1 Introduction

1.1 Why FDTD?

With the continued growth of computing power, modeling and numerical simulation has grown immensely as a tool for understanding and analyzing just about any problem in science. Where in the mid-twentieth century, detailed analyses were required to get any meaningful insight out of complex problems, today we can simply plug the governing differential equations into a computer, the results of which can provide an immense amount of information, which is of course complementary to theoretical analyses. The growth of computing power has brought with it a smorgasbord of modeling methods, applicable in any number of fields. The problem, then, is knowing when to use which method.

In electromagnetic problems, which are of interest to us in this book, there are quite a number of useful numerical methods, including the Method of Moments, Finite Volume methods, Finite Element methods, and Spectral methods, just to name a few. The FDTD method, however, grew to become the method of choice in the 1990s, for a number of reasons. First, it has always had the advantage of being a very simple method; we shall see in Chapter 3 that the derivation of difference equations is very straightforward. However, before the 1990s, the FDTD method was hindered by the need to discretize the simulation space on sub-wavelength scales, with time steps commensurately small. Hence, any reasonable problem would require a large amount of computer memory and time. Since the 1990s, however, with the growth of computing power, the FDTD method has taken off.

As an example, a typical 3D problem would require, at minimum, 100 grid cells in each dimension, or 10^6 grid cells total. With a minimum of six fields to compute (three components each of the electric field $\mathscr{E}$ and magnetic field $\mathscr{H}$), and 2 bytes per value (for 16-bit resolution), we require 12 MB of memory. As for computation time, our simulation might require 1,000 time steps. Each of six equations will have four additions and two multiplications (at minimum, for the free-space algorithm in Chapter 4) at each of the 1 million grid cells, for ~36 billion operations over the time of our simulation. In 1990, 12 MB of memory and 36 billion operations was a significant calculation; today, you could quite easily run this simulation on your mobile phone.[1]

[1] More accurately, this simulation used 48 MB on a desktop PC running Matlab, since it stored the arrays at 8-byte doubles; and it ran 1,000 time steps in under 5 minutes.

The advantages of the FDTD method over other methods include:

- *Short development time*. Thanks to the simple discretization process, a realistic 2D or 3D simulation can be written in only a few minutes in less than 100 lines of code. In other methods, such as the Finite Element method, creating the numerical grid alone can require entire software packages, and understanding of the discretization procedure can be quite convoluted.
- *Ease of understanding*. Again, thanks to the simple discretization procedure, the FDTD method is easily understandable and directly follows from the differential form of Maxwell's equations. The stability and dispersion characteristics of the method also follow from a simple, intuitive understanding of the updating procedure.
- *Explicit nature*. In the traditional explicit FDTD method, no linear algebra or matrix inversions are required, and as such there is no inherent limit to the size of a simulation; computer time is the only limitation.

However, there are also a number of disadvantages of the FDTD method:

- *Stair-stepping edges*. The orthogonal grid structure of the FDTD method implies that edges of structures within the simulation have edges that follow the grid structure. This can become a problem for curved surfaces, for which greater accuracy is sought. Some methods for FDTD have been developed to overcome this limitation, including the subcell structures discussed in Chapter 12, but other methods are generally better suited to these complex geometries.
- *Computational time*. In the FDTD method, the time step at which we advance the solution is limited by the spatial size, and cannot be larger than a certain maximum size, as we will derive in Chapter 5. For simulations with large spaces or multiple scales (largely varying wavelengths), this means the simulation must be run for a very long time. Other methods can often be better at dealing with multiscale problems.

1.2 Other methods

This book covers the FDTD method in detail, with little coverage of other methods. As mentioned above, this is due to the increasing prevalence of the FDTD method in electromagnetic problems. However, any good engineer or scientist should have a good understanding of other available methods, and should develop knowledge of the appropriate conditions under which different methods are used. Here we provide a brief mention of some other methods commonly used in electromagnetic problems.

