238], array generators [239-242], selective mode

251

excitation in waveguides [26], and more. Form birefringent devices have been used in optoelectronic packaging and for optical interconnects and switching [217]. A variety of sensors have been realized as well [243].

Form-birefringent elements can be constructed by joining two anisotropic substrates like calcite or lithium niobate [217, 244, 245]. This approach requires difficult fabrication and alignment. A more attractive approach is to fabricate artificial materials like photonic crystals or subwavelength gratings that exhibit anisotropic properties. This enables devices to be fabricated from standard materials on a single substrate with no alignment steps. Artificial materials often provide stronger birefringence and are more easily varied across the device aperture [221, 246].

In general, all form-birefringent devices simultaneously optimize orientation, period, and degree of birefringence across the device aperture to control the complex amplitude of two orthogonal polarizations. The manner in which the subwavelength features are distributed can be continuous [226, 227, 229, 235] or discrete [26, 221, 240, 241, 247]. Discrete devices use a cell-encoded approach where the element is tiled by an array of uniquely designed cells. These are easier to fabricate and provide greater simultaneous control of polarization, phase, and amplitude of transmitted light. They can provide more consistent birefringence from uniform films because grating period and duty cycle can be made constant. Discrete devices, however, suffer from diffractive effects at the boundaries between cells reducing overall efficiency. Continuous elements offer higher efficiency, but are less able to control phase and polarization at the same time. They are sometimes more difficult to fabricate because feature size can be quite small, particularly near disclinations.

This chapter will focus on form-birefringent devices of the continuous type to generate radial and azimuthal polarized fields. For this application, form-birefringent devices offer a

clever and compact alternative to conventional assemblies of polarizers and waveplates that are larger and more complicated. It will be shown this application involves a tradeoff between transmission efficiency and aspect ratio of the subwavelength gratings. Autocloned structures will be proposed to achieve high efficiency with structural features that are much easier to realize. Physics inherent to the autocloning process will produce consistent layer profiles given almost any initial grating. This enables simplified and very robust fabrication of high performance form-birefringent devices. The string method was used to simulate the autocloning process to predict geometry more accurately. Based on realistic layer profiles, form-birefringent devices were designed and optimized for producing radial and azimuthal polarizations.

12.2. Effective Medium Theory of Subwavelength Binary Gratings

The most popular structure by far in form-birefringent devices is the binary grating. It is easy to fabricate and offers strong birefringence when the period is less than a wavelength. In this regime, gratings behave much like homogeneous uniaxial materials, and can exhibit strong birefringence when index contrast is high [246]. Ruled gratings can be viewed as uniaxial crystals where effective refractive indices are used analogous to ordinary and extraordinary refractive indices along the crystal axes. This concept is illustrated in Figure 12-1. Waves with linear polarization perpendicular to the grooves experience the lowest effective index. Waves polarized parallel to the grooves experience the highest. Origin of this phenomenon can be understood through discussion in Chapter 2 concerning Eq. (2.17). Waves linearly polarized at other angles do not retain linear polarization after transmission. Parallel and perpendicular polarization components of the electric field experience different effective refractive indices and become out of phase after transmission. This produces elliptical polarization on the transmitted side, but special cases exist where linear or circular polarization can be realized.

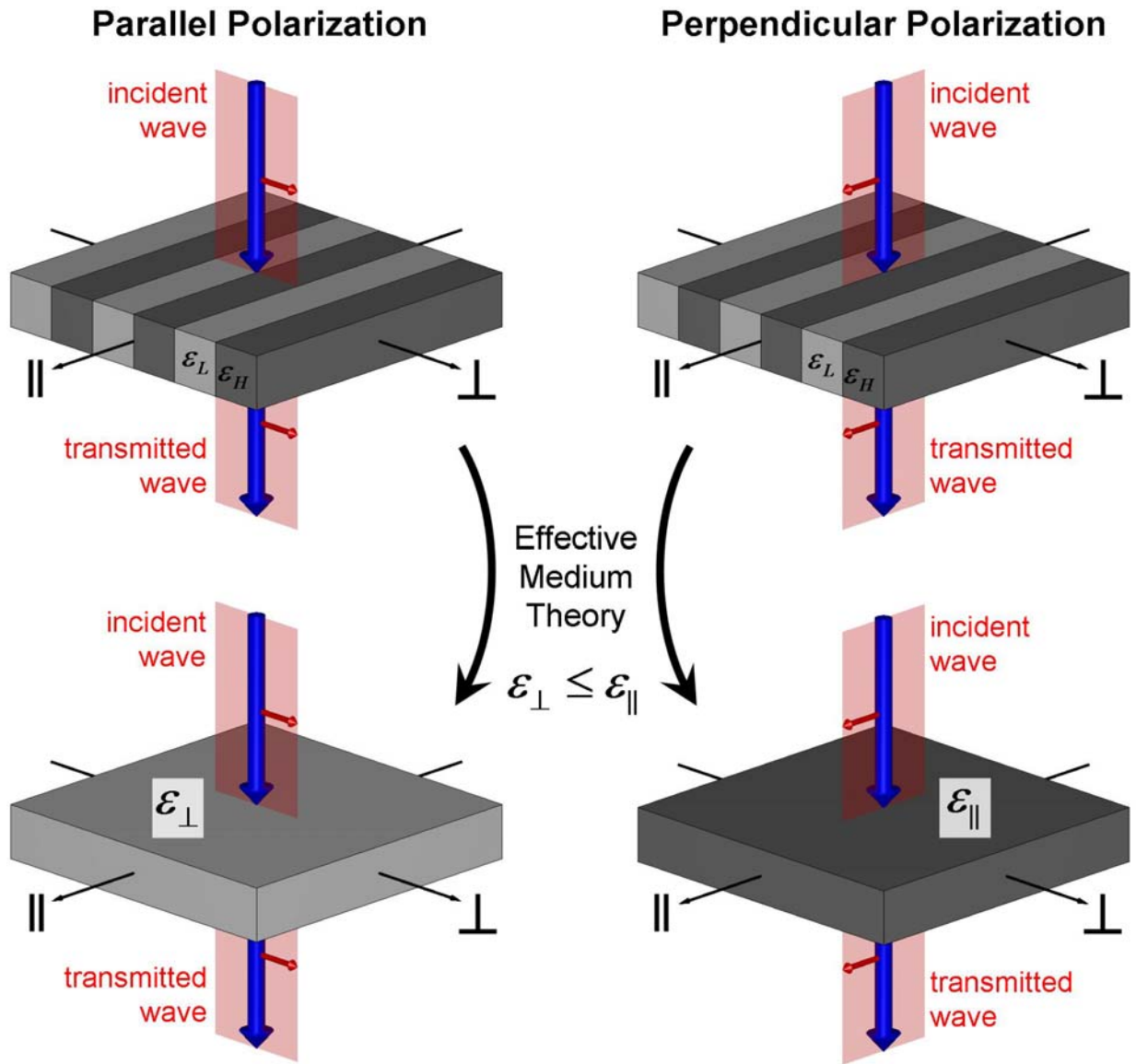**Figure 12-1: Effective medium theory**

Using effective medium theory (EMT) [248-251], a grating may be replaced by a homogeneous uniaxial material without effecting transmission or reflection. Refractive indices of the homogeneous material are the effective indices of the medium. Weiner bounds [252, 253] dictate all effective indices should fall between the effective indices for perpendicular and

parallel polarizations when grating dimensions are much less than a wavelength. In terms of the dielectric function, these are to first order accuracy

$$\varepsilon_\perp \le \varepsilon_{\text{eff}} \le \varepsilon_{\parallel} \qquad \begin{aligned} \varepsilon_\perp &\cong \left[(1-f)\varepsilon_L^{-1} + f\varepsilon_H^{-1}\right]^{-1} \\ \varepsilon_{\parallel} &\cong (1-f)\varepsilon_L + f\varepsilon_H \end{aligned} \qquad (12.1)$$

The parameter $\varepsilon_\perp$ represents the effective dielectric constant for light that is linearly polarized perpendicular to the grating grooves. The parameter $\varepsilon_{\parallel}$ is the effective dielectric constant for light that is linearly polarized parallel to the grating grooves. The grating duty cycle is quantified by parameter f.

A figure of merit often used to characterize phase retardation through a birefringent device is L_π. At this thickness, one polarization component accumulates 180° more phase than the other after transmission through the device. L_π is easily calculated from the effective dielectric constants.

$$L_\pi = \frac{\lambda_0}{2\left(\sqrt{\varepsilon_{\parallel}} - \sqrt{\varepsilon_\perp}\right)} \qquad (12.2)$$

12.2.1. *Binary Gratings as Wave Plates*

Binary gratings can produce strong birefringence where $\Delta n \approx 1.0$ is feasible. This enables them to function as thin wave plates for manipulating polarization [254, 255]. It becomes possible to construct quarter-wave plates for converting between circular polarization (CP) and linear polarization (LP), or half wave plates to reverse handedness of polarization or control phase.

Polarization Converters Using Quarter-Wave Plates
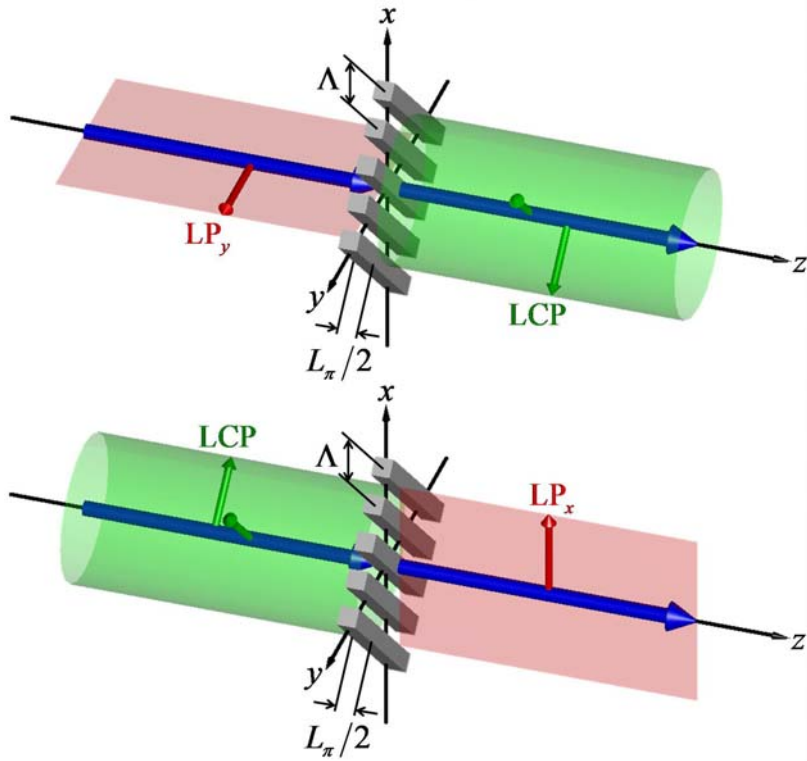
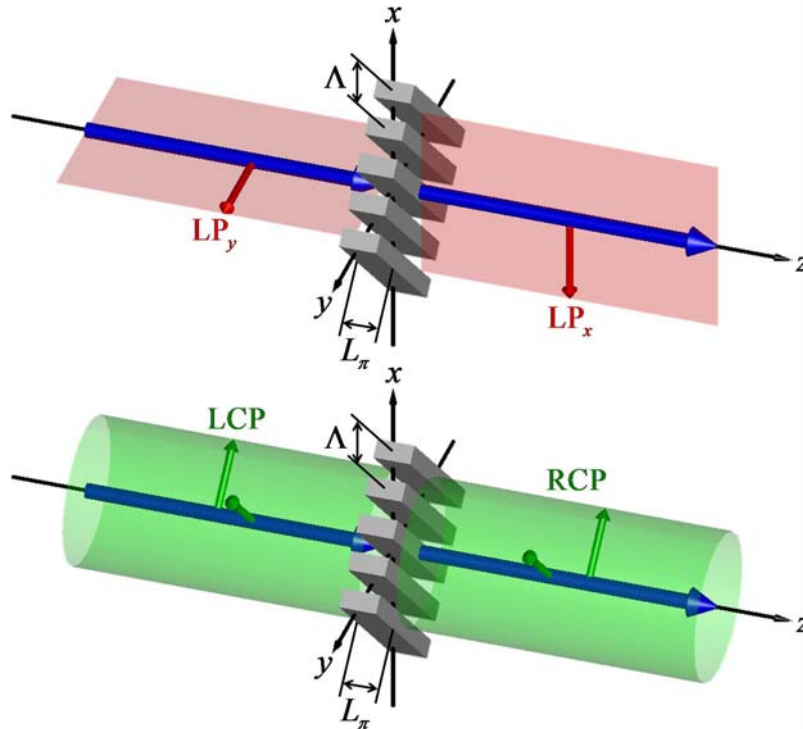## Handedness Converters Using Half-Wave Plates

Figure 12-2: Binary gratings used as half-wave and quarter wave plates

256

Four examples are illustrated in Figure 12-2. The top two diagrams show binary gratings used as quarter-wave plates to convert between linear and circular polarizations. This is accomplished by setting grating thickness such that polarizations along the grating axes become 90° out of phase after transmission; $d=L_\pi/2$. If the grating is turned 90°, output polarization changes handedness. The bottom two diagrams show binary gratings functioning as half-wave plates to reverse handedness of polarization. This is accomplished by setting grating thickness such that polarizations along the grating axes become 180° out of phase after transmission; $d=L_\pi$. If the grating is turned by 90° relative to the incident polarization, the same change in handedness will occur but the transmitted light will be 180° phase shifted. The half wave configuration is particularly useful when illuminated with linear polarization because orientation of the transmitted polarization can be completely controlled through the orientation of the grating with respect to the incident wave.
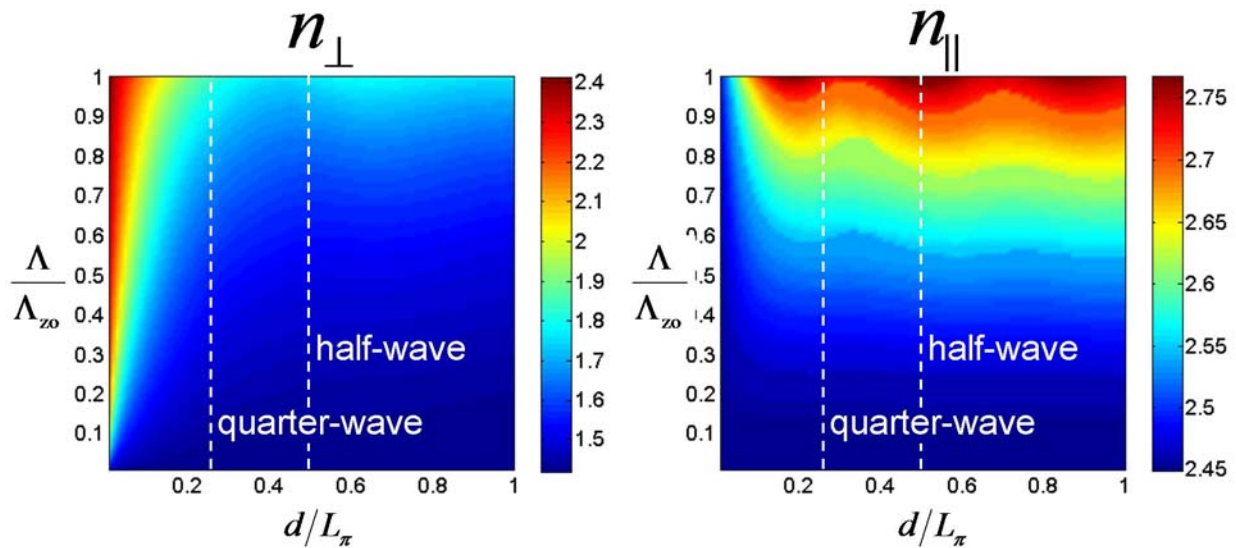

Figure 12-3: Dependence of effective index on grating period and thickness

Equation (12.1) explains how grating duty cycle and refractive index impact effective

indices of the device. It is less intuitive to learn that grating period and grating thickness also

modify the effective refractive indices and therefore modify L_π [251]. This dependence is

quantified in Figure 12-3 where effective indices were calculated over a range of grating period

and thickness. Grating thickness is given relative to L_π. Grating period extends from near zero

to Λ_{zo} defined as the cutoff where a first order diffracted mode appears.

$$\Lambda_{zo} = \lambda_0 / n_g \tag{12.3}$$

Data in Figure 12-3 shows that grating period should be kept very small to realize

consistent effective refractive indices. This enables devices to have uniform thickness across the

aperture and maintain consistent performance. As effective indices drift away from designed

values, device behavior can change because L_π can deviate reducing efficiency. The data also

shows effective refractive indices are more stable for thicker gratings. This suggests half-wave

devices will be more robust to variations in grating period than quarter-wave devices.

12.3. Design of Continuous Form-Birefringent Devices

Using concepts discussed above, it is possible to design devices that produce light of

arbitrary polarization on the transmitted side of a form-birefringent grating. This is

accomplished by varying grating parameters across the device aperture. Of particular interest in

this dissertation are devices that form radial and azimuthal polarization.

The simplest devices to describe are discrete implementations as shown in Figure 12-4.

The device aperture is divided into a series of cells where the grating inside each cell is oriented

to produce the desired polarization after transmission. In this case, linearly polarized light

illuminates the devices.

From the data presented in Figure 12-3, there is some freedom in the choice of grating period that still provides the same effective refractive indices and therefore the same L_π thickness. This freedom can be exploited to produce a form-birefringent device that is smooth and continuous. The following sections outline the design procedure for continuous form-birefringent devices in both quarter-wave and half-wave implementations.
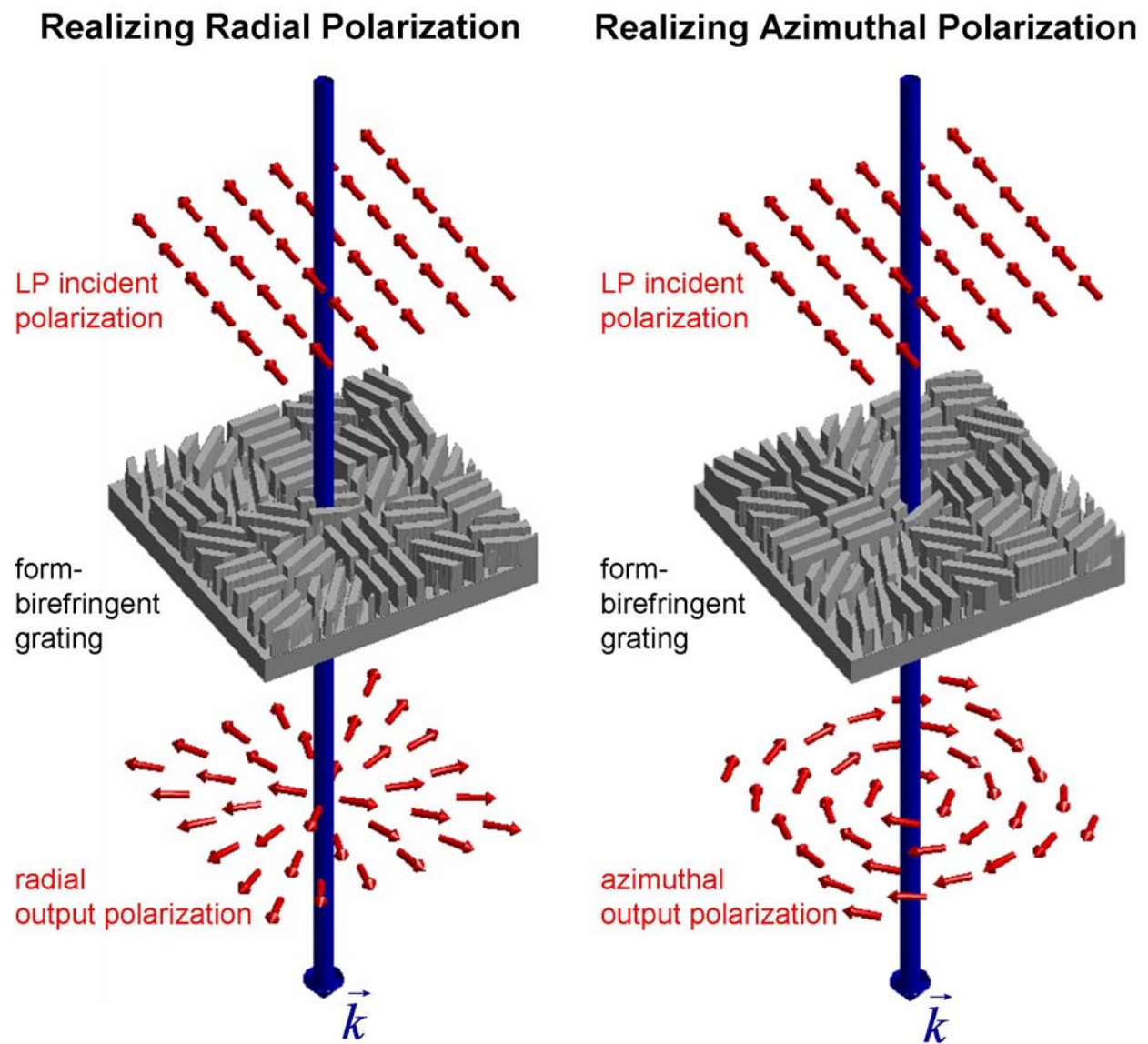

Figure 12-4: Discrete half-wave form-birefringent devices

12.3.1. Quarter-Wave Designs

The quarter-wave plate function of a binary grating may be exploited to convert a collimated and uniform beam of circularly polarized light into a collimated beam with linear polarization. Sign of the transmitted polarization has to be controlled using an additional phase element if that is needed.

