IMPROVED-ACCURACY ALGORITHMS FOR
TIME-DOMAIN FINITE METHODS IN
ELECTROMAGNETICS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the
Graduate School of The Ohio State University

By

Shumin Wang, B.S., M.S.

* * * * *

The Ohio State University

2003

Dissertation Committee:

Professor Fernando L. Teixeira, Co-
Adviser
Professor Robert Lee, Co-Adviser
Professor Jin-Fa Lee

Approved by

Co-Adviser

Department of Electrical
Engineering
Two novel finite-difference time-domain (FDTD) algorithms are proposed to reduce numerical dispersion error. One is to minimize the dispersion error at arbitrary angles. The other is to minimize the maximum dispersion error for all angles. Generic filtering schemes are further developed to improve the performances in broadband simulations. To deal with very large-scale, locally fine problems, the accuracy and stability of FDTD subgridding schemes are numerical studied. A systematic approach is developed to optimize the spatial interpolation coefficients. Unstructured/structured hybrid mesh is further examined as an alternative approach. Composite elements and implicit FDTD method are proposed to facilitate mesh generation and hybridize with Finite-Element Time-Domain (FETD) methods respectively. The stability of FETD/FDTD hybrid methods and the effectiveness of using FDTD as a predictor for FETD, are also examined. Finally, we develop two novel Perfectly Matched Layers (PML) implementations for Alternating-Direction Implicit (ADI) and FETD methods respectively, which exhibit largely reduced reflection errors. In addition, the PML for FETD shows stable late-time behavior in our numerical tests.
Dedicated to my parents and my loving wife.
I wish to express my sincere gratitude to my advisors, true scholars and mentors, Professor Robert Lee and Professor Fernando L. Teixeira. With their patience and invaluable advices, this dissertation finally becomes in shape. I thank Professor Jin-Fa Lee for serving as a committee member and his helpful comments. Special thanks to Professor Ben A. Munk for his physical insights. His diligence always inspires me. Many thanks to everyone supporting and encouraging me during my dissertation work.
VITA

September 9, 1972 ........................................... Born - Qingdao, China

1995 ......................................................... B.S. Physics, Qingdao University,
Qingdao, China.

1998 ......................................................... M.S. Electronics, Beijing University,
Beijing, China.

1998-present .............................................. Graduate Research Associate,
The Ohio State University.

PUBLICATIONS

Research Publications

Refereed Journal Publications

Shumin Wang, Mingzhi Li and Changqing Wang “GRE-FDTD Hybrid Method for
the Computation of Electromagnetic Scattering from 2-D Arbitrarily Shaped Cavities
with Rectangular Terminations”, Acta Electronica Sinica, vol. 28, No. 6, pp. 138-

Shumin Wang, Fernando L. Teixeira, Robert Lee, and Jin-Fa Lee “Optimization
of Sub-gridding Schemes for FDTD”, IEEE Microwave and Wireless Components

Shumin Wang and Fernando L. Teixeira “An Equivalent Electric Field Source for
Wideband FDTD Simulations of Waveguide Discontinuities”, IEEE Microwave and


Conference Publications


**FIELDS OF STUDY**

Major Field: Electrical Engineering

Studies in:

  Computational Electromagnetics  Prof. Fernando L. Teixeira, Prof. Robert Lee
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td></td>
<td>Dedication</td>
<td>iii</td>
</tr>
<tr>
<td></td>
<td>Acknowledgments</td>
<td>iv</td>
</tr>
<tr>
<td></td>
<td>Vita</td>
<td>v</td>
</tr>
<tr>
<td></td>
<td>List of Tables</td>
<td>xi</td>
</tr>
<tr>
<td></td>
<td>List of Figures</td>
<td>xii</td>
</tr>
</tbody>
</table>

**Chapters:**

1. Introduction ........................................... 1

2. Angle-Optimized FDTD Algorithms .......................... 4

   2.1 Introduction ........................................... 4

   2.2 Two-Dimensional AO-FDTD Algorithm .................... 5

      2.2.1 A Conditionally Stable 2-D AO-FDTD Scheme ....... 5
      2.2.2 Numerical Stability Analysis ....................... 7
      2.2.3 Optimal Angle Selection ............................ 8
      2.2.4 Optimal Frequency Band Selection .................. 10
      2.2.5 Angular Behavior of Numerical Dispersion ......... 13
      2.2.6 Numerical Results ................................. 16

   2.3 Three-Dimensional AO-FDTD Algorithm ................... 19

      2.3.1 Update Equations and Stability Analysis .......... 19
      2.3.2 Optimal Angle Selection ............................ 25
      2.3.3 Numerical Experiments ............................. 27

   2.4 Concluding Remarks ..................................... 32
### 3. Dispersion-Relation-Preservation FDTD Schemes

#### 3.1 Introduction

#### 3.2 Two-Dimensional DPR-FDTD Algorithm

- 3.2.1 Methodology
- 3.2.2 Numerical Results

#### 3.3 Three-Dimensional DPR-FDTD Algorithm

- 3.3.1 Methodology
- 3.3.2 Numerical Comparisons
- 3.3.3 Truncation Error

#### 3.4 Concluding Remarks

### 4. FDTD Subgridding Method

#### 4.1 Introduction

#### 4.2 Accuracy

- 4.2.1 Three-point coupling scheme
- 4.2.2 Higher order, five-point coupling scheme

#### 4.3 Stability

- 4.3.1 STS Subgridding Scheme
- 4.3.2 MTS Subgridding Scheme

#### 4.4 Concluding Remarks

### 5. Hybrid Mesh and FETD/INS-FDTD Hybrid Method

#### 5.1 Introduction

#### 5.2 Composite Elements

- 5.2.1 Edge Element Basis Functions for Pyramidal Elements
- 5.2.2 Encapsulation

#### 5.3 FETD/FDTD hybrid methods

#### 5.4 Implicit Non-Staggered FDTD Method

- 5.4.1 Formulation
- 5.4.2 Numerical Dispersion
- 5.4.3 A Different Scheme

#### 5.5 FDTD as a Predictor for FETD

- 5.5.1 Source Prediction
- 5.5.2 Tetrahedral Regions

#### 5.6 Concluding Remarks
6. Time-Domain Perfectly Matched Layer Implementations .... 130
   6.1 Introduction ............................................ 130
   6.2 PML Implementation for ADI-FDTD ....................... 131
      6.2.1 ADI-FDTD Update Equations with PML ............... 131
      6.2.2 Traditional Scheme ............................... 133
      6.2.3 Exponential Differencing Scheme .................... 136
      6.2.4 Proposed Approach ................................ 140
   6.3 Direct PML Implementation for FETD ....................... 145
   6.4 PML Implementation for FETD with Conjugate-Scaled Test and
      Basis Functions ........................................ 150
      6.4.1 Formulation ....................................... 150
      6.4.2 Numerical Results .................................. 157
      6.4.3 Application to Implicit Non-Staggered FDTD Method ... 167
   6.5 Concluding Remarks ..................................... 171

7. Conclusions and Future Work ............................... 173

Bibliography .................................................. 176
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Some coefficients used in 2-D AO-FDTD.</td>
<td>16</td>
</tr>
<tr>
<td>2.2</td>
<td>Some coefficients used in 3-D AO-FDTD.</td>
<td>28</td>
</tr>
<tr>
<td>2.3</td>
<td>Coefficients used in the AO-FDTD with Chebyshev filtering scheme. ( \Delta q = 0.02 ) for all Chebyshev filters.</td>
<td>28</td>
</tr>
<tr>
<td>3.1</td>
<td>Coefficients used in different 2-D (2,4) schemes.</td>
<td>50</td>
</tr>
<tr>
<td>3.2</td>
<td>Coefficients used in different 2-D (2,4) schemes (cont’d).</td>
<td>51</td>
</tr>
<tr>
<td>3.3</td>
<td>Coefficients of Chebyshev filtering schemes with different parameters ( q_c ) and ( \Delta q ).</td>
<td>53</td>
</tr>
<tr>
<td>3.4</td>
<td>Coefficients of Chebyshev filtering schemes with different parameters ( q_c ) and ( \Delta q )(continued).</td>
<td>53</td>
</tr>
<tr>
<td>3.5</td>
<td>Coefficients used in different 3-D (2,4) schemes.</td>
<td>64</td>
</tr>
<tr>
<td>3.6</td>
<td>Coefficients used in different 3-D (2,4) schemes (cont’d).</td>
<td>64</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Isotropic term of the numerical dispersion error $\delta$</td>
<td>14</td>
</tr>
<tr>
<td>2.2</td>
<td>$\cos(4\theta)$ term of the numerical dispersion error $\delta$</td>
<td>15</td>
</tr>
<tr>
<td>2.3</td>
<td>Normalized phase velocity of AO-FDTD scheme optimized at $15^\circ$ using Butterworth filter</td>
<td>17</td>
</tr>
<tr>
<td>2.4</td>
<td>Normalized phase velocity of AO-FDTD scheme optimized at $0^\circ$ using Butterworth filter</td>
<td>18</td>
</tr>
<tr>
<td>2.5</td>
<td>Normalized phase velocity of ordinary FDTD method</td>
<td>19</td>
</tr>
<tr>
<td>2.6</td>
<td>Normalized phase velocity of (2,4) FDTD scheme</td>
<td>20</td>
</tr>
<tr>
<td>2.7</td>
<td>Normalized phase velocity of ordinary FDTD method in a finer mesh.</td>
<td>21</td>
</tr>
<tr>
<td>2.8</td>
<td>Normalized phase velocity of (2,4) FDTD scheme in a finer mesh.</td>
<td>22</td>
</tr>
<tr>
<td>2.9</td>
<td>Normalized phase velocity of AO-FDTD scheme optimized at $22.5^\circ$ using non-filtered scheme</td>
<td>23</td>
</tr>
<tr>
<td>2.10</td>
<td>Comparison of Butterworth, Chebyshev and non-filtered AO-FDTD schemes at $15^\circ$.</td>
<td>24</td>
</tr>
<tr>
<td>2.11</td>
<td>Relative error of normalized phase velocity around $0^\circ$ for the AO-FDTD using Chebyshev filter</td>
<td>25</td>
</tr>
<tr>
<td>2.12</td>
<td>Normalized phase velocity of Butterworth AO-FDTD scheme at $\theta = 90^\circ$ and $\phi = 0^\circ$ with different specification of center frequencies</td>
<td>29</td>
</tr>
</tbody>
</table>
2.13 Normalized phase velocity of Chebyshev AO-FDTD scheme at $\theta = 90^\circ$ and $\phi = 0^\circ$ with different specification of center frequencies .......................... 30

2.14 Normalized phase velocity of Butterworth, Chebyshev and non-filtered AO-FDTD schemes at $\theta = 90^\circ$ and $\phi = 0^\circ$. For the filtered schemes, $q_c = 0.09$ .......................... 31

2.15 Normalized phase velocity of ordinary FDTD and (2,4) scheme at $\theta = 90^\circ$ and $\phi = 0^\circ$. ........................................... 32

2.16 Comparison of the normalized phase velocity at different angles of Butterworth AO-FDTD scheme. The center frequency is such that $q_c = 0.09$ 33

2.17 Simulated results of a first-order differentiated Gaussian pulse propagating at $\theta = 90^\circ$ and $\phi = 0^\circ$ ........................................... 33

3.1 Comparison of the maximum value of the dispersion error $\delta_2(\Gamma_x, \Gamma_y, \theta)$ for all angles when using analytical solution, $(\Gamma_x, \Gamma_y) = (\gamma_1(\infty), \gamma_2(\infty))$, second order approximation $(\Gamma_x, \Gamma_y) = (\gamma_1^{(2)}, \gamma_2^{(2)})$ and Yee’s scheme $(\Gamma_x, \Gamma_y) = (1, 1)$ .......................... 41

3.2 Comparison of the largest $\delta_4$ among all angles by using different DRP schemes for the fourth-order stencil. See text for details. ............... 48

3.3 Comparison of the maximum (for all angles) phase error per wavelength using Yee’s, a traditional (2,4) scheme and the non-filtered DRP (2,2) scheme. ........................................... 51

3.4 Comparison of the maximum (for all angles) phase error per wavelength using different DRP (2,4) schemes and the Deveze (4,4) scheme. .... 52

3.5 Comparison of the maximum (for all angles) accumulated phase error over the largest wavelength in DRP schemes using Chebyshev filtering schemes with various $\Delta q$ while fixing $q_c = 0.1$. .......................... 54

3.6 Comparison of the maximum (for all angles) local phase error per wavelength using DRP schemes and Chebyhev filters with different parameters. ........................................... 55

3.7 Comparison of the maximum (for all angles) phase error per wavelength using Yee’s and the traditional (2,4) scheme. .......................... 65
4.7 Timing procedure of MTS subgridding scheme, superscripts denote time steps .................................. 90

4.8 System amplification matrix of MTS subgridding scheme. .............. 91

4.9 Dominant eigenvalues of MTS subgridding scheme. .................. 92

4.10 Bursting points of different computation domain sizes and CFL numbers. 92

5.1 A Tetrahedral Element ........................................... 97

5.2 A Pyramidal Element ............................................ 97

5.3 Type 0 composite element ...................................... 100

5.4 Type 1 composite element ...................................... 100

5.5 Type 2 composite element ...................................... 101

5.6 Type 3 composite element ...................................... 101

5.7 Type 4 composite element ...................................... 102

5.8 Type 5 composite element ...................................... 102

5.9 Two interface arrangements .................................... 105

5.10 Stability test of the NSCC approach ............................. 110

5.11 Numerical dispersion errors of different schemes evaluated at $\chi = 1.0$ 115

5.12 Anisotropy of numerical dispersion errors of INS-FDTD scheme evaluated at $\chi = 1.0$ ........................................ 115

5.13 Anisotropy of numerical dispersion errors of INS-FDTD and Yee’s FDTD evaluated at $\chi = 1.0$ and $\theta = 45^\circ$ .................. 116

5.14 Numerical dispersion errors of INS-FDTD scheme evaluated at different $\chi$ values .............................................. 117
5.15 Numerical dispersion errors of FETD with cube elements evaluated at
different $\chi$ values ........................................ 117

5.16 Numerical dispersion of the different scheme. .......................... 119

5.17 Observed electrical field in an empty waveguide by using FETD method
and the different scheme with different $\beta/\alpha$ values. .............. 119

5.18 Observed electrical fields in an empty waveguide. ................. 121

5.19 Initial residuals of different initial guesses. ......................... 122

5.20 Number of iterations required by different initial guesses. ....... 122

5.21 Initial residuals when cubes and tetrahedra co-exist. .............. 125

5.22 Required number of iterations when cubes and tetrahedra co-exist. 125

5.23 Initial residual when the source is in the tetrahedral region. ....... 126

5.24 Number of iterations when the source is in the tetrahedral region. 127

5.25 Initial residual when the source is in the cube region. ............. 128

5.26 Number of iterations when the source is in the cube region. ....... 129

6.1 Reflection errors from the PML-ADI-FDTD method with the usual
PML implementation (I). ........................................ 135

6.2 Reflection errors from the PML-ADI-FDTD method with the usual
PML implementation (II). ........................................ 136

6.3 Reflection errors from the PML-ADI-FDTD method with exponential
time differencing (I). ........................................... 140

6.4 Reflection errors from the PML-ADI-FDTD method with exponential
time differencing (II). ........................................... 141

6.5 Comparison of reflection errors from the PML-ADI-FDTD method
with exponential time differencing and the traditional approach (I). 141
6.6 Comparison of reflection errors from the PML-ADI-FDTD method with exponential time differencing and the traditional approach (II). 142

6.7 Reflection error from the PML-ADI-FDTD method with the proposed PML implementation (I). 144

6.8 Reflection error from the PML-ADI-FDTD method with the proposed PML implementation (II). 144

6.9 Late-time instability of direct implementation of PML for FETD. 150

6.10 PML region consisted of cubes. 152

6.11 Group tetrahedrons into a cube in the PML region. 152

6.12 Reflections from a single interface when the theoretical reflection error is -20dB at normal incidence. 158

6.13 Reflections from a single interface when the theoretical reflection error is -40dB at normal incidence. 159

6.14 Reflections from a single interface when the theoretical reflection error is -60dB at normal incidence. 159

6.15 Reflections from a single interface when the theoretical reflection error is -80dB at normal incidence. 160

6.16 Comparison of the observations. 161

6.17 Reflection errors in time domain. 162

6.18 Reflection errors for ten layers of PML with the fourth order conductivity profile. 162

6.19 Reflection errors for 20 layers of PML with the fourth order conductivity profile. 163

6.20 Late-time performance of the proposed PML implementation: empty waveguide. 164

6.21 PEC structure of the loaded waveguide with stratified dielectric medium. 165
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.22</td>
<td>Late-time performance of the proposed PML implementation: loaded</td>
<td>165</td>
</tr>
<tr>
<td></td>
<td>structure.</td>
<td></td>
</tr>
<tr>
<td>6.23</td>
<td>Condition numbers associated with the PML in the FETD.</td>
<td>166</td>
</tr>
<tr>
<td>6.24</td>
<td>Condition numbers associated with the PML in the FETD.</td>
<td>167</td>
</tr>
<tr>
<td>6.25</td>
<td>Time domain reflection errors: $</td>
<td>\Gamma</td>
</tr>
<tr>
<td>6.26</td>
<td>Time domain reflection errors: $</td>
<td>\Gamma</td>
</tr>
<tr>
<td>6.27</td>
<td>Time domain reflection errors: $</td>
<td>\Gamma</td>
</tr>
<tr>
<td>6.28</td>
<td>Time domain reflection errors: $</td>
<td>\Gamma</td>
</tr>
<tr>
<td>6.29</td>
<td>Late-time performance of the proposed PML for the implicit FDTD.</td>
<td>170</td>
</tr>
<tr>
<td>6.30</td>
<td>Condition numbers associated with the PML in the INS-FDTD.</td>
<td>171</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

Macro-scope electromagnetic (EM) phonomania are governed by Maxwell’s equations. Computational Electromagnetics (CEM) seeks to solve complex electromagnetic problems, otherwise intractable by analytical approaches, via numerical simulations. CEM could be generally categorized into time-domain and frequency-domain methods. Time-domain methods solve problems in time domain directly and frequency-domain results of interest can then be obtained by a Fourier Transformation or eigendecomposition methods subsequently. Compared to frequency-domain methods, time-domain methods have the following major advantages. First of all, some problems are of intrinsic time-domain nature, e.g. transient problems. Second, broadband response can be obtained in a single time-domain run.

This dissertation is devoted to novel algorithm developments for time-domain CEM. Two types of problem are being focused, i.e. electrically large problems and very large-scale problems with locally fine structures. Also, we shall improve the Perfectly Matched Layers (PML) performances for different time-domain schemes.

Electrically large problems typically present numerical dispersion issues for differential-equation-based methods. In Chapter 2 and Chapter 3, we propose two novel schemes, the Angle-Optimized FDTD (AO-FDTD) and the Dispersion Relation Preservation 1
FDTD (DRP-FDTD), to reduce dispersion errors. The AO-FDTD is designed to reduce dispersion error in an arbitrarily specified angular sector in a wide frequency band. It is specially useful for problems with strong angle preference, like elongated problems, printed circuit board simulations, or some indoor/outdoor propagation channel problems. In contrast, the DRP-FDTD is designed to reduce dispersion errors for all angles in a wide frequency band, and therefore suitable in more general cases.

It is common for fine structures to exist locally in a very large-scale problem. A popular approach in this case is to use FDTD subgridding method. In Chapter 4, we propose a systematic way of optimizing the spatial interpolation coefficients used in subgridding schemes and numerically study the stability issue.

In Chapter 5, we tackle the same problem by using unstructured grids, i.e. tetrahedrons, to model fine structures and structured grids, i.e. bricks, elsewhere. To combine tetrahedrons and pyramids, we propose cube-shaped composite elements, which are consist of tetrahedrons and pyramids. Therefore, the mesh generation process is simplified and structured and unstructured regions can be connected smoothly. With this hybrid mesh, hybrid FDTD/FETD method becomes a natural choice. A common problem of existing hybrid FDTD/FETD methods is observed instability. To deal with the stability issue, we proposed an implicit FDTD method, which is based on finite-difference concepts but developed in the same way as finite-element time-domain methods. This method is developed specifically to work with FETD methods. This implicit FDTD/FETD hybrid method is very useful substitute for FETD method with brick elements. Furthermore, we discuss using FDTD as a predictor for FETD.
In Chapter 6, Perfectly Matched Layers (PML) boundary truncation methods for two time-domain schemes are discussed. We first turn our attention to ADI methods. The traditional implementation suggests huge reflection errors for large Courant numbers. We propose a novel implementation that uses either forward or backward differencing instead of central differencing for conduction terms. The resulting PML performance remain at the same level for either large or small Courant numbers. We next propose a new implementation of PML for FETD. Old implementations have large reflection errors and/or observed late-time instability. By applying conjugate-scaled test and basis functions in PML regions, a late-time stable implementation with small reflection error is obtained.
CHAPTER 2

ANGLE-OPTIMIZED FDTD ALGORITHMS

2.1 Introduction

A major source of error in the Finite-Difference Time-Domain (FDTD) method [75] is numerical (grid) dispersion, which manifests itself as a change on the phase velocity according to frequency and propagation angle. For ordinary FDTD (Yee’s scheme), the numerical dispersion is minimal at 45° and largest at 0° [63]. To reduce the numerical dispersion, it is customary to decrease the cell size, or to use higher order schemes, or to employ some filtering schemes [45]. All these approaches are designed to improve the dispersion properties at all angles and cause significantly increase in the computational burden. For some problems however, the reduction of the dispersion error is really necessary only around a limited angular sector. This is especially true, for instance, in (electrically large) problems involving highly elongated FDTD domains [18] and/or involving waves propagation mostly limited to a certain pre-assigned angular span, e.g., high directivity antennas or some wireless propagation scenarios.
In this chapter, we present a less costly alternative to reduce dispersion errors in such problems by controlling the angular sector of minimum dispersion error. Numerical results demonstrate that the dispersion error around any preassigned angle and central frequency can be reduced significantly with little computational overhead.

This chapter is organized as follows. Section 2.2 introduces the basic scheme of the proposed angle-optimized FDTD (AO-FDTD) scheme for 2-D case. Section 2.3 further extend this algorithm to 3-D simulations.

2.2 Two-Dimensional AO-FDTD Algorithm

2.2.1 A Conditionally Stable 2-D AO-FDTD Scheme

For a 2-D $TE_z$ wave in the frequency domain, we can write the semi-discrete equations for $E_x$, $E_y$, and $H_z$ in a staggered spatial grid as

$$-j\omega E_x^{m+\frac{1}{2},n} = \frac{\alpha + \beta \omega^2 \Delta y^2}{\epsilon \Delta y}(H_z^{m+\frac{1}{2},n+\frac{1}{2}} - H_z^{m+\frac{1}{2},n-\frac{1}{2}})$$  \hspace{1cm} (2.1)

$$-j\omega E_y^{m,n+\frac{1}{2}} = \frac{\alpha + \beta \omega^2 \Delta x^2}{\epsilon \Delta x}(H_z^{m-\frac{1}{2},n+\frac{1}{2}} - H_z^{m+\frac{1}{2},n+\frac{1}{2}})$$  \hspace{1cm} (2.2)

$$-j\omega H_z^{m+\frac{1}{2},n+\frac{1}{2}} = \frac{\alpha + \beta \omega^2 \Delta y^2}{\epsilon \Delta y}(E_x^{m+\frac{1}{2},n+1} - E_x^{m+\frac{1}{2},n}) + \frac{\alpha + \beta \omega^2 \Delta x^2}{\mu \Delta x}(E_y^{m,n+\frac{1}{2}} - E_y^{m+1,n+\frac{1}{2}})$$  \hspace{1cm} (2.3)

where the superscripts indicate the locations of the corresponding field component in the staggered grid, $\alpha$ and $\beta$ are additional degrees of freedom to be determined. These factors control phase correcting (higher-order) time derivative terms in the equations (in the form of $\omega^2$ terms). Such artificial terms are designed so as to introduce artificial dispersion effects which compensate for the numerical dispersion effects. The compensation mechanism will be detailed later. Transforming Eqs. (2.1)–(2.3) to the time domain and using the Helmholtz equation to write the second-order time
derivatives in terms of second-order space derivatives, we obtain

\[
\frac{\partial E^{m+\frac{1}{2},n}}{\partial t} = \frac{\alpha - \beta \Delta y^2 \nabla^2}{\epsilon \Delta y} (H^{m+\frac{1}{2},n+\frac{1}{2}} - H^{m+\frac{1}{2},n-\frac{1}{2}}) \quad (2.4)
\]

\[
\frac{\partial E^{m,n+\frac{1}{2}}}{\partial t} = \frac{\alpha - \beta \Delta x^2 \nabla^2}{\epsilon \Delta x} (H^{m-\frac{1}{2},n+\frac{1}{2}} - H^{m+\frac{1}{2},n+\frac{1}{2}}) \quad (2.5)
\]

\[
\frac{\partial H^{m+\frac{1}{2},n+\frac{1}{2}}}{\partial t} = \frac{\alpha - \beta \Delta y^2 \nabla^2}{\mu \Delta y} (E^{m+\frac{1}{2},n+1} - E^{m+\frac{1}{2},n}) + \frac{\alpha - \beta \Delta x^2 \nabla^2}{\mu \Delta x} (E^{m,n+\frac{1}{2}} - E^{m+1,n+\frac{1}{2}}) \quad (2.6)
\]

Note that the Helmholtz equation in this form is only valid for homogenous medium.

For inhomogeneous medium, more general vector wave equation should be used. The full discrete equations are obtained by using a leap-frog scheme as

\[
E^{m+\frac{1}{2},n} = E^{m+\frac{1}{2},n} + \frac{\Delta t}{\epsilon \Delta y} \{[\alpha + \beta (3 + 2 \Delta y^2 \Delta x^2)] (H^{m+\frac{1}{2},n+\frac{1}{2}} - H^{m+\frac{1}{2},n-\frac{1}{2}}) + \beta \Delta y^2 \Delta x^2 \}
\]

\[
(H^{m+\frac{3}{2},n-\frac{1}{2}} - H^{m+\frac{3}{2},n+\frac{1}{2}} + H^{m-\frac{3}{2},n-\frac{1}{2}} - H^{m-\frac{3}{2},n+\frac{1}{2}} + \beta (H^{m+\frac{1}{2},n-\frac{1}{2}} - H^{m+\frac{1}{2},n+\frac{1}{2}})) \quad (2.7)
\]

\[
E^{m,n+\frac{1}{2}} = E^{m,n+\frac{1}{2}} + \frac{\Delta t}{\epsilon \Delta x} \{[\alpha + \beta (3 + 2 \Delta x^2 \Delta y^2)] (H^{m+\frac{1}{2},n+\frac{1}{2}} - H^{m+\frac{1}{2},n-\frac{1}{2}}) + \beta \Delta x^2 \Delta y^2 \}
\]

\[
(H^{m+\frac{1}{2},n+\frac{3}{2}} - H^{m+\frac{1}{2},n-\frac{3}{2}} + H^{m-\frac{1}{2},n-\frac{3}{2}} - H^{m-\frac{1}{2},n+\frac{3}{2}} + \beta (H^{m+\frac{3}{2},n+\frac{1}{2}} - H^{m+\frac{3}{2},n-\frac{1}{2}})) \quad (2.8)
\]

\[
H^{m+\frac{1}{2},n+\frac{1}{2}} = H^{m+\frac{1}{2},n-\frac{1}{2}} + \frac{\Delta t}{\mu \Delta y} \{[\alpha + \beta (3 + 2 \Delta y^2 \Delta x^2)] (E^{m+\frac{1}{2},n+1} - E^{m+\frac{1}{2},n}) + \beta \Delta y^2 \Delta x^2 \}
\]

\[
(E^{m+\frac{1}{2},n+1} - E^{m+\frac{1}{2},n-1} + E^{m+\frac{3}{2},n-\frac{1}{2}} - E^{m+\frac{3}{2},n+\frac{1}{2}} + \beta (E^{m+\frac{1}{2},n-\frac{1}{2}} - E^{m+\frac{1}{2},n+\frac{1}{2}})) + \frac{\Delta t}{\mu \Delta x} \{[\alpha + \beta (3 + 2 \Delta x^2 \Delta y^2)] (E^{m,n+\frac{1}{2}} - E^{m+1,n+\frac{1}{2}}) + \beta \Delta x^2 \Delta y^2 \}
\]

\[
+E^{m+1,n-\frac{1}{2}} - E^{m+1,n+\frac{1}{2}} + \beta (E^{m+2,n+\frac{1}{2}} - E^{m+1,n+\frac{1}{2}}) \quad (2.9)
\]

where the subscripts denote the corresponding time step.
2.2.2 Numerical Stability Analysis

To analyze the stability of the resulting AO-FDTD, we employ a Von Neumann analysis [63]. The \( E \) and \( H \) fields are expanded into Fourier modes. For each mode

\[
E_{xl}^{\frac{m+1}{2},n} = E_{xl}(t)e^{-j[k_x(m+\frac{1}{2})\Delta x+k_y n\Delta y]}
\]  
(2.10)

\[
E_{yl}^{\frac{m,n+\frac{1}{2}}{2}} = E_{yl}(t)e^{-j[k_x m\Delta x+k_y (n+\frac{1}{2})\Delta y]}
\]  
(2.11)

\[
H_{zl}^{\frac{m+\frac{1}{2},n+\frac{1}{2}}{2}} = H_{zl}(t)e^{-j[k_x(m+\frac{1}{2})\Delta x+k_y(n+\frac{1}{2})\Delta y]}
\]  
(2.12)

Substituting Eqs. (2.10)–(2.12) into Eqs. (2.7)–(2.9), we get

\[
\begin{pmatrix}
E_{xl+1} \\
E_{yl+1} \\
H_{zl+\frac{1}{2}}
\end{pmatrix}
= 
\begin{pmatrix}
1 - \frac{\xi C_x^2}{\Delta y^2} & \frac{\xi C_x C_y}{\Delta y \Delta x} & \frac{2j\xi C_x}{\epsilon \Delta y} \\
\frac{\xi C_x C_y}{\Delta y \Delta x} & 1 - \frac{\xi C_y^2}{\Delta x^2} & -\frac{2j\xi C_y}{\epsilon \Delta x} \\
-\frac{2j\xi C_x}{\mu \Delta y} & -\frac{2j\xi C_y}{\mu \Delta x} & 1
\end{pmatrix}
\begin{pmatrix}
E_{xl} \\
E_{yl} \\
H_{zl-\frac{1}{2}}
\end{pmatrix}
\]  
(2.13)

where

\[
\xi = \frac{4\Delta t^2}{\mu \epsilon}
\]

\[
C_x = \beta \sin\left(\frac{3k_y \Delta y}{2}\right) - [\alpha + 3\beta + 4\beta \frac{\Delta y^2}{\Delta x^2} \sin^2\left(\frac{k_x \Delta x}{2}\right)] \sin\left(\frac{k_y \Delta y}{2}\right)
\]  
(2.14)

\[
C_y = \beta \sin\left(\frac{3k_x \Delta x}{2}\right) - [\alpha + 3\beta + 4\beta \frac{\Delta x^2}{\Delta y^2} \sin^2\left(\frac{k_y \Delta y}{2}\right)] \sin\left(\frac{k_x \Delta x}{2}\right)
\]  
(2.15)

The eigenvalues of the amplification matrix in (2.13) are:

\[
\lambda_1 = 1
\]

\[
\lambda_2 = 1 - \left(\frac{\sqrt{\mu \epsilon \xi \cdot \zeta + \sqrt{\xi^2 \cdot \zeta \mu \epsilon - 16 \Delta x^2 \Delta y^2 \Delta t^2}}}{2 \Delta x^2 \Delta y^2 \sqrt{\mu \epsilon}}\right)
\]

\[
\lambda_3 = 1 - \left(\frac{\sqrt{\mu \epsilon \xi \cdot \zeta - \sqrt{\xi^2 \cdot \zeta \mu \epsilon - 16 \Delta x^2 \Delta y^2 \Delta t^2}}}{2 \Delta x^2 \Delta y^2 \sqrt{\mu \epsilon}}\right)
\]

with

\[
\zeta = C_x^2 \Delta x^2 + C_y^2 \Delta y^2
\]
For the algorithm to be stable, we require $|\lambda_{1,2,3}| \leq 1$. Thus

$$\Delta t \leq \Delta x \Delta y / (v_p \sqrt{\zeta}) \quad (2.16)$$

where $v_p = 1/\sqrt{\mu\varepsilon}$ is the exact phase velocity in the medium. When $\alpha = 1, \beta = 0$, this recovers to the usual Courant condition. Considering $\Delta x = \Delta y$, Eq. (2.16) becomes

$$\Delta t \leq h / (v_p \sqrt{C_x^2 + C_y^2})$$

where $h$ is the cell size. From Eqs. (2.14)–(2.15), the largest possible values of $|C_x|$ and $|C_y|$ are $\alpha + 8\beta$. Thus

$$\Delta t \leq \frac{h}{\sqrt{2} v_p (\alpha + 8\beta)} \quad (2.17)$$

### 2.2.3 Optimal Angle Selection

Assuming a monochromatic wave so that

$$E_{xl}(t) = E_x \exp(j\omega l \Delta t)$$
$$E_{yl}(t) = E_y \exp(j\omega l \Delta t)$$
$$H_{zl+1/2}(t) = H_z \exp[j\omega(l + \frac{1}{2}) \Delta t]$$

then Eqs. (2.7)–(2.9) become

$$\sin(\frac{\omega\Delta t}{2}) E_x = \frac{\Delta t}{\epsilon \Delta y} C_z H_z$$
$$\sin(\frac{\omega\Delta t}{2}) E_y = -\frac{\Delta t}{\epsilon \Delta x} C_y H_z$$
$$\sin(\frac{\omega\Delta t}{2}) H_z = \frac{\Delta t}{\mu \Delta y} C_x E_x - \frac{\Delta t}{\mu \Delta x} C_y E_y$$

Substituting Eqs. (2.18)–(2.19) into Eq. (2.20) yields the dispersion relation

$$\left[ \frac{\sqrt{2}}{\chi} \sin(\frac{\chi k h}{2\sqrt{2}}) \right]^2 = \tilde{C}_x^2 + \tilde{C}_y^2$$

8
where \( \chi = v_p \Delta t \sqrt{2} / h \) is the CFL number, \( k \) is the wavenumber, \( k = \sqrt{k_x^2 + k_y^2} \), and the terms \( \tilde{C}_x \) and \( \tilde{C}_y \) are defined as

\[
\tilde{C}_x = C_x|_{k_x=k \cos(\theta), k_y=k \sin(\theta), \Delta x=\Delta y}
\]
\[
\tilde{C}_y = C_y|_{k_x=k \cos(\theta), k_y=k \sin(\theta), \Delta x=\Delta y}
\]

where \( \theta \) is the wave propagation angle.

Note that in deriving Eq. (2.21), we have assumed \( k = \omega / v_p \). For a given pair \( \theta \) and \( \chi \), the error in the phase velocity can be made zero if \( \alpha \) and \( \beta \) are chosen properly. In other words, we can solve for \( \alpha \) and \( \beta \) from Eq. (2.21) for a given \( \theta \) and \( \chi \). In order to do this, we define the error

\[
\delta = \left( \frac{\sqrt{2}}{\chi} \sin\left(\frac{\chi k h}{2 \sqrt{2}}\right) \right)^2 - (\tilde{C}_x^2 + \tilde{C}_y^2)
\]

Solving Eq. (2.21) is equivalent to letting \( \delta = 0 \). By noting that \( kh = 2\pi h / \lambda \), we may expand \( \delta \) in series in terms of \( q = h / \lambda \) as:

\[
\delta = (1 - \alpha^2) \pi^2 q^2 + \frac{\pi^4 q^4}{12} \left\{ \alpha \left[ 3(\alpha - 32\beta) + \alpha \cos(4\theta) \right] - 2\chi^2 \right\} + \frac{\pi^6 q^6}{180} \left\{ 720\alpha \beta - 5\alpha^2 - 2880\beta^2 + 2\chi^4 - 3\alpha(\alpha - 80\beta) \cos(4\theta) \right\} + \mathcal{O}(q^8)
\]

where \( q \) is the reciprocal of the number of cells per wavelength (\( q < 0.1 \) in practice).

To solve for \( \alpha \) and \( \beta \), we force the first 2 terms in Eq. (2.23) to be zero. The resulting error will be \( \mathcal{O}(q^6) \). Thus

\[
\alpha = 1.0 \quad \beta = \frac{[3 - 2\chi^2 + \cos(4\theta)]}{96}
\]

For \( 0 \leq \beta < 1/24 \), the minimum maximum \( \chi \) for a stable scheme is 0.75. This can be used to minimize the dispersion error for any angle with guaranteed stability.

Note that \( \theta \) only appears in \( \beta \), which controls the higher order time derivative term in Eqs. (2.1)–(2.3) and hence the reason why this AO-FDTD scheme could be
used for any angle selection. For ordinary FDTD, $\beta = 0$ and $\chi = 1$ yield $\theta = 45^\circ$ in the above, as expected.

### 2.2.4 Optimal Frequency Band Selection

Monochromatic problems comprise just a small fraction of problems of interest for FDTD simulations. In most cases, one is confronted with problems involving broadband excitations. Therefore, it is also important to treat and optimize the dispersion error around a frequency band as well as around a propagation angle. This is described next.

The monomials $\{q^n\}$ can be thought as a basis of an infinite dimensional linear space $P_\infty[q]$ and Eq. (2.23) as an expansion of the dispersion error $\delta(q)$ in $P_\infty[q]$. Therefore, we might as well choose another basis to expand Eq. (2.23), e.g., $\{(q-q_0)^n\}$ or $\{T_n[(q - q_0)/\Delta q]\}$, where $q_0$ corresponds to a center frequency, $\Delta q$ corresponds to the frequency range of interest, and $T_n(x)$ is the n-th order first kind Chebyshev polynomial. In this manner, we are able to manipulate the frequency response of the dispersion error $\delta$. The first alternative corresponds to a Butterworth filter and the second one to a Chebyshev filter.