1.2.1 Finite volume time domain

The Finite Volume Time Domain (FVTD) method became popular in modeling electromagnetic problems due to its flexibility in modeling irregular structures [1, 2]. As we will see throughout this book, the FDTD method is somewhat restricted to regular, structured grids, and any curved surfaces become "staircased" when the discretized grid

is formulated. We will see in Chapter 15 that FDTD methods can be developed around irregular grids; however, the FVTD method is another way of working around irregular structures.

In short, the FVTD method defines the fields $\overline{\mathscr{E}}$ and $\overline{\mathscr{H}}$ in small volumes of space, rather than at the nodes of grid cells as in the FDTD method. These small volumes can be arbitrarily defined, but are typically taken to be tetrahedra in 3D or triangles in 2D. These shapes simplify the resulting equations and can be designed around curved and complex structures quite well. The FVTD method then uses the integral forms of Maxwell's equations to conserve the field quantities. For example, in a small volume V_e with surface area A_e, any change in the electric or magnetic flux inside the volume from one time step to the next must be balanced by the flux moving across the boundary area A_e, which moves into (or out of) the adjacent cells.

An introduction to the Finite Volume method for electromagnetics is provided in the book by S. Rao [3], and we will introduce it in some detail in Chapter 15. The method has the obvious advantage that irregular structures can be modeled quite easily. The simulation time is typically very similar to the FDTD method. Disadvantages include the need to create and define an irregular grid of tetrahedral cells, which is quite cumbersome outside of commercial software.

1.2.2 Finite difference frequency domain

Time domain methods such as FDTD are extremely useful when a transient or broadband analysis is required. For example, we may be interested in the scattering pattern of a broadband pulse of energy from a particular scatterer. However, in cases where a steady-state solution is sought only at a single frequency, the FDTD method is rather inefficient. Instead, frequency domain methods can be much more efficient, since they avoid the need to step in time.

The Finite Difference Frequency Domain (FDFD) method is highly applicable, since it maintains the spatial features of the FDTD method, but removes time stepping. Rather, the steady-state solution is found at a single frequency through a matrix inversion process. We will briefly introduce the FDFD method in Chapter 14.

The FDFD method has the additional advantage that dispersive materials become trivial to implement. As we shall see in Chapter 10, dispersive materials in the FDTD method require either convolution terms or auxiliary equations. In the FDFD method, one simply uses the scalar (or vector, in the case of anisotropic dispersive materials) values of ϵ, μ, and σ at the frequency of interest.

Particular problems that are better suited to FDFD are those where the solution is required at only a single frequency. For example, mobile communications typically operate in a narrow bandwidth around a carrier frequency, so that the bandwidth can be approximated by a single frequency. However, the FDFD method can also be used for broadband simulations, by running multiple simulations, one at each frequency of interest. In this way the spectral response of a problem can be determined, with the frequency resolution limited only by the number of simulations one is willing to run. This can be useful for problems involving dispersive media, whose material parameters

vary with frequency in a way that cannot be easily modeled in FDTD. In Chapter 10 we will discuss modeling of dispersive materials in the FDTD method, where methods have been derived for some typical dispersion characteristics.

1.2.3 Finite element methods

The finite element method has become prominent in electromagnetic problems in the past decade or so, but has been around much longer than that, having originated in the 1940s with the work of A. Hrennikoff and R. Courant[2] [4]. The finite element method (often known as Finite Element Analysis or FEA) was developed in the 1950s for airframe and structural analysis.

Like the finite volume method, the finite element method divides the simulation space into small areas or volumes (in 2D and 3D, respectively) which can be arbitrarily shaped and oriented; for this reason, the finite element method is well suited to problems with complex geometry. Also similar to the finite volume method, while the small "subdomains" can have arbitrary shapes, triangles and tetrahedra are most commonly used for their simplicity.

Now, in each of the other methods we have discussed so far, Maxwell's equations are discretized and values of the fields are found which satisfy these equations. In the finite element method, however, the *solution* to Maxwell's equations is approximated over each subdomain with some functional form, usually a low-order polynomial, which is known as a basis function. The solutions in each subdomain are then made to be continuous across their boundaries, and the solution must be made to fit with the global boundary conditions enforced by any scattering structures.