The grating can be described by a grating vector function $\vec{K}$ that is a function of position. For devices with circular symmetry, the grating vector function is best expressed in polar coordinates as

$$\vec{K}(r,\theta) = K_0(r,\theta)\left\{\hat{a}_r \cos\left[f(r,\theta)\right] + \hat{a}_\theta \sin\left[f(r,\theta)\right]\right\} \tag{12.4}$$

$$K_0(r,\theta) = \frac{2\pi}{\Lambda(r,\theta)} \tag{12.5}$$

The scalar function $K_0(r,\theta)$ is the spatial period of the grating, and the function $f(r,\theta)$ characterizes grating orientation. Borrowing from Figure 12-2, radial polarization can be generated from circular polarization if $f(r,\theta) = -45°$ everywhere. This reduces Eq. (12.4) to

$$\vec{K}(r,\theta) = \frac{K_0(r,\theta)}{\sqrt{2}}\left(\hat{a}_r - \hat{a}_\theta\right) \tag{12.6}$$

To suppress diffraction that would occur if the grating were not continuous, the grating vector function is forced to be continuous by requiring

$$\nabla \times \vec{K} = 0 \tag{12.7}$$

A governing equation for the grating period is derived by substituting Eq. (12.6) into Eq. (12.7).

$$\frac{1}{r}\left\{\frac{\partial}{\partial r}\left[rK_0(r,\theta)\right] + \frac{\partial}{\partial \theta}K_0(r,\theta)\right\} = 0 \tag{12.8}$$

For radial symmetry, dependence on θ can be removed. This reduces the differential equation to

$$\frac{\partial}{\partial r} K_0(r) + \frac{1}{r} K_0(r) = 0 \tag{12.9}$$

This is a standard first-order linear differential equation that is easily solved for $K_0(r)$. The solution can be written as

$$K_0(r) = \frac{2\pi}{\Lambda_0} \frac{r_0}{r} \tag{12.10}$$

The grating vector function can be written by substituting this solution into Eq. (12.6) to obtain

$$\vec{K}(r) = \frac{1}{r} \frac{2\pi}{\sqrt{2}} \frac{r_0}{\Lambda_0} (\hat{a}_r - \hat{a}_\theta) \tag{12.11}$$

The phase function $\phi(r,\theta)$ of the grating is related to the grating vector function through the following equation.

$$\nabla \phi(r,\theta) = \vec{K}(r,\theta) \tag{12.12}$$

Substituting Eq. (12.11) into this equation leads to a differential equation governing the phase function.

$$\frac{\partial \phi}{\partial r} \hat{a}_r + \frac{1}{r} \frac{\partial \phi}{\partial \theta} \hat{a}_\theta = \frac{1}{r} \frac{2\pi}{\sqrt{2}} \frac{r_0}{\Lambda_0} (\hat{a}_r - \hat{a}_\theta) \tag{12.13}$$

Two scalar differential equations can be written by equating vector components on both sides of this equation.

$$\frac{\partial \phi}{\partial r} = \frac{1}{r} \frac{2\pi}{\sqrt{2}} \frac{r_0}{\Lambda_0} \tag{12.14}$$

$$\frac{\partial \phi}{\partial \theta} = -\frac{2\pi}{\sqrt{2}} \frac{r_0}{\Lambda_0} \tag{12.15}$$

These are both standard first order differential equations that are easily solved. The solutions are

$$\phi_1(r,\theta) = \frac{2\pi}{\sqrt{2}} \frac{r_0}{\Lambda_0} \ln\left(\frac{r}{r_0}\right) \tag{12.16}$$

$$\phi_2(r,\theta) = -\frac{2\pi}{\sqrt{2}} \frac{r_0}{\Lambda_0} \theta \tag{12.17}$$

The overall solution to Eq. (12.13) is the sum of both solutions.

$$\phi(r,\theta) = \phi_1(r,\theta) + \phi_2(r,\theta) = \frac{2\pi}{\sqrt{2}} \frac{r_0}{\Lambda_0} \left[\ln\left(\frac{r}{r_0}\right) - \theta \right] \tag{12.18}$$

Continuity in the radial direction requires $\phi(r,\theta) - \phi(r,\theta + 2\pi) = m(2\pi)$ where m is any integer.

Applying this to Eq. (12.18) leads to an important constraint on how r_0 and Λ_0 are chosen. This

is

$$\frac{2\pi}{\sqrt{2}} \frac{r_0}{\Lambda_0} = m \qquad m = \ldots, -2, -1, 0, 1, 2, \ldots \tag{12.19}$$

A relief function $d(r,\theta)$ for a binary grating can be constructed from the phase function $\phi(r,\theta)$

according to

$$d(r,\theta) = \begin{cases} 0 & \cos\left[\phi(r,\theta)\right] \le 0 \\ L_\pi/2 & \cos\left[\phi(r,\theta)\right] > 0 \end{cases} \tag{12.20}$$

A similar procedure can be presented to realize azimuthal polarization. This requires

$f(r,\theta) = 45°$ everywhere. Following same procedure as above, the grating vector function and

phase function are found to be

$$\vec{K}(r,\theta) = \left(\frac{2\pi}{\sqrt{2}} \frac{r_0}{\Lambda_0} \right)(\hat{a}_r + \hat{a}_\theta) \tag{12.21}$$

$$\phi(r,\theta) = \left(\frac{2\pi}{\sqrt{2}}\frac{r_0}{\Lambda_0}\right)\left[\ln\left(\frac{r}{r_0}\right)+\theta\right] \qquad (12.22)$$
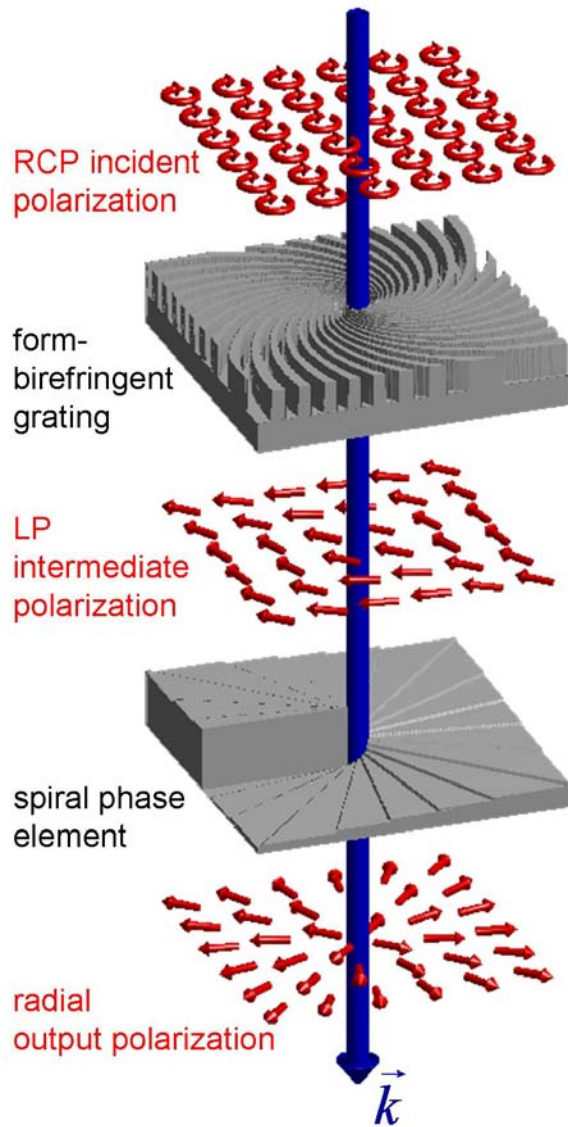
Figure 12-5: Continuous quarter-wave form-birefringent devices

Figure 12-5 illustrates how quarter-wave designs have to be implemented to realize azimuthal or radial polarization from right-hand circular polariz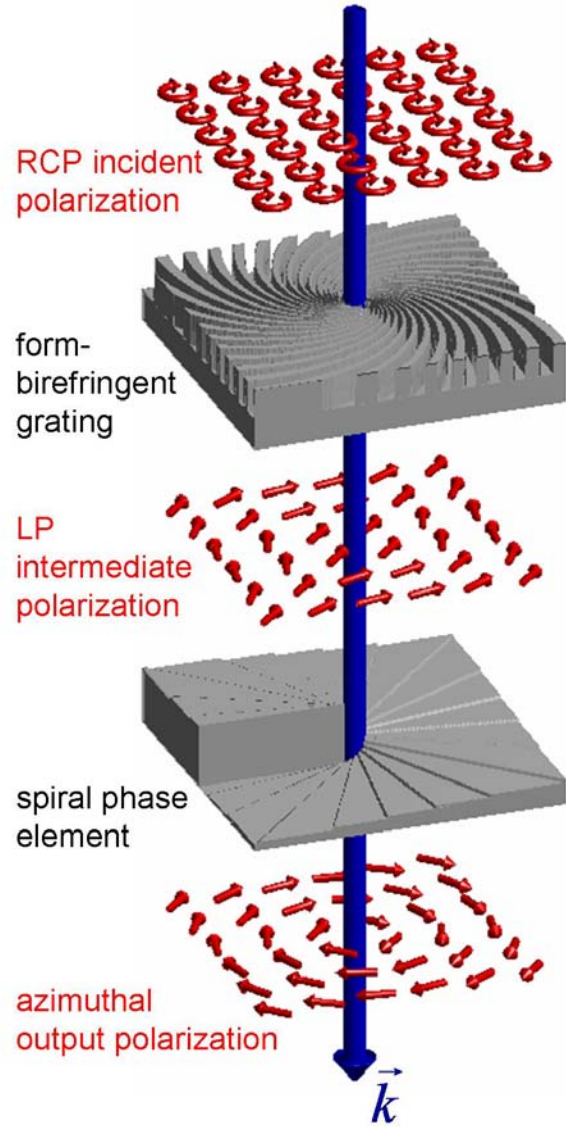ation (RCP). While continuous quarter-wave plates can convert RCP to linear polarization (LP), they cannot do this and control phase simultaneously. An additional spiral phase element (0 to 2π phase) must be added to realize radial or azimuthal polarization. It is clear the only difference between the form-birefringent devices is the direction of the spiral. This indicates reversing polarization handedness will reverse the function of the device.

12.3.2. Half-Wave Designs

A half-wave plate function may be used to convert a collimated and uniform beam of linearly polarized light into a collimated beam with linear polarization of arbitrary orientation throughout the column. This approach does not require an additional phase element to produce the required phase, but the grating must be twice as thick. High aspect ratio structures are difficult to fabricate and distortion in grating geometry leads to poor efficiency. This issue will be mitigated by forming high aspect ratio devices using autocloning.

Starting from Eq. (12.4), the grating vector function to convert linear polarization to azimuthal polarization is

$$\vec{K}(r,\theta) = K_0(r,\theta)\left[\hat{a}_r \sin(\theta/2) + \hat{a}_\theta \cos(\theta/2)\right] \tag{12.23}$$

This is a more complicated function than Eq. (12.6), but is treated using the same procedure. To enforce continuity, this new expression is substituted into Eq. (12.7) to obtain

$$\frac{1}{r}\left\{\frac{\partial}{\partial r}\left[rK_0(r,\theta)\cos(\theta/2)\right] - \frac{\partial}{\partial\theta}\left[K_0(r,\theta)\sin(\theta/2)\right]\right\} = 0 \tag{12.24}$$

It will be assumed that K_0 has no θ dependence. This reduces the differential equation to

$$\frac{\partial}{\partial r} K_0(r) + \frac{1}{2r} K_0(r) = 0 \qquad (12.25)$$

While this appears similar to Eq. (12.9), the factor of 2 multiplying r is very significant. The solution to this first-order linear differential equation can be written as

$$K_0(r) = \frac{2\pi}{\Lambda_0} \sqrt{\frac{r_0}{r}} \qquad (12.26)$$

The overall grating vector function is constructed by substituting this result into Eq. (12.23).

$$\vec{K}(r,\theta) = \frac{2\pi}{\Lambda_0} \sqrt{\frac{r_0}{r}} \left[\hat{a}_r \sin\left(\frac{\theta}{2}\right) + \hat{a}_\theta \cos\left(\frac{\theta}{2}\right) \right] \qquad (12.27)$$

The phase function is derived by substituting this expression into Eq. (12.12) to obtain

$$\frac{\partial \phi}{\partial r} \hat{a}_r + \frac{1}{r} \frac{\partial \phi}{\partial \theta} \hat{a}_\theta = \frac{2\pi}{\Lambda_0} \sqrt{\frac{r_0}{r}} \left[\hat{a}_r \sin\left(\frac{\theta}{2}\right) + \hat{a}_\theta \cos\left(\frac{\theta}{2}\right) \right] \qquad (12.28)$$

From this, two differential equations can be written by equating polarization components on both sides of the equation.

$$\frac{\partial \phi}{\partial r} = \frac{2\pi}{\Lambda_0} \sqrt{\frac{r_0}{r}} \sin\left(\frac{\theta}{2}\right) \qquad (12.29)$$

$$\frac{1}{r} \frac{\partial \phi}{\partial \theta} = \frac{2\pi}{\Lambda_0} \sqrt{\frac{r_0}{r}} \cos\left(\frac{\theta}{2}\right) \qquad (12.30)$$

Both of these first-order linear differential equations are easily solved. The solutions are

$$\phi_1(r,\theta) = \frac{4\pi r_0}{\Lambda_0} \sqrt{\frac{r}{r_0}} \sin\left(\frac{\theta}{2}\right) \qquad (12.31)$$

$$\phi_2(r,\theta) = \frac{4\pi r_0}{\Lambda_0} \sqrt{\frac{r}{r_0}} \sin\left(\frac{\theta}{2}\right) \qquad (12.32)$$

The overall solution is the sum of both solutions above. Adding these together yields

$$\phi(r,\theta) = \phi_1(r,\theta) + \phi_2(r,\theta) = \frac{8\pi r_0}{\Lambda_0} \sqrt{\frac{r}{r_0}} \sin\left(\frac{\theta}{2}\right) \tag{12.33}$$

It should not be surprising that rotating the grating by 90° will produce radial polarization. Similarly, radial polarization could be produced by changing handedness of the incident polarization. It follows that the grating vector function and phase function to produce radial polarization from the same incident polarization is

$$\vec{K}(r,\theta) = \frac{2\pi}{\Lambda_0} \sqrt{\frac{r_0}{r}} \left[\hat{a}_r \cos\left(\frac{\theta + 90°}{2}\right) + \hat{a}_\theta \sin\left(\frac{\theta + 90°}{2}\right) \right] \tag{12.34}$$

$$\phi(r,\theta) = \frac{8\pi r_0}{\Lambda_0} \sqrt{\frac{r}{r_0}} \sin\left(\frac{\theta + 90°}{2}\right) \tag{12.35}$$

Figure 12-6 shows continuous binary grating geometries that produce radial and azimuthal polarizations from linear polarization using a single optical element. The grating relief depth is twice that of the quarter-wave designs, but no additional phase element is needed.
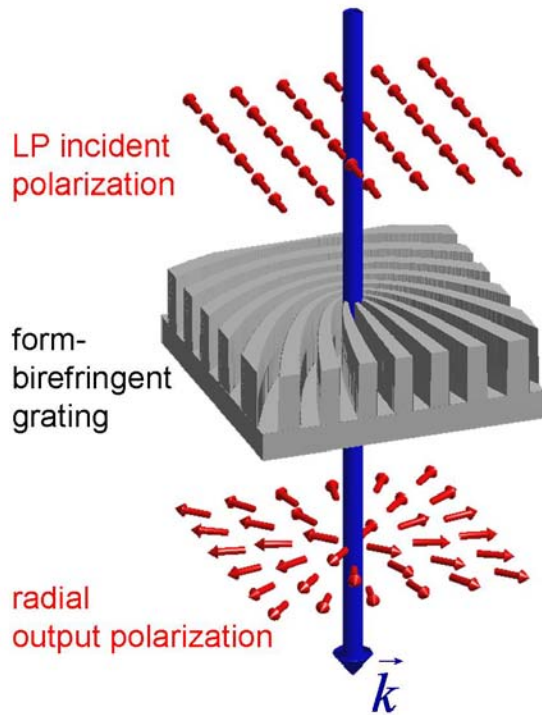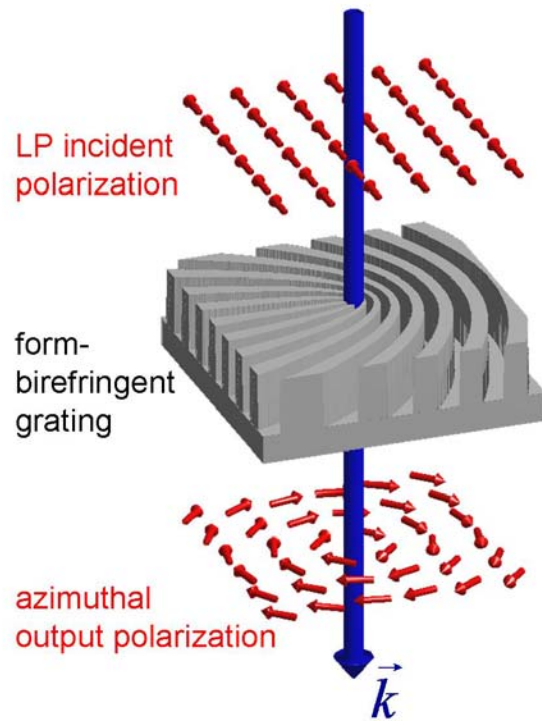
Figure 12-6: Continuous half-wave form-birefringent devices

12.4. Autocloned Birefringent Structures

Concepts and devices discussed so far require an underlying structure to produce

birefringence. These are typically binary gratings. An optimal structure provides high

transmission efficiency with high birefringence over a broad range of wavelengths. In

Ref. [256], a comparison was made between triangular and rectangular grating profiles. It was

shown that the optimum profile was a hybrid that can be described as a rectangular grating with

positively sloped sidewalls. Despite this, the rectangular binary grating remains the dominant

profile due to its easy fabrication and simple analysis.
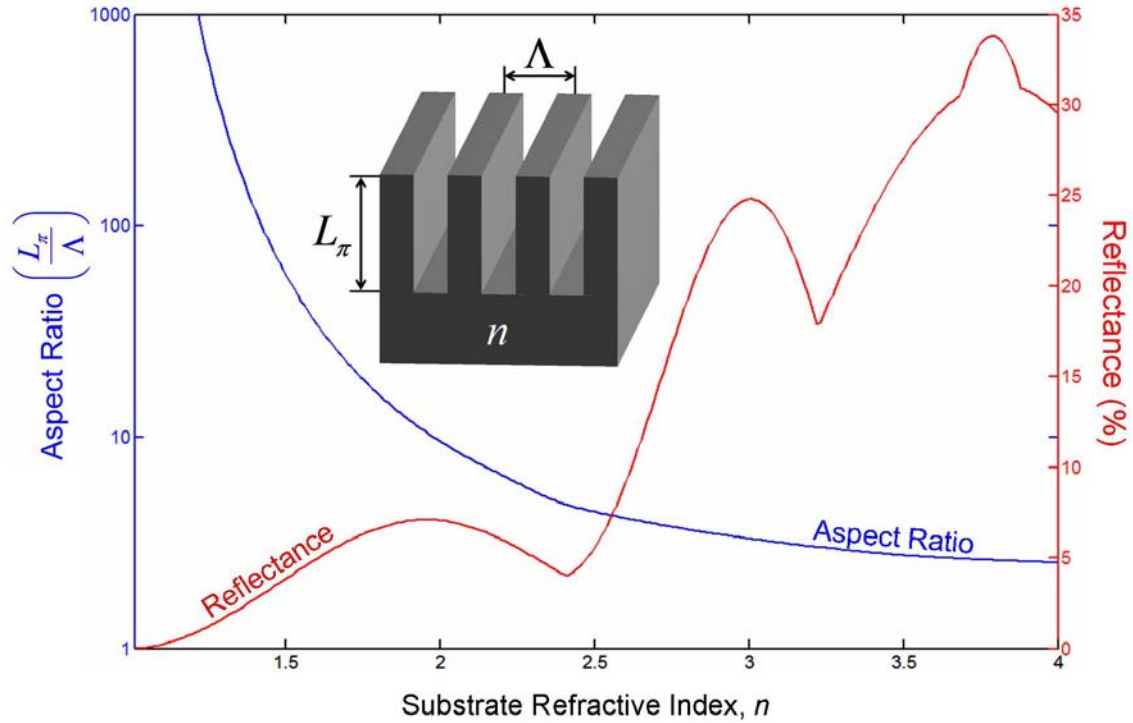
Figure 12-7: Tradeoff between aspect ratio and reflectance for binary gratings

