ACCURATE AND EFFICIENT METHODS FOR MULTISCALE AND MULTIPHYSICS ANALYSIS

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Kaiyuan Zeng

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

May 2019

Purdue University

West Lafayette, Indiana

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF DISSERTATION APPROVAL

Dr. Dan Jiao, Chair

School of Electrical and Computer Engineering

Dr. Peter Bermel

School of Electrical and Computer Engineering

Dr. Byunghoo Jung

School of Electrical and Computer Engineering

Dr. Saeed Mohammadi

School of Electrical and Computer Engineering

Approved by:

Dr. Pedro Irazoqui

Head of the School Graduate Program

To my family.

ACKNOWLEDGMENTS

First of all, I would like to express my sincere thanks to my advisor Professor Dan Jiao for providing me the opportunity to work with her. She has been instrumental in everything from my first exposure to the computational EM equations to the final part of this dissertation. I really appreciate her inspiring words, which help me a lot to regain my confidence and to become a better version of myself.

I would also like to thank the other members of my PhD advisory committee: Professor Peter Bermel, Professor Byunghoo Jung and Professor Saeed Mohammadi for their support and suggestions for this dissertation.

Besides, many thanks to the On-Chip Electromagnetic group members: Miaomiao Ma, Li Xue, Shuzhan Sun, Chang Yang, Yifan Wang, Michael Hayashi, Zhangchao Wei, Ping Li and Yu Zhao, for their helpful research discussion and friendship.

Last but not least, I would like to thank my parents and my wife, for their love and support during all these years.

TABLE OF CONTENTS

Page

Page

Page

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Zeng, Kaiyuan Ph.D., Purdue University, May 2019. Accurate and Efficient Methods for Multiscale and Multiphysics Analysis. Major Professor: Dan Jiao.

Multiscale and multiphysics have been two major challenges in analyzing and designing new emerging engineering devices, materials, circuits, and systems. When simulating a multiscale problem, numerical methods have to overcome the challenges in both space and time to account for the scales spanning many orders of magnitude difference. In the finite-difference time-domain (FDTD) method, subgridding techniques have been developed to address the multiscale challenge. However, the accuracy and stability in existing subgridding algorithms have always been two competing factors. In terms of the analysis of a multiphysics problem, it involves the solution of multiple partial differential equations. Existing partial differential equation solvers require solving a system matrix when handling inhomogeneous materials and irregular geometries discretized into unstructured meshes. When the problem size, and hence the matrix size, is large, existing methods become highly inefficient.

In this work, a symmetric positive semi-definite FDTD subgridding algorithm in both space and time is developed for fast transient simulations of multiscale problems. This algorithm is stable and accurate by construction. Moreover, the method is further made unconditionally stable, by analytically finding unstable modes, and subsequently deducting them from the system matrix. To address the multiphysics simulation challenge, we develop a matrix-free time domain method for solving thermal diffusion equation, and the combined Maxwell-thermal equations, in arbitrary unstructured meshes. The counterpart of the method in frequency domain is also developed for fast frequency-domain analysis. In addition, a generic time marching

scheme is proposed for simulating unsymmetrical systems to guarantee their stability in time domain.

1. INTRODUCTION

1.1 Background of Multiscale and Multiphysics Analysis

A broad range of electromagnetic engineering problems rely on theoretical developments and computational techniques to solve problems spanning different scales or physical properties, making multiscale and multiphysics two major challenges in analyzing and design new emerging engineering devices, materials, circuits, and systems.

The finite-difference time-domain (FDTD) method is widely used as time-domain method for general electromagnetic analysis [1], [2]. It has the advantages as being simple and matrix-free, namely, free of matrix solution. The stability analysis of the conventional FDTD method is straightforward because the numerical system is symmetric positive semi-definite (SPD) while a uniform orthogonal grid is required. However, if there are fine features in a structure, a small space step must be utilized in the discretization to capture the quick variance of the fields. Due to the uniform orthogonal grid property, such small space step has to be applied to regions where there are no fine features. This requirement increases the number of unknowns to be solved, especially the global time step is also restricted by the smallest space step, making the conventional FDTD method not efficient for solving multiscale problems.

Subgridding is an effective method to refine a grid locally in the FDTD method, and this is especially useful when simulating a multiscale problem. Ideally, when simulating a multiscale problem, a subgridding method in both space and time can not only reduce the number of unknowns, but also permit the use of a local time step. In other words, the time step in a base grid is not restricted by that in the subgrid for a stable explicit time marching. Each grid region can be simulated stably using its local time step, and hence greatly accelerating an FDTD simulation.

The electrical-thermal co-analysis is of critical importance in advanced integrated circuit (IC) design, where emerging interconnect solutions have been intensively pursued to overcome the shortcoming of existing copper-based interconnects in performance and reliability. The design of the new interconnect solutions typically involves many physics such as circuits, electromagnetics, materials, electron transport, and thermal diffusion in a broad band of frequencies. To understand the entire physical process happening in an advanced IC design, a rigorous and efficient multiphysics simulation is required.

The analysis of a multiphysics problem involves the solution of multiple partial differential equations (PDEs). Existing solvers for solving PDEs generally cast the original physical problem into a matrix equation of $\mathbf{A}x = b$ to solve [3, 4], where $\mathbf{A}$ can be either dense or sparse. The solution of such a matrix equation can be computationally expensive, i.e., requiring prohibitively large memory and/or long CPU runtime, in unstructured meshes, as system matrix $\mathbf{A}$ is, in general, not diagonal. If a numerical method for solving PDEs can be made matrix-free, i.e., free of a matrix solution, then much larger problems can be solved using the same computational resources.

1.2 Recent Progress in Multiscale and Multiphysics Analysis

In an FDTD subgridding method, the fields at the interface between base grid and subgrid are typically estimated through certain interpolation scheme. Such interpolation can ruin the positive semi-definite property of the original FDTD system, thus causing instability. Since the smallest space step comes from the fine grid part, the allowed largest time step will be restricted by such fine feature, which makes the whole simulation take a long time to finish due to the fact that this time step need to be used in both base grid and subgrid to guarantee stability. Meanwhile, the space interpolation between base grid and subgrid may result in a worse solution accuracy. Especially for inhomogeneous materials, simple linear interpolation will suffer from

accuracy issue since varying material property changes the fields distribution and linear interpolation is not accurate enough.

In literature, extensive work has been done to tackle the FDTD subgridding problem. In [5], a variable step size method (VSSM) was developed, providing a direct interpolation scheme to update fields in both base grid and subgrid. A mesh refinement algorithm (MRA) was presented in [6], requiring less computational memory and time. Accurate interpolation techniques can always be done to obtain the field unknowns at the interface between a base grid and a subgrid. However, instability especially late-time instability has been observed in many of the subgridding algorithms. Various methods have been proposed to fix the instability issue [7–10]. In [11–13], enforcing reciprocity of the fields has been proposed to ensure stability, but accuracy is compromised. In [14], the subgrid was arranged in a special way in order to ensure the resultant numerical system to be symmetric.

Recently, in [15], an accurate unsymmetrical FDTD subgridding method is developed, whose stability is also guaranteed by a new time marching scheme. However, the subgridding is only achieved in space not in time. A series expansion is also involved in the time marching, which makes the computational cost higher than a traditional explicit time marching. In [16], a systematic approach is developed to make an FDTD subgridding algorithm symmetric positive semi-definite regardless of the grid ratio and the grid arrangement, while retaining the FDTD's second-order accuracy. As a result, the resultant time marching is ensured to be stable and accurate. In [16], a 3-D subgridding operator is demonstrated with a capability of local time stepping. However, the solution of the base grid unknowns is decoupled from that of the subgrid to make the resulting system matrix SPD. Such a scheme suffers from inaccuracy when subgrid regions involve strong inhomogeneity. The 2-D subgridding operator in [15] can handle inhomogeneous problems accurately, however, it is only achieved in space, not in time. As a result, a local time stepping is not permitted, and the time step in the base grid is restricted by that in the subgrid, and vice versa.

For multiphysics analysis, a rigorous and efficient multiphysics simulation is required to understand the entire physical process happening in an advanced IC design. However, existing methods all require a matrix solution, which limits their use in large-scale design and analysis. The FDTD method has its merit in being simple and free of a system matrix solution. However, it has been difficult to extend the FDTD to arbitrary unstructured meshes [17]. Nonorthogonal FDTD methods [18–22] generally require a dual mesh, which is not straightforward to construct for a primary mesh that must capture arbitrarily shaped material discontinuities in 3-D settings. Other different attempts have been developed to generalize the FDTD to other grid types such as the curvilinear FDTD [23–25], the discrete integral equation (DSI) methods [26], and generalized Yee (GY) methods [27–30]. Unfortunately, as mentioned in [31], those methods and similar methods can sometimes suffer from late time instabilities. Interpolations and projections are often employed in these methods, and stability and accuracy may not be simultaneously guaranteed. Recently, a matrix-free time-domain method (MFTD) has been developed for solving Maxwell's PDE equations in unstructured meshes [32–34]. This method has a naturally diagonal system matrix independent of the element shape used for discretization, and hence the need for numerically finding the matrix solution is completely eliminated. Despite the success of solving Maxwell's equations, however, a matrix-free time-domain method has not been developed for thermal analysis as well as electrical-thermal co-design and analysis.

In existing methods for thermal simulation and electrical-thermal co-simulation [35–42], based on either static or full-wave Maxwell's equations, a matrix solution is required when dealing with unstructured meshes. These meshes are often necessary in discretizing irregularly shaped geometries and materials. They also help greatly reduce the number of unknowns as compared to a grid-based discretization. Since the material property changes with time due to thermal effects, the system matrix resulting from the discretization of Maxwell's equations is time dependent in an electrical-thermal co-simulation. As a result, at each time instant when the

matrix changes, one has to re-factorize or solve the matrix, which is time consuming especially in analyzing large-scale problems. Although the MFTD has a flexible framework, it cannot be directly applied to perform a thermal analysis since it is formulated for vectors while the thermal diffusion equation is a scalar equation. In addition, the co-simulation of the coupled Maxwell's and thermal equations results in a nonlinear system of equations, the stability of which has not been investigated in existing methods.

Another challenge is how to simulate the unsymmetrical systems in a stable way. For subgridding problems, it has been shown in [15] that the root cause of the instability is the unsymmetrical system matrix resulting from the various subgridding schemes. And the underlying discretized curl-curl operator of the MFTD method is unsymmetrical as well. For an unsymmetrical matrix, it has been proven that a traditional leap-frog or central-difference based explicit time marching is absolutely unstable [32, 43]. This is because an unsymmetrical matrix can support complex eigenvalues, when these eigenvalues exist, no time step can be found to make the explicit time marching stable. To overcome the stability problem while retaining the advantage of a diagonal mass matrix, in [32], a backward difference scheme is employed for time marching. This results in a system matrix to solve, which is an implicit scheme. Although the inverse of the system matrix is made explicit in [32], thus avoiding a matrix solution, the computation of a k-term series expansion is still required, which is equivalent to k sparse matrix-vector multiplications. In contrast, a truly explicit scheme only requires one matrix-vector multiplication.

The stability and accuracy of the aforementioned methods have always been two competing factors. To explain, in order to ensure the accuracy of these methods, the resultant system matrix, in general, cannot be made symmetric. As a result, there is no theoretical guarantee of the stability of the resultant time-domain simulation. If the imaginary part of the eigenvalues of the unsymmetrical system matrix are negligible as compared to the real part such as a subgridding with few interface unknowns, or nonuniform gridding in small local areas, an explicit time marching may

manage to be stable within a certain time window. However, this is not the case when the system matrix is highly unsymmetrical. For example, in the MFTD, we observe an immediate divergence after starting the time marching if a traditional explicit scheme is used. On the other hand, one can construct a symmetric numerical system to guarantee stability, however, accuracy and flexibility are often sacrificed in the space discretization of the curl operators. If we can find a systematic way to handle complex-valued eigenvalues, and explicitly simulate an unsymmetrical numerical system with guaranteed stability, then not only the MFTD method can be made truly explicit, but also other unsymmetrical time-domain methods can be marched on in time explicitly without the need for concerning about their stability.

1.3 Contribution of This Work

In Chap. 2, we propose a systematic approach to develop SPD FDTD subgridding operators in both space and time regardless of the grid settings, and accurate for analyzing both 2- and 3-D inhomogeneous problems. This approach and resultant 2- and 3-D subgridding operators successfully remove the problems encountered in existing SPD FDTD algorithms. The subgridding is achieved in both space and time, and meanwhile it is accurate for solving inhomogeneous problems. We also provide an explanation of the proposed operators in the original differencing equation based FDTD, and show how to implement it in the original FDTD without using matrix operators. This explanation provides many new insights of this work. Extensive numerical experiments have been carried out, which demonstrate the accuracy, efficiency and stability of the proposed work. It is also shown to outperform state-of-the-art subgridding algorithms like [15, 16].

To further improve the efficiency, we develop a technique to make the aforementioned FDTD subgridding algorithm unconditionally stable. To do this, in Chap. 3, we first discuss how to deal with unstable modes in a conventional FDTD algorithm in an efficient way. In existing explicit and unconditionally stable time-domain

methods, the root cause of instability has been identified to be the eigenmodes of the governing system matrix whose eigenvalues are too high to be accurately simulated by a given time step. The unstable modes have been found numerically, and subsequently deducted from the system matrix to permit an unconditionally stable time marching, but such a numerical method may cost a very long time for solving large scale problems. In our work, we show that such unstable modes as well as stable modes can be found for the FDTD method in an analytical way, which can save a lot of time compared to the numerical approach for finding unstable modes.

After finding the analytical unstable modes for the conventional FDTD algorithm, we proceed to make the explicit and symmetric positive semi-definite FDTD subgridding algorithm unconditionally stable in Chap. 4. The unstable modes are identified analytically using the method developed in Chap. 3, and removed from the numerical system based on prescribed accuracy. The resultant explicit subgridding algorithm not only permits a local time stepping in the base grid and the subgrid, but also is unconditionally stable regardless of the space step in the fine subgrid.

To address the multiphysics simulation challenge and the shortcoming of existing methods, we develop a matrix-free algorithm for solving full-wave Maxwell's equations and the thermal diffusion equation simultaneously in Chap. 5. The proposed new algorithm is made naturally free of matrix solutions. Hence, it has a potential of being much more efficient in time and memory than solvers requiring solving matrices. The matrix-free property of the proposed method is achieved regardless of the element shape used for discretization, thus suitable for both regular grid-based discretizations and unstructured meshes. Although the matrix-free time-domain method has a flexible framework, it cannot be directly applied to perform a thermal analysis since it works on vectors while the thermal diffusion equation is a scalar equation. In this work, we overcome this difficulty by proposing a vector representation of the scalar-based temperature, and subsequently transforming the thermal diffusion equation into two equivalent vector equations. Thus, the thermal diffusion equation can be solved in a matrix-free manner with the same ease as Maxwell's equations. We

then develop a matrix-free algorithm for solving full-wave Maxwell's equations and the thermal diffusion equation simultaneously in time domain. The stability of the coupled nonlinear system of equations is also analyzed in detail, and found to be ensured with a correct choice of time step in explicit time marching.

The counterpart of the matrix-free time-domain method in frequency domain is also developed for fast frequency-domain analysis in Chap. 6. This method, having a diagonal mass matrix, is developed for analyzing general electromagnetic problems in arbitrary unstructured meshes in frequency-domain. The spectral radius of the resultant system matrix is further made less than 1, by deducting the modes outside of the unit circle from the system matrix. The condition number is also made controllable by choosing which set of modes to remove. The property of the diagonal mass matrix is utilized to efficiently find the modes to be deducted, and also compute a fast iterative solution that converges in a small number of iterations.

The unsymmetrical numerical systems are commonly encountered in multiscale and multiphysics problems, and for those systems, a traditional explicit time marching is absolutely unstable because an unsymmetrical matrix can have complex-valued eigenvalues, and no time step can be found to make its explicit time marching stable. In Chap. 7, we overcome this barrier and successfully develop a generic truly explicit time marching method and theoretically prove that this scheme is guaranteed to be stable. With this approach, we are able to make the MFTD truly explicit, and the advantage of the diagonal mass matrix is accentuated. Unsymmetric FDTD subgridding method, having no theoretical guaranteed of the stability of a traditional central-difference based time marching, can be stably simulated by using the proposed method as well. As a result, we eliminate the need for a backward-difference-based implicit scheme to support unsymmetrical system matrix, and thereby the series expansion required for obtaining an explicit inverse of the system matrix, greatly improving the computational efficiency of the simulation without compromising its accuracy.

2. SYMMETRIC POSITIVE SEMI-DEFINITE FDTD SUBGRIDDING ALGORITHMS IN BOTH SPACE AND TIME FOR ACCURATE ANALYSIS OF INHOMOGENEOUS PROBLEMS

2.1 Introduction

Compared with the conventional finite-difference time-domain (FDTD) method, FDTD subgridding is an effective method to locally refine a grid for solving multiscale problems. However, such a scheme generally results in an unsymmetrical numerical system to preserve accuracy for arbitrary subgrid settings. The resultant time-domain simulation cannot be guaranteed in stability using conventional schemes. Symmetric positive semi-definite (SPD) subgridding algorithms have been developed to address the stability problem. However, existing SPD subgridding is generally achieved in space, not in time; meanwhile the accuracy is also limited when handling general inhomogeneous problems.

In this chapter, we develop a systematic approach to derive SPD FDTD subgridding operators in both space and time for analyzing general inhomogeneous problems in an accurate fashion. The operators are symmetric positive semi-definite by construction, which is also made truly independent of the grid ratio. The resultant explicit time marching is guaranteed to be stable because such subgridding operators have only nonnegative real eigenvalues. Furthermore, the use of a time step local to the base grid and the subgrid is permitted without sacrificing stability and accuracy. Moreover, the algorithm takes the subgrid information into account to accurately analyze inhomogeneous problems. In addition, we provide an interpretation of the proposed subgridding operators and show how to implement them in the original difference equation based FDTD framework. Surprisingly as it has not been attempted

before, to obtain the interface unknown in the base grid using the time step local to the base grid, only a partial curl operation is performed using the field inside the base grid. This yields one component of the interface unknown. The other component is obtained from the subgrid using the time step local to the subgrid. Such a splitting of the interface field allows for a subgridding in time in addition to space, whose stability is also theoretically proved. Extensive numerical experiments involving both 2- and 3-D subgrids with various grid ratios have demonstrated the stability, accuracy and efficiency of the proposed new SPD subgridding algorithms.

Part of the contents of this chapter has been extracted and revised from the following manuscript: Kaiyuan Zeng and Dan Jiao, "Symmetric Positive Semi-Definite FDTD Subgridding Algorithms in Both Space and Time for Accurate Analysis of Inhomogeneous Problems," submitted to the IEEE Transactions on Antennas and Propagation.

2.2 Preliminaries

First, we provide a brief review of the patch-based single-grid FDTD formulation, which is developed in [44]. It is used in this work to facilitate the development of a generic SPD subgridding algorithm, as this formulation reveals clearly how the equations in different regions are assembled in an FDTD to build a global system of equations.

The formulation is valid for both 2- and 3-D grids. Let $\{e\}$ be a global electric field unknown vector of length N_e, and $\{h\}$ being a global magnetic field unknown vector of length N_h. The FDTD can be written into the following form:

$$\mathbf{S}_e\{e\} = -\mathbf{D}_\mu\{\dot{h}\}, \tag{2.1}$$

$$\mathbf{S}_h\{h\} = \mathbf{D}_\epsilon\{\dot{e}\} + \mathbf{D}_\sigma\{e\} + \{j\}, \tag{2.2}$$

where a dot above a letter denotes the first-order time derivative, $\{j\}$ represents a current source vector, and $\mathbf{D}_\mu$, $\mathbf{D}_\sigma$ and $\mathbf{D}_\epsilon$ are diagonal matrices of permeability, conductivity, and permittivity respectively.

Based on the patch-based single-grid formulation, each row of $\mathbf{S}_e$ in (2.1) corresponds to one patch in the grid, and when multiplied by $\{e\}$, it produces the magnetic field located at the patch center and normal to the patch. Take the i-th row of $\mathbf{S}_e$ as an example, it can be written as

$$\mathbf{S}_e^{(i)} = \{-\frac{1}{L_i}, \frac{1}{L_i}, \frac{1}{W_i}, -\frac{1}{W_i}\} \oplus zeros(1, N_e), \tag{2.3}$$

which has only four nonzero elements, and L_i and W_i are the two side lengths of the i-th patch. A reference normal direction is defined for every patch, which is also **H**'s reference direction on the patch. Using the right hand rule, with the right thumb pointing to the reference normal direction, if the electric field edge's direction is along the direction encircling the normal direction, then a plus sign is used; otherwise, a negative sign appears in (2.3). The $\oplus$ denotes an extended addition by adding the four nonzero elements upon a zero vector of length N_e, based on the global indexes of the four electric field unknowns on the patch. Similarly, for the i-th patch, we generate a column vector

$$\mathbf{S}_h^{(i)} = \{-\frac{1}{L_i}, \frac{1}{L_i}, \frac{1}{W_i}, -\frac{1}{W_i}\}^T \oplus zeros(N_e, 1), \tag{2.4}$$

which is nothing but the transpose of (2.3), thus

$$\mathbf{S}_h = \mathbf{S}_e^T. \tag{2.5}$$

As can be seen, a column i of $\mathbf{S}_h$ has also at most four nonzero entries, located at the rows corresponding to the four electric fields of patch i.

Eliminating $\{h\}$ from (2.1) and (2.2), we obtain

$$\mathbf{D}_\epsilon\{\ddot{e}\} + \mathbf{D}_\sigma\{\dot{e}\} + \mathbf{S}\{e\} = -\{\dot{j}\}, \tag{2.6}$$

where $\mathbf{S}$ can be represented as

$$\mathbf{S} = \mathbf{S}_h\mathbf{D}_{\mu^{-1}}\mathbf{S}_e = \sum_{i=1}^{N_h} \mu_i^{-1}(\mathbf{S}_h^{(i)})_{N_e\times 1}(\mathbf{S}_e^{(i)})_{1\times N_e}, \tag{2.7}$$

which is a sum of the rank-1 matrix $\mathbf{S}_h^{(i)}\mathbf{S}_e^{(i)}$ over all the patches.

A leap-frog-based time marching of (2.1) and (2.2) is equivalent to a central-difference-based explicit time marching of (2.6), which can be readily proved. The stability of the resulting explicit marching is guaranteed if the following condition is satisfied:

$$\Delta t \leq \frac{2}{\sqrt{\lambda_{max}}}, \tag{2.8}$$

where λ_{max} stands for the largest eigenvalue of $\mathbf{D}_\epsilon^{-1}\mathbf{S}$ [45]. In the conventional FDTD for a uniform grid, $\mathbf{S}$ is symmetric positive semi-definite as evident from (2.7) and (2.5). Since $\mathbf{D}_\epsilon$ is also symmetric positive definite, the eigenvalues of $\mathbf{D}_\epsilon^{-1}\mathbf{S}$ are nonnegative real. As a result, a real valued time step can always be found to satisfy (2.8). However, in an FDTD subgridding scheme, due to the mismatch between the base grid and the subgrid. Various unsymmetrical interpolations are used to obtain the unknown fields at the subgrid interface, making the resulting rank-1 matrix of each patch not symmetric. As a result, $\mathbf{S}$ is unsymmetrical in general, which involves complex or negative eigenvalues in nature. When such eigenvalues exist, a traditional explicit time marching is absolutely unstable, which has been proved in [32], and that is why many FDTD subgridding algorithms cannot ensure stability.

2.3 Systematic Approach for Developing SPD Subgridding Algorithms in Both Space and Time

In this section, we present a systematic approach for developing SPD FDTD subgridding algorithms in both space and time, which is SPD by construction. This approach is also algebraic and generic, suitable for general subgridding settings in both 2- and 3-D problems.

From (2.6) and (2.7), we can analyze how the equations in different domains are assembled in the FDTD to simulate the entire problem. Consider two domains, (2.6) can be rewritten as

$$\mathbf{D}_\epsilon\{\ddot{e}\} + \mathbf{D}_\sigma\{\dot{e}\} = -\left[\mathbf{S}_{h,1}\mathbf{D}_{\mu_1^{-1}}\mathbf{S}_{e,1} + \mathbf{S}_{h,2}\mathbf{D}_{\mu_2^{-1}}\mathbf{S}_{e,2}\right]\{e\}, \tag{2.9}$$

where $\mathbf{S}_{h,1(2)}$ has all the column vectors generated from the patches in domain 1 (2), and $\mathbf{S}_{e,1(2)}$ comprises all the row vectors from the patches in domain 1 (2). Here, the source term is omitted to focus on the assembling mechanism in the FDTD. As can been seen, the total $\mathbf{S}$ is an addition of each domain's $\mathbf{S}$, which is the same as the assembling procedure in a finite-element method (FEM). Each patch's $\mathbf{S}$ is assembled to obtain a global $\mathbf{S}$ based on the index of a global unknown vector. However, the $\mathbf{D}_\epsilon$ and $\mathbf{D}_\sigma$ are not added up from each domain's contribution. They are diagonal matrices, whose entries are the permittivity or conductivity at the corresponding e's location. Shall they be assembled from each patch's contribution like that in a finite-element method, then the diagonal entry would be a multiple of the permittivity or conductivity. This unique assembling procedure in the FDTD, originated from the use of a dual grid, also renders its SPD subgridding scheme more difficult to develop, as compared to the FEM in handling a nonconformal mesh.

Based on (2.9), we can express the electric filed unknown as the addition of two contributions: one is from domain 1, expressed by the first term of the right hand side of (2.9); the other is from domain 2, represented by the second term. Hence, we can rewrite (2.9) as a two-row system:

$$\mathbf{D}_\epsilon\{\ddot{e}\}_1 + \mathbf{D}_\sigma\{\dot{e}\}_1 = -\mathbf{S}_{h,1}\mathbf{D}_{\mu_1^{-1}}\mathbf{S}_{e,1}\{e\} \tag{2.10}$$

$$\mathbf{D}_\epsilon\{\ddot{e}\}_2 + \mathbf{D}_\sigma\{\dot{e}\}_2 = -\mathbf{S}_{h,2}\mathbf{D}_{\mu_2^{-1}}\mathbf{S}_{e,2}\{e\}, \tag{2.11}$$

with

$$\{e\} = \{e\}_1 + \{e\}_2, \tag{2.12}$$

which stitches the two domains together. Neither $\{e\}_1$ nor $\{e\}_2$ provides a complete solution of $\{e\}$. This is because for an interface e unknown between domain 1 and domain 2, (2.10) yields the curl of $\mathbf{H}$ from domain 1 patches, and (2.11) generates the curl of $\mathbf{H}$ from domain 2 patches, and the addition shown in (2.12) is required to complete the whole curl of $\mathbf{H}$ operation to produce the electric field on the interface.

If the two regions are a base grid, and a subgrid, respectively, the same principle applies to add the equations from each region. In what follows, we denote the base-

grid region by c, as a $c$oarse mesh is often used in the base grid; and the subgrid region by f, standing for a $f$iner mesh in the subgrid. As illustrated in Fig. 2.1 , we use $\{e_{fb}\}$, $\{e_{fi}\}$, $\{e_{cb}\}$, and $\{e_{ci}\}$ to denote the electric field unknowns (edges) on the subgrid boundary, inside the subgrid, on the base grid boundary that overlaps with the subgrid boundary, and inside the base grid respectively. Obviously, one $\{e_{cb}\}$ overlaps with multiple $\{e_{fb}\}$.

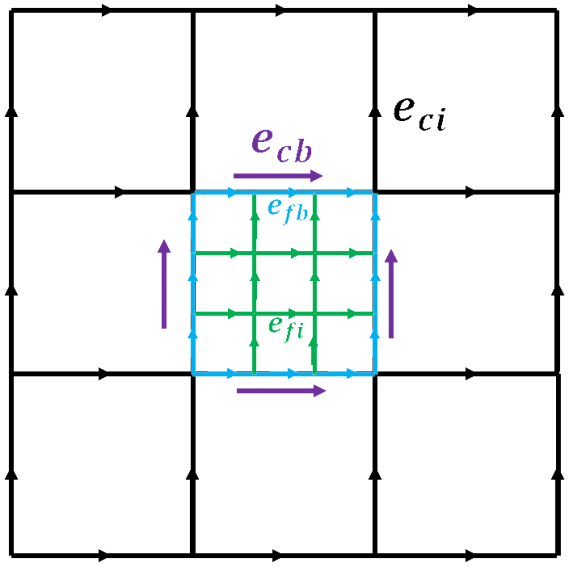


Fig. 2.1. Illustration of a subgrid embedded in a base grid, and different kinds of unknowns.

To maintain the field tangential continuity at the interface between the base grid and the subgrid, we should use only one set of **E** unknowns between $\{e_{cb}\}$ and $\{e_{fb}\}$. If we use $\{e_{cb}\}$ as the set, the global unknown **E** vector is composed of

$$\{e\}_{N_e\times 1} = \{e_{ci}, e_{fi}, e_{cb}\}^T, \tag{2.13}$$

with a total number of **E** unknowns being

$$N_e = \#e_{ci} + \#e_{fi} + \#e_{cb}. \tag{2.14}$$

If we use $\{e_{fb}\}$, we have the following $\{e_{sub}\}$ being the global **E** vector:

$$\{e_{sub}\}_{N_{e,sub}\times 1} = \{e_{ci}, e_{fi}, e_{fb}\}^T, \tag{2.15}$$

whose length is

$$N_{e,sub} = \#e_{ci} + \#e_{fi} + \#e_{fb}. \tag{2.16}$$

If the vector in (2.13) is used as a global unknown vector, the e's solution contributed from the base grid, defined as $\{e\}_p$, can be expressed as

$$\mathbf{D}_\epsilon\{\ddot{e}\}_p + \mathbf{D}_\sigma\{\dot{e}\}_p = -\mathbf{S}_{h,c}\mathbf{D}_{\mu_c^{-1}}\mathbf{S}_{e,c}\{e\}, \tag{2.17}$$

where $\mathbf{S}_{e,c}$ is obtained from all of the patches in the base grid, and $\mathbf{S}_{h,c}$ is the transpose of $\mathbf{S}_{e,c}$. Similarly, the e's solution contributed from the subgrid, defined as $\{e\}_m$, can be written into

$$\mathbf{D}_{\epsilon,sub}\{\ddot{e}_{sub}\}_m + \mathbf{D}_{\sigma,sub}\{\dot{e_{sub}}\}_m = -\mathbf{S}_{h,f}\mathbf{D}_{\mu_f^{-1}}\mathbf{S}_{e,f}\{e_{sub}\}, \tag{2.18}$$

where $\mathbf{S}_{e,f}$ is obtained from all of the patches in the subgrid, and $\mathbf{S}_{h,f}$ is its transpose.

To combine (2.17) and (2.18), we can represent $\{e_{sub}\}$ in terms of $\{e\}$. Since only the subgrid boundary unknowns need such a transformation, we can write

$$\{e_{fb}\} = \mathbf{P}_{fc}\{e_{cb}\}, \tag{2.19}$$

where $\mathbf{P}_{fc}$ is used to interpolate $\{e_{fb}\}$ from $\{e_{cb}\}$. Subsequently, we can write

$$\{e_{sub}\}_{N_{e,sub}\times 1} = \mathbf{P}\{e\}_{N_e\times 1}, \tag{2.20}$$

and $\mathbf{P}$, whose size is $N_{e,sub} \times N_e$, has the following form

$$\mathbf{P} = \begin{bmatrix} (\mathbf{I})_{\#e_{ci}\times\#e_{ci}} & 0 & 0 \\ 0 & (\mathbf{I})_{\#e_{fi}\times\#e_{fi}} & 0 \\ 0 & 0 & (\mathbf{P}_{fc})_{\#e_{fb}\times\#e_{cb}} \end{bmatrix}, \tag{2.21}$$

in which $\mathbf{I}$ denotes an identity matrix, and the subscripts denote matrix dimensions.

Using the above transformation, (2.18) can be rewritten as

$$\mathbf{D}_{\epsilon,sub}\mathbf{P}\{\ddot{e}\}_m + \mathbf{D}_{\sigma,sub}\mathbf{P}\{\dot{e}\}_m = -\mathbf{S}_{h,f}\mathbf{D}_{\mu_f^{-1}}\mathbf{S}_{e,f}\mathbf{P}\{e\}. \tag{2.22}$$

To obtain a symmetric system of equations, we multiply both sides of (2.22) by $\mathbf{P}^T$, obtaining

$$\mathbf{P}^T\mathbf{D}_{\epsilon,sub}\mathbf{P}\{\ddot{e}\}_m + \mathbf{P}^T\mathbf{D}_{\sigma,sub}\mathbf{P}\{\dot{e}\}_m = -\mathbf{P}^T\mathbf{S}_{h,f}\mathbf{D}_{\mu_f^{-1}}\mathbf{S}_{e,f}\mathbf{P}\{e\}. \tag{2.23}$$

Combining (2.23) and (2.17) with the following

$$\{e\} = \{e\}_p + \{e\}_m, \tag{2.24}$$

we obtain the global system of equations for solving e when subgrids are present.

In [16], the 2-D subgridding algorithm does not permit a local time stepping, i.e., allowing for the use of a time step in the base grid local to the base grid regardless of the subgrid. As a result, one has to use a smaller time step restricted by the finer space step in the subgrid for the time marching in the base grid. To permit a local time stepping, in this work, we propose to solve the following $\{e\}_p$ and $\{e\}_m$ systems of equations instead of $\{e\}$ directly:

$$\begin{bmatrix} \mathbf{D}_\epsilon & 0 \\ 0 & \mathbf{P}^T\mathbf{D}_{\epsilon,sub}\mathbf{P} \end{bmatrix} \begin{bmatrix} \ddot{e}_p \\ \ddot{e}_m \end{bmatrix} + \begin{bmatrix} \mathbf{D}_\sigma & 0 \\ 0 & \mathbf{P}^T\mathbf{D}_{\sigma,sub}\mathbf{P} \end{bmatrix} \begin{bmatrix} \dot{e}_p \\ \dot{e}_m \end{bmatrix} + \begin{bmatrix} \mathbf{S}_{cc} & \mathbf{S}_{cc} \\ \mathbf{P}^T\mathbf{S}_{ff}\mathbf{P} & \mathbf{P}^T\mathbf{S}_{ff}\mathbf{P} \end{bmatrix} \begin{bmatrix} e_p \\ e_m \end{bmatrix} = b, \tag{2.25}$$

where

$$\begin{aligned} \mathbf{S}_{cc} &= \mathbf{S}_{h,c}\mathbf{D}_{\mu_c^{-1}}\mathbf{S}_{e,c} \\ \mathbf{S}_{ff} &= \mathbf{S}_{h,f}\mathbf{D}_{\mu_f^{-1}}\mathbf{S}_{e,f}. \end{aligned} \tag{2.26}$$

If the new system (2.25) can be stably simulated, we can prove the time step for simulating $\{e\}_p$ is determined by the base grid, instead of the subgrid, as the following. Rewriting (2.25) in frequency domain, we have

$$\begin{bmatrix} -\omega^2\mathbf{D}_\epsilon + \mathbf{S}_{cc} & \mathbf{S}_{cc} \\ \mathbf{P}^T\mathbf{S}_{ff}\mathbf{P} & -\omega^2\mathbf{D}_\epsilon + \mathbf{P}^T\mathbf{S}_{ff}\mathbf{P} \end{bmatrix} \begin{bmatrix} e_p \\ e_m \end{bmatrix} = b(\omega). \tag{2.27}$$

Eliminating $\{e\}_m$, the $\{e\}_p$ satisfies

$$[-\omega^2\mathbf{D}_\epsilon + \mathbf{S}_{cc}(\mathbf{I} - \mathbf{A}_{ff}^{-1}\mathbf{S}_{ff})]e_p = b, \tag{2.28}$$

where $\mathbf{I}$ is an identity matrix and $\mathbf{A}_{ff} = -\omega^2\mathbf{D}_\epsilon + \mathbf{P}^T\mathbf{S}_{ff}\mathbf{P}$. The $\mathbf{S}$ part in (2.28) is $\mathbf{S}_{cc}$ right multiplied by another matrix, and hence the field solution e_p is still spanned

in the space of $\mathbf{S}_{cc}$'s eigenvectors (from the base grid), which guarantees that the base grid time step is not restricted by the subgrid region. Similarly it can be proven that the subgrid can be simulated using the time step allowed for stably simulating the $\mathbf{S}_{ff}$ modes, thus local to the subgrid.

In our conference paper [46], we prove that the $\mathbf{S}$-related matrix in (2.25) is positive semi-definite. This is because the eigenvalues λ and the eigenvectors x of the $\mathbf{S}$-related matrix satisfy

$$\mathbf{S}_{new}x = \begin{bmatrix} \mathbf{S}_{cc} & \mathbf{S}_{cc} \\ \mathbf{P}^T\mathbf{S}_{ff}\mathbf{P} & \mathbf{P}^T\mathbf{S}_{ff}\mathbf{P} \end{bmatrix} \begin{Bmatrix} x1 \\ x2 \end{Bmatrix} = \lambda \begin{Bmatrix} x1 \\ x2 \end{Bmatrix}, \tag{2.29}$$

hence, $\lambda(x1 + x2) = (\mathbf{S}_{cc} + \mathbf{P}^T\mathbf{S}_{ff}\mathbf{P})(x1 + x2)$. Since both $\mathbf{S}_{cc}$ and $\mathbf{P}^T\mathbf{S}_{ff}\mathbf{P}$ are symmetric positive semi-definite, the $\mathbf{S}_{new}$'s eigenvalues are also non-negative real. However, this proof does not lead to the proof that the $\mathbf{D}_{new}^{-1}\mathbf{S}_{new}$ also has non-negative real eigenvalues, where $\mathbf{D}_{new}$ is the block diagonal matrix in front of the second and the first order time derivative of $\{e\}$ in (2.25). Depending on the choice of $\mathbf{P}$ and also the inhomogeneity of the problem being studied, the $\mathbf{D}_{new}^{-1}\mathbf{S}_{new}$ may have complex eigenvalues and the resulting explicit time marching would not be stable.