The finite element method in electromagnetics has the disadvantage of being rather complicated, and as such we will provide only a brief introduction to the method in Chapter 15. However, Finite Element Time Domain (FETD) is the state-of-the-art for time domain solutions of electromagnetic problems. Many books have been written on the method, and we refer the reader to those books in Chapter 15.

Discontinuous Galerkin methods

One of the drawbacks of the finite element method in electromagnetic problems is that it requires some level of *global* knowledge of the simulation space. The basis functions used are local, defined in each grid element, but to enforce continuity at element boundaries, a large, sparse matrix must be solved, which can heavily increase the computational cost.

More recently, discontinuous Galerkin methods have moved to the forefront of electromagnetic simulation. These methods enforce strict locality by relaxing the requirement of continuity between elements. Discontinuous Galerkin methods borrow ideas from finite volume methods to connect elements together at their boundaries, and result in explicit, local, and highly accurate algorithms. We will provide a brief introduction to discontinous Galerkin methods in Section 15.4.

[2] The same Courant for whom the Courant-Friedrichs-Lewy (CFL) condition in FDTD is named.

1.2.4 Spectral methods

In each of the methods described above, the discretization of space requires on the order of 10 or more grid cells per wavelength, for the smallest wavelength in the simulation, in order to achieve reasonable accuracy in the results. As we have mentioned, for large problems or multiscale problems, this restriction becomes cumbersome and leads to long simulation times. Spectral methods take advantage of the Nyquist theorem, which states that only two points are needed per wavelength to perfectly reconstruct a wave; indeed, spectral methods have been shown to require only two grid cells per wavelength.

In spectral methods, the simulation space is broken into grid cells as usual, but the solution at a given time step is approximated by a function covering the entire simulation space; the methods are "spectral" because the functional form is usually a Fourier decomposition. This type of spectral method is very similar, in fact, to the finite element method; the primary difference is that the finite element method is *local*: the functional forms are assumed to be piecewise continuous over small subdomains; whereas the spectral methods are *global*, where the functional forms cover the entire simulation space. The largest frequency used in the Fourier representation of the solution defines the smallest wavelength, and in turn the grid cell size.

The spectral method of choice in recent years for time domain simulations of Maxwell's equations has been the Pseudospectral Time Domain (PSTD) method. In this method, the spatial derivatives are approximated by taking the Fast Fourier Transform (FFT) of the spatial distribution of fields along an axis; multiplying by jk to achieve the spatial derivative; then taking the inverse FFT to get back to the spatial domain. These spatial derivatives are then used directly in the update equations, and time marching proceeds as in the FDTD method.

These methods provide the stated advantage that a far more coarse grid is required; similar accuracy can be achieved compared to the FDTD method for considerably fewer grid cells. Liu [5, 6] reports a reduction in computer storage and time of a factor of 8^D, where D is the dimensionality, compared to the FDTD algorithm. The costs of the PSTD methods are a slightly stricter stability criterion (by a factor of $\pi/2$) and slightly increased numerical dispersion.

In the interest of brevity, we will not provide a detailed overview of spectral methods or the PSTD method in this book. The interested reader is referred to Chapter 17 of [7] for a good overview of PSTD methods, and the book by Hesthaven et al. [8] for an introduction to spectral methods for time-domain problems.

1.3 Organization

This book is intended to be used as a teaching tool, and has thus been written in the order that we feel is most appropriate for a one-semester course on the FDTD method. The book can be thought of as loosely organized into three sections.

Chapters 2 to 6 introduce the basics required to create, and understand, a simple FDTD problem. Chapter 2 provides a review of Maxwell's equations and the elements

of electromagnetic theory that are essential to understanding the FDTD method, and numerical electromagnetics in general. Chapter 3 describes the methods by which partial differential equations (PDEs) are discretized and transformed into finite difference equations (FDEs). The Yee cell and the FDTD algorithm are introduced in Chapter 4, in one, two, and three dimensions, as well as in other coordinate systems and in lossy materials. The stability and accuracy of the FDTD method are discussed in Chapters 5 and 6, respectively. An understanding of the accuracy of an FDTD simulation is extremely crucial; too often modelers simply run a simulation for stability, and ignore the loss of accuracy that comes with many of the inherent assumptions.