From Figure 12-3, grating period should be kept smaller than a half wavelength to operate the device in a region with near constant birefringence as grating parameters are varied. Thickness of the grating for half-wave designs is required to be L_π. Aspect ratio of a grating is defined as grating relief depth divided by grating period. High aspect ratios are very difficult to fabricate. Substrates with higher refractive indices provide stronger birefringence leading to lower aspect ratio devices, but reflectance is higher. Substrates with lower refractive indices have lower reflection, but provide weaker birefringence and require higher aspect ratio features to compensate. Figure 12-7 summarizes this tradeoff between aspect ratio and reflectance for binary gratings. Grating period was set to $\lambda_0/2n$. This figure suggests an optimum refractive index is around 2.5. Regardless, aspect ratios exceeding 5:1 are needed.
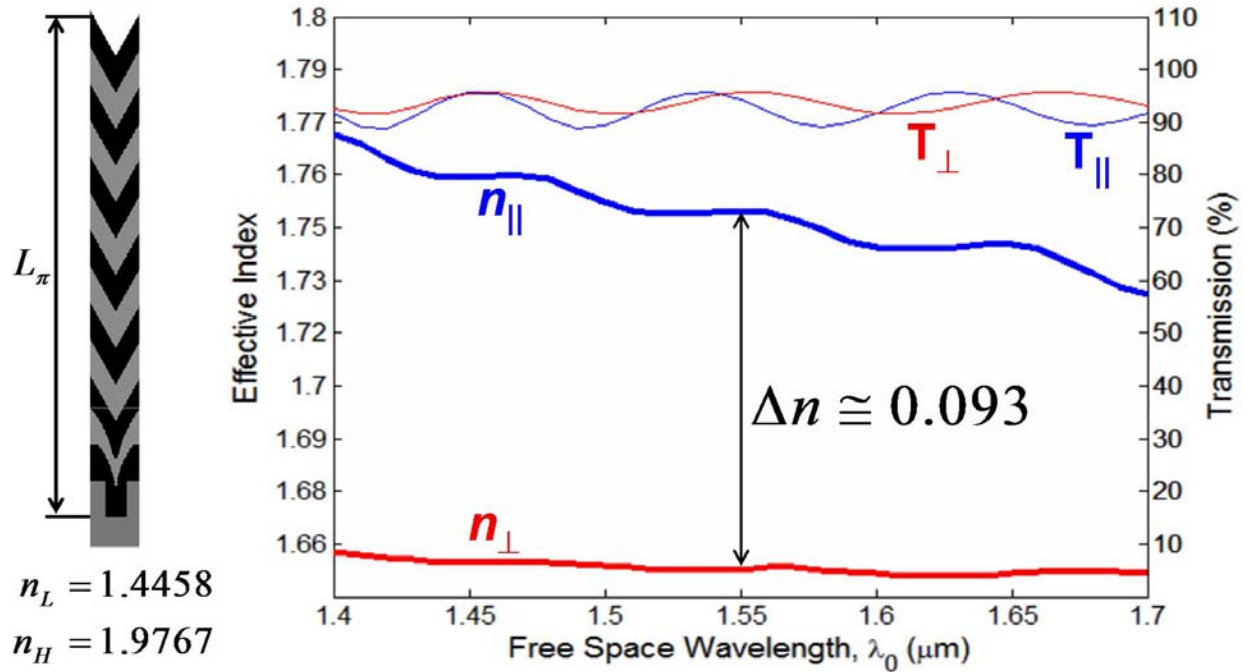
Figure 12-8: Birefringence and transmittance of autocloned structure

To realize both low reflectance and easy fabrication, this dissertation investigated a birefringent structure formed by autocloning lower index materials onto a low aspect ratio binary grating [144]. Reflectance is less due to lower refractive index and triangular surface profile produced by autocloning. To compensate for lower birefringence, devices with very high effective aspect ratios can be formed because virtually an unlimited number of layers can be deposited onto the surface while maintaining a consistent triangular surface profile. Birefringence and transmittance of an autocloned structure is depicted in Figure 12-8. For this structure, L_π=8.3 μm resulting in a structure with an effective aspect ratio just over 10:1. For this device, transmittance remains above 90% at all wavelengths.

12.5. Optimization of Birefringence

It is clear from Figure 12-3 that grating period should be kept much less than a half wavelength in the substrate region. For binary gratings, birefringence depends on refractive index and duty cycle and can be made relatively insensitive to grating thickness and period if period is made sufficiently small. While it is easy to show that grating period should be kept much less than the wavelength, it is less clear how other parameters should be chosen. Duty cycle of the surface profile will always converge to 50% due to physics inherent in the autocloning process, so it is fixed. Slope of the surface profile is also fundamental to the physics of autocloning, so it is fixed. Parameters that can be adjusted are refractive indices and layer thicknesses. For work performed in this dissertation, materials used were SiO_2 (n_1=1.4458), and SiN (n_2=1.9767). While these values are fixed, it is straight forward to conclude that birefringence is directly proportional to the difference between these refractive indices.

Layer thicknesses can be described in terms of a longitudinal period Λ_z and a longitudinal duty cycle f_z. Birefringence in an autocloned structure as a function of these parameters is provided in Figure 12-9. From this data is can be concluded that the alternating layers should be of equal thickness, but as thin as possible. A sharp increase in birefringence is observed as layer thickness approaches zero. Longitudinal period Λ_z is given relative to the transverse period Λ_x.
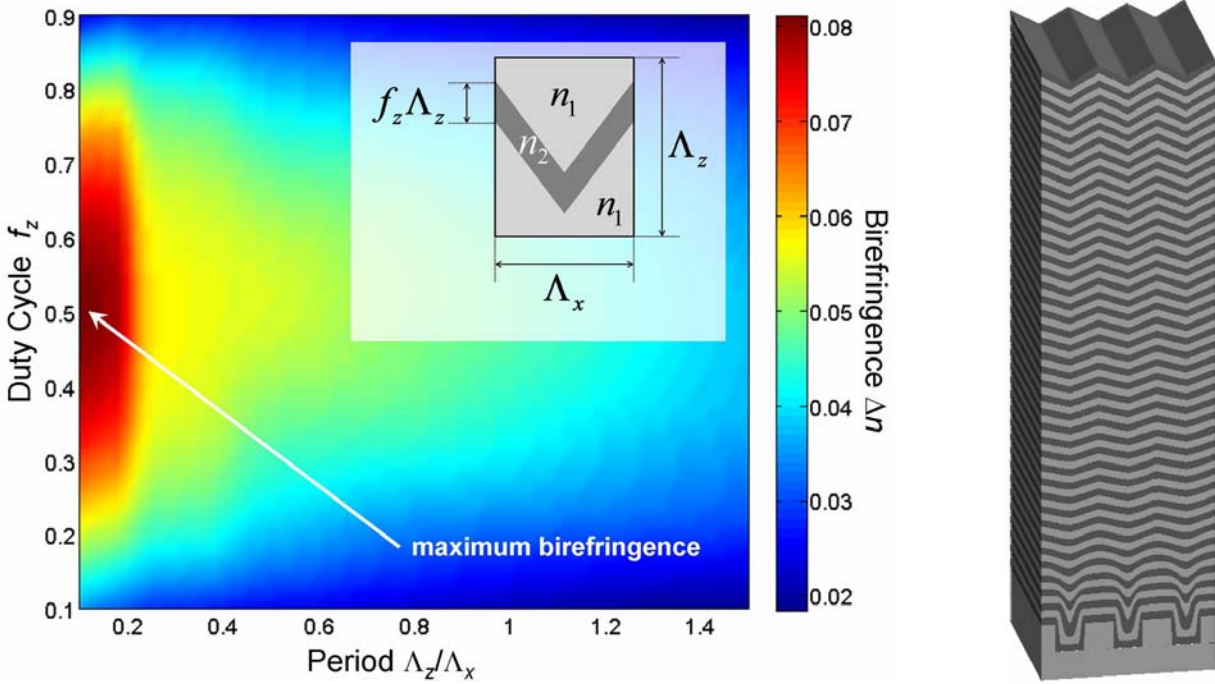
Figure 12-9: Optimization of autocloned structure for birefringence

12.6. Device Concept

Combining the concepts discussed in this chapter, a form-birefringent device constructed by autocloning onto a shallow grating may look like that shown in Figure 12-10. This is the central portion of a larger device with an effective aspect ratio of 10:1. The autocloned structure is birefringent throughout its volume and higher aspect ratios are easily achieved.
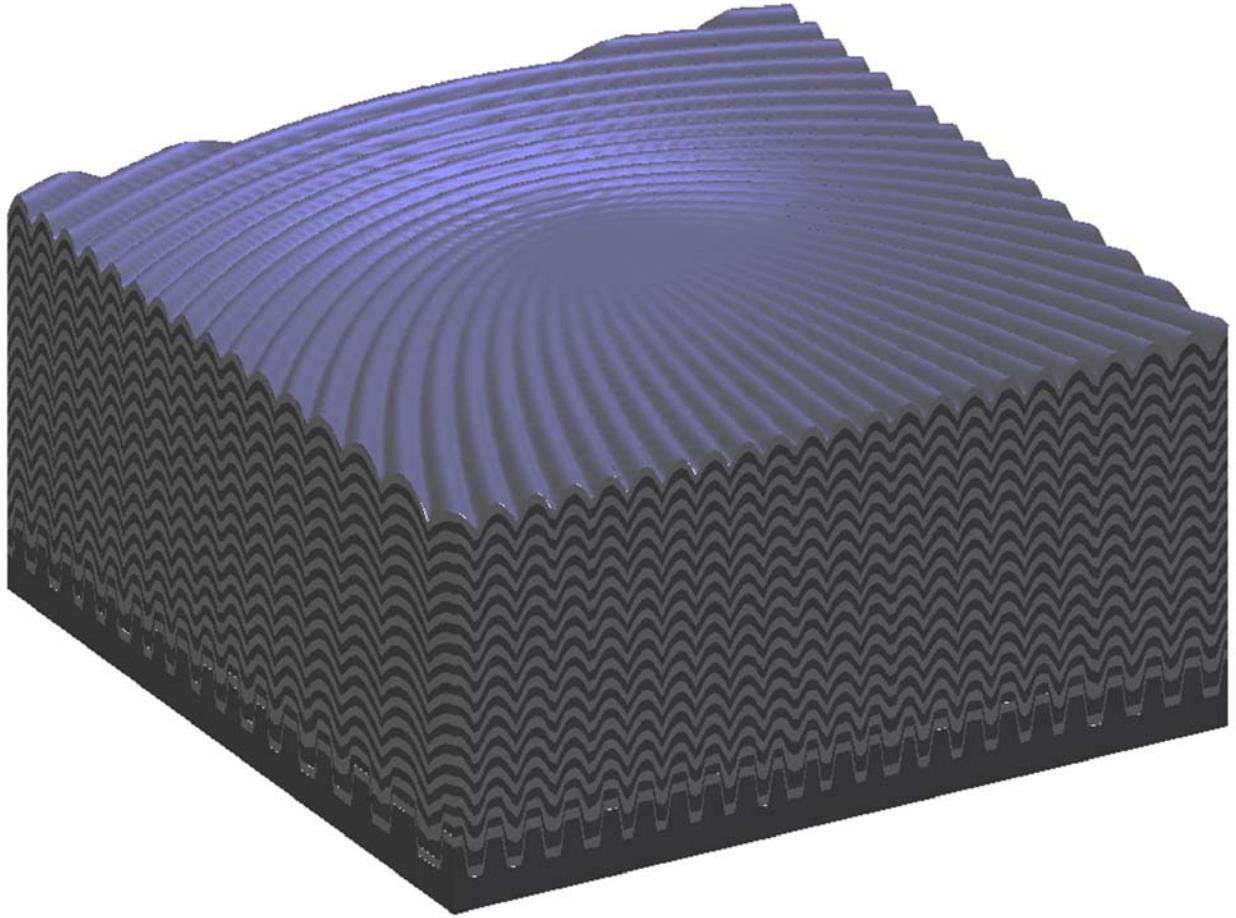
Figure 12-10: Autocloned form-birefringent optical element

12.7. Conclusion

This chapter gave an overview and description of form birefringent devices. It outlined a design procedure for both discrete and continuous devices made from binary gratings. It was shown that binary gratings are often needed with very high aspect ratio. A fundamental tradeoff was identified between device efficiency and aspect ratio of the grating grooves.

This chapter investigated the use of autocloning to build form-birefringent elements from lower refractive index materials to maximize transmittance. In this case birefringence is low, but

birefringent structures with extremely high aspect ratios can be fabricated by autocloning. The string method was used to model the process and was incorporated into optimization codes that found birefringence was maximum when film thickness was minimum. A concept diagram of a functional element was provided to illustrate the concept.

CHAPTER 13
METALLIC PHOTONIC CRYSTALS

13.1. Overview

Metallic photonic crystals were the first periodic structures to be studied and implemented for controlling electromagnetic wave propagation [73, 74]. They were the first materials where the bulk properties of an artificial dielectric were engineered [74]. Metallic photonic crystals offer some important advantages over their dielectric counterparts. The presence of free charges in metals leads to stronger interaction with photons and properties such as dispersion and complete photonic band gaps are more easily achieved. Fill factor of metallic lattices can be less than 1% enabling devices to be smaller and more lightweight than could be achieved with dielectrics alone. While it is possible to exhibit negative refraction in a dielectric photonic crystal, it is not known how to realize a negative refractive index or negative permeability without metals.

Behavior of metallic lattices depends heavily on light polarization and whether the metallic elements are continuous or disconnected. In lattices where the electric field is polarized parallel to the continuous metallic elements, the crystal exhibits a "plasmon-like" band gap that extends all the way to zero-frequency. These materials behave analogous to bulk metals, as first proposed by Pendry [257]. When the metallic structures are discontinuous or electric field is polarized perpendicular to the elements, the materials behave much more like dielectric photonic crystals. Smaller discontinuities lead to smaller bands of transmission at low frequencies. Three-dimensional lattices with continuous elements will always exhibit plasmon-like photonic

band gaps regardless of light polarization, while discontinuous lattices will always behave more like dielectric photonic crystals.

Metallic photonic crystals can be very lossy and their properties tend to change abruptly at high frequencies. This restricts most applications to frequencies below where metals become lossy. This is typically at infrared wavelengths and longer. This chapter will explain general behavior observed in metallic photonic crystals and theories will be presented to estimate their properties.

13.2. Continuous Verses Discontinuous Metallic Elements

There is a profound difference between the optical behavior of continuous metallic structures and discontinuous. A comparison of their photonic band structures is provided in Figure 13-1 for a simple cubic lattice. In the left diagram, the metallic bars are continuous and the crystal exhibits a complete plasmon-like photonic band gap that extends to zero frequency. For low frequencies, optical behavior can be approximated using an effective medium theory developed by Pendry [257].

The right diagram shows the same crystal where the continuous bars have been disconnected at the interface between unit cells. The crystal exhibits a complete photonic band gap, but not a plasmon-like band gap. In this sense, discontinuous metallic lattices behave more like dielectric photonic crystals than continuous metallic crystals. In this case, optical behavior at low frequencies can be approximated using Maxwell-Garnett and Bruggeman effective medium theories [258].

275
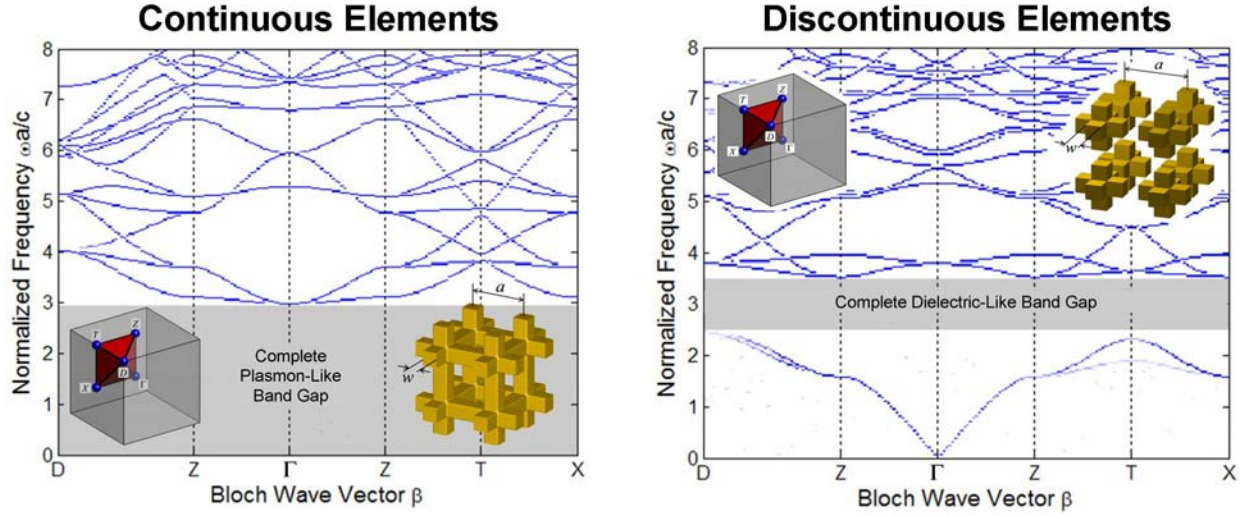
Figure 13-1: Comparison of lattices with continuous and discontinuous metallic elements

13.2.1.   *Pendry Effective Medium Theory*

Consider the periodic array of continuous metallic wires illustrated in Figure 13-2. When illuminated by an incident wave polarized parallel to the wires, currents are induced which generate magnetic fields that interact with the applied wave. From this configuration, Pendry derived simple analytical expressions for effective atomic density and effective electron mass [257]. These equations remain a good approximation even when wire thickness is relatively large [87].

$$N_{\text{eff}} = \frac{\pi r^2 N}{a^2} \qquad (13.1)$$

$$m_{\text{eff}} = \frac{\mu_0 \pi r^2 N q^2}{2\pi} \ln\left(\frac{a}{2r}\right) \qquad (13.2)$$

Substituting these expressions into Eq. (1.52), the effective plasma frequency of the crystal is

$$\omega^2_{p,\text{eff}} = \frac{2\pi c^2}{a^2 \ln(a/2r)} \tag{13.3}$$
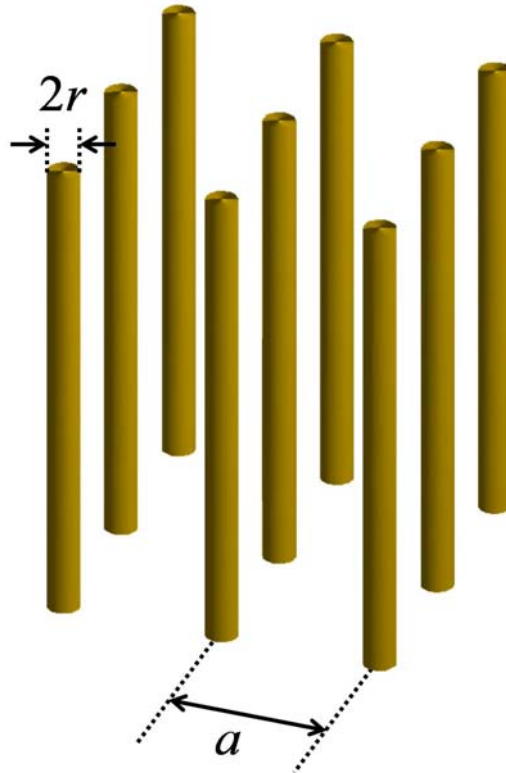
Figure 13-2: Square array of thin metallic wires

Waves at frequencies below the effective plasma frequency are forbidden to propagate in the periodic lattice. The analogous behavior of bulk metals reflecting all frequencies below the plasma frequency leads to this phenomenon being called a "plasmon-like" photonic band gap. A useful feature of the effective plasma frequency is that it can be controlled through the structural

parameters *a* and *r*. For very low frequencies above the plasma cutoff, metallic photonic crystals behave like homogeneous dielectrics exhibiting a controllable plasma resonance.