The expansion for $\delta$ up to order $p = 6$ in Eq. (2.23) can be rewritten as

$$\delta = q^2 (d_1, 0, d_3, 0, d_5) \cdot V^T + O(q^8)$$  \hfill (2.25)

$$d_1 = (1 - \alpha^2)\pi^2$$

$$d_3 = \pi^4 \{\alpha[3(\alpha - 32\beta) + \alpha \cos(4\theta)] - 2\chi^2\}/12$$

$$d_5 = \pi^6 [720\alpha\beta - 5\alpha^2 - 2880\beta^2 + 2\chi^4 - 3\alpha(\alpha - 80\beta) \cos(4\theta)]/180$$

$$V = (1, q, q^2, q^3, q^4)$$
For the new basis \{ (q - q_0)^n \}, the following relationship holds

\[
Q^T = A \cdot V^T
\]

\[
Q = (1, (q - q_0), (q - q_0)^2, (q - q_0)^3, (q - q_0)^4)
\]

\[
A = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 \\
-q_0 & 1 & 0 & 0 & 0 \\
q_0^2 & -2q_0 & 1 & 0 & 0 \\
-q_0^3 & 3q_0^2 & -3q_0 & 1 & 0 \\
q_0^4 & -4q_0^3 & 6q_0^2 & -4q_0 & 1
\end{pmatrix}
\]

and, therefore, Eq. (2.25) can be rewritten as

\[
\delta = q^2 (d_1, 0, d_3, 0, d_5) \cdot A^{-1} \cdot Q^T + \mathcal{O}(q^8)
\]

\[
= q^2 (\tilde{d}_1, \tilde{d}_2, \tilde{d}_3, \tilde{d}_4, \tilde{d}_5) \cdot Q^T + \mathcal{O}(q^8) \tag{2.26}
\]

with

\[
\tilde{d}_1 = \pi^2 \left[ 1 - \frac{(36 - 9\pi^2 q_0^2 + \pi^4 q_0^4)\alpha^2}{36} \right] + \frac{\pi^4 q_0^2}{180} \{ 720(\pi^2 q_0^2 - 2)\alpha \beta - 30\chi^2 + 2\pi^2 \chi q_0^2(\chi^4 - 1440\beta^2) - 3\alpha(\pi^2 q_0^2 - 5)\alpha - 80\pi^2 q_0^2\beta \} \cos(4\theta) \}
\]

\[
\tilde{d}_2 = \frac{\pi^4 q_0^2}{90} \{ (45 - 10\pi^2 q_0^2)\alpha^2 + 1440(\pi^2 q_0^2 - 1)\alpha \beta - 30\chi^2 + 4\pi^2 q_0^2(\chi^4 - 1440\beta^2) + 3\alpha(5 - 2\pi^2 q_0^2)\alpha + 160\pi^2 q_0^2\beta \} \cos(4\theta) \}
\]

\[
\tilde{d}_3 = \frac{\pi^4}{60} \{ (15 - 10\pi^2 q_0^2)\alpha^2 + 480(3\pi^2 q_0^2 - 1)\alpha \beta - 10\chi^2 + 4\pi^2 q_0^2(\chi^4 - 1440\beta^2) + \alpha(5 - 6\pi^2 q_0^2)\alpha + 480\pi^2 q_0^2\beta \} \cos(4\theta) \}
\]

\[
\tilde{d}_4 = -\frac{\pi^6 q_0^2}{45} [ 5\alpha^2 - 720\alpha \beta + 2880\beta^2 - 2\chi^4 + 3\alpha(\alpha - 80\beta) \cos(4\theta) ]
\]

\[
\tilde{d}_5 = -\frac{\pi^6}{180} [ 5\alpha^2 - 720\alpha \beta + 2880\beta^2 - 2\chi^4 + 3\alpha(\alpha - 80\beta) \cos(4\theta) ]
\]

To solve for \( \alpha \) and \( \beta \), we force \( \tilde{d}_1 \) and \( \tilde{d}_2 \) to be 0. The explicit solutions are lengthy and not shown here. Now \( \delta = q^2 [\tilde{d}_3(q - q_0)^2 + \tilde{d}_4(q - q_0)^3 + \tilde{d}_5(q - q_0)^4] + \mathcal{O}(q^8) \).

At the center frequency, \( q = q_0 \) and \( \delta = \mathcal{O}(q^8) \). The remainder corresponds to a \( \mathcal{O}(q^8) \) error and because of this, the real optimal (center) frequency will shift slightly.
towards low frequencies. Note that the filter is fixed to be of second order. Around the center frequency, the error is dominated by the $q^2\tilde{d}_3(q - q_0)^2$ term.

Suppose the optimized angle is $\theta_0$ and we consider $\theta = \theta_0 + \Delta \theta$. Then $\Delta \tilde{d}_1 \propto \sin(4\theta)\Delta \theta$ and $\Delta \tilde{d}_2 \propto \sin(4\theta)\Delta \theta$. The numerical dispersion around $\theta_0$ is very smooth. This is important for the structures we intend to model because the main lobe always has certain angular width. Indeed, $\delta$ is the most flat around $\theta_0 = 0^\circ$. Furthermore, $\delta$ is periodic due to the $\cos(4\theta)$ term and whenever it is optimized for some angle $\theta_0$, other 7 angles ($\pi/2 \pm \theta_0, \pi \pm \theta_0, 3\pi/2 \pm \theta_0$ and $2\pi - \theta_0$) are automatically optimized.

Now let us choose $\{T_n[(q - q_0)/\Delta q]\}$ as a new basis (Chebyshev filtering). The $Q$ vector and the $A$ matrix become

$$Q = (1, \frac{q - q_0}{\Delta q}, T_2(\frac{q - q_0}{\Delta q}), T_3(\frac{q - q_0}{\Delta q}), T_4(\frac{q - q_0}{\Delta q}))$$

$$A = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 \\
\frac{-q_0}{\Delta q} & \frac{1}{\Delta q} & 0 & 0 & 0 \\
\frac{2q_0^2-\Delta q^2}{\Delta q^2} & \frac{-4q_0}{\Delta q^2} & \frac{2}{\Delta q^2} & 0 & 0 \\
\frac{3q_0\Delta q^2-4q_0^3}{\Delta q^3} & \frac{-12q_0^3-3\Delta q^2}{\Delta q^3} & \frac{-12q_0}{\Delta q^3} & \frac{1}{\Delta q^3} & 0 \\
\frac{8q_0^3-8q_0\Delta q^2+\Delta q^4}{\Delta q^4} & \frac{16q_0\Delta q^2-32q_0^3}{\Delta q^4} & \frac{48q_0^2-8\Delta q^2}{\Delta q^4} & \frac{-32q_0}{\Delta q^4} & \frac{8}{\Delta q^4}
\end{pmatrix}$$

Following the same procedure as Butterworth filter case, we can solve for the coefficients $\alpha$ and $\beta$. The numerical dispersion around optimized angle $\theta_0$ follows the same pattern but the behavior around center frequency is different. Here

$$\delta = q^2\{\tilde{d}_2T_2[(q - q_0)/\Delta q] + \tilde{d}_3T_3[(q - q_0)/\Delta q] + \tilde{d}_4T_4[(q - q_0)/\Delta q]\} + \mathcal{O}(q^8)$$

where $\tilde{d}_2$, $\tilde{d}_3$ and $\tilde{d}_4$ denote the coefficients of the corresponding Chebyshev polynomials. Since $T_2[(q - q_0)/\Delta q]$ takes the minimum and $T_3[(q - q_0)/\Delta q]$ takes zero at $q_0$, the minimum numerical dispersion will occur at the center without any slight shift towards low frequency. For both Butterworth and Chebyshev filters, a suitable $\chi$ is determined to be 0.74.
2.2.5 Angular Behavior of Numerical Dispersion

We now look at the anisotropy of the AO-FDTD scheme. This tells us its behavior in the whole angular range. To study this, we can expand the dispersion error $\delta$ given by Eq. (2.22) in Fourier series in terms of $\sin(n\theta)$ and $\cos(n\theta)$. Or we may start from another form equivalent to Eq. (2.22)

$$\delta = \left[ \sqrt{2} \frac{\sin\left(\frac{\chi \pi q}{\sqrt{2}}\right)}{\chi} \right]^2 - (F_x^2 + F_y^2) \quad (2.27)$$

where

$$F_x = (\alpha + 4\beta \pi^2 q^2) \sin^2(\pi q \sin(\theta))$$

$$F_y = (\alpha + 4\beta \pi^2 q^2) \sin^2(\pi q \cos(\theta))$$

It could be derived directly from Eqs. (2.1)–(2.3). Expanding Eq. (2.28) in a Fourier series, we have

$$\delta = \left[ \sqrt{2} \frac{\sin\left(\frac{\chi \pi q}{\sqrt{2}}\right)}{\chi} \right]^2 - (\alpha + 4\beta \pi^2 q^2)^2(1 - J_0(2\pi q)) + (\alpha + 4\beta \pi^2 q^2)^2[2\pi q(4\pi^2 q^2 - 24) J_0(2\pi q) - 8(4\pi^2 q^2 - 6) J_1(2\pi q)]/(4\pi^3 q^3) \cdot \cos(4\theta) + O(\cos(8\theta)) \quad (2.28)$$

where $J_n(x)$ denotes the $n$-th order first kind Bessel function. The DC (isotropic) term in Eq. (2.28) introduces a uniform error for all angles and the $\cos(4\theta)$ term leads to anisotropic errors. The contribution of the higher order $\cos(n\theta)$ terms is much smaller compared to the $\cos(4\theta)$ term, and therefore, the isotropic term plus the $\cos(4\theta)$ terms essentially determine the angular behavior of the numerical dispersion.

As an example, Fig. 2.1 and Fig. 2.2 show the isotropic term and the $\cos(4\theta)$ term of the dispersion error $\delta$, when inserting Eq. (2.24) into Eq. (2.28), optimized for $15^\circ$. By changing the optimized angle, the isotropic term changes from nearly zero at $22.5^\circ$ to nearly 0.0085 at $45^\circ$ while the $\cos(4\theta)$ term remains about the same as the one depicted in Fig. 2.2.
In fact, in terms of the inf-sup (minimum maximum) numerical dispersion, the AO-FDTD optimized at 22.5° is an optimized one. This can be shown by making the term \((\alpha + 4\beta\pi^2q^2)\) to be a unknown, say \(\nu\), and solve for it by enforcing the isotropic term in Eq. (2.28) to be zero, which leads to

\[
\nu = \sqrt{\frac{\cos(\sqrt{2}\pi q\chi) - 1}{\chi^2(J_0(2\pi q) - 1)}}
\]

By expanding \(\nu\) in a Taylor series in terms of \(q\), we obtain

\[
\nu = 1 + \frac{(3 - 2\chi^2)\pi^2q^2}{24}
\]

Note that at 15°, the anisotropic term will almost cancel out the isotropic term. For the filtered schemes, the overall behavior is quite similar. The ordinary Yee’s FDTD scheme is a special case when \(\alpha = 1\) and \(\beta = 0\) so that at \(\chi = 1.0\), it will be automatically optimized (zero dispersion) for 45° from Eq. (2.24). Indeed, the
anisotropy of the AO-FDTD scheme is expected to be the same order as the ordinary FDTD (no improvement on this aspect).

For the higher-order (2,4) FDTD scheme introduced in [13], we could obtain a dispersion error similar to Eq. (2.28) and proceed with a similar analysis as before. The dispersion error for this (2,4) FDTD scheme is expressed as

$$\delta = \left[ \frac{\sqrt{2}}{\chi} \sin \left( \frac{\pi q}{\sqrt{2}} \right) \right]^2 + \frac{1}{576} \left[ 9 \cdot \sin^2 \left( \frac{\pi q}{\sqrt{2}} \right) \right] \cdot \cos(4\theta) + O(\cos(8\theta))$$

(2.29)

Fig. 2.1 and Fig. 2.2 also depict the isotropic and the cos(4\theta) terms of the (2,4) FDTD dispersion error as a function of q when evaluating Eq. (2.29) for 0.01 \leq q \leq 0.1 at \chi = 0.85 [13]. For this higher-order scheme, we observe that the cos(4\theta) term is (designed to be) much smaller than the isotropic term.
2.2.6 Numerical Results

In this section, we compare in more detail the AO-FDTD method with a regular FDTD method and the higher-order (2,4) scheme. Both Butterworth, Chebyshev and non-filtered angle selection schemes are tested at 0°, 15° and 22.5° with $\chi = 0.74$. We use a Butterworth filter with $\delta$ expanded to the $p = 6$ order. The central frequency is such that $q_0 = 0.08$. Furthermore, for the Chebyshev filter we use $\Delta q = 0.02$. The corresponding coefficients are:

\begin{table}[h]
\centering
\begin{tabular}{|l|c|c|}
\hline
 & $\alpha$ & $\beta$ \\
\hline
Butterworth $\theta_0 = 0^\circ$ & 0.999774 & 0.032036 \\
Butterworth $\theta_0 = 15^\circ$ & 0.999843 & 0.0262875 \\
Chebyshev $\theta_0 = 0^\circ$ & 0.99978 & 0.0321267 \\
Chebyshev $\theta_0 = 15^\circ$ & 0.999847 & 0.02635 \\
Non-filtered $\theta_0 = 15^\circ$ & 1.0 & 0.02505 \\
Non-filtered $\theta = 22.5^\circ$ & 1.0 & 0.0198417 \\
\hline
\end{tabular}
\caption{Some coefficients used in 2-D AO-FDTD.}
\end{table}

The phase velocity in free space is directly simulated in a 2-D FDTD grid. The FDTD domain is large enough so that any spurious reflections from the grid boundary are causally isolated from the results presented here. Moreover, the results are normalized to $c = 1/\sqrt{\mu_0\epsilon_0}$, and $\chi = 1$ for ordinary FDTD and $\chi = 0.85$ for the (2,4) scheme.

From Fig. 2.3 and Fig. 2.4, we observe that the AO-FDTD scheme yields almost no numerical dispersion at the optimized angle on a wide range of frequencies around the central frequency. For comparison, Fig. 2.5 depicts the normalized phase velocity of ordinary FDTD for the same frequency range. On the other hand, we see from Fig. 2.6 that the (2,4) scheme yields a highly isotropic result, but with larger absolute errors. Moreover, through a comparison of AO-FDTD results with ordinary FDTD

16
results in finer meshes (Fig. 2.7 vs. Figs. 2.9–2.10), we observe that the cell size would need to be reduced around at least 5 times to obtain a dispersion error around the optimized angle as small as the non-filtered AO-FDTD. In the case of (2,4) scheme, a mesh approximately 3 times finer would be required (Fig. 2.8 vs. Figs. 2.9–2.10) to reduce to error to the levels presented by the non-filtered AO-FDTD. For filtered AO-FDTD these factors will be even larger.

Fig. 2.9 shows the results for a non-filtered AO-FDTD scheme optimized at 22.5°. The numerical dispersion is almost symmetric around this angle. By comparing this result with Figs. 2.3–2.5, we see the anisotropic behavior follows the same pattern, but with a smaller maximum error, as discussed in the Section VI.

Fig. 2.10 compares the relative error on the normalized phase velocity for the AO-FDTD with Butterworth filter, Chebyshev filter and the non-filtered AO-FDTD
Figure 2.4: Normalized phase velocity of AO-FDTD scheme optimized at $0^\circ$ using Butterworth filter.

scheme. In this figure, the relative errors are the absolute differences between normalized phase velocities and 1.0. Notice that the real (optimal) center frequency for Butterworth filter shifts slightly towards the low frequency end, as explained before. The Chebyshev filter performs best and its actual optimal frequency is almost exactly as the center frequency specified theoretically, $q_0 = 0.08$.

Finally, Fig. 2.11 presents the relative error on the normalized phase velocity for four different angles around the optimized angle. A Chebyshev filter optimized at $0^\circ$ is used. In the vicinity of the optimized angle, we see that the AO-FDTD with Chebyshev filtering works very well: an error no larger that 0.09% is obtained in an angular sector as large as $20^\circ$ in this case.
2.3 Three-Dimensional AO-FDTD Algorithm

2.3.1 Update Equations and Stability Analysis

We introduce a modified set of Maxwell’s curl equations as follows

\[
\begin{align*}
([T] \cdot \nabla) \times \vec{E} &= -\frac{\partial \vec{B}}{\partial t} \\
([T] \cdot \nabla) \times \vec{H} &= \frac{\partial \vec{D}}{\partial t}
\end{align*}
\] (2.30)

where \([T] \cdot \nabla\) denotes the dot product of vector differential operator \(\nabla\) and an artificial correction tensor \([T]\) defined as

\[
[T] = \begin{bmatrix}
\alpha - \beta \frac{\Delta x^2 \partial^2}{v_p^2} & 0 & 0 \\
0 & \alpha - \beta \frac{\Delta y^2 \partial^2}{v_p^2} & 0 \\
0 & 0 & \alpha - \beta \frac{\Delta z^2 \partial^2}{v_p^2}
\end{bmatrix}
\] (2.31)

in which \(\Delta x\), \(\Delta y\) and \(\Delta z\) are the cell size along \(x\), \(y\) and \(z\) directions respectively, \(v_p = 1/\sqrt{\mu\varepsilon}\) is the exact (continuum) phase velocity in the medium, and \(\alpha\) and \(\beta\) are additional degrees of freedom in the equations. These parameters are to be
Figure 2.6: Normalized phase velocity of (2,4) FDTD scheme

determined in order to control phase correcting (higher-order) time derivative terms. Such an artificial correction tensor is designed so as to introduce artificial dispersion effects which compensate for the numerical dispersion.

Second order time derivatives terms can also be written in terms of the second order space derivatives via the Helmholtz equation [8]. Again, vector wave equations should be used in general. Thus Eq. (2.31) becomes

\[
[T] = \begin{bmatrix}
\alpha - \beta \Delta x^2 \nabla^2 & 0 & 0 \\
0 & \alpha - \beta \Delta y^2 \nabla^2 & 0 \\
0 & 0 & \alpha - \beta \Delta z^2 \nabla^2
\end{bmatrix}
\] (2.32)

Implementing Eq. (2.32) in the staggered FDTD grid with central differencing in space and a leap-frog discretization in time, we obtain fully discrete update equations for the AO-FDTD. For example, the \( E_x \) update is written as

\[
E_{x_i+\frac{1}{2},m,n}^{t+1} = E_{x_i}^{t+\frac{1}{2},m,n} + \frac{\Delta t}{\varepsilon \Delta y} \left\{ \left[ \alpha + \beta \left( 3 + 2 \frac{\Delta y^2}{\Delta x^2} + 2 \frac{\Delta y^2}{\Delta z^2} \right) \right] (H_{z_i+\frac{1}{2},m+\frac{1}{2},n}^{t+\frac{1}{2}} - H_{z_i+\frac{1}{2},m-\frac{1}{2},n}^{t+\frac{1}{2}}) - \beta \frac{\Delta y^2}{\Delta x^2} \right\}
\]
Figure 2.7: Normalized phase velocity of ordinary FDTD method in a finer mesh.

\[
(H_z^{l+\frac{3}{2},m+\frac{3}{2},n} - H_z^{l+\frac{1}{2},m+\frac{1}{2},n}) + \frac{\Delta y^2}{\Delta z^2}(H_z^{l+\frac{1}{2},m+\frac{1}{2},n+1} - H_z^{l+\frac{1}{2},m+\frac{1}{2},n-1}) + \frac{\Delta z^2}{\Delta y^2}(H_y^{l+\frac{3}{2},m,n+\frac{1}{2}} - H_y^{l+\frac{1}{2},m,n-\frac{1}{2}}) - \frac{\Delta t}{\epsilon \Delta z} \{[\alpha + \beta (3+2\frac{\Delta z^2}{\Delta x^2})] - \beta \frac{\Delta z^2}{\Delta x^2}(H_y^{l+\frac{1}{2},m,n+\frac{1}{2}} - H_y^{l+\frac{1}{2},m,n-\frac{1}{2}}) + H_y^{l+\frac{1}{2},m,n+\frac{1}{2}} - H_y^{l+\frac{1}{2},m,n-\frac{1}{2}} + H_y^{l+\frac{1}{2},m,n+\frac{1}{2}} - H_y^{l+\frac{1}{2},m,n-\frac{1}{2}} + \frac{\Delta z^2}{\Delta y^2}(H_y^{l+\frac{1}{2},m+1,n+\frac{1}{2}} - H_y^{l+\frac{1}{2},m+1,n-\frac{1}{2}})
\]

where the superscripts denote the spatial location of the field components and the subscripts denote the time step. The other update equations are similar.

To analyze the stability of the AO-FDTD method, we employ a Von Neumann analysis [63]. The \(E\) and \(H\) fields are expanded into Fourier modes. For each mode
Substituting Eq. (2.33) into the discrete update equations, we obtain the system amplification matrix as

\[
\begin{bmatrix}
1 - \xi \left( C_{xy}^2 + C_{xz}^2 \right) & \xi C_{xz} C_{yz} & \xi C_{xy} C_{yz} & 0 & \zeta C_{xy} & -\zeta C_{xz} \\
\zeta C_{xz} C_{yz} & 1 - \xi \left( C_{xy}^2 + C_{yz}^2 \right) & \xi C_{xy} C_{xz} & 0 & \zeta C_{xy} & -\zeta C_{yz} \\
\xi C_{xy} C_{yz} & \xi C_{xy} C_{xz} & 1 - \xi \left( C_{yz}^2 + C_{xz}^2 \right) & -\zeta C_{xy} & 0 & \zeta C_{xy} \\
0 & -\gamma C_{xy} & -\gamma C_{yz} & \gamma C_{xz} & 1 & 0 \\
\gamma C_{xy} & 0 & 0 & -\gamma C_{yz} & 0 & 1 \\
-\gamma C_{xz} & \gamma C_{yz} & 0 & 0 & 0 & 1
\end{bmatrix}
\] (2.34)
Figure 2.9: Normalized phase velocity of AO-FDTD scheme optimized at 22.5° using non-filtered scheme

where

\[ \xi = \frac{4 \Delta t^2}{\mu c}, \quad \zeta = 2 j \frac{\Delta t}{\epsilon}, \quad \gamma = 2 j \frac{\Delta t}{\mu} \]

\[ C_{xy} = \frac{1}{\Delta x} \left[ \alpha + \beta \left( 3 + 2 \frac{\Delta x^2}{\Delta y^2} + 2 \frac{\Delta x^2}{\Delta z^2} \right) \right] \sin \left( \frac{k_x \Delta z}{2} \right) - \frac{\beta}{\Delta z} \left[ \sin \left( \frac{3k_x \Delta z}{2} \right) + 2 \sin \left( \frac{k_x \Delta z}{2} \right) \right] \]

\[ C_{xz} = \frac{1}{\Delta y} \left[ \alpha + \beta \left( 3 + 2 \frac{\Delta x^2}{\Delta y^2} + 2 \frac{\Delta x^2}{\Delta z^2} \right) \right] \sin \left( \frac{k_x \Delta y}{2} \right) - \frac{\beta}{\Delta y} \left[ \sin \left( \frac{3k_x \Delta y}{2} \right) + 2 \sin \left( \frac{k_x \Delta y}{2} \right) \right] \]

\[ C_{yz} = \frac{1}{\Delta x} \left[ \alpha + \beta \left( 3 + 2 \frac{\Delta y^2}{\Delta x^2} + 2 \frac{\Delta y^2}{\Delta z^2} \right) \right] \sin \left( \frac{k_y \Delta z}{2} \right) - \frac{\beta}{\Delta z} \left[ \sin \left( \frac{3k_y \Delta z}{2} \right) + 2 \sin \left( \frac{k_y \Delta z}{2} \right) \right] \]

The eigenvalues of the system amplification matrix in Eq. (2.34) are

\[ \lambda = 1, \quad \lambda = 1 - \frac{\Xi \pm \sqrt{\Xi^2 - 4\zeta}}{2} \]
with

\[ \Xi = C_{xy}^2 + C_{xz}^2 + C_{yz}^2 \]

For the algorithm to be stable, we require \(|\lambda| \leq 1\). Thus

\[ \Delta t \leq \frac{1}{v_p \sqrt{\Xi}} \] (2.38)

When \(\alpha = 1, \beta = 0\), this reduces to the usual Courant condition. Considering \(\Delta x = \Delta y = \Delta z = h\), the largest possible value of \(|C_{xy}|, |C_{xz}|\) and \(|C_{yz}|\) is \((\alpha + 12\beta)/h\). Thus

\[ \Delta t \leq \frac{h}{\sqrt{3} v_p (\alpha + 12\beta)} \] (2.39)
2.3.2 Optimal Angle Selection

Assuming a monochromatic wave propagating in a uniform grid, then the time dependent parts of Eq. (2.33) can be written as

\[ \vec{E}_i(t) = \vec{E} e^{j\omega(i\Delta t)} \quad \vec{H}_{i+\frac{1}{2}}(t) = \vec{H} e^{j\omega((i+\frac{1}{2})\Delta t)} \]  

\[ (2.40) \]

Substituting Eq. (2.33) and Eq. (2.40) again into the update equations and eliminating all \( H \) field variables, we get

\[
\begin{bmatrix}
\frac{\sin^2(\frac{\omega \Delta t}{2})}{v_p^2 \Delta t^2} - \tilde{C}_{xz}^2 - \tilde{C}_{xy}^2 & \tilde{C}_{xz} \tilde{C}_{yz} & \tilde{C}_{xy} \tilde{C}_{yz} \\
\tilde{C}_{xz} \tilde{C}_{yz} & \frac{\sin^2(\frac{\omega \Delta t}{2})}{v_p^2 \Delta t^2} - \tilde{C}_{xy}^2 & \tilde{C}_{xy} \tilde{C}_{xz} \\
\tilde{C}_{xy} \tilde{C}_{yz} & \tilde{C}_{xy} \tilde{C}_{xz} & \frac{\sin^2(\frac{\omega \Delta t}{2})}{v_p^2 \Delta t^2} - \tilde{C}_{xz}^2 - \tilde{C}_{yz}^2
\end{bmatrix}
\begin{bmatrix}
\vec{E}_x \\
\vec{E}_y \\
\vec{E}_z
\end{bmatrix} = 0
\]

\[ (2.41) \]
where

\[ \tilde{C}_{xy} = C_{xy} \mid x = k \sin(\theta) \cos(\phi), y = k \sin(\theta) \sin(\phi), z = k \cos(\theta), \Delta x = \Delta y = \Delta z = h \]

\[ \tilde{C}_{xz} = C_{xz} \mid x = k \sin(\theta) \cos(\phi), y = k \sin(\theta) \sin(\phi), z = k \cos(\theta), \Delta x = \Delta y = \Delta z = h \]

\[ \tilde{C}_{yz} = C_{yz} \mid x = k \sin(\theta) \cos(\phi), y = k \sin(\theta) \sin(\phi), z = k \cos(\theta), \Delta x = \Delta y = \Delta z = h \]

\[ k = \sqrt{k_x^2 + k_y^2 + k_z^2} \text{ and } (\theta, \phi) \text{ are polar angles along which the wave propagates.} \]

For \( E_x, E_y \) and \( E_z \) to admit nontrivial, the determinant of the matrix in Eq. (2.41) must be zero. Thus we obtain the dispersion relation

\[ \left( \frac{\sqrt{3}}{k} \sin \left( \frac{\chi k h}{2\sqrt{3}} \right) \right)^2 = h^2 (\tilde{C}_{xy}^2 + \tilde{C}_{xz}^2 + \tilde{C}_{yz}^2) \quad (2.42) \]

where \( \chi = v_p \Delta t \sqrt{3}/h \) is the Courant (or CFL) number.

Note that in deriving Eq. (2.42), we have implicitly assumed \( k = \omega/v_p \). Therefore for given \((\theta, \phi)\) and \( \chi \), the error in the discrete phase velocity can be made zero if \( \alpha \) and \( \beta \) are properly chosen. In other words, we can solve for \( \alpha \) and \( \beta \) from Eq. (2.42) for given \((\theta, \phi)\) and \( \chi \). In order to do this, we first define the error

\[ \delta = \left( \frac{\sqrt{3}}{k} \sin \left( \frac{\chi k h}{2\sqrt{3}} \right) \right)^2 - h^2 (\tilde{C}_{xy}^2 + \tilde{C}_{xz}^2 + \tilde{C}_{yz}^2) \quad (2.43) \]

Solving Eq. (2.42) is equivalent to letting \( \delta = 0 \). By noting that \( k h = 2\pi h/\lambda \), we may expand \( \delta \) in series in terms of the reciprocal of the number of cells per wavelength, \( q = h/\lambda \), as:

\[ \delta = (1 - \alpha^2) \pi^2 q^2 + \frac{\pi^4 q^4}{288} \{ 63 \alpha^2 - 2304 \alpha \beta - 32 \chi^2 + 3 \alpha^2 [4 \cos(2\theta) + 7 \cos(4\theta)] \\
+ 8 \cos(4\phi) \sin^4(\theta) \} + O(q^6) \quad (2.44) \]

where, in practice, \( q \leq 0.1 \). To solve for \( \alpha \) and \( \beta \), we force the first two terms in Eq. (2.44) to be zero. The resulting error will be \( O(q^6) \). Thus

\[ \alpha = 1.0, \quad \beta = \frac{63 - 32 \chi^2 + 12 \cos(2\theta) + 21 \cos(4\theta) + 24 \cos(4\phi) \sin^4(\theta)}{2304} \quad (2.45) \]
For $0 \leq \beta < 1/24$, the minimum (for all angles) maximum $\chi$ for a stable scheme is found from Eq. (2.39) and Eq. (2.45) to be $2/3$. This is the CFL number which can be used to minimize the dispersion error for any angle with guaranteed stability. In reality, the CFL number can be made larger than $2/3$ depending on the specific $\alpha$ and $\beta$. Note that the angular dependency on $(\theta, \phi)$ only appears in $\beta$, which controls the higher order time derivative term and hence the reason why this AO-FDTD scheme could be used for arbitrary angle selection.

By using the filtering technique discussed in the previous section, we can also improve the results. This will be shown in the next subsection.

### 2.3.3 Numerical Experiments

For a particular choice of $\alpha$ and $\beta$ in Eq. (2.31), the actual phase velocity $v_p$ of AO-FDTD simulations can be solved as function of frequency and propagation angle directly from Eq. (2.42), as traditionally done in FDTD dispersion analysis. In this way, the AO-FDTD will be compared with the regular FDTD method (Yee’s scheme) and a scheme with second order of accuracy in time and fourth order of accuracy in space ((2,4) scheme) [12]. Butterworth, Chebyshev and non-filtered AO-FDTD schemes are tested at $\theta = 90^\circ$ and $\phi = 0^\circ$ (i.e., positive $x$-direction), with $\chi = 2/3$. Moreover, the results are normalized to $c = 1/\sqrt{\mu\epsilon}$. $\chi = 1$ is used for ordinary FDTD and $\chi = 6/7$ for the (2,4) scheme (maximum respective CFL numbers). The coefficients used here are listed in Table 3.2 and Table 3.3.

Figs. 2.12 and 2.13 show the normalized phase velocity of Butterworth and Chebyshev AO-FDTD schemes at $\theta = 90^\circ$ and $\phi = 0^\circ$ with different specification of center

\(^1\)As in 2-D case, it can be shown that the dispersion error along all positive and negative axes are simultaneously optimized.

27
Table 2.2: Some coefficients used in 3-D AO-FDTD.

<table>
<thead>
<tr>
<th></th>
<th>Non-filtered</th>
<th>Butterworth $q_c = 0.1$</th>
<th>Butterworth $q_c = 0.09$</th>
<th>Butterworth $q_c = 0.08$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>1</td>
<td>0.999319</td>
<td>0.999562</td>
<td>0.999732</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.0354938</td>
<td>0.0389119</td>
<td>0.0382103</td>
<td>0.0376044</td>
</tr>
</tbody>
</table>

Table 2.3: Coefficients used in the AO-FDTD with Chebyshev filtering scheme. $\Delta q = 0.02$ for all Chebyshev filters.

<table>
<thead>
<tr>
<th></th>
<th>Chebyshev $q_c = 0.1$</th>
<th>Chebyshev $q_c = 0.09$</th>
<th>Chebyshev $q_c = 0.08$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.999368</td>
<td>0.999591</td>
<td>0.999748</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.0388736</td>
<td>0.0382207</td>
<td>0.0376482</td>
</tr>
</tbody>
</table>

frequencies. The results are similar for these two schemes. In consonance with filter theory [49], we observe that the Butterworth filter gives slightly more accurate results than the Chebyshev filter at the specified center frequencies, while the Chebyshev filter produces a smaller maximum error than the Butterworth filter in frequency band shown, $0.01 \leq q \leq 0.1$ (and indeed not limited to $\Delta q = 0.02$ from the filter specification).

Note that in Fig. 2.12, the normalized phase velocities are not exactly 1.0 at the center frequencies and this discrepancy tends to increase as the center frequency increases. This phenomena is due to using a finite number of terms in the transformation of basis from $q^n$ to $\{(q - q_c)^n\}$. Theoretically, this error could be minimized by increasing the computation overhead when building the filters. Alternatively, we can in practice simply shift the center frequency slightly downward (e.g., optimizing at $q_c = 0.09$ instead of at $q_c = 0.1$).

An interesting point to observe in connection with Figs. 2.12-2.13 is that the local dispersion error is actually smaller for higher frequencies (closer to $q_c$) than for lower
Figure 2.12: Normalized phase velocity of Butterworth AO-FDTD scheme at $\theta = 90^\circ$ and $\phi = 0^\circ$ with different specification of center frequencies.

Optimized at $q_c = 0.09$.

Fig. 2.14 shows the comparison of different AO-FDTD schemes, where all filtering schemes are optimized with $q_c = 0.09$. We observe that filtering schemes indeed yield much better results than non-filtered scheme around designed frequencies, where the numerical dispersion is the most critical. Moreover, in the whole frequency range being considered, the maximum errors from both filtering schemes are also smaller than non-filtered scheme.

Fig. 2.15 shows the normalized phase velocity of ordinary FDTD and (2,4) schemes. Compared with non-filtered AO-FDTD scheme, approximately 5% error is observed at high frequencies (as opposed to traditional FDTD schemes). This is highly desirable in practice since at high frequencies the computational domain (of fixed physical size) is electrically larger than at low frequencies and hence the accumulated phase error form the FDTD simulations tends to be larger. With this in mind, $q_c$ may also be chosen close to the maximum frequency of interest.
times and 3 times finer meshes would be required for the same accuracy at \( \theta = 90^\circ \) and \( \phi = 0^\circ \), respectively. If compared with filtering schemes, even finer meshes would be required.

Finally, Fig. 2.16 illustrates the behavior of the dispersion error as a function of angle (anisotropy). A Butterworth filter optimized at \( q_c = 0.09 \) is used. The normalized phase velocity is presented for four different elevation angles close to a given optimized angle \( \theta_c = 90^\circ \). The azimuthal angle is fixed at \( \phi_c = 0^\circ \). Note that the numerical dispersion here is rotational symmetric about each coordinate axis, and the results for \( \theta_c - \theta \) are the same as \( \theta - \theta_c \). In the vicinity of the optimized angle, we see that the AO-FDTD works as intended: an error no larger that 0.06% is obtained in a angular sector as large as 15° around the optimized angle.
We next simulate a ultra wideband pulse propagation to illustrate the effectiveness of the angle optimization in reducing the pulse distortion caused by the numerical dispersion. The source is a \( \hat{z} \) directed electric dipole excited by a first-order differentiated Gaussian pulse centered at 7.5 GHz and having -10 dB points at about 1.5 GHz and 16 GHz. The medium is free-space and the computation domain is large enough so that any spurious reflection from the grid boundaries are causally isolated from the results presented below. The discretization cell size corresponds to 10 cells per wavelength at 10 GHz. The observation point is at 38 cells away from the source at an angle \( \theta = 90^\circ \) and \( \phi = 0^\circ \), for which the optimization is chosen for the 3-D AO-FDTD. The co-polarized component of the electric field at the observation point is illustrated in Fig.2.17. The largest possible Courant numbers are used in each case, i.e. 1 for Yee’s scheme and 2/3 for AO-FDTD scheme. As a result of the much smaller

Figure 2.14: Normalized phase velocity of Butterworth, Chebyshev and non-filtered AO-FDTD schemes at \( \theta = 90^\circ \) and \( \phi = 0^\circ \). For the filtered schemes, \( q_c = 0.09 \).
2.4 Concluding Remarks

Although not detailed here, the modification on the dispersion behavior of Maxwell’s equations given by Eq. (2.30) could also have been effected, in a dual formulation, entirely in terms of a change on the constitutive parameters of the background medium. In this dual formulation, Maxwell’s equations are not changed, and the background medium becomes a dispersive and anisotropic medium. The dispersive and anisotropic properties of the artificial background media is such that they act to compensate for the dispersive and anisotropic properties of wave propagation caused by the discrete grid.
Figure 2.16: Comparison of the normalized phase velocity at different angles of Butterworth AO-FDTD scheme. The center frequency is such that $q_c = 0.09$

Figure 2.17: Simulated results of a first-order differentiated Gaussian pulse propagating at $\theta = 90^\circ$ and $\phi = 0^\circ$
CHAPTER 3

DISPERSION-RELATION-PRESERVATION FDTD SCHEMES

3.1 Introduction

In recent years, several techniques have been developed to reduce numerical dispersion error in finite-difference solutions. Among these, the Cole’s non-standard finite-difference (NSFD) \cite{9} and Forgy’s scheme \cite{15} have particularly low dispersion error for narrow-band simulations \cite{59}. Moreover, various high order schemes based on traditional Yee’s staggered grids have been proposed for broadband simulations. Traditional fourth order of accuracy in time and fourth order of accuracy in space schemes [so called (4,4) schemes], including the ones proposed independently by Fang \cite{12} and Deveze \cite{11} are particularly attractive for such problems \cite{23},\cite{58},\cite{59}.

Since the ability to solve broadband problems is a main motivation for using time domain methods, it is important to develop FDTD algorithms with reduced dispersion error in a broadband range. On this aspect, it is important to notice that, for many practical FDTD simulations, the minimization of the local and accumulated dispersion error over a finite range of frequencies is often more relevant than the theoretical order of accuracy of the FDTD scheme itself, where order of accuracy refers to the (truncation) error of the scheme as the grid spacing goes to zero (very
low frequency or long wavelength limit). This is because of two (interconnected) main reasons: (1) The very low frequency limit is often not that relevant in practice, unless in the case of quasi-static simulations or in simulations where some kind of very fine adaptive grid refinement is employed. Indeed, in the majority of FDTD simulations of wave propagation problems, the spectral energy is mostly within wavelengths on the order of $10^4 \Delta - 10^2 \Delta$, where $\Delta$ is the grid discretization size. (2) FDTD simulations involve computational domains of fixed physical size. In the very low frequency limit, the computational domain is electrically smaller and, therefore, the accumulated phase error (which grows linearly with the electric size of the domain) is less of a problem.

One relevant question then arises on how to construct optimal FDTD algorithms with dispersion-relation-preserving (DRP) properties on a given (possibly wide) frequency band. Because of its cumulative effect, high frequency components are subject to a larger cumulative phase error than low frequency components. Thus it is also of interest to investigate the possibility of constructing FDTD algorithms such that the (local) dispersion error incurred is lower for high frequencies than for low frequencies (in traditional schemes usually the opposite is true).

In this chapter, we shall describe a methodology to develop DRP schemes for both 2-D and 3-D FDTD algorithms with these observations in mind. Although our main motivation is for time-domain broadband problems, the proposed methodology is suited to minimize the dispersion error at specific frequencies.
3.2 Two-Dimensional DPR-FDTD Algorithm

3.2.1 Methodology

Because the dispersion error is in general a function of both frequency, propagation angle, and the particular Courant-Friedrichs-Lewy (CFL) number, several definitions of minimum dispersion error [45], [23] are possible. We define it here in the minimax (minimum maximum) sense, i.e., for a given CFL number, the maximum dispersion error for all angles is minimized up to a certain frequency. The DRP procedure here consists of three main steps: 1) For a scheme with given “order of accuracy in space”, the dispersion error is expanded in a Fourier series in terms of the propagation angle, and analytical expressions for the DRP coefficients are subsequently derived as a function of frequency. 2) These analytical expressions are then cast into a form implementable in the fully-discrete problem (FDTD method) by using polynomial expansions. 3) If necessary, filtering schemes (maximally flat or Chebyshev) are used to fine-tune the DRP 2-D FDTD coefficients for a pre-assigned frequency range.

The first step consists in considering a scheme with a given “order of accuracy in space”. Higher order of accuracy in time is introduced in the second step as we expand the analytical solution in series. The term “order of accuracy in space” is borrowed here from traditional higher order schemes employing Taylor expansions only [12], [11]. Since this work treats the problem from a different standpoint, this term does not retain its original meaning. The term order of accuracy should be considered here as referring to a class of spatial stencil sizes and not necessarily to the precise order of the truncation error as the discretization cell size goes to zero. As we will see, the major difference between the DRP higher order schemes derived
here and traditional higher order schemes resides in the coefficients and not on the particular FDTD stencil.