To systematically develop a subgridding algorithm which is SPD by construction, we carried out the following analysis. First, we should realize that since e_{fb} and e_{cb} overlap, they share the same material parameter, and hence

$$\mathbf{P}^T\mathbf{D}_{\epsilon,sub}\mathbf{P} = \mathbf{P}^T\mathbf{P}\mathbf{D}_{\epsilon} \tag{2.30}$$

$$\mathbf{P}^T\mathbf{D}_{\sigma,sub}\mathbf{P} = \mathbf{P}^T\mathbf{P}\mathbf{D}_{\sigma}. \tag{2.31}$$

As a result, (2.23) can be rewritten as

$$\mathbf{D}_{\epsilon}\{\ddot{e}\}_m + \mathbf{D}_{\sigma}\{\dot{e}\}_m = -(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\mathbf{S}_{h,f}\mathbf{D}_{\mu_f^{-1}}\mathbf{S}_{e,f}\mathbf{P}\{e\}. \tag{2.32}$$

Adding the above upon (2.17), the whole system of equations for solving $\{e\}$ can be written as

$$\mathbf{D}_{\epsilon}\{\ddot{e}\} + \mathbf{D}_{\sigma}\{\dot{e}\} = -\left[\mathbf{S}_{cc} + (\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\mathbf{S}_{ff}\mathbf{P}\right]\{e\}, \tag{2.33}$$

which can be written in short as

$$\mathbf{D}_\epsilon\{\ddot{e}\} + \mathbf{D}_\sigma\{\dot{e}\} = -\mathbf{S}_{total}\{e\}, \tag{2.34}$$

where

$$\mathbf{S}_{total} = \mathbf{S}_{cc} + (\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\mathbf{S}_{ff}\mathbf{P}. \tag{2.35}$$

The stability of (2.33), which is also that of (2.25), is governed by the property of $\mathbf{D}_\epsilon^{-1}\mathbf{S}_{total}$. Since $\mathbf{D}_\epsilon$ is symmetric and positive definite, the stability is governed by the property of $\mathbf{S}_{total}$. If $\mathbf{S}_{total}$ is symmetric and positive semi-definite, then $\mathbf{D}_\epsilon$ and $\mathbf{S}_{total}$ form a SPD eigenvalue problem, whose eigenvalues are known to be nonnegative real. Unfortunately, the $\mathbf{S}_{total}$ shown in (2.35) is not symmetric for general $\mathbf{P}$. However, if $\mathbf{P}^{-1} = \mathbf{P}^T$, then obviously, $\mathbf{S}_{total}$ would become symmetric. Since $\mathbf{P}$ is not a square matrix because the subgrid does not match the base grid, $\mathbf{P}^{-1}$ does not exist. However, if we can make $\mathbf{P}^T\mathbf{P}$ be an identity matrix, or an identity matrix scaled by a positive coefficient, then (2.35) is also symmetric and positive semi-definite. Let $\mathbf{P}^T\mathbf{P} = d\mathbf{I}$, we have

$$\mathbf{S}_{total} = \mathbf{S}_{cc} + d^{-1}\mathbf{P}^T\mathbf{S}_{ff}\mathbf{P}, \tag{2.36}$$

which is symmetric and positive semi-definite.

The aforementioned provides a theoretical framework to develop an FDTD subgridding algorithm that is SPD by construction. The only thing left for consideration is accuracy, i.e., we have to make sure the field solution obtained from the resulting system is accurate. Based on the ideas presented in this section, we successfully develop 2- and 3-D SPD subgridding algorithms in both space and time, which are also accurate, the details of which are presented in the following section.

2.4 Symmetric Positive Semi-Definite 2-D Subgridding Algorithm in Space and Time

In this section, we first present an accurate 2-D SPD subgridding operator in space and time, and then provide an interpretation of the operator in the original FDTD framework.

2.4.1 2-D Subgridding Operator

For 2-D cases, let the grid ratio be n, then there are n subgrid boundary unknowns e_{fb} overlapping with the e_{cb} from the base grid, as illustrated in Fig. 2.1.

We need to find a good choice of $\mathbf{P}$ that can result in a system matrix shown in (2.36). The $\mathbf{P}$ in [16] satisfies this requirement after certain modifications. Specifically, the following $\mathbf{P}_{fc}$ is employed

$$\mathbf{P}_{fc} = [1 \quad 1 \quad 1 \quad ... \quad 1]^T_{n\times 1}. \tag{2.37}$$

For such a choice of $\mathbf{P}_{fc}$, $(\mathbf{P}^T\mathbf{P})^{-1}$ is a diagonal matrix of

$$(\mathbf{P}^T\mathbf{P})^{-1} = \begin{bmatrix} \mathbf{I}_{\#e_{ci}} & 0 & 0 \\ 0 & \mathbf{I}_{\#e_{fi}} & 0 \\ 0 & 0 & n^{-1}\mathbf{I}_{\#e_{cb}} \end{bmatrix}, \tag{2.38}$$

where $\mathbf{I}$ denotes an identity matrix whose size is specified by the subscript, and $\#$ denotes the cardinality of a set.

The aforementioned choice of $\mathbf{P}$ results in a clear geometrical meaning of $(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T$ in (2.33). For an e_{cb} row of equation, the $\mathbf{P}^T$ multiplication is to add the curl of $\mathbf{H}$ from the n subgrid patches that have the n segments of the e_{cb}, and the $(\mathbf{P}^T\mathbf{P})^{-1}$, which is $1/n$, is to take the average of the sum of the curl of $\mathbf{H}$ contributed by the n patches. Thus the $(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T$ operation provides the curl of $\mathbf{H}$ from the subgrid patches that share the e_{cb}.

Although (2.38) is diagonal, it cannot be written as an identity matrix scaled by a constant. However, we can change the diagonal block corresponding to the e_{ci} from $\mathbf{I}$ to $n^{-1}\mathbf{I}$. This would not change (2.33) since the $\mathbf{S}_{ff}$ term is zero in the rows/columns corresponding to e_{ci}. In other words, $\mathbf{S}_{ff}$ does not involve unknowns in the base grid. We can also change the diagonal block corresponding to the e_{fi} from $\mathbf{I}$ to $n^{-1}\mathbf{I}$. This will change the $\mathbf{S}_{ff}$ term. However, we can scale the left hand side matrices corresponding to the e_{fi} in the same way. As a result, the solution would not be changed. After the modifications, we obtain the following system of equations

$$\mathbf{D}\mathbf{D}_{\epsilon}\{\ddot{e}\} + \mathbf{D}\mathbf{D}_{\sigma}\{\dot{e}\} = -\left[\mathbf{S}_{cc} + n^{-1}\mathbf{P}^T\mathbf{S}_{ff}\mathbf{P}\right]\{e\}, \tag{2.39}$$

where $\mathbf{D} = \text{diag}\{\{1\}_{\#e_{ci}}, \{1/n\}_{\#e_{fi}}, \{1\}_{\#e_{cb}}\}$, a diagonal matrix made of the entries shown in the braces. The above system is clearly SPD, whose solution is also the same as (2.33), thus not changed.

Although (2.39) is SPD, we find it is not very accurate. This is because the cell size is different for the subgrid patches and the base grid patches. We need to use an average length to calculate the electric field of e_{cb} from the curl of $\mathbf{H}$. This is not an issue for a conventional FDTD in a uniform grid, because the interface edge is shared by the patches from different grids with the same grid size. However, this is not the case when subgridding exists. To see this point clearly, consider one e_{cb} unknown whose global index is i, its corresponding row of equation in (2.39) can be written as

$$\epsilon_i e_{cb,i} + \sigma_i e_{cb,i} = -\left[\frac{1}{L_c}\dot{h}_{base} - \frac{1}{nL_f}\sum_{j=1}^{n}\dot{h}_{s,j}\right], \tag{2.40}$$

in which ϵ_i and σ_i denote the permittivity, and conductivity respectively at the point of $e_{cb,i}$, L_c denotes the base grid cell size, and L_f is the subgrid one, the h_{base} is the magnetic field at the base grid patch that owns the $e_{cb,i}$, whereas $h_{s,j}$ denotes the j-th subgrid patch with one edge falling onto the $e_{cb,i}$. Since L_c is different from L_f, the above is inaccurate in computing the curl of $\mathbf{H}$ to generate $e_{cb,i}$.

For better accuracy, we should use an average length $L_{2D,ave}$ of the two patches along the direction perpendicular to the e_{cb} to evaluate the curl of $\mathbf{H}$, which is

$$L_{2D,ave} = \frac{L_c + L_f}{2}. \tag{2.41}$$

To utilize the average length, the row entry of $\mathbf{S}_h^{(i)}$ of the patches involving e_{cb} should be changed from the original $1/L_i$ (L_i being L_c or L_f) to $1/L_{2D,ave}$. We hence replace $1/L_i$ by $1/L_{2D,ave}$ for the rows corresponding to the e_{cb} unknowns, obtaining

$$\mathbf{S}_{total} = \mathbf{D}_1\left(\mathbf{S}_{cc} + \frac{L_f}{nL_c}\mathbf{P}^T\mathbf{S}_{ff}\mathbf{P}\right), \tag{2.42}$$

where

$$\mathbf{D}_1 = \text{diag}\{\{1\}_{\#e_{ci}}, \{1\}_{\#e_{fi}}, \{L_c/L_{2D,ave}\}_{\#e_{cb}}\}. \tag{2.43}$$

In this way, (2.40) becomes

$$\epsilon_i e_{cb,i} + \sigma_i e_{cb,i} = -\left[\frac{1}{L_{2D,ave}}\dot{h}_{base} - \frac{1}{nL_{2D,ave}}\sum_{j=1}^{n}\dot{h}_{s,j}\right], \tag{2.44}$$

which is accurate. However, since $\mathbf{S}_{ff}$ also involves e_{fi}, the constant coefficient $\frac{L_f}{nL_c} = \frac{1}{n^2}$ in (2.42) scales the right hand side corresponding to the e_{fi} as well, and hence the equation is changed. Again, we can scale the left hand side of (2.39) in the same way to make the equation correct.

As a result, (2.39) becomes

$$\mathbf{D}_s\mathbf{D}_\epsilon\{\ddot{e}\} + \mathbf{D}_s\mathbf{D}_\sigma\{\dot{e}\} = -\mathbf{S}_{total}\{e\}, \tag{2.45}$$

in which

$$\mathbf{D}_s = \mathrm{diag}\{\{1\}_{\#e_{ci}}, \{1/n^2\}_{\#e_{fi}}, \{1\}_{\#e_{cb}}\}, \tag{2.46}$$

and

$$\mathbf{S}_{total} = \mathbf{D}_1\mathbf{S}_t, \tag{2.47}$$

where

$$\mathbf{S}_t = \mathbf{S}_{cc} + (1/n^2)\mathbf{P}^T\mathbf{S}_{ff}\mathbf{P}. \tag{2.48}$$

The stability of (2.45) is governed by the eigenvalues of $\mathbf{D}_\epsilon^{-1}\mathbf{D}_s^{-1}\mathbf{D}_1\mathbf{S}_t$. Since $\mathbf{D}_\epsilon^{-1}\mathbf{D}_s^{-1}\mathbf{D}_1$ is diagonal and symmetric positive definite, and $\mathbf{S}_t$ is symmetric positive semi-definite, the eigenvalues of $\mathbf{D}_\epsilon^{-1}\mathbf{D}_s^{-1}\mathbf{D}_1\mathbf{S}_t$ are non-negative real. As a result, the resultant subgridding scheme is guaranteed to be stable. Meanwhile, the accuracy is also ensured via the choice of $\mathbf{P}$, and the use of average length across the base grid and the subgrid.

To realize a local time stepping, we transform (2.45) to the following system:

$$\begin{bmatrix} \mathbf{D}_\epsilon & 0 \\ 0 & \mathbf{D}_s\mathbf{D}_\epsilon \end{bmatrix}\begin{bmatrix} \ddot{e}_p \\ \ddot{e}_m \end{bmatrix} + \begin{bmatrix} \mathbf{D}_\sigma & 0 \\ 0 & \mathbf{D}_s\mathbf{D}_\sigma \end{bmatrix}\begin{bmatrix} \dot{e}_p \\ \dot{e}_m \end{bmatrix} + \begin{bmatrix} \mathbf{D}_1\mathbf{S}_{cc} & \mathbf{D}_1\mathbf{S}_{cc} \\ n^{-2}\mathbf{D}_1\mathbf{P}^T\mathbf{S}_{ff}\mathbf{P} & n^{-2}\mathbf{D}_1\mathbf{P}^T\mathbf{S}_{ff}\mathbf{P} \end{bmatrix}\begin{bmatrix} e_p \\ e_m \end{bmatrix} = b. \tag{2.49}$$

An explicit time marching is then performed on the e_p and e_m separately as follows:

$$\mathbf{D}_\epsilon \ddot{e}_p + \mathbf{D}_\sigma \dot{e}_p + \mathbf{D}_1 \mathbf{S}_{cc}\{e\} = b_c \tag{2.50}$$

$$\mathbf{D}_1^{-1}\mathbf{D}_s\mathbf{D}_\epsilon \ddot{e}_m + \mathbf{D}_1^{-1}\mathbf{D}_s\mathbf{D}_\epsilon \dot{e}_m + n^{-2}\mathbf{P}^T\mathbf{S}_{ff}\mathbf{P}\{e\} = \mathbf{D}_1^{-1} b_f \tag{2.51}$$

$$\{e\} = e_p + e_m, \tag{2.52}$$

which allows for a local time stepping without affecting stability.

2.4.2 Interpretation and Implementation in the Original FDTD Difference Equation Based Framework

The formulas provided in the section above may appear abstract since matrix representations are used to derive the algorithm. However, if we do not use matrix operators, one cannot see a system-level picture clearly. Indeed, it is difficult to see how the change of one row of equation affects the stability of the whole numerical system, if one stays in the original framework of FDTD composed of many differencing equations. In contrast, once the SPD operator is developed, we can readily explain it using the language of the original FDTD. Next, we provide an interpretation and show the implementation of the proposed algorithm in the original FDTD. We do so also for both global time stepping, i.e., the same time step (restricted by the finest grid) is used in both base grid and subgrid; and local time stepping, i.e., each grid is stably simulated using its own local time step.

Global time stepping

- For e_{ci} unknowns, which are inside the base grid, they are solved in the same way as in the original FDTD. In other words, each e_{ci} at the current time step is obtained from the curl of $\mathbf{H}$ operation using the two $\mathbf{H}$ fields at previous time step located at the two patches that share the e_{ci}.
- For magnetic field unknowns inside the base grid, they are also solved in the same way as in the original FDTD.

- For e_{cb} unknowns, which reside on the interface between the subgrid and the base grid, we use the average of the **H** fields at the n subgrid patches adjacent to the e_{cb}, and the **H** field at the base grid patch having e_{cb}, to perform a curl of **H** operation. The length averaged from the base grid cell size and subgrid cell size is used for better accuracy.

- The e_{fb} unknowns are obtained from $\mathbf{P}_{fc}e_{cb}$, which is nothing but to set them the same as the e_{cb} where the e_{fb} unknowns reside.

- The e_{fi} and the magnetic field unknowns inside the subgrid are solved in the same way as that in the original FDTD.

As can be seen, only the e_{cb} and e_{fb} are different from the original FDTD in their generation. The time step of the above scheme is restricted by the smallest space step in the grid, which we term a global time step.

Local time stepping

For local time stepping, i.e., subgridding in time, all the unknowns, except for e_{cb}, are generated in the same way as above. For the e_{cb} unknown, we should do the following. At each time step of time marching, from the base grid, we know the **H** field at previous time step at the patch having e_{cb}, we use it to perform only a partial curl of **H** operation, which is $h_{base}/L_{2D,ave}$. In other words, we do not complete the curl of **H** operation using the **H** fields from the subgrid. The partial curl of **H** evaluated from the base grid only provides one component of the e_{cb}, denoted by $e_{cb,p}$, but it allows all unknowns in the base grid to be generated using a large time step. For the subgrid, we use the time step restricted by the subgrid space step, and obtain the other half of the curl of **H** operation by evaluating the **H** fields in every subgrid patch, and then taking the average of the n **H** fields on the n subgrid patches adjacent to e_{cb}. Let such an average field be h_{sub}, we perform $-h_{sub}/L_{2D,ave}$ to obtain the other component of e_{cb}, denoted by $e_{cb,m}$. Then adding the $e_{cb,p}$ and $e_{cb,m}$ makes

the total e_{cb}. Since two different time steps are used, the addition is performed at the time instants of the smaller time step, where $e_{cb,p}$ is interpolated to provide a value at the desired time instants. Then the total e_{cb} is also known for the time instants of the larger time step. The time step ratio is the grid ratio, hence for one time step marching in the base grid, n steps of marching are performed in the subgrid.

2.5 Symmetric Positive Semi-Definite 3-D Subgridding Algorithm in Space and Time

2.5.1 3-D Subgridding Operator

Different from 2-D cases, in addition to $\{e_{ci}\}$, $\{e_{cb}\}$, $\{e_{fi}\}$ unknowns, there are two sets of $\{e_{fb}\}$ unknowns. One set is located along the edges that overlap with $\{e_{cb}\}$, which we denote by $\{e_{fb,e}\}$; and the other set are on the faces of the interface between the base grid and the subgrid, which we denote by $\{e_{fb,f}\}$, as illustrated in Fig. 2.2. This set of $\{e_{fb,f}\}$ unknowns is unique for 3-D cases, which does not appear in 2-D scenarios. Since $\{e_{cb}\}$ is tangential to the subgrid interface, the twelve e_{cb} unknowns on a subgrid interface make a complete set to interpolate both $\{e_{fb,e}\}$ and $\{e_{fb,f}\}$ unknowns. Thus, we have

$$e_{fb} = \begin{bmatrix} e_{fb,f} \\ e_{fb,e} \end{bmatrix} = \mathbf{P}_{fc} e_{cb}, \tag{2.53}$$

where

$$\mathbf{P}_{fc} = \begin{bmatrix} \mathbf{P}_{fc,f} \\ \mathbf{P}_{fc,e} \end{bmatrix}, \tag{2.54}$$

with the upper part used to interpolate $\{e_{fb,f}\}$, and the lower part for interpolating $\{e_{fb,e}\}$. If a linear interpolation is used, for an arbitrary grid ratio n, the i-th column

of $\mathbf{P}_{fc}$ has the following nonzero entries located at the rows corresponding to the $\{e_{fb,f}\}$ and the $\{e_{fb,e}\}$ unknowns interpolated from the i-th e_{cb}

$$\mathbf{P}_{fc,i} = [\underbrace{\{(n-1)/n\}_n, \{(n-2)/n\}_n, ..., \{1/n\}_n}_{e_{fb,f} \text{ on face 1 of } e_{cb} \text{ and parallel to } e_{cb}},$$
$$\underbrace{\{(n-1)/n\}_n, \{(n-2)/n\}_n, ..., \{1/n\}_n}_{e_{fb,f} \text{ on face 2 of } e_{cb} \text{ and parallel to } e_{cb}},$$
$$\underbrace{\{1\}_n}_{e_{fb,e} \text{ that overlaps with } e_{cb}} \quad]^T, \tag{2.55}$$

in which the subscript of each set denotes the number of entries in the set. For example, the last set, $\{1\}_n$ is a set having n ones, each of which corresponds to one fine edge unknown $e_{fb,e}$ located along the e_{cb} edge. This set is, in fact, the $\mathbf{P}_{fc,e}$ in (2.54), which is the same as the $\mathbf{P}_{fc}$ in the 2-D scheme. The other two groups of entries of (2.55) make the part of $\mathbf{P}_{fc,f}$, yielding the $e_{fb,f}$ located on the two faces that share the e_{cb}. The first group from $\{(n-1)/n\}_n$ to $\{1/n\}_n$ corresponds to the $e_{fb,f}$ falling onto the first face and parallel to the e_{cb}, while the second group corresponds to the $e_{fb,f}$ falling onto the second face and parallel to the e_{cb}.

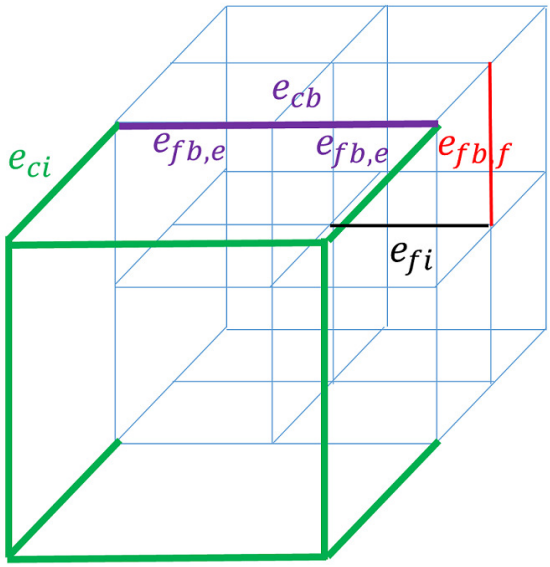


Fig. 2.2. Illustration of a subgrid embedded in a base grid in a 3-D grid.

The $\mathbf{P}$ for 3-D problems is thus (2.21) but with the new 3-D $\mathbf{P}_{fc}$ shown in (2.55). Different from 2-D settings, $(\mathbf{P}^T\mathbf{P})^{-1}$ is not diagonal any more. In [16], we only use $\mathbf{P}_{fc,e}^T$ instead of a complete $\mathbf{P}_{fc}^T$ to develop a 3-D subgridding algorithm. Although the resulting $\mathbf{P}_{fc,e}^T\mathbf{P}_{fc,e}$ is diagonal, the information inside the subgrid cannot be utilized to solve the whole problem. In other words, the base grid solution is decoupled from the subgrid solution, which is inaccurate when strong inhomogeneity exists in the subgrid. Although the subgrid solution still depends on the base grid solution in the 3-D scheme of [16], the base grid solution is the same regardless of the content of the subgrid, which is its source of inaccuracy.

Based on the findings in 2-D cases, the $(\mathbf{P}^T\mathbf{P})^{-1}$ plays a role of averaging the contribution from the multiple subgrid patches to produce an accurate curl of $\mathbf{H}$ to generate e_{cb}. Therefore, it is feasible that we modify $(\mathbf{P}^T\mathbf{P})^{-1}$ to a diagonal matrix that makes the final system matrix SPD, and meanwhile let it perform an accurate operation of averaging. The details are as follows.

For each e_{cb} unknown, there is one column vector in $\mathbf{P}_{fc}$ to interpolate the e_{fb} unknowns. Consider the i-th e_{cb}, thus, the i-th row of $\mathbf{P}^T\mathbf{S}_{ff}\mathbf{P}$. It can be written as

$$\mathbf{P}_i^T\mathbf{S}_{ff}\mathbf{P} = \mathbf{P}_i^T\left[\mathbf{S}_{ff,g1}^{\mathrm{F1(i)}} + \mathbf{S}_{ff,g2}^{\mathrm{F1(i)}} + \mathbf{S}_{ff,g1}^{\mathrm{F2(i)}} + \mathbf{S}_{ff,g2}^{\mathrm{F2(i)}}\right]\mathbf{P}, \tag{2.56}$$

where the superscripts $^{\mathrm{F1(i)}}$ and $^{\mathrm{F2(i)}}$ denote face 1, and face 2 respectively that share the i-th e_{cb}, as illustrated in Fig. 2.3. Using our patch-based formulation, the $\mathbf{S}_{ff}$ is nothing but the summation of the rank-1 matrix of each patch in the subgrid region. Hence, when evaluating (2.56), we only need to identify which patch is selected by the row vector $\mathbf{P}_i^T$, i.e., involved in the multiplication with $\mathbf{P}_i^T$, and hence contributing to the product of $\mathbf{P}_i^T\mathbf{S}_{ff}$. Based on the expression of $\mathbf{P}_{fc,i}$ shown in (2.55), clearly, all patches that have an edge located on the two faces that share e_{cb}, and also parallel to the e_{cb}, will be selected. In addition, these patches can also be classified into two groups: one group falls onto the face 1 and face 2, while the other group of patches are perpendicular to the face 1, and face 2 respectively. In Fig. 2.3, the first group that falls onto face 2 (right face) of e_{cb} is shown and colored in red; while in the second group, the patches perpendicular to the face 1 of e_{cb} are shown and colored in blue.

The contribution from the first group of patches is denoted by $\mathbf{S}_{ff,g1}$, while the other is denoted by $\mathbf{S}_{ff,g2}$ in (2.56). For each of the two faces where the e_{cb} resides, both contributions exist.

The $\mathbf{S}_{ff,g1}$ term can be evaluated on each face as the following:

$$\begin{aligned}\mathbf{P}_i^T\mathbf{S}_{ff,g1}\mathbf{P}\{e\} = \mathbf{P}_i^T \Big[&(\mathbf{S}_{h,11}\dot{h}_{11} + ... + \mathbf{S}_{h,n1}\dot{h}_{n1}) + \\ &(\mathbf{S}_{h,12}\dot{h}_{12} + ... + \mathbf{S}_{h,n2}\dot{h}_{n2}) + ... \\ &(\mathbf{S}_{h,1n}\dot{h}_{1n} + ... + \mathbf{S}_{h,nn}\dot{h}_{nn})\Big],\end{aligned} \tag{2.57}$$

in which the $\mathbf{S}_h$ are from the first group of patches, which are those falling onto the face. For a grid ratio of n, it is evident that there are n^2 such patches, each of which is denoted by a row and a column index of the patch using subscripts. For example, h_{n1} denotes the magnetic field on the patch located in the first column (closest to the $e_{cb,i}$) and the n-th row. Based on the expression of $\mathbf{P}_{fc,i}$, (2.57) can be readily evaluated and found as

$$\mathbf{P}_i^T\mathbf{S}_{ff,g1}\mathbf{P}\{e\} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\dot{h}_{ij}}{L_c}. \tag{2.58}$$

Clearly, the numerator represents the sum of all of the normal magnetic fields at the patches residing on the face. If we divide the above by n^2, then the numerator represents the average magnetic field located at the center of the face, i.e., the red point shown in Fig. 2.4. Thus, we have

$$\mathbf{P}_i^T\mathbf{S}_{ff,g1}\mathbf{P}\{e\} = n^2\frac{\dot{h}_{cnt}^{\mathrm{F}}}{L_c}, \tag{2.59}$$

in which h_{cnt}^{F} stands for the magnetic field at the face center normal to the face, and along the reference direction of the face.

Similarly, we can quantitatively evaluate the second group's contribution in (2.56), which is

$$\begin{aligned}\mathbf{P}_i^T\mathbf{S}_{ff,g2}\mathbf{P}\{e\} = \mathbf{P}_i^T \Big[&(\mathbf{S}_{h,11}^{\perp}\dot{h}_{11}^{\perp} + ... + \mathbf{S}_{h,n1}^{\perp}\dot{h}_{n1}^{\perp}) + \\ &(\mathbf{S}_{h,12}^{\perp}\dot{h}_{12}^{\perp} + ... + \mathbf{S}_{h,n2}^{\perp}\dot{h}_{n2}^{\perp}) + ... \\ &(\mathbf{S}_{h,1,n-1}^{\perp}\dot{h}_{1,n-1}^{\perp} + ... + \mathbf{S}_{h,n,n-1}^{\perp}\dot{h}_{n,n-1}^{\perp})\Big].\end{aligned} \tag{2.60}$$

Here, the $\mathbf{S}_h^{\perp}$ is generated from patches that are perpendicular to the face, as those colored in blue in Fig. 2.3. There are $n-1$ columns of such patches on either of the two faces sharing the e_{cb}, and in each column there are n patches, and hence two subscripts are used for h in the above to denote the patch's, thus h's location. Again, based on the expressions of $\mathbf{P}_{fc,i}$ and $\mathbf{S}_h$, we find the above to be

$$\mathbf{P}_i^T \mathbf{S}_{ff,g2} \mathbf{P}\{e\} = \frac{n-1}{nL_f} \sum_{i=1}^{n} \dot{h}_{i1}^{\perp} + \frac{n-2}{nL_f} \sum_{i=1}^{n} \dot{h}_{i2}^{\perp} + ... + \frac{1}{nL_f} \sum_{i=1}^{n} \dot{h}_{i,n-1}^{\perp}. \tag{2.61}$$

If we divide the above by $n(n-1)/2$, then the summation can be viewed as an average magnetic field in the direction perpendicular to the patches, thus parallel to the face where the patches are attached. Since the patches involved in (2.61) are immediately adjacent to the face, the resulting magnetic field after averaging can be viewed as a field located at the point whose distance to the e_{cb} is $L_f/2$, as shown by the blue point in Fig. 2.4, and on the face parallel to the patch. Let this field be h_{sub}. We have

$$\mathbf{P}_i^T \mathbf{S}_{ff,g2} \mathbf{P}\{e\} = \frac{n(n-1)}{2L_f} \dot{h}_{sub}. \tag{2.62}$$

Take face 2, occupied by all red patches shown in Fig. 2.3, as an example. On this face, the first group of patches' contribution, as shown in (2.59), after averaging, produces a magnetic field located at the center of the face, and normal to the face. This point is the red point shown in Fig. 2.4. The magnetic field along the same direction is also produced by the second group of patches on face 1, i.e., $\mathbf{P}_i^T \mathbf{S}_{ff,g2}^{\mathrm{F1(i)}} \mathbf{P}\{e\}$. As shown in (2.62), this component, after averaging, can be viewed as the magnetic field located at the point whose distance to the e_{cb} is $L_f/2$, thus the blue point in Fig. 2.4. We hence can use these two magnetic fields to do an average to obtain the magnetic field located at the midpoint of the two fields, whose distance to the e_{cb} is $L_f{}' = (L_f + L_c)/2$. This magnetic field together with the magnetic field at the center of the adjacent base grid patch, marked as a green point in Fig. 2.4, can then be

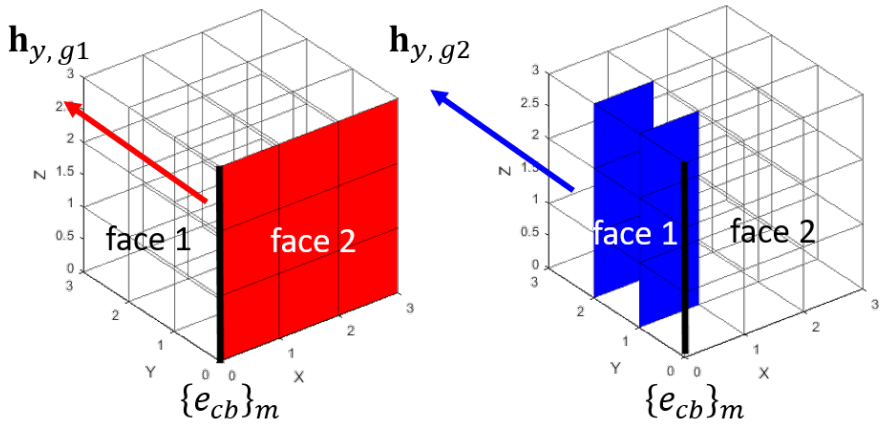


Fig. 2.3. Illustration of one e_{cb} (colored in black), the two faces sharing e_{cb}, and the two groups of subgrid patches whose magnetic fields are used to generate e_{cb}.

used to generate an accurate curl of $\mathbf{H}$ to produce e_{cb}. Similar to the treatment in 2-D, we need to use the average length of L_c and $L_f{}'$ to achieve a better accuracy in computing the curl of $\mathbf{H}$. Thus, the average length, L_{ave} is

$$L_{ave} = \frac{L_c + L_f{}'}{2} = \frac{3L_c + L_f}{4}. \tag{2.63}$$

Based on the aforementioned, we compute e_{cb} in the following way

$$\begin{aligned} &\epsilon_i \ddot{e}_{cb,i} + \sigma_i \dot{e}_{cb,i} = \\ &- \left[\alpha \mathbf{P}_i^T \mathbf{S}_{cc} \mathbf{P} + \beta \mathbf{P}_i^T \mathbf{S}_{ff,g1} \mathbf{P} + \gamma \mathbf{P}_i^T \mathbf{S}_{ff,g2} \mathbf{P}\right] \{e\}, \end{aligned} \tag{2.64}$$

where

$$\alpha = \frac{L_c}{L_{ave}}, \tag{2.65}$$

$$\beta = \frac{L_c}{2n^2 L_{ave}}, \tag{2.66}$$

and

$$\gamma = \frac{L_f}{n(n-1)L_{ave}}. \tag{2.67}$$

Substituting (2.59) and (2.62) into (2.64), it can be seen that (2.64) is nothing but to compute

$$\epsilon_i \ddot{e}_{cb,i} + \sigma_i \dot{e}_{cb,i} = \frac{\dot{h}_{base}^{\mathrm{F1(i)}} - 0.5(\dot{h}_{cnt}^{\mathrm{F1(i)}} + \dot{h}_{sub}^{\mathrm{F1(i)}})}{L_{ave}} + \frac{\dot{h}_{base}^{\mathrm{F2(i)}} - 0.5(\dot{h}_{cnt}^{\mathrm{F2(i)}} + \dot{h}_{sub}^{\mathrm{F2(i)}})}{L_{ave}}, \tag{2.68}$$

where the h_{base} is the magnetic field at the center of the base grid patch that has e_{cb}, and its superscript denotes the subgrid face that is on the same plane as the base grid patch. Hence, (2.64) produces an accurate curl of $\mathbf{H}$ for generating e_{cb}.

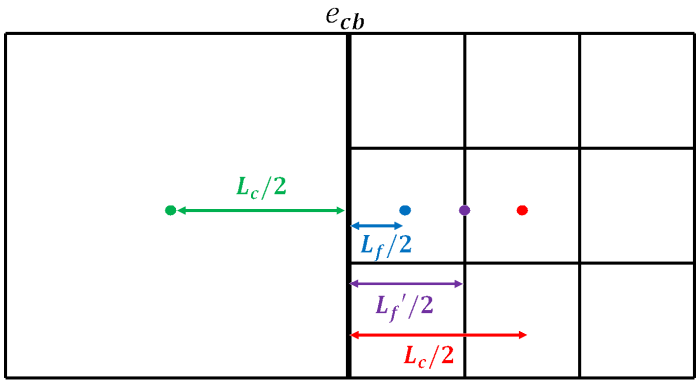


Fig. 2.4. Illustration of the magnetic field point corresponding to each term, and the average length for different patches.

Based on (2.64), now we can write the whole system of equations for solving all unknowns as

$$\mathbf{D}_s\mathbf{D}_\epsilon\{\ddot{e}\} + \mathbf{D}_s\mathbf{D}_\epsilon\{\dot{e}\} = -\mathbf{S}_{total}\{e\}, \tag{2.69}$$

in which

$$\mathbf{D}_s = \mathrm{diag}\{\{1\}_{\#e_{ci}}, \{\gamma/\alpha\}_{\#e_{fi}}, \{1\}_{\#e_{cb}}\}, \tag{2.70}$$

and

$$\mathbf{S}_{total} = \mathbf{D}_1 \left(\mathbf{S}_{h,c}\mathbf{D}_{\mu_c^{-1}}\mathbf{S}_{e,c} + \frac{\beta}{\alpha}\mathbf{P}^T\mathbf{S}_{ff,g1}\mathbf{P} + \frac{\gamma}{\alpha}\mathbf{P}^T(\mathbf{S}_{ff,g2} + \mathbf{S}_{ff,g3})\mathbf{P} \right). \tag{2.71}$$

with

$$\mathbf{D}_1 = \text{diag}\{\{1\}_{\#e_{ci}}, \{1\}_{\#e_{fi}}, \{\alpha\}_{\#e_{cb}}\}. \tag{2.72}$$

In (2.71), the subscript $g3$ denotes the third group of patches, which are inside the subgrid. The $\mathbf{D}_s$ is used in (2.69), because when $\mathbf{S}_{ff,g2}$ term is scaled by $\frac{\gamma}{\alpha}$, the field internal to the subgrid, i.e., e_{fi}, is also scaled by this coefficient. Hence, by left scaling the entire system by $\mathbf{D}_s$, we keep the solution of e_{fi} the same as before. It is obvious that the matrix in the big parenthesis of (2.71) is symmetric positive semi-definite, and hence the final numerical system remains to have nonnegative real eigenvalues. Meanwhile, we have taken the accuracy into account in generating the e_{cb} unknowns.

2.5.2 Interpretation and Implementation in the Original FDTD Difference Equation Based Framework

The proposed 3-D subgridding operator can be interpreted, and thereby implemented in the original FDTD as follows.

Global time stepping

- For e_{ci} and magnetic field unknowns, which are inside the base grid, they are solved in the same way as in the original FDTD.

- For an e_{cb} unknown, we obtain the $\mathbf{H}$ fields at the n^2 subgrid patches on one face (for example, face 1) having the e_{cb}, and then take its average. Let this be h_{cnt}. We then obtain the weighted sum of the magnetic fields shown in the right hand side of (2.61) for the other face (face 2) that has e_{cb}. Dividing the sum by $n(n-1)/(2)$, we obtain h_{sub}. Using the $\mathbf{H}$ field at the base grid patch that has the e_{cb} on the same plane as face 1, denoted by h_{base}, we can perform a curl of $\mathbf{H}$ operation as shown in (2.68) to obtain e_{cb}, together with the h_{cnt} generated on face 2, the h_{sub} at face 1, and the h_{base} on the other base grid patch.

- The e_{fb} unknowns are obtained from $\mathbf{P}_{fc}e_{cb}$.

- The e_{fi} and the magnetic field unknowns inside the subgrid are solved in the same way as that in the original FDTD.

Similar to 2-D, only the e_{cb} and e_{fb} are generated differently.

Local time stepping

For local time stepping, i.e., subgridding in time, all the unknowns, except for e_{cb}, are generated in the same way as above. For the e_{cb} unknown, we do the following. At each time step of time marching, from the base grid, we know the **H** field at previous time step at the two patches having e_{cb}, we use it to perform only a partial curl of **H** operation, which is $(h^1_{base} + h^2_{base})/L_{ave}$. In other words, we do not complete the curl of **H** operation using the **H** fields from the subgrid. The partial curl of **H** evaluated from the base grid only provides one component of the e_{cb}, denoted by $e_{cb,p}$, but it allows all unknowns in the base grid to be generated using a large time step. For the subgrid, we use the time step restricted by the subgrid space step, and obtain the other half of the curl of **H** operation by evaluating the **H** fields in every subgrid patch, and then obtaining h_{sub} and h_{cnt} on both faces inside the subgrid that has e_{cb}, using which we obtain the other component of e_{cb}, $e_{cb,m}$, which correspond to the right hand side of (2.68) with the two base-grid fields excluded. Then adding the $e_{cb,p}$ and $e_{cb,m}$ makes the total e_{cb}.

2.5.3 An Unsymmetrical Subgridding Implementation

In 3-D case, we also tried an interpolation scheme based on the magnetic field interpolation of the other patches from base grid to calculate $\{e_{fb,f}\}$, which is illustrated in Fig. 2.5.