13.2.2.   *Maxwell-Garnett and Bruggeman Effective Medium Theories*

The method of Pendry requires metallic elements to be continuous. When they are not, this condition is no longer valid and a transmission band will exist at low frequencies. When the wavelength is much larger than the lattice constant, virtually all discontinuous geometries can be approximated by either Maxwell-Garnett or Bruggeman models [258, 259]. These are illustrated in Figure 13-3.

Figure 13-3: Maxwell-Garnet and Bruggeman composite material models

The Maxwell-Garnet model addresses the case where nanoparticles of material ε_1 are dispersed in a continuous host material ε_2. This could be a solution of metallic nanoparticles

suspended in a dielectric fluid. Given the volume fill factor f of material ε_1, the effective dielectric constant ε_{MG} of the overall medium satisfies the following equation.

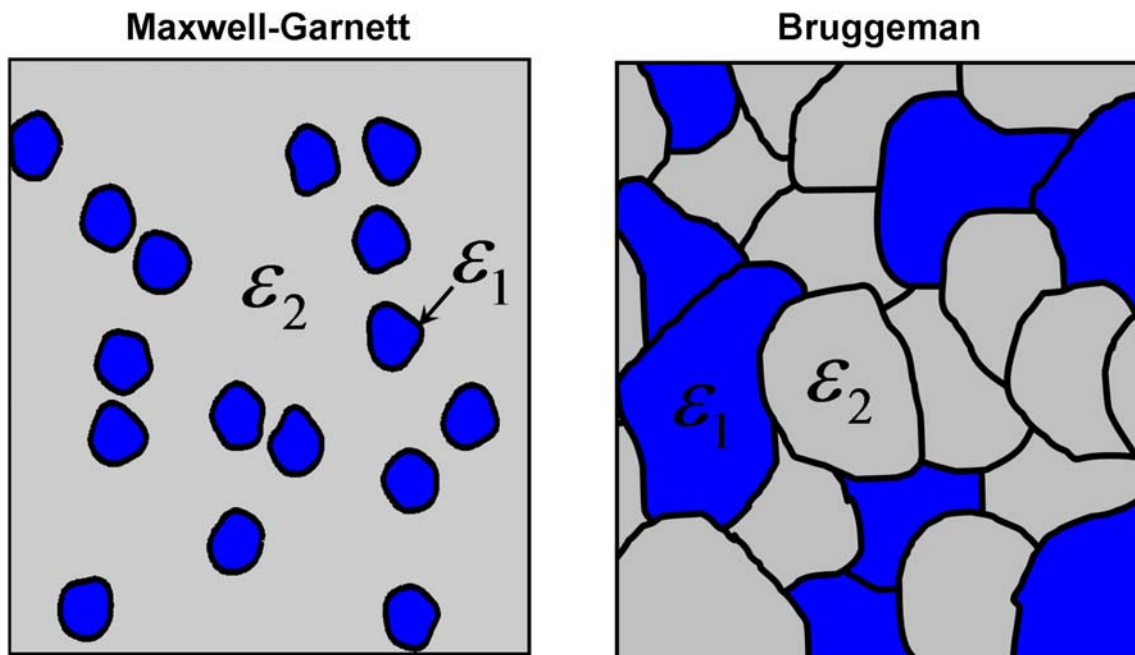$$\frac{\varepsilon_{MG} - \varepsilon_2}{\varepsilon_{MG} + 2\varepsilon_2} = f \frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1 + 2\varepsilon_2} \tag{13.4}$$

The Bruggeman model can be viewed as a very dense case of Maxwell-Garnett. The parameter f is better interpreted as the probability that any block of material will contain ε_1. The effective dielectric constant ε_{Br} of the overall medium satisfies the following equation.

$$f \frac{\varepsilon_1 - \varepsilon_{Br}}{\varepsilon_1 + 2\varepsilon_{Br}} + (1 - f) \frac{\varepsilon_2 - \varepsilon_{Br}}{\varepsilon_2 + 2\varepsilon_{Br}} = 0 \tag{13.5}$$

An extension of these theories to three-component composite materials can be found in Ref. [259].

13.3. Dielectric Verses Metallic Lattices

Similarities and differences between dielectric and metallic photonic crystals can be identified from the data presented in Figure 13-4. This figure compares the reflection from five periods of similar photonic crystals where the rods in one is composed of pure dielectric ($n=1.53$) and other of pure silver. The crystals are constructed from square rods of width 800 nm to form a simple cubic lattice with lattice constant 3.2 µm. Normal incidence was assumed.

Dielectric photonic crystals always exhibit a transmission band at low frequencies because loss is negligible and waves experience an average dielectric constant. Scattering is low so resonances cannot be sufficiently established to prohibit propagation. High reflection is exhibited within the band gap associated with the Bragg condition [260]. The amount of suppression depends on how many lattice periods are present. In this case, a peak reflection just

below 60% is achieved with just five periods. Beyond the first band gap, the dielectric crystal

shows a series of transmission and reflection bands with decreasing contrast between the two.
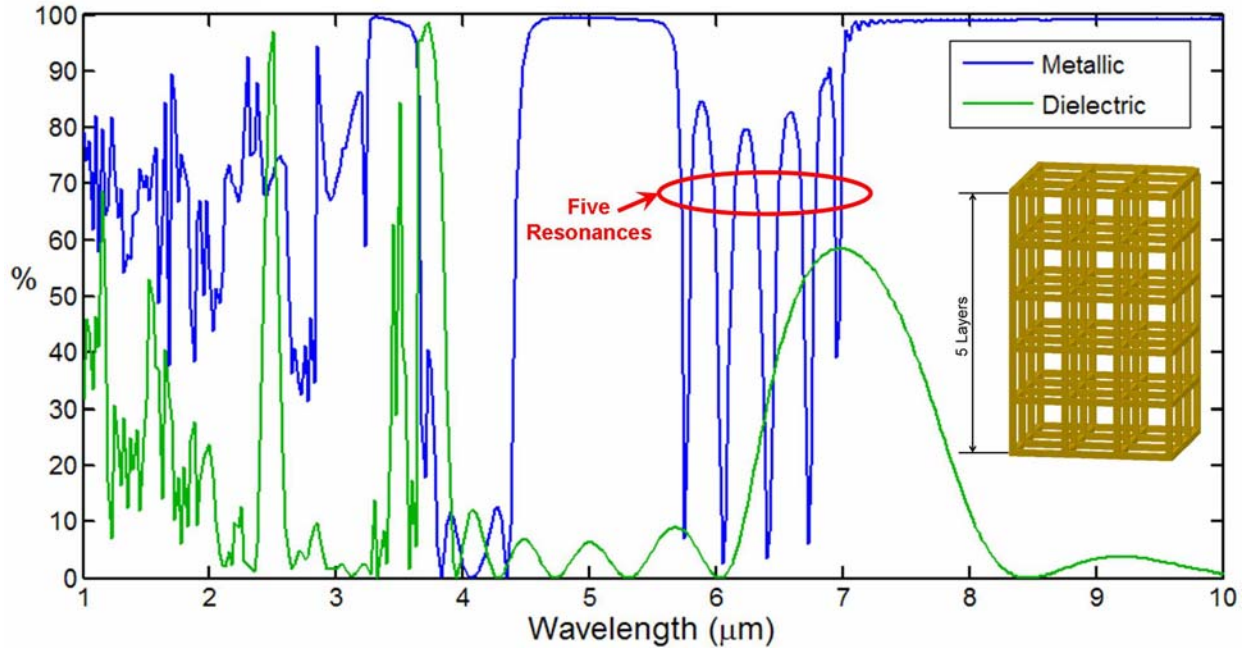
Figure 13-4: Reflection spectra from similar dielectric and metallic photonic crystals

The metallic photonic crystal exhibits a high reflection band to low frequencies because

of the plasmon-like photonic band gap. The first transmission band occurs at the short

wavelength side of the dielectric band gap. Here, the metallic lattice behaves as an ensemble of

coupled Fabry-Perot cavities. Five resonances can be counted in the transmission band

corresponding to the five lattice periods in the longitudinal direction. At frequencies above the

first transmission band, the spectrum shows a similar alternation of transmission and reflection

bands as the dielectric photonic crystal.
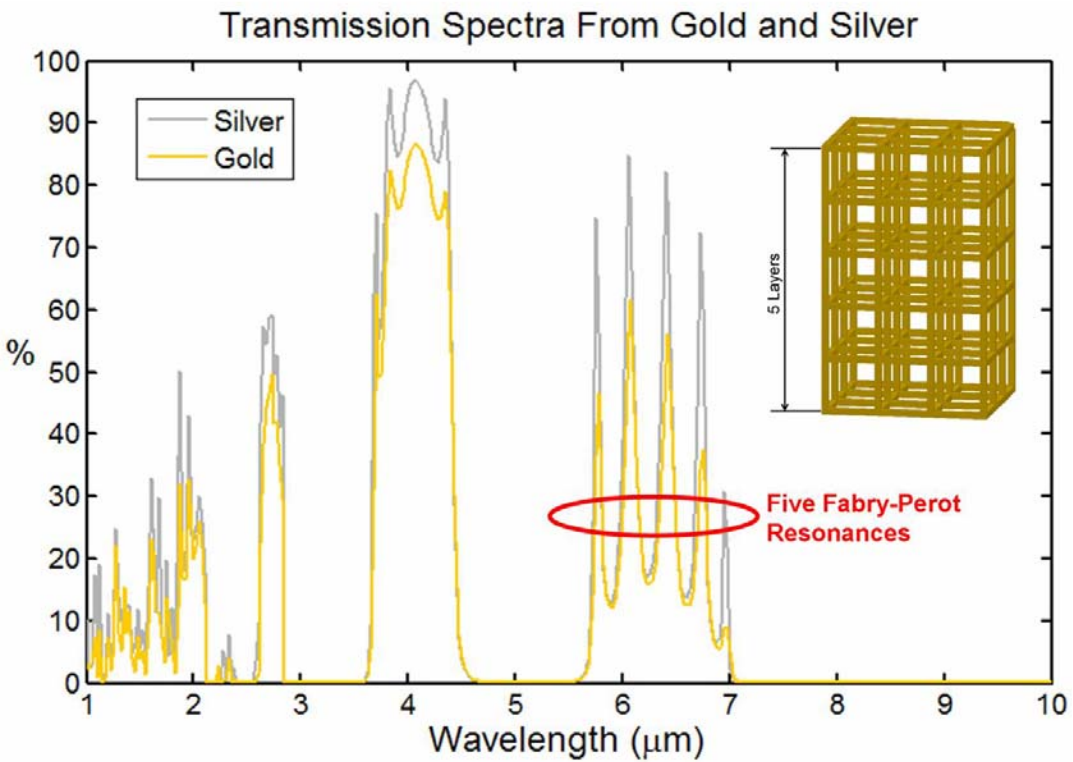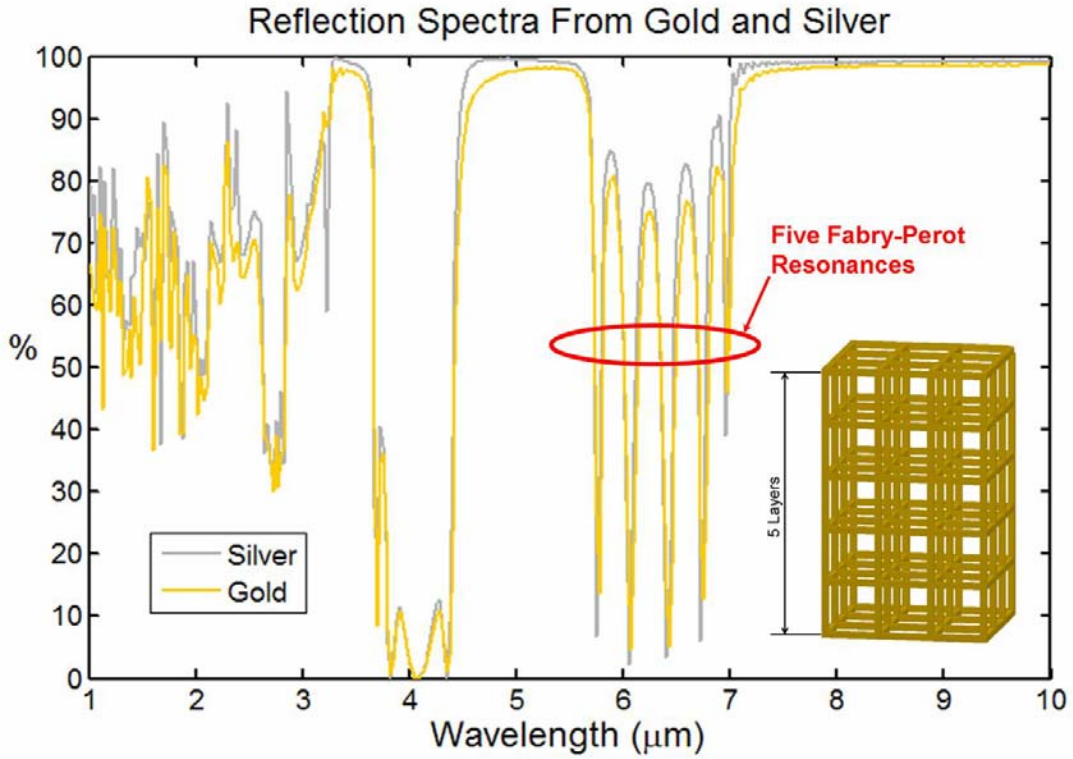
Figure 13-5: Gold versus silver metallic coatings

13.4. Gold Verses Silver Metallic Coatings

To study the impact of different metallic coatings on the same photonic crystal, the structure discussed above was modeled with gold and silver rods. Transmission and reflection spectra are compared in Figure 13-5. In the wavelength range for this simulation, little difference can be observed other than gold producing more loss.

13.5. Photonic Band Diagram of Negative Refractive Index Materials

The photonic band diagram was calculated for the negative refractive index (NRI) material described in Ref. [32] and is depicted in Figure 13-6. It shows both a complete 3D band gap and an effective negative refractive index for the bands just above the band gap.

In dielectric photonic crystals, the lowest order bands are typically V-shaped to be consistent with the light line at low frequencies. The next higher bands typically have an inverted-V shape because they are continuations of the lowest bands, but folded due to the lattice being periodic. For the structure considered here, the upper bands are not inverted indicating the material has an effective negative refractive index over a large band of frequencies.
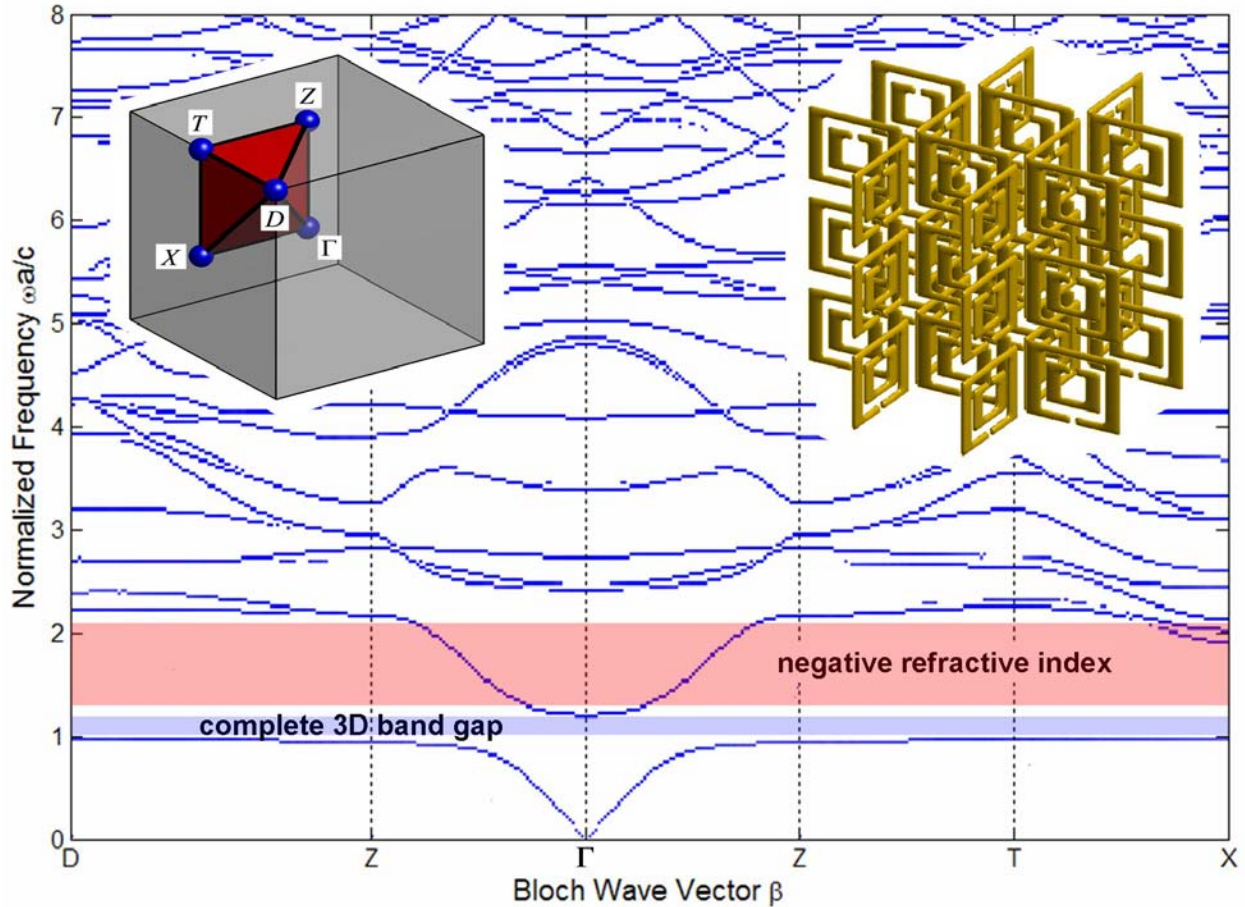
Figure 13-6: Photonic band diagram of Pendry NRI material

13.6. Investigation of Realistic Geometry

Figure 13-7 shows a scanning electron microscope (SEM) image of a photonic crystal formed in a polymer resin by two-photon lithography [261]. A metallic crystal is formed after the structure is coated with 100 nm of silver or gold using a wet deposition process. The rods are noticeably rounded with some degree of bulging at the lattice joints.
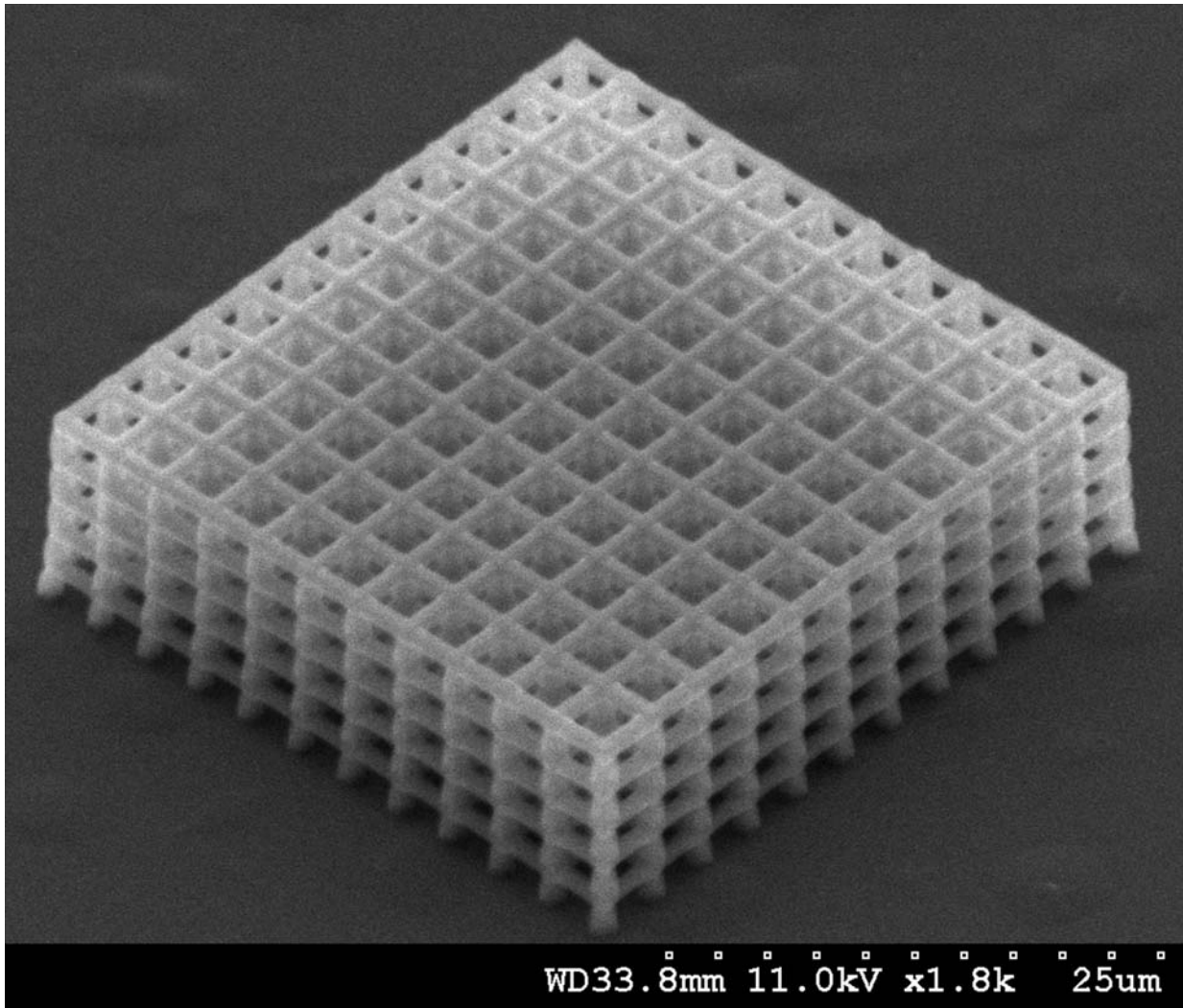
Figure 13-7: Simple cubic photonic crystal formed by two-photon lithography