Non-filtered, DRP (2,2) Schemes

A traditional leap-frog scheme with second order of accuracy in space can be written in general as:

\[
E_{x_{m+\frac{1}{2},n}}^{l+1} = E_{x_{m+\frac{1}{2},n}}^{l} + \frac{\Delta t}{\epsilon \Delta y} \Gamma_x (H_{z_{m+\frac{1}{2},n+\frac{1}{2}}}^{l+\frac{1}{2}} - H_{z_{m+\frac{1}{2},n-\frac{1}{2}}}^{l+\frac{1}{2}})
\] (3.1)

\[
E_{y_{m,n+\frac{1}{2}}}^{l+1} = E_{y_{m,n+\frac{1}{2}}}^{l} + \frac{\Delta t}{\epsilon \Delta x} \Gamma_y (H_{z_{m-\frac{1}{2},n+\frac{1}{2}}}^{l+\frac{1}{2}} - H_{z_{m+\frac{1}{2},n+\frac{1}{2}}}^{l+\frac{1}{2}})
\] (3.2)

\[
H_{z_{m+\frac{1}{2},n+\frac{1}{2}}}^{l+\frac{1}{2}} = H_{z_{m+\frac{1}{2},n+\frac{1}{2}}}^{l-\frac{1}{2}} + \frac{\Delta t}{\mu \Delta y} \Gamma_x (E_{x_{m+\frac{1}{2},n+1}}^{l} - E_{x_{m+\frac{1}{2},n}}^{l}) - \frac{\Delta t}{\mu \Delta x} \Gamma_y (E_{y_{m,n+\frac{1}{2}}}^{l} - E_{y_{m+1,n+\frac{1}{2}}}^{l})
\] (3.3)

where the subscripts denote the spatial location and the superscripts denote the time step. We treat \(\Gamma_x\) and \(\Gamma_y\) as unknown coefficients for the moment. The \(E\) and \(H\) fields can be expanded into a discrete set of Fourier modes. For each mode

\[
E_{x_{m+\frac{1}{2},n}}^{l} = \mathcal{E}_x e^{j\omega \Delta t} e^{-j[k_x (m+\frac{1}{2})\Delta x + k_y n\Delta y]} \] (3.4)

\[
E_{y_{m,n+\frac{1}{2}}}^{l} = \mathcal{E}_y e^{j\omega \Delta t} e^{-j[k_x m\Delta x + k_y (n+\frac{1}{2})\Delta y]} \] (3.5)

\[
H_{z_{m+\frac{1}{2},n+\frac{1}{2}}}^{l} = \mathcal{H}_z e^{j\omega (l+\frac{1}{2}) \Delta t} e^{-j[k_x (m+\frac{1}{2})\Delta x + k_y (n+\frac{1}{2})\Delta y]} \] (3.6)

Substituting (3.4)-(3.6) into (3.1)-(3.3) and noticing that \(\mathcal{E}_x = -\mathcal{E}\sin(\theta)\), \(\mathcal{E}_y = \mathcal{E}\cos(\theta)\) and \(\mathcal{H}_z = \mathcal{H}\), we have

\[
\sin\left(\frac{\omega \Delta t}{2}\right)\mathcal{E}\sin(\theta) = \frac{\Delta t}{\epsilon \Delta y} \Gamma_x \mathcal{H}\sin\left[\frac{k \sin(\theta) \Delta y}{2}\right]
\] (3.7)

\[
\sin\left(\frac{\omega \Delta t}{2}\right)\mathcal{E}\cos(\theta) = \frac{\Delta t}{\epsilon \Delta x} \Gamma_y \mathcal{H}\sin\left[\frac{k \cos(\theta) \Delta x}{2}\right]
\] (3.8)

\[
\sin\left(\frac{\omega \Delta t}{2}\right)\mathcal{H} = \frac{\Delta t}{\mu \Delta y} \Gamma_x \mathcal{E}\sin\left[\frac{k \sin(\theta) \Delta y}{2}\right] \sin(\theta) + \frac{\Delta t}{\mu \Delta x} \Gamma_y \mathcal{E}\sin\left[\frac{k \cos(\theta) \Delta x}{2}\right] \cos(\theta)
\] (3.9)
where \( k_x = k \cos(\theta), k_y = k \sin(\theta) \). The numerical dispersion relationship can be derived from (3.7)-(3.9) as

\[
\frac{1}{(v_p \Delta t)^2} \sin^2 \left( \frac{\omega \Delta t}{2} \right) = \frac{\Gamma_x^2}{\Delta y^2} \sin^2 \left[ \frac{k \sin(\theta) \Delta y}{2} \right] + \frac{\Gamma_y^2}{\Delta x^2} \sin^2 \left[ \frac{k \cos(\theta) \Delta x}{2} \right]
\]  

(3.10)

where \( v_p = \sqrt{\frac{\mu}{\epsilon}} \). For a given set of coefficients \( \Gamma_x \) and \( \Gamma_y \) (in Yee’s scheme they are equal to unity), the above equation is traditionally used to analyze the discrete dispersion in the FDTD grid. The amount by which the discrete dispersion relation deviates from the continuum limit gives the local dispersion error. In this work, we shall adopt the reverse standpoint. That is, we shall enforce the exact relation between frequency and wavenumber, viz. \( \omega = v_p k \), for (3.10), and then solve for \( \Gamma_x \) and \( \Gamma_y \) as the unknowns (in an approximate sense to be clear later on). Ideally, the exact solutions for \( \Gamma_x \) and \( \Gamma_y \) should depend on both the frequency and propagation angle. We shall first expand the dispersion error in terms of a Fourier series in terms of the angular variable \( \theta \) and enforce coefficients on the series to be zero (the number of coefficients made equal to zero give the “order” of the method). In this manner, the maximum dispersion error for all angles is minimized simultaneously, and \( \Gamma_x \) and \( \Gamma_y \) become a function of frequency only. This latter property allows for incorporation of the DRP coefficients in FDTD algorithms after a polynomial expansion.

A simpler but equivalent way of solving for \( \Gamma_x \) and \( \Gamma_y \) in (3.10) is to enforce \( \omega = v_p k \) into (3.7) and (3.8) and solve for \( \Gamma_x \) and \( \Gamma_y \), respectively\(^2\). We start by defining an error functional proportional to the difference between l.h.s. and r.h.s. of (3.7)

\[
\delta_2(\Gamma_x, \theta) = \frac{\sqrt{2}}{\chi_y} \sin \left( \frac{\pi q_y \chi_y}{\sqrt{2}} \right) \sin(\theta) - \Gamma_x \sin \left[ \pi q_y \sin(\theta) \right]
\]  

(3.11)

\(^2\)Note that (from the symmetry of the problem) by letting \( \Delta x \rightarrow \Delta y \) and \( \theta \rightarrow \theta + \pi/2 \), (3.8) reduces to (3.7). In particular if \( \Delta x = \Delta y \), we should of course expect \( \Gamma_x = \Gamma_y \) after these coefficients are reduced to functions of frequency only.
where \( \chi_y = \sqrt{2}v_p\Delta t/\Delta y \), \( q_y = \Delta y/\lambda \). Solving (3.7) with \( \omega = v_p k \) is therefore equivalent to letting \( \delta_2(\Gamma_x, \theta) = 0 \) and enforcing \( \mathcal{E}/\eta = \mathcal{H} \).

We expand \( \delta_2(\Gamma_x, \theta) \) in a Fourier series in terms of \( \theta \) and use the following identities

\[
\int_0^{2\pi} \sin(x \sin(\theta)) \cos(n\theta) d\theta = 0 \\
\int_0^{2\pi} \sin(x \sin(\theta)) \sin(n\theta) d\theta = 2\pi J_n(x)
\]

for odd \( n \) and

\[
\int_0^{2\pi} \sin(x \sin(\theta)) \cos(n\theta) d\theta = 0 \\
\int_0^{2\pi} \sin(x \sin(\theta)) \sin(n\theta) d\theta = 0
\]

for even \( n \), where \( J_n(x) \) is the \( n \)-th order first kind Bessel function, to force the first non-zero (dominant) term of the series to be zero. This leads to

\[
\Gamma_x = \gamma_x^{(\infty)} \equiv \frac{\sqrt{2}}{\chi_y} \sin(\frac{\pi q_y \chi_y}{\sqrt{2}})/\left[2J_1(\pi q_y)\right] \\
\tag{3.12}
\]

Substituting (3.12) in (3.11), the residual error \( \delta_{2e} \equiv \delta_2(\Gamma_x, \theta) - \delta_2(\gamma_x^{(\infty)}, \theta) \) is given by

\[
\delta_{2e} = \Gamma_x J_3(\pi q_y) \sin(3\theta) + \Gamma_x J_5(\pi q_y) \sin(5\theta) + ...
\]

Note that, since \( |J_n(x)| < 1/(2^n \cdot n!) \) for \( |x| < 1 \), (the inequality \( \pi q_y < 1 \) is satisfied in practical FDTD simulations with \( q_y < 0.1 \)), the coefficients in this series decrease very rapidly, as expected. The magnitude of each coefficient in the series represents the maximum dispersion error from the corresponding angular mode (either at \( \sin(n\theta) = 1 \) or at \( \cos(n\theta) = 1 \)).

It is clear that the solution \( \gamma_x^{(\infty)} \) given in (3.12) cannot be implemented in a time domain method. Because of this, we expand \( \gamma_x^{(\infty)} \) in a Taylor series around \( q_y = 0 \) and retain the lowest order terms, i.e., \( \gamma_x^{(\infty)} = \gamma_x^{(2)} + O(q_y^4) \) with

\[
\gamma_x^{(2)} = 1 + \frac{1}{96} 4\pi^2 q_y^2 (3 - 2\chi_y^2) \\
\tag{3.13}
\]

39
The above can be easily transformed back to time domain through $\omega^2 \rightarrow -\partial^2 / \partial t^2$. However, if straightforward time discretization schemes are employed directly on the resulting equations (with third order time derivatives), the update becomes unconditionally unstable [11]. Alternatively, the second order time derivative can be further cast as a combination of spatial derivatives as $v_p^2 \nabla^2$ (Helmholtz equation) and discretized as such (modified equation method [8]). This latter transformation is valid for staggered grids as long as $\epsilon$ is uniform in the local stencil. In this manner, (3.13) becomes

$$
\gamma^{(2)}_x = 1 - \frac{(3 - 2\gamma^2_x) \Delta^2 y \nabla^2}{96}
$$

(3.14)

Following a similar procedure for $\Gamma_y$, we find

$$
\gamma^{(2)}_y = 1 - \frac{(3 - 2\gamma^2_x) \Delta^2 x \nabla^2}{96}
$$

(3.15)

where $\chi_x = \sqrt{2} v_p \Delta t / \Delta x$. If only the first order terms in (3.14) and (3.15) are taken, the Yee’s scheme is recovered. Note that the second terms in (3.14) and (3.15) are the analogous to third order time derivative terms in traditional schemes with fourth order of accuracy in time [11].

To investigate the resulting numerical dispersion, we define a dispersion error functional $\delta_2(\Gamma_x, \Gamma_y, \theta)$ proportional to the difference between the l.h.s. and the r.h.s. of (3.10), i.e.,

$$
\delta_2(\Gamma_x, \Gamma_y, \theta) = \left[ \frac{\sqrt{2}}{\chi} \sin \left( \frac{\pi q \chi}{\sqrt{2}} \right) \right]^2 - h^2 \left[ \frac{\Gamma_x^2 \sin^2 \left( \frac{k \sin(\theta) \Delta y}{2} \right)}{\Delta y^2} + \frac{\Gamma_y^2 \sin^2 \left( \frac{k \cos(\theta) \Delta x}{2} \right)}{\Delta x^2} \right]
$$

(3.16)

where $\chi = \sqrt{2} v_p \Delta t / h$, $q = h / \lambda$, and $h = min(\Delta x, \Delta y)$. By substituting $\Gamma_x = \gamma_x^{(\infty)}$ and $\Gamma_y = \gamma_y^{(\infty)}$ into (3.16), we obtain a limit value for the error $\delta_2^{(\infty)}(\theta) = \delta_2(\gamma_x^{(\infty)}, \gamma_y^{(\infty)}, \theta)$. In practical time domain simulations employing (3.14) and (3.15),
we have an error $\tilde{\delta}_2^{(2)}(\theta) = \tilde{\delta}_2(\gamma_x^{(2)}, \gamma_y^{(2)}, \theta)$ instead, which inevitably introduces additional errors at high frequencies. The function $\tilde{\delta}_2^{(\infty)}(\theta)$ therefore serves as a inferior theoretical limit when an infinite Taylor expansion is considered. However, as we shall see in the next subsection, by expanding $\Gamma_x$ and $\Gamma_y$ in a different basis (e.g., Chebychev polynomials) this limit can indeed be overcome over some finite, pre-assigned frequency range.

Figure 3.1: Comparison of the maximum value of the dispersion error $\tilde{\delta}_2(\Gamma_x, \Gamma_y, \theta)$ for all angles when using analytical solution, $(\Gamma_x, \Gamma_y) = (\gamma_x^{(\infty)}, \gamma_y^{(\infty)})$, second order approximation $(\Gamma_x, \Gamma_y) = (\gamma_x^{(2)}, \gamma_y^{(2)})$ and Yee’s scheme $(\Gamma_x, \Gamma_y) = (1, 1)$.

Fig.3.1 shows the maximum value for all angles of $\tilde{\delta}_2^{(\infty)}(\theta)$ (analytical), $\tilde{\delta}_2^{(2)}(\theta)$ (second order), and $\tilde{\delta}_2(1, 1, \theta)$ (Yee’s scheme), as a function of the number of wavelengths per cell (or, equivalently, the frequency). In these plots, $\Delta x = \Delta y$ and the CFL number $\chi = 3/4 \left( \chi_x = \chi_y = \chi \right)$. From this figure, we observe that, for all the

41
frequency range such that \( q_y \leq 0.1 \), the second order approximation already gives results almost as accurate as using the analytical expression (3.12), and starts to deviate only slightly at high frequencies close to \( q_y = 0.1 \).

**Non-filtered, DRP (2,4) Schemes**

Traditional FDTD schemes with fourth order of accuracy in space employ larger stencils for the \( E_x \) and \( E_y \) update equations. In this case, the \( E_x \) update can be rewritten as follows (compare with (3.1))

\[
E_{x_{l+1}}^{m+\frac{1}{2},n} = E_x^l + \frac{\Delta t}{\epsilon \Delta y} \Gamma_{x1}(H_{z_{l+1}}^{m+\frac{1}{2},n+\frac{1}{2}} - H_{z_{l+1}}^{m+\frac{1}{2},n-\frac{1}{2}}) + \frac{\Delta t}{\epsilon \Delta y} \Gamma_{x2}(H_{z_{l+1}}^{m+\frac{1}{2},n+\frac{3}{2}} - H_{z_{l+1}}^{m+\frac{1}{2},n-\frac{3}{2}})
\]  

(3.17)

We proceed as before and \( \Gamma_{x1}, \Gamma_{x2} \) can be solved from

\[
\sin\left(\frac{\nu_{pl}k\Delta t}{2}\right)E \sin(\theta) = \frac{\Delta t}{\epsilon \Delta y} \Gamma_{x1} \mathcal{H} \sin\left[\frac{k \sin(\theta) \Delta y}{2}\right] + \frac{\Delta t}{\epsilon \Delta y} \Gamma_{x2} \mathcal{H} \sin\left[\frac{3k \sin(\theta) \Delta y}{2}\right]
\]

(3.18)

To obtain \( \Gamma_{x1} \) and \( \Gamma_{x2} \), we define an error functional in terms of the difference between the l.h.s. and r.h.s. of (3.18)

\[
\delta_4(\Gamma_{x1}, \Gamma_{x2}, \theta) = \sqrt{2} \sin\left(\frac{\pi q_y \chi y}{\sqrt{2}}\right) \sin(\theta) - \Gamma_{x1} \sin[\pi q_y \sin(\theta)] - \Gamma_{x2} \sin[3\pi q_y \sin(\theta)]
\]

(3.19)

Similarly as before, we expand \( \delta_4(\Gamma_{x1}, \Gamma_{x2}, \theta) \) in a Fourier series in terms of \( \theta \). However, since there are two unknowns \( \Gamma_{x1} \) and \( \Gamma_{x2} \), we now may force the first two non-zero terms of the series to be zero (\( \sin(\theta) \) and \( \sin(3\theta) \) terms). The solutions are given by

\[
\Gamma_{x1} = \gamma_{x1}^{(\infty)} \equiv \frac{\sqrt{2} \sin\left(\frac{\pi q_y \chi y}{\sqrt{2}}\right) J_3(3\pi q_y)}{2\chi y[J_1(\pi q_y)J_3(3\pi q_y) - J_1(3\pi q_y)J_3(\pi q_y)]}
\]

(3.20)

\[
\Gamma_{x2} = \gamma_{x2}^{(\infty)} \equiv \frac{\sqrt{2} \sin\left(\frac{\pi q_y \chi y}{\sqrt{2}}\right) J_3(\pi q_y)}{2\chi y[J_1(3\pi q_y)J_3(\pi q_y) - J_1(\pi q_y)J_3(3\pi q_y)]}
\]

(3.21)
With these coefficients, the residual error defined as $\delta_{4e} \equiv \delta_4(\Gamma_{x_1}, \Gamma_{x_2}, \theta) - \delta_4(\gamma^{(\infty)}_{x_1}, \gamma^{(\infty)}_{x_2}, \theta)$ becomes

$$\delta_{4e} = \Gamma_{x_1} J_5(\pi q_y) \sin(5\theta) + \Gamma_{x_2} J_5(3\pi q_y) \sin(5\theta) + ...$$

For incorporation into the FDTD update, $\gamma^{(\infty)}_{x_1}$ and $\gamma^{(\infty)}_{x_2}$ are approximated as

$$\gamma^{(\infty)}_{x_1} \approx \gamma^{(2)}_{x_1} = \gamma_{x_11} - \gamma_{x_12} \Delta y^2 \nabla^2$$

$$\gamma^{(\infty)}_{x_2} \approx \gamma^{(2)}_{x_2} = \gamma_{x_21} - \gamma_{x_22} \Delta y^2 \nabla^2$$

(3.22)

(3.23)

with

$$\gamma^{(2)}_{x_11} = \frac{9}{8}$$

$$\gamma^{(2)}_{x_12} = -\frac{3(3 - 4\chi^2_y)}{512}$$

$$\gamma^{(2)}_{x_21} = -\frac{1}{24}$$

$$\gamma^{(2)}_{x_22} = -\frac{(27 - 4\chi^2_y)}{4608}$$

The lowest order terms in (3.22) and (3.23) recover Fang’s (2,4) FDTD scheme [12].

Similarly, $\gamma^{(2)}_{y_1}$ and $\gamma^{(2)}_{y_2}$ write as

$$\gamma^{(2)}_{y_1} = \gamma^{(2)}_{y_11} - \gamma^{(2)}_{y_12} \Delta x^2 \nabla^2$$

(3.24)

$$\gamma^{(2)}_{y_2} = \gamma^{(2)}_{y_21} - \gamma^{(2)}_{y_22} \Delta x^2 \nabla^2$$

(3.25)

with coefficients $\gamma^{(2)}_{y_{ij}}$ written as $\gamma^{(2)}_{x_{ij}}$ above with $\chi_y$ replaced by $\chi_x$.

The implementation of $\gamma^{(2)}_{x_1}$, $\gamma^{(2)}_{x_2}$, $\gamma^{(2)}_{y_1}$ and $\gamma^{(2)}_{y_2}$ in staggered grids are straightforward. For example, the fully discrete form of (3.17) is

$$E^{t+\frac{1}{2}, n}_{x_{m+\frac{1}{2}, n}} = E^t_{x_{m+\frac{1}{2}, n}} + \frac{\Delta t}{\epsilon \Delta y} \left[ (\gamma^{(2)}_{x_11} + \gamma^{(2)}_{x_12}(3 + 2 \frac{\Delta y^2}{\Delta x^2}) - \gamma^{(2)}_{x_22}) \right]$$

$$+ \left[ \gamma^{(2)}_{x_21} + 2 \gamma^{(2)}_{x_22}(1 + \frac{\Delta y^2}{\Delta x^2}) - \gamma^{(2)}_{x_12} \right]$$

43
\[
(H_{z_{m+\frac{1}{2},n+\frac{1}{2}}} - H_{z_{m+\frac{1}{2},n-\frac{1}{2}}}) - \gamma_{x12} \frac{\Delta y^2}{\Delta x^2} \\
(H_{z_{m-\frac{1}{2},n+\frac{1}{2}}} + H_{z_{m+\frac{1}{2},n+\frac{1}{2}}} - H_{z_{m-\frac{1}{2},n-\frac{1}{2}}} - H_{z_{m+\frac{1}{2},n-\frac{1}{2}}}) \\
- \gamma_{x22}(H_{z_{m+\frac{1}{2},n+\frac{1}{2}}} - H_{z_{m+\frac{1}{2},n-\frac{1}{2}}}) - \gamma_{x22} \frac{\Delta y^2}{\Delta x^2} \\
(H_{z_{m-\frac{1}{2},n+\frac{1}{2}}} + H_{z_{m+\frac{1}{2},n+\frac{1}{2}}} - H_{z_{m-\frac{1}{2},n-\frac{1}{2}}} - H_{z_{m+\frac{1}{2},n-\frac{1}{2}}}) \}
\]

The stability condition can be derived in a standard way [63] and the result is

\[
\Delta t \leq \frac{1}{v_p \sqrt{\frac{D_x^2}{\Delta y^2} + \frac{D_y^2}{\Delta x^2}}} \tag{3.26}
\]

where

\[
D_x = 2 \frac{\Delta y^2}{\Delta x^2} [\gamma_{x12} \sin\left(\frac{k_y \Delta y}{2}\right) + \gamma_{x22} \sin\left(\frac{3k_y \Delta y}{2}\right)] \cos(k_x \Delta x) \\
+ \gamma_{x22} \sin\left(\frac{5k_y \Delta y}{2}\right) - \left[\gamma_{x11} + \gamma_{x12} (3 + 2 \frac{\Delta y^2}{\Delta x^2}) - \gamma_{x22} \right] \sin\left(\frac{k_y \Delta y}{2}\right) \\
- \left[\gamma_{x21} + 2 \gamma_{x22} (1 + \frac{\Delta y^2}{\Delta x^2}) - \gamma_{x12} \right] \sin\left(\frac{3k_y \Delta y}{2}\right) \tag{3.27}
\]

and similarly for \(D_y\). For \(\Delta x = \Delta y\), (3.26) becomes

\[
\Delta t \leq \frac{h \chi}{v_p \sqrt{2}} = \frac{h}{v_p \sqrt{2|D_{x,y}|_{\text{max}}}}
\]

where \(|\tilde{D}_{x,y}|_{\text{max}}\) denotes the maximum possible module of \(D_x\) and \(D_y\). The minimum maximum CFL number \(\chi\) with guaranteed stability is found to be 48/65. Note that this is a conservative bound. In reality, we have found \(\chi = 0.75\) to produce stable updates in our tests.

To estimate the optimization quality, an error functional \(\delta_4(\Gamma_{x1}, \Gamma_{x2}, \Gamma_{y1}, \Gamma_{y2}, \theta)\) is constructed in an analogous way as the error functional in (3.16). Fig. 3.2 shows the comparison of the largest \(\delta_4\) (among all angles) by using different schemes when \(\Delta x = \Delta y\) and \(\chi = 48/65\). In this figure, the analytical result refers to the (ideal)
choice \((\Gamma_{x1}, \Gamma_{x2}, \Gamma_{y1}, \Gamma_{y2}) = (\gamma_{x1}^{(\infty)}, \gamma_{x2}^{(\infty)}, \gamma_{y1}^{(\infty)}, \gamma_{y2}^{(\infty)})\), whereas the second order result refers to the (implementable) approximation for the coefficients given by (3.22)-(3.25),

\((\Gamma_{x1}, \Gamma_{x2}, \Gamma_{y1}, \Gamma_{y2}) = (\gamma_{x1}^{(2)}, \gamma_{x2}^{(2)}, \gamma_{y1}^{(2)}, \gamma_{y2}^{(2)}).\) Maximally flat and Chebyshev refer to (implementable) approximations using filtering schemes to be detailed in the next section.

**Filtering**

Instead of possibly using even higher order terms in the Taylor series or other sophisticated (and more costly) time integration schemes, it is possible to improve the results of the previous section (while maintaining the same spatial stencil sizes and computational cost) by using filters adjusted to some pre-assigned, finite frequency range.

The filters are designed to obtain both \(\Gamma_x\) and \(\Gamma_y\) (second order stencil), or \(\Gamma_{x1}, \Gamma_{x2}, \Gamma_{y1}\) and \(\Gamma_{y2}\) (fourth order stencil). The procedure is essentially the same for all coefficients and therefore, we will describe only the \(\Gamma_{x1}\) case in detail.

**Maximally Flat (Butterworth) Filters**

Expanding (3.20) in a Taylor series, and approximating (3.20) in a finite series as in (3.22) noting that \(\Delta y^2 \nabla^2 = -4\pi^2 q_y^2\), we have

\[
\gamma_{x1}^{(\infty)} - \gamma_{x1}^{(m)} = \left(\frac{9}{8} - \gamma_{x11}^{(m)}\right) - \left[\frac{3(4\chi_y^2 - 3)\pi^2}{128} + 4\pi^2 \gamma_{x12}^{(m)}\right] q_y^2 + \frac{3\pi^4 \chi_y^2 (2\chi_y^2 - 5)}{2560} q_y^4 + \mathcal{O}(q_y^6)
\]

(3.28)

The superscript \((m)\) in \(\gamma_{x11}^{(m)}\) and \(\gamma_{x12}^{(m)}\) above refer to maximally flat coefficients. Previously, we have simply chosen \(\gamma_{x11}^{(m)} = \gamma_{x11}^{(2)}\) and \(\gamma_{x12}^{(m)} = \gamma_{x12}^{(2)}\) so as to make the first and second terms of the right hand side of the above identically zero (and obtain a \(\mathcal{O}(q_y^4)\) truncation error). These coefficients are treated as unknowns for the moment.
The above equation can be rewritten as

\[ \gamma_{x1}^{(\infty)} - \gamma_{x1}^{(m)} = (d_0, 0, d_2, 0, d_4) \cdot V^T + \mathcal{O}(q_y^6) \quad (3.29) \]

where

\[ d_0 = \frac{9}{8} - \gamma_{x11}^{(m)} \]
\[ d_2 = -\left[ \frac{3(4\chi_y^2 - 3)^2}{128} + 4\pi^2 \gamma_{x12}^{(m)} \right] \]
\[ d_4 = \frac{3\pi^4 \chi_y^2 (2\chi_y^2 - 5)}{2560} \]
\[ V = (1, q_y, q_y^2, q_y^3, q_y^4) \]

By expanding the above in a new basis \( \{ (q_y - q_y^c)^n \} \), where \( q_y^c \) refers to a center frequency of interest, the following relationship holds

\[ Q^T = A \cdot V^T \]

with

\[ Q = (1, (q - q_y^c), (q - q_y^c)^2, (q - q_y^c)^3, (q - q_y^c)^4) \]
\[ A = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 \\
-q_y^c & 1 & 0 & 0 & 0 \\
q_y^c & -2q_y^c & 1 & 0 & 0 \\
-q_y^c & 3q_y^c & -3q_y^c & 1 & 0 \\
q_y^c & -4q_y^c & 6q_y^c & -4q_y^c & 1
\end{pmatrix} \quad (3.31) \]

and, therefore, (3.29) can be rewritten as

\[ \gamma_{x1}^{(\infty)} - \gamma_{x1}^{(m)} = (d_0, 0, d_2, 0, d_4) \cdot A^{-1} \cdot Q^T + \mathcal{O}(q_y^6) \]
\[ = (\tilde{d}_0, \tilde{d}_1, \tilde{d}_2, \tilde{d}_3, \tilde{d}_4) \cdot Q^T + \mathcal{O}(q_y^6) \quad (3.32) \]

where

\[ \tilde{d}_0 = \left\{ 2880 - 2560\gamma_{x11}^{(m)} + 3\pi^4 q_y^4 \chi_y^2 (2\chi_y^2 - 5) \right. \]
\[ -20q_y^c \left[ 512\pi^2 \gamma_{x12}^{(m)} + 3\pi^2 (4\chi_y^2 - 3) \right] \}/2560 \]
\[ \hat{d}_1 = q_y^c \pi^2 \{90 + 3 \chi^2 q_y^c (2 \chi^2 - 5) - 40\} - 5120 \gamma_{x12}^{(m)} \}/640 \]
\[ \hat{d}_2 = \pi^2 \{90 + 3 \chi^2 q_y^c (2 \chi^2 - 5) - 5120 \gamma_{x12}^{(m)}\}/1280 \]
\[ \hat{d}_3 = 3 \pi^4 \chi^2 q_y^c (2 \chi^2 - 5)/640 \]
\[ \hat{d}_4 = 3 \pi^4 \chi^2 (2 \chi^2 - 5)/2560 \]

To solve for \( \gamma_{x11}^{(m)} \) and \( \gamma_{x12}^{(m)} \), we force \( \hat{d}_0 \) and \( \hat{d}_1 \) to be zero and obtain
\[ \gamma_{x11}^{(m)} = [2880 + 3 \pi^4 \chi^2 q_y^c (5 - 2 \chi^2)]/2560 \quad (3.33) \]
\[ \gamma_{x12}^{(m)} = 3 \{30 + \chi^2 q_y^c (2 \chi^2 - 5)\}/5120 \quad (3.34) \]

which are functions of \( q_y^c \). Note that, if we let \( q_y^c = 0 \), we recover \( \gamma_{x11}^{(m)} = \gamma_{x11}^{(2)} \) and \( \gamma_{x12}^{(m)} = \gamma_{x12}^{(2)} \).

In this manner,
\[ \gamma_{x1}^{(\infty)} - \gamma_{x1}^{(m)} = [\hat{d}_2 (q_y - q_y^c)^2 + \hat{d}_3 (q_y - q_y^c)^3 + \hat{d}_4 (q_y - q_y^c)^4] + \mathcal{O}(q_y^6) \]

At the center frequency, \( q_y = q_y^c \) and \( \delta = \mathcal{O}(q_y^6) \). The remainder corresponds to a \( \mathcal{O}(q_y^6) \) error. Around the center frequency, the error is dominated by the \( \hat{d}_2 (q_y - q_y^c)^2 \) term. In the above, we have illustrated the derivation of \( \gamma_{x1}^{(m)} \) by using a \( p = 4 \) order polynomial for the Taylor expansion in (3.29). Theoretically, we can increase indefinitely the polynomial order \( p \) at the (one-time) cost of inverting a larger \( A \) and hence make the response as close to an exact second-order maximally flat filter response as desired. This has the (one-time) cost of inverting an increasingly larger matrix \( A \). In practice, we observe a fast convergence to the exact filter response since only negligible improvements are obtained for \( p \geq 6 \). We employ \( p = 10 \) throughout our numerical simulations. Also in practice, the design frequency \( q_y^c \) may be chosen as the highest frequency of interest.
Fig. 3.2 includes the $\bar{\delta}_4$ error of the DRP scheme with maximally flat filters designed at $q^c = 0.1$ ($\Delta x = \Delta y$). From this figure, we observe that near the central frequency ($q^c = 0.1$), the $\bar{\delta}_4$ using a second-order (with $p = 10$) maximally flat filter is almost indistinguishable from the one using the analytical solutions (3.20) and (3.21).

Chebyshev Filters

We may choose $\{T_n[(q_y - q^c_y)/\Delta q_y]\}$ as a new basis expansion in (3.30) instead, where $T_n(x)$ is the $n$-th order first-kind Chebyshev polynomial, and $\Delta q_y$ corresponds to the frequency band of interest. Still using $p = 4$ as an example, Eqs. (3.30) and (3.31) now become

$$Q = (1, \frac{q_y - q^c_y}{\Delta q_y}, T_2(\frac{q_y - q^c_y}{\Delta q_y}), T_3(\frac{q_y - q^c_y}{\Delta q_y}), T_4(\frac{q_y - q^c_y}{\Delta q_y}))$$
\[
A = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 \\
-\frac{q_y^c}{\Delta q_y} & 1 & 0 & 0 & 0 \\
\frac{2q_y^c - \Delta q_y^2}{\Delta q_y^2} & -\frac{4q_y^c}{\Delta q_y^2} & \frac{2}{\Delta q_y^2} & 0 & 0 \\
\frac{3q_y^c \Delta q_y^2 - 4q_y^6}{\Delta q_y^6} & \frac{12q_y^c - 3\Delta q_y^2}{\Delta q_y^4} & -\frac{12q_y^c}{\Delta q_y^4} & \frac{4}{\Delta q_y^4} & 0 \\
\frac{8q_y^c - 8q_y^c \Delta q_y^2 + \Delta q_y^4}{\Delta q_y^4} & \frac{16q_y^c \Delta q_y^2 - 32q_y^6}{\Delta q_y^4} & \frac{48q_y^c - 8 \Delta q_y^2}{\Delta q_y^4} & \frac{32q_y^c}{\Delta q_y^4} & \frac{8}{\Delta q_y^4}
\end{pmatrix}
\]

Following the same procedure as maximally flat filter case, we can solve for the coefficients and obtain \(\gamma_{x1}^{(c)}\) and \(\gamma_{x2}^{(c)}\) of an approximate second-order Chebyshev filter.

Here

\[
\gamma_{x1}^{(\infty)} - \gamma_{x1}^{(c)} = \hat{d}_2 T_2[(q_y - q_y^c)/\Delta q_y] + \hat{d}_3 T_3[(q_y - q_y^c)/\Delta q_y] + \hat{d}_4 T_4[(q_y - q_y^c)/\Delta q_y] + \mathcal{O}(q_y^6)
\]

where \(\hat{d}_2, \hat{d}_3\) and \(\hat{d}_4\) denote the coefficients of the corresponding Chebyshev polynomials. Since \(T_2[(q_y - q_y^c)/\Delta q_y]\) takes the minimums at \(q_y^c\) and \(T_3[(q_y - q_y^c)/\Delta q_y]\) is zero at \(q_y^c\), we can obtain an even better accuracy than using the analytical solutions given by (3.20) and (3.21) at a particular design frequency. Fig. 3.2 (with \(q_y = q_x = q\)) clearly shows that \(\gamma_{x1}^{(c)}\) can be smaller than employing (3.20) and (3.21) around the design frequency. For the Chebyshev filter approximation, we have also set \(p = 10\).

Since the Chebyshev filter is set to be of second order, larger ripples are expected to occur in the passband \(q^c - \Delta q \leq q \leq q^c + \Delta q\) if a larger \(\Delta q\) is used [48]. In usual filter design, large ripples are undesired, but in our context, \(\Delta q\) represents an extra degree of freedom that can be explored to reduce the accumulated phase error. This will be illustrated in the next section.

\(\Delta q = 0.02\) is used here as an example. As shown later, \(\Delta q\) can be fine-tuned for a better performance.
3.2.2 Numerical Results

We compare the DRP (2,2) and (2,4) schemes against the Yee’s scheme, a traditional (2,4) scheme, and Deveze’s (4,4) scheme. Both non-filtered, and the maximally flat and Chebyshev filtered versions of the DRP schemes are considered. The phase velocity in free space is solved from the transcendental dispersion relation by assuming an uniform 2-D FDTD grid with $\Delta x = \Delta y$, or $\Gamma_x = \Gamma_y$, and, hence, the coordinate subscript in the coefficients is dropped in what follows.

The CFL number $\chi = 1$ is used for ordinary FDTD, while $\chi = 6/7$ for the (2,4) scheme, $\chi = 3/4$ for the (2,2) scheme and $\chi = 48/65$ for all (2,4) schemes. In fact, there is an optimal CFL number for each algorithm, which yields the minimum dispersion error among all possible CFL numbers. When comparing different algorithms, we employ their largest possible CFL numbers instead of that optimal one. This choice is justified since the largest CFL number minimizes the computational cost and is the most often used in practice.

Table 4.1 and Table 4.2 give some coefficients used in the DRP (2,4) schemes, where the maximally flat filter is designed with $q_c = 0.1$, and the Chebyshev filter is designed with $q_c = 0.1$ with $\Delta q = 0.02$. Table 4.3 and Table 4.4 give the coefficients using Chebyshev filtering schemes for various pairs $(q_c, \Delta q)$.

<table>
<thead>
<tr>
<th>Scheme</th>
<th>$\Gamma_{11}$</th>
<th>$\Gamma_{12}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deveze</td>
<td>1.125</td>
<td>-1.136094e-2</td>
</tr>
<tr>
<td>Non-filtered DRP</td>
<td>1.125</td>
<td>4.797060e-3</td>
</tr>
<tr>
<td>Maximally Flat DRP</td>
<td>1.125026</td>
<td>4.669305e-3</td>
</tr>
<tr>
<td>Chebyshev DRP</td>
<td>1.125025</td>
<td>4.665140e-3</td>
</tr>
</tbody>
</table>

Table 3.1: Coefficients used in different 2-D (2,4) schemes.
Deveze & -4.166667e-2 & 4.207758e-4 \\
Non-filtered DRP & -4.166667e-2 & -5.386002e-3 \\
Maximally Flat DRP & -4.159725e-2 & -5.733253e-3 \\
Chebyshev DRP & -4.159842e-2 & -5.744642e-3 \\

Table 3.2: Coefficients used in different 2-D (2,4) schemes (cont’d).

Fig. 3.3 and Fig. 3.4 show the maximum (for all angles) phase error per wavelength, defined as \((\hat{k}/k - 1) \times 360\) with \(\hat{k}\) denotes the discrete wavenumber obtained by solving the transcendental dispersion relation and \(k\) refers to the exact (continuum) wavenumber. As we see from Fig. 3.3, the DRP (2,2) scheme is about as accurate as the (2,4) scheme (with opposite sign). Fig. 3.4 shows that filtering schemes do reduce the dispersion error considerably around the specified frequency. We also observe that the Chebyshev filter performs better than the maximally flat filter in the full...
Figure 3.4: Comparison of the maximum (for all angles) phase error per wavelength using different DRP (2,4) schemes and the Deveze (4,4) scheme.

frequency range considered. Note that Fig. 3.4 does not correspond to Fig. 3.2 well after the crossover since Fig. 3.2 is just an analytical estimation for design purposes, while Fig. 3.4 depicts the actual performance of the algorithms.

We also note from Fig. 3.4 that, for the DRP schemes with filtering, the dispersion error at high frequencies can be made smaller than for low frequencies (contrary to non-filtered DRP and traditional schemes). As mentioned in the Introduction, this is a desirable characteristic because for a given computational domain (fixed physical size), high frequencies correspond to an electrically larger problem and hence are more impacted by the accumulated phase error. Nevertheless, unless the local dispersion error decreases faster than linearly with frequency, the largest accumulated phase error in the computational domain is still given by the highest frequency components. This issue will be further elaborated next.
<table>
<thead>
<tr>
<th>$q_c = 0.1$, $\Delta q = 0.02$</th>
<th>$\Gamma_{11}$</th>
<th>$\Gamma_{12}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.125025</td>
<td>4.665140e-3</td>
</tr>
<tr>
<td>$q_c = 0.1$, $\Delta q = 0.046$</td>
<td>1.125024</td>
<td>4.647156e-3</td>
</tr>
<tr>
<td>$q_c = 0.1$, $\Delta q = 0.037$</td>
<td>1.125024</td>
<td>4.655009e-3</td>
</tr>
<tr>
<td>$q_c = 0.1$, $\Delta q = 0.058$</td>
<td>1.125023</td>
<td>4.633954e-3</td>
</tr>
<tr>
<td>$q_c = 0.114$, $\Delta q = 0.053$</td>
<td>1.125025</td>
<td>4.665140e-3</td>
</tr>
</tbody>
</table>

Table 3.3: Coefficients of Chebyshev filtering schemes with different parameters $q_c$ and $\Delta q$.

<table>
<thead>
<tr>
<th>$q_c = 0.1$, $\Delta q = 0.02$</th>
<th>$\Gamma_{21}$</th>
<th>$\Gamma_{22}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-4.159842e-2</td>
<td>-5.744642e-3</td>
</tr>
<tr>
<td>$q_c = 0.1$, $\Delta q = 0.046$</td>
<td>-4.160251e-2</td>
<td>-5.793854e-3</td>
</tr>
<tr>
<td>$q_c = 0.1$, $\Delta q = 0.037$</td>
<td>-4.160091e-2</td>
<td>-5.772360e-3</td>
</tr>
<tr>
<td>$q_c = 0.1$, $\Delta q = 0.058$</td>
<td>-4.160454e-2</td>
<td>-5.830010e-3</td>
</tr>
<tr>
<td>$q_c = 0.114$, $\Delta q = 0.053$</td>
<td>-4.159842e-2</td>
<td>-5.744642e-3</td>
</tr>
</tbody>
</table>

Table 3.4: Coefficients of Chebyshev filtering schemes with different parameters $q_c$ and $\Delta q$ (continued).