As is shown in Fig. 2.5, to calculate the interface edge on the interface patch, we require magnetic field from an outside small patch which does not exist in either base grid or subgrid. To obtain the most accurate h field, we do interpolations in all three

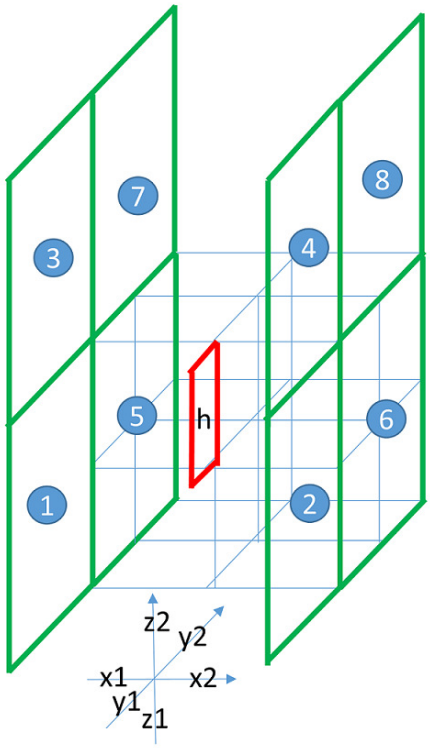


Fig. 2.5. Illustration of interpolation of base grid patch magnetic field.

directions which involves eight patches in total. To explain this interpolation scheme, let the magnetic field for these base grid patches to be $\mathbf{H}_i$, where i rangers from 1 to 8. First we do interpolation for x direction:

$$\begin{aligned}
\mathbf{H}_{12} &= (\frac{x_2}{dx}\mathbf{H}_1 + \frac{x_1}{dx}\mathbf{H}_2), \\
\mathbf{H}_{34} &= (\frac{x_2}{dx}\mathbf{H}_3 + \frac{x_1}{dx}\mathbf{H}_4), \\
\mathbf{H}_{56} &= (\frac{x_2}{dx}\mathbf{H}_5 + \frac{x_1}{dx}\mathbf{H}_6), \\
\mathbf{H}_{78} &= (\frac{x_2}{dx}\mathbf{H}_7 + \frac{x_1}{dx}\mathbf{H}_8),
\end{aligned} \tag{2.73}$$

where x_1, x_2 are distance labeled in Fig. 2.5, denoting distance between position of base grid patch field $\mathbf{H}_i$ and position of h, and dx is the distance between two base grid patches in x direction. Then we do interpolation for z direction:

$$\begin{aligned}\mathbf{H}_{1234} &= (\frac{z_2}{dz}\mathbf{H}_{12} + \frac{z_1}{dz}\mathbf{H}_{34}),\\ \mathbf{H}_{5678} &= (\frac{z_2}{dz}\mathbf{H}_{56} + \frac{z_1}{dz}\mathbf{H}_{78}),\end{aligned} \tag{2.74}$$

where z_1, z_2 are distance labeled in Fig. 2.5, and dz is the distance between two base grid patches in z direction. Finally, we consider interpolation in y direction:

$$h = (\frac{y_2}{dy}\mathbf{H}_{1234} + \frac{y_1}{dy}\mathbf{H}_{5678}), \tag{2.75}$$

where y_1, y_2 are distance labeled in Fig. 2.5, and dy is the distance between two base grid patches in y direction. By introducing this interpolation, we can accurately handle the inhomogeneous problem as well, but as a trade-off, the original symmetric positive semi-definite property will be affected, and it has long term stability issue if central-difference-based explicit time marching is used, making this implementation not as good as the proposed SPD 3-D subgridding operator.

2.6 Numerical Results

In this section, we simulate a variety of 2- and 3-D examples with uniform or highly inhomogeneous materials to examine the performance of the proposed subgridding algorithms in stability, accuracy, and efficiency.

2.6.1 2-D Free-Space Wave Propagation

We first simulate a free-space wave propagation problem in a 2-D region of size 0.5 m by 0.5 m. The base grid size is $L_c = 0.1$ m, and the subgrid is located at the center of the base grid. The grid ratio n ranges from 2, 5, 20, to 100. In Fig. 2.6(a), the grid for a grid ratio of $n = 5$ is shown. The time step used in

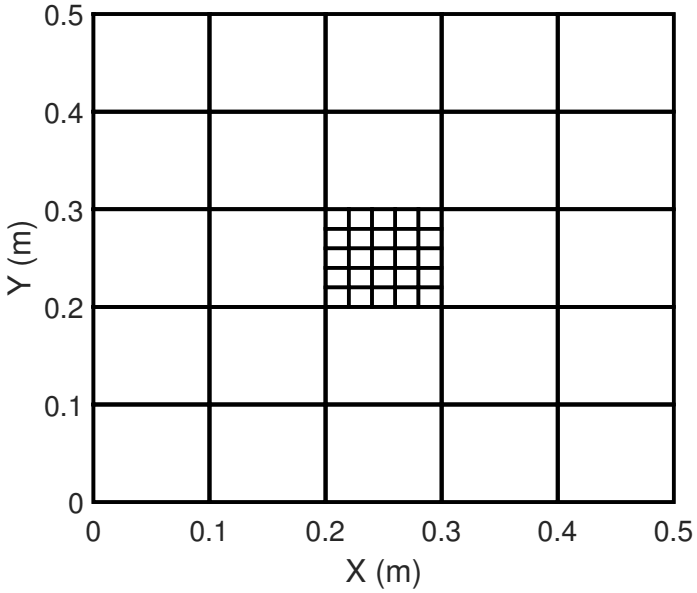
the base grid is $dt_c = 1.9 \times 10^{-10}$ s, which is determined by L_c, and the time step for the subgrid region allowed by the stability condition is $dt_f = dt_c/n$. The $\mathbf{E}^{inc}$ is $\hat{y}2(t-t_0-x/c)e^{(t-t_0-x/c)^2/\tau^2}$ with $c = 3\times10^8$ m/s, $\tau = 2\times10^{-8}$ s and $t_0 = 4\tau$. All the boundaries are terminated by exact absorbing boundary conditions, i.e., known fields for the given problem. The entire solution error at each time step as compared to the analytical solution, $\|\{e\} - \{e\}_{anal}\|/\|\{e\}_{anal}\|$, is plotted versus time in Fig. 2.6(b) for different grid ratios, where $\{e\}$ denotes the vector of all electric field unknowns in the grid. It is clear to see that the simulated fields agree with the analytical solution very well. The center peak error is due to a comparison with zero. We also simulate the same problem using the conventional FDTD method, and compare the CPU run time of the two methods in Table 2.1 for different grid ratios. It is obvious that the proposed method is more efficient since the number of unknowns to solve is greatly reduced, and the time step used in the base grid is also significantly enlarged.

Table 2.1.
CPU Time Comparison of 2-D Example for Different Grid Ratios

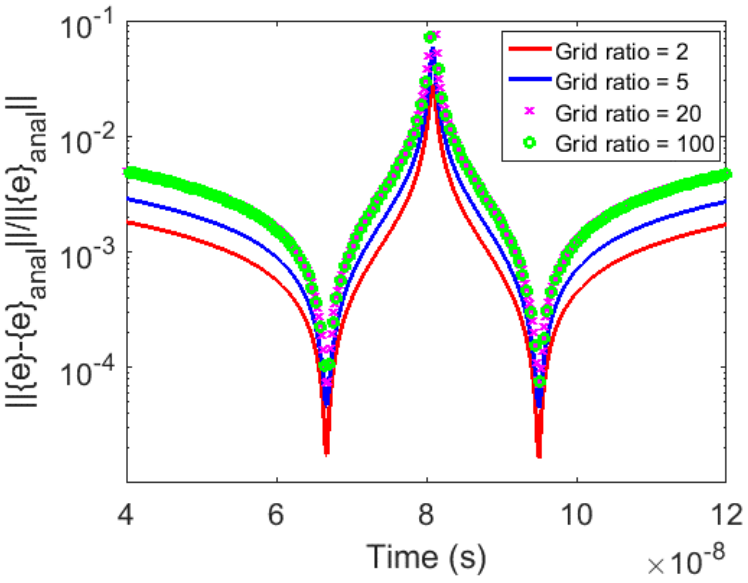
Grid ratio	2	5	20	100
Time (s) of FDTD	0.0487	0.5458	37.8527	5522.14
Time (s) of this method	0.0324	0.0555	0.4359	38.47
Speedup	1.50	9.83	86.84	143.54

2.6.2 2-D PEC Cavity with Conducting Fins

Next, a PEC cavity with two conducting fins separated by a thin gap, as illustrated in Fig. 2.7(a), is simulated. The conductivity of the fins is 5.8×10^7 S/m. A current source is launched at the middle of the fin gap vertically, and with a Gaussian derivative pulse of $-\tau^2\exp(-(t-t_0)^2/\tau^2)$, with $\tau = 2\times10^{-12}$ s, and $t_0 = 4\tau$. The coarse grid size is $L_c = 0.1$ mm, and the subgrid region, having a grid ratio of



(a)



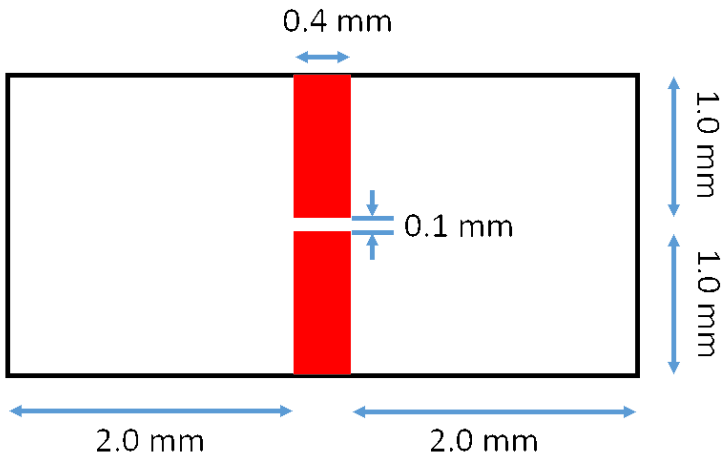
(b)

Fig. 2.6. Simulation of a 2-D wave propagation problem. (a) Grid (for the case of $n = 5$). (b) Entire solution error versus time for different grid ratios ($n = 2, 5, 20, 100$).

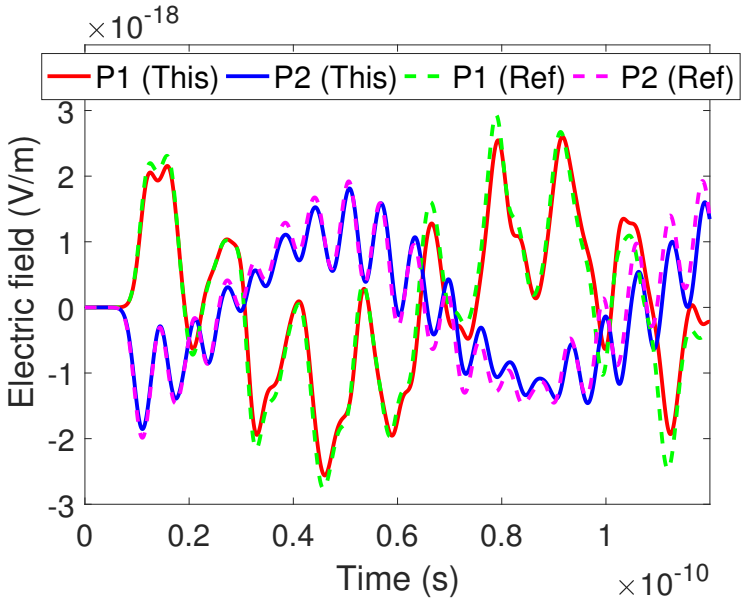
$L_c/L_f = 4$, is located between the two fins. The time step used in the subgrid is $dt_f = 4.2 \times 10^{-14}$ s, which is determined by L_f. The fields simulated from this method at two points, (1, 1.05) and (2.85, 1.5) mm, are plotted in Fig. 2.7(b) and compared with the conventional FDTD results. Very good agreement is observed. The conventional FDTD method, using a uniform grid, takes 6.68 s to finish the simulation while the proposed subgridding method only takes 0.84 s.

2.6.3 3-D Free-Space Wave Propagation

The third example is a free-space wave propagation problem in a 3-D box. The size of the computational domain in each direction is 0.5 m. Along all directions, the coarse space step is $L_c = 0.1$ m. The subgrid is located at the center having a grid ratio n ranging from 2, 4, 10, to 20, making the fine space step $L_f = L_c/n$. The $\mathbf{E}^{inc}$ is $\hat{y}2(t-t_0-x/c)e^{(t-t_0-x/c)^2/\tau^2}$ with $c = 3\times10^8$ m/s, $\tau = 2\times10^{-8}$ s and $t_0 = 4\tau$. Again, all of the boundaries are terminated by exact absorbing boundary conditions. In Fig. 2.8, we first plot the simulated electric fields at two observation points in comparison with the analytical solution for grid ratio $n = 4$. Point 1 is at $(0.1, 0.05, 0.1)$ m and it is inside the base grid, while point 2 is at $(0.225, 0.225, 0.2125)$ m, which is inside the subgrid. As can be seen, the electric fields solved from the proposed method have an excellent agreement with analytical results. We also plot the entire solution error at each time step as compared to the analytical solution, $\|\{e\} - \{e\}_{anal}\|/\|\{e\}_{anal}\|$, versus time in Fig. 2.9 for different grid ratios. As can be observed, the proposed method is not only accurate at selected observation points, but also accurate at all other points. The center peak error is again due to a comparison with zero. In Fig. 2.10, we compare the accuracy of the proposed local time stepping with that of the global one. A grid ratio of 4 is considered. Instead of using the smallest time step dt_f everywhere, we use dt_c in the base grid, which is chosen to be 2, 3, and up to n times larger of the dt_f. As can be seen from Fig. 2.10, the stability is maintained, and the accuracy is not sacrificed.



(a)



(b)

Fig. 2.7. Simulation of a PEC cavity with conducting fins separated by a thin gap. (a) Structure. (b) Simulated electric fields.

In this example, we also compare the difference between the proposed SPD operator shown in (2.69) which has accuracy taken into account, and the preliminary one shown in (2.33), where $(\mathbf{P}^T\mathbf{P})^{-1}$ is kept as it is without modifications. We find

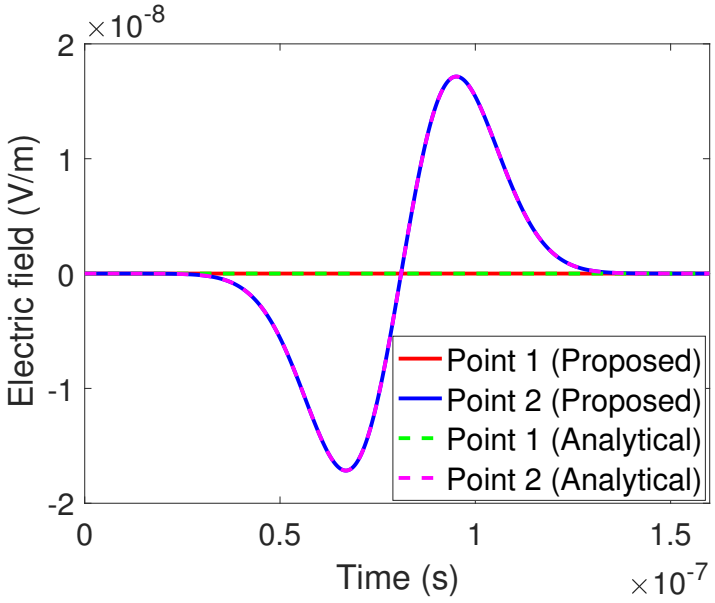


Fig. 2.8. Simulated electric fields at two observation points in comparison with reference analytical solutions.

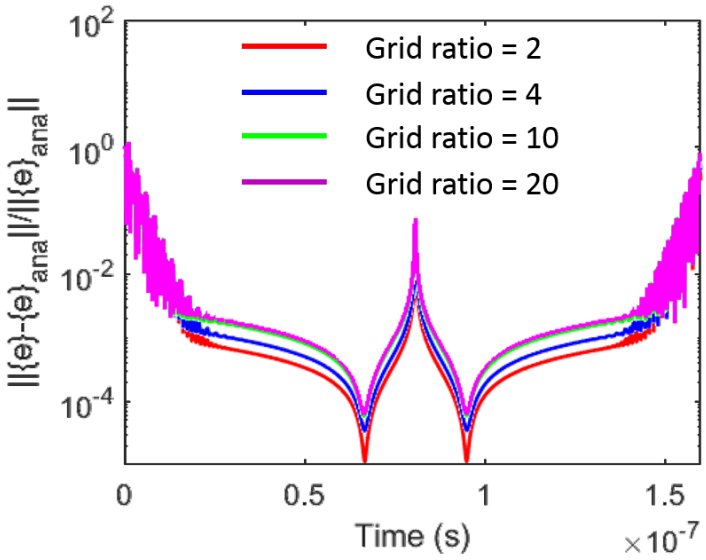


Fig. 2.9. Entire solution error versus time for different grid ratios.

complex eigenvalues such as $3.9384e19 \pm 1.6543e4i$ from (2.33), thus the stability cannot be guaranteed although in this case the imaginary part is small. Meanwhile, the

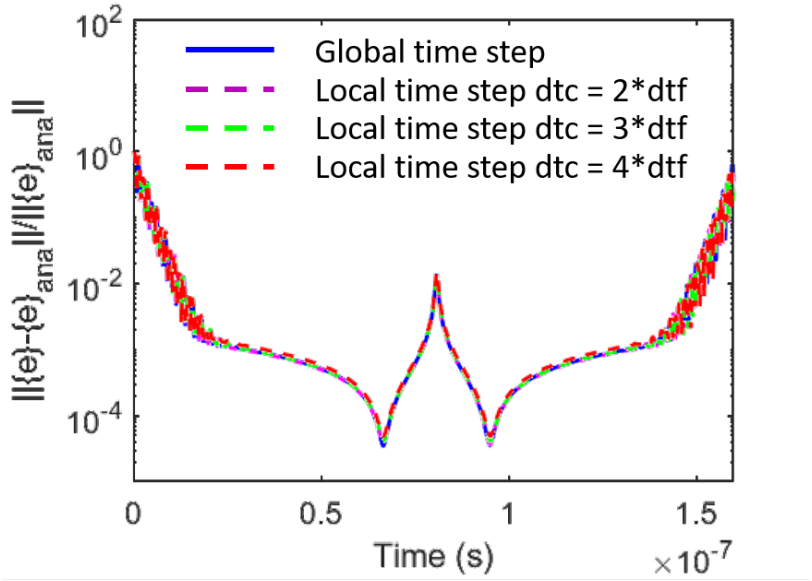


Fig. 2.10. Entire solution error versus time using local time stepping, with different time steps in the base grid.

accuracy is not as good as (2.69), where $(\mathbf{P}^T\mathbf{P})^{-1}$ is modified to perform an accurate averaging for mismatched grid sizes. The accuracy comparison can be seen from Fig. 2.11.

2.6.4 3-D PEC Cavity with an Inhomogeneous Subgrid Region

Next, we simulate a 3-D cavity excited by a current source, which has an inhomogeneous subgrid region, to examine the accuracy of the proposed algorithms in such a setting. The cavity is 1 cm long in all directions and terminated by a PEC boundary condition. The base grid size along each direction is 1 mm, except for the small cube centered at $(4.5, 4.5, 4.5)$ mm, which is illustrated in Fig. 2.12. This center cube is 1 mm long in all directions and filled with inhomogeneous materials, while the base grid has a dielectric constant of 3. The center cube is further subdivided with a grid size of 0.2 mm. Such subgridding results in 125 fine cells. To examine the capability of the proposed work in handling inhomogeneity, each subgrid cell's dielectric constant

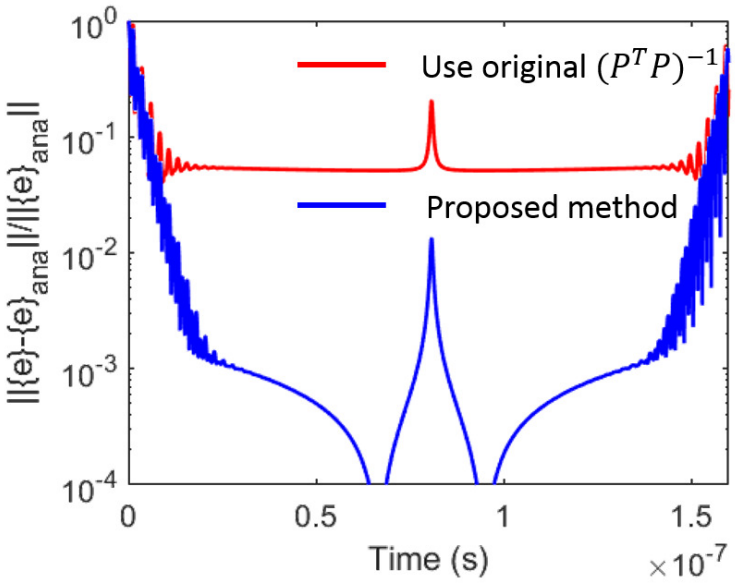


Fig. 2.11. Accuracy comparison between using (2.69) and using (2.33).

is set as a random number in the range of 1 to 125. A current probe is placed at $(3, 3, 3.5)$ mm. The current has a Gaussian pulse whose waveform is $\hat{z}\tau e^{-(t-t_0)^2/\tau^2}$ with $\tau = 2.0 \times 10^{-11}$ s and $t_0 = 4\tau$. As a reference, we also simulate the same problem using the method in [15], which is an unsymmetrical subgridding method but can handle such inhomogeneous problems accurately. The time step allowed by the method in [15] is $dt = 3.8{\times}10^{-13}$ s. In Fig. 2.13, the electric fields sampled at Point 1 $(2, 2, 1.5)$ mm and Point 2 $(8, 8, 7.5)$ mm are plotted in comparison with the reference solution generated using the method of [15] (labeled as Ref1 Fig. 2.13). It can be seen that the accuracy of the proposed method is very good. This shows the proposed new SPD method is able to handle inhomogeneous problems accurately. For comparison, the symmetric semi-definite method in [16] is used to simulate the same problem. As can be seen from Fig. 2.14, the results of [16] (labeled as Ref2) are not as accurate as the proposed method when compared to the reference solution. This is due to the fact that the electric fields shared by the base grid and the subgrid are

solved from the base grid only in [16] to build an SPD system. This scheme can not capture the variation of fields resulting from the inhomogeneous subgrid region.

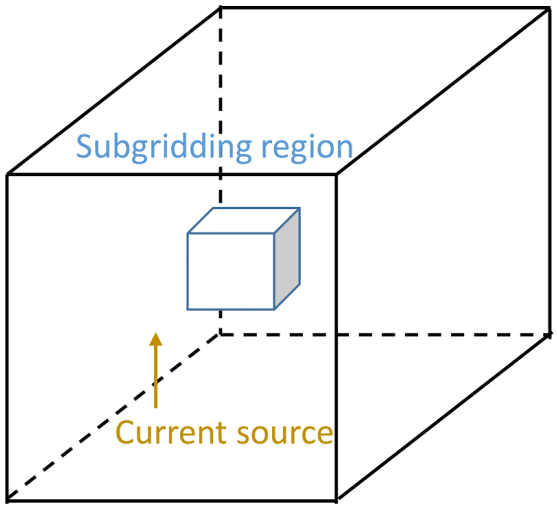


Fig. 2.12. Structure details of a 3-D cavity excited by a current source.

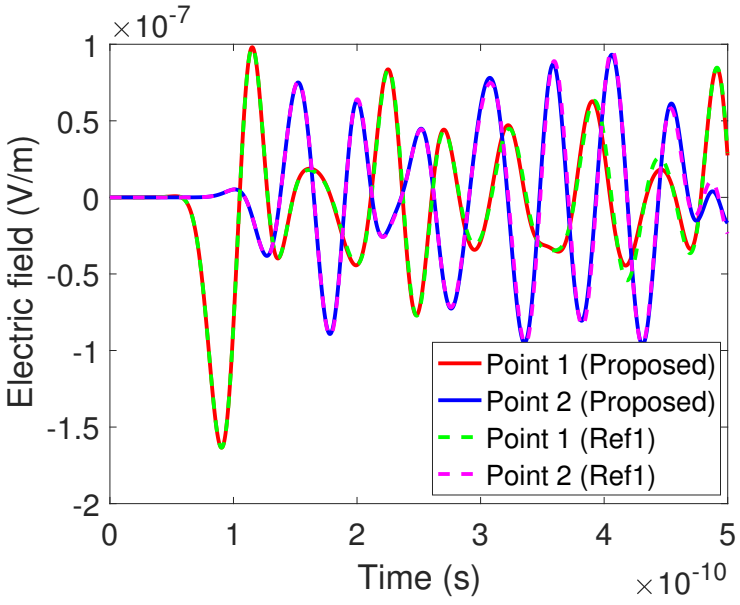


Fig. 2.13. Electric fields at two observation points using the proposed method versus reference results.

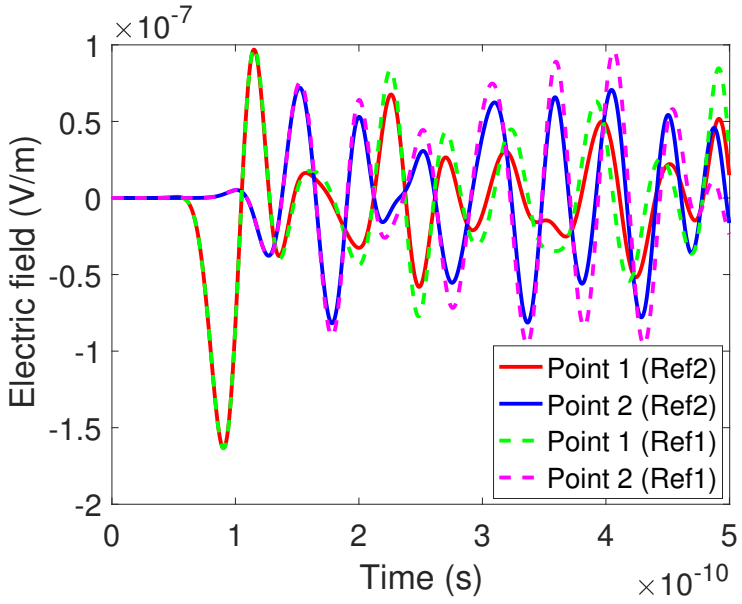


Fig. 2.14. Electric fields at two observation points using method in [16] versus reference results.

2.6.5 Inhomogeneous 3-D Phantom Head beside a Wire Antenna

In this example, we simulate a large-scale phantom head beside a wire antenna, which involves many inhomogeneous materials. The size of the phantom head is $28.16 \times 28.16 \times 17.92$ cm. The permittivity distribution of the head at $z = 2.8$ cm is shown in Fig. 2.15. All the boundaries are truncated by perfect magnetic conducting conditions. The wire antenna is located at $(3.52, 3.52, 2.52)$ cm, with a current pulse of $\hat{z}2(t - t_0)e^{-(t-t_0)^2/\tau^2}$ with $\tau = 5.0 \times 10^{-10}$ s and $t_0 = 4\tau$. The base grid size along x-, y-, z-directions is 4.4, 4.4 and 5.6 mm, respectively. To capture fine tissues, the base grid cell centered at $(14.3, 14.3, 9.24)$ cm is subdivided into subgrid cells in all directions with a grid ratio of 4, making the subgrid grid size along x-, y-, z-directions is 1.1, 1.1 and 1.4 mm, respectively. Again, we use the unsymmetrical subgridding method [15] as our reference. Due to the existence of the subgrid, the unsymmetrical subgridding method [15] must use a time step of 2.2×10^{-12} s across the whole grid to ensure stability. In contrast, the proposed new method allows for a larger time

step $dt_c = 8.8 \times 10^{-12}$ s in the base grid. In Fig. 2.16, the electric fields at two observation points whose locations are $(3.52, 3.52, 15.96)$ cm and $(24.64, 3.52, 15.96)$ cm are plotted in comparison with the reference results obtained using a global time step. In Fig. 2.17, the electric fields obtained from a local time stepping at the same observation points are plotted. It is clear that the two sets of results agree well. The unsymmetrical subgridding method [15] uses 549 s to finish the simulation. In contrast, the proposed subgridding method only costs 345 s using a global time step, and 209 s when using a local time step.

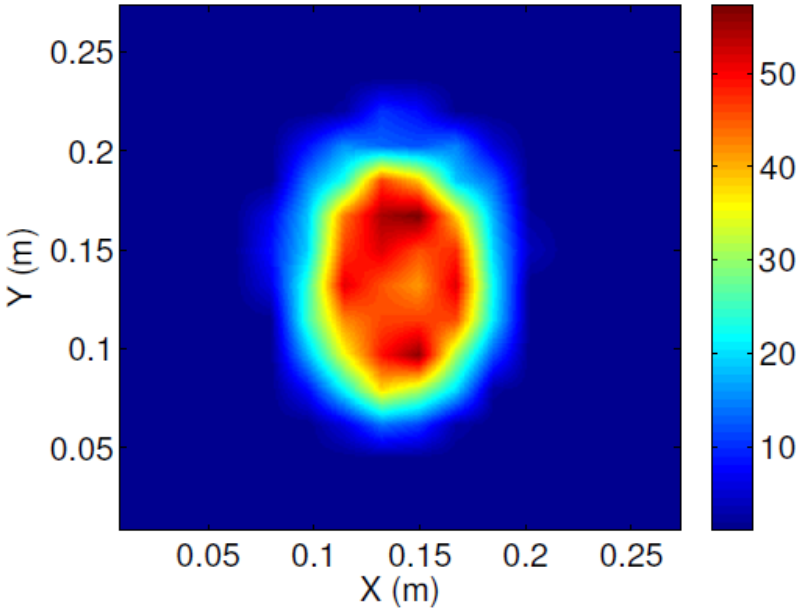


Fig. 2.15. Relative permittivity distribution in a cross section of the phantom head.

2.7 Conclusion

In this chapter, a symmetric positive semi-definite FDTD subgridding method in both space and time is developed for fast FDTD simulations. First, we provide an algebraic method to systematically derive an SPD subgridding operator for the FDTD in both space and time. This method yields a framework for developing a

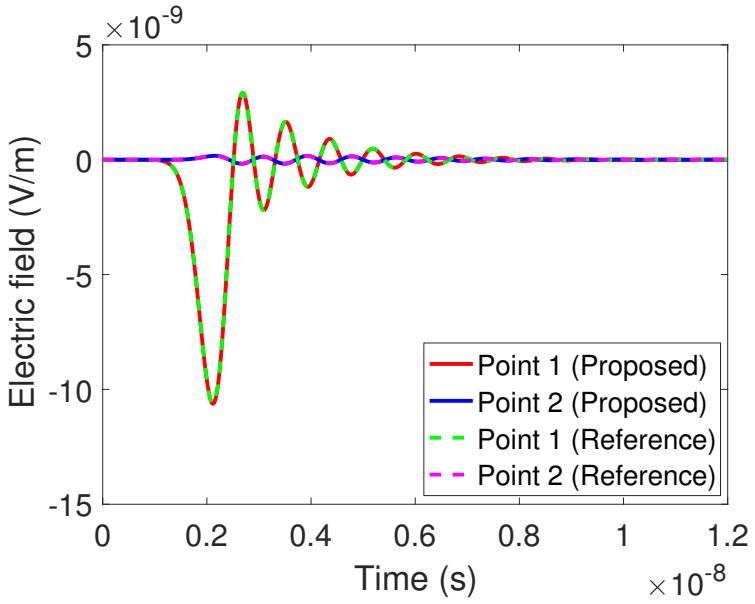


Fig. 2.16. Simulated electric fields at two observation points in comparison with reference results when a global time step is used.

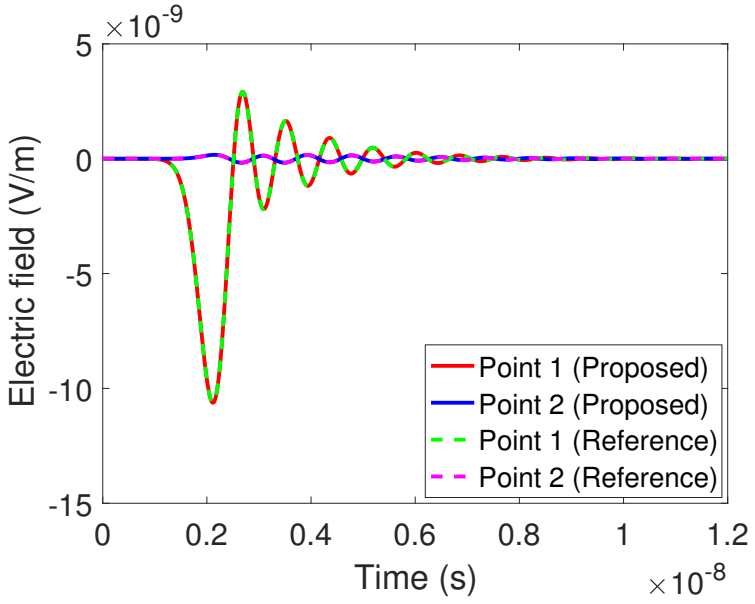


Fig. 2.17. Simulated electric fields at two observation points in comparison with reference results when a local time step is used.

series of SPD subgridding operators. We then take the accuracy into account, and develop 2- and 3-D subgridding algorithms which are not only SPD for arbitrary grid ratios, but also accurate for analyzing general inhomogeneous problems. The stability is guaranteed by construction, because the eigenvalues of the resulting SPD system matrix are nonnegative real. Furthermore, the time step of each grid is determined by the time step local to the grid. Thus, the base grid time step is not restricted by the subgrid region, further accelerating the simulation. We also provide an interpretation of the proposed algorithms in the original FDTD framework, and show how to implement it easily. Numerous numerical experiments have been carried out. Comparisons with both analytical solutions and state-of-the-art subgridding algorithms have demonstrated the accuracy, efficiency, and stability of the proposed new subgridding algorithms.

3. EXPLICIT AND UNCONDITIONALLY STABLE FDTD WITH ANALYTICAL METHOD FOR IDENTIFYING UNSTABLE MODES

3.1 Introduction

As can be seen in Chap. 2, the stability of an explicit time marching requires the time step be restricted by the space step (smallest subgrid size for subgridding algorithm). The root cause of the instability has been identified in [47]. It is of our interest to identify unstable modes and remove them so that we can use larger time step. A new class of explicit yet unconditionally stable time-domain method has been developed. In these methods, one finds either the stable modes, or the unstable ones. Both of them, so far, have been computed numerically, whose cost can be high for large-scale analysis. In this chapter, starting from the finite-difference time-domain (FDTD) system of equations, we show that the eigenvalue solutions of the conventional FDTD governing system in a uniform grid can be analytically obtained. As a result, both unstable modes and the stable ones can be found analytically. Numerical experiments have demonstrated the accuracy and efficiency of the proposed method.

Part of the contents of this chapter has been extracted and revised from the following publication: Kaiyuan Zeng and Dan Jiao, "Explicit and Unconditionally Stable FDTD with Analytical Method for Identifying Unstable Modes," 2018 IEEE International Symposium on Antennas and Propagation and USNC/URSI National Radio Science Meeting [48].

3.2 Review of the FDTD Solution of Maxwell's Equations

The FDTD solution of Maxwell's equations can be written as

$$\{h\}^{n+\frac{1}{2}} = \{h\}^{n-\frac{1}{2}} - \Delta t \mathbf{S}_e \{e\}^n \tag{3.1}$$

$$\{e\}^{n+1} = \{e\}^n + \Delta t \mathbf{S}_h \{h\}^{n+\frac{1}{2}} - \Delta t \mathbf{D}_{\frac{1}{\epsilon}} \{j\}^{n+\frac{1}{2}}, \tag{3.2}$$

where superscripts n, $n+1$, and $n \pm \frac{1}{2}$ denote time instants, Δt represents time step, $\{e\}$ is the vector of electric field unknowns, while $\{h\}$ is the vector of magnetic field unknowns, and $\{j\}$ represents a current source vector. The $\mathbf{S}_e\{e\}$ denotes a discretized $\frac{1}{\mu}\nabla \times \mathbf{E}$ operation, and $\mathbf{S}_h\{h\}$ stands for a discretized $\frac{1}{\epsilon}\nabla \times \mathbf{H}$ operation. The $\mathbf{D}_{\frac{1}{\epsilon}}$ in (3.2) is diagonal and its diagonal entry is $\frac{1}{\epsilon}$. By eliminating $\{h\}$ from (3.1) and (3.2), we can obtain a second-order equation in time for $\{e\}$ as follows

$$\frac{\{e\}^{n+1} - 2\{e\}^n + \{e\}^{n-1}}{\Delta t^2} + \mathbf{S}_h \mathbf{S}_e \{e\}^n = \{f\}^n, \tag{3.3}$$

where $\{f\}^n$ denotes the terms moved to the right hand side when deriving (3.3). Let $\mathbf{S} = \mathbf{S}_h \mathbf{S}_e$. Based on [47], the eigenmodes of $\mathbf{S}$, whose eigenvalues λ_i satisfy the following condition, can always be stably simulated by the given time step Δt

$$\lambda_i \leq 4/\Delta t^2. \tag{3.4}$$

The root cause of instability is the eigenmodes whose eigenvalues λ_i are greater than $4/\Delta t^2$, which are termed unstable modes.

Here, we still build $\mathbf{S}$ by the patch-based approach based on [44] in a matrix based form, to facilitate the eigenvalue problem analysis. Let the total number of $\mathbf{E}$ unknowns be N_e, and the total number of $\mathbf{H}$ unknowns be N_h. We loop over all the patches in a 2- or 3-D grid, for each patch, we generate a row vector, which is

$$\mathbf{S}_e^{(i)} = \frac{1}{\mu}\{\frac{1}{L_i}, -\frac{1}{L_i}, -\frac{1}{W_i}, \frac{1}{W_i}\} \oplus zeros(1, N_e), \tag{3.5}$$

which has only four nonzero elements with L_i and W_i being the two side lengths of the i-th patch. The $\oplus$ denotes an extended addition by adding the four nonzero elements

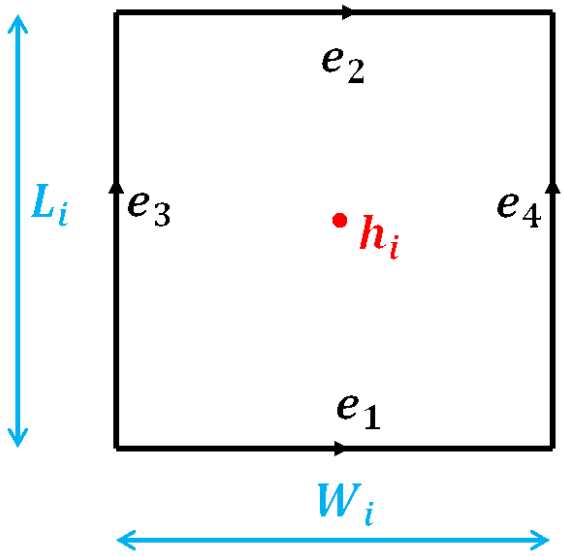


Fig. 3.1. Illustration of a patch-based discretization.

upon a zero vector of length N_e, based on the global indexes of the four $\mathbf{E}$ unknowns on the patch. Similarly, we generate a column vector

$$\mathbf{S}_h^{(i)} = \frac{1}{\epsilon}\{\frac{1}{L_i}, -\frac{1}{L_i}, -\frac{1}{W_i}, \frac{1}{W_i}\}^T \oplus zeros(N_e, 1), \tag{3.6}$$

which is nothing but the transpose of (3.5). The $\mathbf{S}$ is a sum of the rank-1 matrix $\mathbf{S}_h^{(i)}\mathbf{S}_e^{(i)}$ over all the patches.