To assess the impact of fabrication on optical behavior, the photonic band diagrams shown in Figure 13-8 were calculated. Bulges observed at the lattice joints were incorporated in the geometry and was found to shift the edge of the plasmon-line photonic band gap to higher frequencies. This is most likely due to increased fill factor. Higher frequency bands are quite altered indicating the dispersive properties of the lattice have changed dramatically at these wavelengths. A much larger partial band gap is observed for light at near normal incidence.
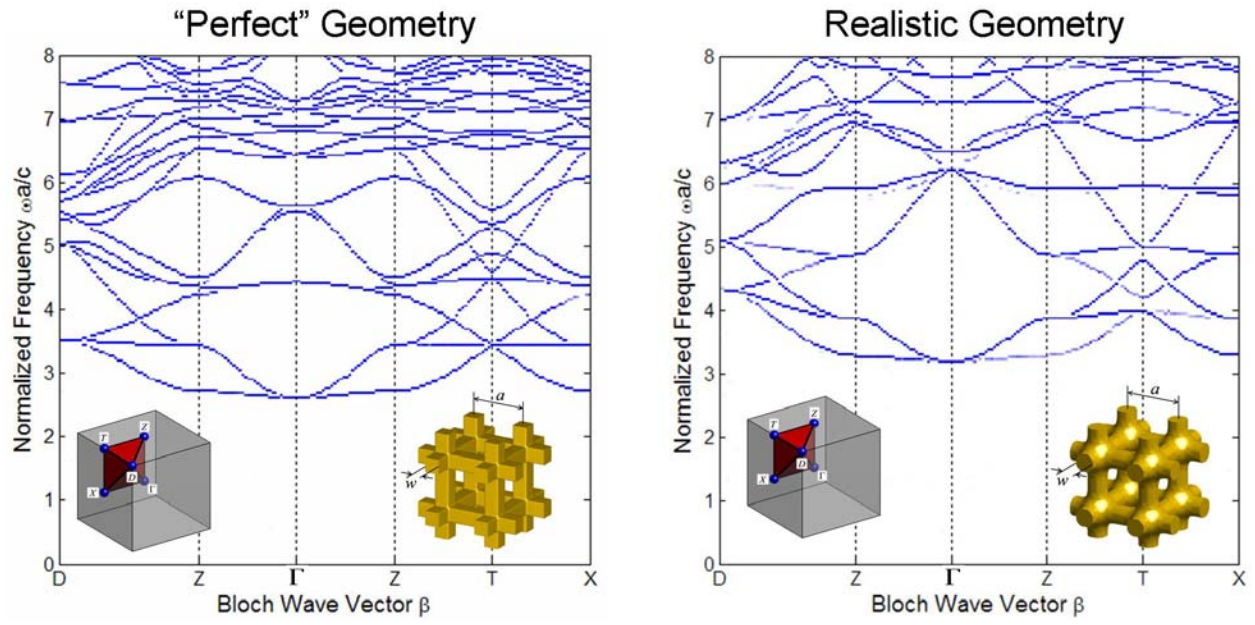
Figure 13-8: Comparison between "perfect" lattice and realistic lattice

CHAPTER 14
CONCLUSION

14.1. Background

This dissertation designed and optimized several nano-optical elements by considering how fabrication affects their optical behavior. Background was given to understand device theory. Numerical tools were discussed and detailed formulations were provided for each along with block diagrams summarizing their implementation. Key results obtained in this dissertation are summarized below.

To better understand electromagnetic theory of nano-optical elements and concepts used to develop numerical models, background information was provided in the first two chapters. CHAPTER 1 introduced Maxwell's equations from which wave equations were derived. These were used to explain scaling theories and introduce the concept of left-handed materials. Microscopic models of the response of materials were presented in the form of the Lorentz oscillator model for dielectrics and the Drude model to metals. From these, basic properties of bulk optical materials were identified and described.

CHAPTER 2 provided an introduction to photonic crystals and the tools used to analyze them. Photonic crystal symmetry was introduced in terms of the fourteen Bravais lattices. A mathematical framework was presented using lattice vectors to quantify their symmetry. Concepts of reciprocal lattices, Wigner-Seitz cells, and Brillouin zones were described. The wave equation was presented as an eigen-value problem and important properties arising from this interpretation were identified. Constructing and interpreting photonic band diagrams was

explained in detail while avoiding the mathematical details. Physical origin of the photonic band gap was explained and dispersive behavior was compared to the response of real dielectrics.

14.2. Numerical Methods

Numerical methods appropriate for modeling fabrication and optical behavior of nano-optical elements must be able to handle large scale problems with high volumetric complexity. Typical run-time of the models was on the order of minutes to days. Simple dielectric structures could be modeled with RCWA in minutes, while metallic structures modeled with FDTD often took several days. The process models all ran within minutes. All simulations were performed using 2 GHz Pentium IV processors.

The finite-difference frequency-domain method proved very powerful for visualizing fields, modeling structures containing metals, or modeling structures with large transverse dimensions. An improved formulation of the method using matrix operators for derivatives on a Yee grid was presented. A new method for incorporating arbitrary sources was outlined. While not discussed in this dissertation, the general approach of matrix operators on the Yee was extended to model full vector hybrid modes of waveguides, and used to formulate beam propagation methods (BPM), method of lines (MOL), and more.

The finite-difference time-domain method was formulated to account for material dispersion. This was done using material polarization and polarization current as intermediate parameters. The potential exists to visualize these parameters and derive physical meaning from them. It was also shown how to use the method to calculate photonic band diagrams. FDTD is a particularly powerful method when metals are considered.

The plane wave expansion method proved very powerful for constructing photonic band diagrams of dielectric structures and calculating effective properties of waves in periodic media.

A unique formulation was presented to construct the eigen-value problem explicitly. This offers easier formulation, easer implementation, and code that is more compact and easer to read. It enabled easier truncation of spatial harmonics and straightforward incorporation of tensor material properties.

RCWA proved to be an incredibly fast and efficient method to model scattering from periodic media where transverse dimensions were on the order of a wavelength. The method made of use of the enhanced transmittance matrix approach to cascade multiple layers because the structures being modeled did not warrant scattering matrices.

To model fabrication, this dissertation made use of the string method for deposition, etching, and autocloning. The FMM was used to model developing. Earlier work used the cell volume method to simulate the developing process, but FMM proved to be orders of magnitude faster, more accurate, and easier to implement. It is somewhat restricted because the rate function can only be a function of position, and the surface can only progress in a single direction. For this reason, the string method was used to model deposition and etching where the rate function is highly dependant on additional factors like visibility to a target and visibility to other points on the surface. The string method is inherently unstable and requires considerable effort to make it stable. Therefore, it has trouble resolving complicated surfaces and does not extend well to three dimensions. In this dissertation, 3D topologies were constructed from the string method by assuming circular symmetry. This is a valid assumption unless hole size is large enough that adjacent holes interact.

14.3. Design and Optimization

It was initially proposed to use genetic algorithms (GA) [262-264] or particle swarm optimization (PSO) [265] to design and optimize nano-optical elements. For the structures

addressed in this dissertation, that approach was not warranted because the degrees of design freedom were sufficiently understood or restricted.

CHAPTER 9 focused on design and analysis of a fabrication process instead of a specific device. This was near-field nano-patterning and was shown to offer high immunity to mechanical vibration and enable partially coherent light to be used. A method for synthesizing a four-beam exposure to realize any of the fourteen Bravais lattices was described. A grating design approach was outlined that showed grating symmetry controlled the radial orientation of diffracted modes, grating period controlled the angle of diffracted modes within the diffraction planes, and the pattern and relief depth of the grating controlled diffraction efficiency. Two example gratings were designed and implemented in the lab to realize 3D photonic crystals. Numerical tools from previous chapters were assembled to construct a comprehensive model of holographic lithography. This is the first known effort to incorporate physics of lithography into simulations of photonic crystals. It was predicted that holographic lithography will always form chirped photonic crystals where fill factor varies with depth. A numerical study was outlined that predicted NFNP could be performed on a standard mask aligner. Experimental results validated this predication when 3D photonic crystals were formed on a Quintell Ultraline 7000.

A space-variant photonic crystal filter was designed and optimized in CHAPTER 10 to enable a single CCD to detect color. Two material systems were analyzed based on a simplified fabrication process of etching holes through hetero-structured substrates. A defect was introduced into the lattice by modifying thickness of one layer. A transmission peak appeared at the center of the photonic band gap. Position of the peak in frequency could be tuned by varying hole size in the transverse direction. This enables custom tuning of the transmission peak at each pixel in the CCD array using a single mask process. Device geometry was optimized and

performance at oblique angles of incidence was studied. It was shown that if hole radius varied with depth, performance of the device was seriously degraded. It was shown this could be partially mitigated by adjusting thickness of each layer.

CHAPTER 11 addressed design and optimization of guided-mode resonance filters to fabricate devices with accurately placed resonances. The GMR filter is very sensitive to structural deformations making it very difficult to achieve an accurately placed resonance without active tuning during operation. A device geometry was devised that was sensitive to etching or deposition. These processes were simulation and devices were designed that enabled them to be tuned to the correct resonance at the time of fabrication. It was shown the tune-by-etching process yielded devices with better performance.

Design and optimization of form-birefringent structures was addressed in CHAPTER 12. Most form birefringent devices use subwavelength binary gratings to provide birefringence. It was shown that there is a fundamental tradeoff between transmission efficiency and aspect ratio of the grating grooves. It was proposed in this dissertation to use lower refractive index materials to improve efficiency and exploit the autocloning process to realize birefringent structures with very high aspect ratio. It was shown that birefringence could be optimized when layer thickness is made as small as possible. Design methods were outlined and equations were derived for constructing continuous form-birefringent devices to convert linear polarization into fields with radial or azimuthal polarization.

CHAPTER 13 introduced metallic photonic crystals and described some of the key theories. It was shown that lattices with continuous metallic elements behaved like bulk metals with a controllable plasma resonance, while lattices with discontinuous elements behaved more like dielectric photonic crystals. Lattices comprised only of dielectric were compared to similar

lattices made of metal. Metallic lattices of gold and silver were compared showing gold to be lossier. Preliminary results were presented for investigating the impact of fabrication on the performance of a metallic photonic crystal formed by two-photon lithography. It was shown that bulges at the lattice joints had a striking effect on optical behavior.

14.4. Recommendations for Future Work

The finite-difference frequency-domain method is an incredibly powerful method and easy to implement. It is less feasible for full vector 3D simulations due to large matrix size. Future efforts may focus on simple techniques that enable the method to be easily extended to 3D problems.

The finite-difference time-domain method is likely the most commonly used numerical technique in existence for rigorous electromagnetic simulation. A serious limitation is the inability to handle oblique angles of incidence when periodic boundaries are used while maintaining the broadband nature of FDTD. Future work may focus on methods of accomplishing this.

The plane wave expansion method may be the most popular method for calculating photonic band diagrams of dielectric photonic crystals. The method is not directly able to account for material dispersion because frequency is made the eigen-value. Future work on this method could focus on efficient methods of incorporating material dispersion.

Rigorous coupled-wave analysis is an incredibly fast and efficient method for rigorous simulation of diffractive structures with dimensions on the order of a wavelength. The method is less efficient for larger structures or structures with high volumetric complexity mainly due to calculation of eigen-modes in each layer. It is particularly problematic for TM polarization in such media. Future work could address these issues.

The near-field nano-patterning process enables fabrication of nano-optical structures on standard mask aligners that virtually all cleanroom facilities are equipped. Holographic lithography is likely to see most applications where large volume photonic crystals are needed. To make NFNP a viable technology, it will have to be performed on very thick films. In this dissertation, film thickness never exceeded 7 μm. The angular spectrum of the light will be a more significant problem for thicker films and methods for addressing this issue will likely be needed. Future work may also focus on incorporating additional optics into the phase mask to produce defects in the photonic crystals. Using the method to form nano-optical structures other than photonic crystals could be equally fertile areas of research.

Space-variant filters addressed in CHAPTER 10 only perform chromatic filtering. Polarization sensitive imaging systems could also be envisioned. Future work may entail devising other photonic crystal structures that manipulate polarization or improve performance at oblique angles of incidence.

There is much room for future work with regard to the guided-mode resonance filters discussed in CHAPTER 11. The device consisted of a single dielectric layer functioning as the grating and guiding layer. Future work may focus on improving suppression away from resonance by considering multilayer structures. This will likely entail tuning intermediate layers at intermediate wavelengths. It was suggested in this chapter that the doubly periodic geometry produced by deposition may be exploited to improve or tailor the performance of the device.

Autocloning dielectric layers onto a formed surface relief grating was shown to provide very high aspect ratio birefringent structures. Future work may address autocloning materials with higher index contrast or investigate deposition methods that are not strictly autocloning. No published literature can be found detailing a rigorous simulation of the performance of a form-

birefringent structure that encompasses transmission through the entire device aperture. Developing numerical methods to address this would enable rigorous analysis of field with singularities such as radial and azimuthal polarizations.

Metallic structures will likely find more applications at infrared and longer wavelengths due to high loss at shorter wavelengths. Future work may use metallic structures to achieve strong spectral response from very small devices. Negative refractive index and left-handed materials are receiving considerable attention at this time, but the future of concepts like "perfect" imaging is unclear. Other aspects of metallic structures like the existence of longitudinal modes and incorporation of chiral structures are equally interesting. Numerical modeling of metallic structures is particularly challenging. Fast and efficient means of simulating large-scale and arbitrarily shaped structures is needed.

APPENDIX A

USEFUL GEOMETRY FOR PHOTONIC CRYSTALS

A.1 Geometry of Hexagonal Unit Cell

The hexagonal unit cell is often of great interest in two dimensional arrays due to the high degree of symmetry. Figure A-1 summarizes the geometry of the unit cell where the lattice constant is a.
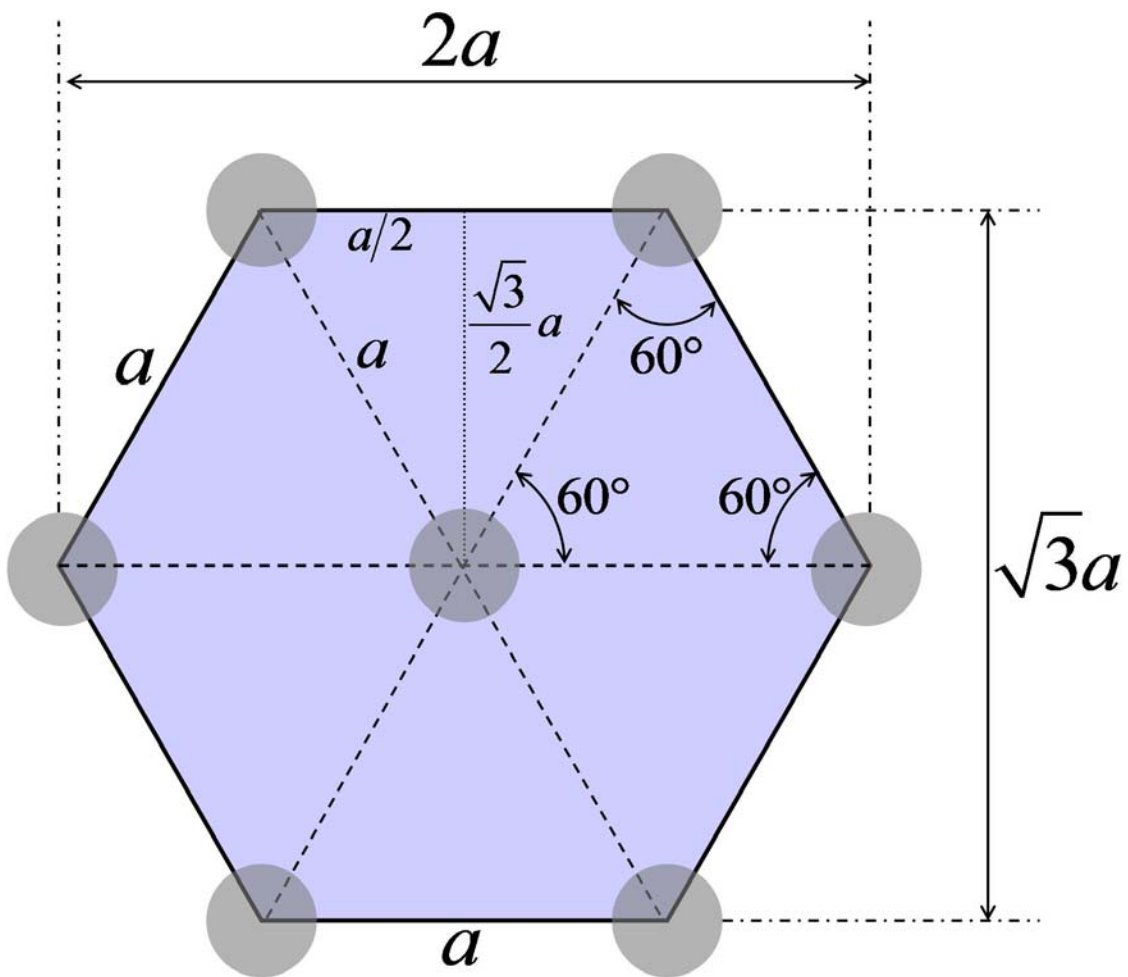
Figure A-1: Geometry of hexagonal unit cell

A.2. Vector Bisection Operation

A useful geometric operation for calculating key points of symmetry in a Brillouin zone is a vector bisection as illustrated in Figure A-2. The bisecting vector $\vec{C}$ is defined as the vector in the same plane as $\vec{A}$ and $\vec{B}$, but with a magnitude that would reach the intersection of the planes perpendicular to $\vec{A}$ and $\vec{B}$. The operation can be written symbolically as

$$\vec{C} = \text{BSV}\left[\vec{A}, \vec{B}\right] \qquad\qquad (A.1)$$
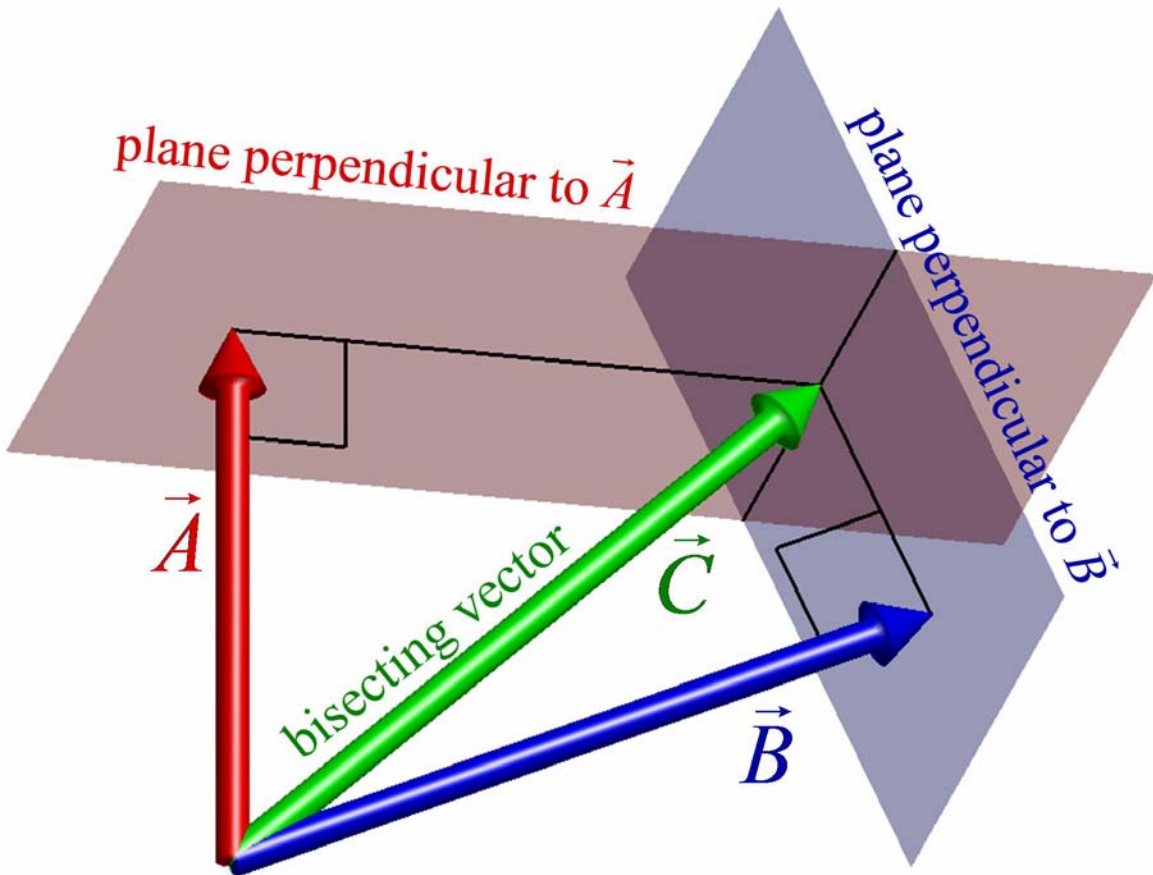
Figure A-2: Bisecting vector