Chapters 7 to 11 provide the next level of understanding required for FDTD simulations. Chapter 7 describes methods by which sources are introduced into the simulation, including the total-field / scattered-field formulation. Chapter 8 introduces some analytical boundary conditions, used to absorb fields at the edge of the simulation space. While these methods are introduced partially for historical and mathematical interest, in many cases they are still the best choice in certain scenarios. The perfectly matched layer (PML) boundary condition is discussed in Chapter 9; the PML is the state of the art in absorbing boundary conditions. Chapter 10 describes methods for simulating wave propagation in dispersive (frequency-dependent) materials, and Chapter 11 describes the FDTD method in anisotropic materials, including materials that are both dispersive and anisotropic.

Chapters 12 to 15 introduce some more advanced topics. We have chosen topics that should be of interest to the general audience, rather than choosing particular applications. Chapter 12 describes a variety of topics, including modeling periodic structures; modeling structures that are smaller than the grid cell size; the bodies of revolution (BOR) method for modeling cylindrical structures; and the near-to-far field transformation for calculating the far-field pattern of a scattering or radiation problem. Chapter 13 introduces implicit FDTD methods, which circumvent the stability restriction of the classic explicit method. The finite difference frequency domain method, mentioned briefly above, is introduced in some detail in Chapter 14. Finally, Chapter 15 provides an overview of nonuniform, nonorthogonal, and irregular grids, which can be used to improve accuracy and efficiency near complex structures, as well as brief introductions to the finite volume, finite element, and discontinuous Galerkin methods for electromagnetics.

References

[1] V. Shankar and W. F. Hall, "A time domain differential solver for electromagnetic scattering," in *URSI National Meeting, Boulder, CO*, 1988.

[2] N. K. Madsen and R. W. Ziolkowski, "Numerical solutions of Maxwell's equations in the time domain using irregular nonorthogonal grids," *Waves Motion*, vol. 10, pp. 538–596, 1988.

[3] S. M. Rao, ed., *Time Domain Electromagnetics*. Academic Press, 1999.

[4] G. Pelosi, "The finite-element method, part i: R. I. Courant," *IEEE Antennas and Propagation Magazine*, vol. 49, pp. 180–182, 2007.

[5] Q. H. Liu, "The PSTD algorithm: A time-domain method requiring only two cells per wavelength," *Microwave Opt. Technol. Lett.*, vol. 15, pp. 158–165, 1997.

[6] O. H. Liu, "Large-scale simulations of electromagnetic and acoustic measurements using the pseudospectral time-domain (PSTD) algorithm," *IEEE Micro. Guided Wave Lett.*, vol. 37, pp. 917–926, 1999.

[7] A. Taflove and S. Hagness, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 3rd edn. Artech House, 2005.

[8] J. S. Hesthaven, S. Gottlieb, and D. Gottlieb, *Spectral Methods for Time-Dependent Problems*. Cambridge University Press, 2007.

2 Review of electromagnetic theory

A study of Numerical Electromagnetics must rely on a firm base of knowledge in the foundations of electromagnetics as stated in Maxwell's equations. Accordingly, we undertake in this chapter a review of Maxwell's equations and associated boundary conditions.

All classical electromagnetic phenomena are governed by a compact and elegant set of fundamental rules known as *Maxwell's equations*. This set of four coupled partial differential equations was put forth as the complete classical theory of electromagnetics in a series of brilliant papers[1] written by James Clerk Maxwell between 1856 and 1865, culminating in his classic paper [2]. In this work, Maxwell provided a mathematical framework for Faraday's primarily experimental results, clearly elucidated the different behavior of conductors and insulators under the influence of fields, imagined and introduced the concept of displacement current [3, Sec. 7.4], and inferred the electromagnetic nature of light. A most fundamental prediction of this theoretical framework is the existence of electromagnetic waves, a conclusion to which Maxwell arrived in the absence of experimental evidence that such waves can exist and propagate through empty space. His bold hypotheses were to be confirmed 23 years later (in 1887) in the experiments of Heinrich Hertz [4].[2]