3.3 Proposed Analytical Method for Finding Unstable Modes

We find that the eigenvalue λ_t and eigenvector v_t of $\mathbf{S}_t = \mathbf{S}_e\mathbf{S}_h$ can be found analytically, from which the eigenvalue λ and eigenvector v of $\mathbf{S}$ can also be obtained without any computation. According to (3.5) and (3.6), the ij-th element of $\mathbf{S}_t$ can be calculated as

$$\mathbf{S}_t^{(ij)} = \mathbf{S}_e^{(i)}\mathbf{S}_h^{(j)}, \tag{3.7}$$

whose value is determined by the relation between the i-th and j-th patch as the following

$$\mathbf{S}_t^{(ij)} = \frac{1}{\mu\epsilon}\begin{cases} \frac{2}{L^2}+\frac{2}{W^2} & i=j \\ -\frac{1}{W^2} & \text{neighbors sharing edge of length } L \\ -\frac{1}{L^2} & \text{neighbors sharing edge of length } W \\ 0 & i\text{-th}, j\text{-th patches are isolated,} \end{cases} \tag{3.8}$$

where L and W are the two side lengths of a cell.

The whole $\mathbf{S}_t$ matrix can be presented as a Kronecker sum of discrete Laplacians, in which case all of its eigenvalues and eigenvectors can be explicitly calculated. For example, in an n by n uniform grid with a PMC boundary condition, $\mathbf{S}_t$ can be written as

$$\mathbf{S}_t = \frac{1}{\mu\epsilon}\left[\mathbf{I}\otimes\mathbf{T}_1 + \mathbf{T}_2\otimes\left(-\frac{1}{L^2}\right)\mathbf{I}\right], \tag{3.9}$$

where the $\otimes$ denotes a Kronecker product, $\mathbf{I}$ is an identity matrix, and

$$\mathbf{T}_1 = \begin{bmatrix} \frac{2}{L^2}+\frac{2}{W^2} & -\frac{1}{W^2} & 0 & 0 \\ -\frac{1}{W^2} & \frac{2}{L^2}+\frac{2}{W^2} & \ddots & 0 \\ 0 & \ddots & \ddots & -\frac{1}{W^2} \\ 0 & 0 & -\frac{1}{W^2} & \frac{2}{L^2}+\frac{2}{W^2} \end{bmatrix}_{n\times n}, \tag{3.10}$$

$$\mathbf{T}_2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & \ddots & 0 \\ 0 & \ddots & \ddots & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}_{n\times n}. \tag{3.11}$$

According to [49], tridiagonal matrices such as $\mathbf{T}_1$ and $\mathbf{T}_2$ have analytical eigenvalue solutions. If $\lambda_{\mathbf{T}_1}$ is an eigenvalue of $\mathbf{T}_1$, and $(v_{\mathbf{T}_1,1}, v_{\mathbf{T}_1,2}, ..., v_{\mathbf{T}_1,n})^T$ is its corresponding eigenvector, then

$$\lambda_{\mathbf{T}_1,k_1} = \left(\frac{2}{W^2}+\frac{2}{L^2}\right) - 2\frac{1}{W^2}\cos\frac{k_1\pi}{n+1}, \tag{3.12}$$

for $k_1 \in \{1, 2, ..., n\}$, and the corresponding $v^{(k_1)}_{\mathbf{T}_1,j_1}$ is

$$v^{(k_1)}_{\mathbf{T}_1,j_1} = \sin \frac{k_1 j_1 \pi}{n+1},\ j_1 = 1, 2, ..., n. \tag{3.13}$$

If $\lambda_{\mathbf{T}_2}$ is an eigenvalue of $\mathbf{T}_2$, and $(v_{\mathbf{T}_2,1}, v_{\mathbf{T}_2,2}, ..., v_{\mathbf{T}_2,n})^T$ is its corresponding eigenvector, then

$$\lambda_{\mathbf{T}_2,k_2} = 2 \cos \frac{k_2 \pi}{n+1}, \tag{3.14}$$

for $k_2 \in \{1, 2, ..., n\}$, and the corresponding $v^{(k_2)}_{\mathbf{T}_2,j_2}$ is

$$v^{(k_2)}_{\mathbf{T}_2,j_2} = \sin \frac{k_2 j_2 \pi}{n+1},\ j_2 = 1, 2, ..., n. \tag{3.15}$$

The $\mathbf{S}_t$'s eigenvalue λ_t can then be obtained as

$$\lambda_t = \frac{1}{\mu\epsilon} \left[\lambda_{\mathbf{T}_1,k_1} + \left(-\frac{1}{L^2} \right) \lambda_{\mathbf{T}_2,k_2} \right], \tag{3.16}$$

with corresponding eigenvector v_t as the vectorization of $v^{(k_1)}_{\mathbf{T}_1} (v^{(k_2)}_{\mathbf{T}_2})^T$. Next we show that $\mathbf{S}$ also has an analytical eigenvalue solution.

If v_t is an eigenvector of $\mathbf{S}_t = \mathbf{S}_e \mathbf{S}_h$ for a nonzero λ_t, then $\mathbf{S}_h v_t \neq 0$, and

$$\lambda_t(\mathbf{S}_h v_t) = \mathbf{S}_h(\lambda_t v_t) = \mathbf{S}_h(\mathbf{S}_e \mathbf{S}_h v_t) = (\mathbf{S}_h \mathbf{S}_e)\mathbf{S}_h v_t = \mathbf{S}(\mathbf{S}_h v_t), \tag{3.17}$$

namely $\mathbf{S}_h v_t$ is an eigenvector for $\mathbf{S} = \mathbf{S}_h \mathbf{S}_e$ with the same eigenvalue λ_t.

Let $\mathbf{V}_h$ denote the matrix formed by all unstable modes. Then we use $\mathbf{V}_h$ to directly change the original system matrix $\mathbf{S}$ to a new system matrix $\mathbf{S}_l$

$$\mathbf{S}_l = \mathbf{S} - \mathbf{V}_h {\mathbf{V}_h}^T \mathbf{S}, \tag{3.18}$$

and perform a time marching on the updated new system $\mathbf{S}_l$,

$$\frac{\partial^2 \{e\}}{\partial t^2} + \mathbf{S}_l \{e\} = \{f\} \tag{3.19}$$

After obtaining e^{n+1} at every step, we add the following treatment

$$e^{n+1} = e^{n+1} - \mathbf{V}_h {\mathbf{V}_h}^T e^{n+1} \tag{3.20}$$

to ensure the solution is free of $\mathbf{V}_h$ modes.

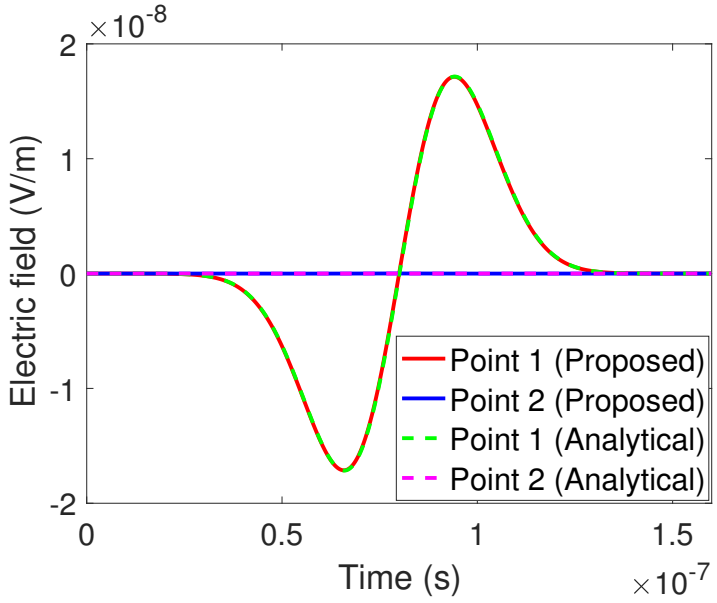


Fig. 3.2. Two electric fields in comparison with analytical results.

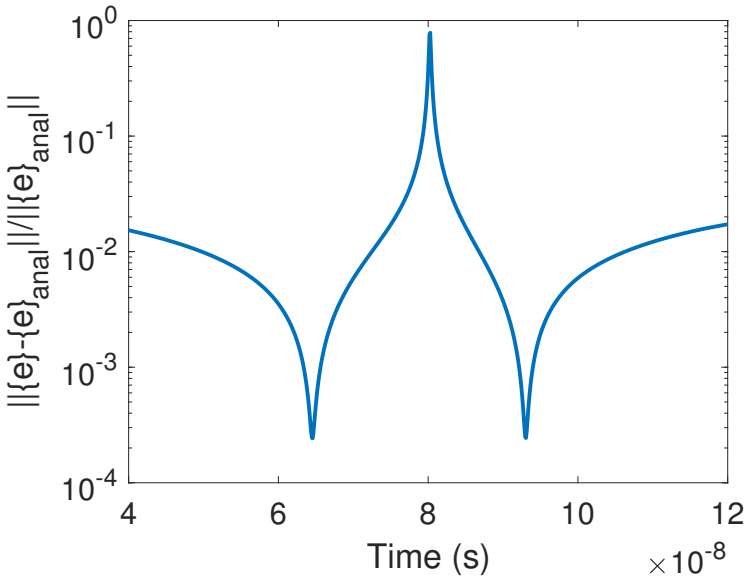


Fig. 3.3. Entire solution error versus time.

3.4 Numerical Results

3.4.1 2-D Free-Space Wave Propagation

A free-space wave propagation problem in a 2-D region of size 0.1 m by 0.1 m is investigated, with smallest grid size of 0.02 m. The $\mathbf{E}^{inc}$ is $\hat{y}2(t-t_0-x/c)e^{(t-t_0-x/c)^2/\tau^2}$

with $c = 3 \times 10^8\ m/s$, $\tau = 2 \times 10^{-8}\ s$ and $t_0 = 4\tau$. The unknown region is surrounded by an exact absorbing boundary layer. The time step Δt is 9.4×10^{-11} s with $\mathbf{V}_h$ containing 21 high modes, otherwise Δt has to be 4.8×10^{-11} s with the original system matrix. The unstable modes are not required by accuracy, and hence they can be deducted from the system matrix without affecting accuracy. In Fig. 3.2, the electric fields at two randomly selected points are plotted in comparison with analytical data. The entire solution error at each time step is compared to the analytical solution, $\|\{e\} - \{e\}_{anal}\|/\|\{e\}_{anal}\|$, is plotted versus time in Fig. 3.3. To demonstrate the efficiency of the proposed method for finding $\mathbf{V}_h$, we also use numerical method (MATLAB function eigs) to find $\mathbf{V}_h$ and compare the CPU time. The simulation parameters are summarized in Table 3.1.

Table 3.1.
Comparison of CPU Time for Finding $\mathbf{V}_h$

# of N_h	25	100	400	2500
Time (s) using eigs	0.0522	0.0737	0.4871	79.1175
Time (s) using this method	0.0043	0.0046	0.0075	0.0841
Speedup	12.14	16.02	64.95	940.75

3.4.2 2-D Grid with Current Source as Excitation

Another example with a current source excitation is investigated. The 2-D region size is 0.1 m by 0.1 m, with smallest grid size of 0.02 m. A current source is provided at one edge, and the Gaussian derivative pulse is $2(1-2(t-t_0)^2/\tau^2)\exp(-(t-t_0)^2/\tau^2)$, with $\tau = 2.4 \times 10^{-8}$ s, and $t_0 = 4\tau$. Without $\mathbf{V}_h$ removal, the Δt is 4.8×10^{-11} s with the original system matrix. To double the time marching step, we remove 22 high modes to make the Δt to be 9.6×10^{-11} s. Two random points are selected to compare their electric field with large Δt and small Δt simulation, which is shown in

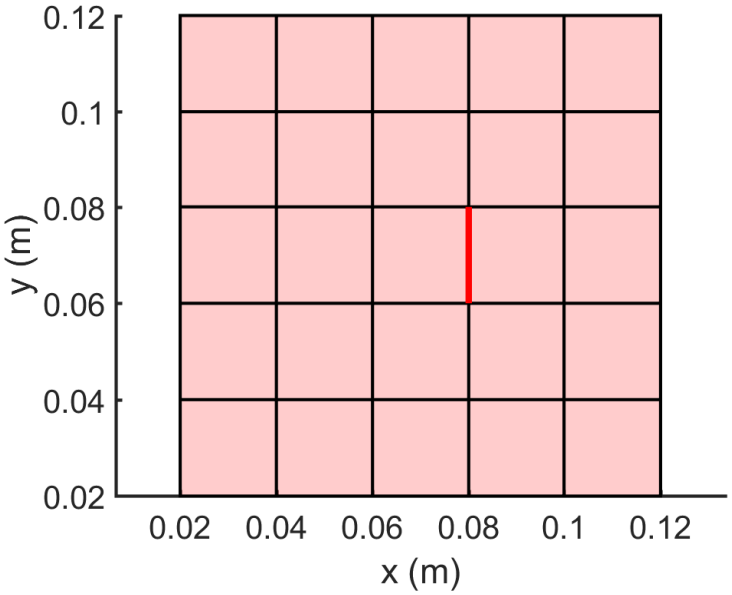


Fig. 3.4. 2-D grid with current source labeled by red edge.

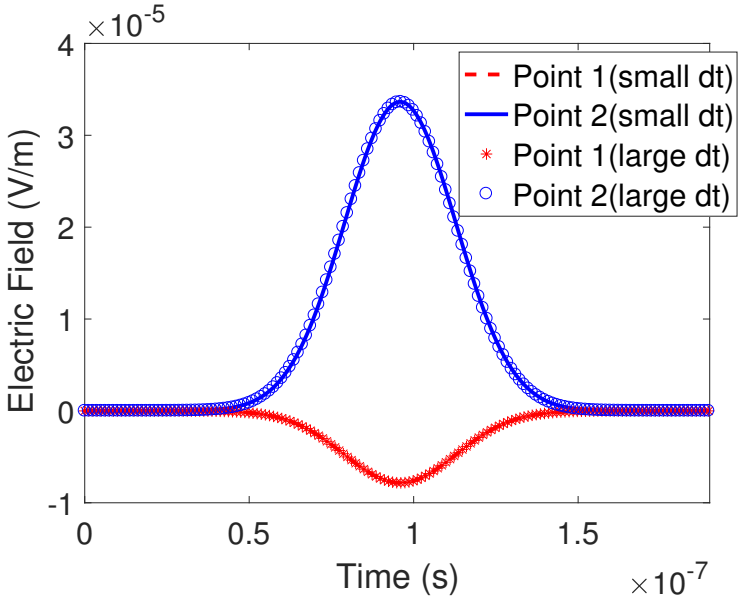


Fig. 3.5. Two electric fields with simulation of different time step.

Fig. 3.5. And the entire solution error between large Δt and small Δt simulation is plotted in Fig. 3.6 as a function of time. Excellent agreement can be observed.

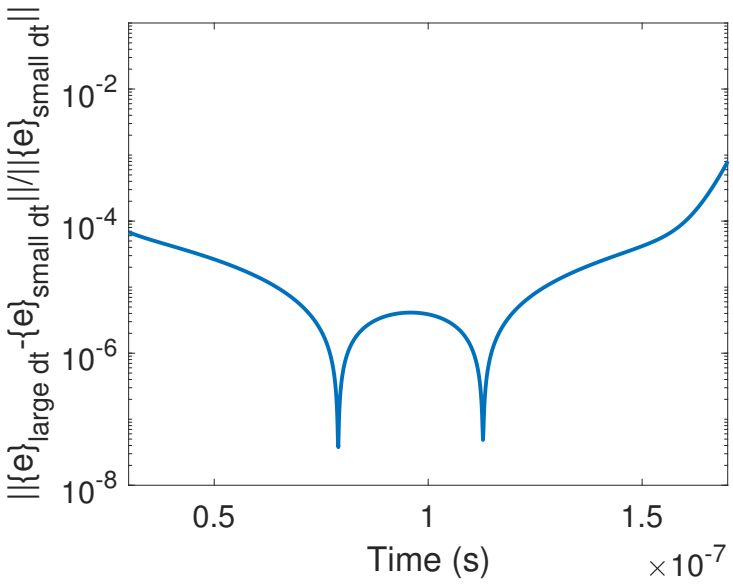


Fig. 3.6. Entire solution error versus time for different time step.

3.5 Conclusion

In existing explicit and unconditionally stable time-domain methods, the root cause of instability has been identified to be the eigenmodes of the governing system matrix whose eigenvalues are too high to be accurately simulated by a given time step. In this work, we show that such unstable modes as well as stable modes can be found for the conventional FDTD method in an analytical way. Numerical results have validated the proposed method.

4. EXPLICIT UNCONDITIONALLY STABLE SYMMETRIC POSITIVE SEMI-DEFINITE FDTD SUBGRIDDING ALGORITHM WITH ANALYTICAL REMOVAL OF UNSTABLE MODES

4.1 Introduction

In Chap. 2 we introduce a symmetric positive semi-definite subgridding algorithm, which not only preserves stability and satisfies accuracy, but also permits the use of a local time step. Nonetheless, the time step in a subgrid is still restricted by the smallest space step inside the subgrid region. In Chap. 3, we show the time step of a conventional explicit finite-difference time-domain (FDTD) method is restricted by space step for stability, and we propose an analytical method to identify such unstable modes.

In this chapter, by combining the two techniques, an explicit and symmetric positive semi-definite FDTD subgridding algorithm is made unconditionally stable for fast FDTD simulations. The root cause of instability in a subgridding algorithm is found to be the eigenmodes of the subgrid whose eigenvalues are too high to be accurately simulated by a given time step. These unstable modes are then identified analytically and removed from the numerical system based on prescribed accuracy. The resultant explicit subgridding algorithm not only permits a local time stepping in the base grid and the subgrid, but also is unconditionally stable regardless of space step in the fine subgrid. Numerical experiments have validated the accuracy and efficiency of the proposed method.

Part of the contents of this chapter has been extracted and revised from the following publication: Kaiyuan Zeng and Dan Jiao, "Explicit Unconditionally Stable Symmetric Positive Semi-Definite FDTD Subgridding Algorithm with Analytical Re-

moval of Unstable Modes," 2019 IEEE International Symposium on Antennas and Propagation and USNC/URSI National Radio Science Meeting [50].

4.2 Proposed Method

Based on contents in Chap. 2, when subgrids exist, we solve

$$\begin{bmatrix} \mathbf{D}_\epsilon & 0 \\ 0 & \mathbf{P}^T\mathbf{D}_\epsilon\mathbf{P} \end{bmatrix} \begin{bmatrix} \ddot{e}_p \\ \ddot{e}_m \end{bmatrix} + \begin{bmatrix} \mathbf{S}_{cc} & \mathbf{S}_{cc} \\ \mathbf{S}_{ff} & \mathbf{S}_{ff} \end{bmatrix} \begin{bmatrix} e_p \\ e_m \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix}, \tag{4.1}$$

where total solution $\{e\} = \{e\}_p + \{e\}_m$, $\mathbf{P}$ is used to connect subgrid interface unknowns with base grid unknowns, $\mathbf{S}_{cc} = \mathbf{S}_{h,c}\mathbf{D}_{\mu_c^{-1}}\mathbf{S}_{e,c}$ is assembled from base grid, $\mathbf{S}_{ff} = \mathbf{P}^T\mathbf{D}_f\mathbf{S}_{h,f}\mathbf{D}_{\mu_f^{-1}}\mathbf{S}_{e,f}\mathbf{P} = \mathbf{P}^T\mathbf{D}_f\mathbf{S}_{ff\_local}\mathbf{P}$ is from subgrid, and $\mathbf{D}_f$ is a diagonal matrix of scaling factors for interface unknowns. The new system (4.1) can allow local time stepping. However, the time step for simulating subgrid unknowns is still restricted by the smallest space step in the subgrid.

The high eigenmodes (corresponding to the largest eigenvalues) of $\mathbf{D}_\epsilon^{-1}\mathbf{S}$ are identified as the root cause of stability [44], where $\mathbf{S} = \mathbf{S}_{cc} + \mathbf{S}_{ff}$. Based on [44], the high eigenmodes are localized in the fine region, and hence they can be found accurately from the eigenmodes of $\mathbf{S}_{ff}$. Using this fact, in this work, we first find the eigenvectors and eigenvalues of $\mathbf{S}_{ff}$. In [44], this is done numerically; in this work, we do it analytically, and hence greatly saving CPU run time. We then develop a scheme to determine which high modes can be removed without sacrificing prescribed accuracy. The detailed procedure is as follows.

First, we find the eigenvalues λ and eigenvectors $\mathbf{V}$ of $\mathbf{S}_{ff\_local}$ using the analytical approach we developed in Chap. 3. The unstable modes correspond to the $\mathbf{V}$ whose eigenvalues are the largest. Denoting such a mode by $\mathbf{V}_{hi}$. We compute $\mathbf{P}^T\mathbf{D}_f\mathbf{V}_{hi}$ and then orthogonalize it to obtain $\widetilde{\mathbf{V}}_{hi}$. Not all $\widetilde{\mathbf{V}}_{hi}$ can be removed, since not all of them are accurate eigenvectors of the original global system. To determine which

$\widetilde{\mathbf{V}}_{hi}$ to remove, we develop the following criterion. Given an accuracy threshold ϵ, if $\widetilde{\mathbf{V}}_{hi}$ satisfies

$$\epsilon_{acc} = \frac{\|\mathbf{S}\widetilde{\mathbf{V}}_{hi} - \widetilde{\mathbf{V}}_{hi}\widetilde{\mathbf{V}}_{hi}^T\mathbf{S}\widetilde{\mathbf{V}}_{hi}\|}{\|\mathbf{S}\widetilde{\mathbf{V}}_{hi}\|} < \epsilon, \tag{4.2}$$

it can be removed. This is because if $\widetilde{\mathbf{V}}_{hi}$ is an accurate eigenvector of $\mathbf{S}$, $\mathbf{S}\widetilde{\mathbf{V}}_{hi} \approx \widetilde{\mathbf{V}}_{hi}y_1$ should satisfy. Writing

$$\mathbf{S}\widetilde{\mathbf{V}}_{hi} = \widetilde{\mathbf{V}}_{hi}y_1 + \widetilde{\mathbf{V}}_{hi}^{\perp}y_2, \tag{4.3}$$

where $\widetilde{\mathbf{V}}_{hi}^{\perp}$ is orthogonal to $\widetilde{\mathbf{V}}_{hi}$, it can be seen that (4.2) is an effective measure of whether the difference between $\mathbf{S}\widetilde{\mathbf{V}}_{hi}$ and $\widetilde{\mathbf{V}}_{hi}y_1$ is sufficiently small.

To see this more clearly, we multiply $\widetilde{\mathbf{V}}_{hi}^T$ on both sides of (4.3), the weight of $\widetilde{\mathbf{V}}_{hi}$, y_1, can be obtained as $\widetilde{\mathbf{V}}_{hi}^T\mathbf{S}\widetilde{\mathbf{V}}_{hi}$ because we have $\widetilde{\mathbf{V}}_{hi}^T\widetilde{\mathbf{V}}_{hi} = \mathbf{I}$ and $\widetilde{\mathbf{V}}_{hi}^T\widetilde{\mathbf{V}}_{hi}^{\perp} = 0$.

We then remove the identified unstable modes from the original system (4.1) by formulating

$$\begin{bmatrix} \mathbf{D}_\epsilon & 0 \\ 0 & \mathbf{P}^T\mathbf{D}_\epsilon\mathbf{P} \end{bmatrix} \begin{bmatrix} \ddot{e}_p \\ \ddot{e}_m \end{bmatrix} + \begin{bmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{I} - \widetilde{\mathbf{V}}_{hi}\widetilde{\mathbf{V}}_{hi}^T \end{bmatrix} \begin{bmatrix} \mathbf{S}_{cc} & \mathbf{S}_{cc} \\ \mathbf{S}_{ff} & \mathbf{S}_{ff} \end{bmatrix} \begin{bmatrix} e_p \\ e_m \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix}.$$

After obtaining e_m^{n+1} at every step, we add the following

$$e_m^{n+1} = e_m^{n+1} - \widetilde{\mathbf{V}}_{hi}\widetilde{\mathbf{V}}_{hi}^T e_m^{n+1} \tag{4.4}$$

to ensure the solution is free of $\widetilde{\mathbf{V}}_{hi}$ modes.

4.3 Numerical Results

4.3.1 2-D Free-Space Wave Propagation

First, a wave propagation problem in a 2-D region is considered. The details of the grid are shown in Fig. 4.1. The base grid size is $L_c = 0.1$ m, and a subgrid region, having a grid ratio of $L_c/L_f = 4$, is introduced to examine the unconditional stability

of the proposed method. The incident field $\mathbf{E}^{inc}$ is $\hat{y}2(t-t_0-x/c)e^{-(t-t_0-x/c)^2/\tau^2}$ with $c = 3 \times 10^8$ m/s, $\tau = 2 \times 10^{-8}$ s, and $t_0 = 4\tau$. The boundaries are terminated by an exact absorbing boundary condition. Without removing unstable modes, the time step allowed in the subgrid is $dt_f = 5.28 \times 10^{-11}$ s, which is determined by L_f.

In Table 4.1, we list the accuracy ϵ_{acc} as a function of the number of $\widetilde{\mathbf{V}}_{hi}$ removed. Clearly, as theoretically predicted, the modes having larger eigenvalues have better accuracy.

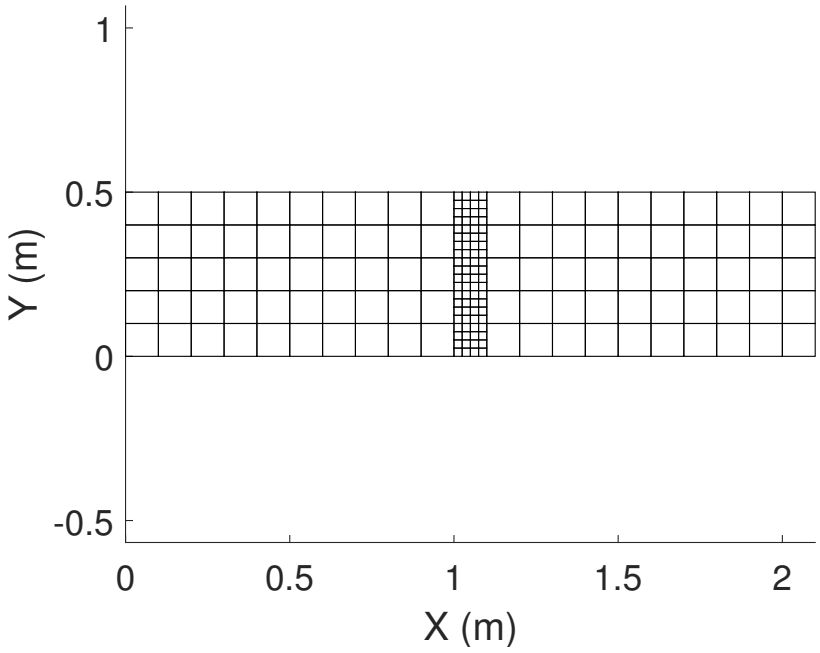


Fig. 4.1. Grid of a 2-D wave propagation problem.

Table 4.1.
Accuracy as a Function of the Number of Removed High Modes

number of modes	1	10	30	50	80
ϵ_{acc}	0.0487	0.0594	0.0842	0.1038	0.1425

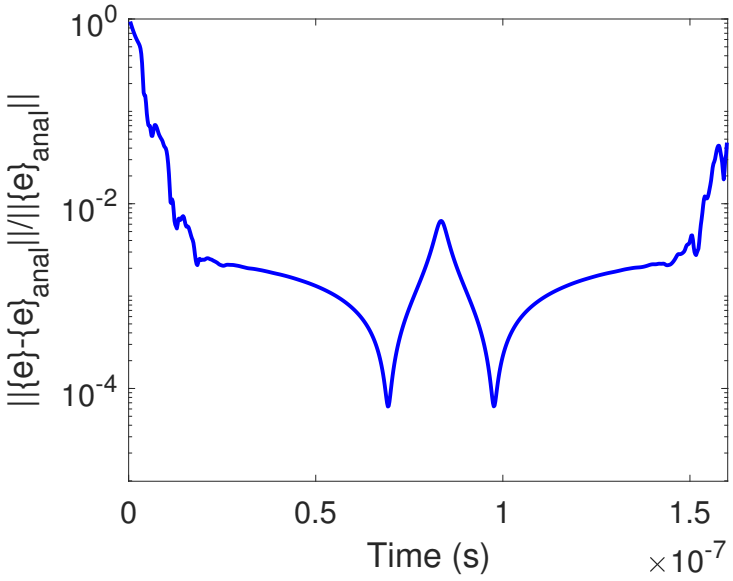


Fig. 4.2. Entire solution error versus time of the wave propagation problem.

Based on Table 4.1, we remove 29 highest $\widetilde{\mathbf{V}}_{hi}$ modes, the dt_f is hence enlarged to 7.04×10^{-11} s from the original 5.28×10^{-11} s, while the time step in the base grid is the same as before. The resultant entire solution error as compared to the analytical solution, $\|\{e\} - \{e\}_{anal}\|/\|\{e\}_{anal}\|$, is plotted versus time in Fig. 4.2, which reveals good accuracy. The large error at early and late time is due to a comparison with zero fields. Without removing unstable modes, it takes the subgridding method 0.40 s to finish the simulation. In contrast, the proposed method only takes 0.21 s, including the time for finding the unstable modes.

4.3.2 2-D PEC Cavity with Conducting Fins

Next, a PEC cavity with two conducting fins separated by a thin gap, as illustrated in Fig. 4.3, is simulated. The conductivity of the fins is 5.8×10^7 S/m. A current source is supplied at the middle of the fin gap vertically, with a Gaussian derivative pulse of $-\tau^2 \exp(-(t-t_0)^2/\tau^2)$, with $\tau = 2 \times 10^{-12}$ s, and $t_0 = 4\tau$. The base grid size is $L_c = 0.1$ mm, and the subgrid region, having a grid ratio of $L_c/L_f = 4$, is located

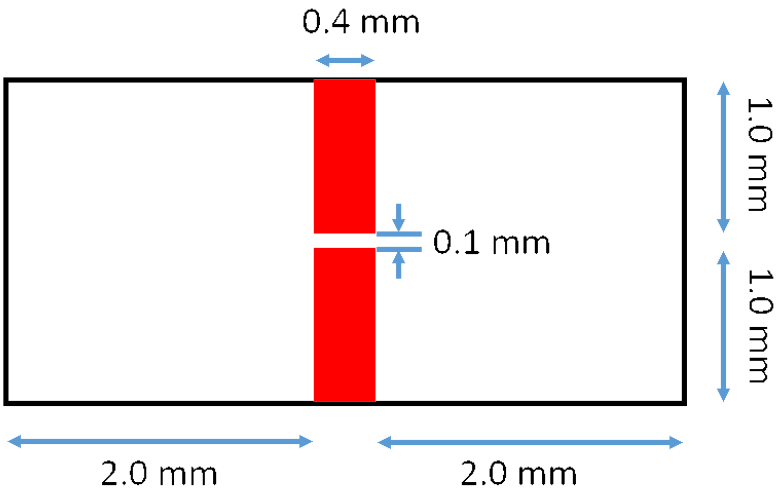


Fig. 4.3. Geometry of a 2-D fin structure.

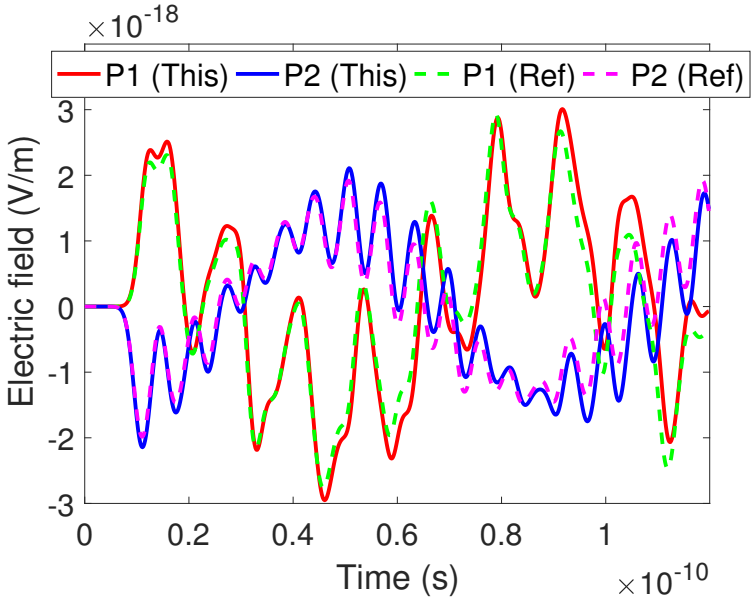


Fig. 4.4. Simulated electric fields of the fin structure.

between the two fins. Without removing the unstable modes, the time step allowed in the subgrid is $dt_f = 4.20 \times 10^{-14}$ s, which is determined by L_f. By removing 32 high modes, dt_f is enlarged to 8.40×10^{-14} s, while dt_c is kept the same as before. The fields simulated from this method at two points, (1, 1.05) and (2.85, 1.5) mm,

are plotted in Fig. 4.4. They agree very well with the reference FDTD solution. The reference FDTD method, using a uniform grid, takes 6.68 s to finish the simulation.

The subgridding method takes 0.84 s without high modes removal. The proposed unconditionally stable method only takes 0.47 s including the time for finding the unstable modes.

4.4 Conclusion

In this work, we analytically identify the unstable modes from the subgrid by combining our previous work in Chap. 2 and Chap. 3. These unstable modes are the root cause of the instability for an explicit time marching. We also develop an error-controlled scheme to determine the number of modes to be removed without sacrificing the accuracy required for time-domain simulation. These modes are then eradicated from the system to enlarge the time step in the subgrid. With the enlarged time step and the flexible local time stepping, the proposed subgridding algorithm can be used to greatly accelerate FDTD simulations when analyzing multiscaled problems.

5. MATRIX-FREE METHOD FOR TRANSIENT MAXWELL-THERMAL CO-SIMULATION IN ARBITRARY UNSTRUCTURED MESHES

5.1 Introduction

Existing multiphysics co-simulation methods require solving a system matrix when handling inhomogeneous materials and irregular geometries discretized into unstructured meshes. In this chapter, a matrix-free method is developed for co-simulating full-wave Maxwell's equations and the thermal diffusion equation in time domain. The method is free of matrix solutions regardless of the shape of the element used for space discretization. A theoretical stability analysis is also developed for the coupled electrical and thermal analysis, which is nonlinear in nature. Numerical experiments on both unstructured tetrahedral and triangular prism element meshes have validated the accuracy and efficiency of the proposed method.

The contents of this chapter have been extracted and revised from the following publication: Kaiyuan Zeng and Dan Jiao, "Matrix-Free Method for Transient Maxwell-Thermal Co-Simulation in Arbitrary Unstructured Meshes," IEEE Transactions on Microwave Theory and Techniques, 2018 [51].

5.2 Equations Governing Electrical-Thermal Co-simulation

The electrical performance of a physical structure is governed by the Maxwell's equations from DC to high frequencies [32]:

$$\nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t}, \tag{5.1}$$

$$\nabla \times \mathbf{H} = \epsilon \frac{\partial \mathbf{E}}{\partial t} + \sigma \mathbf{E} + \mathbf{J}, \tag{5.2}$$

where **E** is electric field intensity, **H** is magnetic field intensity, **J** is input (supply) current density, μ, ϵ, and σ are permeability, permittivity, and conductivity respectively.

The thermal performance is dictated by the well-known thermal diffusion equation [35]:

$$\tilde{\rho} c_p \frac{\partial T}{\partial t} - \nabla \cdot (k \nabla T) = P_{joule} + P_0, \tag{5.3}$$

where k is the thermal conductivity, c_p is the specific heat capacity, $\tilde{\rho}$ denotes the mass density of the material, T is the temperature, P_{joule} represents the heat source

$$P_{joule} = \mathbf{J} \cdot \mathbf{E} = \sigma E^2, \tag{5.4}$$

and P_0 denotes other heat sources. The conductivity is a function of temperature, which obeys

$$\sigma = \frac{\sigma_0}{1 + \alpha(T - T_0)}, \tag{5.5}$$

in which σ_0 is the conductivity at temperature T_0, and α is the temperature coefficient of the material.

Equations (5.1), (5.2), (5.3), and (5.5) can be co-simulated as follows to obtain the electrical and thermal performance of a structure. Starting from an initial temperature, and hence an initial conductivity of the material, (5.1) and (5.2) can be solved to find electric field distribution in the entire structure. This will provide a heat source P_{joule} to (5.3). Eqn. (5.3) can then be simulated to find the temperature distribution. The temperature distribution is subsequently used to update the conductivity of the material through (5.5). Eqns. (5.1) and (5.2) are then simulated again with the updated σ. The entire simulation repeats until the desired time or a steady state is reached.

5.3 Proposed Work

In this section, we present a matrix-solution free method for solving the coupled electrical-thermal equations.

5.3.1 Matrix-Free Time-Domain Method for Solving Maxwell's Equations

Consider a physical structure discretized into either a regular grid or an unstructured mesh consisting of arbitrarily shaped elements. Based on [32], to discretize Faraday's law, we expand the electric field $\mathbf{E}$ in each element by vector bases, yielding $\mathbf{E} = \sum_{j=1}^{m} u_j \mathbf{N}_j$, where u_j is the j-th basis's unknown coefficient, and m is the basis number in each element. First-order vector bases are used as $\mathbf{N}_j$ such that they can produce second-order accurate magnetic fields anywhere to facilitate an accurate discretization of Ampere's law (5.2). Substituting the expansion of $\mathbf{E}$ into Faraday's law (5.1), evaluating $\mathbf{H}$ at the point individuated by the distance vector $\mathbf{r}_{hi}$, and then taking the dot product of the resultant with unit vector $\hat{h}_i$, we obtain

$$\mathbf{S}_e\{u\} = -\text{diag}\left(\{\mu\}\right)\frac{\partial\{h\}}{\partial t}, \tag{5.6}$$

where u denotes a global $\mathbf{E}$-unknown vector of length N_e consisting of all u_j coefficients, and $\mathbf{S}_e$ is a sparse matrix whose ij-th entry is

$$\mathbf{S}_{e,ij} = \hat{h}_i \cdot \{\nabla \times \mathbf{N}_j\}(\mathbf{r}_{hi}), \tag{5.7}$$

and the h is a global H-unknown vector of length N_h, whose i-th entry is $h_i = \mathbf{H}(\mathbf{r}_{hi}) \cdot \hat{h}_i$. The number of nonzero elements in each row of $\mathbf{S}_e$ is the number of basis functions in each element, which is a small constant. The $\text{diag}\left(\{\mu\}\right)$ in (5.6) is a diagonal matrix of magnetic permeability μ. The $\mathbf{H}$-points and directions are chosen along a rectangular loop perpendicular to each E-unknown, and centering the E-unknown, as shown in Fig. 5.1. In this way, the resultant H fields can, in turn, generate desired E accurately. Unlike the FDTD method, here, the $\mathbf{H}$ points and directions do not form a dual mesh. Only a single mesh is needed. No interpolations and projections are required either.