The bisecting vector can be calculated through the following vector equations. The unit vector $\hat{n}$ is perpendicular to $\vec{A}$, $\vec{B}$, and $\vec{C}$. The unit vectors $\hat{a}$ and $\hat{b}$ point from the tips of $\vec{A}$ and $\vec{B}$ respectively toward the tip of $\vec{C}$. Given the distance factors α and β, vector $\vec{C}$ can be calculated as.

$$\vec{C} = \vec{A} + \alpha\hat{a} = \vec{B} + \beta\hat{b} \tag{A.2}$$

This equation can be solved for α and β through the following computations.

$$\hat{n} = \frac{\vec{A} \times \vec{B}}{\left|\vec{A} \times \vec{B}\right|} \tag{A.3}$$

$$\hat{a} = \frac{\hat{n} \times \vec{A}}{\left|\hat{n} \times \vec{A}\right|} \tag{A.4}$$

$$\hat{b} = \frac{\vec{B} \times \hat{n}}{\left|\vec{B} \times \hat{n}\right|} \tag{A.5}$$

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \hat{a} & -\hat{b} \end{bmatrix}^{-1} \begin{bmatrix} \vec{B} - \vec{A} \end{bmatrix} = \begin{bmatrix} a_x & -b_x \\ a_y & -b_y \\ a_z & -b_z \end{bmatrix}^{-1} \begin{bmatrix} B_x - A_x \\ B_y - A_y \\ B_z - A_z \end{bmatrix} \tag{A.6}$$

A.3. Calculation of Key Points of Symmetry for FCC Crystals

The Brillouin zone, irreducible Brillouin zone, and key points of symmetry for a face-centered-cubic lattice are depicted in Fig. A-3. Given the lattice vectors of the reciprocal lattice $\vec{T}_1$, $\vec{T}_2$, and $\vec{T}_3$, the key points of symmetry can be calculated using the bisecting vector function $\text{BSV}\begin{bmatrix} \vec{A}, \vec{B} \end{bmatrix}$ described above. These are

$$\vec{\Gamma} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^{\text{T}} \tag{A.7}$$

$$\vec{X} = \left(\vec{T}_2 + \vec{T}_3\right)\big/2 \tag{A.8}$$

297

$$\vec{L} = \left(\vec{T}_1 + \vec{T}_2 + \vec{T}_3\right)\big/2 \qquad\qquad (A.9)$$

$$\vec{K} = \text{BSV}\left[\vec{L}, \tfrac{1}{2}\vec{T}_3\right] \qquad\qquad (A.10)$$

$$\vec{W} = \text{BSV}\left[\vec{X}, \vec{K}\right] \qquad\qquad (A.11)$$

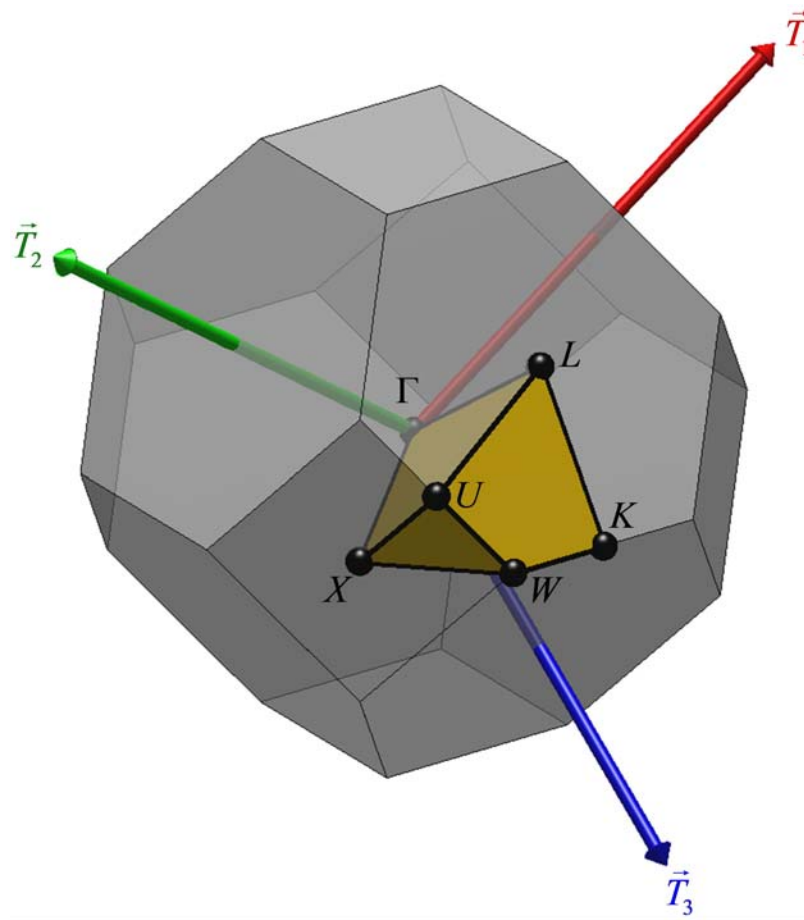$$\vec{U} = \text{BSV}\left[\vec{X}, \vec{L}\right] \qquad\qquad (A.12)$$

Figure A-3: Key points of symmetry for FCC lattice

LIST OF REFERENCES

[1]   R. Rumpf and E. G. Johnson, "Fully three-dimensional modeling of the fabrication and behavior of photonic crystals formed by holographic lithography," *Journal of the Optical Society of America A*, vol. 21, pp. 1703-1713, 2004.

[2]   R. Rumpf and E. G. Johnson, "Comprehensive modeling of near-field nano-patterning," *Optics Express*, vol. 13, 2005.

[3]   R. C. Rumpf and E. G. Johnson, "Modeling the formation of photonic crystals by holographic lithography," presented at Proceedings of SPIE Micromachining Technology for Micro-Optics and Nano-optics III, Bellingham, WA, 2005.

[4]   R. C. Rumpf and E. G. Johnson, "Micro-photonic systems utilizing SU-8," in *MOEMS and Miniaturized Systems IV*, vol. 5346, 2004, pp. 64-72.

[5]   R. C. Rumpf, P. Srinivasan, and E. G. Johnson, "Modeling the fabrication of nano-optical structures," *Proc. SPIE Micromachining Technology for Micro-Optics and Nano-Optics IV*, vol. 6110, 2006.

[6]   R. C. Rumpf, P. Srinivasan, and E. G. Johnson, "Near-Field Nano-Patterning of 3D Structures," presented at Proceedings of OSA Annual Meeting, Tucson, AZ, 2005.

[7]   P. Srinivasan, R. C. Rumpf, and E. G. Johnson, "Fabrication of 3-D photonic crystals by two-step dry etching of layered media," presented at SPIE Micromachining Technology for Micro-Optics and Nano-Optics IV, San Jose, CA, 2006.

[8]   Z. Li and X. Zhang, "Fragility of photonic band gaps in inverse-opal photonic crystals," *Phys. Rev. B*, vol. 62, pp. 1516-1519, 2000.

[9]   H. P. Herzig, *Micro-optics: Elements, systems and applications*. Philadelphia: Taylor & Francis Inc., 1998.

[10]   D. L. Dickensheets, "Imaging performance of off-axis planar diffractive lenses," *J. Opt. Soc. Am. A*, vol. 13, pp. 1849-1858, 1996.

[11]   K. Hirayama, E. N. Glytsis, and T. K. Gaylord, "Rigorous electromagnetic analysis of diffractive cylindrical lenses," *J. Opt. Soc. Am. A*, vol. 13, pp. 2219-2231, 1996.

[12] K. Yamada, W. Watanabe, Y. Li, and K. Itoh, "Multilevel phase-type diffractive lenses in silica glass induced by filamentation of femtosecond laser pulses," *Opt. Lett.*, vol. 29, pp. 1846-1848, 2004.

[13] C. Gimkiewicz, D. Hagedorn, J. Jahns, E. B. Kley, and F. Thoma, "Fabrication of microprisms for planar optical interconnections by use of analog gray-scale lithography with high-energy-beam-sensitive glass," *Applied Optics*, vol. 38, pp. 2986-2990, 1999.

[14] J. M. Bendickson, E. N. Glytsis, and T. K. Gaylord, "Metallic surface-relief on-axis and off-axis focusing diffractive cylindrical mirrors," *J. Opt. Soc. Am. A*, vol. 16, pp. 113-130, 1999.

[15] N. Bokor and N. Davidson, "Aberration-free imaging with an aplanatic curved diffractive element," *Appl. Opt.*, vol. 40, pp. 5825-5829, 2001.

[16] D. Mawet, P. Riaud, J. Surdej, and J. Baudrand, "Subwavelength surface-relief gratings for stellar coronagraphy," *Appl. Opt.*, vol. 44, pp. 7313-7321, 2005.

[17] Y. Zhang, S. Poonja, and A. Roorda, "MEMS-based adaptive optics scanning laser ophthalmoscopy," *Opt. Lett.*, vol. 31, pp. 1268-1270, 2006.

[18] J. Rasanen and K. E. Peiponen, "On-line measurement of the thickness and optical quality of float glass with a sensor based on a diffractive element," *Appl. Opt.*, vol. 40, pp. 5034-5039, 2001.

[19] S. P. Simonaho and R. Silvennoinen, "Sensing of wood density by laser light scattering pattern and diffractive optical element based sensor," *J. Opt. Technol.*, vol. 73, pp. 170-174, 2006.

[20] V. Emiliani, D. Cojoc, E. Ferrari, V. Garbin, C. Durieux, M. Coppey-Moisan, and E. D. Fabrizio, "Wave front engineering for microscopy of living cells," *Opt. Exp.*, vol. 13, pp. 1395-1405, 2005.

[21] E. D. Fabrizio, D. Cojoc, and S. Cabrini, "Diffractive optical elements for differential interference contrast x-ray microscopy," *Opt. Exp.*, vol. 11, pp. 2278-2288, 2003.

[22] E. Dai, C. Zhou, P. Xi, and L. Liu, "Multifunctional double-layered diffractive optical element," *Opt. Lett.*, vol. 28, pp. 1513-1515, 2003.

[23]   Y. Orihara, W. Klaus, M. Fujino, and K. Kodate, "Optimization and application of hybrid-level binary zone plates," *Appl. Opt.*, vol. 40, pp. 5877-5885, 2001.

[24]   H. Sasaki, I. Fukuzaki, Y. Katsuki, and T. Kamijoh, "Design considerations of stacked multilayers of diffractive optical elements for optical network units in optical subscriber-network applications," *Appl. Opt.*, vol. 37, pp. 3735-3745, 1998.

[25]   S. Jeon, G. P. Wiederrecht, and J. A. Rogers, "Photonic systems formed by proximity field nanopatterning," *Proceedings of the SPIE on Micromachining Technology for Micro-Optics and Nano-Optics III*, vol. 5720, pp. 187-195, 2005.

[26]   W. S. Mohammed, A. Mehta, M. Pitchumani, and E. G. Johnson, "Selective Excitation of the TE01 Mode in Hollow-Glass Waveguide Using a Subwavelength Grating," *IEEE Phot. Tech. Lett.*, vol. 17, pp. 1441-1443, 2005.

[27]   M. Rossi and T. Hessler, "Stray-light effects of diffractive beam-shaping elements in optical microsystems," *Appl. Opt.*, vol. 38, pp. 3068, 1999.

[28]   A. Schilling, H. P. Herzig, L. Stauffer, U. Vokinger, and M. Rossi, "Efficient beam shaping of linear, high-power diode lasers by use of micro-optics," *Appl. Opt.*, vol. 40, pp. 5852-5859, 2001.

[29]   M. R. Wang and X. G. Huang, "Subwavelength-resolvable focused non-Gaussian beam shaped with a binary diffractive optical element," *Appl. Opt.*, vol. 38, pp. 2171-2176, 1999.

[30]   Y. Zhao, Y. P. Li, and Q. G. Zhou, "Vector iterative algorithm for the design of diffractive optical elements applied to uniform illumination," *Opt. Lett.*, vol. 29, pp. 664-666, 2004.

[31]   E. Yablonovitch, "Photonic Band-gap structures," *J. Opt. Soc. Am. B*, vol. 10, pp. 283-295, 1993.

[32]   R. A. Shelby, D. R. Smith, and S. Schultz, "Experimental Verifcation of a Negative Index of Refraction," *Science*, vol. 292, pp. 77-79, 2001.

[33]   A. Chutinan and S. Noda, "Effects of structural fluctuations on the photonic band gap during fabrication of a photonic crystal: a study of a photonic crystal with a finite number of periods," *J. Opt. Soc. Am. B*, vol. 16, pp. 1398-1402, 1999.

[34]   S. Ogawa, K. Tomoda, and S. Noda, "Effects of structural fluctuations on three-dimensional photoniccrystals operating at near-infrared wavelengths," *Journal of Applied Physics*, vol. 91, pp. 513-515, 2002.

[35]   Z. Li, X. Zhang, and Z. Zhang, "Disordered photonic crystals understood by a perturbation formalism," *Phys. Rev. B*, vol. 61, pp. 15738-15748, 2000.

[36]   M. M. Sigalas, C. M. Soukoulis, C. T. Chan, and D. Turner, "Localization of electromagnetic waves in two-dimensional disordered systems," *Physics Review B*, vol. 53, pp. 8340-8348, 1996.

[37]   S. Fan, P. R. Villeneuve, and J. D. Joannopoulos, "Theoretical investigation of fabrication-related disorder on the properties of photonic crystals," *J. Appl. Phys.*, vol. 78, pp. 1415-1418, 1995.

[38]   M. M. Sigalas, C. M. Soukoulis, C. T. Chan, R. Biswas, and K. M. Ho, "Effect of disorder on photonic band gaps," *Physics Review B*, vol. 59, pp. 12767-12770, 1999.

[39]   S. Kawakami, T. Kawashima, and T. Sato, "Mechanism of shape formation of three-dimensional periodic nanostructures by bias sputtering," *Applied Physics Letters*, vol. 74, pp. 463-465, 1999.

[40]   J. Jin, *The Finite Element Method in Electromagnetics*, Second ed. New York: John Wiley & Sons, Inc., 2002.

[41]   K. Kawano and T. Kitoh, *Introduction to Optical Waveguide Analysis: Solving maxwell's equations and the Schrodinger equation*. New York: Wiley & Sons, 2001.

[42]   J. B. Pendry, "Photonic band structures," *J. Mod. Opt.*, vol. 41, pp. 209-229, 1994.

[43]   D. M. Sullivan, *Electromagnetic Simulation Using the FDTD Method*. Piscataway, NJ: IEEE Press, 2000.

[44]   A. Taflove and S. C. Hagness, *Computational Electrodynamics: the Finite-Difference Time-Domain Method, 2nd Edition*. Norwood, MA: Artech House, Inc., 2000.

[45] S. Wu and N. Glytsis, "Finite-number-of-periods holographic gratings with finite-width incident beams: analysis using the finite-difference frequency-domain method," *J. Opt. Soc. Am. A*, vol. 19, pp. 2018-2029, 2002.

[46] S. Wu and N. Glytsis, "Volume holographic grating couplers: rigorous analysis by use of the finite-difference frequency-domain method," *Appl. Opt.*, vol. 43, pp. 1009-1023, 2004.

[47] C. Yu and H. Chang, "Compact finite-difference frequency-domain method for the analysis of two-dimensional photonic crystals," *Opt. Exp.*, vol. 12, pp. 1397-1408, 2004.

[48] Z. Li and L. Lin, "Photonic band structures solved by a plane-wave-based transfer-matrix method," *Phys. Rev. E*, vol. 67, pp. 046607, 2003.

[49] M. Auslender and S. Hava, "Scattering-matrix propagation algorithm in full-vectorial optics of multilayer grating structures," *Opt. Lett.*, vol. 21, pp. 1765-1767, 1996.

[50] P. Lalanne, "Improved formulation of hte coupled-wave method for two-dimensional gratings," *J. Opt. Soc. Am. A*, vol. 14, pp. 1592-1598, 1997.

[51] M. G. Moharam, E. B. Grann, and D. A. Pommet, "Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings," *J. Opt. Soc. Am. A*, vol. 12, pp. 1068-1076, 1995.

[52] M. G. Moharam, E. B. Grann, D. A. Pommet, 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*, vol. 12, pp. 1068-1076, 1995.

[53] S. Tibuleac, "Reflection and transmission guided-mode resonance filters," *J. Opt. Soc. Am. A*, vol. 14, pp. 1617-1626, 1997.

[54] J. A. Sethian, "Curvature and the Evolution of Fronts," *Comm. in Math. Phys.*, vol. 54, pp. 425-499, 1985.

[55] S. Tazawa, S. Matsuo, and K. Saito, "A General Characterization and Simulation Method for Deposition and Etching Technology," *IEEE Trans. Semi. Man*, vol. 5, pp. 27-33, 1992.

[56]   A. J. Chorin, "Flame Advection and Propagation Algorithms," *J. Comp. Phys.*, vol. 35, pp. 1-11, 1980.

[57]   C. W. Hirt and B. D. Nicholls, "Volume of Fluid (VOF) Method for Dynamics of Free Boundaries," *J. Comp. Phys.*, vol. 39, pp. 201-225, 1981.

[58]   W. Noh and P. Woodward, "A Simple Line Interface Calculation," presented at Fifth International Conference on Fluid Dynamics, Springer-Verlag, 1976.

[59]   E. G. Puckett, "A Volume-of-Fluid Interface Tracking Algorithm with Applications to Computing Shock Wave Refraction," presented at 4th International Symposium on Computational Fluid Dynamics, Davis, CA, 1991.

[60]   E. W. Scheckler, N. N. Tam, A. K. Pfau, and A. R. Neureuther, "An efficient volume-removal algorithm for practical three-dimensional lithography simulation with experimental verification," *IEEE Transactions on Computer-Aided Design of Integrated Circuts and Systems*, vol. 12, pp. 1345-1356, 1993.

[61]   S. Osher and R. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces*. New York: Springer-Verlag New York, 2003.

[62]   J. A. Sethian, *Level Set Methods and Fast Marching Methods: Evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*. New York: Cambridge University press, 1996.

[63]   J. A. Sethian, *Level Set Methods and Fast Marching Methods: Evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*. New York, New York: Cambridge University Press, 1999.

[64]   C. A. Balanis, *Advanced Engineering Electromagnetics*. New York: Wiley and Sons, 1989.

[65]   V. G. Veselago, *Usp. Fiz. Nauk*, vol. 92, pp. 517, 1964.

[66]   V. G. Veselago, *Sov. Phys. Usp.*, vol. 10, pp. 509, 1968.

[67]   M. N. O. Sadiku, *Elements of Electromagnetics*. Orlando, FL: Sanders College Publishing, 1989.

[68]   J. H. Simmons and K. S. Potter, *Optical Materials*. San Diego, CA: Academic Press, 2000.

[69]   A. D. Rakic, A. B. Djurisic, J. M. Elazar, and M. L. Majewski, "Optical properties of metallic films for vertical-cavity optoelectronic devices," *Appl. Opt.*, vol. 37, pp. 5271-5283, 1998.

[70]   M. J. Weber, *Handbook of Optical Materials*. New York: CRC Press, 2003.