We assume a FDTD simulation in a (typical) frequency range such that the lowest frequency corresponds to $q = 0.01$ and the highest frequency corresponds to $q = 0.1$. Now consider the phase error accumulated in a distance which equals to the largest wavelength being considered. Fig. 3.5 shows the largest (for all angles) phase error accumulated in such a distance by employing Chebyshev filtered DRP algorithms with different $\Delta q$ (while $q_c$ is kept fixed as 0.1). For $\Delta q = 0.02$, the usual notion that the largest accumulated phase error is dominated by the highest frequency is still valid. However, this is not true anymore if larger $\Delta q$ are considered. In fact, when $\Delta q = 0.046$, the largest accumulated phase error is now determined by lower frequency components. Indeed, there is an optimal $\Delta q$ around $\Delta q = 0.058$ for which the maximum accumulated magnitude of the phase error for all frequencies is a minimum.

In the case of narrowband simulations, we can obtain a pair $(q_c, \Delta q)$ which yields a minimum possible maximum phase error at some specific frequency. As illustrated
Figure 3.5: Comparison of the maximum (for all angles) accumulated phase error over the largest wavelength in DRP schemes using Chebyshev filtering schemes with various $\Delta q$ while fixing $q^c = 0.1$.

In Fig. 3.6, if we specify $q^c = 0.1$ and $\Delta q = 0.046$, the maximum phase error is minimum at $q = 0.087$. This suggests that to obtain a better performance at a specific frequency, we can simply design filters with a slight larger $q_c$ and a fine-tuned $\Delta q$ accordingly. Fig. 3.6 shows another example with $q = 0.114$ and $\Delta q = 0.053$ which is intended to improve the performance at $q = 0.1$. As we see, a maximum phase error less than $5 \times 10^{-3}$ is obtained at $q = 0.1$, while a good bandwidth is maintained.
Figure 3.6: Comparison of the maximum (for all angles) local phase error per wavelength using DRP schemes and Chebyhev filters with different parameters.

3.3 Three-Dimensional DPR-FDTD Algorithm

3.3.1 Methodology

3-D FDTD Equations

FDTD equations using a (2,4) stencil are written as

\[ \mathbf{E}^{i+1} = \mathbf{E}^i + \frac{\Delta t}{\epsilon} \mathbf{S} \times \mathbf{H}^{i+\frac{1}{2}} \]

\[ \mathbf{H}^{i+\frac{1}{2}} = \mathbf{H}^{i-\frac{1}{2}} - \frac{\Delta t}{\mu} \mathbf{S} \times \mathbf{E}^i \]

where the superscripts indicate the time step. The operator \( \mathbf{S} \) is defined as

\[ \mathbf{S} = \frac{1}{\Delta x}(C_{1x}S_x^+ + C_{2x}S_x^{++})\hat{x} + \frac{1}{\Delta y}(C_{1y}S_y^+ + C_{2y}S_y^{++})\hat{y} + \frac{1}{\Delta z}(C_{1z}S_z^+ + C_{2z}S_z^{++})\hat{z} \]

(3.37)
where $C_{1s}$ and $C_{2s}$, $s = x, y, z$ are coefficients to be determined, and $S^x_+$ and $S^{++}_x$ are displacement operators \[15\] defined by

$$S^x_+ E^i_y(l+1/2, m+1/2, n) = E^i_y(l+1, m+1/2, n) - E^i_y(l,m+1/2, n)$$

$$S^{++}_x E^i_y(l+1/2, m+1/2, n) = E^i_y(l+2, m+1/2, n) - E^i_y(l-1, m+1/2, n)$$

and similarly for the other components. In the above, the subscripts indicate the spatial location. These operators should be understood as being applied to field components collocated to the components on the left hand side of (3.35) and (3.36), so that their image produces field components at the usual positions on the staggered FDTD grid. For example, the $H_z$ update is given by

$$H^i_z(l+1/2, m+1/2, n) = H^i_z(l+1/2, m+1/2, n) + \Delta t \mu \left( \frac{1}{\Delta y} [C_{1y}(E^i_x(l+1/2, m+1, n) - E^i_x(l+1/2, m, n)) + C_{2y}(E^i_y(l+1/2, m+2, n) - E^i_y(l+1/2, m-1, n)) \right)$$

$$- \frac{1}{\Delta x} [C_{1x}(E^i_x(l+1, m+1/2, n) - E^i_x(l, m+1/2, n)) + C_{2x}(E^i_y(l+2, m+1/2, n) - E^i_y(l-1, m+1/2, n))]$$

(3.40)

In the FDTD grid, the $E$ and $H$ fields may be expanded into a discrete set of Fourier modes as below

$$E^i_x(l+1/2, m, n) = E^i_x(t) e^{-j[k_x(l+1/2)\Delta x + k_y m \Delta y + k_z n \Delta z]}$$

$$E^i_y(l, m+1/2, n) = E^i_y(t) e^{-j[k_x l \Delta x + k_y (m+1/2) \Delta y + k_z n \Delta z]}$$

$$E^i_z(l, m, n+1/2) = E^i_z(t) e^{-j[k_x l \Delta x + k_y m \Delta y + k_z (n+1/2) \Delta z]}$$

$$H^i_x(l+1/2, m, n+1/2) = H^i_x(t) e^{-j[k_x l \Delta x + k_y (m+1/2) \Delta y + k_z (n+1/2) \Delta z]}$$

$$H^i_y(l+1/2, m+1/2, n) = H^i_y(t) e^{-j[k_x (l+1/2) \Delta x + k_y m \Delta y + k_z (n+1/2) \Delta z]}$$

$$H^i_z(l+1/2, m+1/2, n) = H^i_z(t) e^{-j[k_x (l+1/2) \Delta x + k_y (m+1/2) \Delta y + k_z n \Delta z]}$$

(3.41)

where

$$\tilde{E}^i(t) = \mathcal{E} e^{j\omega (i\Delta t)}, \quad \tilde{H}^i(t) = \mathcal{H} e^{j\omega (i+\frac{1}{2}) \Delta t}$$

(3.42)

Substituting (3.41) and (3.42) into (3.35) and (3.36), we obtain

$$\tilde{E}\sin\left(\frac{\omega \Delta t}{2}\right) = -\frac{\Delta t}{\epsilon} \tilde{F} \times \mathcal{H}$$

(3.43)
\[ \vec{H}\sin\left(\frac{\omega \Delta t}{2}\right) = \frac{\Delta t}{\mu} \vec{F} \times \vec{E} \]  

(3.44)

with

\[ \vec{F} = F_x \hat{x} + F_y \hat{y} + F_z \hat{z} \]

\[ F_x = \frac{1}{\Delta x} [C_{1x} \sin\left(\frac{k_x \Delta x}{2}\right) + C_{2x} \sin\left(\frac{3k_x \Delta x}{2}\right)] \]

\[ F_y = \frac{1}{\Delta y} [C_{1y} \sin\left(\frac{k_y \Delta y}{2}\right) + C_{2y} \sin\left(\frac{3k_y \Delta y}{2}\right)] \]

\[ F_z = \frac{1}{\Delta z} [C_{1z} \sin\left(\frac{k_z \Delta z}{2}\right) + C_{2z} \sin\left(\frac{3k_z \Delta z}{2}\right)] \]

Either \( \vec{E} \) or \( \vec{H} \) can be eliminated from (3.43) and (3.44) to obtain a (discrete) dispersion relation. For a given set of coefficients \( C_{1s} \) and \( C_{2s} \) (recall that in a traditional (2,4) scheme, \( C_{1s} = 9/8 \) and \( C_{2s} = -1/24 \)), this relation is used to analyze the numerical dispersion in the FDTD grid. In this work we will adopt an inverse standpoint and use it to actually design modified FDTD schemes with minimized dispersion error. Specifically, we will enforce the exact (continuum) relation between frequency and wavenumber, viz. \( \omega = v_p k \), in the dispersion relation, and then solve (in an approximate sense to be clear later on) for \( C_{1s} \) and \( C_{2s} \), \( s = x, y, z \) as unknowns.

We consider a plane wave (Fourier mode) propagating in the \((\theta, \phi)\) direction such that \( k_x = k \sin \theta \cos \phi \), \( k_y = k \sin \theta \sin \phi \), and \( k_z = k \cos \theta \). Any such wave can be decomposed into a \( TE_z \) polarization \( (E_z = 0) \), and \( TM_z \) polarization \( (H_z = 0) \). A possible way to obtain \( C_{1s} \) and \( C_{2s} \) is to enforce \( \omega = v_p k \) directly in (3.43) and (3.44) and solve for the coefficients. In this way, \( C_{1s} \) and \( C_{2s} \) are obtained in terms of the frequency \( \omega \) and propagation angle \((\theta, \phi)\), for given \( \Delta x, \Delta y, \Delta z \), and \( \Delta t \). The caveat on this approach is that zero dispersion error is approached only for a particular \( \omega \) and \((\theta, \phi)\), with no guarantee of a small dispersion error at other angles. It is clear that, for most practical purposes, optimal coefficients should not depend on a particular
propagation direction in the FDTD grid. Instead, we will obtain here coefficients which minimize the maximum dispersion error for all angles (minimax sense). This will be done by expanding the dispersion error in a (rapidly convergent) series in terms of the elevation, \( \theta \), and azimuth, \( \phi \), angles, and by enforcing the dominant terms to be identically zero. Since this enforcement involves only the series coefficients (still a function of frequency but not angle anymore), the dispersion error can be reduced for all angles simultaneously in a controllable manner.

We consider the \( H \) update for the \( TE_z \) polarization so that

\[
E_x = -E \sin \phi, \quad E_y = E \cos \phi, \quad E_z = 0 \tag{3.45}
\]

\[
\mathcal{H}_x = -\mathcal{H} \cos \theta \cos \phi, \quad \mathcal{H}_y = -\mathcal{H} \cos \theta \sin \phi, \quad \mathcal{H}_z = \mathcal{H} \sin \theta \tag{3.46}
\]

Substituting (3.45) and (3.46) into (3.44), we obtain two independent equations

\[
\mathcal{H} \cos \theta \sin \left( \frac{\omega \Delta t}{2} \right) = \frac{\Delta t}{\mu \Delta z} \left[ C_{1z} \sin \left( \frac{k_z \Delta z}{2} \right) + C_{2z} \sin \left( \frac{3k_z \Delta z}{2} \right) \right] \mathcal{E} \tag{3.47}
\]

\[
\mathcal{H} \sin \theta \sin \left( \frac{\omega \Delta t}{2} \right) = \frac{\Delta t}{\mu \Delta z} \left[ \frac{1}{\Delta x} \cos \phi \left[ C_{1x} \sin \left( \frac{k_x \Delta x}{2} \right) + C_{2x} \sin \left( \frac{3k_x \Delta x}{2} \right) \right] + \frac{1}{\Delta y} \sin \phi \left[ C_{1y} \sin \left( \frac{k_y \Delta y}{2} \right) + C_{2y} \sin \left( \frac{3k_y \Delta y}{2} \right) \right] \right] \mathcal{E} \tag{3.48}
\]

The other polarization gives similar equations for the \( E \) update. We can in principle use (3.47) to find optimized \( C_{1z} \) and \( C_{2z} \), and (3.48) to find optimized \( C_{1x,y} \) and \( C_{2x,y} \).

Eq. (3.48) contains four unknowns coefficients and depends on both \( \theta \) and \( \phi \), while (3.47) contains only two coefficients and does not depend on the azimuth angle \( \phi \). This is because the dispersion error regarded as a function of \( \theta \) and \( \phi \) is invariant with respect to \( \phi \) for any choice for \( C_{1z} \) and \( C_{2z} \) (for fixed \( C_{1x,y} \) and \( C_{2x,y} \)). From the symmetry of the problem, we can work solely with (3.47) to obtain \( C_{1z} \) and \( C_{2z} \) and the other coefficients are simultaneously determined.
Noticing that $\mathcal{E} = \eta \mathcal{H}$, and denoting the Courant-Friedrichs-Lewy (CFL) number as $\chi_s = \sqrt{3}v_p \Delta t/\Delta s$, we rewrite (3.47) as

$$\frac{\sqrt{3}}{\chi_s} \sin\left(\frac{\pi q_s \chi_s}{\sqrt{3}}\right) \cos \theta = C_{1s} \sin(\pi q_s \cos \theta) + C_{2s} \sin(3\pi q_s \cos \theta)$$

(3.49)

where, again, $s = x, y$ or $z$, and $q_s = \Delta s/\lambda$ denotes the number of wavelengths per cell size, $\Delta s$.

We define an error functional proportional to the difference between the l.h.s. and r.h.s. of (3.49)

$$\delta(C_{1s}, C_{2s}, \theta, \phi) = \frac{\sqrt{3}}{\chi_s} \sin\left(\frac{\pi q_s \chi_s}{\sqrt{3}}\right) \cos \theta - \left[ C_{1s} \sin(\pi q_s \cos \theta) + C_{2s} \sin(3\pi q_s \cos \theta) \right]$$

(3.50)

Solving for $C_{1s}$ and $C_{2s}$ in (3.49) is equivalent to letting $\delta(C_{1s}, C_{2s}, \theta, \phi) = 0$. In order to find proper $C_{1s}$ and $C_{2s}$ not depending on $(\theta, \phi)$, we first expand $\delta(C_{1s}, C_{2s}, \theta, \phi)$ in a series of spherical harmonics $Y_{l,m}(\theta, \phi)$, with coefficients $c_{l,m}$ are given by

$$c_{l,m} = \int_0^{2\pi} \int_0^\pi \delta(C_{1s}, C_{2s}, \theta, \phi) Y_{l,m}(\theta, \phi) \sin \theta d\theta d\phi$$

Strictly speaking, since the functional in (3.50) is not a function of $\phi$, we know a priori that the spherical harmonics will involve only $m = 0$ azimuthal harmonics. We will nevertheless start from this more general form, which would accommodate the case of $\Delta x \neq \Delta y \neq \Delta z$ (not pursued further here though).

We use the following conventions for spherical harmonics

$$Y_{l,m}(\theta, \phi) = \Theta_{l|m|}(\theta) \Phi_{m}^{re}(\phi)$$

$$\Phi_{0}^{re}(\phi) = \frac{1}{\sqrt{2\pi}}$$

$$\Phi_{|m|}^{re}(\phi) = \frac{1}{\sqrt{\pi \cos m\phi}}$$

59
$$\Phi_{-|m|}(\phi) = \frac{1}{\sqrt{\pi}} \sin m\phi$$

$$\Theta_{|m|}(\theta) = \sqrt{\frac{(2l + 1)(l - |m|)!}{2(l + |m|)!}} P_{l}^{(|m|)}(\cos \theta)$$

where $P_{l}^{m}(\cos \theta)$ are associated Legendre functions (with $\cos \theta$ argument) with $l \geq 0$, and $-l \leq m \leq l$. As a result, we can rewrite the first term of the r.h.s. of (3.50) as

$$\sqrt{3} \chi_{s} \sin(\pi q_{s} \chi_{s} \sqrt{3}) \cos \theta = 2 \sqrt{\pi} \chi_{s} \sin(\pi q_{s} \chi_{s} \sqrt{3}) Y_{1,0}(\theta, \phi)$$

To expand of the last term of the r.h.s. of (3.50) we use the identity below

$$\int_{0}^{2\pi} \int_{0}^{\pi} \sin(\alpha \cos \theta) Y_{l,m}(\theta, \phi) \sin \theta d\theta d\phi = \sqrt{\pi(2l + 1)} \int_{0}^{\pi} \sin(\alpha \cos \theta) P_{l}(\cos \theta) \sin \theta d\theta$$

$$= \sqrt{\pi(2l + 1)} \int_{0}^{1} \sin(\alpha t) P_{l}(t) dt$$

$$\equiv \sqrt{\pi(2l + 1)} I_{l}(\alpha)$$ (3.51)

Since $P_{l}(-t) = (-1)^{l} P_{l}(t)$, the above integral is zero for $l$ even. Bearing this in mind and using the fact that $P_{l+1}(t) - P_{l-1}(t) = (2l + 1) P_{l}(t)$ and $P_{l}(-1) = (-1)^{l}$, we calculate $I_{l}(\alpha)$ via integration by parts to obtain

$$I_{l}(\alpha) = \left[ \frac{2(2l - 3)}{2l - 5} - \frac{(2l - 1)(2l - 3)}{\alpha^2} \right] I_{l-2}(\alpha) - \frac{2l - 1}{2l - 5} I_{l-4}(\alpha) \quad \text{for odd } l \quad (3.52)$$

The first two terms are readily found as

$$I_{1}(\alpha) = \frac{2(\sin \alpha - \alpha \cos \alpha)}{\alpha^2}$$

$$I_{3}(\alpha) = \frac{6(2\alpha^2 - 5) \sin \alpha - 2\alpha(\alpha^2 - 15) \cos \alpha}{\alpha^4}$$

All other terms can be evaluated from these two by (3.52).

So far, we have obtained the following expansion

$$\delta(C_{1s}, C_{2s}, \theta, \phi) = \sum_{l=0}^{\infty} \sqrt{\pi(4l + 3)} [C_{1s} I_{2l+1}(\pi q_{s}) + C_{2s} I_{2l+1}(3\pi q_{s})] Y_{2l+1,0}(\theta, \phi)$$

$$+ \frac{2\sqrt{\pi}}{\chi_{s}} \sin(\frac{\pi q_{s} \chi_{s}}{\sqrt{3}}) Y_{1,0}(\theta, \phi)$$ (3.53)
Next, we force the first two non-zero terms in (3.53), i.e., $Y_{1,0}(\theta, \phi)$ and $Y_{3,0}(\theta, \phi)$ terms, to be identically zero and solve for $C_{1s}$ and $C_{2s}$. As mentioned before, in this way the maximum dispersion error for all angles is minimized simultaneously, and $C_{1s}$ and $C_{2s}$ become a function of frequency only. As we will demonstrate, this latter property allows for incorporation of the modified coefficients in FDTD algorithms after a polynomial expansion in terms of $\omega$ (or, equivalently, $q_s$). The solutions are given by

$$C_{1s} = C_{1s}^{(\infty)} \equiv \frac{\pi^2 q_s^2 [3\pi q_s (5 - 3\pi^2 q_s^2) \cos(3\pi q_s) + (18\pi^2 q_s^2 - 5) \sin(3\pi q_s)] \sin \left(\frac{\pi q_s \chi_s}{\sqrt{3}}\right)}{\Upsilon}$$

$$C_{2s} = C_{2s}^{(\infty)} \equiv \frac{27\pi^2 q_s^2 [\pi q_s (\pi^2 q_s^2 - 15) \cos(\pi q_s) + 3(5 - 2\pi^2 q_s^2) \sin(\pi q_s)] \sin \left(\frac{\pi q_s \chi_s}{\sqrt{3}}\right)}{\Upsilon}$$

where

$$\Upsilon = 20\sqrt{3}\chi_s \{(1 + 3\pi^2 q_s^2) \cos(2\pi q_s) - \cos(4\pi q_s) + \pi q_s \{3\pi q_s [\cos(4\pi q_s) - 2\pi q_s \cos(\pi q_s) \sin(\pi q_s)^3] + 2[\sin(2\pi q_s) - 2\sin(4\pi q_s)]\}\}$$

Substituting (3.54)–(3.55) into (3.43), the residue error

$$\delta_e \equiv \delta(C_{1s}, C_{2s}, \theta, \phi) - \delta(C_{1s}^{(\infty)}, C_{2s}^{(\infty)}, \theta, \phi)$$

is given by

$$\delta_e = \sum_{l=2}^{\infty} \sqrt{\pi(4l + 3)} [C_{1s}^{(\infty)} I_{2l+1}(\pi q_s) + C_{2s}^{(\infty)} I_{2l+1}(3\pi q_s)] Y_{2l+1,0}(\theta, \phi)$$

(3.56)

It is easy to verify that $I_l(\alpha)$ given in (3.52) decrease very rapidly as $l$ increase when $\alpha < 1$ (notice that the inequality $3\pi q_s < 1$ should be satisfied in practical FDTD simulations, where $q_s < 0.1$), thus the coefficients in (3.56) also decrease very rapidly as expected (the magnitude of each coefficient in the series represents the maximum dispersion error from the corresponding angular mode).
The coefficients $C_{1s}^{(∞)}$ and $C_{2s}^{(∞)}$ given in (3.54) and (3.55) cannot be directly implemented in the time domain. To circumvent this, we expand $C_{1s}^{(∞)}$ and $C_{2s}^{(∞)}$ in a Taylor series around $q_s = 0$ and retain the two lowest order terms, i.e., $C_{1s}^{(∞)} = C_{1s}^{(2)} + O(q_s^4)$ and $C_{2s}^{(∞)} = C_{2s}^{(2)} + O(q_s^4)$ with

$$C_{1s}^{(2)} = \frac{9}{8} - \frac{1}{64} 4\pi^2 q_s^2 (\chi_s^2 - 1) \quad (3.57)$$

$$C_{1s}^{(2)} = -\frac{1}{24} + \frac{1}{1728} 4\pi^2 q_s^2 (\chi_s^2 - 9) \quad (3.58)$$

The above can be easily transformed back to time domain through $\omega^2 \rightarrow -\partial^2/\partial t^2$. However, if straightforward time discretization schemes are employed directly on the resulting equations (with third order time derivatives), the update becomes unconditionally unstable [11]. Alternatively, the second order time derivative can be further cast as a combination of spatial derivatives as $v_p^2 \nabla^2$ (Helmholtz equation) and discretized as such (modified equation method [8]). This latter transformation is valid for staggered grids as long as $\mu$ and $\epsilon$ are uniform in the local stencil. In this manner, (3.57) and (3.58) becomes

$$C_{1s}^{(2)} = \frac{9}{8} + \frac{1}{64} (\chi_s^2 - 1) \Delta s^2 \nabla^2 \equiv C_{11s} - C_{12s} \Delta s^2 \nabla^2 \quad (3.59)$$

$$C_{1s}^{(2)} = -\frac{1}{24} - \frac{1}{1728} (\chi_s^2 - 9) \Delta s^2 \nabla^2 \equiv C_{21s} - C_{22s} \Delta s^2 \nabla^2 \quad (3.60)$$

If only the first order terms in (3.59) and (3.60) are considered, the traditional (2,4) scheme is recovered. The second terms in (3.59) and (3.60) are analogous to third order time derivative terms in traditional schemes with fourth order of accuracy in time [11]. The difference resides in multiplicative factors. In this sense, we denote the scheme a (non-filtered, minimax) optimized (2,4) scheme.

The implementation of $C_{1s}^{(2)}$ and $C_{2s}^{(2)}$ in staggered grids is straightforward. The stability condition can be derived in a standard way [63] from the update and the
result is

$$\Delta t \leq \frac{1}{v_p \sqrt{\frac{D^2_{xy}}{\Delta x^2} + \frac{D^2_{xz}}{\Delta y^2} + \frac{D^2_{yz}}{\Delta z^2}}} \tag{3.61}$$

where

$$D_{xy} = [C_{11x}^{(2)} + C_{12y}^{(2)}(3 + 2\frac{\Delta x^2}{\Delta y^2} + 2\frac{\Delta z^2}{\Delta y^2}) - C_{22x}^{(2)}] \sin\left(\frac{k_y \Delta x}{2}\right) + [C_{21x}^{(2)} + C_{22y}^{(2)}(2 + 2\frac{\Delta x^2}{\Delta y^2}) + 2\frac{\Delta x^2}{\Delta y^2} - C_{12x}^{(2)}] \sin\left(\frac{3k_y \Delta x}{2}\right) - 2\left[\frac{\Delta x^2}{\Delta y^2} \cos(k_y \Delta x) + \frac{\Delta z^2}{\Delta y^2} \cos(k_y \Delta z)\right]\left[C_{12x}^{(2)} \sin\left(k_y \Delta x\right) + C_{22x}^{(2)} \sin\left(\frac{5k_y \Delta x}{2}\right)\right]$$

$$D_{xz} = [C_{11x}^{(2)} + C_{12y}^{(2)}(3 + 2\frac{\Delta x^2}{\Delta y^2} + 2\frac{\Delta z^2}{\Delta x^2}) - C_{22y}^{(2)}] \sin\left(\frac{k_z \Delta x}{2}\right) + [C_{21x}^{(2)} + C_{22y}^{(2)}(2 + 2\frac{\Delta x^2}{\Delta y^2}) + 2\frac{\Delta x^2}{\Delta y^2} - C_{12x}^{(2)}] \sin\left(\frac{3k_z \Delta x}{2}\right) - 2\left[\frac{\Delta x^2}{\Delta y^2} \cos(k_z \Delta x) + \frac{\Delta y^2}{\Delta z^2} \cos(k_z \Delta z)\right]\left[C_{12x}^{(2)} \sin\left(k_z \Delta x\right) + C_{22x}^{(2)} \sin\left(\frac{5k_z \Delta x}{2}\right)\right]$$

$$D_{yz} = [C_{11x}^{(2)} + C_{12y}^{(2)}(3 + 2\frac{\Delta x^2}{\Delta y^2} + 2\frac{\Delta z^2}{\Delta x^2}) - C_{22x}^{(2)}] \sin\left(\frac{k_z \Delta x}{2}\right) + [C_{21x}^{(2)} + C_{22y}^{(2)}(2 + 2\frac{\Delta x^2}{\Delta y^2}) + 2\frac{\Delta x^2}{\Delta y^2} - C_{12x}^{(2)}] \sin\left(\frac{3k_z \Delta x}{2}\right) - 2\left[\frac{\Delta x^2}{\Delta y^2} \cos(k_y \Delta y) + \frac{\Delta z^2}{\Delta y^2} \cos(k_z \Delta z)\right]\left[C_{12x}^{(2)} \sin\left(k_z \Delta x\right) + C_{22x}^{(2)} \sin\left(\frac{5k_z \Delta x}{2}\right)\right]$$

For $\Delta x = \Delta y = \Delta z = h$, we can drop all the subscripts in (3.61) and the above bound can be simplified to

$$\Delta t \leq \frac{h}{v_p \sqrt{3}} \chi = \frac{h}{v_p \sqrt{3} |D|_{max}} \tag{3.62}$$

where $|D|_{max} = (C_{11}^{(2)} - C_{21}^{(2)}) + 12(C_{12}^{(2)} - C_{22}^{(2)})$ denotes the maximum magnitude of $D_{xy}$ and $D_{xz}$ and $D_{yz}$. Substituting $C_{11}^{(2)}$, $C_{21}^{(2)}$, $C_{12}^{(2)}$ and $C_{22}^{(2)}$ (referring to (3.59) and (3.60)) into $|D|_{max}$, we have

$$|D|_{max} = max\{\frac{7}{6} + \frac{36 - 28\chi^2}{144}\}$$

which takes the maximum when $\chi = 0$. Thus the CFL number $\chi$ with guaranteed stability is found to be $12/17$. Note that this is a conservative bound, $\chi$ always larger than 0 in practice. In fact, $\chi = 0.75$ was found to produce stable results (even for schemes with filtering as described ahead).

Rather than employing costly alternatives such as additional terms in the Taylor expansion or sophisticated time integration schemes, we can also improve the previous
Results by using filtering schemes. This technique has been discussed in great details in the previous section and we'll only show the results here.

### 3.3.2 Numerical Comparisons

We compare the non-filtered, Butterworth, and Chebyshev optimized (2,4) schemes against Fang’s (4,4) and Deveze’s (4,4) schemes. The phase velocity in free space is solved from the transcendental dispersion relation for $\chi = 12/17$ (maximum CFL number from the stability bound). Since we assume a uniform 3-D FDTD grid with cell size $h$, we drop the coordinate subscript of the coefficients in the following.

Table 4.5 and Table 4.6 give the coefficients used in the (2,4) schemes\textsuperscript{4}. The Butterworth filter and the Chebyshev filter are optimized at $\Delta q = 0.1$. The scheme labelled as “Chebyshev1” is optimized with $\Delta q = 0.02$, while “Chebyshev2” is optimized with $\Delta q = 0.035$.

\begin{table}[h]
\begin{center}
\begin{tabular}{|l|c|c|}
\hline
 & $C_{11}$ & $C_{12}$ \\
\hline
Deveze & 1.125 & -1.136094e-2 \\
Non-filtered & 1.125 & 7.839533e-3 \\
Butterworth & 1.125011 & 7.785720e-3 \\
Chebyshev1 & 1.125011 & 7.783932e-3 \\
Chebyshev2 & 1.125010 & 7.780166e-3 \\
\hline
\end{tabular}
\end{center}
\caption{Coefficients used in different 3-D (2,4) schemes.}
\end{table}

\begin{table}[h]
\begin{center}
\begin{tabular}{|l|c|c|}
\hline
 & $C_{21}$ & $C_{22}$ \\
\hline
Deveze & -4.166667e-2 & 4.207758e-4 \\
Non-filtered & -4.166667e-2 & -4.919983e-3 \\
Butterworth & -4.161038e-2 & -5.202035e-3 \\
Chebyshev1 & -4.161134e-2 & -5.211189e-3 \\
Chebyshev2 & -4.161315e-2 & -5.230449e-3 \\
\hline
\end{tabular}
\end{center}
\caption{Coefficients used in different 3-D (2,4) schemes (cont’d).}
\end{table}

\textsuperscript{4}The dispersion relation of Fang’s scheme is taken directly from [58].
Figs. 3.7–3.9 depicts the maximum (for all angles) phase error per wavelength, defined as \((\hat{k}/k-1)\times360\), where \(\hat{k}\) denotes the discrete wavenumber obtained by solving the transcendental dispersion relation and \(k\) refers to the continuum wavenumber.

Fig. 3.7 illustrates, for reference, the dispersion error levels from the low-order FDTD schemes (Yee’s and (2,4) schemes). Fig. 3.8 shows that the non-filtered optimized scheme has a better performance than both Deveze’s (4,4) and Fang’s (4,4) schemes. Moreover, Fig. 3.9 shows that the filtering schemes are able to further reduce the dispersion error around the central frequency.

Fig. 3.10 and Fig. 3.11 show the dispersion error produced by specifying different \(q^c\) in schemes with Butterworth and Chebyshev filtering, respectively. It is observed

---

5These curves are generated using the largest CFL number from the stability bound of each algorithm, viz. \(\chi = 1\) for Yee’s scheme and \(\chi = 6/7\) for the traditional (2,4) scheme.
Figure 3.8: Comparison of the maximum (for all angles) phase error per wavelength using a non-filtered optimized (2,4) scheme, Fang’s (4,4) scheme and the Deveze (4,4) scheme.

that the Chebyshev filter produces a smaller error than the Butterworth filter in the full frequency range considered.

Note that for the schemes with filtering, the dispersion error at high frequencies can be actually made smaller than the error for low frequencies (contrary to traditional schemes). As mentioned in the Introduction, this is a highly desirable characteristic since, for a given FDTD domain size, high frequencies correspond to an electrically larger problem and hence their accumulated phase error tends to be larger. Unless the local dispersion error decreases faster than linearly with frequency, the largest accumulated phase error in the computational domain is determined by the highest frequency components. To elaborate this issue in more detail, consider a 3-D FDTD simulation in a (typical) frequency range such that the lowest frequency corresponds
to $q = 0.01$ and the highest frequency corresponds to $q = 0.1$. Fig. 3.12 shows the largest (for all angles) phase error accumulated in a distance equal to the largest wavelength considered by employing the different optimized (2,4) schemes.

As it can be seen from this figure, although the local dispersion error from the Butterworth and Chebyshev schemes of Fig. 3.9 is larger at low frequencies than at high frequencies, the largest accumulated phase error is still dominated by high frequencies. The Chebyshev2 scheme is the one that comes closer to producing an uniform accumulated phase error at all frequencies (and also the smallest).

Fig. 3.13 shows the accumulated phase error at two angles with the largest relative error difference, $(\theta, \phi) = (0^\circ, 0^\circ)$ and $(\theta, \phi) = (45^\circ, 0^\circ)$ for different $\Delta q$. From this Figure, we observe that by increasing $\Delta q$, the discrete $\hat{k}$ is reduced, but this has no
discernible effect on the dispersion anisotropy. By choosing an appropriate $\Delta q$, the magnitude of the maximum accumulated phase error can be further minimized by having near symmetric curves with respect to the zero error level. As in this case, we find that $\Delta q = 0.035$ exhibit more closely this characteristic, with the largest accumulated phase error of $0.2^\circ$ at $q = 0.1$.

Fig. 3.14 illustrates the anisotropy of the accumulated phase error for such a Chebyshev filtering scheme. Fig. 3.15 illustrates the frequency dependency of the anisotropy of the accumulated phase error when $\phi = 0^\circ$, showing the increase on the anisotropy at high frequencies.

Fig. 3.13 shows the comparison of the normalized anisotropy of the accumulated phase error by employing different $\Delta q$ in Chebyshev filtering schemes with fixed
\[ q^c = 0.1 \] This shows that despite the increase on the anisotropy at larger frequencies, the magnitude of the total accumulated phase error is everywhere smaller than 0.2°.

### 3.3.3 Truncation Error

We now study the truncation error of DRP schemes. Let us consider the \( H_z \) update (3.40) in uniform grids and plug (3.59) and (3.60) into (3.40), we have

\[
H_z^{i+\frac{1}{2},(l\frac{1}{2},m\frac{1}{2},n\frac{1}{2})} = H_z^{i\frac{1}{2},(l\frac{1}{2},m\frac{1}{2},n\frac{1}{2})} + \frac{\Delta t}{\mu} \left\{ \frac{1}{3} \left[ \left( \frac{9}{8} + \frac{1}{16} (\chi_s^2 - 1) \Delta s^2 \nabla^2 \right) (E_x^{i,(l\frac{1}{2},m\frac{1}{2},n\frac{1}{2})}) + \left( \frac{1}{24} - \frac{1}{1728} (\chi_s^2 - 9) \Delta s^2 \nabla^2 \right) (E_x^{i,(l\frac{1}{2},m\frac{1}{2},n\frac{1}{2})}) - E_x^{i,(l\frac{1}{2},m\frac{1}{2},n\frac{1}{2})}] \right\} \\
- \frac{1}{1728} (\chi_s^2 + \frac{1}{64} (\chi_s^2 - 1) \Delta s^2 \nabla^2) (E_x^{i,(l\frac{1}{2},m\frac{1}{2},n\frac{1}{2})}) - E_y^{i,(l\frac{1}{2},m\frac{1}{2},n\frac{1}{2})}) + \left( \frac{1}{24} - \frac{1}{1728} (\chi_s^2 - 9) \Delta s^2 \nabla^2 \right) (E_y^{i,(l+1,m\frac{1}{2},n\frac{1}{2})}) - E_y^{i,(l\frac{1}{2},m\frac{1}{2},n\frac{1}{2})}) \}
\]

(3.63)

The major difference between the above update and Fang’s (2,4) scheme involves spatial derivative terms. The time derivative is treated the same in both approaches,
which results in a temporal truncation error of $O(\Delta t^2)$. In the following, we shall focus on the spatial truncation error.