When most of classical physics was fundamentally revised as a result of Einstein's introduction [6][3] of the special theory of relativity, Maxwell's equations remained intact.[4] To this day, they stand as the most general mathematical statements of fundamental natural laws which govern all of classical electrodynamics. The basic justification and validity of Maxwell's equations lie in their consistency with physical experiments over the entire range of the experimentally observed electromagnetic spectrum, extending from cosmic rays at frequencies greater than 10^{22} Hz to the so-called micropulsations at frequencies of about 10^{-3} Hz. The associated practical applications cover an equally wide range, from the use of gamma rays ($10^{18} - 10^{22}$ Hz) for cancer therapy to use of waves at frequencies

[1] For an excellent account with passages quoted from Maxwell's papers, see [1, Ch. 5].
[2] For a collected English translation of this and other papers by H. Hertz, see [5].
[3] The English translation of [6] is remarkably readable and is available in a collection of original papers [7].
[4] Maxwell's formulation was in fact one of the major motivating factors which led to the development of the theory of special relativity. The fact that Galilean relativity was consistent with classical mechanics but inconsistent with electromagnetic theory suggested either that Maxwell's equations were incorrect or that the laws of mechanics needed to be modified. For discussions of the relationship between electromagnetism and the special theory of relativity, see [8, Sec. 15]; [9, Ch. 10]; [1, Ch. 2]; [10, Ch. 11].

of a few Hz and below for geophysical prospecting. Electromagnetic wave theory as embodied in Maxwell's equations has provided the underpinning for the development of many vital practical tools of our technological society, including broadcast radio, radar, television, cellular phones, optical communications, Global Positioning Systems (GPS), microwave heating and processing, X-ray imaging, and numerous others.

We now continue with a brief review of Maxwell's equations [11, pp. 247–262] and their underlying foundations. Maxwell's equations are based on experimentally established facts, namely Coulomb's law, which states that electric charges attract or repel one another in a manner inversely proportional to the square of the distance between them [12, p. 569]; Ampère's law, which states that current-carrying wires create magnetic fields and exert forces on one another, with the amplitude of the magnetic field (and thus force) depending on the inverse square of the distance [13]; Faraday's law, which states that magnetic fields which vary with time induce electromotive force or electric field [14, pp. 1–109]; and the principle of conservation of electric charge. Discussion of the experimental bases of Maxwell's equations is available elsewhere.[5] The validity of Maxwell's equations is based on their consistency with all of our experimental knowledge to date concerning electromagnetic phenomena. The physical meaning of the equations is better perceived in the context of their integral forms, which are listed below together with their differential counterparts:

1. Faraday's law is based on the experimental fact that time-changing magnetic flux induces electromotive force:

$$\oint_C \overline{\mathscr{E}} \cdot d\mathbf{l} = -\int_S \frac{\partial \overline{\mathscr{B}}}{\partial t} \cdot d\mathbf{s} \qquad \nabla \times \overline{\mathscr{E}} = -\frac{\partial \overline{\mathscr{B}}}{\partial t}, \tag{2.1}$$

   where the contour C is that which encloses the surface S, and where the sense of the line integration over the contour C (i.e., direction of $d\mathbf{l}$) must be consistent with the direction of the surface vector $d\mathbf{s}$ in accordance with the right-hand rule.
2. Gauss's law is a mathematical expression of the experimental fact that electric charges attract or repel one another with a force inversely proportional to the square of the distance between them (i.e., Coulomb's law):

$$\oint_S \overline{\mathscr{D}} \cdot d\mathbf{s} = \int_V \tilde{\rho}\, dv \qquad \nabla \cdot \overline{\mathscr{D}} = \tilde{\rho}, \tag{2.2}$$

   where the surface S encloses the volume V. The volume charge density is represented with $\tilde{\rho}$ to distinguish it from its phasor form ρ used in the time-harmonic version of Maxwell's equations.
3. Maxwell's third equation is a generalization of Ampère's law, which states that the line integral of the magnetic field over any closed contour must equal the total current