To discretize Ampere's law, we apply the law at $\mathbf{r}_{ei}$ points, and then take the dot product of the resultant with unit vector $\hat{e}_i$ at each point, obtaining

$$\mathbf{S}_h\{h\} = \text{diag}\left(\{\epsilon\}\right)\frac{\partial\{e\}}{\partial t} + \text{diag}\left(\{\sigma\}\right)\{e\} + \{j\}, \tag{5.8}$$

in which $\mathbf{S}_h$ is a sparse matrix of size N_e by N_h. Each row of $\mathbf{S}_h$ has only four nonzero elements, whose column index corresponds to the global index of the four H-points associated with one E-unknown. In (5.8), the diag $(\{\epsilon\})$ and diag $(\{\sigma\})$ are the diagonal matrices of permittivity, and conductivity respectively.

With $\{u\} = \{e\}$, (5.6) and (5.8) are connected, and they can be solved in a leapfrog way which is free of matrix solutions [1]. We can also eliminate $\mathbf{H}$ and solve $\mathbf{E}$ as the following

$$\frac{\partial^2 \{e\}}{\partial t^2} + \text{diag}\left(\left\{\frac{1}{\epsilon}\right\}\right)\frac{\partial \{\sigma e\}}{\partial t} + \mathbf{S}\{e\} = -\text{diag}\left(\left\{\frac{1}{\epsilon}\right\}\right)\frac{\partial \{j\}}{\partial t}, \tag{5.9}$$

where

$$\mathbf{S} = \text{diag}\left(\left\{\frac{1}{\epsilon}\right\}\right)\mathbf{S}_h \text{diag}\left(\left\{\frac{1}{\mu}\right\}\right)\mathbf{S}_e. \tag{5.10}$$

In (5.9), σ is not taken out of the time derivative because it is time dependent in an electrical-thermal co-simulation. Obviously, the matrices in front of the second- and first-order time derivatives in (5.9) are both diagonal. Therefore, an explicit marching of (5.9), such as a central-difference-based time marching [3], is free of matrix solutions.

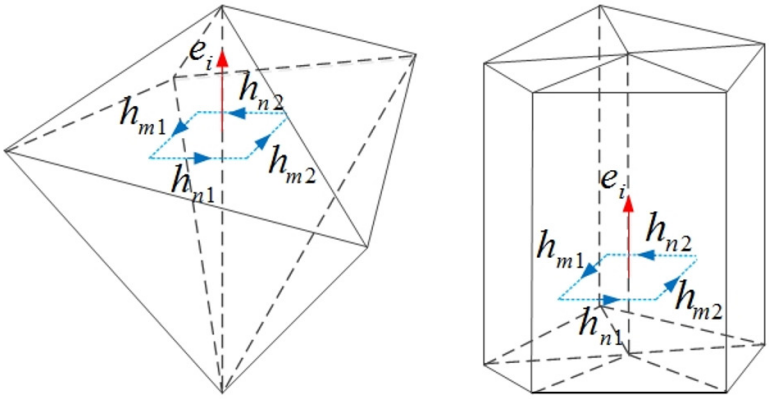


Fig. 5.1. **H** points and directions.

5.3.2 Matrix-Free Time-Domain Method for Solving Thermal Diffusion Equation

Although the thermal diffusion equation is a scalar equation, its matrix-free solution in unstructured meshes does not exist in open literature. Using the finite-element method or other PDE methods, the resultant numerical system involves a sparse matrix to solve. To develop a matrix-free solution of (5.3), we propose to first vectorize the scalar-based (5.3). Although this approach appears to complicate the problem to be solved, the end result is efficient as the number of temperature unknowns to solve remains the same as before.

We append a direction to T, making it a vector $\mathbf{T}$. For example, in (5.3), if we attach a unit vector along z to the right-hand-side heat source $P_{joule} + P_0$, then the $\mathbf{T}$'s z-component solved from a vector-based (5.3) would be the real temperature. To develop a matrix-free solution of (5.3), we also introduce an auxiliary vector $\mathbf{T}_c$ which corresponds to the curl of the $\mathbf{T}$ vector. With the two vector variables, we transform the original thermal diffusion equation (5.3) into the following two vector equations to solve

$$k\nabla \times \mathbf{T} = -\frac{\partial \mathbf{T}_c}{\partial t}, \tag{5.11}$$

$$\nabla \times \mathbf{T}_c = \tilde{\rho} c_p \mathbf{T} - \int (\mathbf{P}_{joule} + \mathbf{P}_0)\, \mathrm{d}t. \tag{5.12}$$

Next, we show the equivalency between (5.3) and the above two equations. Consider a source-free region with uniform thermal constants for convenience. Starting from (5.11) and (5.12), we have

$$\tilde{\rho} c_p \nabla \cdot \mathbf{T} = \nabla \cdot (\nabla \times \mathbf{T}_c) = 0. \tag{5.13}$$

Using the curl of the curl property [52], we obtain

$$\nabla \times \nabla \times \mathbf{T} = \nabla (\nabla \cdot \mathbf{T}) - \nabla^2 \mathbf{T} = -\nabla^2 \mathbf{T}. \tag{5.14}$$

By taking a curl of (5.11) and substituting (5.12) into the resultant, we have

$$\tilde{\rho} c_p \frac{\partial \mathbf{T}}{\partial t} + k\nabla \times \nabla \times \mathbf{T} = \mathbf{P}_{joule} + \mathbf{P}_0. \tag{5.15}$$

The equation above is the same as (5.3) by utilizing the relationship of (5.14). As a result, solving the two vector equations (5.11) and (5.12) simultaneously is equivalent to solving (5.3).

Obviously, (5.11) has the same form as Faraday's law, while (5.12) has a form similar to Ampere's law. Hence, the matrix-free time-domain method can be applied to solve (5.11) and (5.12) without any need for solving a matrix equation. First, we can expand $\mathbf{T}_c$ using a set of first-order vector bases, then evaluate (5.12) at $\mathbf{r}_{ti}$ along direction $\hat{h}_{ti}$ $(i = 1, 2, \ldots, N_t)$. Therefore, (5.12) can be discretized as

$$\mathbf{S}_e\{T_c\} = \text{diag}\left(\{\tilde{\rho}c_p\}\right)\{T\} - \{P\}, \tag{5.16}$$

where $\{P\}$ denotes the vector associated with heat source's time integration. On the other hand, we can discretize (5.11) as

$$\text{diag}\left(\{k\}\right)\mathbf{S}_h\{T\} = -\frac{\partial\{T_c\}}{\partial t}, \tag{5.17}$$

the accuracy of which is guaranteed by (5.16), since the T therein is generated at the points and along the directions that ensure the second-order accuracy of (5.17). In (5.16) and (5.17), both $\mathbf{S}_e$ and $\mathbf{S}_h^T$ are sparse. Their sizes are $N_t \times N_c$ where N_t is the number of $\mathbf{T}$ unknowns, while N_c is the number of $\mathbf{T}_c$ unknowns. The diag$\{k\}$ and diag$\{\tilde{\rho}c_p\}$ are diagonal matrices of k_i, and $\tilde{\rho}_i c_{pi}$ respectively. Vector $\{T\}$ contains all the $\mathbf{T}$ unknowns, while vector $\{T_c\}$ contains all the $\mathbf{T}_c$ unknowns.

Eqns. (5.16) and (5.17) can be solved without any matrix solution using a forward difference scheme. We can also eliminate $\mathbf{T}$ and solve for $\mathbf{T}_c$ first as follows

$$\frac{\partial\{T_c\}}{\partial t} + \mathbf{S}_{t_c}\{T_c\} = \{b\}, \tag{5.18}$$

where

$$\mathbf{S}_{t_c} = \text{diag}\left(\{k\}\right)\mathbf{S}_h\text{diag}\left(\left\{\frac{1}{\tilde{\rho}c_p}\right\}\right)\mathbf{S}_e \tag{5.19}$$

$$\{b\} = -\text{diag}\left(\{k\}\right)\mathbf{S}_h\text{diag}\left(\left\{\frac{1}{\tilde{\rho}c_p}\right\}\right)\{P\}, \tag{5.20}$$

and (5.18) can be discretized in time as the following

$$\{T_c\}^{n+1} = \{T_c\}^n - \Delta t \mathbf{S}_{t_c} \{T_c\}^n + \{b\}^n. \tag{5.21}$$

Once $\{T_c\}$ is solved at each time step, $\{T\}$ can be obtained readily from (5.16). Obviously, no matrix equation needs to be solved in (5.21), thus a linear (optimal) complexity is achieved in computation. Alternatively, we can also eliminate $\mathbf{T}_c$ and directly solve for $\mathbf{T}$ as follows

$$\frac{\partial \{T\}}{\partial t} + \mathbf{S}_t \{T\} = \text{diag}\left(\left\{\frac{1}{\tilde{\rho} c_p}\right\}\right) \{P_{joule} + P_0\}, \tag{5.22}$$

where the i-th entry of $\{P_{joule} + P_0\}$ is the heat source at the i-th temperature point, and

$$\mathbf{S}_t = \text{diag}\left(\left\{\frac{1}{\tilde{\rho} c_p}\right\}\right) \mathbf{S}_e \text{diag}\left(\{k\}\right) \mathbf{S}_h. \tag{5.23}$$

The aforementioned approach obviously is very different from prevailing methods for solving the thermal equation, where the temperature unknown is expanded into certain scalar basis functions, and then the equation is tested also by certain basis functions. The resultant numerical system matrix is not diagonal, and hence must be solved at each time step. In contrast, in the proposed matrix-free method, by introducing a vector-based representation of temperature unknown, and its curl vector, we can interleave the two unknowns in both time and space, and develop an explicit time-domain solution of the thermal diffusion equation that is free of matrix solutions. Apparently, by vectorizing T, we complicate the problem. In fact, the size of the resulting numerical system for T, as shown in (5.22), is the same as before, which is the number of temperature unknowns in the discretized structure.

As mentioned before, the real temperature is only obtained from the direction which was attached to the right-hand-side heat source $P_{joule} + P_0$. A post processing operation for temperature is needed to calculate the cell temperature, while the temperature unknown T is generated at the points and along the directions that ensure the second-order accuracy of (5.17). We pick amplitude of the temperature vector along the defined direction as the real temperature in each cell, and we calculate such

amplitude at the center position of each cell. Such cell temperature will affect the cell conductivity to be used for solving maxwell's equations in the next time step.

5.3.3 System for Electrical-Thermal Co-simulation and Stability Analysis

After using the proposed matrix-free time-domain method, we obtain the following system of equations for electrical-thermal co-simulation

$$\frac{\partial^2 \{e\}}{\partial t^2} + \operatorname{diag}\left(\left\{\frac{1}{\epsilon}\right\}\right)\frac{\partial \{\sigma e\}}{\partial t} + \mathbf{S}\{e\} = - \operatorname{diag}\left(\left\{\frac{1}{\epsilon}\right\}\right)\frac{\partial \{j\}}{\partial t} \tag{5.24}$$

$$\frac{\partial \{T\}}{\partial t} + \mathbf{S}_t\{T\} = \operatorname{diag}\left(\left\{\frac{1}{\tilde{\rho}c_p}\right\}\right)\{\sigma e^2\} \tag{5.25}$$

$$\sigma = \frac{\sigma_0}{1+\alpha(T-T_0)}, \tag{5.26}$$

where P_{joule} is present as the heat source. The above equations are clearly coupled. If we define a global vector of

$$x = \{e\ h\}^T, \tag{5.27}$$

and

$$y = \{T - T_0\}, \tag{5.28}$$

the coupled electrical-thermal system can be rewritten as

$$\begin{aligned} x'(t) = &- \begin{bmatrix} \operatorname{diag}(\{\epsilon\}) & 0 \\ 0 & \operatorname{diag}(\{\mu\}) \end{bmatrix}^{-1} \\ &\begin{bmatrix} \operatorname{diag}\left(\{\frac{\sigma_0}{1+\alpha y}\}\right) & -\mathbf{S}_h \\ \mathbf{S}_e & 0 \end{bmatrix} x + f \\ y'(t) = &-\mathbf{S}_t y + \begin{bmatrix} \operatorname{diag}\left\{\dfrac{\sigma_0}{\tilde{\rho}c_p(1+\alpha y)}\right\} & 0 \end{bmatrix} x^2, \end{aligned} \tag{5.29}$$

where superscript $'$ denotes a time derivative, and $f = \{-j/\epsilon\ 0\}^T$ is the source term.

Eqn. (5.29) constitutes a nonlinear system of equations. The stability of a nonlinear system can be analyzed by finding its Jacobian matrix at each solution point

(x_n, y_n). Take the equilibrium point as an example, this is the point (x_0, y_0) at which the right hand side of (5.29) becomes zero, when the source f is vanished. This point can be readily found as $x_0 = 0,\ y_0 = 0$. Evaluating the Jacobian matrix of (5.29) at (x_0, y_0), we obtain

$$\mathcal{J} = \begin{bmatrix} -\begin{bmatrix} \mathrm{diag}\left(\{\sigma_0\epsilon^{-1}\}\right) & -\mathrm{diag}\left(\{\epsilon^{-1}\}\right)\mathbf{S}_h \\ \mathrm{diag}\left(\{\mu^{-1}\}\right)\mathbf{S}_e & 0 \end{bmatrix} & 0 \\ 0 & -\mathbf{S}_t \end{bmatrix}. \tag{5.30}$$

In a regular grid, $\mathbf{S}_t$ is positive semi-definite, as $\mathbf{S}_h = \mathbf{S}_e^T$. As for the first diagonal block in (5.30), its eigenvalues λ and eigenvectors v satisfy

$$-\begin{bmatrix} \mathrm{diag}\left(\{\sigma_0\epsilon^{-1}\}\right) & -\mathrm{diag}\left(\{\epsilon^{-1}\}\right)\mathbf{S}_h \\ \mathrm{diag}\left(\{\mu^{-1}\}\right)\mathbf{S}_e & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \lambda \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}. \tag{5.31}$$

Substituting the second subsystem of equations of the above into the first, we obtain

$$\lambda^2 v_1 + \lambda \mathrm{diag}\left(\left\{\frac{\sigma_0}{\epsilon}\right\}\right) v_1 + \mathbf{S} v_1 = 0, \tag{5.32}$$

where $\mathbf{S}$ is the same as shown in (5.10).

In a regular gird, $\mathbf{S}$ is positive semi-definite. Since $\mathrm{diag}\left(\left\{\frac{\sigma_0}{\epsilon}\right\}\right)$ is positive semi-definite as well, the eigenvalues of (5.32) have a non-positive real part. Hence, the eigenvalues of the Jacobian matrix $\mathcal{J}$ have a non-positive real part. Therefore, based on the stability theory of a nonlinear system [53], an explicit marching of (5.29) is stable at (x_0, y_0).

The Jacobian matrix of (5.29) is time dependent. At an arbitrary solution point (x_n, y_n), the Jacobian matrix of (5.29) can be written as

$$\mathcal{J}_n = \begin{bmatrix} -\begin{bmatrix} \mathrm{diag}\left(\{\sigma_n\epsilon^{-1}\}\right) & -\mathrm{diag}\left(\{\epsilon^{-1}\}\right)\mathbf{S}_h \\ \mathrm{diag}\left(\{\mu^{-1}\}\right)\mathbf{S}_e & 0 \end{bmatrix} & 0 \\ 2\left[\mathrm{diag}\left\{\frac{\sigma_n}{\tilde{\rho}c_p}\right\} \quad 0\right] x_n & -\mathbf{S}_t \end{bmatrix}, \tag{5.33}$$

in which σ_n, and x_n represent the σ, and x at the n-th time step respectively. Since the above matrix is triangular, the eigenvalues are determined by the eigenvalues of

the diagonal blocks. Hence, the property of the eigenvalues is the same as that of (5.30). Thus, an explicit time-marching is stable.

From (5.30) and (5.33), it can also be seen that the Maxwell's subsystem and the thermal one are governed by different eigenvalues, with the magnitude of the latter much smaller. To see this point clearly, (5.29) can be split into the following two subsystems at an arbitrary n-th time step

$$x'_n = \mathcal{J}_n(1,1)x_n + f_n \tag{5.34}$$

$$y'_n = \mathcal{J}_n(2,2)y_n + \left[\text{diag}\left\{\frac{\sigma_n}{\tilde{\rho}c_p}\right\} \quad 0\right] x_n^2, \tag{5.35}$$

where the eigenvalues of $\mathcal{J}_n(1,1)$ (the first diagonal block of $\mathcal{J}_n$) are related to Maxwell's equations, and the eigenvalues of $\mathcal{J}_n(2,2) = -\mathbf{S}_t$ are solely related to the thermal equation. Hence, the choice of time step can be made different. In this work, we employ a forward difference to discretize the thermal equation. Performing a stability analysis of the forward differencing of (5.35), it can be readily found that the time step should satisfy

$$\Delta t_t < \frac{2Re(\lambda_{ther})}{|\lambda_{ther}|^2}, \tag{5.36}$$

where λ_{ther} is the eigenvalue of $\mathbf{S}_t$ whose magnitude is the largest. The $|\lambda_{ther}|$ can be analytically estimated as

$$|\lambda_{ther}| = \frac{k}{(\tilde{\rho}c_p) \times (4/\Delta_{min}^2)}, \tag{5.37}$$

where Δ_{min} is the smallest space step.

In unstructured meshes, using the proposed matrix-free method, $\mathbf{S}$ is not symmetric, however, with the time-marching method of [32], the explicit marching of (5.34) is equally stable. The resulting time step, Δt_m, is the same as that used in a conventional explicit time marching of Maxwell's equations such as the CFL condition. It satisfies

$$\Delta t_m < \frac{2}{\sqrt{|\xi|_{max}}}, \tag{5.38}$$

where $|\xi|_{max}$ is the maximum magnitude of the eigenvalues of $\mathbf{S}$, which is proportional to the inverse of the square of the smallest space step. For the examples simulated

in this chapter, it is found that Δt_t is much larger than Δt_m. Hence, the thermal equation can be simulated using a much larger step, and the conductivity σ in between the two time instants of the thermal equation is interpolated to obtain that at the time instants for solving Maxwell's equations.

The procedure of the co-simulation algorithm is illustrated in Fig. 5.2. The electrical and thermal analyses are coupled through the temperature-dependent material properties. When two different time steps are used, within one step of a thermal simulation, there exist $ratio = \Delta t_t / \Delta t_m$ steps of electrical simulations, and n_{total} represents the predefined maximum number of simulation steps.

5.4 Numerical Results

In this section, we first validate the proposed method in performing a thermal simulation as well as an electrical-thermal co-simulation. We then apply the proposed method to solve a number of coupled electrical-thermal problems in a variety of structures and meshes. Both FDTD and the time-domain finite element method (TDFEM) are used as the reference methods for comparison.

The conductivity σ_0 of copper is 5.8×10^7 S/m. The heat conduction parameters for copper are $k = 398$ W/(m $\cdot$ K), $c_p = 386$ J(kg $\cdot$ K), $\tilde{\rho} = 8930$ kg/m^3 and $\alpha = 0.0039$. All these simulations are conducted with Intel Xeon CPU E5-2690 v2 @ 3.00 GHz having 128 GB memory.

5.4.1 Thermal Analysis: Node Basis and Vector Basis

First, we examine the correctness of the proposed matrix-free method for thermal analysis, which involves a vectorization of a scalar thermal equation. Three methods, node basis based finite difference method, node basis based FEM method and vector based FDTD method, are used as references. In this example, we consider a piece of copper plane whose side length is 0.3 m. The temperature on one side of the plane is 200 °C while being 100 °C on other sides. In Fig. 5.3(a), we plot the temper-

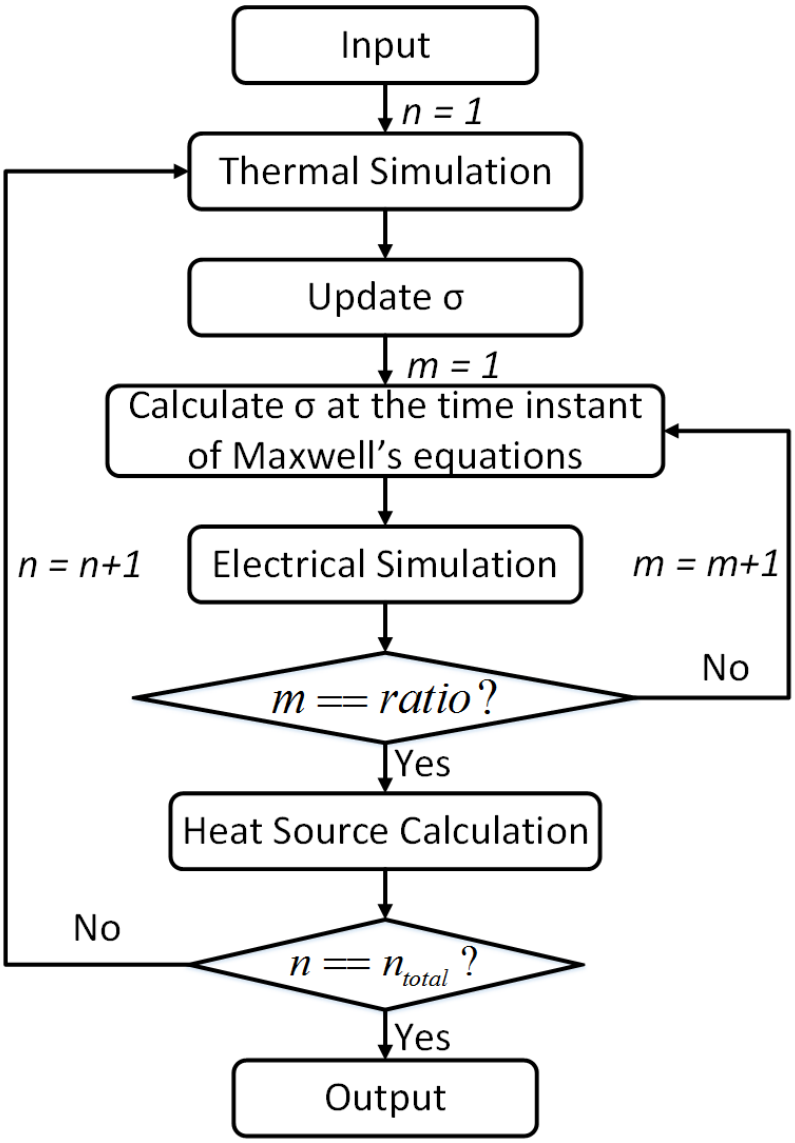
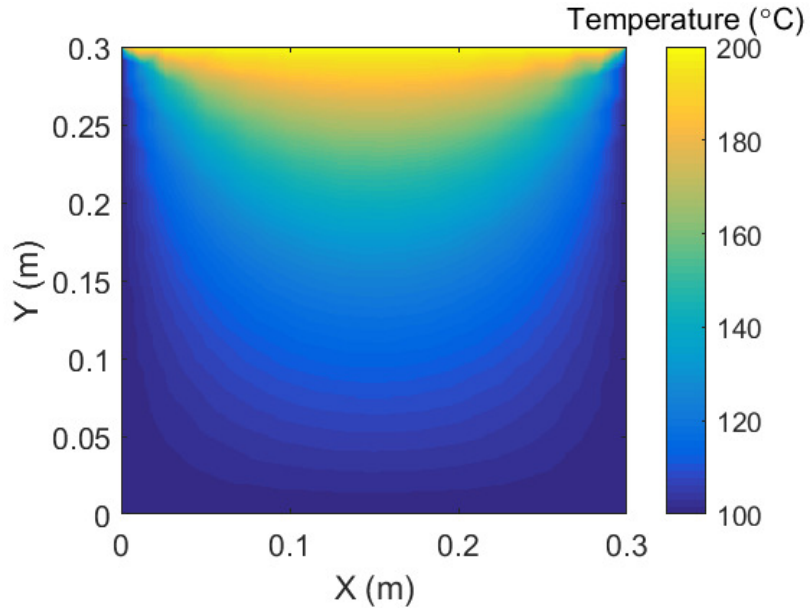
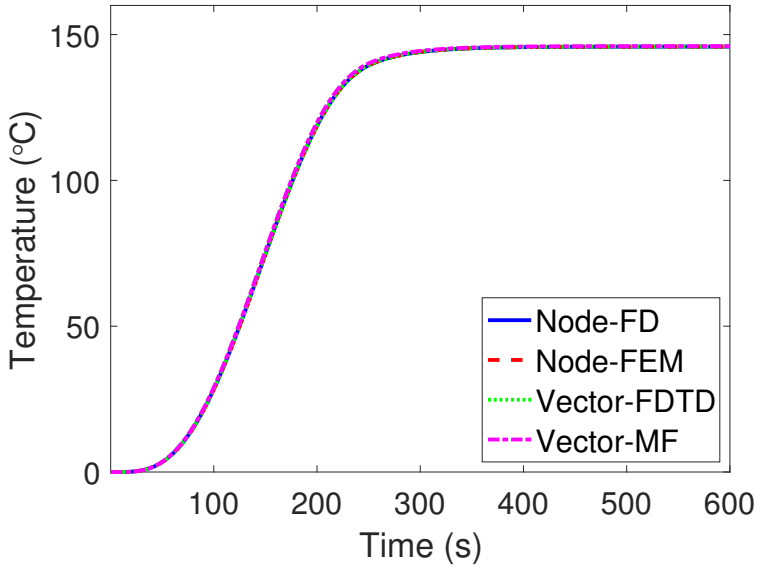


Fig. 5.2. Flowchart of the co-simulation algorithm.

ature distribution across the whole plane at the steady state. In Fig. 5.3(b), the temperatures generated from the proposed method and the three reference methods at point (0.2443, 0.2443) m are plotted versus time. Clearly, the temperature at this point gradually grows, and finally reaches its expected steady-state value. The results from four different methods are on top of each other, validating the correctness of the proposed vectorization of the thermal equation and its matrix-free solution.



(a)



(b)

Fig. 5.3. Thermal simulation of a 2-D problem: (a) Temperature distribution at steady state. (b) Transient temperature at an observation point with four different methods.

5.4.2 Thermal Analysis in a Tetrahedral Mesh

We then simulate a heat conduction problem, which only requires solving the thermal equation. A copper conductor of size $1 \times 0.5 \times 0.75$ m^3 is considered. It is discretized into a tetrahedral mesh, which is illustrated in Fig. 5.4. The temperature on the outermost boundary of the cube is set to be 100 °C. To guarantee the stability of the proposed method, we choose a time step of $\Delta t = 0.8$ s.

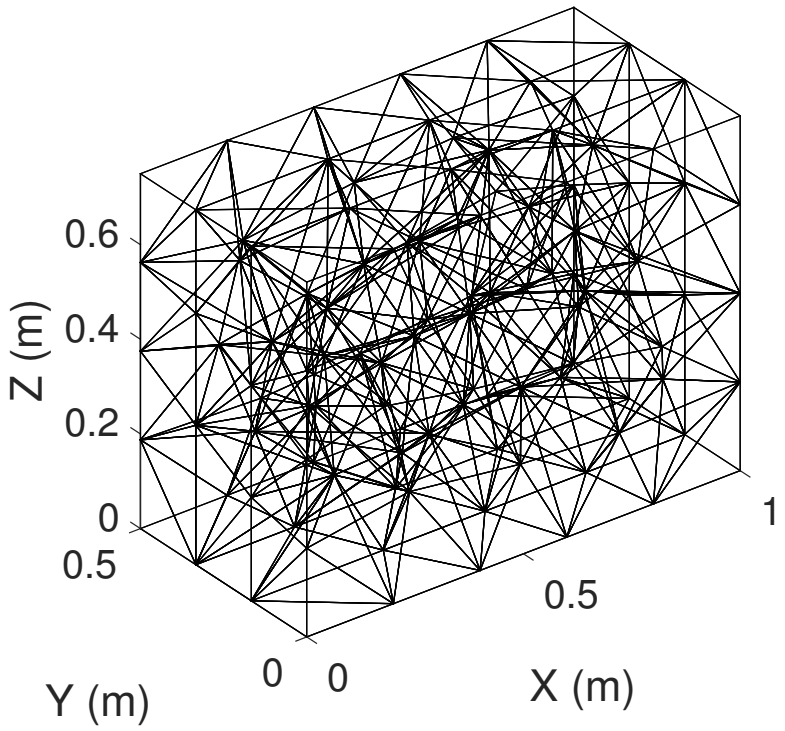


Fig. 5.4. 3-D tetrahedron discretization of a copper conductor.

In Fig. 5.5(a), the temperature at point (0.4747, 0.2197, 0.6826) m is plotted versus time. Clearly, the temperature at this point gradually grows, and finally reaches its expected steady-state value of 100 °C. It is also shown to agree well with the temperature obtained from the TDFEM. In Fig. 5.5(b), we plot the entire solution error as compared to the TDFEM, which is shown to be 2.78% when temperature reaches the steady state. This error is measured by $\|\{T\} - \{T\}_{ref}\|/\|\{T\}_{ref}\|$, where

$\{T\}$ contains the temperatures of all cells simulated from the proposed method, and $\{T\}_{ref}$ is the reference solution from the TDFEM. As can be seen from Fig. 5.5(b), the proposed method is accurate for solving the thermal diffusion equation in an unstructured tetrahedral mesh. The larger difference at the early time is due to a sudden jump from zero to nonzero in the source setup, the high frequency components of which are captured by numerical methods in a different way. We also compare the computational efficiency of the proposed method with the TDFEM whose discretization results in a similar number of unknowns. It is shown that the proposed method takes 0.53 s to finish the whole simulation, whereas the TDFEM costs 1.24 s.

5.4.3 Electrical-thermal Co-simulation of Copper Cube in Tetrahedral Mesh

Next, we validate the proposed co-simulation method using the same copper conductor illustrated in Fig. 5.4. The conductor is excited by a current source, whose waveform is $2\times10^4(t-t_0)\exp(-(t-t_0)^2/\tau^2)$, with $\tau = 6.0\times10^{-9}$ s, and $t_0 = 4\tau$. The time step used in the Maxwell part of the co-simulation is $\Delta t_m = 2.4\times10^{-11}$ s, while that in the thermal part is $\Delta t_t = 2.4\times10^{-9}$ s $= 100\Delta t_m$. The temperature on the outermost boundary of the cube is set to be 0 °C. The simulated temperatures in all of the tetrahedral elements are plotted versus time in Fig. 5.6(a). The relative error of these temperatures, measured by $\|\{T\}-\{T\}_{ref}\|/\|\{T\}_{ref}\|$, is plotted in Fig. 5.6(b), where $\{T\}$ contains the temperatures of all elements simulated from the proposed method, and $\{T\}_{ref}$ is the reference TDFEM solution. Good accuracy is observed across the entire window. In Fig. 5.7(a), we plot the electric field at an observation point (0.1971, 0.0556, 0.0662) m obtained from the co-simulation in comparison with the result obtained from a Maxwell-only simulation. It is clear that the thermal effect is observable in this example, and the co-simulation is able to capture the combined electrical-thermal effects. We also plot the TDFEM results in Fig. 5.7(a). Excellent agreement is observed between TDFEM and the proposed method. Furthermore, we

have compared the entire electric field solution with those of the TDFEM, by evaluating the total error of $\|\{e\} - \{e\}_{ref}\|/\|\{e\}_{ref}\|$. The $\{e\}$ containing all e_i unknowns is from the proposed method, and $\{e\}_{ref}$ is from the TDFEM solution. The error is shown in Fig. 5.7(b) as a function of time, validating the accuracy of the proposed method for electrical-thermal co-simulation. Again, to demonstrate the efficiency of the proposed method, we compare our matrix-free method with the TDFEM method. In the co-simulation, TDFEM need to do LU factorization at each time when the material properties are updated, while the proposed method has no factorization cost since it is free of matrix solution. In contrast to the 81.39 s cost by TDFEM, the proposed method only takes 6.85 s to finish the whole simulation.

5.4.4 U-type Conductor Discretized into Tetrahedral Elements

A 3-D u-type resistor is discretized into tetrahedral elements as shown in Fig. 5.8. These tetrahedral elements are generated from brick elements so that FDTD can be used for comparison. The σ_0 of all metal is 5.8×10^7 S/m, and the surrounding material has relative permittivity of 4. The two ends of the conductor are excited by a current source, which is depicted by the red line in Fig. 5.8. The boundary conditions are PMC (perfect magnetic conductor) on the left, right, front, back sides, and PEC (perfect electric conductor) on the top and at the bottom, with the temperature set to be 0 °C. The source waveform is a Gaussian derivative pulse of $5.7 \times 10^3(t - t_0)\exp(-(t - t_0)^2/\tau^2)$, with $\tau = 1 \times 10^{-10}$ s, and $t_0 = 4\tau$. In Fig. 5.9, we plot the electric field at point (0.5, 0.75, 0.5) μm along a y-direction edge in the conductor obtained with the Maxwell-thermal co-simulation versus that from a Maxwell-only simulation. With peak cell temperature of 82.17 °C, the thermal effect on electrical performance can be clearly observed. For validation purpose, the electric field simulated at the same point along the same direction from the FDTD method is also plotted in Fig. 5.9. Good agreement is observed.

5.4.5 Electrical-thermal Co-simulation of Coaxial Cylinder in Prism Mesh

We have examined the capability of the proposed method in handling irregular prism meshes as well. This example has an irregular triangular prism mesh, the top view of which is shown in Fig. 5.10. The structure has two layers of triangular prism elements (into the paper) with each layer being 0.05 m thick. The discretization results in 3092 edges and 1038 triangular prisms. The conductor is excited by a current source, whose waveform is $2.5 \times 10^4 (t - t_0)\exp(-(t - t_0)^2/\tau^2)$, with $\tau = 5.0 \times 10^{-8}$ s, and $t_0 = 4\tau$. The temperature on both the innermost and the outermost boundary is set to be 0 °C. The time step used in both the Maxwell part and the thermal part of the co-simulation is $\Delta t_m = 2.0 \times 10^{-10}$ s. In Fig. 5.11, we plot the electric field at an observation point (0.1845, 0.7067, 0.0250) m obtained from the co-simulation in comparison with the result obtained from a Maxwell-only simulation. The thermal effect is clearly observable in this example, and the co-simulation is able to capture the combined electrical-thermal effects. We also plot the TDFEM results in Fig. 5.11. Excellent agreement is observed between TDFEM and the proposed method.

5.4.6 Lossy Package Inductor with Triangular Prism Elements

In this example, we simulate a package inductor made of lossy conductors of initial conductivity 5.8×10^7 S/m, which is embedded in a dielectric material of relative permittivity 3.4. Its geometry and material parameters are illustrated in Fig. 5.12. The inductor is discretized into layers of triangular prism elements. The top view of the mesh is shown in Fig. 5.13. The boundary conditions are PMC (perfect magnetic conductor) on the left, right, front, back sides, and PEC (perfect electric conductor) on the top and at the bottom, with the temperature set to be 0 °C. A current source is launched at one end of the inductor, which is marked with blue color. The source waveform is a Gaussian derivative pulse of $0.5 \times 10^4 (t - t_0)\exp(-(t - t_0)^2/\tau^2)$, with $\tau = 0.5 \times 10^{-10}$ s, and $t_0 = 4\tau$. In Fig. 5.14, we plot the electric field along z-direction at point $(-600, -25, 37.5)$ μm in the inductor obtained with the Maxwell-thermal co-

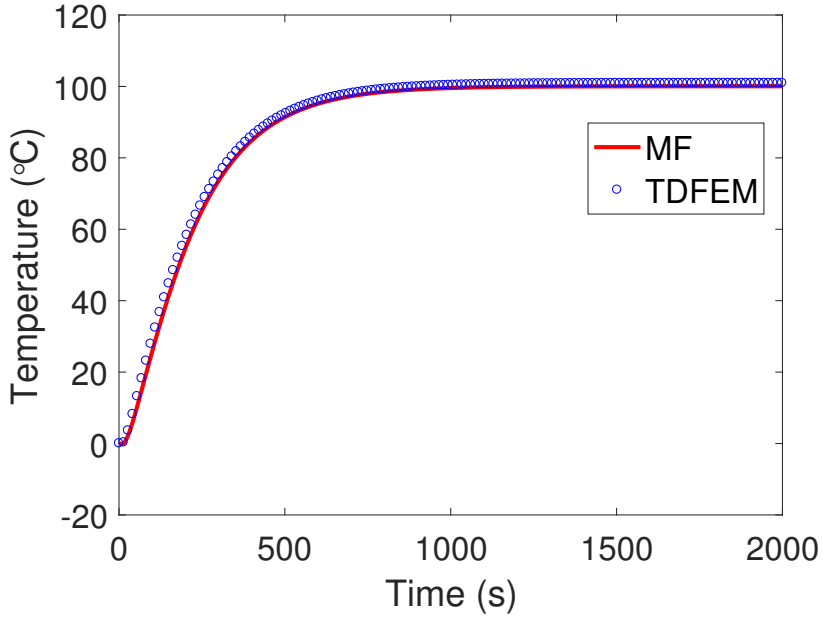
simulation versus that from a Maxwell-only simulation. With peak cell temperature of 51.3 °C, the thermal effect on electrical performance can be clearly observed. For validation purpose, the electric field simulated from the TDFEM is also plotted for comparison. Good agreement is observed.