Let us consider the spatial derivative term involving $E_x$ only, because the $E_y$ term is treated the same as $E_x$ in (3.63). We have

$$
\frac{1}{\Delta s} \left[ \frac{9}{8} + \frac{1}{64} (\chi_s^2 - 1) \Delta s^3 \nabla^2 \right] \left( E_{x,\ell+1/2,m+1,n}^i - E_{x,\ell+1/2,m,n}^i \right) + \left[ -\frac{1}{24} - \frac{1}{1728} (\chi_s^2 - 9) \Delta s^3 \nabla^2 \right] \\
(E_{x,\ell+1/2,m+2,n}^i - E_{x,\ell+1/2,m-1,n}^i)
$$

(3.64)

By expanding $E_{x,\ell+1/2,m+1,n}^i$ in Taylor series as

$$
E_{x,\ell+1/2,m+1,n}^i = E_{x,\ell+1/2,m+1/2,n}^i + \frac{\partial E_{x,\ell+1/2,m+1/2,n}^i}{\partial y} \left( \frac{\Delta s}{2} \right) + \frac{\partial^2 E_{x,\ell+1/2,m+1/2,n}^i}{2 \partial y^2} \left( \frac{\Delta s}{2} \right)^2 \\
+ \frac{\partial^3 E_{x,\ell+1/2,m+1/2,n}^i}{6 \partial y^3} \left( \frac{\Delta s}{2} \right)^3 + \frac{\partial^4 E_{x,\ell+1/2,m+1/2,n}^i}{24 \partial y^4} \left( \frac{\Delta s}{2} \right)^4 \\
+ \frac{\partial^5 E_{x,\ell+1/2,m+1/2,n}^i}{120 \partial y^5} \left( \frac{\Delta s}{2} \right)^5 + O(\Delta s^6)
$$

(3.65)
Figure 3.13: Comparison of the anisotropy of the accumulated phase error over the largest wavelength by employing different $\Delta q$ in Chebyshev filtering schemes with fixed $q^c = 0.1$.

and the other terms similarly, we find the following relationship holds

$$
\frac{1}{\Delta s^2} \left[ \left( E^i_x(l + \frac{1}{2}, m, n) - E^i_x(l + \frac{1}{2}, m + 1, n) \right) - \frac{1}{2} \left( E^i_x(l + \frac{1}{2}, m + 2, n) - E^i_x(l + \frac{1}{2}, m - 1, n) \right) \right] = \partial_x E^i_x(l + \frac{1}{2}, m + 1, n) + O(\Delta s^4) \tag{3.66}
$$

The remaining term in (3.64) that needs to be considered is

$$
\frac{1}{\Delta s^2} \left[ \chi_s^2 - 1 \right] \Delta s^2 \nabla^2 \left[ E^i_x(l + \frac{1}{2}, m, n) - E^i_x(l + \frac{1}{2}, m - 1, n) \right] - \frac{1}{1728} \left( \chi_s^2 - 9 \right) \Delta s^2 \nabla^2 \left[ E^i_x(l + \frac{1}{2}, m + 2, n) - E^i_x(l + \frac{1}{2}, m - 1, n) \right] \tag{3.67}
$$

The $\nabla^2$ operators in (3.67) are approximated by centered difference, e.g.

$$
\nabla^2 E^i_x(l + \frac{1}{2}, m, n) \approx \frac{E^i_x(l - \frac{1}{2}, m + 1, n) - 2 E^i_x(l + \frac{1}{2}, m + 1, n) + E^i_x(l + \frac{1}{2}, m + 1, n)}{\Delta s^2} - \frac{E^i_x(l + \frac{1}{2}, m + 1, n) - 2 E^i_x(l + \frac{1}{2}, m + 1, n) + E^i_x(l + \frac{1}{2}, m + 1, n)}{\Delta s^2} \tag{3.68}
$$
After performing Taylor expansions similar to (3.65), it was found that

$$\nabla^2 E_x^{i(l+\frac{1}{2},m+1,n)} = \frac{\partial^2 E_x^{i(l+\frac{1}{2},m+1,n)}}{\partial x^2} + \frac{\partial^2 E_x^{i(l+\frac{1}{2},m+1,n)}}{\partial y^2}$$

$$+ \frac{\partial^2 E_x^{i(l+\frac{1}{2},m+1,n)}}{\partial z^2} + \frac{\partial^4 E_x^{i(l+\frac{1}{2},m+1,n)}}{12\partial x^4} (\Delta s^2) + \frac{\partial^4 E_x^{i(l+\frac{1}{2},m+1,n)}}{12\partial y^4} (\Delta s^2) + \frac{\partial^4 E_x^{i(l+\frac{1}{2},m+1,n)}}{12\partial z^4} (\Delta s^2) + O(\Delta s^4)$$

By using the following expressions

$$\frac{\partial^2 E_x^{i(l+\frac{1}{2},m+1,n)}}{\partial x^2} = \frac{\partial^2 E_x^{i(l+\frac{1}{2},m+\frac{1}{2},n)}}{\partial x^2} + \frac{\partial^3 E_x^{i(l+\frac{1}{2},m+\frac{1}{2},n)}}{\partial x^2 \partial y^2} (\frac{\Delta s}{2})^2 + \frac{\partial^3 E_x^{i(l+\frac{1}{2},m+\frac{1}{2},n)}}{\partial x^2 \partial z^2} (\frac{\Delta s}{2})^2 + \frac{\partial^4 E_x^{i(l+\frac{1}{2},m+\frac{1}{2},n)}}{\partial x^4} (\Delta s^2) + O(\Delta s^4)$$

and

$$\frac{\partial^4 E_x^{i(l+\frac{1}{2},m+1,n)}}{\partial x^4} = \frac{\partial^4 E_x^{i(l+\frac{1}{2},m+\frac{1}{2},n)}}{\partial x^4} + \frac{\partial^5 E_x^{i(l+\frac{1}{2},m+\frac{1}{2},n)}}{\partial x^4 \partial y^2} (\frac{\Delta s}{2})^2 + \frac{\partial^5 E_x^{i(l+\frac{1}{2},m+\frac{1}{2},n)}}{\partial x^4 \partial z^2} (\frac{\Delta s}{2})^2 + \frac{\partial^6 E_x^{i(l+\frac{1}{2},m+\frac{1}{2},n)}}{\partial x^6} (\Delta s^2) + O(\Delta s^5)$$

72
Figure 3.15: The anisotropy of the accumulated phase error over the largest wavelength for an optimized FDTD scheme with Chebyshev filter designed at $q^c = 0.1$ and $\Delta q = 0.035$. $\phi$ is fixed to be $0^\circ$.

we have

$$
\frac{1}{64}(\chi_s^2 - 1) \Delta s^2 \nabla^2 (E_x^{i(l+\frac{1}{2},m+1,n)} - E_x^{i(l+\frac{1}{2},m,n)}) = \frac{1}{64}(\chi_s^2 - 1) \Delta s^2 \left[ \frac{\partial^3 E_x^{i(l+\frac{1}{2},m+\frac{1}{2},n)}}{\partial x^3 \partial y^2} \Delta s + 2 \frac{\partial^2 E_x^{i(l+\frac{1}{2},m+\frac{1}{2},n)}}{\partial x^2 \partial y^3} \left( \frac{\Delta s}{2} \right)^3 + \frac{\partial^3 E_x^{i(l+\frac{1}{2},m+\frac{1}{2},n)}}{\partial y^3} (\Delta s)^3 \right]$$

(3.72)

and

$$
-\frac{1}{1728}(\chi_s^2 - 9) \Delta s^2 \nabla^2 (E_x^{i(l+\frac{1}{2},m+2,n)} - E_x^{i(l+\frac{1}{2},m-1,n)}) = -\frac{1}{1728}(\chi_s^2 - 9) \Delta s^2 \left[ \frac{\partial^3 E_x^{i(l+\frac{1}{2},m+\frac{1}{2},n)}}{\partial x^3 \partial y^2} (3\Delta s) + 2 \frac{\partial^2 E_x^{i(l+\frac{1}{2},m+\frac{1}{2},n)}}{\partial x^2 \partial y^3} (3\Delta s)^3 + \frac{\partial^3 E_x^{i(l+\frac{1}{2},m+\frac{1}{2},n)}}{\partial y^3} (3\Delta s)^3 \right]$$

(3.73)

Thus (3.67) becomes

$$
\frac{\chi_s^2}{72} \left[ \frac{\partial^3 E_x^{i(l+\frac{1}{2},m+\frac{1}{2},n)}}{\partial x^2 \partial y} + \frac{\partial^3 E_x^{i(l+\frac{1}{2},m+\frac{1}{2},n)}}{\partial y^3} + \frac{\partial^3 E_x^{i(l+\frac{1}{2},m+\frac{1}{2},n)}}{\partial z^2 \partial y} \right] \Delta s^2 + O(\Delta s^4)
$$

(3.74)
Together with (3.66), (3.64) is written as

\[
\begin{align*}
\frac{1}{\Delta x}[\frac{9}{8} + \frac{1}{64} (\chi_s^2 - 1)\Delta s^2 \nabla^2] [E^i_{x(t+\frac{1}{2},m+1,n)} - E^i_{x(t+\frac{1}{2},m,n)}] + \left[-\frac{1}{24} - \frac{1}{1728} (\chi_s^2 - 9)\Delta s^2 \nabla^2 \right] \\
\frac{1}{\Delta x}[\frac{9}{8} + \frac{1}{64} (\chi_s^2 - 1)\Delta s^2 \nabla^2] (E^i_{x(t+\frac{1}{2},m+2,n)} - E^i_{x(t+\frac{1}{2},m-1,n)})
\end{align*}
\]

\[
= \frac{\partial E^i_{x(t+\frac{1}{2},m+\frac{1}{2},n)}}{\partial y} + \frac{\chi_s^2}{12} \left[ \frac{\partial^2 E^i_{x(t+\frac{1}{2},m+\frac{1}{2},n)}}{\partial x^2 \partial y} + \frac{\partial^3 E^i_{x(t+\frac{1}{2},m+\frac{1}{2},n)}}{\partial x^3} + \frac{\partial^3 E^i_{x(t+\frac{1}{2},m+\frac{1}{2},n)}}{\partial y^3} \right] \Delta s^2 + O(\Delta s^4)
\]

(3.75)

Therefore, the truncation error of DRP schemes is in general \(O(\Delta t^2) + O(\Delta s^2)\). In the extreme case where the CFL number \(\chi_s\) approaches to zero, the truncation error recovers to \(O(\Delta t^2) + O(\Delta s^4)\) (from (3.75)). This is no surprise because the DRP schemes solve a different system other than that described by Maxwell’s equations.

The phase velocity correction terms in DRP schemes actually introduce additional truncation errors if being used to approximate Maxwell’s equations. However, since it utilizes Fang’s (2,4) scheme stencil and those extra terms are essentially related to time derivatives, it should recover the truncation error of Fang’s (2,4) scheme in the extreme case where \(\chi_s \to 0\), i.e. \(\Delta t \to 0\).

### 3.4 Concluding Remarks

We have described a dispersion-relation-preserving (DRP) methodology to optimize higher order FDTD schemes for large-scale problems. The dispersion error can be minimized in a minimax sense for all propagation angles. Time-domain DRP schemes are obtained by expanding the resulting frequency dependent coefficients in a Taylor series, and by employing either Butterworth or Chebyshev filters. In particular, the dispersion characteristics can be adjusted to yield a smaller local dispersion error at high frequencies than at low frequencies so that less accumulated phase error (up to a pre-assigned frequency) results.
CHAPTER 4

FDTD SUBGRIDDING METHOD

4.1 Introduction

A major problem in EM modelling is local fine structures. These structures usually require a much finer cell size than elsewhere and overload computational costs. An effective way is subgridding, i.e. only using fine mesh where needed, but coarse mesh otherwise. The early version applies FDTD twice in different regions. The results in coarse regions are used as boundary conditions for fine regions [35]. Due to the lack of timely feedback from fine regions, this approach only yields better fine region results and the enhancements in coarse regions are limited. To address this issue, fine regions are embedded into coarse regions dynamically [30]. Fine region boundary values and their timely feedback to coarse regions are provided by field coupling at the coarse-fine region boundary. Two methods are commonly used to accommodate the stability requirement imposed by fine region cell size. One is to use the same time step (STS), which is dictated by fine region cell size, in both regions. The other is to use different time steps determined by coarse and fine region cell sizes respectively. Thus one time step in coarse regions corresponds to multiple time steps (MTS) in fine regions. In the latter case, temporal interpolation is needed for field coupling at the coarse-fine region boundary.
In this chapter, we shall study two fundamental respects of subgridding schemes: accuracy and stability. Accuracy is studied instead of *convergency* because the major error is thought to be the reflection from the coarse-fine region boundary, while the order of convergency in each region remains the same. In the second section, we shall numerically examine the stability of both STS and MTS subgridding schemes with enforcing spatial reciprocity due to the claimed stability [34][69]. The dominant eigenvalue of the system amplification matrices will be examined and the corresponding numerical simulations will run for sufficient time steps to check the late-time behavior.

### 4.2 Accuracy

Because of the hyperbolic nature of time-domain equations, the fine and coarse mesh regions used in subgridding methods need to be adequately coupled at their respective boundaries. Usually, spurious reflections are produced at those boundaries due to the mismatch in both the discrete wave numbers and discrete impedances at each region (both of which depend on the cell size). We shall introduce a systematic way to optimize the coupling coefficients for 2D-subgridding algorithms in order to minimize the spurious reflections over a broad range of frequencies.

Fig. 4.1 depicts the $TE_z$ case of the 2D–subgridding algorithm with a coarse-to-fine refinement ratio of 3:1. Capital letters denote coarse region fields and lower case letters denote fine region fields. Although not strictly necessary, odd ratios are convenient to yield collocated $H_z$ fields and preserve dual grid symmetry. Moreover, to avoid large impedance mismatching, we focus on the 3:1 case. If a higher refinement ratio is needed, the algorithm could be applied successively.
The subgridding algorithm consists of five steps. (1) At time $t_i$, $E_y(t_i)$ is calculated in the coarse region up to the location indexed by 0 in Fig.4.1. (2) Using spatial interpolation, the missing $h_z$ are obtained at time $t_i - \delta t/2$ at the location indexed by 1.5. For collocated field components, the coarse region values are used directly. (3) The ordinary Yee’s algorithm is applied to update the $e$ and $h$ fields in the fine region from the locations indexed by 2 and above at time $t_i$ and $t_i + \delta t/2$, respectively. (4) Reciprocity is used to calculate $H_z(t_i + \delta t/2)$ at the location indexed by 1.5. Other $H_z(t_i + \delta t/2)$ in the coarse region are also calculated by ordinary Yee’s algorithm. (5) Let $i = i + 1$.

![Figure 4.1: Coarse-fine region boundary](image)

4.2.1 Three-point coupling scheme.

In the above subgridding scheme, mesh coupling occurs when obtaining $h_z$ at the location indexed by 1.5 in Fig.4.1. We assume a nearest neighbor three-point coupling
for \( e_y \) at location indexed by 2 in the general form:

\[
e_y = C_1 H_{z1} + C_2 H_{z2} + C_3 h_{z3} \tag{4.1}
\]

Inserting the plane wave solutions (eigenfunctions in the continuum limit) of Maxwell’s equations, we obtain:

\[
H_{z1} = \eta^{-1} E_0 e^{jkh \cos(\theta)/6} e^{jkh \sin(\theta)/3} \tag{4.2}
\]

\[
H_{z2} = \eta^{-1} E_0 e^{jkh \cos(\theta)/6} e^{-2jkh \sin(\theta)/3} \tag{4.3}
\]

\[
h_{z3} = \eta^{-1} E_0 e^{-jkh \cos(\theta)/6} \tag{4.4}
\]

\[
e_y = E_0 \cos \theta \tag{4.5}
\]

where \( h \) is the cell size in coarse region, \( k \) is the wave number, \( E_0 \) is the magnitude of the electric field, \( \eta \) is the wave impedance and \( \theta \) is the incident angle with respect to the positive x-axis. The relative location of \( H_{z1}, H_{z2}, \) and \( h_{z3} \) is depicted in Fig. 4.1.

Substituting Eqs. (4.2)-(4.5) into Eq. (4.1), we obtain:

\[
\eta \cos \theta = C_1 e^{jkh \cos(\theta)/6} e^{jkh \sin(\theta)/3} + C_2 e^{jkh \cos(\theta)/6} e^{-2jkh \sin(\theta)/3} + C_3 e^{-jkh \cos(\theta)/6} \tag{4.6}
\]

To solve for the coefficients in Eq. (4.6), we employ the ansatz below

\[
\int_0^{2\pi} \delta(\theta) d\theta = 0, \quad \int_0^{2\pi} \delta(\theta) \sin(\theta) d\theta = 0, \quad \int_0^{2\pi} \delta(\theta) \cos(\theta) d\theta = 0 \tag{4.7}
\]

where

\[
\delta(\theta) = \eta \cos \theta - [e^{jkh \cos(\theta)/6} (C_1 e^{jkh \sin(\theta)/3} + C_2 e^{-2jkh \sin(\theta)/3} + C_3 e^{-jkh \cos(\theta)/6})] \tag{4.8}
\]

defines the coupling error. Since \( \delta(\theta) \) can be expanded in a Fourier series in terms of \( \cos(n\theta) \) and \( \sin(n\theta) \), it is clear that Eq. (4.7) is equivalent to enforcing the first three terms in the series to be zero. The remaining terms in the series are proportional to
integrals on the form $\int_0^{2\pi} e^{j[\alpha \cos(\theta) + \beta \sin(\theta)]} \sin(n\theta)d\theta$ and $\int_0^{2\pi} e^{j[\alpha \cos(\theta) + \beta \sin(\theta)]} \cos(n\theta)d\theta$
and can be expressed in terms of first kind Bessel functions $J_n(x)$ with real argument $|x| < 1/2$. This is because

$$\int_0^{2\pi} e^{j[\alpha \cos(\theta) + \beta \sin(\theta)]} \sin(n\theta)d\theta = \begin{cases} j2\pi \cos(n\theta_0)J_n\left(\sqrt{\alpha^2 + \beta^2}\right) & \text{for odd } n \\ -2\pi \sin(n\theta_0)J_n\left(\sqrt{\alpha^2 + \beta^2}\right) & \text{for even } n \end{cases}$$

$$\int_0^{2\pi} e^{j[\alpha \cos(\theta) + \beta \sin(\theta)]} \cos(n\theta)d\theta = \begin{cases} j2\pi \sin(n\theta_0)J_n\left(\sqrt{\alpha^2 + \beta^2}\right) & \text{for odd } n \\ 2\pi \cos(n\theta_0)J_n\left(\sqrt{\alpha^2 + \beta^2}\right) & \text{for even } n \end{cases}$$

where $\theta_0 = \arctan(\alpha/\beta)$. Moreover, $\sqrt{\alpha^2 + \beta^2} \leq 1/2$ since the condition $\kappa \leq \lambda/10$ is usually met in practical FDTD simulations. Also, since $|J_n(x)| < 1/(2^n \cdot n!)$ for $|x| < 1$, the contributions of the higher order neglected terms decrease rapidly as the order increases, i.e., if additional $H_z$ and/or $h_z$ are used in the interpolation scheme.

By using two values of $H_z$ and one value of $\kappa_z$, the first three most significant terms in the coupling error are made equal to zero, as enforced by Eq. (4.7).

After solving the integrals in (4.31), we obtain

$$\begin{bmatrix} J_0(\sqrt{5}\kappa/6) & J_0(\sqrt{17}\kappa/6) & J_0(\kappa/6) \\ J_1(\sqrt{5}\kappa/6)/\sqrt{5} & -2J_1(\sqrt{17}\kappa/6)/\sqrt{17} & 0 \\ J_1(\sqrt{5}\kappa/6)/\sqrt{5} & J_1(\sqrt{17}\kappa/6)/\sqrt{17} & -J_1(\kappa/6) \end{bmatrix} \cdot \begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -i\eta/2 \end{bmatrix}$$

(4.9)

Solving for $C_1$, $C_2$, and $C_3$ yields

$$C_1 = -j\sqrt{85\eta}J_1\left(\sqrt{17}\kappa/6\right)J_0(\kappa/6)/17J_0\left(\sqrt{17}\kappa/6\right)J_1(\kappa/6)J_1\left(\sqrt{5}\kappa/6\right) + 2\sqrt{85}J_0\left(\sqrt{5}\kappa/6\right)J_1(\kappa/6)J_1\left(\sqrt{17}\kappa/6\right)J_1(\sqrt{17}\kappa/6)J_0(\kappa/6)J_1\left(\sqrt{5}\kappa/6\right)
$$

$$C_2 = -j\sqrt{85\eta}J_0(\kappa/6)J_1\left(\sqrt{5}\kappa/6\right)/2\sqrt{85}J_0\left(\sqrt{17}\kappa/6\right)J_1(\kappa/6)J_1\left(\sqrt{5}\kappa/6\right) + 80J_0\left(\sqrt{5}\kappa/6\right)J_1(\kappa/6)J_1\left(\sqrt{17}\kappa/6\right)J_0(\kappa/6)J_1\left(\sqrt{5}\kappa/6\right)
$$

$$C_3 = j\eta\sqrt{17}J_0\left(\sqrt{17}\kappa/6\right)J_1\left(\sqrt{5}\kappa/6\right) + 2\sqrt{5}J_1\left(\sqrt{17}\kappa/6\right)J_0\left(\sqrt{5}\kappa/6\right)/2\sqrt{17}J_0\left(\sqrt{17}\kappa/6\right)J_1(\kappa/6)J_1\left(\sqrt{5}\kappa/6\right) + 6J_1(\sqrt{17}\kappa/6)J_0(\kappa/6)J_1\left(\sqrt{5}\kappa/6\right)$$

(4.10)
By expanding Eq. (4.10) in a Taylor series on $kh$ and transforming the results to time domain, we obtain:

$$
\begin{align*}
 j\omega \epsilon C_1/3 & \rightarrow 2/3 - \left(1/24\chi^2\right)\Delta t^2 \partial_t^2 + O((kh)^4) \\
 j\omega \epsilon C_2/3 & \rightarrow 1/3 - \left(7/144\chi^2\right)\Delta t^2 \partial_t^2 + O((kh)^4) \\
 j\omega \epsilon C_3/3 & \rightarrow -1 - \left(1/48\chi^2\right)\Delta t^2 \partial_t^2 + O((kh)^4)
\end{align*}
$$

(4.11)

where $\chi$ is the coarse region CFL number, $\Delta t$ is the discrete time step, $\omega$ is the angular frequency and $\epsilon$ is the permittivity. Note that in deriving the above equations, we have applied the following identities

$$
\chi = \sqrt{2} \frac{c \Delta t}{h}, \quad kh = \frac{c}{\epsilon} h = \omega \sqrt{2} \frac{\chi}{\Delta t}
$$

(4.12)

The coefficients in Eq. (4.11) are normalized so that the right hand sides can be inserted in the FDTD update directly. The second order time derivatives in Eq. (4.11) can also be replaced by second order space derivatives via Helmholtz equation. The derivative terms can be thought of as dispersion correction terms which compensate for the difference in propagation constants and impedances between the fine and coarse regions. By dropping the second order time derivatives in $C_1$, $C_2$ and $C_3$, the scheme would reduce to a simple linear interpolation of an elliptic equation (static limit or conventional subgridding).

### 4.2.2 Higher order, five-point coupling scheme.

Higher order subgridding schemes utilize more than three points for the nearest neighbor coupling in Eq. (4.1). We have also examined five-point layout schemes. The optimized coefficients for a given five-point layout scheme can be derived in a similar way as done before for the three-point scheme. While different five-point coupling layouts are possible, a satisfactory one uses two extra values for $H_z$ in the coarse region, viz., one at the grid point directly below $H_{z1}$ and the other at the
grid point directly above $H_{z,2}$ in Fig. 1. An equation similar to Eq. (4.6) (now with five coefficients to solve for) can then be obtained, whose coefficients are solved using an ansatz similar to Eq. (4.7), now also involving projections on $\sin(2\theta)$ and $\cos(2\theta)$ also. As mentioned before, the use of more points allows for these additional terms in the Fourier expansion of the coupling error to be made equal to zero, and hence produce a smaller overall residual error. The obvious drawbacks are loss of sparsity and increased computational cost.

4.2.3 Numerical Results

We have run simulations on a coarse-fine region boundary of large FDTD domains avoiding interference from the corners and the mesh termination. Figs. 4.2 and 4.3 show the spurious relative reflection levels from the subgridding boundary as a function of frequency and angle. The cell size is equal to $\lambda/10$ at the highest frequency considered, $f = 10$ GHz, and the CFL number is $\chi = 1.0$ (fine region value).

From Fig. 4.2 we observe that, at small incidence angles, the three-point linear interpolation has a performance comparable to the five-point linear interpolation. This is expected because for plane wave normal incidence a larger stencil on the transverse direction is not important. For oblique incidence angles, the performance of the three-point linear interpolation deteriorates. On the other hand, if the second order derivative terms are used, the performance is better than linear interpolation at 45°, but deteriorates at 0°, exhibiting complementary behavior according to the incidence angle. This behavior can be explored when implementing subgridding schemes in different regions of a FDTD grid. As expected, the five-point linear interpolation has a smaller spurious reflection overall and largest reflection levels at 0°. Note that the
ansatz in Eq. (4.7) is not unique or not necessarily optimal in any problem. Other possibilities could also be explored. In particular, any a priori information about the geometry of the problem, if available, can be used in designing different schemes.

4.3 Stability

The stability of FDTD subgridding schemes is first studied by their system amplification matrices [63]. Let us write a general time-domain method as

\[
\begin{pmatrix}
\vec{E}^n \\
\vec{H}^{n+\frac{1}{2}}
\end{pmatrix} = [A]
\begin{pmatrix}
\vec{E}^{n-1} \\
\vec{H}^{n-\frac{1}{2}}
\end{pmatrix}
\] (4.13)

where \([A]\) denotes the system amplification matrix. To check out the late-time behavior, we run simulations inside Perfect Electric Conducting (PEC) boxes of various sizes. A magnetic point source is set close to the lower left corner of the box with

Figure 4.2: Reflection levels at 0°.
maximum magnitude less than unity. The electric field at a location other than the source and some geometric centers is observed. By setting up a differentiated Gaussian source, the total energy inside the box is limited. Whenever the observation is larger than unity, we check out energy conservation. If energy is not conserved, we claim that a simulation is unstable and that time step is recorded as the “bursting point”. Otherwise, if the observation remains less than unity for at least ten million time steps, we check out where the recently recorded observations tend to grow. Only those simulations with conserved energy and non-growing history will be claimed stable.

The above criterion is nor strict in the following senses: 1) The maximum possible value is different according to different PEC box sizes. A larger PEC box has smaller maximum value. The bursting point only indicates that a simulation is unstable,
while its value should not be used to predict the level of amplification. Also, being
stable in ten million time steps may not imply real stability. However, it is more than
sufficient for most practical simulations. 2) The observation is possibly at a “null”
point. To avoid this situation, two points close to the PEC boundary and separated
far enough without sharing any geometrical similarities are actually observed.

4.3.1 STS Subgridding Scheme

For STS subgridding schemes, the system can be described by

$$\begin{pmatrix}
\vec{U}^{n+\frac{1}{2}} \\
\vec{u}^{n+\frac{1}{2}}
\end{pmatrix} =
\begin{pmatrix}
A_{CC} & A_{CF} \\
A_{FC} & A_{FF}
\end{pmatrix}
\begin{pmatrix}
\vec{U}^{n-\frac{1}{2}} \\
\vec{u}^{n-\frac{1}{2}}
\end{pmatrix}
$$

(4.14)

where \(\vec{U}^{n+\frac{1}{2}}\) contains all coarse region unknowns and \(\vec{u}^{n+\frac{1}{2}}\) contains all fine region
unknowns, both at the current time step. \(A_{CC}\) represents how the current coarse
region values use the previous coarse region values, \(A_{CF}\) represents how the current
coarse region values use the previous fine region values, \(A_{FC}\) represents how the
current fine region values use the previous coarse region values and \(A_{FF}\) represents
how the current fine region values use the previous fine region values.

Figs.4.4 and 4.5 depict the ordinary coarse-fine region boundary and a corner
respectively. In order to calculate the fine region boundary electric field values (e.g.
\(E_{y1}\)), we need the coarse region magnetic field values which is missing because of the
boundary truncation (e.g. \(h'_{z1}\)). Here, a linear interpolation scheme is employed, i.e.

$$h'_{z1} = \frac{2}{3}H_{z1} + \frac{1}{3}H_{z2}, \quad h'_{z2} = \frac{1}{3}H_{z1} + \frac{2}{3}H_{z2}, \quad h'_{z3} = H_{z2}$$

(4.15)

When calculating the coarse region boundary magnetic field values (e.g. \(H_{z2}\)), we
need the electric field values along the coarse-fine region boundary (e.g. \(E_{y2}\)). By

84
enforcing spatial reciprocity, we have [73]

\[ E_{y2} = \frac{1}{3} \left( \frac{1}{3} (e_{y1} + e_{y5}) + \frac{2}{3} (e_{y2} + e_{y4}) + e_{y3} \right) \] (4.16)

To keep at least first order spatial accuracy, we apply the following unbalanced differencing scheme at the boundary

\[ e_{y1}^n = e_{y1}^{n-1} + \frac{\Delta t}{2\Delta h} \left( h_{z1}^{n-\frac{1}{2}} - h_{z1}^{n-\frac{3}{2}} \right) \] (4.17)

where \( \Delta h \) is the fine region cell size.

Note that the coarse region magnetic field values closest to the coarse-fine region boundary are calculated by central differencing scheme in this layout. Another layout
is to expand the fine region 1/3 cell (coarse) toward left. So the coarse region magnetic field values closest to the boundary are calculated by unbalanced differencing scheme and the fine region electric field values along the boundary are calculated by central differencing scheme. Since the unbalanced differencing scheme is of first order accuracy, choosing fine region electric field values along the boundary to be calculated by unbalanced differencing scheme is obviously a better choice.

\[ h'_{z1} = \frac{2}{3}H_{z1} + \frac{1}{3}H_{z3}, \quad h''_{z1} = \frac{2}{3}H_{z2} + \frac{1}{3}H_{z3} \]  

(4.18)

then by reciprocity, we should write

\[ E_{x3} = E_{x3} + \frac{1}{3}e_{x1}, \quad E_{y3} = E_{y3} + \frac{1}{3}e_{y1} \]  

(4.19)

Obviously, this introduces errors when calculating \( H_{z3} \) (also \( H_{z1} \) and \( H_{z2} \)). Other corner layouts may be applied, which primarily involve shifting the location of the
fine region field components. For different layouts, we need to choose between keeping first order accuracy for $H_{z3}$ (also $H_{z1}$ and $H_{z2}$) and enforcing reciprocity. Here, we shall keep reciprocity because stability is our main concern.

Alternative to Eq. (4.18), we may simply ignore corners and write

$$h'_{z1} = H_{z1}, \quad h''_{z1} = H_{z2}$$

(4.20)

Then $h'_{z1}$ and $h''_{z1}$ will lose accuracy in spacial interpolation. This may not present a severe problem if $H_{z1}$ and $H_{z2}$ are understood as the average in those cells. In fact, the choice between these two treatments of corners is made upon stability. In the following, we refer to the first scheme as “reciprocal with corners” and the second scheme as “reciprocal without corners”.

**Reciprocal with corners**

The computational domain is characterized by $m*n$, where $m$ refers to the distance form the coarse-fine region boundary to the PEC boundary and $n$ refers to the fine region total size, both are in terms of coarse region cell size. $m$ and $n$ range from three to ten in our tests, thus the total computation size ranges from $9 \times 9$ to $30 \times 30$.

The system amplification matrix is depicted in Fig.4.6, which shows an example for a $9 \times 9$ PEC box with subgridding region of $3 \times 3$ ($9 \times 9$ if measured in terms of fine region cell size) at the center. For this system, the largest eigenvalue drops from 1.420754 as we decrease the CFL number from 1.03. When the CFL number drops to 1.01, the largest eigenvalue becomes unity and keeps that value for smaller CFL numbers. For larger computational sizes, the largest CFL numbers corresponding to unity dominant eigenvalues are 1.0.
Figure 4.6: Example of the system amplification matrix of “reciprocal with corners” STS subgridding scheme.
Numerical tests show that this scheme is unstable for all computational domain sizes being tested. The instability is not of late-time type. For example, the fastest bursting point is 360, corresponding to a $3 \times 3$ PEC box and the slowest bursting point is 3689, corresponding to a $9 \times 10$ PEC box.

**Reciprocal without corners**

The system amplification matrix is a little sparser than the previous and looks similar to Fig.4.6. The same computational domain sizes were analyzed. For the smallest computation domain size, the largest eigenvalue drops from 1.420752 as we decrease the CFL number from 1.03. When the CFL number is equal to 1.01, the largest eigenvalue becomes 1.0 and keeps that value for subsequent smaller CFL numbers. Contrary to the previous case, FDTD simulations show that this scheme is stable for all sizes being tested within ten million time steps without any instability observed.

### 4.3.2 MTS Subgridding Scheme

The timing of MTS schemes is illustrated in Fig.4.7. When calculating $e^n$ and $e^{n+\frac{1}{2}}$ in the fine region, we lack of $H^{n-\frac{1}{6}}$ and $H^{n+\frac{1}{6}}$ in the coarse region for spacial interpolation. Also, when we calculate $H^{n+\frac{1}{2}}$ by imposing reciprocity, we should use $e^n$ in the fine region. These two requirements make accurate temporal interpolation (obtaining $H^{n-\frac{1}{6}}$ and $H^{n+\frac{1}{6}}$ by using $H^{n+\frac{1}{2}}$ and $H^{n-\frac{1}{2}}$) hard to implement if spacial reciprocity is also required. One solution is to use temporal extrapolation (obtaining $H^{n-\frac{1}{6}}$ and $H^{n+\frac{1}{6}}$ by using $H^{n-\frac{1}{2}}$ and $H^{n-\frac{3}{2}}$ etc.). Another is to use $H^{n-\frac{1}{2}}$ as $H^{n-\frac{5}{6}}$ and $H^{n+\frac{1}{6}}$ [73], which is customarily referred to as Zeroth Order Hold (ZOH) in digital signal processing [48]. ZOH inevitably introduces spectrum distortion to input signals.
While in reality, temporal extrapolation is found to be much more unstable than ZOH. For this reason, we choose ZOH.

\[
\begin{array}{cccc}
H_n^{n-\frac{1}{2}} & E_n & H_n^{n+\frac{1}{2}} \\
\dot{h}_n^{n-\frac{1}{2}} & \dot{e}_n^{n-\frac{1}{2}} & \dot{h}_n^{n+\frac{1}{2}} \ \\
E_n & \dot{e}_n & E_n^{n+\frac{1}{2}}
\end{array}
\]

Figure 4.7: Timing procedure of MTS subgridding scheme, superscripts denote time steps.

The timing procedures could be described mathematically by Eqs. (4.21)-(4.24)

\[
\begin{align*}
\dot{u}_n^{n-\frac{1}{6}} &= A_{FF} \dot{u}_n^{n-\frac{1}{2}} + A_{FC} \dot{U}_n^{n-\frac{1}{2}} \\
\dot{u}_n^{n+\frac{1}{6}} &= A_{2FF} \dot{u}_n^{n-\frac{1}{2}} + (A_{FF} + 1) \cdot A_{FC} \dot{U}_n^{n-\frac{1}{2}} \\
\dot{u}_n^{n+\frac{1}{2}} &= A_{3FF} \dot{u}_n^{n-\frac{1}{2}} + (A_{2FF} + A_{FF} + 1) \cdot A_{FC} \dot{U}_n^{n-\frac{1}{2}} \\
\dot{U}_n^{n+\frac{1}{2}} &= A_{CC} \dot{U}_n^{n-\frac{1}{2}} + A_{CF} \dot{u}_n^{n-\frac{1}{6}} \\
&= (A_{CC} + A_{CF} \cdot A_{FC}) \dot{U}_n^{n-\frac{1}{2}} + A_{CF} \cdot A_{FF} \dot{u}_n^{n-\frac{1}{2}}
\end{align*}
\]

where all symbols have the same meaning as before. Therefore, the system amplification matrix is

\[
A = \begin{bmatrix}
A_{CC} + A_{CF} \cdot A_{FC} & A_{CF} \cdot A_{FF} \\
(A_{2FF} + A_{FF} + 1) \cdot A_{FC} & A_{3FF}
\end{bmatrix}
\]

Fig.4.8 illustrates the amplification matrix of a 9×9 PEC box with a 3×3 (9×9 if measured in fine cell size) fine region at the center. Fig.4.9 shows dominant eigenvalues vs CFL numbers, where the computational domain is also characterized by \( m \star n \). As can be seen, the dominant eigenvalues are always larger than unity regardless of the CFL numbers. As the computational domain increases, the dominant eigenvalues for a given CFL number decrease but never reach unity. This is verified by recorded
bursting points of corresponding simulations shown in Fig.4.10. The instability is not very well understood now, while we believe that it relates to the lack of temporal reciprocity during the interaction between the coarse-fine boundary values.

4.4 Concluding Remarks

In this Chapter, we presented an approach to optimize the spacial interpolation coefficients of FDTD subgridding schemes. However, there is no guarantee on the stability yet, especially for the most useful MTS scheme. Although late-time unstable, the MTS scheme can still be useful as long as late-time instability does not show up before terminating simulation. The biggest problem of late-time instability is that it
Figure 4.9: Dominant eigenvalues of MTS subgridding scheme.

Figure 4.10: Bursting points of different computation domain sizes and CFL numbers.
is hard to estimate when it will occur for an arbitrary problem. Due to this reason, we shall pursue alternative solutions to very large-scale, locally fine problems in the next chapter.
CHAPTER 5

HYBRID MESH AND FETD/INS-FDTD HYBRID METHOD

5.1 Introduction

Vector based Finite-Element Time-Domain (FETD) methods have many advantages over FDTD methods. First of all, FETD is capable of modelling complex structures and rapid field variations by unstructured grids [29],[60]. Second, higher order accuracy is easy to achieve by hierarchal vector basis functions at the presence of inhomogeneous medium [61]. As to FDTD, the application of higher order schemes to inhomogeneous medium without decreasing global accuracy is still an open question. Finally, FETD can be made unconditionally stable by using the Generalized-Θ method [38],[39]. Although FDTD can also be made unconditionally stable by Alternating-Direction Implicit (ADI) methods [16],[43],[76], or the Crank-Nicolson scheme [70], they become more expensive for higher order schemes and it is still not clear how to apply them to unstructured grids.

However, the major drawback of FETD is the computational cost required to solve large linear systems. This problem can be mitigated in some specific problems. In this chapter, we shall focus on large-scale problems involving locally fine structures and possibly highly resonant structures. This type of problem has several characteristics:
1) Unstructured grids are not necessary everywhere. 2) Memory requirement should be kept at a minimum. 3) Late-time instability is not allowed.

The first implication suggests using brick elements wherever possible within the grid. Brick elements have two obvious advantages over tetrahedral elements in this case. First of all, tetrahedral elements increase the number of unknowns by 50% compared to brick elements due to the fact that one brick element consists of at least five tetrahedra. Second, brick elements are easier to generate.

A natural way to connect bricks and tetrahedra is using pyramidal elements [79],[47]. In section §5.2, we shall introduce a new type of element - composite elements - to alleviate the difficulty of mesh generation involving both bricks and tetrahedra.

To reduce memory requirements, we shall generate the matrix on-the-fly in the brick region. Thus, matrix preconditioning methods have limited applicability. Because many efficient preconditioning methods, such as the incomplete-LU family, need storage no less than that required by the matrix. Sections §5.3 - §5.4 will describe several attempts to solve this problem.

## 5.2 Composite Elements

To connect brick and tetrahedral elements, it is natural to use pyramids [10]. However, pyramids are rarely used in FETD and most mesh generators do not accommodate such elements. To transfer from tetrahedra to bricks, we can approach in two different ways. One is to design general-purpose algorithms to generate pyramids. Since only one layer of pyramids are needed, this approach is expensive. On the other hand, it is much simpler if we can encapsulate pyramids in a way that
the computational domain only see the need to generate tetrahedra and bricks. The latter is the idea behind composite elements.

5.2.1 Edge Element Basis Functions for Pyramidal Elements

Tangential vector basis functions are based on the theory of Nedelec [44]. In a three-dimensional vector space, let us define $\mathcal{H}(\text{curl})$ to be

$$\mathcal{H}(\text{curl}) = \{ u \in (L^2(\Omega))^3 | \nabla \times u \in (L^2(\Omega))^3 \}$$

(5.1)

All tangential vector basis functions are in $\mathcal{H}(\text{curl})$. This requires that the basis functions and their tangential components are continuous.

The edge element basis functions are the lowest order tangential basis functions. For tetrahedral elements, let $\lambda_i$ and $\lambda_j$ denote the nodal basis functions associated with nodes $i$ and $j$

$$\lambda_i = \frac{\begin{vmatrix} 1 & x & y & z \\ 1 & x_{i+1} & y_{i+1} & z_{i+1} \\ 1 & x_{i+2} & y_{i+2} & z_{i+2} \\ 1 & x_{i+3} & y_{i+3} & z_{i+3} \\ 1 & x_i & y_i & z_i \\ 1 & x_{i+1} & y_{i+1} & z_{i+1} \\ 1 & x_{i+2} & y_{i+2} & z_{i+2} \\ 1 & x_{i+3} & y_{i+3} & z_{i+3} \end{vmatrix}}{\begin{vmatrix} 1 & x_{i+1} & y_{i+1} \\ 1 & x_{i+2} & y_{i+2} \\ 1 & x_{i+3} & y_{i+3} \end{vmatrix}}$$

(5.2)

where the subscripts are cyclic indices and their meanings are illustrated in Fig. 5.1.

The basis function associated with the edge connecting nodes $i$ and $j$ is

$$\vec{w}_{ij} = \lambda_i \nabla \lambda_j - \lambda_j \nabla \lambda_i$$

(5.3)

For pyramidal elements, the nodal basis functions defined in a master element are rational functions [80]

$$\lambda_1 = \frac{1}{4} \left[ (1 + u_1)(1 + u_2) - u_3 + \frac{u_1 u_2 u_3}{1 - u_3} \right]$$

(5.4)
Figure 5.1: A Tetrahedral Element

Figure 5.2: A Pyramidal Element
\[ \lambda_2 = \frac{1}{4} \left[ (1 + u_1)(1 - u_2) - u_3 - \frac{u_1 u_2 u_3}{1 - u_3} \right] \quad (5.5) \]

\[ \lambda_3 = \frac{1}{4} \left[ (1 - u_1)(1 - u_2) - u_3 + \frac{u_1 u_2 u_3}{1 - u_3} \right] \quad (5.6) \]

\[ \lambda_4 = \frac{1}{4} \left[ (1 - u_1)(1 + u_2) - u_3 - \frac{u_1 u_2 u_3}{1 - u_3} \right] \quad (5.7) \]

\[ \lambda_5 = u_3 \quad (5.8) \]

where \( u_1, u_2, u_3 \) and nodes 1-5 are defined in Fig. 5.2. Note that the limits of basis functions \( \lambda_1-4 \) are zero at the tip. The edge basis functions are defined slightly different from Ref. [10]. For horizontal edges, they are

\[ \vec{W}_1 = -2\tilde{W}_{12} = -2[\lambda_1 \nabla_u (\lambda_2 + \lambda_3) - \lambda_2 \nabla_u (\lambda_4 + \lambda_1)] \quad (5.9) \]

\[ \vec{W}_2 = -2\tilde{W}_{23} = -2[\lambda_2 \nabla_u (\lambda_3 + \lambda_4) - \lambda_3 \nabla_u (\lambda_1 + \lambda_2)] \quad (5.10) \]

\[ \vec{W}_3 = 2\tilde{W}_{34} = 2[\lambda_3 \nabla_u (\lambda_4 + \lambda_1) - \lambda_4 \nabla_u (\lambda_2 + \lambda_3)] \quad (5.11) \]

\[ \vec{W}_4 = 2\tilde{W}_{41} = 2[\lambda_4 \nabla_u (\lambda_1 + \lambda_2) - \lambda_1 \nabla_u (\lambda_3 + \lambda_4)] \quad (5.12) \]

For oblique edges, they are defined similarly to tetrahedral elements by

\[ \vec{W}_5 = \sqrt{3}\tilde{W}_{15} = \sqrt{3}(\lambda_1 \nabla_u \lambda_5 - \lambda_5 \nabla_u \lambda_1) \quad (5.13) \]

\[ \vec{W}_6 = -\sqrt{3}\tilde{W}_{25} = -\sqrt{3}(\lambda_2 \nabla_u \lambda_5 - \lambda_5 \nabla_u \lambda_2) \quad (5.14) \]

\[ \vec{W}_7 = \sqrt{3}\tilde{W}_{35} = \sqrt{3}(\lambda_3 \nabla_u \lambda_5 - \lambda_5 \nabla_u \lambda_3) \quad (5.15) \]

\[ \vec{W}_8 = -\sqrt{3}\tilde{W}_{45} = -\sqrt{3}(\lambda_4 \nabla_u \lambda_5 - \lambda_5 \nabla_u \lambda_4) \quad (5.16) \]

where \( \tilde{W}_{ij} \) represents the originally defined edge basis function with the positive direction pointing from node \( i \) to node \( j \) [10], \( \tilde{W}_k \) represents the modified basis function whose subscript is its local edge index.
The major differences between the above definitions and the original ones are the scaling factors and the definition of positive directions. The scaling factors are defined in such a way that the following condition is satisfied in a master element

\[ \vec{W}_k \cdot \hat{n}_{ij} = \delta_{ij} \quad (5.17) \]

where \( \hat{n}_{ij} \) is the unit vector pointing from node \( i \) to node \( j \), and \( k \) indicates the corresponding basis function index. The redefined positive directions are compatible with the positive directions of those tetrahedral and brick elements generated by popular commercial mesh generating libraries, such as the MODULEF [3]. For an arbitrarily shaped pyramid, the edge basis functions in terms of \( (x_1, x_2, x_3) \) are obtained by

\[ \vec{w}_k = [G]^{-1} \vec{W}_k \quad (5.18) \]

and

\[ \nabla_x \times \vec{w}_k = \frac{[G]^T}{\text{det}[G]} \nabla_u \times \vec{W}_k \quad (5.19) \]

where \( G \) is the Jacobian matrix defined as

\[ [G]_{ij} = \frac{\partial x_j}{\partial u_i} \quad (5.20) \]

Eqs.(5.18)–(5.20) are readily verified by considering a single component.