[71]   E. Yablonovitch, "Inhibited spontaneous emission in solid-state physics and electronics," *Phys. Rev. Lett.*, vol. 58, pp. 2059-2062, 1987.

[72]   S. John, "Strong localization of photons in certain disordered dielectric superlattices," *Phys. Rev. Lett.*, vol. 58, pp. 2486-2489, 1987.

[73]   G. Marconi and C. Franklin, "Reflector For Use In Wireless Telegraphy And Telephony." US Patent 1301473, 1919.

[74]   A. F. Harvey, "Periodic and Guiding Structures at Microwave Frequencies," *IRE Trans. Microwave Theory and Techniques*, vol. 8, pp. 30-61, 1960.

[75]   C. Elachi, "Waves in active and passive periodic structures: A review," *Proc. IEEE*, vol. 64, pp. 1666-1698, 1976.

[76]   M. J. A. Dood, A. Polman, and J. G. Fleming, "Modified spontaneous emission from erbium-doped, photonic, layer-by-layer crystals," *Phys. Rev. B*, vol. 67, pp. 115106, 2003.

[77]   M. A. Kaliteevski, J. M. Martinez, D. Cassagne, and J. P. Albert, "Appearance of photonic minibands in disordered photonic crystals," *Journal of Physics: Condensed Matter*, vol. 15, pp. 785-790, 2003.

[78]   T. Baba, N. Fukaya, and J. Yonekura, "Observation of light propagation in photonic crystal optical waveguides with bends," *Elect. Lett.*, vol. 35, pp. 654-655, 1999.

[79]   A. Mekis, J. C. Chen, I. Kurland, S. Fan, P. R. Villeneuve, and J. D. Joannopoulos, "High transmission through sharp bends in photonic crystal waveguides," *Phys. Rev. Lett.*, vol. 77, pp. 3787-3790, 1996.

[80] A. Talneau, L. Gouezigou, N. Bouadma, M. Kafesaki, C. M. Soukoulis, and M. Agio, "Photonic-crystal ultrashort bends with improved transmission and low reflection in 1.55 µm," *Appl. Phys. Lett.*, vol. 80, pp. 547-549, 2002.

[81] T. D. Engeness, M. Ibanescu, S. G. Johnson, O. Weisberg, M. Skorobogatiy, S. Jacobs, and Y. Fink, "Dispersion tailoring and compensation by modal interactions in OmniGuide fibers," *Opt. Exp.*, vol. 11, pp. 1175-1196, 2003.

[82] D. R. Solli, C. F. McCormick, R. Y. Chiao, and J. M. Hickmann, "Experimental observation of superluminal group velocities in bulk two-dimensional photonic band gap crystals," *IEEE Journal of Select Topics in Quantum Electronics*, vol. 9, pp. 40-42, 2003.

[83] C. Luo, S. G. Johnson, and J. D. Joannopoulos, "Negative refraction without negative index in metallic photonic crystals," *Opt. Exp.*, vol. 11, pp. 746-754, 2003.

[84] N. W. Ashcroft and N. D. Mermin, *Solid State Physics*. USA: Harcourt College Publishers, 1976.

[85] Y. S. Chan, C. T. Chan, and Z. Y. Liu, "Photonic Band Gaps in Two Dimensional Photonic Quasicrystals," *Phys. Rev. Lett.*, vol. 80, pp. 956-959, 1998.

[86] M. E. Zoorob, M. D. B. Charlton, G. J. Parker, J. J. Baumberg, and M. C. Netti, "Complete and absolute photonic bandgaps in highly symmetric photonic quasicrystals embedded in low refractive index materials," *Mat. Sci. Eng. B*, vol. 74, pp. 168-174, 2000.

[87] J. M. Lourtioz, *Photonic Crystals: Towards Nanoscale Photonic Devices*. New York: Springer, 2005.

[88] J. D. Joannopoulos, R. D. Meade, and J. N. Winn, *Photonic Crystals: Modling the flow of light*. New Jersey: Princeton University Press, 1995.

[89] S. G. Johnson and J. D. Joannopoulos, "Introduction to Photonic Crystals: Bloch's Theorem, Band Diagrams, and Gaps (But No Defects)," in *Lectures on photonic crystals*: http://ab-initio.mit.edu/photons/tutorial/photonic-intro.pdf, 2003.

[90] J. Hwang, "A Compact 2-D FDFD Method for Modeling Microstrip Structures with Nonuniform Grids and Perfectly Matched Layer," *IEEE Trans. Microwave Theory and Techniques*, vol. 53, pp. 653-659, 2005.

[91]   W. Liu, "Vector diffraction from subwavelength optical disk structures: Two-dimensional near-field profiles," *Opt. Exp.*, vol. 2, pp. 191-197, Kowarz, M. W.

[92]   E. A. Marengo, C. M. Rappaport, and E. L. Miller, "Optimum PML ABC Conductivity Profile in FDFD," *IEEE Trans. Magnetics*, vol. 35, pp. 1506-1509, 1999.

[93]   C. M. Rappaport and B. J. McCartin, "FDFD Analysis of Electromagnetic Scattering in Anisotropic Media Using Unconstrained Triangular Meshes," *IEEE Trans. Antennas and Propagation*, vol. 39, pp. 345-349, 1991.

[94]   M. A. Sharkawy, V. Demir, and A. Z. Elsherbeni, "Iterative multiregion technique for large-scale electromagnetic scattering problems: Two-dimensional case," *Radio Science*, vol. 40, RS5002, pp. 1-10, 2005.

[95]   W. Sun, K. Liu, and A. Balanis, "Analysis of Singly and Doubly Periodic Absorbers by Freqeuncy-Domain Finite-Difference Method," *IEEE Trans. Antennas and Propagation*, vol. 44, pp. 798-805, 1996.

[96]   M. N. O. Sadiku, *Numerical Techniques in Electromagnetics*. Boca Raton, FL: CRC Press, 2001.

[97]   K. S. Yee, "Numerical solution of the initial boundary value problems involving Maxwell's equations in isotropic media," *IEEE Trans. Antennas and Propagation*, vol. 14, pp. 302-307, 1966.

[98]   S. Dey and R. Mittra, "A conformal finite-difference time-domain technique for modeling cylindrical dielectric resonators," *IEEE Trans. Microwave Theory and Techniques*, vol. 47, pp. 1737-1739, 1999.

[99]   J. Fang, "Generalized Perfectly Matched Layer for the Absorption of Propagating and Evanescent Waves in Lossless and Lossy Media," *IEEE Trans. Microwave Theory and Techniques*, vol. 44, pp. 2216-2222, 1996.

[100] Y. Chen, K. Sun, B. Beker, and R. Mittra, "Unified Matrix Presentation of Maxwell's and Wave Equations Using Generalized Differential Matrix Operators," *IEEE Trans. Education*, vol. 41, pp. 61-69, 1998.

[101] S. Guo, F. Wu, and S. Albin, "Photonic band gap analysis using finite-difference frequency-domain method," *Opt. Exp.*, vol. 12, pp. 1741-1746, 2004.

[102] T. Itoh, *Numerical Techniques for Microwave and Millimeter-Wave Passive Structures*. New York: John Wiley & Sons, 1989.

[103] K. Yasumoto, *Electromagnetic Theory and Applications for Photonic Crystals*. New York: CRC Taylor & Francis, 2006.

[104] K. R. Umashankar and A. Taflove, "A novel method to analyze electromagnetic scattering of complex objects," *IEEE Trans. Electromagnetic Compatibility*, vol. 24, pp. 397-405, 1982.

[105] R. Holland, L. Simpson, and K. S. Kunz, "Finite-difference analysis of EMP coupling to lossy dielectric structures," *IEEE Trans. Electromagnetic Compatibility*, vol. 22, pp. 203-209, 1980.

[106] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing, Principles, Algorithms, and Applications*, third ed. Upper Saddle River: Prentice-Hall, Inc., 1996.

[107] S. Fan, P. R. Villeneuve, and J. D. Joannopoulos, "Large omnidirectional band gaps in metallodielectric photonic crystals," *Phys. Rev. B*, vol. 54, pp. 11245-11251, 1996.

[108] K. Sakoda, N. Kawai, T. Itoh, A. Chutinan, S. Noda, T. Mitsuyu, and K. Hirao, "Photonic bands of metallic systems. I. Principle of calculation and accuracy," *Phys. Rev. B*, vol. 64, pp. 045116-, 2001.

[109] A. J. Ward and J. B. Pendry, "A program for calculating the photonic band structures and Green's functions unsing a non-orthogonal FDTD method," *Comp. Phys. Commun.*, vol. 112, pp. 23-41, 1998.

[110] J. R. Marek and J. MacGillivray, "A method for reducing run-times of out-of-core FDTD problems," presented at 9th Annual Review of Progress in Applied Computational Electromagnetics, Monterey, CA, 1993.

[111] P. H. Harms, J. A. Roden, J. G. Maloney, M. P. Kesler, E. J. Kuster, and S. D. Gedney, "Numerical analysis of periodic structures using the split-field algorithm," presented at 13th Annual Review of Progress in Applied Computational Electromagnetics, Monterey, CA, 1997.

[112] W. Ko and R. Mittra, "Implementation of Floquet boundary condition in FDTD for FSS analysis," presented at IEEE Antennas and Propagation Society International Symposium, Ann Arbor, MI, 1993.

[113] J. A. Roden, S. D. Gedney, M. P. Kesler, and J. G. Maloney, "Time-domain analysis of periodic structures at oblique incidence: Orthogonal and nonorthogonal FDTD implementations," *IEEE Trans. Microwave Theory and Techniques*, vol. 46, pp. 420-427, 1998.

[114] M. E. Veysoglu, R. T. Shin, and J. A. Kong, "A finite-difference time-domain analysis of wave scattering from periodic surfaces: oblique incidence case," *J. Electromagnetic Waves and Applications*, vol. 7, pp. 1595-1607, 1993.

[115] S. Guo and S. Albin, "Simple plane wave implementation of photonic crystal calculations," *Opt. Exp.*, vol. 11, pp. 167-175, 2003.

[116] S. G. Johnson and J. D. Joannopoulos, "Block-iterative frequency-domain methods for Maxwell's equations in a planewave basis," *Opt. Exp.*, vol. 8, pp. 173-190, 2001.

[117] K. M. Leung and Y. F. Liu, "Photon band structures: The plane-wave method," *Phys. Rev. B*, vol. 41, pp. 10188-10190, 1990.

[118] R. D. Meade, A. M. Rappe, K. D. Brommer, and J. D. Joannopoulos, "Accurate theoretical analysis of photonic band-gap materials," *Phys. Rev. B*, vol. 48, pp. 8434-8437, 1993.

[119] S. Shouyuan, C. Chen, and D. W. Prather, "Plane-wave expansion method for calculating band structure of photonic crystal slabs with perfectly matched layers," *J. Opt. Soc. Am. A*, vol. 21, pp. 1769-1775, 2004.

[120] H. S. Sozuer and J. P. Dowling, "Photonic band calculations for woodpile structures," *J. Mod. Opt.*, vol. 41, pp. 231-239, 1994.

[121] S. G. Johnson and J. D. Joannopoulos, *Photonic Crystals: the road from theory to practice*. New York: Springer Science+Business Media, 2002.

[122] D. C. Dobson, "An efficient method for band structure calculations in 2D photonic crystals," *J. Comp. Phys.*, vol. 149, pp. 363-376, 1999.

[123] A. Greenwell and M. G. Moharam, "personal communications," 2005.

[124] L. Li, "New formulation of the Fourier modal method for cross surface-relief gratings," *J. Opt. Soc. Am. A*, vol. 14, pp. 2758-2767, 1997.

[125] M. Neviere and E. Popov, *Light Propagation in Periodic Media: Differential theory and design*. New York: Marcel Dekker, 2003.

[126] C. B. Burckhardt, "Diffraction of a Plane Wave at a Sinusoidally Stratified Dielectric Grating," *J. Opt. Soc. Am.*, vol. 56, pp. 1502-1509, 1966.

[127] N. Chateau and J. P. Hugonin, "Algorithm for the rigorous coupled-wave analysis of grating diffraction," *J. Opt. Soc. Am. A*, vol. 11, pp. 1321-1331, 1994.

[128] T. K. Gaylord and M. G. Moharam, "Analysis and Applications of Optical Diffraction by Gratings," *Proc. IEEE*, vol. 73, pp. 894-937, 1985.

[129] P. Lalanne and G. M. Morris, "Highly improved convergence of the coupled-wave method for TM polarization," *J. Opt. Soc. Am. A*, vol. 13, pp. 779-784, 1996.

[130] M. G. Moharam and T. K. Gaylord, "Rigorous coupled-wave analysis of planar-grating diffraction," *J. Opt. Soc. Am.*, vol. 71, pp. 811-818, 1981.

[131] M. G. Moharam and T. K. Gaylord, "Diffraction analysis of dielectric surface-relief gratings," *J. Opt. Soc. Am. A*, vol. 72, pp. 1385-1392, 1982.

[132] M. G. Moharam and T. K. Gaylord, "Rigorous coupled-wave analysis of metallic surface-relief gratings," *J. Opt. Soc. Am. A*, vol. 3, pp. 1780-1787, 1986.

[133] S. T. Peng, T. Tamir, and H. L. Bertoni, "Theory of Periodic Dielectric Waveguides," *IEEE Trans. Microwave Theory and Techniques*, vol. MTT-23, pp. 123-133, 1975.

[134] G. Granet and B. Guizal, "Efficient implementation of the coupled-wave method for metallic lamellar gratings in TM polarization," *J. Opt. Soc. Am. A*, vol. 13, pp. 1019-1023, 1996.

[135] L. Li and C. W. Haggans, "Convergence of the coupled-wave method for metallic lamellar diffraction gratings," *J. Opt. Soc. Am. A*, vol. 10, pp. 1184-1189, 1993.

310

[136] J. Turunen, M. Kuittinen, and P. Vahimaa, "Form-birefringence limits of Fourier-expansion methods in grating theory: arbitrary angle of incidence," *J. Opt. Soc. Am. Comm.*, pp. 2314-2316, 1997.

[137] L. Li, "Use of Fourier series in the analysis of discontinuous periodic structures," *J. Opt. Soc. Am. A*, vol. 13, pp. 1870-1876, 1996.

[138] E. Popov, M. Neviere, B. Gralak, and G. Tayeb, "Staircase approximation validity for arbitrary-shaped gratings," *J. Opt. Soc. Am. A*, vol. 19, pp. 33-42, 2002.

[139] J. A. Sethian, "An Analysis of Flame Propagation," *Ph.D. dissertation, Dept. Mathematics, University of California, Berkeley, CA*, 1982.

[140] S. Kawakami, T. Sato, K. Miura, Y. Ohtera, T. Kawashima, and H. Ohkubo, "3-D Photonic-Crystal Heterostructures: Fabrication and In-Line Resonator," *IEEE Phot. Tech. Lett.*, vol. 15, pp. 816-818, 2003.

[141] T. Kawashima, T. Sato, Y. Ohtera, and S. Kawakami, "Tailoring of the Unit Cell Structure of Autocloned Photonic Crystals," *IEEE J. Quant. Elect.*, vol. 38, pp. 899-903, 2002.

[142] K. Miura, Y. Ohtera, H. Ohkubo, N. Akutsu, and S. Kawakami, "Reduction of propagation and bending losses of heterostructured photonic crystal waveguides by use of a high-delta structure," *Opt. Lett.*, vol. 28, pp. 734-736, 2003.

[143] Y. Ohtera, T. Kawashima, Y. Sakai, T. Sato, I. Yokohama, A. Ozawa, and S. Kawakami, "Photonic crystal waveguides utilizing a modulated lattice structure," *Opt. Lett.*, vol. 27, pp. 2158-2160, 2002.

[144] T. Sato, K. Miura, N. Ishino, Y. Ohtera, T. Tamamura, and S. Kawakami, "Photonic crystals for the visible range fabricated by autocloning technique and their application," *Opt. & Quant. Elect.*, vol. 34, pp. 63-70, 2002.

[145] S. Osher and J. A. Sethian, *J. Comput. Phys.*, vol. 79, pp. 12-49, 1988.

[146] J. A. Sethian and D. Adalsteinsson, "An Overview of Level Set Methods for Etching, Deposition, and Lithography Development," *IEEE Trans. Semi. Man.*, vol. 10, pp. 167-184, 1997.

[147] A. Sheikholeslami, C. Heitzinger, F. Badrieh, H. Puchner, and S. Selberherr, "Three-Dimensional Topography Simulation Based on a Level Set Method," *27th Int'l Spring Seminar on Electronics Technology*, pp. 263-265, 2004.

[148] M. Khan, B. Bollepalli, and F. Cerrina, "A semi-empirical resist dissolution model for submicron lithographies," presented at MSM98: Technical Proceedings of the 1998 International Conference on Modeling and Simulation of Microsystems, http://www.cr.org/index.html, 1998.

[149] J. A. Sethian, "A fast marching level set method for monotonically advancing fronts," *Proc. Natl. Acad. Sci.*, vol. 93, pp. 1591-1595, 1996.

[150] J. A. Sethian, "Curvature Flow and Entropy Conditions Applied to Grid Generation," *J. Comp. Phys.*, vol. 115, pp. 440-454, 1994.

[151] R. Malladi and J. A. Sethian, "Image Processing via Level Set Curvature Flow," *Proc. Natl. Acad. of Sci.*, vol. 92, pp. 7046-7050, 1995.

[152] R. Malladi and J. A. Sethian, "Level Set Methods for Curvature Flow, Image Enhancement, and Shape Recovery in Medical Images," presented at Proc. of Conf. on Visualization and Mathematics, Berlin Germany, 1995.

[153] R. Malladi and J. A. Sethian, "Image Processing: Flows under Min/Max Curvature and Mean Curvature," *Graphical Models and Image Processing*, vol. 58, pp. 127-141, 1996.

[154] R. Malladi and J. A. Sethian, "A Unified Approach to Noise Removal, Image Enhancement, and Shape Recovery," *IEEE Trans. on Image Processing*, vol. 5, pp. 1554-1568, 1996.

[155] R. Kimmel and J. A. Sethian, "Fast Marching Methods for Computing Distance Maps and Shortest Paths," *CPAM Report 669, Univ. of California, Berkeley*, 1996.

[156] R. Kimmel and J. A. Sethian, "Fast Marching Methods for Robotic Navigation with Constraints," *Center for Pure and Applied Mathematics Report, Univ. of California, Berkeley*, 1998.

[157] L. Z. Cai, X. L. Yang, and Y. R. Wang, "All fourteen Bravais lattices can be formed by interference of four noncoplanar beams," *Opt. Lett.*, vol. 27, pp. 900-902, 2002.

[158] Z. Ling, K. Lian, and L. Jian, "Improved patterning quality of SU-8 microstructures by optimizing the exposure parameters," *Advances in Resist Techology and Processing XVII*, vol. 3999, pp. 1019-1027, 2000.

[159] Y. V. Miklyaev, D. C. Meisel, A. Blanco, G. Freymann, K. Busch, W. Kock, C. Enrich, M. Deubel, and M. Wegener, "Three-dimensional face-centered-cubic photonic crystal templates by laser holography: fabrication, optical characterization, and band-structure calculations," *Appl. Phys. Lett.*, vol. 82, pp. 1284-1286, 2003.

[160] J. M. Shaw, J. D. Gelorme, N. C. LaBianca, W. E. Conley, and S. J. Holmes, "Negative photoresists for optical lithography," *IBM J. Res. Dev.*, vol. 41, pp. 81-94, 1997.

[161] I. Divliansky and T. S. Mayer, "Three-dimensional low-index-contrast photonic crystals fabricated using a tunable beam splitter," *Nanotechnology*, vol. 17, pp. 1241-1244, 2006.

[162] I. Divliansky, T. S. Mayer, and K. S. C. V. H. Holliday, "Fabrication of three-dimensional polymer photonic crystal structures using single diffraction element interference lithography," *Appl. Phys. Lett.*, vol. 82, pp. 1667-1669, 2003.

[163] A. Bertsch, H. Lorenz, and P. Renaud, "Combining Microstereolithography and thick resist UV lithography for 3D microfabrication," *Proc. IEEE Eleventh Annual International Workshop on Micro Electro Mechanical Systems*, pp. 18-23.

[164] M. Despont, H. Lorenz, N. Fahrni, J. Brugger, P. Renaud, and P. Vettiger, "High-aspect-ratio, ultrathick, negative-tone near-UV photoresist for MEMS applications," *Proc. IEEE Tenth Annual International Workshop on Micro Electro Mechanical Systems*, pp. 518-522, 1997.

[165] S. K. Sampath, L. St. Clair, X. Wu, D. V. Ivanov, Q. Wang, C. Ghosh, and K. R. Farmer, "Rapid MEMS prototyping using SU-8, wafer bonding and deep reactive ion etching," *Proc. IEEE Fourteenth Biennial University/government/Industry Microelectronics Symposium*, pp. 158-161, 2001.