5.4.7 3-D On-Chip Power Grid Discretized into Tetrahedral Mesh

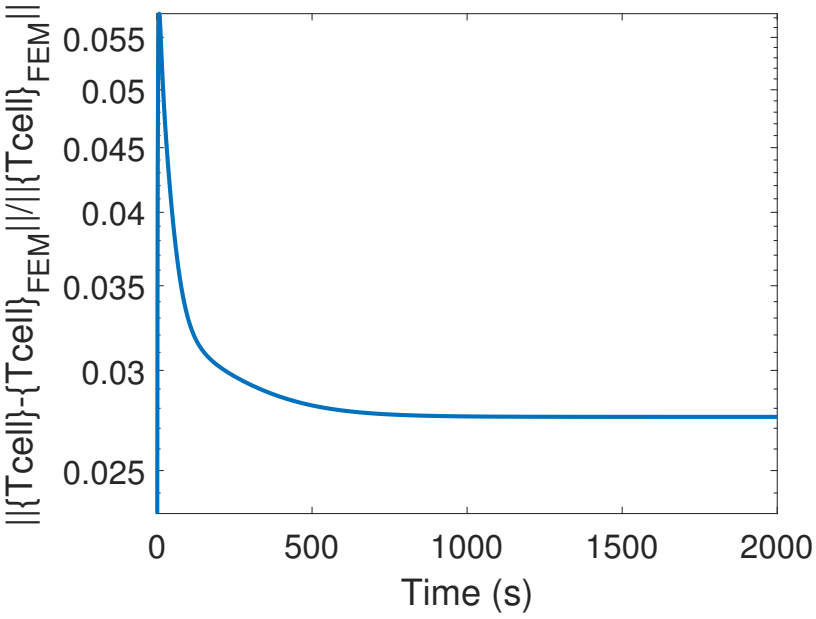
Next, we simulate a 3-D on-chip power grid, illustrated in Fig. 5.15(a) and 5.15(b). The power grid is discretized into a tetrahedral mesh to model non-uniform materials and process variations with fewer unknowns. The σ_0 of all metal is 5.8×10^7 S/m, and that of the silicon substrate is 10^4 S/m. The boundary conditions are PMC (perfect magnetic conductor) on the left, right, front, back sides, and PEC (perfect electric conductor) on the top and at the bottom, with the temperature set to be 0 °C. The near end between a pair of power and ground wires in layer M1 is excited by a current source, which is depicted by the red line in Fig. 5.15(a). The source waveform is a Gaussian derivative pulse of $3.8 \times 10^3(t - t_0)\exp(-(t - t_0)^2/\tau^2)$, with $\tau = 1 \times 10^{-10}$ s, and $t_0 = 4\tau$. In Fig. 5.16, we plot the electric field along y-direction at point (6, 0.75, 1.5) μm in M1 layer obtained with the Maxwell-thermal co-simulation versus that from a Maxwell-only simulation. TDFEM results are also shown for comparison. With peak cell temperature of 78.84 °C, the thermal effect on electrical performance can be clearly observed. If using TDFEM to handle the same tetrahedral mesh and using the same time step for the Maxwell and thermal simulations, the TDFEM costs 47.1 s per time step, while the proposed method costs 0.29 s only. If using two different time steps for Maxwell and thermal simulations and only update conductivity at thermal time step, the CPU time cost at one thermal time step, which includes one thermal simulation and 100 steps of Maxwell's simulations, is 12.17 s in the proposed method, and 102.54 s for the TDFEM, demonstrating the efficiency of the proposed method.

5.5 Conclusion

In this work, we develop a matrix-free algorithm for solving full-wave Maxwell's equations and the thermal diffusion equation simultaneously in time domain. The thermal equation is vectorized to develop a matrix-free solution without increasing the problem size. The matrix-free property is achieved independent of the element shape used for discretizing Maxwell's or thermal equations. The stability of the coupled nonlinear system of equations is also analyzed in detail, and found to be ensured with a correct choice of time step in explicit time marching. Numerical experiments have demonstrated the accuracy and efficiency of the proposed matrix-free method for simulating thermal as well as coupled electrical-thermal problems.

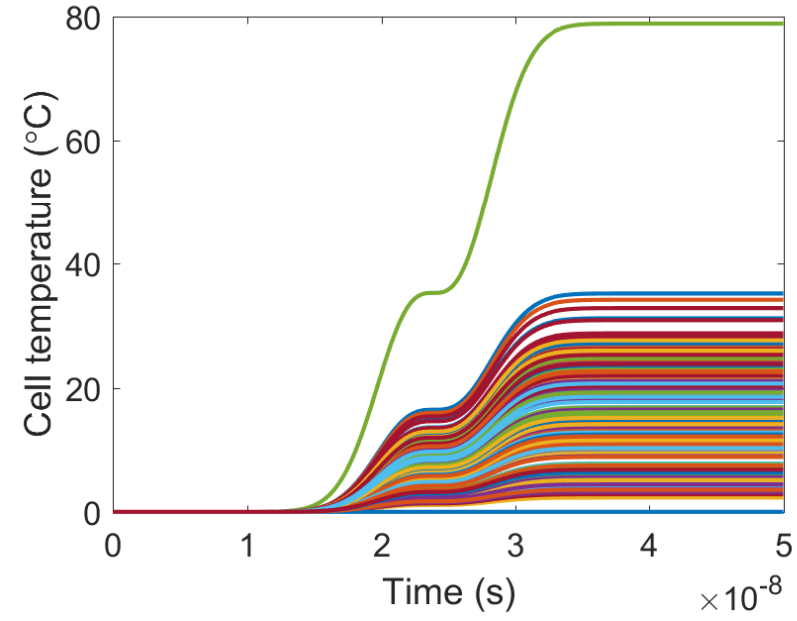


(a)

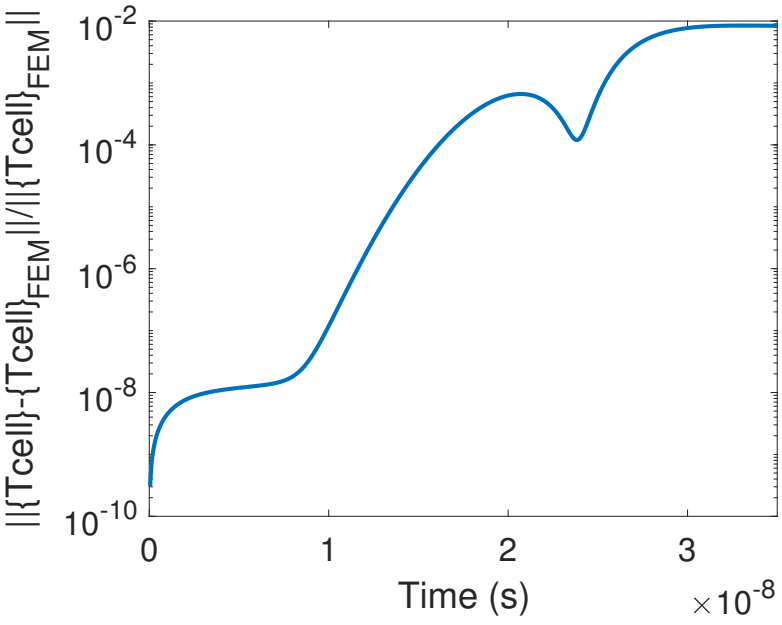


(b)

Fig. 5.5. Thermal simulation of a 3-D problem: (a) Transient temperature at one observation point. (b) Entire solution error of the transient temperature at *all* points.

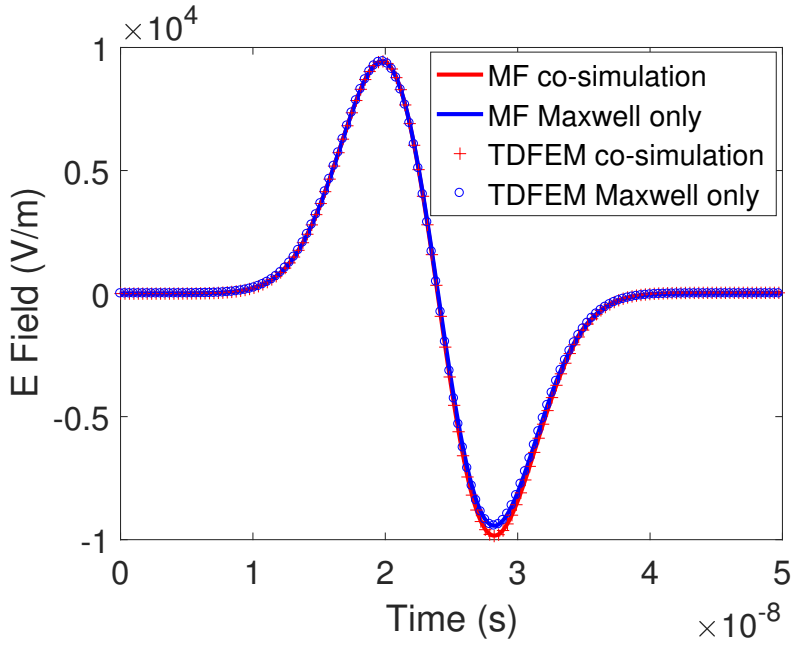


(a)

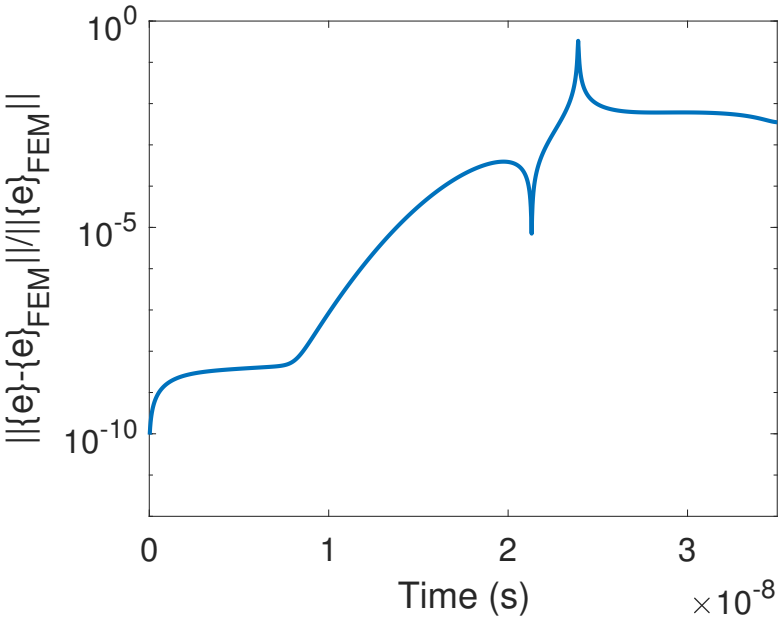


(b)

Fig. 5.6. Copper cube co-simulation: (a) Temperature v.s. time at all points. (b) Entire T solution error as a function of time.



(a)



(b)

Fig. 5.7. Copper cube co-simulation: (a) Simulated electric field at one point. (b) Entire **E** solution error as a function of time.

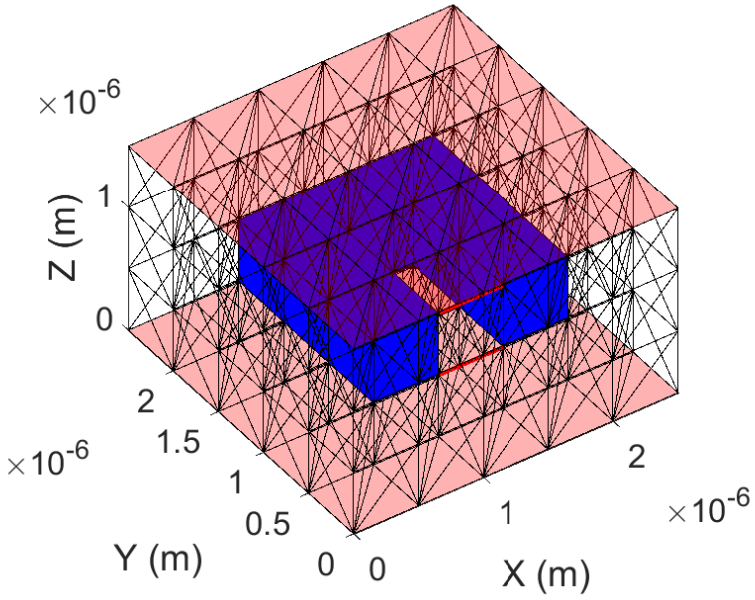


Fig. 5.8. 3-D view of a u-type resistor.

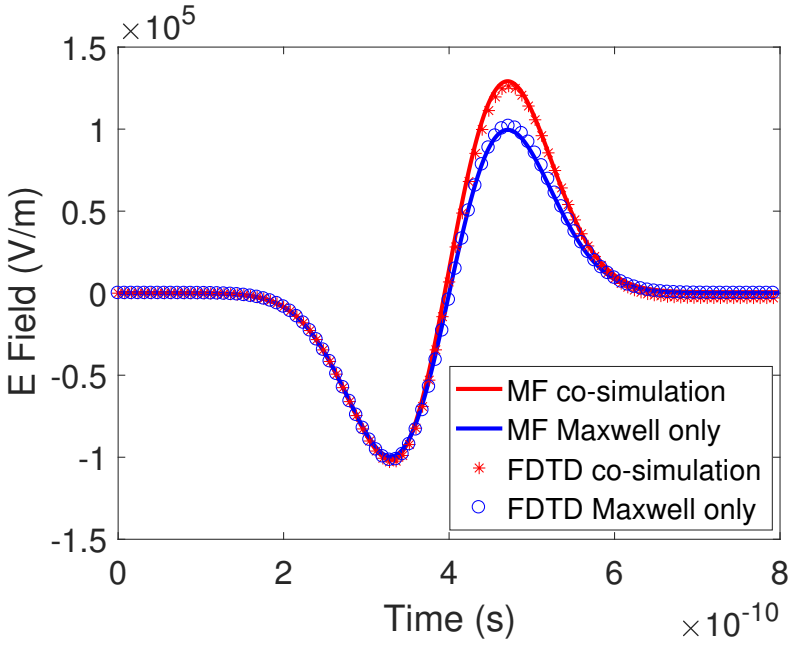


Fig. 5.9. U-type conductor electrical-thermal co-simulation: electric field at an observation point.

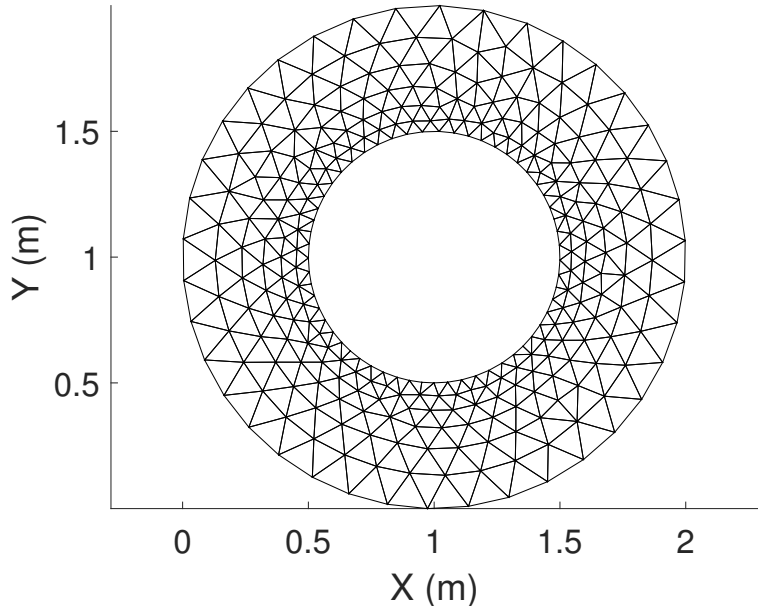


Fig. 5.10. Top view of the triangular prism mesh of an coaxial cylinder structure.

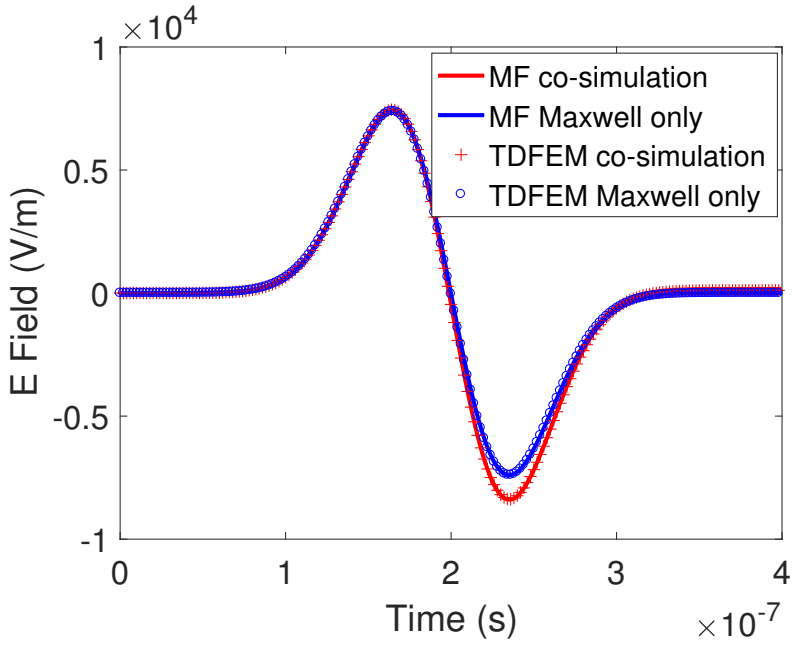


Fig. 5.11. Simulated electric field at one point in coaxial cylinder co-simulation.

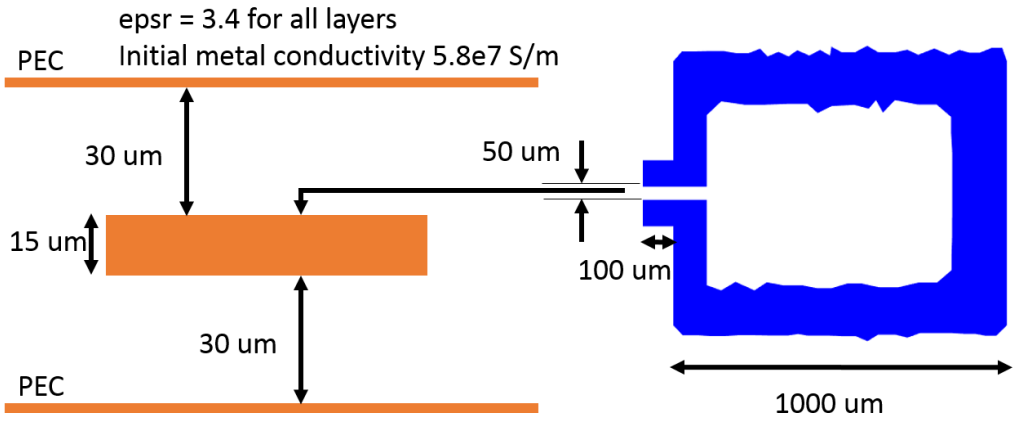


Fig. 5.12. Illustration of materials and geometry of a package inductor.

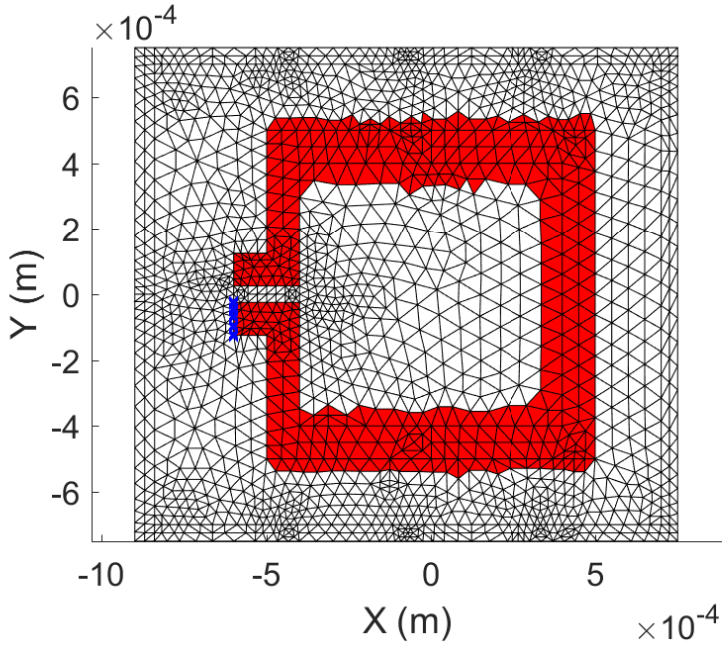


Fig. 5.13. Top view of the triangular prism element mesh.

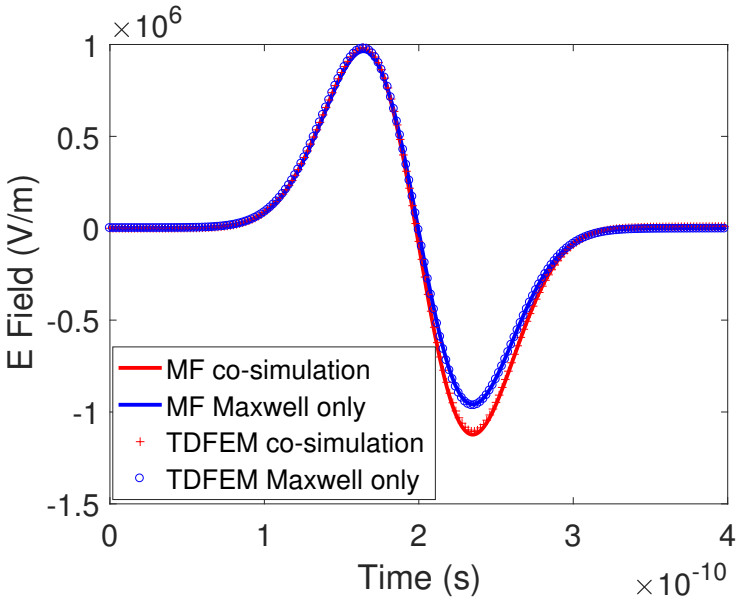
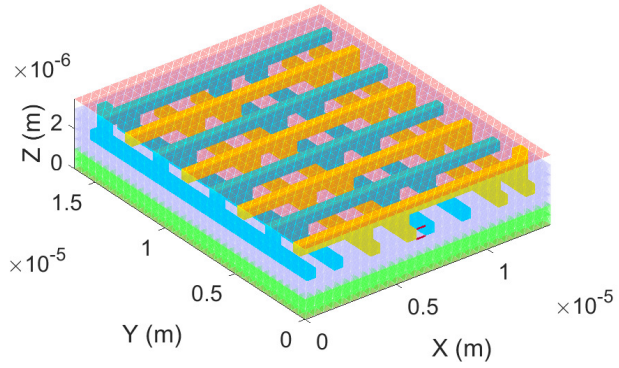
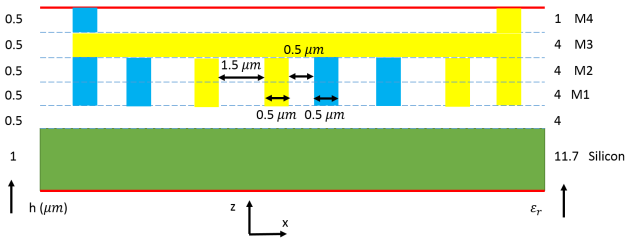


Fig. 5.14. Package inductor electrical-thermal co-simulation: electric field at an observation point.



(a)



(b)

Fig. 5.15. On-chip power grid: (a) 3-D view. (b) Geometry and cross-sectional view.

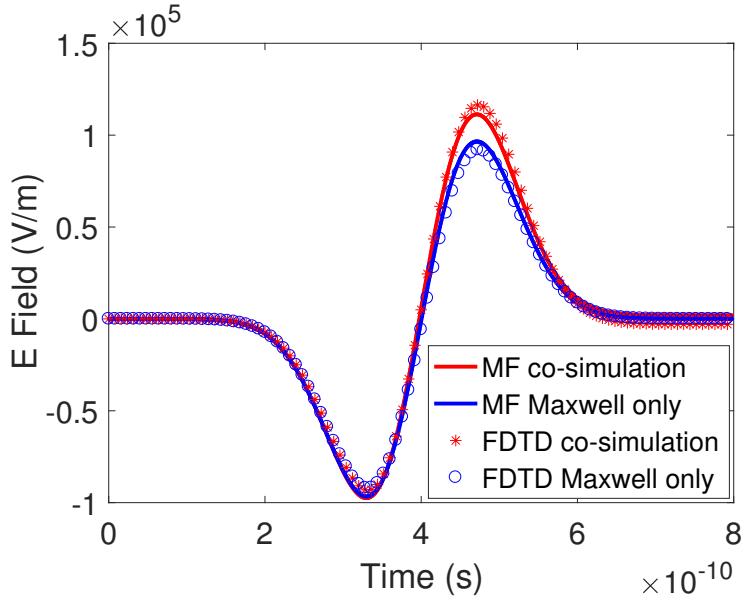


Fig. 5.16. Power grid electrical-thermal co-simulation: electric field at an observation point.

6. FREQUENCY-DOMAIN METHOD HAVING A DIAGONAL MASS MATRIX IN ARBITRARY UNSTRUCTURED MESHES FOR EFFICIENT ELECTROMAGNETIC ANALYSIS

6.1 Introduction

In Chap. 5 we present how to use matrix-free time-domain method solving electromagnetic problems in time domain. In this chapter, as a counterpart, a new frequency-domain method having a diagonal mass matrix is developed for analyzing general electromagnetic problems. The matrix-free time-domain method is independent of the element shape used for discretization, and it has a diagonal mass matrix in nature. Although the mass matrix is only one component of a frequency-domain system matrix, we show that the diagonal property of the mass matrix can be utilized to develop a fast solution. Numerical results have validated the accuracy and efficiency of the proposed new method.

Part of the contents of this chapter has been extracted and revised from the following publication: Frequency-domain Method Having a Diagonal Mass Matrix in Arbitrary Unstructured Meshes for Efficient Electromagnetic Analysis," 2017 IEEE International Symposium on Antennas and Propagation and USNC/URSI National Radio Science Meeting [54].

6.2 Proposed Method

We first find the modes that make the spectral radius of the system matrix greater than 1. We then deduct them directly from the system matrix. As a result, an iterative solution of the updated system matrix can converge in a small number of

iterations. The condition number can also be flexibly controlled by choosing which set of modes to remove. Since the mass matrix is diagonal, no matrix solutions are required in the whole solution procedure. Only a small number of sparse matrix-vector multiplications need to be performed. After solving the updated system matrix, we add the contribution from the deducted modes back to obtain the true solution, the cost of which is negligible.

Consider a general electromagnetic problem discretized into arbitrarily shaped elements. Based on [32], we have following equations in time-domain:

$$\mathbf{S}_e e = -\text{diag}(\{\mu\})h', \tag{6.1}$$

$$\mathbf{S}_h h = \text{diag}(\{\epsilon\})e' + j, \tag{6.2}$$

$$e'' + \mathbf{S}\{e\} = -\text{diag}(\{1/\epsilon\})j', \tag{6.3}$$

$$\mathbf{S} = \text{diag}(\{\frac{1}{\epsilon}\})\mathbf{S}_h\text{diag}(\{\frac{1}{\mu}\})\mathbf{S}_e, \tag{6.4}$$

where e is a global vector of **E**-field unknowns whose length is N_e, and h is a global vector of **H**-field unknowns, whose number is N_h. Eqn. (6.1) represents a discretization of Faraday's law, and (6.2) is a discretization of Ampere's law. The superscript $'$ denotes a time derivative, $\text{diag}(\{\mu\})$, $\text{diag}(\{\epsilon\})$ are, respectively, diagonal matrices of permeability, and permittivity, and j is a current source vector. Eliminating h unknowns, we can obtain a second-order equation (6.3) from (6.1) and (6.2) for e, whose stiffness matrix is shown in (6.4). It is evident that the mass matrix in (6.3) is naturally diagonal.

The aforementioned matrix-free method in time domain can be readily converted to its counterpart in frequency domain as the following

$$(-\omega^2\mathbf{I} + \mathbf{S})e = b, \tag{6.5}$$

where ω is an angular frequency, $\mathbf{I}$ is an identity matrix, which is different from the mass matrix in an FEM method, and $\mathbf{S}$ represents a discretized curl-curl operator. Although the mass matrix is diagonal, the combined system matrix is not. To develop

a fast solution to (6.5), we first did a theoretical study on the eigenvalue solution of $\mathbf{S}$ since its eigenvectors constitute a complete space to represent the field solution. In a general unstructured mesh, $\mathbf{S}$ resulting from the matrix-free method is unsymmetrical. Hence, its eigenvalues can be both real- and complex-valued. The smallest one is $\lambda_{min} = 0$, and the largest eigenvalue magnitude, $|\lambda|_{max}$, is proportional to $(\pi^2 c^2)/{s_{min}}^2$, where s_{min} is the smallest mesh size, and c is the speed of light. Since generally the mesh size is chosen no greater than half of the wavelength in a full-wave analysis, a relative relationship of $\lambda_{min} < \omega^2 < |\lambda|_{max}$ can be deduced. For a typical choice of ten points per wavelength, $|\lambda|_{max}$ is approximately $25\omega^2$. As a result, the system shown in (6.5) is indefinite, whose spectral radius is greater than one. The condition number can also be large, rendering an iterative solution difficult to converge [55].

To develop a fast solution, firstly we rewrite (6.5) as

$$(-\omega_0^2\mathbf{I} + \mathbf{S} + \omega_0^2\mathbf{I} - \omega^2\mathbf{I})e = b \tag{6.6}$$

where ω_0^2 is chosen to be slightly larger than $|\lambda|_{max}$. The above can be denoted in short as

$$(\mathbf{B} - \mathbf{A})e = b \tag{6.7}$$

with $\mathbf{A} = -\mathbf{S} + \omega_0^2\mathbf{I}, \mathbf{B} = (\omega_0^2 - \omega^2)\mathbf{I}$. The solution of (6.7) is governed by a new eigenvalue problem of

$$\mathbf{A}x = \lambda_{new}\mathbf{B}x = (\omega_0^2 - \omega^2)\lambda_{new}x \tag{6.8}$$

whose eigenvalues can be written as $\lambda_{new} = (\omega_0^2 - \lambda)/(\omega_0^2 - \omega^2)$. When $|\omega_0^2 - \lambda| < \omega_0^2 - \omega^2$, $0 < |\lambda_{new}| < 1$; and when $|\omega_0^2 - \lambda| > \omega_0^2 - \omega^2$, $|\lambda_{new}| > 1$. Hence, the original smallest eigenvalues of $\mathbf{S}$ now become the largest ones. If we deduct the eigenmodes whose $|\lambda_{new}| > 1$ from $(\omega_0^2 - \omega^2)^{-1}\mathbf{A}$, the remaining eigenvalues of $(\omega_0^2 - \omega^2)^{-1}\mathbf{A}$ would satisfy $0 < |\lambda_{new}| < 1$, making the solution converge within a small number of iterations. By (6.6), we flip the eigenvalue spectrum because the number of modes is smaller in the range of $|\omega_0^2 - \lambda| > \omega_0^2 - \omega^2$. Hence, the number of modes to be

deducted is smaller. Since the mass matrix is diagonal, and $\mathbf{A}$ is sparse, finding the k largest eigenpairs of (6.8) only costs $O(k)$ sparse matrix-vector multiplications.

Let the eigenmodes whose $|\lambda_{new}| > 1$ be $\mathbf{U}_h$. Since $\mathbf{S}$ is not symmetrical, the eigenvectors of $(\omega_0^2 - \omega^2)^{-1}\mathbf{A}$ are not orthogonal. We hence orthogonalize $\mathbf{U}_h$ to obtain $\mathbf{V}_h$. The complete solution e can be expanded as $e = e_h + e_l = \mathbf{V}_h y_h + \mathbf{V}_l y_l$, where $\mathbf{V}_l$ is orthogonal to $\mathbf{V}_h$. Then we update (6.7) to the following new system of equations

$$((\omega_0^2 - \omega^2)\mathbf{I} - \mathbf{A} + \mathbf{V}_h\mathbf{V}_h^H\mathbf{A})e_l = b. \tag{6.9}$$

After obtaining e_l from (6.9), we perform $e_l = e_l - \mathbf{V}_h\mathbf{V}_h^H e_l$ to ensure the solution is free of $\mathbf{V}_h$ modes. To find e_h, we multiply (6.5) by $\mathbf{V}_h^H$, then we obtain $\mathbf{Q}y_h = b'$, where $b' = \mathbf{V}_h^H[b - (-\omega^2\mathbf{I} + \mathbf{S})e_l]$, and $\mathbf{Q} = \mathbf{V}_h^H(-\omega^2\mathbf{I} + \mathbf{S})\mathbf{V}_h$. This is a small system of size k (the number of $\mathbf{V}_h$ modes). The total solution can then be obtained as $e = e_h + e_l = \mathbf{V}_h y_h + e_l$, which is the same as that obtained by solving (6.6) as it is.

To prove the field solution obtained from the proposed method is the same as that of (6.7), we can substitute $e = \mathbf{V}_h y_h + e_l$ into (6.7), and multiply the result by $\mathbf{V}_l^H$. Since $\mathbf{V}_h$ is orthogonalized from eigenvector matrix $\mathbf{U}_h$, $\mathbf{V}_l^H((\omega_0^2 - \omega^2)\mathbf{I} - \mathbf{A})\mathbf{V}_h = 0$ is true. Hence, we obtain $\mathbf{V}_l^H((\omega_0^2 - \omega^2)\mathbf{I} - \mathbf{A})e_l = \mathbf{V}_l^H b$. Multiplying both sides by $\mathbf{V}_l$, and recognizing $\mathbf{V}_l\mathbf{V}_l^H = \mathbf{I} - \mathbf{V}_h\mathbf{V}_h^H$, we obtain

$$((\omega_0^2 - \omega^2)\mathbf{I} - \mathbf{A} + \mathbf{V}_h\mathbf{V}_h^H\mathbf{A})e_l = (\mathbf{I} - \mathbf{V}_h\mathbf{V}_h^H)b, \tag{6.10}$$

which is the same solution as obtained from (6.9) and (3.20).

To prove (6.9) has a spectral radius less than 1, we can rewrite it as as $(\omega_0^2 - \omega^2)(\mathbf{I} - \mathbf{Y}_0)$, where $\mathbf{Y}_0 = \mathbf{V}_l\mathbf{V}_l^H(\omega_0^2 - \omega^2)^{-1}\mathbf{A}$. Let the eigenvectors of $(\omega_0^2 - \omega^2)^{-1}\mathbf{A}$ be $\mathbf{U} = [\mathbf{U}_h, \mathbf{U}_l]$. Since $(\omega_0^2 - \omega^2)^{-1}\mathbf{A} = \mathbf{U}\Lambda_{new}\mathbf{U}^{-1}$, and $\mathbf{V}_l^H\mathbf{U}_h = 0$, $\mathbf{Y}_0$ can be written as $\mathbf{Y}_0 = \mathbf{V}_l\mathbf{V}_l^H[\mathbf{U}_h\Lambda_{new,h}(\mathbf{U}^{-1})_h + \mathbf{U}_l\Lambda_{new,l}(\mathbf{U}^{-1})_l] = \mathbf{V}_l\mathbf{V}_l^H[\mathbf{U}_l\Lambda_{new,l}(\mathbf{U}^{-1})_l]$, where $(\mathbf{U}^{-1})_{h/l}$ denotes the rows of $\mathbf{U}^{-1}$ corresponding to the $\Lambda_{new,h/l}$ part. The spectral radius of $\mathbf{Y}_0$ is bounded by that of $\Lambda_{new,l}$, which satisfies $|\lambda_{new}| < 1$. Furthermore, it can be controlled by the choice of $\mathbf{V}_h$.

To summarize, in the proposed method, we change the spectrum of the original system (6.5) to a new representation (6.9), and then obtain the solution to the original problem by adding back the contribution of the deducted components. One point worth mentioning is that the original system matrix $\mathbf{S}$ has a nullspace whose size can grow with unknown size N, and this nullspace can be bypassed in the GMRES procedure. Specifically, the zero eigenvalue of (6.5) is related to the largest eigenvalue ξ of (6.7), and the ξ is analytically known as $(\omega_0^2)/(\omega_0^2 - \omega^2)$, which is slightly larger than 1. Let λ_r be the largest magnitude of eigenvalue of $\mathbf{Y}_0$, which has been proved to be less than 1. Therefore, we just need to deduct $\mathbf{V}_h$ modes whose magnitudes of eigenvalues are between ξ and λ_r, and by controlling λ_r to a desired constant number, we can efficiently solve the problem by GMRES in a small number of iterations.

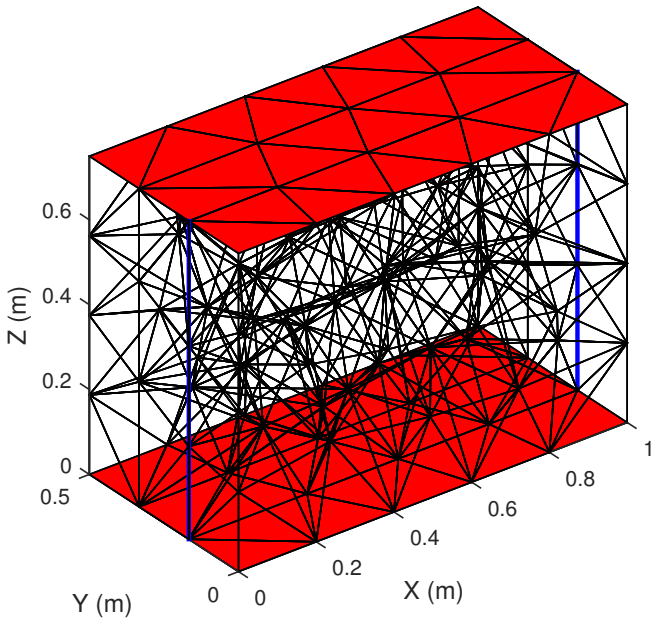


Fig. 6.1. Illustration of the tetrahedron mesh of the 3-D parallel plate.

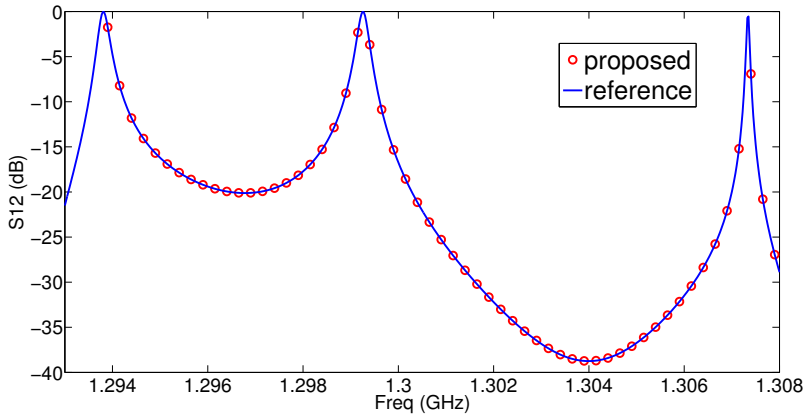


Fig. 6.2. S-parameters of the 3-D parallel plate.