5.2.2 Encapsulation

A composite element consists of seven elementary elements: five pyramids and two tetrahedra. To surround a tetrahedral region, six basic types of composite elements are needed. These are illustrated in Figs. 5.3-5.8. Note that each type in turn splits into two subtypes. The reason is numerical accuracy and we shall discuss it latter.

The composite elements are always cube-shaped. Thus they can be generated by mesh generators for structured grids. In finite-element methods, the global matrices
Figure 5.3: Type 0 composite element

Figure 5.4: Type 1 composite element
Figure 5.5: Type 2 composite element

Figure 5.6: Type 3 composite element
Figure 5.7: Type 4 composite element

Figure 5.8: Type 5 composite element
are constructed by mapping local mass and stiffness matrices. Thus a composite element can be defined by its local mass and stiffness matrices. In the following, we shall use Type 11 as an example to illustrate how to construct a composite element for FETD methods.

The mass matrix \([T]\) is defined by

\[
[T]_{ij} = \int \int \int_{v} \vec{w}_i \cdot \vec{w}_j dv
\]

and the stiffness matrix \([S]\) is defined by

\[
[S]_{ij} = \int \int \int_{v} \nabla \times \vec{w}_i \cdot \nabla \times \vec{w}_j dv
\]

By assuming cube-shaped composite elements, all entries of the mass and stiffness matrices of a pyramidal element can be evaluated analytically. Let us assume that the edge length of a composite element is \(l\) and study the pyramid that points to the right in Fig. 5.4. By arranging the coordinate axes as follows

\[
\hat{u}_1 = \hat{x}, \quad \hat{u}_2 = -\hat{z}, \quad \hat{u}_3 = \hat{y}
\]

the Jacobian matrix is

\[
[G] = \begin{bmatrix}
l & 0 & 0 \\
0 & 0 & -l \\
0 & l & 0 \\
\end{bmatrix}
\]

Substitute Eq.(5.23) into Eq.(5.18) and Eq.(5.19), and then in turn substitute the results in Eq.(5.21) and Eq.(5.22), we obtain the following mass and stiffness matrices

\[
[T] = \frac{l^3}{4}
\]

\[
\begin{bmatrix}
\frac{8}{35} & \frac{1}{90} & \frac{4}{15} & \frac{1}{90} & \frac{1}{80\sqrt{3}} & \frac{1}{80\sqrt{3}} & \frac{1}{80\sqrt{3}} & \frac{1}{80\sqrt{3}} \\
\frac{8}{35} & \frac{1}{90} & \frac{4}{15} & \frac{1}{90} & \frac{1}{80\sqrt{3}} & \frac{1}{80\sqrt{3}} & \frac{1}{80\sqrt{3}} & \frac{1}{80\sqrt{3}} \\
\frac{8}{35} & \frac{1}{90} & \frac{4}{15} & \frac{1}{90} & \frac{1}{80\sqrt{3}} & \frac{1}{80\sqrt{3}} & \frac{1}{80\sqrt{3}} & \frac{1}{80\sqrt{3}} \\
\frac{8}{35} & \frac{1}{90} & \frac{4}{15} & \frac{1}{90} & \frac{1}{80\sqrt{3}} & \frac{1}{80\sqrt{3}} & \frac{1}{80\sqrt{3}} & \frac{1}{80\sqrt{3}} \\
\frac{8}{35} & \frac{1}{90} & \frac{4}{15} & \frac{1}{90} & \frac{1}{80\sqrt{3}} & \frac{1}{80\sqrt{3}} & \frac{1}{80\sqrt{3}} & \frac{1}{80\sqrt{3}} \\
\frac{8}{35} & \frac{1}{90} & \frac{4}{15} & \frac{1}{90} & \frac{1}{80\sqrt{3}} & \frac{1}{80\sqrt{3}} & \frac{1}{80\sqrt{3}} & \frac{1}{80\sqrt{3}} \\
\frac{8}{35} & \frac{1}{90} & \frac{4}{15} & \frac{1}{90} & \frac{1}{80\sqrt{3}} & \frac{1}{80\sqrt{3}} & \frac{1}{80\sqrt{3}} & \frac{1}{80\sqrt{3}} \\
\frac{8}{35} & \frac{1}{90} & \frac{4}{15} & \frac{1}{90} & \frac{1}{80\sqrt{3}} & \frac{1}{80\sqrt{3}} & \frac{1}{80\sqrt{3}} & \frac{1}{80\sqrt{3}} \\
\end{bmatrix}
\]
and

\[
[S] = \frac{l}{4} \begin{bmatrix}
\frac{34}{9} & -\frac{2}{9} & \frac{14}{9} & \frac{2}{9} & -\frac{16}{3} & \frac{16}{3} & \frac{8}{3} & \frac{8}{3} \\
\frac{34}{9} & \frac{2}{9} & \frac{14}{9} & \frac{8}{3} & \frac{8}{3} & \frac{16}{3} & \frac{16}{3} & \frac{8}{3} \\
\frac{34}{9} & -\frac{2}{9} & \frac{14}{9} & \frac{8}{3} & \frac{8}{3} & \frac{16}{3} & \frac{16}{3} & \frac{8}{3} \\
\frac{34}{9} & \frac{2}{9} & \frac{14}{9} & \frac{8}{3} & \frac{8}{3} & \frac{16}{3} & \frac{16}{3} & \frac{8}{3} \\
\frac{34}{9} & -\frac{2}{9} & \frac{14}{9} & \frac{8}{3} & \frac{8}{3} & \frac{16}{3} & \frac{16}{3} & \frac{8}{3} \\
\frac{34}{9} & \frac{2}{9} & \frac{14}{9} & \frac{8}{3} & \frac{8}{3} & \frac{16}{3} & \frac{16}{3} & \frac{8}{3} \\
\frac{34}{9} & -\frac{2}{9} & \frac{14}{9} & \frac{8}{3} & \frac{8}{3} & \frac{16}{3} & \frac{16}{3} & \frac{8}{3} \\
\frac{34}{9} & \frac{2}{9} & \frac{14}{9} & \frac{8}{3} & \frac{8}{3} & \frac{16}{3} & \frac{16}{3} & \frac{8}{3}
\end{bmatrix}
\] (5.25)

where both \([T]\) and \([S]\) are symmetric and only half of them are shown. The mass and stiffness matrices of the other four pyramids are computed analytically in the same way. The mass and stiffness matrices of the two tetrahedra are computed by Eqs.(5.2)-(5.3) and Eqs.(5.21)-(5.22). Because their shapes are fixed, we can also evaluate the entries analytically. For example, the mass and stiffness matrices of the tetrahedron at the lower-left corner are

\[
[T] = \frac{l^3}{4} \begin{bmatrix}
\frac{1}{5} & \frac{1}{10} & -\frac{\sqrt{3}}{30} & 0 & -\frac{\sqrt{3}}{24} & -\frac{\sqrt{3}}{24} \\
\frac{1}{5} & \frac{1}{10} & -\frac{\sqrt{3}}{30} & 0 & -\frac{\sqrt{3}}{24} & -\frac{\sqrt{3}}{24} \\
\frac{1}{5} & \frac{1}{10} & -\frac{\sqrt{3}}{30} & 0 & -\frac{\sqrt{3}}{24} & -\frac{\sqrt{3}}{24} \\
\frac{1}{5} & \frac{1}{10} & -\frac{\sqrt{3}}{30} & 0 & -\frac{\sqrt{3}}{24} & -\frac{\sqrt{3}}{24} \\
\frac{1}{5} & \frac{1}{10} & -\frac{\sqrt{3}}{30} & 0 & -\frac{\sqrt{3}}{24} & -\frac{\sqrt{3}}{24} \\
\frac{1}{5} & \frac{1}{10} & -\frac{\sqrt{3}}{30} & 0 & -\frac{\sqrt{3}}{24} & -\frac{\sqrt{3}}{24} \\
\frac{1}{5} & \frac{1}{10} & -\frac{\sqrt{3}}{30} & 0 & -\frac{\sqrt{3}}{24} & -\frac{\sqrt{3}}{24} \\
\frac{1}{5} & \frac{1}{10} & -\frac{\sqrt{3}}{30} & 0 & -\frac{\sqrt{3}}{24} & -\frac{\sqrt{3}}{24}
\end{bmatrix}
\] (5.26)

and

\[
[S] = l \begin{bmatrix}
\frac{1}{3} & -\frac{2\sqrt{3}}{3} & \frac{2\sqrt{3}}{3} & \frac{2\sqrt{3}}{3} \\
\frac{1}{3} & -\frac{2\sqrt{3}}{3} & \frac{2\sqrt{3}}{3} & \frac{2\sqrt{3}}{3} \\
\frac{1}{3} & -\frac{2\sqrt{3}}{3} & \frac{2\sqrt{3}}{3} & \frac{2\sqrt{3}}{3} \\
\frac{1}{3} & -\frac{2\sqrt{3}}{3} & \frac{2\sqrt{3}}{3} & \frac{2\sqrt{3}}{3} \\
\frac{1}{3} & -\frac{2\sqrt{3}}{3} & \frac{2\sqrt{3}}{3} & \frac{2\sqrt{3}}{3} \\
\frac{1}{3} & -\frac{2\sqrt{3}}{3} & \frac{2\sqrt{3}}{3} & \frac{2\sqrt{3}}{3} \\
\frac{1}{3} & -\frac{2\sqrt{3}}{3} & \frac{2\sqrt{3}}{3} & \frac{2\sqrt{3}}{3} \\
\frac{1}{3} & -\frac{2\sqrt{3}}{3} & \frac{2\sqrt{3}}{3} & \frac{2\sqrt{3}}{3}
\end{bmatrix}
\] (5.27)

where both \([T]\) and \([S]\) are symmetric and only half of them are shown.

After obtaining the local mass and stiffness matrices of each element, we can map the entries to a group mass and stiffness matrices. The group mass and stiffness matrices serve as an intermediate abstract to construct global matrices. They are
21 \times 21 matrices and will be shown here due to limited spaces. The other composite elements are constructed in the same way.

Before leaving this section, we shall briefly discuss how to use composite elements. There are two major different arrangements of composite elements on the interface between tetrahedra and cubes, as shown in Fig. 5.9. The one-directional arrangement only needs one subtype of composite elements and is easy to use. The diamond arrangement needs two subtypes of composite elements and is relatively difficult to arrange. However, the diamond arrangement tends to generate more randomly oriented tetrahedra. This is preferred because numerical dispersion errors tend to cancel each other if tetrahedra are randomly oriented [74]. Therefore, the diamond arrangement is expected to be more accurate than the one-directional arrangement.

5.3 FETD/FDTD hybrid methods

Because cube elements used in FETD and FDTD methods have no difference in terms of shape, FETD/FDTD hybrid methods become a choice if we want to solve
the cube region efficiently. Various FETD/FDTD hybrid methods could be divided into two types based on the FDTD method being used: those using traditional Yee’s scheme belong to the first type and those using non-staggered FDTD scheme [1] belong to the second type. Since both of them use the curl-curl FETD method [39], we distinct the first type as Yee-Curl-Curl (YCC) approach and the second type as Non-Staggered-Curl-Curl (NSCC) approach in this context. In this section, we shall take a look at both of them.

**Yee-Curl-Curl Approach**

Most implementations belong to the YCC type [14],[25],[26],[42],[31]. YCC approach uses traditional Yee’s scheme in the FDTD region. In the FETD region, different elements can be used, e.g., Ref. [26] and Ref. [31] use pentahedra, Ref. [25] uses tetrahedra. A common problem of the YCC approach is the coupling between the FETD and the FDTD results. The curl-curl FETD method only solves for either electric or magnetic fields (in most cases, electric fields are solved). Using electric field as an example, the implicit procedure solves for $E$ as a boundary value problem with the boundary values known at the present time step. In the FETD/FDTD hybrid method, these boundary values are provided by the results of the enclosing FDTD region at the current time step, where the $E$ values at the current time step are calculated by their neighboring $H$ values at $1/2$ time step earlier.

Due to the staggered nature of Yee’s scheme, it is easy to see that there must be an overlapping region between the FDTD and the FETD regions. A typical update procedure consists of three steps. The boundary $E$ values in the FETD region are calculated first by FDTD using the $H$ values in the overlapping region. Then the $E$ values inside the FETD region are calculated by using the boundary $E$ values as
its Dirichlet boundary condition. Finally, the $H$ values in the overlapping region are updated by FDTD using the $E$ values obtained in previous steps.

As a matter of fact, late-time instability is a common problem in this approach. The exact reason is still unclear at this moment. Practical treatments include using filters schemes [25] like those being used in FDTD subgridding methods.

**Non-Staggered-Curl-Curl Approach**

The NSCC approach uses the non-staggered FDTD method [1]. To relate the NSCC approach to FETD method, let us start from the vector wave equation in free space

$$\nabla \times \frac{1}{\mu_r} \nabla \times \vec{E} + \frac{\epsilon_r}{c^2} \frac{\partial^2 \vec{E}}{\partial t^2} = 0 \quad (5.28)$$

and express the electrical field with vector basis functions $\vec{N}_i$ as

$$\vec{E} = \sum_{i=1}^{N} e_i \vec{N}_i \quad (5.29)$$

By applying Galerkin's method, we obtain the following weak form according to each testing function $\vec{N}_j$ and basis function $\vec{N}_i$

$$\int \int \int_v \left( \frac{1}{\mu_r} \nabla \times \vec{N}_i \right) \cdot \left( \nabla \times \vec{N}_j \right) e_i + \frac{1}{\epsilon_r} (\epsilon_r \vec{N}_i) \cdot \vec{N}_j \frac{\partial^2 e_i}{\partial t^2} dv = 0 \quad (5.30)$$

Using matrix notations, this set of linear equations can be written as

$$[S]\vec{e} + [T] \frac{1}{c^2} \frac{\partial^2 \vec{e}}{\partial t^2} = 0 \quad (5.31)$$

where $[S]$ is the global stiff matrix, $[T]$ is the global mass matrix, and $\vec{e}$ represents all unknown electric field components. By the Generalized-Θ method [39], Eq.(5.31) is discretized in time domain as

$$\vec{e}^{n+1} \{ [T] + \Delta t^2 (2\theta_2 + \theta_1) [S] \} - \vec{e}^n \{ 2[T] + \Delta t^2 (4\theta_2 - 1) [S] \} + \vec{e}^{n-1} \{ [T] + \Delta t^2 (2\theta_2 - \theta_1) [S] \} = 0 \quad (5.32)$$
where superscripts denote time steps. If \([S]\) is semi-positive definite and \([T]\) is positive definite, then \(\theta_1 \geq 0\) and \(\theta_2 \geq 1/8\) guarantee unconditional stability \([39]\). To obtain minimum numerical dispersion errors, \(\theta_1 = 0\) and \(\theta_2 = 1/8\) are commonly used. Therefore, Eq. (5.32) becomes

\[
\vec{e}^{n+1}\{[T] + \frac{\Delta t^2}[S]\} - \vec{e}^n\{2[T] - \frac{\Delta t^2}[S]\} + \vec{e}^{n-1}\{[T] + \frac{\Delta t^2}[S]\} = 0 \tag{5.33}
\]

The Non-staggered FDTD method can be derived by taking \(\theta_1 = 0\) and \(\theta_2 = 0\) in Eq. (5.32) to obtain

\[
\vec{e}^{n+1}[T] - \vec{e}^n\{2[T] - \Delta t^2[S]\} + \vec{e}^{n-1}[T] = 0 \tag{5.34}
\]

and then followed by mass-lumping \([T]\) matrix.

We assume that tetrahedral and brick elements co-exist in the problem domain. In this circumstance, several schemes are possible by applying either Eq. (5.33) or Eq. (5.34), to either the tetrahedral or the cube region. Bearing in mind that mass-lumping and conditional stability are not practical for tetrahedral elements \([39]\), and applying Eq. (5.33) to both regions is nothing but the regular unconditionally stable FETD method, there is only one scheme left to be considered. It applies Eq. (5.34) to the brick region and Eq. (5.33) to the tetrahedral region \([53],[54]\). This seems to be a promising hybrid scheme at first glance and we shall examine it carefully.

Let us use subscript “2” to denote the matrices and unknowns associated with the region consisting of both tetrahedral and pyramidal elements, and “1” to denote

\[6\] The conclusion with regard to which FDTD method was used in Refs. \([53],[54]\) is inferred from the context of these papers. Since not being specified clearly, Yee’s scheme might also be possible. However, whether these papers use non-staggered FDTD does not affect our discussion.

108
the region consisting of brick elements only. Combining Eqs. (5.33) and (5.34) yields

\[
\begin{bmatrix}
\frac{T_{11}}{2} + \frac{c^2 \Delta t^2}{4} S_{21} & \frac{T_{12}}{2} + \frac{c^2 \Delta t^2}{4} S_{22} \\
\frac{T_{21}}{2} - \frac{c^2 \Delta t^2}{4} S_{21} & \frac{T_{22}}{2} - \frac{c^2 \Delta t^2}{4} S_{22}
\end{bmatrix}
\begin{bmatrix}
e_n^1 \\
e_n^2
\end{bmatrix}
= 2
\begin{bmatrix}
\frac{T_{11}}{2} & 0 \\
\frac{T_{12}}{2} & \frac{T_{22}}{2} - \frac{c^2 \Delta t^2}{4} S_{22}
\end{bmatrix}
\begin{bmatrix}
e_n^e \\
e_n^e
\end{bmatrix}
\]

(5.35)

By mass-lumping the mass matrix in the brick region, the above transforms to

\[
\begin{bmatrix}
\frac{T_{11}^D}{2} + \frac{c^2 \Delta t^2}{4} S_{21} & \frac{T_{12}^D}{2} + \frac{c^2 \Delta t^2}{4} S_{22} \\
\frac{T_{21}^D}{2} - \frac{c^2 \Delta t^2}{4} S_{21} & \frac{T_{22}^D}{2} - \frac{c^2 \Delta t^2}{4} S_{22}
\end{bmatrix}
\begin{bmatrix}
e_n^1 \\
e_n^2
\end{bmatrix}
= 2
\begin{bmatrix}
\frac{T_{11}^D}{2} & 0 \\
\frac{T_{12}^D}{2} & \frac{T_{22}^D}{2} - \frac{c^2 \Delta t^2}{4} S_{22}
\end{bmatrix}
\begin{bmatrix}
e_n^{e-1} \\
e_n^{e-1}
\end{bmatrix}
\]

(5.36)

where superscript “\(D\)” denotes the resultant diagonal matrix. To find the solution, we can either inverse the entire matrix or decouple the solving procedure into two steps. First, we solve the unknowns in the cube region (this is just the non-staggered FDTD). Then we solve the rest of unknowns implicitly by using those solved in the first step as the Dirichlet boundary condition. Because we actually choose a hybrid method to inverse the matrix, this approach is still a FETD approach.

The resultant system amplification matrix is written as

\[
\begin{bmatrix}
\frac{T_{11}^D}{2} + \frac{c^2 \Delta t^2}{4} S_{21} & \frac{T_{12}^D}{2} + \frac{c^2 \Delta t^2}{4} S_{22} \\
\frac{T_{21}^D}{2} - \frac{c^2 \Delta t^2}{4} S_{21} & \frac{T_{22}^D}{2} - \frac{c^2 \Delta t^2}{4} S_{22}
\end{bmatrix}^{-1}
\begin{bmatrix}
\frac{T_{11}^D}{2} - \frac{c^2 \Delta t^2}{4} S_{11} & \frac{T_{12}^D}{2} - \frac{c^2 \Delta t^2}{4} S_{12} \\
\frac{T_{21}^D}{2} - \frac{c^2 \Delta t^2}{4} S_{21} & \frac{T_{22}^D}{2} - \frac{c^2 \Delta t^2}{4} S_{22}
\end{bmatrix}^{-1}
\begin{bmatrix}
e_n^1 \\
e_n^2
\end{bmatrix}
\]

followed by the procedure described in Ref. [70]. Because it is not normal, we have to prove/disprove the diagonalizability of the system amplification matrix before reaching firm conclusions on the stability (based on the requirements of the stability of time-domain schemes [70]). However, this is much involved analytically due to the complicated matrix structure. Instead, we shall use numerical example to show the late-time stability.
We run a simulation inside the same waveguide as that in Ref. [53]. A first order differentiated Gaussian pulse is used. Fig. 5.10 illustrates the observed electrical field inside the waveguide, where late-time instability shows up. In fact, we have run several different size rectangular resonators and the results all blow up sooner or later. It should be mentioned that this late-time instability may not be detrimental in some applications. The practical issues are the same as associated FDTD subgridding schemes.

5.4 Implicit Non-Staggered FDTD Method

5.4.1 Formulation

The implicit Non-Staggered FDTD (INS-FDTD) method is aimed to directly reduce the bandwidth of global matrices. For brick elements with the lowest order
vector basis functions, the element matrices in finite-element methods are [29]

\[
[S]^e = \begin{bmatrix}
[S_{11}]^e & -\frac{\mu}{6}[K_3]^T & -\frac{\mu}{6}[K_3]^T \\
-\frac{\mu}{6}[K_3] & [S_{22}]^e & -\frac{\mu}{6}[K_3]^T \\
-\frac{\mu}{6}[K_3] & -\frac{\mu}{6}[K_3]^T & [S_{33}]^e
\end{bmatrix}
\]  

(5.37)

\[
[T]^e = \frac{l_x l_y l_z}{36} \begin{bmatrix}
[K_4] & 0 & 0 \\
0 & [K_4] & 0 \\
0 & 0 & [K_4]
\end{bmatrix}
\]  

(5.38)

with

\[
[S_{11}]^e = \frac{l_x l_y}{6l_y} [K_1] + \frac{l_x l_z}{6l_z} [K_2]
\]  

(5.39)

\[
[S_{22}]^e = \frac{l_x l_y}{6l_z} [K_1] + \frac{l_y l_z}{6l_z} [K_2]
\]  

(5.40)

\[
[S_{33}]^e = \frac{l_x l_z}{6l_x} [K_1] + \frac{l_x l_z}{6l_y} [K_2]
\]  

(5.41)

\[
[K_1] = \begin{bmatrix}
2 & -2 & 1 & -1 \\
-2 & 2 & -1 & 1 \\
1 & -1 & 2 & -2 \\
1 & 1 & -2 & 2
\end{bmatrix}
\]  

(5.42)

\[
[K_2] = \begin{bmatrix}
2 & 1 & -2 & -1 \\
1 & 2 & -1 & -2 \\
-2 & -1 & 2 & 1 \\
-1 & -2 & 1 & 2
\end{bmatrix}
\]  

(5.43)

\[
[K_3] = \begin{bmatrix}
2 & 1 & -2 & -1 \\
-2 & -1 & 2 & 1 \\
1 & 2 & -1 & -2 \\
-1 & -2 & 1 & 2
\end{bmatrix}
\]  

(5.44)

\[
[K_4] = \begin{bmatrix}
4 & 2 & 2 & 1 \\
2 & 4 & 1 & 2 \\
2 & 1 & 4 & 2 \\
1 & 2 & 2 & 4
\end{bmatrix}
\]  

(5.45)

where \( l_x, l_y, \) and \( l_z \) are edge lengths of each element along \( \hat{x}, \hat{y}, \) and \( \hat{z} \) directions respectively.

INS-FDTD method uses the following building blocks for stiffness and mass matrices instead.
\[ K_1 = \begin{bmatrix} 3 & -3 & 0 & 0 \\ -3 & 3 & 0 & 0 \\ 0 & 0 & 3 & -3 \\ 0 & 0 & -3 & 3 \end{bmatrix} \] (5.46)

\[ K_2 = \begin{bmatrix} 3 & 0 & -3 & 0 \\ 0 & 3 & 0 & -3 \\ -3 & 0 & 3 & 0 \\ 0 & -3 & 0 & 3 \end{bmatrix} \] (5.47)

\[ K_3 = \begin{bmatrix} 3 & 0 & -3 & 0 \\ -3 & 0 & 3 & 0 \\ 0 & 3 & 0 & -3 \\ 0 & -3 & 0 & 3 \end{bmatrix} \] (5.48)

\[ K_4 = \begin{bmatrix} 9 & 0 & 0 & 0 \\ 0 & 9 & 0 & 0 \\ 0 & 0 & 9 & 0 \\ 0 & 0 & 0 & 9 \end{bmatrix} \] (5.49)

It is ready to verify that the corresponding stiffness matrix is semi-positive definite and the mass matrix is positive definite. Therefore, we can use \( \theta_1 \geq 0 \) and \( \theta_2 \geq 1/8 \) in Eq. (5.32) to make the scheme fully implicit and unconditionally stable. Eqs. (5.46)-(5.49) have the advantage that many non-zero entries generated by Eqs. (5.42)-(5.45) are zeros now. When the elemental matrices are assembled, the bandwidth of the global matrix drops by more than 60% from 33 to 13. It should be pointed out that this scheme is essentially a non-staggered FDTD variant. Whether it is also a FETD approach depends on the existence of qualified test and basis functions. We shall leave this topic to future research.

### 5.4.2 Numerical Dispersion

Without losing generality, dispersion errors of cube element are discussed for simplicity. Numerical dispersions of finite-element methods have been studied by different authors in both 2-D and 3-D cases [37],[74]. It is evaluated by assuming a
monochromatic wave propagating in an infinite region with the numerical solution of the electrical field given by

$$\vec{E} = (E_x \hat{x} + E_y \hat{y} + E_z \hat{z})e^{j(\omega n \Delta t - \beta \hat{\beta} \cdot \vec{r})} \quad (5.50)$$

where $\hat{\beta} = \sin \theta \cos \phi \hat{x} + \sin \theta \sin \phi \hat{y} + \cos \theta \hat{z}$ represents the wave propagating direction, $\vec{r} = h(m_x \hat{x} + m_y \hat{y} + m_z \hat{z})$ is the location of the electric field component and $\beta = 2\pi/\lambda$ is the numerical wave number. By defining $q = h/\lambda$ as the reciprocal of number of cells per wavelength and $p = v_p/\bar{v}_p$ as the ratio of the exact phase velocity to the numerical phase velocity, the phase term in Eq.(5.50) is reformed as

$$\omega n \Delta t - 2\pi pq(m_x \sin \theta \cos \phi + m_y \sin \theta \sin \phi + m_z \cos \theta) \quad (5.51)$$

To be able to compare with the dispersion property of Yee’s scheme, we write $\Delta t = \chi h/\sqrt{3}v_p$ with $\chi$ being the CFL number. Also $\theta_1 = 0$ and $\theta_2 = 1/8$ are used as an example.

With the above specifications, we obtain the following linear system for $E_x$, $E_y$ and $E_z$ by substituting Eqs.(5.50)-(5.51) into Eq.(5.31)

$$\begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \cdot \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} = 0 \quad (5.52)$$

with

$$A_{11} = \frac{4}{3}(1 - \gamma + \gamma^2) - \frac{(1 + \gamma)^2}{6}[\cos(2\xi_y) + \cos(2\xi_z)] \quad (5.53)$$

$$A_{22} = \frac{4}{3}(1 - \gamma + \gamma^2) - \frac{(1 + \gamma)^2}{6}[\cos(2\xi_x) + \cos(2\xi_z)] \quad (5.54)$$

$$A_{33} = \frac{4}{3}(1 - \gamma + \gamma^2) - \frac{(1 + \gamma)^2}{6}[\cos(2\xi_x) + \cos(2\xi_y)] \quad (5.55)$$

$$A_{12} = A_{21} = -\frac{(1 + \gamma)^2}{3} \sin \xi_x \sin \xi_y \quad (5.56)$$

$$A_{13} = A_{31} = -\frac{(1 + \gamma)^2}{3} \sin \xi_x \sin \xi_z \quad (5.57)$$
\[ A_{23} = A_{32} = -\frac{(1 + \gamma)^2}{3} \sin \xi_y \sin \xi_z \] (5.58)

where

\[ \gamma = e^{\frac{i\pi z}{2}} \]

\[ \xi_x = \frac{\nu}{q} \sin \theta \cos \phi \]

\[ \xi_y = \frac{\nu}{q} \sin \theta \sin \phi \]

\[ \xi_z = \frac{\nu}{q} \cos \theta \]

For \( E_x, E_y \) and \( E_z \) to have non-trivial solutions, there must be

\[ \text{Det} \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} = 0 \] (5.59)

This results in a transcendental equation which is solved numerically.

Fig. 5.11 shows the largest (all-angle) dispersion errors of Yee’s scheme, INS-FDTD method, and the FETD method using brick elements. The dispersion error is defined by \( 360(\tilde{\beta}/\beta - 1) \). With the same CFL number \( \chi = 1.0 \), INS-FDTD method exhibits relatively large error, which is about 10.5°/λ at ten cells per wavelength. The FETD method has the smallest error, which is about 1.7°/λ at ten cells per wavelength. The numerical dispersion of Yee’s scheme is about 4°/λ.

Fig. 5.12 shows the anisotropy of INS-FDTD evaluated at \( \chi = 1.0 \). Fig. 5.13 further illustrates the comparison of dispersion errors of INS-FDTD and Yee’s scheme evaluated at \( \chi = 1.0 \) and \( \theta = 45^\circ \). As we can see, INS-FDTD has about the same anisotropy as Yee’s scheme. This suggests that filter schemes, introduced in Ref. [45] for Yee’s scheme, can also be used to effectively reduce the dispersion error to about 2°/λ at ten cells per wavelength.

Without the aforementioned optimization procedure, the cell size used in INS-FDTD method should be kept smaller than Yee’s scheme. The appropriate cell size for INS-FDTD method depends on the electrical size of the entire problem domain, because the dispersion error is spatially cumulative. If we adopt that ten cells per
Figure 5.11: Numerical dispersion errors of different schemes evaluated at $\chi = 1.0$

Figure 5.12: Anisotropy of numerical dispersion errors of INS-FDTD scheme evaluated at $\chi = 1.0$
wavelength should be kept in Yee’s scheme, 16 cells per wavelength is suggested for INS-FDTD method, as illustrated in Fig.5.11.

As we change the CFL numbers, the dispersion error of INS-FDTD method changes proportionally. Fig. 5.14 shows that the dispersion errors when increasing the CFL numbers. This is because large CFL numbers correspond to large temporal discretization error. The same thing happens to FETD methods and the results are shown in Fig. 5.15.

5.4.3 A Different Scheme

This scheme is intended to reduce the dispersion error of INS-FDTD method further. Inside the general framework of INS-FDTD method, this scheme differs from
Figure 5.14: Numerical dispersion errors of INS-FDTD scheme evaluated at different \( \chi \) values

Figure 5.15: Numerical dispersion errors of FETD with cube elements evaluated at different \( \chi \) values
the INS-FDTD method only by using a different mass matrix building block

$$[K_4] = \frac{9}{\alpha + 2\beta} \begin{bmatrix} \alpha & \beta & \beta & 0 \\ \beta & \alpha & 0 & \beta \\ \beta & 0 & \alpha & \beta \\ 0 & \beta & \beta & \alpha \end{bmatrix}$$

(5.60)

where $\alpha$ and $\beta$ are unknown coefficients to be determined. The above equation reflects the following considerations. First of all, the mass matrix should not generate non-zero global matrix entries in addition to those introduced by the stiff matrix. Since the bandwidth of the stiff matrix is already 13, this guarantees that this different scheme is the same efficient as INS-FDTD method. Second, the sum of all entries is 9. Thus INS-FDTD method could be naturally recovered by mass lumping this scheme. In general, this scheme could be thought of taking weighted average of the neighboring field values when constructing the mass matrix.

What really counts in Eq. (5.60) is not the absolute values of $\alpha$ and $\beta$, but their ratio. Note that when $\beta/\alpha = 0$, this scheme recovers the INS-FDTD scheme. The eigenvalues of $[K_4]$ in Eq. (5.60) are

$$\lambda_{1,2} = 9, \quad \lambda_3 = 9 \frac{\alpha}{\alpha + 2\beta}, \quad \lambda_4 = 9 \frac{\alpha - 2\beta}{\alpha + 2\beta}$$

To guarantee unconditional stability, they must be larger than zero. Thus, $\beta/\alpha < 0.5$ must be satisfied.

Fig. 5.16 shows the dispersion error in the first quadrant when $\beta/\alpha = 0.49$, which is very broken. The problem is that there are always two phase velocities at each propagation angle for any $\beta/\alpha$ value, except for $\beta/\alpha = 0$, which is the INS-FDTD scheme. Fig. 5.16 actually only catches one of the two phase velocities at each angle. At this moment, we believe that this is because this scheme is neither a rigorous FDTD method nor a valid FETD approach. So there is no guarantee on one phase velocity at each propagation angle.
Figure 5.16: Numerical dispersion of the different scheme.

Figure 5.17: Observed electrical field in an empty waveguide by using FETD method and the different scheme with different $\beta/\alpha$ values.
Fig. 5.17 further shows the observed electrical field in an empty waveguide by using FETD method and this scheme with different $\beta/\alpha$ values. Obviously, there are two different phase velocities. When $\beta/\alpha = 0.4$, the result is very close to that of the FETD.

5.5 FDTD as a Predictor for FETD

The FETD/INS-FDTD hybrid method is useful when the dispersion error caused by INS-FDTD method is tolerable or correctable. But there are cases where INS-FDTD can not be used. In the following, we shall further discuss the efficiency issues of pure FETD with iterative solvers.

The results of FDTD and the lowest-order FETD methods can be very close. This is illustrated in Fig. 5.18, where we show the observed electrical fields in an empty waveguide by using FETD and FDTD methods respectively. Therefore, it is natural to consider using FDTD to obtain accurate initial guesses for CG solvers, i.e. as a predictor.

The implementation of FDTD predictor is similar to FETD/FDTD hybrid method. In this section, we shall focus on evaluating the efficiency. To do so, we use the following definition of residual error

$$
\epsilon = \frac{||A \cdot x_{new} - b||_2}{||A||_2||x_{new}||_2 + ||b||_2} \quad (5.61)
$$

in CG solver for solving problem

$$
[A] \cdot x = b
$$

where $|| \cdot ||_2$ represents 2-norm. Whenever this value is below a threshold, say $10^{-7}$ in our tests, we terminate the CG solver. By theory [21], the resultant forward error
bound is
\[ ||x' - x||_2 \leq \epsilon ||A^{-1}||_2 \cdot (||A||_2||x'||_2 + ||b||_2) \] (5.62)
where \( x \) denotes the exact answer and \( x' \) the CG solution. There are two things we should comment on the above criterion. First, it is accurate but also expensive. We only use it for evaluation purpose. Second, the forward error bound actually depends on the condition number of \( [A] \), as can be seen from Eq. (5.62). Thus for systems with different condition numbers, the same stopping criterion does not imply the same accuracy.

5.5.1 Source Prediction

We first consider an empty waveguide constructed by cube elements only. A \( TE_{10} \) current source with second order differentiated Gaussian pulse is excited at one end.

Figure 5.18: Observed electrical fields in an empty waveguide.
Figure 5.19: Initial residuals of different initial guesses.

Figure 5.20: Number of iterations required by different initial guesses.
During simulation, the unknowns are computed by non-staggered FDTD method at the beginning of each time step. The results are then fed into the CG solver as initial guesses.

The major problem is source excitation. As FETD and FDTD methods interpret source excitation quite differently, we did not actually excite the FDTD at each time step. The FDTD solver simply generates the results at the current time step from the previous two steps without source. The resulting initial $\epsilon$ and the number of iterations for $\epsilon$ to converge under $10^{-7}$ are shown in Fig. 5.19 and Fig. 5.20 respectively. The line denoted by “original” represents the results when using previous time step values as initial guesses, and the line denoted by “FDTD predic. no src” represents the results when using FDTD predictor as aforementioned.

Obviously, the initial errors are much smaller even without source predictions. Correspondingly, it requires an average of three iterations to converge when the source is turned off. However, the initial errors are still large when the source is turned on. To combat this, let us consider the curl-curl equation when the current source is present. Eq. (5.28) then becomes

$$\nabla \times \frac{1}{\mu_r} \nabla \times \vec{E} + \frac{\epsilon_r}{c^2} \frac{\partial^2 \vec{E}}{\partial t^2} + \mu_0 \frac{\partial \vec{J}}{\partial t} = 0$$

(5.63)

where $\vec{J}$ represents the current source. Accordingly, Eq. (5.31) becomes

$$[S][\vec{E}] + [T] \frac{1}{c^2} \frac{\partial^2 \vec{E}}{\partial t^2} + \mu_0[T] \frac{\partial \vec{j}}{\partial t} = 0$$

(5.64)

Note that $\vec{j}$ is only non-zero where the current exists. The third term is the only difference between Eq. (5.64) and Eq. (5.31). Therefore, we can use FDTD to predict the results without source, which correspond to Eq. (5.31), and then add the
prediction to the third term in Eq. (5.64). The sum will be the prediction when the current source is turned on.

Because the current source does not change its location in each time step, we can further write the third term in Eq. (5.64) as

$$\mu_0 [T] \cdot \tilde{j} \cdot \frac{\partial p(t)}{\partial t}$$

where $\tilde{j}$ is the space-dependent part and $p(t)$ is the time-dependent part, or the impulse shape. To predict the source contributions, we first compute and store the results of $\mu_0 [T] \cdot \tilde{j}$. Then at each time step, we simply compute $\frac{\partial p(t)}{\partial t}$ and multiply it with $\mu_0 [T] \cdot \tilde{j}$ to obtain the prediction. Note that $\frac{\partial p(t)}{\partial t}$ is computed analytically at each time step and this causes errors in the prediction. The results of this proposed prediction are shown in Fig. 5.19 and Fig. 5.20. Clearly, both the initial residual and the number of iterations are greatly reduced.

### 5.5.2 Tetrahedral Regions

When tetrahedral elements are used, things are quite different. We use linear prediction in the tetrahedral region instead. That is, the current initial guesses are the linear extrapolations of the results of the previous two time steps. Fig. 5.21 and Fig. 5.22 illustrate the initial residual and the number of iterations required for $\epsilon$ to converge to $10^{-7}$. The example is again an empty waveguide. Half of the waveguide is constructed by cubes and the other half by tetrahedra. The source is a $TE_{10}$ current source with second order differentiated Gaussian pulse excited at the cube end. The results denoted by “Linear predict” are obtained by applying linear prediction in both cube and tetrahedral regions. The results denoted by “FDTD predict” are obtained by applying linear prediction in the tetrahedral region and FDTD prediction in the
Figure 5.21: Initial residuals when cubes and tetrahedra co-exist.