[5] Numerous books on fundamental electromagnetics have extensive discussion of Coulomb's law, Ampère's law, Faraday's law, and Maxwell's equations. For a recent reference that provides a physical and experimentally based point of view, see [3, Ch. 4–7].

enclosed by that contour:

$$\oint_C \overline{\mathcal{H}} \cdot d\mathbf{l} = \int_S \overline{\mathcal{J}} \cdot d\mathbf{s} + \int_S \frac{\partial \overline{\mathcal{D}}}{\partial t} \cdot d\mathbf{s} \qquad \nabla \times \overline{\mathcal{H}} = \overline{\mathcal{J}} + \frac{\partial \overline{\mathcal{D}}}{\partial t}, \tag{2.3}$$

where the contour C is that which encloses the surface S, and $\overline{\mathcal{J}}$ is the electrical current density (see below). Maxwell's third equation expresses the fact that time-varying electric fields produce magnetic fields. This equation with only the first term on the right-hand side (also referred to as the conduction-current term) is Ampère's law, which is a mathematical statement of the experimental findings of Oersted, whereas the second term, known as the displacement-current term, was introduced theoretically by Maxwell in 1862 and verified experimentally many years later (1888) in Hertz's experiments [4].

4. Maxwell's fourth equation is based on the fact that there are no magnetic charges (i.e., magnetic monopoles) and that, therefore, magnetic field lines always close on themselves:

$$\oint_S \overline{\mathcal{B}} \cdot d\mathbf{s} = 0 \qquad \nabla \cdot \overline{\mathcal{B}} = 0, \tag{2.4}$$

where the surface S encloses the volume V. This equation can actually be derived [3, Sec. 6.5–6.7] from the Biot-Savart law, so it is not completely independent.[6]

The continuity equation, which expresses the principle of conservation of charge in differential form, is contained in Maxwell's Equations and in fact can be readily derived by taking the divergence of Equation (2.3) and using Equation (2.2). For the sake of completeness, we give the integral and differential forms of the continuity equation:

$$-\oint_S \overline{\mathcal{J}} \cdot d\mathbf{s} = \frac{\partial}{\partial t} \int_V \tilde{\rho}\, dv \qquad \nabla \cdot \overline{\mathcal{J}} = -\frac{\partial \tilde{\rho}}{\partial t}, \tag{2.5}$$

where the surface S encloses the volume V. The fact that the continuity equation can be derived from Equations (2.2) and (2.3) indicates that Maxwell's Equations (2.2) and (2.3) are not entirely independent, if we accept conservation of electric charge as a fact; i.e., using Equations (2.3) and (2.5), one can derive Equation (2.2).

Note that for all of the Equations (2.1) through (2.5), the differential forms can be derived from the integral forms (or vice versa) by using either Stokes's or the divergence

[6] Note that Equation (2.4) can be derived from Equation (2.1) by taking the divergence of the latter and using the vector identity of $\nabla \cdot (\nabla \times \overline{\mathcal{G}}) \equiv 0$, which is true for any vector $\overline{\mathcal{G}}$. We find

$$\nabla \cdot (\nabla \times \overline{\mathcal{E}}) = -\nabla \cdot \left(\frac{\partial \overline{\mathcal{B}}}{\partial t} \right) \quad \rightarrow \quad 0 = -\frac{\partial (\nabla \cdot \overline{\mathcal{B}})}{\partial t} \quad \rightarrow \quad \text{const.} = \nabla \cdot \overline{\mathcal{B}}$$

The constant can then be shown to be zero by the following argument. If we suppose that the $\overline{\mathcal{B}}$ field was produced a finite time ago, i.e., it has not always existed, then, if we go back far enough in time, we have $\overline{\mathcal{B}} = 0$ and therefore $\nabla \cdot \overline{\mathcal{B}} = 0$. Hence it would appear that

$$\nabla \cdot \overline{\mathcal{B}} = 0 \quad \text{and} \quad \oint_S \overline{\mathcal{B}} \cdot d\mathbf{s} = 0$$