6.3 Numerical Results

First, a 3-D parallel plate waveguide discretized into tetrahedral elements is simulated. Fig. 6.2 shows the $\mathbf{S}_{12}$ computed from the proposed method when $|\lambda_{new}| > 0.99$ modes are removed. Excellent agreement with the reference result from [32] is observed. The iteration number of the proposed method is 39 to reach an accuracy of 1e-4. In contrast, a brute-force GMRES solution requires a large iteration number of 1,815 for the same accuracy. To find $\mathbf{V}_h$ modes, a traditional FEM-based method requires 77.27 s, while the proposed method only takes 5.23 s due to the diagonal mass matrix. Next, a 2-mm-long microstrip line discretized into tetrahedral elements is simulated, with $|\lambda_{new}| > 0.99$ modes removed. Fig. 6.4 compares the simulated $\mathbf{S}_{12}$ with reference data. The proposed method has an iteration number of 40 to reach 1e-4 accuracy while a brute-force solution requires 11,801 iterations.

6.4 Conclusion

The matrix-free time-domain method has a diagonal mass matrix in nature. Although the mass matrix is only one component of a frequency-domain system matrix, in this chapter we show that the diagonal property of the mass matrix can be utilized

to develop a fast solution. Numerical experiments have demonstrated the accuracy and efficiency of the proposed new frequency-domain method for solving Maxwell's equations.

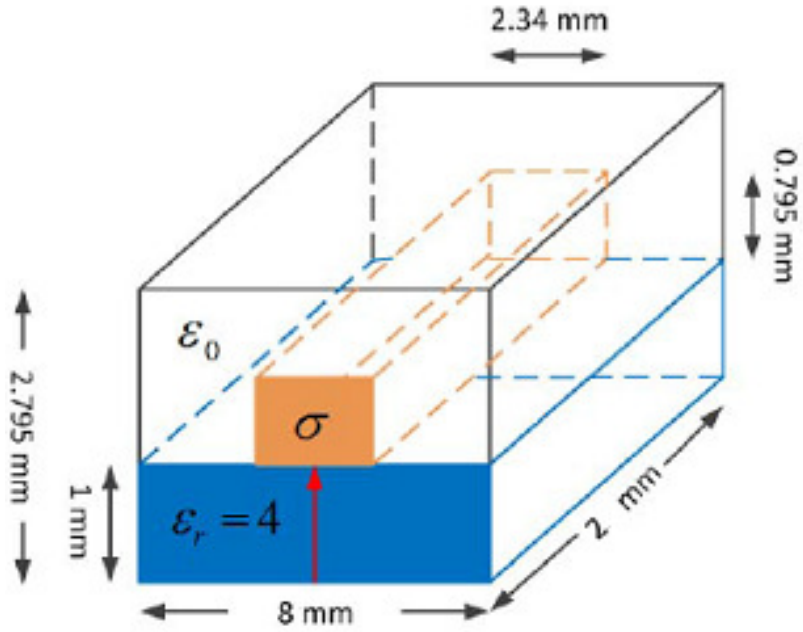


Fig. 6.3. Structure of a lossy microstrip line.

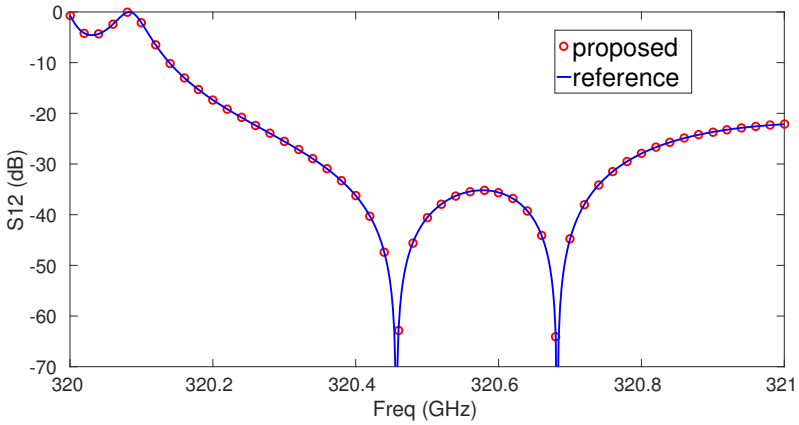


Fig. 6.4. S-parameters of the microstrip line.

7. A TRULY EXPLICIT MATRIX-FREE TIME-DOMAIN METHOD IN UNSTRUCTURED MESHES AND ITS APPLICATION TO EXPLICIT SIMULATION OF GENERAL UNSYMMETRICAL NUMERICAL SYSTEMS

7.1 Introduction

In this chapter, we overcome the barrier of simulating an unsymmetrical numerical system whose traditional explicit simulation is known to be absolutely unstable, and successfully develop a truly explicit matrix-free time-domain (MFTD) method, requiring no matrix solution in arbitrary unstructured meshes. The core of this innovation is a new explicit time marching scheme for simulating an unsymmetrical system, whose stability is theoretically guaranteed. Meanwhile, the accuracy of the time marching is not sacrificed; and the time step size allowed by a traditional explicit method is not reduced to ensure the stability of the new explicit scheme. As a result, we greatly improve the computational efficiency of the MFTD method without compromising its accuracy. Extensive numerical experiments on both unstructured triangular and tetrahedral meshes, and comparisons with the original MFTD method and analytical results have validated the accuracy, efficiency, and stability of the proposed new explicit MFTD method.

In addition to the MFTD, unsymmetrical numerical systems are also frequently encountered in other numerical methods such as subgridding methods, non-orthogonal FDTD methods, etc. In this work, we show the proposed new explicit method is a general method for stably simulating unsymmetrical systems. Hence, it can be utilized in other unsymmetrical methods to guarantee their stability in explicit time domain simulation. Numerical results are provided to demonstrate this capability.

Part of the contents of this chapter has been extracted and revised from the following manuscript: Kaiyuan Zeng and Dan Jiao, "A Truly Explicit Matrix-Free Time-Domain Method in Unstructured Meshes and Its Application to Explicit Simulation of General Unsymmetrical Numerical Systems with Guaranteed Stability," submitted to the IEEE Transactions on Microwave Theory and Techniques.

7.2 Analysis of the Problem

In this section, we provide the background of this work, and also analyze the problem encountered in the transient simulation of an unsymmetrical numerical system. We start from the MFTD in unstructured meshes, then proceed to a state-of-the-art unsymmetrical FDTD subgridding method, and then general unsymmetrical methods for solving Maxwell's equations.

7.2.1 MFTD in Unstructured Meshes and Underlying Unsymmetrical System

Given an irregular structure with inhomogeneous materials discretized into an arbitrary unstructured mesh, based on [32], the electric field $\mathbf{E}$ in each element is expanded into vector bases whose order is no less than one such that the resulting curl $\mathbf{E}$ is not a constant in the element. As a result, the $\mathbf{H}$ field's space dependence in a single element can be captured, as its order would be higher than zero.

The expansion into vector bases yields $\mathbf{E} = \sum_{j=1}^{m} e_j \mathbf{N}_j$, where e_j is the j-th basis's unknown coefficient, and m is the basis number in each element. Using the modfied higher-order vector bases shown in [32], each e_j denotes $\mathbf{E}(\mathbf{r}_{ej}) \cdot \hat{e}_j$, i.e., $\mathbf{E}$ field at point $\mathbf{r}_{ei}$ along the direction of unit vector $\hat{e}_j$. Substituting the expansion of $\mathbf{E}$ into Faraday's law, evaluating $\mathbf{H}$ at a point $\mathbf{r}_{hi}$, and then taking the dot product of the resultant with a unit vector $\hat{h}_i$, we obtain

$$-\text{diag}\left(\{\mu\}\right) \frac{\partial\{h\}}{\partial t} = \mathbf{S}_e\{e\}, \tag{7.1}$$

where e denotes a global **E**-unknown vector of length N_e consisting of all e_j coefficients, $\mathbf{S}_e$ is a sparse matrix denoting a discretized curl operation on **E**, and the h is a global H-unknown vector of length N_h, whose i-th entry is $h_i = \mathbf{H}(\mathbf{r}_{hi}) \cdot \hat{h}_i$. The diag $(\{\mu\})$ in (7.1) denotes a diagonal matrix of permeability μ.

Then we choose the **H**-points, $\mathbf{r}_{hi}$, and the **H**-directions, $\hat{h}_i$, along a rectangular loop perpendicular to each **E**-unknown, and centering the **E**-unknown. In this way, the resultant **H** fields can, in turn, generate **E** required in (7.1) accurately at next time step, via the discretization of Ampere's law as follows,

$$\text{diag}\left(\{\epsilon\}\right)\frac{\partial\{e\}}{\partial t} + \{j\} = \mathbf{S}_h\{h\} \tag{7.2}$$

in which $\{j\}$ denotes a current source vector, $\mathbf{S}_h$ is a sparse matrix of size N_e by N_h, representing a curl operation on **H**. Each row of $\mathbf{S}_h$ obviously has only four nonzero elements, whose column index corresponds to the global index of the four **H**-points associated with one **E**-unknown. In (7.2), the diag $(\{\epsilon\})$ is a diagonal matrix of permittivity. From (7.1) and (7.2), we can also eliminate **H** and solve **E** as the following

$$\text{diag}\left(\{\epsilon\}\right)\frac{\partial^2\{e\}}{\partial t^2} + \mathbf{S}\{e\} = -\frac{\partial\{j\}}{\partial t}, \tag{7.3}$$

where

$$\mathbf{S} = \mathbf{S}_h \text{diag}\left(\left\{\frac{1}{\mu}\right\}\right)\mathbf{S}_e. \tag{7.4}$$

.

When $\mathbf{S}_e = \mathbf{S}_h^T$, **S** is symmetrical and positive semi-definite, thus having non-negative real eigenvalues only. Here, however, since the curl of **E** and the curl of **H** are carried out in a completely different way, the $\mathbf{S}_e \neq \mathbf{S}_h^T$, and the resultant **S** is highly unsymmetrical.

7.2.2 Unsymmetrical FDTD Subgridding

In [15], an unsymmetrical FDTD subgridding method is developed, which utilizes the strength of an unsymmetrical discretization of curl operators in accuracy to build

an accurate and flexible subgridding method for arbitrary grid ratios and subgrid arrangements. In this method, a row vector is generated for every patch in the grid, regardless of a 2- or 3-D grid. This row vector represents a curl of **E** operation in the patch, which produces the time derivative of the normal **H** at the patch center. Let this row vector be denoted by $\mathbf{S}_e^{(i)}$ for patch i. A column vector, denoted by $\mathbf{S}_h^{(i)}$, is also generated for every patch in the grid. It represents how the normal **H** field at the patch center is used to generate the electric fields. The product of $\mathbf{S}_h^{(i)}$ and $\mathbf{S}_e^{(i)}$ makes a rank-1 matrix. The sum of this rank-1 matrix divided by the μ of the patch over all patches in the grid makes the total **S**. In a regular grid, $\mathbf{S}_e^{(i)}$ has only four nonzero elements, and $\mathbf{S}_h^{(i)} = (\mathbf{S}_e^{(i)})^T$. Hence, the symmetric and positive semi-definiteness of **S** is ensured.

However, when subgrids exist, as shown in [15], the $\mathbf{S}_h^{(i)} = (\mathbf{S}_e^{(i)})^T$ is not satisfied. For the subgridding method in [15], three irregular patch types, hence three irregular kinds of $\mathbf{S}_e^{(i)}$ and $\mathbf{S}_h^{(i)}$ are identified. The first type of patches has one or multiple sides residing on the interface between the base grid and the subgrid. For such a patch, $\mathbf{S}_e^{(i)}$ has more than four nonzero elements, since the subgrid electric fields along the base-grid sides are used to perform the curl of **E** to obtain the **H** field at the patch center. Therefore, the number of nonzero entries in $\mathbf{S}_e^{(i)}$ is the number of base-grid and subgrid electric fields along the loop of the i-th patch. The $\mathbf{S}_h^{(i)}$ has also more than four nonzero elements. These entries are located at the rows corresponding to the electric fields generated by the **H** at the i-th patch's center. The value of each entry is the coefficient of the **H** field used to generate the corresponding **E** field, which is not the same as the coefficient of the **E** used to generate **H**. In addition, the **H** field at patch i's center is used to generate not only the **E** fields along the four sides of patch i, but also some other **E** fields for interpolation accuracy. Therefore, the number of nonzero elements in $\mathbf{S}_h^{(i)}$ is also different from that of $\mathbf{S}_e^{(i)}$. As a result, the $\mathbf{S}_h^{(i)}$ is by no means the transpose of $\mathbf{S}_e^{(i)}$. The same is true for the other two irregular patch types. The final system matrix **S** in [15] is again a sum of the rank-1 matrix over

every patch, but now mixed with both symmetric and unsymmetric rank-1 matrices, and hence being unsymmetrical.

7.2.3 General Unsymmetrical Systems Arising from Other Methods

The unsymmetrical system is also observed in many other methods such as the non-orthogonal FDTD methods and other subgridding methods different from [15]. Although their ways of discretizing Maxwell's equations are different, and hence the resultant $\mathbf{S}_e$ and $\mathbf{S}_h$ are different from those in the MFTD, they all can be cast in the format of (7.1) and (7.2). As long as the two curl operations are not made reciprocal of each other, which is the case to ensure the accuracy of the space discretization in an nonorthogonal grid or a grid with subgrids, the resultant numerical system is unsymmetrical.

In a second-order system where only one field unknown is solved, and in a single mesh setup, like that in the finite element method, when the basis function for expanding the field unknown is different from the testing function, the resulting system matrix is also unsymmetrical.

The final second-order system of equations for solving Maxwell's equations using various methods can be written into the following form:

$$\mathbf{D}_\epsilon \frac{\partial^2 \{e\}}{\partial t^2} + \mathbf{S}\{e\} = b(t), \tag{7.5}$$

where $\mathbf{D}_\epsilon$ is associated with the permittivity, which can be either diagonal or nondiagonal, symmetric or unsymmetrical, depending on the method used for discretizing Maxwell's equations. The $\mathbf{S}$ represents a discretized $\nabla\mu^{-1}\nabla\times$ operator. For an unsymmetrical treatment of the curl of $\mathbf{E}$ and the curl of $\mathbf{H}$, $\mathbf{S}$ is unsymmetrical. Using different basis and testing functions, $\mathbf{S}$ is also unsymmetrical. Certainly, there exist other scenarios such as non-reciprocal materials, which can make $\mathbf{S}$ unsymmetrical as well.

7.2.4 Stability Analysis of an Unsymmetrical System

The time marching of (7.1) and (7.2) is usually performed in a leap-frog way, like the one in the FDTD method. This is, in fact, equivalent to a central-difference based explicit time marching of (7.3) or (7.5). This point can be readily proved by eliminating one field unknown from the leap-frog based time discretization of (7.1) and (7.2). However, when $\mathbf{S} \neq \mathbf{S}^T$, such an explicit time marching is absolutely unstable. To see this clearly, we first discretize (7.5) using a central-difference based explicit time marching, obtaining

$$\mathbf{D}_\epsilon \left(\{e\}^{n+1} - (2\{e\}^n - \{e\}^{n-1} \right) + \Delta t^2 \mathbf{S} \{e\}^n = \{b\}^n, \tag{7.6}$$

where $\{f\}$ represents the terms related to the excitation.

Removing the source term since it has nothing to do with the stability, and performing a z-transform of (7.6), we obtain

$$z^2 - (2 - \Delta t^2 \lambda) z + 1 = 0, \tag{7.7}$$

where Δt is the time step, and λ is the eigenvalue of $\mathbf{D}_\epsilon^{-1}\mathbf{S}$. If $\mathbf{S}$ is unsymmetrical, the $\mathbf{D}_\epsilon^{-1}\mathbf{S}$ can have complex-valued and even negative eigenvalues. From (7.7), it can be seen that the two roots satisfy

$$|z_1 z_2| = 1 \tag{7.8}$$

Since neither $|z_1| = 1$ nor $|z_2| = 1$ is satisfied for complex λ, because

$$z_{1,2} = \frac{2 - \Delta t^2 \lambda \pm \sqrt{\Delta t^2 \lambda \left(\Delta t^2 \lambda - 4 \right)}}{2}, \tag{7.9}$$

One of the roots must be greater than 1 in magnitude. As a result, the traditional explicit scheme for simulating an unsymmetrical system is absolutely unstable.

To overcome this stability problem, in [32], a backward-difference based discretization of (7.3) is performed, which requires a lower bound of the time step

$$\Delta t > \frac{2 \left| \mathrm{Im}(\sqrt{\lambda}) \right|}{\left(\left| \sqrt{\lambda} \right|^2 \right)}, \tag{7.10}$$

where $\mathrm{Im}(\cdot)$ denotes the imaginary part of $(\cdot)$. To avoid solving a matrix in time marching, a series expansion is used to derive an explicit inverse of the system matrix. The $\{e\}$ at the $(n+1)$th time step is hence computed as

$$\{e\}^{n+1} = \left(\mathbf{I} - \tilde{\mathbf{M}} + \tilde{\mathbf{M}}^2 - \cdots + (-\tilde{\mathbf{M}})^k\right)\{f\}, \tag{7.11}$$

where $\{f\} = 2\{e\}^n - \{e\}^{n-1} - \Delta t^2 \mathbf{D}_\epsilon^{-1} b^{n+1}$ and $\tilde{\mathbf{M}} = \Delta t^2 \mathbf{D}_\epsilon^{-1}\mathbf{S}$. For the series expansion to converge, an upper bound of the time step is imposed

$$\Delta t < \frac{1}{\sqrt{\|\mathbf{D}_\epsilon^{-1}\mathbf{S}\|}}, \tag{7.12}$$

which is similar to the time step used in a leap-frog scheme. However, such a series expansion with k terms requires k sparse matrix-vector multiplications. Although k is not large, which is, in general, around 10, the resulting time marching is not as efficient as performing only one matrix-vector multiplication in the right hand side, like that in a traditional explicit FDTD method.

7.3 New Explicit Method for Simulating Unsymmetrical Systems with Guaranteed Stability

Based on the stability analysis in Section 7.2.4, it appears that there is no way forward to make an explicit time marching stable for simulating an unsymmetrical system. However, we found the following explicit method worked out.

The method can be used to solve both first-order Maxwell's equations, and the second-order one in a stable fashion. Consider the first-order system. We discretize (7.1) in the following way:

$$\mathbf{S}_e(2\{e\}^n - \{e\}^{n-1}) = -\mathrm{diag}\left(\{\mu\}\right)\frac{\{h\}^{n+\frac{1}{2}} - \{h\}^{n-\frac{1}{2}}}{\Delta t} \tag{7.13}$$

where we change the $\{e\}^n$ after $\mathbf{S}_e$ used in a traditional explicit method to $2\{e\}^n - \{e\}^{n-1}$. Obviously, the discretization is explicit since we use the field solution at

previous time step to obtain the field at current time step. As for the discretization of the Ampere's law, we keep it the same as before, and hence obtaining

$$\mathbf{S}_h\{h\}^{n+\frac{1}{2}} = \mathrm{diag}\left(\{\epsilon\}\right)\frac{\{e\}^{n+1}-\{e\}^{n}}{\Delta t} + \{j\}^{n+\frac{1}{2}}, \tag{7.14}$$

where the time instants for $\{e\}$ and $\{h\}$, denoted by superscripts, are staggered by half.

Consider the second-order system shown in (7.5). We perform the following time discretization,

$$\begin{aligned}&\mathbf{D}_\epsilon\left(\{e\}^{n+1}-2\{e\}^{n}+\{e\}^{n-1}\right)+\\ &\Delta t^2\mathbf{S}\left(2\{e\}^{n}-\{e\}^{n-1}\right)=\{f\}^{n},\end{aligned} \tag{7.15}$$

where again the $\{e\}^n$ after $\mathbf{S}$ used in a traditional explicit method is replaced by $(2\{e\}^n - \{e\}^{n-1})$. In the MFTD, $\mathbf{D}_\epsilon = \mathrm{diag}\left(\{\epsilon\}\right)$.

Now if we carry out a stability analysis, we find a totally different result! Setting the excitation to be zero as it is irrelevant to stability, and performing a z-transform of (7.15), we obtain

$$z^2 - 2(1-\Delta t^2\lambda)z + (1-\Delta t^2\lambda) = 0 \tag{7.16}$$

where λ are the eigenvalues of $\mathbf{D}_\epsilon^{-1}\mathbf{S}$. Since $\mathbf{S}$ is unsymmetrical, λ can be either real or complex.

However, comparing (7.16) with (7.7), now we have

$$|z_1 z_2| = |1-\Delta t^2\lambda|, \tag{7.17}$$

which is not 1 any more, and it can be made less than 1 via an appropriate choice of time step! Hence, it becomes feasible now to make the magnitude of both roots less than 1, and therefore making the time marching stable. In the following, we quantitatively derive a stability criterion for (7.15).

Denoting $\Delta t^2\lambda$ by

$$\Delta t^2\lambda = a + jb, \tag{7.18}$$

where a is the real part, and b denotes the imaginary part:

$$a = \Delta t^2 \mathrm{Re}(\lambda)$$
$$b = \Delta t^2 \mathrm{Im}(\lambda). \tag{7.19}$$

The two roots of (7.16) can be found as

$$z_1 = 1 - (a + jb) + \sqrt{-(a + jb) + (a + jb)^2}, \tag{7.20}$$

$$z_2 = 1 - (a + jb) - \sqrt{-(a + jb) + (a + jb)^2}. \tag{7.21}$$

Let

$$\xi = -(a + jb) + (a + jb)^2. \tag{7.22}$$

Using the property of

$$\sqrt{\xi} = \sqrt{\frac{|\xi| + \mathrm{Re}(\xi)}{2}} + \mathrm{sign}(\mathrm{Im}(\xi)) j \sqrt{\frac{|\xi| - \mathrm{Re}(\xi)}{2}}, \tag{7.23}$$

we obtain

$$z_{1,2} = 1 - (a + jb) \pm \left(\sqrt{\frac{|\xi| + \mathrm{Re}(\xi)}{2}} + \mathrm{sign}(\mathrm{Im}(\xi)) j \sqrt{\frac{|\xi| - \mathrm{Re}(\xi)}{2}} \right). \tag{7.24}$$

If the product of the two roots is greater than 1 in magnitude, then at least one of the roots has a greater than 1 magnitude. Therefore, to ensure stability, the following condition should be satisfied:

$$|z_1 z_2| = |1 - \Delta t^2 \lambda| = |1 - (a + jb)| < 1, \tag{7.25}$$

which results in

$$2a - a^2 - b^2 > 0. \tag{7.26}$$

We notice that in terms of z_1 (associated with the plus sign), $|z_1|^2 < 1$ provides

$$2(1 - a) \sqrt{\frac{|\xi| + \mathrm{Re}(\xi)}{2}} - 2|b| \mathrm{sign}(2a - 1) \sqrt{\frac{|\xi| - \mathrm{Re}(\xi)}{2}} < 2a - a^2 - b^2 - |\xi|. \tag{7.27}$$

In terms of z_2 (associated with the minus sign), $|z_2|^2 < 1$ provides

$$-2(1-a)\sqrt{\frac{|\xi|+\mathrm{Re}(\xi)}{2}}+2|b|\mathrm{sign}(2a-1)\sqrt{\frac{|\xi|-\mathrm{Re}(\xi)}{2}} < 2a-a^2-b^2-|\xi|\,. \quad (7.28)$$

In (7.27), when $a < 1/2$, both signs of the two terms on the left hand side are positive; in (7.28), when $a > 1$, both signs of the two terms on the left hand side are positive. Hence we obtain

$$2|1-a|\sqrt{\frac{|\xi|+\mathrm{Re}(\xi)}{2}}+2|b|\sqrt{\frac{|\xi|-\mathrm{Re}(\xi)}{2}} < 2a-a^2-b^2-|\xi|\,, \quad (7.29)$$

which is a general form for the maximum magnitude when $a < 1/2$ and $a > 1$. For a in between $1/2$ and 1, the two terms, coming from either (7.27) or (7.28)'s left hand side, have opposite signs, which denote the difference between the two.

For the above to hold true, the right hand side must be greater than 0. Otherwise, since the left hand side is positive, the above condition can never be satisfied. This yields

$$2a-a^2-b^2 > |\xi|\,, \quad (7.30)$$

which results in the following condition:

$$b^2 < \frac{3a^2-2a^3}{2a+1}. \quad (7.31)$$

We then take a square operation on both sides of (7.29) to remove the square root when $a < 1/2$ and $a > 1$, and we obtain

$$4(1-2a+a^2)\frac{|\xi|+\mathrm{Re}(\xi)}{2}+4b^2\frac{|\xi|-\mathrm{Re}(\xi)}{2} +8|1-a||b|\sqrt{\frac{|\xi|^2-|\mathrm{Re}(\xi)|^2}{4}}$$

$$< 4a^2-4a\,|\xi|+|\xi|^2+a^4+b^4+2a^2b^2-2(a^2+b^2)(2a-|\xi|). \quad (7.32)$$

Realizing $|\mathrm{Im}(\xi)| = \sqrt{|\xi|^2 - |\mathrm{Re}(\xi)|^2}$ and grouping some terms, we can further simplify the equation as:

$$\begin{aligned} -|\xi|^2 + 2|\xi| + (2a^2 - 2b^2 - 4a + 2)\mathrm{Re}(\xi) & \\ + 4|1-a||b|\,|\mathrm{Im}(\xi)| & \\ < a^4 + b^4 - 4a^3 - 4ab^2 + 2a^2b^2 + 4a^2. & \quad (7.33) \end{aligned}$$

Utilizing $|\xi|^2 = (a^2 - b^2 - a)^2 + b^2(2a-1)^2$, we obtain

$$2|\xi| + 4|1-a||b|\,|\mathrm{Im}(\xi)| - 8a^2b^2 + a^2 - 3b^2 - 2a + 12ab^2 < 0. \quad (7.34)$$

Based on (7.22), we replace $|\xi|$ and $|\mathrm{Im}(\xi)|$ in terms of a and b. To determine the sign of the term $4|1-a||b|\,|\mathrm{Im}(\xi)|$, either $a < 1/2$ or $a > 1$ yields $4|1-a||b|\,|\mathrm{Im}(\xi)| = -4b^2(1-a)(2a-1)$, and we obtain

$$2|\xi| < 2a - a^2 - b^2. \quad (7.35)$$

When a is in between $1/2$ and 1, we take a square operation on both sides of (7.27) or (7.28), and we obtain a similar expression as (7.34) except that the cross-talk term is $-4|1-a||b|\,|\mathrm{Im}(\xi)|$. This will lead to $-4b^2(1-a)(2a-1)$ as well for $1/2 < a < 1$, namely for all cases, (7.35) is the condition to be satisfied. This will require the right hand side of (7.35) to be positive and it is naturally satisfied based on (7.26). Taking a square on both sides, we obtain

$$3a^4 + 3b^4 - 4a^3 + 6a^2b^2 - 4ab^2 + 4b^2 < 0, \quad (7.36)$$

which can be expressed in terms of the real part and the imaginary part of λ as

$$\begin{aligned} 3(|\mathrm{Re}(\lambda)|^2 + |\mathrm{Im}(\lambda)|^2)^2\Delta t^4 & \\ - 4\mathrm{Re}(\lambda)(|(\mathrm{Re}(\lambda)|^2 + |\mathrm{Im}(\lambda)|^2)\Delta t^2 + 4|\mathrm{Im}(\lambda)|^2 < 0 & \quad (7.37) \end{aligned}$$

Solving the above inequality, we find that the time step needs to satisfy the following condition

$$\sqrt{\frac{2\mathrm{Re}(\lambda)-2\sqrt{|\mathrm{Re}(\lambda)|^2-3|\mathrm{Im}(\lambda)|^2}}{3|\lambda|^2}}<\Delta t < \sqrt{\frac{2\mathrm{Re}(\lambda)+2\sqrt{|\mathrm{Re}(\lambda)|^2-3|\mathrm{Im}(\lambda)|^2}}{3|\lambda|^2}}. \quad (7.38)$$

If $|\mathrm{Re}(\lambda)|^2-3|\mathrm{Im}(\lambda)|^2>0$ is not satisfied, then no real-valued time step can be found to satisfy the above condition, and hence the scheme becomes unstable. Therefore, the following condition is required:

$$b^2<\frac{a^2}{3}, \quad (7.39)$$

which is in general satisfied since the imaginary part of a complex eigenvalue is small compared to the real part.

When the imaginary part is zero or negligible, the left hand side of (7.38) becomes zero, thus this condition is naturally satisfied. However, when the imaginary part of $\mathbf{S}$'s eigenvalues cannot be ignored, (7.38) yields a lower bound of the time step, and we need to go through all the complex eigenvalues to define such a bound. When the imaginary part does not exist or is negligible for the largest eigenvalue, the right hand side of (7.38) becomes

$$\Delta t<\sqrt{\frac{4}{3}}\frac{1}{\sqrt{\lambda_{max}}}, \quad (7.40)$$

where λ_{max} is the largest eigenvalue of $\mathbf{D}_\epsilon^{-1}\mathbf{S}$. This upper bound is, in fact, larger than (7.12), thus a larger time step is allowed in the proposed method compared with [32]. More importantly, this new time marching scheme is truly explicit. No series expansion is required. Hence, the proposed method is more efficient than the original MFTD method. Moreover, as can be seen from the aforementioned analysis, the proposed method is applicable to other unsymmetrical methods to make them stable, since the curl operations represented by $\mathbf{S}_h$ and $\mathbf{S}_e$ in (7.13) and (7.14), and the $\mathbf{S}$ in (7.15) can be arbitrary.

One may wonder why we made a choice of using $2\{e\}^n - \{e\}^{n-1}$ to replace $\{e\}^n$, and whether this choice is accurate or not, although it makes the time marching stable. In fact, the $2\{e\}^n - \{e\}^{n-1}$ constitutes an accurate approximation of $\{e\}^{n+1}$, and the $\{e\}^{n+1}$ is the value used in a backward difference scheme. Since the backward difference has been shown to be accurate in simulating an unsymmetrical system [15, 32], the accuracy of the proposed explicit time marching is also ensured. In a general differencing scheme, a backward or forward difference appears to be less accurate than a central difference. However, each of the three difference schemes produces the same result in a time interval where the field variation is at most linear, as they all are equal to the slope of the field line in this time interval. This is indeed the case in our time domain simulations, since the time step required by the sampling theorem is $1/(2f_{max})$, with f_{max} being the maximum frequency, while what is used in practice is $1/(10f_{max})$ or even smaller so that the accuracy is satisfactory. In such a time interval, a backward difference produces the same result as a central difference.

7.4 Numerical Results

In this section, we first demonstrate the performance the proposed explicit MFTD for simulating unstructured meshes. A number of irregular meshes are simulated, and the accuracy and efficiency of the new method are compared with that of the original MFTD. We then apply the proposed explicit method to the recently developed unsymmetrical FDTD subgridding method [15] to make its explicit simulation stable. All simulations are carried out on an Intel Xeon CPU E5-2690 v2 @ 3.00 GHz.

7.4.1 MFTD in a 2-D Irregular Triangular Mesh

The first example is a highly irregular 2-D mesh shown in Fig. 7.1. The discretization results in 2,081 edges and 1,325 triangular patches. Due to the highly irregular mesh, $\mathbf{S}$ is highly unsymmetrical which has many complex eigenvalues whose imaginary parts are not small as compared to the real parts. In Table 7.1, we list some of

Table 7.1.
Illustration of Complex Eigenvalues of the Triangular Mesh Example

	1	2	3	4
Real	5.150e+27	1.314e+28	2.257e+28	4.781e+28
Imaginary	±1.235e+25	±3.774e+25	±1.282e+26	±1.0542e+27

the representative eigenvalues of **S**. As can be seen, the imaginary part can be quite significant as compared to the real part. Because of this, when we tried to perform a leap-frog scheme directly on the MFTD system of equations, the simulation immediately becomes unstable. Hence, a new explicit method like the proposed is necessary for simulating this example.

To investigate the accuracy of the proposed method in such a mesh, we set up a free-space wave propagation problem so that an analytical solution is available for comparison. Specifically, since **E** is known, we impose an analytical boundary condition, i.e., the known value of tangential **E**, on the outermost boundary of the mesh. We then numerically simulate the fields inside the mesh and correlate the results with the analytical solution.

The incident **E** is set to be $\mathbf{E} = \hat{y} f(t - x/c)$, where $f(t) = 2(t - t_0)\exp(-(t - t_0)^2/\tau^2)$, with $\tau = 2.0 \times 10^{-12}$ s, and $t_0 = 4\tau$, and c denotes the speed of light. This is also the total **E** in a free-space wave propagation problem. The time step used in the proposed method is $\Delta t = 2.42 \times 10^{-16}$ s, which is determined by (7.40). Notice that the **S**'s norm can be analytically estimated from the smallest space step, and (7.40) has a good correlation with the CFL condition. In Fig. 7.2(a), we plot the electric fields of the 6,811-th and 6,812-th entry randomly selected from the unknown $\{e\}$ vector, and compare them with analytical solutions as well as those from the original MFTD method [32]. It can be seen clearly that the electric fields solved from the proposed method have an excellent agreement with analytical results and the results of [32]. To verify the accuracy everywhere in the computational domain, we evaluate

the total solution error by calculating $\|\{e\} - \{e\}_{ana}\|/\|\{e\}_{ana}\|$, where $\{e\}$ contains all e_i unknowns solved from the proposed method, and $\{e\}_{ana}$ is from the analytical solution. As can be seen from Fig. 7.2(a), the proposed method is not only accurate at selected observation points as shown in Fig. 7.2(a), but also accurate at all other points. The center peak error in Fig. 7.2(b) is due to the comparison with close-to-zero fields. In addition, the error plot of the proposed method is on top of that of the original MFTD method, thus the proposed method does not sacrifice the accuracy of the original scheme. Furthermore, it greatly shortens the CPU run time, as the method of [32] takes 88.05 s to finish the simulation, while the proposed method only costs 15.02 s.

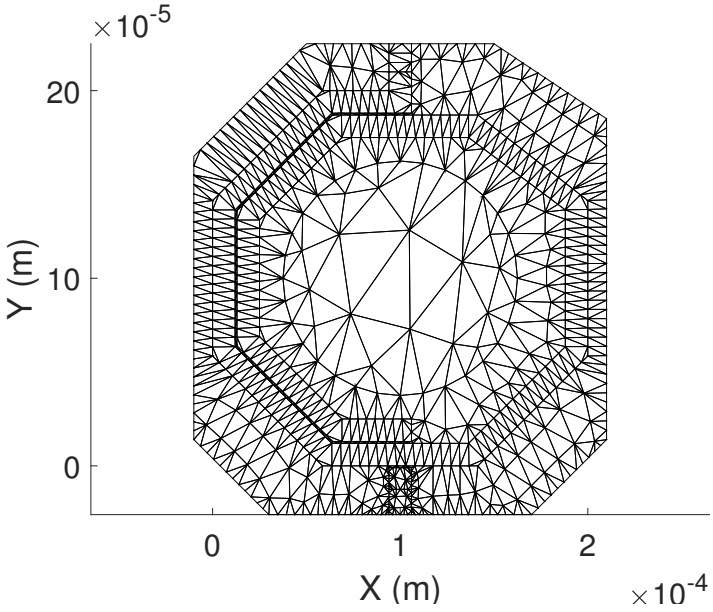
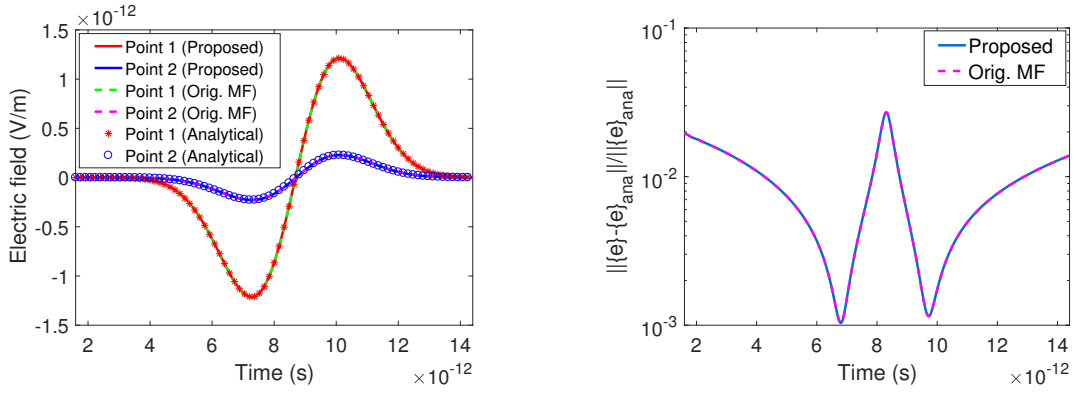


Fig. 7.1. Irregular 2-D triangular mesh.

7.4.2 MFTD in a 3-D Box Discretized into Tetrahedral Mesh

The second example is a 3-D box discretized into tetrahedral elements shown in Fig. 7.3. The discretization results in 544 edges and 350 elements. We set up a free-space wave propagation problem in the given mesh to validate the accuracy



(a) (b)

Fig. 7.2. Simulation of a 2-D triangular mesh. (a) Simulated two electric fields. (b) Entire **E** field solution error as a function of time.

of the proposed method against analytical results. The incident **E** has the same form as that of the previous example, but with $\tau = 6.0 \times 10^{-9}$ s in accordance with the new 3-D structure's dimension. The time step used in the proposed method is $\Delta t = 2.77 \times 10^{-11}$ s, which is determined from (7.40). This also correlates well with the traditional CFL condition. In Fig. 7.4(a), we plot the electric fields of the 1st and the 1,832-th entry from the unknown $\{e\}$ vector, and compare them with analytical solutions and the method of [32]. Excellent agreement can be observed. We also plot the entire solution error shown in Fig. 7.4(b) versus time. It is evident that the proposed method is not just accurate at certain points, but accurate at all points in the computational domain for all time instants simulated. The center peak in Fig. 7.4(b) is due to the comparison with close to zero fields. It takes the proposed method 0.18 s only to finish the simulation without sacrificing accuracy, as can be seen from Fig. 7.4(b), while the method of [32] takes five times longer.