Figure 5.22: Number of iterations to converge with a normalized residual $10^{-7}$, 2nd order Diff. Gaussian Src.
Figure 5.23: Initial residual when the source is in the tetrahedral region.

cube region. As can be seen, the initial residual of the FDTD prediction is about 20 dB less than that of the linear prediction before wave entering the tetrahedral region. After that, the initial residual is dominated by the part corresponding to the tetrahedral region. The FDTD prediction in the cube region only makes it slightly better than linear prediction. Accordingly, the required numbers of iterations are almost the same in both cases, as shown in Fig. 5.22.

Because we have assumed that the tetrahedral regions are small (locally fine structures) and there are enough memory to store the associated matrices, more effective preconditioners can be used, such as incomplete LU preconditioners. To assure that such preconditioners indeed reduce the number of iterations, we study an extreme case in the following.
In this extreme case, we actually solve the unknowns in the tetrahedral regions by LU solver. If there are $M$ unknowns to be solved, $M^2$ storage units are expected for complete LU factorization and the memory requirement for incomplete LU preconditioners fall between $M^2$ and $M$. Also, when $M^2$ storage units are required, it does not matter whether we use them to represent the matrix inverse or performing LU decomposition. In this sense, we call our approach an extreme case of incomplete LU preconditioners.

The tests are carried out in an empty waveguide. The source is first in the tetrahedral region. Because LU solver actually solves the problem when wave is confined in the tetrahedral region, there is no initial residual and no iteration needed. This is shown in Fig. 5.23 and Fig. 5.24. After wave entering the cube region, the error
caused by the FDTD predictor becomes dominant. Because portion of the total energy is still in the tetrahedral region, the initial residual errors are smaller than those in Fig. 5.19 and less iterations are required to converge. Next, we put the source in the cube region and the results are shown in Fig. 5.25 and Fig. 5.26. It is clear that before wave entering the tetrahedral region, the initial error and the required number of iterations are similar to those in Fig. 5.19 and Fig. 5.20. After wave entering the tetrahedral region, the initial residual and required number of iterations drop to a level similar to those in Fig. 5.23 and Fig. 5.24. In both cases, the initial residual errors and the required number of iterations are dominated by FDTD predictor. This indicates that incomplete LU preconditioners can be effectively used as long as global residual errors are dominated by FDTD predictor.

Figure 5.25: Initial residual when the source is in the cube region.
Figure 5.26: Number of iterations when the source is in the cube region.

5.6 Concluding Remarks

This chapter presented a new way to facilitate generating unstructured/structured hybrid mesh and discussed three different methods in conjunction with the hybrid mesh. Implicit Non-Staggered FDTD method is proposed to be hybridized with FETD methods. When numerical dispersion is not critical – either due to small enough cell size or the application of optimization scheme, this approach guarantees stability and is more efficient than pure FETD methods. Using FDTD as a predictor for FETD was also discussed for efficiency improvement.
CHAPTER 6

TIME-DOMAIN PERFECTLY MATCHED LAYER IMPLEMENTATIONS

6.1 Introduction

In various finite-difference and finite-element methods, the computational domain must be truncated by Absorbing Boundary Conditions (ABCs) when modelling infinite structures or open problems. There are two main categories of ABCs in time-domain methods, i.e. absorbing boundary operators [4] and Perfectly Matched Layer (PML) [2]. Absorbing boundary operators are easy to implement but have large reflection errors. PML is first proposed by Berenger in a split-field form in 1994 for FDTD [2]. Different variants are proposed subsequently, which primarily include stretched coordinate PML [7], anisotropic PML [55], and Complex Frequency Shift (CFS) PML [36]. PMLs generate much smaller reflection errors than absorbing boundary operators [2],[17]. However, its implementations are not always satisfactory. In Ref. [40], large reflection errors are observed in the ADI-FDTD implementation. In Ref. [71], large reflection errors are also observed in the FETD implementation. The implementation in Ref.[28] has observed late-time instability. In this chapter, we shall focus on solving these problems.
6.2 PML Implementation for ADI-FDTD

Alternating-Direction Implicit Finite Difference Time Domain (ADI-FDTD) method is an attractive alternative to the standard FDTD due to its unconditional stability with moderate computational overhead [16],[43],[62],[76]. The ADI-FDTD can be particularly useful for problems involving devices with fine geometric features that are much smaller than the wavelengths of interest. In order to be applied for unbounded problems, the ADI-FDTD requires the use of appropriate absorbing boundary conditions. The PML has been first extended to the ADI-FDTD method in [40], where the split-field PML performance deteriorates for large Courant numbers.

Because the ADI-FDTD method is most necessary for large (greater than unity) Courant numbers (as long as the accuracy of the results is not impacted by the potential increase on the time step truncation error [16]), a PML implementation that does not deteriorate for large Courant numbers is highly desirable. In this section, we introduce a novel PML implementation for the ADI-FDTD for this purpose. For simplicity, we consider a 2-D TE problem, but the basic approach equally applies to the 3-D problem.

6.2.1 ADI-FDTD Update Equations with PML

In the split-field PML formulation [2], the $E_x$, $E_y$, $H_{zx}$, and $H_{zy}$ field components satisfy the following relations

$$\epsilon_0 \frac{\partial E_x}{\partial t} + \sigma_y E_x = \frac{\partial (H_{zx} + H_{zy})}{\partial y}$$  \hspace{1cm} (6.1)

$$\epsilon_0 \frac{\partial E_y}{\partial t} + \sigma_x E_y = -\frac{\partial (H_{zx} + H_{zy})}{\partial x}$$  \hspace{1cm} (6.2)

$$\mu_0 \frac{\partial H_{zx}}{\partial t} + \frac{\sigma_x}{\epsilon_0} H_{zx} = -\frac{\partial E_y}{\partial x}$$  \hspace{1cm} (6.3)
\[
\frac{\partial H_{xy}}{\partial t} + \frac{\sigma_y}{\epsilon_0} H_{zy} = \frac{\partial E_x}{\partial y}
\]  \hfill (6.4)

In the ADI-FDTD method, the update at each time step is divided into two sub-steps. First, assuming that \( E_y \) is computed implicitly at the time step \((n + \frac{1}{2})\), Eqs.(6.1)-(6.4) are discretized by \([43],[40]\).

Substituting Eqs.(6.5), (6.7), and (6.8) into Eq. (6.6), an implicit update for \( E_y \) is obtained:

\[
\begin{align*}
\beta_x(i + \frac{1}{2}, j + \frac{1}{2}) E^{n+\frac{1}{2}}_y(i + 1, j + 1) &- c_1^n E^{n+\frac{1}{2}}_y(i, j + \frac{1}{2}) + \beta_y(i - \frac{1}{2}, j + \frac{1}{2}) \\
E^{n+\frac{1}{2}}_x(i - 1, j + \frac{1}{2}) &- \alpha_x(i - \frac{1}{2}, j + \frac{1}{2}) E^{n+\frac{1}{2}}_x(i, j + \frac{1}{2}) + \beta_x(i + \frac{1}{2}, j + \frac{1}{2}) \\
- \alpha_y(i - \frac{1}{2}, j + \frac{1}{2}) &+ \beta_y(i - \frac{1}{2}, j + \frac{1}{2}) E^{n+\frac{1}{2}}_x(i + 1, j + 1) \\
- \beta_x(i + \frac{1}{2}, j + \frac{1}{2}) &- \beta_y(i + \frac{1}{2}, j + \frac{1}{2}) E^n_x(i + \frac{1}{2}, j + 1) - E^n_x(i + \frac{1}{2}, j) 
\end{align*}
\]  \hfill (6.9)

where the superscripts denote time steps. The eight coefficients \( \alpha_x, \alpha_y, \alpha_x, \alpha_y, \beta_x, \beta_y, \beta_x, \beta_y \) depend on the local PML conductivities \( \sigma_x \) or \( \sigma_y \). Eqs. (6.5), (6.7), and (6.8) into Eq. (6.6), an implicit update for \( E_y \) is obtained:

\[
\begin{align*}
\beta_x(i + \frac{1}{2}, j + \frac{1}{2}) E^{n+\frac{1}{2}}_y(i + 1, j + 1) &- c_1^n E^{n+\frac{1}{2}}_y(i, j + \frac{1}{2}) + \beta_y(i - \frac{1}{2}, j + \frac{1}{2}) \\
E^{n+\frac{1}{2}}_x(i - 1, j + \frac{1}{2}) &- \alpha_x(i - \frac{1}{2}, j + \frac{1}{2}) E^{n+\frac{1}{2}}_x(i, j + \frac{1}{2}) + \beta_x(i + \frac{1}{2}, j + \frac{1}{2}) \\
- \alpha_y(i - \frac{1}{2}, j + \frac{1}{2}) &+ \beta_y(i - \frac{1}{2}, j + \frac{1}{2}) E^{n+\frac{1}{2}}_x(i + 1, j + 1) \\
- \beta_x(i + \frac{1}{2}, j + \frac{1}{2}) &- \beta_y(i + \frac{1}{2}, j + \frac{1}{2}) E^n_x(i + \frac{1}{2}, j + 1) - E^n_x(i + \frac{1}{2}, j) 
\end{align*}
\]  \hfill (6.9)

with

\[
\begin{align*}
c_1^n &= \frac{1}{\beta_y(i, j + \frac{1}{2})} + \beta_x(i + \frac{1}{2}, j + \frac{1}{2}) + \beta_y(i - \frac{1}{2}, j + \frac{1}{2}) \\
c_2^n &= -\frac{\beta_y(i, j + \frac{1}{2})}{\beta_x(i, j + \frac{1}{2})}
\end{align*}
\]

where it is clear that the left hand side (l.h.s.) of Eq. (6.9) forms a tri-diagonal matrix, and the associated linear system can be solved with \( \mathcal{O}(N) \) complexity [62].
In the second sub-step, the fields at the \((n+1)\) time step are obtained. The discretized equations are [43],[40]

\[
E^{n+1}_x(i + \frac{1}{2}, j) = \alpha^x(i + \frac{1}{2}, j)E^{n+\frac{1}{2}}_x(i + \frac{1}{2}, j) + \beta^x(i + \frac{1}{2}, j)[H^{n+1}_{xx}(i + \frac{1}{2}, j) + \frac{1}{2}]
\]

\[
-H^{n+1}_{xx}(i + \frac{1}{2}, j - \frac{1}{2}) + H^{n+\frac{1}{2}}_{zy}(i + \frac{1}{2}, j + \frac{1}{2}) - H^{n+1}_{zy}(i + \frac{1}{2}, j - \frac{1}{2})
\]

(6.10)

\[
E^{n+1}_y(i, j + \frac{1}{2}) = \alpha^y(i, j + \frac{1}{2})E^{n+\frac{1}{2}}_y(i, j + \frac{1}{2}) - \beta^y(i, j + \frac{1}{2})[H^{n+\frac{1}{2}}_{xx}(i + \frac{1}{2}, j) + \frac{1}{2}]
\]

\[-H^{n+\frac{1}{2}}_{xx}(i - \frac{1}{2}, j + \frac{1}{2}) + H^{n+\frac{1}{2}}_{zy}(i + \frac{1}{2}, j + \frac{1}{2}) - H^{n+\frac{1}{2}}_{zy}(i - \frac{1}{2}, j + \frac{1}{2})
\]

(6.11)

\[
H^{n+1}_{xx}(i + \frac{1}{2}, j + \frac{1}{2}) = \alpha^h_x(i + \frac{1}{2}, j + \frac{1}{2})H^{n+\frac{1}{2}}_{xx}(i + \frac{1}{2}, j + \frac{1}{2}) - \beta^h_x(i + \frac{1}{2}, j + \frac{1}{2})
\]

\[-H^{n+\frac{1}{2}}_{xx}(i + 1, j + \frac{1}{2}) - E^{n+\frac{1}{2}}_y(i, j + \frac{1}{2})
\]

(6.12)

\[
H^{n+1}_{zy}(i + \frac{1}{2}, j + \frac{1}{2}) = \alpha^h_y(i + \frac{1}{2}, j + \frac{1}{2})H^{n+\frac{1}{2}}_{zy}(i + \frac{1}{2}, j + \frac{1}{2}) + \beta^h_y(i + \frac{1}{2}, j + \frac{1}{2})
\]

\[-E^{n+\frac{1}{2}}_x(i + 1, j + \frac{1}{2}) - E^{n+\frac{1}{2}}_x(i + \frac{1}{2}, j + 1)
\]

(6.13)

from which the implicit update for \(E_x\) can be derived by

\[
\beta^h_y(i + \frac{1}{2}, j + \frac{1}{2})E^{n+1}_x(i + \frac{1}{2}, j + 1) - c^2 E^{n+1}_x(i + \frac{1}{2}, j) + \beta^h_y(i + \frac{1}{2}, j - \frac{1}{2})
\]

\[
E^{n+1}_x(i + \frac{1}{2}, j - 1) = c^2 E^{n+\frac{1}{2}}_x(i + \frac{1}{2}, j) - [\alpha^h_x(i + \frac{1}{2}, j + \frac{1}{2})H^{n+\frac{1}{2}}_{xx}(i + \frac{1}{2}, j + \frac{1}{2})
\]

\[-\alpha^h_x(i + \frac{1}{2}, j - \frac{1}{2})H^{n+\frac{1}{2}}_{xx}(i + \frac{1}{2}, j - \frac{1}{2}) + \alpha^h_x(i + \frac{1}{2}, j + \frac{1}{2})H^{n+\frac{1}{2}}_{xx}(i + \frac{1}{2}, j + \frac{1}{2})
\]

\[-A^h_y(i + \frac{1}{2}, j - \frac{1}{2})H^{n+\frac{1}{2}}_{zy}(i + \frac{1}{2}, j - \frac{1}{2}) - \{\beta^h_y(i + \frac{1}{2}, j - \frac{1}{2})[E^{n+\frac{1}{2}}_y(i + 1, j - \frac{1}{2})
\]

\[-E^{n+\frac{1}{2}}_y(i, j - \frac{1}{2}) - \beta^h_y(i + \frac{1}{2}, j + \frac{1}{2})[E^{n+\frac{1}{2}}_x(i + 1, j + \frac{1}{2}) - E^{n+\frac{1}{2}}_x(i + \frac{1}{2}, j)]
\]

(6.14)

with

\[
c^2_1 = \frac{1}{\beta^h_x(i + \frac{1}{2}, j)} + \beta^h_y(i + \frac{1}{2}, j + \frac{1}{2}) + \beta^h_y(i + \frac{1}{2}, j - \frac{1}{2})
\]

\[
c^2_2 = -\frac{\alpha^h_x(i + \frac{1}{2}, j)}{\beta^h_x(i + \frac{1}{2}, j)}
\]

\[
6.2.2 \quad \text{Traditional Scheme}
\]

In the traditional scheme[40], the coefficients \(\alpha^e_x, \alpha^e_y, \alpha^h_x, \alpha^h_y, \beta^e_x, \beta^e_y, \beta^h_x\) and \(\beta^h_y\)

are written as

\[
\alpha^e_x = \frac{(\alpha_0 - \sigma_x)}{(\alpha_0 + \sigma_x)} \quad \beta^e_x = \left[(\alpha_0 \sigma_x + \gamma_x) \Delta h\right]^{-1}
\]

\[
\alpha^e_y = \frac{(\alpha_0 - \sigma_y)}{(\alpha_0 + \sigma_y)} \quad \beta^e_y = \left[(\alpha_0 \sigma_y + \gamma_y) \Delta h\right]^{-1}
\]

\[
\alpha^h_x = \frac{(\alpha_0 - \sigma_x)}{(\alpha_0 + \sigma_x)} \quad \beta^h_x = \left[(\alpha_0 \sigma_x \sigma_y) \Delta h\right]^{-1}
\]

\[
\alpha^h_y = \frac{(\alpha_0 - \sigma_y)}{(\alpha_0 + \sigma_y)} \quad \beta^h_y = \left[(\alpha_0 \sigma_y \sigma_x) \Delta h\right]^{-1}
\]

(6.15)
where \( \eta_0 = \sqrt{\mu_0/\epsilon_0} \), \( \Delta t \) is the time step and \( \Delta h \) is the cell size. Note that the same set of coefficients is used for both Eqs. (6.5)-(6.8) and Eqs. (6.10)-(6.13).

To derive these coefficients, we can assume a central differencing approximation in time for the l.h.s. involving the conduction terms in Eqs. (6.1)-(6.4). For example, in the first sub-step from Eq. (6.1):

\[
\epsilon_0 \frac{\partial E_x}{\partial t} + \sigma_y E_x \Rightarrow \epsilon_0 \frac{E_x^{n+\frac{1}{2}}(i+\frac{1}{2},j) - E_x^n(i+\frac{1}{2},j)}{2\Delta t} + \sigma_y \frac{E_x^{n+\frac{1}{2}}(i+\frac{1}{2},j) - E_x^n(i+\frac{1}{2},j)}{2}
\]

\[\Rightarrow (\frac{2\epsilon_0}{\Delta t} + \frac{\sigma_y}{2}) E_x^{n+\frac{1}{2}}(i+\frac{1}{2},j) + (\frac{2\epsilon_0}{\Delta t} - \frac{\sigma_y}{2}) E_x^n(i+\frac{1}{2},j)\]

Now it is clear to see from Eq. (6.5) that

\[
\alpha^c_x = \frac{(\frac{2\epsilon_0}{\Delta t} - \frac{\sigma_y}{2})}{(\frac{2\epsilon_0}{\Delta t} + \frac{\sigma_y}{2})} \quad \beta^c_x = \frac{[(\frac{2\epsilon_0}{\Delta t} + \frac{\sigma_y}{2})\Delta h]^{-1}}{\delta^4}
\]

For the second sub-step, the same procedure is taken and we obtain the same coefficients. Note that this central differencing scheme is essential to the bad performance of the traditional approach and we shall deliberate this later.

We simulate the radiation of a point source in free space using this algorithm. The whole computation domain, including the PML region, contains \( 42 \times 42 \) cells with uniform cell size equal to 1.33 mm. Ten layers of PML are used in both \( x \) and \( y \) directions with a polynomial profile given by

\[
\sigma = \frac{\sigma_{max}|l - l_0|^4}{\delta^4}
\]

where \( \delta \) is the thickness of the PML absorber and \( l_0 \) represents the interface to free space. The excitation is located at the center of the computation region. The observation point is located ten cells away from the source and one cell away from the PML interface. The source is differentiated Gaussian pulse centered at 3.175 GHz and with a half bandwidth of 3.175 GHz also, applied to the \( H_z \) component. Fig. 6.1
Figure 6.1: Reflection errors from the PML-ADI-FDTD method with the usual PML implementation (I).

shows the relative reflection error from the PML, defined as

$$20 \log_{10} \left( \frac{|H_t - H_{ref}^t|}{\max |H_{ref}^t|} \right)$$

where the Courant number $\chi$ is defined as usual by $\Delta t = \chi \Delta h / (\sqrt{2} v_p)$ with $v_p$ representing the phase velocity. $H_{ref}^t$ in the equation above is the reference value calculated in a grid large enough so that any reflection from the boundary is causally isolated.

From Fig. 6.1 and Fig. 6.2, we observe that for small Courant numbers, the PML works well and the largest reflection error is about -75 dB. However, for large Courant numbers, the performance gets progressively worse. For $\chi = 5$, the largest reflection error is about -40 dB. This agrees very well with those reported in Ref. [40].
6.2.3 Exponential Differencing Scheme

It is natural to apply the exponential differencing scheme introduced by Berenger [2] to improve the PML performance. Experiences show that the exponential differencing scheme performs better at the presence of conductive material in some cases. Before applying it to the PML, let us examine its nature first.

Let us consider the following equation

$$\frac{\partial E_y}{\partial t} = -\frac{1}{\varepsilon_0 + \sigma_x E_y} \frac{\partial (H_{xx} + H_{xy})}{\partial x}$$  \hspace{1cm} (6.16)

as an example and reform it to

$$\varepsilon_0 \frac{\partial E_y}{\partial t} + \sigma_x E_y = \Upsilon^z_x$$  \hspace{1cm} (6.17)
where we use the abbreviation $\Upsilon^z_x$ to represent $-\frac{\partial (H_{xx} + H_{xy})}{\partial x}$. Performing the Laplace transformation to Eq. (6.17) followed by inverse Laplace transformation, we get

$$(s + \frac{\sigma_x}{\varepsilon_0}) E_y = \frac{\Upsilon^z_x}{\varepsilon_0}$$

$$\Rightarrow E_y = \frac{1}{s + \frac{\sigma_x}{\varepsilon_0}} \frac{\Upsilon^z_x}{\varepsilon_0}$$

$$E_y(t) = \int_0^t e^{-\frac{\sigma_x}{\varepsilon_0} \tau} \frac{\Upsilon^z_x (t - \tau)}{\varepsilon_0} d\tau$$

(6.18)

For leap-frog schemes, let us assume that all electrical field values are evaluated at time step $(n + \frac{1}{2})\Delta t$ and all magnetic field values are evaluated at time step $n\Delta t$. Then Eq. (6.18) is written as

$$E_y((n + \frac{1}{2})\Delta t) = \int_0^{(n + \frac{1}{2})\Delta t} e^{-\frac{\sigma_x}{\varepsilon_0} \tau'} \frac{\Upsilon^z_x ((n + \frac{1}{2})\Delta t - \tau)}{\varepsilon_0} d\tau'$$

(6.19)

Let $\tau = \tau' + \Delta t$, then

$$E_y((n + \frac{1}{2})\Delta t) = \int_{-\Delta t}^{(n - \frac{1}{2})\Delta t} e^{-\frac{\sigma_x}{\varepsilon_0} (\tau' + \Delta t)} \frac{\Upsilon^z_x ((n - \frac{1}{2})\Delta t - \tau')}{\varepsilon_0} d\tau'$$

$$+ \int_{-\Delta t}^0 e^{-\frac{\sigma_x}{\varepsilon_0} \tau'} \frac{\Upsilon^z_x ((n - \frac{1}{2})\Delta t - \tau')}{\varepsilon_0} d\tau'$$

$$= e^{-\frac{\sigma_x}{\varepsilon_0} \Delta t} E_y((n - \frac{1}{2})\Delta t) + \int_{-\Delta t}^0 e^{-\frac{\sigma_x}{\varepsilon_0} \tau'} \frac{\Upsilon^z_x ((n - \frac{1}{2})\Delta t - \tau)}{\varepsilon_0} d\tau'$$

$$= e^{-\frac{\sigma_x}{\varepsilon_0} \Delta t} E_y((n - \frac{1}{2})\Delta t) - \frac{1}{\sigma_x} \int_{-\Delta t}^0 \Upsilon^z_x ((n + \frac{1}{2})\Delta t - \tau) e^{-\frac{\sigma_x}{\varepsilon_0} \tau'} d\tau'$$

$$= e^{-\frac{\sigma_x}{\varepsilon_0} \Delta t} E_y((n - \frac{1}{2})\Delta t) - \frac{1}{\sigma_x} \left[ e^{-\frac{\sigma_x}{\varepsilon_0} \tau} \Upsilon^z_x ((n + \frac{1}{2})\Delta t - \tau) \int_{-\Delta t}^0 e^{-\frac{\sigma_x}{\varepsilon_0} \tau'} d\Upsilon^z_x ((n + \frac{1}{2})\Delta t - \tau) \right]$$

$$= e^{-\frac{\sigma_x}{\varepsilon_0} \Delta t} E_y((n - \frac{1}{2})\Delta t) - \frac{1}{\sigma_x} \left[ e^{-\frac{\sigma_x}{\varepsilon_0} \Delta t} \Upsilon^z_x ((n - \frac{1}{2})\Delta t) - \Upsilon^z_x ((n + \frac{1}{2})\Delta t) - \int_{-\Delta t}^0 e^{-\frac{\sigma_x}{\varepsilon_0} \tau'} d\Upsilon^z_x ((n + \frac{1}{2})\Delta t - \tau) \right]$$

(6.20)

Now let us focus on the integration $\int_0^{\Delta t} e^{-\frac{\sigma_x}{\varepsilon_0} \tau'} d\Upsilon^z_x ((n + \frac{1}{2})\Delta t - \tau)$. Note that $\Upsilon^z_x = -\frac{\partial (H_{xx} + H_{xy})}{\partial x}$ and $H$ are evaluated at time step $n\Delta t$. More importantly, the
FDTD method assumes an averaged value at the middle of each time step, i.e.

$$\Upsilon^z_x((n - \frac{1}{2})\Delta t) = \frac{\Upsilon^z_x(n\Delta t) + \Upsilon^z_x((n - 1)\Delta t)}{2}$$

In our case, $\Delta t$ is very small and the above relation could be expressed as

$$\Upsilon^z_x(t) = \Upsilon^z_x((n - 1)\Delta t) + [\Upsilon^z_x(n\Delta t) - \Upsilon^z_x((n - 1)\Delta t)] \eta(t - (n - \frac{1}{2})\Delta t) \quad (6.21)$$

for $(n - \frac{1}{2})\Delta t < t \leq (n + \frac{1}{2})\Delta t$ and $\eta(t)$ represents the Heaviside step function as

$$\eta(t) = \begin{cases} 1 & t > 0 \\ \frac{1}{2} & t = 0 \\ 0 & t < 0 \end{cases}$$

Thus

$$\Upsilon^z_x((n + \frac{1}{2})\Delta t - \tau) = \Upsilon^z_x((n - 1)\Delta t) + [\Upsilon^z_x(n\Delta t) - \Upsilon^z_x((n - 1)\Delta t)] \eta(\Delta t - \tau)$$

and

$$\frac{d\Upsilon^z_x((n + \frac{1}{2})\Delta t - \tau)}{d\tau} = -[\Upsilon^z_x(n\Delta t) - \Upsilon^z_x((n - 1)\Delta t)] \delta(\tau - \Delta t)$$

$$\int_0^{\Delta t} e^{-\frac{2\pi}{\epsilon_0} \tau} d\Upsilon^z_x((n + \frac{1}{2})\Delta t - \tau) = -\int_0^{\Delta t} e^{-\frac{2\pi}{\epsilon_0} \tau} [\Upsilon^z_x(n\Delta t) - \Upsilon^z_x((n - 1)\Delta t)] \delta(\tau - \Delta t) d\tau$$

$$= -e^{-\frac{2\pi}{\epsilon_0} \Delta t} [\Upsilon^z_x(n\Delta t) - \Upsilon^z_x((n - 1)\Delta t)] \quad (6.22)$$

Let us change the integration limit in Eq. (6.19) to

$$E_y((n + \frac{1}{2})\Delta t) = \int_{0+}^{(n + \frac{1}{2})\Delta t} e^{-\frac{2\pi}{\epsilon_0} \tau} \Upsilon^z_x((n + \frac{1}{2})\Delta t - \tau) d\tau$$

we subsequently have $\Upsilon^z_x((n + \frac{1}{2})\Delta t)^-$ and $\Upsilon^z_x((n - \frac{1}{2})\Delta t)^-$ instead of $\Upsilon^z_x((n + \frac{1}{2})\Delta t)$ and $\Upsilon^z_x((n - \frac{1}{2})\Delta t)$, where the superscript “-” sign represent the evaluation when approaching from the negative side. Also when using Eq. (6.21), we only consider

7This is valid in FDTD because its initial value can always be set to zero.
Υ\textsuperscript{z\times} evaluated at time steps \((n - 1)\Delta t\), \((n - \frac{1}{2})\Delta t\), and \(n\Delta t\) and ignore all details in between. Bearing this in mind, we can write \(Υ\textsuperscript{z\times}((n + \frac{1}{2})\Delta t)^{-} = Υ\textsuperscript{z\times}(n\Delta t)\) and \(Υ\textsuperscript{z\times}((n - \frac{1}{2})\Delta t)^{-} = Υ\textsuperscript{z\times}((n - 1)\Delta t)\). Now substitute Eq. (6.22) into the subsequently modified Eq. (6.20) and we obtain

\[
E_y((n + \frac{1}{2})\Delta t) = e^{-\frac{\sigma_y}{\epsilon_0}\Delta t}E_y((n - \frac{1}{2})\Delta t) - \frac{1}{\sigma_x}\left\{ e^{-\frac{\sigma_x}{\epsilon_0}\Delta t}Υ\textsuperscript{z\times}((n - 1)\Delta t) - Υ\textsuperscript{z\times}(n\Delta t) \right\}
+ e^{-\frac{\sigma_x}{\epsilon_0}\Delta t}[Υ\textsuperscript{z\times}(n\Delta t) - Υ\textsuperscript{z\times}((n - 1)\Delta t)]
\]

\[
= e^{-\frac{\sigma_x}{\epsilon_0}\Delta t}E_y((n - \frac{1}{2})\Delta t) + \frac{1-e^{-\frac{\sigma_x}{\epsilon_0}\Delta t}}{\sigma_x}Υ\textsuperscript{z\times}(n\Delta t)
\]

(6.23)

It is clear to see that Eq. (6.23) is the exponential differencing scheme. This proof shows that exponential differencing could be derived directly from the continuum Maxwell’s equation without further assumptions. Compared to the central differencing scheme, they have the same continuum correspondence. In other words, their behavior should be similar when time steps or conductivities are small.

It should be pointed out that \(E_y(t)\) in Eq. (6.18) is the convolution of \(Υ\textsuperscript{z\times}\) and Eq. (6.23) is nothing but recursive convolution [41]. Exponential differencing scheme is the same as recursive convolution. Because mid-point rule is also of second order accuracy, the exponential differencing and the central differencing schemes are expected to be the same in terms of order of accuracy.

The resultant coefficients by applying exponential differencing schemes to the PML-ADI are listed below:

\[
\begin{align*}
α_x^e & = e^{-\frac{\sigma_y}{\epsilon_0}\Delta t} & β_x^e & = 1-e^{-\frac{\sigma_y}{\epsilon_0}\Delta t} \\
α_y^e & = e^{-\frac{\sigma_x}{\epsilon_0}\Delta t} & β_y^e & = 1-e^{-\frac{\sigma_x}{\epsilon_0}\Delta t} \\
α_x^h & = e^{-\frac{\sigma_y}{\epsilon_0^2}\Delta t} & β_x^h & = 1-e^{-\frac{\sigma_y}{\epsilon_0^2}\Delta t} \\
α_y^h & = e^{-\frac{\sigma_x}{\epsilon_0^2}\Delta t} & β_y^h & = 1-e^{-\frac{\sigma_x}{\epsilon_0^2}\Delta t}
\end{align*}
\]

(6.24)
To test the performances, the same problem as in the previous subsection is simulated. Fig. 6.3 and Fig. 6.4 show that for the Courant number up to 5, the results are pretty much the same as the traditional implementation.

Fig. 6.3 and Fig. 6.4 further show some comparisons with the traditional implementation. As we see, for very small Courant numbers, such as $\chi = 0.5$, there are virtually no differences. For $\chi$ as large as 4, the exponential differencing scheme performs better by about 1dB. At $\chi = 6$ (the largest Courant number being tested), the improvements are within 3dB.

### 6.2.4 Proposed Approach

Because the ADI-FDTD method is most useful for large (i.e., greater than unity) Courant numbers (as long as the accuracy of the results is not impacted by the
Figure 6.4: Reflection errors from the PML-ADI-FDTD method with exponential time differencing (II).

Figure 6.5: Comparison of reflection errors from the PML-ADI-FDTD method with exponential time differencing and the traditional approach (I).
increased numerical dispersion in this regime, the recommended largest value of the Courant number when the cell size is between 10 cells/λ to 20 cells/λ is χ = 4), PML implementation that does not deteriorate for large Courant numbers is highly desirable. In this section, we introduce a novel PML implementation for the ADI-FDTD for this purpose.

As mentioned before, the coefficients in Eq. (6.15) are obtained by assuming central difference approximation in time on the l.h.s. of Eqs. (6.1)-(6.4). For the first sub-step of the algorithm, this implies that the r.h.s. of Eqs. (6.1)-(6.4) need to be evaluated at the time step \( n + \frac{1}{4} \), otherwise the l.h.s. and r.h.s. will not be collocated in time. For the second sub-step, the r.h.s. of Eqs. (6.1)-(6.4) similarly need to be evaluated at the time step \( n + \frac{3}{4} \). However, in the ADI method, the r.h.s. of
Eqs. (6.1)-(6.4) are evaluated either at the time step \((n + \frac{1}{2})\) or at the time step \(n\) as clear from Eqs. (6.5)-(6.8). This suggests that either forward or backward differencing approximations should be used instead for the l.h.s. of Eqs. (6.1)-(6.4) in order to be collocated in time with the spatial derivative terms of their respective r.h.s. Similar conclusion holds for the second procedure.

Therefore, in the improved PML implementation, two sets of coefficients are used, one for each sub-step of the ADI-FDTD update. In the first sub-step the following set of coefficients are used

\[
\begin{align*}
\alpha^e_x &= (2\epsilon_0 - \sigma_y \Delta t)/2\epsilon_0 \quad \beta^e_x = \Delta t/(2\epsilon_0 \Delta h) \\
\alpha^e_y &= 2\epsilon_0/(2\epsilon_0 + \sigma_x \Delta t) \quad \beta^e_y = \Delta t/[(2\epsilon_0 + \sigma_x \Delta t) \Delta h] \\
\alpha^h_x &= 2\epsilon_0/(2\epsilon_0 + \sigma_y \Delta t) \quad \beta^h_x = \Delta t/[(2\mu_0 + \sigma_x \eta_0^2 \Delta t) \Delta h] \\
\alpha^h_y &= (2\epsilon_0 - \sigma_y \Delta t)/2\epsilon_0 \quad \beta^h_y = \Delta t/(2\mu_0 \Delta h)
\end{align*}
\]

In the second sub-step, a different set of coefficients are used

\[
\begin{align*}
\alpha^e_x &= 2\epsilon_0/(2\epsilon_0 + \sigma_y \Delta t) \quad \beta^e_x = \Delta t/[(2\epsilon_0 + \sigma_y \Delta t) \Delta h] \\
\alpha^e_y &= (2\epsilon_0 - \sigma_x \Delta t)/2\epsilon_0 \quad \beta^e_y = \Delta t/(2\epsilon_0 \Delta h) \\
\alpha^h_x &= (2\epsilon_0 - \sigma_x \Delta t)/2\epsilon_0 \quad \beta^h_x = \Delta t/(2\mu_0 \Delta h) \\
\alpha^h_y &= 2\epsilon_0/(2\epsilon_0 + \sigma_y \Delta t) \quad \beta^h_y = \Delta t/[(2\mu_0 + \sigma_y \eta_0^2 \Delta t) \Delta h]
\end{align*}
\]

These coefficients are derived by assuming either forward or backward differencing in time for the l.h.s. of Eqs. (6.1)-(6.4). Fig. 6.7 and Fig. 6.8 shows the relative reflection error from this proposed PML from the same problem as before. From these figures, we observe that the proposed new PML has similar performances as the traditional one for small Courant numbers and it does not deteriorate for large Courant numbers. In particular, an improvement of more than 35 dB in the maximum reflection error is obtained for Courant factor \(\chi = 5\), compared to Fig. 6.2.

These results show that our approach is more appropriate than the traditional approach. Now it is not difficult to understand why the traditional approach performs worse. When using Eq.(6.15), we equivalently use the r.h.s. of Eqs.(6.1)-(6.4) evaluated either at time step \((n + \frac{1}{2})\) or at time step \(n\) to approximate their values at time
Figure 6.7: Reflection error from the PML-ADI-FDTD method with the proposed PML implementation (I).

Figure 6.8: Reflection error from the PML-ADI-FDTD method with the proposed PML implementation (II).
step \((n+\frac{1}{4})\). Because Eq.(15) is also used in Eqs.(6.10)-(6.13), a similar argument that we equivalently use the r.h.s. of Eqs.(6.1)-(6.4) evaluated either at time step \((n+\frac{1}{2})\) or at time step \((n+1)\) to approximate their values at time step \((n+\frac{3}{4})\) also hold. This approximation is good as long as the difference between any field component evaluated at a \(\Delta t/4\) time interval is small enough. When larger Courant numbers, therefore larger \(\Delta t\), are used, Eq.(6.15) is valid but the above approximation tends to be worse and worse.

### 6.3 Direct PML Implementation for FETD

Although the anisotropic PML is first proposed for FEM [55], its time-domain implementation enjoys much more success in FDTD. The FETD implementation only began to emerge recently [71],[28]. Ref.[71] suggested a reflection error of -50dB for 20 layers of PML absorbers. Also, a thorough numerical stability test, especially the late-time one, is not available. In this section, we shall first take a look at a direct implementation of the anisotropic PML for FETD.