[166] I. R. Matias, I. Villar, F. J. Arregui, and R. O. Claus, "Comparative study of modeling of three-dimensional photonic band gap structures," *J. Opt. Soc. Am. A*, vol. 20, pp. 644-654, 2003.

[167] X. Ao and S. He, "Three-dimensional photonic crystal of neative refraction achieved by interference lithography," *Opt. Lett.*, vol. 29, pp. 2542-2544, 2004.

[168] N. D. Lai, W. P. Liang, J. H. Lin, and C. C. Hsu, "Fabrication of two- and three-dimensional periodic structures by multi-exposure of two-beam intereference technique," *Opt. Express*, vol. 13, pp. 9605-9611, 2005.

[169] D. N. Sharp, A. J. Turberfield, and R. G. Denning, "Holographic photonic crystals with diamond symmetry," *Phys. Rev. B*, vol. 68, 2003.

[170] H. M. Su, Y. C. Zhong, X. Wang, X. G. Zheng, J. F. Xu, and H. Z. Wang, "Effects of polarization on laser holography for microstructure fabrication," *Phys. Rev. E*, vol. 67, 2003.

[171] X. Yang, L. Cai, and Q. Liu, "Polarization optimization in the interference of four umbrellalike symmetric beams for making three-dimensional periodic microstructures," *Appl. Opt.*, vol. 41, pp. 6894-6900, 2002.

[172] A. Fernandez and D. W. Phillion, "Effects of phase shifts on four-beam interference patterns," *Appl. Opt.*, vol. 37, pp. 473-478, 1998.

[173] M. Pitchumani, H. Hockel, W. Mohammed, and E. G. Hohnson, "Additive lithography for fabrication of diffractive optics," *Appl. Opt.*, vol. 41, pp. 6176, 2002.

[174] F. H. Dill, "Positive Optical Lithography," *IEEE International Conference on Solid State Circuits*, pp. 54-55, 1975.

[175] F. H. Dill, W. P. Hornberger, P. S. Hauge, and J. M. Shaw, "Characterization of Positive Photoresist," *IEEE Trans. on Elect. Dev.*, vol. ED-22, pp. 445-452, 1975.

[176] F. H. Dill, A. R. Neureuther, J. A. Tuttle, and E. J. Walker, "Modeling projection printing of positive photoresists," *IEEE Trans. Elect. Dev.*, vol. ED-22, pp. 456-464, 1975.

[177] A. Erdmann, W. Henke, S. Robertson, E. Richter, B. Tollkuhn, W. Hoppe, C. A. Mack, and T. Stevenson, "Comparison of simulation approaches for chemically amplified resists," in *Lithography for Semiconductor Manufacturing II*, vol. 4404, 2001, pp. 99-110.

[178] S. Robertson, E. Pavelchek, W. Hoppe, R. Wildfeuer, and F. M. Houlihan, "An Improved Notch Model for Resist Dissolution in Lithography Simulation," in *Advances in Resist Technology and Processing XVIII*, vol. 4345, 2001, pp. 912-920.

[179] A. Erdmann and N. Kachwala, "Enhancements in rigorous simulation of light diffraction from phase shift masks," presented at Optical Microlithography XV, 2002.

[180] A. Erdmann, C. Kalus, T. Schmoller, and A. Wolter, "Efficient simulation of light diffraction from 3-dimensional EUV-masks using field decomposition techniques," presented at Emerging Lithographic Technologies VII.

[181] A. Vial, A. Erdmann, T. Schmoeller, C. Kalus, and H. Kawahira, "Modification of boundary conditions in the FDTD algorithm for EUV masks modeling," in *Photomask and Next-Generation Lithography Mask Technology IX*, vol. 4754, Anonymous, Ed., 2002, pp. 890-899.

[182] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: principles, algorithms, and applications, 3rd edition*. Upper Saddle River, New Jersey: Prentice-Hall, Inc., 1996.

[183] Internet, "www.bandwidthsemi.com," *Bandwidth Semiconductor, LLC*.

[184] Internet, "www.zeonchemicals.com," *Zeon Chemicals*.

[185] Internet, "www.leica-microsystems.com," *Leica Microsystems*.

[186] Internet, "www.transene.com," *Transene Company, Inc.*

[187] G. R. Fowles, *Introduction to Modern Optics*. New York, NY: Dover Publications, 1975.

[188] R. W. Wood, *Proc. Roy. Soc. (London) XVIII*, pp. 396, 1902.

[189] L. Rayleigh, *Phil. Mag.*, vol. 14, pp. 60, 1907.

[190] A. Hessel and A. Oliner, "Wood's anomalies and leaky waves," presented at 1962 Symposium on Electromagnetic Theory and Antennas, Copenhagan, Denmark, 1962.

[191] L. Mashev and E. Popov, "Zero order anomaly of dielectric coated gratings," *Opt. Comm.*, vol. 55, pp. 377-380, 1985.

[192] G. Niederer, M. Salt, H. P. Herzig, T. Overstolz, W. Noell, and N. F. Rooij, "Resonant grating filter for a MEMS based add-drop device at oblique incidence," *IEEE/LEOS International Conference on Optical MEMS*, pp. 99-100, 2002.

[193] J. A. Cox, R. A. Morgan, R. Wilke, and C. Ford, "Guided-mode grating resonant filter for VCSEL applications," *Proc. SPIE in Diffractive and Holographic Device Technologies and Applications V*, vol. 3291, pp. 70-76, 1998.

[194] R. Magnusson and S. S. Wang, "New principle for optical filters," *Appl. Phys. Lett.*, vol. 61, pp. 1022-1024, 1992.

[195] S. M. Norton, "Resonant grating structures: theory, design, and applications." Rochester, NY: University of Rochester, 1997.

[196] M. T. Gale, K. Knop, and R. Morf, "Zero-order diffractive microstructures for security applications," presented at Optical Security and Anticounterfeiting Systems, Proc. SPIE, p. 83-89 1210, 1990.

[197] J. E. Roman and K. A. Winick, "Waveguide Grating Filters for Dispersion Compensation and Pulse Compression," *IEEE J. Quant. Elect.*, vol. 29, pp. 975-982, 1993.

[198] T. Vallius, P. Vahimaa, and J. Turunen, "Pulse deformations at guided-mode resonance filters," *Opt. Exp.*, vol. 10, pp. 840-843, 2002.

[199] S. Tibuleac, R. Magnusson, T. A. Maldonado, P. P. Young, and T. R. Holzheimer, "Dielectric Frequency-Selective Structures Incorporating Waveguide Gratings," *IEEE Trans. on Microwave Theory and Techniques*, vol. 48, pp. 553-561, 2000.

[200] S. S. Wang and R. Magnusson, "Theory and applications of guided-mode resonance filters," *Appl. Opt.*, vol. 32, pp. 2606-2613, 1993.

[201] S. S. Wang, R. Magnusson, J. S. Bagby, and M. G. Moharam, "Guided-mode resonances in planar dielectric-layer diffraction gratings," *J. Opt. Soc. Am. A*, vol. 7, pp. 1470-1474, 1990.

[202] C. Zuffada, T. Cwik, and C. Ditchman, "Synthesis of novel all-dielectric grating filters using genetic algorithms," presented at Antennas and Propagation Society International Symposium 1997, Montreal, Cananda, 1997.

[203] R. R. Boye and R. K. Kostuk, "Investigation of the effect of finite grating size on the performance of guided-mode resonance filters," *Appl. Opt.*, vol. 39, pp. 3649-3653, 2000.

[204] A. Fehrembach, D. Maystre, and A. Sentenac, "Phenomenological theory of filtering by resonant dielectric gratings," *J. Opt. Soc. Am. A*, vol. 19, pp. 1136-1144, 2002.

[205] A. Fehrembach and A. Sentenac, "Study of waveguide grating eigenmodes for unpolarized filtering applications," *J. Opt. Soc. Am. A*, vol. 20, pp. 481-488, 2003.

[206] A. Mizutani, H. Kikuta, K. Nakajima, and K. Iwata, "Nonpolarizing guided-mode resonant grating filter for oblique incidence," *J. Opt. Soc. Am. A*, vol. 18, pp. 1261-1266, 2001.

[207] S. Peng and G. M. Morris, "Experimental demonstration of resonant anomalies in diffraction from two-dimensional gratings," *Opt. Lett.*, vol. 21, pp. 549-551, 1996.

[208] S. T. Peng and G. M. Morris, "Resonant scattering from two-dimensional gratings," *J. Opt. Soc. Am. A*, vol. 13, pp. 993-1005, 1996.

[209] D. Lacour, P. Plumey, G. Granet, and A. M. Ravaud, "Resonant waveguide grating: Analysis of polarization independent filtering," *Optcal and Quantum Electronics*, vol. 33, pp. 451-470, 2001.

[210] R. E. Collin, *Field Theory of Guided Waves*, Second ed. New York: IEEE Press, 1991.

[211] K. Okamoto, *Fundamentals of Optical Waveguides*. New York: Academic Press, 2000.

[212] Y. Ding and R. Magnusson, "Doubly resonant single-layer bandpass optical filters," *Opt. Lett.*, vol. 29, pp. 1135-1137, 2004.

[213] C. Zuffada, T. Cwik, and C. Ditchman, "Synthesis of Novel All-Dielectric Grating Filters Using Genetic Algorithms," *IEEE Trans. Antennas and Propagation*, vol. 46, pp. 657-663, 1998.

[214] F. Lemarchand, A. Sentenac, and H. Giovannini, "Increasing the angular tolerance of resonant grating filters with doubly periodic structures," *Opt. Lett.*, vol. 23, pp. 1149-1151, 1998.

[215] W. H. Wong, E. Y. B. Pun, and K. S. Chan, "Electron Beam Direct-Write Tunable Polymeric Waveguide Grating Filter," *IEEE Phot. Tech. Lett.*, vol. 15, pp. 1731-1733, 2003.

[216] V. Soifer, V. Kotlyar, and L. Doskolovich, *Iterative Methods for Diffractive Optical Elements Computation*. Bristol, PA: Taylor and Francis, Inc., 1997.

[217] F. Xu, J. Ford, and Y. Fainman, "Polarization-selective computer-generated holograms: design, fabrication, and applications," *Appl. Opt.*, vol. 34, pp. 256-266, 1995.

[218] F. Xu, R. Tyan, P. Sun, Y. Fainman, C. Cheng, and A. Scherer, "Form-birefringent computer-generated holograms," *Opt. Lett.*, vol. 21, pp. 1513-1515, 1996.

[219] E. Hasman, V. Kleiner, G. Biener, and A. Niv, "space-variant polarization-state manipulation with computer-generated subwavelength gratings," presented at Micromachining Technology for Micro-optics and Nano-optics, San Jose, CA.

[220] H. Hertz, *Electric Waves*: Macmillan and Company, Ltd., 1893.

[221] M. Mirotznik, D. Pustai, D. Prather, and J. Mait, "Design of two-dimensional polarization-selective diffractive optical elements with form-birefringent microstructures," *Appl. Opt.*, vol. 43, pp. 5947-5954, 2004.

[222] J. N. Mait, A. Scherer, O. Dial, D. W. Prather, and X. Gao, "Diffractive lens fabricated with binary features less than 60 nm," *Opt. Lett.*, vol. 25, pp. 381-383, 2000.

[223] S. Quabis, R. Dron, M. Eberler, O. Glockl, and G. Leuchs, "Focusing light to a tighter spot," *Opt. Commun.*, vol. 179, pp. 1-7, 2000.

[224] G. Biener, A. Niv, V. Kleiner, and E. Hasman, "Near-field Fourier transform polarimetry by use of a discrete space-variant subwavelength grating," *J. Opt. Soc. Am. A*, vol. 20, pp. 1940-1948, 2003.

[225] D. Mawet, J. Baudrand, C. Lenaerts, V. Moreau, P. Riaud, D. Rouan, and J. Surdej, "Birefringent achromatic phase shifters for nulling interferometry and phase coronagraphy," presented at Towards Other Earths: DARWIN/TPF and the Search for Extrasolar Terrestrial Planets, Heidelberg, Germany, 2003.

[226] Z. Bomzon, G. Biener, V. Kleiner, and E. Hasman, "Radially and azimuthally polarized beams generated by space-variant dielectric subwavelength gratings," *Opt. Lett.*, vol. 27, pp. 285-287, 2002.

[227] Z. Bomzon, V. Kleiner, and E. Hasman, "Computer-generated space-variant polarization elements with subwavelength metal stripes," *Opt. Lett.*, vol. 26, pp. 33-35, 2001.

[228] U. Levy, C. Tsai, L. Pang, and Y. Fainman, "Engineering space-variant inhomogeneous media for polarization control," *Opt. Lett.*, vol. 29, pp. 1718-1720, 2004.

[229] A. Niv, G. Biener, V. Kleiner, and E. Hasman, "Formation of linearly polarized light with axial symmetry by use of space-variant subwavelength gratings," *Opt. Lett.*, vol. 28, pp. 510-512, 2003.

[230] C. Tsai, U. Levy, L. Pang, and Y. Fainman, "Form-birefringent space-variant inhomogeneous medium for shaping point-spread function," *Appl. Opt.*, vol. 45, pp. 1777-1784, 2006.

[231] S. Babin, H. Haidner, P. Kipfer, A. Lang, J. T. Sheridan, W. Stork, and N. Streibl, "Artificial index surface relief diffraction optical elements," presented at Photo-Opt. Instrum. Eng., 1992.

[232] F. T. Chen and H. G. Craighead, "Diffractive phase elements based on two-dimensional artificial dielectrics," *Opt. Lett.*, vol. 20, pp. 121-123, 1995.

[233] P. Lalanne, S. Astilean, and P. Chavel, "Blazed binary subwavelength gratings with efficiencies larger than those of conventional echelette gratings," *Opt. Lett.*, vol. 23, pp. 1081-1083, 1998.

[234] S. Aoyama and T. Yamashita, "Grating beam splitting polarizer using multi-layer resist method," *Proc. Soc. Photo-Opt. Instrum. Eng.*, vol. 1545, pp. 241-250, 1991.

[235] Z. Bomzon, G. Biener, V. Kleiner, and E. Hasman, "Space-variant Pancharatnam-Berry phase optical elements with computed generated subwavelength gratings," *Opt. Lett.*, vol. 27, pp. 1141-1143, 2002.

[236] K. Shiraishi, T. Sato, and S. Kawakami, "Experimental verification of a form-birefringent polarization splitter," *Appl. Phys. Lett.*, vol. 58, pp. 211-212, 1991.

[237] R. Tyan, A. A. Salvekar, H. Chou, C. Cheng, and A. Scherer, "Design, fabrication, and characterization of form-birefringent multilayer polarizing beam splitter," *J. Opt. Soc. Am. A*, vol. 14, pp. 1627-1636, 1997.

[238] R. Tyan, P. Sun, and Y. Fainman, "Polarizing beam splitters constructed of form-birefringent multilayer gratings," presented at Diffractive and Holographic Optics Technology III, 1995.

[239] M. Honkanen, V. Kettunen, J. Tervo, and J. Turunen, "Fourier array illuminators with 100% efficiency: analytical Jones-matrix construction," *J. Mod. Opt.*, vol. 47, pp. 2351-2359, 2000.

[240] U. Levy, H. C. Kim, C. H. Tsai, and Y. Fainman, "Near-infrared demonstration of computer-generated holograms implemented by using subwavelength gratings with space-variant orientation," *Opt. Lett.*, vol. 30, pp. 2089-2091, 2005.

[241] U. Levy, C. H. Tsai, H. C. Kim, and Y. Fainman, "Design, fabrication and characterization of subwavelength computer-generated holograms for spot array generation," *Opt. Exp.*, vol. 12, pp. 5345-5355, 2004.

[242] J. Tervo and J. Turunen, "Paraxial-domain diffractive elements with 100% efficiency based on polarization gratings," *Opt. Lett.*, vol. 25, pp. 785-786, 2000.

[243] B. O, R. Liu, Y. Li, M. Sailor, and Y. Fainman, "Vapor Sensor Realized in an Ultracompact Polarization Interferometer Built of a Freestanding Porous-Silicon Form Birefringent Film," *Phot. Tech. Lett.*, vol. 15, pp. 834-836, 2003.

[244] N. Nieuborg, A. Kirk, B. Morlion, H. Thienpont, and I. Veretennicoff, "Polarization-selectrive diffractive optical elements with an index-matching gap material," *Appl. Opt.*, vol. 36, pp. 4681-4686, 1997.

[245] N. Nieuborg, A. Kirk, B. Morlion, H. Thienpont, and I. Veretennicoff, "Highly polarization-selective diffractive optical elements in calcite with an index-matching gap material," presented at SPIE Diffractive and Holgoraphic Device Technologies and Applications IV, 1997.

[246] E. B. Grann, M. G. Moharam, and D. A. Pommet, "Artificial uniaxial and biaxial dielectrics with use of two-dimensional subwavelength binary gratings," *J. Opt. Soc. Am. A*, vol. 11, pp. 2695-2703, 1994.

320

[247] D. Mawet, V. Moreau, D. Rouan, and J. Surdej, "Achromatic Four Quadrants Phase Mask using the Dispersion of form Birefringence," presented at High-Contrast Imaging for Exo-Planet Detection, 2002.

[248] C. W. Haggans and L. Li, "Effective-medium theory of zeroth-order lamellar gratings in conical mountings," *J. Opt. Soc. Am. A*, vol. 10, pp. 2217-2225, 1993.

[249] H. Kikuta, T. Ohira, H. Kubo, and K. Iwata, "Effective medium theory of two-dimensional subwavelength gratings in the non-quasi-static limit," *J. Opt. Soc. Am. A*, vol. 15, pp. 1577-1585, 1998.

[250] P. Lalanne and J. P. Hugonin, "High-order effective-medium theory of subwavelength gratings in classical mounting: application to volume holograms," *J. Opt. Soc. Am. A*, vol. 15, pp. 1843-1851, 1998.

[251] P. Lalanne and D. Lemercier-Lalanne, "Depth dependance of the effective properties of subwavelength gratings," *J. Opt. Soc. Am. A*, vol. 14, pp. 450-458, 1997.

[252] D. E. Aspnes, "Local-field effects and effective-medium theory: A microscopic perspective," *Am. J. Phys.*, vol. 50, pp. 704-709, 1982.

[253] A. J. Jaaskelainen, K. E. Peiponen, J. Raty, U. Tapper, O. Richard, E. I. Kauppinen, and K. Lumme, "Estimation of the refractive index of plastic pigments by Wiener bounds," *Opt. Eng.*, vol. 39, pp. 2959-2963, 2000.

[254] L. H. Cescato, E. Gluch, and N. Streibl, "Holographic quarterwave plates," *Appl. Opt.*, vol. 29, pp. 3286, 1990.

[255] G. P. Nordin and P. C. Deguzman, "Broadband form birefringent quarter-wave plate for the mid-infrared wavelength region," *Opt. Exp.*, vol. 5, pp. 163-168, 1999.

[256] I. Richter, P. C. Sun, F. Xu, and Y. Fainman, "Design considerations of form birefringent microstructures," *Appl. Opt.*, vol. 34, pp. 2421-2429.

[257] J. B. Pendry, A. J. Holden, W. J. Stewart, and I. Youngs, "Extremely Low Frequency Plasmons in Metallic Mesostructures," *Phys. Rev. Lett.*, vol. 76, pp. 4773-4776, 1996.

[258]   G. A. Niklasson, C. G. Granqvist, and O. Hunderi, "Effective medium models for the optical properties of inhomogeneous materials," *Appl. Opt.*, vol. 20, pp. 26-30, 1981.

[259]   R. Luo, "Effective medium theories for the optical properties of three-component composite materials," *Appl. Opt.*, vol. 36, pp. 8153-8158, 1997.

[260]   B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*. New York: John Wiley & Sons, Inc., 1991.

[261]   A. Tal, S. Chen, and S. M. Kuebler, "personal communications." Orlando, 2006.

[262]   E. G. Johnson and M. A. G. Abushagar, "Micro-genetic algorithm optimization methods applied to dielectric gratings," *J. Opt. Soc. Am. A*, vol. 12, pp. 1152-1160, 1995.

[263]   E. G. Johnson, M. A. G. Abushagar, and A. Kathman, "Phase grating optimization using genetic algorithms," *OSA Digest of Optical Design for Photonics*, vol. 9, pp. 71-72, 1993.

[264]   N. Yoshikawa, M. Itoh, and T. Yatagai, "Qauntized phase optimization of two-dimensional Fourier kinoforms by a genetic algorithm," *Opt. Lett.*, vol. 20, 1995.

[265]   J. Robinson and Y. Rahmat-Samii, "Particle swarm optimization in electromagnetics," *IEEE Trans. Antennas and Propagation*, vol. 52, pp. 397-407, 2004.