7.4.3 3-D Sphere Discretized into a Tetrahedral Mesh

The third MFTD example is a sphere discretized into tetrahedral elements in free space, whose 3-D mesh is shown in Fig. 7.5. The discretization results in 3,183 edges

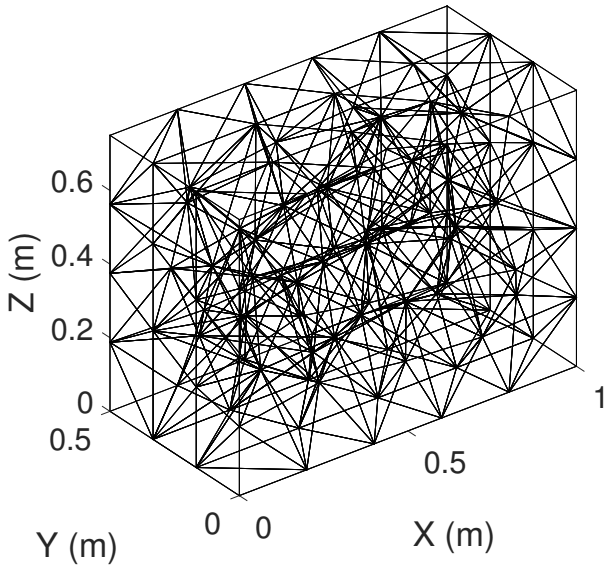
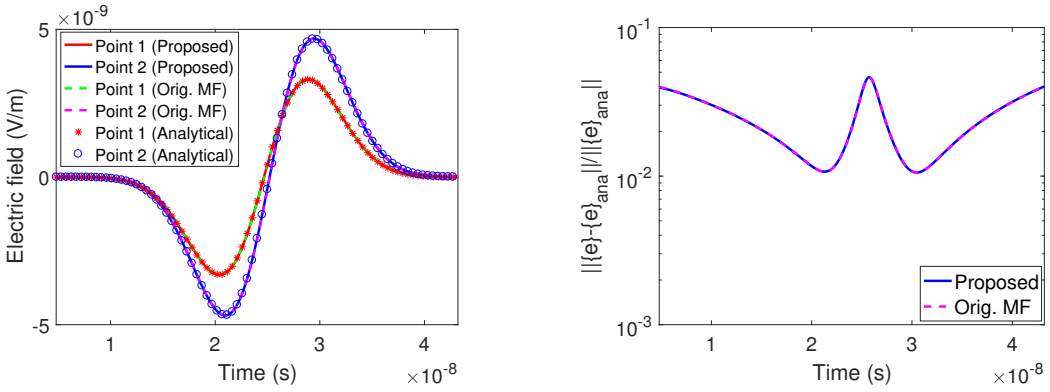


Fig. 7.3. Tetrahedron mesh of a 3-D box.



(a) (b)

Fig. 7.4. Simulation of a 3-D box discretized into tetrahedral elements. (a) Simulated two electric fields in comparison with analytical results. (b) Entire **E** field solution error as a function of time.

and 1,987 tetrahedrons. Similar to previous examples, **S** is highly unsymmetrical having many complex eigenvalues, for some of which the imaginary part is even similar to the real part such as $9.672e+21 \pm j1.923e+21$. The traditional leap-frog scheme is found to be absolutely unstable in simulating this example no matter how

the time step is chosen. The structure is illuminated by a plane wave having the same form as that in the first example but with $\tau = 2.0 \times 10^{-9}$ s. The time step used in the proposed method is $\Delta t = 3.12 \times 10^{-12}$ s, which is selected to satisfy (7.40). Two electric fields, whose indices in vector $\{e\}$ are 1 and 2,942, respectively, are plotted in Fig. 7.6(a) in comparison with analytical data. In Fig. 7.6(c), we plot the entire solution error versus time, where the accuracy of the proposed method is demonstrated. We also simulate this example using the method in [32], whose $\Delta t = 2.70 \times 10^{-12}$ s, which is restricted by (7.12). It takes 16.58 s to finish the simulation. In contrast, the proposed method only takes 3.37 s while achieving the same accuracy as can be seen from Fig. 7.6(b). The original number of time step for the simulation with the proposed time marching scheme is 5,132. We also run a very long time by enlarging the number of time step to be 1,026,400. No late time instability is observed.

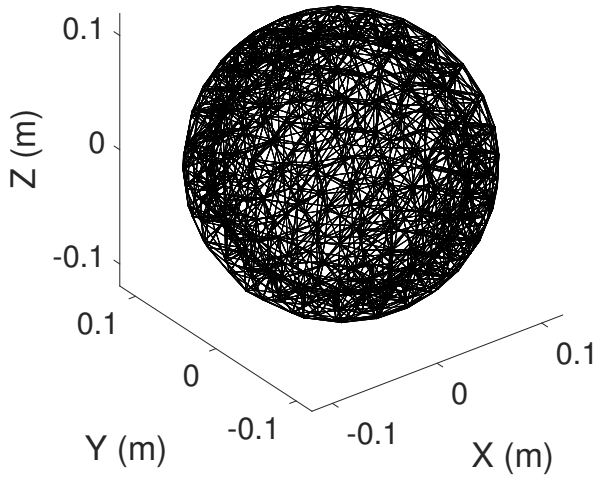
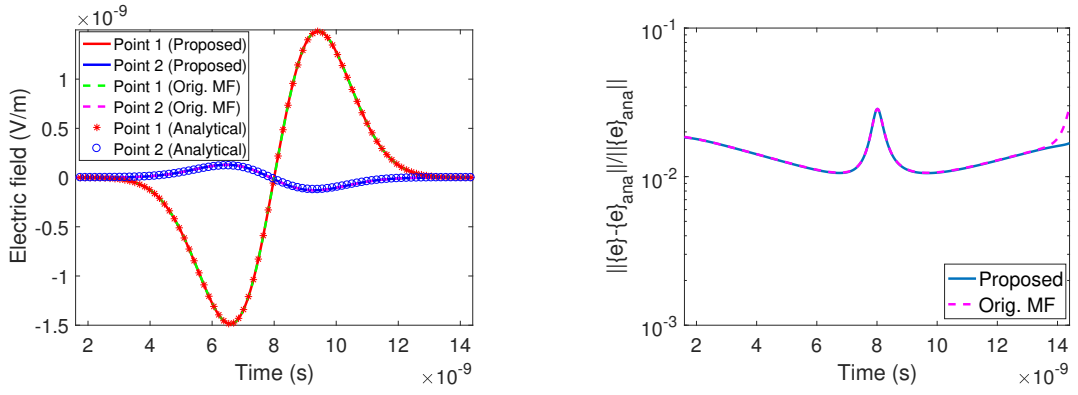


Fig. 7.5. Tetrahedron mesh of a sphere.



(a) (b)

Fig. 7.6. Simulation of a 3-D sphere. (a) Simulated two electric fields. (b) Entire **E** field solution error as a function of time.

7.4.4 Unsymemetric FDTD Subgridding: 2-D Grid with Multiple Subgrids

A 2-D wave propagation problem involving multiple subgrids is considered, whose grid is shown in Fig. 7.7(a). The coarse grid size is $L_c = 0.1$ m, and the blue regions are subdivided into fine grids where the grid ratio $CR = L_c/L_f = 3$. The $\mathbf{E}^{inc}$ is $\hat{y}2(t - t_0 - x/c)e^{(t-t_0-x/c)^2/\tau^2}$ with $c = 3 \times 10^8$ m/s, $\tau = 2 \times 10^{-8}$ s and $t_0 = 4\tau$. All the boundaries are terminated by an exact absorbing boundary condition. The time step used is $\Delta t = 1/\sqrt{\|\mathbf{S}\|} = 4.4 \times 10^{-11}$ s, which satisfies the stability criterion of a traditional explicit method. In Fig. 7.7(b) we plot the electric fields the 1st and 340th entry. It can be seen that both traditional explicit method and the proposed new explicit method can generate accurate results.

However, when we run a long time simulation, late time instability is observed from the traditional explicit time marching, as shown in Fig. 7.8(a). This is because of the existence of complex eigenvalues, and no Δt can satisfy both $|z_1| = 1$ and $|z_2| = 1$ as mentioned in (7.7). For example, the 102nd eigenvalue of the system matrix is $1.5018e20 \pm 2.1538e16i$, making $|z_1| = 1.00004$. Although the imaginary part is small compared to the real part, as long as such a complex eigenvalue exists,

the time domain simulation cannot be made stable. The instability can be clearly observed when 1,818,000 time steps are simulated. In contrast, using the proposed explicit method, no such instabilities are observed in the same late time simulation, which can be clearly seen from Fig. 7.8(b).

7.4.5 Unsymemetric FDTD Subgridding: 3-D Cube with Two Subgridding Cells

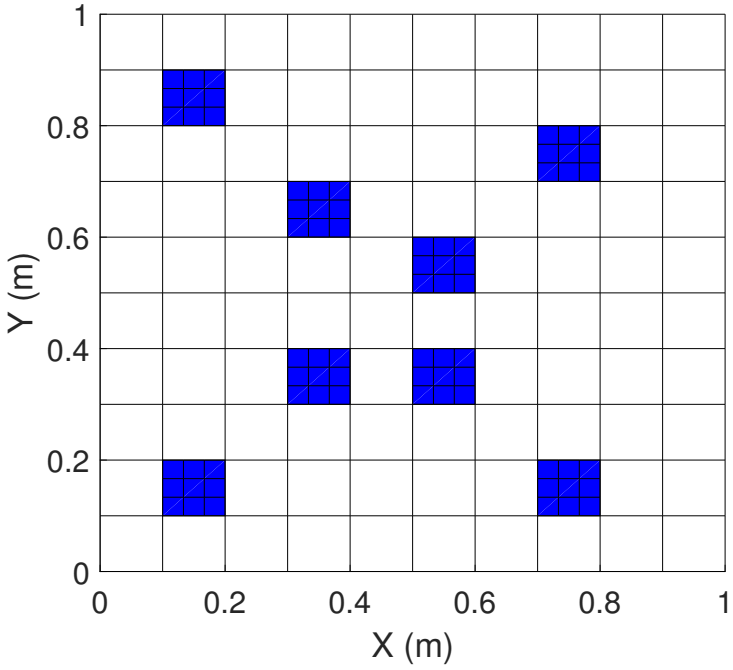
The second subgridding example is a free-space wave propagation problem in a 3-D Cube. The size of the simulation domain in each direction is 1.0 m with the coarse grid size is $L_c = 0.1$ m. Two cells which are centered at (0.15, 0.15, 0.15) m, and (0.45, 0.45, 0.45) m respectively are subdivided into fine grids with a grid ratio of $CR = L_c/L_f = 3$. The grid is shown with the blue subgrids in Fig. 7.9(a). We apply the same incident field and boundary condition as those in the previous example. The time step used is $\Delta t = 1/\sqrt{\|\mathbf{S}\|} = 3.5 \times 10^{-11}$ s. In Fig. 7.9(b) we plot the electric fields at the 1st and the 2,694th entry. It can be seen that both the proposed new explicit method and the traditional explicit method agree very well with the analytical solution.

However, similar to previous example, when we run a long time simulation, the traditional explicit method is shown to be unstable, as shown in Fig. 7.10(a), where the number of time steps is 1,142,750. Complex eigenvalue pairs are observed from the unsymmetrical system matrix of the subgridding method. For example, the 965th eigenvalue is 5.4851e21+2.0377e16i, making $|z_1| = 1.00005$. In contrast, using the proposed new method, no such instabilities are observed in the late time simulation, and the results compared with the reference analytical data are shown in Fig. 7.10(b).

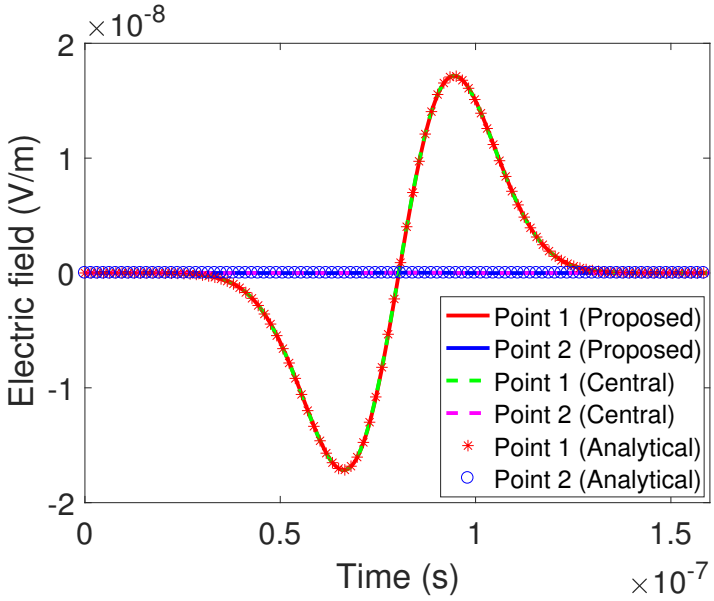
7.5 Conclusion

In this work, we develop a truly explicit time marching scheme for solving unsymmetrical numerical systems in time domain. The proposed method is theoretically

proved to be stable despite the unsymmetrical system matrix. Meanwhile, the accuracy of the time marching is not sacrificed, and the time step size is not reduced. As a result, we make the MFTD method truly matrix-free, and hence accentuating its advantage of having a diagonal mass matrix irrespective of the element shape used for discretization. We also successfully apply the proposed method to make the explicit time marching of a state-of-the-art FDTD subgridding method stable. The proposed method is generic, and hence it can be used in other unsymmetrical methods to guarantee their stability in explicit time-domain simulation.

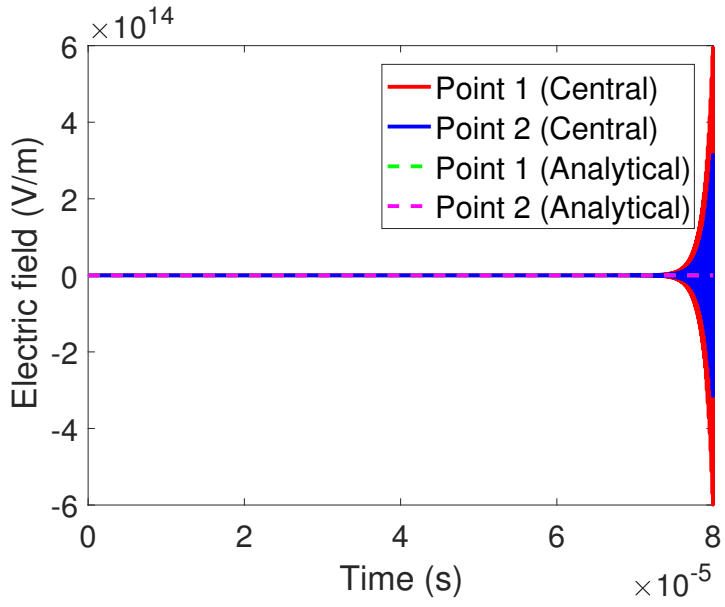


(a)

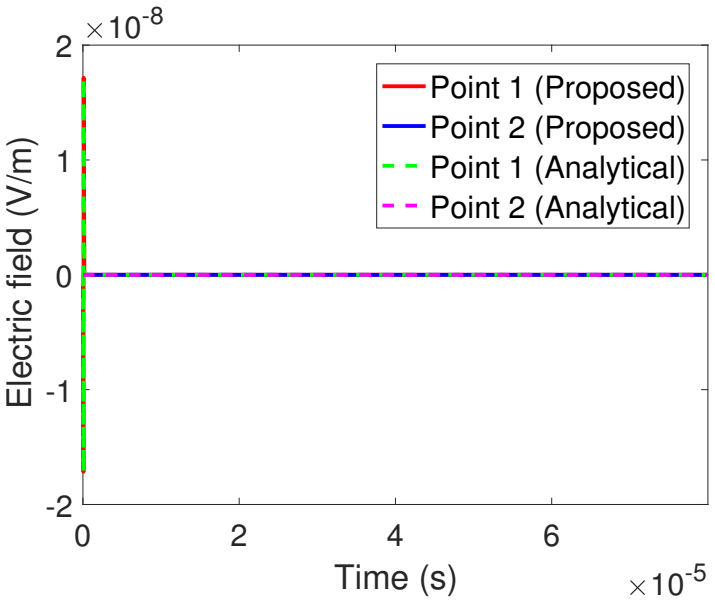


(b)

Fig. 7.7. Simulation of a 2-D subgridding problem. (a) Grid details. (b) Simulated two electric fields in comparison with traditional method and analytical results.

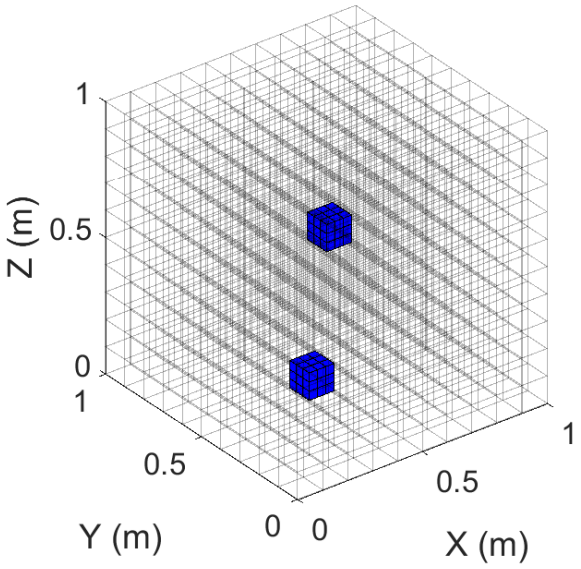


(a)

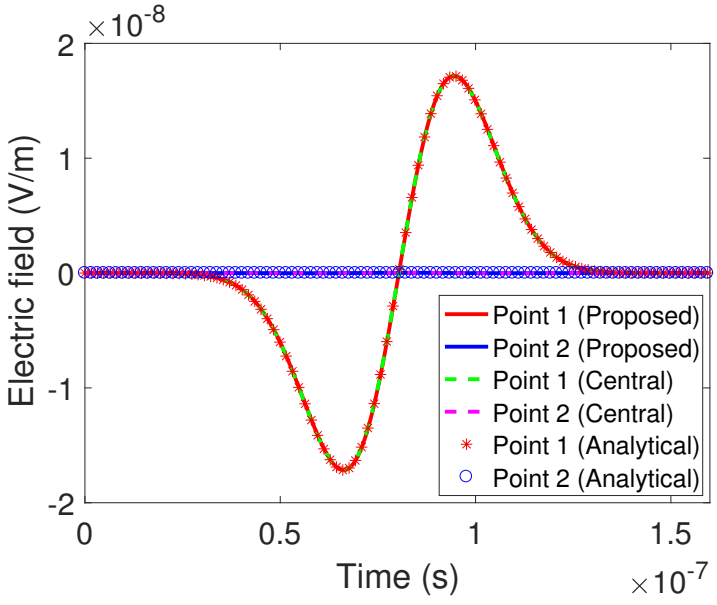


(b)

Fig. 7.8. Long term simulation of a 2-D subgridding problem. (a) Long term simulation using traditional central difference vs. analytical results. (b) Long term simulation using proposed method vs. analytical results.

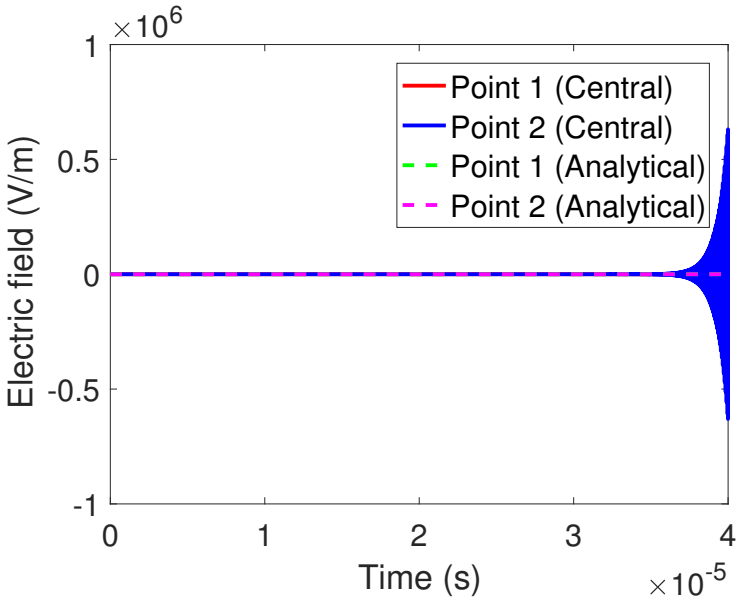


(a)

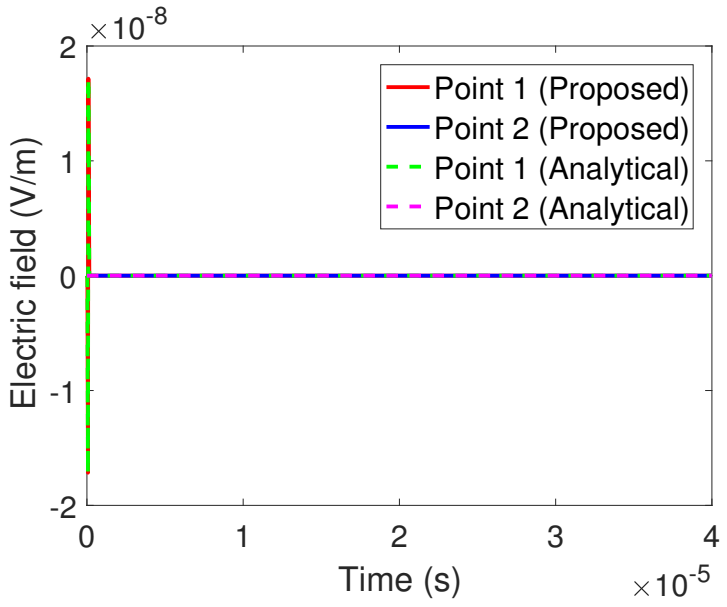


(b)

Fig. 7.9. Simulation of a 3-D cube with subgridding cells. (a) Grid details. (b) Simulated two electric fields in comparison with traditional method and analytical results.



(a)



(b)

Fig. 7.10. Long term simulation of a 3-D cube with subgridding cells. (a) Long term simulation using traditional central difference vs. analytical results. (b) Long term simulation using proposed method vs. analytical results.

8. CONCLUSIONS AND FUTURE WORK

In this dissertation, we focus on two major topics in solving new emerging engineering problems: multiscale and multiphysics. First, we develop a symmetric positive semi-definite finite-difference time-domain (FDTD) subgridding algorithm in both space and time for fast transient simulations of multiscale problems. This algorithm is stable and accurate by construction. To further improve the efficiency, we propose an analytical method to identify the unstable modes in a uniform grid, and remove such modes subsequently to make the subgridding method unconditionally stable.

To address the multiphysics simulation challenge, we develop a matrix-free time domain method for solving thermal diffusion equation, and the combined Maxwell-thermal equations in arbitrary unstructured meshes. Since the matrix-free time-domain method has a diagonal mass matrix in nature, we take advantage of the diagonal property of the mass matrix and develop a method for fast frequency-domain analysis. Furthermore, we propose a new time marching scheme to make matrix-free method truly explicit. And this scheme is generic, which can be used to stably simulate other unsymmetrical systems as well.

To create these new methods, we have considered the following three aspects that are equally important.

- *Accuracy*: No matter how fast the algorithm can be, accuracy should be guaranteed first. Since our subgridding algorithm involves both space and time, a careful consideration is taken in implementing the subgridding framework. We separate base grid from the subgrid, and the interpolation matrix connecting these two parts is important to ensure the accuracy. Especially on a 3-D grid, we have the challenge in calculating the electric field on the interface accurately. Instead of directly using a linear interpolation from the base grid electric field,

we take into account the contribution from both base grid and subgrid regions to make sure that we can get accurate results.

While developing the method free of matrix solution to solve Maxwell and thermal equations, accuracy must be ensured as well. As mentioned, when solving Maxwell's equations, the **H** field is sampled in such a way that **E** unknowns can be accurately calculated from sampled **H** fields, which can be located at any point inside the element. Therefore, to obtain these **H** fields accurately, we use a first order basis when choosing the vector basis functions for **E**. And we solve the thermal diffusion equation in a similar manner. The field/temperature at any point along any direction can be obtained accurately.

- *Efficiency*: With the accuracy issue addressed, we need to come up with some ideas to speed up the algorithm. First, the root cause of low efficiency of existing methods is investigated, and our algorithm is designed in a way to remove the root cause. In terms of multiscale problems, FDTD subgridding techniques have been developed to address the multiscale challenge and they are faster compared with conventional FDTD without subgrids. However, since we have both the base grid and the subgrid, the time step is limited by the smallest space in the fine feature. In the proposed symmetric positive semi-definite FDTD subgridding algorithm, such a restriction is removed, and it permits the use of a time step local to the base grid and the subgrid, thus the efficiency is much improved.

  In multiphysics problems, existing methods require solving a system matrix. As a result, at each time instant when the matrix changes, one has to re-factorize or solve the matrix, which is time consuming especially in analyzing large-scale problems. To improve efficiency, we develop a matrix-free time-domain method to solve Maxwell's equations and the thermal diffusion equation simultaneously in time domain. This method has a naturally diagonal system matrix, and hence the need for numerically finding the matrix solution is completely eliminated.

- *Time marching stability*: Even though very accurate interpolation techniques can be developed to obtain the field unknowns at the interface in an FDTD subgridding algorithm, the stability of the resultant time marching cannot be guaranteed. The fundamental reason is that the system matrices with such interpolations are unsymmetric. In our symmetric positive semi-definite FDTD subgridding algorithm, stability issue is investigated and it is proved that although there is a coupling between base grid and subgrid, our combined system matrix is symmetric positive semi-definite, and hence the algorithm is stable by construction.

  The co-simulation of the coupled Maxwell's and thermal equations results in a nonlinear system of equations, and the stability of which has not been investigated in existing methods. We rigorously analyze the stability of the coupled nonlinear system of equations, and it is found to be ensured with a correct choice of time step in explicit time marching.

  Our new time marching scheme developed in the last chapter is a systematic approach to overcome the absolute instability of the explicit simulating an unsymmetrical numerical system. With this approach, we are able to remove the need for using the backward difference and the series expansion to deal with unsymmetrical systems.

The future research potentials of this work include but not limited to

- *Unconditionally stable 3-D FDTD subgridding algorithm*: In Chap. 4, we combine the proposed symmetric positive semi-definite FDTD subgridding algorithm with the analytical unstable modes removal technique in some 2-D examples. In fact, the analytical unstable modes removal technique can be readily expanded to 3-D cases. As a result, the subgridding algorithm with the analytical unstable modes removal can have a broader application in 3-D analysis.

- *Larger scale Maxwell-thermal co-simulation*: The proposed matrix-free time-domain method has been applied to solve examples with both tetrahedral meshes

and layers of triangular prism elements. In the co-simulation, we can solve Maxwell's equations even faster by using the new time marching scheme developed in the last chapter. For a large signal and power integrity problem which involves many more unknowns, to further improve the efficiency, an unconditional stable version of this method need to be considered to enlarge the time step, in order to finish the whole simulation in a practical time.

- *Extensions to other partial differential equations*: The proposed matrix-free time-domain method provides a flexible framework for solving multiphysics problems, not only Maxwell's equations, thermal diffusion equation, but also other equations. For example, the Boltzmann equation is also a first-order partial differential equation, and a similar treatment can be performed to avoid solving a matrix equation at each time step suing the proposed matrix-free method.

REFERENCES

REFERENCES

[1] K. Yee, "Numerical solution of initial boundary value problems involving maxwells equations in isotropic media," *IEEE Transactions on Antennas and Propagation*, vol. 14, no. 3, pp. 302–307, 1966.

[2] A. Taflove and S. C. Hagness, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*. Artech House, 2000.

[3] J. Jin, *The finite element method in electromagnetics*. John Wiley & Sons, 2002.

[4] M. Feliziani and F. Maradei, "Hybrid finite-element solutions as time dependent maxwell's curl equations," *IEEE Transactions on Magnetics*, vol. 31, no. 3, pp. 1330–1335, 1995.

[5] S. S. Zivanovic, K. S. Yee, and K. K. Mei, "A subgridding method for the time-domain finite-difference method to solve maxwells equations," *IEEE Transactions on Microwave Theory and Techniques*, vol. 39, no. 3, pp. 471–479, 1991.

[6] D. T. Prescott and N. V. Shuley, "A method for incorporating different sized cells into the finite-difference time-domain analysis technique," *IEEE Microwave and Guided Wave Letters*, vol. 2, no. 11, pp. 434–436, 1992.

[7] M. J. White, M. F. Iskander, and Z. Huang, "Development of a multigrid fdtd code for three-dimensional applications," *IEEE Transactions on Antennas and Propagation*, vol. 45, no. 10, pp. 1512–1517, 1997.

[8] M. W. Chevalier, R. J. Luebbers, and V. P. Cable, "Fdtd local grid with material traverse," *IEEE Transactions on Antennas and Propagation*, vol. 45, no. 3, pp. 411–421, 1997.

[9] M. J. White, Z. Yun, and M. F. Iskander, "A new 3d fdtd multigrid technique with dielectric traverse capabilities," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, no. 3, pp. 422–430, 2001.

[10] M. Okoniewski, E. Okoniewska, and M. A. Stuchly, "Three-dimensional subgridding algorithm for fdtd," *IEEE Transactions on Antennas and Propagation*, vol. 45, no. 3, pp. 422–429, 1997.

[11] P. Thoma and T. Weiland, "A consistent subgridding scheme for the finite difference time domain method," *International Journal of Numerical Modeling: Electronic Networks, Devices and Fields*, vol. 9, no. 5, pp. 359–374, 1996.

[12] O. Podebrad, M. Clemens, and T. Weiland, "New flexible subgridding scheme for the finite integration technique," *IEEE Transactions on Magnetics*, vol. 39, no. 3, pp. 1662–1665, 2003.

[13] L. Kulas and M. Mrozowski, "Reciprocity principle for stable subgridding in the finite difference time domain method," in *2007 The international conference on "Computer as a Tool"*. IEEE, 2007, pp. 106–111.

[14] K. Xiao, D. J. Pommerenke, and J. L. Drewniak, "A three-dimensional fdtd subgridding algorithm with separated temporal and spatial interfaces and related stability analysis," *IEEE Transactions on Antennas and Propagation*, vol. 55, no. 7, pp. 1981–1990, 2007.

[15] J. Yan and D. Jiao, "An unsymmetric fdtd subgridding algorithm with unconditional stability," *IEEE Transactions on Antennas and Propagation*, vol. 66, no. 8, pp. 4137–4150, 2018.

[16] ——, "Symmetric positive semidefinite fdtd subgridding algorithms for arbitrary grid ratios without compromising accuracy," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 12, pp. 5084–5095, 2017.

[17] D. Mavriplis, "On convergence acceleration techniques for unstructured meshes," in *29th AIAA, Fluid Dynamics Conference*. AIAA, 1998, pp. 1–20.

[18] M. Cinalli and A. Schiavoni, "Stable and consistent generalization of the fdtd technique to nonorthogonal unstructured grids," *IEEE Transactions on Antennas and Propagation*, vol. 54, no. 5, pp. 1503–1512, 2006.

[19] H. X. Zheng and K. W. Leung, "A nonorthogonal adi-fdtd algorithm for solving two dimensional scattering problems," *IEEE Transactions on Antennas and Propagation*, vol. 57, no. 12, pp. 3891–3902, 2009.

[20] M. M. Rana and A. S. Mohan, "Nonorthogonal lod-fdtd method for em scattering from two-dimensional structures," *IEEE Transactions on Electromagnetic Compatibility*, vol. 55, no. 4, pp. 764–772, 2013.

[21] R. T. Lee, J. G. Maloney, and D. W. Landgren, "Fdtd in curvilinear coordinates using a rectangular fdtd formulation," in *IEEE International Symposium on Antennas and Propagation*. IEEE, 2011, pp. 2326–2329.

[22] L. Bernard and L. Pichon, "Generalized finite difference scheme using mainly orthogonal and locally barycentric dual mesh for electromagnetic problems," *The European Physical Journal-Applied Physics*, vol. 52, no. 2, pp. 1–4, 2010.

[23] R. Holland, "Finite-difference solution of maxwell's equations in generalized nonorthogonal coordinates," *IEEE Transactions on Nuclear Science*, vol. 30, no. 6, pp. 4589–4593, 1983.

[24] M. Fusco, "Fdtd algorithm in curvilinear coordinates [em scattering]," *IEEE Transactions on Antennas and Propagation*, vol. 38, no. 1, pp. 76–89, 1990.

[25] J. F. Lee, R. Palandech, and R. Mittra, "Modeling three-dimensional discontinuities in waveguides using nonorthogonal fdtd algorithm," *IEEE Transactions on Microwave Theory and Techniques*, vol. 40, no. 2, pp. 346–352, 1992.

[26] N. K. Madsen, "Divergence preserving discrete surface integral methods for maxwell equations using nonorthogonal grids," *Journal of Computational Physics*, vol. 119, no. 1, pp. 34–45, 1995.

[27] C. Chan, J. Elson, and H. Sangani, "An explicit finite-difference time-domain method using whitney element," in *1994 IEEE International Symposium on Antennas and Propagation*. IEEE, 1994, pp. 1–2.

[28] S. Gedney, F. Lansing, and D. Rascoe, "A full-wave analysis of passive monolithic integrated circuit devices using a generalized yee-algorithm," *IEEE Transactions on Microwave Theory and Techniques*, vol. 44, no. 8, pp. 1393–1400, 1996.

[29] C. F. Lee, B. J. McCartin, T. Shin, R, and J. A. Kong, "A triangle grid finite-difference time-domain method for electromagnetic scattering problems," *Journal of Electromagnetic Waves and Applications*, vol. 8, no. 4, pp. 1429–1438, 1994.

[30] M. Hano and T. Itoh, "Three-dimensional time-domain method for solving maxwells equations based on circumcenters of elements," *IEEE Transactions on Magnetics*, vol. 32, no. 3, pp. 946–949, 1996.

[31] J. A. Roden, *Broadband electromagnetic analysis of complex structures with the finite-difference time-domain technique in general curvilinear coordinates*. Ph.D. Dissertation, University of Kentucky, Lexington, KY, 1997.

[32] J. Yan and D. Jiao, "Accurate and stable matrix-free time-domain method in 3-d unstructured meshes for general electromagnetic analysis," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 12, pp. 4201–4214, 2015.

[33] ——, "Matrix-free time-domain method for general electromagnetic analysis in 3-d unstructured meshes–modified-basis formulation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 8, pp. 2371–2382, 2016.

[34] ——, "Time-domain method having a naturally diagonal mass matrix independent of element shape for general electromagnetic analysis–2-d formulations," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 3, pp. 1202–1214, 2017.

[35] T. Lu and J. Jin, "Electrical-thermal co-simulation for analysis of high-power rf/microwave components," *IEEE Transactions on Electromagnetic Compatibility*, vol. 59, no. 1, pp. 93–102, 2017.

[36] N. Li, J. Mao, W. S. Zhao, M. Tang, W. Chen, and W. Y. Yin, "Electrothermal cosimulation of 3-d carbon-based heterogeneous interconnects," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 6, no. 4, pp. 518–526, 2016.

[37] X. Zhang, Z. Chen, and Y. Yu, "An unconditional stable meshless adi-rpim for simulation of coupled transient electrothermal problems," *IEEE Journal on Multiscale and Multiphysics Computational Techniques*, vol. 1, pp. 98–106, 2016.

[38] Y. Nakatani, T. Sekine, and H. Asai, "Three-dimensional iterative electrical-thermal co-simulation (3-d ietc) method for power/thermal integrity analysis," in *IEEE CPMT Symposium Japan*. IEEE, 2016, pp. 167–168.

[39] T. Casper, H. D. Gersem, R. Gillon, T. Gotthans, T. Kratochvil, P. Meuris, and S. Schops, "Electrothermal simulation of bonding wire degradation under uncertain geometries," in *2016 Design, Automation and Test in Europe Conference and Exhibition*. IEEE, 2016, pp. 1297–1302.

[40] A. Leggieri, D. Passi, F. D. Paolo, B. Spataro, and E. Dyunin, "Design of a sub-millimetric electron gun with analysis of thermomechanical effects on beam dynamics," *Vaccum*, vol. 122, pp. 103–116, 2015.

[41] J. Xie and M. Swaminathan, "Electrical-thermal cosimulation with nonconformal domain decomposition method for multiscale 3-d integrated systems," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 4, no. 4, pp. 588–601, 2014.

[42] F. Torres and B. Jecko, "Complete fdtd analysis of microwave heating processes in frequency dependent and temperature-dependent media," *IEEE Transactions on Microwave Theory and Techniques*, vol. 45, no. 1, pp. 108–117, 1997.

[43] S. D. Gedney and J. A. Roden, "Numerical stability of nonorthogonal fdtd methods," *IEEE Transactions on Antennas and Propagation*, vol. 48, no. 2, pp. 231–239, 2000.

[44] J. Yan and D. Jiao, "Fast explicit and unconditionally stable fdtd method for electromagnetic analysis," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 65, no. 8, pp. 2698–2710, 2017.

[45] M. Gaffar and D. Jiao, "An explicit and unconditionally stable fdtd method for electromagnetic analysis," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 11, pp. 2538–2550, 2014.

[46] K. Zeng and D. Jiao, "Symmetric positive semi-definite fdtd subgridding algorithm in both space and time," in *2018 IEEE International Symposium on Antennas and Propagation*. IEEE, 2018, pp. 1–2.

[47] M. Gaffar and D. Jiao, "Alternative method for making explicit fdtd unconditionally stable," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 63, no. 12, pp. 4215–4224, 2015.

[48] K. Zeng and D. Jiao, "Explicit and unconditionally stable fdtd with analytical method for identifying unstable modes," in *2018 IEEE International Symposium on Antennas and Propagation*. IEEE, 2018, pp. 1–2.

[49] W.-C. Yueh, "Eigenvalues of several tridiagonal matrices," in *Applied Mathematics E-notes*, 2005, pp. 5–66.

[50] K. Zeng and D. Jiao, "Explicit unconditionally stable symmetric positive semidefinite fdtd subgridding algorithm with analytical removal of unstable modes," in *2019 IEEE International Symposium on Antennas and Propagation*. IEEE, 2019, pp. 1–2.

[51] ——, "Matrix-free method for transient maxwell-thermal cosimulation in arbitrary unstructured meshes," *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, no. 12, pp. 5439–5448, 2018.

[52] E. C. Yang, *Vector and Tensor Analysis*. Dekker, 1973.

[53] G. Teschl, *Ordinary Differential Equations and Dynamical Systems*. American Mathematical Society, 2012.

[54] K. Zeng and D. Jiao, "Frequency-domain method having a diagonal mass matrix in arbitrary unstructured meshes for efficient electromagnetic analysis," in *2017 IEEE International Symposium on Antennas and Propagation*. IEEE, 2017, pp. 1–2.

[55] L. N. Trefethen and D. Bau III, *Numerical linear algebra*. Siam, 1997.

VITA

VITA

Kaiyuan Zeng received the B.S. degree in Electronic Information Engineering from the University of Science and Technology of China, Hefei, China, in 2012. Since then, he has been working toward the Ph.D. degree in electrical engineering at Purdue University, West Lafayette, IN, USA. His research is focused on computational electromagnetics and multiphysics modeling.

He received Honorable Mention Award from the 2018 IEEE AP-S International Symposium on Antennas and Propagation. He also won Best Student Paper Award (1st place winner) from the 2015 IEEE Wireless and Microwave Technology Conference.