Let us start from the two curl equations

\[
\nabla \times \vec{E} = -j\omega \mu_r[\Lambda]\vec{H} \\
\nabla \times \vec{H} = j\omega \epsilon_r[\Lambda]\vec{E}
\]

(6.27) (6.28)

where \(\mu_r\) and \(\epsilon_r\) are the relative permeability and the relative permittivity respectively.

\[
[\Lambda] = \begin{bmatrix}
\frac{s_x s_y}{s_x} & \frac{s_y s_z}{s_y} \\
\frac{s_y s_z}{s_y} & \frac{s_z s_x}{s_z} \\
\frac{s_z s_x}{s_z} & \frac{s_x s_y}{s_x}
\end{bmatrix}
\]

(6.29)

and

\[
s_x = \alpha_x + \frac{\sigma_x}{j\omega};
\]

(6.30)
\[ s_y = \alpha_y + \frac{\sigma_y}{j\omega}; \quad (6.31) \]
\[ s_z = \alpha_z + \frac{\sigma_z}{j\omega}; \quad (6.32) \]

Note that in the above expressions, \( \sigma_{x,y,z} \) are already normalized by \( \epsilon_0 \). For simplicity, we shall assume that evanescent waves are not being considered for absorption and let \( \alpha_{x,y,z} = 1 \). From Eq.(6.27), we have

\[ \nabla \times \frac{1}{\mu_r[A]} \nabla \times \vec{E} = -j\omega \nabla \times \vec{H} \quad (6.33) \]

Substituting Eq.(6.28) into Eq.(6.33) yields

\[ \nabla \times \frac{1}{\mu_r[A]} \nabla \times \vec{E} - \omega^2 \epsilon_r[A] \vec{E} = 0 \quad (6.34) \]

By applying Galerkin’s method to Eq.(6.34), we obtain

\[
\int \int \int_{\Omega_e} \left[ \left( \nabla \times \frac{1}{\mu_r[A]} \nabla \times \vec{E} \right) \cdot \vec{N}_i - \omega^2 \epsilon_r[A] \vec{E} \cdot \vec{N}_i \right] dv = 0
\]

where \( \vec{N}_i \) is the usual vector test function, \( \Omega_e \) represents the volume of an element. By applying Eq.(6.27), the above equation is written as

\[
\int \int \int_{\Omega_e} \left[ \left( \frac{1}{\mu_r[A]} \nabla \times \vec{E} \right) \cdot (\nabla \times \vec{N}_i) - \nabla \cdot (\vec{N}_i \times \frac{1}{\mu_r[A]} \nabla \times \vec{E}) - \omega^2 \epsilon_r[A] \vec{E} \cdot \vec{N}_i \right] dv = j\omega \int \int_{S_e} \vec{N}_i \cdot (\hat{n} \times \vec{H}) ds \quad (6.36)
\]

where \( S_e \) represents the surface of an element and \( \hat{n} \) the outer normal of \( S_e \). If we assume that no free surface currents exist on element boundaries, tangential \( \vec{H} \) will be continuous and the contribution of the \( \hat{n} \times \vec{H} \) terms of adjacent elements will cancel each other. This is due to the property that \( \vec{N}_i \) is curl conforming. So globally, Eq.(6.36) looks like

\[
\int \int \int_{\Omega_e} \left[ \left( \frac{1}{\mu_r[A]} \nabla \times \vec{E} \right) \cdot (\nabla \times \vec{N}_i) - \omega^2 \epsilon_r[A] \vec{E} \cdot \vec{N}_i \right] dv = 0 \quad (6.37)
\]
We next expand $\vec{E}$ as

$$\vec{E} = \sum_j e_j \vec{N}_j$$  \hspace{1cm} (6.38)

and Eq.(6.11) becomes

$$\int \int \int_{\Omega_e} \left[ \left( \frac{1}{\mu_r[\Lambda]} \nabla \times \vec{N}_j \right) \cdot (\nabla \times \vec{N}_i) - \omega^2 \epsilon_r[\Lambda] \vec{N}_j \cdot \vec{N}_i \right] d\vec{v} \cdot e_j = 0$$  \hspace{1cm} (6.39)

for each unknown coefficient $e_j$ in each element $\Omega_e$. To facilitate derivation, we shall transform to the Laplace domain, i.e., substitute the $j\omega$ in Eq.(6.39) with $s$. Thus Eq.(6.39) is transformed to

$$\int \int \int_{\Omega_e} \left[ \left( \frac{1}{\mu_r[\hat{\Lambda}]} \nabla \times \vec{N}_j \right) \cdot (\nabla \times \vec{N}_i) + \epsilon_r[\hat{\Lambda}] \vec{N}_j \cdot \vec{N}_i \right] d\vec{v} \cdot e_j = 0$$  \hspace{1cm} (6.40)

where

$$[\hat{\Lambda}] = \frac{1}{[\Lambda]} = \begin{bmatrix} \frac{s(s+\sigma_y)}{(s+\sigma_y)(s+\sigma_z)} & \frac{s(s+\sigma_z)}{(s+\sigma_x)(s+\sigma_y)} & \frac{s(s+\sigma_x)}{(s+\sigma_z)(s+\sigma_y)} \\ \frac{s(s+\sigma_x)}{(s+\sigma_y)(s+\sigma_z)} & \frac{s(s+\sigma_y)}{(s+\sigma_x)(s+\sigma_z)} & \frac{s(s+\sigma_y)}{(s+\sigma_x)(s+\sigma_z)} \\ \frac{s(s+\sigma_z)}{(s+\sigma_x)(s+\sigma_y)} & \frac{s(s+\sigma_x)}{(s+\sigma_y)(s+\sigma_z)} & \frac{s(s+\sigma_x)}{(s+\sigma_y)(s+\sigma_z)} \end{bmatrix}$$  \hspace{1cm} (6.41)

and

$$[\tilde{\Lambda}] = s^2[\Lambda] = \begin{bmatrix} \frac{s(s+\sigma_y)(s+\sigma_z)}{(s+\sigma_x)(s+\sigma_y)} & \frac{s(s+\sigma_z)(s+\sigma_x)}{(s+\sigma_y)(s+\sigma_z)} & \frac{s(s+\sigma_y)(s+\sigma_z)}{(s+\sigma_x)(s+\sigma_y)} \\ \frac{s(s+\sigma_x)(s+\sigma_y)}{(s+\sigma_z)(s+\sigma_x)} & \frac{s(s+\sigma_y)(s+\sigma_z)}{(s+\sigma_x)(s+\sigma_y)} & \frac{s(s+\sigma_x)(s+\sigma_y)}{(s+\sigma_z)(s+\sigma_x)} \\ \frac{s(s+\sigma_z)(s+\sigma_x)}{(s+\sigma_y)(s+\sigma_z)} & \frac{s(s+\sigma_y)(s+\sigma_z)}{(s+\sigma_x)(s+\sigma_y)} & \frac{s(s+\sigma_x)(s+\sigma_y)}{(s+\sigma_z)(s+\sigma_x)} \end{bmatrix}$$  \hspace{1cm} (6.42)

Now let us factorize $[\hat{\Lambda}]$ and $[\tilde{\Lambda}]$. For $[\hat{\Lambda}]_{11}$, we have

$$[\hat{\Lambda}]_{11} = 1 + \frac{s(\sigma_y - \sigma_z - \sigma_x)}{(s+\sigma_y)(s+\sigma_z)}$$  \hspace{1cm} (6.43)

The other two terms could be factorized in a similar way and we obtain

$$[\hat{\Lambda}] = [I] + [\hat{\Lambda}]_1$$  \hspace{1cm} (6.44)

where $[I]$ is the identity matrix and

$$[\hat{\Lambda}]_1 = \begin{bmatrix} \frac{s(\sigma_x - \sigma_y - \sigma_z)}{(s+\sigma_y)(s+\sigma_z)} & \frac{s(\sigma_y - \sigma_z - \sigma_x)}{(s+\sigma_z)(s+\sigma_y)} & \frac{s(\sigma_z - \sigma_x - \sigma_y)}{(s+\sigma_x)(s+\sigma_y)} \\ \frac{s(\sigma_y - \sigma_z - \sigma_x)}{(s+\sigma_z)(s+\sigma_y)} & \frac{s(\sigma_z - \sigma_x - \sigma_y)}{(s+\sigma_x)(s+\sigma_y)} & \frac{s(\sigma_x - \sigma_y - \sigma_z)}{(s+\sigma_y)(s+\sigma_z)} \\ \frac{s(\sigma_z - \sigma_x - \sigma_y)}{(s+\sigma_x)(s+\sigma_y)} & \frac{s(\sigma_x - \sigma_y - \sigma_z)}{(s+\sigma_y)(s+\sigma_z)} & \frac{s(\sigma_y - \sigma_z - \sigma_x)}{(s+\sigma_z)(s+\sigma_x)} \end{bmatrix}$$  \hspace{1cm} (6.45)
For $[\hat{\Lambda}]_{11}$, we have

$$[\hat{\Lambda}]_{11} = s^2 - (\sigma_x - \sigma_y - \sigma_z)s + \sigma_x^2 - \sigma_x(\sigma_y + \sigma_z) + \sigma_y\sigma_z - \frac{\sigma_x[\sigma_x^2 - \sigma_x(\sigma_y + \sigma_z) + \sigma_y\sigma_z]}{s + \sigma_x}$$

(6.46)

The other two terms could be factorized in a similar way and we have

$$[\hat{\Lambda}] = s^2[I] + s[\hat{\Lambda}]_1 + [\hat{\Lambda}]_2 + [\hat{\Lambda}]_3$$

(6.47)

where

$$[\hat{\Lambda}]_1 = \begin{bmatrix} \sigma_y + \sigma_z - \sigma_x & \sigma_x + \sigma_y - \sigma_z \\ \sigma_x + \sigma_y - \sigma_z & \sigma_z + \sigma_y - \sigma_x \end{bmatrix}$$

(6.48)

$$[\hat{\Lambda}]_2 = \begin{bmatrix} \sigma_x^2 - \sigma_x(\sigma_y + \sigma_z) + \sigma_y\sigma_z & \sigma_y^2 - \sigma_y(\sigma_z + \sigma_x) + \sigma_x\sigma_y \\ \sigma_y^2 - \sigma_y(\sigma_z + \sigma_x) + \sigma_x\sigma_y & \sigma_z^2 - \sigma_z(\sigma_x + \sigma_y) + \sigma_x\sigma_y \end{bmatrix}$$

(6.49)

$$[\hat{\Lambda}]_3 = \begin{bmatrix} \frac{-\sigma_x[\sigma_x^2 - \sigma_x(\sigma_y + \sigma_z) + \sigma_y\sigma_z]}{s + \sigma_x} & \frac{-\sigma_y[\sigma_y^2 - \sigma_y(\sigma_x + \sigma_z) + \sigma_x\sigma_y]}{s + \sigma_y} \\ \frac{-\sigma_y[\sigma_y^2 - \sigma_y(\sigma_x + \sigma_z) + \sigma_x\sigma_y]}{s + \sigma_y} & \frac{-\sigma_z[\sigma_z^2 - \sigma_z(\sigma_x + \sigma_y) + \sigma_x\sigma_y]}{s + \sigma_z} \end{bmatrix}$$

(6.50)

Now we are ready to perform inverse Laplace transformations and obtain the time-domain formulas. The $s$ and $s^2$ terms are transformed to

$$s \to \frac{\partial}{\partial t}$$

(6.51)

$$s^2 \to \frac{\partial^2}{\partial^2 t}$$

(6.52)

As to the rational terms, it was found convenient to utilize the Heaviside formula, which states that for

$$F(s) = \frac{P(s)}{Q(s)}$$

where $P(s)$ and $Q(s)$ are polynomials and the order of $P(s)$ is smaller than that of $Q(s)$, then the inverse transform of $F(s)$ is

$$\mathcal{L}^{-1}(F) = \sum_k \frac{P(s_k)}{Q'(s_k)} e^{skt}$$

(6.53)
where \( s_k \) are zeros of \( Q(s) \). With the above formula, we get

\[
L^{-1}([\tilde{\Lambda}]_3) = \text{diag}\left\{ \frac{\sigma_x}{\sigma_x - \sigma_y} e^{-\sigma_y t}, \frac{\sigma_y}{\sigma_y - \sigma_x} e^{-\sigma_x t}, \frac{\sigma_z}{\sigma_z - \sigma_x} e^{-\sigma_x t}, \frac{\sigma_z}{\sigma_z - \sigma_y} e^{-\sigma_y t} \right\}
\]

(6.54)

\[
L^{-1}([\hat{\Lambda}]_1) = \text{diag}\left\{ \frac{(\sigma_x - \sigma_y)}{\sigma_x - \sigma_z} e^{-\sigma_x t}, \frac{(\sigma_y - \sigma_z)}{\sigma_y - \sigma_x} e^{-\sigma_y t}, \frac{(\sigma_z - \sigma_x)}{\sigma_z - \sigma_y} e^{-\sigma_z t} \right\}
\]

(6.55)

and their corresponding time-domain contributions are

\[
\int \int \int_{\Omega} \epsilon_x L^{-1}([\tilde{\Lambda}]_3) \ast e_j \vec{N}_j \cdot \vec{N}_i \, dv
\]

(6.56)

\[
\int \int \int_{\Omega} \left( \frac{1}{\mu_r} L^{-1}([\hat{\Lambda}]_1) \right) \ast e_j \nabla \times \vec{N}_j \cdot \left( \nabla \times \vec{N}_i \right) \, dv
\]

(6.57)

where \( \ast \) represents convolution. For example

\[
e^{-\sigma_x t} \ast e_j = \int_{-\infty}^{t} e^{-\sigma_x \tau} e_j(t - \tau) \, d\tau
\]

(6.58)

As always the case in the FETD, the spatial integration in Eq.(6.57) is performed first to generate a coefficient matrix and the convolution is performed later in a recursive fashion at each time step. Up to now, we have all the necessities needed for time-domain implementations. Because this method directly use inverse Laplace transformations, as in Eqs.(6.54)–(6.55), we call it a direct implementation.

The major problem of the direct approach is stability. Stability is hard to prove or disprove analytically because the matrices are too complicated and the system amplification matrix, not each individual matrix, is not symmetric. To show the late-time instability, we run a simulation in an empty waveguide. The waveguide is 120-cell long and its cross-section is modelled by \( 8 \times 2 \) cells. The right end is terminated by ten layers of PML with piece-wise constant parabolic profile. The left end is simply terminated by PEC. A first-order differentiated Gaussian current

149
Figure 6.9: Late-time instability of direct implementation of PML for FETD.

source is located right next to the left end. The observation is five cells away from the PML interface. To avoid as much unwanted errors as possible, we use LU solver to solve the matrices. Fig. 6.9 shows the results of fifty thousand time steps. As can be seen, late-time instability became obvious from about thirty thousand time steps. The late-time instability is very detrimental to practical applications because we never know when it will happen beforehand.

6.4 PML Implementation for FETD with Conjugate-Scaled Test and Basis Functions

6.4.1 Formulation

We shall derive a novel implementation of PML for FETD methods with element-wise constant profile. 
*Element-wise constant profile* means that the PML conductivity remains constant inside each element. So the conductivity terms can be safely
pulled out from integration and differentiation inside each element. The proposed approach uses a different set of test and basis function. Instead of using $\tilde{N}_i$ as the test function and $\tilde{N}_j$ as the basis function in the direct approach, we let the test function $\tilde{N}_i$ to be

$$\tilde{N}_i = [S] \cdot \tilde{N}_i$$  \hspace{1cm} (6.59)$$

and the basis function $\tilde{N}_j$ to be

$$\tilde{N}_j = [S] \cdot \tilde{N}_j$$  \hspace{1cm} (6.60)$$

where

$$[S] = \begin{bmatrix} s_x \\ s_y \\ s_z \end{bmatrix} .$$

It is easy to see that outside the PML region, the above test and basis functions degenerate to the usual test and basis functions. We obtain the following equation inside each element $\Omega_e$ if we go through standard minimized residual approach

$$\sum_j \int \int_{\Omega_e} \left[ \left( \frac{1}{\mu_e} \nabla \times \tilde{N}_j \right) \cdot \left( \nabla \times \tilde{N}_i \right) - \omega^2 \varepsilon_r[A] \tilde{N}_j \cdot \tilde{N}_i \right] dv E_j = j \omega \int \int_{S_e} \tilde{N}_i \cdot (\hat{n} \times \vec{H}) ds$$  \hspace{1cm} (6.61)$$

For the above equation to be a valid finite-element approach, the question is whether the test and basis functions are curl conforming. Let us only look at the basis function $\tilde{N}_j$ because $\tilde{N}_i$ and $\tilde{N}_j$ have the same continuity properties. For element-wise constant profile, it is clear to see from Eq. (6.60) that $\tilde{N}_j$ is curl-conforming when cube elements are used. This is further illustrated in Fig. 6.10. For the edge being considered, its surrounding elements could only have different $\sigma_x$ and $\sigma_z$ values by the very nature of how PML conductivities are assigned. $\sigma_y$ is constant in all four elements since it only changes along the $\hat{y}$ direction. For other edges, the same argument hold.
Figure 6.10: PML region consisted of cubes.

Figure 6.11: Group tetrahedrons into a cube in the PML region.
For tetrahedral elements, $\tilde{N}_j$ is not curl conforming in general. However, if we group tetrahedra into cubes and assign the same $\sigma_x$, $\sigma_y$ and $\sigma_z$ values to all tetrahedra inside one cube, $\tilde{N}_j$ is again curl conforming. This is illustrated in Fig. 6.11. If tetrahedron 1 and tetrahedron 2 have different $\sigma_x$ and $\sigma_z$ values, then $\tilde{N}_j$ corresponding to edge 12 is not continuous. Because $\tilde{N}_j$ has both $\hat{x}$ and $\hat{z}$ components and from Eq. (6.60), its $\hat{x}$ and $\hat{z}$ components are multiplied by different $\sigma_x$ and $\sigma_z$ values in tetrahedra 1 and 2 respectively. But if $\sigma_x$ and $\sigma_z$ are the same in both tetrahedra 1 and 2, the corresponding $\tilde{N}_j$ will be continuous. Since only $\sigma_y$ changes along $\hat{y}$ direction, curl conforming remains when other tetrahedral elements containing edge 12 are being considered. Therefore, by specially arranging the tetrahedra, the proposed test and basis functions are valid for tetrahedral elements. As long as tetrahedra are grouped into layers in the longitudinal direction and the longitudinal conductivity remains constant inside each layer, the test and basis functions remain valid. For pyramids and pentahedra, we can arrange them similarly to achieve curl conforming.

In general, we have the following with regard to the curl of the proposed basis function

$$\nabla \times ([S] \cdot \tilde{N}_j) = \begin{bmatrix} 0 & -\frac{\partial N_{jy}}{\partial z} & -\frac{\partial N_{jz}}{\partial y} \\ -\frac{\partial N_{jx}}{\partial y} & 0 & -\frac{\partial N_{jz}}{\partial x} \\ -\frac{\partial N_{jx}}{\partial x} & \frac{\partial N_{jy}}{\partial x} & 0 \end{bmatrix} \cdot \begin{bmatrix} s_x \\ s_y \\ s_z \end{bmatrix} \quad (6.62)$$

For simplicity, we shall derive formulas for brick elements only. Others follow the same procedure and will not be repeated here. In this case, Eq. (6.62) reduces to

$$\nabla \times ([S] \cdot \tilde{N}_j) = s_j \nabla \times \tilde{N}_j \quad (6.63)$$

where $s_j$ represents either $s_x$, or $s_y$, or $s_z$ depending on the direction of $\tilde{N}_j$. For example, if $\tilde{N}_j$ is $\hat{x}$ directed, then $s_j = s_x$. Bearing in mind that the contributions to the right hand side of Eq. (6.61) from different elements cancel out due to the
tangential continuity of $\tilde{N}_i^*$, we obtain the following weak form

$$\sum_j \int \int \int_{\Omega_e} \left[ \frac{1}{\mu_r[\Lambda]} s_j \nabla \times \tilde{N}_j \cdot (s_i \nabla \times \tilde{N}_i) - \omega^2 \epsilon_r[\Lambda] s_j \tilde{N}_j \cdot s_i \tilde{N}_i \right] dv E_j = 0 \quad (6.64)$$

Now, let us look at how to construct elemental matrices. For the second term in Eq. (6.64), we have

$$-\omega^2 \epsilon_r[\Lambda] s_j \tilde{N}_j \cdot s_i \tilde{N}_i = -\omega^2 \epsilon_r s_x s_y s_z \tilde{N}_j \cdot \tilde{N}_i$$

because it is non-zero only when $\tilde{N}_i$ and $\tilde{N}_j$ pointing to the same direction. Note that

$$\int \int \int_{\Omega_e} \epsilon_r \tilde{N}_j \cdot \tilde{N}_i dv$$

resembles the ordinarily defined mass matrix, the second term in Eq. (6.64) is written in the frequency domain as

$$[(j\omega)^2 + j\omega(\sigma_x + \sigma_y + \sigma_z) + (\sigma_x \sigma_y + \sigma_x \sigma_z + \sigma_y \sigma_z) + \frac{\sigma_x \sigma_y \sigma_z}{j\omega}] [T] \quad (6.65)$$

As to the first term in Eq. (6.64), it is ready to prove its equivalence to the following matrix combination

$$\begin{bmatrix}
\frac{\epsilon_x l_x^c \epsilon_y l_y^c}{\sigma_y l_y^c} [K_1] + \frac{\epsilon_y l_y^c \epsilon_z l_z^c}{\sigma_z l_z^c} [K_2]
& -\frac{\sigma_x l_x^c}{6 \sigma_x l_x^c} [K_3]
& -\frac{\sigma_y l_y^c}{6 \sigma_y l_y^c} [K_3]^T

-\frac{\sigma_y l_y^c}{6 \sigma_y l_y^c} [K_3]^T
& \frac{\epsilon_y l_y^c \epsilon_z l_z^c}{\sigma_z l_z^c} [K_1] + \frac{\epsilon_z l_z^c \epsilon_y l_y^c}{\sigma_y l_y^c} [K_2]
& -\frac{\sigma_y l_y^c}{6 \sigma_y l_y^c} [K_3]

\frac{1}{j\omega} - \frac{\epsilon_x l_x^c \epsilon_y l_y^c}{\sigma_y l_y^c} [K_2]
& 0
& 0

\frac{\epsilon_y l_y^c \epsilon_z l_z^c}{\sigma_z l_z^c} [K_1]
& 0
& 0

\frac{1}{j\omega} - \frac{\epsilon_y l_y^c \epsilon_z l_z^c}{\sigma_z l_z^c} [K_2]
& 0
& 0

\frac{(\sigma_x - \sigma_y)(\sigma_x - \sigma_y)}{\sigma_y l_y^c}
& \frac{\sigma_y}{\sigma_y + j\omega}
& - \frac{\sigma_y l_y^c}{6 \sigma_y l_y^c} [K_1]

\frac{\sigma_x \sigma_y}{\sigma_y + j\omega}
& 0
& \frac{\sigma_x l_x^c \epsilon_y l_y^c}{\sigma_y l_y^c} [K_1]

\frac{\sigma_x \sigma_x}{\sigma_x + j\omega}
& 0
& 0

\frac{\sigma_x l_x^c \epsilon_y l_y^c}{\sigma_y l_y^c} [K_2]
& 0
& 0

\frac{\sigma_y l_y^c}{\sigma_y + j\omega}
& 0
& \frac{\sigma_y l_y^c}{\sigma_y l_y^c} [K_1]
\end{bmatrix} + [S] \quad (6.66)$$

154
where \( l_x, l_y, \) and \( l_z \) are the edge length in each direction, \([S]\) is the usual stiffness matrix defined by

\[
[S] = \begin{bmatrix}
\frac{l_x l_z}{\nu} [K_1] + \frac{l_y l_z}{\nu} [K_2] & -\frac{l_x}{\nu} [K_3] & -\frac{l_y}{\nu} [K_3]^T \\
-\frac{l_x}{\nu} [K_3]^T & \frac{l_y l_z}{\nu} [K_1] + \frac{l_y l_z}{\nu} [K_2] & -\frac{l_y}{\nu} [K_3] \\
-\frac{l_y}{\nu} [K_3] & -\frac{l_y}{\nu} [K_3]^T & \frac{l_y l_z}{\nu} [K_1] + \frac{l_z l_x}{\nu} [K_2]
\end{bmatrix}
\]

and

\[
[K_1] = \begin{bmatrix}
2 & -2 & 1 & -1 \\
-2 & 2 & -1 & 1 \\
1 & -1 & 2 & -2 \\
-1 & 1 & -2 & 2
\end{bmatrix}
\]

\[
[K_2] = \begin{bmatrix}
2 & 1 & -2 & -1 \\
1 & 2 & -1 & -2 \\
-2 & -1 & 2 & 1 \\
-1 & -2 & 1 & 2
\end{bmatrix}
\]

\[
[K_3] = \begin{bmatrix}
2 & 1 & -2 & -1 \\
-2 & -1 & 2 & 1 \\
1 & 2 & -1 & -2 \\
-1 & -2 & 1 & 2
\end{bmatrix}
\]

Eqs. (6.65)-(6.66) are all we need for the proposed PML implementation in frequency domain. Actually, these formulas reuse and only use building blocks utilized by ordinary FETD methods. Its implementation could be easily adapted from the non-PML case. Also, Eq. (6.66) is for the most general case (more exactly, for the corner regions) but fails when any one of \( \sigma_x, \sigma_y \) and \( \sigma_z \) is zero. In contrast, Eq. (6.65) has no singular problem in any region. In cases where any one of \( \sigma_x, \sigma_y \) and \( \sigma_z \) is zero, we should go back to Eq. (6.60) and start derivation from there. For example, when \( \sigma_x \neq 0, \sigma_y = 0 \) and \( \sigma_z = 0 \), the definition of \([\Lambda]\)

\[
[\Lambda] = \begin{bmatrix}
s_y s_x s_z & 0 & 0 \\
0 & s_x s_z & 0 \\
0 & 0 & s_x
\end{bmatrix}
\]

degenerates to

\[
[\Lambda] = \begin{bmatrix}
\frac{1}{s_y} & 0 & 0 \\
0 & s_x & 0 \\
0 & 0 & s_x
\end{bmatrix}
\]
From Eq. (6.71), the first term in Eq. (6.64) is equivalent to the combination of following matrices

\[
[S] + \begin{bmatrix}
\frac{\partial \mathbf{E}}{\partial t} \mathbf{K}_1 + \frac{\partial \mathbf{E}}{\partial x} \mathbf{K}_2 & 0 & 0 \\
0 & -\frac{\partial \mathbf{E}}{\partial y} \mathbf{K}_3 & 0 \\
0 & -\frac{\partial \mathbf{E}}{\partial z} \mathbf{K}_3^T & \frac{\partial \mathbf{E}}{\partial t} \mathbf{K}_1
\end{bmatrix}
\frac{\sigma_x}{\sigma_x + j\omega} \\
\begin{bmatrix}
0 & 0 & 0 \\
0 & \frac{\partial \mathbf{E}}{\partial t} \mathbf{K}_2 & 0 \\
0 & 0 & \frac{\partial \mathbf{E}}{\partial t} \mathbf{K}_1
\end{bmatrix}
\frac{\sigma_x}{\sigma_x + j\omega}
\]

(6.72)

Other cases follow the same rule.

To transform the frequency-domain equations to time domain, we need to apply the inverse Fourier transformation. The major difference between the PML region and the non-PML region is the fractional terms, which will be transformed to convolutions and simple integrations in time domain. In the following, we shall focus on the time-domain implementations of these convolutions and simple integrations.

The following formula is found useful for the inverse Fourier transformation

\[
\mathcal{F}^{-1}\left(\frac{\sigma_i}{\sigma_i + j\omega} : \mathcal{E}\right) = \sigma_i \int_0^t e^{-\sigma_i \tau} E(t - \tau) d\tau
\]

(6.73)

where \(\sigma_i\) could be \(\sigma_x, \sigma_y\) or \(\sigma_z\). The above convolution can be evaluated in a recursive manner. If we define the convolution at time step \(n\) as

\[
I^n_i = \sigma_i \int_0^{n\Delta t} e^{-\sigma_i \tau} E(n\Delta t - \tau) d\tau
\]

(6.74)

the recursive convolution is expressed as

\[
I^n_i = e^{-\sigma_i \Delta t} I^{n-1}_i + (1 - e^{-\sigma_i \Delta t})[E^n + e^{-\sigma_i \Delta t} E^{n-1}]
\]

(6.75)

To achieve unconditionally stability, we use the Generalized-\(\Theta\) method with \(\theta_1 = 0\) and \(\theta_2 = 1/8\). By doing so, \(I^n_i\) is replaced by

\[
\frac{1}{4} I^{n+1}_i + \frac{1}{2} I^n_i + \frac{1}{4} I^{n-1}_i = \frac{1}{4} (1 + e^{-\sigma_i \Delta t})^2 I^{n-1}_i + \frac{1}{4} (1 - \frac{\sigma_i \Delta t}{2}) E^{n+1}_i + \frac{1}{4} (1 - \frac{\sigma_i \Delta t}{2}) (e^{-\sigma_i \Delta t} + 2 + e^{-\sigma_i \Delta t}) E^n + \frac{1}{4} (1 - \sigma_i \Delta t/2) (1 - e^{-\sigma_i \Delta t}) (2 + e^{-\sigma_i \Delta t}) E^{n-1}
\]

(6.76)
Besides convolution terms, simple integration terms defined by

\[ I^n = \int_0^{n\Delta t} E(\tau) d\tau \]  

(6.77)

are also encountered. Its recursive form is

\[ I^n = I^{n-1} + \frac{E^n + E^{n-1}}{2} \Delta t \]  

(6.78)

By applying the Generalized-\( \Theta \) method, \( I^n \) is replaced by

\[ \frac{1}{4} I^{n+1} + \frac{1}{2} I^n + \frac{1}{4} I^{n-1} = I^{n-1} + \frac{\Delta t}{8} E^{n+1} + \frac{\Delta t}{2} E^n + \frac{3\Delta t}{8} E^{n-1} \]  

(6.79)

In the edge regions, another integration will be used. It is in the form of

\[ I^n_2 = \int_0^{n\Delta t} \int_0^\tau E(\kappa) d\kappa d\tau \]  

(6.80)

Its recursive form is

\[ I^n_2 = I^{n-1}_2 + I^{n-1} \Delta t + \frac{E^n + E^{n-1}}{4} \Delta t^2 \]  

(6.81)

where \( I^n \) is defined by Eq. (6.77). By the Generalized-\( \Theta \) method, \( I^n_2 \) is replaced by

\[ \frac{1}{4} I^{n+1}_2 + \frac{1}{2} I^n_2 + \frac{1}{4} I^{n-1}_2 = I^{n-1}_2 + I^{n-1} \Delta t + \frac{\Delta t^2}{16} E^{n+1} + \frac{3\Delta t^2}{8} E^n + \frac{5\Delta t^2}{16} E^{n-1} \]  

(6.82)

Eqs. (6.76), (6.79) and (6.82) are all specially needed for the proposed implementation of PML. In these equations, the convolutions and integrations do not enter as new unknowns to the system, which keep the extra memory consumption low.

### 6.4.2 Numerical Results

To demonstrate its performance, let us first consider a waveguide problem. The waveguide cross-section is modelled by 9×3 cells with cell size being 0.74925cm. The source is the incident/scattered field type with a modulated Gaussian pulse, whose
Figure 6.12: Reflections from a single interface when the theoretical reflection error is -20dB at normal incidence.

Center frequency and bandwidth are 3.75GHz and 2.5GHz respectively. The power at 2.5GHz and 5GHz is about -20dB down with respect to that at 3.75GHz. The observation is three cells away from the source and also three cells away from the PML interface. One end of the waveguide is terminated with ten layers of PML. The other end is set to be long enough to isolate any spurious reflection. The reference solution is obtained by setting both ends to be long enough. To assure accuracy, LU solver is used.

We first set PML conductivities corresponding to a theoretical reflection at normal incidence and compare the reflection error by applying different conductivity profiles. The results are shown in Figs. 6.12-6.15.

It is clear that the proposed PML performances correspond to the theoretical values very well when the conductivities are small, as shown in Figs. 6.12-6.13.
Figure 6.13: Reflections from a single interface when the theoretical reflection error is -40dB at normal incidence.

Figure 6.14: Reflections from a single interface when the theoretical reflection error is -60dB at normal incidence.
Figure 6.15: Reflections from a single interface when the theoretical reflection error is -80dB at normal incidence.

For the cubic and fourth order profiles, it performs even slightly better than the theoretical limit. The same has also been observed in the FDTD. The results shown in Figs. 6.14-6.15 are not as smooth as those in small conductivity cases. This is due to two reasons. First, the finite truncation of the time-domain results (300 time steps in our tests) causes ripples during the Fourier transformation. When reflection errors are comparable to ripples’ level, ripples become pronouncing. If enough time steps are being executed, we expect smoother results as in Figs. 6.12-6.13. Second, there is a limit of how small the reflection error could practically be.

From these four figures, we see that the fourth order profile performs the best. This is the same as the FDTD. Fig. 6.16 shows the comparison between the reference results and the observation when the PML is present. The PML conductivity is such that the theoretical reflection is -80dB at normal incidence. As can be seen, virtually
no difference is visible. Fig. 6.17 further shows the reflection in time domain. In fact, the maximum reflection errors correspond to the minimum reflection errors in frequency domain very well (see Figs. 6.12-6.15). In very low reflection error cases, we consider this as a better measurement since it is immune to the ripples mentioned above.

Next, we shall find out what is the lowest reflection error we can get with respect to ten layers of PML. For obvious reasons, the fourth order conductivity profile is employed. Fig. 6.18 shows the results. As we increase the maximum conductivity, the maximum reflection error goes down and then pops up. The optimum value is about $\sigma_{\text{max}} = 0.8$ (Siemens) where -64dB reflection error is obtained. When the PML absorbers increase to 20 layers, the optimum value is about $\sigma_{\text{max}} = 0.5$ (Siemens) where the lowest reflection error is about -80dB.
Figure 6.17: Reflection errors in time domain.

Figure 6.18: Reflection errors for ten layers of PML with the fourth order conductivity profile.
Finally, we show the late-time performances of the proposed PML implementation. The first example is an empty waveguide, which is exactly the same as in section §6.2. The only difference is the PML implementations. The second example illustrated in Fig.6.21 is a $18 \times 10 \times 8$ waveguide-like structure loaded with inhomogeneous material and a PEC rod. Five layers of PML are set at the right end and a current source is set at the left end right next to the PEC wall. The observation is eight cells away from the left end. To excite possible unstable modes, the relative permittivity varies as $\epsilon_r = 1 + \sqrt{z}$, where $z$ varies from zero to eight. The PEC rod is “L” shaped and measured three cells in height and two cells in width. It sits two cells away from the bottom and one cell from the side PEC wall. Also, the rod starts from half way of the structure and goes all the way down to the end of the PML region.

Figure 6.19: Reflection errors for 20 layers of PML with the fourth order conductivity profile.
Figure 6.20: Late-time performance of the proposed PML implementation: empty waveguide.

The results for both the empty waveguide and the loaded structure are shown in Fig.6.20 and Fig.6.22 respectively. For the empty waveguide, there is no sign of instability even after one million time steps. For the loaded structure, Fig.6.22 shows the results of one million time steps. It is interesting to see the linear growth after 200,000 time steps. Indeed, the linear growth could be different if the same program is compiled and run on different operating systems. It is caused by the amplified round-off errors and is not due to the stability of the proposed PML itself. It should be pointed out that the linear growth is common to FETD methods. A practical treatment can be found in Ref. [25].

In reality, iterative solves are more preferable than direct solvers for the FETD method. This is especially true for large-scale problems where direct solvers are not even possible. Next, the condition numbers of the matrices (as denoted in Eq.
Figure 6.21: PEC structure of the loaded waveguide with stratified dielectric medium.

Figure 6.22: Late-time performance of the proposed PML implementation: loaded structure.
(6.61)) associated with the PML regions are plotted as a function of the maximum conductivities in Fig. 6.24. The matrices are generated by filling a 12×9×3 PEC box with quadratic-profiled PML absorbers. The minimum is at the left end and the maximum appears at the right end. Note that when the maximum reflection is set to be 1, it simply corresponds to free space. Fig. 6.24 shows that the condition number indeed grows continuously as the conductivity gets larger. But the growth is not large and the condition numbers are still reasonably small.

Fig. 6.24 shows that the condition numbers of other three problems. “Full PML: 12 cells total” indicates the results of a 12×9×3 PEC box fully filled with constant-profiled PML absorbers. “Half PML: 12 cells total” indicates the results of a 12×9×3 PEC box filled with 6-layer constant-profiled PML absorbers. “Half PML: 24 cells total” indicates the results of a 24×9×3 PEC box filled with 12-layer constant-profiled
PML absorbers. This figure implies that for real problems, where computational domains are relatively larger and only partially filled with PML absorbers, the presence of PML absorbers will further increase the condition numbers, but not largely.

6.4.3 Application to Implicit Non-Staggered FDTD Method

The only difference between an implicit non-staggered FDTD (INS-FDTD) program and a FETD program using cube elements is applying Eqs. (6.46)–(6.49) instead of Eqs. (6.42)–(6.45). Therefore, the element-wise constant PML developed for the FETD in this section can be naturally extended to the INS-FDTD.

To test reflection errors, late-time stability, and CG solvers’ performances, the simulations are set in exactly the same way as described in the previous subsection. Figs. 6.25-6.28 demonstrate the time-domain reflection errors of ten layers of PML for different conductivity profiles and different absorption levels. When the maximum
conductivity is relatively small, fourth order profile performs the best (Fig. 6.25). When the maximum conductivity increases, the cubic profile is the best choice as shown in Figs. 6.26-6.28. The minimum reflection error is about -58 dB when cubic profile is used and the maximum conductivity is such that the theoretical reflection error is -60 dB at normal incidence(Fig. 6.26).

The late-time stability is shown in Fig. 6.29. Instead of testing a 120-cell long waveguide structure, a 60-cell long waveguide with ten layers of PML at the right end is tested. Because the waveguide is shortened, a larger portion of it is filled with PML and better wave absorption is obtained. As a result, linear growth caused by numerical noises is seen in this simulation.

The condition numbers of the matrices associated with the PML regions are plotted as a function of the maximum conductivities in Fig. 6.30. They are smaller
Figure 6.26: Time domain reflection errors: $|\Gamma| = -60$ dB at normal incidence.

Figure 6.27: Time domain reflection errors: $|\Gamma| = -80$ dB at normal incidence.
Figure 6.28: Time domain reflection errors: $|\Gamma| = -100$ dB at normal incidence.

Figure 6.29: Late-time performance of the proposed PML for the implicit FDTD.
compared to the corresponding FETD cases. In fact, it is observed that the number of iterations required by the INS-FDTD is about 60% of that required by the FETD when the same stopping criterion is applied.

6.5 Concluding Remarks

In this chapter, we have improved the PML performances for both ADI-FDTD and FETD methods. The novel PML implementation for FETD is also naturally extended to the INS-FDTD method. Although the PML performances for FETD are good in general, they are not as good as the FDTD, where -80 dB to -100 dB reflection errors can be obtained for ten layers of PML. One possible reason is that the PML conductivity actually changes cell by cell in the FETD implementation, while it changes every half cell in the FDTD. The degraded PML performance for FETD is
thought to be at least partially due to this “coarseness”. A new PML which supports continuous conductivity profiles is currently under investigation.
CHAPTER 7

CONCLUSIONS AND FUTURE WORK

Computational Electromagnetics thrived in past decades. Large research efforts were put on making main stream methods, such as FDTD, FEM, and MoM, more accurate and more efficient, either in general or to specific problems. In this dissertation, we have proposed and discussed several schemes intended for accuracy and/or efficiency improvement.

In Chapter 2 and 3, we developed two higher order leap-frog schemes. They all start from the numerical dispersion error expressions and determine their unknown coefficients by certain criteria. Chapter 2 proposed the Angle-Optimized FDTD method, which allows the dispersion error to be minimized at any propagation angle. Chapter 3 proposed the Dispersion-Relation-Preservation FDTD method, which minimizes the dispersion errors for all angles. The significance of these two methods also relies on the fact that they provide a new look on the FDTD method and its higher order variants from the optimization perspective.

Large-scale problems with locally fine structures are tackled in Chapter 4 and Chapter 5. The major problem with FDTD subgridding methods is late-time instability. Because of this, we choose to model local fine structures with unstructured mesh and other parts with structured mesh. To connect unstructured and structured
regions, we first proposed cube-shaped composite elements combining pyramids and tetrahedra. Based on this type of mesh, the FDTD/FETD hybrid methods are numerically examined for their stabilities. To overcome the observed late-time instability, we proposed the implicit non-staggered FDTD method, which is specially useful when working with FETD to form a hybrid method with guaranteed stability.

Time-domain PML implementations are considered in Chapter 6. An improved ADI-FDTD implementation is proposed by using either forward or backward differences in the two sub-steps for conduction terms. Compared to the traditional approach, the reflection error is greatly reduced (by 40 dB in our tests). PML for FETD is also considered in Chapter 6. Previously, PML implementations for FETD suffer from either bad performances or late-time instability. These problems are solved in our proposed approach, where a conjugate-scaled test and basis functions are used in conjunction with a minimized residual approach. Numerical tests show stable late-time performances and reduced reflection error (by 30 dB compared to Ref. [71]).

Looking into the future, there are several things to do. Higher order FDTD schemes enjoy much success in homogenous problems without infinite thin PEC structures. For inhomogeneous material or infinite thin PEC conductors, there are still lack of elegant solutions. To extend the usefulness of higher order methods, these problems must be solved.

Composite elements present a different type of mesh generation problem. Here, the boundary of tetrahedral regions is given before generating tetrahedra. A way to generate this type of mesh is Advancing Front Method [20]. How to use this method to generate qualified meshes for hyperbolic PDE problems still needs research.
The absorbing boundary conditions for FETD methods will continue to be a topic. Smaller reflection error is always preferable. One promising way is to explore PML supporting continuous conductive profiles.
BIBLIOGRAPHY


[24] T. Hirono, L. Wayne, S. Seki, and Y. Yoshikuni  A three-dimensional fourth-
order finite-difference time-domain scheme using a symplectic integrator propaga-

[25] C-T Hwang and R-B Wu  Treating Late-Time Instability of Hybrid Finite-
Element/Finite-Difference Time-Domain Method, IEEE Trans. Antennas Propa-

[26] C-T Hwang, S-G Mao, R-B Wu and C-H Chen  Partially Prism-Gridded FDTD
Analysis for Layered Structures of Transversely Curved Boundary, IEEE Trans.

[27] J. M. Hyman and M. Shashkov  The orthogonal decomposition theorem for

simulation of three-dimensional scattering and radiation problems using perfectly


[30] I. S. Kim and W. J. R. Hoefer  A local mesh refinement for the time-domain
finite-difference method using Maxwell’s curl equations, IEEE Trans. Microwave


2000.

[33] P. R. Kotiuga  Hodge decompositions and computational electromagnetics, Ph.D.

[34] K. M. Krishnaiah, C. J. Railton  Passive equivalent circuit of FDTD: An appli-

[35] K. S. Kunz and L. Simpson  A technique for increasing the resolution of finite-
difference solution of the Maxwell equation., IEEE Trans. Electromagn. Compat.,


[53] T. Rylander, A. Bondeson *Application of Stable FEM-FDTD hybrid to Scattering Problems*,


182